AI-driven design of poly(ethylene terephthalate) replacement copolymers

Chiho Kim,[1,2,*] Wei Xiong,[1] Akhlak Mahmood,[2] Rampi Ramprasad,[1,2] and Huan Tran[1,2]
[1]*School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, USA*
[2]*Matmerize, Inc., Atlanta, Georgia 30308, USA*

(Received 21 August 2025; accepted 25 February 2026; published 23 March 2026)

Poly(ethylene terephthalate) (PET), a widely used thermoplastic in packaging, textiles, and engineering applications, is valued for its strength, clarity, and chemical resistance. Increasing environmental impact concerns and regulatory pressures drive the search for alternatives with comparable or superior performance. We present an AI-driven polymer design pipeline employing virtual forward synthesis (VFS) to generate PET-replacement copolymers. Inspired by the esterification route of PET synthesis, we systematically combined a down-selected set of Toxic Substances Control Act (TSCA)-listed monomers to create 12 100 PET-like polymers. Machine learning models predicted glass transition temperature (Tg), band gap, and tendency to crystallize, for all designs. Multi-objective screening identified 1108 candidates predicted to match or exceed PET in Tg and band gap, including the "rediscovery" of other known commercial PET-alternate polymers (e.g., PETG, Tritan) that provide retrospective validation of our design pipeline, demonstrating a capability to rapidly design experimentally feasible polymers at a scale. Furthermore, selected, entirely new (previously unknown) candidates designed here have been synthesized and characterized, providing a definitive validation of the design framework.

DOI: 10.1103/wjcq-jv1c

I. INTRODUCTION

Designing synthetic polymers has been a central focus of polymer and plastics R&D in both industry and academia for over a century, driving the development of materials with targeted properties for diverse applications [1]. Landmark examples such as Kevlar, Teflon, and Nylon, having originated from laboratory innovations reached commercialization through extensive optimization, validation, and feasibility testing [2–4]. Recent global regulatory efforts, including the push for PFAS-free materials [5], the U.S. Toxic Substances Control Act (TSCA) [6], and the European REACH regulation, have made the development of new polymers increasingly complex [7]. These initiatives demand materials that are nontoxic, sustainable, and environmentally friendly, further constraining the allowable chemical space.

Traditionally, polymer discovery has relied on trial-and-error experiments, which are time consuming and expensive. The vastness of the chemical space, including combinations of homopolymers and copolymers, neat resins and composites, as well as formulation and processing variables, makes efficient exploration and optimization challenging. To address these limitations, artificial intelligence and machine learning (AI/ML) have been actively adopted [8–15]. These approaches connect data collection and curation, numerical representation of polymer structures (fingerprinting) [14,15], model training [11], and property prediction to guide experimental pathway. While early efforts focused on building predictive models and robust fingerprinting schemes, current trends are shifting toward AI-driven generation and design of new candidate materials [16–20].

This work presents a polymer design pipeline that employs virtual forward synthesis (VFS) [19] within an AI-assisted design pipeline to craft alternative copolymers for poly(ethylene terephthalate) (PET), one of the widely used thermoplastics in the polyester family. The PET is commonly found in transparent and thermally resistant plastics used for bottles and packaging. In this use case, the design objective is to identify candidate copolymers that exceed PET's performance in key properties, specifically, the target criteria include the glass transition temperature (Tg) higher than PET's Tg range of 65 °C to 81 °C [21], and a band gap as a proxy for optical transparency greater than PET's measured band gap 3.9 eV [22]. The transparency is also associated with amorphous phase in PET as the crystallinity tends to increase light scattering [23]. Thus, the tendency to crystallization is used as a third design objective. The goal is to identify materials with lower tendency to crystallize than PET, which typically exhibits a crystallinity of less than 40% (Table I). Two selected candidate copolymers were successfully synthesized and experimentally tested, confirming their predicted properties and demonstrating the practical feasibility of the proposed design approach.

