Electronic Supplementary Material (ESI) for Molecular Systems Design & Engineering.

Supporting information

Machine Learning based Epoxy Resins Properties Prediction

Huiwon Jang[1], Dayoung Ryu[1†], Wonseok Lee[1], Geunyeong Park[2] and Jihan Kim[1]*

1 Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea

2 KOLON One&Only TOWER, 110, Magokdong-ro Gangseo-gu Seoul 07793, Republic of Korea

*Email : jihankim@kaist.ac.kr

Simulation Details

The epoxy monomers and hardeners were initially loosely placed in the simulation box. The ratio of epoxy monomers to hardeners was generally adjusted to maximize the degree of curing as possible. Subsequently, they underwent a curing process, leading to the formation of cross-links between components. Following the curing process, well-equilibrated cross-linked epoxy resins were produced, resulting appropriately sized simulation box for epoxy resins, as depicted below. The specific details of the MD simulation process employed for the curing process can be found in the previous study in Lee et al(1). Various previous studies have used the GAFF force field for epoxy resin and validated that GAFF has an advantage in accurately reproducing its properties (2-8). Therefore, we chose the GAFF force field to model the epoxy resin in our work.

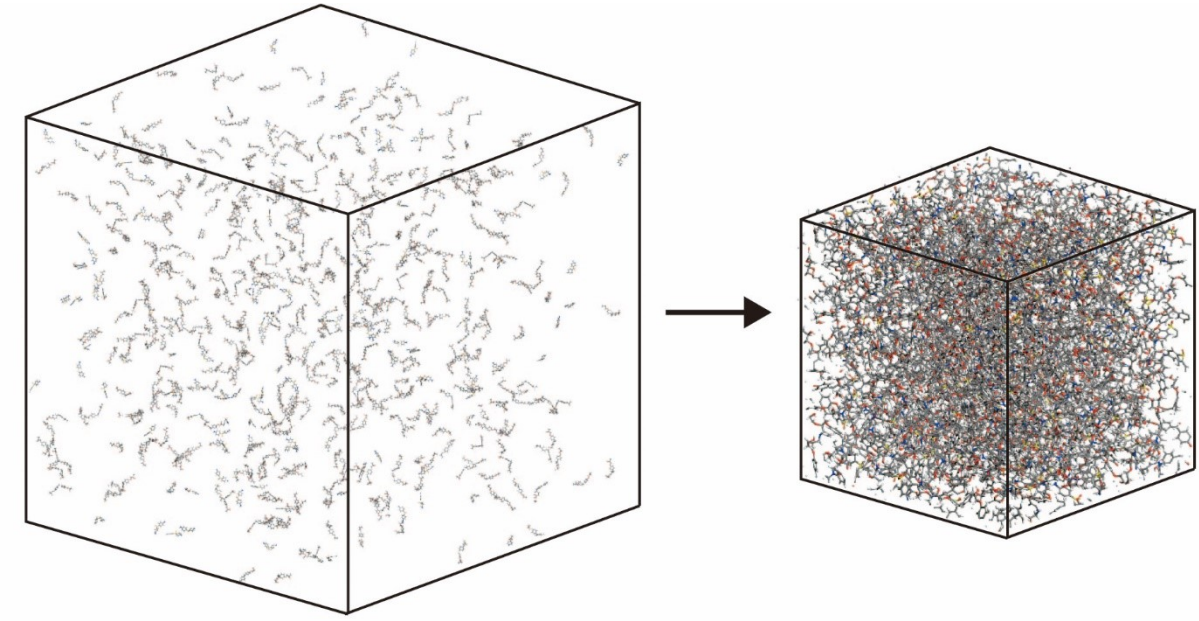

Figure S1. Simulation snapshots depicting the initial and final configuration of epoxy resins.

Following the generation of cross-linked polymers, various properties were computed. The glass transition temperature (T_g) and volumetric coefficient of thermal expansion (CTE) were determined from the Temperature-Density plot. T_g was estimated using the piecewise regression

on the Temperature-Density plot. To enhance the accuracy of T_g calculations, the entire temperature range was divided into glassy, glass transition, and rubbery regions as visualized in Figure S2. Regression analysis was performed on each data point within the glassy and rubbery states to derive the coefficients of a first-order polynomial. The intersection of the regression lines from the glassy and rubbery regions was then calculated to determine the potential T_g value. The transition regions encompassed all possible temperature intervals within the calculated range. The potential T_g values were calculated for each transition region, and the final value was determined by averaging all these potential points.

