Modelling and
Simulation in Materials
Science and
Engineering

PAPER • **OPEN ACCESS**

Feature-based prediction of properties of cross-linked epoxy polymers by molecular dynamics and machine learning techniques

To cite this article: Sindu B S and Jan Hamaekers 2025 *Modelling Simul. Mater. Sci. Eng.* **33** 065010

OPEN ACCESS
IOP Publishing
Modelling and Simulation in Materials Science and Engineering
Modelling Simul. Mater. Sci. Eng. **33** (2025) 065010 (18pp) https://doi.org/10.1088/1361-651X/adf56c

Feature-based prediction of properties of cross-linked epoxy polymers by molecular dynamics and machine learning techniques

Sindu B S[1] **and Jan Hamaekers**[2,*]

[1] Special and Multifunctional Structures Laboratory, CSIR-Structural Engineering Research Centre, Taramani, Chennai 600113, Tamil Nadu, India
[2] Fraunhofer Institute for Algorithms and Scientific Computing, Schloss Birlinghoven, D-53757 Sankt Augustin, Germany

E-mail: jan.hamaekers@scai.fraunhofer.de

Received 9 May 2025; revised 27 June 2025
Accepted for publication 29 July 2025
Published 11 August 2025

Abstract

The conventional way of developing epoxy polymers may not be fully efficient as it is constrained by fixed compositions that limit performance. To overcome this, we introduce a machine learning (ML)-based approach that accurately predicts mechanical properties from its basic structural features, enabling broader design exploration. The results from molecular dynamics simulations have been used to derive the ML model. The salient feature of our work is that for the development of epoxy polymers based on EPON-862, several new hardeners were explored in addition to the conventionally used ones. The influence of additional parameters like the proportion of curing agent used and the extent of curing on the mechanical properties of epoxy polymers were also investigated. This method can be further extended by providing the epoxy polymer with the desired properties through knowledge of the structural characteristics of its constituents. The findings of our study can thus lead toward development of efficient design methodologies for epoxy polymeric systems.

Keywords: epoxy polymers, mechanical properties, molecular dynamics, machine learning, feature importance

* Author to whom any correspondence should be addressed.

1. Introduction

Thermoset epoxy polymers are widely used as matrix/adhesive material in structural, aeronautical and marine applications. They are formed due to reaction between two base monomers (resin and hardener) and the reaction leads to formation of a complex three-dimensional network. Thus, the three-dimensional network and the properties of the resulting polymer depend upon several factors like the type of base monomers used, their composition, parameters under which curing is carried out, extent of curing and the like. Over several decades, the design of epoxy polymers has been carried out using conventional experiments with preset resin/hardener combination and curing process which has resulted in the epoxy polymer space with limited performance characteristics. Hence, in order to expand the epoxy polymer space with high and multi-functional performance characteristics for futuristic applications, advanced techniques should be used to understand the nature/interactions of epoxy polymers at fundamental length scale and identify their correlation with properties/performance at the application length scale.

For this purpose, molecular dynamics (MDs) approaches were used, in which the reactions (cross-linking processes) were virtually simulated with different resin/hardener combinations at different curing levels, and their thermo-mechanical properties such as elastic modulus, density, and glass transition temperature (T_g) were determined [1–3]. MD simulations with the *ReaxFF* potential [4] were used to simulate the bond dissociation and reorganization phenomena, thus enabling understanding of plastic deformation and fracture process as well as predicting the gel-point, volumetric shrinkage and yield strength [5, 6]. MD simulations were also used to determine the mechanisms of moisture diffusion, thermal conductivity and hygrothermal degradation in cross-linked epoxy polymers by investigating the inter/intra-molecular interactions (radial distribution function, polar interactions, hydrogen bonding activity, dipole moment fluctuations) [7–9]. Moreover, MD simulations were carried out to gain insights on mechanisms of water/ion migration at the composite interfaces (epoxy-cement, epoxy-fiber) as in the case of chemical attack and to evaluate their influence on the inter-facial integrity [10, 11]. Furthermore, it was demonstrated that MD can also be used for design, property prediction and understanding interactions between various constituents in nano-engineered/smart/multifunctional epoxy polymer systems [12–15]. While MD helps in simulating polymer systems with new combinations and new process parameters, and also contributes to the understanding of the fundamental phenomena that occur in various macroscopic processes, the method is limited because it is time consuming and computationally intensive.

In recent times, machine learning (ML) techniques were also applied for design, optimization and property prediction of epoxy polymers. For example, ML based active learning and Bayesian optimization were used to predict and optimize the process parameters for development of epoxy polymers with high adhesive strength. In particular, this approach lead towards significantly less experimental trials (32 as against 256 possible combinations) to design epoxy polymer with very high adhesive strength [16]. A similar approach was also used for the design of multi-component, bio-based epoxy polymers with high and low T_g [17]. In addition, artificial neural networks were used to develop shape memory epoxy polymers (SMPs) with optimized performance characteristics (T_g, flexural strength, strain fixation rate, and strain recovery rate), where the input of the ML model are the constituents of the epoxy polymer and the output is the polymer property [18]. A method involving Raman spectroscopy in conjunction with ML techniques (random forest and partial least square regression) was used to determine the optimum molar composition of epoxy polymers which is one of the most determining factor that dictates the performance and properties of epoxy polymers [19]. ML based technique was also applied to predict and quantify the effects of environmental factors

(thermal exposure—different temperatures and time periods) on the degradation in properties of epoxy polymers [20]. Recently, BigSMILES notation (string based representation of molecules which can also support stochastic molecular structures) was employed as an input for convolutional neural network (CNN) models for polymer property prediction. A CNN model was developed which can predict both the T_g and recovery stress of SMP. The model was further used to design new SMPs which had higher recovery stress than the known ones [21]. ML also proved to be a powerful tool in understanding correlations between fundamental structural features/genes and properties of epoxy polymers. ML ensemble model was used to predict the T_g of epoxy polymers with different resin/hardener combinations (16 resins and 19 hardeners) using the molecular descriptors determined from Mordred [22] and RDKit [23] in combination with principal component analysis. Using this approach, it has been identified that T_g is expected to increase for smaller, less polarizable and more hydrophobic resins [24]. A ML assisted materials genome approach combining concepts of graph convolutional networks (attention- and gate-augmented), classical gel theory and transfer learning was used for design of epoxy polymers with high mechanical properties (elastic modulus, tensile strength and toughness) based on experimental results collected for epoxy polymers and polyamides from which the gene structures affecting these properties were also identified [25]. While ML is promising for designing new polymers with unusual properties and understanding relationships that are not possible with conventional techniques, to create an accurate model, it needs to be trained with more data, which in the case of epoxy polymers is limited.

In order to overcome the limitations of individual techniques and to utilize their benefits, in recent times, appropriate combinations of these two methods are being used, wherein, MD is used to simulate new polymeric systems with different blends and process parameters, the results from MD is used for developing ML models to get deeper insights. For example, MD was used to simulate epoxy polymers (30 different types) with different resin/hardener/toughening agent combinations including different proportions and their results were further used to develop ML model for optimizing the constituents to design high performance epoxy polymer with high elastic modulus, ultimate tensile strength, density and T_g [26]. Similarly, MD was used in conjunction with ML for predicting the properties of self-healing epoxy polymers and for optimizing their constituents [27]. These methods were further extended to determine the correlation between constituent and polymer property using Pearson coefficient to identify the constituent responsible for each of the property (cohesive energy density, elastic modulus and T_g) [28]. Coarse grained simulations and ML based methods were used to predict the viscosity of different epoxy polymer clusters (categorized using K-means clustering) with different diluent proportions and at various temperatures [29]. While the combination of both methods seem more promising, very limited studies are carried out using these techniques together and are mostly limited to composition optimization with a preset resin/hardener combination or identifying the constituent-property relationship within the limited blend combinations inspite of wide range of possibilities. It was demonstrated that, shape memory polymers with improved performance can be developed from manually designed monomers by extending these techniques together with feature engineering [30].

