Prediction of Prop-2-enoate Polymer and Styrene Polymer Glass Transition Using Artificial Neural Networks

G. Astray,*,† A. Cid,† J. A. Ferreiro-Lage,† J. F. Gálvez,‡ J. C. Mejuto,*,† and O. Nieto-Faza§

Department of Physical Chemistry, Faculty of Sciences, University of Vigo, 32004, Ourense, Spain, Department of Informatics, ESEI, University of Vigo, 32004, Ourense, Spain, and Department of Organic Chemistry, Faculty of Sciences, University of Vigo, 32004, Ourense, Spain

This paper was withdrawn on January 11, 2011 (DOI: 10.1021/je200072e).

In this article, the molecular average polarizability α, the energy of the highest occupied molecular orbital E_{HOMO}, the total thermal energy $E_{thermal}$, and the total entropy S were used to correlate with the glass transition temperature T_g for 113 polymers. The quantum chemical descriptors obtained directly from polymer monomers can represent the essential factors that are governing the nature of glass transition in polymers. Stepwise multiple linear regression (MLR) analysis and the back-propagation artificial neural network (ANN) were used to generate the model. The final optimum neural network with 4-[4-4-1]$_3$-1 structure produced a training set root-mean-square error (RMSE) of 11 K ($R = 0.973$) and a validation set RMSE of 17 K ($R = 0.955$). The results show that the ANN model obtained in this paper is accurate in the prediction of T_g values for polymers.

Introduction

The glass transition temperature, T_g, is one of the best-known properties of polymeric materials and composites. Its importance is marked by the ability of processing and use of these materials and is undoubtedly a necessary condition to determine mechanical and thermodynamic properties.[1] T_g can be obtained by different experimental techniques;[2] nevertheless, it is difficult to determine it.[1−3] For this reason, a lot of people have focused their research on predicting T_g for polymers on the basis of quantitative structure−property relationships (QSPRs), both from empirical and theoretical approaches.[4]

Empirical methods are used in the correlation of T_g with other physical or chemical properties of polymers.[5] Bicerano et al.[6] have developed the theoretical reference model, which corresponds with a regression model. This model relates T_g with the solubility parameter and weighted sum of 13 variables for the data set of 320 polymers ($R = 0.9749$, $S = 24.65$ K). Katritzky et al.[7] implemented a model with R^2 of 0.928 for 22 polymers using 4 variables. Subsequently, Katritzky et al.[4] generated by a five-parameter QSPR model 88 linear homopolymers with a standard error of 32.9 K for T_g. Cao and Lin[8] tested the same set of QSPRs obtaining a polymer with a coefficient of determination $R^2 = 0.9056$ and a standard error of 20.86 K.

Artificial neural networks (ANNs) are a complete statistical tool for data analysis[8] which try to reproduce artificially the human ability of making decisions, simulating the human brain's basic unit, the neuron, and the interconnections between the neurons that allow them to work together and save experience's information.[9−11] It is a flexible structure, capable of making a nonlinear mapping between input and output spaces.[12] ANNs were abstract simulations of the biological brain systems, composed by an assembly of units called "neurons" (or "nodes") connected between them. The advantage of ANNs consists of their ability to learn from real cases and relearn when new data are input into the system. They are particularly useful in managing different aspects. In recent years, ANNs have been extended successfully to very different fields, from hydrology to finance.[13−21] The use of ANNs was assayed by Chen et al.[22] and Wanqiang et al.[23] The former had implemented an ANN model with 28 variables, trained with the 65 polymers and tested with 6 polymers. The obtained results show training RMSEs of 17 K ($R^2 = 0.95$) and a validation average error of 17 K ($R^2 = 0.85$). The latter had presented an ANN that predicts the values of T_g of 113 polymers using quantum chemistry variables calculated from the corresponding monomers, but they present more than 30 K of absolute errors in the prediction. In the present paper, a more robust QSPR model has been obtained, and we propose a new ANN, which facilitates the calculation of T_g.

Materials and Methods

Data Set. T_g values, taken from the literature,[6,23,24] of 113 polyprop-2-enoates (polyacrylates) and polystyrenes are listed in Table 1, have large T_g values (198 K ∼ 389 K), and are characterized by a high degree of structural variety. The polymers in Table 1 are divided among the 58 polymers used to train the ANN and 55 used to verify the proper operation thereof.

Computational Methods: Quantum Properties Determination. Density functional theory (DFT)[25,26] has been used to optimize the geometry of the parent monomers and calculate their electronic and thermal properties. All the calculations have been performed with the Gaussian 03 suite of programs[27] using Becke's three-parameter exchange functional[28] and the nonlocal correlation functional of Lee, Yang, and Parr[29] (B3LYP) and Pople's 6-31G(d,p) basis set, which includes a polarization function for every atom.

All minima on the potential energy surface were characterized by harmonic analysis, and the computed frequencies were used to obtain zero-point energies and thermodynamic parameters,

* Corresponding author. E-mail: xmejuto@uvigo.es.
† Department of Physical Chemistry.
‡ Department of Informatics.
§ Department of Organic Chemistry.

