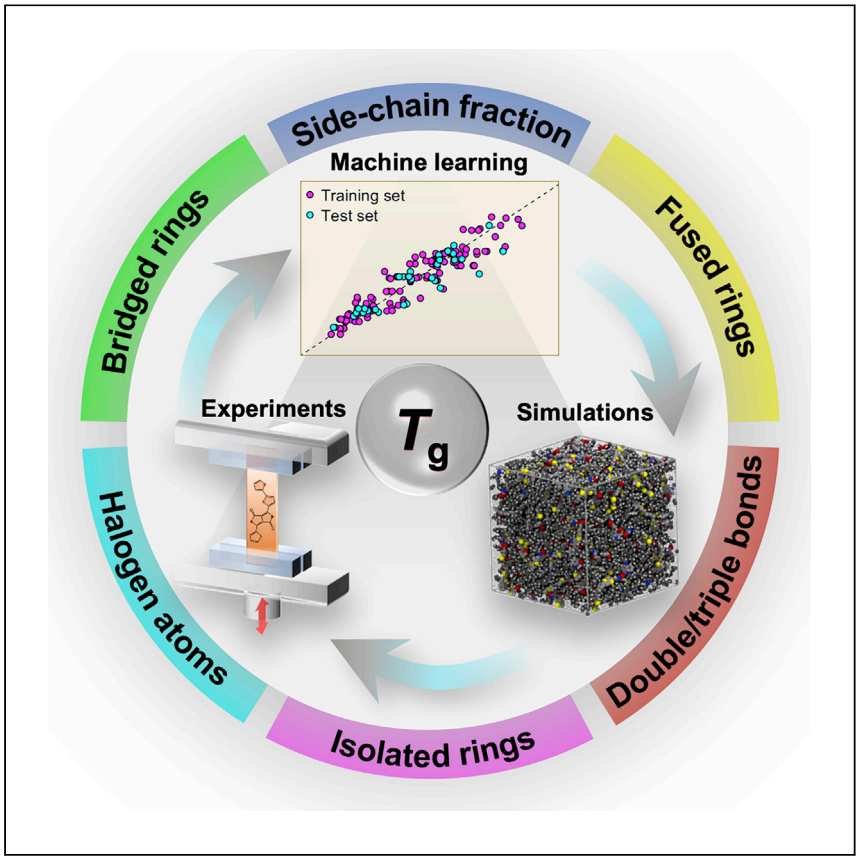
Cell Reports
Physical Science

CellPress
OPEN ACCESS

Article

# Machine learning prediction of glass transition temperature of conjugated polymers from chemical structure




Alesadi et al. present an integrated framework to predict the glass transition temperature of conjugated polymers having diverse chemistry through the integration of machine learning, molecular dynamics simulations, and experiments. The predictive model takes simplified "geometry" of six key chemical building blocks as molecular features to predict glass transition temperature.

Amirhadi Alesadi, Zhiqiang Cao, Zhaofan Li, Song Zhang, Haoyu Zhao, Xiaodan Gu, Wenjie Xia

xiaodan.gu@usm.edu (X.G.)
wenjie.xia@ndsu.edu (W.X.)

## Highlights

Machine learning approach is employed to predict glass transition temperature ($T_g$)

Aromatic rings and alkyl side chains are dominant building blocks to determine $T_g$

Molecular dynamics simulations unravel role of diverse building blocks in dynamics

Experimental measurements of $T_g$ confirm the predictive performance of the model

Alesadi et al., Cell Reports Physical Science 3, 100911
June 15, 2022 
https://doi.org/10.1016/j.xcrp.2022.100911

Cell Reports
Physical Science

CellPress
OPEN ACCESS

Article

# Machine learning prediction of glass transition temperature of conjugated polymers from chemical structure

Amirhadi Alesadi,[1,3] Zhiqiang Cao,[2,3] Zhaofan Li,[1,3] Song Zhang,[2] Haoyu Zhao,[2] Xiaodan Gu,[2,*] and Wenjie Xia[1,4,*]

## SUMMARY

**Predicting the glass transition temperature ($T_g$) is of critical importance as it governs the thermomechanical performance of conjugated polymers (CPs). Here, we report a predictive modeling framework to predict $T_g$ of CPs through the integration of machine learning (ML), molecular dynamics (MD) simulations, and experiments. With 154 $T_g$ data collected, an ML model is developed by taking simplified "geometry" of six chemical building blocks as molecular features, where side-chain fraction, isolated rings, fused rings, and bridged rings features are identified as the dominant ones for $T_g$. MD simulations further unravel the fundamental roles of those chemical building blocks in dynamical heterogeneity and local mobility of CPs at a molecular level. The developed ML model is demonstrated for its capability of predicting $T_g$ of several new high-performance solar cell materials to a good approximation. The established predictive framework facilitates the design and prediction of $T_g$ of complex CPs, paving the way for addressing device stability issues that have hampered the field from developing stable organic electronics.**

## INTRODUCTION

Semiconducting conjugated polymers (CPs) are attractive organic electronic materials for a wide range of applications due to their unique properties, such as easy processability,[1] tunable electrical performance, and mechanical flexibility with diverse chemistry.[2–9] Through the engineering of molecular structure and chain architecture, these organic semiconductors have shown great promise in advanced stretchable and flexible electronic devices and wearable electronics.[10–14] In particular, thermomechanical behaviors of CPs are directly related to the glass transition temperature ($T_g$) that governs the dynamics and thermal performance of polymer chains. Despite tremendous efforts, design and prediction of $T_g$ remain notably challenging for CPs due to their complex chain architecture associated with diverse "donor-acceptor" (D-A) chemical building blocks. The highly rigid backbone and semi-crystalline nature of CPs complicate tracking the heat flow and accurate assessment of glass transition.[15] More importantly, due to the endless possibilities of CPs' building blocks, experimental measurement of $T_g$ for new designs of CPs can be time -consuming, if not impossible. So far, only a limited number of CPs have been tested and reported for their $T_g$ through experiments.[16] Hence, a computational predictive framework to predict $T_g$ directly from the geometry of chemical building blocks, before any synthesis process, is profoundly advantageous for CPs in applications of the next-generation electronic devices.

[1]Department of Civil, Construction and Environmental Engineering, North Dakota State University, Fargo, ND 58108, USA

[2]School of Polymer Science and Engineering, Center for Optoelectronic Materials and Devices, The University of Southern Mississippi, Hattiesburg, MS 39406, USA

[3]These authors contributed equally

[4]Lead contact

*Correspondence: xiaodan.gu@usm.edu (X.G.), wenjie.xia@ndsu.edu (W.X.)

https://doi.org/10.1016/j.xcrp.2022.100911

Many physical parameters are found to play a vital role in influencing complex molecules, such as chain rigidity, intermolecular interactions, segmental structure, and chain topology.[16–24] For most glass-forming polymers, $T_g$ would raise by increasing the chain stiffness of the backbone due to the higher hindrance for translational and rotational motion;[17,25–27] consistently, for different CPs, $T_g$ is found to be correlated to persistence length and rotational volume.[22,28,29] Alkyl side groups largely influence the chain mobility, where increasing side-chain length causes $T_g$ to drop due to a larger volume fraction of the side chain relative to the polymer backbones. This causes increase in the free volume, which could be quantified in terms of the Debye-Waller factor at a picosecond timescale.[30] Our previous study has shown that increase of alkyl side-chain length of CPs leads to higher Debye-Waller factor and mobility compared with the backbone, unraveling the origin of their decreasing influence on the $T_g$.[31] In addition, modifications in the topology of the polymer chain, e.g., changing the grafting density, have been found to considerably affect the $T_g$.[16,17,31] The coupled influence of these many structural and physical features provokes the idea of quantifying $T_g$ directly from the chemical building block and fingerprints by delineating the underlying "structure-property" relationship.

To address this issue, the cheminformatics-based quantitative structure-property relationship (QSPR) modeling approach has been utilized as a predictive computational framework that links the physicochemical properties of compounds to their chemical structure.[32] The QSPR aims to find a mathematical relationship between the target variable and molecular descriptors that are extracted from the chemical structure of the molecules. QSPR methods have been successfully employed to predict $T_g$ of non-conjugated polymers with an appropriate level of confidence. Katritzky and coworkers[33] developed a four-parameter QSPR model with a $R^2$ of 0.928 for predicting the $T_g$ values of 22 different linear-chain homopolymers and copolymers. On a larger dataset,[34] they developed a five-descriptor QSPR model where the predictive model showed an $R^2$ of 0.946. However, despite the predictive capability of QSPR methodology, interpreting and quantifying a large number (i.e., over a few thousand) of molecular descriptors, e.g., molar refractivity, dipole moment, polarizability, and constitutional descriptors, could be very challenging, limiting its practical usage for the community.

In recent years, machine learning (ML) algorithms have been increasingly employed to develop simplified surrogate models with excellent predictive capability.[35–37] By extracting molecular features directly from the geometry of the monomer, ML-based surrogate modeling could be advantageous over QSPR modeling by avoiding the aforementioned complexities of feature descriptors. If successful, such simplified models are much more desired for the experimental community to predict $T_g$ for newly designed CPs before any synthesis process. More recently, Xie and coworkers[38] proposed a single-parameter empirical model to predict $T_g$ based on the mobility parameter of different functional groups of the repeat unit. Their model showed an excellent $T_g$ prediction for 32 experimentally tested semi-flexible (mostly conjugated) polymers with alkyl side chains, where a root-mean-square error (RMSE) of 13°C was observed. However, without any out-of-sample testing for validation, it is still questionable that such a simplified model is capable of $T_g$ prediction over diverse CPs with a much-enlarged dataset.

