The Journal of Chemical Physics

RESEARCH ARTICLE | AUGUST 21 2023

Determining glass transition in all-atom acrylic polymeric melt simulations using machine learning

Special Collection: Machine Learning Hits Molecular Simulations

Atreyee Banerjee ; Aysenur Iscen ; Kurt Kremer ; Oleksandra Kukharenko

J. Chem. Phys. 159, 074108 (2023)
https://doi.org/10.1063/5.0151156

The Journal of Chemical Physics ARTICLE pubs.aip.org/aip/jcp

Determining glass transition in all-atom acrylic polymeric melt simulations using machine learning

Cite as: J. Chem. Phys. 159, 074108 (2023); doi: 10.1063/5.0151156
Submitted: 19 March 2023 • Accepted: 27 July 2023 •
Published Online: 21 August 2023

Atreyee Banerjee,[a] Aysenur Iscen, Kurt Kremer, and Oleksandra Kukharenko[b]

AFFILIATIONS

Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany

Note: This paper is part of the JCP Special Topic on Machine Learning Hits Molecular Simulations.
[a]**Electronic mail:** banerjeea@mpip-mainz.mpg.de
[b]**Author to whom correspondence should be addressed:** kukharenko@mpip-mainz.mpg.de

ABSTRACT

The functionality of many polymeric materials depends on their glass transition temperatures (T_g). In computer simulations, T_g is often calculated from the gradual change in macroscopic properties. Precise determination of this change depends on the fitting protocols. We previously proposed a robust data-driven approach to determine T_g from the molecular dynamics simulation data of a coarse-grained semi-flexible polymer model. In contrast to the global macroscopic properties, our method relies on high-resolution microscopic details. Here, we demonstrate the generality of our approach by using various dimensionality reduction and clustering methods and apply it to an atomistic model of acrylic polymers. Our study reveals the explicit contribution of the side chain and backbone residues in influencing the determination of the glass transition temperature.

I. INTRODUCTION

Most polymeric materials in application are used in their glassy state. Therefore, functionality and the applicability of these materials depend on the glass transition temperature, T_g. This necessitates an accurate prediction of T_g to better control the properties of the material for specific applications. The static properties, such as radial distribution function and structure factor, do not show any notable change around T_g; in contrast, the dynamic properties, such as viscosity and relaxation time, substantially increase in a super-Arrhenius manner.[1–4] Experimentally, T_g of polymer melts is calculated from the change in heat capacity using differential scanning calorimetry (DSC),[5] thermal expansion coefficient using thermomechanical analysis (TMA),[6] or viscoelastic behavior of polymers using dynamic mechanical analysis (DMC).[7] However, the experimentally relevant cooling rates are inaccessible in computer simulations, and this leads to ambiguity while connecting computational models of polymeric materials to experimentally observed properties. Computationally, T_g is determined from changes in the macroscopic properties such as specific volume, density, or energy. The reliable predictions of T_g are challenging, particularly for systems where these changes occur gradually.[8–12] Recent studies have shown attempts to define T_g based on the changes in molecular structures of polymeric materials. For example, by quantifying the changes in specific dihedral angles and transitions between states defined by those angles.[10,13] Baker *et al.*[14] have observed that the glass transition temperatures for polymer melts with different chain lengths can be collapsed on a master curve using the local intra-chain conformational dynamics with packing. Machine learning (ML) methods hold great promise to automatize the determination of structural properties of the glassy systems that can reflect the changes in T_g from molecular dynamics (MD) simulation data, but their application for polymeric materials is limited.[10,15,16] Iwaoka and Takano[15] applied principal component analysis (PCA)[17] to Cartesian coordinates of short polymer chains in a melt. They showed a change in eigenvalue distributions before and after glass transition and connected this change to approximate conformational entropy differences.

Recently, we proposed a new unsupervised machine-learning approach to estimate T_g from molecular dynamics simulation trajectories.[18] Our data-driven methodology takes into account the structural information of individual chains and makes use of high-resolution information obtained from molecular dynamics simulations. By combining principal component analysis[17] and a density-based clustering algorithm (DBSCAN),[19] we identified T_g at the asymptotic limit even from relatively short-time trajectories. The PCA captured the change in nature of the fluctuations in the system: from conformational fluctuation above T_g to localized rearrangements below T_g. Our method was introduced on a coarse-grained model of polymer melts containing weakly semiflexible polymer chains.[20] In this work, we extend it and show the generality of our approach by using nonlinear dimensionality reduction, agglomerative hierarchical clustering[21] and apply it to other glass-forming liquids.

Here, we analyze the simulation trajectories of all-atom acrylic polymers[22] found in acrylic paints: poly(methyl methacrylate) (PMMA), poly(ethyl acrylate) (PEA), and poly(n-butyl acrylate) (PnBA); see Fig. 1. The glass transition temperatures for acrylic polymers are close to room temperature, i.e., 333–387 K[23–26] for PMMA, 249 K[27] or 231 K[28] for PEA, and 223 K[25] for PnBA. This proximity to room temperature is relevant for both, the stability of a painting and the applicability during the painting process. However, it also means that the paints can suffer from crack formation at low temperatures while becoming sticky at high temperatures. Consequently, the temperature has a significant impact on the degradation of acrylics. As our analysis relies on the microscopic observables, we explicitly investigate the role of the side chain and backbone atoms in determining the glass transition temperature. Extending our approach to another nonlinear multi-dimensional scaling approach enables us to apply it to a very small number of observations per temperature, which is the case for simulation data obtained with, e.g., continuous cooling protocols.[9] Here, we employ agglomerative clustering that requires minimum prior knowledge about the system and hyperparameter tuning. Overall, with various dimensionality reduction and clustering techniques, the generality of our approach is tested for these acrylic polymer melts.

The paper is organized as follows: In Sec. II, we present the methods and simulation details. The results obtained from different methods are given in Sec. III. In Sec. IV, we provide an overall discussion of the methods. Finally, we conclude our results in Sec. V.

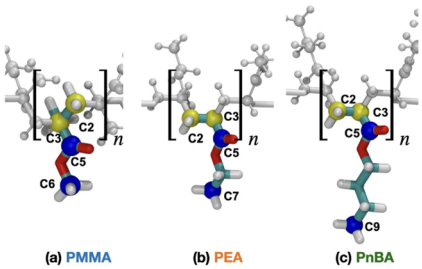


FIG. 1. Chemical structure of the polymers used in this study. Only the selected carbon atoms in the backbone (yellow) and side chain (blue) are used in the calculation of T_g for (a) poly(methyl methacrylate), (b) poly(ethyl acrylate), and (c) poly (n-butyl acrylate). In our model, each polymer chain is composed of $n = 15$ monomers.

II. METHODS AND SIMULATION DETAILS

A. Simulation details and data preparation

1. Atomistic simulations

Molecular dynamics simulations of bulk homopolymers (PMMA, PEA, and PnBA) were presented in Ref. 22. They were performed using the general Amber force field[29] with Gromacs 2019 software.[30,31] The isotactic 15-mer polymer chains were constructed with AmberTools,[32] and $n_{ch} = 100$ polymer chains were placed randomly in a box with initial dimensions $9 \times 9 \times 9$ nm^3 using Packmol.[33] Each system was minimized for 1000 steps by using the steepest descent algorithm. Following the minimization, we equilibrated for 200 ps using the NVT (constant number of particles, volume, and temperature) at 600 K. We further equilibrated the system with NPT (constant number of particles, pressure, and temperature) ensemble for 10 ns to maintain temperature (600 K) and pressure (1 bar), respectively. After this short equilibration at 600 K, the temperature was decreased to 100 K using a cooling rate of 20 K/ns to calculate the glass transition temperature and obtain correct polymer density at each temperature. The coordinates of the system at temperature intervals of 50 K were saved and further equilibrated for 10 ns using NPT ensemble to study the temperature-dependent properties of the polymers. Atomic coordinates were saved every 100 ps for the trajectory analysis resulting in 100 frames per temperature. Further information on the simulation details is given in Ref. 22.

