

Predicting Glass Transition of Amorphous Polymers by Application of Cheminformatics and Molecular Dynamics Simulations

Anas Karuth^{1[+]}, Amirhadi Alesadi^{2[+]}, Wenjie Xia^{2*}, Bakhtiyor Rasulev^{1*}

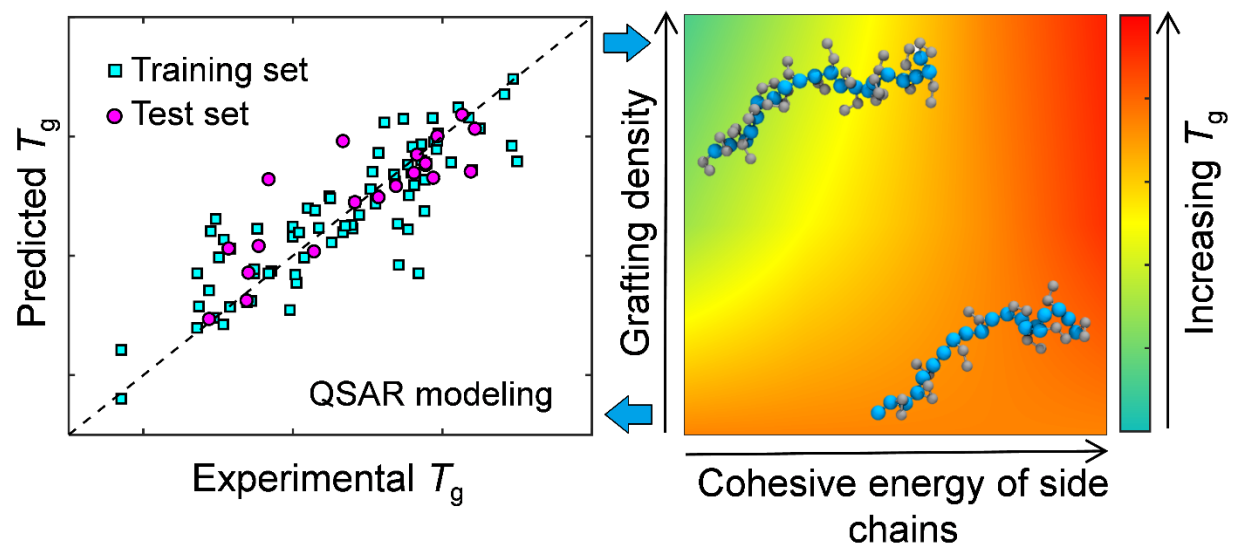
¹Department of Coatings and Polymeric Materials, North Dakota State University, Fargo, ND
58108, United States

²Department of Civil and Environmental Engineering, North Dakota State University, Fargo, ND
58108, United States

[+]These authors contributed equally to this work.

*To whom correspondence should be addressed.

Contact information: bakhtiyor.rasulev@ndsu.edu, wenjie.xia@ndsu.edu



ABSTRACT

Predicting glass-transition temperature (T_g) of glass-forming polymers is of critical importance as it governs the thermophysical properties of polymeric materials, such as relaxation dynamics, modulus, specific heat, dielectric properties. The cheminformatics approaches based on machine learning algorithms are becoming very useful in predicting the quantitative relationships between key molecular descriptors and various physical properties of materials. In this work, we developed a modeling framework by integrating cheminformatics methods and coarse-grained molecular dynamics (CG-MD) simulations to predict T_g of diverse sets of polymers. The best predictive machine learning-based QSPR model identified the most prominent molecular descriptors influencing the T_g of a hundred of polymers. Informed by the model, CG-MD simulations are performed to further delineate mechanistic interpretation and systematic dependence of these influential molecular features on T_g by investigating three major CG model parameters, namely the cohesive interaction, chain stiffness, and grafting density. The CG-MD simulations reveal that the higher intermolecular interaction and chain stiffness elevates the T_g of CG polymers, where their relative influences are coupled with the existence of side chains grafted on the backbone. This synergistic modeling approach provides valuable insights into the roles of key molecular features influencing the T_g of polymers, paving the way to establish a materials-by-design framework for polymeric materials via molecular engineering.

Keywords: Glass-transition temperature, polymeric materials, QSPR, cheminformatics, coarse-grained modeling, molecular dynamics simulations.

1. INTRODUCTION

Understanding and predicting the glass-transition temperature (T_g) of amorphous polymers is of critical importance due to its crucial role in governing their thermal and mechanical behaviors. Consequently, T_g is often considered as one of the most important descriptors of a polymeric material in various engineering and technological applications. The dramatic change in dynamics of polymers during the glass-forming process is found to be closely relevant to various thermodynamic and kinetic attributes (e.g. rapid loss in entropy and cooling rate dependence).¹⁻⁴ Quantifying T_g is important to classify the polymers to the different categories of elastomer, plastics, and hard plastics for purposes of practical usages. Experimentally, T_g can be measured by various techniques, such as differential scanning calorimetry (DSC), thermomechanical analysis (TMA), and dynamical mechanical analysis (DMA).⁵ In parallel with experiments, computational assessment, and prediction of T_g become increasingly demanding to accelerate design and development of polymer materials. More importantly, it provides valuable insights into the physical mechanisms of complex glass-forming behaviors of polymers by delineating the underlying relationships between T_g and essential chemical features at a fundamental molecular level.

Due to the rapid development of computational capacity, various computational techniques have been used to assess the physical properties of polymeric materials, including physics-based and data-driven based modeling. Among these techniques, the machine learning-based Quantitative Structure-Activity/Property Relationship (QSAR/QSPR) modeling approach has drawn considerable attention for the development of a predictive mathematical relationship linking chemical structure and activity/property of a series of chemical compounds. They are based upon various machine learning techniques, e.g. regression analysis, genetic algorithms, artificial neural

networks, to train the dataset for making predictions of various properties.⁶ QSPR models can predict the properties of chemical compounds and help in streamlining the subsequent process, such as synthesizing and testing, thereby saving resources and time. Various physical properties of polymeric materials have been predicted with a significant level of confidence using QSAR/QSPR methods, such as refractive index,⁷ dielectric constant,⁸ intrinsic viscosity,⁹ and diffusion constant.¹⁰ Moreover, these methods have been successfully applied in the pharmaceutical industry for the development of new drugs and drug delivery agents.¹¹ Such success further provokes applications of QSPR to study T_g for diverse set polymeric materials for the design and selection of new materials.

Over the past decade, numerous attempts have been made towards the prediction of T_g of polymers using the QSPR approach. Bicerano developed a QSPR-enabled predictive model for T_g with a high-level fidelity (i.e., R^2 of 0.95) for a dataset of about 320 different polymers by evaluating the solubility parameter and topological bond connectivity parameters of the monomer structures.¹² Using the Comprehensive Descriptors for Structural and Statistical Analysis (CODESSA) program, Katritzky *et al.*¹³ have developed a four-parameter QSPR model with a R^2 of 0.928 for a set of T_g values for a small set of 22 different linear-chain homopolymers and copolymers with medium molecular weight.¹³ In a later study,¹⁴ they developed a five-descriptor QSPR model for T_g values of 88 high molecular weight homopolymers, where the descriptors are independent of the molecular weight. However, it is important to note that the models in studies¹²⁻¹⁴ are not properly validated since no external dataset was used. Several other studies have adopted similar approaches based on QSPR to develop models using extended descriptors and diverse dataset for improvement of T_g predictions.¹⁵⁻¹⁹ Integrations of QSPR and other computational techniques, such as quantum chemistry, have been proposed to better the estimate of descriptors

and predict the T_g of diverse set of polymers with improved accuracy.^{8,20} A very interesting recently published two works by Missio and Schwartz^{17,18} are reporting the application of artificial neural networks (ANN) to predict T_g of polymers. In first work¹⁷, authors applied convolutional fully connected neural networks to a set of 100 polymers, mainly polystyrenes and polyacrylates. To encode polymers they used SMILES strings applied with relatively low error of prediction, about 8%. The second work¹⁸ is applying similar connected ANN method to investigate a larger set of polymers, more than 200 atactic polyacrylates. The obtained model had a good predictive power with relative error of prediction about 3%. The second work¹⁸ was focused not only on statistical significance of the models as it was in work¹⁷, but also on structural fragments contribution to T_g values.