Built upon recent efforts, in this study, we employ an integrated experimental-computational approach to predict $T_g$ for a diverse set of over 94 polymers (mostly CPs) with a drastic difference in their chemical structures (i.e., backbones and side chains). As illustrated in Figure 1, built upon the backward elimination technique

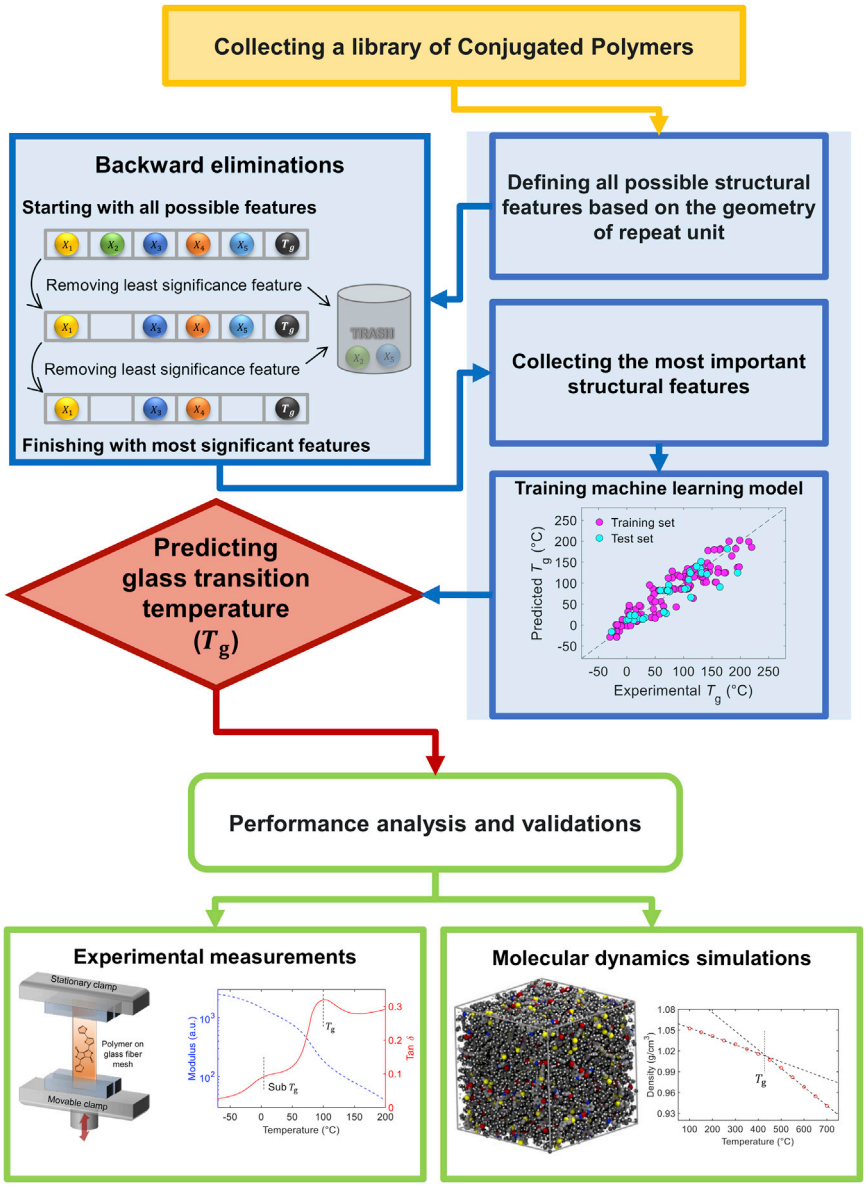



**Figure 1. Machine learning model**

Flow chart of the current integrated machine learning framework, molecular dynamics simulations, and experimental techniques to predict glass transition temperature of conjugated polymers.

and regression analysis, a surrogate ML model is proposed that takes the geometry of the chemical building blocks within a repeat unit as simplified molecular features to make the $T_g$ prediction. In addition to the existing $T_g$ data collected from the literature,[16] we perform dynamic mechanical analysis (DMA), differential scanning calorimetry (DSC) experiments, and molecular dynamics (MD) simulations to improve the diversity of the dataset, which is further used to train the predictive model and perform external validation. The simplified ML results reveal that $T_g$ is strongly governed by the chemical structure of the key building block of the CPs. Furthermore, MD simulations allow us to gain valuable insights into the dynamic heterogeneity and chain dynamics at a fundamental molecular level, which is experimentally verified by performing quasi-elastic neutron scattering on CPs. Our study establishes an

integrated predictive framework for $T_g$s of diverse CPs, paving the way for materials by design for high-performance CPs and relevant materials via molecular engineering.

## RESULTS AND DISCUSSION

### Predicting $T_g$ from the chemical building blocks

In this section, we will introduce an ML-based surrogate model by employing the ordinary least squares (OLS) multiple linear regression algorithm. Among 154 collected $T_g$ data points, 119 values are from previously published literature,[16,31,38] 17 are from our experimental measurements (i.e., DMA and DSC), and 18 are obtained via MD simulations. It should be noted that, for some CPs, multiple $T_g$ values are reported from different groups, and we consider all those data points in our ML model. Overall, we collect $T_g$ data points of over 94 diverse polymers and acceptor units, reported in Tables S1 and S2. Then, we define around 30 structural and molecular features (see Note S1) directly from the geometry of the chemical building blocks within a repeat unit. Informed by the QSPR modeling,[17] these structural features can be categorized into four major groups, namely the side chains, aromatic rings, different kinds of atoms, and double and triple bonds that are not part of aromatic rings. In particular, the aromatic ring features based on their chemical environment include isolated, fused, and bridged rings. After defining all features, a backward elimination[39] technique is utilized to perform the significance level test (i.e., significance is used to determine whether the relationship between features and $T_g$ exists or not) and find the most important structural features that are closely correlated to $T_g$. In the end, six structural features are identified to develop the ML model for the prediction of $T_g$ of diverse CPs. Accordingly, $T_g$ is found to be largely influenced by the side-chain fraction, number of isolated, fused, and bridged aromatic rings; number of halogenated atoms, such as "fluorine" or "chlorine;" and number of double and triple bonds that are not part of aromatic rings. It should be noted that all aforementioned features are normalized by the total number of atoms in the repeat unit (i.e., excluding hydrogens). The multiple linear regression model based on structural features can be expressed as follows:

$$T_g = \beta_0 + \sum \beta_i X_i, \quad X_i = [NSC,\ NFCL, NDTB, NIR, NFR, NBR]/N, \qquad \text{(Equation 1)}$$

where $\beta_0$ is a constant value and $\beta_i$ is the contribution coefficient of each feature in the regression model. $N$ is the total number of atoms in the repeat unit (excluding hydrogens). *NSC* stands for the number of carbon atoms in the alkyl side chains, *NFCL* is the total number of fluorine and chlorine atoms in the monomer, *NDTB* is the number of double and triple bonds that are not part of aromatic rings, *NIR* is the number of free/isolated rings either in the backbone or side groups, *NFR* stands for the number of fused rings in which adjacent carbon atoms are shared, and *NBR* is the number of bridged rings where adjacent rings are bridged via valence bonds. A list of these selected structural features, their descriptions, and example of relevant moieties are presented in Table S3.

Figure 2 shows a graphical illustration of different kinds of aromatic rings in representative CPs. To simplify the feature calculations consistently, four assumptions (or rules) have been made to count the number of aromatic rings prior to training the ML model:

(1) There is no difference between different kinds of rings, for example, thiophene, phenyl, and pyrrole compounds are treated in the same way.

(2) If a ring is fused on one side and bridged on the other side, that compound would be counted twice: one for fused rings (*NFR*) and one for bridged rings (*NBR*) since those kinds of hindrances significantly increase the chain stiffness and subsequently $T_g$ (see Y2 acceptor unit in Figure 2).
(3) The pentagonal rings with grafted side chains on the tip that bridges two adjacent compounds on both sides are not counted as any features, such as fused or bridged rings. The examples of these rings, excluded from the features, are highlighted with a purple "X" label in Figure 2.
(4) The number of successive fused rings is weighted differently before counting the overall *NFR* for each repeat unit. According to our collected data, when more than two rings are fused, $T_g$ of CPs exhibits an excessive tendency to increase. To address this phenomenon, the weighted number of successive fused rings is calculated as $2 + w_a(n - 2)$, where $n$ is the number of successive fused rings and $w_a$ is a weighting factor, which needs to be determined. For example, for the PTB7-Th polymer in Figure 2, there are two fused rings groups: for those two fused thiophenes, the weighted number would be $2 + w_a(2 - 2) = 2$, implying that no weighting is needed when just two rings are fused; however, that phenyl ring fused to two thiophenes at both sides form three successive fused rings, which need to be weighted as $2 + w_a(3 - 2) = 2 + w_a$. Then, the weighted values of all fused groups will be summed to determine the value of *NFR* of that polymer. Another example for naphthalenediimide (NDI)-based polymers, four successive fused rings need to be weighted as $2 + w_a(4 - 2) = 2 + 2w_a$. Then, the optimum value of $w_a$ is determined by evaluating the coefficient of determination ($R^2$) and RMSE of the ML models as a function of $w_a$, where the highest $R^2$ or lowest RMSE values lead to $w_a = 3.5$ (see Figure S1), which will be used for training the ML model.

Then, the contribution of each feature ($\beta_i$) is determined by fitting the regression model to the training set, and the ML model for $T_g$ prediction can be described as follows:

$$T_g(^\circ\mathrm{C}) = 115.3 + \frac{-239.8\,NSC - 598.3\,NFCL + 170\,NDTB + 456.6\,NIR + 666.5\,NFR + 1323.9\,NBR}{N}. \quad \text{(Equation 2)}$$

A comparison between predicted $T_g$ and experimental and computational values is presented in Figure 3A for both training and test sets. Despite having a very diverse library of CPs and acceptor units with different chemical structures, assessment of the model shows excellent predictive performance as measured by $R^2$ of 0.86 and RMSE of 22.4°C for the training set. Furthermore, ML model prediction for the external validation of the test set yields $R^2$ of 0.83 and RMSE of 27.8°C, showing a satisfactory performance based on an out-of-sample dataset. The evaluation of the results demonstrates that the trained ML model is capable of predicting $T_g$ from the "simplified" features based on the geometry of a few chemical building blocks. In particular, for those high-$T_g$ CPs with complex D-A units, such as NDI-based polymers, poly[(2,6-(4,8-bis(5-(2-ethylhexyl-3-fluoro)thiophen-2-yl)-benzo[1,2-b:4,5-b′]dithiophene))-alt-(5,5-(1′,3′-di-2-thienyl-5′,7′-bis(2ethylhexyl)benzo[1′,2′-c:4′,5′-c′]dithiophene-4,8-dione)] (PM6), poly[(2,6-(4,8-bis(5-(2-ethylhexyl-3-chlore)thiophen-2-yl)-benzo[1,2-b:4,5-b′]dithiophene))-alt-(5,5-(1′,3′-di-2-thienyl-5′,7′-bis(2ethylhexyl)benzo[1′,2′-c:4′,5′-c′]dithiophene-4,8-dione)] (PM7), and several other polymers and acceptor units with aromatic rings, our ML model shows an excellent agreement between predicted and measured $T_g$.

