

Prediction of Polymer Glass Transition Temperatures Using a General Quantitative Structure–Property Relationship Treatment

Alan R. Katritzky,^{*,§} Peter Rachwal,[§] Kam Wah Law,[‡] Mati Karelson,[⊥] and Victor S. Lobanov[⊥]

Department of Chemistry, Florida Center for Heterocyclic Compounds, University of Florida,
P.O. Box 117200, Gainesville, Florida 32611-7200, 3M Company, St. Paul, Minnesota, 55144-1000,
and the Department of Chemistry, University of Tartu, 2 Jakobi Str., Tartu EE 2400, Estonia

Received October 13, 1995[⊗]

A novel approach to the prediction of the physical properties of polymers is presented. A QSPR study, involving the use of a newly developed statistical package, CODESSA, is described for the T_g of a set of 22 low molecular weight polymers which gave a four-parameter equation with $R^2 = 0.928$. The physical significance of the descriptors selected is discussed.

INTRODUCTION

Quantitative structure–activity relationship (QSAR) correlations have been widely applied for biological activities over several decades.^{1–7} Also, many applications of quantitative structure–property relationships (QSPR) are known in analytical chemistry.^{3,8–14} For instance, we recently successfully used¹⁵ our CODESSA (*Comprehensive Descriptors for Structural and Statistical Analysis*) QSPR program¹⁶ to achieve the first comprehensive correlation of both GC retention times and response factors for a large and widely diversified set of organic compounds. The QSPR equations developed¹⁵ allow the prediction of each of these two quantities for unknown compounds with significant confidence.

Considerable attention has been devoted to establishing quantitative structure–property relationships in polymers, as recently reviewed.¹⁷ As regards glass transition temperatures, the most successful approach for homopolymers has been that of Koehler–Hopfinger which relates T_g by multiple linear regression to the intramolecular flexibility represented by linear contributions of conformational entropies of the repeat units computed from torsional rotations about backbone and side-chain bonds and to intermolecular interactions represented by the dispersion, positive electrostatic, and negative electrostatic intermolecular energies.^{18,19} This model was recently simplified using the Tripos 5.2 Force field and for a set of 12 alkyl acrylates R^2 of 0.902 was established with four descriptors [93MCMS261]. However, this method is not applicable to copolymers, and the scatter is greater for less closely related sets.

Earlier work relied on group additive methods^{21,22} which are not applicable to predictions for polymers containing previously noninvestigated groups. The effect of molecular weight on T_g was rationalized by a “corresponding states” model.²³ Another approach to the prediction of T_g for linear polymers, random copolymers, and cross-linked polymers based on the tabulated atomic incremental volumes has been published by Wiff, Altieri, and Goldfarb in preliminary form only.²⁴

We now report that a QSPR treatment based on CODESSA methodology can successfully rationalize the glass transition temperature (T_g) values of a set of low and medium molecular weight homo- and copolymers and that the derived multilinear dependencies can be used for the prediction of the T_g values for unknown polymers of similar structure.

PROCEDURE

Glass Transition Temperatures of Polymers. A set of 22 linear, medium MW polymers and copolymers was used in the QSPR treatment (cf. Table 1). The glass transition temperatures of all the samples, except sample **12**, were measured by a Perkin-Elmer DSC 7. The temperature range was from -150 °C to 20 °C. The scanning rate was 20 °C/min, and the sample size was ~ 10 mg. The onset temperature of the transition was taken as the T_g . The experimental error was estimated by measuring sample **11** five times and was calculated as 0.7 °C.

The T_g of sample **12** was measured by TSC with a Solomat TSC/RMA 91000. Sample **13** was used as a reference. The area and the thickness of the sample were 38 mm² and 0.1 mm, respectively. The temperature range was from -160 to -60 °C, and the scan rate was 7 °C/min. The temperature at the peak of relaxation was taken as T_g .