10.1021/je100573n
Published on Web 08/18/2010

Table 1. T_g Values for Each Polymer

training data		validation data	
polymer	T_g/K	polymer	T_g/K
poly(3-methoxypropyl prop-2-enoate)	198	poly(4-thiahexyl prop-2-enoate)	197
poly(3-thiapentyl prop-2-enoate)	202	poly(4-thiapentyl prop-2-enoate)	208
poly(5-thiahexyl prop-2-enoate)	203	poly(heptyl prop-2-enoate)	213
poly(1*H*,1*H*-tridecafluoro-4-oxaoctyl prop-2-enoate)	205	poly(6-(2-cyanoethylthio)hexyl prop-2-enoate)	214
poly(3-thiabutyl prop-2-enoate)	213	poly(nonyl prop-2-enoate)	215
poly(3-(2-cyanoethylthio)propyl prop-2-enoate)	215	poly(pentyl prop-2-enoate)	216
poly(hexyl prop-2-enoate)	216	poly(2-ethoxyethyl prop-2-enoate)	223
poly(3-methoxybutyl prop-2-enoate)	217	poly(2-ethylbutyl prop-2-enoate)	223
poly(3-ethoxypropyl prop-2-enoate)	218	poly(2-methoxyethyl prop-2-enoate)	223
poly(2,2,3,3,5,5,5-heptafluoro-4-oxapentyl prop-2-enoate)	218	poly(5,5,6,6,7,7,7-heptafluoro-3-oxaheptyl prop-2-enoate)	228
poly(butyl prop-2-enoate)	219	poly(4-cyanobutyl prop-2-enoate)	233
poly(5-cyano-3-thiapentyl prop-2-enoate)	223	poly(7,7,8,8-tetrafluoro-3,6-dioxaoctyl prop-2-enoate)	233
poly(1*H*,1*H*-nonafluoro-4-oxahexyl prop-2-enoate)	224	poly(5-cyano-3-oxapentyl prop-2-enoate)	250
poly(3-methylbutyl prop-2-enoate)	228	poly(1*H*,1*H*,3*H*-hexafluorobutyl prop-2-enoate)	251
poly(1*H*,1*H*-undecafluorohexyl prop-2-enoate)	234	poly(propan-2-yl prop-2-enoate)	270
poly(2-heptyl prop-2-enoate)	235	poly(methyl prop-2-enoate)	283
poly(2-methylpentyl prop-2-enoate)	235	poly(4-ethoxyl-carbonyl-phenyl prop-2-enoate)	310
poly(5,5,5-trifluoro-3-oxapentyl prop-2-enoate)	235	poly(*p*-totyl acrylate prop-2-enoate)	316
poly(1*H*,1*H*-nonafluoropentyl prop-2-enoate)	236	poly(2-hexyloxycarbonylstyrene)	318
poly(propyl prop-2-enoate)	236	poly(2- methoxycarbonylphenyl prop-2-enoate)	319
poly(1*H*,1*H*,5*H*-octafluoropentyl prop-2-enoate)	238	poly[4-(2-hydroxybutoxymethyl)styrene]	319
poly(2-methylbutyl prop-2-enoate)	241	poly(3-dimethylaminophenyl prop-2-enoate)	320
poly(1*H*,1*H*-heptafluorobutyl prop-2-enoate)	243	poly(4-octanoylstyrene)	323
poly(1*H*,1*H*-pentafluoropropyl prop-2-enoate)	247	poly(phenyl prop-2-enoate)	330
poly(ethyl prop-2-enoate)	249	poly(2,4-dichlorophenyl prop-2-enoate)	333
poly(2-methylpropyl prop-2-enoate)	249	poly(2-butoxycarbonylstyrene)	339
poly(butan-2-yl prop-2-enoate)	250	poly(4-hexanoylstyrene)	339
poly(4,4,5,5-tetrafluoro-3-oxapentyl prop-2-enoate)	251	poly(4-hexyloxycarbonylstyrene)	339
poly(4-methylpentan-2-yl prop-2-enoate)	258	poly(2-isopentyloxycarbonylstyrene)	341
poly(2,2,2trifluoroethyl prop-2-enoate)	263	poly(4-propoxystyrene)	343
poly(3-pentyl prop-2-enoate)	267	poly(pentanoicstyrene)	343
poly(dodecyl prop-2-enoate)	270	poly(4-butyrylstyrene)	347
poly(2-phenylethyl prop-2-enoate)	270	poly(2-ethoxymethylstyrene)	347
poly(2-cyanoethyl prop-2-enoate)	277	poly(2-methoxystyrene)	348
poly(benzyl acrylate prop-2-enoate)	279	poly(4-butoxycarbonylstyrene)	349
poly(heptafluoro-2-propyl prop-2-enoate)	283	poly(4-methoxymethylstyrene)	350
poly(*p*-carbobutoxyphenyl prop-2-enoate)	286	poly(2-isopentyloxymethylstyrene)	351
poly(fluoromethyl prop-2-enoate)	288	poly(4-phenylacetylstyrene)	351
poly(3-ethoxyl carbonyl phenyl prop-2-enoate)	297	poly(4-methoxy-2-methylstyrene)	358
Poly(tetradecyl prop-2-enoate)	297	poly(4-*s*-butylstyrene)	359
poly(*m*-tolyl prop-2-enoate)	298	poly(4-ethoxymethylstyrene)	359
poly(2-ethoxyl carbonyl phenyl prop-2-enoate)	303	poly(5-*tert*-butyl-2-methystyrene)	360
poly(hexadecyl prop-2-enoate)	308	poly(2-isopropoxymethylstyrene)	361
poly(3-methoxycarbonylphenyl prop-2-enoate)	311	poly(4-isobutoxycarbonylstyrene)	363
poly(4-cyanophenyl acrylate prop-2-enoate)	317	poly[4-(1-hydroxy-1-methylhexyl)styrene]	364
poly(2-isobutyl cyano prop-2-enoate)	324	poly(4-propoxycarbonylstyrene)	365
poly(4-methoxyphenyl prop-2-enoate)	324	poly(4-ethoxycarbonylstyrene)	367
poly(*o*-tolyl prop-2-enoate)	325	poly(4-isopropoxycarbonylstyrene)	368
poly(2-chlorophenyl prop-2-enoate)	326	poly(4-benzoylstyrene)	371
poly(4-chlorophenyl prop-2-enoate)	331	poly(4-*p*-toluoylostyrene)	372
poly(2-cyanoisopropyl prop-2-enoate)	339	poly(4-phenoxystyrene)	373
poly(4-methoxycarbonylphenyl prop-2-enoate)	340	poly(4-diethylcarbamoylstyrene)	375
poly(4-*tert*-butylphenyl prop-2-enoate)	344	poly(4-*p*-anisoylstyrene)	376
poly(2-*tert*-butylphenyl prop-2-enoate)	345	poly(4-[(1-hydroxyimino)-2-phenethyl]styrene)	384
poly(2-cyanohexyl prop-2-enoate)	358	poly(2-cyanoheptyl prop-2-enoate)	389
poly(2-naphthalen-2-yl prop-2-enoate)	358		
poly(4-cyanophenyl prop-2-enoate)	363		
poly(4-biphenyl prop-2-enoate)	383		

