


 Cite this: *Chem. Commun.*, 2025, 61, 14705

 Received 3rd June 2025,  
 Accepted 15th August 2025

DOI: 10.1039/d5cc03072b

[rsc.li/chemcomm](https://rsc.li/chemcomm)

# Predicting the glass transition temperature of conjugated polymers *via* cheminformatics from monomer structures

 Amirhadi Alesadi,<sup>†a</sup> Zhaofan Li,<sup>†b</sup> Amara Arshad<sup>c</sup> and Wenjie Xia<sup>†\*b</sup>

**We present a cheminformatics model for predicting the glass-transition temperature ( $T_g$ ) of conjugated polymers using four interpretable molecular descriptors. The model achieves high predictive accuracy ( $R^2 \approx 0.85$ ), and molecular dynamics simulations validate the descriptor– $T_g$  relationships. This integrated framework enables rational design of conjugated polymers with tailored glass-transition properties.**

Conjugated polymers (CPs) have attracted significant interest in optoelectronic applications due to their tunable electronic, mechanical, and processability properties, which stem from infinite molecular modification possibilities.<sup>1–8</sup> Among their key characteristics, the glass-transition temperature ( $T_g$ ) plays a crucial role in determining processability, mechanical stability, and device performance.<sup>9,10</sup> While experimental measurements of  $T_g$  exist, they are often time-consuming and challenging due to the semi-crystalline nature and rigid backbones of CPs. Thus, an accurate predictive framework is essential for accelerating CP design.

Predictive frameworks for polymer materials generally fall into two categories: models based on geometric features (*e.g.*, connectivity and topology)<sup>9,11</sup> and cheminformatics-based approaches<sup>12</sup> that utilize physicochemical descriptors. In our previous work,<sup>9</sup> we developed a machine learning (ML) model trained on a diverse dataset of CPs, which achieved an  $R^2$  of  $\sim 0.85$  for  $T_g$  prediction by identifying key structural patterns such as side-chain composition and aromatic ring connectivity. However, such ML models often rely on geometric heuristics and may lack physical interpretability.

While quantitative structure-property relationship (QSPR) methods have shown excellent performance in predicting the

thermal and mechanical properties of non-conjugated polymers,<sup>13</sup> their application to CPs has mostly focused on optoelectronic performance and mechanical flexibility. In contrast, the use of QSPR to predict the  $T_g$  of CPs remains limited. It is unclear whether descriptor-based models can reliably capture  $T_g$  trends in CPs, particularly given their complex structures and rigidity effects. This motivates the present study, which combines interpretable molecular descriptors with MD simulation validation to build a physically grounded QSPR model for  $T_g$  prediction in CPs.

In this study, we build on our previous compiled dataset<sup>9</sup> of CPs to develop a QSPR model that predicts  $T_g$  directly from the chemical structure of the building block. The model identifies key descriptors related to backbone flexibility, electronic delocalization, and steric characteristics, offering insights beyond geometric heuristics used in prior ML-based models. We assess predictive performance and compare descriptor relevance to features highlighted in earlier work.<sup>9</sup> To support the physical basis of these descriptors, we also perform molecular dynamics (MD) simulations that elucidate how backbone rigidity influences chain mobility and packing, reinforcing the mechanistic understanding of  $T_g$  in CPs.

The dataset, previously used in our ML model,<sup>9</sup> consists of 154 polymers and small-molecule acceptor units, spanning a  $T_g$  range from  $-30$  °C to  $220$  °C (Fig. 1). It includes both flexible and rigid polymers, ensuring diverse structural representation. The majority of polymers exhibit  $T_g$  values between  $0$  °C and  $170$  °C, encompassing a broad spectrum of structural variations, from relatively flexible backbones influenced by side-chain mobility to more rigid systems with fused-ring structures and strong intermolecular interactions. This diversity enables a comprehensive structure– $T_g$  relationship, ensuring the QSPR model captures key thermal behavior trends. While this dataset represents one of the largest available for  $T_g$  prediction in CPs, we acknowledge its limitations in size and origin. The proposed descriptor-based framework is modular and can be readily expanded as new experimental data become available, which will further improve model accuracy and generalizability.