At the same time, despite the tremendous progress, the predictive capabilities of QSPR models are still limited due to a lack of external validation and a relatively smaller and homologous dataset. For instance, the prediction quality of a QSPR model often varies greatly depending on the size and composition of the test dataset.²¹ More importantly, it is challenging for the QSPR modeling to provide a mechanistic interpretation and systematic dependence of the structural descriptors or features involved in the models. This is of critical importance for T_g prediction, as recent experimental and computational studies have suggested correlations between various molecular descriptors and T_g of polymers. For this purpose, molecular dynamics (MD) simulations can provide valuable insights into the roles of molecular descriptors from QSPR study in T_g index values. In particular, coarse-grained (CG) MD simulations that simplify all-atomistic configuration by grouping a cluster of atoms into CG beads are often utilized to bridge the spatiotemporal limitations and study the complex behaviors of polymers.²² It worth to note, since isobaric or isochoric cooling of a polymer model largely influences the glass formation,^{23–25} T_g calculations

in MD simulations in this study are performed under constant pressure condition where the isobaric process is more relevant to the experimental conditions. The integration of machine learning-based QSPR method with CG-MD simulations offers a unique way towards a fundamental understanding of the parameter-structure-property relationships in the T_g of diverse polymers (as depicted in **Figure 1**).

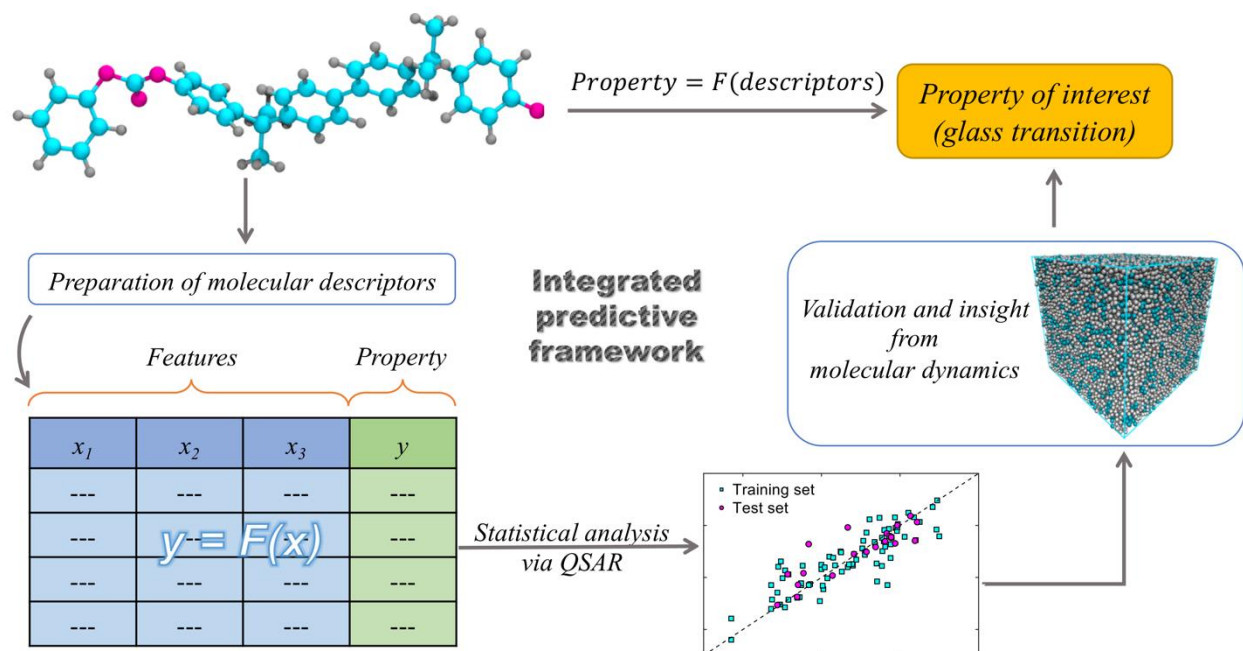


Figure 1. Flow chart of the integrated framework of cheminformatics and coarse-grained molecular dynamics (CG-MD) for prediction of the glass-transition temperature T_g of polymers. In this workflow, the molecular descriptors ($x_1, x_2, x_3 \dots$) and property of interest (y) are collected and analyzed for the development of QSPR models based on machine learning. CG-MD simulations are then carried out to validate the QSPR model and provide physical insights into essential molecular features.

In the present study, we employed the QSPR modeling in combination with CG-MD simulations, to understand the interrelation of key molecular features that influence T_g in a diverse

set of polymers, for further prediction of polymer properties (**Figure 1**). Specifically, a machine-learning based QSPR model is developed by examining the experimental dataset of T_g values for a series of polymers having diverse chemical functionalities, including carboxylates, ethers, halides, cyanides, amides, acetates, alcohols, hydrocarbon chains, and aromatic and nonaromatic rings. In particular, the applied dataset covers a wide range of T_g values (i.e., from 145 K to 482 K) of amorphous polymers, representing a large variety of structural and chemical features. CG-MD simulations are carried out to systematically explore the roles of essential molecular parameters on T_g of polymers informed by the QSPR model. The established predictive framework allows to gain a fundamental understanding of the structural and functional features that influence T_g , paving the way to materials-by-design strategy for polymeric materials' development with tailored performance.

2. MATERIALS AND METHODS

Polymer Dataset. In this work a large set of polymers with a wide range of experimental T_g values was collected. The experimental T_g values of 100 different polymers are collected from diverse sources.²⁶ For predictive model validating purposes the dataset is divided into training and test sets (80% of dataset as the training set and 20% of dataset as test or external set). The splitting of the dataset to training and test set is carried out by ranking the T_g values of structures in ascending order and taking every 5th chemical from the dataset as a test set.

Physico-Chemical and Structural Descriptors. To encode a polymeric structure a series of structural/molecular descriptors are generated based on the chemical structure of the monomer. Each chemical structure is built using a ChemSketch²⁷ software, followed by structure encoding

process within a Dragon 6²⁸ software. The initial set of generated descriptors contained more than 4,500 various descriptors, corresponding to 0D-, 1D-, 2D-, and 3D-structure based descriptor classes. The generated descriptors are represented 20 different classes - constitutional, topological, walk and path counts, connectivity indices, information indices, 2D autocorrelations, edge adjacency indices, Burden eigenvalues, topological charge indices, eigenvalue based indices, Randic molecular profiles, geometrical descriptors, RDF descriptors, 3D-MoRSE descriptors, WHIM descriptors, GETAWAY descriptors, functional groups, atom-centered fragments, charge descriptors, and molecular property descriptors.²⁹ A final set of 2,863 descriptors is selected after filtering out constant and near-constant descriptors to describe each polymeric structure. The list of the initially generated set of 4,500 descriptors and the final set of 2,863 descriptors is represented in the Supplementary Information (SI) file (see **Table S1** and **Table S2**, respectively).

Machine Learning-based QSPR Modeling. In the next step, a machine learning-based QSPR approach is applied to find a quantitative relationship between T_g values and structural/molecular descriptors. The descriptors selection and QSPR model development are made by using a combination of a variable selection Genetic Algorithm (GA)³⁰ and Multiple Linear Regression Analysis (MLRA) methods. The QSPR models are generated by combined GA-MLRA³¹ technique as implemented in QSARINS 2.2 software.³² It is worth noting that a combination of GA and MLRA for variable selection is a powerful tool to address many tasks in QSPR studies where needed a mechanistic interpretation of results.³³ The MLRA method was specifically applied to make the model transparent and mechanistically explainable. In this work, the GA variable selection started with a population of 100 random models and 5,000 iterations to evolution with the mutation probability specified at 85%. The systematic study delineated here consists of several developed QSAR models with 1 to 10 variables in the models, followed by

statistical analysis for models' performance evaluation by a squared correlation coefficient R^2 , root means square error (RMSE) and F -test (measure of dispersion, larger values represent greater dispersion). The following equations are utilized to calculate the correlation coefficient R^2 for training set (Eq. 1) and the root mean square error of calibration (Training) $RMSE_C$, as the measures of goodness-of-fit for each developed model (Eq. 2):

$$R^2_{Training} = 1 - \frac{\sum_{i=1}^n (y_i^{obs} - y_i^{pred})^2}{\sum_{i=1}^n (y_i^{obs} - \bar{y}^{obs})^2} \quad (1)$$

$$RMSE_C = \sqrt{\frac{\sum_{i=1}^n (y_i^{obs} - y_i^{pred})^2}{n}} \quad (2)$$

To ensure the reliability of the models and their robustness, the proper validation of QSPR models is imperative. The internal and external validations of the developed models are performed in this study. In the internal validation process, The best model generated in each variable model using GA-MLR approach is then tested by applying the cross-validation “leave-one-out” technique, $Q^2_{Training}$ (or Q^2_{LOO}). In this statistical technique, each variable is iteratively held-out from the training set used for model development and “predicted” as a new model to verify internal “predictivity”. We calculate the cross-validated coefficient $Q^2_{Training}$ and root-mean-square error of cross-validation (CV) $RMSE_{CV}$ (Eq. 3 and 4), which is utilized to measure the predictive ability of the model by avoiding over-fitting scenarios:

$$Q^2_{Training} = 1 - \frac{\sum_{i=1}^n (y_i^{obs} - y_i^{predcv})^2}{\sum_{j=1}^n (y_j^{obs} - \bar{y}^{obs})^2} \quad (3)$$

$$RMSE_{CV} = \sqrt{\frac{\sum_{i=1}^n (y_i^{obs} - y_i^{predcv})^2}{n}} \quad (4)$$

