JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

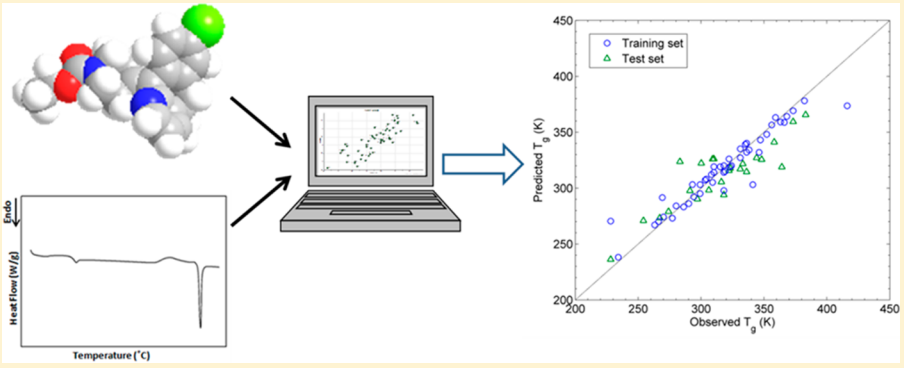
Experimental and Computational Prediction of Glass Transition Temperature of Drugs

Ahmad Alzghoul,[†] Amjad Alhalaweh,[‡] Denny Mahlin,[‡] and Christel A. S. Bergström*[,‡]

[†]Department of Information Technology, Uppsala University, P.O. Box 337, SE- 751 05 Uppsala, Sweden

[‡]Department of Pharmacy, Uppsala University, Uppsala Biomedical Centre, P.O. Box 580, SE-751 23 Uppsala, Sweden

Supporting Information



ABSTRACT: Glass transition temperature (T_g) is an important inherent property of an amorphous solid material which is usually determined experimentally. In this study, the relation between T_g and melting temperature (T_m) was evaluated using a data set of 71 structurally diverse druglike compounds. Further, in silico models for prediction of T_g were developed based on calculated molecular descriptors and linear (multilinear regression, partial least-squares, principal component regression) and nonlinear (neural network, support vector regression) modeling techniques. The models based on T_m predicted T_g with an RMSE of 19.5 K for the test set. Among the five computational models developed herein the support vector regression gave the best result with RMSE of 18.7 K for the test set using only four chemical descriptors. Hence, two different models that predict T_g of drug-like molecules with high accuracy were developed. If T_m is available, a simple linear regression can be used to predict T_g. However, the results also suggest that support vector regression and calculated molecular descriptors can predict T_g with equal accuracy, already before compound synthesis.

INTRODUCTION

Amorphous material is widely investigated in several disciplines including optical, food, material, and pharmaceutical sciences.[1−4] It has been defined simply as a "liquid that lost its ability to flow".[1] Compared to its crystalline counterpart, it lacks long-range order of molecular packing[2] and has higher molecular mobility and different molecular association and distances within the solid.[5,6] The amorphous material is characterized by the glass transition temperature (T_g) below which a decrease in the specific heat and thermal expansion coefficient occur.[3,7] The behavior of the glassy material, including glass transition, enthalpy relaxation, devitrification, fragility, entropy, and free energy of the material, is reflected by its molecular level properties.[5,6] Small organic compounds are considered to be fragile glass formers that exhibit a non-Arrhenius behavior of relaxation as a function of temperature.[2] The T_g (in kelvin) of these fragile systems may be estimated from the melting temperature (T_m). T_m has been found to be 1 to 1.5-fold greater than T_g (or the inverse relationship of T_g over T_m being 0.66−1).[4,8] The relation between T_g and T_m for 20 pharmaceutical compounds was reported to be in the range 0.59−0.84.[9] In another study, a larger data set (46 drug molecules and 18 sugars, vitamins, lipids, carboxylic acids) was investigated for the relation of T_g/T_m, and it was found to be in a similar range (0.59−0.86).[3,4,10] Based on these studies, the mean of T_m over T_g has been found to be 1.36 and this value has been suggested as useful for estimation of the T_g.[4]

T_g can be determined by different experimental approaches, e.g. differential scanning calorimetry (DSC), isothermal calorimetry, and dynamic mechanical analysis, and under different conditions. These methods are invasive, costly, and time-consuming.[2,11−13] Computational models based on molecular descriptors have been developed to predict T_g of non-drug-like molecules.[14,15] These models were based on a data set with limited structural diversity out of which the majority of the substances were organic solvents, and hence, there is a need to explore to which extent e.g. in silico models based on molecular descriptors also can predict the T_g of solid drugs. In this study we therefore investigated models of