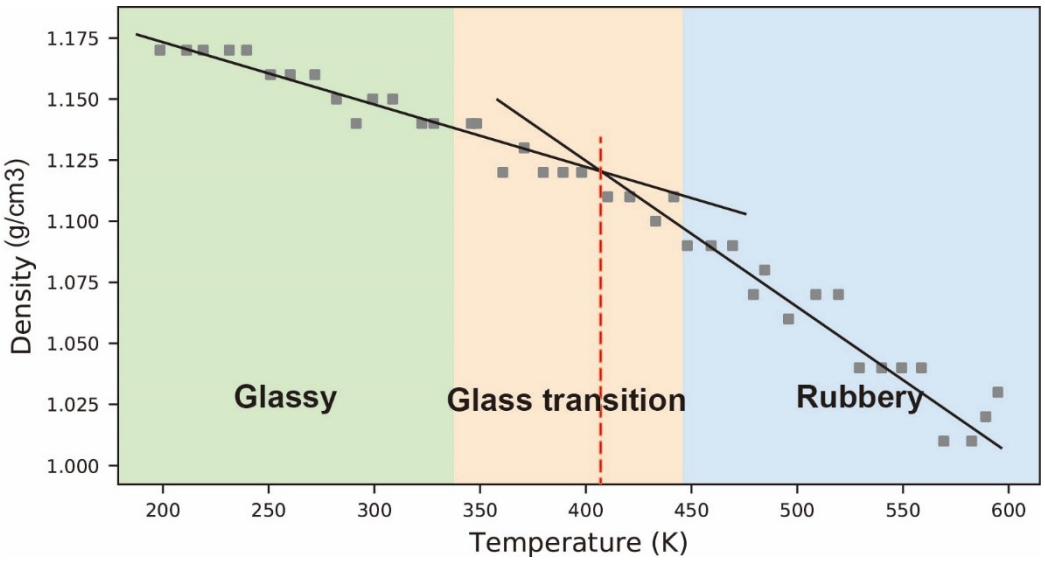

Figure S2. Temperature-density plot illustrating the estimation of T_g with the depiction of three divided regions: glassy, glassy transition, and rubbery region.

We conducted validation of our glass transition temperature (T_g) estimation method using varying temperature intervals (2K, 5K, and 10K). This validation aimed to assess how the T_g values fluctuated with changes in the interval size. The results, illustrated in Figure S3, demonstrate that T_g values calculated with a 10K interval are consistent with those obtained using smaller intervals of 2K and 5K. Such consistency across these intervals ensures the accuracy of our T_g estimations.

Considering the computational costs, which are a crucial factor in efficiently collecting data for machine learning models, we opted for the 10K interval. This approach allows us to effectively

gather a diverse data on epoxy resins, thereby facilitating the development of robust machine learning models.