2. Input data

To identify the liquid-to-glass transition from the simulation trajectories, we extract information from each individual chain independently. As possible descriptors, we use sets of internal pairwise distances. To retain the information about conformational fluctuations of individual polymer chains of 15-mers, we choose three different sets of descriptors—all pairwise distances between (a) the backbone C-atoms (C2, C3) (**bb-bb**), (b) side chain C-atoms (C5, C6 for PMMA; C5, C7 for PEA; C5, C9 for PnBA) (**sc-sc**), and (c) the backbone C-atoms (C2) and the end side chain C-atoms (C6 for PMMA; C7 for PEA; and C9 for PnBA) (**bb-sc**). A snapshot of selected carbon atoms is given in Fig. 1. The yellow beads correspond to the backbone atoms (C2, C3), and blue beads, to the side chain atoms (C5, C6/C7/C9) for each monomer.

We extract this information from 100 snapshots of each NPT run (we get similar results using NVT simulations) of simulation trajectories at 11 temperatures in the range 100–600 K for each chain independently. In this way, for each chain in the melt, we construct the data matrix $\mathbf{X}_{ch} \in \mathbb{R}^{M \times L}$, where $ch = 1, \ldots, n_{ch}$ is a chain index, n_{ch} is a number of chains in the melt, M is the number of simulation snapshots/frames, and L is the number of descriptors: all pairwise distances between side chains C-atoms (**sc-sc**), backbone C-atoms (**bb-bb**), and backbone side chain (**bb-sc**) separately (for our systems here, the length of descriptors $L = 435$ for all three cases).

B. Data-driven identification of T_g

We have recently developed a data-driven protocol[18] to calculate T_g described below. The analysis workflow consists of two independent methods: (I) using combined information from all temperatures [Fig. 2(a)] or (II) individual information from each temperature [Fig. 2(b)].

1. Dimensionality reduction

For both methods I and II, we first perform a dimensionality reduction once the possible sets of the internal descriptors are defined (as described in Sec. II). In the original paper,[18] we used (PCA),[17] which identifies linearly uncorrelated subsets of the data space containing most of the variance of the original data. PCA has been successfully used to characterize different physical phenomena, i.e., the phase transition in conserved Ising spin systems,[34,35] secondary structure prediction of proteins,[36] shape fluctuation in DNA,[37] and glass transition temperature prediction in polymer melts.[15,18] Here, PCA is applied on the internal pairwise distances' matrix $\mathbf{X}_{ch}$ of a randomly selected single chain. Before applying PCA, $\mathbf{X}_{ch}$ is standardized column wise, i.e., it is mean free, and standard deviation is equal to one. First, the covariance matrix $C_{ch} = \mathbf{X}_{ch}^T\mathbf{X}_{ch} \in \mathbb{R}^{L\times L}$ is calculated. Then, the pairs of eigenvalues $\lambda_{ch,i}$ and eigenvectors $\mathbf{v}_{ch,i}$ for $i = 1, 2, 3, \ldots,$ min (L, M) are found for C_{ch} and sorted in the decreasing order of $\lambda_{ch,i}$. The original data $\mathbf{X}_{ch}$ are then projected $\tilde{\mathbf{X}}_{ch,i} = \mathbf{X}_{ch}\mathbf{v}_{ch,i}$ to the new orthogonal basis of principal components (PCs) formed by P leading eigenvectors $\mathbf{v}_{ch,i}$, where $i = 1, \ldots, P$ and $P \le$ min (L, M) is the reduced number of dimensions. For the index notations of the above equations, see Ref. 18. The number of leading principal components (P) to project the input data on is usually chosen to retain a specific amount of variance in the input data reflected by the magnitude of the respective eigenvalues (e.g., see Fig. S4).

PCA is able to perform only linear mapping of the data to a lower-dimensional space. To evaluate whether this linear projection is able to capture all of the important details, here, we also used the recently introduced nonlinear dimensionality reduction method—cc_analysis[38,39]—that was successfully applied for the analysis of protein data.[40,41] It is a variant of multi-dimensional scaling methods[42] in which a dimensionality reduction is performed by minimizing the loss function between the distances (or other metrics) between data in original high- and target low-dimensional spaces. The key feature of cc_analysis is that it minimizes the differences between Pearson correlation coefficients[43] of pairs of high-dimensional datasets and the scalar product of the low-dimensional vectors representing them [see Eq. (1)] projecting the data into a unit sphere,

$$\sum_{i=1}^{M-1}\sum_{j=i+1}^{M}\left(r_{ij} - \mathbf{x}_i \cdot \mathbf{x}_j\right)^2 \to \min. \tag{1}$$

Here, r_{ij} is the correlation coefficient between configurations X_i and X_j in the high-dimensional space (the rows of the matrix $\mathbf{X}_{ch}$), $X_i \in \mathbb{R}^L$, $i, j = 1, \ldots, M$; $\mathbf{x}_i \cdot \mathbf{x}_j$ denotes the dot product of the unit

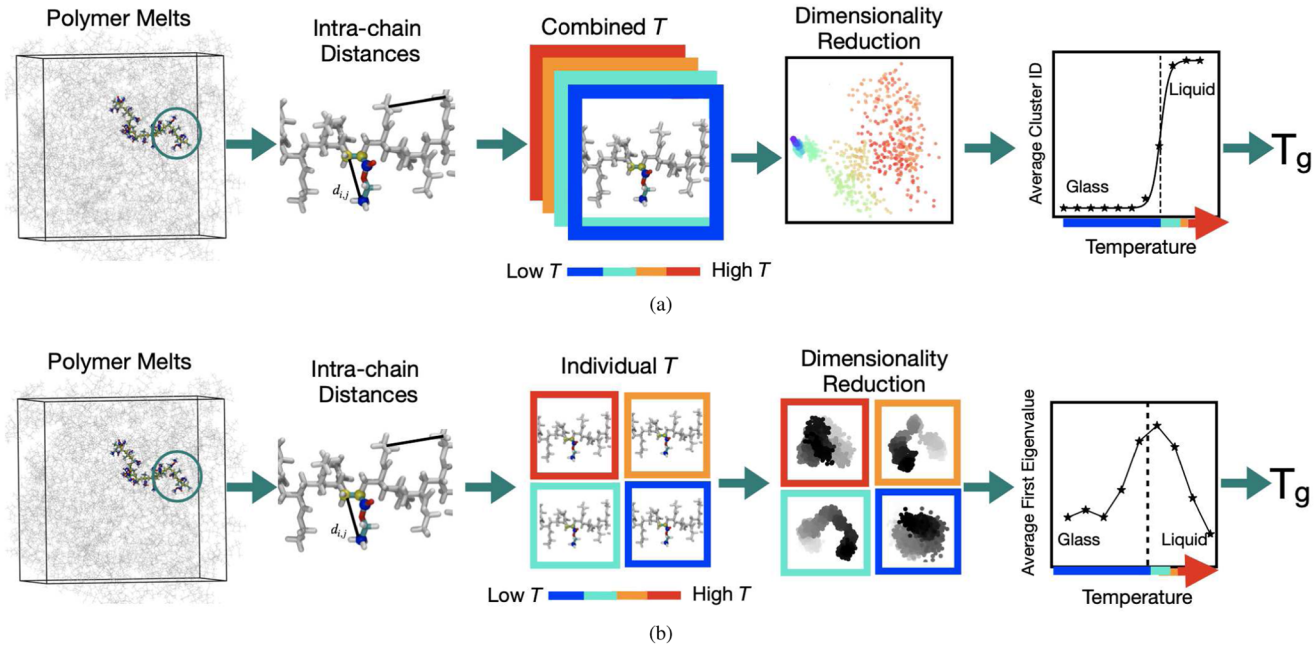


FIG. 2. Schematic representation of the workflow employed in the paper to determine glass transition temperature from combined (a, Method I) and individual (b, Method II) temperature analyses. Both methods start with analysis of the simulations for a single chain and calculation of internal descriptors. Then, in Method I (a), the data from all temperatures are combined, standardized, and followed by dimensionality reduction and clustering. This procedure is repeated for all chains in the melt. An inflection point of averaged cluster indices is assumed to be the T_g. In Method II (b), the data from each temperature are analyzed independently: standardized and followed by dimensionality reduction. An inflection point of first eigenvalues (or participation ratio) averaged over all chains as a function of temperature is used as an indicator for T_g.

