International Journal of
Molecular Sciences

MDPI

Article

Data-Driven Modeling and Design of Sustainable High Tg Polymers

Qinrui Liu [1], Michael F. Forrester [2], Dhananjay Dileep [2], Aadhi Subbiah [2], Vivek Garg [2], Demetrius Finley [2], Eric W. Cochran [2], George A. Kraus [2] and Scott R. Broderick [1,*]

[1] Department of Materials Design and Innovation, University at Buffalo, Buffalo, NY 14260, USA
[2] Department of Chemical and Biological Engineering, Iowa State University, Ames, IA 50011, USA; djay11@iastate.edu (D.D.); asubbiah@iastate.edu (A.S.); dafinley@iastate.edu (D.F.)
* Correspondence: scottbro@buffalo.edu

Abstract: This paper develops a machine learning methodology for the rapid and robust prediction of the glass transition temperature (Tg) for polymers for the targeted application of sustainable high-temperature polymers. The machine learning framework combines multiple techniques to develop a feature set encompassing all relative aspects of polymer chemistry, to extract and explain correlations between features and Tg, and to develop and apply a high-throughput predictive model. In this work, we identify aspects of the chemistry that most impact Tg, including a parameter related to rotational degrees of freedom and a backbone index based on a steric hindrance parameter. Building on this scientific understanding, models are developed on different types of data to ensure robustness, and experimental validation is obtained through the testing of new polymer chemistry with remarkable Tg. The ability of our model to predict Tg shows that the relevant information is contained within the topological descriptors, while the requirement of non-linear manifold transformation of the data also shows that the relationships are complex and cannot be captured through traditional regression approaches. Building on the scientific understanding obtained from the correlation analyses, coupled with the model performance, it is shown that the rigidity and interaction dynamics of the polymer structure are key to tuning for achieving targeted performance. This work has implications for future rapid optimization of chemistries.

Keywords: sustainable polymers; glass transition temperature; topological descriptors; machine learning; graph theory

check for updates

Academic Editor: Marcel Popa

Received: 13 February 2025
Revised: 11 March 2025
Accepted: 17 March 2025
Published: 18 March 2025

Citation: Liu, Q.; Forrester, M.F.; Dileep, D.; Subbiah, A.; Garg, V.; Finley, D.; Cochran, E.W.; Kraus, G.A.; Broderick, S.R. Data-Driven Modeling and Design of Sustainable High Tg Polymers. *Int. J. Mol. Sci.* **2025**, *26*, 2743. https://doi.org/10.3390/ijms26062743

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In recent decades, petroleum-based plastics, sourced from fossil fuels, have dominated the production of commodity plastics; however, only about 9% of plastic waste has been recycled. The increasing ecological and public health concerns have emphasized the need for developing sustainable polymers [1]. Although the definition of sustainable polymers remains somewhat open to interpretation, it generally encompasses polymers derived from biomass-based monomers, those synthesized using green chemistry principles, polymers capable of closed-loop chemical recycling (including monomer recovery, repurposing to alternative materials, or degradation via biological or simulated environmental processes), and those for which a comprehensive life cycle assessment has been performed [2–4].

Bio-renewable polymers, a subset of sustainable polymers, are materials in which at least part of the polymer is produced from renewable raw materials, such as biomass from sugarcane or corn [5]. These polymers can be derived from biochemical processes involving

plant sources (e.g., cellulose-based polymers, alginate, polyisoprene, starch) and animal sources (e.g., polylactic acid, polyhydroxyalkanoates, polybutylene succinate), as well as bacterial fermentation products (e.g., chitin, chitosan, collagen, sericin) [6,7]. Modifying bio-renewable polymers by incorporating different functional groups can yield a variety of derivatives [8,9]. It is important to note that not all bio-renewable polymers are biodegradable, as a polymer's biodegradability is determined by its chemical structure rather than the source of its raw materials [2,3]. The conversion of biomass-based feedstocks into polymers not only reduces reliance on fossil fuels but also offers numerous opportunities to design bio-renewable polymers with tailored properties and functionalities.

In real-world applications, biopolymers face certain limitations due to their relatively low mechanical strength and thermal stability. The glass transition temperature (Tg) is a critical indicator of a polymer's thermal properties, marking the point where a polymer transitions from a rigid, glassy state to a softer, rubbery state. Polymers with higher Tg values exhibit enhanced thermal stability, allowing them to maintain their mechanical integrity even under elevated temperatures. This makes them ideal for use in high-performance applications such as energy, aerospace, biomedical, and electronics [10,11], where materials might withstand extreme thermal environments.

The majority of commonly used biopolymers have Tg values ranging from 55 °C to 75 °C [2,12,13]. Focusing on the discovery of high Tg polymers is essential to meet the growing demands for materials that can withstand challenging operational conditions without compromising functionality, as well as remaining environmentally friendly. Numerous studies have investigated the structural design of backbones and side groups in the pursuit of polymers with higher Tg. Most of these efforts have focused on synthesizing novel polymers through the manipulation of monomer ratios and crosslinks, introducing additional functional groups [14–17], as well as exploring various polymerization methods to synthesize polymers with tunable Tg values [18]. These prior studies suggest that high Tg polymers often have rigid backbones that limit chain flexibility. A common strategy to increase rigidity involves adding non-flexible ring structures to the backbone, as seen in polymers like aromatic polycarbonates, polynorbornenes, and polyimides. Another approach is introducing substituents near the backbone to hinder chain rotation via steric interactions, exemplified by poly (α-methyl styrene) and similar polymers [19–21].

While these investigations have provided valuable insights for further development, they are often constrained by experimental capacity and may not fully capture the complexity of factors that influence polymer properties. Additionally, many of these case-by-case studies include specific types of chemical groups, as the analyses typically rely on comparisons between polymers with similar main-chain structures [12,22]. To overcome these limitations, quantitative and systematic approaches are needed to evaluate the structure–property relationships of polymer candidates. Our goal is to utilize an unbiased approach that examines all topological features derived from monomer chemical structures, in order to systematically identify the key factors governing Tg.

There have been multiple methods developed to establish an efficient descriptor base, including Bicerano's topological indices [23], Van Krevelen's additive group contributions [24], and Askadskii's semi-empirical method based on Van der Waals volume and dispersion interaction [25]. Other relevant recent approaches include the polymer genome database [26,27], polymer quantum chemistry combined with machine learning [28], predictions based on SMILES string of polymers [29], empirical QSAR data for polymers [30,31], and hybrid descriptors [32]. With the advancement of machine learning (ML), an increasing number of studies have been conducted in the material design field integrating these polymer structure descriptors to accelerate the prediction of polymer properties [33–43]. In this paper, we build upon these works, as well as our additional works in ML [44–48], to

develop a model capturing the underlying physics to accelerate the discovery of high Tg renewable polymers.