QSPR Approach. The QSPR method applied by us is based on the correlation of the property under investigation with a limited number of descriptors selected from a large set of descriptors derived solely from the molecular structure of the individual compounds. Such structural descriptors are numerous and fall into various classes (constitutional, topological, energetic, quantum chemical, etc.); they are calculated from the molecular structures using defined mathematical procedures and reflect geometrical or electronic properties of the compound.

Because of software and hardware limitations, the calculations of many types of standard descriptors for high molecular weight polymer molecules (molecular weights over 2000) cannot be carried out directly. Therefore, their descriptors have to be estimated using some approximate procedure. The essence of the present approach is the definition of suitable extrapolated values of the descriptors of polymers by using descriptor values calculated for series of analogous low molecular weight oligomers. We now

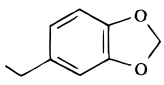
[§] Florida Center for Heterocyclic Compounds.

[‡] 3M Company.

[⊥] University of Tartu.

[⊗] Abstract published in *Advance ACS Abstracts*, May 15, 1996.

Table 1. Structural Formulas and Glass Transition Temperatures (T_g) of the Polymers Studied

compd	formula	MW	comments	T_g (°C)
1	$\text{CH}_3(\text{CHRCH}_2)_n\text{CHRCH}_3$	527	$\text{R} = \text{C}_8\text{H}_{17}$	-93.4
2	$\text{CH}_3(\text{CHRCH}_2)_n\text{CHRCH}_3$	662	$\text{R} = \text{C}_8\text{H}_{17}$	-88.7
3	$\text{CH}_3(\text{CHRCH}_2)_n\text{CHRCH}_3$	753	$\text{R} = \text{C}_8\text{H}_{17}$	-87.6
4	$\text{CH}_3(\text{CHRCH}_2)_n\text{CHRCH}_3$	1705	$\text{R} = \text{C}_8\text{H}_{17}$	-79.1
5	$\text{CH}_3(\text{CHRCH}_2)_n\text{CHRCH}_3$	2750	$\text{R} = \text{C}_8\text{H}_{17}$	-74.6
6	$\text{HO}(\text{CH}_2\text{CH}_2\text{O})_n\text{OH}$	400		-115.0
7	$\text{CH}_3\text{O}(\text{CH}_2\text{CH}_2\text{O})_n\text{OCH}_3$	2000		-126.1
8	$\text{CH}_3\text{O}_2\text{CCF}_2\text{O}(\text{CF}_2\text{CF}_2\text{O})_m(\text{CF}_2\text{O})_n\text{CF}_2\text{CO}_2\text{CH}_3$	2000	$n/m = 1.25$	-128.7
9	$\text{HO}_2\text{CCF}_2\text{O}(\text{CF}_2\text{CF}_2\text{O})_m(\text{CF}_2\text{O})_n\text{CF}_2\text{CO}_2\text{H}$	2000	$n/m = 1.25$	-124.4
10	$\text{RCF}_2\text{O}(\text{CF}_2\text{CF}_2\text{O})_m(\text{CF}_2\text{O})_n\text{CF}_2\text{R}$	2000	$\text{R} = $ 	-126.7
11	$\text{HOCF}_2\text{O}(\text{CF}_2\text{CF}_2\text{O})_m(\text{CF}_2\text{O})_n\text{CF}_2\text{OH}$	1000	$n/m = 1.25$	-122.8
12	$\text{CF}_3\text{CF}_2\text{O}(\text{CF}_2\text{CF}_2\text{O})_m(\text{CF}_2\text{O})_n\text{CF}_2\text{CF}_3$	4000	$n/m = 12.8$	-158.0
13	$\text{CF}_3\text{CF}_2\text{O}(\text{CF}_2\text{CF}_2\text{O})_m(\text{CF}_2\text{O})_n\text{CF}_2\text{CF}_3$	8000	$n/m = 1.37$	-134.9
14	$\text{CF}_3\text{CF}_2\text{O}(\text{CF}_2\text{CF}_2\text{O})_m(\text{CF}_2\text{O})_n\text{CF}_2\text{CF}_3$	9500	$n/m = 1.22$	-131.3
15	$\text{CF}_3\text{CF}_2\text{O}(\text{CF}_2\text{CF}_2\text{O})_m(\text{CF}_2\text{O})_n\text{CF}_2\text{CF}_3$	8000	$n/m = 1.27$	-131.1
16	$\text{CF}_3\text{CF}_2\text{O}(\text{CF}_2\text{CF}_2\text{O})_m(\text{CF}_2\text{O})_n\text{CF}_2\text{CF}_3$	4500	$n/m = 0.67$	-132.8
17	$\text{CF}_3\text{CF}_2\text{CF}_2\text{O}(\text{CF}_2\text{CF}_2\text{CF}_2\text{O})_m\text{CF}_2\text{CF}_3$	2140		-116.5
18	$\text{CH}_3\text{O}_2\text{CC}_3\text{F}_6\text{O}(\text{C}_2\text{F}_4\text{O})_n\text{C}_3\text{F}_6\text{CO}_2\text{CH}_3$	2050		-98.6
19	$\text{CH}_3\text{O}_2\text{CC}_3\text{F}_6\text{O}(\text{C}_4\text{F}_8\text{O})_n\text{C}_3\text{F}_6\text{CO}_2\text{CH}_3$	3300		-100.4
20	$\text{CF}_3\text{CF}_2\text{CF}_2\text{O}(\text{C}(\text{CF}_3)\text{CF}_2\text{O})_n\text{CF}_2\text{CF}_3$	3500		-82.2
21	$\text{F}(\text{CF}_2\text{C}(\text{F})(\text{Cl}))_n\text{CF}_3$	700		-78.9
22	$\text{CF}_3\text{CF}_2\text{O}(\text{CF}_2\text{CF}_2\text{OCF}_2\text{O})_n\text{CF}_2\text{CF}_3$	2000		-128.5