vectors $\mathbf{x}_i$ and $\mathbf{x}_j$ representing the data in the low-dimensional space $\tilde{\mathbf{X}}_{ch}$. Importantly, similar to PCA, for deciding on the number of required dimensions for low-dimensional projection of cc_analysis, one can use the gap in the eigenvalues of the $\mathbf{X}_{ch}\mathbf{X}_{ch}^T \in \mathbb{R}^{M\times M}$ matrix (please note the difference to the C_{ch} of the PCA covariance matrix, which is calculated column wise but will have the same eigenvalues and eigenvectors up to normalization). The dependence of the results on the number of dimensions for Method I using both dimensionality reduction methods is discussed in the results sections.

As a result of dimensionality reduction, each data point in the latent projection corresponds to the configuration of one chain at a given time and temperature.

2. Clustering

For Method I, after the dimensionality reduction using linear or nonlinear methods, we perform clustering on the new lower-dimensional projected space $\tilde{\mathbf{X}}_{ch} \in \mathbb{R}^{M\times P}$ for each polymer chain. In our previous paper,[18] we used a density-based spatial clustering of applications with noise (DBSCAN),[19] which groups together the data points that are close based on two hyperparameters: the Euclidean distance to create a neighborhood and the minimum number of points to form a dense region. The clustering performance depends on the careful choice of the hyperparameters and can produce any number of clusters. In this manuscript, instead of DBSCAN, we use agglomerative clustering.[21] For this clustering method, the number of expected clusters is the main hyperparameter. Since our dataset consists of either glassy or liquid states, our natural choice for the number of clusters to find is 2. As other two parameters, we used Euclidean distance as the metric and Ward linkage.[44] We use agglomerative clustering both on PCA and cc_analysis space to observe the change from liquid to glassy states. As the result of the clustering, each chain ch at each simulation frame and temperature will get a cluster index (ID) n_i, where n_i is an integer (i.e., $n_i \in \{0, 1, \ldots, n_{cluster} - 1\}$, $n_{cluster} = 2$ for agglomerative clustering in this work) for $i = 1, 2, \ldots, n_{ch}M$, n_{ch} is the number of chains and M is the number of frames. Subsequently, the matrix $\tilde{\mathbf{X}}_{ch} \in \mathbb{R}^{M\times P}$ will be transformed to a vector $\tilde{X}_{ch} \in \{n_i | n_i = 0 \text{ or } 1\}^M$. In this work, $n_i = 0$ corresponds to the glassy state, whereas $n_i = 1$ corresponds to the liquid state.

3. Method I

We use the data combined from simulations from all temperatures, resulting in the data matrix $\mathbf{X}_{ch} = \mathbf{X}_{ch}(\bigcup T) \in \mathbb{R}^{1100\times 435}$ for each chain ch, $ch = 1, \ldots, n_{ch}$. We perform the dimensionality reduction on $\mathbf{X}_{ch}$ (linear or nonlinear) obtaining a new reduced projection space $\tilde{\mathbf{X}}_{ch}$. We cluster this new space assigning a cluster index n_i to each configuration of the chain ch and obtaining a vector of cluster indices $\tilde{X}_{ch}$. Then, we repeat dimensionality reduction with subsequent clustering on each chain present in the system for a total of 100 chains. This implies that each chain will be represented as a vector of 1100 cluster IDs (each 100 elements of this vector corresponds to the chain's configuration at the same temperature). Then, we calculate average cluster indices $\overline{\mathrm{ID}}(T)$ as a mean over cluster IDs of all chains that were simulated at the same temperature. At each temperature T, the average cluster index $\overline{\mathrm{ID}}(T)$ is given as

$$\overline{\mathrm{ID}}(T) = \sum_{n_i=0}^{n_{cluster}-1} n_i P(n_i, T), \tag{2}$$

where $P(n_i, T)$ is the probability distribution of cluster IDs for all n_{ch} chains over all simulation frames at each T.

Finally, the sharp change in the average cluster IDs is used to determine the glass transition temperature [see the schematic in Fig. 2(a)]. To quantify the jump, we interpolate the data by a hyperbolic tangent function,

$$g(T) = C(1 - \tanh(sT - d))/2, \tag{3}$$

where s and d are the fitting parameters and C is the gap between the two states at $T \gg T_g$ and $T \ll T_g$, respectively. The inflexion point [a point in which $g''(T) = 0$] defines T_g and is estimated as $T_g = T_{\text{inflextion}} = d/s$.

4. Long time approximation with method I

With the aim to extrapolate glass transition temperature at the long observation time limits, we calculate the average cluster indices for different observation time windows Δt with equal intervals. We choose k_t different observation time windows Δt ($k_t = 9$ with Δt from 2 to 10 ns in this work). For each Δt, we have used $\Delta t/t_{lag}$ consecutive frames with $t_{lag} = 100$ ps. We fit the average cluster ID values using Eq. (3) and calculate the inflexion points at different Δt. Then, we observe the change in $T_g(\Delta t)$ as the inverse of Δt. At relatively larger time windows, it follows a linear behavior and an extrapolation of this linear behavior to $1/\Delta t \to 0$, i.e., infinite observation time, allows us to estimate an asymptotic limit of T_g [$T_g(\Delta t \to \infty)$] from a relatively short trajectory length.

5. Method II

In contrast to Method I, we analyze the data from each temperature independently. In this case, $\mathbf{X}_{ch} \in \mathbb{R}^{2500\times 435}$. Here, we used the atomic coordinates saved every 4 ps for the trajectory analysis resulting in 2500 frames per temperature. In this way, we do not provide any information on individual chain conformations from different temperatures to dimensionality reduction methods. We standardize the data and perform PCA for each chain at each temperature independently obtaining a set of eigenvalues $\lambda_{ch,i}(T)$, where ch is a chain index, T is the temperature at which we performed PCA, and i is the eigenvalue index (they all sorted in descending order and $i = 1$ correspond to the highest eigenvalue), $i = 1, \ldots, L$. Additionally, we calculate the participation ratio (PR) over the set of eigenvalues defined at each temperature as

$$\mathrm{PR}_{ch}(T) = \left(\sum_{i=1}^{L} \lambda_{ch,i}(T)\right)^2 / \sum_{i=1}^{L} \lambda_{ch,i}(T)^2, \tag{4}$$

where $\lambda_{ch,i}(T)$ are eigenvalues of PCA sorted in descending order. PR can be viewed as the effective dimensionality of the data. It reflects the decay rate of eigenvalues: faster decay results in a smaller PR compared with a slower decay rate.

We repeat this analysis for all $ch = 1, \ldots, n_{ch}$ chains in the melt. In Ref. 18, we observed that T_g can also be determined from the change in monotonic behavior of the first eigenvalue of PCA or the participation ratio (PR) averaged over all chains as a function of temperature:

$$\overline{\lambda_1}(T) = \frac{1}{n_{ch}} \sum_{ch=1}^{n_{ch}} \lambda_{ch,1}(T), \tag{5}$$

$$\overline{\mathrm{PR}}(T) = \frac{1}{n_{ch}} \sum_{ch=1}^{n_{ch}} \mathrm{PR}_{ch}(T). \tag{6}$$

$\overline{\mathrm{PR}}(T)$ is a more general measure than $\overline{\lambda_1}(T)$ as it can be applied to the data with different variances. In our case, we standardize the data that allows us to average over the eigenvalues as well. Method II is summarized in Fig. 1.