II. METHODS

All procedures described in this study, including (1) training predictive models for Tg, band gap, and tendency to crystallize, (2) generation of PET-co-(PET-like) copolymer designs through VFS, (3) multi-objective screening based on target criteria, (4) experimental validation of polymers selected based on synthetic accessibility (SA) score [24], and

*Contact author: chiho.kim@gatech.edu

TABLE I. Target criteria for design PET-replacement polymers.

Target property	PET, experimental (predicted)	Design goal (predicted)
Glass transition temperature	65 °C - 81 °C (72 °C)	Greater than 72 °C
Band gap	3.9 eV (3.7 eV)	Greater than 3.7 eV
Tendency to crystallize	Less than 40% (40%)	Less than 40%

(5) feature importance analysis and correlation analysis using SHapley Additive exPlanations (SHAP) [25] analysis and principal component analysis (PCA) [26], were carried out using PolymRize™ platform [27], a standardized polymer informatics software.

Three predictive ML models were developed to predict Tg, band gap, and tendency to crystallize for new polymer designs. The Tg model was trained on 8962 experimentally measured values from a diverse set of homo and copolymers. The band gap model was trained on 562 density functional theory (DFT)-calculated values using atomistic homopolymer models packed in 3D supercells [28]. The tendency to crystallize model was developed using 111 experimental data points, supported by 432 computationally derived values (based on heat of formation) within a multi-target co-learning framework. All models employed Gaussian process regression (GPR) [29] with radial basis function and white noise kernels, using Polymer Genome fingerprinting scheme, implemented in PolymRize™, to numerically represent the generated copolymers. Model performance on test sets averaged over 5-fold cross-validation (CV) showed root mean square errors (RMSEs) of 31 K for Tg, 0.6 eV for the band gap, and 12% for crystallization tendency, while the final models trained on the full dataset exhibited substantially lower overall RMSEs of 13 K, 0.19 eV, and 10%, respectively. In addition, predictive uncertainty from the Gaussian process was incorporated into the screening criteria so that only high-confidence predictions were retained.

To apply VFS for generating synthetically feasible polymers structurally similar to PET, it is critical to understand the underlying reaction chemistry. Two well-established reaction mechanisms are (1) esterification between terephthalic acid (TPA) and ethylene glycol (EG), and (2) transesterification between dimethyl terephthalate (DMT) and EG. This study focuses on the esterification route, which is a straightforward and classical polycondensation reaction, as illustrated in Fig. 1(a). Candidate monomers for TPA and EG analogs were down-selected from the TSCA inventory of 28 289 registered molecules based on the structural criteria essential for esterification. For TPA replacements, 72 monomers were selected based on the presence of two terminal -COOH groups and five- or six-membered carbon rings, either aromatic or aliphatic. For EG replacements, 168 monomers were chosen using the following criteria: terminal -OH groups, composed solely of carbon and hydrogen functional groups, and a maximum of eight carbon atoms to limit the size and feasibility of the generated products. Through combinatorial pairing of these 72 TPA-like and 168 EG-like monomers, 12 100 candidate PET-like polymers were generated via esterification. Two example structures, along with the CAS numbers of their monomer components, are shown in Fig. 1(b). By combining the original PET repeat unit with each PET-like polymer in a 1:1 molar ratio, a set of 12 100 PET-co-(PET-like) copolymers was constructed using the RxnChainer feature of PolymRize [30], and prepared for property prediction.