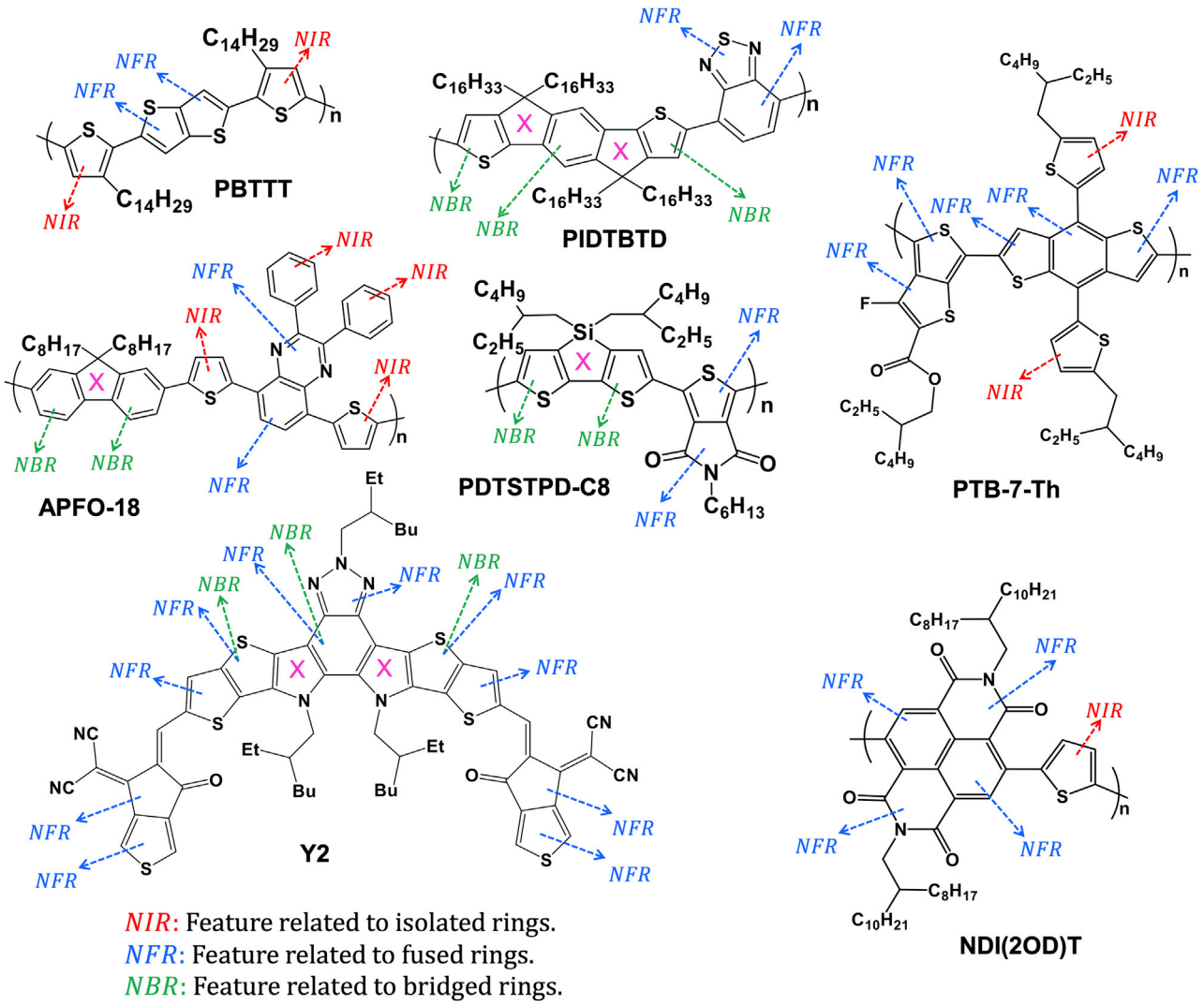



**Figure 2. Graphical illustration of defining features related to the aromatic rings in the repeat unit of conjugated polymers**

Features related to isolated rings (NIR), fused rings (NFR), and bridged rings (NBR) are shown in red, blue, and green fonts, respectively. Those rings with a purple "X" label are not counted as any features, such as fused or bridged rings.

We next compare our model with a semi-empirical model recently developed by Colby and Gomez group (i.e., we call it "Xie's model" here)[38] for the $T_g$ predictions from the chemical structure. In Xie's model, they proposed a single adjustable mobility parameter $\xi$, defined as the summation of assigned atomic mobility to each functional group in the repeat unit to establish a quantitative relationship between $T_g$ and chemical structures of about 30 CPs. To have a better comparison between these two models, we calculate Xie's mobility parameter $\xi$ for the whole dataset utilized in this study and predicted $T_g$ values are reported in Figure 3B. As it is indicated, despite an acceptable performance of Xie's model, especially for lower $T_g$ CPs, such as thiophene- and polydiketopyrrolopyrrole (PDPP)-based systems, it does not yield reliable prediction for the aforementioned high-$T_g$ CPs as measured by a relatively low $R^2$ of 0.57 and a large RMSE of 46.6°C. There are three main differences between Xie's model and our ML approach: first, Xie and coworkers did not consider the influence of successive fused rings when more than two rings are fused, resulting in an underestimated prediction of $T_g$ for CPs

with a larger number of fused rings, such as NDI-based polymers; second, their model does not demonstrate the reducing influence of halogen atoms on $T_g$ of CPs; and third, Xie's model is tested for a limited set of 32 semi-flexible polymers and mobility value $\xi$ does not generalize to our diverse dataset, and as they discussed, modifications of $\xi$ values for specific chemical structures may be needed to improve the predictive performance.

## Relative importance of essential structural features

Based on the current developed ML model, we next discuss the relative importance and contribution of each structural feature to $T_g$ prediction of diverse CPs and other complex polymers to better understand the physical roles of different chemical building blocks. For this purpose, the t-statistic parameter, defined as the feature coefficient normalized by its standard error, is calculated as a measurement of the feature contribution. A larger t-statistic parameter implies the statistically significant contribution of that feature in which positive and negative sign of t-statistic is related to enhancing and reducing influences on $T_g$, respectively.

### *Side chains*

As indicated in Figure 3C, the coefficient related to the side-chain fraction (*NSC*) shows the largest t-statistic that indicates its significant contribution, where the negative value implies the reducing influence on $T_g$. To determine whether the current trend that we observe in our samples remains valid on the larger population, the p value of each feature is calculated. The p value assesses the null hypothesis if the feature does not correlate with the target variable ($T_g$), in which a p value less than the significance level (i.e., 0.05) provides enough evidence to reject the null hypothesis over the entire population. For the side-chain fraction (*NSC*), a $p < 0.001$ is calculated much below the significance level, proving this feature is statistically considerable. Previous efforts have shown that increasing the side-chain length decreases the $T_g$ due to an increase in free volume associated with higher segmental mobility and governs localized relaxations.[31,40,41] For instance, experimental results showed a reduction in the $T_g$ and elastic modulus of DPP-based polymer with increasing the length of alkyl side chains.[30] Consistent with experimental observations, this reducing influence of the alkyl side chains on $T_g$ is captured well by our ML model.

### *Halogen atoms*

The next feature that lowers $T_g$ is the number of fluorine and chlorine atoms (*NFCL*) in the repeat unit. These halogen atoms are highly electronegative, being widely utilized recently to control the energy bandgap and improve the power conversion efficiency of organic solar cells.[42] However, their influence on $T_g$ still remains ambiguous. It has been shown that the C–F bond can form intermolecular interaction with other atoms (e.g., phenyl-perfluorophenyl, C … H, F … F, and C–F … $\pi_F$),[43,44] which inherently can influence the thermomechanical properties of the polymer.[20] Kroon and coworkers[45] showed that the incorporation of two fluorine atoms on the acceptor in poly[6,7-difluoro-2,3-bis(3-(octyloxy)phenyl)qui-noxaline-alt-thiophene] (TQ-F) drops $T_g$ from 100°C to 48°C. For TQ-F, the incorporation of fluorine atoms was found to slightly increase the torsion angle, which increases chain flexibility accompanied by backbone irregularity and thereby lowers $T_g$. Our MD simulations, see Table S2, also show a similar reducing trend of $T_g$ for Y5 acceptor unit and its fluorinated and chlorinated versions (i.e., Y6 and Y7 units, respectively), where both Y6 and Y7 yield lower $T_g$ values compared to Y5. However, $T_g$ of chlorinated version, Y7, is slightly larger than fluorinated Y6 unit. Similar trend for fluorinated

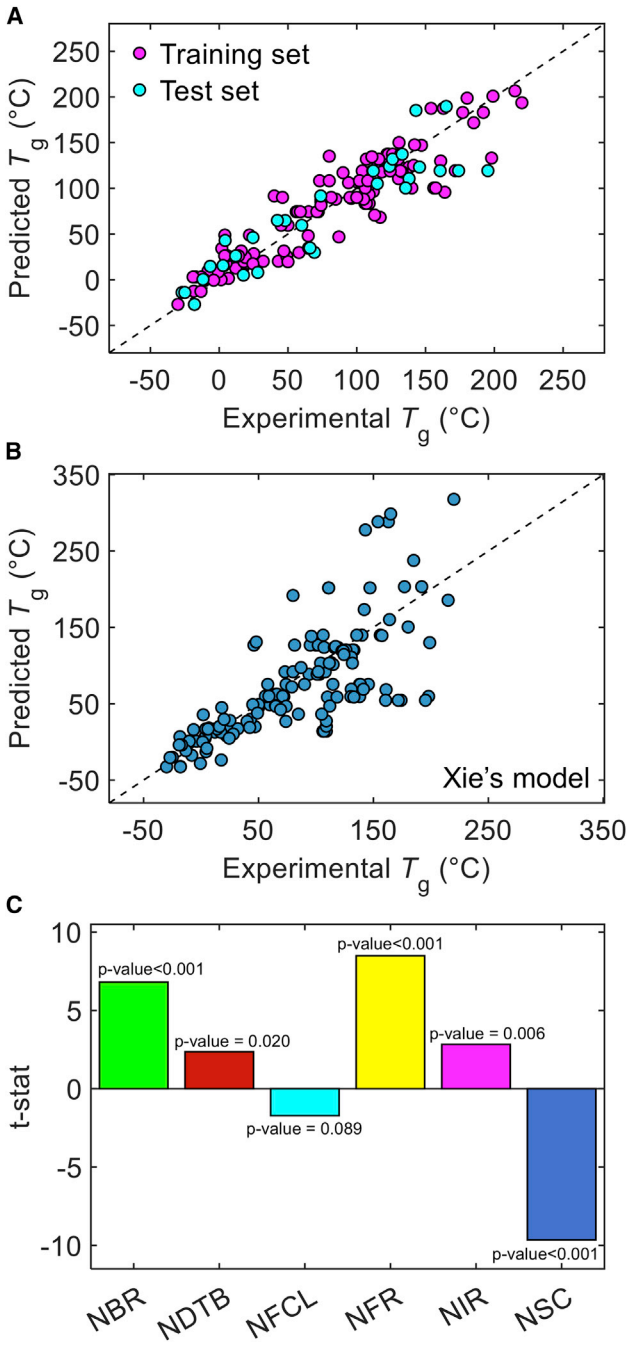



**Figure 3. Predictive performance of the current machine learning model**

(A) Comparison between glass transition temperature ($T_g$) predicted by multiple regression model versus experimental and simulation values. Eighty percent of the dataset is divided into the training set, and 20% is preserved unknown from the model for out-of-sample testing.

(B) Comparison between the $T_g$ predicted by Xie's model versus experimental and simulation values of the whole dataset of conjugated polymers employed in this study.

(C) The contribution of features in the regression model assessed by the value of the t-statistic. Large absolute values of t-statistic mean more noticeable contribution. Side chains and aromatic rings are found to have the dominant role to determine $T_g$. The p -value of each feature is reported inside the figure.

