Article

Not peer-reviewed version

Stalking the Synthetically-Accessible Polymer Universe with RxnChainer: Digital Synthesis, Retrosynthesis and Beyond

Shivank S. Shukla , Rishi Gurnani , Chiho Kim , Rampi Ramprasad , Akhlak Mahmood *

Posted Date: 14 May 2026

doi: 10.20944/preprints202605.0977.v1

Keywords: polymer informatics; machine learning; polymer synthesis; virtual forward synthesis; retrosynthesis; post-polymerization modification

Article

Stalking the Synthetically-Accessible Polymer Universe with RxnChainer: Digital Synthesis, Retrosynthesis and Beyond

Shivank S. Shukla [1,2], Rishi Gurnani [1,2], Chiho Kim [1,2], Rampi Ramprasad [1,2] and Akhlak Mahmood [2,*]

[1] School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, USA
[2] Matmerize, Inc., Atlanta, Georgia 30308, USA
* Correspondence: akhlak.mahmood@matmerize.com

Abstract

Polymers enable countless modern technologies, yet vast regions of their chemical space remain unexplored. Traditional polymer discovery relies on chemical intuition, ingenuity, and experience (with a healthy dose of serendipity), yet it fails to leverage millions of potentially accessible and synthesizable polymer structures. Here, we present RxnChainer, a digital methodology integrating virtual polymer generation, retrosynthetic analysis, and post-polymerization modification to systematically explore synthetically accessible polymer space. Using commercially available monomers from the Toxic Substances Control Act (TSCA) and ChEMBL databases and RxnChainer, we generated over 289 million hypothetical polymers across 44 polymerization pathways spanning 32 polymer classes, including polyamides, polyimides, polyesters, and polyethers. Comparison with the known (i.e., previously synthesized) spectrum of polymers revealed that a significant portion of these new synthesizable structures are novel, i.e., previously unknown and unexplored. We demonstrate the methodology's versatility through automated retrosynthetic planning for 30,000 polyesters and targeted functionalization via four post-polymerization modification pathways incorporating vinyl and nitrile pendant groups. The resulting datasets enable downstream tasks such as property-driven screening, application-specific design, and training of generative models.

Keywords: polymer informatics; machine learning; polymer synthesis; virtual forward synthesis; retrosynthesis; post-polymerization modification

Introduction

Polymers are essential materials that underpin modern technology, from biomedical devices and food packaging to energy storage and electronic systems. [1–6] They exhibit remarkable chemical diversity, ranging from simple homopolymers and copolymers to complex polymer blends, composites, and formulations. [7–10] This extraordinary chemical diversity, coupled with the vast array of possible architectures, creates a polymer universe that is theoretically infinite. [11] Yet the vast majority of this space remains unexplored, and critically, it is unclear which regions are synthetically accessible and what property value ranges are achievable. This knowledge gap represents both a fundamental challenge and an enormous opportunity for accelerated materials discovery.

To meet urgent comtemporary challenges, industries spanning energy, environment, and manufacturing are exploring the replacement of existing materials with safer, higher-performing, and environmentally friendly alternatives. [12] Examples include novel polymers for membrane separation, fuel cells, food packaging, additive manufacturing, and high-energy-density capacitors. [13–19] Systematic and efficient search of the synthetically accessible polymer space offers a pathway to identify candidates optimized for specific applications based on targeted property criteria.

Traditional polymer design relies heavily on expert intuition and trial-and-error experimentation, leading to lengthy development cycles and limited exploration of chemical space. [20] A fundamental challenge is identifying designs that are both synthetically accessible and exhibit the target properties required for specific applications. Meanwhile, emerging virtual design paradigms that leverage artificial intelligence (AI) methods have enabled accelerated polymer discovery and design. [21] Given the vast scale of the polymer chemical space, these AI-driven approaches may be effective to systematically explore synthetically accessible regions and guide the identification of promising candidates.

Among existing computational approaches, virtual forward synthesis (VFS) has generated hypothetical polymer structures by digitally reacting commercially available monomers from databases such as ChEMBL, ZINC-15, and eMolecules via known polymerization pathways. [22–24] Resources including Open Macromolecular Genome (OMG), SMIPOLY, PolyVERSE, and PolyUniverse provide access to millions of such polymers, [19,25–27] with select designs validated through experimental synthesis. [19,28,29] While laboratory synthesis require optimization of reaction conditions depending on the specific monomer pair, the computational generation of the polymers derived from commercially available monomers via established polymerization chemistries provides a valuable starting point for experimental exploration. In addition, computational retrosynthesis offers a complementary capability, identifying commercially available precursors for target polymer structures and streamlining synthetic planning. [30] Beyond these generation and retrosynthesis capabilities, post-polymerization modification (PPM) provides an additional route to chemical diversity, enabling tuning of properties such as mechanical strength, thermal stability, or chemical resistance without redesigning base polymer architectures. [31–35]

Here, we introduce RxnChainer, a unified computational methodology that integrates polymer generation, retrosynthetic analysis, and post-polymerization modification within a single algorithmic framework for systematically exploring synthetically accessible polymer space. The methodology, shown in Figure 1, operates on a core algorithm that filters reactants based on compatibility criteria and applies reaction chains to generate products, supporting three complementary applications: forward synthesis for polymer generation, retrosynthesis for identifying synthetic precursors, and post-polymerization modification for expanding chemical diversity. Using commercially available monomers from the Toxic Substances Control Act (TSCA) and ChEMBL databases, we generated over 289 million hypothetical polymers spanning 44 polymerization pathways across 32 polymer classes. [22,36] Comparison with the PolyInfo database reveals that a substantial fraction of these computationally generated structures are novel yet synthetically accessible, offering the potential to significantly expand knowledge beyond the known polymer chemical space. [37]

A downstream task undertaken here is to determine property ranges accessible by such a comprehensive chemical space exploration. Machine learning models trained on existing polymer datasets predict key properties of the generated polymers including bandgap, glass transition temperature, and dielectric constant, revealing broad property ranges across polymer families. Diversity analysis of the generated polymers through Tanimoto similarity distributions and UMAP projections maps the structural landscape relative to known polymers, revealing regions of high novelty. To demonstrate practical utility, the generated polymers were screened to identify polymer candidates that meet desired target property criteria as attempted before. [13,18,19] Interestingly, many previously designed polymers emerge from this screening, providing retrospective validation of the procedure adopted here. Additional downstream applications of the generated library of polymers would be to train generative models for synthetically-accessible polymer design, including language model pretraining tasks as have been done recently. [38–40]

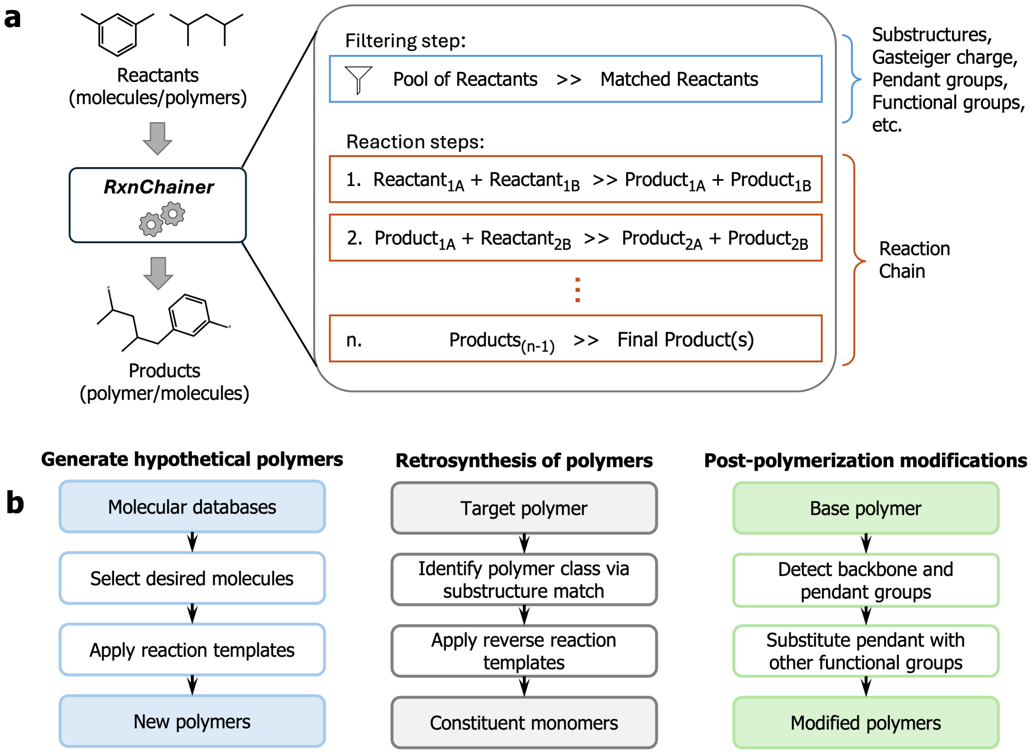


Figure 1. (a) Overview of the RxnChainer methodology, which processes a pool of input reactants by filtering and producing the output products based on a reaction chain containing multiple reaction steps. **(b)** Flow diagrams showing different application examples of RxnChainer, including the generation of hypothetical polymers, the retrosynthesis of hypothetical or known polymers, and post-polymerization modifications of base polymer.

Methods

Molecular Databases. In this work, we utilized the Toxic Substances Control Act (TSCA) inventory, a comprehensive list of chemical substances manufactured or processed in the United States.[36] The TSCA inventory is maintained by the Environmental Protection Agency (EPA) and contains information on the chemical identity, production volume, and use of each substance. Our TSCA inventory dataset (May, 2024 version) contains more than 70 thousand entries.