With this in mind, this study will use MD simulations to simulate a wide range of polymers by extending the conventional blend combinations and employ ML techniques to gain deeper insights into the parameters affecting the mechanical properties of these epoxy polymers. We generate several epoxy polymer systems with different types of hardeners (aromatic rings/aliphatic rings/aliphatic chains), varying proportions of hardeners and also from partially cured to maximum cured systems, see sections 2.1 and 2.2. Then, subject these generated systems to strain simulations, see section 2.3 and from their stress–strain response, key mechanical properties like yield strength and elastic modulus are determined, see section 3. Further, ML

based techniques are employed to predict the properties of different epoxy polymer systems wherein the structural features of the hardeners (since same resin is used for all cases, the features of resin was not considered) are used as input, see section 4. Finally, the structural feature-property correlation is established for thermoset epoxy polymers. The findings of this study would enable efficient design of epoxy polymers with desired mechanical properties by merely selecting constituents with relevant structural features. This study has also paved the way for expanding the epoxy polymer space with more types of combinations in addition to the conventional ones.

2. MDs simulations

In this study, MD simulations are carried out on different epoxy systems formed by atomistic cross-linking procedure. All the epoxy systems have same resin, Diglycidyl Ether of Bisphenol-F (EPON 862) (as shown in figure 1) but different hardeners. A total of 11 amine-based hardeners (as shown in table 1) is considered with different structural configurations (aromatic amines/aliphatic amines, primary/secondary amines). All simulations are carried out using a commercial software (QuantumATK) [31] with OPLS-AA (optimized potentials for liquid simulations) potential [32] built in Tremolo-X Calculator [31]. The functional form of OPLS-AA potential is presented in equation (1), in which the total potential energy of the system is the summation of contributions due to bonded (bond stretching, angle bending, torsional and inversion terms) and non-bonded (Lennard–Jones and electrostatic) interactions,

$$\begin{aligned} E(x) = & \sum_{i}^{\text{bonds}} k_i (r_i - r_0)^2 + \sum_{i}^{\text{angles}} k_i (\theta_i - \theta_0)^2 \\ & + \sum_{i}^{\text{torsions}} \sum_{n=1}^{3} k_{i,n} (1 - \cos(n\phi_i - \delta_{i,n})) + \sum_{i}^{\text{inversion}} k_i (1 - \cos(n\chi_i)) \\ & + \sum_{ij}^{\text{atoms}} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \\ & + \sum_{ij}^{\text{atoms}} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}, \end{aligned} \tag{1}$$

where r, θ, ϕ and χ denote bond distances, angles, torsions and inversions respectively.

2.1. Construction of simulation box

Firstly, the molecular configuration of monomers (resin/hardener) is generated using SMILES (simplified molecular-input line-entry system) [33]. The monomers of resin and hardener are then randomly inserted into the simulation box of size $65 \times 65 \times 65$ Å^3 (number of atoms $\approx$ 15 000 atoms) with a packing density of 0.6 g cc^{-1} using Packmol tool [34]. The number of resin and hardener molecules to be inserted into the simulation box is calculated using the stoichiometric ratio, molecular mass of the monomers and the packing density. The tolerance (intermolecular distance) and buffer (to the cell vectors) during packing is considered as 2 Å [34]. The packed simulation box is then subjected to geometry optimization using LBFGS algorithm with a force tolerance of 0.05 eV Å^{-1}. The optimization is carried out in 500 steps with a step size of 0.2 Å. The optimized simulation box is further equilibrated in three stages

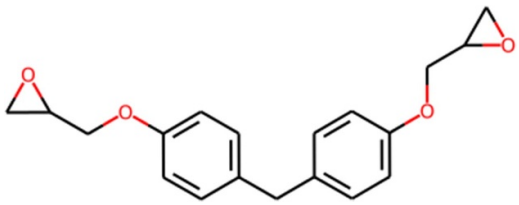
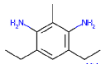
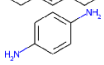
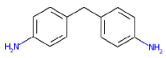
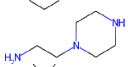
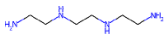
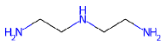
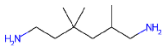
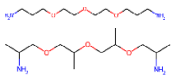


Figure 1. Epoxy resin considered in this study (Diglycidyl Ether of Bisphenol-F (EPON 862)).

Table 1. Hardeners considered in this study.

Hardener	Molecular conf.	Type
Diethyl toluene diamine (DETDA)		Aromatic ring
Phenylene diamine (PPD)		
Diamino diphenyl methane (DDM)		
Diamino cyclohexane (DACH)		Aliphatic ring
Aminoethyl piperazine (PIPER)		
Isophorone diamine (IPDA)		
Triethylene tetramine (TETA)		Aliphatic chain
Diethylene triamine (DETA)		
Trimethyl hexamethylene diamine (TMD)		
Trioxa tridecane diamine(TTD)		
Polyoxypropylene diamine (Jeffamine)		

(NVT @300 K, NPT @1 bar, 300 K and NPT @1 bar, 480 K),during which the temperature and pressure parameters are varied gradually to ensure smooth equilibration. All simulations are carried out with a timestep of 1 fs. Berendsen thermostat and barostat are used in the first two stages of equilibration with 50 000 steps while Martyna–Tobias–Klein barostat is used in the last stage with 100 000 steps where high temperature NPT simulations are taking place.

2.2. Cross-linking procedure

The equilibrated structure is then subjected to a cross-linking process. During this process, reactions are simulated between the epoxide and amine (aromatic/aliphatic) groups in the resin and hardener respectively. More details about the cross-linking process can be found elsewhere [8, 35–37]. Bonds are formed between these reaction groups when they are at a certain distance. The bond search radius is gradually increased from 5 Å to 10 Å with a step

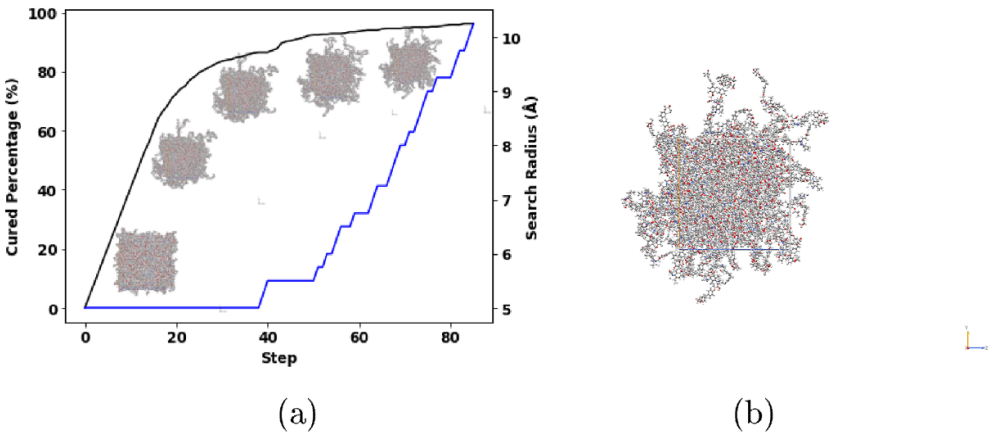


Figure 2. (a) Evolution of the cross-linking process (inscribed figures shows the simulation box at different degrees of curing). (b) Fully cross-linked structure.

size of 0.25 Å. During each cycle, the system is equilibrated using NVT (10 000 steps) followed by NPT simulations (10 000 steps). The evolution of the curing process along with the cross-linked simulation systems is presented in figure 2. Since high temperature leads to better cross-linking degree, the whole process is simulated at 480 K. The simulation box with the cross-linked structure is further equilibrated to bring the polymeric system to the correct density (1.13 g cc^{-1}). As the relaxation of polymeric system usually takes place at a longer timescales, the process is accelerated using 21-step equilibration process (as suggested by [38]) using cycles of high temperature and high pressure simulations and annealing until the desired temperature (300 K) and pressure (1 bar) is reached.

