Macromolecules

pubs.acs.org/Macromolecules Article

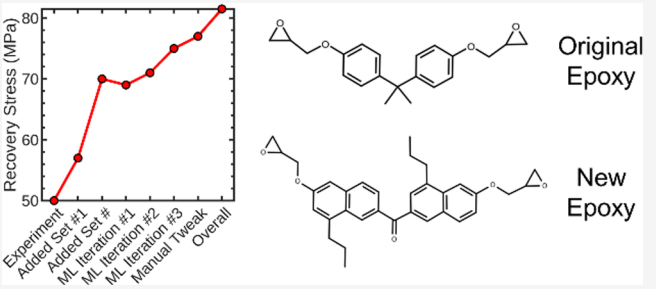
Identification and Design of Better Diamine-Hardened Epoxy-Based Thermoset Shape Memory Polymers: Simulation and Machine Learning

Anwar Shafe, Pouria Nourian, Xiyuan Liu, Guoqiang Li, Collin D. Wick, and Andrew J. Peters*

Cite This: https://doi.org/10.1021/acs.macromol.4c01598 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: An approach for designing thermoset shape memory polymers (TSMPs) with improved shape memory properties through the integration of molecular dynamics (MD) simulation, machine learning (ML), and chemical intuition is presented. We identified key molecular features correlated with desired shape memory properties, and used MD simulations to create an initial data set of TSMPs consisting of commercially available and manually designed monomers. Our prediction set was prepared by employing four different approaches for modifying existing monomers based on chemical intuition and insights gleaned from the literature. We trained our ML model on the initial data set, used it to identify the most promising candidates, evaluated their properties, and added them to our initial data set. To further speed up the process, we identified the most promising candidate after a few cycles and modified its structure to obtain a variant with better properties. Our approach, which capitalizes on the synergy between computational methodologies and human expertise to enable efficient exploration of vast chemical space, resulted in the design of a monomer exhibiting more than 60% increase in the desired recovery stress compared to the highest experimentally validated one.



1. INTRODUCTION

Shape memory polymers (SMPs) are materials that can return to their original shapes after being plastically deformed. This unique property is due to the energy stored in the cross-linked polymer structure. When exposed to an external stimulus like temperature changes,[1−3] light exposure[4] or Joule heating, the stored energy in the cross-linked structure releases and drives the recovery of the original shape. Current applications of shape memory polymers include biomedical devices,[5] actuators,[6] smart textiles,[7] and actuators in the aerospace industry.[8,9] Moreover, SMPs can be integrated with other advanced materials to create novel products with exceptional functionalities, exemplified by using SMP coatings to enhance stent devices.[10] SMPs are also being considered in drug delivery[11,12] and tissue engineering.[13,14] SMPs have several advantages over traditional materials, such as reduced weight and corrosion resistance, making them desirable in industries requiring lightweight and long-lasting materials.[15]

SMPs can be manufactured from both thermoplastics[16] and thermosets.[15] Different plastics affect the material's properties, such as thermal behavior, recovery stress, recovery ratio, and cycle life. Varying chemical composition, molecular weight, and cross-linking density can also affect the properties of shape memory polymers.[15] Further improvements to SMPs will come through identifying and realizing superior monomer chemistry and network structure.[17,18] Simulation and machine learning (ML) offer the potential to accelerate this development process rapidly.

Computationally modeling and screening compounds have resulted in the development of novel polymers,[19−21] as well as photovoltaic,[22] piezoelectric,[23] and CO_2 capture materials.[24] Recently, ML has been used to propose new and improved shape memory polymers. In one study by Yan et al., a Convolutional Neural Network (CNN) was used to accurately predict new shape memory polymers with improved properties, thus significantly reducing the trial-and-error development of new SMPs.[25] Another study by the same authors predicted SMPs with the consideration of molar ratio of compounds based on a limited training data set.[26] The authors concluded that more fingerprints or descriptors must be identified to predict SMPs accurately. Similarly, ML has been used to predict the thermomechanical behavior of SMPs.[27]

Such approaches require feature engineering or "molecular fingerprinting" that uses atomic-level features or properties such as bond stiffness or bond angle to identify unique

Received: July 24, 2024
Revised: September 6, 2024
Accepted: October 14, 2024

Table 1. List of Atomistic Fingerprints

fingerprint	description
epoxy length	The number of bonds that separate the shortest distance between the reacted DGE carbons.
hardener length	The number of bonds that separate the shortest distance of reacted amine nitrogens.
nbackbone	The average number of backbone heavy atoms following complete cross-linking, divided by the initial number of monomers. Please note that this calculation counts all atoms within a ring structure.
bbratio	The ratio of backbone heavy atoms to the overall count of heavy atoms.
stretch	The stiffness of the bond stretching interaction among heavy atoms.
angle	The stiffness of the angular interaction between heavy atoms.
dihedral	Dihedral energy strength between heavy atoms.
vdW size and strength	The van der Waals (vdW) size and strength are characterized by their mean σ and ε values in the heavy atom Lennard-Jones potential, employed to describe vdW interactions.
polarity	The polarity of each atom is determined by the mean squared electron charge (q_i^2) per atom.
R_g	The radius of gyration for the epoxy monomer and the hardener monomer.
max SC	The maximum length of sidechains for heavy atoms. For a particular heavy atom within a side chain, its side chain length is characterized by the number of bonds between itself and the nearest backbone atom.
nring	The number of aromatic and nonaromatic backbone rings.

molecular characteristics of a substance.[28] These fingerprints can be used as inputs to ML algorithms,[28] allowing for the prediction of polymer properties, such as mechanical strength and thermal behavior, based on their molecular structure. Molecular fingerprinting can also be used for polymer classification, discovering relationships between different polymers, and identifying structural features associated with specific properties that may be overlooked without computer-assisted screening or ML.[28]

In our previous study,[29] we calculated twenty-one atomistic fingerprints for systems consisting of 9 epoxies and 22 hardeners, and examined their correlations with epoxies and amine hardeners pairs and shape memory properties (recovery stress and recovery ratio).

This work aims to identify amine-hardened, diglycidyl-ether-based epoxy systems with higher recovery stress by using an optimization scheme that combines a data-driven approach with chemical intuition. Here, we expand the data set from our previous work with a series of new epoxies and hardeners. Newly designed molecules inspired by other literature and our chemical intuition were simulated, and their shape memory properties were obtained. This expanded data set serves as a better foundation for ML algorithms that need wide sampling to accurately predict new structures. A small number of models for predicting the recovery stress of new epoxies and hardeners were tested on the initial data set, and the best model was identified. We then created new epoxies and hardeners using four different approaches inspired by chemical intuition and insights gleaned from previous analyses. The ML algorithm was then used to identify the best epoxies and hardeners, which were subsequently simulated, and the ML algorithm was retrained. This was repeated three times. Finally, the optimization process was pushed even further by manually tweaking the structure of the best molecules, which were then tested. Several new thermoset shape memory polymer (TSMP) candidates with a calculated recovery stress much higher than previously found were ultimately identified.

