



Encoding alternatives for the prediction of polyacrylates glass transition temperature by quantitative structure–property relationships



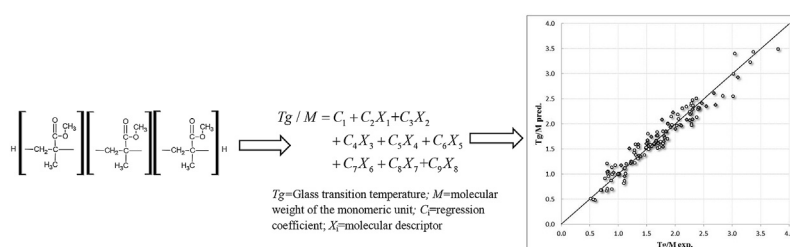
Andrew G. Mercader*, Pablo R. Duchowicz

Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas (INIFTA, UNLP, CCT La Plata-CONICET), Diag. 113 y 64, Sucursal 4, C.C. 16, 1900 La Plata, Argentina

HIGHLIGHTS

- A QSPR model of the glass transition temperature of polyacrylates was build.
- The polymers structure encoding alternatives for this type of studies were explored.
- The validation procedures revealed very good predictive attributes by the model.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 20 July 2015

Received in revised form

3 December 2015

Accepted 23 January 2016

Available online 4 February 2016

Keywords:

Polymers

Glass transitions

Computer modeling and simulation

Computational techniques

ABSTRACT

The glass transition temperature, T_g , is one of the most important properties of amorphous polymers. The ability to predict the T_g value of a polymer prior to its synthesis is of great value. For this reason we performed a predictive Quantitative Structure–Property Relationships (QSPR) analysis of T_g . The study explored the best way to encode the polymers structure for this type of studies, finding that the optimal option is using three monomeric units. The best linear model constructed from 126 molecular structures incorporated eight molecular descriptors and showed very good predictive ability, being a very simple and straight forward method for the prediction of T_g for polyacrylates since three dimensional descriptors were not used.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

The glass transition temperature, T_g , also known as the glass temperature or the glass–rubber transition temperature, is one of the most important properties of amorphous polymers [1].

As the temperature of a polymer drops below T_g , it behaves in an increasingly brittle manner. As the temperature rises above the T_g ,

the polymer becomes a rubber-like material. Thus, knowledge of T_g is essential in the selection of materials for various applications. In general, values of T_g well below room temperature define the domain of elastomers and values above room temperature define rigid structural polymers [1].

In the vicinity of T_g , a polymer experiences a sudden increase in the rate of molecular motions and, as a result, undergoes a series of conformational transformations. The torsional oscillations and/or rotations about most of the backbone bonds are activated, which causes a sharp increase in the free volume of the system as it is converted from the initial rigid (glassy) state to quasi-liquid state

* Corresponding author.

E-mail address: amercader@inifta.unlp.edu.ar (A.G. Mercader).

[2]. As a result of these processes, many physical properties of polymers change dramatically; for example, their coefficients of thermal expansion, heat capacities, and viscosities. The T_g is difficult to determine experimentally and predict theoretically because the transition takes place over a comparatively wide temperature range and is dependent on conditions such as the method of measurement, duration of the experiment, and pressure during the measurement [3,4]. The T_g is also very dependent on the structural (cross-linking, chain stiffness) [5], constitutional (additives, fillers, impurities) [6], and conformational (tacticity) features of polymers [1,4,7]. For these reasons, the discrepancies between reported values of T_g in the literature can be quite high [8].

Numerous researchers have attempted to predict T_g for polymers on the basis of Quantitative Structure Property Relationships (QSPR). According to the view of Katritzky et al., there are two kinds of approaches, the empirical and the theoretical [8]. Empirical methods correlate the studied property with other physical or chemical properties of the polymers, for instance, group additive properties (GAP) [1]. The GAP methodology is an entirely empirical approach, restricted to systems made merely of functional groups that have previously been investigated. It is an approximate method, since it fails to account the presence of neighboring groups or conformational influences. The most extensively referenced model made from theoretical estimations was proposed by Bicerano [4]; this regression model ($R = 0.9749$, $s = 24.65$ K) related the T_g with the solubility and the weighted sum of 13 structural parameters for a data set of 320 polymers; however the model was not tested on an external set, hence its validation was not assured.