PM6 and its chlorinated version (PM7) is also observed in our MD simulations— PM7 shows a higher $T_g$ value compared with PM6. This could be because of higher steric hindrance of chlorine atoms relative to fluorine as it was experimentally reported for a series of halogenated CPs.[42]

The current ML model is trained by the aforementioned available CPs data, which includes halogen atoms, and Figure 3C shows that their decreasing influence on $T_g$ is predicted. However, due to limited data points, this feature slightly falls beyond the significance level test with a p value of 0.089, larger than the threshold value. We decide to keep the current feature in the predictive model, since it captures the $T_g$ trend for halogenated organic semiconducting materials. With the development of new CPs with halogen atoms, the model can be further improved by having more experimental data in the future to train our model.

### *Double and triple bonds*

The other feature that is found to noticeably contribute to the determination of $T_g$ is the number of double and triple bonds (*NDTB*) in the repeat unit, which considers the influence of weakly mobile atoms. Vinylene, carbonyl, and nitrile groups are counted as *NDTB* in our ML model (see Y2 units in Figure 2 as an example). As the t-statistic parameter shows in Figure 3C, the *NDTB* feature has a noticeably enhancing influence on $T_g$, where a small p value of 0.028 indicates this feature passes the significance level test.

### *Aromatic rings*

The structural features of aromatic rings (i.e., *NIR*, *NFR*, and *NBR*) in our predictive model are found to have a significant enhancing influence on $T_g$, as they govern conformational rigidity of the CPs' backbone. As shown in Figure 3C, the *NIR* feature for the isolated rings exhibits a considerable contribution to the model prediction and successfully passes the significance level test, where a small p value of 0.006 is calculated. Hence, incorporating any isolated aromatic rings into the polymer backbone is expected to increase $T_g$. This is consistent with experimental studies, where the addition of the thiophene rings to the DPP-based CPs ended in an increased $T_g$.[31,46] The fused and bridged rings as labeled by *NFR* and *NBR* features, respectively, have the most significant positive contributions to $T_g$, where *NFR* feature yields a slightly larger contribution parameter (t-statistic). Both features have a very small p value (<0.001), indicating a passed significance-level test. Previous experimental studies also confirm this finding that the presence of fused and bridged rings in the backbone largely correlates to stiffer CPs with larger $T_g$ and modulus.[46–49]

## Dynamical heterogeneity of diverse chemical structures

As the glass transition is primarily governed by the segmental relaxation dynamics and chain mobility, we next perform MD simulations, in conjunction with experiments, to better understand the fundamental roles of diverse chemical building blocks in the dynamics of CPs. Informed from the ML modeling and feature analyses, we focus on segmental dynamics of aromatic rings with different chemical configurations and side chains (i.e., the most influential structural features) of four representative CPs (i.e., DPPT, poly[N-9′-heptadecanyl-2,7-carbazole-alt-5,5-(4′,7′-di-2-thienyl-2′,1′,3′-benzothiadiazole)] (PCDTBT), isoindigo (IID-2T), and NDI-T) (see Figures 4 and S2). Specifically, we evaluate the local Debye-Waller factor $\langle u^2 \rangle$, a fast dynamic physical property at a picosecond timescale,[30] by calculating mean-squared displacement (MSD) of the center of mass of each chemical group in MD simulations. In particular, $\langle u^2 \rangle$ can be experimentally measured via the

quasi-elastic neutron scattering (QENS) technique, which is found to be closely related to the local free volume and mobility and inversely related to the local molecular stiffness ($1/\langle u^2 \rangle$) (i.e., inverse local mobility).[30] Previously, QENS was broadly employed to study the dynamics of polycarbonate (PC) films,[50] non-conjugated polymers with alkyl side chains in bulk samples,[51,52] and polyalkylthiophenes CPs.[53–56]

Figure 4C shows the $\langle u^2 \rangle$ results for DPPT with different side-chain lengths from MD simulations. The results show that $\langle u^2 \rangle$ increases with temperature due to the enhanced segmental mobility upon heating; at a higher temperature above $T_g$, $\langle u^2 \rangle$ starts to deviate from a near-linear trend at a lower temperature. Another observation is that, at a lower $T$, there is no significant difference in $\langle u^2 \rangle$ values for different alkyl side-chain lengths; however, above 350 K, the increase in side-chain length is found to increase $\langle u^2 \rangle$ and thus the chain mobility. To validate our MD simulations, we also calculate the MSD of DPPT polymer via QENS measurement (by directly probing the dynamics of the hydrogen atoms due to its large incoherent scattering section), and the results are shown in Figure 4D. For a temperature range up to 430 K, two observations are important for us to examine our MD simulations. First, for both MD simulations and QENS, the range of $\langle u^2 \rangle$ is varied from 0 to 5 $\text{Å}^2$, while the MD simulation yields a slightly lower magnitude. With increasing the side-chain length, $\langle u^2 \rangle$ values obtained via QENS start to increase at a higher rate beyond a critical temperature. Such characteristic behavior is observed for simulation data as well. Considering the nature and control of the two measurements, the consistency between experiments and simulation indicates that our MD simulation could serve as a reliable tool to computationally explore the segmental dynamics of CPs. Below, we focus on analyzing segmental dynamics of each functional group at an elevated $T \approx 1.5\ T_g$ for different CPs.

Figure 4E shows the probability distribution of $1/\langle u^2 \rangle$, a measure of inverse local mobility, for four different building blocks of the DPPT polymers with C2C8C10 side chains, i.e., DPP core (labeled as M1 in blue), thiophene rings tied to DPP core (group M2 marked in green), isolated thiophene moiety that is not tied to DPP core (group M3 marked in red), and alkyl side chains (labeled as M4 in purple). Apparent Gaussian distribution of $1/\langle u^2 \rangle$ is observed for all four functional groups. The fused DPP unit (M1) shows a larger mean of $1/\langle u^2 \rangle = 0.72\ \text{Å}^{-2}$, implying that this compound indicates relatively lowest mobility among all functional groups in this polymer. In contrast, side chains (M4) exhibit the lowest mean of $1/\langle u^2 \rangle = 0.53\ \text{Å}^{-2}$ with a smaller standard deviation due to higher mobility of side chains. The thiophene rings in the DPPT monomer exhibit different inverse local mobility based on their chemical environment—thiophenes directly tied to the DPP core (M2) show a slightly higher mean in the distribution ($1/\langle u^2 \rangle = 0.67\ \text{Å}^{-2}$) compared with M3 thiophenes ($1/\langle u^2 \rangle = 0.65\ \text{Å}^{-2}$). The MD simulation results of segmental dynamics of the DPPT are largely consistent with our feature analyses in the ML model, where the fused rings yield a steeper increase in the $T_g$ compared with the isolated rings. The observed lower inverse local mobility of the side chains is also in line with the negative contribution coefficient of the *NSC* feature in the ML model.

Figure 4F shows the distribution of $1/\langle u^2 \rangle$ of different chemical building blocks of PCDTBT having three different groups of aromatic rings. Similar to DPPT, it can be observed that CDT bridged unit (M1) shows the highest mean value of

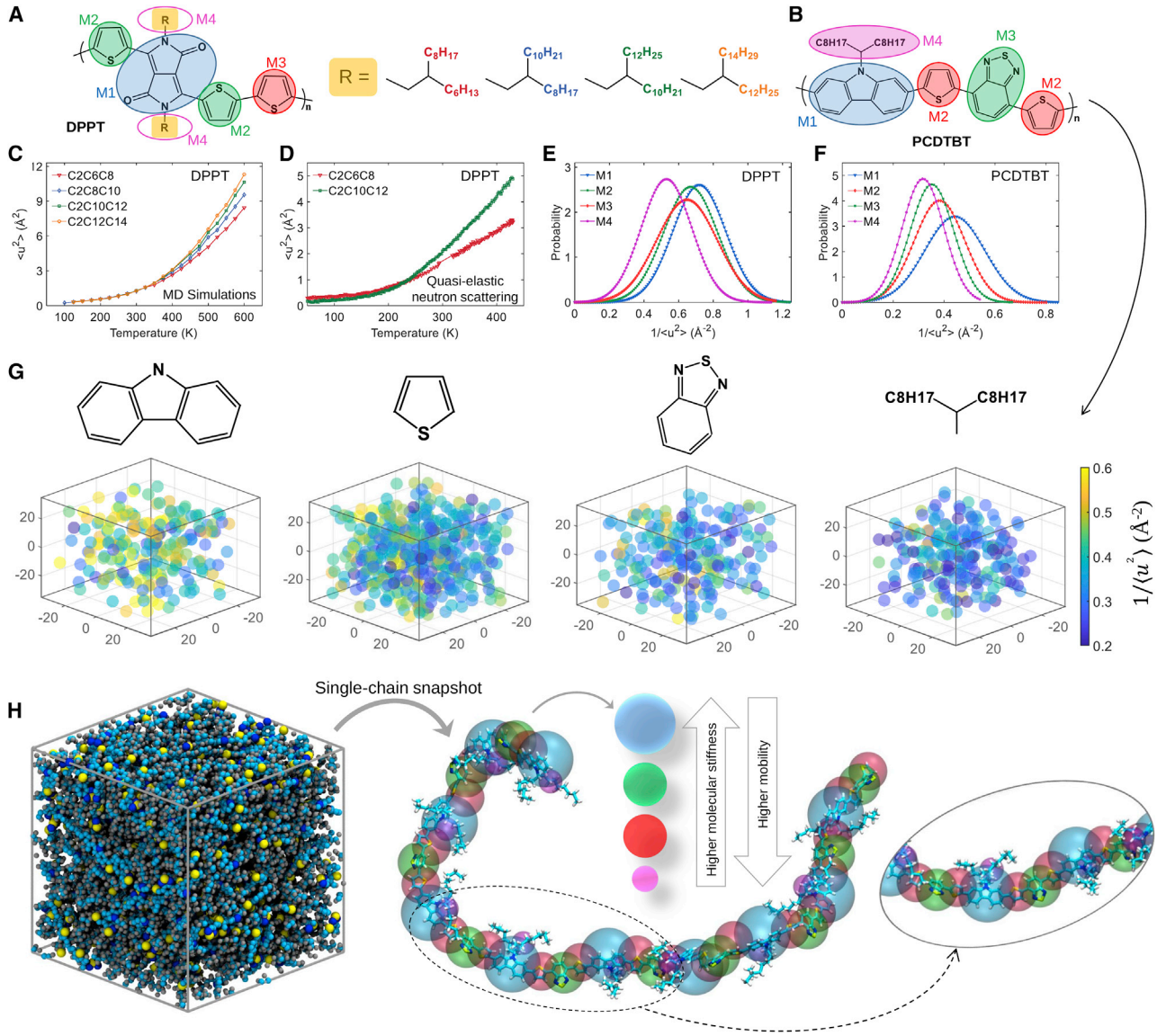



**Figure 4. Segmental dynamics of different building blocks of conjugated polymers**

(A) Chemical structure of DPP-based polymer with four different side-chain lengths.