In the process of model validation, it is important to perform an external validation. The external validation of the model is checked for its ability to predict new structures that are not

included in the training set. This is done by applying the developed model obtained based on the training set, to set of chemicals not included in the training set. The external predictivity of each model is represented as R^2_{Test} (Eq. 5) and root mean square error of test ($RMSE_{Test}$) (Eq. 6).

$$R^2_{Test} = 1 - \frac{\sum_{j=1}^k (y_j^{obs} - y_j^{pred})^2}{\sum_{j=1}^k (y_j^{obs} - \hat{y}^{obs})^2} \quad (5)$$

$$RMSE_{Test} = \sqrt{\frac{\sum_{i=1}^k (y_j^{obs} - y_j^{pred})^2}{k}} \quad (6)$$

where y_j^{obs} is the experimental (observed) value of the property for the i^{th}/j^{th} compound; y_j^{pred} and y_j^{predcv} are the predicted value for i^{th}/j^{th} compound in the training and cross-validation set, respectively; $\tilde{y}$ and $\hat{y}$ are the mean experimental value of the property in the training and Test set, respectively; n and k are the numbers of compounds in the training and test set, respectively.

Additionally, an applicability domain (AD) for the models is calculated using the leverage approach to verify the predictive reliability of each model for the current class of investigated structures.³⁴ The Williams plot is utilized to visualize the AD for each QSPR model. The Williams plot of standardized cross-validated residuals (RES) vs. leverage (Hat diagonal) values (HAT) helps to depict both the response outliers (Y outliers) and structurally influential compounds (X outliers) in each developed model.

CG-MD Simulations. The CG-MD simulations are performed to evaluate the role of key structural and functional features informed by the QSPR models on the T_g of polymers. We employ CG models of polymers having three representative chain structures i.e., linear chain without any grafted side chain, linear chains with two different side-chain grafting densities f . The backbones of all three polymer models consist of 20 CG beads. Each side chain attached to the

backbone consists of three CG beads. The bulk simulation boxes of all three models consist of 500 polymer chains. All CG-MD simulations are carried out using the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS)³⁵ software.

The force field components of the CG models have the contributions of bonded interactions including the bonds and angles, and the non-bonded interaction of the CG beads. The potential energy of the bond stretching is estimated with a harmonic function as follows:

$$U_{bond}(r) = K(r - r_0)^2 \quad (7)$$

where $r_0 = 0.99 \sigma$ is the equilibrium bond length and $K = 2500 \varepsilon/\sigma^2$ is the spring force constant, which is consistent with previously explored branched polymers.³⁶ The potential of the angles is captured via a cosine angle style as follows:

$$U_{angle}(r) = K_\theta[1 + \cos(\theta)] \quad (8)$$

where K_θ is the chain stiffness constant that governs the flexibility of the backbone and side chain. Here, we use a fixed $K_\theta = 0.5 \varepsilon$ for the side chains and $K_\theta = 1.5 \varepsilon$ for the backbone in $f = 0.5$ and $f = 1$ models. However, for the linear chain polymer model ($f = 0$), we systematically vary K_θ of the chains (i.e., $K_\theta = \{0.5, 1.0, 2.5\} \varepsilon$) to explore the effect of the chain stiffness on T_g . The non-bonded interactions are captured by a standard 12-6 Lennard-Jones (LJ) potential function (Eq. 9):

$$U_{nb}(r) = 4\varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] \quad r < r_c \quad (9)$$

where ε and σ denote the energy well depth and distance where U_{nb} is zero, respectively; the cutoff distance is $r_c = 2.5 \sigma$. The energy well depth ε is considered as a key parameter that is related to the cohesive interaction strength of the polymer, which has a significant influence on the

thermomechanical properties. Here, we consider ε_b for backbone beads and ε_s for the side-chain beads separately to explore the effects of cohesive interactions on the T_g .

In the CG modeling, T_g is calculated based on the isothermal-isobaric (NPT) ensemble under constant zero pressure. The density of the simulation box after equilibration is determined systematically over a wide range of temperatures. Two linear lines are fitted to the density data in the low- T and high- T regimes, and intersection marks the magnitude of T_g .

3. RESULTS AND DISCUSSION

Table 1. The list of QSPR models (i.e., 1-10 variable models), statistical performance parameters and associated descriptors for each model.

Model No.	Descriptors	R^2_{Train}	$RMSE_C$	Q^2_{Train}	R^2_{Test}	$RMSE_{Test}$	F
1	Ui	0.34	0.10	0.31	0.47	0.08	40.50
2	AVS_B(e), SsssSiH	0.45	0.09	0.42	0.62	0.07	31.00
3	AVS_B(e), SsssSiH, P1u	0.55	0.08	0.51	0.72	0.06	31.09
4	AVS_B(e), Ke, nOxiranes, F01[C-Si]	0.65	0.08	0.61	0.71	0.06	34.54
5	AVS_B(e), Mor04v, RARS, nOxiranes, SsssSiH	0.68	0.07	0.65	0.70	0.06	32.10
6	AVS_B(e), nCt, RARS, nOxiranes, SsssSiH, F02[N-O]	0.73	0.07	0.70	0.72	0.06	33.20
7	AVS_B(e), Mor06v, nCt, RARS, nOxiranes, SsssSiH, B02[N-O]	0.75	0.06	0.72	0.74	0.06	30.80
8	AVS_B(e), SM14_EA(bo), Mor20v, nCt, RARS, L2s, nOxiranes, SsssSiH,	0.79	0.06	0.74	0.57	0.09	32.70
9	AVS_B(e), SM10_EA(bo), Mor28v, JGI3, RARS, nOxiranes, SsssSiH, H-050, B03[C-N]	0.83	0.05	0.79	0.53	0.10	36.80
10	AVS_B(e), SM08_EA(bo), SpDiam_B(e) Mor28v, JGI3, RARS, nOxiranes, SsssSiH, CATS2D_04_DL, B03[C-N]	0.84	0.05	0.80	0.51	0.11	37.15

First, different QSPR models for T_g prediction were developed and evaluated using the experimental dataset of over 100 polymers. The model building was carried out systematically by selecting the best set of descriptors, followed by assessing the regression coefficients of the training and test sets. The essential molecular descriptors and statistical parameters for each of the tested QSPR models (i.e., 1-10 variable models) are summarized in **Table 1**. From **Figure 2**, it is observed that the $R^2_{Training}$ values show an increasing trend with an increasing number of descriptors in the QSPR models. Similarly, the correlation coefficient of the test set R^2_{Test} exhibits an analogous trend of training set but only up to the 7-variable model and then shows a plummeting behavior, which indicates the reduced predictive capability of QSPR models (8-10 variable models) due to overfitting issues. The correlation coefficients ($R^2_{Training}$ and $Q^2_{Training}$) of the training set for higher variable models (8 to 10) still show an increasing trend, but its predictive capability for the test set R^2_{Test} decreases (**Figure 2**). Based on that, the 7-variable model that yields higher values of both $R^2_{Training}$ and R^2_{Test} has been selected as the most robust model among all ten tested models. The best 7-variable model for prediction of T_g can be analytically described as follows:

$$\begin{aligned} \text{Log } T_g = & 0.42(\pm 0.03)\text{AVS_B(e)} + 0.05(\pm 0.02)\text{Mor06v} + 0.60 (\pm 0.08) \text{RARS} \\ & + 0.05 (\pm 0.01) \text{nCt} - 0.32(\pm 0.05)\text{nOxiranes} + 0.20 (\pm 0.03)\text{SsssSiH} \\ & + 0.11 (\pm 0.03)\text{B02[N - O]} + 0.650(\pm 0.17) \end{aligned} \quad (10)$$

It is worth noting that the 7-variable model (Eq. 10) is a transparent, validated and reproducible model that can be applied to predict T_g values for other polymers (additional information is given in SI file).