2.3. Stress–strain simulations

The final equilibrated structure is then used for the production stage where the actual simulations are performed to obtain the material properties. Strain is applied to the simulation box at a constant rate and the corresponding stress is measured and from the stress–strain relationship, properties like elastic modulus and yield strength are determined. Since epoxy polymers are isotropic materials, the strain simulation is carried out in all three directions (as shown in figure 3 and the average value is taken as the computed material property. The elastic modulus is calculated from the slope of the stress–strain curve when the material is still in elastic stage [39, 40] (a strain level of 0.025). The yield strength is taken as the stress at which the slope of the stress–strain curve changes to negative. In order to get the properties, five different samples (structures) are generated for each case and the results are averaged (from all fives cases and in each case three different directions).

3. Mechanical properties of epoxy polymers from MD simulations

The mechanical properties of epoxy polymers determined from MD simulations are compared with those of the experimental results reported in the literature [6, 37, 41–52]. Since strain rate plays a crucial role on the obtained results, simulations (of a particular combination, DGEBF-DETDA) are carried out with five different strain rates (0.005 ps^{-1}, 0.0015 ps^{-1}, 0.0005 ps^{-1},

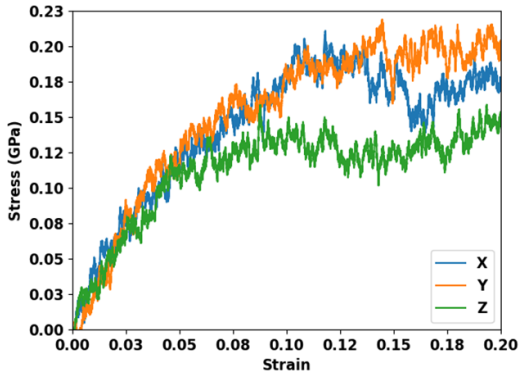


Figure 3. Stress–strain response of DGEBF-DETDA obtained from strain simulations in three orthogonal directions (X), (Y), (Z).

0.000 15 ps^{-1}, 0.000 05 ps^{-1}). The elastic modulus of DGEBF-DETDA system from different strain rate simulations is presented in figure 4(a) where it can be found that a linear relationship exists between the elastic modulus and strain rate. In general, the strain rates used in MD simulations are very high in comparison to that used in actual experiments (0.001 s^{-1}). The comparisons of elastic modulus and yield strength obtained from MD simulations and those reported in experiments are presented in figures 4(b) and (c) respectively and in table 2. It can be observed that in case of elastic modulus, except for the case with the strain rate of 0.005 ps^{-1}, all values from simulation fall within the values (experimental and computational) reported in literature. Simulations with lower strain rates (0.000 15 ps^{-1}, 0.000 05 ps^{-1}) typically require longer simulation time until yielding takes place (to determine yield strength). Hence, in order to strike a balance between accuracy and computational demand, a strain rate of 0.0005 ps^{-1} is used for further simulations. The yield strength obtained from the present study also falls well within the range of values reported (from MD simulations) in literature. However, the yield strength obtained from MD simulations is found to be higher (3 times) than the values reported in literature from experimental investigations. This phenomenon is also found to be logical in the sense that, in usual, the strength is strongly dependent on strain rate, and higher the strain rate, the strength is over-predicted. Further, the influence of additional parameters like the degree of curing and the type of hardener used on the mechanical properties like elastic modulus and yield strength is determined.

3.1. Influence of stoichiometric ratio and curing percentage

Since epoxy polymers are formed by reaction between resin and hardener, inclusion of the right quantity of resin and hardener is very important for proper reaction to take place. Each primary amine ($–NH_2$) of the hardener can react with two epoxide groups and each secondary amine (–NH) can react with one epoxide group of the resin. So the ratio of resin and hardener used for preparation of epoxy polymers should be such that the mix contains right proportion of amine/epoxide groups (stoichiometric ratio $= 1$). If less quantity of resin is present in the mix, the stoichiometric ratio is less than one and if more quantity of resin is present, the stoichiometric ratio is greater than one. The stoichiometric ratio greatly affects the curing percentage (reaction percentage) of the final product, epoxy system. The maximum curing percentage achieved in different epoxy polymer systems with stoichiometric ratio ranging from 0.4 to 1.7 is presented in figure 5(a). The curing percentage, in turn, affects the mechanical properties of

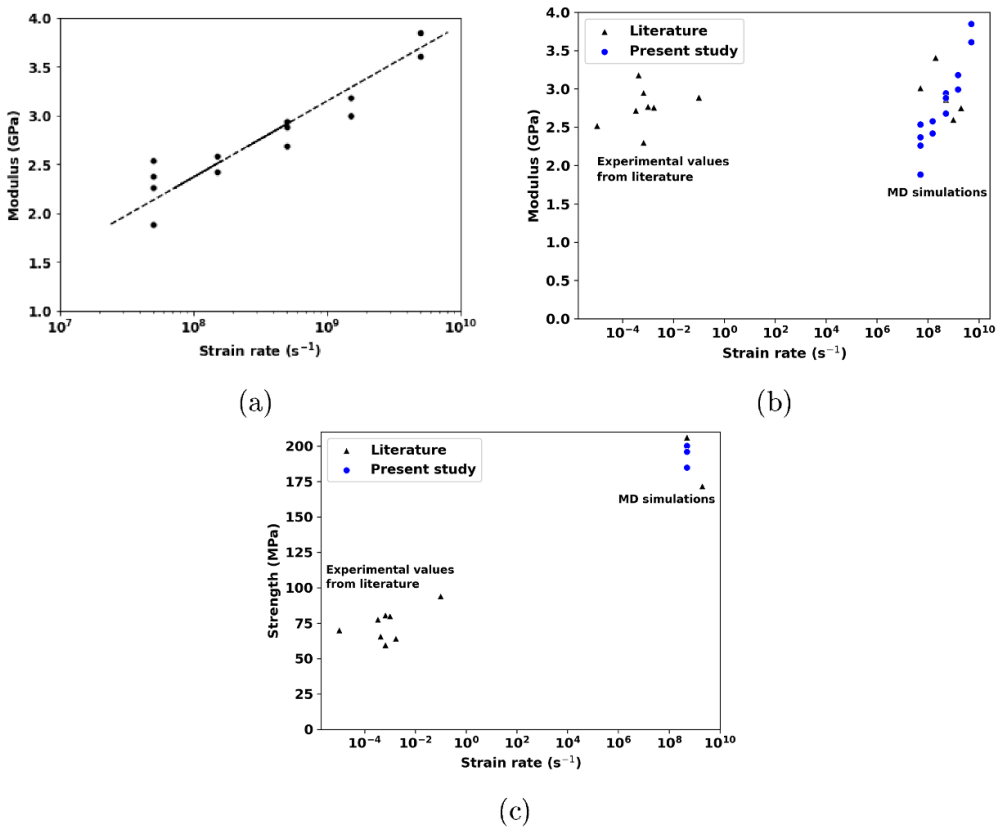


Figure 4. (a) Influence of strain rate on MD results. MD results compared with experimental results from literature [6, 37, 41–50]: (b) elastic modulus. (c) Yield strength.