In addition to the TSCA inventory, we also utilized the ChEMBL database to identify potential reactant molecules for polymerization reactions.[22] The ChEMBL database is maintained by the European Bioinformatics Institute (EBI) and contains information on bioactive compounds, including their chemical structures, biological activities, and pharmacological properties. The ChEMBL database is widely used in drug discovery and development research, as it provides a wealth of information on the chemical properties and biological activities of compounds. We used the database to identify potential reactant molecules for polymerization reactions by filtering for compounds that contain specific functional groups or structural motifs that are compatible with the polymerization reaction chains we implemented. Our ChEMBL database dataset (2015 version) contains more than 1.2 million molecular entries. Both databases contain molecular structures represented in the Simplified Molecular Input Line Entry System (SMILES) format and identification numbers for each molecule.

Filtering of reactant molecules. A crucial step in generating polymers through the RxnChainer methodology was identifying compatible molecules from molecular databases. We utilized the RDKit library and SMILES arbitrary target specification (SMARTS) patterns for substructure searches and screened for compatible molecules.[41] Simpler methods for filtering compatible molecules were also implemented, such as string-based search. Additionally, the RxnChainer methodology also incorporated capabilities to search for molecules with a user-specified Gasteiger charge difference for a

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 May 2026 doi:10.20944/preprints202605.0977.v1

specified SMARTS pattern. This is especially advantageous for AA-BB polycondensation reactions, where a smaller Gasteiger charge difference between the reacting functional groups is a proxy for similar reaction rates.

Polymerization rules through reaction chains. Polymerization reactions were executed within the RxnChainer methodology through the use of reaction SMARTS[42] which serve as the foundation for defining reaction rules, a functionality supported by RDKit. [43] Once a polymerization reaction was chosen for implementation, we applied the reaction rules by obtaining general templates of the reactions from the literature, textbooks, and other sources to form linear homopolymers. These templates involve reacting functional groups and spacer groups with generalized reactions. We then implemented the general template within the RxnChainer methodology by encoding the reaction rules by writing their reaction SMARTS. The number of reaction SMARTS needed to encode a polymerization reaction depends on the specific polymerization mechanism. The list of all reaction chains implemented in this work is provided in the Supplementary Information.

Calculation of Synthetic Accessibility Score. The synthetic accessibility score (SA score) measures the ease of synthesizing a molecule and is calculated using known molecular fragment data and a complexity penalty. [44] In this work, we used the SA score algorithm implemented in the RDKit library to evaluate the feasibility of synthesizing the hypothetical polymers. To calculate the SA score for polymers, we first converted the polymer SMILES to monomer SMILES by replacing the dangling bonds ([*]) with a hydrogen ([H]) atom and then calculated the SA score of the resulting structure. This approach provides an estimate of the synthetic accessibility of the polymer based on the properties of its monomer unit.

Similarity Calculation. The Tanimoto similarity metrics implemented in the RDKit library were used to compare the hypothetical polymers with known polymers from the literature, which measures the overlap between the molecular fingerprints of two compounds, and other structural similarity indices.[45] Morgan fingerprints were implemented for fast similarity calculations.

RxnChainer Modules. The RxnChainer methodology for homopolymer generation, retrosynthesis, and post-polymerization modification developed in this work has been implemented in PolymRize [46], a standardized software platform for polymer informatics.

Results and Discussion

Overview of the RxnChainer Methodology

The RxnChainer workflow (Figure 1a) begins with a pool of reactants and a defined reaction chain. Reactants are filtered based on compatibility criteria including SMARTS patterns, functional groups, Gasteiger charge, molecular weight, and other molecular properties.[42,47] Compatible reactants then proceed through the reaction chain, which may consist of multiple reaction steps, each specified by a reaction SMARTS pattern. The reaction step may input the products of a previous reaction step as reactants for the subsequent step. This sequential approach enables combining multiple reactions for complex polymerization or product generation tasks.

New hypothetical homopolymer designs can be generated using this approach by providing a list of reacting molecules (monomers) and a polymerization reaction template, as illustrated in Figure 1b. In the filtering stage, the input monomers are screened against the defined compatibility criteria to ensure only suitable monomers proceed to the polymerization stage. In the polymerization stage, the specified reaction chain is applied to the filtered monomers to generate hypothetical polymer structures. This approach can be used for various polymerization methods, including polyaddition, polycondensation, and ring-opening polymerization as demonstrated later in this manuscript.

Beyond forward synthesis, the RxnChainer methodology also enables retrosynthesis of hypothetical or known polymers. Given a target polymer structure and a set of reaction chains, this approach identifies the polymer class and then applies class-specific inverse reactions to obtain plausible reactants. The methodology further supports post-polymerization modifications for expanding chemical diversity. Starting from a base polymer, functional groups of the base polymer are identified during filtering to select compatible modification reactions from a predefined library. The group-specific reactions are then applied to the base polymers to generate a set of modified polymers. Additional technical details on the implementation and specific templates used in each stage of the RxnChainer methodology are provided in the Methods and Supplementary Information sections.

Generation of Hypothetical Polymers

In this study, we generated polymer designs utilizing the TSCA inventory [36] and the ChEMBL database [22] via RxnChainer. The TSCA inventory comprises a comprehensive list of chemical substances manufactured or processed in the United States, making it a valuable resource for identifying potential monomers for polymer synthesis. ChEMBL was selected due to its extensive collection of bioactive molecules, which can also serve as potential monomers for polymerization reactions. Considering commercially available molecules from these databases increases the possibility of later synthesis for the generated hypothetical polymers.

We successfully implemented 44 polymerization reaction templates using the RxnChainer methodology. The templates include 22 polycondensation, 13 polyaddition, and 9 ring-opening polymerization reactions. As shown in Figure 2a, the generation workflow begins with 70,000 molecules from TSCA and 1.2 million molecules from ChEMBL. Following filtering for monomers compatible with the reaction templates, 14,470 TSCA molecules (20.7%) and 293,875 ChEMBL molecules (24.5%) were retained as suitable reactants. The number of polymers generated for each class, grouped by their types and templates, are detailed in Table 1. Notably, the polycondensation reactions yield the highest number of hypothetical polymers due to the multiplicative potential arising from two distinct reactants, followed by polyaddition and ring-opening polymerization. A comprehensive list of the implemented polymerization reaction templates, including the reaction steps for representative polymers from each class, is available in the Supplementary Information.

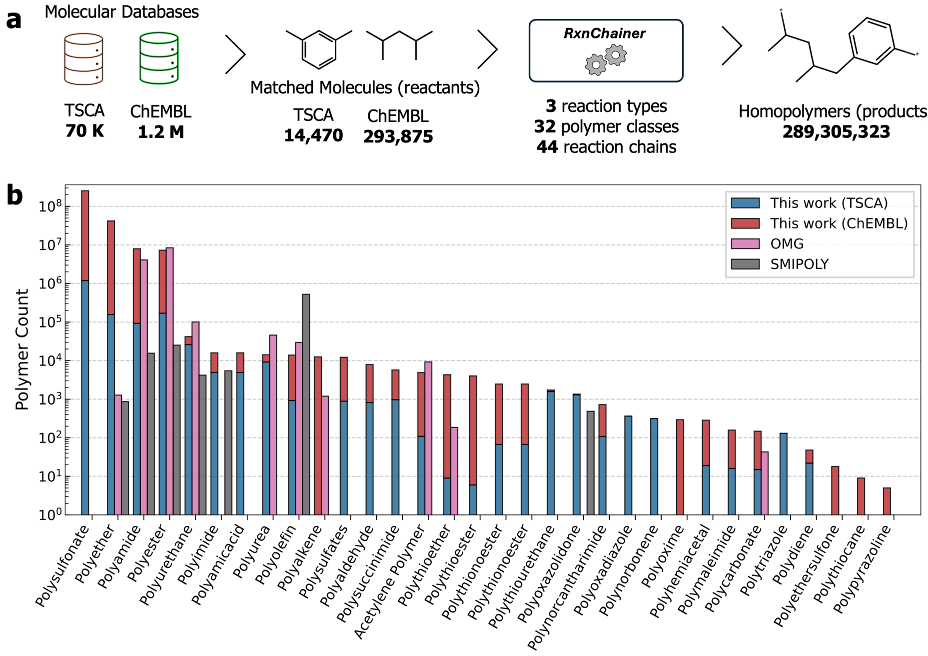


Figure 2. (a) Overview of generated polymers using RxnChainer. We considered 70,000 molecules from the TSCA inventory and 1.2 million molecules from the ChEMBL database. After filtering for relevant monomers suitable for our polymerization routes, 14,470 molecules from TSCA and 293,875 molecules from ChEMBL were retained. These monomers were used to generate 289 million hypothetical homopolymers spanning 32 polymer classes through 46 reaction chains. **(b)** Comparison of the number of hypothetical polymers generated in this work with previous studies (OMG, SMIPOLY) across different polymer classes.

Table 1. Number of hypothetical homopolymers generated using RxnChainer grouped by different reaction types, polymer classes, reaction routes, and monomers from the TSCA inventory and the ChEMBL database. The polymer classes are grouped by their reaction types (polyaddition, polycondensation, and ring-opening). The reaction routes are defined by the monomer classes. The number of corresponding monomers for each class is shown in parentheses.

