15 (2022) 100106

Contents lists available at ScienceDirect

Journal of Non-Crystalline Solids: X

journal homepage: www.sciencedirect.com/journal/journal-of-non-crystalline-solids-x

Estimating glass transition temperature and related dynamics of molecular glass formers combining artificial neural networks and disordered systems theory

Claudia Borredon [a], Luis A. Miccio [a,b,c], Anh D. Phan [d,e], Gustavo A. Schwartz [a,b,*]

[a] Centro de Física de Materiales (CSIC-UPV/EHU), Materials Physics Center (MPC), P. M. de Lardizábal 5, 20018 San Sebastián, Spain
[b] Donostia International Physics Center, P. M. de Lardizábal 4, 20018 San Sebastián, Spain
[c] Institute of Materials Science and Technology (INTEMA), National Research Council (CONICET), Colón 10850, 7600 Mar del Plata, Buenos Aires, Argentina
[d] Faculty of Materials Science and Engineering, Phenikaa University, Hanoi 12116, Viet Nam
[e] Phenikaa Institute for Advanced Study (PIAS), Phenikaa University, Hanoi 12116, Viet Nam

ARTICLE INFO

Keywords:
QSPR
Properties prediction
Artificial neural networks
ECNLE

ABSTRACT

Glass transition temperature and related dynamics play an essential role in amorphous materials research since many of their properties and functionalities depend on molecular mobility. However, the temperature dependence of the structural relaxation time for a given glass former is only experimentally accessible after synthesizing it, implying a time-consuming and costly process. In this work, we propose combining artificial neural networks and disordered systems theory to estimate the glass transition temperature and the temperature dependence of the main relaxation time based on the knowledge of the molecule's chemical structure. This approach provides a way to assess the dynamics of molecular glass formers, with reasonable accuracy, even before synthesizing them. We expect this methodology to boost industrial development, save time and resources, and accelerate the scientific understanding of structure-properties relationships.

1. Introduction

Quantitative structure-property relationships (QSPR) models can boost both materials design and scientific understanding of molecular glass formers. They can correlate the molecular structure with important properties like glass transition temperature and its related dynamics, which are among the most significant issues associated with the behaviour of glass formers. Many challenging problems, especially in the pharmaceutical industry, like the tendency to recrystallize, water solubility and dissolution rate, or the long-term stability [1–4], are related to the structural relaxation dynamics and the glass transition temperature. This molecular relaxation process is usually described by a characteristic relaxation time and its temperature dependence, which can be experimentally measured using broadband dielectric spectroscopy (BDS), dynamic light scattering (DLS), or dynamic mechanical analysis (DMA), among other techniques. However, when designing new molecular glass formers, we do not know their dynamics before synthesizing and characterizing them, which are costly and time-consuming processes. In this sense, QSPR models are able to estimate the desired properties based only on the chemical structure of the molecules.

Some theoretical approaches can help in these challenging tasks. For instance, the elastically collective nonlinear Langevin equation (ECNLE) theory has been recently used to successfully describe the temperature dependence of the relaxation times of different amorphous materials [5]. However, this approach requires the knowledge of the glass transition temperature (T_g) to estimate the molecular dynamics. Although this information is not available for new glass formers until synthesized, recent developments based on artificial neural networks (ANN) allow estimating their glass transition temperature based only on their chemical structure [6–8], without involving any experimental measurements or complex synthesis.

This work proposes a joint theoretical and numerical approach to estimate the glass transition temperature and the temperature dependence of the structural relaxation time for several molecular glass formers (including amorphous drugs and biomolecules), based only on their chemical structure. A neural network approach is firstly used to

* Correspondence to: G. A. Schwartz, Centro de Física de Materiales (CSIC-UPV/EHU), Materials Physics Center (MPC), P. M. de Lardizábal 5, 20018 San Sebastián, Spain.
E-mail address: gustavo.schwartz@csic.es (G.A. Schwartz).

https://doi.org/10.1016/j.nocx.2022.100106
Received 17 March 2022; Received in revised form 1 June 2022; Accepted 11 June 2022
Available online 17 June 2022

estimate the glass transition temperature of "new" compounds from their chemical structure codified into a Simplified Molecular Input Line Entry System (SMILES) representation; then, this information is used in ECNLE theory to estimate the temperature dependence of the relaxation time for the structural relaxation process. In this way, we can estimate the dynamics of molecular glass formers, even before synthesizing them, only knowing their chemical structure.

2. Theoretical background

ECNLE theory describes glass-forming liquids using a hard-sphere fluid [1,5,9–15]. Key characteristics of the fluid are the particle size, d, and the number of particles per volume, ρ, from where the volume fraction is estimated as $\Phi = \rho\pi d^3/6$. Two main factors affecting the mobility of a tagged particle are 1) interactions with its nearest neighbours and 2) cooperative motions of particles beyond the first shell. The local dynamics (or the motion of the tagged particle within a particle cage) is quantified by the dynamic free energy [1,5,9–15], $F_{dyn}(r) = F_{ideal}(r) + F_{caging}(r)$, where r is the displacement. $F_{ideal}(r)$ corresponds to the delocalized or ideal fluid state and $F_{caging}(r)$ characterizes the localized state of the particle via caging forces, which strongly depends on the density and structure of systems.

Fig. 1 shows an example for calculations of $F_{dyn}(r)$ at $\Phi = 0.58$ and indicates physical quantities of the local dynamics. In a sufficiently dense fluid, a reduction of the free volume dynamically restricts the motion of particles and forms a particle cage surrounding a tagged particle. The dynamical constraint is characterized by the emergence of a barrier in $F_{dyn}(r)$. A particle cage radius, r_{cage}, is roughly estimated by the first minimum of the radial distribution function, $g(r)$. Other important length scales of the local dynamics are a localization length, r_L, a barrier position, r_B, a jump distance, $\Delta r = r_B - r_L$, and a local barrier, $F_B = F_{dyn}(r_B) - F_{dyn}(r_L)$. From these, we can calculate $K_0 = \frac{\partial^2 F_{dyn}(r)}{\partial r^2}\big|_{r=r_L}$ and $K_B = \frac{\partial^2 F_{dyn}(r)}{\partial r^2}\big|_{r=r_B}$ corresponding to harmonic curvatures at r_L and r_B, respectively. K_0 can be interpreted as a spring constant at the localization length.