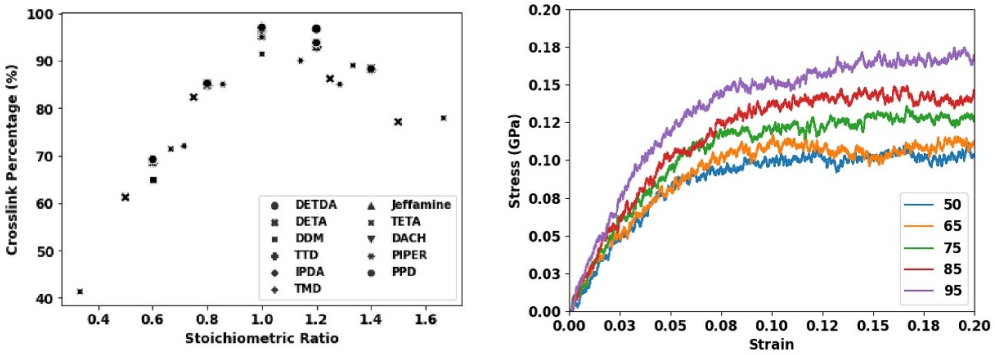
epoxy polymers. The stress–strain response of DGEBF-DETDA at different levels of curing is presented in figure 5(b). It can be observed that both the elastic modulus and yield strength are affected by the level of curing. The results of all epoxy combinations considered in this study at different levels of curing are presented in figures 6(a) and (b), respectively. It can be observed that the elastic modulus of the considered epoxy polymer systems is found to decrease by 15%–31% when the curing percentage is only 50% as against the fully cured system; however, the yield strength is found to decrease by 26%–36%.

3.2. Influence of the type of hardener

The stress–strain response of fully cured (95%) epoxy polymers pertaining to different hardener classes, i.e. hardeners with aromatic rings, aliphatic rings, and aliphatic chains are presented in figures 7(a)–(c) respectively. In the case of hardeners with aromatic rings (figure 7(a)), DDM is found to possess lower mechanical properties compared to DETDA and PPD. A simple visual correlation of the atomic structure (presented in table 1) reveals that DDM possesses two aromatic rings, while DETDA and PPD possess only one aromatic ring. In the case of hardeners with aliphatic rings (figure 7(b)), DACH and IPDA are found to perform better than PIPER. The visual interpretation of their atomic structures reveals that,

Table 2. Validation of mechanical properties, i.e. elastic modulus (EM) and tensile yield strength (TS), with experimental results, where PW stands for our present work.

Epoxy polymer	Strain rate	EM (GPa)	TS (MPa)	Method	References
DGEBF-DETDA	$5.08\,\mathrm{mm\,min^{-1}}$	2.76 ± 0.01	64.1 ± 5.6	Expt.	[41]
DGEBF-DETDA	$2\,\mathrm{mm\,min^{-1}}$	2.95 ± 0.02	80.6 ± 3.1	Expt.	[42]
DGEBF-DETDA	$1.27\,\mathrm{mm\,min^{-1}}$	3.18 ± 0.12	65.63 ± 7.37	Expt.	[43]
DGEBF-DETDA	$1 \times 10^{-5}\,\mathrm{s^{-1}}$	2.52	70	Expt.	[44]
DGEBF-DETDA	$1 \times 10^{-3}\,\mathrm{s^{-1}}$	2.77	80	Expt.	[44]
DGEBF-DETDA	$1 \times 10^{-1}\,\mathrm{s^{-1}}$	2.89	94	Expt.	[44]
DGEBF-DETDA	$1\,\mathrm{mm\,min^{-1}}$	2.72 ± 0.04	77.6 ± 0.9	Expt.	[45]
DGEBF-DETDA	$2\,\mathrm{mm\,min^{-1}}$	2.3	59.7	Expt.	[46]
DGEBF-DETDA	$5 \times 10^{8}\,\mathrm{s^{-1}}$	2.86	206.16	MD	[47]
DGEBF-DETDA	$5 \times 10^{5}\,\mathrm{s^{-1}}$	4.60	—	MD	[48]
DGEBF-DETDA	$5 \times 10^{7}\,\mathrm{s^{-1}}$	3.01	—	MD	[49]
DGEBF-DETDA	$2 \times 10^{9}\,\mathrm{s^{-1}}$	2.75	171.72	MD	[50]
DGEBF-DETDA	$1 \times 10^{9}\,\mathrm{s^{-1}}$	2.60	—	MD	[37]
DGEBF-DETDA	$2 \times 10^{8}\,\mathrm{s^{-1}}$	3.41	—	MD	[6]
DGEBF-DETDA	$5 \times 10^{8}\,\mathrm{s^{-1}}$	2.83 ± 0.14	604.54 ± 12.24	MD	PW
DGEBF-DETA	$5.9 \times 10^{-5}\,\mathrm{s^{-1}}$	2.97	72	Expt.	[51]
DGEBF-DETA	$4.93 \times 10^{-4}\,\mathrm{s^{-1}}$	3.08	81	Expt.	[51]
DGEBF-DETA	$0.5\,\mathrm{mm\,min^{-1}}$	2.29 ± 0.07	62 ± 6	Expt.	[52]
DGEBF-DETA	$5 \times 10^{8}\,\mathrm{s^{-1}}$	2.69 ± 0.18	621.26 ± 12.45	MD	PW



(a) (b)

Figure 5. (a) Influence of stoichiometric ratio on the final cross-linked percentage of different epoxy polymer systems. (b) Influence of curing percentage on the stress–strain response of epoxy polymers (a typical case of DGEBF-DETDA presented).

while DACH and IPDA possess two primary amines, PIPER possess only one primary and one secondary amine that are accessible for reaction. Similarly, in the case of hardeners with aliphatic chains (figure 7(c)), TETA and DETA perform better than the other three hardeners, which can also be attributed to the presence of additional secondary amines in these two cases. From the above findings, it can be concluded that there exists a correlation between the atomic structure/feature with its mechanical properties. Hence, we further use ML based techniques to predict the properties of epoxy polymers from its constituent atomistic features.