Received: August 4, 2014
Published: October 31, 2014

dx.doi.org/10.1021/ci5004834 | J. Chem. Inf. Model. 2014, 54, 3396−3403

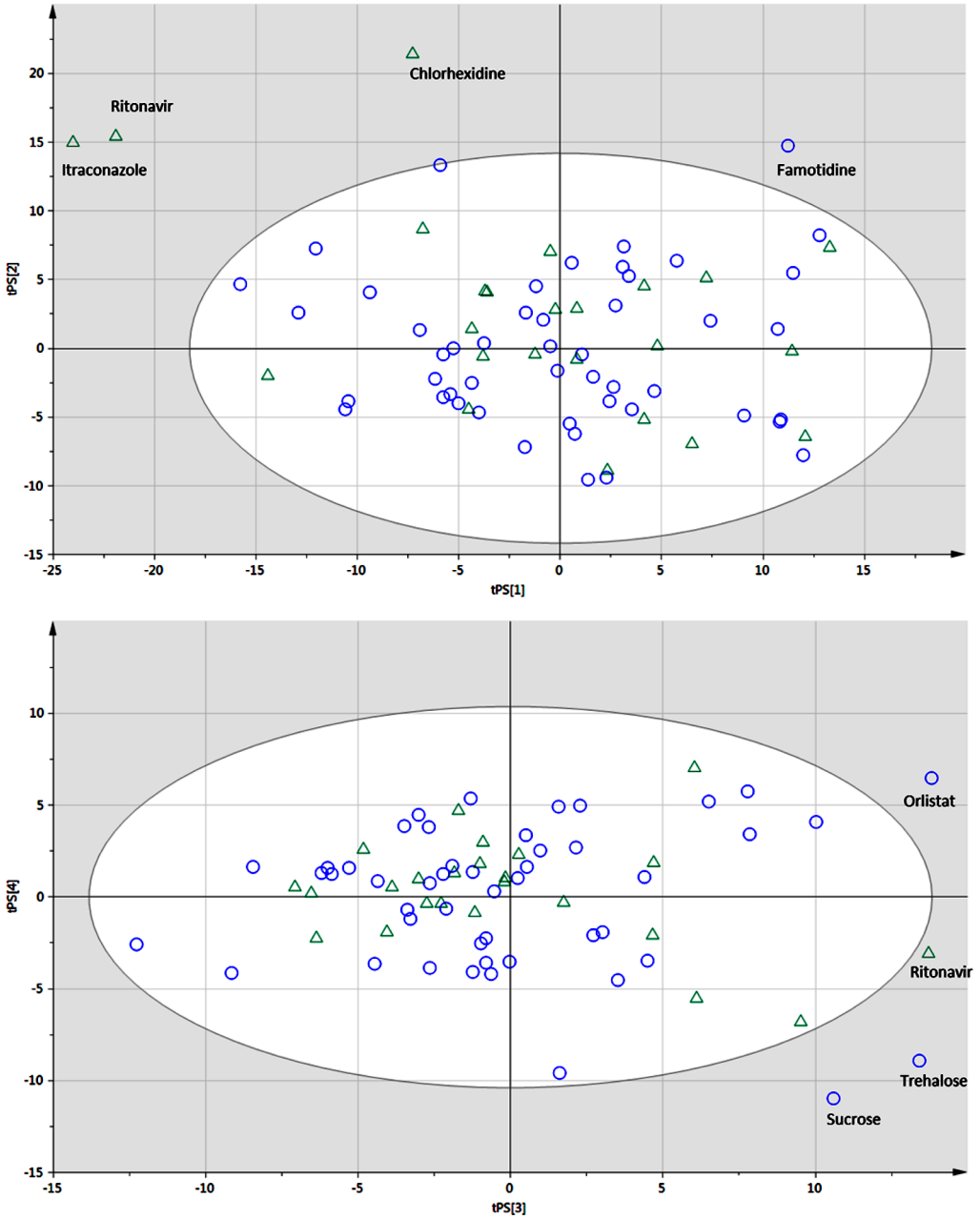


Figure 1. Principal component analysis for selection of training (blue) and test (green) sets. The ellipse shows the 95% confidence interval of the presented principal components. The first four principal components describe 53% of the chemical diversity of the data set.

different levels of complexity for their applicability for T_g predictions making use of a data set of 71 druglike molecules. First the linear relationship between T_m of the drugs and their T_g was analyzed based on this structurally diverse data set. Second, the data set was used for developing in silico models for prediction of T_g from molecular descriptors. Linear (multiple linear regression (MLR), partial least-squares (PLS), principal component regression (PCR)) and nonlinear (neural network (NN), support vector regression (SVR)) regression models were explored to identify the descriptors and algorithm that best predict T_g of drugs.

METHODS

Materials. All chemicals were purchased from standard suppliers and were of high purity (>98%). To enable the development of models with general applicability 71 structural diverse but druglike compounds were selected based on our previous knowledge on glass forming ability[16−19] and an estimated wide range in T_g.

Production of the Amorphous State. Measurements of T_m, amorphization and T_g determination of each compound were performed in a DSC Q2000 (TA Instruments, USA) which was calibrated for temperature and enthalpy using indium. The instrument was equipped with a refrigerated cooling system. Melting point was determined for each compound as received using an amount of 1−3 mg in nonhermetic sealed aluminum pans. The compounds were scanned at a heating rate of 10 °C/min under a continuously purged dry nitrogen atmosphere (50 mL/min). Glass formation was investigated by using 1−3 mg of the compound in nonhermetic sealed pans and heated (using equilibrate function) to around 2 °C above the peak of melting

dx.doi.org/10.1021/ci5004834 | J. Chem. Inf. Model. 2014, 54, 3396−3403

Table 1. Experimentally Measured Glass Transition Temperature (T_g) and Onset of Melting (T_m)