(B) Chemical structure of PCDTBT polymer.

(C and D) Mean-squared displacement (MSD) of the DPP-based polymers with varied alkyl side chains obtained by (C) molecular dynamics (MD) simulations and (D) quasi-elastic neutron scattering.

(E and F) Normal distribution of inverse local mobility ($1/\langle u^2\rangle$) of different functional groups of (E) DPP-based polymers (C2C8C10) and (F) PCDTBT obtained by MD simulations.

(G) Spatial distribution of $1/\langle u^2\rangle$ of different functional groups of PCDTBT polymer. Each point represents the center of mass of the relevant functional group in the bulk system. Yellowish color means higher $1/\langle u^2\rangle$ corresponding to lower mobility. The blueish color represents lower $1/\langle u^2\rangle$ related to higher mobility.

(H) A representative snapshot bulk simulation box and single-chain snapshot of PCDTBT polymer obtained by MD simulations. A larger transparent sphere illustrates higher $1/\langle u^2\rangle$. To have more clear illustration, one repeat unit of the polymer chain is magnified.

$1/\langle u^2\rangle = 0.44\ \text{Å}^{-2}$ with a larger mean of the distribution, while the side chains (M4) show the lowest value ($1/\langle u^2\rangle = 0.44\ \text{Å}^{-2}$). Furthermore, the fused benzothiadiazole (BT) unit (M3) exhibits a lower mean distribution of $1/\langle u^2\rangle = 0.35\ \text{Å}^{-2}$ compared with thiophene rings (M2) ($1/\langle u^2\rangle = 0.38\ \text{Å}^{-2}$), implying that M3 fused moiety has higher mobility than isolated rings (M2). This could be attributed to the fact that

only the phenyl ring of the BT unit is tied to the backbone, and its current vertical orientation makes it easier for out-of-plane rotation, resulting in relatively higher mobility than isolated rings (M2). Our result again suggests that the local dynamics of the functional groups are largely dependent on the chemical environment.

Figure 4G shows the spatial distributions of $1/\langle u^2 \rangle$ of different functional groups of PCDTBT, confirming again that CDT bridged unit and side chains have the highest and lowest magnitude of inverse local mobility, respectively. Each point illustrates the center of mass of the relevant functional group in the bulk system, with the color showing the magnitude of inverse local mobility $1/\langle u^2 \rangle$. Interestingly, the spatial distribution of $1/\langle u^2 \rangle$ reveals a reduced level of dynamical heterogeneity observed from M1 to M4 as their mobility increases, which reflects their packing efficiency at a molecular level; i.e., a higher level dynamical heterogeneity corresponds to a more frustrated molecular packing. Such dynamical heterogeneity associated with different chemical groups of PCDTBT is further illustrated in Figure 4H, where larger transparent spheres indicate higher local $1/\langle u^2 \rangle$. Similar dynamics behaviors are also observed for IID and NDI-based donor-acceptor CPs (Figure S2), where the fused rings (i.e., IID and NDI core) and side chains exhibit the highest and lowest magnitude of inverse local mobility, respectively. The MD simulation results of local dynamics corroborate findings of ML modeling that aromatic rings (i.e., *NIR*, *NFR*, and *NBR* features) and side chains (*NSC*) govern $T_g$ prediction of CPs.

### Application of ML model for predicting $T_g$ of conjugated polymers

In this section, we test the ML model by applying it to predicting $T_g$ of representative CPs particularly related to organic solar cell applications. Device stability is a challenge that severely limits organic solar cells from being widely adopted to replace silicon ones and their inorganic counterparts. Through decades of optimizing the optoelectronic properties of narrow bandgap CPs together with non-fullerene acceptors (NFAs), remarkable progress has been made in organic photovoltaics (OPV) technology, with the power conversion efficiency (PCE) reaching over 18%.[57–59] However, the long-term stability of OPV devices is one of the major hurdles that prevents this technology from being widely implemented, along with the lack of scalable manufacturing techniques to obtain ideal bulk heterojunction (BHJ) morphology. We expect that our ML model developed here can potentially rationalize the stability issue and guide the new CP materials development.

We apply the model to the leading optoelectronic polymers and NFAs (see Figure 5 for their chemical structure) that were recently reported in high-performance solar cells. For the PM6, PM7, and PTB7 polymers, $T_g$ values of 123.2°C, 123.2°C, and 91.7°C are predicted by the ML model, respectively. We further experimentally measure $T_g$ for PM6 and PTB7 polymers, and the results are shown in Figure 5, in which DMA and DSC techniques indicate a $T_g$ of 138.6°C and 114.7°C, respectively, for PM6 and a $T_g$ of 73.8°C for PTB7 (obtained by DMA). Both conventional and fast scanning calorimeter are employed to detect $T_g$ for PM6, where the $T_g$ values were consistent for both methods regarding the difference in cooling rates (e.g., the conventional DSC cooling rate is 0.17 K/s and the flash DSC cooling rate is 0.1 K/s). In addition, MD simulations predict $T_g$ value of 145.6°C for the PM7 polymer close enough to $T_g$ of 123.2°C predicted by our ML framework. For the Y-family NFAs, as indicated in Figure 5 and Table S1, $T_g$ values predicted by the ML model are found in good agreement with those calculated by MD simulations. In addition, for other experimentally measured as well as simulated $T_g$ results for different CPs employed in this study, listed in Figure 5 and Table S2, a good

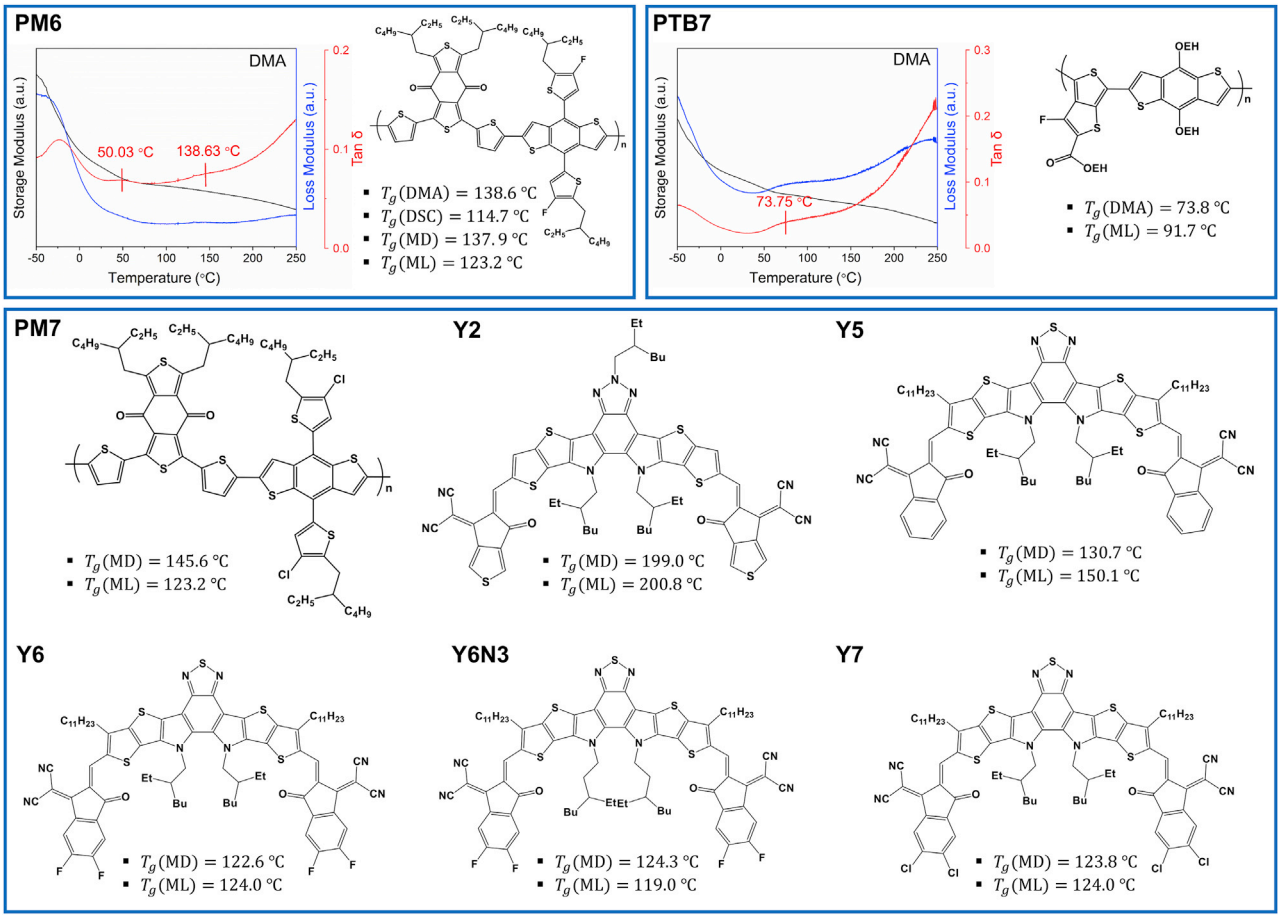



**Figure 5. Glass transition temperature ($T_g$) of high-performance conjugated polymers and acceptor units with aromatic backbones and alkyl side chains**

Results of $T_g$ are obtained from MD simulations, machine learning (ML) prediction, and available experimental techniques (DMA and DSC).

agreement is observed with ML prediction, implying the reliability of our predictive framework.

Based on our discovery here, all PM6, PM7, PTB7, and Y-family NFAs possess a relatively high $T_g$, as compared with other CPs, with the operating temperature of the device at room temperature or up to 80°C for most upper limit in inorganic cells. The $T_g$ values for measured components are slightly higher than the operating temperature; thus, CPs are limited in chain dynamics within the experimental timescale for device degradation. The observed PCE drop over time below $T_g$ thus is highly likely due to changes in the interface property between the BHJ layer with electrodes or due to the diffusion-limited cold crystallization in either donor or acceptors. For the small molecules in Y-family NFAs, other researchers suggested this is possible. In the case of tuning $T_g$ based on our ML model, for example, for the PM6, $T_g$ could be improved by lowering the side-chain length from 2-ethylhexyl to 2-ethylbutyl in order to raise $T_g$ above 150°C. Additional attention should also be paid to the role of cold crystallization with respect to $T_g$ of both donors and acceptors, which have been discussed in previous studies.[60–63] In general, the practical usage of our ML model is to quantitatively estimate the unknown $T_g$ for advanced D-A CP polymers used in organic solar cell field, since $T_g$ determines the dynamics of the possible cold crystallization occurred under operational conditions. Cold

crystallization can form different polymorphs depending on the isothermal annealing temperature higher or lower than nominal $T_g$.[63] The polymorph formed at higher temperature normally impairs the performance, whereas the polymorph formed at lower temperature could improve performance presumably due to the stronger light absorption from the continuous aromatic structures.[62] However, the size of crystallite also influences the performance of organic solar cells that the larger size tends to be detrimental to the performance and stability.[61] Therefore, the observed PCE drop over time below $T_g$ thus is highly likely due to changes in the interface property between the BHJ layer with electrodes or due to the large size of crystallites formed via diffusion-limited cold crystallization in either donor or acceptors.