Escaping of a particle from its cage requires reorganization of both the nearest neighbours and all particles outside the cage to generate the extra space. Thus, the collective motions are strongly coupled to the local dynamics within the cage. The cooperative rearrangement creates a displacement field, $u(r)$, from the surface of the particle cage, that triggers particles beyond the first coordination shell to vibrate as oscillators and radially propagates through the rest medium. By employing Lifshitz's linear continuum mechanics [16], one can analytically calculate the displacement field for $r \geq r_{cage}$ as $u(r) = \frac{\Delta r_{eff} r_{cage}^2}{r^2}$, where Δr_{eff} is the amplitude of the field, whose mathematical expression was reported elsewhere [14,15]. Since $u(r)$ is small, the oscillation of each particle is approximately harmonic and, thus, the elastic energy of an oscillator is $K_0\frac{u^2(r)}{2}$. From this, we quantify collective motion effects on the relaxation process by summing the harmonic elastic energy of particles outside the cage to obtain the collective elastic barrier, which is $F_e = 4\pi\rho\int_{r_{cage}}^{\infty} dr r^2 g(r) K_0\frac{u^2(r)}{2}$.

Inserting the local and elastic components into Kramer's theory gives us the structural relaxation time

$$\frac{\tau_\alpha}{\tau_s} = 1 + \frac{2\pi}{\sqrt{K_0 K_B}}\frac{k_B T}{d^2} exp\left(\frac{F_B + F_e}{k_B T}\right) \quad (1)$$

where τ_s is a short time scale and its analytical form was previously reported [1,5,9–15]. Note that the above calculations provide $\tau_\alpha(\Phi)$. Direct comparisons between theory and experiments need a density-to-temperature conversion. In prior works [1,9–12], based on a thermal expansion process, it was proposed a simple thermal mapping $T = T_g + (\Phi_g - \Phi)/\beta\Phi_0$, where T_g is the dynamic glass transition temperature defined by $\tau_\alpha(T_g) = 100$s, Φ_g is the volume fraction when $\tau_\alpha(\Phi_g \approx 0.6157) = 100$s, $\Phi_0 \approx 0.50$ is a characteristic volume fraction, and $\beta \approx 12 \times 10^{-4}\text{K}^{-1}$ is an effective thermal expansion coefficient considered constant for all amorphous materials [1,9–12].

Fig. 2 shows the theoretical and experimental temperature dependence of structural relaxation time for griseofulvin, nordazepam, celecoxib, tetrazepam, and ibuprofen. Overall, theoretical results quantitatively agree with experimental data over a wide temperature range or timescale without any fitting parameter.

The slight deviation observed for ibuprofen (solid line) is related to the fact that we assume that local and collective dynamics correlate to each other in a universal manner for all materials. F_B and F_e in Eq. (1) are summed with the ratio of prefactor equal to 1. In prior works [12,17], ECNLE calculations were improved by multiplying the elastic barrier with an adjustable parameter a to change the relative importance of collective dynamics in the glass transition. The new adjusted elastic barrier is $F_e \rightarrow a^2 F_e$ and it modifies the structural relaxation time in Eq. (1) as

$$\frac{\tau_\alpha}{\tau_s} = 1 + \frac{2\pi}{\sqrt{K_0 K_B}}\frac{k_B T}{d^2} exp\left(\frac{F_B + a^2 F_e}{k_B T}\right) \quad (2)$$

The value of Φ_g and the thermal mapping are strongly dependent on

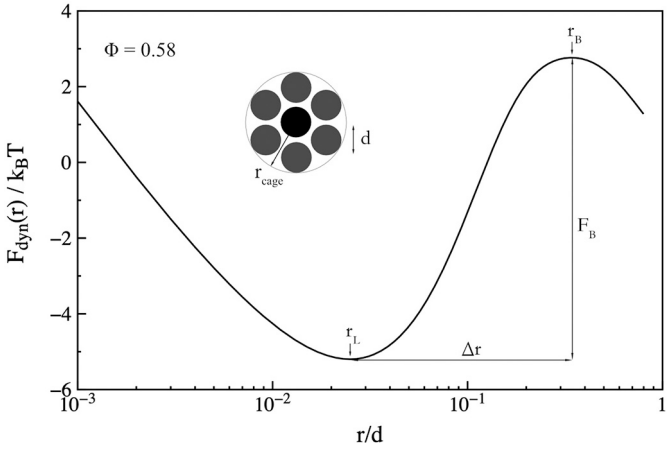

Fig. 1. The free energy profile normalized by k_BT for a hard-sphere fluid with $\Phi = 0.58$, where k_B is the Boltzmann constant and T is the temperature. Characteristic length and energy scales for the local dynamics are defined.

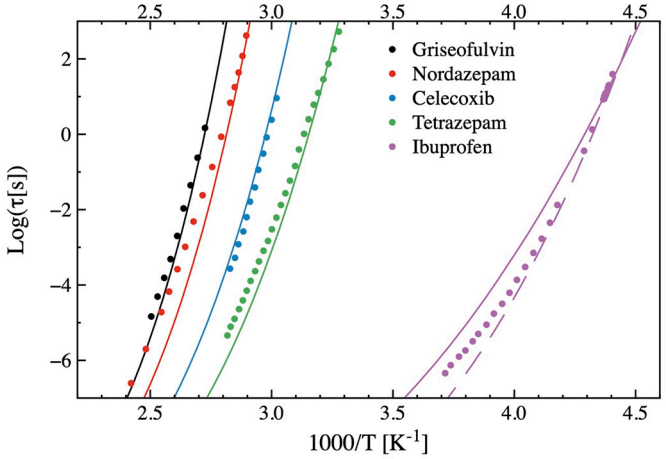

Fig. 2. Temperature dependence of the structural relaxation time calculated using the ECNLE theory (solid lines) and the corresponding experimental values (dots) for several molecular glass formers. The dashed line represents the ECNLE prediction for ibuprofen with the adjustable parameter $a = 2.5$ taken from Fig. 3.

the parameter a, which accounts for the non-universal effects of biological, conformational, and chemical complexities on the collective motions of molecules. It was empirically observed that the scaling parameter (a) typically increases with increasing fragility and depends on the glass transition temperature, as shown in Fig. 3. Although the correlation is not strong, there is a clear trend indicating an increment of the parameter a upon decreasing glass transition temperature. The estimated value of the parameter a for ibuprofen is 2.5 (being the experimentally observed value 2.4). The dashed line in Fig. 2 shows the ECNLE prediction for ibuprofen using this value for the scaling parameter.