compound	T_g (K)	T_m(K)	T_g/T_m
Tenofovir	416	552	0.75
Metolazone	382	539	0.71
Hydroflumethiazide	373	542	0.69
Budesonide	368	530	0.69
Prednisone	366	513	0.71
Aripiprazole	363	526	0.69
Hydrocortisone	359	497	0.72
Bucindolol	356	459	0.78
Danazol	352	500	0.70
Sucrose	347	458	0.76
Bezafibrate	346	457	0.76
Warfarine	341	435	0.78
Ezetimibe	338	437	0.77
Glafenine	336	437	0.77
Glibenclamide	335	445	0.75
Celecoxib	331	436	0.76
D-Salicin	331	474	0.70
Pimozide	327	492	0.66
Omeprazole	324	428	0.76
Famotidine	323	411	0.79
Zolmitriptan	322	410	0.79
Diazepam	319	404	0.79
Felodipine	318	420	0.76
Ketoconazole	318	423	0.75
Nifedipine	318	446	0.71
Carvedilol	315	390	0.81
Testosterone	315	426	0.74
Loratadine	310	409	0.76
Nandrolone	310	397	0.78
Simvastatin	309	412	0.75
Clemastine	308	451	0.68
Chloramphenicol	304	425	0.72
Clotrimazole	303	418	0.72
Acetohexamide	299	463	0.65
Fluorescamine	299	428	0.70
Nimesulide	294	423	0.70
Physostigmine	293	377	0.78
Bifonazole	290	424	0.68
Nizatidine	286	406	0.70
Cinnarizine	280	394	0.71
Captopril	277	380	0.73
Ketoprofen	270	368	0.73
Clofoctol	269	361	0.75
Tinidazole	266	399	0.67
Tamoxifen	263	371	0.71
Procaine	234	335	0.70
Orlistat	228	316	0.72
Hydrochlorothiazide	383	536	0.71
Linaprazan	373	519	0.72
Spironolactone	364	486	0.75
Estradiol	358	451	0.79
Sulindac	348	460	0.76
Emtricitabine	344	426	0.81
Chlorhexidine	336	408	0.82
Albendazole	333	475	0.70
Itraconazole	331	441	0.75
Bicalutamide	323	465	0.69
Ritonavir	322	399	0.81
Indomethacin	318	434	0.73
Isradipine	316	432	0.73
Acemetacin	310	421	0.74
Aprepitant	309	422	0.73
Nilutamide	306	428	0.71
Probucol	300	400	0.75
Acetaminophen	297	443	0.67
Tolazamide	291	445	0.65
Aceclofenac	283	426	0.66
Miconazole	274	359	0.76
Flurbiprofen	267	388	0.69
Fenofibrate	254	354	0.72
Ibuprofen	228	350	0.65
median	318	428	0.73
min	228	316	0.65
max	416	552	0.82

[a]Compounds above the blank line were included in the training set whereas compounds below the line were included in the test set. Compounds are listed in decreasing order of T_g.

temperature under a continuously purged dry nitrogen atmosphere (50 mL/min). The system was kept isothermal for 2 min to ensure complete melting of the system and thereafter cooled to −70 °C at a ramp rate of 20 °C/min. The formation of the glass state was then investigated by performing a second heat cycle at a heating rate of 20 °C/min immediately after cooling. The amorphous formation was indicated by detection of T_g upon heating.

Model Development. The compounds were listed in decreasing order for T_g and every third compound was designated to the test set. The suitability of the training set for model development and the test set for validation was investigated by principal component analysis (PCA) performed in Simca-P v 13 (Umetrics, Sweden). The analysis identified that ritonavir was not well-described by the training set, and this molecule was therefore included in the test set to not overweight the models developed. This resulted in 47 compounds for training and 24 for validation of the final models (Figure 1, Table 1).

Linear Relationship between Glass Transition Temperature and Melting Temperature. The relation between T_m and T_g for the training data set in the study was found to be linear (R = 0.91) and therefore a linear regression model based on T_m was developed. The parameters of the linear model were estimated through the least-squares approach.

Models Based on Calculated Chemical Descriptors. Molecular descriptors were in this work calculated with the software ADMET Predictor (SimulationsPlus, CA). In total a number of 284 descriptors were used as input for the modeling. The first step was to investigate the strength of the linear relationship between each descriptor and the target T_g using the correlation coefficient (R). Unfortunately, none of the 284 descriptors was found to have a high correlation with T_g. The highest correlation with T_g was achieved by the number of hydrogen bond donors descriptor (R of 0.44). However, several descriptors were found to be highly correlated (R > 0.60) with each other. The highly correlated descriptors can be considered as redundant and thus only the descriptor with the highest