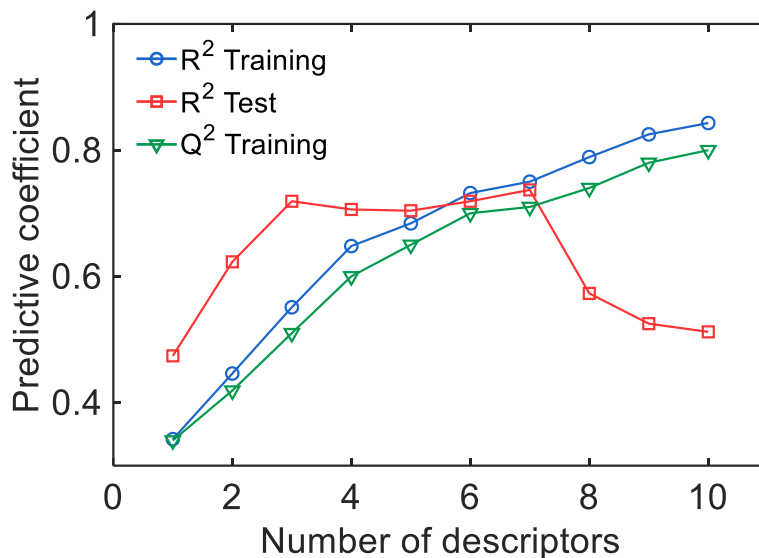


Figure 2. Statistical analyses on R^2 and Q^2 as a function of number of descriptors for QSPR models with 1-10 descriptors for training and external test sets: the regression coefficients of the training set $R^2_{Training}$ (blue circles), the cross validation by “leave-one-out” method $Q^2_{Training}$ (green triangle), and external test set R^2_{Test} (red squares).

Table 2. A list of descriptors included in the 7-variable QSPR model in Eq. 10, and their description.

Descriptor	Description	Type
AVS_B(e)	Average vertex sum from Burden matrix weighted by Sanderson electronegativity	2D matrix-based descriptors
Mor06v	3D-MoRSE signal 06 / weighted by van der Waals volume	3D-MoRSE descriptors
RARS	R matrix average row sum	gateway descriptors
nCt	Number of total tertiary C(sp3)	functional group count
nOxiranes	Number of ethylene oxide groups	functional group count
B02[N-O]	Presence/absence of N-O at topological distance 2	2D Atom pairs
SsssSiH	Sum of sssSiH E-states	atom type Estate indices

It is imperative to explicitly interpret the molecular descriptors in the best QSPR model for T_g predictions. As listed in **Table 2**, the best QSPR model (Eq. 10) can be quantitatively described by seven selected descriptors as a 2D-matrix, 3DMoRSE, gateway, functional, atom pair, and electro-topological index descriptors. AVS_B(e) is a burden matrix weighted by Sanderson electronegativity, which reveals that electronegativity plays a crucial role in T_g of polymers. This descriptor is strongly related to the polar nature of bonds and the positive influence of electronegativity indicates that the presence of electronegative atoms (e.g. fluorine, oxygen, and nitrogen) in the monomer increases T_g of polymer.³⁷ The presence of atom pairs N-O as described by B02 increases the T_g index value, which relates its ability to form intramolecular and intermolecular hydrogen. Another similar contributing descriptor (SsssSiH) for increasing the cohesive energy of polymer is electro topological (E-state) index, a measure of two unified attributes (electronic and topological); SsssSiH or E-state index value increases with a number of electron-rich atoms in a molecule, facilitating enhancement in intermolecular interactions. Several examples of correlations between E-state index value and physical and biological phenomena have been reported before for similarity search of anti-inflammatory drugs and its utility in the toxicity modeling for a set of amide-based herbicides.^{38,39}

The presence of RARS⁴⁰ (gateway descriptors) and nCt⁴¹ (total number of sp^3 carbon) molecular descriptors in the best QSPR model indicates that T_g is sensitive to molecular branching (e.g., existence of side-chain groups). Several attempts were made to understand the effect of molecular branching on T_g of polymers. Zhu *et al.*⁴² showed that T_g first increases with the degree of branching (DB), passes through a maximum, and then decreases sharply for hyper branched polyethers; however, the T_g of polyethylene is observed to decrease with increase of degree of branching.⁴³ These contrast observations suggest that molecular branching shares a complex

relationship with T_g of amorphous polymers. Interestingly, a negative effect on T_g is exhibited by the presence of the ethylene oxide group (nOxiranes) on the backbone of polymer chain. Previous studies have shown that T_g can be suppressed by increasing the number of oxygen atoms in the backbone of polymer chain with reduced stiffness.^{44,45} The next descriptor Mor06v, which represents a weighted van der Waals volume, describes a space occupied by a molecule that is impenetrable to other molecules. Mor06v indicates that the size and volume of the molecule have a positive contribution to T_g . The relative influences of these descriptors on T_g will be further discussed.

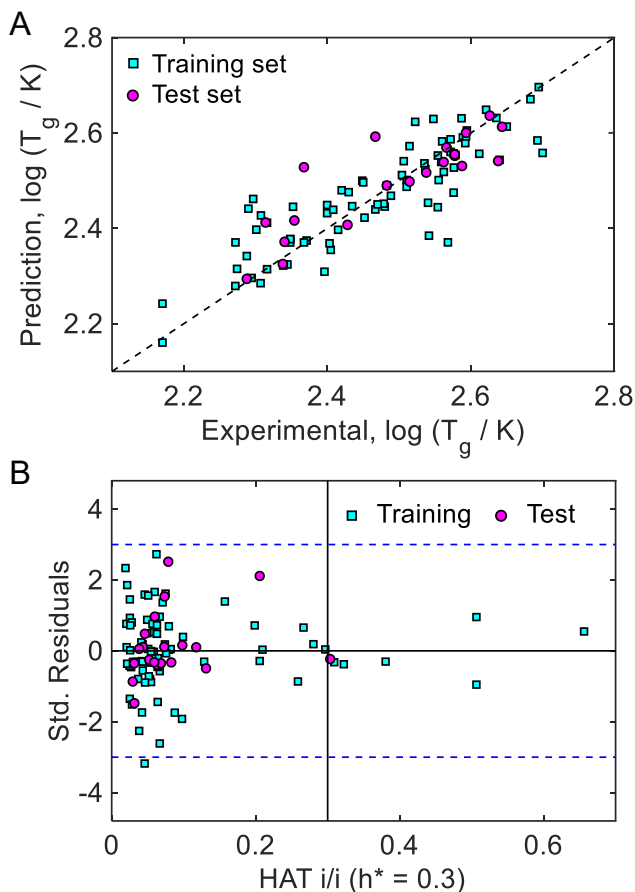


Figure 3. (A) A correlation plot between the observed and predicted values of T_g of polymers in the 7-variable QSPR model. (B) Williams plot of standardized residual versus leverage for

evaluation of the applicability domain for the 7-variable model, where the chemical compounds of training and test sets are shown. The blue dashed and solid black lines correspond to ± 3 standardized residual (σ) and warning leverage value ($h^* = 0.3$), respectively.

It is important to estimate the robustness of the model and its predictive ability. In this regard, several statistical parameters were analyzed to estimate the performance of the models. **Figure 3A** shows a correlation between the experimental T_g and the predicted T_g based on the 7-variable model. Given the complexity of T_g and diversity of applied dataset, the level of correlation between the predicted and observed T_g values in a 7-variable model (Eq. 10) indicates a reasonably good predictive capability of the 7-variable model, with $R^2 = 0.75$ (Training) and $R^2 = 0.73$ (Test) for a diverse set of 100 different polymers. **Figure 3B** shows the Williams plot that represents the applicability domain (AD) of the QSPR model, allowing for a graphical detection of both the outliers for the response and structurally different chemicals in the model. The graph represents standardized residuals (σ), the measure of the strength of the difference between observed and predicted values on the Y-axis and the leverage value on X-axis. Here, the observations with standardized residuals beyond the -3σ to $+3\sigma$ range are considered as outlier responses. A leverage value (HAT) represents the degree of influence of the chemical on the model. A higher leverage value ($HAT > h^*$) means the compound exerts a greater influence on the QSPR model. Thus, the Williams plot shown in **Figure 3B** indicates that nearly all points are located within 3σ of error limit with only one outlier, demonstrating the validity of the 7-variable model. There are several polymeric units that show higher error estimation deviations than others (5, 11, 44, 45, 56, 92 98). Analysis of the impact of descriptors in the model to these polymeric units T_g value shows the following. In most of the cases, there is a combined influence of the descriptors' values on the estimation error of the T_g value of these polymeric units. Thus, in the case of 5, 11, 44, 45, 56 and

92 the larger deviation of AVS_B(e), Mor06v and RARS descriptors resulted in a larger deviation of overall residual value for T_g (**Figures S1-S3** in SI). In the case of polymer unit 45, there is a large contribution of nCt descriptor's error deviation to the T_g value that also led to the larger deviation in the property estimation (**Figure S3**). In the case of polymeric unit 98, a large error estimation in descriptor Mor06v led to the larger error of estimation in T_g value. A detailed information on the contribution of each descriptor's error of estimation value to T_g value's error of estimation is shown in **Table S4** and **Figure S3**.

