Modified Group Contribution Scheme to Predict the Glass-Transition Temperature of Homopolymers through a Limiting Property Dataset

Yang Yang, Xiong Zou,* Haotian Ye, Weixuan Zhu, Hongguang Dong,* and Mingshu Bi

ABSTRACT: Previous studies on glass-transition temperature ($T_g$) prediction mainly focus on developing diverse methods with higher regression accuracy, but very little attention has been paid to the dataset. Generally, a large range of $T_g$ values of a specified polymer could be found in the literature but which one should be selected into a dataset merely depends on the implicit preference rather than a recognized and clear criterion. In this paper, limiting glass-transition temperature ($T_g(\infty)$), a constant value obtained at the infinite number-average molecular weight $M_w$, was validated to be an adequate bridge index in the $T_g$ prediction models. Furthermore, a new dataset containing 198 polymers was established to predict $T_g(\infty)$ using the improved group contribution method and it showed a good correlation ($R^2 = 0.9925$, adjusted $R^2 = 0.9894$). The method could also generate $T_g-M_w$ curves by introducing the $T_g(\infty)$ function and provide more information to polymer scientists and engineers for material selection, product design, and synthesis.

1. INTRODUCTION

Glass-transition temperature ($T_g$) marks the temperature at which an amorphous polymer is cooled from the rubbery to the glassy state. As rheological, mechanical, and dielectric properties change dramatically when the operating temperature passes through $T_g$, $T_g$ plays a vital role in a wide range of both scientific and industrial processes. To explore the correlation between $T_g$ and the chain structure of the polymer, several attempts have been made by experiments,1 molecular dynamics simulation,3,4 and quantitative structure–property relationship.5 Undoubtedly, a reliable and quick property prediction of polymers that are not even synthesized is desirable as it saves time and resources for basic research and industrial product development. In the $T_g$ prediction literature, two types of methods are well-recognized and being developed all the time: theoretical methods and semi-empirical methods. Theoretical methods are based on molecular descriptors of polymers, and many kinds of methods/tools have been introduced to predict $T_g$ with high accuracy, such as connectivity indices,5 CODESSA program,6 and an approach based on considering the polymer repeating unit as a set of anharmonic oscillators.7 In the last few years, neural network-based quantitative structure–property relationship (QSPR) methods are also in vogue,8–11 these methods perform better when many factors influence the targets with large amounts of data.

Among them, the effect of the polymer structure on $T_g$ has been studied systematically by Bicerano5 and Askadskii.7 Bicerano’s approach is based on the principal descriptors of the polymer repeat unit structure, where bond indices are introduced to correlate with $T_g$. The method can predict $T_g$ of polymers containing nine atomic elements C, H, O, N, F, Si, S, Cl, and Br. Furthermore, Askadski’s approach considers a repeat unit structure as a set of anharmonic oscillators that describe the thermal motion of atoms in the range of intra- and intermolecular forces, where atomic constants and empirical parameters which are independent of the polymer chemical structure are introduced in the calculation. The number of atomic elements contained in the predicted polymer extends to fifteen: C, H, O, N, F, Si, S, Cl, Br, I, P, B, As, Sn, and Pb. Subsequently, this method is applied for the prediction of $T_g$ of polymer–solvent systems.12

The most widely referenced semi-empirical method is the group contribution method due to its simplicity and reasonable accuracy. Weyland et al.13 predicted the glass-transition temperature of polymers using weighted additive contribution for their constitutional structure groups, and this method was then improved by integrating the connectivity index method14 or neural network.15 Over the last decades, the group contribution method has been applied for the prediction of

Received: September 14, 2020
Accepted: October 28, 2020
Published: November 9, 2020
polymer melts, polymer diluents, copolymers, and polymer networks, and group contribution-based computer-aided molecular design (CAMD) have been developing and have become a powerful approach for polymer product design. The group contribution method in the CAMD approach was used to screen simple polymer structures, and then used in the field of drug delivery and rubber polymers, which shows its broad application prospect.

Most contributions above focused on method innovation with various descriptors, models, and algorithms, but the dataset selection and validation has received scant attention. Actually, the value of $T_g$ varies over a wide range for a specified polymer in the literature. There are two main reasons: First, different from normal molecules, $T_g$ is highly structure-dependent. Many factors besides the repeat unit structure are confirmed to influence $T_g$: molecular weight distribution, cross-linking, long side chain, chain stiffness, and even the interactions between polymer chains. Second, $T_g$ is more difficult to measure compared with the normal boiling point or melting point because it is also known to heavily depend on both method selection and operation in the test. The cooling/heating rate and measuring device selection will also affect the result even though the measuring method is fixed. So the discrepancy of the reported data points for a specified polymer can also be very high.