dx.doi.org/10.1021/ci5004834 | J. Chem. Inf. Model. 2014, 54, 3396−3403

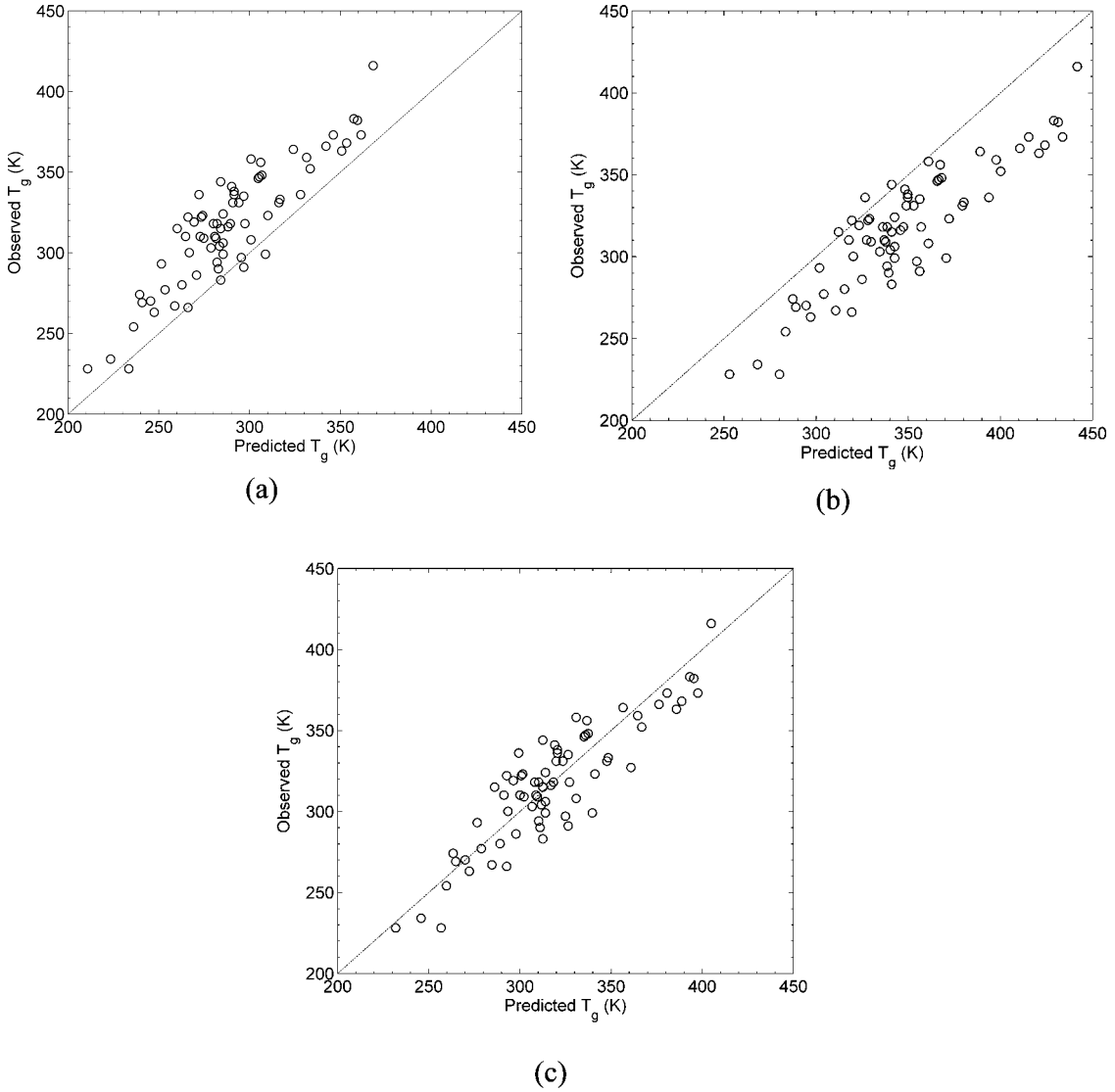


Figure 2. Prediction of T_g using (a) 0.67, (b) 0.80, and (c) 0.73 of T_m. The dashed line represents a line of unity.

relation to T_g was used as input. As a result, the number of descriptors was reduced from 284 to 43.

The next step was to construct regression models using the remaining 43 descriptors. In this work, linear and nonlinear regression models were developed. The linear regression models were based on MLR, PCR, and PLS while the nonlinear models were based on ANN and SVR, each of which is described in detail below.

Development of Linear Models. Linear regression can be used to find relationships between the input variables (the molecular descriptors) and the target (T_g). Linear regression methods aim at finding a target function that can fit the input data with minimum error.[20] In this work, the least-squares error method was used. Advantages of linear regression models are that they are transparent and easy to implement.[21] Since it is not practical to use 43 descriptors in a linear regression model, several algorithms were applied for dimension reduction. The stepwise method was used as a feature selection method, while PCR and PLS were used as feature extraction methods. When applying stepwise regression, the final regression model is obtained by forward selection and backward elimination steps. At each step, a molecular descriptor is added or removed based on the change in the sum of squared error. In comparison, PCR is a regression method that is based on the PCA. It uses the principal components that are calculated from the input variables (i.e., the chemical descriptors) to create the linear regression model, but the calculation of the principal components does not consider the target (i.e., T_g in this work). Normally, using more number of components reduces the prediction error for the training set but may lead to overfitting the model. To avoid the latter, we used cross-validation with four groups. In contrast to PCR, PLS creates a linear regression model based on both the input variables (i.e., the descriptors) and the response (i.e., T_g). Thus, it is expected that the PLS regression model will require less number of components than the PCR. To determine the required number of components for PLS regression model, a cross-validation was used similar to that performed during PCR using four groups.

Development of Nonlinear Models. In addition to the linear models two popular nonlinear regression methods (ANN and SVR) were explored. ANNs can contain several layers of interconnected neurons. These layers represent the input, the output, and one or more hidden layers between them. During the training phase, the weights of the ANN model are adjusted

dx.doi.org/10.1021/ci5004834 | J. Chem. Inf. Model. 2014, 54, 3396−3403

according to the trained data set. Once the neural network is trained, new data points can be applied on the trained network for prediction or classification.[22] In this work, ANN was used to fit the relationship between the descriptors and T_g and the descriptors that had the highest ANN absolute weight were selected for the final model. First, we used one hidden layer with one neuron and trained the neural network with all normalized descriptors. Since we used normalized data, we assumed that the descriptors that have higher absolute weights contributed more to the model and thus were selected for further investigation. The next step was to build an ANN model based on the selected descriptors. The 47 data points (i.e., training data) were divided into three groups: training (70%), validation (15%), and test (15%). Then, the ANN model was trained, validated, and tested. This procedure was iterated using different numbers of neurons, and based on the regression performance of the three groups, the number of neurons in the hidden layer was selected to be 5. Furthermore, we trained and tested the ANN model using the latent variables obtained by PCA to investigate to which extent such an approach may be successful for T_g prediction.

The SVR algorithm is based on the principles of support vector machines (SVM). SVR uses a kernel function to map the input data into a high dimensional feature space where the computation of a linear regression is performed.[23] In this work, the input data, i.e. the molecular descriptors, were mapped into a higher dimensional feature space using a radial basis function. To reduce the number of variables, the Recursive Feature Elimination (RFE) was utilized.[24] The RFE algorithm is a backward elimination method where each step in this work involved the following: (1) SVR was applied to the descriptors, (2) the descriptors were ranked based on their weights in the SVR model, and (3) the descriptors that had the lowest weights were eliminated (the number varied between different iterations). This work flow was then iterated until only the most important variables for the prediction remained. The two stopping criteria were the number of descriptors (target: low) and the prediction performance (target: high). Thus, descriptor(s) were removed even if removing them reduced the prediction performance with around 1%. However, after the RFE process was performed we continued the model development by adding the removed descriptors one by one to analyze to which extent this improved the performance of the model. It was found that the model performance was not improved by the latter.

