QSPR MODELLING: PREDICTION OF REFRACTIVE INDICES AND GLASS TRANSITION TEMPERATURE OF POLYMERS

Pham Van Tat , Pham Nu Ngoc Han
Department of Chemistry, University of Dalat

The structures of a set of 570 polymers were built with Scipolymer system. The quantitative relationships between the structure and their properties have been estimated by a combination of theoretical descriptors and an ordinary multiple regression (OMR) and the genetic algorithms (GA). The glass transition temperature (Tg) and refractive index (RInd) values of 10 polymers were predicted from charge dependent parameters and 3D dependents parameter and 2D and 3D topological descriptors.

Key words: quantitative structure-property relationship (QSPR), cross-validation, ordinary multiple regression (OMR), genetic algorithms (GA).

INTRODUCTION

The computer has been aided in the quantitative structure-activity relationship (QSAR) modelling for drug design (CADD) with multiple linear regression or artificial neural network. Besides, the computer also has been used for the studies for the quantitative relationships between the structure and properties of the chemicals, in which the QSPR modelling was applied on several polymers.

The various polymer properties include a set of topological substructure searching algorithms and molecular descriptors [1, 2]. The calculations are performed via ordinary multiple regression and genetic algorithms. The underlying theory and methods have been described by J. Bicerano in his monograph [4]. Calculations are performed dynamically.
In this work, we used the calculations of molecular descriptors in QSPR modelling for two dependent parameters of the glass transition temperature (Tg) and refractive index (RInd) with multiple linear regression and genetic algorithms and then QSPR models can be used in prediction of Tg and RInd values for new polymers.

MATERIALS AND METHODS

Data: The source of experimental glass transition temperature (Tg) values of 380 polymers and experimental refractive index (RInd) values of 190 polymers taken from the SciPolymer 3.0 [1].

Software: SciPolymer 3.0 [1], BMDP new System 2.0 [3], Polymer-IS [2] .

Cross-validation: There are several ways to run a cross – validation. One way is to create a new regression by randomly removing 10 – 20% of the polymers [2,6].

Ordinary Multiple Regression: Computing the least squares fit in several independent variables to the dependent variable [2, 3].
The form of the regression equation is $Y = b_o + b_1X_1 + b_2X_2 + ... + b_nX_n$ (1)

Genetic algorithms (GA): This is based on the principles of Darwinian evolution, and has been widely used for combinatorial optimization [2].

RESULTS AND DISCUSSION

1. Glass Transition Temperature

The molecular descriptors were used for creating the statistical model. The Connectivity Simple, Subgraph Count, Valence, and E-State descriptors were chosen for this QSPR and these descriptors were the most importance. In this model, 26 descriptors were selected by the GA and 359 polymers were used. Twenty-one outliers were found and removed from this model.

Regression equation

Tg = 121.8dxvp8 - 12.55dsCH + 3.348SssssC + 19.48SsNH$_2$ - 9.032dNH + 14.23SssNH + 10.98tN + 7.98SaaN + 0.5719dO + 44.47SsssSiH - 36.67dsssP + 2.65SsCl + 7.368SsBr + 1.846nxp8 + 79.83nxch6 + 51.21xp3 - 45.86xp5 - 216xp10 - 37.84xv1 + 22.96xv2 + 59.58xvp5 - 304.3xvp8 + 412.9xvch5 + 18.85dxv1 - 51.13dxvp3 + 50.64dxvp6 + 263.149

Regression quality

Multiple R-Squared = 0.9159, Standard Error of estimation = 33.21, F-statistic = 139.1
P-value = 0, Multiple Q-Squared = 1, Cross validation RSS = 0.
By plotting the Tg_{exp} vs. Tg_{pred}, the statistical fit resulted with a R^2 value of 0.92 was shown in figure 3 and a mean absolute error of 33.21.

2. Refractive Indices

Similarly, the connectivity Simple & Valence, E-State, Total topological and 3D General descriptors were chosen for this QSPR. 20 descriptors were determined by the GA and 182 polymers were used.

Regression equation

RInd = -0.068dxv1 - 0.01999SsCH$_3$ - 0.01677dCH$_2$ - 0.005443SssCH$_2$ + 0.002777SaaCH + 0.005356SaaaC + 0.009528SssNH + 0.002948dO + 0.01176SsSH + 0.09136MaxNeg - 0.1103MaxQp - 6.39E-007Polarizability - 0.009507sumI - 0.002454TTd(4)Valence + 0.001225tets2 + 0.04339x2 + 0.03713xp3 - 0.1317xp10 - 0.3567xch6 + 0.01681xvp3 - 0.01739xvpc4 + 0.03259dxv0 + 1.63296

Regression quality

Multiple R-Squared = 0.9602, Standard error of estimation = 0.01583, F-statistic = 174.2
P-value = 0, Multiple Q-Squared = 1, Cross validation RSS = 0.

By plotting the $RInd_{exp}$ vs. $RInd_{pred}$ the statistical fit resulted with a R^2 value of 0.96 was shown in figure 4 and a mean absolute error of 0.016.

10 polymers outside the training groups are shown in figure 1, and figure 2. The glass transition temperature (Tg) and the refractive index (RInd) values of these polymers were predicted by multiple linear regression and shown in table 1.