Polymer Class	Reaction Route	TSCA	ChEMBL
Reaction Type: Addition			
Polyacetylene	Acetylene	109 (109)	4,796 (4,796)
Polyaldehyde	Aldehyde	819 (819)	7,117 (7,117)
Polyamicacid	Dianhydride + Diamine	4,913 (17+289)	10,971 (3+3,657)
Polyamide	Heterocumulene	74 (74)	10 (10)
Polydiene	Diene	22 (22)	26 (26)
Polyethersulfone	Vinylsulfonylethanol	-	18 (18)
Polyhemiacetal	Cyclic Acetal	19 (19)	267 (267)
Polymaleimide	Maleimide	16 (16)	142 (142)
Polyoxazolidone	Diepoxide + Diisocyanate	1,300 (25+52)	58 (29+2)
Polysuccinimide	Dimaleimide + Diamine	972 (9+108)	4,748 (2+2,374)
Polyurea	Diamine + Diisocyanate	9,176 (37+248)	5,012 (2+2,506)
Polyurethane	Diol + Diisocyanate	26,159 (37+707)	15,730 (2+7,865)
polyalkene	Vinyl Monomer	924 (924)	13,004 (13,004)
Reaction Type: Condensation			
Polyamide	Diacid + Diamine	56,165 (239+235)	3,661,596 (2,316+1,581)
Polyamide	Diacid Chloride + Diamine	8,526 (29+294)	-
Polyamide	Dimethyl Ester + Diamine	27,636 (94+294)	4,184,510 (1,355+1,581)
Polyester	Diacid + Diol	112,540 (170+662)	4,924,920 (840+5,863)
Polyester	Diacid Chloride + Diol	18,536 (28+662)	-
Polyester	Dimethyl Ester + Diol	38,340 (60+639)	2,157,805 (373 +5,785)
Polyether	Dichloride + Diol	112,200 (200+561)	19,097,736 (4,123+4,632)
Polyether	Dibromide + Diol	36,465 (65+561)	2,427,168 (524+4,632)
Polyether	Diiodide + Diol	7,293 (13+561)	226,968 (49+4,632)
Polyimide	Dianhydride + Diamine	4,913 (17+289)	10,971 (3+3,657)
Polynorcantharimide	Difuran + Dimaleimide	108 (12+9)	618 (309+2)
Polyoxadiazole	Diacid Chloride + Dihydrazide	135 (27+5)	-
Polyoxadiazole	Diacid + Dihydrazine	230 (230+1)	2,125 (2,125+1)
Polyoxime	Dialdehyde + Diaminoxy	-	294 (147+2)
Polypyrazoline	Tetrazole-Alkene	-	5 (5)
Polysulfates	Diacid	888 (888)	11,294 (11,294)
Polysulfonate	Primary + Diol	1,185,954 (2,114+561)	252,564,432 (54,526+4,632)
Polythioether	Diene + Dithiol	4,991 (161+31)	61,920 (860+72)
Polythioether	Diyne + Dithiol	403 (13+31)	12,456 (173+72)
Polythioether	Dibromo + Dithiol	2,449 (79+31)	141,636 (1,914+74)
Polythiourethane	Diisocyanate + Dithiol	1,581 (51+1)	148 (2+74)
Polytriazole	Diazide + Dialkyne	130 (10+13)	-
Reaction Type: Ring Opening Polymerization (ROP)			
Polyamide	Lactam	178 (178)	56,293 (56,293)
Polycarbonate	Cyclic Carbonate	15 (15)	133 (133)
Polyalkene	Cyclic Alkene	-	12,502 (12,502)
Polyester	Lactone	451 (451)	40,131 (40,131)
Polyether	Cyclic Ether	558 (558)	14,280 (14,280)
Polythiocane	Cyclic Thiocane	-	9 (9)
Polythioester	Cyclic Thioester	6 (6)	3,999 (3,999)
Polythioether	Cyclic Thioether	9 (9)	4,291 (4,291)
Polythionoester	Cyclic Thionoester	67 (67)	2,401 (2,401)
	Total	1,665,253	287,640,070

A comparative analysis of the total number of unique polymers generated across the 32 polymer classes established in this work with those produced in previous studies, including the Open Macromolecular Genome (OMG) and SMIPOLY, is presented in Figure 2b. The comparison reveals that RxnChainer substantially expands the accessible polymer space across nearly all polymer classes. For instance, polysulfonate, the most populated class in this work, accounts for over 2.5 × 10^8 polymers, vastly exceeding the combined contributions from OMG and SMIPOLY. Similarly, for polyether, polyester, and polyamide—classes with significant industrial relevance—this work generates 10^7 to 10^8 polymers, representing 2-3 orders of magnitude improvement over previous databases. The enhanced coverage is particularly pronounced for condensation polymers, where the multiplicative nature of two-reactant systems (e.g., diol + dichloride for polyether, diacid + diamine for polyamide) enables combinatorial expansion of the design space. Even for less-explored classes such as polytriazole, polyoxadiazole, and polythiourethane, RxnChainer provides substantial coverage.

Novelty and Synthetic Feasibility of Generated Polymers

To investigate the novelty of the polymers, we conducted a comparative analysis of the generated polymers from the top 16 polymer classes, randomly sampling a maximum of 10,000 polymers from each class for computational efficiency. We calculated the Tanimoto similarity matrix for the selected polymers and 13,000 known homopolymers sourced from the PolyInfo database [37] utilizing Morgan fingerprints. [45,48] The similarity metrics for each of the polymer classes, based on numerical representation and the randomly selected polymers are presented in Figure 3a. It can be observed that for nearly all polymer classes, there is at least one exact match within the known polymer space, which is anticipated, given our use of commercially available molecules. Interestingly, for four polymer classes, namely polysulfonate, polynorcantharimide, polythiourethane and polytriazole, no matches were identified within the known polymer space. On average, the hypothetical space demonstrates significant distinctiveness from the known polymer space, as evidenced by the low average Tanimoto similarity values and their associated standard deviation.

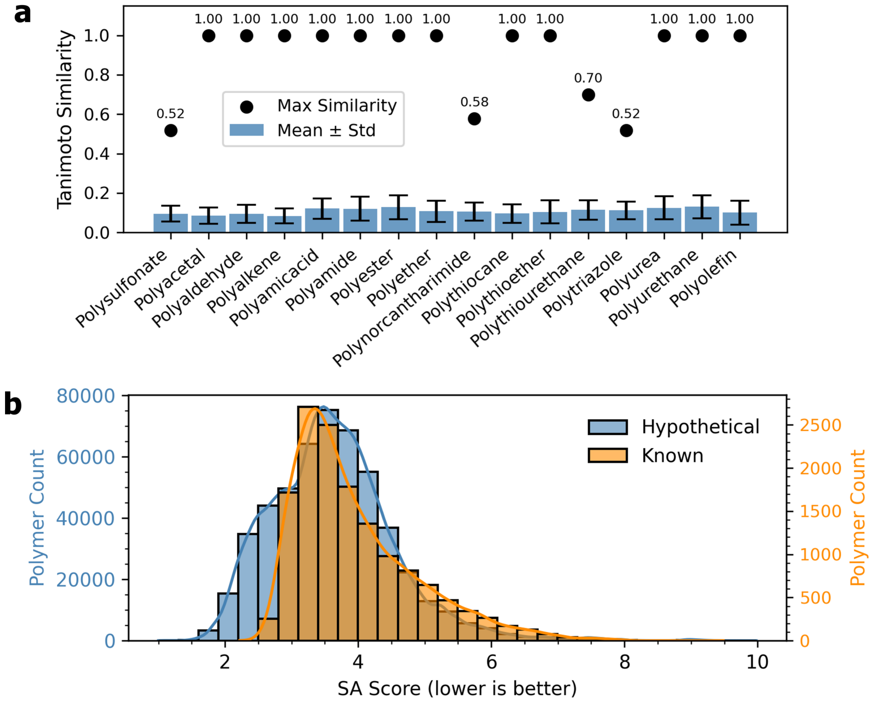


Figure 3. (a) The Tanimoto similarity scores of the generated polymers compared to the known polymers, grouped by the top 16 polymer classes. The low average similarity scores indicate highly novel polymers generated in this work. Some polymers show a perfect similarity of 1.0 with the known polymers. **(b)** The distribution of Synthetic Accessibility (SA) scores of the hypothetical polymers generated in this work compared to 13,000 known polymers. Polymers with lower SA scores are predicted to be more synthetically accessible.