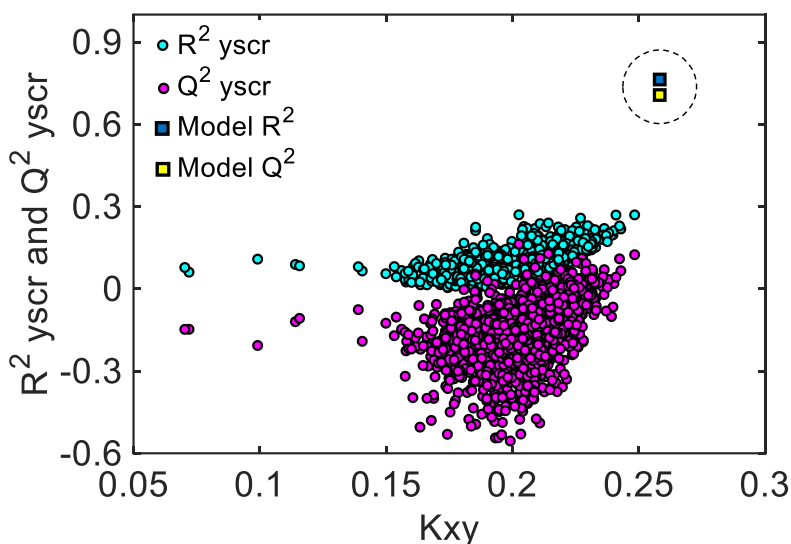


Figure 4. Representation of the y -scrambling plot: blue and yellow dots (highlighted by dashed circle) represent R^2 and Q^2 values of the 7-variable QSPR model, respectively; cyan and purple dots represent R^2_{yscr} and Q^2_{yscr} of models based on randomized data.

In **Figure 4** represented a y -scrambling plot, an important validation technique, which underlines the robustness and uniqueness of the best QSPR model. The y -scrambling plot is obtained by scrambling the y values (i.e., experimental T_g values) in a random fashion. In this

randomization process, 2,000 simulations per model are performed, where none of the models has any acceptable correlation in comparison with the best 7-variable model (Eq. 10). The best model has much higher R^2 and Q^2 values (data points marked by the square symbols in the plot) than those of all other simulated (unrealistic, scrambled) models. This analysis again confirms that the developed QSPR model is a clear case of robustness and not mere coincidence.

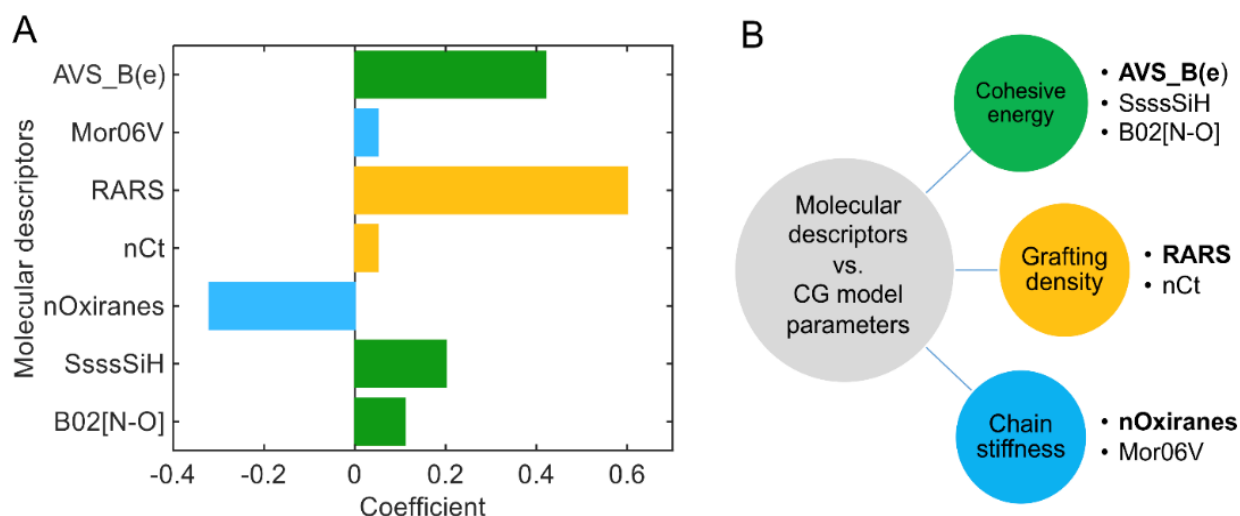


Figure 5. (A) The coefficients of different descriptors in the 7-variable model. The greater absolute magnitude of the coefficient indicates higher relative importance for T_g . All descriptors, except nOxiranes, show a positive influence on T_g index. **(B)** Categorizing best molecular descriptors from the 7-variable model into three physical parameters of CG models. The dominant influential descriptor in the QSPR model in each category is highlighted as bold.

As above mentioned, the molecular descriptors identified by QSPR models suggest that physico-chemical and electro-topological parameters, such as polarity of bonds and hydrogen bonding, branching intensity, presence of bulky atoms or groups, and chain stiffness play pivotal roles in determining T_g of diverse polymers. As quantified by the magnitudes of coefficients, **Figure 5A** shows the relative influence of these descriptors in the QSPR model on T_g index value.

The positive or negative influence of each descriptor on T_g index value is represented by the positive or negative sign of the coefficient's value of each descriptor. All the descriptors in **Figure 5A**, except nOxiranes (i.e., presence of ethylene oxide in repeating unit), show a positive contribution to T_g index value. Among these descriptors, RARS, AVS_B(e), and nOxiranes yield higher degrees of relative importance associated with T_g due to their greater absolute magnitudes of coefficients compared to the others.

To better understand the roles of these key molecular descriptors in governing the T_g of polymers, the descriptors identified by the QSPR model (Eq. 10 and **Figure 5A**) are grouped into three major categories as illustrated in **Figure 5B**:

- Intermolecular cohesive interactions (i.e., AVS_B(e), B02 [N-O], SsssSiH);
- Backbone chain stiffness (i.e., nOxiranes, Mor06V);
- Grafting or branching density of side chain (RARS and nCt).

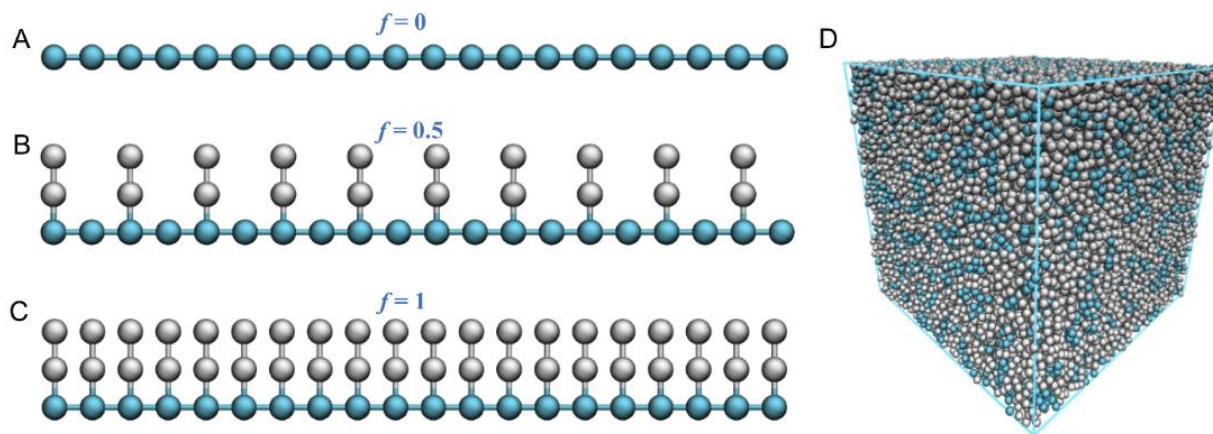


Figure 6. Illustrations of CG models of polymers: (A) a linear polymer chain with $f = 0$, (B) an alternatively branched polymer chain with $f = 0.5$, and (C) a fully branched polymer chain with $f = 1$. (D) Snapshot of the bulk simulation box consisting of polymer chains.

Then, to gain fundamental insights into the roles of descriptors informed from the QSPR model, we employed the CG-MD simulations to explore and validate their influences on T_g of CG polymer models. Specifically, the three major categories of descriptors can be highly associated with the following three CG model parameters: cohesive interaction strength ε (i.e., ε_b for backbone beads and ε_s for the side-chain beads), backbone chain stiffness K_θ , and grafting density f (= number of grafted backbone beads/number of backbone beads). Here, we consider three representative architectural structures of the polymer chains in the CG-MD simulations. As shown in **Figure 6A**, the CG model with $f=0$ is a linear polymer chain without any side chain. The CG models with $f=0.5$ and $f=1$, as illustrated in **Figure 6B** and **C**, are partially and fully branched polymers, respectively, where each side chain comprised of three CG beads is grafted to the backbone. Each backbone chain of the polymer consists of 20 CG beads and bulk simulation boxes of all three models consist of 500 chains.

