Check for updates

GDCh Research Article Angewandte International Edition Chemie www.angewandte.org

How to cite: *Angew. Chem. Int. Ed.* **2025**, *64*, e202411097
doi.org/10.1002/anie.202411097

Polymer Chemistry

Property Prediction of Bio-Derived Block Copolymer Thermoplastic Elastomers Using Graph Kernel Methods

Shannon R. Petersen, David Kohan Marzagão, Georgina L. Gregory,* Yichen Huang, David A. Clifton,* Charlotte K. Williams,* and Clive R. Siviour**

Abstract: Increasing the diversity of bio-based polymers is needed to address the combined problems of plastic pollution and greenhouse gas emissions. The magnitude of the problems necessitates rapid discovery of new materials; however, identification of appropriate chemistries maybe slow using current iterative methods. Machine learning (ML) methods could significantly expedite new material discovery and property identification. Here, PolyAGM, a ML algorithm using graph kernel methods, is introduced and used to predict the properties of block copolymers and identify the responsible structural 'motifs'. It applies a "fingerprinting" method to convert Graph representations of polymers into numerical vectors. The Graphs explicitly encode the entire copolymer of atoms and bonds such that the sequencing of chemical features and polymer chain length are included, alongside relevant stereochemical information. PolyAGM gives predictions for both thermal and mechanical properties that are in good agreement with experimental measurements. This work focuses on predicting the properties of bio-derived ABA-block polymer thermoplastic elastomers, but the general fingerprinting technique of PolyAGM should be relevant to other application fields.

[*] Dr. S. R. Petersen,[+] Dr. G. L. Gregory,[+] Prof. C. K. Williams
Department of Chemistry, University of Oxford, Mansfield Rd, Oxford OX1 3TA, UK
E-mail: georgina.gregory@chem.ox.ac.uk
charlotte.williams@chem.ox.ac.uk

Dr. D. Kohan Marzagão,[+] Prof. D. A. Clifton, Prof. C. R. Siviour
Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, UK
E-mail: david.kohan@chch.ox.ac.uk
clive.siviour@eng.ox.ac.uk
david.clifton@eng.ox.ac.uk

Y. Huang
Department of Computer Science, University of Oxford, 7 Parks Road, Oxford OX1 3QG, UK

© 2024 The Author(s). Angewandte Chemie International Edition published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Introduction

The development of a new polymeric material typically takes 10–20 years.[1] Traditionally, these polymers are discovered by screening large sample sets for specific properties prior to the optimization of a few candidates; the process is effective but time-consuming.[1–2] While combinatorial and high-throughput design methodologies have been used in conjunction with computational simulations to expedite the process, the current design strategy remains inverted, i.e., a material's properties are usually discovered through enumeration rather than explicit design to elicit those features.[1–3] To combat these issues and accelerate development, researchers have turned to machine learning (ML) and polymer informatics.[1–2] Although still in their infancy, such approaches have the potential to rapidly and efficiently explore the vast chemical, structural, and topological polymer design space and uncover relationships between molecular structures and macroscopic properties.[1–2,4]

This is crucial given the majority of currently commercialized polymers are produced from fossil sources of carbon, causing significant greenhouse gas emissions and in need of redesign for recyclability and/or degradability.[5] Therefore, methods to hasten the discovery and development of more sustainable polymers with properties aligned to conventional materials but better end-of-life options are important – polymer informatics could play a role in accelerating solutions.[6] For example, recent reports describe progress in using ML tools to predict (co)polymer thermal properties, dielectric constants, and mechanical properties.[4b,c,e,7] Polymer Genome, for instance, predicts various properties from polymer repeat units represented as SMILES strings.[4a,e] Notably, this is one of the few examples that predict mechanical performances, which is crucial for subsequent applications.[7a,8] Several algorithms report property predictions from structure, though only a small number can identify structures to meet desired specifications.[9] In one example, Mannodi-Kanakkithodi et al. used a genetic algorithm inspired by natural selection processes to identify novel materials with specific band gaps.[4b] More recently, Kuenneth et al. used a multitask deep neural network algorithm to identify new polyhydroxyalkanoates as alternatives to current commodity plastics.[10] Lastly, Batra et al. used variational autoencoders to generate polymer candidates for tolerance under extreme temperatures.[11] Despite these advances, the representation of more complex poly-

 © 2024 The Author(s). Angewandte Chemie International Edition published by Wiley-VCH GmbH

mer structures still poses a challenge as can the availability of sufficient datasets for training ML algorithms.[7d,12]

While the details of these algorithms vary, most polymer informatics development follows a series of distinct steps. First, a database of materials and properties is acquired via experimentation, literature curation, or from existing online databases.[13] Next, each material is encoded as a numerical vector in a process called "fingerprinting".[2a,4a] Fingerprinting transforms polymer structures into numerical vectors with fixed dimensions and typically captures features across the atomic, molecular (i.e., functional groups) and morphological length scales.[2a,4e] In addition to chemical structures or fragments of those structures, fingerprints can also include features such as van der Waals interactions, fraction of rotatable bonds and atoms bonded within cyclic structures, shortest topological distance between rings, and length of side chains, etc.[2a] Selection and inclusion of the necessary features typically require specialist knowledge of particular chemical characteristics at each length scale and predetermination of their significance in the construction of feature vectors.[2a] Regardless of the method of fingerprint selection, these features are used together with the associated material properties to train the algorithm to identify significant features and relationships, which are then utilized to make predictions, allowing optimization of a specific property. Fingerprinting methods for copolymers have been extensively compared by Patel et al., showing how different input methods (one-hot encoding, molecular, and descriptor vectors) affect property predictions.[14] In this work, we aim to add to the fingerprinting set of strategies by connecting it to the field of graph kernel methods.

Here, polymer informatics is applied to predict the properties of block polyester and -carbonate thermoplastic elastomers (TPEs). TPEs represent an important subset of plastics, with a forecast global market production of 5.6 Mt by 2026.[15] They are widely used in transportation, consumer goods, electronics, robotics, healthcare and construction.[5a,16] Typically, TPEs have ABA triblock polymer structures, where A = 'hard' polymer block with a thermal transition above the operating temperature and B = 'soft' polymer block with a thermal transition below room temperature. Microphase separation into A and B domains leads to physically crosslinked materials and hence contributes to mechanical performance. Usually, a minority A-block arranged in spheres or cylinders within a soft B-block phase is associated with elastomeric behavior.[17] As the materials are not covalently cross-linked, product recycling by thermal re-processing is feasible, unlike crosslinked rubbers.[16c,18] Nonetheless, chemical recycling or polymer degradation is still necessary after a certain number of thermal reprocessing cycles. Both are currently challenging with commercial TPEs, which are mostly comprised of hydrocarbon blocks, like polystyrene and -isoprene/-butadiene blocks that are not degradable.[5a,16c] To tackle these challenges, attention has turned to polyester and/or carbonate block polymers since these linkage chemistries facilitate polymer backbone degradation e.g. via transesterification or catalyzed/enzymatic hydrolyses.[19] Pioneering research by Hillmyer and colleagues has resulted in the development of more sustainable TPEs comprising ABA block polyesters, which are prepared using sequential cyclic ester ring-opening polymerization (ROP).[16a,20] Besides thermal reprocessing as an end-of-life option, these TPEs are also hydrolytically degradable due to the ester moieties, and in many cases, the monomers are bio-derived.[5a,21]

Since the first publications, numerous studies have built on this by introducing other polymer chemistries and new bio-based monomers, broadening the accessible material properties and performances. Recently, new types of polyester/carbonate block polymer TPEs were prepared using epoxide ring-opening copolymerization (ROCOP) with anhydrides (polyesters) or CO_2 (polycarbonates).[22] Using catalysts able to switch between heterocycle ROP and epoxide/heteroallene ROCOP allows for the production of many new types of TPEs incorporating lactones, epoxides, anhydrides, CO_2, or cyclic carbonates into a variety of structures and block configurations. In addition to the wide range of monomer combinations, the use of controlled polymerizations to make the TPEs allows for precise tuning of the degree of polymerization (DP) and composition, both of which influence material properties.[22a,23] Indeed, the use of controlled ROP and ROCOP techniques to generate sequence-defined block compositions with narrow molar mass distributions makes them particularly attractive for encoding into ML algorithms.[24] As things currently stand, the synthesis and characterization of all possible block polymers may take lifetimes to complete. Without intervention, it is possible that only a small fraction of possible structures will ever be explored.

This paper introduces Polymer Property Prediction based on the Automatic Generation of Motifs or PolyAGM (Figure 1). Unlike other methods that represent repeating units,[25] or chemical fragments extracted from polymers,[26] PolyAGM encodes the entire copolymer length as a graph. In computer science, graphs are used to represent connections and spatial relationships between objects and consist of "nodes" and "edges". In PolyAGM, entire polymer chains are described by graphs, with atoms encoded as nodes and the bonds between them as edges. This method is applied to block polyester and carbonate TPEs, where each polymer in the dataset has a well-defined block composition and DP. BigSMILES strings are used to transform each high-DP polymer into graphs with hundreds or thousands of nodes, each representing an atom and explicitly encoding how the repeat units are joined. The importance of such explicit accounting for the sequence of chemical features has been highlighted by, for example, Patel et al.[14] Graph kernel methods are then used to automatically extract motifs or fingerprints from graph inputs to form feature vectors that are used to predict polymer thermal and mechanical properties. PolyAGM captures both local structures and patterns, as well as global ones, and allows for the encoding of polymer stereochemistry, which is relevant to prevalent bio-derived poly(lactide) chemistry, amongst others.[20a,b,27] For both the method and some evidence of the impact of succesful capturing of stereochemistry, see Supporting Information Figure S11 and S12.