2. MODELS AND METHODS

2.1. Molecular Dynamics Simulations. The protocol used for calculating recovery stress is identical to that used in our previous work.[29] We used LAMMPS[30] to run molecular dynamics simulations of the thermomechanical cycle of the cross-linked epoxy-hardener system. This cycle for a heat-activated TSMP typically involves three steps: loading, relaxation, and shape recovery. All epoxies and hardeners investigated had two glycidyl ether and two amine groups, respectively. The monomers were modeled using the LigParGen[31−33] software with the optimized potential for liquid simulations (OPLS) forcefield.[34−36] The monomers were then cross-linked and equilibrated to create five independent systems as described in detail in our previous work.[29] To ensure an equal number of diglycidyl ethers (DGE) and amine hydrogens, a simulation box was generated for each system containing two epoxies and one hardener, and their structures were relaxed through conjugate gradient energy minimization. Following this, the systems were replicated six times in the X, Y, and Z directions. The cross-linking process was conducted at 398 K to create a network with a cross-linking percentage of 70% as described in our previous work,[37] which took 600 cross-linking steps to complete.

Once the five independent networks were formed, molecular dynamics simulations were used to calculate their glass transition temperatures (T_g) and recovery stress. For the calculation of T_g, we initially subjected the systems to a heating process, raising the temperature to 598 K, and allowing them to reach equilibrium during a 500 ps period. Subsequently, a gradual cooling procedure spanning 2.5 ns was carried out, lowering the temperature from 598 to 198 K while monitoring density changes. For each of the five independent systems, simulations at multiple temperatures were carried out, ranging from 250 to 550 K in increments of 0.8 K. We conducted two linear regression analyses in each temperature trial: one to find the best-fit straight line for density versus temperature for all temperatures 40 K above the trial temperature and another for temperatures 40 K below it. The intersection point of these two straight lines was identified as the T_g value. We selected the trial that minimized the error between these straight lines and the density/temperature curve to determine the T_g value for each independent system.

To calculate the recovery stress, the systems were first brought to equilibrium using the NPT ensemble at 1 atm pressure, 30 K above their T_g values. They were then programmed by slowly compressing them by 50% for 4 ns in one direction, after which the systems were relaxed for 4 ns. The box directions orthogonal to the programmed directions were allowed to fluctuate according to the NPT pressure controller. They were then cooled to 298 K over 2 ns and given an additional 2 ns to further equilibrate. After fixing the direction of the initial deformation, the system was heated over 2 ns to 30 K over respective T_g values. The recovery stress was then determined by running a further 2 ns simulation at 30 K above the T_g. While the short timescales compared to experiment are known to result in an overpredicted recovery stress compared to experiment, this overprediction should be consistent across various systems as long as the relaxation time for each is comparable.[29]

2.2. Fingerprinting. In this study, we used atomistic fingerprints to describe the epoxy-hardener systems. The atomistic fingerprints describe the atomic level features and thus can be calculated for a single set of epoxy and hardener molecules before replication and

https://doi.org/10.1021/acs.macromol.4c01598
Macromolecules XXXX, XXX, XXX−XXX

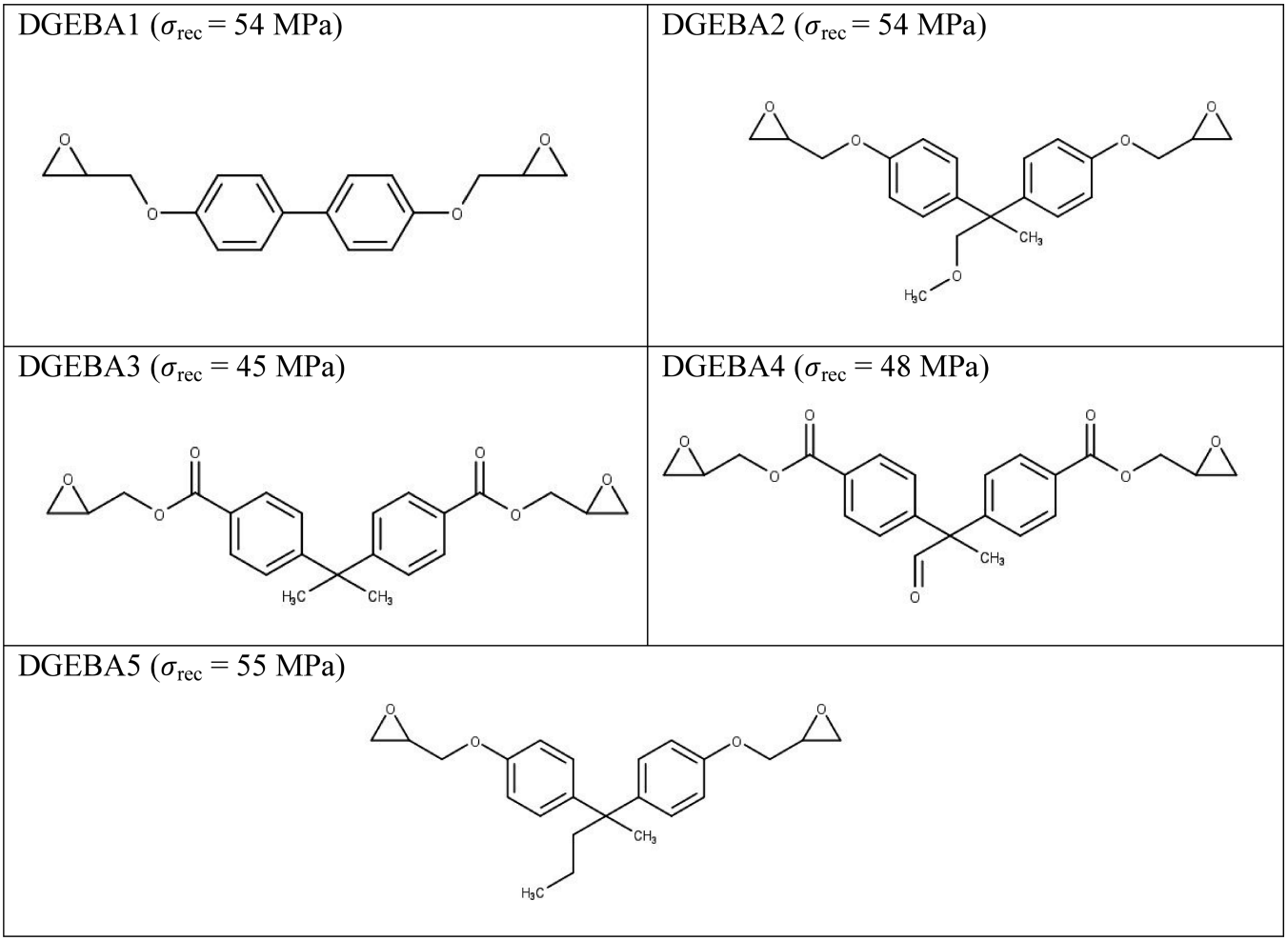


Figure 1. Epoxies modified from DGEBA. σ_{rec} values are all shown with IPD as a hardener.

cross-linking. This allows for fingerprinting of a single system in about one minute. The equations for calculating atomistic fingerprints are described in detail in our previous paper[29] and are summarized in Table 1.

2.3. Training Set Creation. The initial data set, taken from our previous work,[29] only included epoxies and hardeners that are available commercially. As mentioned in the introduction, this study aims to increase the search space of epoxy-hardener systems. To achieve these goals, epoxies and hardeners were modified systematically using insights offered by our previous work. These additions to the training set provided a convenient way of testing hypotheses, which are discussed with each set.