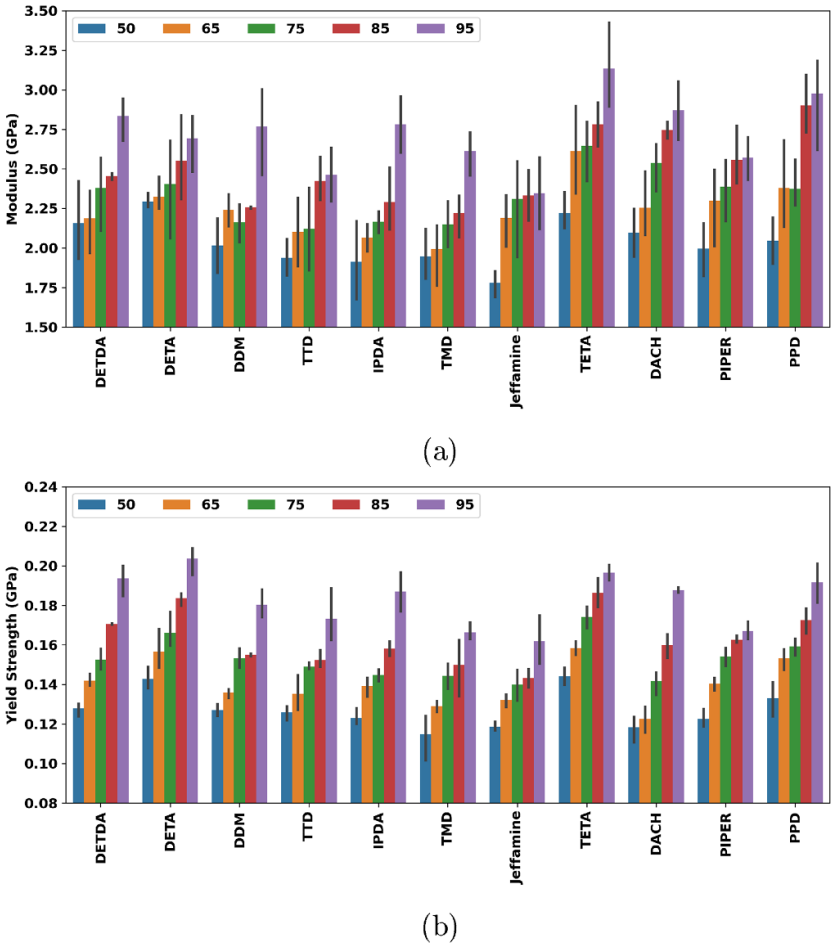


Figure 6. Influence of curing percentage on the mechanical properties of epoxy polymers: (a) elastic modulus. (b) Yield strength.

4. ML to predict mechanical properties from atomistic features

In order to predict the mechanical properties of epoxy polymers from the atomistic structural features, we employ ML techniques. For each of the epoxy combinations, the following features of the hardener are extracted from cheminformatics tool, RDKit (the features of resin are not considered since same resin is used in all cases):

- molecular weight,
- number of carbon atoms/oxygen atoms/SP^3 hybridized carbon atoms,
- number of rotatable bonds,
- number of aromatic rings/aliphatic rings,
- number of secondary amines/primary amines,
- number of radical electrons/valence electrons.

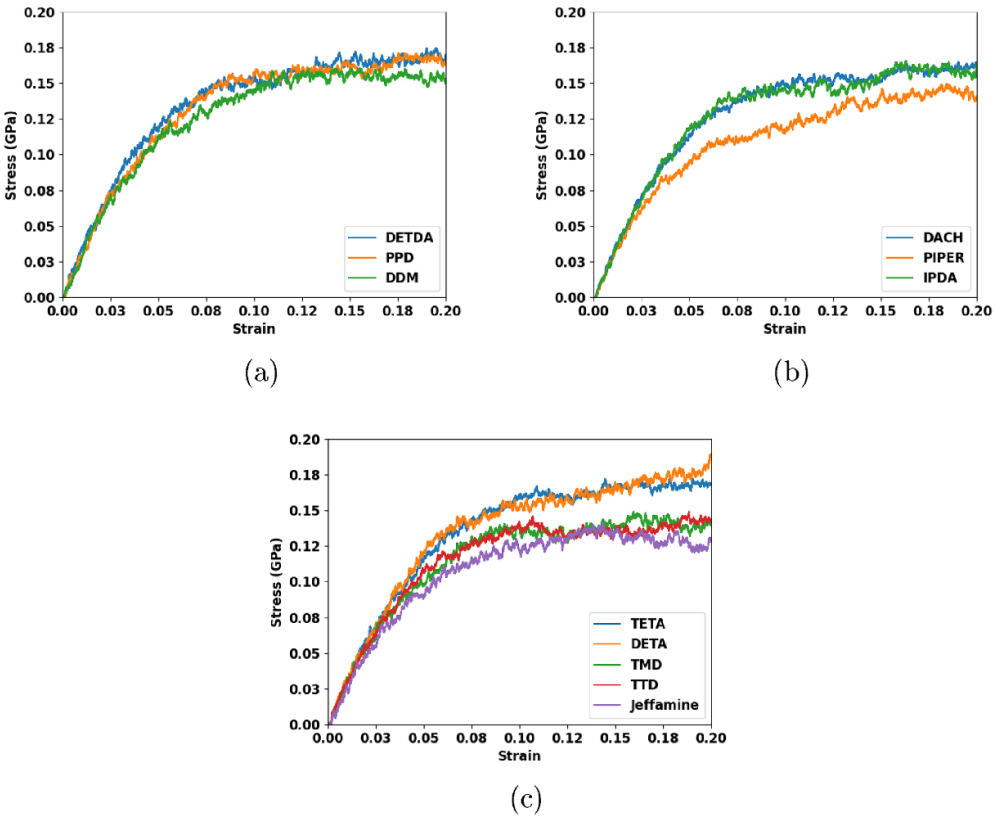


Figure 7. Influence of the type of hardener: (a) hardener with aromatic rings. (b) Hardener with aliphatic rings. (c) Hardener with aliphatic chains.

We apply a Gaussian process regression (GPR) approach to develop a predictive model for the mechanical properties (elastic modulus and yield strength) based on these features, where we label the data using the results of appropriate MD simulations. Note here that GPR is a non-parametric, kernel-based, and probabilistic method that is effective for small datasets with a limited number of features [53]. In total we use 55 cases, which consist of epoxy polymers with different hardeners and curing degrees. A total of 12 structural features, as detailed in figure 8, were initially selected for the feature selection process. To ensure compatibility between the different feature ranges, we normalize the features to have zero mean and unit variance.

Initially, we use a feature selection technique called sequential backward selection (SBS) to identify the important structural features for predicting the elastic modulus and yield strength. SBS works by starting with the full set of features and iteratively removing one feature at a time, the one contributing the least to model performance in each iteration until the specified minimum number of features. The performance of each feature subset is evaluated using the coefficient of determination (R^2 score) which has been provided as the scoring metric during the SBS process.

The variation in prediction accuracy (R^2 score) with feature removal, along with the specific features eliminated at each step, are presented in figure 9 and table 3. We find that the optimum number of features for predicting the elastic modulus is 6 and yield strength is 5 (with