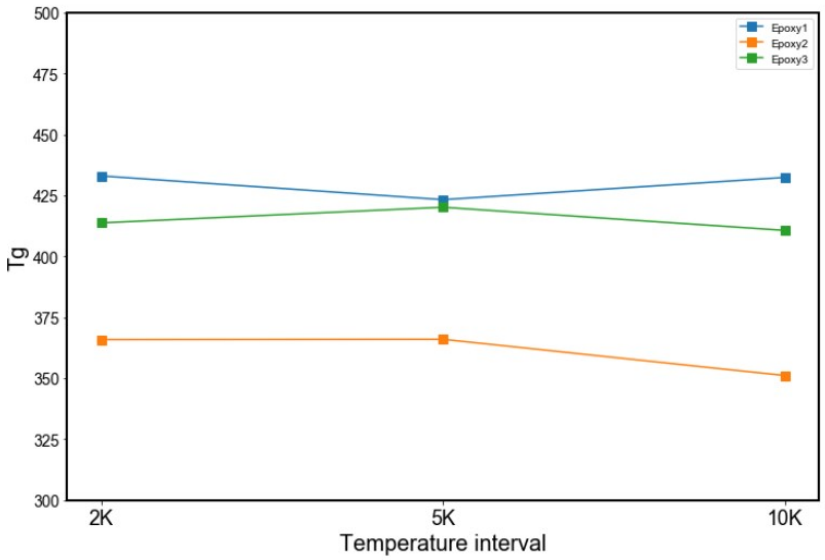

Figure S3. Estimation of T_g at various temperature intervals

Table S1. Details of the chemical space of epoxy resins collected via MD simulation, including the number, range, and average values of the data.

	Number of Data	Range	Average
Density [g/cm^3]	789	1.00 - 1.32	1.12
CTE [ppm/K]	789	18.23 - 212.24	66.16
T_g [K]	789	265.8 - 510.6	391.8
Young's modulus [GPa]	789	0.50 -15.85	5.56

Machine Learning Models

Various ML models were tested to determine the most suitable model for predicting epoxy resin properties. The five models included XGBoost, Gradient Boosting Machine (GBM), Random Forest, Support Vector Machine (SVM), and Single-layer Neural Network. The Single-layer Neural Network exhibited a negative R^2 score, indicating insufficient training, likely due to the relatively small dataset of polymers. While, regression models demonstrated superior performance compared to the deep learning model. The performance of regression models is described below. Among the tested models, XGBoost exhibited the best performance across all four properties. Consequently, it is concluded that the XGBoost model is the most appropriate ML model for predicting epoxy resin properties. The SVM (svm), Random Forest (RandomForestRegressor), and GBM (GradientBoostingRegressor) machine learning models were conducted using scikit-learn (sklearn) package, version 0.23.1 and XGBoost model was conducted using the xgboost package, version 1.5.0. The Single-layer Neural Network was implemented using Tensorflow, version 2.4.1, and featured a single dense layer with 'ReLU' activation and 'RMSprop' as the optimizer.