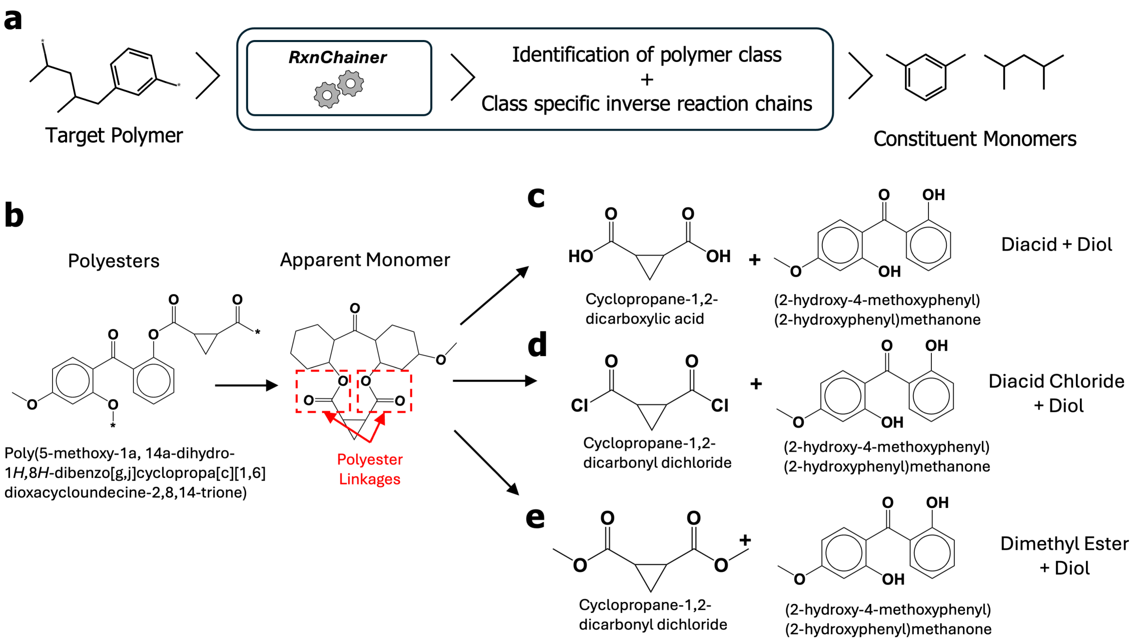
Next we assessed the synthetic feasibility of the hypothetical polymers by employing the Synthetic Accessibility (SA) score implemented in RDKit. The metric evaluates molecules based on fragment similarity and imposes penalties based on the complexity of well-known fragments. [44] In our approach, we initially converted the PSMILES to molecular SMILES by adding hydrogen atoms to the dangling bonds. While this scoring method is an approximation, the same fragment-based penalties are equally applicable to the polymers and provides a relative measure for selecting promising candidates for synthesis. The distribution plots of the SA scores for the hypothetical polymers generated in this study and the known polymers are depicted in Figure 3b. Polymers with low SA scores closer to 1 are deemed more synthesizable. The vast majority of the hypothetical polymers possess an SA score ranging from 2 to 6. In comparison, previously synthesized polymers have SA scores in the range of 2.5 to 7. There is a significant overlap in the distribution plots of SA scores between hypothetical and known polymers, suggesting that a large portion of the generated polymers are potentially synthesizable. This information, combined with polymer design considerations, provides valuable insights regarding which polymers to prioritize for synthesis and testing.

Retrosynthesis of Hypothetical Polymers

The capabilities of RxnChainer also extend to retrosynthetic analysis of polymers, enabling the deconstruction of complex polymer structures into their constituent monomers or reactants. This functionality is particularly valuable for researchers aiming to determine the routes to synthesize existing polymers or to identify the commercially available monomer candidates to create new polymers. The overview of the retrosynthesis workflow is illustrated in Figure 4a. The process begins with the classification of the input polymer via substructure matching using the RxnChainer's filtering capabilities. This approach mirrors established retrosynthetic strategies in polymer chemistry, where the polymer backbone and functional groups guide the selection of likely synthetic routes. Once the polymer class is determined, RxnChainer applies a set of inverse reaction templates specific to that class to deconstruct the input polymer into its constituent monomers or reactants. If a polymer class cannot be determined, the workflow does not yield a result for the input polymer. Multiple retrosynthetic pathways may exist for a single polymer, reflecting the diversity of synthetic strategies available in polymer science.

In this study, we demonstrate retrosynthetic capabilities of RxnChainer for polyesters, providing proof-of-concept examples for the class as illustrated in Figure 4b. First, the apparent monomer is identified by detecting the repeating unit of the input polymer structure. Next, the polymer class of the apparent monomer is determined through substructure matching in the filtering stage. Once the class is established, RxnChainer applies one or more pre-defined inverse reaction chains specific to the identified polymer class to generate the constituent monomers. For polyesters, three inverse reaction chains, each corresponding to an independent retrosynthetic route, have been implemented: the diacid plus diol route, the diacid chloride plus diol route, and the dimethyl ester plus diol route, shown in Figure 4c,d,e. Each polyester can, in principle, be synthesized via any of these three pathways; the choice of which route to pursue in practice depends on the availability of the corresponding monomers or starting materials. For each pathway, reactant molecules are generated by applying the corresponding inverse reaction chain, enabling comprehensive identification of the monomeric precursors.

To assess the accuracy of the retrosynthesic workflow of RxnChainer, 30,000 polyester polymers were randomly selected from the pool of polymers generated specifically from the TSCA inventory. The workflow was then applied to these polymers to recover their constituent monomers. For validation, the identified monomers were subsequently canonicalized and finally compared to the cannonical version of the original monomers found in the TSCA inventory. As shown in Figure 4f, on average, the workflow correctly matched the original monomers for 95% of the selected polyester polymers, with the accuracy varying depending on the specific reaction pathway. Approximately 5% of the polyesters didn't lead to the generation of the original monomers primarily because of the presence of multiple ester linkages in the polymers, making the heuristics-based approach for retrosynthesis prone to errors. Regardless, the high accuracy demonstrates the effectiveness of RxnChainer's retrosynthetic capabilities in identifying plausible monomeric precursors for target polymers. The results also signify that there is a limit to how much a heuristics-based approach using reaction SMARTS can be successfully used to identify polymer reactants with certainty.



f

Polyester Reaction Path	Polymer Count	Retrosynthesis Accuracy
Diacid + Diol	10,000	0.90
Diacid Chloride + Diol	10,000	0.98
Dimethyl Ester + Diol	10,000	0.97

Figure 4. **(a)** Overview of the retrosynthesis pipeline used to identify the constituent monomers of a target polymer. The polymer class is first identified by the RxnChainer and the monomers are identified by inverse reaction chains. **(b)** Identification of the apparent monomer and polymer class. **(c, d, e)** Determination of the constituent monomers identified by different inverse reaction pathways for the identified polymer class. **(f)** Accuracy of the retrosynthesis pipeline in identifying the monomers for 30,000 polyester polymers generated from known monomers.

Post Polymerization Modification

Post-polymerization modification (PPM) is another powerful approach to modify the properties of polymers. This is particularly useful when desired properties of the polymer are not immediately achievable through the initial polymerization reaction. Many existing polymers, such as ion exchange membranes, are modified to achieve the desired transport properties using PPM [49].

RxnChainer, as a general capability, can also be used to implement *in-silico* PPM to modify known or generated polymers. In the RxnChainer PPM workflow, illustrated in Figure 5a, we first identify the class and pendant group of the input polymer using substructure search in the filtering stage. Once the pendant group is identified, a set of pre-defined reaction chains are applied to modify the polymer. This is performed by using a set of group-specific reaction templates that are designed to modify the polymer by adding functional groups or changing the polymer architecture, similar to laboratory experiments.

As a proof of concept, we have defined reaction chains for polyolefins or vinyl polymers to modify the polymer with 4 different functional groups. We selected these functional groups based on their tendency to improve the properties of the base polymer, such as hydrophilicity, hydrophobicity, and reactivity. A primary advantage of the reaction chains is that they are designed to be compatible with vinyl polymer backbone and can be applied to any vinyl polymer. Figure 5b shows an example of a vinyl polymer containing nitrile and vinyl pendant groups. The PPM workflow identifies the compatible pendant groups and applies the corresponding reaction chains to modify the polymer. The nitrile group is modified using a cycloaddition reaction to form a triazole group, which is known to increase the hydrophilicity of polymers. Similarly, the vinyl group is modified using thiol-ene, epoxidation, or halogenation reactions to form different functional groups. The thiol-ene reaction is known to increase the hydrophobicity of polymers, while the epoxidation and halogenation reactions introduce reactive groups to the polymer backbone. The generated polymer structures after functionalization are shown in Figure 5c. These polymers can be further modified by applying the PPM workflow iteratively.

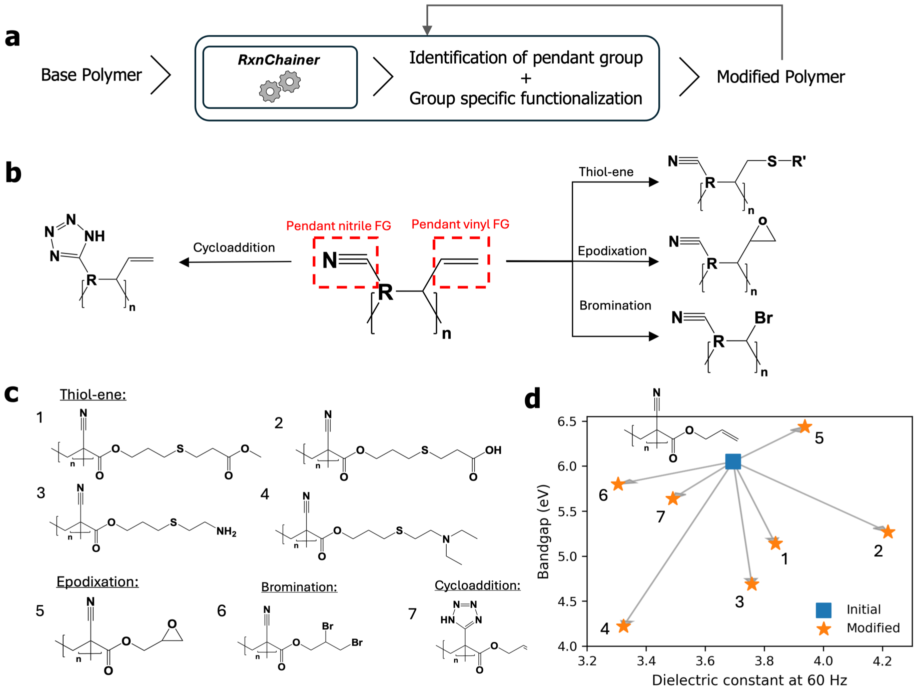


Figure 5. (a) Workflow to perform post-polymerization modification (PPM). First, the pendant groups of the base polymers are identified, then compatible functionalization reaction chains are applied to generate the modified polymers. The generated polymer can be further functionalized iteratively. **(b)** Example functionalizations performed on a base polymer containing nitrile and vinyl pendant groups. **(c)** Generated polymers after functionalizations via different groups. **(d)** Predicted bandgap and dielectric constant of the modified polymers compared to the base polymer.