© 2024 The Author(s). Angewandte Chemie International Edition published by Wiley-VCH GmbH

15213773, 2025, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.202411097 by Test, Wiley Online Library on [11/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

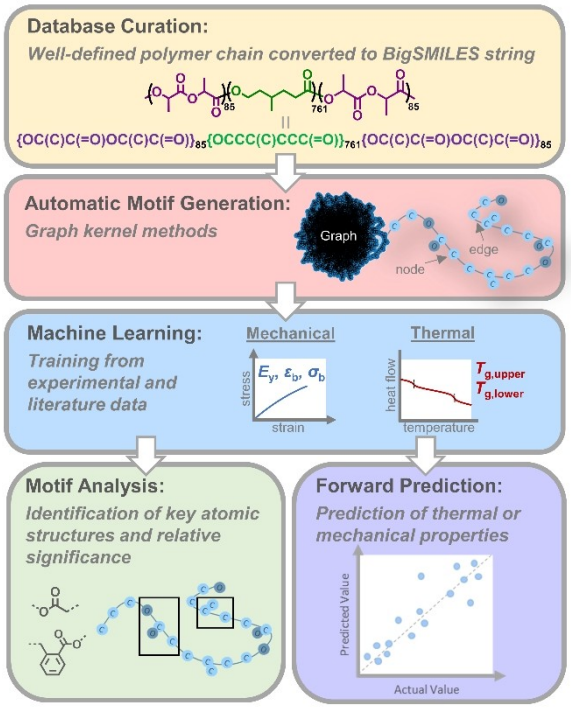

Figure 1. Outline of PolyAGM. Each ABA-block polymer is expressed using BigSMILES strings (shown for poly(lactide-*b*-6mεCL-*b*-lactide) with 20 wt% lactide A-block and overall DP 931; see Figure S9 for BigSMILES stereo-chemistry descriptors). The entire structure is then transformed into a graph, where nodes = atoms and edges = bonds, as well as incorporating features such as stereochemistry in node labels. PolyAGM automatically extracts motifs to yield a feature vector. Training on experimental data allows property predictions and identification of the motifs relevant to each.

Results and Discussion

Block polymer TPEs are ABA-copolymers that have a phase-separated structure (Figures 2 and 3).[18b] The A-block has a high glass transition temperature (T_g) and is considered "hard," while the B-block has a low T_g and is considered "soft." As a result, these materials have two distinct glass transitions with the temperature range between them representing the materials' application window.[18b] This paper focuses on developing more sustainable TPEs made from cyclic ester/carbonate ring-opening polymerization (ROP) or epoxide/heteroallene ring-opening copolymerization (ROCOP) with monomers like lactones, epoxides, anhydrides, cyclic carbonates, and carbon dioxide.[16b,20a,b,22a,b,27a,b,28] These monomers are copolymerized in different block and chain sequences to create oxygenated polymers that degrade through their backbone ester and/or carbonate linkages (Figure 2A).

While recent years have witnessed significant progress in the production of such materials, it is still challenging to produce a sizable database for ML using existing literature-sourced datasets. To broaden the structural types and data set, additional block copolymers were thus synthesized (Figure 2B with new monomers/polymers highlighted in red). These new copolymers were synthesized via bio-based lactone ROP followed by epoxide/anhydride ROCOP (see Supporting Information for experimental details and bio-based monomer sourcing). In some cases, carboxylic acid functional groups were introduced into the hard A-blocks after polymerization using thiol-ene chemistry.[22b,29] In separate work, it was shown that the introduction of carboxylic acid functionalities is a highly effective strategy to significantly moderate both the upper glass transition temperature (A-block, $T_{g\ upper}$) and the materials' stress-at-break (σ_{break}) compared with unfunctionalized analogues.[22b,29] Hence, it was a relevant design strategy to include.

To create the database, each block polymer was expressed in line notation using BigSMILES strings, which were converted into graphs using standard Python libraries (RDkit and PySMILES).[30] The copolymer thermal and mechanical data, including glass transition temperatures ($T_{g\ lower}$, $T_{g\ upper}$), stress-at-break (σ_{break}) and strain-at-break (ε_{break}), together with associated error ranges (where these were reported), from literature sources or experimental measurements were added to the database (see data availability).[16b,20a,b,22a,23,27b,28,31]

As microphase separation is a prerequisite for TPE performance, all the block polymers in the dataset showed phase-separated structures. Experimentally, phase separation is often established by examination of thermal transitions (compared against homopolymers) and, in many cases, using small-angle X-ray scattering (SAXS) experiments (e.g. representative data for these polymers in Figure 3). For each sample, the propensity for phase separation and precise morphology adopted would be determined by the incompatibility of the blocks (expressed as the Flory-Huggins interaction parameter, χ), the overall DP and the fraction of A-block content. Although X is not used by PolyAGM, all the block polymers in the database feature phase-separated block polymers featuring A: B block ratios and overall DPs such that they behave experimentally as elastomers (determined by tensile mechanical measurements). For broader information, the database indicates the particular phase morphology and associated domain spacings, together with χ, for specific polymers and where these were available (Table S2). Most polymers were fully amorphous, so only the lower (B-block) and upper (A-block) T_g values were recorded. A proportion of polymers (~25%) were semicrystalline, and in these cases, crystallisation (T_c) and melting (T_m) temperatures were also incorporated in the database. It is important to note, however, that the PolyAGM method does not require any descriptors beyond the molecular length scale or knowledge of the established theories which govern different length-scale behaviours in order to correlate machine-predicted thermal-mechanical properties with experimental measurements.

Compared to some prior polymer informatics approaches, the resulting dataset composed of 91 different block polymer TPEs is rather small. The small dataset is a consequence of the field size. However, there is literature precedence for ML algorithms to successfully predict outcomes using similarly sized datasets for related tasks.[8c] We

© 2024 The Author(s). Angewandte Chemie International Edition published by Wiley-VCH GmbH

15213773, 2025, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.202411097 by Test, Wiley Online Library on [11/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