Katritzky et al. [9] develop a mode with R^2 of 0.928 using 22 medium molecular weight polymers consisting of four parameters. Not presenting details about the way structures were encoded, only mentioning that tree to five monomeric units were used. Later on, Katritzky et al. [8] used CODESSA to predict the T_g for 88 linear homopolymers using five parameters and generated a QSPR model with a standard error of 32.9 K for T_g . In this case three monomeric units were used but no analysis was done to determine this number. In both these works, no external test sets were used; hence the models were not properly validated. Cao and Lin [10] tested the same set of 88 polymers using five parameters with clear physical meanings, calculated from individual repeating unit structures, finding a model with coefficient of determination of $R^2 = 0.9056$ and a standard error of 20.86 K. Once more, the model was not properly validated by an external test set.

Mattioni and Jurs [11] developed a 10-descriptor model using the structure of the monomer of 165 polymers, to predict T_g values using Artificial Neural Networks, the training set rms error was 10.1 K ($R^2 = 0.98$) and a prediction set (17 polymers) rms error of 21.7 K ($R^2 = 0.92$). In addition, an 11-descriptor model using one repeating unit from 251 different polymers, in this case, the training set rms error was 21.1 K ($R^2 = 0.96$) and a prediction set (25 polymers) rms error 21.9 K ($R^2 = 0.96$). Although the size of the prediction set is rather small, the results indicate that the use of the repeating unit instead of the monomer structure has a better predictive ability. In this article no further trials were done to attempt to determine the best number of repeating units to encode the structures.

A comprehensive neural network model with 28 descriptors was developed by Chen et al. [12] to predict T_g values of 6 randomly selected polymers from a database containing 71 polymers. The network was trained with the remaining 65 polymers, using descriptors calculated from individual repeating unit structures, and had training root mean square error of 17 K ($R^2 = 0.95$) and prediction average error of 17 K ($R^2 = 0.85$). Arriving at a presumably good model, however the number of test set polymers seems

excessively low and the descriptors used excessively high, hence the predictivity of the model is uncertain.

A Support Vector Machine-based QSPR for the Prediction of Glass Transition Temperatures using 77 polymers was done by Yu [2]. Finding a model with root mean square (rms) errors for the training (38 polymers), validation (18 polymers) and prediction set (21 polymers) of 12.13, 15.58, and 16.22 K, respectively. Polymers were represented by one repeating unit end-capped by two hydrogen atoms, to calculate molecular descriptors.

An artificial neural network prediction of glass transition temperature using 113 polymers was done by Liu et al. [13], the final optimum neural network with produced a training set root mean square error (RMSE) of 11 K ($R = 0.973$) and a prediction set RMSE of 17 K ($R = 0.955$). To calculate the descriptors, the polymers were represented by their corresponding monomer.

As can be appreciated, none of the previous studies have evaluated the optimal number of monomeric units to represent the polymer structure in the prediction of T_g .

Recently a study using flexible descriptors successfully modeled a different property, the refractive index, using 234 structurally diverse polymers [14]. In this case the best found alternative was to encode the polymers with two repeating units.

In the case of polymer studies, it is not possible to calculate the molecular descriptors directly from the entire structure, since polymers possess very high molecular weights; moreover the size of the molecular chains may vary from different polymer preparations. Hence, the way to encode the molecules becomes a crucial part of a QSPR study involving polymers.

For that reason, the main objective of the present work is to study the best way to encode polymers in QSPR studies, in order to obtain reliable predictions based on a straight forward method. In order to do so, a dataset consisting of 126 polyacrylates was selected. Only polyacrylates were included in this study aiming to have a structurally similar set, and consequently producing more precise models.

2. Methods

2.1. Data sets

To carry out this study, a total of 126 polyacrylates with experimental T_g were taken from a published compilation [15], to our knowledge this set of molecules was not employed in this type of study before. Only the polyacrylates family was chosen aiming to produce a more specific and precise study. The experimental T_g values along with the SMILES structure representation can be found on Table S1. SMILES notation was chosen as a way of sharing the dataset with any interested reader, since it allows easily copying the text string and entering it in many chemical structure representation software. The data-set was divided into a training set of 84 and a test set of 42 polymers by applying a k-means cluster analysis [16], in order to have representative molecules of the structure diversity of the complete dataset in both training and test sets.

Following the procedure done by of Katritzky et al. [8] where T_g was divided by the molecular weight of the repeating unit (M), and after some preliminary tests that showed that using T_g/M presented better correlation results than using directly T_g ; it was decided to use T_g/M for the study.

The experimental measure of T_g is a difficult task, which is revealed in the dispersion of experimental data for some polymers, complicating the correlation studies since they rely on the quality of the experimental dataset. When more than one value was found for the same polymer an average was used.

Download English Version:

<https://daneshyari.com/en/article/1520820>

Download Persian Version:

<https://daneshyari.com/article/1520820>

[Daneshyari.com](https://daneshyari.com)