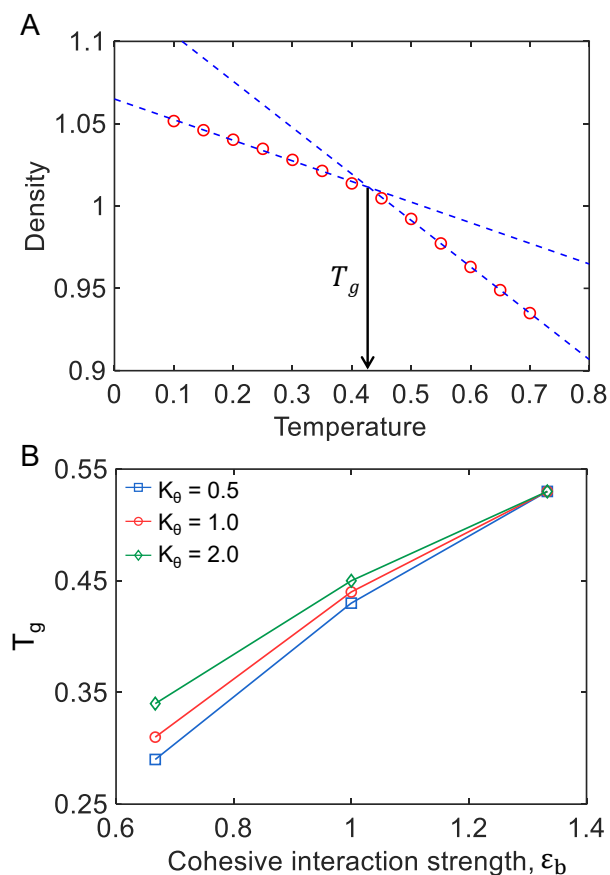


Figure 7. (A) Density versus temperature for calculation of T_g of the CG polymer model with $f=0$, $K_\theta = 0.5$, and $\epsilon_b = 1$. **(B)** T_g as a function of the cohesive interaction strength ϵ_b of the CG model ($f=0$) for different backbone chain stiffness K_θ of the chain.

We begin with investigating the influence of cohesive interaction strength and chain stiffness on T_g of the CG model with $f = 0$. **Figure 7A** shows the typical T_g determination from the CG modeling by linearly fitting the density vs. temperature data in the low and high temperature regimes, respectively, where the intersection marks the T_g . **Figure 7B** shows the T_g variations with different cohesive interaction strengths ϵ_b as well as different backbone chain stiffness K_θ . It is observed that both the cohesive interaction strength parameter ϵ_b and chain stiffness K_θ have considerable influence on the T_g of CG polymer, as it was found by the QSPR

model (Eq. 10). When ε_b becomes larger, T_g of the polymer increases significantly because of a suppression of segmental mobility. Similarly, for a given ε_b , increasing the K_θ causes the T_g to increase. Although increasing both K_θ and ε_b parameters will increase T_g for $f=0$ model, ε_b plays a more dominant role in controlling T_g compared to K_θ . This observation is consistent with a recent CG-MD study by Xu and co-workers,⁴⁶ who showed that the T_g of polymer melts with an ideal chain (i.e., no angles, dihedrals or side chains) tends to increase as cohesive energy becomes larger. In another MD study of glassy polycarbonate (PC),⁴⁷ it has been shown that increasing the cohesive energy significantly enhances mechanical properties (i.e., shear modulus and yielding stress) and dynamical heterogeneity associated with the glass-forming process.^{48,49}

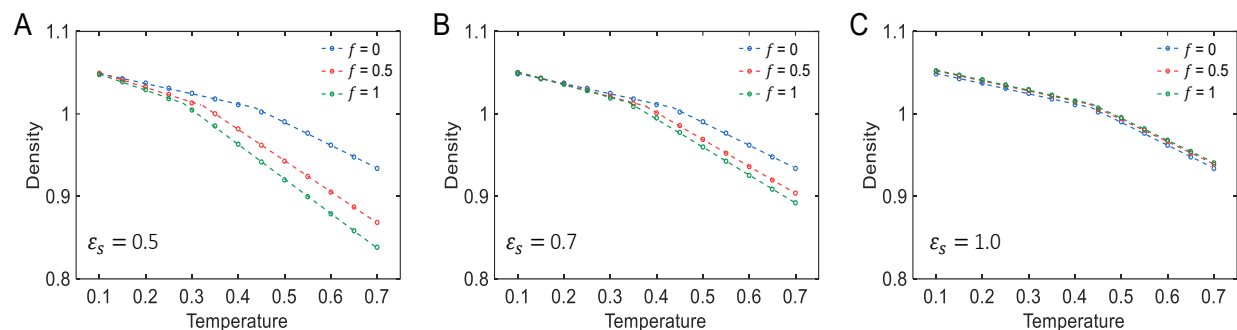


Figure 8. Temperature dependent density for calculations of T_g of CG polymer models with cohesive interaction strength $\varepsilon_b = 1.0$ of backbone and varying cohesive interaction strength ε_s of (A), 0.5 (B), 0.7 and (C) 1.0 of side chains for different grafting density f . A fixed $K_\theta = 0.5 \varepsilon$ for the side chains and $K_\theta = 1.5 \varepsilon$ for the backbone are used in the CG models.

We proceed to explore the influence of grafting density f on the T_g of CG polymers having grafted side chains, which is relevant to the descriptors associated with molecular branching identified through the QSPR modeling. From the simulations (**Figure 8**), it appears that grafting

the side-chain groups to the chain backbone considerably affect T_g of polymer. In particular, as shown in **Figure 8A**, when the cohesive interaction strength ϵ_s of the side-chain beads is much smaller than ϵ_b of the backbone, increasing f will decrease the T_g . This reducing effect of the side groups on the T_g of the current CG bulk models resembled the influence of alkyl side group length on the T_g and mechanical properties of conjugated polymers, where increasing side chain length was found to decrease the T_g and elastic modulus.^{50,51} However, **Figure 8B** and **C** show that when ϵ_s parameter of the side-chain beads becomes larger and gets close to the ϵ_b , the reducing effect of f on T_g will diminish.

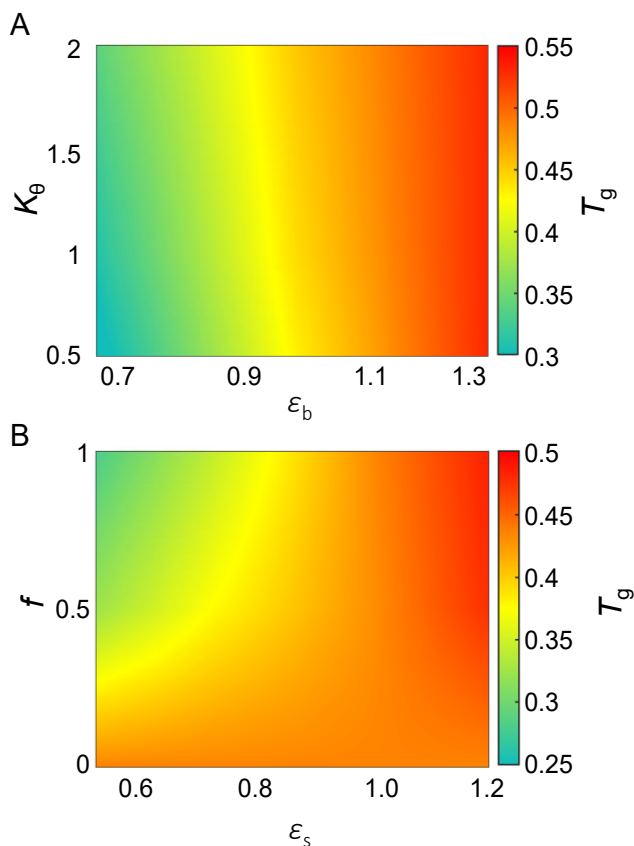


Figure 9. (A) Contour plot of T_g in the plane of chain rigidity K_θ vs. cohesive interaction strength ϵ_b of CG model with $f = 0$. (B) Contour plot of T_g in the plane of grafting density f vs. cohesive interaction strength of the side chains ϵ_s .

To better understand the couple influences of cohesive energy and other explored parameters, we summarized T_g results from MD simulations for both linear and grafted polymer models. **Figure 9A** shows a contour plot of mutual influence of backbone stiffness K_θ and cohesive interaction strength ε_b on the T_g of linear chain model ($f=0$), in which for larger values of ε_b , altering the values of K_θ is found to be insignificant in controlling the T_g . For the grafted polymer models (**Figure 9B**), we observe that as the cohesive energy of the side groups ε_s becomes smaller than that of the backbone ε_b , increasing f tends to decrease the T_g . However, when ε_s becomes larger than ε_b , an opposite trend of T_g with increasing f is observed – more grafted side groups are found to increase the T_g . These results not only demonstrate the complexed influences of these key molecular parameters informed from the QSAR model, but also provide valuable insight into the vital role of cohesive energy in the glass formation of the polymers. In previous investigations based on the generalized entropy theory (GET), Dudowicz and co-workers⁵² showed that depending on the stiffnesses of side chain and backbone, side-chain groups can strongly influence the T_g . For polymers with a flexible backbone and stiff side chain, increasing of the side-chain length cause T_g to increase. However, when flexible side chains are grafted to relatively stiffer backbones, T_g tends to decrease when the side-chain length is increased. This is in good agreement with the findings of the current study, where the backbone chain of the investigated CG model has a relatively high stiffness compared to the grafted side chains. In another relevant study of star polymers by Fan *et al.*,⁵³ it is reported that increasing the arm length leads to a T_g drop as the number of arms exceeding a critical value.