2. Results

2.1. Data

In order to develop a descriptor set that encompasses multiple aspects of the structure in a bias-free manner, the components of the molecular structure were calculated (Table 1). The logic of this approach is that the descriptors should be physically meaningful, while being calculable for any structures, thereby allowing for improved mechanistic understanding and robust modeling of new polymer systems. The descriptor set was built based partially on the descriptors defined by Bicerano [23] and Van Krevelen [24]. On the basis of graph theory, Bicerano defined topological connectivity indices and used them to predict more than 70 polymer properties. Building on this logic, with the inclusion of ML analyses, it is expected that the information gain and property prediction can be further expanded and enhanced. The process for descriptor calculation, as well as a more thorough description of the features, is provided in the Supplementary Materials.

Table 1. The feature set developed and calculated for the relevant polymers. These features are then used in the ML framework for Tg modeling.

Polymer Features	How They Were Defined or Calculated
N	total non-hydrogen atoms in one polymer repeat unit
N_C	number of carbon atoms in one polymeric repeat unit
N_H	number of hydrogen atoms in one polymeric repeat unit
N_ester_n	number of backbone -COO- (non-conjugated with aromatic ring)
N_ester_c	number of backbone -COO- (one-sided conjugation with aromatic ring)
N_aromaticring	number of aromatic rings in one polymeric repeat unit
N_CH2	number of -CH_2 in one polymeric repeat unit
N_ether	number of -O- in a polymeric repeat unit
N_backbone_O	number of backbone oxygen atoms in one polymeric repeat unit
N_O	number of oxygen atoms in a polymeric repeat unit
M	mole weight of one polymer repeat unit(g/mol)
N_alkyl_ether	number of ether (R-O-R′) linkages between two units R and R′ both of which are connected to the alkyl carbon atom
N_rot	total number of rotational degrees of freedom parameter (N_rot = the backbone rotational degrees plus the side group rotational degrees)
N_K	N_K = 5N_amide + 7N_cyanide + 15N_carbonate + 5N_Cl + 13N_Br + 4N_hydroxyl − 3N_(ether) − 5N_C = C + 3N_sulfone − 3N_acrylic ester − 5N_ (isolated saturated aliphatic hydrocarbon rings, i.e., cyclohexyl or cyclopentyl)
N_SP	number of atoms in the shortest path across the backbone of a polymeric repeat unit, N_SP ≤ N_BB
Nmv	Nmv = 2 × N_ester + 3 × N_ether
$^0\chi$	the zeroth-order (atomic) connectivity indices (the first atomic index)
$^0\chi^V$	the zeroth-order (atomic) connectivity indices (the second atomic index)
$^1\chi$	the first-order (bond) connectivity indices (the first bond index)
$^1\chi^V$	the first-order (bond) connectivity indices (the second bond index)

Table 1. *Cont.*

Polymer Features	How They Were Defined or Calculated
BB_index1	backbone index1 is a steric hindrance parameter that reflects the flexibility of the polymer backbone structure, similar to the stiffness of the backbone.
BB_index2	backbone index2 is a steric hindrance parameter that differentiates between backbone atoms with the same (δ/δV) values but different δ values, reflecting variations in the number of non-hydrogen neighbors around each backbone atom.

The values for Tg, which are used in the analysis, were taken from two different sources. The first was from assorted literature values, as compiled in the "Polymer Products from Aldrich" catalog [49]. These data were for commonly available polymers, with the data spanning multiple sources. From these data, we calculated the topological features of approximately 80 polymer chemistries. When multiple Tg values (or a range of values) were provided for a given polymer, the average value was taken. This approximation, along with the combination of various sources, introduces uncertainty and error into the analysis; however, the range of approaches used and the consideration of underlying science provide sufficient confidence for this approach. The second set of Tg values was focused on biobased acrylates [50]. This data set provides a smaller data set; however, by being centered on the acrylate systems, it provides a more focused examination in terms of the relevance to sustainable polymers. The first data set was used in the correlation analyses, while both data sets were independently used in the predictive modeling. The reason for using two data sets is to ensure that the model is robust and not specific to a certain data type or range of systems, or requiring a specific number of measurements. Additionally, comparing the two models allows for assessing if there are differences in applicability between acrylate-focused data and broader data. In this regard, if the broader data provide similar accuracy as unseen acrylate data, then the broader data can be used in the future, as the larger data size will help reduce uncertainty.

One consideration to raise is that polymers with the same chemical composition can have different microstructures and very different Tg values. To account for this, our model incorporates descriptors that capture structural and chemical influences on Tg, including features related to molecular flexibility, intermolecular interactions, and steric effects. While microstructural variations are not explicitly parameterized, their impact is indirectly reflected in the selected features, such as rotational freedom, hydrogen-bonding potential, and aromaticity, all of which influence polymer packing and rigidity. A more explicit characterization of microstructure (e.g., crystallinity levels) could further refine the predictive power of the model and reduce uncertainty.

2.2. Data-Driven Modeling

Two different aspects of modeling were applied: unsupervised learning to understand the underlying correlations and science in the data, and supervised learning to develop predictive models for Tg. The consideration of both aspects is significant because it allows for both a high-throughput model but also provides a science-driven description. An additional novelty of this work is the integration of unsupervised learning and supervised learning. In the case of the correlation analyses, the logic followed an unsupervised approach, but some supervised learning was integrated into the analysis. For the supervised learning approach (predicting Tg), the feature set was first converted into a non-linear parameterization using an unsupervised methodology. Figure 1 summarizes the logic, with the first part (Section 2.3) focused on the correlation and understanding of the features described above with Tg, and the second part (Section 2.4) focused on the development

and testing of a quantitative model. The future application of optimized polymer design is discussed.

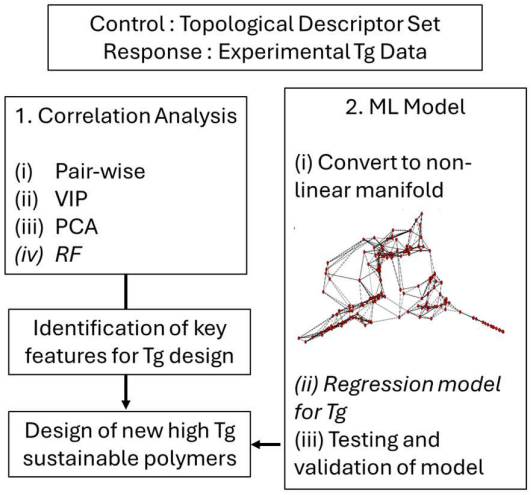


Figure 1. Logic of this paper. The input feature set is based on the topological descriptors, with the work having two aspects: correlation analysis for scientific reasoning and prediction through ML modeling.