III. RESULTS AND DISCUSSION

Out of designed copolymers, 3390 cases met the target criteria of Tg higher than PET's predicted Tg (72 °C) and band gap larger than PET's predicted band gap (3.7 eV), as shown in Fig. 2. Applying a secondary filter based on model confidence, GPR 1-sigma uncertainty of predictions below 30 °C for Tg and 0.5 eV for band gap, narrowed the candidate

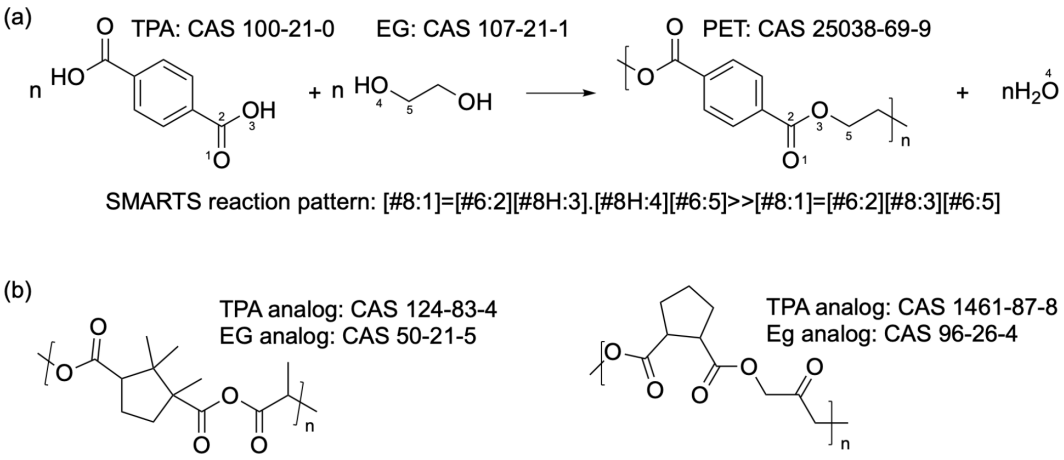

FIG. 1. (a) Reaction of terephthalic acid (TPA) and ethylene glycol (EG) to produce PET, and corresponding SMARTS representing condensation of groups—carboxyl from TPA and hydroxyl from EG. (b) Example PET-like polymers generated using the same reaction mechanism but different monomer reactants.

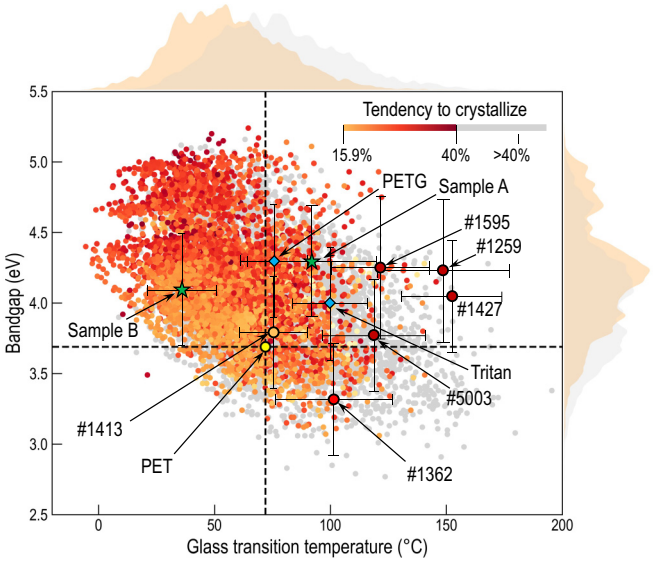

FIG. 2. Distribution of predicted Tg and band gap of 12 100 PET-co-(PET-like) copolymers, color-coded by tendency to crystallize. Two known commercial polymers (e.g., PETG, Tritan) reidentified by VFS fall within the desired design region. Error bars indicate the uncertainty of the GPR model, represented as one standard deviation of the predictions. The predicted Tg values are consistent with reported experimental values, approximately 80 °C for PETG [31] and 119 °C for Tritan [32]. Structure of labeled polymers are shown in Figs. 3 and 4.

set to 1108 polymers. Polymers with a predicted tendency to crystallize equal to or lower than 40% are highlighted in orange-red colormap in Fig. 2. All of 1108 polymers' tendency to crystalize met <40% criterion. Two commercial copolymers well-known as PET replacements, PETG and Tritan were among the predicted candidates. These alternatives have been widely adopted due to their improved processability, reduced brittleness, and enhanced clarity compared to PET. Their re-identification through our pipeline demonstrates that such successful PET alternatives can be readily and systematically recovered, providing clear validation of the approach. Figure 3 presents a selection of randomly chosen polymers from the final candidate pool that meet three screening criteria.