III. RESULTS

A. Method I: Analysis of combined temperatures

1. Results with PCA

We start our analysis for each of the three systems (PMMA, PEA, and PnBA) by performing PCA on a randomly selected single chain of the homopolymer using the input descriptors (**sc-sc**, **bb-bb**, and **bb-sc**) over the 10 ns classical MD simulations concatenated for all temperatures. In Fig. 3, we show the projection of the data onto two leading principal components (PCs) for **sc-sc** descriptors. The projections in the new PCA space can be viewed as linear combinations of all input descriptors. Each point in the plot corresponds to the chain's conformation at a given temperature at each time. The amount of variance explained by eigenvectors for PCA analysis is given in supplementary material, Fig. S4). Even in two-dimensional projection, we observe that for all three homopolymers, the low-temperature states are concentrated in a small region on the PCA projection (Fig. 3). We clustered the data using agglomerative clustering, as described in Sec. II B. In the inset of Fig. 3, we plot the same projection colored based on obtained cluster indices. The agglomerative clustering groups the low temperature ($T < T_g$) part into one cluster with cluster ID = 0 (maroon color) and the liquid state to cluster with ID = 1 (gray color).

To obtain a general estimate of the temperature at which the separation between the liquid and the glassy states (characterized by a change in cluster IDs) occurs, we perform PCA for each chain separately, followed by agglomerative clustering and averaging over obtained cluster IDs as a function of temperature as given in Fig. 4(a). To quantify the jump, we use Eq. (3) for fitting the data (detail in Sec. II B) and define the inflexion points as the glass transition temperatures. Figure S2 shows the results do not change after applying DBSCAN instead of agglomerative clustering to the same data as in Fig. 4(a). Comparing the obtained results for the three systems considered here we observe that T_g decreases with increasing the side chain length of the polymer, which is in good agreement with the previous analysis[22] and shown in Fig. 3. The methods of predicting T_g based on the macroscopic properties such as density or volume are sensitive to the fitting protocols as the transition around T_g is not sharp. For example, when the same data were fitted with a different fitting range, the estimated T_g for PMMA was lower[22] compared with the value we report in Fig. 4(c) or a similar model in Ref. 10. A different choice of the fitting range leads to around 50 K shift in T_g value for PnBA system, see Fig. S1, suggesting that T_g is highly affected by the bilinear fitting uncertainties.

In Fig. 4(b), we plot T_g values obtained for different input descriptors **bb-bb**, **sc-sc**, and **bb-sc** as a function of the number dimensions the data were projected on. The estimated T_g values averaged over different clustering results for each of the descriptors are given in Table I. We do not see notable differences in T_g values when we compare side chain and backbone contributions. Since in Method I the fluctuations of all the temperatures are scaled together, the dominant contribution in fluctuation is expected to come from the high-temperature states suggesting that both the side chain and backbone fluctuate more at high temperatures compared with their low-temperature states. However, the scenario changes completely when we treat each temperature separately in Method II. This will be discussed later in Sec. III B.

It is important to note here, as was already discussed in Ref. 22, the experimental T_g values for the considered polymers are lower than our prediction due to the differences in the cooling rates accessible to MD simulations and experiments (experiential rates are much lower). Several studies have attempted to link those cooling rates[46] and proposed an adjustment to T_g values for acrylic polymers calculated from MD simulations.[22] Our results agree well with the previously calculated T_g values (Table I) from simulations of acrylic

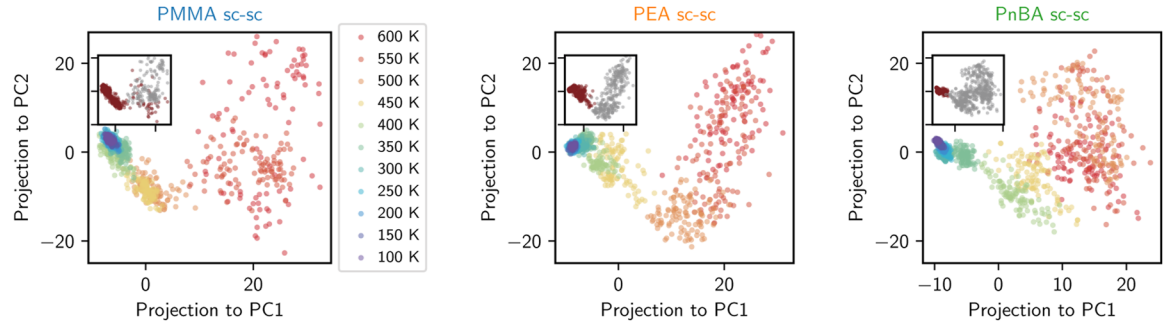
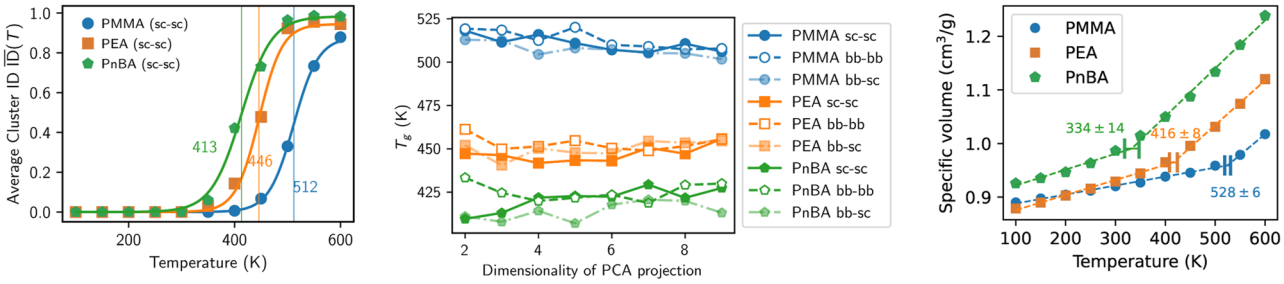


FIG. 3. Projections of a single chain over multiple time frames for (a) PMMA, (b) PEA, and (c) PnBA in the reduced PCA space determined using the **sc-sc** input descriptors. Each point in the plot corresponds to the chain's conformation at a given temperature at each time. Projections are colored varying from red to blue from high temperature to low temperature. Note that the axis values in the PCA embedding do not correspond to any physical quantity. (Inset) The same projection is colored based on clustering IDs. Gray and maroon colors represent cluster IDs 1 (liquid) and 0 (glass), respectively.



(a) (b) (c)

FIG. 4. (a) Agglomerative clustering on the three-dimensional PCA projections for the three different polymers. Results are averaged over all chains, and the average cluster index, $\overline{\mathrm{ID}}(T)$, is plotted for each temperature. To quantify the jump, we interpolate the data by using Eq. (3). The fitting functions are given as the solid curves (fitting parameters are listed in supplementary material, Table S1). The vertical lines correspond to the inflection points, i.e., the glass transition temperatures. T_g decreases with increasing the side chain lengths of the polymer melt. (b) The glass transition temperatures calculated for three different polymers as a function of the number of leading PCs ($P = 2, \ldots, 9$) used for clustering for different input descriptors: **bb-bb**, **sc-sc**, and **bb-sc**. (c) Glass transition temperatures from the bilinear fits of the low- and high-temperature regions of the specific volume. The fitting is done by the pwlf package[45] using the whole data as input and specifying one breakpoint. The error bars are nonlinear standard errors associated with the piece-wise linear fitting.

TABLE I. Comparison of T_g values (in K) obtained with different methods.