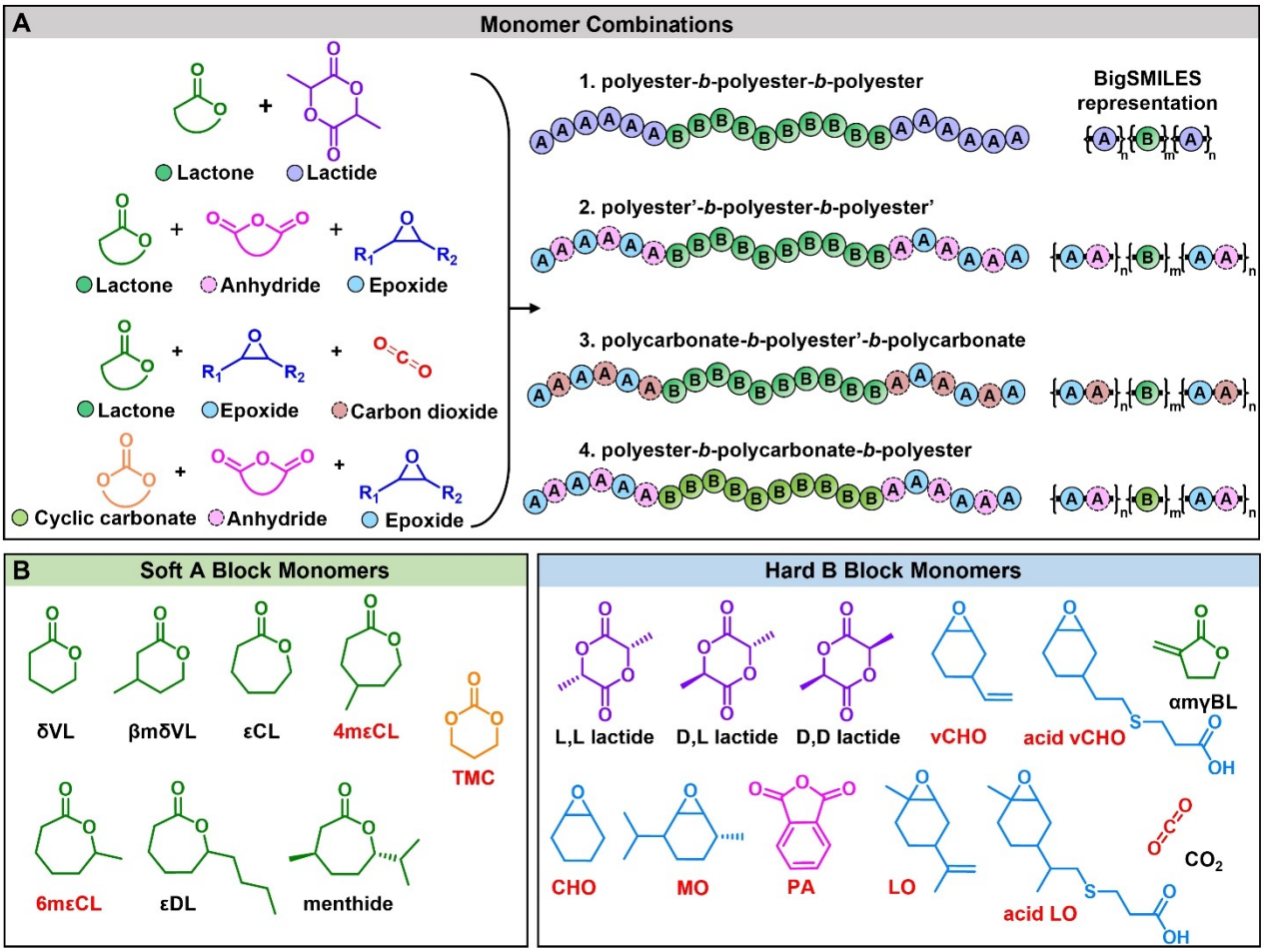

Figure 2. Examples of thermoplastic elastomers in the PolyAGM database. A) Scheme illustrating potential monomer combinations to produce polyester/carbonate ABA block polymers. B) Soft (B-block) and hard (A-block) monomers that are included in the database; in many cases, the monomers are bio-derived. The new monomers and ABA block polymers are labelled in red (Figure S1–S8 and Supporting Information for synthesis and characterisation details). A key to abbreviations and monomer bio-sourcing can be found in Table S1.

also elected to create a database in preference to commercial/open-source databases since the latter typically focus on today's materials, which are predominantly restricted to hydrocarbon polymer backbones. Consequently, the consideration of heteroatom-rich bio-based polymer classes is reportedly beyond the scope of some ML studies.[26] This work specifically sets out to try to accelerate structural identification for polymers featuring degradable ester/carbonate backbones, i.e. oxygenated polymers, and as far as possible, using bio-derived monomers. As such, learning from hydrocarbon polymers may be counterproductive.

In PolyAGM, graph kernel methods are used to infer property predictions.[32] These algorithms are combined with probabilistic regression models to predict values for the polymers' thermal and mechanical properties, i.e. $T_{g\ lower}$, $T_{g\ upper}$, σ_{break}, and ε_{break}. Graph kernel methods compare graphs by learning similarities or differences between given structures. For this series of block polymers, similar graphs might be expected to yield similar polymer properties. To compare the different TPEs, the graph kernel methods identify the frequency of occurrence of various structurally different motifs - recurring patterns or groups of atoms in the polymer structure (Figure 3). Unlike other ML investigations of polymers, PolyAGM does not necessitate that the user predefines these motifs.[1–2] Instead, it explores families of all patterns given by rules based on graph kernel methods. It then identifies which patterns are most significant, with those that are insignificant being automatically discarded.

Here, we show results for when the 'Weisfeiler-Lehman' (WL) graph kernel algorithm is used.[8c] The WL algorithm examines the substructure surrounding a central atom at a distance k from it. For example, in Figure 3, a motif retrieved for $k=2$ for the atom circled in red is illustrated. For each graph, a feature vector is produced such that each dimension (i.e., position in the vector) represents a motif and its frequency of occurrence in the graph gives the value associated with that dimension. The greater the distance k selected, the more motifs there are. In our dataset, the total number of different motifs (dimensions of the feature vectors) for $k=1$, 2, 3, 4, and 5 are 31, 115, 234, 386, and 549, respectively. Of course, not all motifs appear in all polymers, and where motifs are absent, the value associated

© 2024 The Author(s). Angewandte Chemie International Edition published by Wiley-VCH GmbH

15213773, 2025, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.202411097 by Test, Wiley Online Library on [11/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

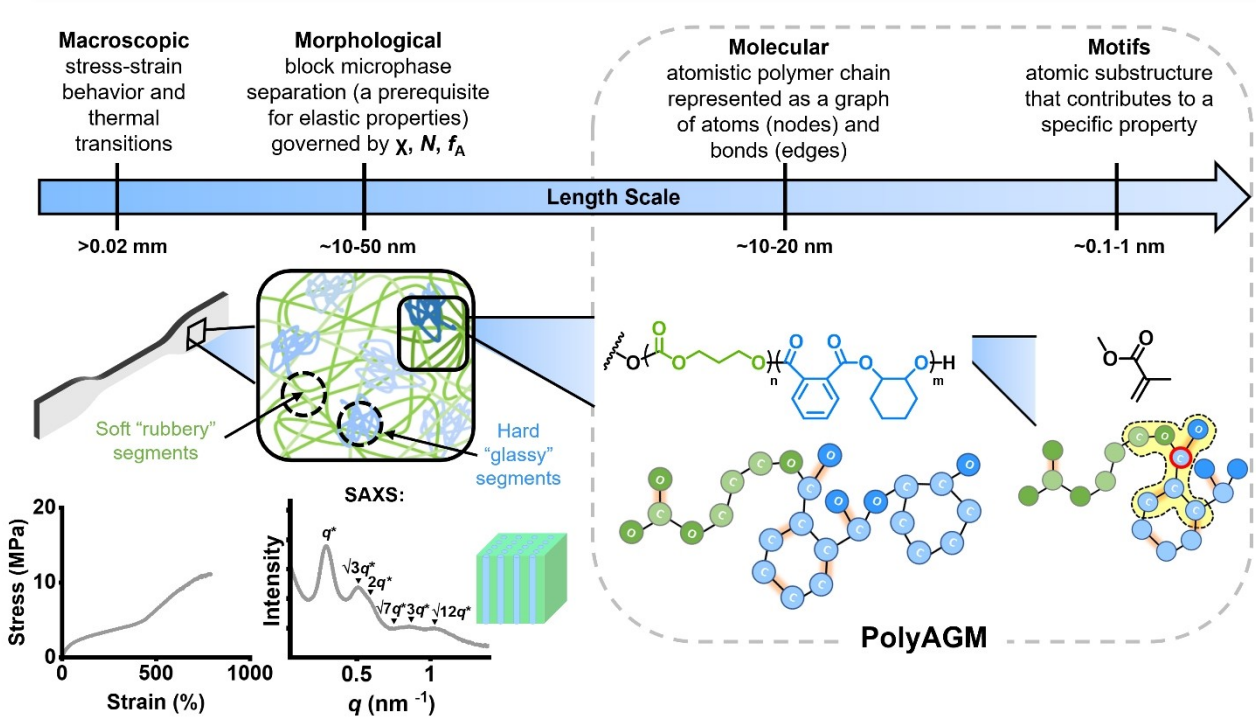

Figure 3. Polymer length scales relevant to conventional understanding of factors influencing thermal-mechanical properties. Polymer science typically applies knowledge of larger length scales (above molecular) to understand macroscopic thermal and mechanical properties. Theories to understand block polymer microphase separation are dominated by the Flory Huggins Interaction Parameter (χ), the volume normalized overall degree of polymerization (N) and A-block volume fraction (f_A). The block polymer phase separated morphology is often confirmed by SAXS experiments. In the diagram, 'green' represents the 'soft' B-block and 'blue' the 'hard' A-block polymer. In this work: PolyAGM only applies descriptors relevant to the molecular length scale (i.e. the sequence of monomers and the relative block compositions). It does not apply information on the specific phase-separated morphology or X and nor does it require the user to define the motifs.

with that motif is zero. After creating feature vectors for all the polymers in the dataset, the values are scaled such that the number of occurrences of each motif is represented within a more manageable range, typically between 0 and 1. For a given polymer and motif, the standardized value is defined by dividing the number of occurrences of a given motif in each polymer by the maximum number of times that motif occurs in all polymers in the dataset. This is standard practice and not negatively affected by the later discovery of additional datasets with even larger numbers of motifs – they simply take a value greater than one. The model allows for both the use of all motifs up to a given distance and the use of motifs at a fixed distance. Finally, any machine learning method that takes feature vectors as inputs can be used. Here, we show results using probabilistic methods, such as Bayesian regression, which is used to learn from the feature vectors and RFE (Recursive Feature Elimination) was employed to discard features which are less significant, i.e., their values have a negligible effect on predictions. Due to the smaller dataset size, a "leave-one-out cross-validation" was used. Each data point was evaluated after training the model with all but that one point, allowing for property predictions for all samples. The quality of the obtained predictions was assessed by RMSE and R^2 values. RMSE is the square root of the average of all squared errors. The R^2 value is reported as a measure of how much of the variance of the predicted variables is captured by the input variables.