applying the free particle, harmonic oscillator, and rigid rotor approximations at the high-temperature limit in a canonical ensemble ($T = 298.15$ K, $p = 1$ atm). Frequency values were uncorrected, based on Scott and Radom estimation of correction coefficients very close to unity for calculations using B3LYP/6-31G(d).[30]

In Table 2, we show the values of the molecular average polarizability α, the energy of the highest occupied molecular orbital E_h, the total thermal energy E, and the total entropy S.

Computational Methods: Artificial Neural Network. For the implementation of ANNs, we used a commercial software package provided by Neural Planner Software Ltd. All of the component parts are implemented as C++ reusable classes to simplify future development. We will pursue a perceptron neural network (Figure 1) which could be described as follows: each neuron from the primary layer collects the data of the input variables and presents them according to an input vector (eq 1), which spreads toward the intermediate layer by means of the following propagation rule (eq 2)

$$x^p = (x_1^p, x_2^p, ..., x_N^p)^T \quad (1)$$

Table 2. Quantum Chemical Descriptors for Each Polymer Obtained by DFT Calculations

polymer	α / $C^2 \cdot m^2 \cdot J^{-1} \cdot 10^{-39}$	E_h / $J \cdot 10^{-18}$	E / $kJ \cdot mol^{-1}$	S / $J \cdot mol^{-1} \cdot K^{-1}$
Training Data				
poly(3-methoxypropyl prop-2-enoate)	1.423	−1.124	518.088	459.039
poly(3-thiapentyl prop-2-enoate)	1.657	−0.974	510.473	482.817
poly(5-thiahexyl prop-2-enoate)	1.828	−0.956	588.685	511.264
poly(1*H*,1*H*-tridecafluoro-4-oxaoctyl prop-2-enoate)	1.972	−1.274	503.603	771.040
poly(3-thiabutyl prop-2-enoate)	1.455	−0.983	431.513	449.654
poly(3-(2-cyanoethylthio)propyl prop-2-enoate)	2.056	−1.046	590.827	559.208
poly(hexyl prop-2-enoate)	1.725	−1.173	659.947	491.382
poly(3-methoxybutyl prop-2-enoate)	1.597	−1.111	595.199	486.331
poly(3-ethoxypropyl prop-2-enoate)	1.613	−1.115	596.157	492.055
poly(2,2,3,3,5,5,5-heptafluoro-4-oxapentyl prop-2-enoate)	1.392	−1.264	384.664	574.233
poly(butyl prop-2-enoate)	1.352	−1.174	503.084	429.831
poly(5-cyano-3-thiapentyl prop-2-enoate)	1.846	−1.042	511.984	526.234
poly(1*H*,1*H*-nonafluoro-4-oxahexyl prop-2-enoate)	1.575	−1.272	424.705	634.144
poly(3-methylbutyl prop-2-enoate)	1.529	−1.174	580.388	455.930
poly(1*H*,1*H*-undecafluorohexyl prop-2-enoate)	1.723	−1.274	449.483	668.256
poly(2-heptyl prop-2-enoate)	1.902	−1.168	736.853	517.184
poly(2-methylpentyl prop-2-enoate)	1.714	−1.176	658.645	491.875
poly(5,5,5-trifluoro-3-oxapentyl prop-2-enoate)	1.422	−1.222	462.395	516.950
poly(1*H*,1*H*-nonafluoropentyl prop-2-enoate)	1.506	−1.266	410.317	606.387
poly(propyl prop-2-enoate)	1.166	−1.176	424.588	398.790
poly(1*H*,1*H*,5*H*-octafluoropentyl prop-2-enoate)	1.510	−1.266	431.107	589.697
poly(2-methylbutyl prop-2-enoate)	1.526	−1.177	580.438	457.144
poly(1*H*,1*H*-heptafluorobutyl prop-2-enoate)	1.328	−1.263	370.569	553.949
poly(1*H*,1*H*-pentafluoropropyl prop-2-enoate)	1.165	−1.271	328.013	461.035
poly(ethyl prop-2-enoate)	0.980	−1.178	346.059	366.179