2.3. Correlations

The comparison between Tg and the features (Table 1) was assessed through a variety of methods to ensure robustness in conclusions and to verify that no conclusion or interpretation is based solely on the methodology applied. As a first step, we want to ensure that Tg is captured by the feature set. If not, then any model would be expected to have little applicability. To test whether Tg is in fact somehow captured in the correlations within the data, PCA was applied with only the features (and not Tg values) and then repeated with Tg included. The purpose of this step is to see if the variability captured in PCA changes significantly with the addition of Tg; in that case, the result would be based primarily on statistical fitting, suggesting Tg is not represented well in the features and additional features are needed. Figure 2 shows the results of this analysis. The first step (Figure 2a) was to apply k-means clustering to the Principal Component (PC) results. The k-means step identifies those chemistries that are most similar, as defined by proximity in PC space. For this analysis, Tg was not input as the purpose was to explore if Tg is captured without being defined. The labels on the figure are the average Tg for each cluster. Significantly different values in Tg are captured in each cluster, suggesting that Tg is in fact represented in the features and captured in the analysis. The standard deviation in Tg for each cluster is also relatively consistent, so the values are not based on outliers or a few large values. The comparison of variance captured by the PCs, both without and with Tg included (Figure 2b,c, respectively), shows the variance is consistent; thus, the inclusion of Tg does not provide significantly new information. These results indicate that the feature set developed is sufficient for assessing and modeling Tg.

Four different approaches for calculating correlations were applied: pair-wise correlation, VIP for five PCs, the visual comparison of points within PC space for the first three PCs, and the importance of the features from the RF model. The results of these correlations are shown in Figure 3. Positive values indicate a positive correlation, while negative values indicate an inverse correlation. It should be noted that the data set here only contained 53 data points; however, the issue is not necessarily in the size of the data, but rather that the broad range of relevant information is contained in the data. Further to ensure that the results were not over-fit to a relatively small data, the correlation results were tested for robustness by removing data and ensuring that the conclusions did not

change. Additionally, in the next section, the development of a reasonable predictive model using these data gives confidence that the relevant underlying relationships are represented in these data.

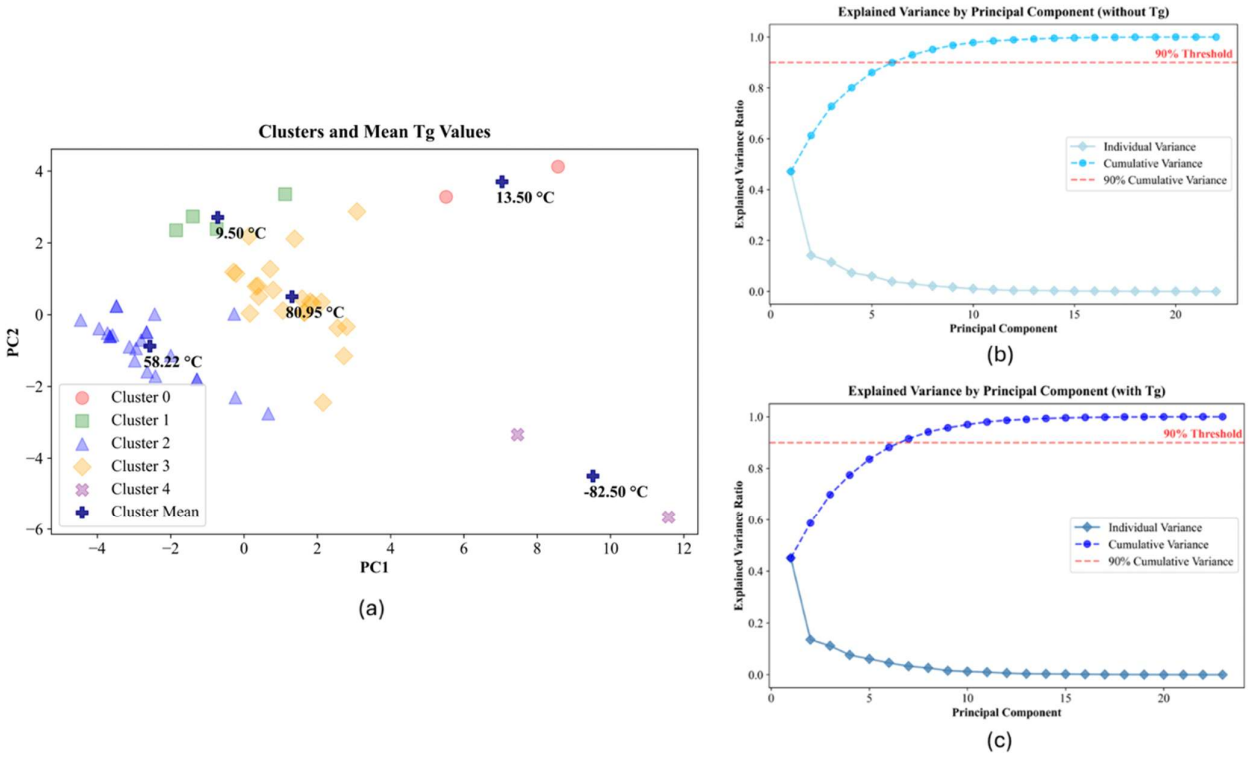
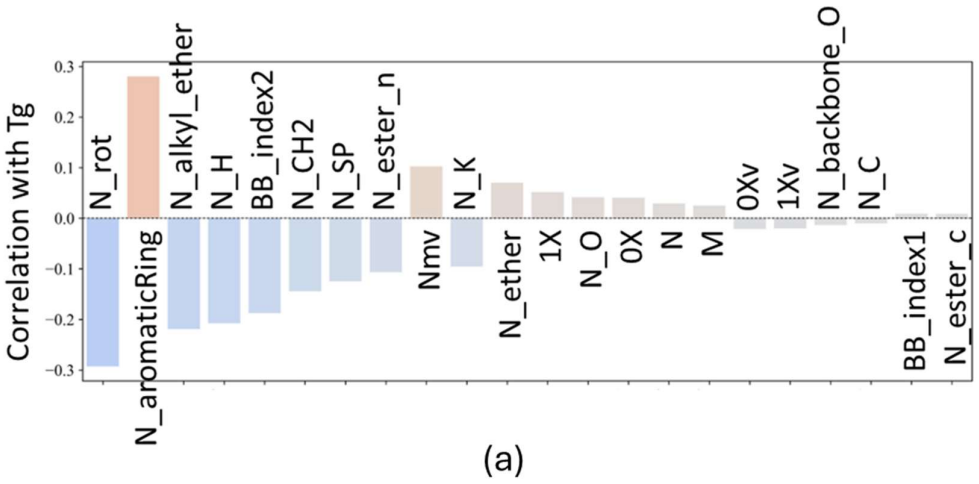