demonstrate that such extrapolated descriptors, when combined with fragment descriptors and directly calculated descriptors, allow the treatment of polymer properties, specifically T_g , by QSPR techniques.

Calculations of Descriptors for Homopolymers. Six main types of molecular descriptors can be calculated for any individual molecular structure using the CODESSA program package.¹⁶ These include the *constitutional*, *geometrical*, *topological*, *electrostatic*, *quantum-chemical*, and *thermodynamic* classes of descriptors. It is important to emphasize that all of these descriptors are calculated solely on the basis of the intrinsic structural information of the molecular species under consideration, without the need for any experimental data. The constitutional descriptors (number of atoms and bonds, absolute and relative numbers of different types of atoms and bonds, number of rings, molecular weight, etc.) are the simplest type: they reflect solely the molecular composition of the compound without incorporating any information regarding its geometry or electronic structure. The geometrical descriptors (moments of inertia, shadow indices,²⁵ molecular volume and molecular surface area) are calculated from the three-dimensional structure of the molecule. Topological descriptors (also called topological indices) describe the atomic connectivity in the molecule.^{3,26} Electrostatic descriptors reflect the characteristics of the charge distribution of the molecule: partial charges in the molecule, and the descriptors derived from them, are calculated in CODESSA either from semiempirical quantum-chemical data or using the empirical approaches proposed by Zefirov^{27,28} and Gasteiger.²⁹ A subclass of electrostatic descriptors, called charged partial surface area (CPSA) descriptors, was introduced by Jurs et al.^{30,31} and describes the charge in terms of the whole surface area of the molecule and in terms of functional group contributions. These CPSA descriptors encode features responsible for polar interactions between molecules; in CODESSA, a set of 26 CPSA descriptors is calculated as a combination of the contributions of atomic partial charges to the total molecular solvent-accessible surface area.

Numerous studies have involved descriptors calculated from quantum chemical data³²⁻³⁷ and a wide variety of quantum-chemically calculated descriptors is available in the CODESSA program. These quantum-chemical descriptors are calculated from the output results of the MOPAC package.³⁸ The thermodynamic descriptors are calculated on the basis of the total partition function of the molecule and its electronic, translational, rotational, and vibrational components.^{39,40} Whereas the constitutional, geometrical, topological, and electrostatic descriptors can be calculated directly proceeding from the three-dimensional structure of the molecule, quantum-chemical and thermodynamic descriptors can be derived only by using the quantum-chemical wave function and potential surface of the molecule.