Property Prediction

Several ML investigations have successfully predicted the thermal transitions of homopolymers and copolymers.[4a,d,e,7b] However, there are not yet reports of these approaches being applied to phase-separated block polymers containing multiple thermal transitions. The PolyAGM predicted vs experimentally measured thermal transitions for the different TPE blocks, $T_{g\ upper}$ and $T_{g\ lower}$ were compared in parity plots and show close agreement with R^2 values >0.95 (Figures 4A–B). The color of each data point represents the monomer combination used to construct the polymer and excellent fits result for all the different monomer combinations in the database. Note that, experimentally, the thermal transitions can be measured using a range of techniques, and T_g values can shift depending on the specific technique and conditions used to measure them. Techniques include differential scanning calorimetry (DSC), dynamic mechanical analysis (DMA), or oscillatory rheology, all of which show the full transition over a range of temperatures. Values in

© 2024 The Author(s). Angewandte Chemie International Edition published by Wiley-VCH GmbH

15213773, 2025, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.202411097 by Test, Wiley Online Library on [11/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

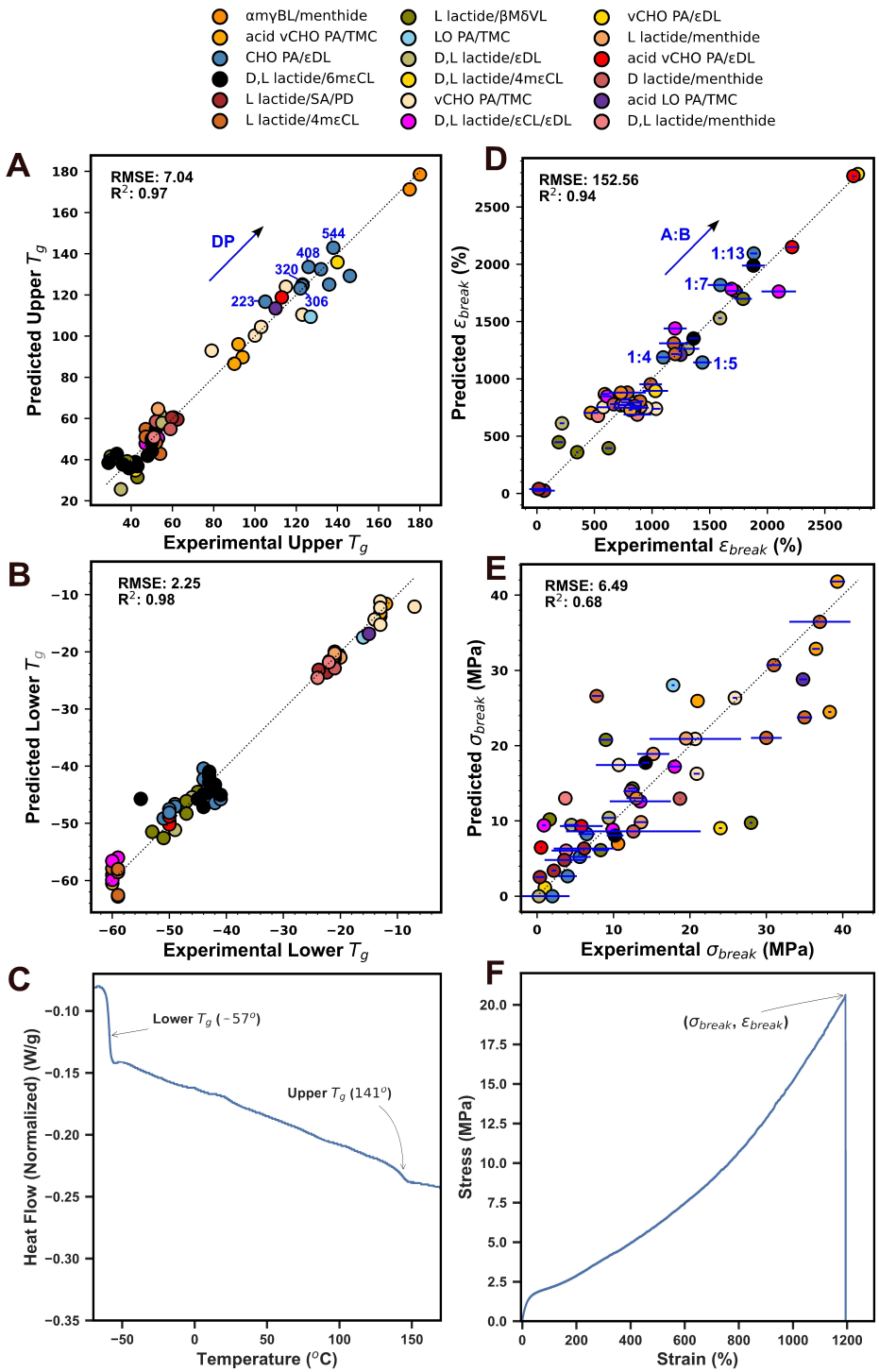

Figure 4. Performance of PolyAGM. Parity plots showing experimental vs predicted values of A) $T_{g\ upper}$ and B) $T_{g\ lower}$. C) Exemplar DSC trace of poly(PA/CHO-*b*-4mεCL-*b*-PA/CHO) showing the experimental upper and lower T_g values that are characteristic of phase-separated structures. Parity plots showing experimental vs predicted values (of unseen data) of D) Elongation-at-break, ε_{break} and E) Stress-at-break, σ_{break}. F) Exemplar stress-strain data showing experimental measurement of ε_{break} and σ_{break} for poly(PA/CHO-*b*-4mεCL-*b*-PA/CHO). All plots correspond to WL-3 with shortest paths and all predictions are of test data. For the molecular structure of each repeat unit (L-lactide, etc) please refer to Figure 2 and Table S1. The dotted lines are parity lines. Inset: Performance of PolyAGM in correlating property trends illustrated through poly(PA/CHO-*b*-εDL-*b*-PA/CHO) datasets[22a] in A) $T_{g\ upper}$ with overall polymer DP (as labelled) and inset D) ε_{break} with varying A- to B-block lengths but same overall DP (see also Figures S9–10).

the dataset were almost invariably recorded by DSC, where the T_g was determined as the midpoint value measured using standard conditions of heating rate and removal of thermal history. Owing to the well-defined, narrow dispersity of the

© 2024 The Author(s). Angewandte Chemie International Edition published by Wiley-VCH GmbH

15213773, 2025, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.202411097 by Test, Wiley Online Library on [11/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

polymers, typical experimental errors spanned a few degrees. The low RMSE values observed for both the lower and upper T_g predictions are consistent with this experimental error and in line with values reported using ML to predict the properties of other polymer types.[4b,c,e,7b] A detailed comparison of PolyAGM with Morgan Fingerprinting, showing the improved results obtained with PolyAGM, is provided in the Supporting Information (Section on Morgan Fingerprints, Table S5–S8 and Figures S14–16).

PolyAGM was also successfully used to predict TPE mechanical properties and parity plots for the predicted and reported values of σ_{break} and ε_{break} also showed good agreements (Figures 4D–E). Experimentally, there are significantly greater errors associated with these types of mechanical measurements and the only consistently reported source is a normal variation between samples (standard deviation). The data used in this study included experimental and literature values reported for films fabricated by both solvent casting and thermal pressing, which were measured using uniaxial tensile testing apparatus. Both characterization and fabrication techniques cause variations in the experimental data that cannot be addressed in the algorithm (see *limitations and future outlook* section). Additionally, other considerations not consistently reported in the literature, including the experimental temperature, humidity and any pre-treatments/storage, may also impact experimental results. With these limitations in mind, the prediction of σ_{break} is very effective in capturing the overall trends in the data as well as for specific families of monomers (Figure 4 inset). In particular, the predicted values alongside those for ε_{break} fell within experimental error, which was of the order of ± 1.2 MPa for σ_{break} and ± 56 % for ε_{break}. The prediction vs experimental data for ε_{break} showed a better fit, with R^2 of 0.94, indicating that the algorithm successfully captures these mechanical properties. Since the measurement of ε_{break} also has the largest numerical span of values, higher RMSE values are expected.

The polymers examined in this work were all synthesized using controlled polymerization methods and so feature narrow dispersity and predictable DPs and block ratios. While PolyAGM may not be best suited to address any differences in DP distributions, the accuracy of the predictions shows the importance of capturing the full DP. To illustrate this point, polymers with identical monomer compositions but varying block or overall DP have different predicted properties, and these are consistent with the experimentally observed and theoretically rationalized/expected trends. For example, PolyAGM correctly matches the experimental trend of increasing $T_{g\ upper}$ observed for a series of block polymers comprised of a similar compositional ratio of PA, CHO and εDL monomers but increasing overall DP (Inset Figure 4A).[22a] Similarly, when the overall DP of ABA polymers made from this monomer set are kept constant, changes in ε_{break} with A:B block length ratio are also correctly captured in the predictions (Figure S9). These experimental trends are attributed to an increase in chain entanglements in the hard block with DP, resulting in high stresses before mechanical failure.[22a] Similar contextualization of the trends in experimental vs predicted results can be made for ε_{break} values (Figure 4D) and other exemplar monomer series in the dataset (Figure S10). Identifying these trends reassures that the PolyAGM approach is suitable.