Figure 5d shows the effects of the post-polymerization modification on the properties of a base vinyl polymer. The results show that the post-polymerization modification significantly alter the bandgap and dielectric constant of the polymer as predicted using a Polymer Genome-based machine learning model,[50] implemented in PolymRize.[46] The modification performed by epoxidation (polymer 5) simultaneously increase the bandgap and dielectric constant of the polymer, which is desirable for dielectric materials. Similarly, the modification performed by the thiol-ene reaction (polymer 2) significantly increase the dielectric constant of the polymer while maintaining a low bandgap, which is desirable for conductive polymers. These results demonstrate the potential of the PPM workflow to modify the properties, which can be directly validated in the laboratory.

This *in-silico* PPM approach implemented here using RxnChainer ensures that the final polymer structures are synthesizable in practice as the same chemical modifications can be performed in the laboratory to improve the properties of a base polymer. The workflow is designed to be flexible and can be extended to other polymer classes by defining new reaction templates. In addition, the workflow can be applied iteratively to further modify the properties of a modified polymer. This approach, combined with an optimization algorithm such as Bayesian optimization or Genetic Algorithm, could be used to automatically search for an optimal polymer chemistry that is directly reproducible in the laboratory.

Downstream Applications for Polymer Discovery

The library of hypothetical polymers generated in this work using the RxnChainer methodology enable multiple downstream utilities. These applications leverage the diversity, synthetic accessibility, and property predictions associated with the generated polymer structures to accelerate both computational and experimental research efforts. We highlight a few use cases below.

I. Accessible Property Values of Generated Polymers

We predicted several key properties for the newly generated polymers, including bandgap, glass transition temperature (Tg), and dielectric constant. To do this, we used machine learning models implemented in PolymRize that were trained on extensive datasets of known polymers with computational or experimentally measured properties. The accuracy of these models have been validated in prior studies, demonstrating their capability to provide reliable property predictions for novel polymer chemistries.[25,38,50] We applied these models to a randomly selected subset of 10 million polymers from our generated library, covering various polymer classes.

As shown in Figure 6, a wide range of property values is covered by different polymer chemistries and they exhibit substantial variation across the generated polymer classes. Bandgap values range from approximately 0 eV to nearly 8 eV, with the majority of polymers falling within the 2-5 eV range. Glass transition temperatures (Tg) span from below 200 K to above 600 K, with a concentration of values between 300 K and 500 K. Notably, polyimides and polyethers exhibit higher Tg distributions relative to other polymer classes, consistent with their known thermal stability. Dielectric constants vary widely from approximately 2 to over 8 at 60 Hz, with polyamides displaying a higher dielectric constant distribution compared to other polymer classes, which can be attributed to their polar chemical structure. These observed trends are consistent with the established characteristics of these polymer classes and validate the predictive capability of the employed machine learning models.

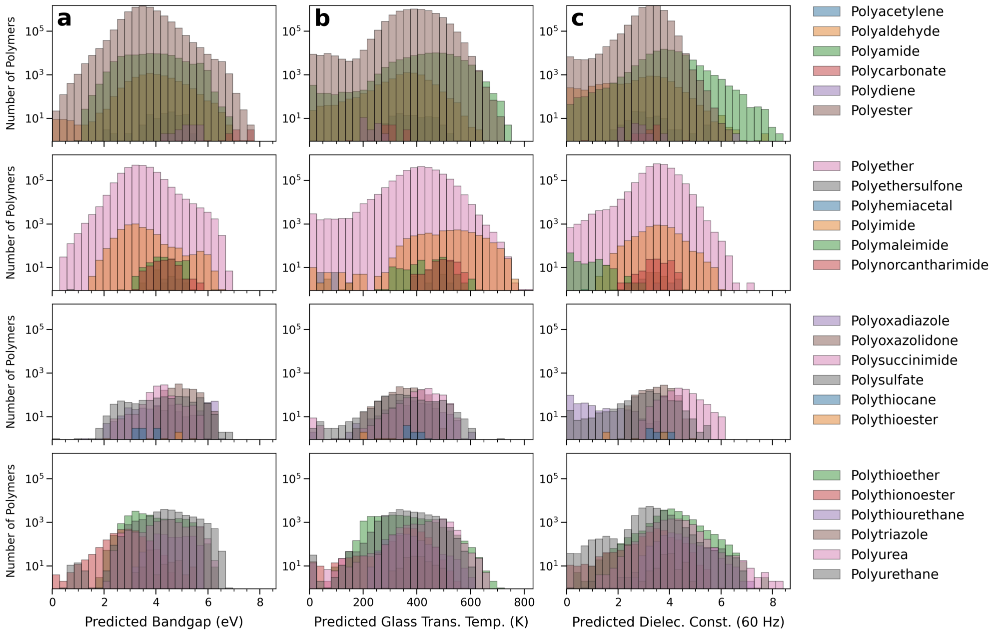


Figure 6. (a) Bandgap, (b) Glass Transition Temperature (Tg), and (c) Dielectric Constant distributions of randomly selected 10 million generated polymers predicted using the PolymRize ML models. The histograms show the number of polymers (y-axis, log scale) versus the predicted property values (x-axis). Different colors represent different polymer classes grouped by a block of 6. Dielectric constant is shown for frequency 60 Hz.

II. Chemical Space Exploration and Diversity Analysis

The systematically generated library of the hypothetical polymers also enables quantitative analysis of accessible polymer chemical space. We computed Uniform Manifold Approximation and Projections (UMAPs) using Morgan fingerprints (see Figure S1), which compare 10,000 generated polymers randomly selected from the top 16 polymer classes against 13,000 known homopolymers from the PolyInfo database. The results suggest that while many of the generated polymers overlap with known space, a significant portion represent entirely novel chemistries warranting investigation. The UMAP projections also reveal regions of high structural novelty where property predictions may be less reliable, guiding prioritization of experimental characterization to fill data gaps. Conversely, regions of high similarity to known polymers provide opportunities for targeted optimization of established chemistries. This dual capability supports both exploratory research into uncharted chemical spaces and iterative improvement of existing materials.

III. Property-Driven Screening: Example of Dielectrics Design

The availability of 289 million synthetically accessible polymer structures with predicted properties enables high-throughput virtual screening for application-specific design. This capability supports the systematic exploration of the structure-property relationships across vast chemical spaces that would be impractical to synthesize and test experimentally. For instance, polymers for high-energy-density capacitors, a critical application for advancing energy storage technologies, must satisfy multiple property constraints simultaneously. High glass transition temperatures ($T_g > 425$ K) ensure thermal stability, wide bandgaps ($E_g > 4$ eV) enable higher breakdown fields, and elevated dielectric constants ($\epsilon > 3$ at 60 Hz) increase energy storage capacity. Since energy density scales as $U \sim \frac{1}{2}\epsilon E^2$, both high dielectric constant and breakdown strength are critical. [19]

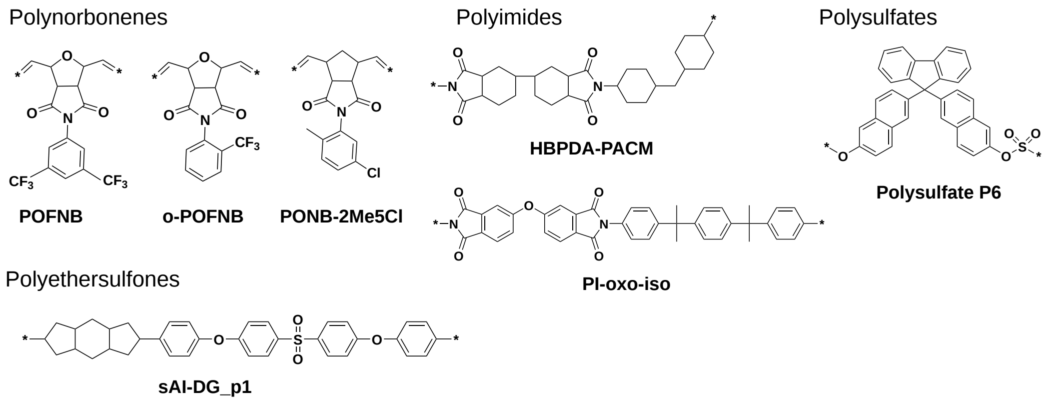


Figure 7. Structures of recently synthesized all-organic homopolymer dielectrics rediscovered in the RxnChainer-generated library, spanning polynorbornenes, polyimides, polysulfates, and polyethersulfones. Corresponding ML-predicted and experimentally measured properties are listed in Table 2.