<sup>a</sup> Department of Civil, Construction and Environmental Engineering, North Dakota State University, Fargo, ND 58108, USA

<sup>b</sup> Department of Aerospace Engineering, Iowa State University, Ames, Iowa, 50011, USA. E-mail: wxia@iastate.edu

<sup>c</sup> Materials and Nanotechnology, North Dakota State University, Fargo, ND 58108, USA

<sup>†</sup> These authors contributed equally to this work.





Fig. 1 Distribution of the experimental glass-transition temperature ( $T_g$ ) values of conjugated polymers used in this study. The solid red line represents the kernel density estimation.

Further details on dataset distribution and molecular structures are provided in the SI (Table S1).

To select and evaluate the performance of the QSPR regression models, we use  $R^2$  and  $Q_{5\text{-fold}}^2$  as statistical measures of model accuracy and generalizability. While  $R^2$  quantifies how well the model fits the training data, it alone is not sufficient to confirm predictive reliability, as adding more descriptors can artificially inflate  $R^2$  due to overfitting. To prevent this issue, we develop multiple QSPR models, each incorporating between 1 and 10 descriptors, and assess their performance using training ( $R^2$ ), cross-validation ( $Q_{5\text{-fold}}^2$ ), and test set ( $R^2$ ) results. As shown in Fig. 2A, increasing the number of descriptors initially improves both  $R^2$  and  $Q_{5\text{-fold}}^2$ , indicating a better model fit and stronger generalizability. However, beyond four descriptors, test set  $R^2$  declines, suggesting that additional descriptors do

not contribute meaningful information and instead introduce noise, leading to overfitting. The model with four descriptors achieves the best balance between predictive power and generalizability, making it the optimal choice for  $T_g$  prediction in CPs:

$$T_g (\text{°C}) = -130.14 - 1240.36\text{RBF} + 238.95\text{SpMin1\_Bh}(i) - 84.18\text{HATS6e} + 27.10\text{B08}[\text{N-S}]$$

Here, RBF is the rotatable bond fraction, determining the number of bonds that allow free rotation.  $\text{SpMin1\_Bh}(i)$  is the smallest eigenvalue of the Burden matrix weighted by ionization potential. HATS6e is the H-GETAWAY descriptor derived from atomic electronegativity distributions.<sup>15,16</sup> B08[N-S] captures the presence/absence of N-S at topological distance of 8.<sup>17</sup> Y-randomization tests and applicability-domain analysis confirm that the QSPR model is statistically robust, avoids spurious correlations, and is reliable for  $T_g$  prediction. Key dataset details and validation metrics are provided in the SI. The following sections further discuss its predictive performance, molecular interpretation, and applicability in the design of CPs.

Fig. 2B shows the correlation between predicted and experimental  $T_g$  values using a 70/30 training/test split. The QSPR model achieves  $R^2 = 0.89$  for training and 0.85 for testing, confirming strong predictive power. With just four descriptors, the model effectively captures structural diversity and predicts  $T_g$  directly from the monomer structure. Compared to our previous ML-based model<sup>9</sup> ( $R^2 = 0.85$  overall, Fig. 2C), the QSPR approach offers slightly higher accuracy, particularly in distinguishing subtle variations in backbone rigidity and electronic effects. Xie's empirical model,<sup>11</sup> based on a single mobility parameter (Fig. 2D), performs well for low-to-moderate  $T_g$  CPs but reaches only  $R^2 \approx 0.4$  on our dataset, especially struggling with high- $T_g$  CPs ( $T_g > 150$  °C). This suggests that inter-chain interactions, backbone rigidity, and steric effects play larger roles in these materials, which single-parameter models may not fully capture. Thus, our QSPR model provides a more comprehensive and accurate framework for predicting  $T_g$  across diverse CPs, making it a reliable tool for materials design. It should be noted that this study is computational in nature and relies on the experimental  $T_g$  values of CPs (including homopolymers and donor-acceptor polymers with similar structural features of alternating copolymers) reported in the literature. While the model demonstrates strong predictive performance relative to established baselines, future validation using newly synthesized CPs would further strengthen its general applicability.

We next discuss each of the identified descriptors and their role in QSPR model  $T_g$  prediction for CPs.