The CG-MD simulations offer valuable insights into the QSPR models by providing a mechanistic interpretation and systematic dependence of key molecular features that govern the T_g . Through the QSPR modeling, seven most essential molecular descriptors are identified via

statistical analysis based on machine learning, which can be described by three major CG model parameters; cohesive energy, chain stiffness, and grafting density. The CG-MD simulations confirm the significant roles of these molecular descriptions in influencing T_g . The integration of machine learning-based QSPR modeling and MD simulations outlined here is an inspiring pathway for predicting and understanding the glass-forming behaviors of polymers having diverse chemistry. It is worth noting that the presented modeling framework can be further improved in the future. For instance, a more robust QSPR model can be developed by applying the ‘feedback loop’ approach, where more physics-based descriptors can be considered and analyzed explicitly. The predictive capability of the developed QSPR model can be improved by employing extended and diverse datasets for both training and validation sets. Besides, the QSPR models can also guide the development of more robust CG models for better mechanistical simulation, which in return can give a valuable feedback to QSAR modeling for a rapid quantitative validation. For the future work, it will be useful to associate QSAR descriptors with the molecular parameters of MD models in a more quantitative way for a better validation of predictive relationships developed by the QSAR modeling. On the other hand, the essential feature descriptors identified from QSAR modeling will improve the development of chemistry specific CG models for polymers, whose physico-chemical properties can be added to the QSPR library to enrich the feature data sets and further improve the QSPR model predictions. We believe that the results of the current study will pave a way for future steps to better integration of these two robust approaches for properties’ predictions of complex materials. By harnessing the power of this unparalleled computational efficiency provided by this framework, it is expected that other relevant properties can be predicted for not only the polymeric materials but also other classes of organic and inorganic materials.

4. CONCLUSION

In this study, the combination of two techniques, machine learning-based cheminformatics and CG-MD modeling, is applied to investigate the factors that affect T_g and validate the predictions for a diverse set of 100 polymers. To this end, a set of 1- to 10-variable models is developed by applying the machine learning-based QSPR modeling. After testing and validating, the 7-variable QSPR model is identified, yielding the best predictive performance, which is further confirmed by additional validation techniques, including $Q^2_{Training}$ and y -scrambling analyses. Among the physico-chemical and electro-topological descriptors in the QSPR model, AVS_B(e), RARS, and nOxiranes are found as the most influential ones that govern the T_g of polymers in the context of structure-property relationships. Informed by the QSPR model, CG-MD simulations are performed to further delineate mechanistic interpretation and systematic dependence of these influential molecular features on T_g by investigating three major CG model parameters, namely the cohesive interaction ε , chain stiffness K_θ , and grafting density f . The CG-MD simulation results demonstrate the importance of these molecular descriptors in T_g of polymers, where their influences are highly coupled. This synergistic framework of integrating the cheminformatics and CG-MD simulations provides valuable insights into the roles of key molecular features influencing the T_g of polymers, which can be further applied for the prediction of different properties of polymers and glass-forming materials.

ACKNOWLEDGEMENTS

A.K. and B.R. thank the Department of Coatings and Polymeric Materials at North Dakota State University (NDSU) for providing the Lowell Wood funding. The authors thank Prof. Paola Gramatica for generously provided a free license for the QSARINS software. A.A. and W.X. acknowledge the support from the North Dakota Established Program to Stimulate Competitive Research (ND EPSCoR Award #IIA-1355466) through the New Faculty Award. A.A. and W.X. acknowledge support from the Department of Civil & Environmental Engineering and College of Engineering at North Dakota State University (NDSU). Supercomputing support from CCAST Thunder HPC System at NDSU is acknowledged.

Author Contributions

A.K. and A.A. contributed equally to the project. B.R. and W.X. conceived the idea and directed the project. A.K. performed the cheminformatics modeling. A.A. conducted molecular dynamics simulations. All authors contributed to writing of the manuscript.

Data Availability

The raw data required to reproduce these findings are available to download from the Supplementary Information (SI) file. The processed data required to reproduce these findings are available to download from Supplementary Information (SI) file.

REFERENCES

- (1) Stillinger, F. H.; Debenedetti, P. G. Glass Transition Thermodynamics and Kinetics. *Annu. Rev. Condens. Matter Phys.* **2013**, 4 (1), 263–285.
- (2) Debenedetti, P. G.; Stillinger, F. H. Supercooled Liquids and the Glass Transition. *Nature* **2001**, 410 (6825), 259–267.
- (3) Angell, C. A.; Ngai, K. L.; McKenna, G. B.; McMillan, P. F.; Martin, S. W. Relaxation in Glassforming Liquids and Amorphous Solids. *J. Appl. Phys.* **2000**, 88 (6), 3113–3157.
- (4) Mauro, J. C.; Smedskjaer, M. M. Statistical Mechanics of Glass. *Journal of Non-Crystalline Solids*. **2014**, pp 41–53.
- (5) Joseph D. Menczel, R. B. P. *Thermal Analysis of Polymer, Fundamentals and Applications*; John Wiley & Sons, **2009**; Vol. 6.
- (6) Mitchell B.O., J. B. O. Machine Learning Methods in Chemoinformatics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, 4 (5), 468–481.
- (7) Katritzky, A. R.; Sild, S.; Karelson, M. Correlation and Prediction of the Refractive Indices of Polymers by QSPR. *J. Chem. Inf. Comput. Sci.* **1998**, 38 (6), 1171–1176.
- (8) Liu, A.; Wang, X.; Wang, L.; Wang, H.; Wang, H. Prediction of Dielectric Constants and Glass Transition Temperatures of Polymers by Quantitative Structure Property Relationships. *Eur. Polym. J.* **2007**, 43 (3), 989–995.
- (9) Gharagheizi, F. QSPR Analysis for Intrinsic Viscosity of Polymer Solutions by Means of GA-MLR and RBFNN. *Comput. Mater. Sci.* **2007**, 40 (1), 159–167.

- (10) Patel, H. C.; Tokarski, J. S.; Hopfinger, A. J. Molecular Modeling of Polymers 16. Gaseous Diffusion in Polymers: A Quantitative Structure-Property Relationship (QSPR) Analysis. *Pharmaceutical Research*. 1997, pp 1349–1354.
- (11) Verma J, Khedkar VM, Coutinho EC. 3D-QSAR in drug design - A Review. *Current Topics in Medicinal Chemistry*. **2010**, 10(1), 95-115.
- (13, 12) Bicerano, J. *Prediction of Polymer Properties.*; CRC press: New York: Marcel Dekke, **2002**.
- (14, 13) Katritzky, A. R.; Rachwal, P.; Law, K. W.; Karelson, M.; Lobanov, V. S. Prediction of Polymer Glass Transition Temperatures Using a General Quantitative Structure-Property Relationship Treatment. *J. Chem. Inf. Comput. Sci.* **1996**, 36 (4), 879–884.
- (15, 14) Katritzky, A. R.; Sild, S.; Lobanov, V.; Karelson, M. Quantitative Structure - Property Relationship (QSPR) Correlation of Glass Transition Temperatures of High Molecular Weight Polymers. *J. Chem. Inf. Comput. Sci.* **1998**, 38 (2), 300–304.
- (16,15) García-Domenech, R.; de Julián-Ortiz, J. V. Prediction of Indices of Refraction and Glass Transition Temperatures of Linear Polymers by Using Graph Theoretical Indices. *J. Phys. Chem. B* **2002**, 106 (6), 1501–1507.
- (17, 16) Mercader, A. G.; Duchowicz, P. R. Encoding Alternatives for the Prediction of Polyacrylates Glass Transition Temperature by Quantitative Structure-Property Relationships. *Mater. Chem. Phys.* **2016**, 172, 158–164.
- (17) Miccio, L. A.; Schwartz, G. A. From chemical structure to quantitative polymer properties