### Table 1. $T_g$ Values of Various Polymers Gathered in the Different Sources

| polymer name          | Van Krevelen value (K) | Katritzky value (K) | Bicerano value (K) | Askadskii value (K) | PoLyInfo range (K) | number of points |
|-----------------------|------------------------|---------------------|--------------------|---------------------|-------------------|------------------|
| poly(1-butene)        | 249                    | 228                 | 213                | 250                 | 150−298           | 32               |
| poly(4-methyl-1-pentene) | 303                   | 302                 | 246                | 272                 | 287−423           | 72               |
| polyisobutylene       | 200                    | 199                 | 184                | 214                 | 184−358           | 60               |
| polystyrene           | 373                    | 373                 | 381                | 376                 | 174−502           | 1590             |
| poly(4-methylstyrene) | 380                    | 374                 | 373                | 400                 | 355−395           | 36               |
| poly(α-methylstyrene) | 441                    | 409                 | 410                | 401                 | 228−463           | 106              |
| poly(vinyl alcohol)   | 372                    | 358                 | 333                | 365                 | 303−466           | 104              |
| poly(acrylic acid)    | 379                    | 379                 | 349                | 374                 | 321−443           | 57               |
| poly(methyl acrylate) | 279                    | 281                 | 308                | 282                 | 270−316           | 67               |
| poly(ethyl acrylate)  | 249                    | 251                 | 284                | 267                 | 230−323           | 83               |
| poly(butyl methacrylate) | 293                  | 293                 | 295                | 287                 | 249−356           | 105              |
| poly(methylene oxide) | 190                    | 218                 | 213                | 197                 | 183−343           | 48               |
| poly(ethylene oxide)  | 206                    | 206                 | 205                | 204                 | 157−368           | 54               |

**Figure 1.** $T_g$–$M_n$ relationships (a−f) are gathered in the literature and database, the ID of each polymer in the legend can be found in the PoLyInfo database (https://polymer.nims.go.jp/PoLyInfo/cgi-bin/p-search.cgi).
the one-to-many mapping nature and the implicit selection preference, inconsistency and conflict inevitably often exist in different datasets, which makes the prediction results seem unconvincing. Therefore, it is necessary to select a more proper property index to predict \( T_g \) based on a reliable dataset with a reasonable number of consistent data points.

Although several theories and models exist for discussing the affecting factors of polymer \( T_g \) such as pressure,\(^{4,6}\) cross-linking,\(^{27}\) and crystallinity.\(^{28}\) Molecular weight distribution still has the most critical influence on property prediction and synthesis/production processes.\(^{29}\) It is now well-established from a variety of studies that the glass-transition temperature of the polymer increases with the number-average molecular weight \((M_n)\), which has been proposed by Fox and Flory\(^{30}\)

\[
T_g = T_g(\infty) - \frac{K_g}{M_n}
\]

where \( T_g(\infty) \) is the limiting value of \( T_g \) for an infinitely high molecular weight. \( K_g \) is the empirically determined constant. eq 1 reflects that when the molecular weight reaches an infinite value, \( T_g \) tends to be stabilized and independent of molecular weight distribution. Thus compared with plenty of \( T_g \) values in the previous data extraction strategy, \( T_g(\infty) \) is consistent and unique, which can be used in the dataset as a more convincing referenced property index. Besides, consider the relationship presented in eq 1, \( T_g \) of a specified polymer at different number-average molecular weights could also be predicted, which means that more valuable information is provided for decision-making in scientific research and industrial development.

In this work, the limiting value \( T_g(\infty) \) is adopted as the bridge index in the \( T_g \) prediction models by establishing one-to-one mapping essentially between the polymer repeat unit structure and properties. A dataset that contains 198 polymers is established through a standard selection criterion, detailed data extraction, and treatment procedures. For the purpose of the demonstration, new regression with this proposed dataset is performed by the group contribution method. Finally, the comparison with other methods and \( T_g=\bar{M}_w \) curves are also discussed.

2. RESULTS AND DISCUSSION

2.1. Determining \( T_g(\infty) \) to be the Referenced Value in Prediction. Previous scholars have tried to gather data from the literature to find relationships between \( T_g \) and \( M_n \).\(^{5,31}\) To further demonstrate the applicability of the relationship in eq 1, we searched \( T_g \) values with \( M_n \) information in the database\(^{32-40}\) and literature.\(^{41-43}\) Seventy-two polymers were found with sufficient data points, in which 32 polymers and regression results are shown in Figure 1 (all polymers are shown in Table S1 in the Supporting Information). Although more polymers have been discovered over the correlation curve between \( T_g \) and \( M_n \),\(^{11-41}\) our results show that \( T_g \) increases asymptotically toward a constant limiting value with increasing \( M_n \) in the overwhelming majority of the cases with sufficient data points. Even though \( T_g(\infty) \) has not been proved rigorously, \( T_g(\infty) \) could still be an adequate index used in the preliminary property prediction for polymer scientists and engineers.