**RBF.** The rotatable bond fraction (RBF) quantifies backbone flexibility by measuring the fraction of freely rotating bonds. A higher RBF value indicates greater conformational freedom, leading to increased chain mobility and free volume, which in turn lowers  $T_g$ . Conversely, a lower RBF reflects a more rigid backbone that restricts segmental motion and increases  $T_g$ . This trend aligns with our previous ML-based  $T_g$  model, where the alkyl side-chain fraction showed a negative correlation with

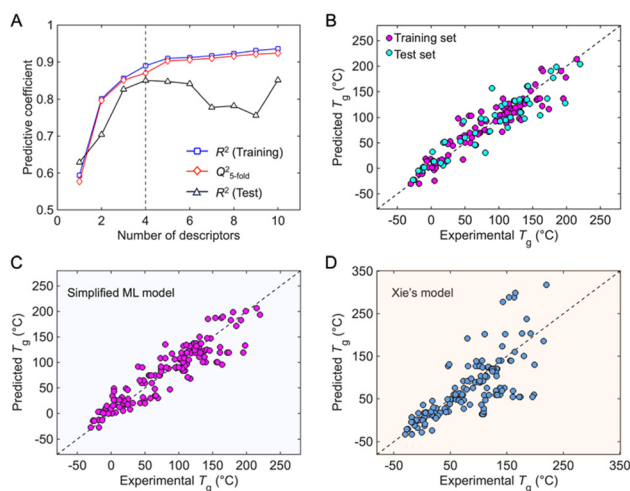


Fig. 2 Fig. 2. (A) Statistical analyses of  $R^2$  and  $Q^2$  as a function of the number of descriptors for QSPR models with 1–10 descriptors for training and test sets. The black dashed line indicates the selected 4-descriptor model. A correlation plot between the observed and predicted values of the  $T_g$  of polymers in (B) the 4-variable QSPR model, (C) Alesadi's ML model,<sup>9</sup> and (D) Xie's predictive model.<sup>11</sup>



$T_g$  due to its disruption of chain packing. However, our QSPR model more comprehensively represents backbone flexibility by accounting for both side-chain effects and steric hindrance from the backbone. The inverse relationship between RBF and  $T_g$  is particularly evident in systems with conjugated aromatic units, fused rings, and sterically hindered groups—all of which reduce the RBF and elevate  $T_g$  by restricting rotational freedom. The extent of this effect depends on whether aromatic rings are isolated, fused, or bridged—while isolated rings allow some dihedral rotation, fused and bridged rings introduce significant steric constraints, further reducing the RBF and increasing  $T_g$ . Prior studies<sup>18–20</sup> have shown that CPs with extended fused-ring backbones exhibit higher  $T_g$  (lower RBF) due to limited segmental motion, reinforcing the RBF's reliability in capturing structural rigidity. Similarly, alkyl side-chains counteract backbone rigidity by introducing rotatable bonds, thereby increasing the RBF and lowering  $T_g$ . Longer side-chains expand free volume and reduce inter-chain interactions, further lowering  $T_g$ .<sup>10</sup>

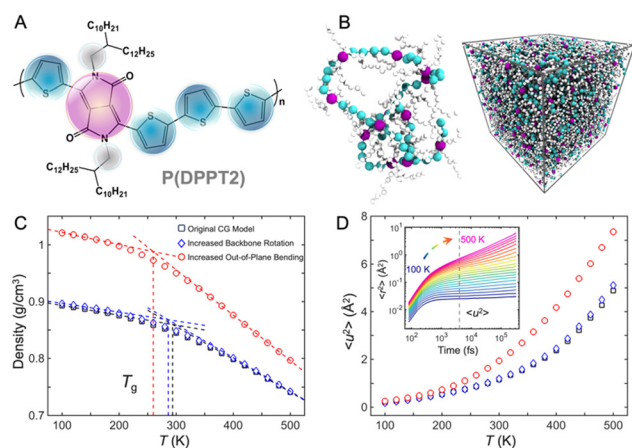
To establish a molecular-level understanding of the relationship between backbone flexibility and  $T_g$ , we conducted coarse-grained molecular dynamics (CG-MD) simulations<sup>21,22</sup> using a chemistry-specific CG model informed by the all-atomistic (AA) model of PDPPT (Fig. 3A and B, detailed in the SI). Notably, PDPPT is included in the QSPR dataset (Fig. 1), enabling a direct conceptual link between descriptor-based predictions and simulation results. The simulation systematically varied backbone rotation ( $K_{\text{rotation}}$ ) and bending rigidity ( $K_{\text{rigidity}}$ ) to quantify their respective effects on thermal behavior. Fig. 3C presents density vs. temperature curves for different backbone flexibility conditions, where  $T_g$  is determined from the