poly(2-methylpropyl prop-2-enoate)	1.343	−1.178	501.703	425.797
poly(butan-2-yl prop-2-enoate)	1.342	−1.171	501.691	421.948
poly(4,4,5,5-tetrafluoro-3-oxapentyl prop-2-enoate)	1.415	−1.242	444.671	535.113
poly(4-methylpentan-2-yl prop-2-enoate)	1.703	−1.167	657.382	477.361
poly(2,2,2-trifluoroethyl prop-2-enoate)	1.160	−1.229	369.623	453.901
poly(3-pentyl prop-2-enoate)	1.514	−1.169	580.488	452.336
poly(dodecyl prop-2-enoate)	2.846	−1.173	1130.617	675.515
poly(2-phenylethyl prop-2-enoate)	1.919	−1.071	571.706	479.951
poly(2-cyanoethyl prop-2-enoate)	1.176	−1.257	347.628	410.522
poly(benzyl acrylate prop-2-enoate)	1.732	−1.086	492.976	455.742
poly(heptafluoro-2-propyl prop-2-enoate)	1.149	−1.312	290.441	504.431
poly(*p*-carbobutoxyphenyl prop-2-enoate)	2.719	−1.078	772.609	612.031
poly(fluoromethyl prop-2-enoate)	0.794	−1.246	250.835	363.937
poly(3-ethoxyl carbonyl phenyl prop-2-enoate)	2.280	−1.081	615.617	542.799
poly(tetradecyl prop-2-enoate)	3.221	−1.172	1287.505	736.392
poly(*m*-tolyl prop-2-enoate)	1.816	−1.024	489.900	452.332
poly(2-ethoxyl carbonyl phenyl prop-2-enoate)	2.211	−1.112	615.274	540.640
poly(hexadecyl prop-2-enoate)	3.576	−1.177	1443.903	799.813
poly(3-methoxycarbonylphenyl prop-2-enoate)	2.092	−1.084	537.661	513.193
poly(4-cyanophenyl acrylate prop-2-enoate)	2.035	−1.151	493.850	491.130
poly(2-isobutyl cyano prop-2-enoate)	1.531	−1.248	502.093	466.190
poly(4-methoxyphenyl prop-2-enoate)	1.890	−0.930	505.348	468.315
poly(*o*-tolyl prop-2-enoate)	1.803	−1.022	490.168	444.529
poly(2-chlorophenyl prop-2-enoate)	1.745	−1.068	391.074	441.052
poly(4-chlorophenyl prop-2-enoate)	1.791	−1.047	390.957	442.554
poly(2-cyanoisopropyl prop-2-enoate)	1.350	−1.251	422.048	422.203
poly(4-methoxycarbonylphenyl prop-2-enoate)	2.139	−1.084	537.728	514.331
poly(4-*tert*-butylphenyl prop-2-enoate)	2.362	−1.004	723.125	522.544
poly(2-*tert*-butylphenyl prop-2-enoate)	2.288	−1.029	723.481	512.891
poly(2-cyanohexyl prop-2-enoate)	1.904	−1.247	660.700	533.301
poly(2-naphthalen-2-yl prop-2-enoate)	2.373	−0.940	542.648	466.278
poly(4-cyanophenyl prop-2-enoate)	1.898	−1.120	414.032	452.341
poly(4- biphenyl prop-2-enoate)	2.758	−0.960	637.847	515.640
Validation Data				
poly(4-thiahexyl prop-2-enoate)	1.843	−0.957	589.061	514.264
poly(4-thiapentyl prop-2-enoate)	1.640	−0.966	510.071	482.503
poly(heptyl prop-2-enoate)	1.911	−1.173	738.384	523.230
poly(6-(2-cyanoethylthio)hexyl prop-2-enoate)	2.421	−1.057	747.618	613.659
poly(nonyl prop-2-enoate)	2.285	−1.173	895.259	584.204
poly(pentyl prop-2-enoate)	1.539	−1.174	581.480	461.483
poly(2-ethoxyethyl prop-2-enoate)	1.426	−1.131	517.582	461.587
poly(2-ethylbutyl prop-2-enoate)	1.696	−1.176	659.603	483.327
poly(2-methoxyethyl prop-2-enoate)	1.236	−1.141	439.525	428.684
poly(5,5,6,6,7,7,7-heptafluoro-3-oxaheptyl prop-2-enoate)	1.765	−1.219	542.464	643.072
poly(4-cyanobutyl prop-2-enoate)	1.550	−1.216	505.034	472.809
poly(7,7,8,8-tetrafluoro-3,6-dioxaoctyl prop-2-enoate)	1.856	−1.207	616.458	628.625
poly(5-cyano-3-oxapentyl prop-2-enoate)	1.621	−1.227	519.448	506.908
poly(1*H*,1*H*,3*H*-hexafluorobutyl prop-2-enoate)	1.312	−1.263	390.610	519.439