The first modification examined the impact of changing the backbone and side chains of Bisphenol A Diglycidyl Ether (DGEBA), which was shown to have the highest recovery stress of the epoxies we studied previously.[29] The specific modifications can be seen in Figure 1. The sidechains differed in polarity and length, and in the case of DGEBA1, they were removed completely. Each of these was tested with Isophorone diamine (IPD) as a hardener. The modified epoxies showed similar recovery stresses to DGEBA (50 MPa[29]). Introducing more atoms along the backbone and lengthening it was detrimental to the recovery stress, as seen by looking at DGEBA3 and DGEBA4. In the case of DGEBA1, removing the central carbons led to an increase in recovery stress. Similar results were also observed by adding more carbon atoms (DGEBA5) or an ether group (DGEBA4) to the side chain coming off the molecule's center.

We then investigated the effect of the position of glycidyl groups around the epoxy monomer core. Resorcinol diglycidyl ether (DGER) and 1,7-dihydroxyl naphthalene diglycidyl ether (DGEDHN) were modified as shown in Figure 2. The simulation results, using IPD as a

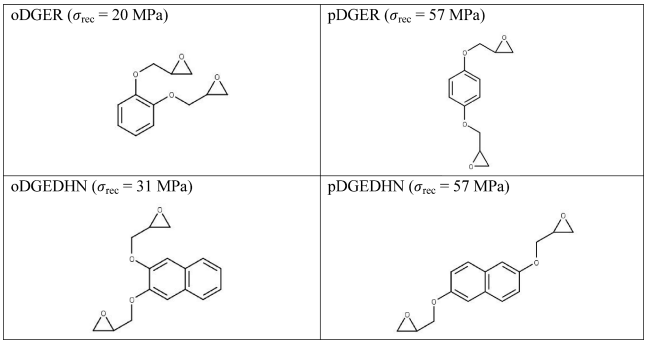


Figure 2. Modifications of DGER and DGEDHN. The shown recovery stresses are those with IPD as a hardener.

hardener, show that when the glycidyl ether groups are directly opposite (as in para position along a six-membered ring), the recovery stress is the highest and decreases as the glycidyl ether groups move nearer to one another. The difference in behavior between the para-like and ortho-like positions in a monomer arises due to the structural constraints imposed by the aromatic ring. When glycidyl ether groups are positioned ortho to the ring, the stiff ring itself is not as significant of a part of the backbone ring as in the para-like position. Consequently, the monomer exhibits greater flexibility within the network when the glycidyl ether groups occupy the ortho-like position. When the glycidyl ether groups fall between the para and ortho positions, as in unmodified DGER and DGEDHN (see Figure S1), the recovery stress is between the two variants − the ring

https://doi.org/10.1021/acs.macromol.4c01598
Macromolecules XXXX, XXX, XXX−XXX

contributes to the backbone, but some is bypassed and so the recovery stress is not as high as when in the para position.

To further expand the training data set, biobased epoxies were examined.[38] These include Methyl-2,4-dihydroxybenzoate diglycidyl ether (MDHB), Vanillyl Alcohol Diglycidyl Ether (DGEVA), Furan Diglycidyl Ether (DGEF), and Bisfuran Diglycidyl Ether (DGEBF), as shown in Figure S2. MDHB and DGEVA are similar to DGER and pDGER but have different sidechains. As with DGER vs pDGER, MDHB exhibited lower recovery stress because of the position of the glycidyl ether groups. In both cases, the recovery stress was lowered by adding sidechains.

In addition to expanding the training set of epoxies, more hardeners were examined as well. New hardeners were taken from PubChem's "Find Similar Structures" feature,[39] where IPD was used as the seed molecule. These hardeners were 6,7-dimethyl-5,6,7,8-tetrahydronaphthalene-2,3-diamine (DTHDA), 4,5-dimethylbenzene-1,2-diamine (DBDA), 3-(3-aminobenzoyl) aniline (ABA), 4-(4-aminophenyl) aniline (APA), 3,5-diaminobenzoic acid (DABA), 3-(2-aminophenoxy) aniline (APOA). The recovery stresses of these hardeners with DGEBA were calculated and are shown in Figure S3. Of note, and shown in Figure 3, are APA and DABA. APA had the highest recovery stress among the hardeners examined due to its stiff structure. DABA also yielded a high recovery stress (67 MPa), even though it has a very similar structure to m-Phenylene Diamine MPDA (σ_{rec} = 42 MPa), but a carboxylic acid group bonded to its ring.

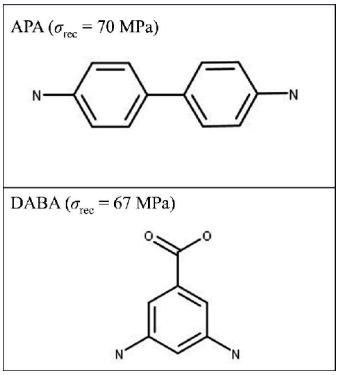


Figure 3. Structures of APA and DABA (recovery stresses are with DGEBA as an epoxy resin).

2.4. Selection of Prediction Algorithm. With the initial training data set expanded, we used the scikit-learn[40] library to perform all of our model training and predictions. For this study, we chose five regression models from the linear model sublibrary of scikit-learn. The models are multiple linear regression, ridge regression, Bayesian ridge regression, Theil Sen regression, and Poisson regression. The initial data set containing 67 data points was split into a training set and a validation set, with 80% of the data points in the training set and the rest in the validation set. The parameters used to train each model are given in Table S2. The fingerprints were used as inputs, and the recovery stress was the output. The trained models were then used to predict the recovery stress of validation set. The mean squared error (MSE) of each prediction is defined as $\mathrm{MSE} = \frac{1}{n}\left\{\sum_{\{i=1\}}^{\{n\}}\left(\hat{y}_i - y_i\right)^2\right\}$, where n is the number of measurements, y is the vector of observed values and $\hat{y}$ is the vector of predicted values. The MSE was calculated for each model and is reported in Figure 4, while the R^2 values and residual plots are shown in Figure S9. Multiple linear regression resulted in the lowest MSE, and highest R^2 and so it is used for the rest of this work. Further optimization of the training process and creation of a larger data set would likely result in more complex methods that perform better, but multiple linear regression was sufficient for identifying shape memory polymers with enhanced recovery stress.

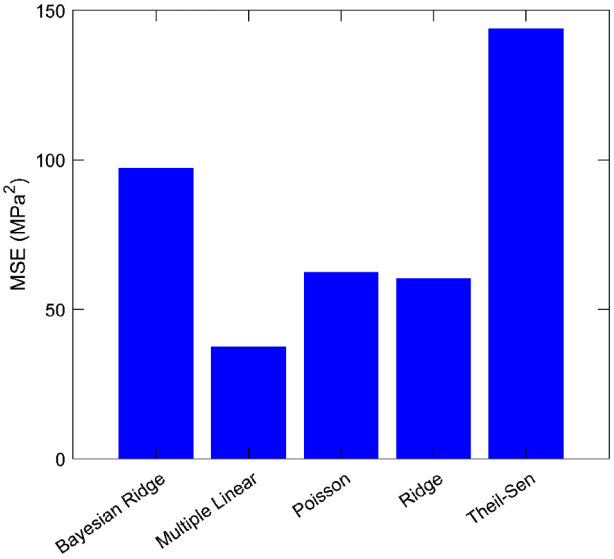


Figure 4. Average mean squared error for different models.