Taking into account the synthetic accessibility (SA) scores of the PET-like polymers and practical considerations from polymer synthesis experience, two candidates (Fig. 4) were selected for experimental validation. Both polymers were synthesized via solution polycondensation of the diacid chloride (TPA analog) and the diol (EG analog) monomers under basic conditions. The reactions were carried out under anhydrous nitrogen at low temperature to control the exothermic step-growth polymerization, followed by room-temperature stirring to complete chain growth. Sample A was prepared from terephthaloyl chloride and nadic acid chloride with EG, whereas Sample B employed terephthaloyl chloride with a mixed diol system of EG and triethylene glycol. The resulting polymers precipitated during the reaction and were isolated by filtration, acid washing, ether precipitation, and vacuum drying to afford purified copolyesters. The chemical structure and composition of the synthesized copolymers were confirmed by ^{1}H nuclear magnetic resonance (NMR) spectroscopy in $CDCl_3$ at room temperature. Characteristic resonances corresponding to both monomer units were clearly identified,

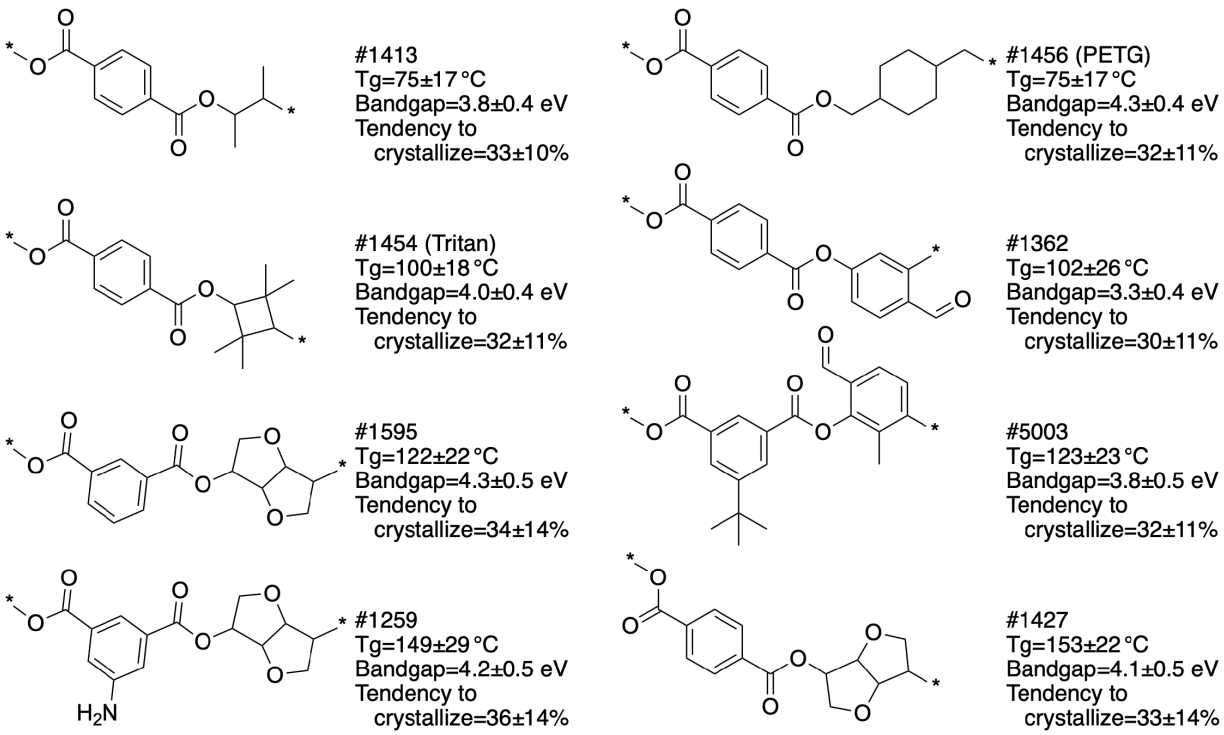

FIG. 3. Copolymers potential to replace PET with higher thermal resistance and comparable transparency. Symbols (*) indicate the endpoints of repeating unit. Only PET-like blocks of PET-co-(PET-like) copolymers are visualized.

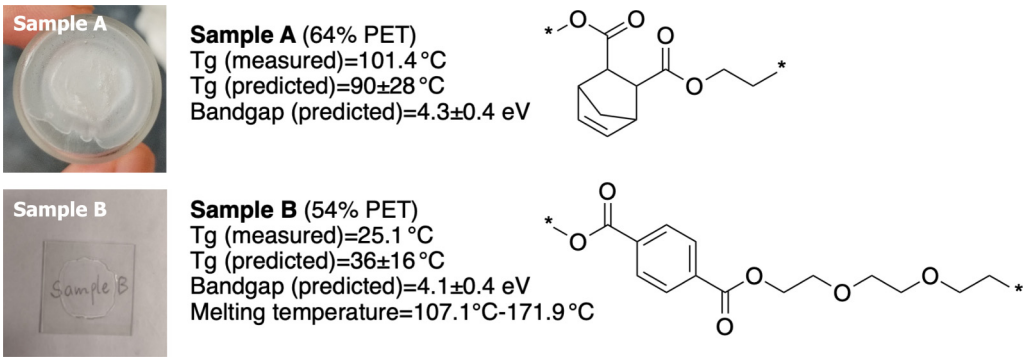

FIG. 4. Photographs, predicted properties, and chemical structures of potential PET replacement copolymers. Sample A exhibits an opaque appearance, indicating partial crystallinity or phase separation, whereas Sample B is fully transparent. The pencil written text visible through Sample B is positioned beneath the film, demonstrating its high optical clarity. Asterisks (*) denote the endpoints of the repeating units. For clarity, only the PET-like segments of the PET-co-(PET-like) copolymers are shown.