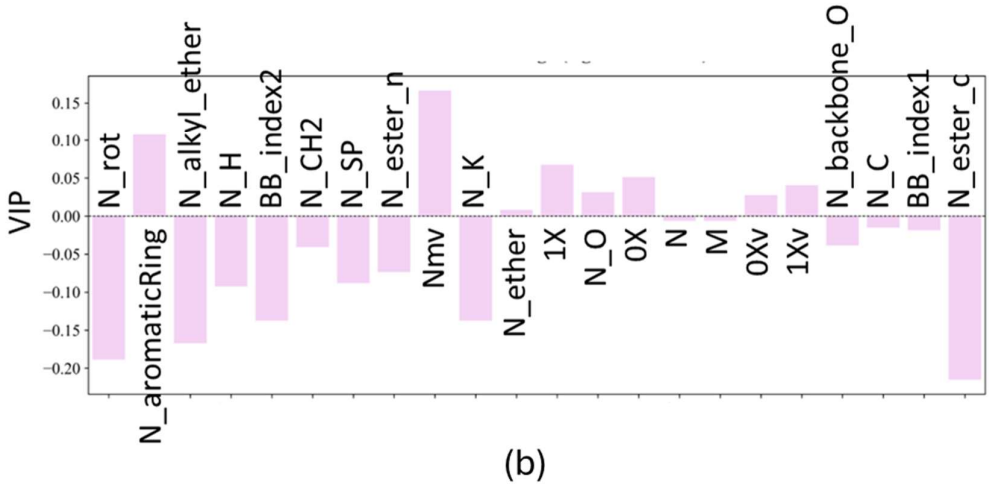
Figure 2. PC analysis to ensure feature set (Table 1) captures changes in Tg. (**a**) Average values of Tg for each cluster defined in the PCs through k-means clustering are shown. These demonstrate that PCA captures the different chemistries resulting in different Tg values. (**b**) Variance of each PC without Tg included in the analysis. (**c**) Variance of each PC with Tg included in the analysis. The similarity of (**b**,**c**) shows that Tg is sufficiently captured in the feature set.

The results across the methods are fairly consistent, with the only significant difference being that VIP identifies N_ester_c as having a large inverse correlation. The inclusion of noise in the lower PCs does not significantly impact the conclusions when comparing Figure 3a,b. The cumulative results of all four approaches are shown in Table 2. In this table, the number of times that a feature was identified among the four approaches is listed, as well as if the correlation was positive (+) or inverse (−).

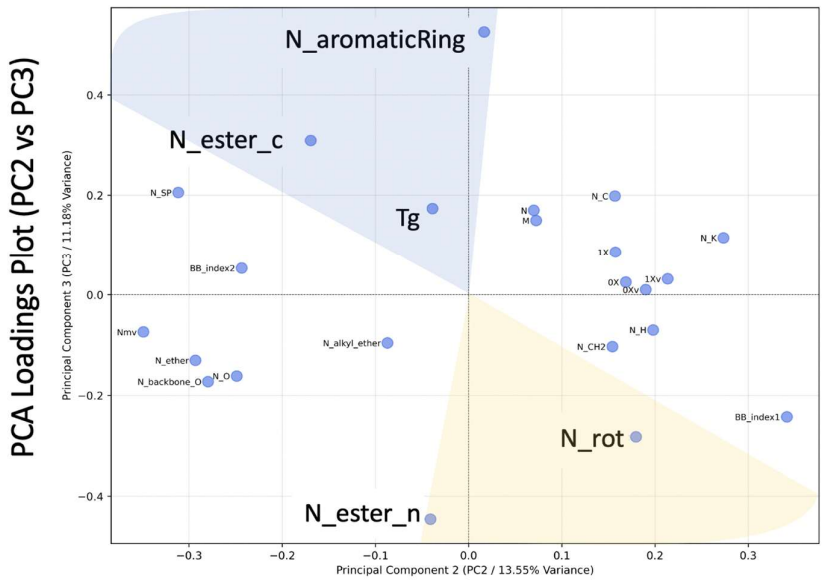
Since each method highlights different aspects of feature importance, we combined the results and counted how frequently each feature was selected across the four techniques. The "Times Selected" column in Table 2 reflects the number of methods that identified a given feature as important. For the pair-wise correlation, VIP, and RF analyses, the top six features with the strongest relationships to Tg were selected, while the PCA loadings features were identified based on their positioning within a predefined shaded cone, capturing those with significant correlation with Tg. While the cut-offs could be defined differently, based on the visual analysis of the resulting correlations, it was determined that this cut-off was sufficient to capture the key correlations without including weak correlations.



(a)



(b)



(c)

Figure 3. *Cont*.

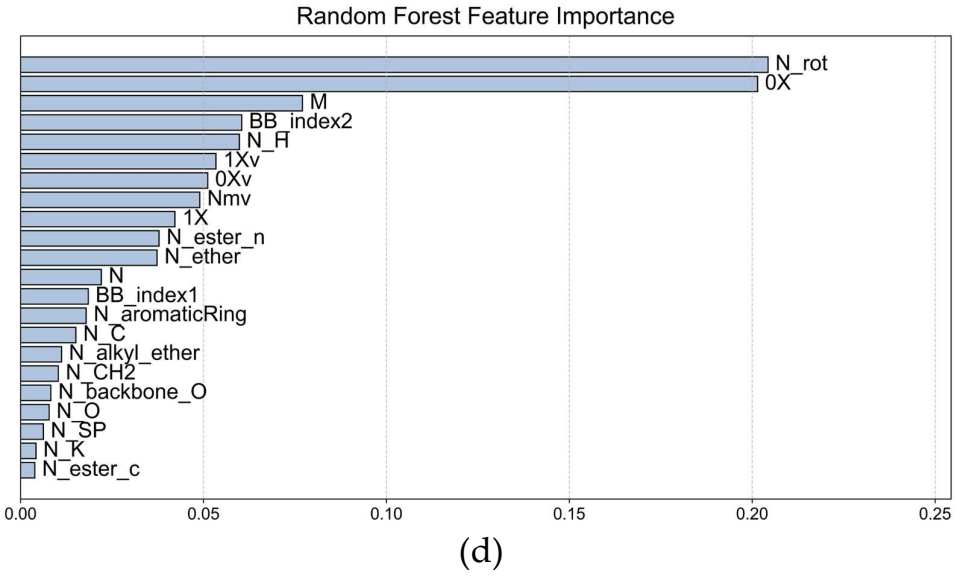


Figure 3. Results from correlation analyses. The labels correspond with Table 1. (**a**) Pair-wise correlations, (**b**) VIP result, (**c**) PCA loading plot with Tg-correlation areas shaded, and (**d**) Random Forest feature importance.

Table 2. Combined results for the four different methods of correlation. 'Feature' corresponds with the labels in Table 1, 'Times Selected' corresponds with the number of methods that identified the feature as important, and 'Correlation' labels whether the correlation was positive or inverse.

Feature	Times Selected	Correlation
N_rot	4	−
BB_index2	3	−
N_alkyl_ether	2	−
N_ether_c	2	+ and −
N_aromaticRing	2	+
N_H	2	−
Nmv	1	+
N_K	1	−
N_ester_n	1	−
$^{0}\chi$	1	
$^{1}\chi^{v}$	1	
M	1	

In terms of conclusions, N_rot was identified as important, with a negative correlation, by all four approaches. BB_index2 was identified as inversely correlated by three of the four models. This would indicate that these two features are most important in controlling Tg. In the case of N_ether_c, it was identified by different models as having different impacts, while the last three features were identified by the RF model, which does not give directionality of correlation. In the Supplementary Materials, heat maps showing the correlations between features are included. From these heat maps, it is found that there are some strong intercorrelations in these key features. N_H, $^{1}\chi^{v}$, and $^{0}\chi^{V}$ are all highly correlated, and thus the information added by including more than one is not that significant. It is noteworthy that many of the topological descriptors are driven by the number of hydrogen atoms. Nmv and N_K also have strong inverse correlations.