In addition to proposing a predictive ML framework, we would like to emphasize the importance of the integrated methodology established in this study through the utilization of ML, experimental measurements, and MD simulations. Although the ML model shows an acceptable predictive performance, the current methodology could be continuously improved with the availability of more reliable $T_g$ data of newly designed CPs, generated from both experiments and simulations. Particularly, for those structural features with limited data (e.g., halogen atoms), the addition of more data points can appropriately tune the contribution of different building blocks on the $T_g$ of CPs.

In summary, we establish a ML model to predict $T_g$ of CPs from the chemical structure of the repeat unit. With 154 $T_g$ values collected from diverse resources, a predictive model is developed where six functional groups in the repeat unit (i.e., side-chain fraction, isolated rings, fused rings, bridged rings, double and triple bonds, and halogen atoms) are identified as essential structural features. Among the selected features, side chains and halogen atoms show a decreasing influence on $T_g$. The t-statistic parameter is employed to quantify the contribution of each structural feature in the prediction of $T_g$, where aromatic rings and side chains are found to have the most significant influence. Our MD simulations further reveal that aromatic rings and side chains have the highest and lowest local mobility, respectively, unraveling their dominant roles in the ML model. Our experimental measurements of $T_g$ as well as those calculated by MD simulations confirm the noticeably good predictive performance of our ML model for high-performance solar cell materials. We believe that the current ML framework provides an effective strategy to design the next generation of organic electronics with tailored chain dynamics.

## EXPERIMENTAL PROCEDURES

### Resource availability

#### *Lead contact*

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Wenjie Xia (wenjie.xia@ndsu.edu).

#### *Materials availability*

This study did not generate new unique materials.

#### *Data and code availability*

All the necessary data supporting the main findings of the paper are available within the main paper and its supplemental information files and from the lead contact upon reasonable request.

### Materials

The materials for $T_g$ measurements were purchased from Ossila Company, including F8BT, PTB7, PTB70Th, PM6, PCPDTBT, TFB, TQ1, NDI(2HD)T, NDI(2HD)2T, NDI(2OD)T, and NDI(2OD)2T.

### Quasi-elastic neutron scattering

QENS measurements were performed using the Backscattering Spectrometer (BASIS)[64] at the Spallation Neutron Source (SNS) at Oak Ridge National Laboratory (ORNL). Timescales for BASIS range from 10 to 1,000 ps.[56] Elastic scan measurements were carried out under vacuum to protect the material from degradation at high temperatures between 50 and 450 K. Scattering data were reduced utilizing the DAVE software[65] to measure MSD.

### Dynamic mechanical analysis

DMA measurements were carried out with a TA Q800 DMA by a modified DMA method.[31,66] For sample preparation, polymer solutions (5 to 10 mg/mL) were drop casted on top of a glass fiber mesh. At a fixed frequency of 1 Hz, temperature ramp experiments with a heating rate of 3°C $min^{-1}$ were measured between −50 and 250°C to measure the $T_g$ of the materials. For more details, please see Figure S3.

### Differential scanning calorimetry

DSC measurements were performed using both Mettler-Toledo fast scanning calorimeter (Flash DSC 2+) and conventional DSC (DSC 3+). The Flash DSC was equipped with an ultra-fast standard chip with heating and cooling rates up to 4,000 K/s. Flash DSC has been utilized to detect thin-film $T_g$ since 2011.[67] The investigation of nanosized polystyrene (PS) thin-film $T_g$ showed a good agreement between conventional DSC and Flash DSC during the overlapped cooling rates (0.1 K/s to 1 K/s).[68] In addition to conventional polymers, the rigid backbone of CPs restricted the $T_g$ measurements by conventional DSC, but Flash DSC was able to probe weak $T_g$ of conjugated polymers.[69] To fairly compare the results from conventional and flash DSC, similar cooling rates were adopted, where 0.17 K/s for conventional DSC and 0.1 K/s for flash DSC were employed, respectively (see Figure S4). For conventional DSC, the indium standard was used to calibrate the heat flow and temperatures. All measurements were done under nitrogen gas.

### Dataset

The experimental values of $T_g$ are collected from the previously published literature, including 119 data points.[16,31,38] To improve the diversity of the dataset, we further experimentally measured an additional 17 $T_g$ data points of different CPs from either DMA and DSC measurements as well as 18 $T_g$ obtained by MD simulations. These additional data cover newly designed high-performance CPs, allowing for the external validation and improvement of the model prediction. Overall, the dataset is composed of 154 data points of 94 polymers covering thiophene-, fluorene-, DDP-, bridged-, and fused-based CPs; newly designed non-fluorine acceptor units; and some non-conjugated polymers with aromatic rings in their backbone. The dataset is provided in Tables S1 and S2. It should be noted that sample quality, control and resolution of the measurement equipment, size of the samples, and many other factors are a matter of concern about the collected dataset. In this work, this concern is alleviated by utilizing a relatively large dataset that covers a wide range of $T_g$ for CPs.

### ML-based surrogate model

Informed from the QSPR modeling and feature analyses, we use the aforementioned dataset of CPs to define the structural features based on the geometry (i.e., chemical building block) of the repeat unit. We first define all possible structural features, including the number of different atoms in the repeat units, bridged units, fused rings, free isolated rings, stiff bonds, heavy atoms, side-chain fraction, etc. Then, backward elimination,[39] a feature selection technique, is utilized to screen those structural features while building the regression models. To perform backward elimination, we introduce a significance level of 5% to test the null hypothesis for each feature based on the p value.[70] It should be noted that the p value threshold is not considered here as a hard rule for the feature selection,[71] and all final selected molecular features are physically interpreted to make sure they are rationally relevant to the $T_g$ of CPs. The regression model is then fitted to the dataset, and the p value of each feature is calculated, which determines whether the feature will generalize any population out of the current sample. Next, the feature with the highest p value is removed from the predictive model, and the process iterates for the remaining features. This mechanism continues until all p values of all remaining features in the dataset are less than the significance level. Then, a regression model is trained to the selected feature to make a $T_g$ prediction.

### MD simulations

All-atomistic MD simulations of the 18 bulk CPs and acceptor units are carried out using the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) software package.[72] The interactions of the MD models are defined via the general amber force field (GAFF), which was originally developed for the organic molecules.[73] The CP models are initially built using Materials Studio software followed by geometry optimization. Then, the antechamber module of the AmberTools20 package[73] is employed for the atom typing and calculation of atomic charges. To equilibrate the system, the total potential energy is first minimized using an iterative conjugate gradient algorithm,[74] followed by the equilibration of the bulk system under the isothermal-isobaric (NPT) ensemble at the melt state and then cooling down to the target temperature. An integration time step of $\Delta t = 1$ fs is implemented in all MD simulations. To determine $T_g$ of bulk CP, the simulation is run further under NPT ensemble to collect the density over a wide range of the temperature from 100 K to 800 K. The obtained density as a function of temperature can be fitted with linear lines at low-temperature glassy and high-temperature melt regimes, where the intersection of fitted lines marks $T_g$. To measure the local mobility of moieties, MSD (i.e., a measurement of the deviation of position of atoms from the reference position over time) $\langle r^2(t)\rangle$ values are obtained for the center of mass functional groups of the repeat unit, $\langle r^2(t)\rangle = \langle |r_i(t) - r_i(0)|^2\rangle$, where $r_i(t)$ is the position of the $i$th atoms at time $t$ and $\langle r^2(t)\rangle$ is obtained from the average of all related functional groups. Then, Debye-Waller factor $\langle u^2\rangle$, a fast-dynamic property at a picosecond timescale,[30] is defined at $t = 4$ ps for all models to quantify local mobility.

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrp.2022.100911.

## ACKNOWLEDGMENTS

A.A., Z.L., and W.X. acknowledge support from the National Science Foundation (NSF) under NSF OIA ND-ACES award no. 1946202. A.A., Z.L., and W.X.

acknowledge support from the North Dakota Established Program to Stimulate Competitive Research (ND EPSCoR) through the New Faculty Award; the Department of Civil, Construction and Environmental Engineering; and the College of Engineering at North Dakota State University (NDSU). This work used supercomputing resources of the CCAST at NDSU, which were made possible in part by NSF MRI award no. 2019077. S.Z. and X.G. thank NSF (DMR-2047689) for providing funding for the experimental characterization of $T_g$. Z.C., H.Z., and X.G. thank the DOE BES program for providing funding to support the neutron scattering experiments under award number DE-SC0022050. Part of the research used resources at the Spallation Neutron Source (SNS) and the Center for Nanophase Materials Sciences (CNMS), DOE Office of Science User Facilities operated by the Oak Ridge National Laboratory. The authors thank Naresh C. Osti (SNS) for assistance during the QENS experiments.

## AUTHOR CONTRIBUTIONS

A.A., Z.C., and Z.L. contributed equally to this work. W.X. and X.G. conceived and directed the project. Z.C., S.Z., and H.Z. performed the DMA and DSC tests. Z.C. performed and analyzed QENS tests. A.A. performed the machine learning analysis. A.A. and Z.L. conducted molecular dynamics simulations and analyzed the simulation results. A.A., Z.C., Z.L., X.G., and W.X. wrote the manuscript, and all the coauthors commented on the manuscript.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 18, 2021
Revised: April 18, 2022
Accepted: May 4, 2022
Published: May 31, 2022

## REFERENCES

1. Gu, X., Shaw, L., Gu, K., Toney, M.F., and Bao, Z. (2018). The meniscus-guided deposition of semiconducting polymers. Nat. Commun. *9*, 534. https://doi.org/10.1038/s41467-018-02833-9.

2. Root, S.E., Savagatrup, S., Printz, A.D., Rodriquez, D., and Lipomi, D.J. (2017). Mechanical properties of organic semiconductors for stretchable, highly flexible, and mechanically robust electronics. Chem. Rev. *117*, 6467–6499. https://doi.org/10.1021/acs.chemrev.7b00003.

3. Someya, T., Bao, Z., and Malliaras, G.G. (2016). The rise of plastic bioelectronics. Nature *540*, 379–385. https://doi.org/10.1038/nature21004.

4. Lipomi, D.J., and Bao, Z. (2017). Stretchable and ultraflexible organic electronics. MRS Bull *42*, 93–97. https://doi.org/10.1557/mrs.2016.325.

5. Bao, Z., and Chen, X. (2016). Flexible and stretchable devices. Adv. Mater. *28*, 4177–4179. https://doi.org/10.1002/adma.201601422.