Table 2. Continued

polymer	α / $C^2 \cdot m^2 \cdot J^{-1} \cdot 10^{-39}$	E_h / $J \cdot 10^{-18}$	E / $kJ \cdot mol^{-1}$	S / $J \cdot mol^{-1} \cdot K^{-1}$
poly(propan-2-yl prop-2-enoate)	1.160	−1.173	423.015	390.576
poly(methyl prop-2-enoate)	0.788	−1.190	268.220	334.143
poly(4-ethoxyl-carbonyl-phenyl prop-2-enoate)	2.332	−1.081	615.701	543.150
poly(*p*-totyl acrylate prop-2-enoate)	1.845	−1.002	489.934	456.395
poly(2-hexyloxycarbonylstyrene)	2.683	−1.031	883.594	596.115
poly(2- methoxycarbonylphenyl prop-2-enoate)	2.030	−1.106	537.385	508.415
poly[4-(2-hydroxybutoxymethyl)styrene]	2.461	−0.947	788.617	556.853
poly(3-dimethylaminophenyl prop-2-enoate)	2.117	−0.872	615.809	492.817
poly(4-octanoylstyrene)	2.997	−1.016	946.266	612.433
poly(phenyl prop-2-enoate)	1.573	−1.046	413.091	412.480
poly(2,4-dichlorophenyl prop-2-enoate)	1.970	−1.078	368.794	471.667
poly(2-butoxycarbonylstyrene)	2.420	−1.001	727.949	533.473
poly(4-hexanoylstyrene)	2.644	−1.016	789.412	549.924
poly(4-hexyloxycarbonylstyrene)	2.893	−1.009	881.857	583.208
poly(2-isopentyloxycarbonylstyrene)	2.483	−1.032	804.537	549.928
poly(4-propoxystyrene)	2.084	−0.876	616.968	464.144
poly(pentanoicstyrene)	2.448	−1.016	711.016	519.226
poly(4-butyrylstyrene)	2.257	−1.017	632.566	487.721
poly(2-ethoxymethylstyrene)	1.944	−0.978	617.885	449.161
poly(2-methoxystyrene)	1.632	−0.903	460.976	394.078
poly(4-butoxycarbonylstyrene)	2.519	−1.009	727.769	542.979
poly(4-methoxymethylstyrene)	1.863	−0.941	538.598	441.052
poly(2-isopentyloxymethylstyrene)	2.515	−0.975	852.151	538.824
poly(4-phenylacetylstyrene)	2.907	−1.022	701.176	531.799
poly(4-methoxy-2-methylstyrene)	1.876	−0.878	538.564	427.931
poly(4-*s*-butylstyrene)	2.167	−0.938	680.109	468.261
poly(4-ethoxymethylstyrene)	1.890	−0.878	538.623	431.513
poly(5-*tert*-butyl-2-methystyrene)	2.299	−0.940	755.325	479.248
poly(2-isopropoxymethylstyrene)	2.121	−0.974	694.753	476.495
poly(4-isobutoxycarbonylstyrene)	2.463	−1.025	725.744	538.498
poly[4-(1-hydroxy-1-methylhexyl)styrene]	2.795	−0.931	928.559	576.919
poly(4-propoxycarbonylstyrene)	2.267	−1.024	648.533	515.607
poly(4-ethoxycarbonylstyrene)	2.133	−1.011	570.953	474.930
poly(4-isopropoxycarbonylstyrene)	2.319	−1.009	647.880	500.691
poly(4-benzoylstyrene)	2.762	−1.010	624.182	488.277
poly(4-*p*-toluoylostyrene)	3.016	−1.002	701.063	543.033
poly(4-phenoxystyrene)	2.560	−0.904	605.563	478.880
poly(4-diethylcarbamoylstyrene)	2.554	−0.975	760.262	526.418
poly(4-*p*-anisoylstyrene)	3.108	−0.990	716.619	541.757
poly(4-[(1-hydroxyimino)-2-phenethyl]styrene)	3.073	−0.948	747.898	560.572
poly(2-cyanoheptyl prop-2-enoate)	2.089	−1.247	739.041	563.606

$$s_i^p = \sum_{j=1}^{N} w_{ji} x_j^p + b_i \qquad (2)$$

where N is the number of the network input neurons; w_{ji} is the weight value of the connection between the neuron j from the input layer and the neuron i from the intermediate layer; and b_i is the value of the "bias" associated to the neuron i. If it is assumed that the activation state of the neuron i is the function of the network input vector, then the network output derives from eq 3