Following these results, several interpretations are made. Achieving a high Tg in polymers involves multiple interconnected strategies that target the rigidity/flexibility and interaction dynamics of the polymer structure. One effective approach is the incorporation of conformationally rigid components, such as aliphatic or aromatic rings, into the polymer backbone. These rigid elements reduce chain flexibility, directly contributing to an increase in Tg [8,51]. Additionally, the choice of side-chain structures plays a critical role; side groups designed to lower chain mobility can stabilize the polymer, further enhancing Tg [2].

Another important factor is the introduction of high conformational barriers along the polymer backbone. Structural modifications, such as additional substituents or sp^2-hybridized carbons, create these barriers, reducing the effective degrees of freedom and elevating Tg [20]. Furthermore, augmenting chain–chain interactions through the use of polar or hydrogen-bonding groups strengthens inter-chain forces, which also contribute to a higher Tg. These interactions are better observed in the connected polymerized structure than in individual monomers [52]. Aromatic segments are particularly advantageous in achieving high Tg due to their intrinsic rigidity and capacity for strong inter-chain interactions. Incorporating "hard" segments with high conformational barriers and robust polar associations, such as π–π stacking, effectively enhances Tg. Compared to aliphatic segments, aromatic components impose stricter conformational constraints and facilitate stronger polar interactions, making them highly effective in boosting thermal stability [2,8,52].

The description of these key concepts aligns with the explanation of free rotational degrees commonly discussed in the literature and research. Increasing the rigidity of components, reducing chain mobility, and raising conformational barriers directly affect the number of bonds that can be classified as 'free' single bonds in terms of rotational degrees. The incorporation of rigid rings, floppy rings, or semi-floppy rings into the polymer backbone or side groups modifies the structural dynamics, as these elements do not contribute rotational freedom in the same manner as single bonds [23]. Notably, these structural changes have a minimal impact on the number of atoms connected to the bonds, preserving overall atomic stability while enhancing rigidity.

2.4. Tg Prediction

Two different models were developed for predicting Tg. The first approach used a larger data set that contained a wide variety of polymer chemistries [49], while the second data set was smaller but focused on acrylate systems that are more relevant to our stated objectives [50]. A comparison of the two defines if we can develop a larger and more generalized model that is applicable to all systems and the level of confidence in using the models to design high-temperature sustainable polymer chemistries.

As discussed, the feature set was first converted into a non-linear manifold and then a multi-linear regression was applied. The results of the larger data are shown in Figure 4a. The model has high accuracy, especially considering the uncertainty within the data. A total of 20% of the data were removed prior to building the model and were used as testing data. As seen in the figure, the accuracy of the model was comparable to the testing data, demonstrating that the model was not over-fit and is robust. The process was repeated with the smaller acrylate data set (Figure 4b). Again, the accuracy was high, with comparable training and test accuracy, as shown in the R^2 values labeled in Figure 4. In both cases, the accuracy between training and test data is similar, showing that the models are not overfitting the data and are driven by the underlying relationships in the data. The lower accuracy is expected, given the smaller data set. However, the application of the same feature set and ML framework were applied in both cases, with similar results, demonstrating that the feature set sufficiently captures the mechanics of Tg for polymers in general and acrylates specifically. Based on this, it is anticipated that the larger (or

combined) data sets can be used in the future, even when focusing on specific polymer types. This agrees with the result of Figure 3a, where it was shown that different regimes of Tg map together. Although in that case, it was PC space and not the non-linear space used in the prediction, the concept of self-organization in systems is still relevant.

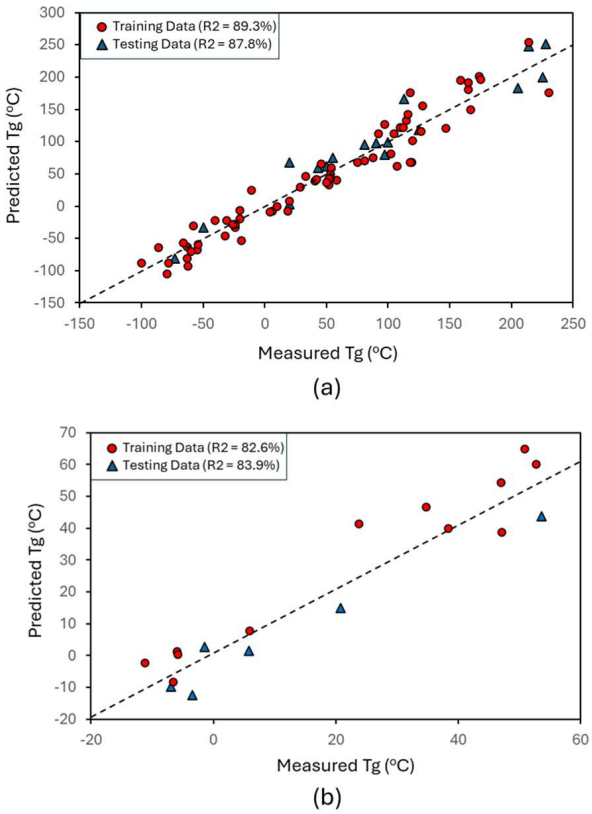


Figure 4. Parity plot from the models, where a regression was applied to the IsoMapping of the feature space. (**a**) Model trained to a larger, more general data set. (**b**) Model trained to a smaller data set focused on acrylates. The similarity in model accuracies demonstrates the robustness of the ML framework. Note that the axes are different within and across the plots for visualization reasons.

3. Discussion

To experimentally validate our model, it was applied to a recently experimentally processed and measured chemistry (Figure 5). The material is available from myo-inositol, a natural material. The polymer was measured to have a remarkable Tg of 210 °C. Relevant experimental details of the process were discussed in our prior work [50]. This chemistry was subsequently predicted using both models. In the case of the larger data model, the Tg value was predicted with high accuracy. In the smaller acrylate data set, the predicted value of the new experimental polymer is not particularly accurate. However, most predictions are not good with extrapolation, as compared with interpolation. That is, models work much better where we have data and are less accurate where we do not have data (such as in this case, where the Tg value is much higher than anything included in the model). However, the model still qualitatively works quite well, as it predicts a much higher Tg for it than any of the other systems. While quantitatively the prediction is not highly accurate, qualitatively it is still accurate and performs within the expected uncertainty in that region, given the limited data included at those Tg ranges. Both results provide a point of experimental validation.