Most types of constitutional descriptors for polymers may be computed directly provided the molecular formula and average molecular weight are available. Atom or bond counts, the relative ratio of a particular atom type, or the molecular weight itself serve as good examples of such descriptors. Because no extrapolation is involved in their calculation, the obtained descriptor values are precise and should therefore be highly reliable although they will be average values and can in principal depend on the molecular weight distribution.

Although constitutional descriptors comprise but a small fraction of the complete descriptor set available in CODESSA, in some cases one may avoid extrapolation by changing focus from the entire molecule to its individual building blocks. Topological descriptors, such as the Wiener,⁴¹ Randić,⁴² and Kier and Hall¹ indices, which describe the branching of molecules, provide virtually identical information when calculated on a per-fragment basis as when extrapolated for entire polymers. Moreover, such topological descriptors need to be calculated only once for each unique fragment, and then the obtained value can be utilized as required for other copolymers containing that same fragment. To reflect the regular dependence of topological indices on polymer length, it is advantageous to multiply the descriptors derived from individual fragments by the logarithm of the

Table 2. Descriptors Involved in the Best Four-Parameter Model (22 Compounds) Derived for T_g , the Corresponding Regression Coefficients X , the Errors ΔX , t -Test Values, and the Successive Correlation Coefficients R^2 and R^2_{cv}

descriptor name	X	ΔX	t -test	R^2 ^a	R^2_{cv} ^b
intercept	-69.4	6.37	-10.9		
DPSA-3 divided by the number of atoms ^c	-39.1	3.55	-11.0	0.753	0.711
Randić $3 \times \log(N)$ ^d	9.14	1.59	5.76	0.786	0.713
number of OH groups	9.53	2.46	3.88	0.850	0.796
f-PNSA-2 ^e	0.0229	0.00532	4.30	0.928	0.890

^a Correlation coefficient squared for the regression including this and previous descriptors in table. ^b Cross-validated correlation coefficient squared for the regression including this and previous descriptors in table. ^c DPSA-3 = difference between atomic charge weighted partial positive and partial negative surface areas. ^d Randić index (order 3), ${}^3\chi = \sum_{(\text{paths of length } 3)} (\delta_i \delta_j \delta_k \delta_l)^{-1/2}$. ^e Partial negative charge surface area weighted by total charge, calculated for a single fragment.

total number of fragments (henceforth called N), provided that the polymer length is sufficiently large. This normalization factor should account for the actual polymer size.

The remaining classes of descriptors need to be individually extrapolated. Descriptors reflecting the minimum or maximum values of some atomic properties (*e.g.*, atomic charges, valencies) generally become independent of the size of a molecule above a certain molecular weight and can be readily obtained by consideration of simple monomeric or oligomeric structures. Other descriptors may vary with N according to a linear, reciprocal, exponential, or logarithmic dependence as predetermined by the physical nature of the descriptor.^{43,44} In such cases, the best fit extrapolation was performed using the descriptor values calculated for the first three to five oligomers and from the equation obtained the descriptor value at actual polymer length (N) was estimated.