intersection of linear fits. According to free volume theory, increasing backbone flexibility by lowering  $K_{\text{rigidity}}$  suppresses chain packing efficiency, increases available free volume, and thereby reduces  $T_g$ .<sup>23</sup> Reduced torsional stiffness slightly increases density and modestly lowers  $T_g$ , while reduced bending rigidity causes a larger density increase and a pronounced  $T_g$  drop. These results align with previous CG-MD simulations by Xu and co-workers,<sup>24,25</sup> who used the generalized entropy theory to show that backbone stiffness modulates  $T_g$  by altering configurational entropy, packing frustration, and segmental relaxation, supporting the entropy-driven mechanism observed here.

To further assess the role of backbone flexibility in segmental mobility, we calculate the Debye–Waller factor ( $\langle u^2 \rangle$ ), which quantifies fast segmental motion at short picosecond time-scales (Fig. 3D). A higher  $\langle u^2 \rangle$  indicates increased local mobility, corresponding to greater free volume and enhanced chain dynamics.<sup>9,14</sup> Across all cases, lowering  $K_{\text{rotation}}$  leads to a moderate increase in  $\langle u^2 \rangle$ , whereas decreasing  $K_{\text{rigidity}}$  significantly enhances local mobility, reflecting a reduction in  $T_g$ -associated molecular constraints. The inset of Fig. 3D shows mean-squared displacement (MSD) curves, where the vertical dashed line at  $t = 4$  ps marks the caging time at which  $\langle u^2 \rangle$  is determined. This trend aligns with the plasticizer-like effect, where increased backbone flexibility promotes molecular rearrangement, leading to looser packing and lower  $T_g$ .<sup>26</sup> These findings confirm that backbone flexibility directly influences  $T_g$  through its effects on molecular packing and segmental mobility, supporting the inverse correlation between the RBF and  $T_g$  observed in our QSPR model.

*SpMin1\_Bh(i)*. This descriptor represents the smallest eigenvalue of the Burden matrix weighted by ionization potential and shows a positive correlation with  $T_g$ . This trend may suggest that higher ionization potential—often associated with more electron-deficient and less polarizable monomers—correlates with reduced chain mobility. One possible explanation is that electron-deficient backbones exhibit greater electronic rigidity or limit  $\pi$ -electron delocalization, which in turn constrains segmental motion and elevates  $T_g$ . Alternatively, this correlation may arise from changes in chain packing or interchain interactions driven by electronic effects. These factors will be investigated separately in future studies to better understand the role of electronic properties in governing polymer thermal behavior.

*HATS6e*. HATS6e is a 3D H-GETAWAY descriptor that encodes a leverage-weighted autocorrelation of atomic Sanderson electronegativities at a six-bond topological lag, capturing both spatial and electronic characteristics of the molecular structure.<sup>27,28</sup> In our model, HATS6e exhibits a negative correlation with  $T_g$ , suggesting that higher charge localization—reflected by greater variation in electronegativity along the polymer backbone—may reduce intermolecular cohesion and increase chain mobility. This effect may result from disrupted  $\pi$ - $\pi$  stacking in conjugated systems or weakened dipolar interactions in polar polymers. This trend is supported by experimental observations where highly fluorinated CPs (extremely electronegative substituents) tend to pack poorly (lower crystallinity) and exhibit greater chain mobility (potentially lower  $T_g$ ).<sup>29</sup> Our results