2.5. Prediction Set Creation. To identify systems with enhanced recovery stress a large number of new systems were created, and the trained model was used to predict their recovery stress. We created two types of epoxy monomers, those that were symmetric and those that were asymmetric, for which the recovery stress was predicted after pairing them with hardeners. Both methods were constructed from large and small fragments that were bonded together to create the entire epoxy. A total of 861 symmetric epoxies and 660 asymmetric epoxies were created and fingerprinted. The full details of the creation method along with figures of the various fragments that were used to build up the epoxies can be found in the SI.

Several hardeners were also created, and the trained model was used to predict their recovery stress with various epoxy monomers. A two-method approach similar to that used for the epoxies was employed to create the hardener test systems. In one method, we moved the sidechains around a base hardener, and in the other, we moved the location of the amines around the hardener. A total of 4615 hardeners were created by moving sidechains around base hardeners and a total of 769 hardeners were generated by moving the amines around the base hardener. The full details can be found in the SI, as well as figures of the base hardeners and sidechains attached to them.

3. RESULTS

3.1. Automated Prediction of Better TSMPs. A linear regression model was trained on the initial 67 training data points. The target variable was recovery stress, and the features were the fingerprints described above. The coefficients for the features are listed in the Supporting Information. Afterward, the model was used to make predictions for our prediction set. Separate predictions were made for each prediction set category: (1) Symmetric epoxies, (2) Asymmetric epoxies, (3) Hardeners with different side chain locations, and (4) Hardeners with different amine locations. The two systems with the highest recovery stress from each set for a total of 8 systems (see Table 2) were then selected and simulated with MD. A comparison of the predicted and calculated recovery stress for the first set of predictions is presented in Table 2. The SMILES strings of the epoxies and hardeners are also in the SI.

The systems with the modified epoxies had a much higher predicted recovery stresses than with the modified hardeners. However, the calculated values, while an improvement over the

https://doi.org/10.1021/acs.macromol.4c01598
Macromolecules XXXX, XXX, XXX−XXX

Table 2. Predicted and Calculated Recovery Stress (σ_{rec}) of the First Set of Predictions

epoxy	hardener	prediction set	predicted σ_{rec} (MPa)	calculated σ_{rec} (MPa)
autoEpoxy2348	IPD	symmetric epoxy	81.39	53
autoEpoxy2256	IPD	symmetric epoxy	80.67	63
autoEpoxy2_409	APA	asymmetric epoxy	195.42	69
autoEpoxy2_545	APA	asymmetric epoxy	194.34	56
DGEBA	autohardener96	hardener with moved side chain	67.61	59
DGEBA	autohardener70	hardener with moved side chain	66.34	55
DGEBA	MPDA_Base4	hardener with moved amine	56.44	60
DGEBA	dMPDA_Base117	hardener with moved amine	56.18	61

Table 3. Predicted and Calculated Recovery Stress (σ_{rec}) of the Second Set of Predictions

epoxy	hardener	prediction set	predicted σ_{rec} (MPa)	calculated σ_{rec} (MPa)
autoEpoxy3084	IPD	symmetric epoxy	61	57
autoEpoxy4694	IPD	symmetric epoxy	61	71
autoEpoxy2_590	APA	asymmetric epoxy	70	68
autoEpoxy2_592	APA	asymmetric epoxy	69	59
DGEBA	autohardener40	hardener with moved side chain	66	50
DGEBA	autohardener142	hardener with moved side chain	66	54
DGEBA	dDAP_Base134	hardener with moved amine	57	39
DGEBA	dMPDA_Base227	hardener with moved amine	56	53

Table 4. Predicted and Calculated Recovery Stress (σ_{rec}) of the Third Set of Predictions

epoxy	hardener	prediction set	predicted recovery stress (MPa)	calculated recovery stress (MPa)
autoEpoxy3038	IPD	symmetric epoxy	60	75
autoEpoxy4510	IPD	symmetric epoxy	60	69
autoEpoxy2_400	APA	asymmetric epoxy	69	67
autoEpoxy2_598	APA	asymmetric epoxy	63	59
DGEBA	autohardener421	hardener with moved side chain	63	51
DGEBA	autohardener539	hardener with moved side chain	63	20
DGEBA	dMPDA_Base213	hardener with moved amine	56	45
DGEBA	MPDA_Base16	hardener with moved amine	55	58

original DGEBA-IPD system with a recovery stress of 50 MPa, were much lower than the ML predicted values. This is especially true of the two newly created asymmetric epoxies, which had unreasonably high predicted recovery stresses of over 190 MPa. This is likely because these new epoxies were much different from the original training data set (i.e., located much further from other systems in the search space) for the model to accurately capture the effect of the specific combinations considered. For example, some of these new epoxies have a backbone length greater than the highest backbone length found in the initial data set. We remedied this by adding these new structures from Table 2 to the training set and making new predictions. Again, the two systems with the highest recovery stress from each set for a total of 8 systems were then selected and simulated with MD. The results from the second predictions are shown in Table 3.

The newly predicted recovery stresses are much more reasonable and better match the calculated recovery stresses, especially for the asymmetric epoxies (autoEpoxy2_590 and autoEpoxy2_592). Furthermore, the highest calculated recovery stress has improved to 71 MPa. A third set of predictions were then made including the systems from Table 3 in the training data set. This third training data set then had 83 data points and a final data set of predictions were made. The results from the third prediction are summarized in Table 4.

The newly predicted epoxies continue to have higher calculated recovery stress, while the hardeners did not show any improvement this time. As can be seen in Figure 5, the accuracy of predictions increased with each ML iteration. Points are drawn as outliers if they are greater than the third quartile plus 1.5 times the inner quartile range, or less than the first quartile minus the interquartile range. The third prediction set has an outlier with a hardener that performed unexpectedly poorly, but the average error was near zero percent.

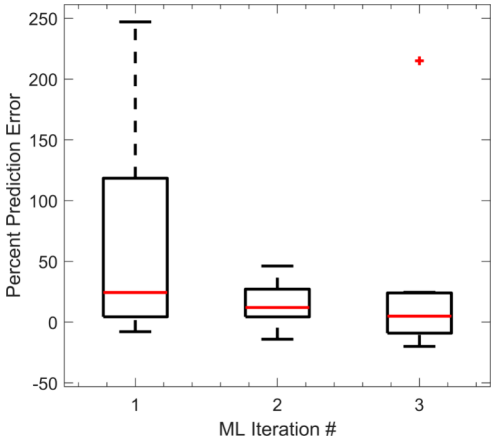


Figure 5. Boxplot of percentage prediction error (defined as $\frac{\sigma_{rec}^{predicted} - \sigma_{rec}^{calcualted}}{\sigma_{rec}^{calculated}}$) for each ML iteration. The red cross indicates an outlier.