In addition to the empirical evidence, the theoretical basis of eq 1 comes from free volume theory.\(^{30}\) The reason why \( T_g \) is found to increase asymptotically with molecular weight can be explained by the impact of end groups or chain entangle-ment.\(^{44}\) End groups exist in the two ends of the polymer chain and introduce additional free volumes, which leads to a decrease in \( T_g \).\(^{31}\) When the molecular weight is low, the effect of end groups is significant, the type of chain end also influences \( T_g \).\(^{45}\) But as the molecular weight increases, the proportion of end groups decreases. And when the molecular weight increases to a certain extent, the proportion can be neglected, \( T_g \) approaches \( T_g(\infty) \). Experiments on polystyrene\(^{46}\) and poly(methyl methacrylate)\(^{46}\) indicated that the type of chain-end has little effect on \( T_g(\infty) \), whether the polymer is linear or cyclic. In previous studies, the group contribution method was used to predict the common \( T_g \) value, and only repeat unit structures were used as molecular descriptors. However, as the effect of the end groups is eliminated at an infinite molecular weight, this method is validated to be more suitable for \( T_g(\infty) \) prediction.

2.2. Dataset Generation. The next most important thing is to allocate a reliable \( T_g(\infty) \) value for each candidate polymer in the dataset. Different from the direct measurement data \( T_g \), \( T_g(\infty) \) requires additional processing of a series of raw \( T_g \) data obtained in a wide range of \( M_n \). On the one hand, searching for the reported \( T_g \) through mountains of papers is a cumbersome and tedious task. On the other hand, information quality and multidimensionality (reporting \( T_g, M_n \), test method, and parameters simultaneously) should be checked to ensure that these generated \( T_g(\infty) \) values are trustworthy. Taking all factors into account, only homopolymers without any postprocessing are studied in this work.

2.2.1. Data Sources. The author holds that scholarly journal articles with rigorous peer review are the most reliable sources. But direct search in Google Scholar and/or SciFinder may sound a bit like searching a needle-in-a-haystack problem. As our guide, the following three sources are adopted in the preanalysis stage before literature retrieval and reading: Bicerano,\(^5\) Katritzky et al.,\(^6\) and PoLyInfo.\(^{25}\) Bicerano gathered \( T_g-M_n \), relationships of 35 polymers and Katritzky reported 88 un-cross-linked homopolymers with high molecular weights. PoLyInfo contains over 16 000 homopolymers with polymer structure, synthesis, and measurement information, and it is the foundation of our searching, sorting, and data analytics. It should be noted that the verification of the original literature for all data is the essential step.

2.2.2. Screening Criteria. According to the definition of the concept, \( T_g(\infty) \) could be generated by two ways: regression and approximation. The former is calculated strictly by calculation and has been considered authors’ favorite, while the latter adds more polymers empirically to meet the requirements of group contribution methods on the number and kind of data points. The information on \( T_g \) and molecular weight has been checked to ensure that any outliers cannot affect the performance of our prediction model. And all of the selected polymers are linear homopolymers, which eliminates the influence of cross-linking. The influence of the measuring method cannot be ignored, and we also record this message in Table S1 in the Supporting Information.

2.2.2.1. Regression-Set 1 and Set 2. For polymers which has more than five number-average molecular weight information data points in our collection, regression based on eq 1 was carried out to determine the relationship and find the limiting values. If \( R^2 \) is higher than 0.8, we assign it to Set 1 (the most convincing data group), while Set 2 will receive the regression values with \( R^2 \) located between 0.5 and 0.8. These values are listed in the “Regression value” column of Table S1.
Only 25 polymers reported by Bicerano\textsuperscript{5} were selected into our dataset because some of them are commercial products which are copolymers or homopolymers with long end groups (like Fomblin series and Denum) and named “Bicerano value” in Table S1. The data of polyethylene were also not selected because it comes from oligomers.

2.2.2.2. Approximation-Set 3. The approximation is performed depending on the characteristics: $T_g(\infty)$ should be higher than the common $T_g$ value, and the values of $T_g$ for polymers with high $M_n$ are very little different from $T_g(\infty)$. However, so far, no unanimous conclusion can be drawn in the literature as to which number could be regarded as a high number. Through a quantitative analysis of 45 polymers in Set 1, the results in Figure 2 indicated that $T_g$ will reach/exceed 0.95, 0.98, and 0.99 times $T_g(\infty)$ with $M_n = 20$ kg/mol in over 80, 53.3, and 31.1% of the cases, respectively. Moreover, the values of $T_g$ for polymer with $M_n = 50$ kg/mol are quite similar to $T_g(\infty)$ (deviation < 2%) in 80% cases. Thus, in scenarios requiring low precision, the value of $T_g$ (or $T_g$/margin factor, 0.95, 0.98, and 0.99 according to the $M_n$) for a polymer with a high $M_n$ could be regarded as an acceptable approximation value for $T_g(\infty)$. Detailed information about this analysis can be found in Table S2.

For polymers that have sufficient data points (more than five points but not always have $M_n$ information) in the database, we will attempt to find the values that are close to the limiting values. If the number of data points for a specified polymer is more than ten, a histogram generated by PoLyInfo was also provided. As the limiting values are higher than experimental values for the given polymers, the maximum values of the data points in the collection would be considered as the candidate limiting values before the $M_n$ check. However, to keep the data clean, the maximum points which are far higher than other points and discontinuous in the same polymer were eliminated. This is because this value may result from the measuring method, pre- or postreatment, or just experimental error. These screened values are listed in the “PoLyInfo value” column of Table S1.