**Fig. 3** (A) Coarse-grained (CG) modeling of PDPPT-based CPs: mapping scheme from the all-atomistic model to the CG representation, where each repeat unit is represented by three bead types. (B) Representative polymer chain configuration (left), and a snapshot of the equilibrated amorphous bulk CG systems (right). (C) Density vs. temperature ( $T$ ), showing the impact of backbone flexibility on thermal behavior. The vertical dashed lines indicate  $T_g$ , determined from linear fits. (D) Debye–Waller factor ( $\langle u^2 \rangle$ ) vs.  $T$ , capturing fast segmental motion. (Inset) Mean-squared displacement ( $\langle r^2 \rangle$ ) at various temperatures (100 K to 500 K) with the vertical dashed line at  $t = 4$  ps marking the caging time for  $\langle u^2 \rangle$  determination.



also confirm previous MD and ML-based<sup>9</sup> findings that the presence of halogen atoms (e.g., fluorine and chlorine) correlates negatively with  $T_g$ , likely due to increased free volume and reduced cohesion.

**B08[N-S].** The B08[N-S] descriptor, a topological measure of nitrogen and sulfur connectivity, shows a positive correlation with  $T_g$ , suggesting that N-S interactions may contribute to polymer rigidity and intermolecular cohesion. While direct evidence is limited, prior studies indicate that polymers with nitrogen and sulfur heteroatoms often exhibit higher  $T_g$ , possibly due to dipole interactions and secondary bonding effects. Our model captures a similar trend, aligning with reports that benzothiadiazole- and thiophene-containing CPs tend to have elevated  $T_g$ .<sup>9</sup> Although heteroatoms have been identified in QSPR studies as factors influencing  $T_g$ , further validation is needed to fully clarify their role in restricting segmental mobility and enhancing intermolecular interactions.

In conclusion, we developed a QSPR model using four interpretable molecular descriptors—RBF, SpMin1\_Bh(i), HATS6e, and B08[N-S]—to accurately predict the glass-transition temperature ( $T_g$ ) of conjugated polymers. These descriptors capture backbone flexibility, electronic properties, charge distribution, and heteroatom interactions. The model achieves strong predictive accuracy ( $R^2 \approx 0.85$ ), comparable to our previous ML-based model.<sup>9</sup> CG-MD simulations support the descriptor- $T_g$  relationships, while the model's generalizability to new polymer classes remains to be validated in future studies. For next-generation CPs with increasingly complex chemistries, this QSPR framework offers improved robustness and interpretability, making it well-suited to guide the rational design of advanced polymer materials.

This work was supported by the National Science Foundation (NSF) under Award No. 2237063.

## Conflicts of interest

There are no conflicts to declare.

## Data availability

Supporting data are provided in the SI. See DOI: <https://doi.org/10.1039/d5cc03072b>