3. Numerical background

This section describes the dataset, the encoding of the chemical structures, and the corresponding T_g prediction using ANN.

3.1. Dataset

The dataset for this work is composed of 216 molecules, including pharmaceutical drugs (like benzocaine or ibuprofen), biological molecules (like sucrose or ribose), and typical additives used in the pharmaceutical industry (like benzophenone), with T_g in the range 200 K – 400 K. We accounted for the chemical and spatial structure of the compounds by using the Simplified Molecular Input Line Entry System (SMILES) [18,19], which codifies the molecules into a string of characters. Table 1 in the Supporting Information (SI) shows the name of the compounds, their corresponding SMILES code, and the experimental and estimated T_g values.

3.2. Data treatment (encoding)

Following the same approach reported in previous works [6–8], we used a one-hot encoding method and an appropriate dictionary with all the existing characters in the SMILES code to convert the SMILES strings (1D) into binary matrices (2D). Thus, row i-th of the matrix is filled with zeros except for the position of the dictionary that coincides with the same character on the i-th position in the SMILES code. A one is placed in this case. Therefore, the number of characters in the dictionary (nd) and the length of each SMILES code (npos) define the columns and the rows of the matrix (as shown in Fig. 1 in the SI).

3.3. ANN's architecture

Based on previous results [6], we used a fully connected neural network for this work. We tried different architectures varying the number of hidden layers, the number of neurons, the dropout probability, and the activation functions to improve the performance of the ANN. Fig. 4 shows a scheme of the optimal network: the inputs to the ANN are the flattened versions of the 2D SMILES matrices; we then pass the input to two fully connected hidden layers, containing 40 neurons each, and a single output regression layer. We used the ELU activation function in the hidden layers, a variation of the most common ReLU activation function, characterized by an exponential contribution [20,21]. In addition, we imposed a 40% dropout probability on each hidden layer in the training phase, whereas the output regression activation function was linear, and its loss function was the mean average percentage error (MAPE), defined by:

$$Loss = \frac{1}{m_x}\sum_{i=1}^{m_x}\frac{\left|T_i - T_i'\right|}{T_i} \tag{3}$$

where m_x represents the number of elements in the x-th mini-batch, T_i represents the experimental T_g value collected for the i-th compound in the mini-batch and T_i' the calculated value using the ANN for the same compound.

3.4. ANN's optimization

The network was trained using the Adam optimization algorithm [22] provided by MATLAB with the default parameters for beta_1, beta_2, and epsilon (0.9, 0.999 and 10^{-8}, respectively) and applying the mini-batch strategy (mini-batch size = 16) to estimate the gradient of the loss function. In addition to the Adam algorithm, we imposed an external drop of the initial learning rate (lr), starting from $lr = 0.01$ and multiplying it by 0.25 every 500 epochs. We found that initializing the hidden layers and the output layers with a bias vector and a weight tensor whose elements are all ones, significantly improved the network's performance, compared to other initialization methods [23], most likely due to the inclusion of the dropout layers during training, which randomize the data transfer from one layer to the next one [24].

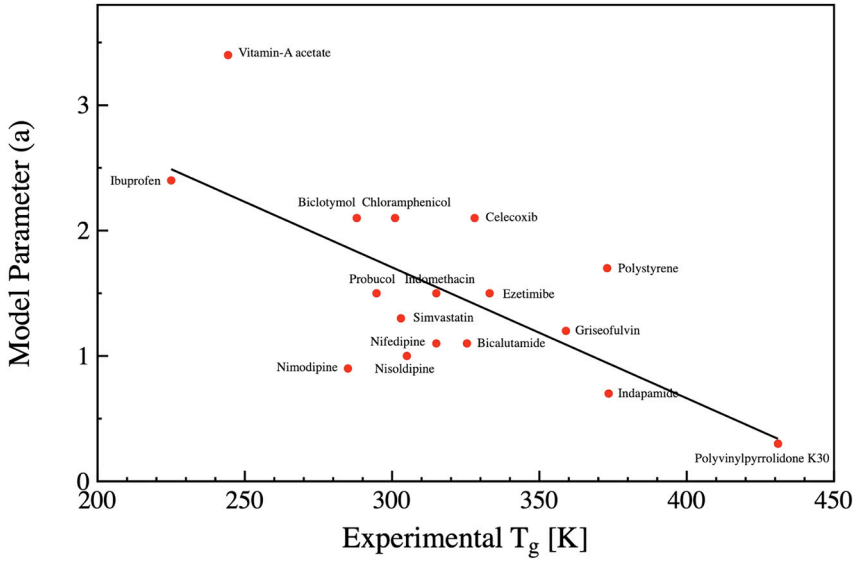

Fig. 3. Glass transition temperature dependence of the model adjustable parameter a for several glass formers. The solid line represents the linear fit of the experimental data. Data were taken from reference 12.