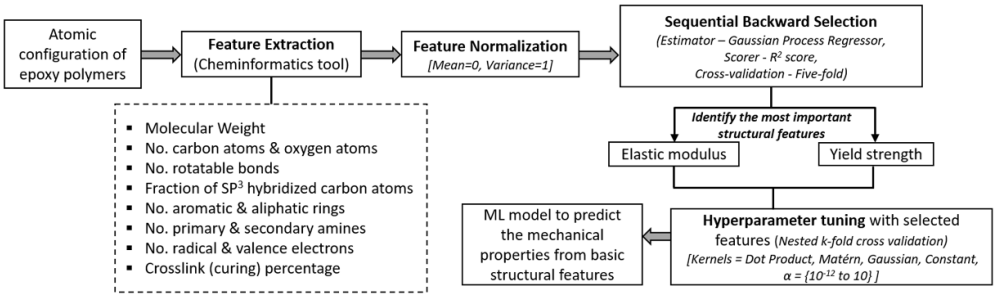
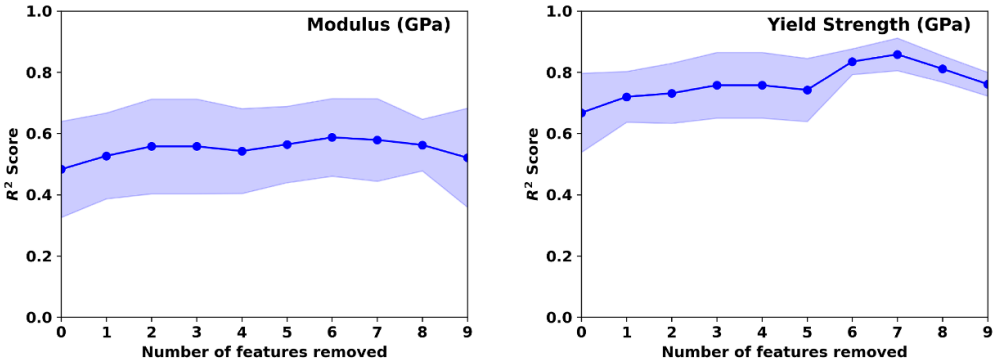


Figure 8. Schematic representation of ML model used in this study.



(a) (b)

Figure 9. R^2 score during sequential feature selection during prediction of: (a) elastic modulus. (b) Yield strength.

Table 3. Prediction accuracy of GPR model during backward sequential feature selection.

Elastic modulus		Yield strength	
Feature removed	R^2 score	Feature removed	R^2 Score
—	0.48	—	0.67
Oxygen atoms	0.53	Rotatable bonds	0.72
Valence electrons	0.56	Secondary amines	0.73
Radical electrons	0.54	Aromatic rings	0.76
Rotatable bonds	0.56	Radical electrons	0.76
Aromatic rings	0.59	Carbon atoms	0.74
Aliphatic rings	0.58	Aliphatic rings	0.83
Secondary amines	0.56	Valence electrons	0.86
Primary amines	0.52	Oxygen atoms	0.81
Molecular weight	0.52	SP^3 carbon atoms	0.76

maximum R^2 score). The important features for predicting the elastic modulus are molecular weight, fraction of carbon atoms that are SP^3 hybridized, total number of carbon atoms, primary amines, secondary amines and curing percentage. Similarly, the important features for

Table 4. Prediction accuracy of GPR model.

Property	MAPE	MAE	RMSE	R^2
Elastic modulus (GPa)	0.0592	0.1428	0.1607	0.79
Yield strength (GPa)	0.0250	0.0036	0.0045	0.95

predicting yield strength are molecular weight, fraction of carbon atoms that are SP^3 hybridized, total number of carbon atoms, primary amines and curing percentage. We can also find a chemical rationale linking these molecular features to the mechanical properties, providing insights into how they influence both elastic modulus and yield strength. For example, a greater number of primary and secondary amines in the system increases the availability of reactive sites for cross-linking reactions during curing. This results in a denser and more tightly connected polymer network, which typically enhances both stiffness and strength. Additionally, SP^3-hybridized carbon atoms are often formed at these cross-linked junctions, introducing three-dimensional connectivity that further contributes to the mechanical integrity of the material.

With the selected features, we use a nested k-fold cross-validation in order to optimize the kernel function and α (a parameter through which the noise level of the target can be specified, the parameter also helps in dealing with numerical instabilities during fitting). In this process, we split the whole dataset into five folds, with one fold kept aside as test set and the remaining four folds taken as validation/training sets and repeat the whole process five times. We provide the following kernel functions along-with their their summations for hyperparametric tuning using grid search method: dot-product kernel [k_{D}] with initial $\sigma_0 = 1$, Matérn kernel [$k_{\mathrm{M}}(\nu = 1.5)$] with initial length scale, $l = 1$, Gaussian kernel [k_{G}] with initial $l = 1$ and constant kernel [k_{C}] with initial $C = 1$ (as shown in equations (2)–(5)) and α in the range 10^{-12} to 10. The terms used in Matérn kernel like $d(.,.)$, $K_\nu(.)$ and $\Gamma(.)$ denote the euclidean distance, modified Bessel function and the gamma function. The parameters of kernel functions like σ_0, l and C are further optimized during the GPR training process.

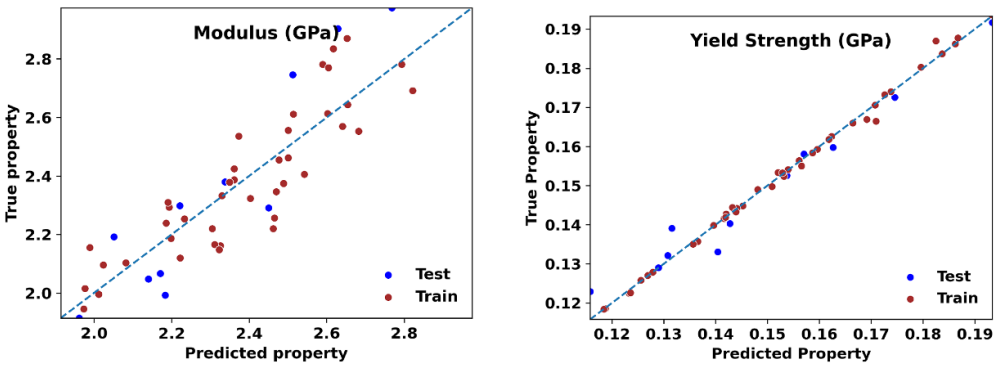
$$k_{\mathrm{D}}(x,y;\sigma_0) = \sigma_0^2 + x \cdot y \tag{2}$$

$$k_{\mathrm{M}}(x,y;\nu,l) = \frac{1}{\Gamma(\nu)\,2^{\nu-1}} \left(\frac{\sqrt{2\nu}}{l} d(x,y) \right)^{\nu} K_\nu \left(\frac{\sqrt{2\nu}}{l} d(x,y) \right) \tag{3}$$

$$k_{\mathrm{G}}(x,y;l) = \exp\left(-\frac{d(x,y)^2}{2l^2} \right) \tag{4}$$

$$k_{\mathrm{C}}(x,y;C) = C \;\forall\, x,y. \tag{5}$$

We evaluate the performance of the model with different hyperparameters by measuring the R^2 score between true value and predicted value on the validation set. In the case of elastic modulus, the model with summation of a constant, dot-product and Matérn kernel, α with 0.359 38 is found to perform best. In the case of yield strength, the model with summation of constant and dot-product, α with 0.012 92 is found to perform best. The accuracy of the trained GPR model with tuned kernel function and α for predicting elastic modulus and yield strength is presented in table 4. It can be observed that the R^2 score improved from 0.48 to 0.79 in case of elastic modulus and from 0.67 to 0.95 in case of yield strength. The comparison of predicted values with the actual values both in the test set and training set is presented in figure 10. We also compare the performance of the current model with other regression models where also feature selection using SBS and hyperparametric tuning is carried out. The MAPE

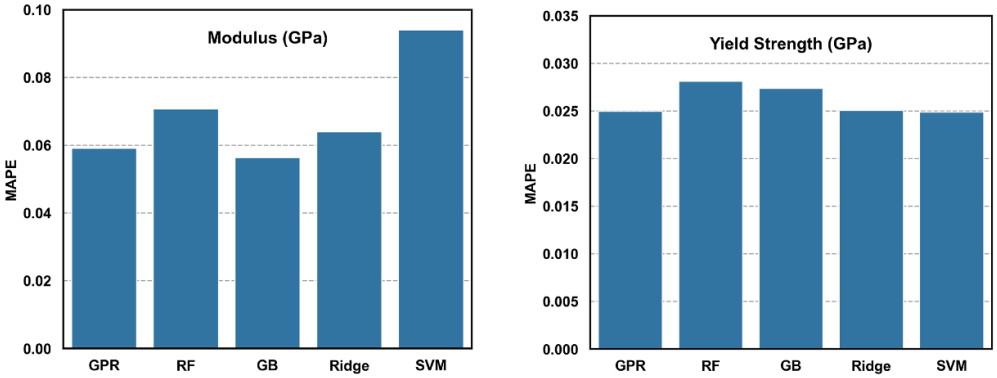


(a) (b)

Figure 10. Comparison of values predicted from GPR with the reference values: (a) elastic modulus (GPa). (b) Yield strength (GPa).