https://doi.org/10.1021/acs.macromol.4c01598
Macromolecules XXXX, XXX, XXX−XXX

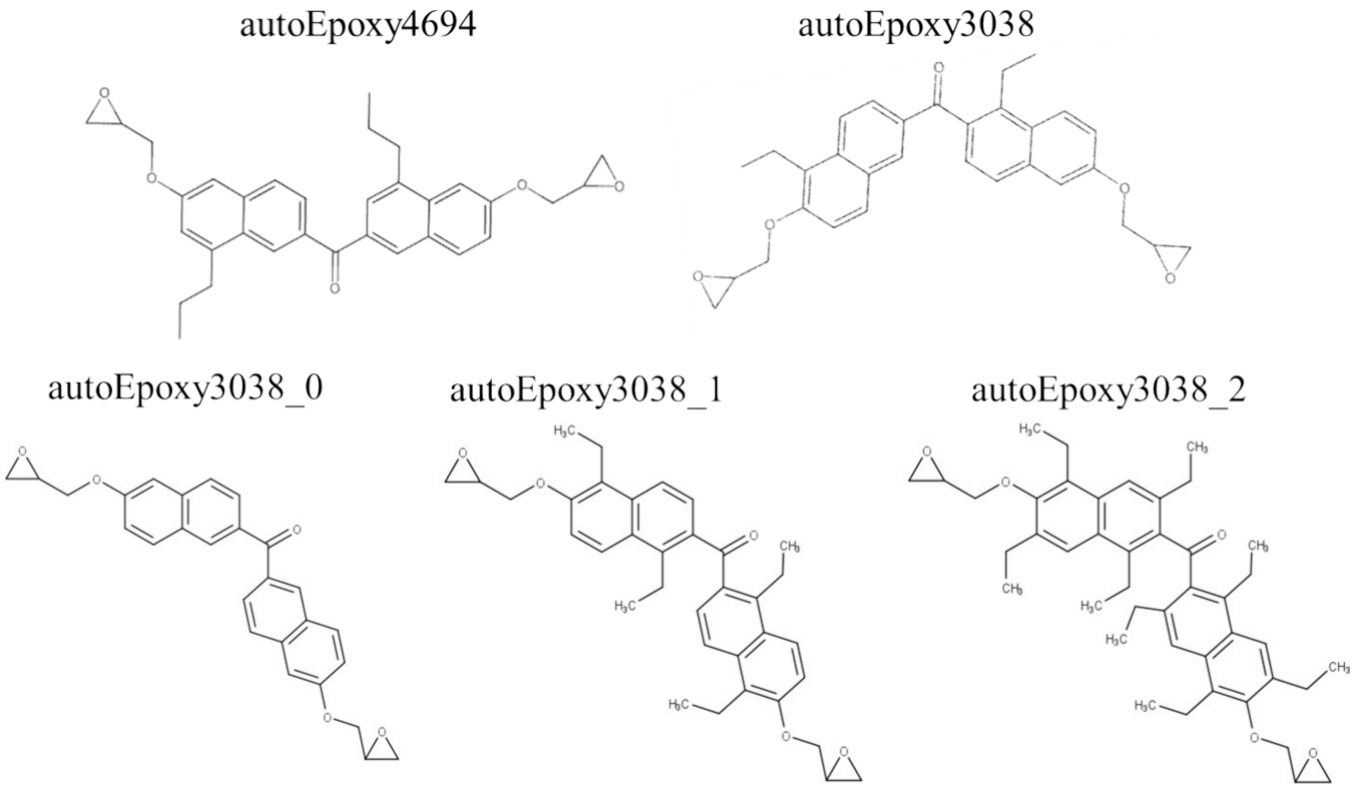


Figure 6. Structures of the best epoxies found from ML (autoEpoxy4694 and autoEpoxy3038) and the side chain modified variants of autoEpoxy3038.

Through this approach, we achieved a recovery stress of 75 MPa. This optimization process can be repeated to gradually move in the feature space toward the theoretical best-performing system. However, while the calculation of fingerprints and prediction of new systems is relatively fast, the calculation of recovery stress takes about a week for each new system. Therefore, we pivot toward leveraging chemical intuition and insights gleaned from our analysis to fine-tune the molecular structure of the best systems found so far for a final improvement.

3.2. Structural Optimization Based on Chemical Intuition. The highest recovery stress systems found with the automated procedure are autoEpoxy3038 w/IPD (75 MPa) and autoEpoxy 4694 w/IPD (71 MPa). AutoEpoxy3038 and autoEpoxy4694 are shown in Figure 6. They are notably very similar in structure, with the only difference being the location and length of the sidechains: autoEpoxy3038 has an ethyl group closer to the glycidyl group, and one carbon closer than autoEpoxy4694. We hypothesize that putting side chains near the two glycidyl groups would increase recovery stress, possibly related to additional steric hindrance. To test this, we selected the epoxy that resulted in the highest recovery stress (autoEpoxy3038 with a recovery stress of 75 MPa) and created 3 variants of it by changing the number of its sidechains. The resulting new epoxies are autoEpoxy3038_0, autoEpoxy3038_1, and autoEpoxy3038_2, all of which are found in Figure 6 AutoEpoxy3038_0 has no sidechains. AutoEpoxy3038_1 has four ethyl groups, two of which are bonded to the carbons next to glycidyl ether groups and two bonded to the carbons next to the middle carbonyl group. AutoEpoxy3038_2 has four more sidechains than autoEpoxy3038_1; four ethyl groups are bonded to the carbons next to the glycidyl ether groups, and the other four are bonded to the carbons next to the carbonyl group. We then simulated these 3 epoxies variants w/ IPD and calculated their recovery stress (see Table 5).

Table 5. Calculated Recovery Stress (σ_{rec}) of autoEpoxy3038 and Its Variants

epoxy	hardener	predicted σ_{rec} (MPa)	calculated σ_{rec} (MPa)
autoEpoxy3038	IPD	60	75
autoEpoxy3038_0	IPD	55	77
autoEpoxy3038_1	IPD	69	56
autoEpoxy3038_2	IPD	63	63

The system made with autoEpoxy3038_0 has a higher recovery stress (77 MPa) compared to the one made with autoEpoxy3038. However, as mentioned earlier, autoEpoxy3038_0 has no side chains. It appears that in contrast to our hypothesis, the addition of extra sidechains reduced the recovery stress instead of increasing it. Moreover, autoEpoxy3038_2 has a higher recovery stress than autoEpoxy3038_1 despite having the highest number of sidechains. Therefore, it seems that more parameters than simply the number and location of sidechains are at play when it comes to the overall recovery stress of the network.

Another modification that can be implemented would be to change the hardener. We noticed that systems with APA as the hardener generally have a higher recovery stress compared to those with IPD. Thus, we calculated the recovery stress of systems with autoEpoxy4694 and autoEpoxy3038 as the epoxy and APA as the hardener (Table 6).

Had the autoEpoxy4694 (see Figure 6) been discarded in the previous step and focused on different systems with

https://doi.org/10.1021/acs.macromol.4c01598
Macromolecules XXXX, XXX, XXX−XXX

Table 6. autoEpoxy4694 and autoEpoxy3038 with APA

epoxy	hardener	predicted σ_{rec} (MPa)	calculated σ_{rec} (MPa)
autoEpoxy4694	APA	66	81.5
autoEpoxy3038	APA	65	77

autoEpoxy3038 and its variants, we could not have identified the system with more than 80 MPa of recovery stress. Given that APA and IPD are both commercially available hardeners, the design optimization of just the epoxy has led us to a system with more than 60% increased recovery stress compared to the best commercially available and experimentally validated system (i.e., DGEBA-IPD).