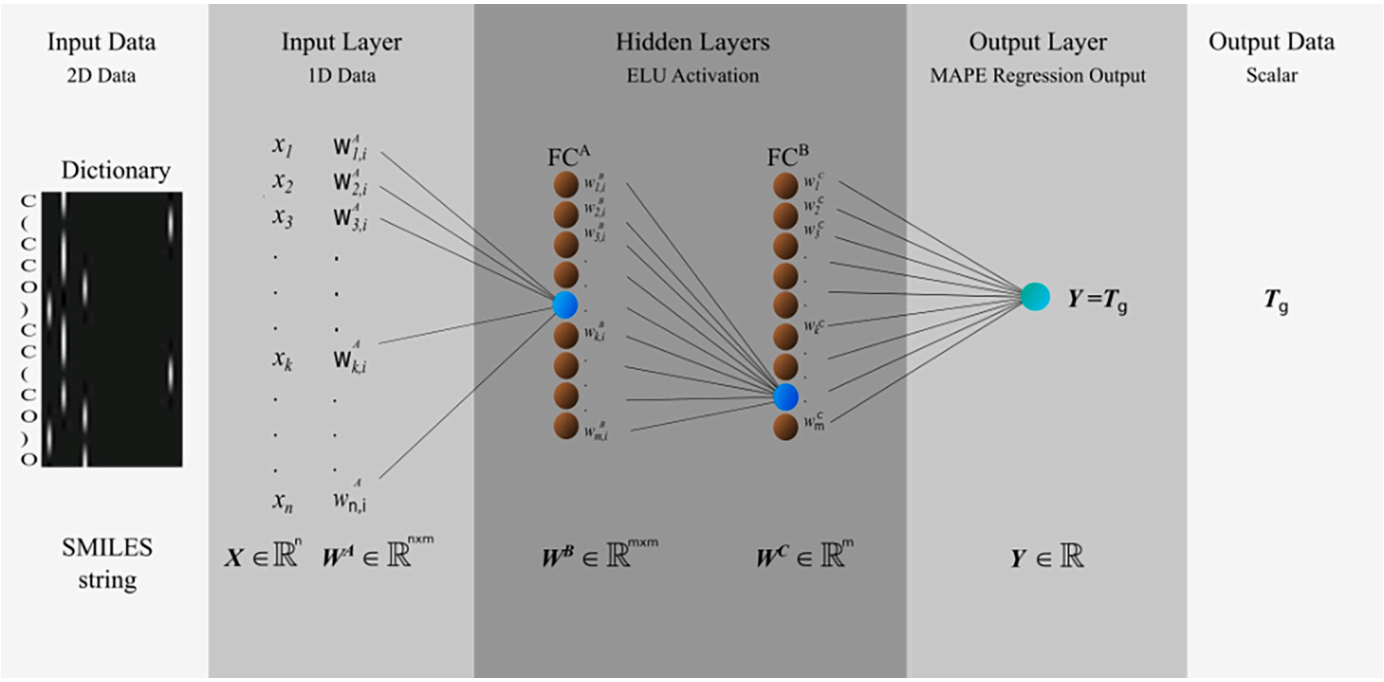

Fig. 4. Schematic picture of the artificial neural network used to predict the glass transition temperature.

3.5. ANN training

The first step consists of training the neural network with known pairs of SMILES strings – glass transition temperatures. This data set is called the *training set*. The examples in the training set are fed into the network, which compares the predicted value with the corresponding experimental T_g. Then, the network adjusts its weights and bias values, using an appropriate learning algorithm, to minimize the average relative error between predicted and known T_g values. In parallel to the training process, the average relative error of the *validation set* is also calculated after each epoch (1 epoch = number of iterations over the mini-batches so that the whole training set is spanned). Since the molecules in the validation set do not participate in the training process, the prediction of their T_g values gives an estimation of the generalization power of the neural network. Once the network can generalize, it is fed with the molecules of the *test set*. This data set corresponds to molecules that were never fed to the network during the training phase, and it is in this sense that we say they are "unknown" (or "new") compounds. Since these molecules do not belong to training or validation sets, we can say that the neural network *predicts* their glass transition temperature.

We divided the data set into 90% training set, 6% validation set, and 4% test set for two main reasons: on the one hand, due to the structural complexity and variability of the molecules in our data set, the network needs to learn many different features and therefore it needs to have as many examples (molecules) as possible in the training set; on the other hand, we selected for test set those molecules for which we could find published experimental measurements of the alpha-relaxation dynamics, in order to have physical feedback to compare our method with. In addition to this, we also wanted to ensure that the T_gs of the test set span over the whole temperature range (200 K–400 K). Once we extracted the test set molecules from the data set, we tried different partitions between training and validation sets, looking for a good representation of the chemical features (in the training set) that minimizes the average percentage error (in the validation set).

4. Results and discussion

4.1. Estimation of the glass transition temperature

Fig. 5 shows the predicted vs. measured glass transition temperature for all the molecules in the data set. The green, blue and red points correspond to training, validation, and test sets, respectively. In addition, the chemical structure of the test set compounds is also shown. We got average relative errors of 7.26% and 7.63% for validation and test sets, respectively. These errors are comparable to similar previous ANN published results [25–27] and are close to half the error obtained with linear regression models. Thus, the ANN does capture the relationship between the chemical structure and the glass transition temperature of molecular glass formers. The observed differences can be rationalized by analysing the chemical features in the training and the test sets. As shown in Table 1 in the SI, the training set contains several examples of molecules with strong intermolecular forces (with several hydrogen bond acceptors and donors in specific molecules), aliphatic cycles, and stereochemistry close to sucrose, lyxose, and trehalose. A somewhat similar situation is observed for sorbitol, where the structure and presence of OH groups are also well represented in the training set (especially in xylitol and meglumine). As a result, the ANN can correctly learn the structure-glass transition temperature relationship of these compounds from sucrose benzoate, galactose, fructose, salicin, xylose, halothane, lactose, meglumine, and ribose (see Fig. 2 in SI).

4.2. Estimation of the temperature dependence of the relaxation times

Figs. 6 and 7 show the temperature dependence of the relaxation times for sucrose, lyxose, salol, trehalose, and sorbitol. Dots represent the experimental values, as measured by BDS and reported elsewhere [28–31], while shaded bands represent the range of relaxation times obtained by ECNLE theory (from ANN's predicted T_g values, including error bands for $T_g \pm 8\%$ corresponding to the average percentage error on the validation set). As shown, the experimental observations are in these cases inside (or very close to) the predicted relaxation region having an excellent agreement for sucrose, lyxose, salol, sorbitol, and trehalose. It is worth reminding here that the only input to the joint