To further expand the dataset, data points from Katritzky with high molecular weights ($M_n > 20$ kg/mol) were also introduced into our dataset. Due to the large and strong electron-withdrawing properties of CF\textsubscript{3}, as well as the simplification of the side chain in our prediction, polymers with CF\textsubscript{3} in the long side chain were eliminated. As a result, 85 polymers were selected, and these values are listed in the “Katritzky value” column of Table S1. All of these polymers were allocated into Set 3, and detailed screening procedures for each polymer in this dataset are also provided in the Supporting Information Part 2.

2.3. Dataset Description and Statistics. This version of the dataset consists of 198 homopolymers and is further divided into three sets with various precisions. As shown in Figure 3, 16 polymer classes including polyolefins, polysty-
$R^2_{\text{adjusted}} = 1 - \frac{(1 - R^2)(n - 1)}{(n - k - 1)}$  

(2)

where $n$ is the number of data points used in the regression and $k$ is the number of polymer groups. The adjusted $R^2$ in this work is 0.9894, this value is slightly lower than the value of $R^2$, proving the appropriate use of the polymer groups.

MRE is used to describe the average relative error, which is the difference between the experimental data $X_{\text{exp}}$ and predicted data $X_{\text{pre}}$ of property $X$ for the given polymers. The equation can be written as follows

$$\text{MRE} (\%) = \frac{1}{N} \times \left( \sum \left| \frac{X_{\text{exp}} - X_{\text{pre}}}{X_{\text{exp}}} \right| \right) \times 100$$  

(3)

Where $N$ is the number of data used in the regression.

As mentioned above because the $T_g$ values in the literature come from different measuring methods under different conditions, although the influence of the molecular weight on $T_g$ is eliminated using limiting values, the error still exists. Different from the normal boiling or melting point, $T_g$ is measured by indirect methods; there are great differences using different measuring methods and duration of experiments, so this error cannot be simply considered as measurement error. Under this circumstance, it is suitable to use $R^2$ to evaluate the prediction performance rather than MRE or root-mean-square error in other literature.

Table 2 shows the prediction performance of various methods in the literature. Katritzky’s method, Bicerano’s method, and Askadskii’s method are all theoretical methods which are based on molecular descriptors of polymer, the $R^2$ of Katritzky’s method is 0.946, and then improved by Bicerano and Askadskii with 0.9749 and 0.998, respectively. Afantitis’s method and Palomba’s method are completed through the neural network based on Katritzky’s method, increasing the $R^2$ to 0.9269 and 0.953, respectively. Gani’s method and our method in this paper are both based on the group contribution method, Gani et al. reported $R^2$ of 0.9374 and 0.948 for improvements achieved by high order groups and connectivity index. An $R^2$ of 0.9925 in our method indicates that the proposed model fits the used dataset very well. It is worth noting that the $R^2$ value is not suitable for direct comparison strictly due to the different datasets used in different studies. The higher $R^2$ value in our paper is not devoted to demonstrate the superiority of the group contribution method. The basic group contribution is a widely used, simple method, which has been proved in the literature that it is not the best performance model. However, as our work has paid attention to the dataset used for prediction, we want to demonstrate that $T_g(\infty)$ is better than $T_g$ used in the model. To better illustrate our idea, we chose this most traditional method rather than a more sophisticated one in this paper. Table 2 indicates that the simple method could also achieve sufficiently good performance using the proper index and dataset. Obviously, other more advanced QSPR methods can further improve the accuracy based on the proposed limiting property dataset.

MRE can be used to reflect the accuracy of different polymer classes in the dataset. Total MRE of $T_g(\infty)$ is 8.09%, there are 96 polymers out of 194, whose MRE is less than 5%, 87 polymers have the MRE ranging from 5 to 20%, and the MRE of 15 polymers is greater than 20%. Regression performance statistics based on the main types of polymers for $T_g(\infty)$ are listed in Figure 5. As shown in the figure, the MRE values for polylefiners are much higher than those of the other polymers. This phenomenon may be caused due to the following reasons: First, we only define the number of backbones in the side chain rather than the detailed side chain structure. This simplification could improve the generalization ability of the proposed model because of fewer independent variables, but the MRE value will increase for the polymers with complicated side chain structures. Second, the literature reported more varied and complicated side chain structures for polylefiners, at least in the

![Figure 4. Parity plot of limiting molar glass-transition function, $Y_g(\infty)$.](image)

![Figure 5. Data points and MRE of each class of polymer in the dataset.](image)
proposed dataset, the polyolefins sub-set has more data points with long and branching side chain structures than other subsets. Thus, the side chain simplification will lead to a higher MRE for polyolefins. Third, main chain and side chain structures do not have an equal effect on $T_g$. Van Krevelen’s work proves that side chain structures have more influence on $T_g$ of polyolefins than that of other polymers. In principle, the more complicated the side chain structures, the more influential the side chain on the main chain, the more likely the simplification will affect.