and quantitative integration of representative signals indicated molar ratios of approximately 1.79:1 for Sample A and 1.16:1 for Sample B (Fig. S1 of the Supplemental Material (SM) [33]). Detailed synthetic procedures and ^{1}H NMR spectra are also provided in the SM [33].

Sample A exhibited two distinct Tg values, observed at 18.3 °C and 101.4 °C, respectively (Fig. S2 of the SM [33]). The presence of two transitions indicates heterogeneous segmental mobility, which is consistent with phase-separated or compositionally distinct domains within the copolymer [34]. The lower Tg suggests that a fraction of the material becomes mobile near ambient temperature, which may enhance ductility and toughness, although it could also increase creep and reduce dimensional stability depending on the service temperature and the fraction of this lower-Tg component. The higher Tg notably exceeds that of commercial PET, indicating enhanced thermal rigidity in one phase. The measured Tg of Sample B was 25.1 °C, which is consistent with the predicted Tg value of 33 ± 16 °C. The melting temperature (Tm) was observed in the range of 107.1 °C–171.9 °C, indicating the presence of semi-crystalline domains. This combination of a low Tg and a broad melting region suggests that Sample B may exhibit rubbery behavior at ambient temperature, making it a promising candidate for use as an elastomeric material.

Understanding the relationships between chemical or structural descriptors of designed polymers and their properties provides valuable insights for expanding the design space and further identifying additional novel candidates. Using the Polymer Genome fingerprinting scheme, 283 descriptors, including atomic features, molecular fragments, and extended length-scale characteristics, were generated for 12 100 polymers. SHAP analysis result (Fig. 5) and PCA plots (Fig. S3 of the SM [33]), highlights the features dominating Tg and band gap. These include Chi2n (normalized molecular connectivity index for paths of length two), -CH2- (methylene groups not in a ring), and substructures such as C3-C3-O1. Here, C3 and O1 represents a carbon atom bonded to three neighboring atoms, and an oxygen atom bonded to one neighboring atom.