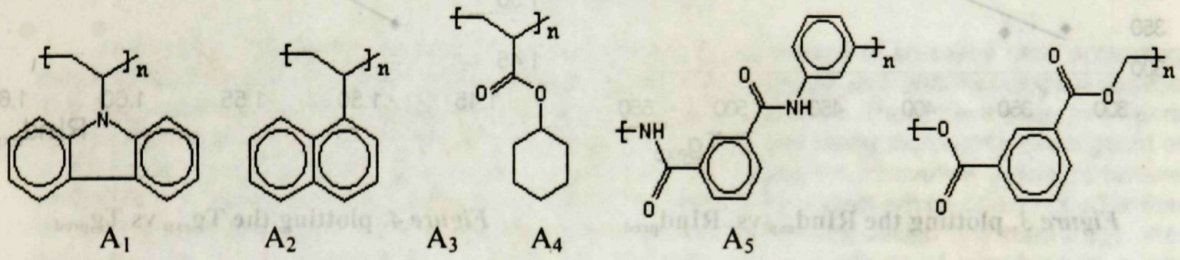

Figure 1. **The polymers were predicted on glass transition temperature (Tg)**

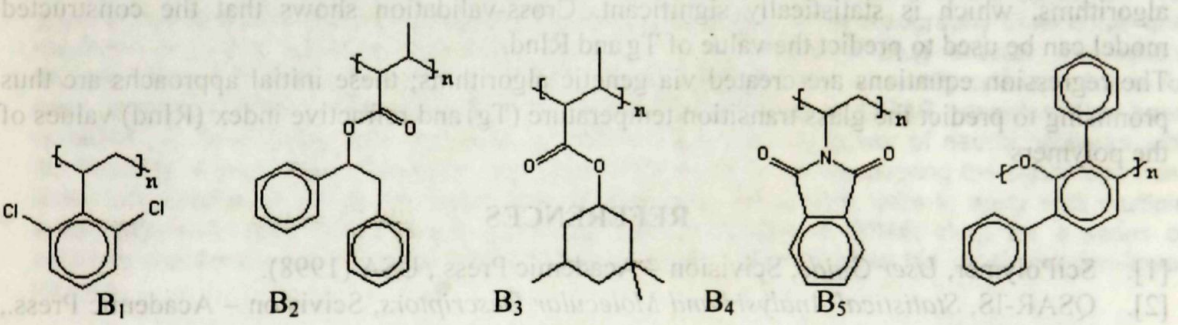

Figure 2. **The polymers were predicted on refractive index (RInd)**

Table 1. **The predicted glass transition temperature (Tg) and refractive index (RInd) values**

No	Tg values (K)			No	RInd values		
	Polymer	Tg_{exp}	Tg_{pred}		Polymer	$RInd_{exp}$	$RInd_{pred}$
1	A_1	423	476.5	1	B_1	1.625	1.630
2	A_2	432	423.1	2	B_2	1.582	1.575
3	A_3	356	350.1	3	B_3	1.485	1.482
4	A_4	545	565.2	4	B_4	1.620	1.635
5	A_5	324	354.6	5	B_5	1.640	1.622

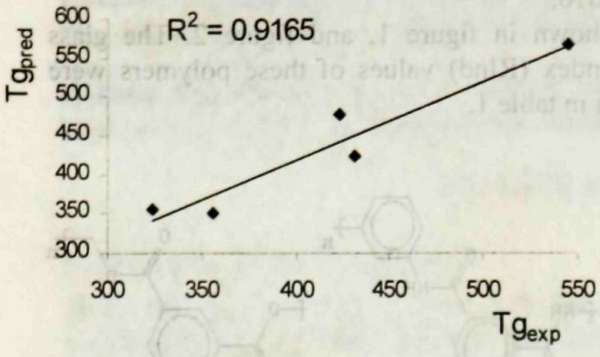

Figure 3. **plotting the $RInd_{exp}$ vs. $RInd_{pred}$**

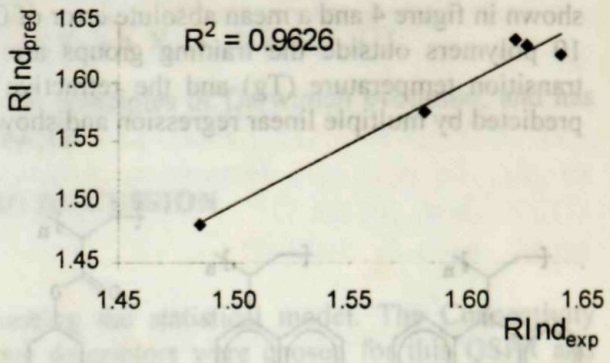

Figure 4. **plotting the Tg_{exp} vs Tg_{pred}**

CONCLUSION

The training set is very well described by the ordinary multiple regression and genetic algorithms, which is statistically significant. Cross-validation shows that the constructed model can be used to predict the value of Tg and RInd.

The regression equations are created via genetic algorithms; these initial approachs are thus promising to predict the glass transition temperature (Tg) and refractive index (RInd) values of the polymers

REFERENCES

[1]. SciPolymer, *User Guide*, Scivision – Academic Press., USA (1998).

[2]. QSAR-IS, *Statistical Analysis and Molecular Descriptors*, Scivision – Academic Press., USA (2000).

[3]. BMDP new system,*Versatile methods for Data Analysis,* Statistical Solutions Ltd., USA (2001).

[4]. J. Bicerano, *Prediction of Polymer Properties*, Marcel Decker Inc., (1993).

[5]. Pham Van Tat, Pham Nu Ngoc Han, *Journal of Chemistry*, Vol. 38, No. 3, P.91, (2000).

[6]. Pham Van Tat, Pham Nu Ngoc Han, *Journal of Chemistry*, Vol. 39, No.1, P.110,(2001).