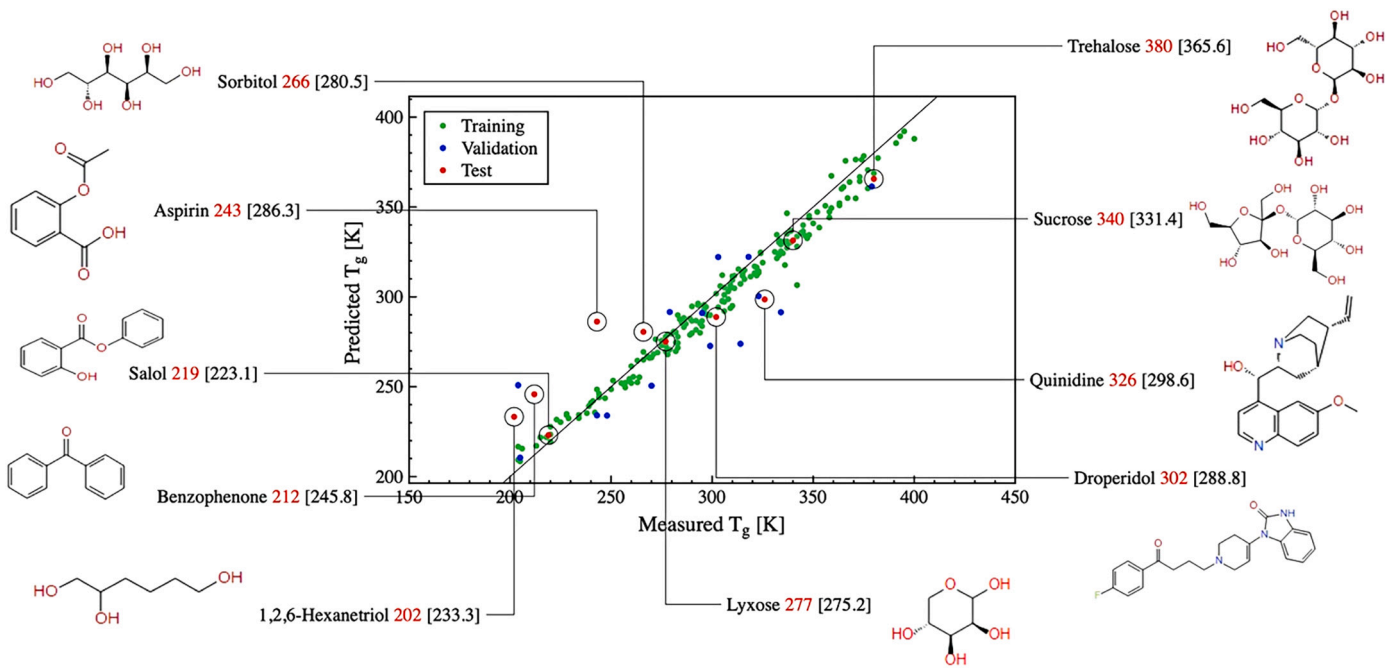

Fig. 5. Predicted vs. measured glass transition temperature for all the molecules in the data set. The green, blue and red points correspond to training, validation, and test sets, respectively. Experimental (in red) and predicted (in brackets) glass transition temperatures are indicated (in kelvin) for all the molecular glass formers in the test set. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

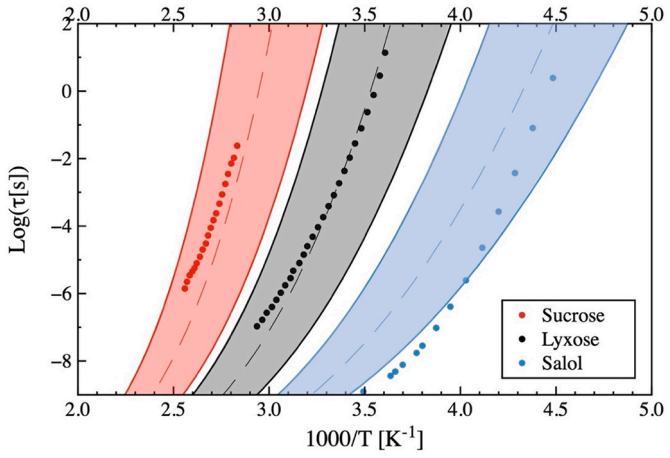

Fig. 6. Relaxation map for sucrose, lyxose and salol. Dots represent experimental values measured by BDS taken from references 28–30. Dashed lines represent the ECNLE prediction for the temperature dependence of the relaxation times based on the glass transition temperature estimated from the ANN. Shaded bands indicate the range of relaxation times as predicted by ECNLE theory based on the prediction error of the ANN ($T_g \pm 8\%$).

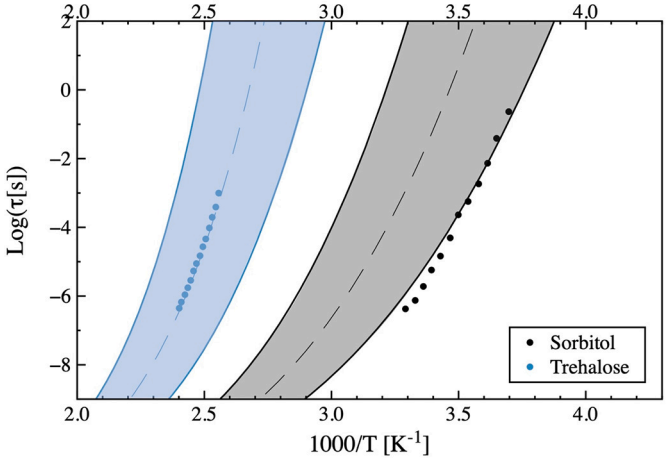

Fig. 7. Relaxation map for sorbitol and trehalose. Dots represent experimental values measured by BDS taken from references 31 and 29. Dashed lines represent the ECNLE prediction for the temperature dependence of the relaxation times based on the glass transition temperature estimated from the ANN. Shaded bands indicate the range of relaxation times as predicted by ECNLE theory based on the prediction error of the ANN ($T_g \pm 8\%$).

numerical-theoretical approach we propose in this work is the chemical structure represented as a SMILES code. Even if the molecule has not been synthesized yet, we can still have a good estimation of its glass transition temperature and the temperature dependence of its main relaxation time.