In this work, we have used the epsilon-SVR method. The epsilon-SVR has two parameters: epsilon and cost. At every step in the RFE process we used four-fold cross-validation and different ranges of epsilon and cost parameters. The parameters that achieved the lowest average error, when using the four-fold cross validation, were used for training the SVR model at that specific step in the RFE process. Note that the parameters may change at each step.

RESULTS AND DISCUSSION

Prediction of Glass Transition Temperature from Melting Temperature. The T_g/T_m for our data set was found to be in the range 0.65 (~2/3) to 0.82 (~4/5) (Table 1) which is in agreement with previous findings.[9,10] The predicted T_g based on 2/3 and 4/5 of T_m is plotted against the experimentally determined T_g in Figure 2a and b, respectively. It was found that neither the 2/3 rule nor the 4/5 rule were successful in predicting this diverse set of drug compounds, as shown in the large RMSE values (33−35 K; Table 2). Using the previous estimated relation of T_m/T_g = 1.36, the T_g was predicted with RMSE value of 17.9 K (Figure 2c).

Table 2. Models Based on T_m and Their Statistical Results

	RMSE (K)		
equation for T_g prediction	training	test	all data
$T_g = 0.67T_m$		33.6	32.6
$T_g = 0.80T_m$		36.2	35.4
$T_g = T_m/1.36$		19.5	17.9
$T_g = 0.73T_m$	16.7	19.5	17.7
$T_g = 0.63T_m + 42$	15.9	19.7	17.3

The relation between T_g and T_m was found to be linear (R = 0.91) and therefore it was decided to establish a linear regression model based on the data set to predict the T_g based on the T_m (Figure 3 and Table 2). The equation derived to predict the T_g based on the training data set was

$$T_g = 0.73T_m \quad (1)$$

eq 1 resulted in R^2 of 0.82 and 0.72 for the training set and test set, respectively, and arriving at the coefficient 0.73 confirms the validity of the previously suggested ratio of T_m and T_g being 1.36. The T_g was estimated with equal accuracy by using a linear model with an intercept:

$$T_g = 0.63T_m + 42 \quad (2)$$

This model resulted in R^2 of 0.82 and 0.72 for the training set and test set, respectively. The RMSE for training and test set for both models are presented in Table 2. The equations were evaluated on the data set published by Kerč and co-workers,[10] and it was found that the RMSE is around 24 K for both eqs 1 and 2. Hence, the equations seem to have general applicability in prediction of organic pharmaceuticals. T_m is a property within reach during early drug development stages and easier to determine than the T_g. The proposed equations will therefore provide early information on T_g, a property of importance for glass forming ability and physical stability of the amorphous state, without the need of transforming the crystalline material to its amorphous counterpart.

Prediction of Glass Transition Temperature from Molecular Descriptors. Although T_m typically can be measured during the early stages of drug development, it would be convenient to predict T_g without the need of experimentally determined properties. We therefore explored at which accuracy calculated molecular descriptors could predict T_g. For this purpose five different algorithms were investigated (Table 3).

Three different linear models were analyzed for their ability to predict T_g based on calculated molecular descriptors. The model derived from stepwise linear regression (Figure 4) included eight descriptors (Table 4). In comparison it was found that nine components were required when PCR was used (Figure 5) whereas only two components were needed in the final PLS (Figure 6) regression model. The results, in terms of R^2 and root-mean-square error (RMSE), of the three methods are summarized in Table 3. Although relatively strong statistics were obtained for the training set of the linear models, none achieved good predictions of the test set revealing that the

dx.doi.org/10.1021/ci5004834 | J. Chem. Inf. Model. 2014, 54, 3396−3403

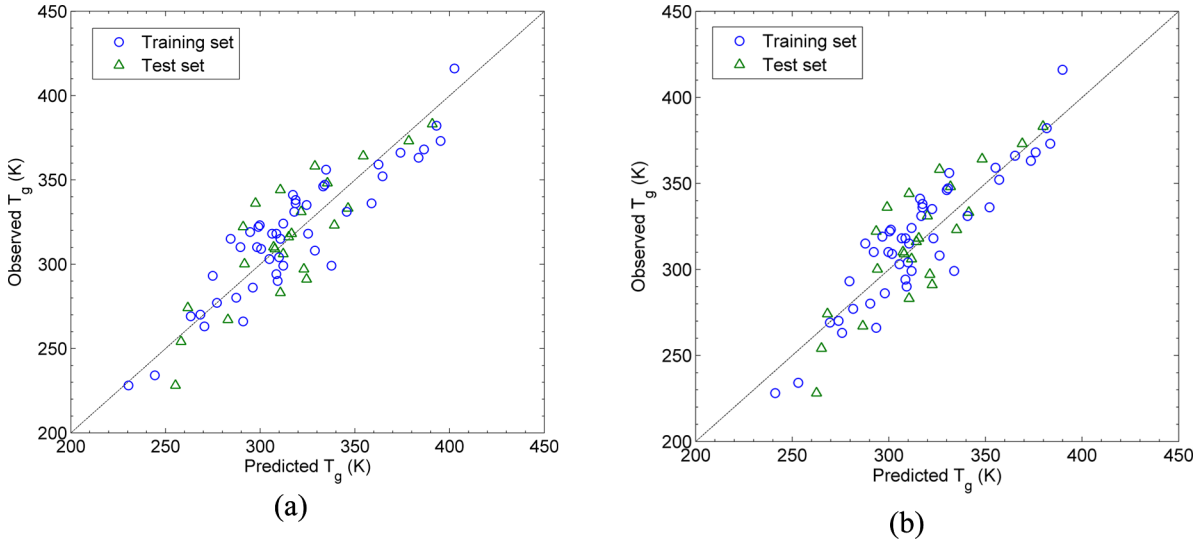


Figure 3. Prediction of T_g from herein derived equations based on T_m. The dashed line indicates a line of unity. (a) Predicted T_g based on eq 1. (b) Predicted T_g based on eq 2.