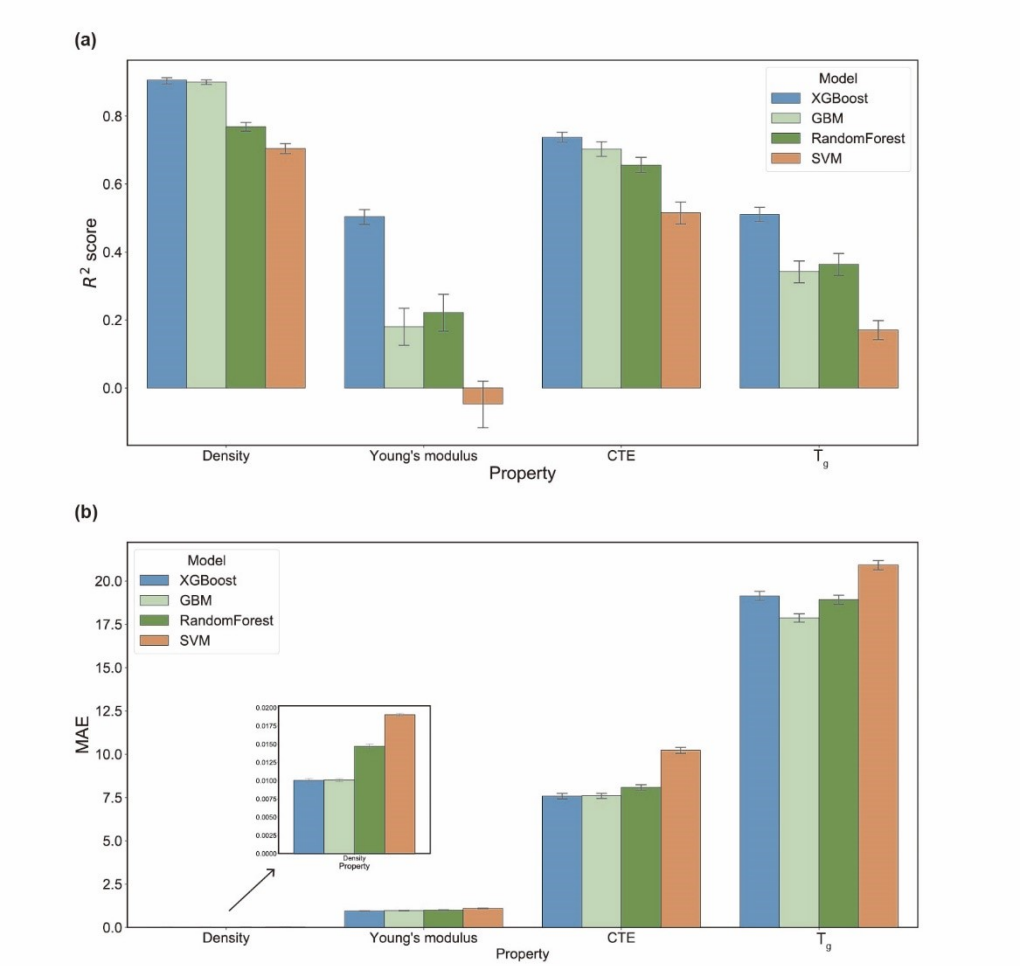

Figure S4. Performance of ML Models Evaluated with (a) R^2 Score and (b) Mean Absolute Error (MAE).

A total of 1463 descriptors from epoxy monomer and 1345 descriptors from the hardener were extracted. Subsequently, an automatic feature selection algorithm, Recursive Feature Elimination (RFE), was implemented to select the optimal subset of descriptors for each property. The number of optimized descriptors varies for each property and listed below. Hyperparameter optimization was conducted using 'GridSearchCV', which evaluates every possible parameter combination to determine the one that yields the highest cross-validation score. We concentrated on the 'n_estimators', 'learning_rate', and 'max_depth' hyperparameters, optimizing each for the respective properties. The optimized hyperparameters are detailed in Table S3.

Table S2. The number of selected descriptors for each property.

Density	CTE	T_g	Young's modulus
129	79	99	79

Table S3. Results of hyperparameter tuning.

	Density	CTE	T_g	Young's modulus
n_estimator	200	100	600	600
Learning_rate	0.1	0.1	0.01	0.01
Max_depth	4	4	4	4