For specific compounds, like hexanetriol (see Fig. 3 in SI) or benzophenone, some deviations between estimated and experimental dynamics are observed. These differences may arise from two sources. On the one hand, due to the complexity and the variability of the molecular structures of the data set, it becomes difficult to obtain an excellent generalization from the ANN: we have to consider that mapping the chemical features during the ANN training determines the chemical structure-glass transition temperature relationship for each compound. Intuitively, we can see that chemical features better represented in the training set are more likely to be accurately mapped (see Fig. 2 in SI), and therefore, the corresponding T_g is better predicted. On the contrary, molecules with non-common features will be underrepresented, so their estimated glass transition temperature will likely present higher uncertainties. Therefore, it is expected that the ANN precision will further improve when more examples (that appropriately represent the chemical features observed in the test set) are added to the training set.

On the other hand, in some cases (see salol in Fig. 6), the glass transition temperature is well predicted by the ANN, but the estimated

dynamics (the temperature dependence of the relaxation time) deviates from the experimental values. This behaviour is most likely related to the assumption that local and collective dynamics (in the ECNLE model) correlate to each other in the same way for all materials. As shown in Fig. 3, the coupling between the two dynamics (parameter a) depends on T_g, and therefore, predicting the dynamics for those compounds with lower values of T_g is less accurate. These deviations can be corrected by considering non-universal effects of chemical and biological complexities (as previously discussed).

In this sense, it is noteworthy that trehalose, with a relatively high T_g (380 K), shows excellent agreement between the experimental and predicted values for the temperature dependence of the relaxation times. For intermediate T_g values, like lyxose (277 K), the predicted temperature dependence of the relaxation times slightly deviates from the experimental values (provided the T_g is well predicted). On the low T_g side, we have the case of salol (219 K), for which the T_g is correctly estimated, but the dynamics departs from the experimental points upon increasing temperature. These differences in the temperature dependence of the relaxation times are expected due to different molecular structures, glass transition temperature, and the different number of H bond donors (OHs in the structure) and acceptors (oxygen atoms in the structure) in the studied compounds [32].

Fig. 8 shows some corrected dynamics calculated, including non-universal effects (a $\neq$ 1) of chemical and biological complexities in molecules like sucrose, lyxose, and salol. According to Fig. 3, we took a = 1.37, 1.96, and 2.5 for sucrose, lyxose, and salol, respectively. Theoretical curves (with the corresponding values of the adjustable parameter) are now closer to the experimental data. We expect that a better understanding of the dependence of this parameter on chemical structure or glass transition temperature further improves the theoretical predictions of the temperature dependence of the relaxation times.

It is important to discuss here some limitations of the proposed approach. For new materials or those without experimental data of $\tau_\alpha(T)$, the parameter a cannot be determined. To zeroth-order approximation, we use a linear function to empirically describe the a-T_g relation as shown in Fig. 3. Thus, combining the T_g value predicted from the chemical structure and ANN network with the empirical a-T_g relation allows us to determine (through ECNLE) $\tau_\alpha(T)$ without any adjustable/fit parameter. However, the linear a-T_g fit means that a given T_g leads to one value of a or dynamic fragility. As a result, our approach deduces that glass-forming materials having the same T_g have the same fragility, and this is not necessarily the case as shown in previous publications [33,34]. A good option to overcome this limitation is to use the ANN not only to predict the T_g, but also the fragility. We tried this approach, but unfortunately, there is a lack of data for the fragility in the literature, that makes it highly inaccurate. We expect that the available amount of data will increase in the next years, allowing the use of ANNs for predicting fragility.

Besides estimating the glass transition temperature, the ANN can also provide a new understanding of molecular glass formers. For instance, Fig. 9 shows the predicted T_g for isomers of lyxose and galactose (except for these two, we could not find the corresponding experimental T_g values for the rest of the molecules in the scientific literature). It is interesting to note that for L-Arabinose and beta-L-Arabinose, which only differ on the position of the upper right OH group, the T_g only changes three degrees. However, the same structural change between galactose and alpha-D-Galactose gives a T_g difference of 33 K. In this case, the presence of the upper left group induces a higher sensitivity of the dynamics to minor structural changes. It is worth mentioning that although the average relative error of the ANN's prediction is about 8%, in the case of lyxose and galactose, the corresponding errors are below 1% (see Table 1 in SI), making sense of the observed differences in Fig. 9. The same analysis can be performed on molecules not even synthesized, boosting the development of new materials with tuned properties.

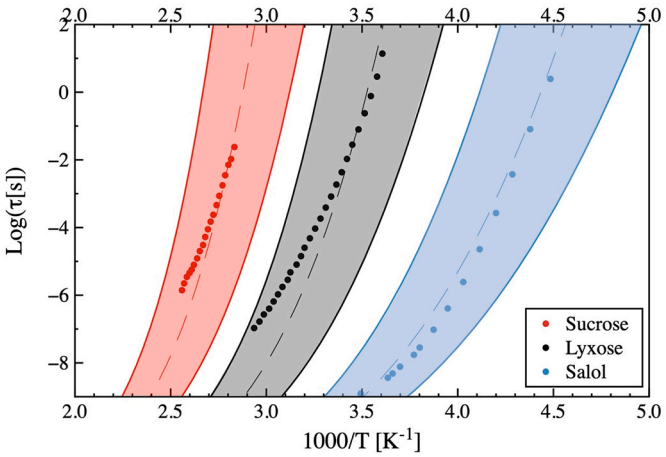

Fig. 8. Relaxation map for sucrose, lyxose and salol. Dots represent experimental values measured by BDS taken from references 28–30, and, respectively. Dashed lines represent the ECNLE prediction for the temperature dependence of the relaxation times based on the glass transition temperature estimated from the ANN and corrected with the adjustable parameter a as taken from Fig. 3. Shaded bands indicate the range of relaxation times predicted by the joint numerical/theoretical approach proposed here.

5. Conclusions

We have presented in this work a new approach that combines numerical methods with theory to estimate the temperature dependence of the structural relaxation time for molecular glass formers. Firstly, we built, optimized and trained an artificial neural network to assess the glass transition temperature of molecular glass formers only based on their chemical structure. Then, we used a theoretical approach based on the elastically collective nonlinear Langevin equation to estimate the full relaxation map. Although there is still some room to improve accuracy and overcome limitations, this first joint theoretical and numerical approach constitutes a suitable tool for giving a reasonable estimation of the dynamics of unknown molecular glass formers based on their chemical structure. This approach will boost materials and drug development by designing molecular glass formers with desired properties and will also increase the understanding of the physical mechanisms related to molecular dynamics.