When applying these screening criteria to our virtual library, numerous candidates passed the property thresholds. Notably, several of these candidates match recently synthesized all-organic homopolymer dielectrics reported in the literature (Table 2). Among these rediscovered polymers, polynorbonene-based dielectrics including POFNB, o-POFNB, and PONB-2Me-5Cl exhibit glass transition temperatures ranging from 459 to 517 K and bandgaps between 4.39 and 5.0 eV, with measured energy densities reaching 5.7 to 8.3 J/cc. [51–53] Polyimides such as HBPDA-PACM and PI-oxo-iso achieve even higher T_g values (527–531 K) and exceptionally wide bandgaps up to 6.32 eV, with energy densities of 2.1 to 5.65 J/cc. [19,54] Polysulfate P6 demonstrates a T_g of 551 K, a bandgap of 3.8 eV, dielectric constant of 3.4 at 473 K, and an energy density of 6.37 J/cc. [55] The polyethersulfone

doi:10.20944/preprints202605.0977.v1

sAI-DG_p1 achieves the highest T_g of 573 K with a bandgap of 4.0 eV, dielectric constant of 3.7 at 473 K, and energy density of 6.2 J/cc. [56] The fact that these recently synthesized high-performance polymers were independently identified through our computational screening demonstrates that RxnChainer can effectively navigate the vast polymer design space to find promising candidates.

Table 2. Recently synthesized all-organic homopolymer dielectrics rediscovered in the RxnChainer-generated library. For each property, the first line shows the ML-predicted value and the second line (in italics) shows the experimentally measured value. Measurement conditions for dielectric constant are noted in parentheses.

Polymer	**Class**	T_g **(K)**	**Bandgap (eV)**	ϵ	**Ref.**
POFNB	Polynorbonene (Polyalkene)	453.5 *459*	4.433 *4.9*	2.543 *2.5*^{a}	[51]
o-POFNB	Polynorbonene (Polyalkene)	503.9 *517*	4.402 *5.0*	2.892 *2.8*^{b}	[52]
PONB-2Me-5Cl	Polynorbonene (Polyalkene)	506.7 *505*	4.659 *4.39*	2.909 *3.0*^{b}	[53]
HBPDA-PACM	Alicyclic polyimide	496.9 *527*	5.616 *6.32*	2.764 *2.8*^{a}	[54]
PI-oxo-iso	Polyimide Polyimide	511.5 *531*	3.308 *4.3*	3.395 —	[19]
Polysulfate P6	Polysulfate Polysulfate	538.0 *551*	3.462 *3.8*	2.913 *3.4*^{b}	[55]
sAI-DG_p1	Polyethersulfone Polyethersulfone	493.7 *573*	3.893 *4.0*	2.678 *3.7*^{b}	[56]

^{a} Measured at 423 K, 1 kHz. ^{b} Measured at 473 K, 1 kHz.

IV. Training Data for Generative Models

Beyond screening existing designs, the generated datasets provide training data for polymer-specific generative models. Large-scale polymer datasets enable pretraining of chemical language models such as polyBERT,[38] polyT5 [40] and polyBART [39], which learn representations of polymer structures and chemical space from SMILES strings. The diversity of polymer classes (32 classes spanning addition, condensation, and ring-opening polymerizations) and structural variety within each class provides a comprehensive coverage of polymer chemistry grammar. These pretrained models can subsequently be fine-tuned for specific tasks including property prediction, inverse design, and reaction outcome prediction. The systematic generation approach ensures that training data encompasses both common polymer architectures and novel structural motifs, improving model generalization.

Conclusion

RxnChainer provides a unified computational methodology integrating polymer generation, retrosynthesis, and post-polymerization modification for systematically exploring synthetically accessible polymer space. Using commercially available monomers from the TSCA and ChEMBL databases, we generated over 289 million hypothetical polymers across 44 polymerization pathways spanning 32 polymer classes. Comparison with the PolyInfo database reveals that a substantial fraction of these structures are novel yet synthetically accessible, significantly expanding the known polymer chemical space. Beyond generation, retrosynthetic analysis of 30,000 polyesters achieved 95% accuracy in identifying constituent monomers, demonstrating automated synthetic route planning. Post-polymerization modification pathways targeting vinyl and nitrile pendant groups enable property tuning through triazole formation, thiol-ene, epoxidation, and halogenation reactions. The generated library enables multiple downstream applications. Machine learning property predictions reveal wide accessible ranges—bandgaps from 0 to 8 eV, glass transition temperatures from 200 to 600 K, and dielectric constants from 2 to 8—supporting high-throughput virtual screening. Property-driven screening independently rediscovered recently synthesized high-performance dielectrics including polynorbornenes, polyimides, polysulfates, and polyethersulfones, validating the approach. The library also provide comprehensive training data for polymer-specific language models. These results demonstrate RxnChainer's capability to bridge computational design with practical polymer discovery.

Supplementary Materials: The supplementary information related to this work is available free of charge at Preprints.org. List of polymerization reaction templates implemented in RxnChainer, and additional figures and tables (PDF). References [57–81] are cited in the supplementary materials

Author Contributions: Conceptualization: R.R., S.S., A.M., Data generation and analysis: S.S., A.M., Visualization: S.S., A.M., C.K., Original draft: S.S., A.M., Review & editing: S.S., A.M., C.K., R.G., R.R.

Data Availability Statement: The datasets used to generate the polymers are available from official TSCA inventory at https://www.epa.gov/tsca-inventory/ and the ChEMBL database at https://www.ebi.ac.uk/chembl/. The known polymer datasets used in this study were derived from the publicly accessible PolyInfo database (https://polymer.nims.go.jp). The RxnChainer methodology has been implemented in the PolymRize platform available at https://polymrize.matmerize.com/ and can be used to reproduce the results presented in this work.

Conflicts of Interest: The authors declare no competing financial interest.

References

1. Siracusa, V.; Rocculi, P.; Romani, S.; Rosa, M. D. Biodegradable polymers for food packaging: a review. *Trends in Food Science & Technology* **2008**, *19*, 634–643, DOI: `https://doi.org/10.1016/j.tifs.2008.07.003`.
2. Cao, Y.; Uhrich, K. E. Biodegradable and biocompatible polymers for electronic applications: A review. *Journal of Bioactive and Compatible Polymers* **2019**, *34*, 3–15, DOI: `10.1177/0883911518818075`.
3. Ulery, B. D.; Nair, L. S.; Laurencin, C. T. Biomedical applications of biodegradable polymers. *Journal of Polymer Science, Part B: Polymer Physics* **2011**, *49*, 832–864, DOI: `10.1002/polb.22259`.
4. Wright, W. Polymers in aerospace applications. *Materials & Design* **1991**, *12*, 222–227, DOI: `https://doi.org/10.1016/0261-3069(91)90169-5`.
5. Ghori, S. W.; Siakeng, R.; Rasheed, M.; Saba, N.; Jawaid, M. In *Sustainable Composites for Aerospace Applications*; Jawaid, M., Thariq, M., Eds.; Woodhead Publishing Series in Composites Science and Engineering; Woodhead Publishing, 2018; pp 19–34, DOI: `https://doi.org/10.1016/B978-0-08-102131-6.00002-5`.
6. Rathod, V. T.; Kumar, J. S.; Jain, A. Polymer and Ceramic Nanocomposites for Aerospace Applications. *Applied Nanoscience* **2017**, *7*, 519–548, DOI: `10.1007/s13204-017-0592-9`.
7. González-Campos, J. B.; Luna-Bárcenas, G.; Zárate-Triviño, D. G.; Mendoza-Galván, A.; Prokhorov, E.; Villaseñor-Ortega, F.; Sanchez, I. C. *Handbook of Polymer Synthesis, Characterization, and Processing*; John Wiley & Sons, Ltd, 2013; Chapter 2, pp 15–39, DOI: `https://doi.org/10.1002/9781118480793.ch2`.
8. Pfaendner, R. *Handbook of Polymer Synthesis, Characterization, and Processing*; John Wiley & Sons, Ltd, 2013; Chapter 11, pp 225–247, DOI: `https://doi.org/10.1002/9781118480793.ch11`.
9. Robledo-Ortíz, J. R.; Fuentes-Talavera, F. J.; González-Núñez, R.; Silva-Guzmán, J. A. *Handbook of Polymer Synthesis, Characterization, and Processing*; John Wiley & Sons, Ltd, 2013; Chapter 26, pp 493–503, DOI: `https://doi.org/10.1002/9781118480793.ch26`.
10. Sánchez-Valdes, S.; Ramos-De Valle, L. F.; Manero, O. *Handbook of Polymer Synthesis, Characterization, and Processing*; John Wiley & Sons, Ltd, 2013; Chapter 27, pp 505–517, DOI: `https://doi.org/10.1002/9781118480793.ch27`.
11. Saldívar-Guerra, E.; Vivaldo-Lima, E. *Handbook of Polymer Synthesis, Characterization, and Processing*; John Wiley & Sons, Ltd, 2013; Chapter 1, pp 1–14, DOI: `https://doi.org/10.1002/9781118480793.ch1`.
12. Pillai, C. K. S. Challenges for Natural Monomers and Polymers: Novel Design Strategies and Engineering to Develop Advanced Polymers. *Designed Monomers and Polymers* **2010**, *13*, 87–121, DOI: `10.1163/138577210X12634696333190`.
13. Ma, R.; Baldwin, A. F.; Wang, C.; Offenbach, I.; Cakmak, M.; Ramprasad, R.; Sotzing, G. A. Rationally Designed Polyimides for High-Energy Density Capacitor Applications. *ACS Applied Materials & Interfaces* **2014**, *6*, 10445–10451, DOI: `10.1021/am502002v`.