$$y_i^p = F_i(s_i^p) \qquad (3)$$

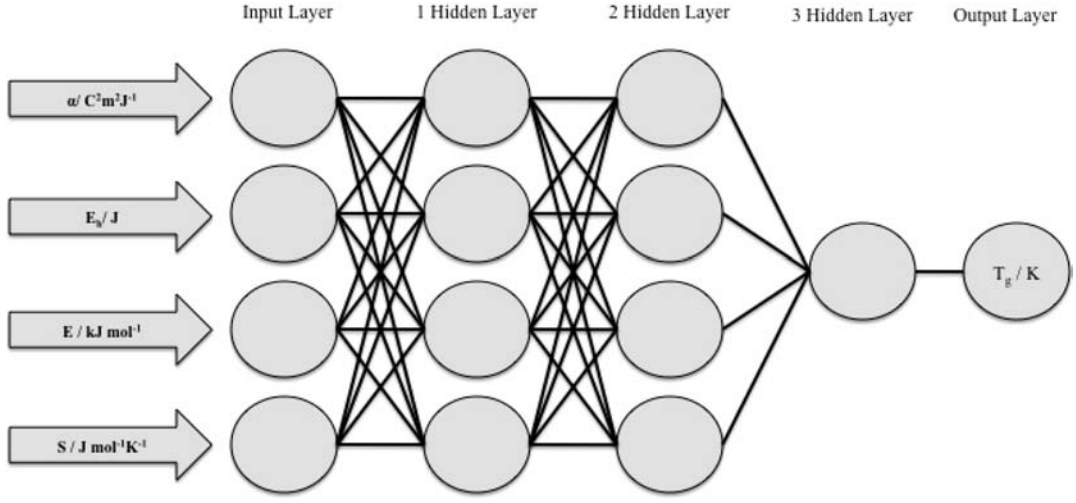

Figure 1. Diagram of a perceptron network constituted by four neurons in the input layer, three hidden layers with four, four, and one neurons, and one output neuron.

Similarly, for any neuron k from the output layer, the equations that determine its activation state are

$$s_k^p = \sum_{i=1}^{L} w_{ik} y_i^p + b_k \quad (4)$$

$$y_k^p = F_k(s_k^p) \quad (5)$$

where L is the neuron number of the intermediate layer; w_{ik} is the weight value of the connection between the neuron i from the intermediate layer and the neuron k from the output layer; and bk is the value of the "bias" associated with the neuron k. The term of error for the output neuron is calculated by means of the following equation

$$E^p = \frac{1}{2} \sum_{k=1}^{M} (d_k^p - y_k^p)^2 \quad (6)$$

and if it adjusts to the previously established value, the training of the neural network finishes here. On the contrary, if it does not adjust to the previously established margins of error, the process would be repeated until reaching the desired error value. The activation equation used in this article is the sigmoid or logistic.

$$F_k(s_k^p) = \frac{1}{1 + e^{-s_k^p}} \quad (7)$$

A back-propagation rule (BP), which is a typical gradient-based learning algorithm, was used as a learning rule in the present work.

$$\Delta_p w_{ik} = -\eta \frac{\partial E^p}{\partial w_{ik}} = \eta (d_k^p - y_k^p) F_k'(s_k^p) y_i^p = \eta \delta_k^p y_i^p \quad (8)$$

This learning rule presents an important limitation, which is the large number of input cases for the training process, but for our study, a large amount of cases were available. This application could be interpreted as an artificial neural network constituted by a primary neural layer (where the data of the input variables would be collected), an output neural layer (where the collected value would be obtained), and one or various intermediate layers (where the convergence work of the neural network would be facilitated) (see Figure 1).

As usual, the architecture of ANN is denoted as the following code

$$N_{in} - [N_{h1} - N_{h2}]_e - N_{out}$$

where N_{in} and N_{out} are the number of neurons in the input layer and output layer, respectively; N_{h1} and N_{h2} are the numbers of neurons in the first and second intermediate layer, respectively; and e is the number of hidden layers.

Results and Discussion

The four variables were used like parameters of the information entrance. The ANNs were trained with data corresponding to 58 polymers. The number of neurons in the intermediate layer was tested between $n/2 + 1$ and $2n + 1$, where n corresponds with the input variables. Once trained, the ANN correct functioning has been tested with the validation data corresponding to 55 polymers.

Table 3. Root Mean Square Errors (RMSE) of Training and Validation Using Different ANN Architectures

		$RMSE_T$	$RMSE_V$	E_{sum}
no.	architectures	K	K	K
1	4-[8-4]$_2$-1	5	30	35
2	4-[6-4]$_2$-1	3	26	29
3	4-[4]$_2$-1	9	21	30
4	4-[4-3]$_2$-1	5	30	35
5	4-[4-2]$_2$-1	7	23	31
6	4-[3]$_2$-1	13	19	32
7	4-[3-2]$_2$-1	8	21	29
8	4-[5]$_2$-1	9	21	30
9	4-[7]$_2$-1	7	25	32
10	4-[9]$_2$-1	5	30	35
11	4-[3-3]$_2$-1	8	20	28
12	4-[4-4-1]$_3$-1	9	17	26

In Table 3 are shown the training RMSE value ($RMSE_T$) and the validation RMSE value ($RMSE_V$) for each ANN architecture. To evaluate the accuracy of the ANN model, we use the sum of root-mean-square errors (E_{sum}) of the training set ($RMSE_T$) and the validation set ($RMSE_V$). The value of E_{sum} (see Table 3) can be expressed as $E_{sum} = RMSE_T + RMSE_V$. As we can see in Table 3, the best ANN architecture consists of four input neurons [(i) the average polarizability (α), (ii) the energy of the highest occupied molecular orbital (E_h), (iii) the total thermal energy (E), and (iv) the total entropy (S)], three middle layers (with four, four, and one neurons, respectively), and one output neuron (see Figure 1). The training of this ANN has established a target error of 0.01 %; the maximum number of training cycles was established as 7000; the learning rate was set at 0.60; and the momentum value was 0.8.