## References

- S. E. Root, S. Savagatrup, A. D. Printz, D. Rodriguez and D. J. Lipomi, *Chem. Rev.*, 2017, **117**, 6467–6499.
- A. Chortos, J. Liu and Z. Bao, *Nat. Mater.*, 2016, **15**, 937–950.
- J. Y. Oh and Z. Bao, *Adv. Sci.*, 2019, **6**, 1900186.
- D. J. Lipomi, B. C. K. Tee, M. Vosgueritchian and Z. Bao, *Adv. Mater.*, 2011, **23**, 1771–1775.
- S. Zhang, A. Alesadi, G. T. G. T. Mason, K.-L. K.-L. Chen, G. Freychet, L. Galuska, Y.-H. Cheng, P. B. J. B. J. Onge, M. U. M. U. Ocheje, G. Ma, Z. Qian, S. Dhakal, Z. Ahmad, C. Wang, Y.-C. Chiu, S. Rondeau-Gagné, W. Xia and X. Gu, *Adv. Funct. Mater.*, 2021, **31**, 2100161.
- M. Xiao, A. Sadhanala, M. Abdi-Jalebi, T. H. Thomas, X. Ren, T. Zhang, H. Chen, R. L. Carey, Q. Wang, S. P. Senanayak, C. Jellett, A. Onwubiko, M. Moser, H. Liao, W. Yue, I. McCulloch, M. Nikolka and H. Sirringhaus, *Adv. Funct. Mater.*, 2021, **31**, 1–11.
- A. Alesadi, W. Xia and D. Kilin, *J. Phys. Chem. C*, 2022, **126**, 12015–12024.
- A. Alesadi, F. Fatima, W. Xia and D. Kilin, *J. Phys. Chem. B*, 2021, **125**, 8953–8964.
- A. Alesadi, Z. Cao, Z. Li, S. Zhang, H. Zhao, X. Gu and W. Xia, *Cell Rep. Phys. Sci.*, 2022, 100911.
- S. Zhang, A. Alesadi, M. Selivanova, Z. Cao, Z. Qian, S. Luo, L. Galuska, C. Teh, M. U. M. U. M. U. Ocheje, G. T. G. T. Mason, P. B. J. B. J. Onge, D. Zhou, S. Rondeau-Gagné, W. Xia, X. Gu, S. Rondeau-Gagné, W. Xia, X. Gu, S. Rondeau-Gagné, W. Xia and X. Gu, *Adv. Funct. Mater.*, 2020, **2002221**, 2002221.
- R. Xie, A. R. Weisen, Y. Lee, M. A. Aplan, A. M. Fenton, A. E. Masucci, F. Kempe, M. Sommer, C. W. Pester, R. H. Colby and E. D. Gomez, *Nat. Commun.*, 2020, **11**, 4–11.
- A. Karuth, A. Alesadi, W. Xia and B. Rasulev, *Polymer*, 2021, **218**, 123495.
- A. R. Katritzky, P. Rachwal, K. W. Law, M. Karelson and V. S. Lobanov, *J. Chem. Inf. Comput. Sci.*, 1996, **36**, 879–884.
- A. Alesadi and W. Xia, *Macromolecules*, 2020, **53**, 2754–2763.
- S. Filipic, D. Ruzic, J. Vucicevic, K. Nikolic and D. Agbaba, *J. Pharm. Biomed. Anal.*, 2016, **127**, 101–111.
- V. Consonni, R. Todeschini and M. Pavan, *J. Chem. Inf. Comput. Sci.*, 2002, **42**, 682–692.
- R. Todeschini and Viviana Consonni, *Molecular Descriptors for Chemoinformatics*, Wiley-VCH, 2nd edn, 2009.
- D. J. Lipomi, H. Chong, M. Vosgueritchian, J. Mei and Z. Bao, *Sol. Energy Mater. Sol. Cells*, 2012, **107**, 355–365.
- B. B. Roth, S. Savagatrup, N. V. De Los Santos, O. Hagemann, J. E. Carlé, M. Helgesen, F. Livi, E. Bundgaard, R. R. Søndergaard, F. C. Krebs, N. V. De Los Santos, O. Hagemann, J. E. Carlé, M. Helgesen, F. Livi, E. Bundgaard, R. R. Søndergaard, F. C. Krebs and D. J. Lipomi, *Chem. Mater.*, 2016, **28**, 2363–2373.
- Y. Li, W. K. Tatum, J. W. Onorato, Y. Zhang and C. K. Luscombe, *Macromolecules*, 2018, **51**, 6352–6358.
- Z. Li, Y. Wang, G. Ma, Y. Liao, X. Gu and W. Xia, *J. Polym. Sci.*, 2024, **62**, 1296–1309.
- Z. Li, S. A. Tolba, Y. Wang, A. Alesadi and W. Xia, *Chem. Commun.*, 2024, **60**, 11625–11641.
- W.-S. Xu, J. F. Douglas and Z.-Y. Sun, *Macromolecules*, 2021, **54**, 3001–3033.
- W.-S. Xu and K. F. Freed, *Macromolecules*, 2014, **47**, 6990–6997.
- W.-S. Xu, J. F. Douglas and X. Xu, *Macromolecules*, 2020, **53**, 4796–4809.
- R. A. Riggelman, J. F. Douglas and J. J. de Pablo, *Soft Matter*, 2010, **6**, 292–304.
- A. G. Mercader, P. R. Duchowicz, F. M. Fernández, E. A. Castro, D. O. Bennardi, J. C. Autino and G. P. Romanelli, *Bioorg. Med. Chem.*, 2008, **16**, 7470–7476.
- S. Filipic, D. Ruzic, J. Vucicevic, K. Nikolic and D. Agbaba, *J. Pharm. Biomed. Anal.*, 2016, **127**, 101–111.
- A. Zhang, C. Xiao, Y. Wu, C. Li, Y. Ji, L. Li, W. Hu, Z. Wang, W. Ma and W. Li, *Macromolecules*, 2016, **49**, 6431–6438.