The dataset contains TPE materials that are strongly influenced by stereochemistry. Poly(lactide), derived from corn starch, can form isotactic A-blocks when prepared from L-lactide monomer or predominantly atactic stereochemistry when racemic mixtures of D- and L-lactide are employed (Figure 4).[16b,20a,b]

One benefit of representing polymers as graphs is the capability to consider this influence. This is achieved by adding a label to nodes (atoms) that are identified as stereocenters (Figure S11). For example, polymers with equivalent B-block chemistry, poly(4mεCL), but only a single L-lactide stereochemistry in the A-block yield TPEs with 25 % higher σ_{break} than those consisting of mixtures of lactide stereochemistry (Figure S12). This difference arises from crystallization in the poly(L-lactide) blocks acting to reinforce the hard blocks and requiring high stresses to break. When a mixture of stereocenters is present in poly(*rac*-lactide), this crystallization behavior is prevented.[20b] Hence, encoding stereochemical information is essential to ensure accurate results.[16b,20a],[20b,20c,26–27,28b] In the past, ML in polymer chemistry has struggled to define stereochemistry or tacticity, and often, these high-level features are ignored.[4e,33]

Motif Analysis

ML is a valuable tool for structure-property optimization and so far guided experimental work and has tended to focus on the prediction of homopolymer thermal transitions, dielectric constants, band gaps, and gas permeability; in all these cases, the monomer chemistry (i.e. repeat unit structure) primarily influences the properties.[2a] That is to say, the fingerprinting techniques employed typically do not consider the higher-order descriptors of stereochemistry and DP.[2a,34] These descriptors were important for the TPE property predictions above, and further structure-property explanations can be provided based on the motif analysis. In PolyAGM, motifs are automatically generated without additional user input from graphs that consider both the chemical structures of polymer repeating units, stereochemistry and DP. In this way, PolyAGM motifs are not the same as repeat unit structures. They allow contributions from pendant functionalities and chain ends to be isolated from the backbone chemistry, and they capture patterns related to junctions between A- and B-blocks and neighbouring monomers in random copolymer blocks. A worked example is provided below (Figure 5).

It is important to note that PolyAGM encodes only molecular information and does not rely on knowledge of complex morphological behaviours.[35] As a worked example then, it was interesting to consider block polymer TPE structures where non-covalent interactions were also at play. In this regard, the block polymers featuring carboxylic acid

© 2024 The Author(s). Angewandte Chemie International Edition published by Wiley-VCH GmbH

15213773, 2025, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.202411097 by Test, Wiley Online Library on [11/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

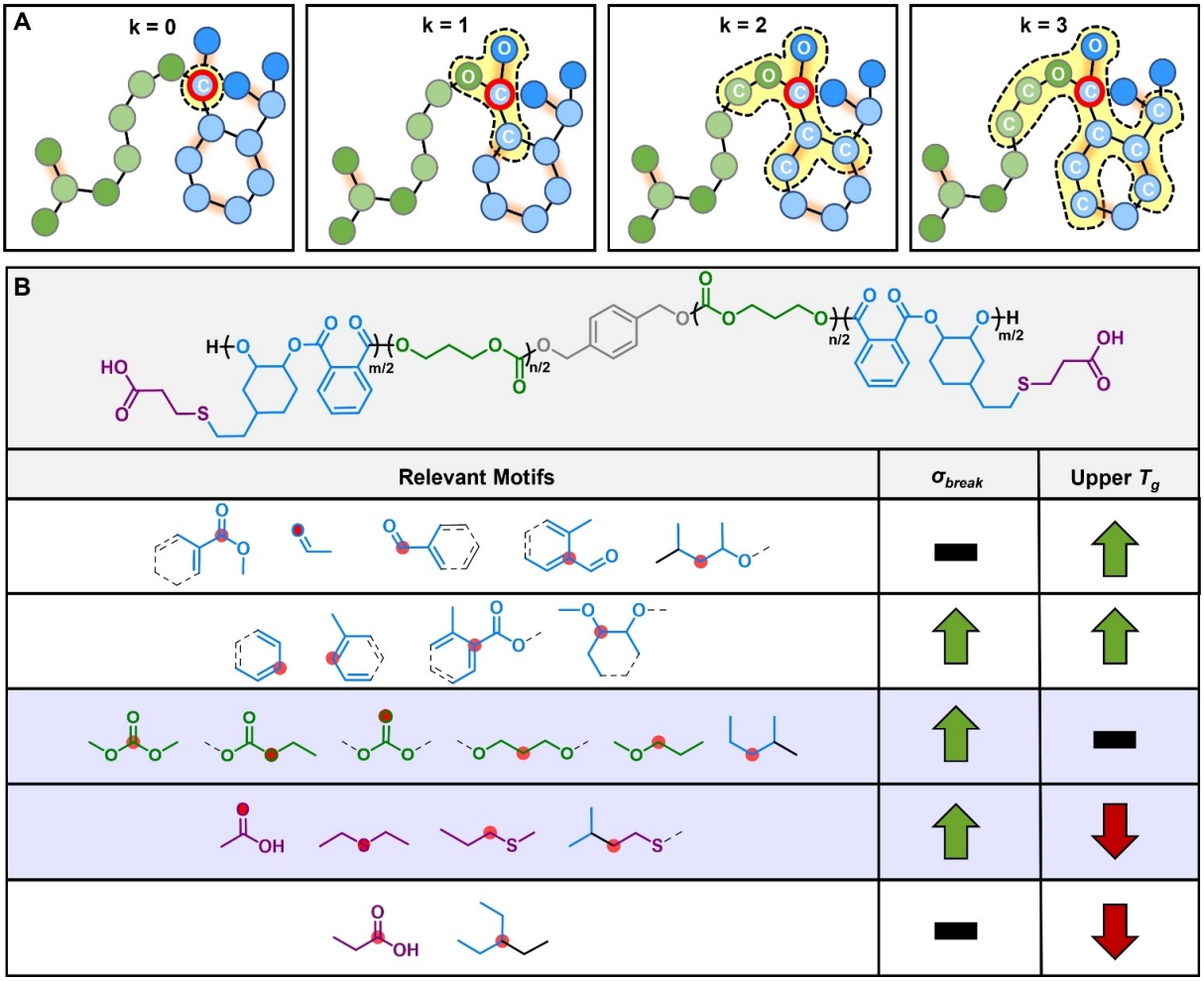

Figure 5. **Motif generation and analysis**. A) Illustration of how 'motifs' are generated in PolyAGM at different distances (excluding hydrogens) from the central atom (outlined in red). Double bonds are highlighted in orange. B) Proof of concept of the motif analysis conducted by PolyAGM using the triblock polymer featuring carboxylic acid functionalities in the ester hard blocks. The motifs are all generated at $k=2$, and the central atom is indicated by a red dot. The influences on $T_{g\ upper}$ and σ_{break} predictions are illustrated with green upward arrows representing an increase in value, red downward arrows representing a decrease in value, and black dashes representing a negligible effect. The set of motifs corresponding to the PTMC block and the carboxylic acid functionality are highlighted in purple.

substituents are significant since these functional groups are known to undergo hydrogen-bonding interactions, which moderate mechanical properties.[22b,29]

A carboxylic acid functionalized poly(ester-*b*-carbonate-*b*-ester) was selected for the motif analysis since it contains such non-covalent interchain features that contribute to its material properties (Figure 5B,, see also Supporting Information section *Motif Analysis* Figure S13, Table S3–S4).[22b] Furthermore, it was established in earlier experimental work that the poly(trimethylene carbonate), P(TMC), midblock undergoes strain-induced crystallization (SIC) at high strain, resulting in an increased σ_{break} compared to other TPEs in the dataset (which are not reported to undergo SIC in this B-block).[22b] This alignment of chains, resulting in crystallization, only occurs above a sufficiently high DP, and neither this behaviour nor the H-bonding between carboxylic acid substituents in the outer polyester hard phases is directly encoded by PolyAGM. Thus, the material was investigated further to understand which motifs were most influential. The polymer structures are well-defined, so the carboxylic acid functional groups are attached to every ring-opened epoxide in the polyester hard blocks. It was revealed experimentally that the installation of the carboxylic acids results in a slight reduction to $T_{g\ upper}$ and an increase to σ_{break} compared to the same polymer before functionalization.[22b] The $T_{g\ upper}$ reduction occurs because the reaction to install the carboxylic acid also introduces a short alkyl chain, which increases segmental motion. Since the carboxylic acid is only present in the polyester hard blocks, its addition has a negligible impact on the $T_{g\ lower}$. The increase in σ_{break} arises from H-bonding between chains, reducing mechanical failure by chain pull-out from the hard blocks compared to the unfunctionalized counterpart.[22b,29]

The important motifs were determined in terms of both the influence on $T_{g\ upper}$ and σ_{break} (Figure 5B). The relative significance of each motif was determined using LIME,[4a,e]

© 2024 The Author(s). Angewandte Chemie International Edition published by Wiley-VCH GmbH

15213773, 2025, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.202411097 by Test, Wiley Online Library on [11/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

which assesses a motif's impact by comparing how the prediction changes when that motif is absent.[33] Such algorithms are also referred to as 'explainers', given their power to capture how a particular motif affects a given prediction. The magnitude of the increase or decrease in a property cannot be accurately attributed to a specific motif; rather, motifs are identified as contributing to a change in properties, which allows for a qualitative analysis of how different atomic structures influence material properties. Naturally, whether the predicted value increases or decreases depends on how the algorithm was trained, which reflects the information available in the dataset. For example, the dataset used in this work provides the ML with comparative polymers with and without carboxylic acid functionality. Therefore, when a motif from the pendant carboxylic acid chain is removed from the polymer, the prediction of the $T_{g\ upper}$ value should increase and its σ_{break} value decrease.