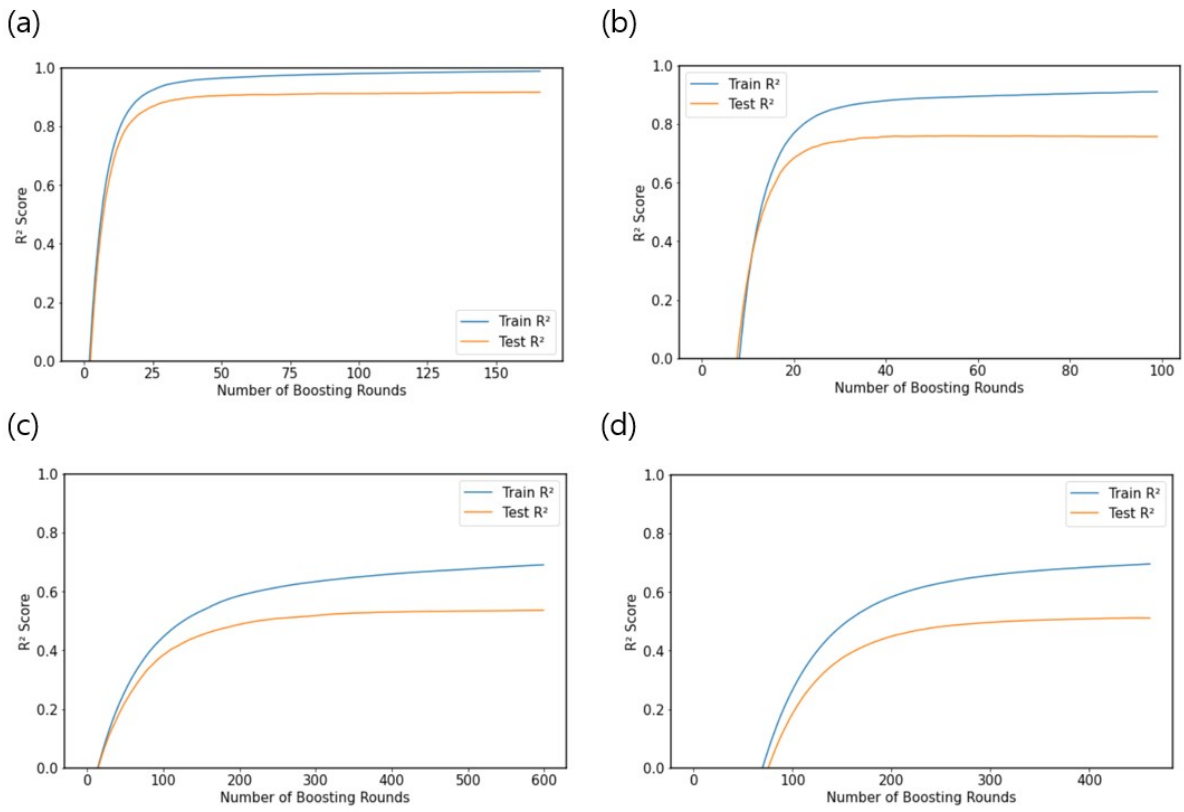

Figure S5. Learning curves of the machine learning model for (a) density, (b) CTE, (c) T_g, and (d) Young's modulus.

New Epoxy Design

Leveraging the structure-property relationship extracted from our ML model, we designed a new epoxy resin to achieve desired properties. The relationships indicate that 'System_mass' and epoxide number significantly impact these key properties. Therefore, we designed the epoxy resin to have high 'System_mass' and epoxide number, anticipating high density, T_g, Young's modulus, and low CTE. To achieve this, we selected (a) epoxy phenol novolac and (b) dicyandiamide. Epoxy phenol novolac has a relatively high epoxide number (8), and the combination of epoxy phenol novolac and dicyandiamide results in a high system mass (6.9g), which fits our design algorithm.

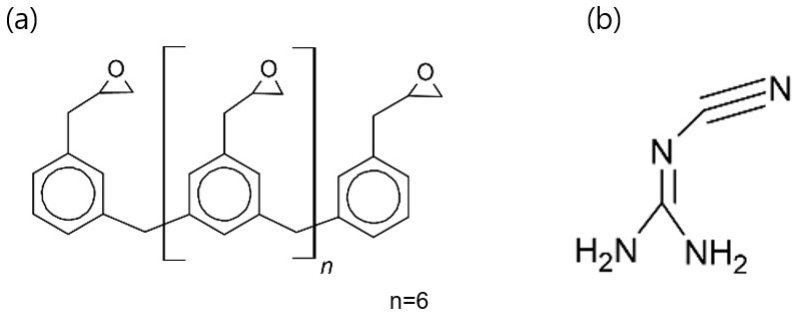

Figure S6. Chemical structures of (a) epoxy phenol novolac, (b) dicyandiamide.

This epoxy resin was synthesized, and its properties were calculated virtually via MD simulation. The properties predicted by our ML model and the actual values calculated from MD simulation were compared in Figure S6. Each property is plotted within the existing property range to visualize the extent to which the epoxy resin's properties fit within our existing dataset. The results show that the error in the properties is relatively low, indicating that our ML model provides accurate predictions. Additionally, the designed epoxy resin achieves the expected properties with relatively high density, T_g, Young's modulus, and low CTE. This strongly

suggests that our predictive model offers accurate predictions and that the structure-property relationship we proposed can assist researchers in designing new epoxy resins with desired properties.