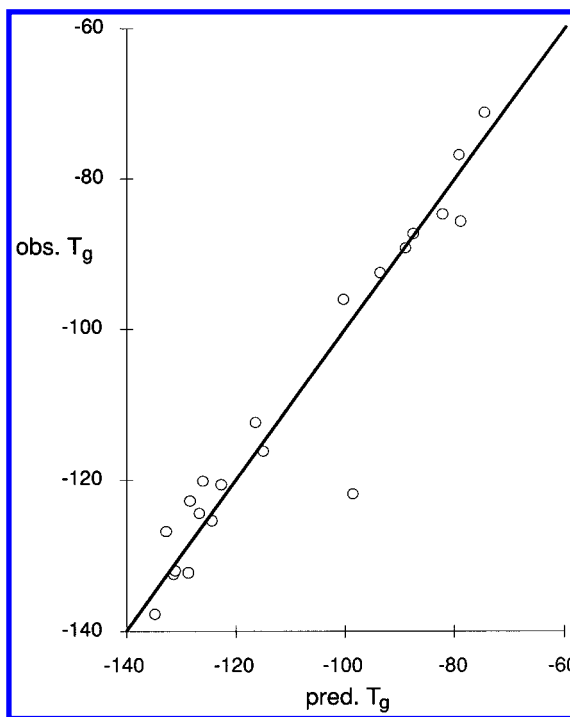
Treatment of Copolymers. A significant fraction (11 out of 22) of the structures used in the study were copolymers, *i.e.*, polymers composed of two distinct repeating fragment types. For our calculations, each copolymer was represented as a pair of polymers, one for each type of repeating fragment; the descriptor values calculated for such pairs were combined into a single value characteristic of the actual copolymer. For descriptors whose values are independent of N , formula 1 was used

$$D = \frac{nD_n + mD_m}{n + m} \quad (1)$$

$$D = D_n + D_m - D_t \quad (2)$$

In formula 1, n and m are the repeating nonterminal unit counts, and D_n and D_m are the descriptor values obtained from the extrapolation of the data for the first three to five oligomers with one or the other fragment, respectively, as their building blocks. For many significant descriptors which vary linearly with N , formula 2 was utilized for the calculation of the descriptor values for the copolymers. In formula 2, D_t denotes the contribution of both terminal groups and is taken as the intercept of the best fit line through the data points for $n = 1, 2$, and 3. It is assumed that this intercept is the same regardless of the fragment used since for $m = n = 0$ both structures would be identical (all the copolymers in the set used possess uniquely defined terminal groups).

Preselection of Descriptors and Derivation of the QSPR. Altogether, 238 CODESSA descriptors were preselected for the final QSPR treatment to obtain the best correlations. Several criteria were applied in the preselection procedure: (i) To avoid misleading correlations because of

**Figure 1.** Plot of observed vs predicted values of T_g .

the collaterality of descriptor scales,⁴⁵ strongly intercorrelated descriptor scales were represented only by one single descriptor in the further QSPR development. (ii) Descriptors expected to have no relation to glass transition temperatures were eliminated. (iii) Descriptors which depend heavily on the conformational shape of the molecule (such as molecular shadow area indices and moments of inertia) were not used because they would require unknown information regarding the conformations of polymers studied. (iv) The importance of each descriptor remaining after the above eliminations was individually evaluated based on the associated t -value in the one-parameter correlation with T_g . Low t -values normally indicate high uncertainty in the respective parameter coefficients. However, some of the t -values increased and the overall stability of the model improved after one or two descriptors were substituted with closely related, though not entirely identical, descriptors.

In practice it is mechanically impossible to evaluate computationally all of the possible multiparameter correlations in the space of 238 descriptors, and a systematic approach for the selection of the best regression line of a given dimensionality was required. We used the following computational scheme, previously proposed by us in the framework of the CODESSA program, to find the final QSPR equations:

Table 3. Descriptors Involved in the Best Four-Parameter Model (21 Compounds, Compound **18** Eliminated) Derived for T_g , the Corresponding Regression Coefficients X , the Errors ΔX , t -Test Values, and the Successive Correlation Coefficients R^2 and R^2_{cv}

descriptor name	X	ΔX	t -test	R^2 ^a	R^2_{cv} ^b
intercept	-74.5	2.37	-20.9		
DPSA-3 divided by the number of atoms ^c	-37.6	1.75	-21.5	0.762	0.718
Randić3 $\times \log(N)$ ^d	10.8	0.812	13.3	0.795	0.720
number of OH groups	10.4	1.21	8.57	0.865	0.812
f-PNSA-2 total charge weighted PNSA ^e	0.0295	0.00276	10.7	0.983	0.974

^a Correlation coefficient squared for the regression including this and previous descriptors in table. ^b Cross-validated correlation coefficient squared for the regression including this and previous descriptors in table. ^c DPSA-3 = difference between atomic charge weighted partial positive and partial negative surface areas. ^d Randić index (order 3), ${}^3\chi = \sum_{(paths\ of\ length\ 3)} (\delta_i \delta_j \delta_k \delta_l)^{-1/2}$. ^e Partial negative charge surface area weighted by total charge, calculated for a single fragment.