Data availability statements

The data that supports the findings of this study are available within the article and in the Supporting Information file (SI).

Credit author statement

Author 1: Claudia Borredon.
Collected the data; Contributed data or analysis tools; Performed the analysis; Wrote the paper.
Author 2: Luis A. Miccio.
Conceived and designed the analysis; Contributed data or analysis tools; Wrote the paper.
Author 3: Anh D. Phan.
Contributed data or analysis tools; Wrote the paper.
Author 4: Gustavo A. Schwartz.
Conceived and designed the analysis; Contributed data or analysis tools; Wrote the paper; other contributions.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

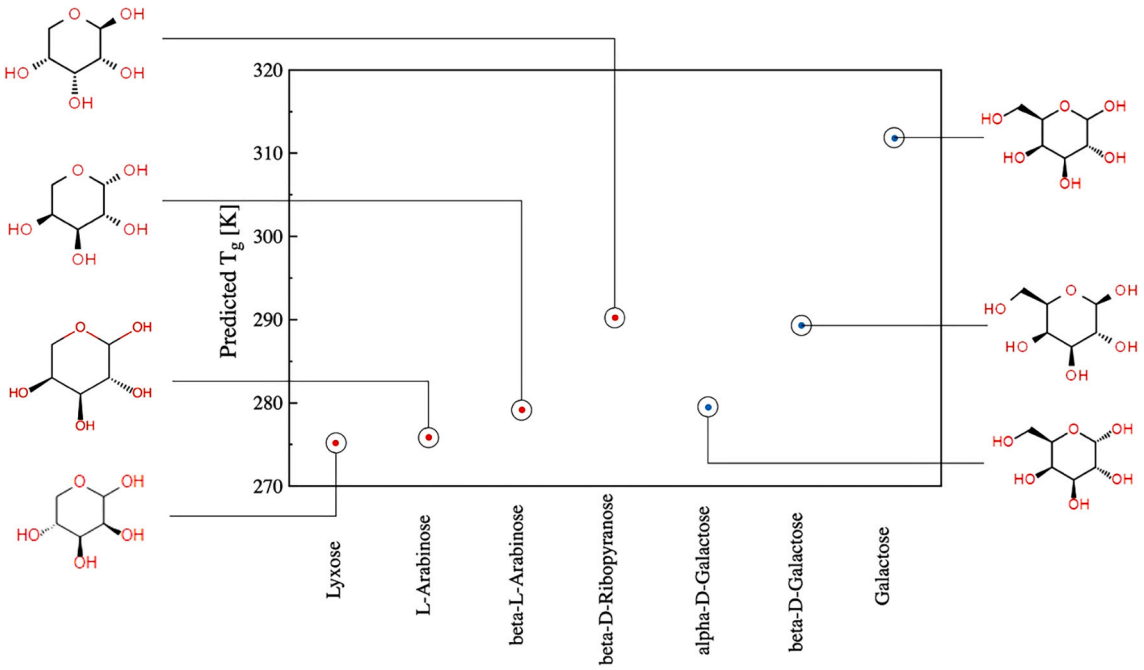

Fig. 9. Predicted glass transition temperature for isomers of lyxose (red) and galactose (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the work reported in this paper.

Acknowledgments

We gratefully acknowledge the financial support from the Spanish Government "Ministerio de Ciencia e Innovación" (PID2019-104650GB-C21) and the Basque Government (IT-1566-22). We also acknowledge support of the publication fee by the CSIC Open Access Publication Support Initiative through its Unit of Information Resources for Research (URICI). This research was also funded by the Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 103.01-2019.318. We also acknowledge the support of NVIDIA Corporation with the donation of two GPUs used for this research.

Appendix A. Supplementary data

Additional information on the glass formers, their glass transition temperatures, chemical structures, and SMILES code can be found in the Supporting Information file (SI). Supplementary data to this article can be found online at https://doi.org/10.1016/j.nocx.2022.100106.

References

[1] A.D. Phan, J. Knapik-Kowalczuk, M. Paluch, T.X. Hoang, K. Wakabayashi, Theoretical model for the structural relaxation time in coamorphous drugs, Mol. Pharm. 16 (2019) 2992–2998, https://doi.org/10.1021/acs.molpharmaceut.9b00230.

[2] W. Tu, J. Knapik-Kowalczuk, K.C. Molecular, Glass transition dynamics and physical stability of amorphous Griseofulvin in binary mixtures with low-Tg excipients, Chem. Rev. (2019), https://doi.org/10.1021/acs.molpharmaceut.9b00476.

[3] K. Grzybowska, A.G. Molecular, Molecular dynamics and physical stability of ibuprofen in binary mixtures with an acetylated derivative of maltose, Chem. Rev. (2020), https://doi.org/10.1021/acs.molpharmaceut.0c00517.

[4] A.D. Phan, Determination of Young's Modulus of Active Pharmaceutical Ingredients by Relaxation Dynamics at Elevated Pressures, 2020, pp. 1–7, https://doi.org/10.1021/acs.jpcb.0c05523.

[5] A.D. Phan, K.S. Schweizer, Elastically collective nonlinear Langevin equation theory of glass-forming liquids: transient localization, thermodynamic mapping, and cooperativity, J. Phys. Chem. B 122 (2018) 8451–8461, https://doi.org/10.1021/acs.jpcb.8b04975.

[6] L.A. Miccio, G.A. Schwartz, Localizing and quantifying the intra-monomer contributions to the glass transition temperature using artificial neural networks, Polymer 203 (2020), 122786, https://doi.org/10.1016/j.polymer.2020.122786.

[7] L.A. Miccio, G.A. Schwartz, Mapping chemical structure–glass transition temperature relationship through artificial intelligence, Macromolecules 54 (2021) 1811–1817, https://doi.org/10.1021/acs.macromol.0c02594.

[8] L.A. Miccio, G.A. Schwartz, From chemical structure to quantitative polymer properties prediction through convolutional neural networks, Polymer 193 (2020), 122341, https://doi.org/10.1016/j.polymer.2020.122341.