The motif analysis correctly identified the correlation between (hetero)cyclic structures in the polymer hard blocks and higher values for the $T_{g\ upper}$ (Figure 5B). This finding is consistent with long-standing polymer theory since heterocyclic structures reduce the chain's rotational freedom. The analysis also correctly identified motifs corresponding to the carboxylic acid as responsible for reduction to the $T_{g\ upper}$ and increasing the σ_{break}. Nonetheless, PolyAGM incorrectly proposed several sulfur-containing motifs as contributing to the increased σ_{break}. This likely arises because of the thiol-ene reaction used to install the carboxylic acid substituents, which means that sulfur substituents always accompany carboxylic acids. Since every acid-functionalized polymer in the training set was prepared using thiol-ene reactions, PolyAGM has no way of separating or interpreting different effects from the substituents. This limitation could be resolved with additional training, either by using a data set containing other functional groups added via thiol-ene reactions or by adding carboxylic acid groups to the polymer backbone by other synthetic methods (oxidations or protecting group chemistry using functionalized monomers). The motif analysis also correctly identified motifs within the PTMC block as contributing to the increased σ_{break}, which is consistent with the experimentally observed SIC.[22b] Importantly, it correctly identified that PTMC motifs were not contributing to the decreased $T_{g\ upper}$ value.

Polymer and PolyAGM Comparisons

The motivation behind this work was to accelerate the development of bio-derived, recyclable and degradable alternatives to commodity plastics. By comparing the predicted TPE σ_{break} and ε_{break} values to specifications for commercial elastomers, it was observed that many of these property spaces can be matched (Figure 6). For example, styrenic block copolymers (SBS, SIS and SEBS) are the largest TPE market by volume. Block copolymers composed of polymenthide and polylactide derived from cumin/thyme and corn starch, respectively, match the requirements of SBS. Those based on decalactone (εDL), phthalic anhydride

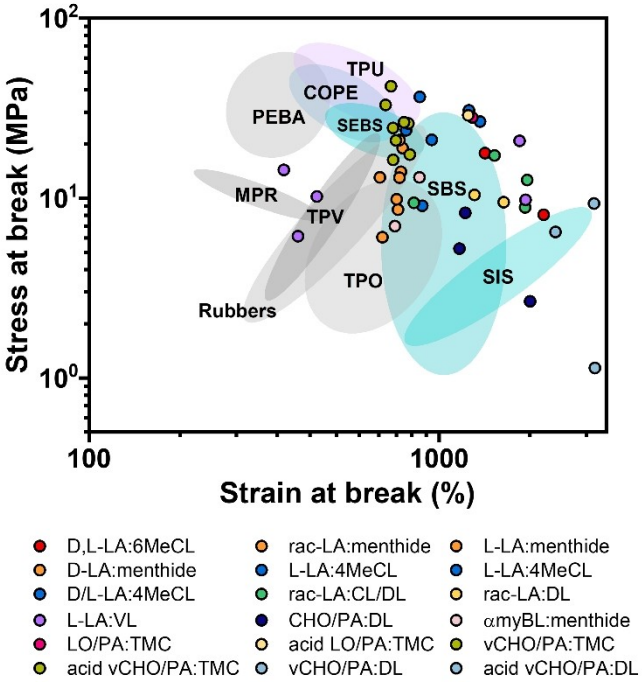

Figure 6. Ashby Plot of σ_{break} vs ε_{break} for commercial elastomers and PolyAGM predicted TPE properties. Commercial elastomer property spaces encompass multiple data points based on literature and commercial datasheets. Polymers synthesised as part of this work are marked in red in Figure 2 above, where the chemical structure of repeating units can also be found. Styrenic block copolymers are poly(styrene-*b*-butadiene-*b*-styrene) (SBS), poly(styrene-b-isoprene-b-styrene) (SIS) and poly(styrene-b-ethylene-butylene-b-styrene) (SEBS); TPO = Thermoplastic Polyolefins; TPV = Thermoplastic Vulcanisates; MPR = Melt Processable Rubber; PEBA = Thermoplastic Polyether Block Amides; COPE = Thermoplastic Copolyesters; TPU = Thermoplastic Polyurethanes.

(PA) and cyclohexene oxide (CHO) derivable from castor oil, corn stover and triglycerides fall within the space for SIS elastomers. Although this could be identified from the experimental data, PolyAGM motif analysis allows the relevant chemical features important to these property spaces to be considered. For example, polystyrene and phthalic anhydride-derived polyesters impart aromatic ring structures to the polymer backbone, and alongside polylactide, they all exhibit high T_g values. Compared to styrenics, these copolyesters can be degraded and/or chemically recycled. Though obvious to the trained polymer scientist, the automatic generation of and learning from these motifs, without expert input knowledge, forms the basis of the potential for non-specialists to accelerate the discovery of structure-property correlations and materials development.

To evaluate PolyAGM's potential compared to other ML methods, we compared it with Morgan Fingerprints and, specifically, the model described by Rogers and Hahn and implemented by RDKit.[36] We apply the same ML algorithms (identical parameters) to feature vectors derived by both approaches. One might also consider Graph Neural Networks.[37] Although Neural Networks normally require significantly more data, there have been advances in

© 2024 The Author(s). Angewandte Chemie International Edition published by Wiley-VCH GmbH

15213773, 2025, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.202411097 by Test, Wiley Online Library on [11/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

techniques for datasets of around hundreds of data points.[38] Most crucially, however, PolyAGM allows a more straightforward motif analysis using explainability methods than Graph Neural Networks. The consequence is that PolyAGM produces results that can be scrutinised without prior chemical and material understanding of TPEs. This is important considering the still limited data on bio-derived, recyclable, and degradable polymers and the need to act quickly to address the plastic waste problem.

Experimentally, polymer chain length (DP) is essential to understanding block polymer thermal and mechanical properties and must be accounted for in any machine-learning method. Indeed, Patel et al.[14] recently reviewed a range of polymer ML methods and emphasised the importance of developing methods to accurately encode for polymer DP, particularly for more complex structures, such as block polymers where it is also essential to ensure the correct monomer ratios. One detraction for Morgan Fingerprints is that they ignore polymer descriptors such as stereochemistry and DP. Whilst Tao et al.[34a] were able to predict high and low T_g values from chemical structures using Morgan Fingerprints; this was only for homopolymers. Overall, then PolyAGM addresses challenges such as accounting for polymer DP, being applicable to copolymers with variable block fractions, including stereochemical descriptors, and automated fingerprint generation. As mentioned, most current polymer informatics is restricted to homopolymers, and prior efforts using copolymers have struggled to account for their complexity.[2a]

Another challenge is the creation of universal predictions and ML tools capable of operating across multiple classes of materials. Feature vectors typically include descriptors across the length scales relevant to a specific property. Thus, these features must be modified if a different property is investigated, limiting universal applicability.[1–2] PolyAGM allows any polymer or copolymer to be represented as long as it features a set of chemical repeat unit structures and has a defined DP. Hence, it should be applicable to more complex architectures, such as stars, branched, or brush polymers, that are also known to significantly impact mechanical properties. Importantly, PolyAGM obviates the need for specialist chemical knowledge input about these architectures to predict properties and allows predictions free from user restriction or bias.

While more work is necessary to convert PolyAGM into a user-based platform, the simple methodology, accurate predictions, and significant facility to adapt it to other structures or properties appear significant. For comparison, better TPE property predictions were obtained using PolyAGM than using the same dataset with a method derived from the original Morgan Fingerprints (see SI).[36] In this investigation, the WL graph kernel framework was applied, but in the future, PolyAGM can work using other graph kernel methods. An interesting avenue will be to develop a comprehensive analysis of how such methods can be combined (e.g. ensemble methods) and identify how best to compare method performances. For example, there were improved results by including alternative graph kernel methods in the prediction of σ_{break}, suggesting that combining different kernel methods may lead to improved results (for a more detailed discussion, see Methods).

Limitations and Future Outlook

Focused on the development of ML tools to aid sustainable TPE discovery, a focused dataset comprising exclusively phase-separated triblock polyesters and polycarbonates (ABA-type), prioritizing bio-derived monomer enchainment, was constructed from existing literature and new experimental data. Despite the limited data size, PolyAGM was able to generate accurate predictions of thermal and mechanical properties and is expected to be more broadly applicable to other classes of oxygenated polymers. However, more experimental data would improve the scope and range of its predictions. We note that there are some systematic sources of error because PolyAGM does not account for the dispersity of the samples which impacts the precision in presentation of DP values. However, the use of our curated database is important since the samples dispersity values are generally narrow, indeed the average dispersity is 1.31 across the 91 samples. Although PolyAGM is successful in accounting for the stereochemistry in each repeat unit (where relevant) and this feature does impact the thermal and mechanical properties for those polymers, we cannot guarantee that the experimental methods used to make isotactic polymer blocks occurred without any epimerization. In particular, further development of PolyAGM would aim to generate new bio-based copolymer candidates for targeted properties (forward predictions), possibly aided by robotics automation.[11,39] Future investigations using PolyAGM will target more complex questions relating to the structures and properties of block polymers, e.g. predicting DPs and block ratios required for phase-separation, which phase-separated morphologies should be accessed for different block polymer compositions or predicting χ for new block polymer structures. Sequence distribution of copolymers is also of key importance and should be further explored with these chemistries.[7c,24c,25] A particular challenge apparent from the outset in predicting TPE tensile mechanical properties and, mentioned earlier, are the influences of different sample processing, characterization and pre-treatment protocols. For example, in processing the polymers into samples for mechanical testing, a range of methods, including solvent-cast films or hot-pressed (compression molded) films, were applied. In characterization, either DMA or tensiometer methods were used, following ISO protocols, but with variable numbers of repeat experiments, strain rates and somewhat subjective reporting of stress and strain-at-break. PolyAGM does not explicitly account for these factors; one consequence is that the experimental mechanical properties typically show larger errors than the thermal transitions.