Table 3. Prediction of T_g from Molecular Descriptors

	R^2		RMSE (K)	
model	training	test	training	test
MLR	0.80	0.38	16.6	32.7
PLS	0.78	0.39	17.3	29.5
PCR	0.69	0.40	20.7	28.9
ANN	0.83	0.61	15.3	26.2
SVR	0.91	0.77	12.2	18.7

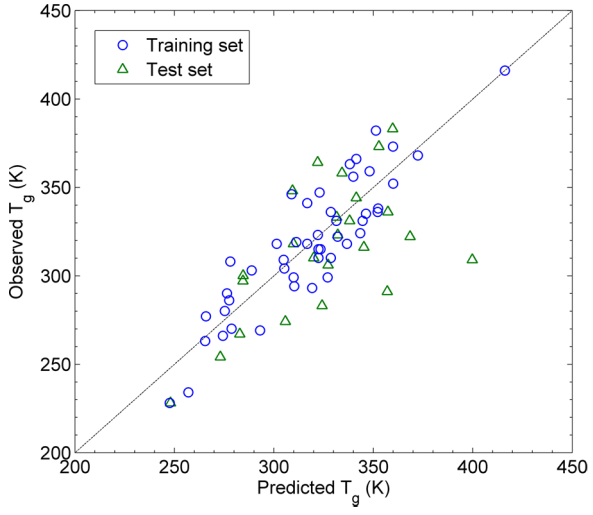


Figure 4. Prediction of T_g using the MLR model derived from molecular descriptors. The dashed line indicates a line of unity.

Table 4. Prediction of T_g from Molecular Descriptors

model	descriptors used in the model
MLR	N_Rings (number of rings)
	SaasN (atom-type E-state index for aaN-groups (e.g., substituted imidazole))
	TerAmine_>N- (number of tertiary amine groups)
	Nitro___-NO2 (number of nitro groups)
	HBD (number of hydrogen bond donors)
	EEM_XFh (maximum sigma Fukui index on H)
	EqualChi (equalized molecular electronegativity)
	FZwitter (portion of FUnion contributed by zwitterionic species (does not depend on pH))[a]
ANN	N_Rings (Number of rings)
	M_RNG (indicator variable for the presence of ring structures except benzene and its condensed rings (aromatic, heteroaromatic, and hydrocarbon rings))
	PolASA__3D (polar solvent accessible surface area in A^2)
	PEoEDIIb3D (proximity effects of electron donors of type II excluding atoms with hydrogens)[b]
	NPA_MinQ (minimal estimated NPA partial atomic charge)
	EqualChi (equalized molecular electronegativity)
SVR	N_Rings (Number of rings)
	N_AlipR (number of aliphatic rings)
	HBD (number of hydrogen bond donors)
	EEM_MaxF (maximum sigma Fukui index)

[a]FUnion: cumulative contribution of all species with zero formal charge to fraction ionized at specified pH (default 7.4). [b]PEoEDIIb3D is a descriptor derived from Seelig's work on P-glycoprotein substrate recognition patterns.[25]

models obtained were not generally applicable. To summarize the different linear regression models showed poor performance in the prediction of T_g of the test set from molecular descriptors. Thus, it was concluded that the molecular descriptors and T_g were not strongly linearly correlated. As a result of the poor predictions using linear algorithms two nonlinear models (ANN and SVR) were developed and analyzed for their applicability for T_g predictions. The final ANN model involved six descriptors (Figure 7 and Table 4) whereas the SVR model took use of four descriptors (Figure 8 and Table 4). The statistics is shown in Table 3. The ANN model achieved a good prediction with R^2 = 0.83 for the training set but similar to the linear models ANN performed poor when applied to the test set. The ANN method was also tested using the first nine principle components. It achieved an RMSE of 17.6 K and R^2 of 0.78 for the training data set, and an RMSE of 27.4 K and R^2 of 0.45 for the test data set. Thus, using the latent variables obtained by principal component analysis did not improve the results of ANN. More encouraging was the performance of the SVR. This model predicted T_g with high accuracy both for the training and the test set (Table 3) and the good performance of the SVR model may be due to the applied

dx.doi.org/10.1021/ci5004834 | J. Chem. Inf. Model. 2014, 54, 3396−3403

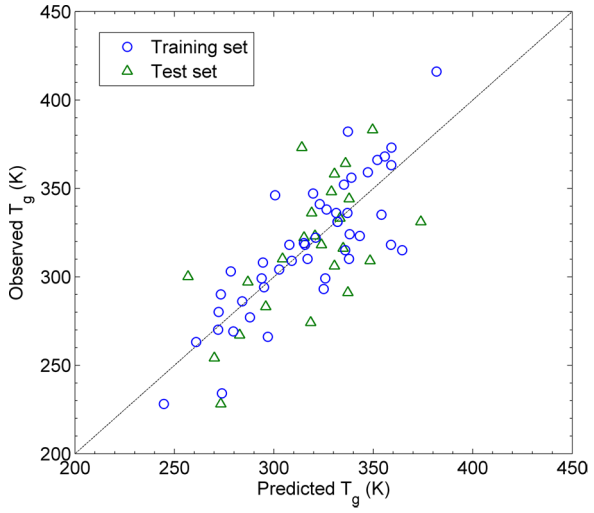


Figure 5. Prediction of T_g using the PCR model derived from molecular descriptors. The dashed line indicates a line of unity.