	MMA	EA	NBA
Specific volume fitting[a]	478 ± 8,[22] 526,[10] 524 ± 6[b]	416 ± 8[22]	334 ± 14[22]
PCA[c] (sc-sc)	511 ± 4	447 ± 4	421 ± 6
PCA (bb-bb)	513 ± 5	453 ± 4	425 ± 5
PCA (bb-sc)	507 ± 4	450 ± 4	414 ± 5
cc_analysis (sc-sc)	521 ± 2	457 ± 1	422 ± 3
cc_analysis (bb-bb)	537 ± 1	465 ± 1	444 ± 2
cc_analysis (bb-sc)	520 ± 2	457 ± 1	421 ± 3
$\Delta t \to \infty$	506 ± 5	430 ± 5	402 ± 2

[a]Error values are nonlinear standard errors associated with the piece-wise linear fitting calculated using the Delta method.[45]
[b]This study.
[c]The error values are the standard deviations associated with the number of reduced dimensions used for clustering, detail in Fig. 4(b).

polymers or coarse-grained bead-spring polymer model.[18] Overall, our result also shows a good qualitative agreement with the experimental findings, where T_g values decrease with increasing side chain length.

One of the limitations of all-atom MD simulation is the short time that can be accessed due to the large number of atoms. This presents a challenge when calculating properties, such as glass transition temperatures, because the dynamics of the polymer melt are slow. Therefore, a long equilibration (hundreds of nanoseconds) is necessary. Here, the simulation data we have are only 10 ns for each temperature, which is considered short for equilibration of macroscopic properties, such as specific volume (Fig. 3). In Sec. II, we provide a method to extrapolate T_g to long time limits from a relatively short trajectory. We perform PCA followed by clustering at nine different observation time windows Δt from 2 to 10 ns with 1 ns interval and interpolate the data using Eq. (3). The inflexion point, i.e., T_g, is now calculated for different lengths of time, Δt, along the trajectory. Taking into account, this finite-time effect, we found a linear dependency of T_g and Δt for the bead-spring entangled polymer model that allows estimating an asymptotic limit of T_g from a relatively short trajectory length. We extended this approach to acrylic polymers. In Fig. 5, we plot the inflection points at different Δt values and find linear-type behavior at larger observation times. Although the fitting uncertainty in the all-atom system is relatively high compared with the CG model, one can still extrapolate the T_g values in the asymptotic limit within some error bars of the fitting. The obtained values are $T_g \approx 506 \pm 5(K)$ for MMA, $430 \pm 5(K)$ for EA, and $402 \pm 2(K)$; however, one requires longer simulation runs for reliable linear fitting. We can observe the linear tendency to the lower T_g values with an increase in simulation time, nonetheless, obtained estimates are within the cooling step range.

2. Results with cc_analysis

In this section, we use cc_analysis for dimensionality reduction instead of PCA to identify the glass transition. We apply it with the same input descriptors (**sc-sc**, **bb-bb**, and **bb-sc**). The projections are shown in Figs. 6(a)–6(c) for the three systems for **sc-sc** input descrip-

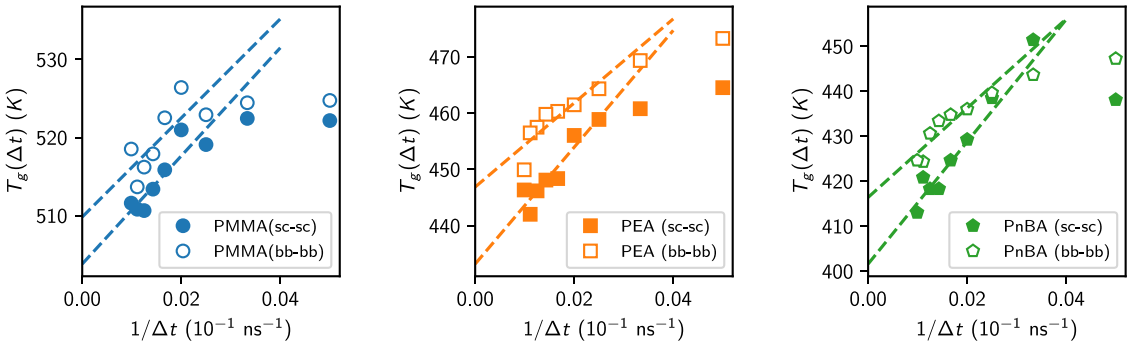


FIG. 5. The inflexion points of the hyperbolic curve fitting at different observation time windows. Extrapolating to $\Delta t \to \infty$, we obtain $T_g \approx 506 \pm 5(K)$ for PMMA, $430 \pm 5(K)$ for PEA, and $402 \pm 2(K)$ for PnBA system for *sc-sc* descriptor. The **bb-bb** descriptor overestimates the T_g predictions.

tor. The state separation results are clearer than the PCA projections [Figs. 3(a)–3(c)]. To quantify the jump in cluster ID, we use the same procedure with Eq. (3) for fitting the data and observe that inflection points, i.e., the glass transition temperatures, decrease on increasing the side chain lengths of the polymer melt. The data are tabulated in Table I. cc_analysis results are less sensitive to the choice of dimensions [Fig. 6(e)] compared with PCA [Fig. 4(b)]. We observe that the backbone descriptor systematically overestimates the

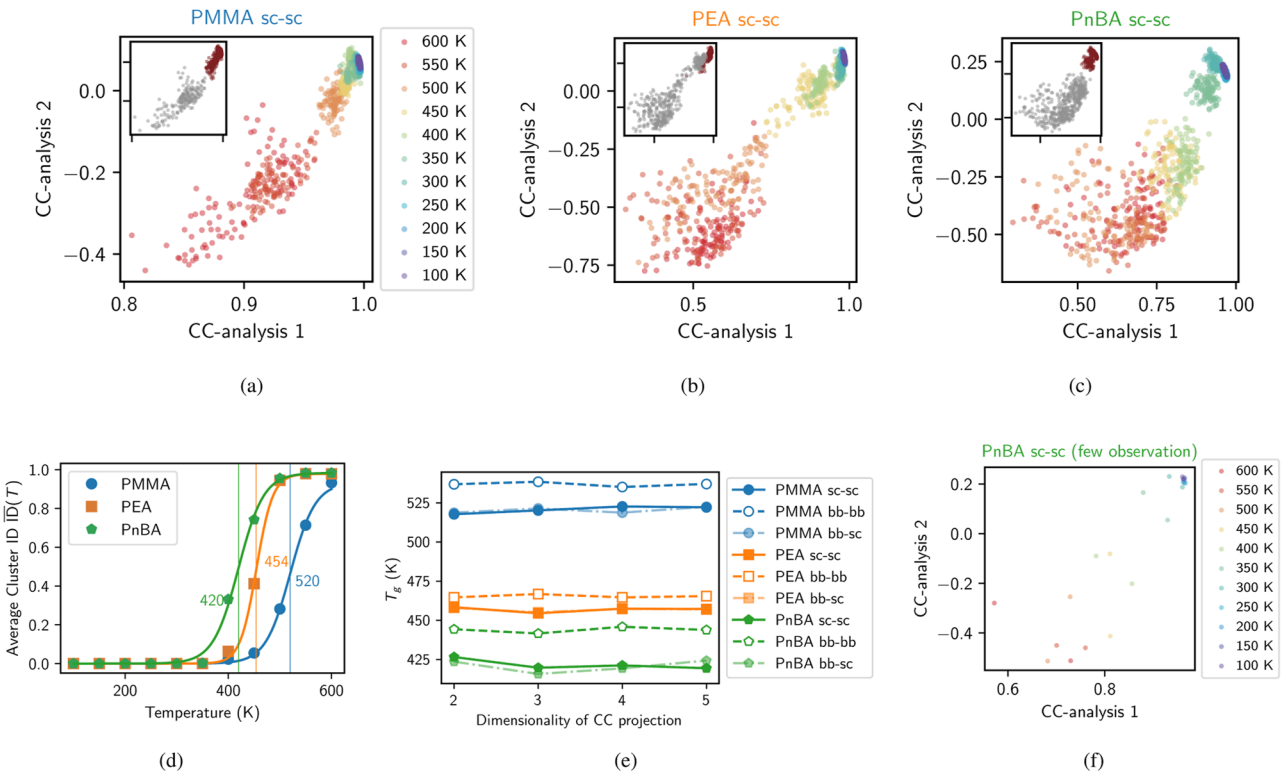


FIG. 6. Projections using cc_analysis of a single chain over multiple time frames (the same as in Fig. 3): (a) PMMA, (b) PEA, and (c) PnBA. Each point in the plot corresponds to the chain's conformation at a given temperature at each time. Projections are colored varying from red to blue from high temperatures to low temperatures. (d) Agglomerative clustering on the cc_analysis projections for the three different polymers. Results are averaged over all chains, and the average cluster index $\overline{\mathrm{ID}(T)}$ is plotted for each temperature. The vertical lines correspond to the determined glass transition temperatures. (e) The glass transition temperatures were calculated for three different polymers as a function of the number of leading cc-components used for clustering for different input descriptors: **bb-bb**, **sc-sc**, and **bb-sc**. (f) Projections of concatenated data from all T for a single chain with only a single frame at each T in the new reduced space determined using cc_analysis.