Root mean square errors (RMSEs) are 9 K ($R = 0.98$) for the training set and 17 K ($R = 0.96$) for the prediction set (Figure 2). In comparison with previous models,[4,6,22,24] the present ANN model shows better results.

Our results indicate that ANN developed from the training set of polyacrylates could make a prediction for polyacrylates and polystyrenes. By convention, the structures of the prediction set cover the range of the structures of the training set. However, in this paper, the unusual principles for separation into training sets and test sets are carried out, and the results demonstrate the ANN model could be extrapolated.

Table 4 shows the importance of the variables selected for the ANN, and this value is the sum of weights of the input neurons to all intermediate neurons.

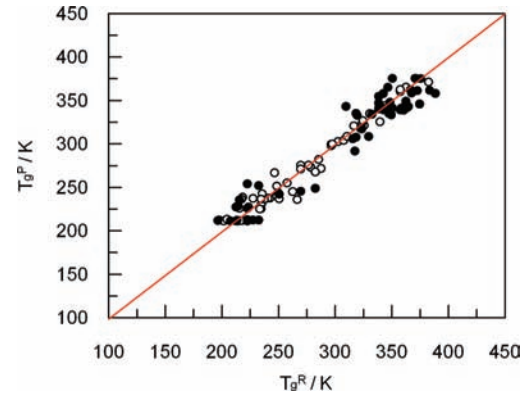

Figure 2. Plot of experimental T_g^R (K) versus calculated T_g^P (K) for ○, training values and ●, validation values.

Table 4. Weight Matrix of the ANN

−19.12	13.74	11.05	15.67
−6.72	4.76	−4.96	−13.54
26.89	4.28	3.38	−22.36
−12.26	15.18	−18.40	20.00
4.66	7.66	10.94	−21.68
17.58	2.15	−2.08	−5.48
−9.66	−0.78	22.43	−0.11
−11.56	−0.72	22.95	−3.57
−2.19	2.55	−1.51	−8.08
−5.78			

Table 5. Importance of Variables Considered for ANN

value	importance	value	importance
S	71.57	E_h	37.96
α	64.99	E	37.79

Table 6. Adjustment for the Validation Values after Training and Validation of ANN, *R*, and RMSE

value	R	RMSE
training	0.9822	9
validation	0.9602	17

Table 5 shows the importance of the variables selected for the ANN, and this value is the sum of weights of the input neurons to all intermediate neurons.

Once the good functioning of the ANN was verified, the previously reserved validation data corresponding to 55 polymers were used to check it. As shown in Figure 2, the data predicted by the ANN were confronted with the previously reserved validation data. The correlation coefficient was 0.96.

Conclusions

A QSPR model has been implemented to predict the values of T_g of 113 polymers. Four variables of quantum chemistry were used, α, E_h, E, and S, obtained from the monomer structures in the theory of the density function (DFT).

Of all the ANNs developed for this study, the one consisting of an input layer with four neurons, three intermediate layers formed by four, four, and one neurons, respectively, and one neuron in the output layer showed the best results obtained (Table 6). T_g values predicted on training cases had a RMSE of 9 K ($R = 0.98$) and on validation cases a RMSE of 17 K ($R = 0.96$).

We must quote that both acrylate polymers and styrene polymers do not contain chemically complicated functional groups. This means that more simple methods for prediction would be used. In the literature, there are a large amount of studies of T_g estimations using QSPR methods.[1,6,7] Our result yields more accurate values of T_g than the QSPR methods, and they present the advantage of a lower number of variables for the prediction. In fact, a QSPR method with 13 variables and 320 polymers yields an *R*-value of 0.97[6] and a RMSE of 24.6 K (which is bigger than the value of 17 found in our case).

QSPR methods with fewer variables imply a significant decrease of the *R*-value. A model with 4 variables and 22 polymers yields an *R*-value of 0.93.[7]

The results encourage the use of this type of modeling for the determination of T_g in other polymers.

Literature Cited

(1) Bicerano, J. *Encyclopedia of polymer science and technology;* Wiley: New York, 2003.

(2) Tracht, U.; Wilhelm, M.; Heuer, A.; Feng, H.; Schmidt-Rohr, K.; Spiess, H. W. Length scale of dynamic heterogeneities at the glass transition determined by multidimensional nuclear magnetic resonance. *Phys. Rev. Lett.* **1998**, *81*, 2727–2730.

(3) Anderson, P. W. Interlayer tunneling mechanism for high-Tc superconductivity: Comparison with c axis infrared experiments. *Science* **1995**, *267*, 1154–1155.

(4) Katritzky, A. R.; Sild, S.; Lobanov, V.; Karlson, M. Quantitative Structure - Property Relationship (QSPR) correlation of glass transition temperatures of high molecular weight polymers. *J. Chem. Inf. Comput. Sci.* **1998**, *38* (2), 300–304.