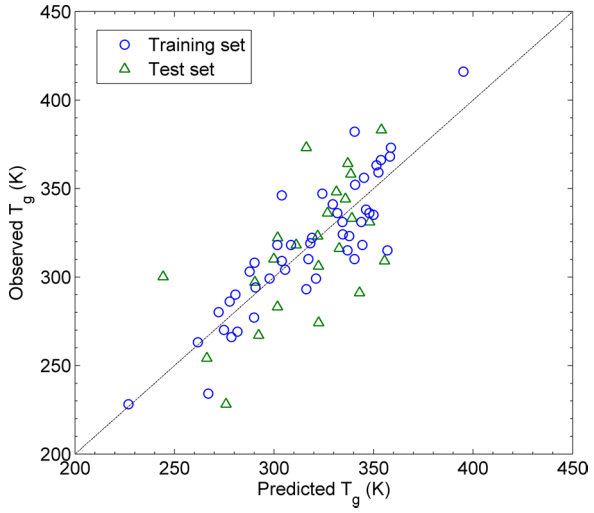


Figure 6. Prediction of T_g using the PLS model derived from molecular descriptors. The dashed line indicates a line of unity.

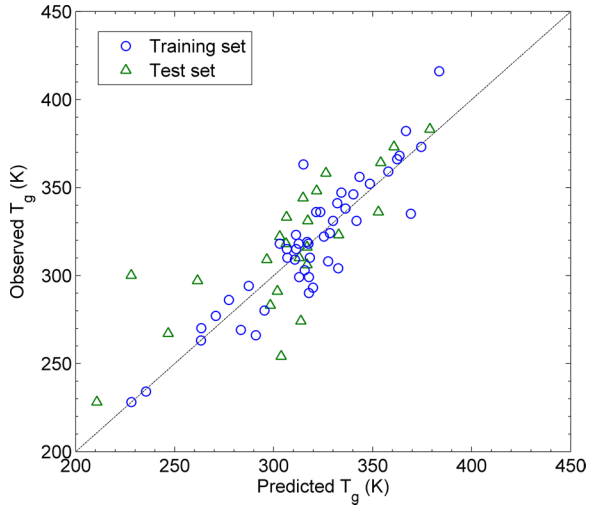


Figure 7. Prediction of T_g using the ANN model derived from molecular descriptors. The dashed line indicates a line of unity.

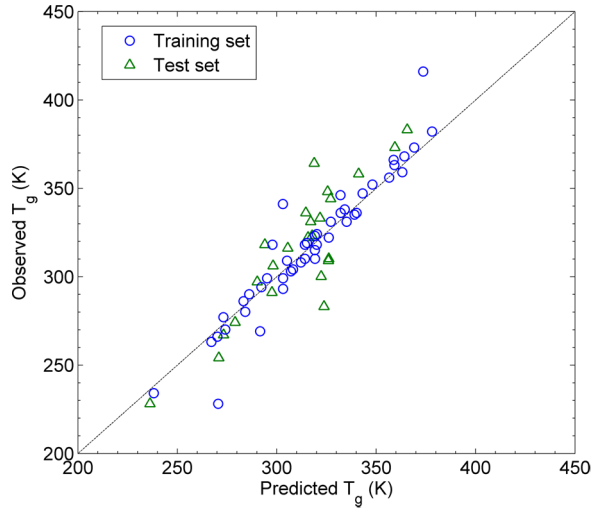


Figure 8. Prediction of T_g using the SVR model derived from molecular descriptors. The dashed line indicates a line of unity.

kernel method and the use of the RFE method. Thus, one could interpret the poor performance of the remaining algorithms as a result of the complex relation between the molecular descriptors and T_g.

The SVR model makes it possible to predict the T_g of new drugs before compound synthesis. This property is important for the crystallization tendency (e.g., during compound synthesis or during storage of the solid material) as well as an indicator of the possibility to take use of amorphous solid formulations as a means to increase dissolution of poorly soluble drugs. To further validate the SVR method, we compiled a literature data set that was used as an ad hoc external test set. It was difficult to find measured T_g for drug-like molecules that were not already explored in the data set established in our laboratory. In addition, the SVR model is applicable to the chemical space of drugs and hence, is not suitable for the prediction of e.g. polymers or excipients for which typically T_g can be found. These limitations resulted in that a smaller data set of 12 compounds were extracted from the literature (Supporting Information, Table S1) and it was found that the SVR model was able to predict these compounds with RMSE of 19.3 K and R^2 of 0.69. The overall accuracy of the SVR model when applied to both test sets (i.e., the 24 test set in Table 1 and the 12 compounds extracted from the literature) was RMSE of 19.0 K of and R^2 of 0.72.

The descriptors found to be predictive for T_g were number of rings (both total number and number of aliphatic rings), number of hydrogen bond donors, and the maximum sigma Fukui index. Interestingly similar descriptors have previously been identified as important descriptors also for glass-forming ability and crystallization tendency[16−19] and here were also shown to capture the kinetic property associated with the transition from rubber to glass in the solid state. Furthermore, the hydrogen bond donors had the highest correlation with T_g and were also identified as an important descriptor for T_g by the stepwise regression and RFE feature selection methods.

CONCLUSION

In this study experimental and computational prediction models of T_g of druglike compounds were developed. It was confirmed that the physical property T_m can be used to predict the T_g with high accuracy. More interestingly the novel in silico

dx.doi.org/10.1021/ci5004834 | J. Chem. Inf. Model. 2014, 54, 3396−3403

model developed herein show that SVR and calculated descriptors can predict T_g with equal accuracy as when an experimentally determined input variable is used. This allows estimations of T_g without the need of having the solid sample available and will hence early inform the drug discovery and development stages about crystallization tendencies, glass-forming ability, and physical stability of amorphous material of new chemical entities.