Descriptors such as Chi2n [35], number of rings in main chain, and C = C double bonds reduce backbone flexibility, contributing to higher Tg, whereas flexible units like -CH2- groups and short side chains lower Tg by enhancing chain mobility. For the band gap, structural motifs with more SP^2-hybridized (three-coordinated) carbons (e.g., C3-C3-O1 and C3-C3-O2) enhances π-electron delocalization, which reduces the band gap. SP^3-rich environments (e.g., C4-C3-O2 and C4-C4-C4) disrupt conjugation and results in a larger band gap. Fewer aromatic rings composed mainly 3-folded carbons correlate with wider band gap.

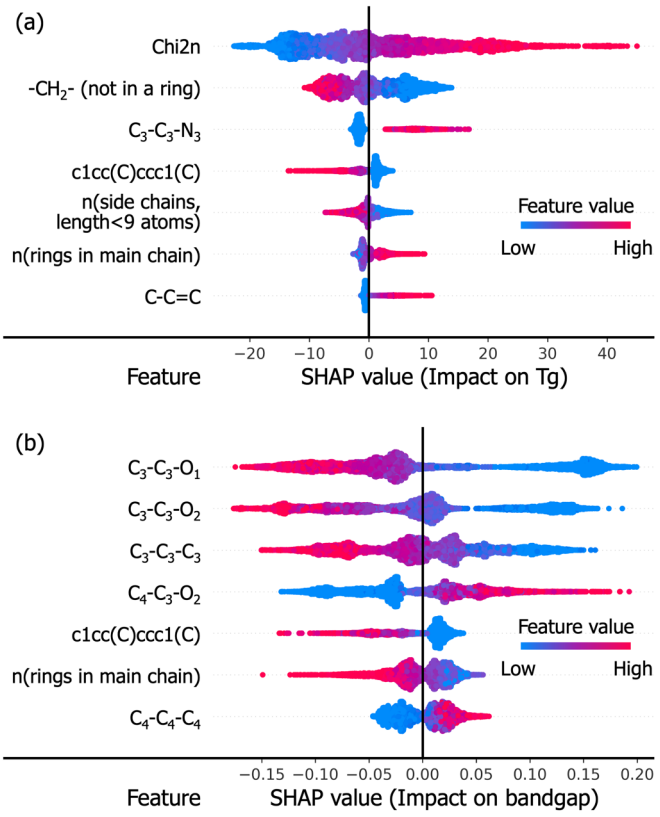

FIG. 5. SHAP analysis showing top 7 most important features influencing (a) Tg and (b) band gap for 12 100 PET-co-(PET-like) copolymers.

IV. CONCLUSIONS

This proof-of-concept study demonstrates a workflow, generating large number of new polymers inspired by the reaction mechanism of an existing polymer, applying multi-objective screening using ML-predicted properties, and validating synthetic feasibility and performance by lab-test. While commercial PET alternatives with excellent thermal properties already exist, the newly designed copolymers offer promising PET replacements. This study considered only a 1:1 compositional ratio of PET and PET-like units of copolymer. Investigation on broader compositional ranges will help better capturing property trends. The presented framework is generalizable and can be applied to the design of replacements for other polymers. Incorporating practical evaluation factors such as mechanical properties, chemical resistance, toxicity and sustainability, as well as raw material cost and production expenses, would make the study more applicable to real-world scenarios and facilitate the development of scalable, commercially viable designs.

ACKNOWLEDGMENTS

This work is financially supported by the Office of Naval Research through a Multidisciplinary University Research Initiative (MURI) grant (N00014-20-1-2586).