Table 5. Comparison of prediction accuracy (MAPE) of GPR model with other regression models.

Model	Elastic modulus (GPa)	Yield strength (GPa)
GPR	0.0592	0.0250
Random forest (RF)	0.0708	0.0281
Gradient boosting (GB)	0.0563	0.0274
Support vector machine (SVM)	0.0940	0.0249
Ridge regression (Ridge)	0.0640	0.0251



(a) (b)

Figure 11. Prediction accuracy of GPR with other conventional regression models: (a) elastic modulus (GPa). (b) Yield strength (GPa).

of the considered models is presented in table 5 and figure 11. It can be observed that the elastic modulus is predicted better than GPR using gradient boosting and the yield strength is predicted at par with ridge regression and SVM.

5. Conclusions

In this study, we used MDs simulations and ML techniques to get in-depth understanding on the structural features and process parameters affecting the mechanical properties of epoxy polymers. We employed MD cross-linking simulations to design EPON-862-based epoxy polymers with 11 different curing agents and systematically investigated the effects of hardener type, stoichiometric ratio, and curing percentage on their mechanical properties, specifically elastic modulus and yield strength. Using SBS, we identified key structural features such as molecular weight, fraction of SP^3-hybridized carbon atoms, total carbon count, primary amines, and curing percentage as the most important for accurately predicting the elastic modulus and yield strength of epoxy polymers. The developed model predicts elastic modulus and yield strength with good accuracy (MAPE of 0.0592 and 0.025, and R^2 scores of 0.79 and 0.95, respectively), performing comparably to conventional regression models for both properties. The proposed method can thus be extended to simulate a larger number of epoxy polymeric systems virtually and determine their properties in an efficient way. This will further lead towards increased performance space of the epoxy polymers and also aid in designing the polymers by merely selecting the constituents with appropriate structural features. The main limitation of this study is the limited number of data points, though the developed methodology remains broadly applicable to various design problems provided sufficient features are available.

Data availability statement

The data cannot be made publicly available upon publication because the cost of preparing, depositing and hosting the data would be prohibitive within the terms of this research project. The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgment

The authors would like to thank the help received Rick Oerder while developing the model. The first author (Sindu B S) acknowledges the support in the form of Post Doctoral Industrial Fellowship received from Indo-German Science and Technology Centre (IGSTC) to carry out research in Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Germany. This paper has been assigned the registration number CSIR-SERC-1091/2023. Moreover, this work was supported in part by the BMBF-project 05M2AAA MaGriDo (Mathematics for Machine Learning Methods for Graph-Based Data with Integrated Domain Knowledge).

ORCID iDs

Sindu B S 0000-0001-7138-7917
Jan Hamaekers 0000-0001-5399-043X

References

[1] Orselly M, Devemy J, Bouvet-Marchand A, Dequidt A, Loubat C and Malfreyt P 2022 Molecular simulations of thermomechanical properties of epoxy-amine resins *ACS Omega* **7** 30040–50

[2] Odegard G M, Patil S U, Deshpande P P, Kanhaiya K, Winetrout J J, Heinz H, Shah S P and Maiaru M 2021 Molecular dynamics modeling of epoxy resins using the reactive interface force field *Macromolecules* **54** 9815–24
[3] Kundalwal S and Kumar S 2016 Multiscale modeling of stress transfer in continuous microscale fiber reinforced composites with nano-engineered interphase *Mech. Mater.* **102** 117–31
[4] Senftle T P *et al* 2016 The ReaxFF reactive force-field: development, applications and future directions *npj Comput. Mater.* **2** 1–14
[5] Konrad J, Meißner R H, Bitzek E and Zahn D 2021 A molecular simulation approach to bond reorganization in epoxy resins: from curing to deformation and fracture *ACS Polym. Au* **1** 165–74
[6] Patil S U, Shah S P, Olaya M, Deshpande P P, Maiaru M and Odegard G M 2021 Reactive molecular dynamics simulation of epoxy for the full cross-linking process *ACS Appl. Polym. Mater.* **3** 5788–97
[7] Liu X and Rao Z 2020 A molecular dynamics study on heat conduction of crosslinked epoxy resin based thermal interface materials for thermal management *Comput. Mater. Sci.* **172** 109298
[8] Sindu B and Sasmal S 2022 Atomistic to continuum scale investigations on mechanical properties of epoxy bonded fiber reinforced polymer composite systems under hygro-thermal exposures *Modelling Simul. Mater. Sci. Eng.* **30** 035012
[9] Guha R D, Idolor O and Grace L 2020 An atomistic simulation study investigating the effect of varying network structure and polarity in a moisture contaminated epoxy network *Comput. Mater. Sci.* **179** 109683
[10] Tam L-H, He L and Wu C 2019 Molecular dynamics study on the effect of salt environment on interfacial structure, stress and adhesion of carbon fiber/epoxy interface *Compos. Interfaces* **26** 431–47
[11] Hou D, Yu J, Liu Q-F, Dong B, Wang X, Wang P and Wang M 2020 Nanoscale insight on the epoxy-cement interface in salt solution: a molecular dynamics study *Appl. Surf. Sci.* **509** 145322
[12] Jian W, Wang X, Lu H and Lau D 2021 Molecular dynamics simulations of thermodynamics and shape memory effect in CNT-epoxy nanocomposites *Compos. Sci. Technol.* **211** 108849
[13] Yang X, Wan Y, Wang X, Fu Y, Huang Z and Xie Q 2019 Molecular dynamics studies of the mechanical behaviors and thermal conductivity of the DGEBA/MTHPA/CNB composites *Composites B* **164** 659–66
[14] Alian A, Kundalwal S and Meguid S 2015 Multiscale modeling of carbon nanotube epoxy composites *Polymer* **70** 149–60
[15] Sindu B and Sasmal S 2015 Evaluation of mechanical characteristics of nano modified epoxy based polymers using molecular dynamics *Comput. Mater. Sci.* **96** 146–58
[16] Pruksawan S, Lambard G, Samitsu S, Sodeyama K and Naito M 2019 Prediction and optimization of epoxy adhesive strength from a small dataset through active learning *Sci. Technol. Adv. Mater.* **20** 1010–21
[17] Albuquerque R Q, Rothenhäusler F and Ruckdäschel H 2023 Designing formulations of bio-based, multicomponent epoxy resin systems via machine learning *MRS Bull.* **49** 1–12
[18] Liu B, Jin K, Tao J, Wang H, He D and Li H 2022 Performance optimization of shape memory epoxy polymers based on machine learning *Polym. Adv. Technol.* **33** 1222–32
[19] Guan Q, Guo K, Tan W and Zhou Y 2019 Rapid decomposition of epoxy resins via raman spectrometry in combination with machine learning algorithms *J. Bioresour. Bioprod.* **4** 130–4
[20] Doblies A, Boll B and Fiedler B 2019 Prediction of thermal exposure and mechanical behavior of epoxy resin using artificial neural networks and Fourier transform infrared spectroscopy *Polymers* **11** 363
[21] Yan C, Feng X, Wick C, Peters A and Li G 2021 Machine learning assisted discovery of new thermoset shape memory polymers based on a small training dataset *Polymer* **214** 123351
[22] Moriwaki H, Tian Y-S, Kawashita N and Takagi T 2018 Mordred: a molecular descriptor calculator *J. Cheminf.* **10** 1–14
[23] Landrum G *et al* 2024 RDKit: Open-source cheminformatics (https://doi.org/10.5281/zenodo.10893044)
[24] Meier S, Albuquerque R Q, Demleitner M and Ruckdäschel H 2022 Modeling glass transition temperatures of epoxy systems: a machine learning study *J. Mater. Sci.* **57** 13991–4002
[25] Hu Y, Zhao W, Wang L, Lin J and Du L 2022 Machine-learning-assisted design of highly tough thermosetting polymers *ACS Appl. Mater. Interfaces* **14** 55004–16