Compared with the previous $T_g$ prediction method, $T_g$ with $M_n$ information can also be predicted in this work. Figure 6 displays the previously predicted values of $T_g$ of four polymers, and $T_g(\infty)/T_g-M_n$ curves predicted using our method. As shown, these predicted curves show good agreements with experimental values extracted from the database and the fitting result using the Fox–Flory equation (eq 1). These results also indicate that predicting $T_g$ without molecular weight information (we put these values in the $T_g$ axis) may puzzle the user with a large deviation between the experimental and predicted value. Thus predicting $T_g-M_n$ curves can provide practical guidelines for the follow-up polymer synthesis and help achieve the target $T_g$ value with fewer experiments.

3. CONCLUSIONS

Because of the large range of glass-transition temperature $T_g(\infty)$, the single constant value for each polymer was introduced in the $T_g$ prediction model, and a new dataset that consists of 198 polymers through a detailed selection procedure was established. To build the structure–property relationship, the group contribution method was adopted, where polymers with long side chains can be predicted using the combination of the basic group and number of backbones in the side chain. After this, 58 structural groups were used in the regression, and a good correlation was achieved ($R^2 = 0.9925$, adjusted $R^2 = 0.9894$, MRE = 8.09%), which was found to be higher than the previous work. Finally, $T_g-M_n$ curves could also be generated by the proposed method, which provides an additional dimension of average molecular weight information.

At present, it is still not easy to find molecular weight information in the literature, which leads to difficulty in determining $T_g(\infty)$ and $T_g-M_n$ relationship for more polymers. The authors indicate that multi-dimensional experimental data can not only convincingly support the researchers’ own conclusions but also lay a foundation for follow-up data mining and knowledge discovery. Besides, the proposed methodology could also be extended to other molecular weight-dependent properties such as melting temperature, surface tension, or heat capacity, as well as other structural parameters such as degree of crystallinity, cross-linking, and branching.

Figure 6. $T_g$ experimental values of (a) poly(N-vinylcarbazole), (b) poly(N-vinylpyrrolidone), (c) poly(vinyl chloride), and (d) poly(4-vinylpyridine) from the database and predicted values obtained from refs 5 and 6. Comparison of the fitting experimental values using Fox–Flory equation represented by the solid blue curve and predicting $T_g-M_n$ curves in this work using eqs 6 and 9, represented by the solid red curve. $T_g(\infty)$ predicted value is represented by a red star and dashed line.
4. COMPUTATIONAL METHODS

4.1. Group Contribution Method. Several methods currently exist for the prediction of $T_g$ among these, the group contribution method is simple and frequently-used by polymer chemists and engineers. To demonstrate the effectiveness of the proposed methodology and dataset in a practical way, the group contribution method proposed by Van Krevelen is introduced to predict $T_g(\infty)$ with slight modifications. The general form of this method for predicting $T_g$ can be written as follows

$$f(T_g) = \sum_i N_i Y_g$$

(4)

where $Y_g$ is the glass-transition temperature contribution of the $i$ group. However, in this work, to predict the limiting value of glass-transition temperature, eq 4 should be modified

$$f(T'_g(\infty)) = Y'_g(\infty) = \sum_i N_i Y'_g$$

(5)

where $Y'_g(\infty)$ is the limiting glass-transition temperature contribution of the $i$ group, and $N_i$ is the occurrence number of the $i$ group. Similarly, $Y'_g$ and $N'_i$ are the contribution and occurrence number of backbone in the side chain. $Y_g(\infty)$ is defined as the limiting molar glass-transition function.

The following issue is about the form of the left side of eq 5. For eq 4, both Satyanarayana and Van Krevelen have discussed the left function, and the difference between them is whether the additional adjustable parameter should be considered in the function (in the Satyanarayana method, $f(T_g) = T_g M_w - Y_g(\infty)$ in the Van Krevelen method, $f(T_g) = T_g M_w^*$, where $M_w$ is the formula weight of the repeat unit. For polypropylene, $M_w = 42$ g/mol). The Satyanarayana method is meaningful in polymers where $Y_g$ can be regarded as the influence of end groups, especially those with low molecular weights. However, as end groups can be varied, the value of the additional adjustable parameter $Y_g$ is not a constant. In the Van Krevelen method, end groups are thought to play a minor role in the whole polymer chain, so the repeat unit structure is used to express additive properties. In our work, we chose not to use $Y_g(\infty)$ although it will help improve the accuracy because for limiting the glass-transition temperature, the influence of the end groups is eliminated, as proved above. Then the group contribution method for estimating $T_g(\infty)$ can be written as

$$f(T'_g(\infty)) = Y'_g(\infty) = T_g(\infty) \times M_w = \sum_i N_i Y'_g(\infty)$$

(6)

$$+ N'_i Y'_g$$

4.2. Structure Description and Group Definition Criteria. Both the main chain and side chain exist in the polymer repeat unit structure, as shown in Figure 7. As the effect of the side chain on polymer properties is different from that of the main chain, it is unwise to regard them as equal when defining groups for group contribution.