T_g value. Moreover, it performs equally well with only a few observations per temperature [Fig. 6(f)].

B. Method II: Analysis of individual temperatures

In this section, we use Method II as described in Sec. II B where we perform PCA on single polymer chains at different temperatures independently and use averaged eigenvalues [Eq. (5)] and the participation ratio [Eq. (6)] as described in Sec. II B to determine T_g [see schematic Fig. 2(b)]. With this method, the information on single chain conformations from different temperatures is not accessible to the dimensionality reduction method, rather we investigate whether the data-driven protocol itself can distinguish different temperature inputs. Examples of resulting projections with **sc-sc** input descriptors for three of the systems are shown in Fig. S5 in supplementary material. We observe a change from the completely random distribution of points in the projection at high temperatures to a more "clustered" projection around the glass transition temperature. Similar behavior is also observed for the bead-spring polymer model near T_g.[18] Below T_g, only small random fluctuations dominate the behavior of the chain, and as a result, the projections look scattered. Here, we would like to emphasize that due to the standardization of the data, those fluctuations at different temperatures have the same magnitude, resulting in visually similar projections below and above T_g in Fig. S5.

The magnitude of the PCA eigenvalues can be used to quantify the observed behavior. In general, for independently projected data, this magnitude does not have a uniform value, but in our case, all distances are standardized. As a result, we could average the first eigenvalue, $\overline{\lambda_1}(T)$, across all projections [see Fig. 7(a)]. As more general criteria, we plot the participation ratio, $\overline{\mathrm{PR}}(T)$, for the three systems in Fig. 7(b). We observe a non-monotonic behavior for all the systems on lowering the temperature. We argue that the change in $\overline{\mathrm{PR}}(T)$ [or $\overline{\lambda_1}(T)$] is connected with a change in nature of the fluctuations in the system: from local configurational rearrangements above T_g to only local rearrangements along the chain below T_g. As a result, more dimensions are required for the random motion description below T_g. With this analysis, we also confirm that the point at which the sudden change occurs (i.e., T_g) shifts toward lower values on increasing side chain length (Fig. 7).

Previously, we showed that the results obtained with Method I did not depend on the choice of descriptors, i.e., the atoms chosen for the analysis. However, the backbone diffuses slower than the side chains (Fig. S3); hence, different descriptors can play a role in this

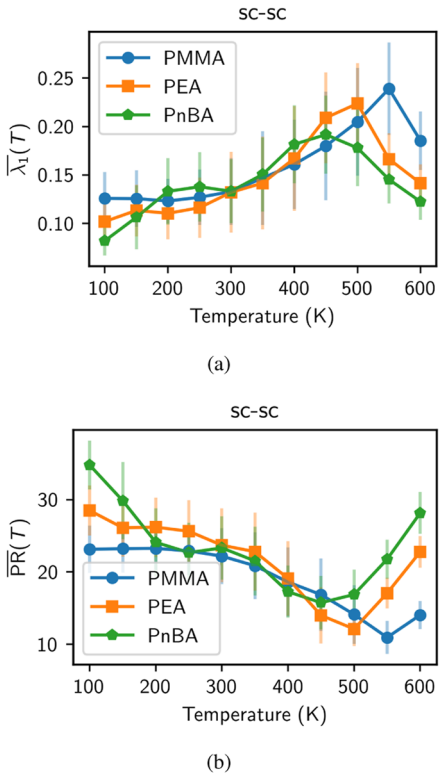


FIG. 7. Analysis of each temperature independently. (a) The magnitude of the first eigenvalues and (b) the participation ratios for PMMA, PEA, and PnBA polymers. We used the side chain distances as input features to PCA, and the results are averaged over all chains. Around T_g, the monotonic behavior of either PR or the first eigenvalue changes, and the temperature at which the change occurs decreases on increasing side chain length.

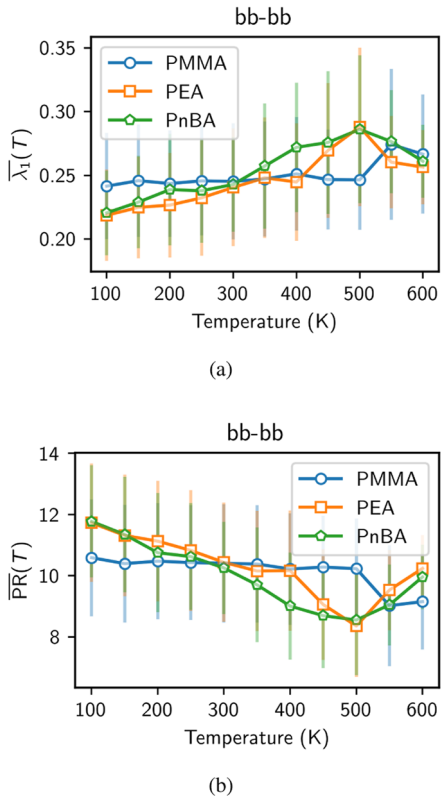


FIG. 8. Analysis of each temperature independently using only the backbone distances as input features to PCA. (a) The magnitude of the first eigenvalues, and (b) the participation ratios for PMMA, PEA, and PnBA polymers. The results are averaged over all chains. The plots suggest that it is essential to include the contributions from side chains to get the correct nature of T_g.

method where each temperature is treated independently. Therefore, we repeat the same analysis using Method II with the **bb-bb** input descriptors for all three systems. Remarkably, the projections do not show "clustering" around T_g (Fig. S6) as well as there are no clear changes in $\overline{\lambda_1}(T)$ and $\overline{\mathrm{PR}}(T)$ (Fig. 8). Thus, our result suggests that the side chains play an important role in determining T_g compared with the backbone contribution. Such difference is more prominent for PnBA system that has the longest side chain. As mentioned in Sec. II B, eigenvalues of cc_analysis and PCA are the same (up to normalization); hence, we show results of Method II only with PCA.

In our combined temperature analysis, we do not see notable differences in T_g values when we compare side chain and backbone contribution. Since fluctuations of all the temperatures are scaled together in the combined temperature analysis, the dominant contribution is expected to come from the high-temperature states. On the contrary, here, each temperature is considered separately, resulting in the explicit role of the side chain and backbone contribution to T_g.

IV. DISCUSSION

Both methods have their advantages and drawbacks that we would like to summarize shortly, but at the same time, they complement each other. In contrast to Method II, Method I is performed on concatenated data from all temperatures; as a result, there are no eigenvalues/eigenvectors associated with a specific temperature.

Method I requires following the same chain over all temperatures, meaning it would not allow us to compare the data of the same systems with independently generated configurations from different simulation runs (as typically done for the glass-forming liquids simulations), whereas Method II does not have this limitation as each chain is analyzed independently at each temperature. One more important consequence of such an independent application is that one can test the simulation results after each cooling step. After observing the non-monotonic behavior in $\overline{\lambda_1}(T)$ or $\overline{\mathrm{PR}}(T)$, no further simulations at lower temperatures are required, in contrast to Method I or the conventional bilinear fitting. However, Method II requires relatively long NVT/NPT simulation runs at each temperature for a certain time and cannot be applied to continuous cooling simulations in contrast to Method I, which in combination with, e.g., cc_analysis, can be used having only one snapshot per temperature.