Table 4. Values of Descriptors Plus Experimental and Predicted values of T_g for Each of the 22 Polymers Used in the Study

ID	DPSA-3/no. of atoms	Randić3 $\times \log(N)$	no. of OH	f-PNSA-2	$T_g^{exp.}$	$T_g^{pred. a}$	exp. -pred ^a	$T_g^{pred. b}$	exp. -pred ^b
1	0.795	1.143	0	-107.9	-93.4	-92.5	-0.9	-95.2	1.8
2	0.796	1.506	0	-107.9	-88.7	-89.2	0.5	-91.3	2.6
3	0.792	1.700	0	-107.9	-87.6	-87.3	-0.3	-89.0	1.4
4	0.787	2.824	0	-107.9	-79.1	-76.8	-2.3	-76.7	-2.4
5	0.781	3.423	0	-107.9	-74.6	-71.1	-3.5	-70.0	-4.6
6	1.808	0.690	2	-57.4	-115.0	-116.1	1.1	-116.0	1.0
7	1.574	1.328	0	-54.1	-126.1	-120.1	-6.0	-120.9	-5.2
8	1.990	3.045	0	-557.9	-128.7	-132.2	3.5	-132.9	4.2
9	2.109	2.275	2	-587.4	-124.4	-125.5	1.1	-125.8	1.4
10	1.581	2.194	0	-578.8	-126.7	-124.4	-2.3	-127.3	0.6
11	2.042	2.299	2	-499.8	-122.8	-120.7	-2.1	-120.4	-2.4
12	2.105	1.118	0	-460.1	-158.0	-152.0	-6.0	-155.2	-2.8
13	2.051	2.773	0	-587.5	-134.9	-137.7	2.8	-139.0	4.0
14	2.057	3.397	0	-598.3	-131.3	-132.5	1.2	-132.8	1.5
15	2.071	3.501	0	-594.7	-131.1	-132.0	0.9	-132.0	0.9
16	1.979	3.833	0	-652.6	-132.8	-126.7	-6.1	-126.7	-6.1
17	1.600	5.988	0	-1537.4	-116.5	-112.4	-4.1	-115.1	-1.4
18	1.488	3.178	0	-1013.4	-98.6	-121.7	23.1	-125.9	27.3
19	1.220	6.287	0	-1591.3	-100.4	-96.1	-4.3	-99.2	-1.2
20	1.363	7.391	0	-1290.6	-82.2	-84.7	2.5	-83.7	1.5
21	1.207	3.871	0	-193.0	-78.9	-85.7	6.8	-83.7	4.8
22	1.488	4.980	0	-1775.6	-128.5	-122.7	-5.8	-128.9	0.4

^a Based on the model listed in Table 2. Average deviation = 3.96 K (3.04 K with structure **18** removed). ^b Based on the model listed in Table 3. Average deviation = 3.61 K (2.48 K with structure **18** removed).

(1) First, the intercorrelations between all of the 238 selected descriptors were calculated and 10 755 orthogonal pairs of descriptors i and j (with $R_{ij}^2 < 0.1$) were determined.

(2) The statistical analysis of the property (T_g) started with the calculation of the two-parameter regression with each of the 10 755 orthogonal pairs of descriptors obtained in step 1. Out of these, the 400 descriptor pairs with the highest regression correlation coefficients were selected for the higher-order regression treatments.

(3) For each of the 400 descriptor pairs selected in the previous step a noncollinear descriptor scale, k (with $R_{ik}^2 < 0.4$ and $R_{kj}^2 < 0.4$), was added, and the respective three-parameter regression was calculated. This procedure was repeated with all noncollinear scales to a given pair of descriptors. Out of the descriptor triplets obtained, the 400 with the highest regression correlation coefficients were submitted to the next step.