For the main chain structure, the definition of groups is referred to Van Krevelen’s work, which prefers bivalent and composed groups. In this paper, 36 of 57 main chain groups are identical to those in Van Krevelen’s work, which are marked in Table S3. As some polymer repeat unit structures in our dataset cannot be fully described using these groups, 18 new groups are created based on Van Krevelen’s rules. A long chain may consist mainly of bivalent groups. 2. It is better to regard a composed unit as one structural group. Besides, there are still 3 special main chain groups in the group set, which are discussed later in this section.

The effect of structural groups on $T_g(\infty)$ is different in the polymer main chains and side chains, and many authors have different strategies for defining the side chain. When the side chain is small, the influence on $T_g(\infty)$ can be eliminated by combining it into the main chain. When the side chain is large, treatments are demonstrated to define the side chain: a long side chain group contains two parts, the basic group, which is from the main chain group set, and the side chain. The number of backbone atom number in the side chain $N_b$ is introduced to describe the length of the side chain, and hydrogen atoms are not calculated into $N_b$. The schematic diagram is shown in Figure 8. When meeting groups with a long side chain, their contributions can be calculated using the combinations of the main chain and side chain.

Figure 7. Example of the main chain group and side chain group that belong to the repeat unit structure of poly[oxycarbonothioyloxy-1,4-phenylene(isobutylmethylmethylene)-1,4-phenylene], ID in PoLyInfo: P150086.

Figure 8. Illustration of actual and basic groups.

In particular, for polyoxides or polyethers, the value of contribution from the same groups may be different because of their neighboring atoms or groups. Considering the repeat unit structure of polyethers, several special groups are introduced: $-O-(\text{end})$, $-O-(\text{oxide})$, and $-(\text{CH}_2)_n-(\text{oxide})$. Polyether usually exists with an end oxygen group in the repeat unit structure, as shown in Figure 9a. When calculating $T_g(\infty)$, this kind of oxygen group $-O-(\text{end})$ has a unique contribution value. Specifically, for polyoxides, which consist of only an oxygen group and several saturated hydrocarbon groups.
4.3. \( T_g-M_n \) Relationship. Many correlations have been proposed to describe the relationship between \( T_g \) and \( M_n \) through calculation and simulation. It should be noted that Askadskii deduced a more rigorous \( T_g-M_n \) correlation (as shown in eq 7) based on Lin’s theory considering the effect of flow temperature \( T_f \) and mechanical segment \( M_s \).

\[
\frac{T_g}{T_f} = 1 + \frac{1}{26} \ln \left( \frac{M_s}{M} \right)^{1.5} \left( \left[ \left( \frac{2.4}{M_s/M} \right)^{0.5} - 1 \right] + \frac{1}{2} \right)
\]

Among these eq 1 is simple and widely used. However, \( K_g \) is a constant which is also related to the polymer structure, eq 1 can be applied only to the polymer where \( K_g \) of this polymer has been reported or calculated. For a polymer, for which little \( T_g \) is known, data from several studies suggest that because of chain stiffness, \( K_g \) of high \( T_g(\infty) \) is higher than that of low \( T_g(\infty) \), and \( K_g \) is proportional to a power of \( T_g(\infty) \).

Bicerano gathered data from the literature and analyzed statistically. The relationship is shown in eq 8

\[
K_g = 0.002715 T_g(\infty)^3
\]

Substituting eq 8 into eq 1, we can obtain at the desired expression eq 9. Once the values of \( T_g(\infty) \) and \( M_n \) are obtained, \( T_g \) can be predicted using the equation above, and this is why we choose this equation in this work.

\[
T_g = T_g(\infty) - 0.002715 \times \frac{T_g(\infty)^3}{M_n}
\]

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.0c04499. Dataset containing the limiting values and resources of 198 polymers, \( T_g(\infty) \) and \( K_g \) values of polymers in Set 1 and their \( M_n \) values when the \( T_g \) reaches 95% \( T_g(\infty) \), 98% \( T_g(\infty) \), and 99% \( T_g(\infty) \), 58 structural groups and their formula weights and \( T_g(\infty) \) contributions and the comparison between the original and predicted \( T_g(\infty) \) for polymers are presented in the Supporting Information Part I. Detailed \( T_g(\infty) \) determination for each polymer is presented in the Supporting Information Part II (PDF).

AUTHOR INFORMATION

Corresponding Authors

Xiong Zou — School of Chemical Engineering and School of Energy and Power Engineering, Dalian University of Technology, Dalian, Liaoning 116024, China; Email: zouxiong@dlut.edu.cn

Hongguang Dong — School of Chemical Engineering, Dalian University of Technology, Dalian, Liaoning 116024, China; orcid.org/0000-0001-6544-1526; Email: hgdong@dlut.edu.cn

Authors

Yang Yang — School of Chemical Engineering, Dalian University of Technology, Dalian, Liaoning 116024, China

Haotian Ye — School of Chemical Engineering, Dalian University of Technology, Dalian, Liaoning 116024, China

Weixuan Zhu — School of Chemical Engineering, Dalian University of Technology, Dalian, Liaoning 116024, China

Mingshu Bi — School of Chemical Engineering, Dalian University of Technology, Dalian, Liaoning 116024, China

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.0c04499

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support from the National Natural Science Foundation of China, under Grant No. 21276039.