Note that we performed PCA on a single polymer chain, followed by taking an average over all chains in the system. Performing PCA on 100 chains combined, we only observe the same Gaussian-like distribution within fluctuations, resulting from different chains, which is essentially independent of the temperatures [Fig. S7(a)]. This result is similar to the observation that the radius of gyration (R_g) or end-to-end distance (R_e) distributions over all the chains are independent of temperature[22] [Figs. S7(b) and S7(c)].

The value of T_g from Method II can be defined only within the cooling step range (e.g., between 400 and 450 K for PEA system), whereas Method I allows for more precise prediction as well as extrapolations to the long time limits. To check the influence of the cooling step range, we performed additional simulations with the smaller temperature gap (25 K instead of 50 K) for PnBA system and obtained the same T_g (Fig. S8).

Naturally, both methods are sensitive to the input descriptors (as shown in Fig. 8) as well as parameters [but in this case, we show that the results are robust for a big range of parameters, e.g., Fig. 4(b)]. Compared with the projections shown in Ref. 18, the separation between glassy and liquid states in two-dimensional projection (Fig. 3) is not as clear for the considered system. In the general case, it can be either due to the analysis routine (choice of descriptors, dimensionality of the projection, dimensionality reduction algorithms, and clustering parameters) or due to internal properties of the considered polymers. In the latter case, the choice of the clustering method, which takes the number of clusters as input parameters, might not be optimal and require a deeper understanding of the considered system. Nonetheless, for considered systems, both clustering algorithms with completely different input hyperparameters (agglomerative and DBSCAN) show a sharp change in average cluster indices around T_g (see Fig. S2).

Our analysis shows that the internal degrees of freedom in the polymer chains help to employ the intra-monomer distances within a single polymer chain as an input descriptor. In the future, we plan to check our analysis with different sets of input descriptors (including those that can account for intermolecular interactions explicitly) such as local bond orientation order parameters,[47] softness,[48] dihedral angles,[10] local chemical environment descriptors,[49,50] etc.

V. CONCLUSIONS

In summary, we extend and validate our recently developed data-driven approach to determine the glass transition temperature from all-atom and coarse-grained molecular dynamics simulation data. Our approach utilizes high-resolution microscopic details available from simulations and considers conformational fluctuations of polymer melts over time at the level of individual chains. Here, we apply it to all-atom simulations of acrylic homopolymers of different side chain lengths. Our result qualitatively agrees well with those of other experimental studies where T_g values decrease with the growing length of the polymer side chains. By using different dimensionality reduction methods and clustering algorithms, we show the generality of our approach, which can be applied to a wide variety of systems ranging from coarse-grained polymer models to all-atom systems and, therefore, is robust. Finally, we provide a way to quantify the role of the side chain and backbone in determining the glass transition.

This method could be applied to other systems with "soft" glass transitions such as organic light-emitting diodes where precise prediction of T_g is debatable.

SUPPLEMENTARY MATERIAL

The supplementary material contains plots for determining glass transition temperatures from the bilinear fits with different fitting ranges (Fig. S1); glass transition temperature determined using another clustering method (DBSCAN) (Fig. S2); comparison of the diffusion values for center of mass, backbone atoms, and side chain atoms of the polymer chains (Fig. S3); fitting parameters of $g(T)$ used to calculate T_g shown in Fig. 4(a) (Table S1); explained variance ratio for PCA (Fig. S4); temperature dependent PCA projections of one selected chain with pairwise distances between all side chains

(Fig. S5) and backbone (Fig. S6) atoms; PCA projections of all chains (Fig. S7); and glass transition temperatures obtained with our methods for the simulations with additional cooling steps (Fig. S8).

ACKNOWLEDGMENTS

We acknowledge open-source packages MDTraj,[51] Numpy,[52] Matplotlib,[53] and Scikit-learn[54] used in this work. A.I. and K.K. are supported by the European Union Horizon 2020 APACHE (Active and Intelligent Packaging Materials and Display Cases as a Tool for Preventive Conservation of Cultural Heritage), under Horizon 2020 Research and Innovation Program Grant Agreement No. AMD-814496-10. A.B. thanks Rituparno Mandal for discussion. We acknowledge Sharon Volpe and Denis Andrienko for critical reading of our manuscript.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Atreyee Banerjee: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Methodology (equal); Software (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Aysenur Iscen**: Data curation (lead); Formal analysis (supporting); Visualization (supporting); Writing – original draft (equal); Writing – review & editing (equal). **Kurt Kremer**: Conceptualization (supporting); Supervision (lead); Writing – review & editing (equal). **Oleksandra Kukharenko**: Conceptualization (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

Raw simulation data are available at https://doi.org/10.17617/3.4JHOMW. Derived data supporting the findings of this study and all relevant calculations are available at https://gitlab.mpcdf.mpg.de/banerjeea/acrylic_polymers.