6. Oh, J.Y., and Bao, Z. (2019). Second skin enabled by advanced electronics. Adv. Sci. *6*, 1900186. https://doi.org/10.1002/advs.201900186.

7. Lipomi, D.J., Tee, B.C.K., Vosgueritchian, M., and Bao, Z. (2011). Stretchable organic solar cells. Adv. Mater. *23*, 1771–1775. https://doi.org/10.1002/adma.201004426.

8. Chortos, A., Liu, J., and Bao, Z. (2016). Pursuing prosthetic electronic skin. Nat. Mater. *15*, 937–950. https://doi.org/10.1038/nmat4671.

9. Zhang, S., Alesadi, A., Mason, G.T., Chen, K.L., Freychet, G., Galuska, L., Cheng, Y.H., Onge, P.B.J.S., Ocheje, M.U., Ma, G., et al. (2021). Molecular origin of strain-induced chain alignment in PDPP-based semiconducting polymeric thin films. Adv. Funct. Mater. *31*, 2100161. https://doi.org/10.1002/adfm.202100161.

10. Oh, J.Y., Rondeau-Gagné, S., Chiu, Y.C., Chortos, A., Lissel, F., Wang, G.J.N., Schroeder, B.C., Kurosawa, T., Lopez, J., Katsumata, T., et al. (2016). Intrinsically stretchable and healable semiconducting polymer for organic transistors. Nature *539*, 411–415. https://doi.org/10.1038/nature20102.

11. Wang, S., Xu, J., Wang, W., Wang, G.J.N., Rastak, R., Molina-Lopez, F., Chung, J.W., Niu, S., Feig, V.R., Lopez, J., et al. (2018). Skin electronics from scalable fabrication of an intrinsically stretchable transistor array. Nature *555*, 83–88. https://doi.org/10.1038/nature25494.

12. Xu, J., Wang, S., Wang, G.J.N., Zhu, C., Luo, S., Jin, L., Gu, X., Chen, S., Feig, V.R., To, J.W.F., et al. (2017). Highly stretchable polymer semiconductor films through the nanoconfinement effect. Science *355*, 59–64. https://doi.org/10.1126/science.aah4496.

13. Xu, J., Wu, H., Mun, J., Ning, R., Wang, W., Wang, G.N., Nikzad, S., Yan, H., Gu, X., Luo, S., et al. (2021). Tuning conjugated polymer chain packing for stretchable semiconductors. Adv. Mater. 2104747. https://doi.org/10.1002/adma.202104747.

14. Maller, O., Drain, A.P., Barrett, A.S., Borgquist, S., Ruffell, B., Zakharevich, I., Pham, T.T., Gruosso, T., Kuasne, H., Lakins, J.N., et al. (2021). Tumour-associated macrophages drive stromal cell-dependent collagen crosslinking and stiffening to promote breast cancer aggression. Nat. Mater. *20*, 548–559. https://doi.org/10.1038/s41563-020-00849-5.

15. Qian, Z., Galuska, L., McNutt, W.W., Ocheje, M.U., He, Y., Cao, Z., Zhang, S., Xu, J., Hong, K., Goodman, R.B., et al. (2019). Challenge and solution of characterizing glass transition temperature for conjugated polymers by differential scanning calorimetry. J. Polym. Sci.

Part B Polym. Phys. *57*, 1635–1644. https://doi.org/10.1002/polb.24889.

16. Qian, Z., Cao, Z., Galuska, L., Zhang, S., Xu, J., and Gu, X. (2019). Glass transition phenomenon for conjugated polymers. Macromol. Chem. Phys. *220*, 1900062. https://doi.org/10.1002/macp.201900062.

17. Karuth, A., Alesadi, A., Xia, W., and Rasulev, B. (2021). Predicting glass transition of amorphous polymers by application of cheminformatics and molecular dynamics simulations. Polymer, 123495. https://doi.org/10.1016/j.polymer.2021.123495.

18. Xie, S.J.J., Qian, H.J.J., and Lu, Z.Y.Y. (2015). The glass transition of polymers with different side-chain stiffness confined in free-standing thin films. J. Chem. Phys. *142*, 074902. https://doi.org/10.1063/1.4908047.

19. Xu, W.S., and Freed, K.F. (2014). Influence of cohesive energy and chain stiffness on polymer glass formation. Macromolecules *47*, 6990–6997. https://doi.org/10.1021/ma501581u.

20. Alesadi, A., and Xia, W. (2020 Mar 27). Understanding the role of cohesive interaction in mechanical behavior of a glassy polymer. Macromolecules *53*, 2754–2763. https://doi.org/10.1021/acs.macromol.0c00067.

21. Xu, W.S., Douglas, J.F., and Freed, K.F. (2016). Influence of cohesive energy on the thermodynamic properties of a model glass-forming polymer melt. Macromolecules *49*, 8341–8354. https://doi.org/10.1021/acs.macromol.6b01503.

22. Xiao, M., Sadhanala, A., Abdi-Jalebi, M., Thomas, T.H., Ren, X., Zhang, T., Chen, H., Carey, R.L., Wang, Q., Senanayak, S.P., et al. (2021). Linking glass-transition behavior to photophysical and charge transport properties of high-mobility conjugated polymers. Adv. Funct. Mater. *31*, 2007359–2007411. https://doi.org/10.1002/adfm.202007359.

23. Borges, R.M., Colby, S.M., Das, S., Edison, A.S., Fiehn, O., Kind, T., Lee, J., Merrill, A.T., Merz, K.M., Metz, T.O., et al. (2021). Quantum chemistry calculations for metabolomics. Chem. Rev. *121*, 5633–5670. https://doi.org/10.1021/acs.chemrev.0c00901.

24. Vargas-Lara, F., Starr, F.W., and Douglas, J.F. (2017). Molecular rigidity and enthalpy-entropy compensation in DNA melting. Soft Matter *13*, 8309–8330. https://doi.org/10.1039/c7sm01220a.

25. Xu, W.S., Douglas, J.F., and Xu, X. (2020). Molecular dynamics study of glass formation in polymer melts with varying chain stiffness. Macromolecules *53*, 4796–4809. https://doi.org/10.1021/acs.macromol.0c00731.

26. Cao, Z., Leng, M., Cao, Y., Gu, X., and Fang, L. (2021). How rigid are conjugated non-ladder and ladder polymers? J. Polym. Sci. *60*, 298–310. https://doi.org/10.1002/pol.20210550.

27. Zhang, W., Gomez, E.D., and Milner, S.T. (2014). Predicting chain dimensions of semiflexible polymers from dihedral potentials. Macromolecules *47*, 6453–6461. https://doi.org/10.1021/ma500923r.

28. McCulloch, B., Ho, V., Hoarfrost, M., Stanley, C., Do, C., Heller, W.T., and Segalman, R.A. (2013). Polymer chain shape of poly(3-alkylthiophenes) in solution using small-angle neutron scattering. Macromolecules *46*, 1899–1907. https://doi.org/10.1021/ma302463d.

29. Müller, C., Esmaeili, M., Riekel, C., Breiby, D.W., and Inganäs, O. (2013). Micro X-ray diffraction mapping of a fluorene copolymer fibre. Polymer *54*, 805–811. https://doi.org/10.1016/j.polymer.2012.10.059.

30. Pazmiño Betancourt, B.A., Hanakata, P.Z., Starr, F.W., Douglas, J.F., Betancourt, B.A.P., Hanakata, P.Z., Starr, F.W., Douglas, J.F., Pazmiño Betancourt, B.A., Hanakata, P.Z., et al. (2015). Quantitative relations between cooperative motion, emergent elasticity, and free volume in model glass-forming polymer materials. Proc. Natl. Acad. Sci. U S A *112*, 2966–2971. https://doi.org/10.1073/pnas.1418654112.

31. Zhang, S., Alesadi, A., Selivanova, M., Cao, Z., Qian, Z., Luo, S., Galuska, L., Teh, C., Ocheje, M.U., Mason, G.T., et al. (2020). Toward the prediction and control of glass transition temperature for donor–acceptor polymers. Adv. Funct. Mater. *30*, 2002221. https://doi.org/10.1002/adfm.202002221.

32. Katritzky, A.R., Kuanar, M., Slavov, S., Hall, C.D., Karelson, M., Kahn, I., and Dobchev, D.A. (2010). Quantitative correlation of physical and chemical properties with chemical structure: utility for prediction. Chem. Rev. *110*, 5714–5789. https://doi.org/10.1021/cr900238d.

33. Katritzky, A.R., Rachwal, P., Law, K.W., Karelson, M., and Lobanov, V.S. (1996). Prediction of polymer glass transition temperatures using a general quantitative structure-property relationship treatment. J. Chem. Inf. Comput. Sci. *36*, 879–884. https://doi.org/10.1021/ci950156w.

34. Katritzky, A.R., Sild, S., Lobanov, V., and Karelson, M. (1998). Quantitative Structure - property Relationship (QSPR) correlation of glass transition temperatures of high molecular weight polymers. J. Chem. Inf. Comput. Sci. *38*, 300–304. https://doi.org/10.1021/ci9700687.

35. Doan Tran, H., Kim, C., Chen, L., Chandrasekaran, A., Batra, R., Venkatram, S., Kamal, D., Lightstone, J.P., Gurnani, R., Shetty, P., et al. (2020). Machine-learning predictions of polymer properties with Polymer Genome. J. Appl. Phys. *128*, 171104. https://doi.org/10.1063/5.0023759.

36. Jackson, N.E., Webb, M.A., and de Pablo, J.J. (2019). Recent advances in machine learning towards multiscale soft materials design. Curr. Opin. Chem. Eng. *23*, 106–114. https://doi.org/10.1016/j.coche.2019.03.005.

37. Sivaraman, G., Jackson, N.E., Sanchez-Lengeling, B., Vázquez-Mayagoitia, Á., Aspuru-Guzik, A., Vishwanath, V., and de Pablo, J.J. (2020). A machine learning workflow for molecular analysis: application to melting points. Mach. Learn. Sci. Technol. *1*, 025015. https://doi.org/10.1088/2632-2153/ab8aa3.

38. 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., et al. (2020). Glass transition temperature from the chemical structure of conjugated polymers. Nat. Commun. *11*, 893–911. https://doi.org/10.1038/s41467-020-14656-8.

39. Hocking, R.R. (1976). A Biometrics invited paper. The analysis and selection of variables in linear regression. Biometrics *32*, 1–49. https://doi.org/10.2307/2529336.

40. Chiang, Y.C., Wu, H.C., Wen, H.F., Hung, C.C., Hong, C.W., Kuo, C.C., Higashihara, T., and Chen, W.C. (2019). Tailoring carbosilane side chains toward intrinsically stretchable semiconducting polymers. Macromolecules *52*, 4396–4404. https://doi.org/10.1021/acs.macromol.9b00589.