4. DISCUSSION

4.1. Improvement through the Process. Figure 7 summarizes the performance of our monomer design optimization campaign by showcasing the gradual improvement of the target variable (recovery stress) throughout the process. Our goal was to design monomers that would result in a TSMP with a better recovery stress than that of DGEBA-IPD, the best previously available and experimentally validated TSMP,[41] with a recovery stress of 50 MPa.[37] To that end, we evaluated commercially available monomers as well as other identified yet commercially unavailable monomers. We found a set with a recovery stress of 57 MPa among commercially available ones, and another with a recovery stress of 70 MPa among commercially unavailable ones. We then created a prediction set via modifying existing monomers, trained a ML algorithm on the data set consisting of commercially available and unavailable monomers, and used it to predict the recovery stress of monomers in the prediction set. At each iteration, we evaluated the properties of candidates with the highest predicted recovery stress and added them to the training data set. The first ML iteration did not offer any improvement over the initial data set and suffered from relatively higher prediction inaccuracies. However, incorporation of the best candidates identified in each iteration into the training data set allowed for the data set to encompass more of the search space in each iteration, resulting in an improvement in the accuracy of prediction after each iteration (Figure 5). Following the first ML iteration, all the subsequent steps resulted in continuous improvement in the recovery stress.

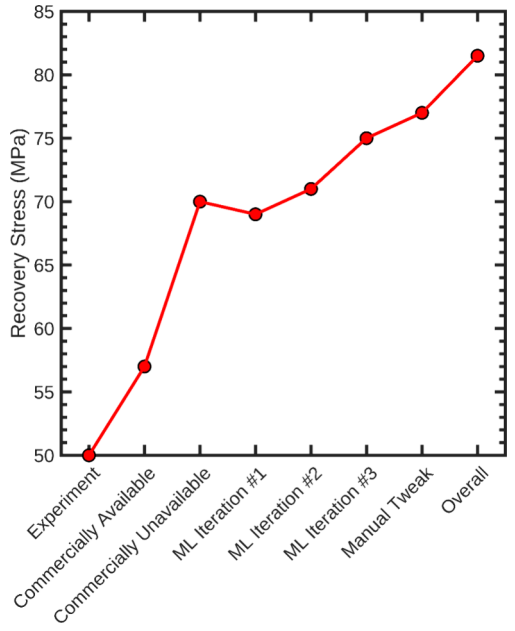


Figure 7. Representation of the best-design-so-far for identification and design of new monomers.

Only 3 ML iterations coupled with manual tweaking of the best identified epoxies resulted in a set with a recovery stress of 77 MPa. Finally, pairing one of the best epoxies identified with a hardener shown to have high recovery stress with DGEBA resulted in the identification of autoEpoxy4694-APA system with a recovery stress of 81.5 MPa (more than 60% greater than that of DGEBA-IPD). It is worth noting that our initial data set had only 67 data points, and throughout the ML study, we only added 24 more data points to the set. Our approach serves as a proof of concept that data-driven approaches to optimization, when coupled with human intuition, can perform reasonably well even with limited data sets. Our approach can be implemented to further optimize the design of the hardener molecule as well as to achieve even better recovery stress, which we intend to do in future studies.

4.2. Fingerprint Importance. The SHapley Additive exPlanations (SHAP)[42−44] method was used to quantify the importance of each of the fingerprints used in the predictive model. SHAP is based on Shapley values from cooperative game theory, which provides a way to attribute the contribution of each fingerprint to the model's predictions. The SHAP value was calculated using a random 80-20 train-set data split 100 times using the default linear regressor in sklearn. Figure S10 shows the average SHAP values for each fingerprint. A higher SHAP indicates a fingerprint is more important when distinguishing between the candidate chemistries. In this case, the number of backbone atoms and the length of the epoxy monomer are the most important fingerprints. It is not obvious that this alone is the most important factor in creating a high recovery stress material, but other factors coinciding with that fingerprint could explain their importance. For example, as already mentioned the para position gave a much higher recovery stress than the ortho, because much stiffer aromatic ring mattered in carrying stress along the backbone. This feature difference would result in a high number of backbone atoms. Furthermore, it was commonly found that the napthalene-like structure found in Figures S4b and S6b often resulted in high recovery stress. Adding more of these to the epoxy also raises nbackbone and the length of the epoxy monomer. Other important features include the size of the heavy atoms in the backbone, the bond stretching spring constant, and the polarity of side chain atoms. Sidechain polarity and backbone atom size are ways to measure atomic composition and differentiate kinds of atoms of the same type (i.e., aromatic vs alkane carbons). A high importance on the polarity of side chain atoms suggests that the functionality of the said sidechains may be especially important, while the importance of Van der Waals size may suggest that the specific functional groups are important, since those different functional groups will have different atoms types or carbon types which are reflected in the Van der Waals size. A high bond stretch coefficient in the backbone would naturally result in a stronger push back upon compression, that is, a higher recovery stress. This analysis provides a hypothesis which could be tested in future works.

4.3. Potential Future Improvements. While we have demonstrated the effectiveness of the fingerprinting model, the

https://doi.org/10.1021/acs.macromol.4c01598
Macromolecules XXXX, XXX, XXX−XXX

generations of new potential chemistries, and identified new and promising shape memory polymers, improvements to the prediction method could result in a more efficient traversal of the design space. Increasingly complex approaches have the potential to create a more efficient process by reducing the number of additional simulations required to optimize the material. While our initial tests concluded that multiple linear regression resulted in the lowest MSE and so was chosen, improved hyper-parameter choices could result in more accurate predictions for other methods, as could an increased data set size. A more in-depth k-fold validation could also be done in case model performance is dependent on specific train/test splits. Beyond improving the regression model, Bayesian optimization,[45,46] which uses Bayesian regression models and then uses decision theory to suggest the next best sample, could be used. This approach would use regression models like Gaussian process regression[47] (though the large number of fingerprints and small data set may limit its efficiency), which shares similarities with kernel ridge regression (KRR) and regression with radial basis functions, and ensemble learning models,[48] which consider the predictions of a number of underlying models, to create the predictive model and more efficiently traverse the compositional space.[45,46] Given the large number of inputs and limited observations, the Randomized Conditional Independence Test (RCIT) in causal discovery could also be a useful method for identifying potential fingerprinting models.[49]

This method of creating new potential chemistries will do a very good job of sampling a specific phase space and generating valid structures with the correct cross-linking sites. However, this approach has limitations. For example, the core of the hardener molecules is limited by the bases selected. Generating these in a manner similar to that done for the epoxies would widen the set of potential chemistries. Similarly, specific common groups were chosen to build the epoxies from. A wider array of potential structures would also widen the phase space, though doing so would potentially require filtering out many structures that are impossible, unstable, or synthetically very challenging.

5. CONCLUSIONS

We created an optimization strategy based on MD simulation and ML to identify monomers that maximized the recovery stress of TSMPs made from diglycidyl ether epoxies and diamine hardeners. The best experimentally tested system is DGEBA-IPD with a computationally measured recovery stress of 50 MPa. We expanded our data set using our chemical intuition and insights obtained from our previous work to modify existing epoxies, as well as using hardeners found via PubChem's 'Find Similar Structures' feature.

Twenty-one atomistic fingerprints were used as features to predict recovery stresses. Using 80% of the data set to train ML models and the rest to test those models, it was found that a multiple linear regression model was suitable for predicting TSMPs with high recovery stress within a prediction set. The best results from each prediction were incorporated into the previous data set to expand the search space further and train the model, which yielded better results than the previous iteration.