The future expansion of PolyAGM and similar algorithms will heavily rely on community engagement and robust, consistent data reporting. For this reason, the database used in this study is openly available on GitHub for others to expand and add to.[40] We envisage the current

© 2024 The Author(s). Angewandte Chemie International Edition published by Wiley-VCH GmbH

15213773, 2025, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.202411097 by Test, Wiley Online Library on [11/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

database as a basis for future investigations to add to and feed into developing ML algorithms. Sustainability is a priority for plastics, and it may be decades before a sufficient volume of data is experimentally generated. Thus, working with the available data allows for the acceleration of research in this area. The data addition and curation allow for the deposition of results that may not exhibit desired properties and thus may never be published. It also provides a route to input new materials and properties where online databases are not currently available. Many new polymers are not featured in such repositories, and we do not have time to wait for bio-sourced polymer databases to be generated on a comparable scale. Therefore, others are invited to populate, use, and improve the database provided in this work.

Conclusion

A new property predictive machine learning tool, PolyAGM, an algorithm that predicts properties based only on molecular-level motifs generated by representing polymers as graphs and applying graph kernel methods, is reported. It was applied to predict the thermal and mechanical properties of ABA block polymer thermoplastic elastomers featuring -ester and/or -carbonate linkages. Where possible, the monomers used were selected to have viable bio-based routes to production. The polymers all feature physically cross-linked structures, making them more amenable to recycling through reprocessing and, through the ester/carbonate linkages, to chemical/bio-chemical degradations. PolyAGM showed a strong ability to predict glass transition thermal properties, namely the $T_{g\ upper}$ and $T_{g\ lower}$ of the respective blocks in block copolymers, as well as tensile mechanical properties like strain and stress-at-break. It was also used to identify 'key' chemical features correlating with changes to the thermal and mechanical properties. PolyAGM is openly available and designed to be easily modified in future to examine other bio-derived, degradable polymers and to test for new properties.

Supporting Information

Experimental details and characterization of new polymers synthesized in this work are provided in the SI. The database used in this study is available directly from https://github.com/davidkmarzagao/polyAGM/blob/main/database.xlsx. All Python code (including the database) required to replicate the research is also available on GitHub: https://github.com/davidkmarzagao/polyAGM.

Acknowledgements

The Engineering and Physical Sciences Research Council (EP/S018603/1; EP/V003321/1, EP/Z532782/1), Research England (RED, RE-P-2020-04) and the Faraday Institution (FIRG-007, FIRG-056) are acknowledged for research funding. DAC was supported by an NIHR Research Professorship, an RAEng Research Chair, the InnoHK Hong Kong Centre for Cerebrocardiovascular Health Engineering (COCHE), and the Pandemic Sciences Institute at the University of Oxford.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are openly available in github at https://github.com/davidkmarzagao/polyAGM, reference number 40.

Keywords: polymers · machine learning · property prediction · bio-derived · graph kernel

[1] D. J. Audus, J. J. de Pablo, *ACS Macro Lett.* **2017**, *6*, 1078–1082.
[2] a) L. Chen, G. Pilania, R. Batra, T. D. Huan, C. Kim, C. Kuenneth, R. Ramprasad, *Mater. Sci. Eng. R Rep.* **2021**, *144*, 100595; b) R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, C. Kim, *Npj Comput. Mater.* **2017**, *3*, 54.
[3] a) R. Hoogenboom, M. A. R. Meier, U. S. Schubert, *Macromol. Rapid Commun.* **2003**, *24*, 15–32; b) M. A. R. Meier, R. Hoogenboom, U. S. Schubert, *Macromol. Rapid Commun.* **2004**, *25*, 21–33; c) S. Baudis, M. Behl, *Macromol. Rapid Commun.* **2022**, *43*, 2100400; d) A. M. Mroz, V. Posligua, A. Tarzia, E. H. Wolpert, K. E. Jelfs, *J. Am. Chem. Soc.* **2022**, *144*, 18730–18743; e) K. E. Jelfs, *Ann. N. Y. Acad. Sci.* **2022**, *1518*, 106–119.
[4] a) C. Kim, A. Chandrasekaran, T. D. Huan, D. Das, R. Ramprasad, *J. Phys. Chem. C* **2018**, *122*, 17575–17585; b) A. Mannodi-Kanakkithodi, G. Pilania, T. D. Huan, T. Lookman, R. Ramprasad, *Sci. Rep.* **2016**, *6*, 20952; c) L. Chen, C. Kim, R. Batra, J. P. Lightstone, C. Wu, Z. Li, A. A. Deshmukh, Y. Wang, H. D. Tran, P. Vashishta, G. A. Sotzing, Y. Cao, R. Ramprasad, *Npj Comput. Mater.* **2020**, *6*, 61; d) C. Kuenneth, A. C. Rajan, H. Tran, L. Chen, C. Kim, R. Ramprasad, *Patterns* **2021**, *2*, 100238; e) H. Doan 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, R. Ramprasad, *J. Appl. Phys.* **2020**, *128*, 171104; f) K. Hara, S. Yamada, A. Kurotani, E. Chikayama, J. Kikuchi, *Sci. Rep.* **2022**, *12*, 10558; g) J. A. Pugar, C. Gang, C. Huang, K. W. Haider, N. R. Washburn, *ACS Appl. Mater. Inter.* **2022**, *14*, 16568–16581.
[5] a) Y. Zhu, C. Romain, C. K. Williams, *Nature* **2016**, *540*, 354–362; b) F. Vidal, E. R. van der Marel, R. W. F. Kerr, C. McElroy, N. Schroeder, C. Mitchell, G. Rosetto, T. T. D. Chen, R. M. Bailey, C. Hepburn, C. Redgwell, C. K. Williams, *Nature* **2024**, *626*, 45–57.
[6] a) A. N. Wilson, P. C. St John, D. H. Marin, C. B. Hoyt, E. G. Rognerud, M. R. Nimlos, R. M. Cywar, N. A. Rorrer, K. M. Shebek, L. J. Broadbelt, G. T. Beckham, M. F. Crowley, *Macromolecules* **2023**, *56*, 8547–8557; b) K. A. Fransen, S. H. M. Av-Ron, T. R. Buchanan, D. J. Walsh, D. T. Rota, L. Van Note, B. D. Olsen, *Proc. Natl. Acad. Sci.* **2023**, *120*, e2220021120; c) N. H. Park, D. Y. Zubarev, J. L. Hedrick, V.

© 2024 The Author(s). Angewandte Chemie International Edition published by Wiley-VCH GmbH