- prediction through convolutional neural networks. *Polymer*, **2020**, 193, 122341
- (18) Miccio, L. A.; Schwartz, G. A. Localizing and quantifying the intra-monomer contributions to the glass transition temperature using artificial neural networks. *Polymer*, **2020**, 203, 122786.
- (19) Chen, M.; Jabeen, F.; Rasulev, B.; Ossowski, M.; Boudjouk, P. A Computational Structure–Property Relationship Study of Glass Transition Temperatures for a Diverse Set of Polymers. *J. Polym. Sci. Part B Polym. Phys.* **2018**, 56 (11), 877–885.
- (20) Liu, W.; Yi, P.; Tang, Z. QSPR Models for Various Properties of Polymethacrylates Based on Quantum Chemical Descriptors. *QSAR Comb. Sci.* **2006**, 25 (10), 936–943.
- (21) Roy, K.; Mitra, I.; Kar, S.; Ojha, P. K.; Das, R. N.; Kabir, H. Comparative Studies on Some Metrics for External Validation of QSPR Models. *J. Chem. Inf. Model.* **2012**, 52 (2), 396–408.
- (22) Izvekov, S.; Voth, G. A. A Multiscale Coarse-Graining Method for Biomolecular Systems. *J. Phys. Chem. B* **2005**, 109 (7), 2469–2473.
- (23) Buchholz, J.; Paul, W.; Varnik, F.; Binder, K. Cooling Rate Dependence of the Glass Transition Temperature of Polymer Melts: Molecular Dynamics Study. *J. Chem. Phys.* **2002**, 117 (15), 7364–7372.
- (24) Yang, L.; Srolovitz, D. J.; Yee, A. F. Molecular Dynamics Study of Isobaric and Isochoric Glass Transitions in a Model Amorphous Polymer. *J. Chem. Phys.* **1999**, 110 (14), 7058–7069.
- (25) Rigby, D.; Roe, R. Molecular Dynamics Simulation of Polymer Liquid and Glass. I. Glass

Transition. *J. Chem. Phys.* **1987**, 87 (12), 7285–7292.

- (26) J. Brandrup, E.H. Immergut, E. A. G. *Polymer Handbook*, 2nd ed.; John Wiley and Sons, 1975.
- (27) Advanced Chemistry Development, I. ChemSketch. Toronto, On, Canada, 2014.
- (28) Todeschini, R.; Consonni, V.; Pavan, M. DRAGON—Software for the Calculation of Molecular Descriptors. Release 1.12 for Windows. *Web version*. Milan, Italy, 2001.
- (29) Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim, NY, 2000; Vol. 11. <https://doi.org/10.1002/9783527613106>.
- (30) J. Devillers. *Genetic Algorithms in Molecular Modeling*; Academic Press: London, San Diego, CA, 1996.
- (31) Ahmed, L.; Rasulev, B.; Turabekova, M.; Leszczynska, D.; Leszczynski, J. Receptor- and Ligand-Based Study of Fullerene Analogues: Comprehensive Computational Approach Including Quantum-Chemical, QSAR and Molecular Docking Simulations. *Org. Biomol. Chem.* **2013**, 11 (35), 5798–5808.
- (32) Gramatica, P.; Cassani, S.; Chirico, N. QSARINS-Chem: Insubria Datasets and New QSAR/QSPR Models for Environmental Pollutants in QSARINS. *J. Comput. Chem.* **2014**, 35 (13), 1036–1044.
- (33) Rasulev, B. F.; Saidkhodzhaev, A. I.; Nazrullaev, S. S.; Akhmedkhodzhaeva, K. S.; Khushbaktova, Z. A.; Leszczynski, J. Molecular Modelling and QSAR Analysis of the Estrogenic Activity of Terpenoids Isolated from Ferula Plants. *SAR QSAR Environ. Res.* **2007**, 18 (7–8), 663–673.

- (34) Gramatica, P.; Giani, E.; Papa, E. Statistical External Validation and Consensus Modeling: A QSPR Case Study for Koc Prediction. *J. Mol. Graph. Model.* **2007**, *25* (6), 755–766.
- (35) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117* (1), 1–19.
- (36) Chremos, A.; Douglas, J. F. A Comparative Study of Thermodynamic, Conformational, and Structural Properties of Bottlebrush with Star and Ring Polymer Melts. *J. Chem. Phys.* **2018**, *149* (4), 44904.
- (37) Cao, C.; Lin, Y. Correlation between the Glass Transition Temperatures and Repeating Unit Structure for High Molecular Weight Polymers. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (2), 643–650.
- (38) Gough, J. D.; Hall, L. H. Modeling the Toxicity of Amide Herbicides Using the Electrotopological State. *Environ. Toxicol. Chem.* **1999**, *18* (5), 1069–1075.
- (39) Hall, L. H.; Kier, L. B. The E-State as the Basis for Molecular Structure Space Definition and Structure Similarity. *J. Chem. Inf. Comput. Sci.* **2000**, *40* (3), 784–791.
- (40) Consonni, V.; Todeschini, R.; Pavan, M. Structure/Response Correlations and Similarity/Diversity Analysis by GETAWAY Descriptors. 1. Theory of the Novel 3D Molecular Descriptors. *J. Chem. Inf. Comput. Sci.* **2002**, *42* (3), 682–692.
- (41) Khan, P. M.; Roy, K. QSPR Modelling for Prediction of Glass Transition Temperature of Diverse Polymers. *SAR QSAR Environ. Res.* **2018**, *29* (12), 935–956.
- (42) Zhu, Q.; Tu, C.; Shi, Y.; Zhu, X.; Yan, D.; Wu, J.; He, L.; Wang, R. Role of Branching

- Architecture on the Glass Transition of Hyperbranched Polyethers. *J. Phys. Chem. B* **2009**, *113* (17), 5777–5780.
- (43) Luo, X.; Xie, S.; Liu, J.; Hu, H.; Jiang, J.; Huang, W.; Gao, H.; Zhou, D.; Lü, Z.; Yan, D. The Relationship between the Degree of Branching and Glass Transition Temperature of Branched Polyethylene: Experiment and Simulation. *Polym. Chem.* **2014**, *5* (4), 1305–1312.
- (44) Sugiyama, F.; Kleinschmidt, A. T.; Kayser, L. V.; Rodriguez, D.; Finn, M.; Alkhadra, M. A.; Wan, J. M. H.; Ramírez, J.; Chiang, A. S. C.; Root, S. E.; Savagatrup, S.; Lipomi, D. J. Effects of Flexibility and Branching of Side Chains on the Mechanical Properties of Low-Bandgap Conjugated Polymers. *Polym. Chem.* **2018**, *9* (33), 4354–4363.
- (45) Imrie, C. T.; Karasz, F. E.; Attard, G. S. Effect of Backbone Flexibility on the Transitional Properties of Side-Chain Liquid-Crystalline Polymers. *Macromolecules* **1993**, *26* (15), 3803–3810.
- (46) Xu, W.-S.; Douglas, J. F.; Freed, K. F. Influence of Cohesive Energy on the Thermodynamic Properties of a Model Glass-Forming Polymer Melt. *Macromolecules* **2016**, *49* (21), 8341–8354.
- (47) Alesadi, A.; Xia, W. Understanding the Role of Cohesive Interaction in Mechanical Behavior of a Glassy Polymer. *Macromolecules* **2020**, *53* (7), 2754–2763.
- (48) Dudowicz, J.; Freed, K. F.; Douglas, J. F. The Glass Transition Temperature of Polymer Melts. *J. Phys. Chem. B* **2005**, *109* (45), 21285–21292.
- (49) Betancourt, B. A. P.; Hanakata, P. Z.; Starr, F. W.; Douglas, J. F. Quantitative Relations

- between Cooperative Motion, Emergent Elasticity, and Free Volume in Model Glass-Forming Polymer Materials. *Proc. Natl. Acad. Sci.* **2015**, *112* (10), 2966–2971.
- (50) Xie, R.; Weisen, A. R.; Lee, Y.; Aplan, M. A.; Fenton, A. M.; Masucci, A. E.; Kempe, F.; Sommer, M.; Pester, C. W.; Colby, R. H.; Gomez, E. D. Glass Transition Temperature from the Chemical Structure of Conjugated Polymers. *Nat. Commun.* **2020**, *11* (1), 893. <https://doi.org/10.1038/s41467-020-14656-8>.
- (51) Zhang, S.; Alesadi, A.; Selivanova, M.; Cao, Z.; Qian, Z.; Luo, S.; Galuska, L.; Teh, C.; Ocheje, M. U.; Mason, G. T.; St. Onge, P. B. J.; Zhou, D.; Rondeau-Gagné, S.; Xia, W.; Gu, X. Toward the Prediction and Control of Glass Transition Temperature for Donor–Acceptor Polymers. *Adv. Funct. Mater.* **2020**, *2002221*, 2002221. <https://doi.org/10.1002/adfm.202002221>.
- (52) Dudowicz, J.; Freed, K. F.; Douglas, J. F. Generalized Entropy Theory of Polymer Glass Formation. *Adv. Chem. Phys.* **2008**, *137*, 125.
- (53) Fan, J.; Emamy, H.; Chremos, A.; Douglas, J. F.; Starr, F. W. Dynamic Heterogeneity and Collective Motion in Star Polymer Melts. *J. Chem. Phys.* **2020**, *152* (5), 54904.