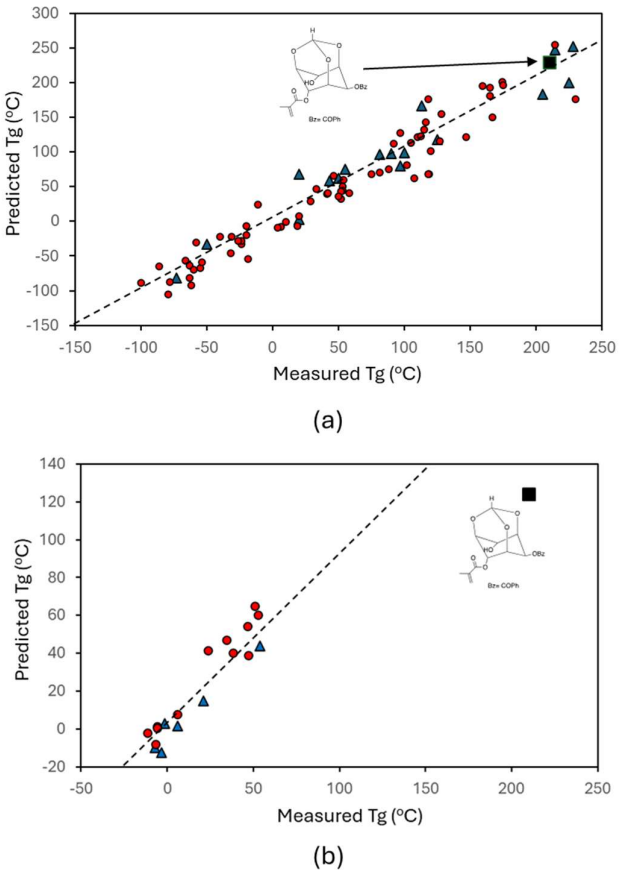


Figure 5. Predictive models corresponding with Figure 3, but with an additional new experimental point added, with the monomer chemistry shown in the figures and the chemistry shown as a square. The model with the larger data (**a**) predicts this chemistry well, while the smaller acrylate data (**b**) predicts the behavior well qualitatively.

4. Materials and Methods

4.1. Unsupervised Learning

To develop correlations between the features and Tg, multiple approaches were used, with principal component analysis (PCA) being the main tool considered. PCA converts the data into linear combinations of the features, with the combinations based on capturing the largest amount of independent variance in the data. In this way, the correlations and information in the data are captured in fewer dimensions. Since the feature set developed in this work is largely correlated, using it as is would result in a model that has large correlations; therefore, the model would be based solely on statistical relationships and not capture the underlying science. To assess the correlations, a pair-wise correlation between each feature and Tg is made. This is followed by calculating the dot product between Tg and each feature after being converted into PC space, providing a so-called Variable Importance Projection (VIP). The VIP serves as akin to a pair-wise correlation except in only a limited number of PCs. Finally, correlations are made based on a supervised learning approach, described in that section.

An unsupervised methodology is also integrated into the predictive modeling. Through the use of IsoMap, the data are converted onto a non-linear manifold. This approach is akin to PCA; however, by considering the geodesic distances, non-linear relationships can be captured. The primary reason for including this step is that it was found to significantly improve model accuracy. This indicates that the relationships between features and Tg are better captured in non-linear space. Additionally, given the model accuracy, we

can conclude that the relationships described through the various correlative analyses are still maintained in this approach. The negative aspect of this step though is that we lose interpretability as to what the model is capturing. However, since we are including the previous step of assessing correlations in the data through different correlative analyses, we still have a scientific assessment of the models.

4.2. Supervised Learning

The final models developed were built solely through the use of multiple linear regression. The key step in the model was the conversion of the features first into the non-linear parameterization with IsoMap. The logic in selecting multiple linear regression is that the simplest model that provides high-quality results is the most likely to maintain robustness. Therefore, this approach is efficient, accurate, and applicable to a broad range of chemistries and structures. Additionally, random forest (RF) analysis was applied between the features and Tg. The model performed relatively poorly, as compared to the multiple linear approach applied to IsoMap parameterization. However, it is still suitable for an additional measure of correlations, where the features that most impact the RF model are identified and compared with the correlation measures through the unsupervised aspect.

5. Conclusions

In this paper, a variety of unique aspects were combined into a single machine learning framework for the prediction and design of sustainable high-Tg polymers. This framework included the development of a feature space based on topological descriptors that can be calculated for any chemistry. The key features controlling Tg were identified and reasonably understood, while it was also verified that the features contained relative information on Tg, thereby leading to an analysis based on underlying science and not statistical fitting. The assessment of features was based on an integration of unsupervised and supervised approaches, while a separate aspect focused on the quantitative prediction of Tg. Using two different data sets, the Tg was well predicted and was additionally experimentally validated with a new, high-Tg sustainable polymer chemistry. This work has significant future implications, first for the automation of the process, and secondly for the optimization of polymers for target properties. While this paper has focused on Tg, it can be expanded in the future to consider other properties and the underlying trade-offs. Therefore, this paper has introduced a machine learning approach for assessing and predicting high-temperature polymers, based on both scientific basis and predictions, while laying out future applications for accelerated design and optimization of multi-functional polymer chemistries.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms26062743/s1.

Author Contributions: Conceptualization, M.F.F., E.W.C., G.A.K. and S.R.B.; Formal analysis, Q.L. and S.R.B.; Investigation, D.D., A.S., V.G., D.F. and S.R.B.; Methodology, Q.L.; Supervision, S.R.B.; Validation, M.F.F.; Writing—original draft, Q.L.; Writing—review and editing, S.R.B. All authors have read and agreed to the published version of the manuscript.

Funding: Research was supported by the Army Research Laboratory under Cooperative Agreement Number W911NF-23-2-0087. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes not with-standing any copyright notation herein.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are available by reasonable email request to the corresponding author.

Conflicts of Interest: The authors declare no competing interests.