15213773, 2025, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.202411097 by Test, Wiley Online Library on [11/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Kiyek, C. Corbet, S. Lottier, *Macromolecules* **2020**, *53*, 10847–10854.
[7] a) B. Rajabifar, G. F. Meyers, R. Wagner, A. Raman, *Macromolecules* **2022**, *55*, 8731–8740; b) C. Kuenneth, W. Schertzer, R. Ramprasad, *Macromolecules* **2021**, *54*, 5957–5961; c) L. Tao, J. Byrnes, V. Varshney, Y. Li, *iScience* **2022**, *25*; d) T. B. Martin, D. J. Audus, *ACS Polym. Au* **2023**, *3*, 239–258; e) C. Yan, G. Li, *Adv. Intell. Sys.* **2023**, *5*, 2200243; f) N. Andraju, G. W. Curtzwiler, Y. Ji, E. Kozliak, P. Ranganathan, *ACS Appl. Mater. Inter.* **2022**, *14*, 42771–42790.
[8] a) E. Kazemi-Khasragh, J. P. Fernández Blázquez, D. Garoz Gómez, C. González, M. Haranczyk, *Int. J. Solids Struct.* **2024**, *286–287*, 112547; b) A. S. Fard, J. Moebus, G. Rodriguez, *J. Adv. Manuf. Process.* **2023**, *5*, e10156; c) A. Menon, J. A. Thompson-Colón, N. R. Washburn, *Front. Mater.* **2019**, *6*; d) S. Park, K. P. Marimuthu, G. Han, H. Lee, *Int. J. Mech. Sci.* **2023**, *246*, 108162.
[9] G. Pilania, *Comput. Mater. Sci.* **2021**, *193*, 110360.
[10] C. Kuenneth, J. Lalonde, B. L. Marrone, C. N. Iverson, R. Ramprasad, G. Pilania, *Commun. Mater.* **2022**, *3*, 96.
[11] R. Batra, H. Dai, T. D. Huan, L. Chen, C. Kim, W. R. Gutekunst, L. Song, R. Ramprasad, *Chem. Mater.* **2020**, *32*, 10489–10500.
[12] T. K. Patra, *ACS Polym. Au* **2022**, *2*, 8–26.
[13] a) S. Otsuka, I. Kuwajima, J. Hosoya, Y. Xu, M. Yamazaki, in *2011 International Conference on Emerging Intelligent Data and Web Technologies*, **2011**, pp. 22–29; b) in *Polymer Property Predictor and Database*, https://pppdb.uchicago.edu/, Centre for Hierarchical materials Design (CHI-MaD), **2024**; c) N. J. Rebello; d) **2024**.
[14] R. A. Patel, C. H. Borca, M. A. Webb, *Mol. Syst. Des. Eng.* **2022**, *7*, 661–676.
[15] P. Ellis, *The Future of Thermoplastic Elastomers to 2026*, **2021**, Smithers Market Report (accessed 10/06/2024).
[16] a) C. Creton, *Macromolecules* **2017**, *50*, 8297–8316; b) D. K. Schneiderman, E. M. Hill, M. T. Martello, M. A. Hillmyer, *Polym. Chem.* **2015**, *6*, 3641–3651; c) D. K. Schneiderman, M. A. Hillmyer, *Macromolecules* **2017**, *50*, 3733–3749.
[17] a) P. Maji, K. Naskar, *J. Appl. Polym. Sci.* **2022**, *139*, e52942; b) W. Wang, W. Lu, A. Goodwin, H. Wang, P. Yin, N.-G. Kang, K. Hong, J. W. Mays, *Prog. Polym. Sci.* **2019**, *95*, 1–31; c) G. Holden, in *Rubber Technology* (Ed.: M. Morton), Springer US, Boston, MA, **1987**, pp. 465–481.
[18] a) S. R. Petersen, H. Prydderch, J. C. Worch, C. J. Stubbs, Z. Wang, J. Yu, M. C. Arno, A. V. Dobrynin, M. L. Becker, A. P. Dove, *Angew. Chem.* **2022**, *61*, e202115904; b) R. J. Spontak, N. P. Patel, *Curr. Opin. Colloid Interface Sci.* **2000**, *5*, 333–340.
[19] G. W. Coates, Y. D. Y. L. Getzler, *Nature. Rev. Mater.* **2020**, *5*, 501–516.
[20] a) C. L. Wanamaker, M. J. Bluemle, L. M. Pitet, L. E. O'Leary, W. B. Tolman, M. A. Hillmyer, *Biomacromolecules* **2009**, *10*, 2904–2911; b) A. Watts, N. Kurokawa, M. A. Hillmyer, *Biomacromolecules* **2017**, *18*, 1845–1854; c) D. C. Batiste, M. S. Meyersohn, A. Watts, M. A. Hillmyer, *Macromolecules* **2020**, *53*, 1795–1808.
[21] a) Z. Li, Y. Shen, Z. Li, *Macromolecules* **2024**, *57*, 1919–1940; b) C. M. Plummer, L. Li, Y. Chen, *Macromolecules* **2023**, *56*, 731–750.
[22] a) G. L. Gregory, G. S. Sulley, L. P. Carrodeguas, T. T. D. Chen, A. Santmarti, N. J. Terrill, K.-Y. Lee, C. K. Williams, *Chem. Sci.* **2020**, *11*, 6567–6581; b) G. L. Gregory, G. S. Sulley, J. Kimpel, M. Łagodzińska, L. Häfele, L. P. Carrodeguas, C. K. Williams, *Angew. Chem. Int. Ed.* **2022**, *61*, e202210748; c) T. T. D. Chen, L. P. Carrodeguas, G. S. Sulley, G. L. Gregory, C. K. Williams, *Angew. Chem. Int. Ed.* **2020**, *59*, 23450–23455; d) J. M. Longo, M. J. Sanford, G. W. Coates, *Chem. Rev.* **2016**, *116*, 15167–15197; e) S. Paul, Y. Zhu, C. Romain, R. Brooks, P. K. Saini, C. K. Williams, *Chem. Commun.* **2015**, *51*, 6459–6479.
[23] M. A. Hillmyer, W. B. Tolman, *Acc. Chem. Res.* **2014**, *47*, 2390–2396.
[24] a) A. C. Deacy, G. L. Gregory, G. S. Sulley, T. T. D. Chen, C. K. Williams, *J. Am. Chem. Soc.* **2021**, *143*, 10021–10040; b) S. L. Perry, C. E. Sing, *ACS Macro Lett.* **2020**, *9*, 216–225; c) D. Bhattacharya, D. C. Kleeblatt, A. Statt, W. F. Reinhart, *Soft Matter* **2022**, *18*, 5037–5051.
[25] M. Aldeghi, C. W. Coley, *Chem. Sci.* **2022**, *13*, 10486–10498.
[26] C. Kuenneth, R. Ramprasad, *Nature Commun.* **2023**, *14*, 4099.
[27] a) M. T. Martello, M. A. Hillmyer, *Macromolecules* **2011**, *44*, 8537–8545; b) M. T. Martello, D. K. Schneiderman, M. A. Hillmyer, *ACS Sustain. Chem. Eng.* **2014**, *2*, 2519–2526; c) F. Auriemma, C. De Rosa, M. R. Di Caprio, R. Di Girolamo, W. C. Ellis, G. W. Coates, *Angew. Chem. Int. Ed.* **2015**, *54*, 1215–1218.
[28] a) M. Xiong, D. K. Schneiderman, F. S. Bates, M. A. Hillmyer, K. Zhang, *Proc. Natl. Acad. Sci.* **2014**, *111*, 8357–8362; b) J. Shin, Y. Lee, W. B. Tolman, M. A. Hillmyer, *Biomacromolecules* **2012**, *13*, 3833–3840; c) Y. Huang, R. Chang, L. Han, G. Shan, Y. Bao, P. Pan, *ACS Sustain. Chem. Eng.* **2016**, *4*, 121–128.
[29] G. L. Gregory, C. K. Williams, *Macromolecules* **2022**, *55*, 2290–2299.
[30] T.-S. Lin, C. W. Coley, H. Mochigase, H. K. Beech, W. Wang, Z. Wang, E. Woods, S. L. Craig, J. A. Johnson, J. A. Kalow, K. F. Jensen, B. D. Olsen, *ACS Cent. Sci.* **2019**, *5*, 1523–1531.
[31] a) L. Zahir, T. Kida, R. Tanaka, Y. Nakayama, T. Shiono, N. Kawasaki, N. Yamano, A. Nakayama, *Polym. Degrad. Stab.* **2020**, *181*, 109353; b) C. Fang, X. Wang, X. Chen, Z. Wang, *Polym. Chem.* **2019**, *10*, 3610–3620.
[32] K. Borgwardt, E. Ghisu, F. Llinares-López, L. O'Bray, B. Rieck, *Trends. Mach. Learn.* **2020**, *13*, 531–712.
[33] M. T. Ribeiro, S. Singh, C. Guestrin, in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, **2016**, pp. 1135–1144.
[34] a) L. Tao, G. Chen, Y. Li, *Patterns* **2021**, *2*, 100225; b) E. R. Antoniuk, P. Li, B. Kailkhura, A. M. Hiszpanski, *J. Chem. Inf. Model.* **2022**, *62*, 5435–5445.
[35] a) T. Aoyagi, *Comp. Mater. Sci.* **2021**, *188*, 110224; b) R. Hosoya, H. Morita, *Macromolecules* **2023**, *56*, 6692–6703.
[36] D. Rogers, M. Hahn, *J. Chem. Inf. Model.* **2010**, *50*, 742–754.
[37] C. Chen, W. Ye, Y. Zuo, C. Zheng, S. P. Ong, *Chem. Mater.* **2019**, *31*, 3564–3572.
[38] B. Dou, Z. Zhu, E. Merkurjev, L. Ke, L. Chen, J. Jiang, Y. Zhu, J. Liu, B. Zhang, G.-W. Wei, *Chem. Rev.* **2023**, *123*, 8736–8780.
[39] M. J. Tamasi, R. A. Patel, C. H. Borca, S. Kosuri, H. Mugnier, R. Upadhya, N. S. Murthy, M. A. Webb, A. J. Gormley, *Adv. Mater.* **2022**, *34*, 2201809.
[40] S.R. Petersen, D.Kohan Marzagão, G.L. Gregory, Y. Huang, D.A. Clifton, C.K. Williams, C.R. Siviour, polyAGM: Package for Property Prediction of Polymers based on Automatically Generated Motifs, 2024, GitHub repository, https://github.com/davidkmarzagao/polyAGM

Manuscript received: June 12, 2024
Accepted manuscript online: November 29, 2024
Version of record online: December 11, 2024

© 2024 The Author(s). Angewandte Chemie International Edition published by Wiley-VCH GmbH

15213773, 2025, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.202411097 by Test, Wiley Online Library on [11/02/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License