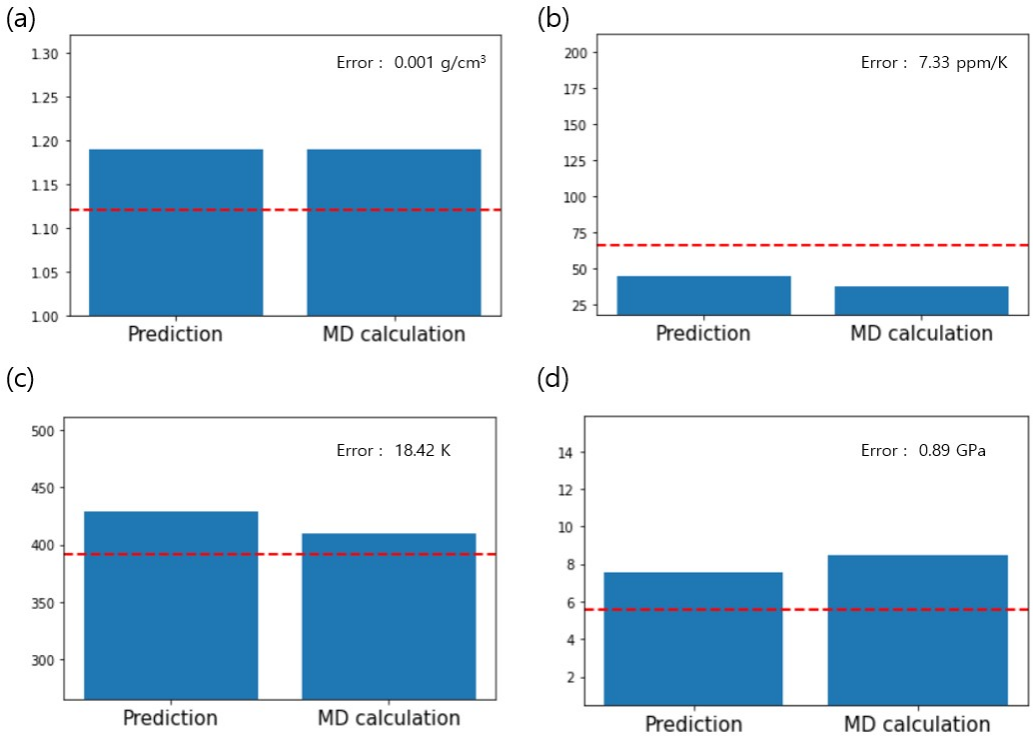

Figure S6. Comparison between ML prediction and MD simulation values of epoxy resin synthesized from epoxy phenol novolac and dicyandiamide for (a) density, (b) CTE, (c) T_g, and (d) Young's modulus. The red line denotes the average value of the existing dataset.

REFERENCES

1. Lee W, Chong S, Kim J. Graph-Network-Based Predictive Modeling for Highly Cross-Linked Polymer Systems. arXiv preprint arXiv:240106152. 2023.
2. Fan J, Anastassiou A, Macosko CW, Tadmor EB. Molecular dynamics predictions of thermomechanical properties of an epoxy thermosetting polymer. Polymer. 2020;196:122477.
3. Makarov G, Borodina O, Makarova T, Ignatova A, Olivenko N, Bartashevich E, et al. Molecular Dynamics Study of Cured ED-20 Epoxy Resin for Predicting the Glass Transition Temperature and Relationship with Structure Features. The Journal of Physical Chemistry A. 2023;127(17):3894-905.
4. Gavrielides A, Duguet T, Aufray M, Lacaze-Dufaure C. Model of the DGEBA-EDA Epoxy Polymer: Experiments and Simulation Using Classical Molecular Dynamics. International Journal of Polymer Science. 2019;2019:9604714.
5. Sridhar AS. Effect of stoichiometry on crosslinked epoxy resin characteristics: structural heterogeneities, topological defects, properties, free volume and segmental mobility. Soft Matter. 2022;18(12):2354-72.
6. Lin P-H, Khare R. Molecular Simulation of Cross-Linked Epoxy and Epoxy−POSS Nanocomposite. Macromolecules. 2009;42(12):4319-27.
7. Miura T, Shimoi Y, Hasegawa K, Takagi K, Yamazaki N, Matsuki T, et al. Simulation Study of the Effects of Interfacial Bonds on Adhesion and Fracture Behavior of Epoxy Resin Layers. The Journal of Physical Chemistry B. 2021;125(39):11044-57.
8. Khare KS, Phelan FR, Jr. Quantitative Comparison of Atomistic Simulations with Experiment for a Cross-Linked Epoxy: A Specific Volume–Cooling Rate Analysis. Macromolecules. 2018;51(2):564-75.