REFERENCES

(1) Shen, J.; Caydamli, Y.; Gurarslan, A.; Li, S.; Tonelli, A. E. The glass transition temperatures of amorphous linear aliphatic polyesters. Polymer 2017, 124, 235-245.

(2) Mohammadi, M.; fæzi, H.; karevan, M.; Davoodi, J. The glass transition temperature of PMMA: A molecular dynamics study and comparison of various determination methods. Eur. Polym. J. 2017, 91, 121-133.

(3) McKechnie, D.; Cree, J.; Wadkin-Snaith, D.; Johnston, K. Glass transition temperature of a polymer thin film: Statistical and fitting uncertainties. Polymer 2020, 195, No. 122433.

(4) Chen, M.; Jabeen, F.; Rasulev, B.; Oscowski, M.; Boudjouk, P. A computational structure-property relationship study of glass transition temperatures for a diverse set of polymers. J. Polym. Sci., Part B: Polym. Phys. 2018, 56, 877-885.

(5) Bicerano, J. Prediction of Polymer Properties; 3rd ed.; Marcel Dekker: New York, 2002.

(6) Katritzky, A. R.; Sild, S.; Lobanov, V.; Karelson, M. Quantitative structure-property relationship (QSPR) correlation of glass transition temperatures of high molecular weight polymers. J. Chem. Inf. Comput. Sci. 1998, 38, 300-304.

(7) Askadskii, A. A. Computational Materials Science of Polymers; Cambridge International Science Publishing: Cambridge, 2003.

(8) Almntis, A.; Melagraki, G.; Makridina, K.; Alexandridis, A.; Sarimevis, H.; Igelisi-Markopoulou, O. Prediction of high weight polymers glass transition temperature using RBF neural networks. J. Mol. Struct.: THEOCHEM 2005, 716, 193–198.

(9) Palomba, D.; Vazquez, G. E.; Diaz, M. F. Novel descriptors from main and side chains of high-molecular-weight polymers applied to prediction of glass transition temperatures. J. Mol. Graphics Modell. 2012, 38, 137–147.

(10) Miccio, L. A.; Schwartz, G. A. Localizing and quantifying the intra-monomer contributions to the glass transition temperature using artificial neural networks. Polymer 2020, 203, No. 122786.

(11) Miccio, L. A.; Schwartz, G. A. From chemical structure to quantitative polymer properties prediction through convolutional neural networks. Polymer 2020, 193, No. 122341.

(12) Matseevich, T. A.; Goleneva, L. M.; Afanas’ev, E. S.; Kurskaya, E. A.; Kovriga, O. V.; Askadskii, A. A. The dependence of glass transition temperature and flow temperature on solvent or plasticiser concentration. Int. Polym. Sci. Technol. 2017, 44, 17–22.

(13) Weyland, H. G.; Hoftyzer, P. J.; Van Krevelen, D. W. Prediction of the glass transition temperature of polymers. Polymer 1970, 11, 79–87.
weight dependence of chain structure on the glass transition temperature. Part 3. Molecular

1988

ACS Omega http://pubs.acs.org/journal/acsodf

Amsterdam, 2009.

and Prediction from Additive Group Contributions

Database (PoLyInfo). 2020, https://polymer.nims.go.jp/PoLyInfo/

Fluid Phase Equilib. modeling using grid technology for design of structured products.

polymers and diluents.

Ind. J. Intell. Syst. 2013, 23, 89−104.

Camacho-Zuniga, C.; Ruiz-Treviño, F. A. A new group contribution scheme to estimate the glass transition temperature for polymers and diluents. Ind. Eng. Chem. Res. 2003, 42, 1530−1534.

Hamedi, M.; Danner, R. P.; Duda, J. L. A lattice-fluid, group-contribution treatment of the glass transition of homopolymers, copolymers, and polymer solutions. J. Appl. Polym. Sci. 2003, 89, 697−705.

Preparation of dimethacrylate polymer networks by a group contribution approach. Int. J. Polym. Anal. Charact. 2013, 18, 93−104.

Optical computer-aided molecular design: A polymer design case study. Ind. Eng. Chem. Res. 1996, 35, 3403−3414.

Pavurala, N.; Achenie, L. E. K. Identifying polymer structures for oral drug delivery - A molecular design approach. Comput. Chem. Eng. 2014, 71, 734−744.

Liang, X.; Zhang, X.; Zhang, L.; Liu, L.; Du, J.; Zhu, X.; Ng, K. M. Computer-aided polymer design: Integrating group contribution and molecular dynamics. Ind. Eng. Chem. Res. 2019, 58, 15542−15552.

Schneider, H. A. Polymer class specificity of the glass transition temperature. Polymer 2005, 46, 2230−2237.

Van Krevelen, D. W.; Te Nijenhuis, K. Properties of Polymers: Their Correlation with Chemical Structure; Their Numerical Estimation and Prediction from Additive Group Contributions; 4th ed.; Elsevier: Amsterdam, 2009.