[9] A.D. Phan, K. Wakabayashi, Theory of structural and secondary relaxation in amorphous drugs under compression, Pharm 12 (2020) 177, https://doi.org/10.3390/pharmaceutics12020177.

[10] A.D. Phan, T.T.T. Thuy, N.T.K. An, J. Knapik-Kowalczuk, M. Paluch, K. Wakabayashi, Molecular relaxations in Supercooled liquid and glassy states of amorphous Gambogic acid: dielectric spectroscopy, calorimetry, and theoretical approach, AIP Adv. 10 (2020), 025128, https://doi.org/10.1063/1.5139101.

[11] A.D. Phan, A. Jedrzejowska, M. Paluch, K. Wakabayashi, Theoretical and experimental study of compression effects on structural relaxation of glass-forming liquids, Acs Omega 5 (2020) 11035–11042, https://doi.org/10.1021/acsomega.0c00860.

[12] A.D. Phan, K. Wakabayashi, M. Paluch, V.D. Lam, Effects of cooling rate on structural relaxation in amorphous drugs: elastically collective nonlinear Langevin equation theory and machine learning study, RSC Adv. 9 (2019) 40214–40221, https://doi.org/10.1039/c9ra08441j.

[13] S. Mirigian, K.S. Schweizer, Elastically cooperative activated barrier hopping theory of relaxation in viscous fluids. II. Thermal liquids, J. Chem. Phys. 140 (2014), 194507, https://doi.org/10.1063/1.4874843.

[14] S. Mirigian, K.S. Schweizer, Unified theory of activated relaxation in liquids over 14 decades in time, J. Phys. Chem. Lett. 4 (2013) 3648–3653, https://doi.org/10.1021/jz4018943.

[15] S. Mirigian, K.S. Schweizer, Elastically cooperative activated barrier hopping theory of relaxation in viscous fluids. I. General formulation and application to hard sphere fluids, J. Chem. Phys. 140 (2014), 194506, https://doi.org/10.1063/1.4874842.

[16] L.D. Landau, E.M. Lifshitz, Theory of Elasticity, Permagon Press, London, 1975. ISBN 9780080570693.

[17] S.-J. Xie, K.S. Schweizer, Nonuniversal coupling of cage scale hopping and collective elastic distortion as the origin of dynamic fragility diversity in glass-forming polymer liquids, Macromolecules 49 (2016) 9655–9664, https://doi.org/10.1021/acs.macromol.6b02272.

[18] N.M. O'Boyle, Towards a universal SMILES representation - a standard method to generate canonical SMILES based on the InChI, Aust. J. Chem. 4 (2012) 22, https://doi.org/10.1186/1758-2946-4-22.

[19] D. Weininger, SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules, J. Chem. Inf. Model. 28 (1988) 31–36, https://doi.org/10.1021/ci00057a005.

[20] C. Nwankpa, W. Ijomah, A. Gachagan, S. Marshall, Activation functions: comparison of trends in practice and research for deep learning, 2nd International Conference on Computational Sciences and Technology, Jamshoro, Pakistan. (2021) 124–133.

[21] D.-A. Clevert, T. Unterthiner, S. Hochreiter, Fast and accurate deep network learning by exponential linear units (ELUs), 4th International Conference on Learning Representations (2016) 1–14.
[22] D.P. Kingma, Jimmy Ba, Adam: a method for stochastic optimization, 3rd International Conference on Learning Representations (2015) 1–14.
[23] X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in: Proceedings of the 13th International Conference on Artificial Intelligence and Statistics *9*, **2010**, pp. 249–256.
[24] D. Hendrycks, K. Gimpel, Adjusting for dropout variance in batch normalization and weight initialization, Workshop track - 5th International Conference on Learning Representations (2017) 1–10.
[25] W. Liu, 2010 prediction of glass transition temperatures of aromatic heterocyclic polyimides using an ANN model, Wiley Online Library 50 (2010) 1547–1557, https://doi.org/10.1002/pen.21670.
[26] L. Ning, Artificial neural network prediction of glass transition temperature of fluorine-containing polybenzoxazoles, J. Mater. Sci. 44 (2009) 3156–3164, https://doi.org/10.1007/s10853-009-3420-0.
[27] W. Liu, C. Cao, Artificial neural network prediction of glass transition temperature of polymers, Colloid Polym. Sci. 287 (2009) 811–818, https://doi.org/10.1007/s00396-009-2035-y.
[28] P.S. Lokendra, A. Alegría, J. Colmenero, Broadband dielectric spectroscopy and calorimetric investigations of D-Lyxose, Carbohydr. Res. 346 (2011) 2165, https://doi.org/10.1016/j.carres.2011.06.029.
[29] K. Kaminski, K. Adrjanowicz, D.Z. Molecular, Dielectric studies on molecular dynamics of two important disaccharides: sucrose and trehalose, Chem. Rev. (2012), https://doi.org/10.1021/mp2004498.
[30] J. Bartoš, M. Iskrová, M. Köhler, R. Wehn, O. Šauša, P. Lunkenheimer, J. Krištiak, A. Loidl, Positron annihilation response and broadband dielectric spectroscopy: salol, European Phys J E 34 (2011) 104, https://doi.org/10.1140/epje/i2011-11104-x.
[31] A. Faivre, G. Niquet, M. Maglione, J. Fornazero, J.F. Jal, L. David, Dynamics of sorbitol and maltitol over a wide time-temperature range, The Eur. Phys. J. B - Condensed Matter Complex Syst. 10 (1999) 277–286, https://doi.org/10.1007/s100510050856.
[32] Q. Qian, Gregory B. McKenna, Gregory correlation between dynamic fragility and glass transition temperature for different classes of glass forming liquids, J. Non-Cryst. Solids 352 (2006) 2977, https://doi.org/10.1016/j.jnoncrysol.2006.04.014.
[33] K. Kunal, C.G. Robertson, S. Pawlus, S.F. Hahn, A.P. Sokolov, Role of chemical structure in fragility of polymers: a qualitative picture, Macromolecules 41 (2008) 7232–7238, https://doi.org/10.1021/ma801155c.
[34] W.-S. Xu, K.F. Freed, Influence of cohesive energy and chain stiffness on polymer glass formation, Macromolecules 47 (2014) 6990–6997, https://doi.org/10.1021/ma501581u.