14. Fazio, A.; Caroleo, M. C.; Cione, E.; Plastina, P. Novel Acrylic Polymers for Food Packaging: Synthesis and Antioxidant Properties. *Food Packaging and Shelf Life* **2017**, *11*, 84–90, DOI: `10.1016/j.fpsl.2017.01.002`.
15. Ilinitch, O.; Semin, G.; Chertova, M.; Zamaraev, K. Novel Polymeric Membranes for Separation of Hydrocarbons. *Journal of Membrane Science* **1992**, *66*, 1–8, DOI: `10.1016/0376-7388(92)80085-X`.
16. Kerres, J.; Ullrich, A.; Meier, F.; Häring, T. Synthesis and Characterization of Novel Acid–Base Polymer Blends for Application in Membrane Fuel Cells. *Solid State Ionics* **1999**, *125*, 243–249, DOI: `10.1016/S0167-2738(99)00181-2`.
17. Wang, Y.; Ding, Y.; Yu, K.; Dong, G. Innovative polymer-based composite materials in additive manufacturing: A review of methods, materials, and applications. *Polymer Composites* **2024**, *45*, 15389–15420, DOI: `https://doi.org/10.1002/pc.28854`.
18. Wang, Y.; Zhou, X.; Chen, Q.; Chu, B.; Zhang, Q. Recent development of high energy density polymers for dielectric capacitors. *IEEE Transactions on Dielectrics and Electrical Insulation* **2010**, *17*, 1036–1042, DOI: `10.1109/TDEI.2010.5539672`.
19. Gurnani, R.; Shukla, S.; Kamal, D.; Wu, C.; Hao, J.; Kuenneth, C.; Aklujkar, P.; Khomane, A.; Daniels, R.; Deshmukh, A. A.; Cao, Y.; Sotzing, G.; Ramprasad, R. AI-assisted Discovery of High-Temperature Dielectrics for Energy Storage. *Nature Communications* **2024**, *15*, 6107, DOI: `10.1038/s41467-024-50413-x`.
20. Dangayach, R.; Jeong, N.; Demirel, E.; Uzal, N.; Fung, V.; Chen, Y. Machine Learning-Aided Inverse Design and Discovery of Novel Polymeric Materials for Membrane Separation. *Environmental Science & Technology* **2025**, *59*, 993–1012, DOI: `10.1021/acs.est.4c08298`.
21. Ramprasad, R.; Batra, R.; Pilania, G.; Mannodi-Kanakkithodi, A.; Kim, C. Machine Learning in Materials Informatics: Recent Applications and Prospects. *npj Computational Materials* **2017**, *3*, 54, DOI: `10.1038/s41524-017-0056-5`.
22. Gaulton, A.; Hersey, A.; Nowotka, M.; Bento, A. P.; Chambers, J.; Mendez, D.; Mutowo, P.; Atkinson, F.; Bellis, L. J.; Cruz, C. M.; others ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Research* **2017**, *45*, D945–D954, DOI: `10.1093/nar/gkw1074`.
23. Sterling, T.; Irwin, J. J. ZINC15 – Ligand discovery for everyone. *Journal of Chemical Information and Modeling* **2015**, *55*, 2324–2337, DOI: `10.1021/acs.jcim.5b00559`.
24. eMolecules Database. https://www.emolecules.com, 2025; Accessed: 2025-08-28.
25. Kim, S.; Schroeder, C. M.; Jackson, N. E. Open Macromolecular Genome: Generative Design of Synthetically Accessible Polymers. *ACS Polymers Au* **2023**, *3*, 318–330, DOI: `10.1021/acspolymersau.3c00003`.
26. Ohno, M.; Hayashi, Y.; Zhang, Q.; Kaneko, Y.; Yoshida, R. SMiPoly: Generation of a Synthesizable Polymer Virtual Library Using Rule-Based Polymerization Reactions. *Journal of Chemical Information and Modeling* **2023**, *63*, 5539–5548, DOI: `10.1021/acs.jcim.3c00329`.
27. Yue, T.; He, J.; Li, Y. Polyuniverse: Generation of a Large-Scale Polymer Library Using Rule-Based Polymerization Reactions for Polymer Informatics. *Digital Discovery* **2024**, *3*, 2465–2478, DOI: `10.1039/D4DD00196F`.
28. Kern, J.; Su, Y.-L.; Gutekunst, W. R.; Ramprasad, R. An Informatics Framework for the Design of Sustainable, Chemically Recyclable, Synthetically Accessible, and Durable Polymers. *npj Computational Materials* **2025**, *11*, 182, DOI: `10.1038/s41524-025-01683-6`.
29. Phan, B. K.; Kim, C.; Nistane, J.; Xiong, W.; Chen, H.; Jang, W. J.; Gholami, F.; Su, Y.; Qi, J.; Lively, R.; Gutekunst, W.; Ramprasad, R. AI-assisted design of chemically recyclable polymers for food packaging. **2025**,
30. Chen, L.; Kern, J.; Lightstone, J. P.; Ramprasad, R. Data-Assisted Polymer Retrosynthesis Planning. *Applied Physics Reviews* **2021**, *8*, 031405, DOI: `10.1063/5.0052962`.
31. Gauthier, M.; Gibson, M.; Klok, H.-A. Synthesis of Functional Polymers by Post-Polymerization Modification. *Angewandte Chemie International Edition* **2009**, *48*, 48–58, DOI: `https://doi.org/10.1002/anie.200801951`.
32. Urban, V. M.; Machado, A. L.; Vergani, C. E.; Giampaolo, E. T.; Pavarina, A. C.; de Almeida, F. G.; Cass, Q. B. Effect of Water-Bath Post-Polymerization on the Mechanical Properties, Degree of Conversion, and Leaching of Residual Compounds of Hard Chairside Reline Resins. *Dental Materials* **2009**, *25*, 662–671, DOI: `10.1016/j.dental.2008.10.017`.
33. Farmer, T. J.; Comerford, J. W.; Pellis, A.; Robert, T. Post-polymerization modification of bio-based polymers: maximizing the high functionality of polymers derived from biomass. *Polymer International* **2018**, *67*, 775–789, DOI: `https://doi.org/10.1002/pi.5573`.
34. Lv, A.; Cui, Y.; Du, F.-S.; Li, Z.-C. Thermally Degradable Polyesters with Tunable Degradation Temperatures via Postpolymerization Modification and Intramolecular Cyclization. *Macromolecules* **2016**, *49*, 8449–8458, DOI: `10.1021/acs.macromol.6b01325`.

35. Agar, S.; Baysak, E.; Hizal, G.; Tunca, U.; Durmaz, H. An emerging post-polymerization modification technique: The promise of thiol-para-fluoro click reaction. *Journal of Polymer Science Part A: Polymer Chemistry* **2018**, *56*, 1181–1198, DOI: `https://doi.org/10.1002/pola.29004`.
36. U.S. Environmental Protection Agency TSCA Chemical Substance Inventory. https://www.epa.gov/tsca-inventory, 2025; Accessed: 2025-08-28.
37. Otsuka, S.; Kuwajima, I.; Hosoya, J.; Xu, Y.; Yamazaki, M. PoLyInfo: Polymer Database for Polymeric Materials Design. 2011 International Conference on Emerging Intelligent Data and Web Technologies. 2011; pp 22–29, DOI: `10.1109/EIDWT.2011.13`.
38. Kuenneth, C.; Ramprasad, R. polyBERT: A Chemical Language Model to Enable Fully Machine-Driven Ultrafast Polymer Informatics. *14*, 4099, DOI: `10.1038/s41467-023-39868-6`.
39. Savit, A.; Sahu, H.; Shukla, S.; Xiong, W.; Ramprasad, R. polyBART: A Chemical Linguist for Polymer Property Prediction and Generative Design. 2025; https://arxiv.org/abs/2506.04233.
40. Sahu, H.; Xiong, W.; Savit, A.; Shukla, S. S.; Ramprasad, R. An Encoder-Decoder Foundation Chemical Language Model for Generative Polymer Design. 2025; https://arxiv.org/abs/2510.18860.
41. rdkit/rdkit: 2025_03_5 (Q1 2025) Release. 2025; https://doi.org/10.5281/zenodo.16439048.
42. Daylight Chemical Information Systems, I.; Weininger, D. SMILES, a chemical language and information system. 3. SMARTS: A language for describing molecular patterns. *Journal of Chemical Information and Computer Sciences* **1989**, *29*, 97–101, DOI: `10.1021/ci00062a008`.
43. Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of chemical information and computer sciences* **1988**, *28*, 31–36.
44. Ertl, P.; Schuffenhauer, A. Estimation of Synthetic Accessibility Score of Drug-like Molecules Based on Molecular Complexity and Fragment Contributions. *Journal of Cheminformatics* **2009**, *1*, 8, DOI: `10.1186/1758-2946-1-8`.
45. Willett, P.; Barnard, J. M.; Downs, G. M. Chemical Similarity Searching. *Journal of Chemical Information and Computer Sciences* **1998**, *38*, 983–996, DOI: `10.1021/ci9800211`.
46. PolymRize 0.24.0 by Matmerize, Inc. 2025; https://polymrize.matmerize.com.
47. Landrum, G. RDKit: Open-source cheminformatics. http://www.rdkit.org, 2006; Accessed: 2025-08-28.
48. Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *Journal of Chemical Information and Modeling* **2010**, *50*, 742–754, DOI: `10.1021/ci100050t`.
49. Chen, J.; Yan, J.; Yang, W.; Xu, Y.; Chi, R.; Zhang, Q.; Yan, Y. Cobaltocenium-Containing Poly(Carbazole)s towards Alkaline-Stable Anion Exchange Membranes via Post-Polymerization Modification. *Sustainable Energy Fuels* **2024**, *8*, 4767–4771, DOI: `10.1039/D4SE00888J`.
50. Kim, C.; Chandrasekaran, A.; Huan, T. D.; Das, D.; Ramprasad, R. Polymer Genome: A Data-Powered Polymer Informatics Platform for Property Predictions. *The Journal of Physical Chemistry C* **2018**, *122*, 17575–17585, DOI: `10.1021/acs.jpcc.8b02913`.
51. Wu, C.; Deshmukh, A. A.; Li, Z.; Chen, L.; Alamri, A.; Wang, Y.; Ramprasad, R.; Sotzing, G. A.; Cao, Y. Flexible Temperature-Invariant Polymer Dielectrics with Large Bandgap. *Advanced Materials* **2020**, *32*, 2000499, DOI: `https://doi.org/10.1002/adma.202000499`.
52. Deshmukh, A. A. et al. Flexible polyolefin dielectric by strategic design of organic modules for harsh condition electrification. *Energy Environ. Sci.* **2022**, *15*, 1307–1314, DOI: `10.1039/D1EE02630E`.
53. Wang, R.; Zhu, Y.; Fu, J.; Yang, M.; Ran, Z.; Li, J.; Li, M.; Hu, J.; He, J.; Li, Q. Designing Tailored Combinations of Structural Units in Polymer Dielectrics for High-Temperature Capacitive Energy Storage. *14*, 2406, DOI: `10.1038/s41467-023-38145-w`.
54. Huang, W.; Wan, B.; Yang, X.; Cheng, M.; Zhang, Y.; Li, Y.; Wu, C.; Dang, Z.-M.; Zha, J.-W. Alicyclic Polyimide With Multiple Breakdown Self-Healing Based on Gas-Condensation Phase Validation for High Temperature Capacitive Energy Storage. *Advanced Materials* **2024**, *36*, 2410927, DOI: `https://doi.org/10.1002/adma.202410927`.
55. Li, H. et al. Machine Learning-Accelerated Discovery of Heat-Resistant Polysulfates for Electrostatic Energy Storage. *10*, 90–100, DOI: `10.1038/s41560-024-01670-z`.
56. Ren, W.; Tong, H.; Cao, S.; Zhao, S.; Yang, M.; Li, X.; Pan, J.; Sun, N.; Xiao, Y.; Xu, E.; Nan, C.-W.; Shen, Y. Semi-Alicyclic Dipolar Glass Dielectric Polymer Capacitors for Superior High-Temperature Capacitive Energy Storage. *Advanced Materials* **2025**, *37*, e05296, DOI: `https://doi.org/10.1002/adma.202505296`.
57. Young, R. J.; Lovell, P. A. *Introduction to Polymers*, 3rd ed.; CRC Press: Boca Raton, FL, 2011; pp 21–60.