National Institute for Materials Science (NIMS) Polymer Database (PoLyInfo). 2020, https://polymer.nims.go.jp/PoLyInfo/cgi-bin/p-search.cgi.

Ransom, T. C.; Oliver, W. F. Glass transition temperature and density scaling in cumene at very high pressure. Phys. Rev. Lett. 2017, 119, 052502.

Chiou, B. S.; Schoen, P. E. Effects of crosslinking on thermal and mechanical properties of polyurethanes. J. Appl. Polym. Sci. 2002, 83, 213−223.

Askadskii, A.; Popova, M.; Matseveich, T.; Kurskaya, E. The influence of the degree of crystallinity on the glass transition temperature of polymers. Adv. Mater. Res. 2013, 864−867, 751−754.

Chen, X.; Shao, Z.; Gu, X.; Feng, L.; Biegler, L. T. Process intensification of polymerization processes with embedded molecular weight distribution models: An advanced optimization approach. Ind. Eng. Chem. Res. 2019, 58, 6113−6145.

Fox, T. G.; Flory, P. J. Second-order transition temperatures of low molecular weight polystyrene by initiator fragments located at chain ends. Macromolecules 2016, 49, 2387−2398.

Danusso, F.; Levi, M.; Gianotti, G.; Turri, S. End unit effect on the glass transition temperature of low-molecular-weight polymers and copolymers. Polymer 1993, 34, 3687−3693.

Zhang, L.; Torkelson, J. M. Enhanced glass transition temperature of low molecular weight poly(methyl methacrylate) by initiator fragments located at chain ends. Polymer 2017, 122, 194−199.

Askadskii, A. A.; Matseveich, T. A.; Markov, V. A. Determination of glass-transition temperatures of polymers: A modified computational scheme. Polym. Sci. Ser. A 2016, 58, 506−516.

Fox, T. G.; Loshak, S. Influence of molecular weight and degree of crosslinking on the specific volume and glass temperature of polymers. J. Polym. Sci. 1955, 15, 371−390.

Dibenedetto, A. T.; Dilandro, L. Correlation of glass transition temperature and molecular weight: A model based on the principle of corresponding states. J. Polym. Sci., Part B: Polym. Phys. 1989, 27, 1405−1417.

Xia, W.; Mishra, S.; Keten, S. Substrate vs. free surface: Competing effects on the glass transition of polymer thin films. Polymer 2013, 54, 5942−5951.

Wu, C. Multiscalar modeling of glass transition in polymeric films: Application to stereoregular poly(methyl methacrylate). Polymer 2018, 146, 91−100.

Lin, Y. H. Linear viscoelasticity and translational diffusion of linear flexible polymers. Macromolecules 1991, 24, 5346−5350.

Lu, X.; Jiang, B. Glass transition temperature and molecular parameters of polymer. Polymer 1991, 32, 471−478.

Hintermeyer, J.; Herrmann, A.; Kahlau, R.; Goicouanu, C.; Rössler, E. A. Molecular weight dependence of glassy dynamics in linear polymers revisited. Macromolecules 2008, 41, 9335−9344.

Larrin, R.; Tagle, L. H.; Diaz, F. R. Glass transition temperature-molecular weight relation for poly(hexamethylene perchloroterephthalamide). Polym. Bull. 1981, 4, 487−490.

Malhotra, S. L.; Lessard, P.; Blanchard, L. P. The thermal decomposition and glass transition temperature of poly(p-terbutylisoprene). J. Macromol. Sci. Part A - Chem. 1981, 15, 121−141.

Malhotra, S. L.; Lessard, P.; Blanchard, L. P. The thermal decomposition and glass transition temperature of poly-p-methoxystyrene. J. Macromol. Sci. Part A: Chem. 1981, 15, 301−321.

Montserrat, S.; Colomer, P. The effect of the molecular weight on the glass transition temperature in amorphous poly(ethylene terephthalate). Polym. Bull. 1984, 12, 173−180.

Faucher, J. A.; Koleske, J. V.; Santee, E. R.; Stratta, J. J.; Wilson, C. W. Glass transitions of ethylene oxide polymers. J. Appl. Phys. 1966, 37, 3962−3964.

Faucher, J. A. The dependence of glass transition temperature on molecular weight for poly(propylene oxide) and poly(butylen oxide). J. Polym. Sci., Part B: Polym. Lett. 1965, 3, 143−145.

Gao, L.; Oh, J.; Yu, T.; Chang, T.; Li, C. Y. Glass transition temperature of cyclic polystyrene and the linear counterpart contamination effect. Polymer 2019, 170, 198−203.

Zhang, L.; Marsiglio, J. A.; Lan, T.; Torkelson, J. M. Dramatic tunability of the glass transition temperature and fragility of low molecular weight polystyrene by initiator fragments located at chain ends. Macromolecules 2016, 49, 130, No. 025702.

Fox, T. G.; Flory, P. J. Second-order transition temperatures of polymers: A modified computational scheme. Polym. Sci. Ser. A 2016, 58, 506−516.