(5) van Krevelen, D. W. *Properties of polymers, their estimation and correlation with chemical Structure*; Elsevier: Amsterdam, 1976.

(6) Bicerano, J. *Prediction of polymers properties*; Marcel Dekker: New York, 1996.

(7) Katrizky, A. R.; Rachwal, P.; Law, K. W.; Karelson, M.; Lobanov, V. S. Prediction of polymer glass transition temperatures using a general quantitative structure-property relationship treatment. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 879–884.

(8) Bishop, M. C. *Neural Networks for Pattern Recognition*: Oxford University Press: New York, 1995.

(9) Rosenblatt, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psych. Rev.* **1958**, *65*, 386–408.

(10) Haykin, S. *Neural Networks*; Macmillan College Publishing Company (ed.).: New York, 1994.

(11) Xu, K.; Xie, M.; Tang, L.; Ho, S. L. Application of neural networks in forecasting engine system reliability. *App. Soft. Comput.* **2003**, *2*, 255–268.

(12) Rumelhart, D. E.; McClelland, J. L. *Parallel distributed processing: Exploration in the microstructure of cognition*: MIT Press: Cambridge, USA, 1986.

(13) Castillo, E.; Gutiérrez, J. M.; Hadi, A. S.; Lacruz, B. Some applications of functional networks in statistics and engineering. *Technometrics* **2001**, *43*, 10–24.

(14) Arizmendi, C. M.; Sánchez, J. R.; Ramos, N. E.; Ramos, G. J. Time series predictions with neural sets: application to airborne pollen forecasting. *Int. J. Biometeorol.* **1993**, *37*, 139–144.

(15) Castellano-Méndez, M.; Aira, M. J.; Iglesias, I.; Jato, V.; González-Manteiga, W. Artificial neural networks as a useful tool to predict the risk level of Betula pollen in the air. *Int. J. Biometeorol.* **2005**, *49*, 310–316.

(16) Grinn-Gofron, A.; Strzelczak, A. Artificial neural network models of relationships between Cladosporium spores and meteorological factors in Szczecin (Poland). *Grana* **2008**, *47*, 305–315.

(17) Sánchez Mesa, J. A.; Galán, C.; Hervás, C. The use of discriminant analysis and neural networks to forecast the severity of the Poaceae pollen season in a region with a typical Mediterranean climate. *Int. J. Biometeorol.* **2005**, *49*, 355–362.

(18) Sánchez Mesa, J. A.; Galán, C.; Martínez, J. A.; Hervás, C. The use of a neural network to forecast daily grass pollen concentration in a Mediterranean region: the southern part of the Iberian Peninsula. *Clin. Exp. Allergy* **2002**, *32*, 1606–1612.

(19) Rodríguez-Rajo, F. J.; Astray, G.; Ferreiro-Lage, J. A.; Aira, M. J.; Jato-Rodriguez, V.; Mejuto, J. C. Evaluation of atmospheric Poaceae pollen concentration using a neural network applied to a coastal Atlantic climate region. *Neural Networks* **2010**, *23*, 419–425.

(20) Astray, G.; Castillo, X.; Ferreiro-Lage, J. A.; Gálvez, J. F.; Mejuto, J. C. Artificial neural networks: a promising tool to evaluate the authenticity of wine. *J. Food* **2010**, *8*, 79–86.

(21) Astray, G.; Caderno, P. V.; Ferreiro-Lage, J. A.; Gálvez, J. F.; Mejuto, J. C., Prediction of Ethene + Oct-1-ene Copolymerization Ideal Conditions Using Artificial Neuron Networks. *J. Chem. Eng. Data*, DOI: 10.1021/je1001973.

(22) Chen, X.; Sztandera, L.; Cartwright, H. M. A neural network approach to prediction of glass transition temperature of polymers. *Int. J. Intell. Syst.* **2008**, *23*, 22–32.

(23) Wanqiang, L.; Chenzhong, C. Artificial neural network prediction of glass transition temperature of polymers. *Colloid Polym. Sci.* **2009**, *287*, 811–818.

(24) Brandrup, J.; Immergut, E. H.; Grulke, E. A. *Polymer Handbook*, 4th ed.; Wiley: New York, 1999.

(25) Parr, R. G.; Yang, W. *Density functional theory of atoms and molecules*; Oxford: New York, 1989.

(26) Johnson, B. G.; Gill, P. M. W.; Pople, J. A. The performance of a family of density functional methods. *J. Chem. Phys.* **1993**, *98*, 5612–5626.

(27) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas,

O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision B.01; Gaussian, Inc.: Pittsburgh PA, 2003.

(28) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(29) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789.

(30) Scott, A. P.; Radom, L. J. Harmonic Vibrational Frequencies: An Evaluation of Hartree-Fock, Møller-Plesset, Quadratic Configuration Interaction, Density Functional Theory, and Semiempirical Scale Factors. *J. Phys. Chem.* **1996**, *100*, 16502–16513.

Received for review May 28, 2010. Accepted August 8, 2010. Astray thanks the "Ministerio de Educación" of Spain for a FPU grant P.P. 0000 421S 14006.

JE100573N