41. Balar, N., Rech, J.J., Siddika, S., Song, R., Schrickx, H.M., Sheikh, N., Ye, L., Megret Bonilla, A., Awartani, O., Ade, H., et al. (2021). Resolving the molecular origin of mechanical relaxations in donor–acceptor polymer semiconductors. Adv. Funct. Mater. *32*, 2105597. https://doi.org/10.1002/adfm.202105597.

42. Olla, T., Ibraikulov, O.A., Ferry, S., Boyron, O., Méry, S., Heinrich, B., Heiser, T., Léveque, P., and Leclerc, N. (2019). Benzothiadiazole halogenation impact in conjugated polymers, a comprehensive study. Macromolecules *52*, 8006–8016. https://doi.org/10.1021/acs.macromol.9b01760.

43. Reichenbächer, K., Süss, H.I., and Hulliger, J. (2005). Fluorine in crystal engineering—"the little atom that could. Chem. Soc. Rev. *34*, 22–30. https://doi.org/10.1039/b406892k.

44. Desiraju, G.R. (2007). Crystal engineering: a holistic view. Angew. Chem. Int. Ed. *46*, 8342–8356. https://doi.org/10.1002/anie.200700534.

45. Kroon, R., Gehlhaar, R., Steckler, T.T., Henriksson, P., Müller, C., Bergqvist, J., Hadipour, A., Heremans, P., and Andersson, M.R. (2012). New quinoxaline and pyridopyrazine-based polymers for solution-processable photovoltaics. Sol. Energy Mater. Sol. Cells *105*, 280–286. https://doi.org/10.1016/j.solmat.2012.06.029.

46. Zhang, S., Ocheje, M.U., Huang, L., Galuska, L., Cao, Z., Luo, S., Cheng, Y.H., Ehlenberg, D., Goodman, R.B., Zhou, D., et al. (2019). The critical role of electron-donating thiophene groups on the mechanical and thermal properties of donor–acceptor semiconducting polymers. Adv. Electron. Mater. *5*, 1800899–1800911. https://doi.org/10.1002/aelm.201800899.

47. Roth, B.B., Savagatrup, S., de los Santos N, V., Hagemann, O., Carlé, J.E., Helgesen, M., Livi, F., Bundgaard, E., Søndergaard, R.R., Krebs, F.C., et al. (2016). Mechanical properties of a library of low-band-gap polymers. Chem. Mater. *28*, 2363–2373. https://doi.org/10.1021/acs.chemmater.6b00525.

48. Lipomi, D.J., Chong, H., Vosgueritchian, M., Mei, J., and Bao, Z. (2012). Toward mechanically robust and intrinsically stretchable organic solar cells: evolution of photovoltaic properties with tensile strain. Sol. Energy Mater. Sol. Cells *107*, 355–365. https://doi.org/10.1016/j.solmat.2012.07.013.

49. Li, Y., Tatum, W.K., Onorato, J.W., Zhang, Y., and Luscombe, C.K. (2018). Low elastic modulus and high charge mobility of low-crystallinity indacenodithiophene-based semiconducting polymers for potential applications in stretchable electronics.

Macromolecules *51*, 6352–6358. https://doi.org/10.1021/acs.macromol.8b00898.

50. Soles, C.L., Douglas, J.F., Wu, W.L., Peng, H., and Gidley, D.W. (2004). Comparative specular x-ray reflectivity, positron annihilation lifetime spectroscopy, and incoherent neutron scattering measurements of the dynamics in thin polycarbonate films. Macromolecules *37*, 2890–2900. https://doi.org/10.1021/ma035579a.

51. Sakai, V.G., and Arbe, A. (2009). Quasielastic neutron scattering in soft matter. Curr. Opin. Colloid Interface Sci. *14*, 381–390. https://doi.org/10.1016/j.cocis.2009.04.002.

52. Gerstl, C., Schneider, G.J., Fuxman, A., Zamponi, M., Frick, B., Seydel, T., Koza, M., Genix, A.C., Allgaier, J., Richter, D., et al. (2012). Quasielastic neutron scattering study on the dynamics of poly (alkylene oxide)s. Macromolecules *45*, 4394–4405. https://doi.org/10.1021/ma3003399.

53. Obrzut, J., and Page, K.A. (2009). Electrical conductivity and relaxation in poly (3-hexylthiophene). Phys. Rev. B *80*, 195211. https://doi.org/10.1103/PhysRevB.80.195211.

54. Guilbert, A.A.Y., Urbina, A., Abad, J., Díaz-Paniagua, C., Batallán, F., Seydel, T., Zbiri, M., García-Sakai, V., and Nelson, J. (2015). Temperature-dependent dynamics of polyalkylthiophene conjugated polymers: a combined neutron scattering and simulation study. Chem. Mater. *27*, 7652–7661. https://doi.org/10.1021/acs.chemmater.5b03001.

55. Zhan, P., Zhang, W., Jacobs, I.E., Nisson, D.M., Xie, R., Weissen, A.R., Colby, R.H., Moulé, A.J., Milner, S.T., Maranas, J.K., and Gomez, E.D. (2018). Side chain length affects backbone dynamics in poly (3-alkylthiophene) s. J. Polym. Sci. Part B Polym. Phys. *56*, 1193–1202. https://doi.org/10.1002/polb.24637.

56. Wolf, C.M., Kanekal, K.H., Yimer, Y.Y., Tyagi, M., Omar-Diallo, S., Pakhnyuk, V., Luscombe, C.K., Pfaendtner, J., and Pozzo, L.D. (2019). Assessment of molecular dynamics simulations for amorphous poly (3-hexylthiophene) using neutron and X-ray scattering experiments. Soft Matter *15*, 5067–5083. https://doi.org/10.1039/C9SM00807A.

57. Meng, L., Zhang, Y., Wan, X., Li, C., Zhang, X., Wang, Y., Ke, X., Xiao, Z., Ding, L., Xia, R., et al. (2018). Organic and solution-processed tandem solar cells with 17.3% efficiency. Science *361*, 1094–1098. https://doi.org/10.1126/science.aat2612.

58. Wang, Z., Gao, K., Kan, Y., Zhang, M., Qiu, C., Zhu, L., Zhao, Z., Peng, X., Feng, W., Qian, Z., et al. (2021). The coupling and competition of crystallization and phase separation, correlating thermodynamics and kinetics in OPV morphology and performances. Nat. Commun. *12*, 332–414. https://doi.org/10.1038/s41467-020-20515-3.

59. Liu, Q., Jiang, Y., Jin, K., Qin, J., Xu, J., Li, W., Xiong, J., Liu, J., Xiao, Z., Sun, K., et al. (2020). 18% Efficiency organic solar cells. Sci. Bull. *65*, 272–275. https://doi.org/10.1016/j.scib.2020.01.001.

60. Hultmark, S., Paleti, S.H.K., Harillo, A., Marina, S., Nugroho, F.A.A., Liu, Y., Ericsson, L.K.E., Li, R., Martín, J., Bergqvist, J., et al. (2020). Suppressing co-crystallization of halogenated non-fullerene acceptors for thermally stable ternary solar cells. Adv. Funct. Mater. *30*, 2005462. https://doi.org/10.1002/adfm.202005462.

61. Hu, H., Ghasemi, M., Peng, Z., Zhang, J., Rech, J.J., You, W., Yan, H., and Ade, H. (2020). The role of demixing and crystallization kinetics on the stability of non-fullerene organic solar cells. Adv. Mater. *32*, 2005348. https://doi.org/10.1002/adma.202005348.

62. Marina, S., Scaccabarozzi, A.D., Gutierrez-Fernandez, E., Solano, E., Khirbat, A., Ciammaruchi, L., Iturrospe, A., Balzer, A., Yu, L., Gabirondo, E., et al. (2021). Polymorphism in non-fullerene acceptors based on indacenodithienothiophene. Adv. Funct. Mater. *31*, 2103784. https://doi.org/10.1002/adfm.202103784.

63. Yu, L., Qian, D., Marina, S., Nugroho, F.A.A., Sharma, A., Hultmark, S., Hofmann, A.I., Kroon, R., Benduhn, J., Smilgies, D.M., et al. (2019). Diffusion-limited crystallization: a rationale for the thermal stability of non-fullerene solar cells. ACS Appl. Mater. Interfaces *11*, 21766–21774. https://doi.org/10.1021/acsami.9b04554.

64. Mamontov, E., and Herwig, K.W. (2011). A time-of-flight backscattering spectrometer at the Spallation Neutron Source, BASIS. Rev. Sci. Instrum. *82*, 085109. https://doi.org/10.1063/1.3626214.

65. Azuah, R.T., Kneller, L.R., Qiu, Y., Tregenna-Piggott, P.L.W., Brown, C.M., Copley, J.R.D., and Dimeo, R.M. (2009). DAVE: a comprehensive software suite for the reduction, visualization, and analysis of low energy neutron spectroscopic data. J. Res. Natl. Inst. Stand. Technol. *114*, 341. https://doi.org/10.6028/jres.114.025.

66. Sharma, A., Pan, X., Campbell, J.A., Andersson, M.R., and Lewis, D.A. (2017). Unravelling the thermomechanical properties of bulk heterojunction blends in polymer solar cells. Macromolecules *50*, 3347–3354. https://doi.org/10.1021/acs.macromol.7b00430.

67. Mathot, V., Pyda, M., Pijpers, T., Vanden Poel, G., Van de Kerkhof, E., Van Herwaarden, S., Van Herwaarden, F., and Leenaers, A. (2011). The Flash DSC 1, a power compensation twin-type, chip-based fast scanning calorimeter (FSC): first findings on polymers. Thermochim. Acta *522*, 36–45. https://doi.org/10.1016/j.tca.2011.02.031.

68. Gao, S., Koh, Y.P., and Simon, S.L. (2013). Calorimetric glass transition of single polystyrene ultrathin films. Macromolecules *46*, 562–570. https://doi.org/10.1021/ma3020036.

69. Martín, J., Stingelin, N., and Cangialosi, D. (2018). Direct calorimetric observation of the rigid amorphous fraction in a semiconducting polymer. J. Phys. Chem. Lett. *9*, 990–995. https://doi.org/10.1021/acs.jpclett.7b03110.

70. McClave, J.T., and Sincich, T. (2016). Statistics (Pearson Education).

71. Grabowski, B. (2016). "P< 0.05" might not mean what you think: American Statistical Association clarifies P values. JNCI J. Natl. Cancer Inst. *108*, djw194. https://doi.org/10.1093/jnci/djw194.

72. Plimpton, S. (1995). Fast parallel algorithms for short-range molecular dynamics. J. Comput. Phys. *117*, 1–19. https://doi.org/10.1006/jcph.1995.1039.

73. Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A., and Case, D.A. (2004). Development and testing of a general Amber force field. J. Comput. Chem. *25*, 1157–1174. https://doi.org/10.1002/jcc.20035.

74. Payne, M.C., Teter, M.P., Allan, D.C., Arias, T.A., and Joannopoulos, J.D. (1992). Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients. Rev. Mod. Phys. *64*, 1045–1097. https://doi.org/10.1103/RevModPhys.64.1045.