The modifications made to obtain the molecules in the prediction were limited, given the fact the only a few fragments and base molecules were considered for the modification of epoxies and hardeners, respectively. Nonetheless, from our predicted set, we managed to identify two epoxies that had very high recovery stresses with IPD hardeners of 75 and 71 MPa, respectively, an almost 50% increase from DGEBA-IPD. This was further improved by using APA as a hardener, which was found to yield high recovery stress with DGEBA, yielding a recovery stress of 81.5 MPa.

In summary, we demonstrated that an optimization approach that combines smart feature engineering, machine learning, and human intuition can effectively be used to identify and/or design new molecules for creating novel polymers with superior properties.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.macromol.4c01598.

Structures and calculated recovery stresses of monomers added to training set, details of prediction set creation, regression model parameters, residual plots and R^2 values of tested models, and SHAP plot and details (PDF)

Complete list of the initial training set monomers, their calculated recovery stresses, glass transition temperatures, and their fingerprints (XLSX)

Complete list of symmetric epoxies generated and their fingerprints (XLSX)

Complete list of asymmetric epoxies generated and their fingerprints (XLSX)

Complete list of hardeners generated by moving amines and their fingerprints (XLSX)

Complete list of hardeners generated by moving sidechains and their fingerprints (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Andrew J. Peters − *College of Engineering & Science, Louisiana Tech University, Ruston, Louisiana 71270, United States;* orcid.org/0000-0001-5031-2828; Phone: (318) 257-511; Email: apeters@latech.edu

Authors

Anwar Shafe − *College of Engineering & Science, Louisiana Tech University, Ruston, Louisiana 71270, United States*

Pouria Nourian − *College of Engineering & Science, Louisiana Tech University, Ruston, Louisiana 71270, United States*

Xiyuan Liu − *College of Engineering & Science, Louisiana Tech University, Ruston, Louisiana 71270, United States*

Guoqiang Li − *Mechanical & Industrial Engineering Dept., Louisiana State University, Baton Rouge, Louisiana 70803, United States;* orcid.org/0000-0002-7004-6659

Collin D. Wick − *College of Engineering & Science, Louisiana Tech University, Ruston, Louisiana 71270, United States;* orcid.org/0000-0002-0261-0780

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.macromol.4c01598

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the US National Science Foundation under grant number OIA-1946231 and the

Louisiana Board of Regents for the Louisiana Materials Design Alliance (LAMDA). The high-performance computing resources provided by the Louisiana Optical Network Infrastructure (https://loni.org) were used for this work.

REFERENCES

(1) Hager, M. D.; Bode, S.; Weber, C.; Schubert, U. S. Shape Memory Polymers: Past, Present and Future Developments. *Prog. Polym. Sci.* **2015**, *49*−*50*, 3−33.
(2) Zhao, Q.; Qi, H. J.; Xie, T. Recent Progress in Shape Memory Polymer: New Behavior, Enabling Materials, and Mechanistic Understanding. *Prog. Polym. Sci.* **2015**, *49*−*50*, 79−120.
(3) Berg, G. J.; McBride, M. K.; Wang, C.; Bowman, C. N. New Directions in the Chemistry of Shape Memory Polymers. *Polymer* **2014**, *55* (23), 5849−5872.
(4) Lee, K. M.; Koerner, H.; Vaia, R. A.; Bunning, T. J.; White, T. J. Light-Activated Shape Memory of Glassy, Azobenzene Liquid Crystalline Polymer Networks. *Soft Matter* **2011**, *7* (9), 4318.
(5) Rokaya, D.; Skallevold, H. E.; Srimaneepong, V.; Marya, A.; Shah, P. K.; Khurshid, Z.; Zafar, M. S.; Sapkota, J. Shape Memory Polymeric Materials for Biomedical Applications: An Update. *J. Compos. Sci.* **2023**, *7* (1), 24.
(6) Chen, L.; Wei, X.; Wang, F.; Jian, S.; Yang, W.; Ma, C.; Duan, G.; Jiang, S. In-Situ Polymerization for Mechanical Strong Composite Actuators Based on Anisotropic Wood and Thermoresponsive Polymer. *Chin. Chem. Lett.* **2022**, *33* (5), 2635−2638.
(7) Gu, J. F.; Gorgutsa, S.; Skorobogatiy, M. Soft Capacitor Fibers Using Conductive Polymers for Electronic Textiles. *Smart Mater. Struct.* **2010**, *19* (11), No. 115006.
(8) Liu, Y.; Du, H.; Liu, L.; Leng, J. Shape Memory Polymers and Their Composites in Aerospace Applications: A Review. *Smart Mater. Struct.* **2014**, *23* (2), No. 023001.
(9) Li, F.; Liu, Y.; Leng, J. Progress of Shape Memory Polymers and Their Composites in Aerospace Applications. *Smart Mater. Struct.* **2019**, *28* (10), 103003.
(10) Govindarajan, T.; Shandas, R. A Survey of Surface Modification Techniques for Next-Generation Shape Memory Polymer Stent Devices. *Polymers (Basel).* **2014**, *6* (9), 2309−2331.
(11) Zainal, M. A.; Ahmad, A.; Mohamed Ali, M. S. Frequency-Controlled Wireless Shape Memory Polymer Microactuator for Drug Delivery Application. *Biomed. Microdevices* **2017**, *19* (1), 8.
(12) Wischke, C.; Neffe, A. T.; Steuer, S.; Lendlein, A. Evaluation of a Degradable Shape-Memory Polymer Network as Matrix for Controlled Drug Release. *J. Controlled Release* **2009**, *138* (3), 243−250.
(13) Pfau, M. R.; Grunlan, M. A. Smart Scaffolds: Shape Memory Polymers (SMPs) in Tissue Engineering. *J. Mater. Chem. B* **2021**, *9* (21), 4287−4297.
(14) Xie, M.; Wang, L.; Ge, J.; Guo, B.; Ma, P. X. Strong Electroactive Biodegradable Shape Memory Polymer Networks Based on Star-Shaped Polylactide and Aniline Trimer for Bone Tissue Engineering. *ACS Appl. Mater. & Interfaces* **2015**, *7* (12), 6772−6781.
(15) Leng, J.; Lu, H.; Liu, Y.; Huang, W. M.; Du, S. Shape-Memory Polymers—A Class of Novel Smart Materials. *MRS Bull.* **2009**, *34* (11), 848−855.
(16) Imran Khan, M.; Zagho, M. M.; Shakoor, R. A. A Brief Overview of Shape Memory Effect in Thermoplastic Polymers. In *Springer Series on Polymer and Composite Materials*; Springer International Publishing: 2017; pp 281−301.
(17) Nourian, P.; Wick, C. D.; Li, G.; Peters, A. J. Correlation between Cyclic Topology and Shape Memory Properties of an Amine-Based Thermoset Shape Memory Polymer: A Coarse-Grained Molecular Dynamics Study. *Smart Mater. Struct.* **2022**, *31* (10), 105014.
(18) Nourian, P.; Wick, C. D.; Peters, A. J. Effect of Crosslinking Fraction, Hardener Functionality and Topological Quality on Stress Recovery of Thermoset Shape Memory Polymers: A Coarse-Grained Molecular Dynamics Study. *Smart Mater. Struct.* **2023**, *32* (11), 115001.
(19) Mannodi-Kanakkithodi, A.; Chandrasekaran, A.; Kim, C.; Huan, T. D.; Pilania, G.; Botu, V.; Ramprasad, R. Scoping the Polymer Genome: A Roadmap for Rational Polymer Dielectrics Design and Beyond. *Mater. Today* **2018**, *21* (7), 785−796.
(20) Mannodi-Kanakkithodi, A.; Huan, T. D.; Ramprasad, R. Mining Materials Design Rules from Data: The Example of Polymer Dielectrics. *Chem. Mater.* **2017**, *29* (21), 9001−9010.
(21) Mannodi-Kanakkithodi, A.; Pilania, G.; Huan, T. D.; Lookman, T.; Ramprasad, R. Machine Learning Strategy for Accelerated Design of Polymer Dielectrics. *Sci. Rep.* **2016**, *6* (1), 20952.
(22) Yu, L.; Zunger, A. Identification of Potential Photovoltaic Absorbers Based on First-Principles Spectroscopic Screening of Materials. *Phys. Rev. Lett.* **2012**, *108* (6), No. 068701.
(23) Armiento, R.; Kozinsky, B.; Fornari, M.; Ceder, G. Screening for High-Performance Piezoelectrics Using High-Throughput Density Functional Theory. *Phys. Rev. B* **2011**, *84* (1), No. 014103.
(24) Lin, L.-C.; Berger, A. H.; Martin, R. L.; Kim, J.; Swisher, J. A.; Jariwala, K.; Rycroft, C. H.; Bhown, A. S.; Deem, M. W.; Haranczyk, M.; et al. In Silico Screening of Carbon-Capture Materials. *Nat. Mater.* **2012**, *11* (7), 633−641.
(25) Yan, C.; Feng, X.; Wick, C.; Peters, A.; Li, G. Machine Learning Assisted Discovery of New Thermoset Shape Memory Polymers Based on a Small Training Dataset. *Polymer* **2021**, *241*, No. 123351.
(26) Yan, C.; Feng, X.; Li, G. From Drug Molecules to Thermoset Shape Memory Polymers: A Machine Learning Approach. *ACS Appl. Mater. & Interfaces* **2021**, *13* (50), 60508−60521.
(27) Segura Ibarra, D.; Mathews, J.; Li, F.; Lu, H.; Li, G.; Chen, J. Deep Learning for Predicting the Thermomechanical Behavior of Shape Memory Polymers. *Polymer* **2022**, *261*, No. 125395.
(28) 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* (31), 17575−17585.
(29) Shafe, A.; Wick, C. D.; Peters, A. J.; Liu, X.; Li, G. Effect of Atomistic Fingerprints on Thermomechanical Properties of Epoxy-Diamine Thermoset Shape Memory Polymers. *Polymer* **2022**, *242*, No. 124577.
(30) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in 't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; et al. LAMMPS - a Flexible Simulation Tool for Particle-Based Materials Modeling at the Atomic, Meso, and Continuum Scales. *Comput. Phys. Commun.* **2022**, *271*, No. 108171.
(31) Jorgensen, W. L.; Tirado-Rives, J. Potential Energy Functions for Atomic-Level Simulations of Water and Organic and Biomolecular Systems. *Proc. Natl. Acad. Sci.* **2005**, *102* (19), 6665−6670.
(32) Dodda, L. S.; Vilseck, J. Z.; Tirado-Rives, J.; Jorgensen, W. L. 1.14* CM1A-LBCC: Localized Bond-Charge Corrected CM1A Charges for Condensed-Phase Simulations. *J. Phys. Chem. B* **2017**, *121* (15), 3864−3870.
(33) Dodda, L. S.; De Vaca, I. C.; Tirado-Rives, J.; Jorgensen, W. L. LigParGen Web Server: An Automatic OPLS-AA Parameter Generator for Organic Ligands. *Nucleic Acids Res.* **2017**, *45* (W1), W331−W336.
(34) Jorgensen, W. L.; Maxwell, D. S.; Tirado-rives, J.; Haven, N. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118* (45), 11225−11236.
(35) Jorgensen, W. L.; Madura, J. D.; Swenson, C. J. Optimized Intermolecular Potential Functions for Liquid Hydrocarbons. *J. Am. Chem. Soc.* **1984**, *106* (22), 6638−6646.
(36) Jorgensen, W. L.; Tirado-Rives, J. The OPLS [Optimized Potentials for Liquid Simulations] Potential Functions for Proteins, Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* **1988**, *110* (6), 1657−1666.
(37) Wick, C. D.; Peters, A. J.; Li, G. Quantifying the Contributions of Energy Storage in a Thermoset Shape Memory Polymer with High