REFERENCES

[1] C. A. Angell, "Perspective on the glass transition," J. Phys. Chem. Solids **49**, 863–871 (1988).
[2] C. A. Angell, "Formation of glasses from liquids and biopolymers," Science **267**, 1924–1935 (1995).
[3] K. Binder, J. Baschnagel, and W. Paul, "Glass transition of polymer melts: Test of theoretical concepts by computer simulation," Prog. Polym. Sci. **28**, 115–172 (2003).
[4] L. Berthier and G. Biroli, "Theoretical perspective on the glass transition and amorphous materials," Rev. Mod. Phys. **83**, 587–645 (2011).
[5] M. J. O'Neill, "Measurement of specific heat functions by differential scanning calorimetry," Anal. Chem. **38**, 1331–1336 (1966).
[6] R. B. Bird, R. C. Armstrong, and O. Hassager, *Dynamics of Polymeric Liquids, Vol. 1: Fluid Mechanics* (John Wiley and Sons Inc., New York, 1987).
[7] R. P. Chartoff, J. D. Menczel, and S. H. Dillman, *Dynamic Mechanical Analysis (DMA)* (Wiley San Jose, 2009), Vol. 387.
[8] M. A. F. Afzal, A. R. Browning, A. Goldberg, M. D. Halls, J. L. Gavartin, T. Morisato, T. F. Hughes, D. J. Giesen, and J. E. Goose, "High-throughput molecular dynamics simulations and validation of thermophysical properties of polymers for various applications," ACS Appl. Polym. Mater. **3**, 620–630 (2020).
[9] K.-H. Lin, L. Paterson, F. May, and D. Andrienko, "Glass transition temperature prediction of disordered molecular solids," npj Comput. Mater. **7**, 179 (2021).
[10] T. Jin, C. W. Coley, and A. Alexander-Katz, "Molecular signatures of the glass transition in polymers," Phys. Rev. E **106**, 014506 (2022).
[11] W. Chu, M. A. Webb, C. Deng, Y. J. Colón, Y. Kambe, S. Krishnan, P. F. Nealey, and J. J. de Pablo, "Understanding ion mobility in P2VP/NMP^{+}I^{-} polymer electrolytes: A combined simulation and experimental study," Macromolecules **53**, 2783–2792 (2020).
[12] C. Deng, M. A. Webb, P. Bennington, D. Sharon, P. F. Nealey, S. N. Patel, and J. J. de Pablo, "Role of molecular architecture on ion transport in ethylene oxide-based polymer electrolytes," Macromolecules **54**, 2266–2276 (2021).
[13] F. Godey, A. Fleury, and A. Soldera, "Local dynamics within the glass transition domain," Sci. Rep. **9**, 9638 (2019).
[14] D. L. Baker, M. Reynolds, R. Masurel, P. D. Olmsted, and J. Mattsson, "Cooperative intramolecular dynamics control the chain-length-dependent glass transition in polymers," Phys. Rev. X **12**, 021047 (2022).
[15] N. Iwaoka and H. Takano, "Conformational fluctuations of polymers in a melt associated with glass transition," J. Phys. Soc. Jpn. **86**, 035002 (2017).
[16] Y. Shimizu, T. Kurokawa, H. Arai, and H. Washizu, "Higher-order structure of polymer melt described by persistent homology," Sci. Rep. **11**, 2274 (2021).
[17] H. Abdi and L. J. Williams, "Principal component analysis," Wiley Interdiscip. Rev.: Comput. Stat. **2**, 433–459 (2010).
[18] A. Banerjee, H.-P. Hsu, K. Kremer, and O. Kukharenko, "Data-driven identification and analysis of the glass transition in polymer melts," ACS Macro Lett. **12**, 679–684 (2023).
[19] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proceedings of the Second International Conference on KDDM, KDD'96* (AAAI Press, 1996), pp. 226–231.
[20] H.-P. Hsu and K. Kremer, "A coarse-grained polymer model for studying the glass transition," J. Chem. Phys. **150**, 091101 (2019).
[21] F. Nielsen, "Hierarchical clustering," in *Introduction to HPC with MPI for Data Science* (Springer International Publishing, Cham, 2016), pp. 195–211.
[22] A. Iscen, N. C. Forero-Martinez, O. Valsson, and K. Kremer, "Acrylic paints: An atomistic view of polymer structure and effects of environmental pollutants," J. Phys. Chem. B **125**, 10854–10865 (2021).
[23] K. Ute, N. Miyatake, and K. Hatada, "Glass transition temperature and melting temperature of uniform isotactic and syndiotactic poly(methyl methacrylate)s from 13mer to 50mer," Polymer **36**, 1415–1419 (1995).
[24] A. F Behbahani, S. M. Vaez Allaei, G. H Motlagh, H. Eslami, and V. A. Harmandaris, "Structure and dynamics of stereo-regular poly(methyl-methacrylate) melts through atomistic molecular dynamics simulations," Soft Matter **14**, 1449–1464 (2018).
[25] A. Buzin, M. Pyda, P. Costanzo, K. Matyjaszewski, and B. Wunderlich, "Calorimetric study of block-copolymers of poly(*n*-butyl acrylate) and gradient poly(*n*-butyl acrylate-co-methyl methacrylate)," Polymer **43**, 5563–5569 (2002).
[26] H. Zhang, K. Lamnawar, and A. Maazouz, "Rheological modeling of the diffusion process and the interphase of symmetrical bilayers based on PVDF and PMMA with varying molecular weights," Rheol. Acta **51**, 691–711 (2012).
[27] S. Krause, J. J. Gormley, N. Roman, J. A. Shetter, and W. H. Watanabe, "Glass temperatures of some acrylic polymers," J. Polym. Sci., Part A: Gen. Pap. **3**, 3573–3586 (1965).
[28] L. Andreozzi, V. Castelvetro, M. Faetti, M. Giordano, and F. Zulli, "Rheological and thermal properties of narrow distribution poly(ethyl acrylate)s," Macromolecules **39**, 1880–1889 (2006).
[29] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field," J. Comput. Chem. **25**, 1157 (2004).
[30] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen, "GROMACS: A message-passing parallel molecular dynamics implementation," Comput. Phys. Commun. **91**, 43–56 (1995).

[31]M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," SoftwareX **1–2**, 19–25 (2015).
[32]D. Case, I. Ben-Shalom, S. Brozell, D. Cerutti, T. Cheatham III, V. Cruzeiro, T. Darden, R. Duke, D. Ghoreishi, M. Gilson, H. Gohlke, A. Goetz, D. Greene, R. Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R. Walker, J. Wang, H. Wei, R. Wolf, X. Wu, L. Xiao, D. M. York, and P. Kollman, AMBER, 2018.
[33]L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez, "PACKMOL: A package for building initial configurations for molecular dynamics simulations," J. Comput. Chem. **30**, 2157–2164 (2009).
[34]L. Wang, "Discovering phase transitions with unsupervised learning," Phys. Rev. B **94**, 195105 (2016).
[35]C. Wang and H. Zhai, "Machine learning of frustrated classical spin models. I. Principal component analysis," Phys. Rev. B **96**, 144432 (2017).
[36]L.-W. Yang, E. Eyal, I. Bahar, and A. Kitao, "Principal component analysis of native ensembles of biomolecular structures (PCA_NEST): Insights into functional dynamics," Bioinformatics **25**, 2147 (2009).
[37]A. E. Cohen and W. Moerner, "Principal-components analysis of shape fluctuations of single DNA molecules," Proc. Natl. Acad. Sci. U. S. A. **104**, 12622–12627 (2007).
[38]K. Diederichs, "Dissecting random and systematic differences between noisy composite data sets," Acta Crystallogr., Sect. D: Struct. Biol. **73**, 286–293 (2017).
[39]See https://strucbio.biologie.uni-konstanz.de/xdswiki/index.php/Cc_analysis for the cc_analysis code was used for the analysis.
[40]K. Su, O. Mayans, K. Diederichs, and J. R. Fleming, "Pairwise sequence similarity mapping with PaSiMap: Reclassification of immunoglobulin domains from titin as case study," Comput. Struct. Biotechnol. J. **20**, 5409–5419 (2022); bioRxiv:2022.05.13.491469 (2022).
[41]S. Hunkler, K. Diederichs, O. Kukharenko, and C. Peter, "Fast conformational clustering of extensive molecular dynamics simulation data," J. Chem. Phys. **158**, 144109 (2023).
[42]G. Young and A. S. Householder, "Discussion of a set of points in terms of their mutual distances," Psychometrika **3**, 19–22 (1938).
[43]K. Pearson and F. Galton, "VII. Note on regression and inheritance in the case of two parents," Proc. R. Soc. London **58**, 240–242 (1895).
[44]J. H. Ward, Jr., "Hierarchical grouping to optimize an objective function," J. Am. Stat. Assoc. **58**, 236–244 (1963).
[45]C. F. Jekel and G. Venter, "pwlf: A Python library for fitting 1D continuous piecewise linear functions," https://github.com/cjekel/piecewise_linear_fit_py (2019).
[46]A. Soldera and N. Metatla, "Glass transition of polymers: Atomistic simulation versus experiments," Phys. Rev. E **74**, 061803 (2006).
[47]E. Boattini, S. Marín-Aguilar, S. Mitra, G. Foffi, F. Smallenburg, and L. Filion, "Autonomously revealing hidden local structures in supercooled liquids," Nat. Commun. **11**, 5479 (2020).
[48]S. S. Schoenholz, E. D. Cubuk, E. Kaxiras, and A. J. Liu, "Relationship between local structure and relaxation in out-of-equilibrium glassy systems," Proc. Natl. Acad. Sci. U. S. A. **114**, 263–267 (2016).
[49]C. Caruso, A. Cardellini, M. Crippa, D. Rapetti, and G. M. Pavan, J. Chem. Phys. **158**, 214302 (2023).
[50]C. Caruso, A. Cardellini, M. Crippa, D. Rapetti, and G. M. Pavan, "Detecting dynamic domains and local fluctuations in complex molecular systems via time-lapse neighbors shuffling," Proc. Natl. Acad. Sci. U. S. A. **120**(30), e2300565120 (2023).
[51]R. T. McGibbon, K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernández, C. R. Schwantes, L.-P. Wang, T. J. Lane, and V. S. Pande, "MDTraj: A modern open library for the analysis of molecular dynamics trajectories," Biophys. J. **109**, 1528–1532 (2015).
[52]C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with NumPy," Nature **585**, 357–362 (2020).
[53]J. D. Hunter, "Matplotlib: A 2D graphics environment," Comput. Sci. Eng. **9**, 90–95 (2007).
[54]F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," J. Mach. Learn. Res. **12**, 2825–2830 (2011); arXiv:1201.0490 (2012).