■ ASSOCIATED CONTENT

Supporting Information

Tables S1 and S2 and Figure S1. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: christel.bergstrom@farmaci.uu.se. Phone: +46-18 471 4118. Fax: +46-18 471 4223.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Financial support to this project from the Swedish Research Council (Grants 621-2008-3777 and 621-2011-2445) is gratefully acknowledged. We are thankful to Elisabeth and Alfred Ahlqvist for the post doc grant to Amjad Alhalaweh. We also thank Simulations Plus (Lancaster, CA) for providing the Drug Delivery group at the Department of Pharmacy, Uppsala University, with a reference site license for the software ADMET Predictor.

■ REFERENCES

(1) Angell, C. A. Formation of glasses from liquids and biopolymers. *Science* **1995**, *267* (5206), 1924−1935.

(2) Yu, L. Amorphous pharmaceutical solids: preparation, characterization and stabilization. *Adv. Drug Delivery Rev.* **2001**, *48* (1), 27−42.

(3) Kauzmann, W. The Nature of the Glassy State and the Behavior of Liquids at Low Temperatures. *Chem. Rev.* **1948**, *43* (2), 219−256.

(4) Hancock, B. C.; Zografi, G. Characteristics and significance of the amorphous state in pharmaceutical systems. *J. Pharm. Sci.* **1997**, *86* (1), 1−12.

(5) Taylor, L. S.; Zografi, G. Spectroscopic characterization of interactions between PVP and indomethacin in amorphous molecular dispersions. *Pharm. Res.* **1997**, *14* (12), 1691−1698.

(6) Kaushal, A. M.; Chakraborti, A. K.; Bansal, A. K. FTIR studies on differential intermolecular association in crystalline and amorphous states of structurally related non-steroidal anti-inflammatory drugs. *Mol. Pharmaceutics* **2008**, *5* (6), 937−945.

(7) Angell, C. Perspective on the glass transition. *J. Phys. Chem. Solids* **1988**, *49* (8), 863−871.

(8) Angell, C.; Monnerie, L.; Torell, L. In *Strong and fragile behavior in liquid polymers*; MRS Proceedings; Cambridge Univ Press, 1990; p 3.

(9) Fukuoka, E.; Makita, M.; Yamamura, S. Glassy state of pharmaceuticals. III. Thermal properties and stability of glassy pharmaceuticals and their binary glass systems. *Chem. Pharm. Bull.* **1989**, *37* (4), 1047−1050.

(10) Kerč, J.; Srcic, S. Thermal analysis of glassy pharmaceuticals. *Theor. Chim. Acta* **1995**, *248*, 81−95.

(11) Liem, H.; Cabanillas-Gonzalez, J.; Etchegoin, P.; Bradley, D. Glass transition temperatures of polymer thin films monitored by Raman scattering. *J. Phys.: Condens. Matter.* **2004**, *16* (6), 721.

(12) Kalichevsky, M.; Jaroszkiewicz, E.; Ablett, S.; Blanshard, J.; Lillford, P. The glass transition of amylopectin measured by DSC, DMTA and NMR. *Carbohydr. Polym.* **1992**, *18* (2), 77−88.

(13) Royall, P. G.; Craig, D. Q.; Doherty, C. Characterisation of the glass transition of an amorphous drug using modulated DSC. *Pharm. Res.* **1998**, *15* (7), 1117−1121.

(14) Preiss, U. P.; Saleh, M. I. An augmented volume-based model of the glass transition temperature of 209 molecular liquids. *J. Pharm. Sci.* **2013**, *102* (6), 1970−1980.

(15) Yeong, S.; Jae, H.; Jung, S. Prediction of glass transition temperature (Tg) of some compounds in organic electroluminescent devices with their molecular properties. *J. Chem. Inf. Comput. Sci.* **2002**, *42* (1), 75−81.

(16) Baird, J. A.; Van Eerdenbrugh, B.; Taylor, L. S. A classification system to assess the crystallization tendency of organic molecules from undercooled melts. *J. Pharm. Sci.* **2010**, *99* (9), 3787−3806.

(17) Mahlin, D.; Bergström, C. A. Early drug development predictions of glass-forming ability and physical stability of drugs. *Eur. J. Pharm. Sci.* **2013**, *49* (2), 323−332.

(18) Mahlin, D.; Ponnambalam, S.; Heidarian Höckerfelt, M.; Bergström, C. A. Toward in silico prediction of glass-forming ability from molecular structure alone: a screening tool in early drug development. *Mol. Pharmaceutics* **2011**, *8* (2), 498−506.

(19) Alhalaweh, A.; Alzghoul, A.; Kaialy, W.; Mahlin, D.; Bergström, C. A. Computational Predictions of Glass-Forming Ability and Crystallization Tendency of Drug Molecules. *Mol. Pharmaceutics* **2014**, *11* (9), 3123−3132.

(20) Pang-Ning, T.; Steinbach, M.; Kumar, V. *Introduction to data mining*; Addison-Wesley, 2006.

(21) Alzghoul, A.; Löfstrand, M.; Backe, B. Data stream forecasting for system fault prediction. *Comput. Ind. Eng.* **2012**, *62* (4), 972−978.

(22) Medsker, L. R. Microcomputer applications of hybrid intelligent systems. *J. Netw.Comput. Appl.* **1996**, *19* (2), 213−234.

(23) Basak, D.; Pal, S.; Patranabis, D. C. Support vector regression. *Neural Information Processing−Letters and Reviews* **2007**, *11* (10), 203−224.

(24) Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene selection for cancer classification using support vector machines. *Machine Learning* **2002**, *46* (1−3), 389−422.

(25) Seelig, A. A general pattern for substrate recognition by P-glycoprotein. *Eur. J. Biochem.* **1998**, *251*, 252−261.