[26] Jin K, Luo H, Wang Z, Wang H and Tao J 2020 Composition optimization of a high-performance epoxy resin based on molecular dynamics and machine learning *Mater. Des.* **194** 108932
[27] Luo H, Jin K, Tao J and Wang H 2021 Properties prediction and design of self-healing epoxy resin combining molecular dynamics simulation and back propagation neural network *Mater. Res. Express* **8** 045308
[28] Choi J, Kang H, Lee J H, Kwon S H and Lee S G 2022 Predicting the properties of high-performance epoxy resin by machine learning using molecular dynamics simulations *Nanomaterials* **12** 2353
[29] Qiu H, Zhao W, Pei H, Li J and Sun Z-Y 2022 Highly accurate prediction of viscosity of epoxy resin and diluent at various temperatures utilizing machine learning *Polymer* **256** 125216
[30] Shafe A, Nourian P, Liu X, Li G, Wick C D and Peters A J 2024 Identification and design of better diamine-hardened epoxy-based thermoset shape memory polymers: simulation and machine learning *Macromolecules* **57** 9933–42
[31] Schneider J, Hamaekers J, Chill S T, Smidstrup S, Bulin J, Thesen R, Blom A and Stokbro K 2017 ATK-ForceField: a new generation molecular dynamics software package *Modelling Simul. Mater. Sci. Eng.* **25** 085007
[32] Jorgensen W L, Maxwell D S and Tirado-Rives J 1996 Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids *J. Am. Chem. Soc.* **118** 11225–36
[33] Weininger D 1988 SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules *J. Chem. Inf. Comput. Sci.* **28** 31–36
[34] Martínez L, Andrade R, Birgin E G and Martínez J M 2009 PACKMOL: a package for building initial configurations for molecular dynamics simulations *J. Comput. Chem.* **30** 2157–64
[35] Smidstrup S *et al* 2019 QuantumATK: an integrated platform of electronic and atomic-scale modelling tools *J. Phys.: Condens. Matter* **32** 015901
[36] Varshney V, Patnaik S S, Roy A K and Farmer B L 2008 A molecular dynamics study of epoxy-based networks: cross-linking procedure and prediction of molecular and material properties *Macromolecules* **41** 6837–42
[37] Kallivokas S V, Sgouros A P and Theodorou D N 2019 Molecular dynamics simulations of EPON-862/DETDA epoxy networks: structure, topology, elastic constants and local dynamics *Soft Matter* **15** 721–33
[38] Abbott L J, Hart K E and Colina C M 2013 Polymatic: a generalized simulated polymerization algorithm for amorphous polymers *Theor. Chem. Acc.* **132** 1–19
[39] Griebel M and Hamaekers J 2004 Molecular dynamics simulations of the elastic moduli of polymer–carbon nanotube composites *Comput. Methods Appl. Mech. Eng.* **193** 1773–88
[40] Sindu B and Sasmal S 2020 Molecular dynamics simulations for evaluation of surfactant compatibility and mechanical characteristics of carbon nanotubes incorporated cementitious composite *Constr. Build. Mater.* **253** 119190
[41] Sun L, Warren G L, O'Reilly J Y, Everett W N, Lee S M, Davis D, Lagoudas D and Sue H-J 2008 Mechanical properties of surface-functionalized SWCNT/epoxy composites *Carbon* **46** 320–8
[42] Bajpai A and Wetzel B 2019 Tensile testing of epoxy-based thermoset system prepared by different methods Preprints (https://doi.org/10.20944/preprints201907.0143.v1)
[43] Shinde D K, Emmanwori L and Kelkar A D 2014 Comparison of mechanical properties of EPON 862/W with and without TEOS electrospun nanofibers in nanocomposite *SAMPE Seattle* pp 59–4040
[44] Littell J D, Ruggeri C R, Goldberg R K, Roberts G D, Arnold W A and Binienda W K 2008 Measurement of epoxy resin tension, compression and shear stress–strain curves over a wide range of strain rates using small test specimens *J. Aerospace Eng.* **21** 162–73
[45] King J A, Klimek D R, Miskioglu I and Odegard G M 2015 Mechanical properties of graphene nanoplatelet/epoxy composites *J. Compos. Mater.* **49** 659–68
[46] Gollins K, Chiu J, Delale F, Liaw B and Gursel A 2014 Comparison of manufacturing techniques subject to high speed impact *ASME Int. Mechanical Engineering Congress and Exposition* vol 46583 (American Society of Mechanical Engineers) p V009T12A019
[47] Li C and Strachan A 2011 Molecular dynamics predictions of thermal and mechanical properties of thermoset polymer EPON862/DETDA *Polymer* **52** 2920–8
[48] Kelkar A, Komuves F, Mohan R and Kelkar V 2011 Prediction of mechanical properties of EPON 862 (DGEBF) cross-linked with curing agent W (DETDA) and SWCNT using MD simulations *52nd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conf. 19th AIAA/ASME/AHS Adaptive Structures Conf. 13t* p 1921

[49] Wan X, Demir B, An M, Walsh T R and Yang N 2021 Thermal conductivities and mechanical properties of epoxy resin as a function of the degree of cross-linking *Int. J. Heat Mass Transfer* **180** 121821
[50] Li C and Strachan A 2015 Evolution of network topology of bifunctional epoxy thermosets during cure and its relationship to thermo-mechanical properties: a molecular dynamics study *Polymer* **75** 151–60
[51] Fard M Y, Liu Y and Chattopadhyay A 2012 Characterization of epoxy resin including strain rate effects using digital image correlation system *J. Aerospace Eng.* **25** 308–19
[52] Jakubinek M B, Ashrafi B, Martinez-Rubi Y, Rahmat M, Yourdkhani M, Kim K S, Laqua K, Yousefpour A and Simard B 2018 Nanoreinforced epoxy and adhesive joints incorporating boron nitride nanotubes *Int. J. Adhes. Adhes.* **84** 194–201
[53] Rasmussen C E *et al* 2006 *Gaussian Processes for Machine Learning* vol 1 (Springer)