References

1. Tang, C.; Ryu, C.Y. (Eds.) *Sustainable Polymers from Biomass*; Wiley-VCH: Weinheim, Germany, 2017; Chapter 1; pp. 36–40.
2. Nguyen, H.T.H.; Qi, P.; Rostagno, M.; Feteha, A.; Miller, S.A. The Quest for High Glass Transition Temperature Bioplastics. *J. Mater. Chem. A* **2018**, *6*, 9298–9331. [CrossRef]
3. Fagnani, D.E.; Tami, J.L.; Copley, G.; Clemons, M.N.; Getzler, Y.D.Y.L.; McNeil, A.J. 100th Anniversary of Macromolecular Science Viewpoint: Redefining Sustainable Polymers. *ACS Macro Lett.* **2021**, *10*, 41–53. [CrossRef] [PubMed]
4. Muelhaupt, R. Green Polymer Chemistry and Bio-Based Plastics: Dreams and Reality. *Macromol. Chem. Phys.* **2013**, *214*, 159–174. [CrossRef]
5. Papageorgiou, G.Z. Thinking Green: Sustainable Polymers from Renewable Resources. *Polymers* **2018**, *10*, 952. [CrossRef]
6. Huang, Y.; Kormakov, S.; He, X.; Gao, X.; Zheng, X.; Liu, Y.; Sun, J.; Wu, D. Conductive Polymer Composites from Renewable Resources: An Overview of Preparation, Properties, and Applications. *Polymers* **2019**, *11*, 187. [CrossRef]
7. Zhu, Y.; Romain, C.; Williams, C.K. Sustainable Polymers from Renewable Resources. *Nature* **2016**, *540*, 354–362. [CrossRef]
8. Yu, X.; Jia, J.; Xu, S.; Lao, K.U.; Sanford, M.J.; Ramakrishnan, R.K.; Nazarenko, S.I.; Hoye, T.R.; Coates, G.W.; DiStasio, R.A. Unraveling Substituent Effects on the Glass Transition Temperatures of Biorenewable Polyesters. *Nat. Commun.* **2018**, *9*, 2880. [CrossRef]
9. Plass, C.; Adebar, N.; Hiessl, R.; Kleber, J.; Grimm, A.; Langsch, A.; Otter, R.; Liese, A.; Gröger, H. Structure-Performance Guided Design of Sustainable Plasticizers from Biorenewable Feedstocks. *Eur. J. Org. Chem.* **2021**, *2021*, 6086–6096. [CrossRef]
10. Ding, J.; Gong, J.; Bai, H.; Li, L.; Zhong, Y.; Ma, Z.; Svrcek, V. Constructing Honeycomb Micropatterns on Nonplanar Substrates with High Glass Transition Temperature Polymers. *J. Colloid Interface Sci.* **2012**, *380*, 99–104. [CrossRef]
11. Sun, H.-S.; Chiu, Y.-C.; Chen, W.-C. Renewable Polymeric Materials for Electronic Applications. *Polym. J.* **2017**, *49*, 61–73. [CrossRef]
12. Zhou, J.; Zhang, H.; Deng, J.; Wu, Y. High Glass-Transition Temperature Acrylate Polymers Derived from Biomasses, Syringaldehyde, and Vanillin. *Macromol. Chem. Phys.* **2016**, *217*, 2402–2408. [CrossRef]
13. Park, P.; Jonnalagadda, S. Predictors of Glass Transition in the Biodegradable Polylactide and Poly-Lactide-Co-Glycolide Polymers. *J. Appl. Polym. Sci.* **2006**, *100*, 1983–1987. [CrossRef]
14. Ni, H.; Liu, J.; Yang, S. Preparation and Characterization of Inherently Heat-Sealable Polyimides with High Glass Transition Temperatures. *J. Appl. Polym. Sci.* **2016**, *133*, 43058. [CrossRef]
15. Li, Y.; Guo, H. Crosslinked Poly(Methyl Methacrylate) with Perfluorocyclobutyl Aryl Ether Moiety as Crosslinking Unit: Thermally Stable Polymer with High Glass Transition Temperature. *RSC Adv.* **2020**, *10*, 1981–1988. [CrossRef]
16. Zhou, Q.; Yan, C.; Li, H.; Zhu, Z.; Gao, Y.; Xiong, J.; Tang, H.; Zhu, C.; Yu, H.; Lopez, S.P.G.; et al. Polymer Fiber Rigid Network with High Glass Transition Temperature Reinforces Stability of Organic Photovoltaics. *Nano-Micro Lett.* **2024**, *16*, 224. [CrossRef]
17. Liu, X.; Gao, Y.; Zhu, X.; Shang, Y.; Cui, Z.; Yan, Q.; Zhang, H. Design and Synthesis of Poly(Arylene Ether Sulfone)s with High Glass Transition Temperature by Introducing Biphenylene Groups. *Polym. Int.* **2020**, *69*, 1267–1274. [CrossRef]
18. Boukis, A.C.; Llevot, A.; Meier, M.A.R. High Glass Transition Temperature Renewable Polymers via Biginelli Multicomponent Polymerization. *Macromol. Rapid Commun.* **2016**, *37*, 643–649. [CrossRef]
19. Odian, G. *Principles of Polymerization*, 4th ed.; John Wiley & Sons Inc.: Hoboken, NJ, USA, 2004; Chapter 1; pp. 620–626.
20. Chee, K.K. Dependence of Glass Transition Temperature on Chain Flexibility and Intermolecular Interactions in Polymers. *J. Appl. Polym. Sci.* **1991**, *43*, 1205–1208. [CrossRef]
21. Solunov, C.A. Cooperative Molecular Dynamics and Strong/Fragile Behavior of Polymers. *Eur. Polym. J.* **1999**, *35*, 1543–1556. [CrossRef]
22. Brunet, J.; Collas, F.; Humbert, M.; Perrin, L.; Brunel, F.; Lacôte, E.; Montarnal, D.; Raynaud, J. High Glass-Transition Temperature Polymer Networks Harnessing the Dynamic Ring Opening of Pinacol Boronates. *Angew. Chem. Int. Ed.* **2019**, *58*, 12216–12222. [CrossRef]
23. Bicerano, J. *Prediction of Polymer Properties*, 2nd ed.; M. Dekker: New York, NY, USA, 1996.
24. van Krevelen, D.W.; te Nijenhuis, K. *Properties of Polymers: Their Correlation with Chemical Structure: Their Numerical Estimation and Prediction from Additive Group Contributions*, 4th ed.; Elsevier: Amsterdam, The Netherlands, 2009.
25. Askadskii, A.A. *Computational Materials Science of Polymers*; Cambridge International Science Pub.: Cambridge, UK, 2003.