DATA AVAILABILITY

The data that support the findings of this article are not publicly available upon publication because it is not technically feasible and/or the cost of preparing, depositing, and hosting the data would be prohibitive within the terms of this research project. The data are available from the authors upon reasonable request.

[1] R. Batra, L. Song, and R. Ramprasad, Emerging materials intelligence ecosystems propelled by machine learning, Nat. Rev. Mater. **6**, 655 (2021).

[2] DuPont, "What is Kevlar," https://www.dupont.com/what-is-kevlar.html.

[3] W. E. Hanford and R. M. Joyce, Polytetrafluoroethylene, J. Am. Chem. Soc. **68**, 2082 (1946).

[4] P. K. Vagholkar, Nylon (Chemistry, Properties and Uses), Int. J. Sci. Res. **5**, 349 (2016).

[5] S. E. Fenton, A. Ducatman, A. Boobis, J. C. DeWitt, C. Lau, C. Ng, J. S. Smith, and S. M. Roberts, Per- and polyfluoroalkyl substance toxicity and human health review: Current state of knowledge and strategies for informing future research, Environ. Toxicol. Chem. **40**, 606 (2021).

[6] US Environmental Protection Agency, "TSCA Chemical Substance Inventory," https://www.epa.gov/tsca-inventory.

[7] European Commission, "REACH Regulation," https://environment.ec.europa.eu/topics/chemicals/reach-regulation_en.

[8] H. Tran, C. Kim, R. Gurnani, O. Hvidsten, J. DeSimpliciis, R. Ramprasad, K. Gadelrab, C. Tuffile, N. Molinari, D. Kitchaev, and M. Kornbluth, Polymer composites informatics for flammability, thermal, mechanical and electrical property predictions, Polym. Chem. **16**, 3459 (2025).

[9] S. Agarwal, A. Mahmood, and R. Ramprasad, Polymer solubility prediction using large language models, ACS Mater. Lett. **7**, 2017 (2025).

[10] C. Kuenneth, J. Lalonde, B. L. Marrone, C. N. Iverson, R. Ramprasad, and G. Pilania, Bioplastic design using multitask deep neural networks, Commun. Mater. **3**, 96 (2022).

[11] C. Kuenneth and R. Ramprasad, polyBERT: A chemical language model to enable fully machine-driven ultrafast polymer informatics, Nat. Commun. **14**, 4099 (2023).

[12] P. Shetty, A. C. Rajan, C. Kuenneth, S. Gupta, L. P. Panchumarti, L. Holm, C. Zhang, and R. Ramprasad, A general-purpose material property data extraction pipeline from large polymer corpora using natural language processing, npj Comput. Mater. **9**, 52 (2023).

[13] R. Gurnani, C. Kuenneth, A. Toland, and R. Ramprasad, Polymer informatics at scale with multitask graph neural networks, Chem. Mater. **35**, 1560 (2023).

[14] H. Tran, C. Kim, L. Chen, A. Chandrasekaran, R. Batra, S. Venkatram, D. Kamal, J. P. Lightstone, R. Gurnani, P. Shetty, M. Ramprasad, J. Laws, M. Shelton, and R. Ramprasad, Machine-learning predictions of polymer properties with Polymer Genome, J. Appl. Phys. **128**, 171104 (2020).

[15] T. D. Huan, A. Mannodi-Kanakkithodi, and R. Ramprasad, Accelerated materials property predictions and design using motif-based fingerprints, Phys. Rev. B **92**, 014106 (2015).

[16] J. Nistane, R. Datta, Y. J. Lee, H. Sahu, S. S. Jang, R. Lively, and R. Ramprasad, Polymer design for solvent separations by integrating simulations, experiments and known physics via machine learning, npj Comput. Mater. **11**, 187 (2025).

[17] J. Kern, Y.-L. Su, W. Gutekunst, and R. Ramprasad, An informatics framework for the design of sustainable, chemically recyclable, synthetically accessible, and durable polymers, npj Comput. Mater. **11**, 182 (2025).