58. Hill, J. W.; McCreary, T. W. Addition Polymerization — "One + One + One + ... Gives One". 2022; https://chem.libretexts.org/Bookshelves/Introductory_Chemistry/Chemistry_for_Changing_Times_(Hill_and_McCreary)/10%3A_Polymers/10.03%3A_Addition_Polymerization_-_One__One__One__..._Gives_One, Accessed: 2025-10-03.
59. Ophardt, C.; Padleckas, H. A. Rubber Polymers. 2023; https://chem.libretexts.org/Bookshelves/Organic_Chemistry/Supplemental_Modules_(Organic_Chemistry)/Polymers/Rubber_Polymers, Accessed: 2025-12-18.
60. Mison, P.; Sillion, B. In *Progress in Polyimide Chemistry I*; Kricheldorf, H. R., Ed.; Advances in Polymer Science; Springer: Berlin, Heidelberg, 1999; Vol. 140; pp 65–100, DOI: `10.1007/3-540-49815-X_5`.
61. Nie, Z.; Korre, A.; Durucan, S. Life cycle modelling and comparative assessment of the environmental impacts of oxy-fuel and post-combustion CO_2 capture, transport and injection processes. *Energy Procedia* **2011**, *4*, 2510–2517, DOI: `10.1016/j.egypro.2011.02.147`.
62. Mitchell, S. M.; Sachinthani, K. A. N.; Pulukkody, R.; Pentzer, E. B. 100th Anniversary of Macromolecular Science Viewpoint: Polymerization of Cumulated Bonds: Isocyanates, Allenes, and Ketenes as Monomers. *ACS Macro Letters* **2020**, *9*, 1046–1059, DOI: `10.1021/acsmacrolett.0c00396`.
63. Sastri, V. R. In *Plastics in Medical Devices*, 2nd ed.; Sastri, V. R., Ed.; William Andrew Publishing, 2014; pp 121–172, DOI: `10.1016/B978-1-4557-3201-2.00007-0`.
64. Hermanson, G. T. In *Bioconjugate Techniques*, 3rd ed.; Hermanson, G. T., Ed.; Academic Press, 2013; pp 275–298, DOI: `10.1016/B978-0-12-382239-0.00005-4`.
65. Odian, G. *Principles of Polymerization*, 4th ed.; Wiley-Interscience: Hoboken, NJ, 2004; pp 39–197.
66. Hoyle, C. E.; Bowman, C. N. Thiol–ene Click Chemistry. *Angewandte Chemie International Edition* **2010**, *49*, 1540–1573, DOI: `10.1002/anie.200903924`.
67. Hoyle, C. E.; Bowman, C. N. Thiol–yne Click Chemistry. *Chemical Society Reviews* **2010**, *39*, 1355–1387, DOI: `10.1039/B917102A`.
68. Zhang, Y.; Chen, G.; Lin, Y.; Zhao, L.; Yuan, W. Z.; Lu, P.; Jim, C. K. W.; Zhang, Y.; Tang, B. Z. Thiol–bromo Click Polymerization for Multifunctional Polymers: Synthesis, Light Refraction, Aggregation-Induced Emission and Explosive Detection. *Polymer Chemistry* **2015**, *6*, 97–105, DOI: `10.1039/C4PY01164C`.
69. González, G.; Sardon, H. Polythiourethanes: Synthesis, Applications, and Opportunities. *Progress in Polymer Science* **2024**, *145*, 101487, DOI: `10.1016/j.progpolymsci.2023.101487`.
70. Geng, Z.; Shin, J. J.; Xi, Y.; Hawker, C. J. Click Chemistry Strategies for the Accelerated Synthesis of Functional Macromolecules. *Journal of Polymer Science* **2021**, *59*, 963–1042, DOI: `10.1002/pol.20210126`.
71. Wang, H.; Zhou, F.; Ren, G.; Zheng, Q.; Chen, H.; Gao, B.; Klivansky, L.; Liu, Y.; Wu, B.; Xu, Q.; Lu, J.; Sharpless, K. B.; Wu, P. SuFEx-Based Polymers with Tunable Physical Properties from Sulfur(VI) Fluoride Exchange Click Chemistry. *Angewandte Chemie International Edition* **2017**, *56*, 11203–11208, DOI: `10.1002/anie.201701160`.
72. Collins, J.; Xiao, Z.; Espinosa-Gomez, A.; Fors, B. P.; Connal, L. A. Extremely Rapid and Versatile Synthesis of High Molecular Weight Step Growth Polymers via Oxime Click Chemistry. *Polymer Chemistry* **2016**, *7*, 2581–2588, DOI: `10.1039/C6PY00372A`.
73. Kumar, G. S.; Lin, Q. Light-Triggered Click Chemistry. *Chemical Reviews* **2021**, *121*, 6991–7031, DOI: `10.1021/acs.chemrev.0c00799`.
74. Hizal, G.; Tunca, U.; Sanyal, A. Discrete macromolecular constructs via the Diels–Alder "Click" reaction. *Journal of Polymer Science Part A: Polymer Chemistry* **2011**, *49*, 4103–4120, DOI: `https://doi.org/10.1002/pola.24835`.
75. Odian, G. *Principles of Polymerization*, 4th ed.; Wiley-Interscience: Hoboken, NJ, 2004; pp 544–618.
76. Li, H.; Guillaume, S. M.; Carpentier, J.-F. Polythioesters Prepared by Ring-Opening Polymerization of Cyclic Thioesters and Related Monomers. *Chemistry – An Asian Journal* **2022**, *17*, e202200641, DOI: `10.1002/asia.202200641`.
77. Kern, J.; Su, Y.-L.; Gutekunst, W.; Ramprasad, R. An Informatics Framework for the Design of Sustainable, Chemically Recyclable, Synthetically Accessible, and Durable Polymers. *npj Computational Materials* **2025**, *11*, 182, DOI: `10.1038/s41524-025-01683-6`.
78. Kamber, N. E.; Jeong, W.; Waymouth, R. M.; Pratt, R. C.; Lohmeijer, B. G. G.; Hedrick, J. L. Organocatalytic Ring-Opening Polymerization. *Chemical Reviews* **2007**, *107*, 5813–5840, DOI: `10.1021/cr068415b`.
79. Kern, J.; Su, Y.; Gutekunst, W.; Ramprasad, R. An Informatics Framework for the Design of Sustainable, Chemically Recyclable, Synthetically-Accessible and Durable Polymers. 2024.

80. Yang, H.; Sun, A.; Chai, C.; Huang, W.; Xue, X.; Chen, J.; Jiang, B. Synthesis and Post-Functionalization of a Degradable Aliphatic Polyester Containing Allyl Pendent Groups. *Polymer* **2017**, *121*, 256–261, DOI: `10.1016/j.polymer.2017.06.029`.
81. Sruthi, P. R.; Anas, S. An overview of synthetic modification of nitrile group in polymers and applications. *Journal of Polymer Science* **2020**, *58*, 1039–1061, DOI: `https://doi.org/10.1002/pol.20190190`.