26. Kim, C.; Chandrasekaran, A.; Huan, T.D.; Das, D.; Ramprasad, R. Polymer Genome: A Data-Powered Polymer Informatics Platform for Property Predictions. *J. Phys. Chem. C* **2018**, *122*, 17575–17585. [CrossRef]
27. Song, C.; Gu, H.; Zhu, L.; Jiang, W.; Weng, Z.; Zong, L.; Liu, C.; Hu, F.; Pan, Y.; Jian, X. A Polymer Genome Approach for Rational Design of Poly(Aryl Ether)s with High Glass Transition Temperature. *J. Mater. Chem. A* **2023**, *11*, 16985–16994. [CrossRef]
28. Hickey, K.; Feinstein, J.; Sivaraman, G.; MacDonell, M.; Yan, E.; Matherson, C.; Coia, S.; Xu, J.; Picel, K. Applying Machine Learning and Quantum Chemistry to Predict the Glass Transition Temperatures of Polymers. *Comput. Mater. Sci.* **2024**, *238*, 112933. [CrossRef]
29. Liu, D.-F.; Feng, Q.-K.; Zhang, Y.-X.; Zhong, S.-L.; Dang, Z.-M. Prediction of High-Temperature Polymer Dielectrics Using a Bayesian Molecular Design Model. *J. Appl. Phys.* **2022**, *132*, 014901. [CrossRef]
30. Miccio, L.A.; Borredon, C.; Schwartz, G.A. A Glimpse Inside Materials: Polymer Structure–Glass Transition Temperature Relationship as Observed by a Trained Artificial Intelligence. *Comput. Mater. Sci.* **2024**, *236*, 112863. [CrossRef]
31. Casanola-Martin, G.M.; Karuth, A.; Pham-The, H.; González-Díaz, H.; Webster, D.C.; Rasulev, B. Machine Learning Analysis of a Large Set of Homopolymers to Predict Glass Transition Temperatures. *Commun. Chem.* **2024**, *7*, 226–229. [CrossRef]
32. Tao, L.; Chen, G.; Li, Y. Machine Learning Discovery of High-Temperature Polymers. *Patterns* **2021**, *2*, 100225. [CrossRef]
33. Li, X.; Petersen, L.; Broderick, S.; Narasimhan, B.; Rajan, K. Identifying Factors Controlling Protein Release from Combinatorial Biomaterial Libraries via Hybrid Data Mining Methods. *ACS Comb. Sci.* **2011**, *13*, 50–58. [CrossRef]
34. Mullis, A.S.; Broderick, S.R.; Phadke, K.S.; Peroutka-Bigus, N.; Bellaire, B.H.; Rajan, K.; Narasimhan, B. Data Analytics-Guided Rational Design of Antimicrobial Nanomedicines Against Opportunistic, Resistant Pathogens. *Nanomedicine* **2023**, *48*, 102647. [CrossRef]
35. Zhang, Y.; Xu, X. Machine Learning Glass Transition Temperature of Polymers. *Heliyon* **2020**, *6*, e05055. [CrossRef]
36. Phanse, Y.; Puttamreddy, S.; Loy, D.; Ramirez, J.V.; Ross, K.A.; Alvarez-Castro, I.; Mogler, M.; Broderick, S.; Rajan, K.; Narasimhan, B.; et al. RNA Nanovaccine Protects Against White Spot Syndrome Virus in Shrimp. *Vaccines* **2022**, *10*, 1428. [CrossRef] [PubMed]
37. Mullis, A.S.; Broderick, S.R.; Binnebose, A.M.; Peroutka-Bigus, N.; Bellaire, B.H.; Rajan, K.; Narasimhan, B. Data Analytics Approach for Rational Design of Nanomedicines with Programmable Drug Release. *Mol. Pharm.* **2019**, *16*, 1917–1928. [CrossRef] [PubMed]
38. Phanse, Y.; Carrillo-Conde, B.R.; Ramer-Tait, A.E.; Roychoudhury, R.; Broderick, S.; Pohl, N.; Rajan, K.; Narasimhan, B.; Wannemuehler, M.J.; Bellaire, B.H. Functionalization Promotes Pathogen-Mimicking Characteristics of Polyanhydride Nanoparticle Adjuvants. *J. Biomed. Mater. Res. A* **2017**, *105*, 2762–2771. [CrossRef] [PubMed]
39. Ross, K.; Adams, J.; Loyd, H.; Ahmed, S.; Sambol, A.; Broderick, S.; Rajan, K.; Kohut, M.; Bronich, T.; Wannemuehler, M.J.; et al. Combination Nanovaccine Demonstrates Synergistic Enhancement in Efficacy Against Influenza. *ACS Biomater. Sci. Eng.* **2016**, *2*, 368–374. [CrossRef]
40. Rumman, A.H.; Sahriar, M.A.; Islam, M.T.; Shorowordi, K.M.; Carbonara, J.; Broderick, S.; Ahmed, S. Data-Driven Design for Enhanced Efficiency of Sn-Based Perovskite Solar Cells Using Machine Learning. *APL Mach. Learn.* **2023**, *1*, 046117. [CrossRef]
41. Tao, L.; Varshney, V.; Li, Y. Benchmarking Machine Learning Models for Polymer Informatics: An Example of Glass Transition Temperature. *J. Chem. Inf. Model.* **2021**, *61*, 5395–5413. [CrossRef]
42. Ma, R.; Zhang, H.; Xu, J.; Sun, L.; Hayashi, Y.; Yoshida, R.; Shiomi, J.; Wang, J.; Luo, T. Machine Learning-Assisted Exploration of Thermally Conductive Polymers Based on High-Throughput Molecular Dynamics Simulations. *Mater. Today Phys.* **2022**, *28*, 100850. [CrossRef]
43. Kim, C.; Chandrasekaran, A.; Jha, A.; Ramprasad, R. Active-Learning and Materials Design: The Example of High Glass Transition Temperature Polymers. *MRS Commun.* **2019**, *9*, 860–866. [CrossRef]
44. Islam, M.T.; Liu, Q.; Broderick, S. Machine Learning Accelerated Design of High-Temperature Ternary and Quaternary Nitride Superconductors. *Appl. Sci.* **2024**, *14*, 9196. [CrossRef]
45. Broderick, S.R.; Rajan, K. Designing a Periodic Table for Alloy Design: Harnessing Machine Learning to Navigate a Multiscale Information Space. *JOM* **2020**, *72*, 4370–4379. [CrossRef]
46. Giles, S.A.; Sengupta, D.; Broderick, S.R.; Rajan, K. Machine-Learning-Based Intelligent Framework for Discovering Refractory High-Entropy Alloys with Improved High-Temperature Yield Strength. *NPJ Comput. Mater.* **2022**, *8*, 235. [CrossRef]
47. Broderick, S.; Dongol, R.; Rajan, K. Exploring the Shape of Data for Discovering Patterns in Crystal Chemistry. *MRS Commun.* **2021**, *11*, 811–817. [CrossRef]
48. Dasgupta, A.; Gao, Y.; Broderick, S.R.; Pitman, E.B.; Rajan, K. Machine Learning-Aided Identification of Single Atom Alloy Catalysts. *J. Phys. Chem. C* **2020**, *124*, 14158–14166. [CrossRef]
49. Aldrich Polymer Products. *Reference: Polymer Properties*; Sigma-Aldrich: St. Louis, MO, USA; 52–53.
50. Goyal, S.; Lin, F.-Y.; Forrester, M.; Henrichsen, W.; Murphy, G.; Shen, L.; Wang, T.; Cochran, E.W. Glycerol Ketals as Building Blocks for a New Class of Biobased (Meth)acrylate Polymers. *ACS Sustain. Chem. Eng.* **2021**, *9*, 10620–10629. [CrossRef]

51. Kanbargi, N.; Damron, J.T.; Gao, Y.; Kearney, L.T.; Carrillo, J.M.; Keum, J.K.; Sumpter, B.G.; Naskar, A.K. Amplifying Nanoparticle Reinforcement through Low Volume Topologically Controlled Chemical Coupling. *ACS Macro Lett.* **2024**, *13*, 280–287. [CrossRef]
52. Nguyen, H.N.; Lu, L.-H.; Huang, C.-J. Aromatic Disulfide Cross-Linkers for Self-Healable and Recyclable Acrylic Polymer Networks. *ACS Appl. Polym. Mater.* **2024**, *6*, 4615–4624. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.