https://doi.org/10.1021/acs.macromol.4c01598
Macromolecules XXXX, XXX, XXX−XXX

Stress Recovery: A Molecular Dynamics Study. *Polymer* **2021**, *213*, No. 123319.
(38) Ng, F.; Couture, G.; Philippe, C.; Boutevin, B.; Caillol, S. Bio-Based Aromatic Epoxy Monomers for Thermoset Materials. *Molecules* **2017**, *22* (1), 149.
(39) *PubChem Structure Search*. https://pubchem.ncbi.nlm.nih.gov/search/search.cgi.
(40) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12* (85), 2825−2830.
(41) Fan, J.; Li, G. High Enthalpy Storage Thermoset Network with Giant Stress and Energy Output in Rubbery State. *Nat. Commun.* **2018**, *9* (1), 642.
(42) Lundberg, S. M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17*; Curran Associates Inc.: Red Hook, NY, USA, 2017; pp 4768−4777.
(43) Shapley, L. S.; Arrow, K. J.; Barankin, E. W.; Blackwell, D.; Bott, R.; Dalkey, N.; Dresher, M.; Gale, D.; Gillies, D. B.; Glicksberg, I.;et al.A Value for N-Person Games. In *Contributions to the Theory of Games (AM-28), Volume II*; Princeton University Press: 1953; pp 307−318.
(44) Shapley, L. S.; Arrow, K. J.; Barankin, E. W.; Blackwell, D.; Bott, R.; Dalkey, N.; Dresher, M.; Gale, D.; Gillies, D. B.; Glicksberg, I.;et al.Quota Solutions of N-Person Games. In *Contributions to the Theory of Games (AM-28), Volume II*; Princeton University Press: 1953; pp 343−360.
(45) Frazier, P. I.; Wang, J. *Bayesian Optimization for Materials Design BT - Information Science for Materials Discovery and Design*; Lookman, T.; Alexander, F. J.; Rajan, K., Eds.; Springer International Publishing: Cham, 2016; pp 45−75.
(46) Packwood, D. *Overview of Bayesian Optimization in Materials Science BT - Bayesian Optimization for Materials Science*; Packwood, D., Ed.; Springer Singapore: Singapore, 2017; pp 1−10.
(47) Deringer, V. L.; Bartók, A. P.; Bernstein, N.; Wilkins, D. M.; Ceriotti, M.; Csányi, G. Gaussian Process Regression for Materials and Molecules. *Chem. Rev.* **2021**, *121* (16), 10073−10141.
(48) Polikar, R. *Ensemble Learning BT - Ensemble Machine Learning: Methods and Applications*; Zhang, C.; Ma, Y., Eds.; Springer New York: New York, NY, 2012; pp 1−34.
(49) Strobl, E. V.; Zhang, K.; Visweswaran, S. Approximate Kernel-Based Conditional Independence Tests for Fast Non-Parametric Causal Discovery. *J. Causal Infer.* **2019**, *7* (1), 20180017.

https://doi.org/10.1021/acs.macromol.4c01598
Macromolecules XXXX, XXX, XXX−XXX