[18] H. Tran, R. Gurnani, C. Kim, G. Pilania, H.-K. Kwon, R. P. Lively, and R. Ramprasad, Design of functional and sustainable polymers assisted by artificial intelligence, Nat. Rev. Mater. **9**, 866 (2024).

[19] R. Gurnani, S. Shukla, D. Kamal, C. Wu, J. Hao, C. Kuenneth, P. Aklujkar, A. Khomane, R. Daniels, A. A. Deshmukh, Y. Cao, G. Sotzing, and R. Ramprasad, MAI-assisted discovery of high-temperature dielectrics for energy storage, Nat. Commun. **15**, 6107 (2024).

[20] H. Tran, K.-H. Shen, S. Shukla, H.-K. Kwon, and R. Ramprasad, Informatics-driven selection of polymers for fuel-cell applications, J. Phys. Chem. C **127**, 977 (2023).

[21] N. A. Tarazona, R. Wei, S. Brott, L. Pfaff, U. T. Bornscheuer, A. Lendlein, and R. Machatschek, Rapid depolymerization of poly(ethylene terephthalate) thin films by a dual-enzyme system and its impact on material properties, Chem Catal. **2**, 3573 (2022).

[22] S. Siddhartha, S. Aarya, K. Dev, S. K. Raghuvanshi, J. B. M. Krishna, and M. A. Wahab, Effect of gamma radiation on the

structural and optical properties of poly(ethylene terephthalate) (PET) polymer, Radiat. Phys. Chem. **81**, 458 (2012).

[23] I. Karacan, An in-depth study of crystallinity, crystallite size and orientation measurements of a selection of poly(ethylene terephthalate) fibers, Fibers Polym. **6**, 186 (2005).

[24] P. Ertl and A. Schuffenhauer, Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions, J. Cheminform. **1**, 8 (2009).

[25] S. M. Lundberg and S.-I. Lee, A unified approach to interpreting model predictions, in *Advances in Neural Information Processing Systems, Vol. 30* (NIPS, San Diego, 2017), pp. 4765–4774.

[26] F. L. Gewers, G. R. Ferreira, H. F. de Arruda, F. N. Silva, C. H. Comin, D. R. Amancio, and L. da F. Costa, Principal component analysis: A natural approach to data exploration, ACM Comput. Surv. **54**, 1 (2021).

[27] Matmerize Inc., "PolymRize™–A standardized software for polymer informatics," https://polymrize.matmerize.com/.

[28] A. Patra, R. Batra, A. Chandrasekaran, C. Kim, H. Tran, and R. Ramprasad, A multi-fidelity information-fusion approach to machine learn and predict polymer band gap, Comput. Mater. Sci. **172**, 109286 (2019).

[29] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (MIT Press, Cambridge, MA, 2006).

[30] S. S. Shukla, C. Kim, R. Gurnani, R. Ramprasad, and A. Mahmood, Stalking the polymer universe using Rxn Chainer (unpublished).

[31] Eastman Chemical Company, Eastar Copolymer G (PETG) Technical Data Sheet, https://productcatalog.eastman.com/tds/ProdDatasheet.aspx?product=71040786.

[32] Eastman Chemical Company, Eastman Tritan™ Copolyester MP200 Technical Data Sheet, https://productcatalog.eastman.com/tds/ProdDatasheet.aspx?product=71103173.

[33] See Supplemental Material at http://link.aps.org/supplemental/10.1103/wjcq-jv1c for experimental characterization data including 1H NMR spectra of synthesized polymers, DSC curves, and principal component analysis (PCA) plots used for pattern analysis.

[34] R. J. Angelo, R. M. Ikeda, and M. L. Wallach, Multiple glass transitions of block polymers, Polymer **6**, 141 (1965).

[35] D. Bonchev and N. Trinajstic, Information theory, distance matrix, and molecular branching, J. Chem. Phys. **67**, 4517 (1977).