(4) In a similar fashion, an additional noncollinear descriptor scale was added to each of the 400 descriptor subsets selected in the previous step, and the respective four-parameter regression treatment was performed. The best four-parameter correlation found was superior, according to the Fisher criterion at the given probability level, F , to any higher order correlations, and as it also proved to be the best according to the cross-validated correlation coefficient, it was therefore chosen as the final equation. It is essential to note

that the cross-validation technique⁴⁶ also provides an estimate of the stability of the obtained regression model, i.e., the sensitivity of the model to the elimination of any single data point.

It has to be noted that the rigorous selection of the uniquely best parameters for a correlation equation from a large descriptor set remains an unsolved statistical problem and existing procedures do not guarantee finding the overall best correlation.

Comparisons with the Previous QSPR Approaches Used To Correlate T_g . All previously reported QSPR studies of glass transition temperature employed limited sets of structural parameters. Within the framework of the group additivity approach the correlation equation is predetermined^{17,21,24} or chosen from among a few similar variants,²² and it usually involves only one or two structural parameters. The least-squares method is used to tabulate these group contributions (e.g., per-atom, per-bond, per-repeating unit values). Hence, the utility of these QSPRs rely on the extent of the domain for which chosen equation is correct and on the availability of tabulated increments.

The Hopfinger and Koehler approach to estimate T_g is based on seven structural parameters calculated by means of a molecular modeling technique.^{18,19} Originally all seven parameters were included into the linear correlation equation and regression coefficients were determined by multidimen-

sional linear regression analysis.¹⁹ However, the authors noted that some of the parameters have only minimal contributions¹⁹ and therefore applied genetic function approximation and evolutionary programming techniques to create correlation equations with optimal combinations of the seven structural parameters.^{47,48} Apparently only those structural parameters were tested.

The approach reported here provides an enhanced QSPR treatment of T_g of polymers. The large collection of structural descriptors examined and a thorough correlation selection procedure have led to a successful correlation model, which employs solely structural parameters of perceived significance.

ANALYSIS AND PHYSICAL INTERPRETATION OF THE QSPR

The optimum four-parameter model thus derived for T_g is shown in Table 2. The corresponding predicted values are plotted in Figure 1 vs experimental. The R^2 value for the T_g correlation was 0.928 with $R_{cv}^2 = 0.890$. However, there is an obvious outlier at structure **18**, and with this structure removed, the R^2 increases to 0.983 and the cross-validated R_{cv}^2 likewise increases to 0.974. The regression coefficients, however, remain practically unchanged for the set of 21 compounds (Table 3) versus the set of 22.

The glass transition temperatures (T_g) of the polymers studied are strongly influenced by the difference between the positive and the negative partial surface areas normalized by the number of atoms (DPSA). As expected, the polymers with larger DPSA values have stronger intermolecular electrostatic interactions and therefore display higher glass transition temperatures. The next most important descriptor is the topological Randić index⁴² computed for a repeating unit and then extrapolated through multiplication by $\log N$. The Randić index generally reflects the branching level of a molecule. According to the QSPR model now developed, a higher degree of branching in the repeating fragment structure elevates the glass transition temperature. The third parameter, the number of OH groups, is of the expected significance because it accounts for the presence of hydrogen bonds in the polymer matrix. In the polymer set now studied, hydroxyl groups appear only in the terminal groups. Therefore, the hydrogen bonding effect should be relatively more important for lower molecular weight polymers. The partial negative surface area of the fragment completes the model and like the DPSA, this parameter again reflects the electrostatic interactions between the polymer chains.

CONCLUSION

It has been demonstrated that a four parameter correlation with strongly orthogonal descriptors can predict the glass transition temperatures of a variety of low and medium molecular weight polymers with substantial confidence. The standard error of this correlation equation was 3.05 °C and the square of the cross-validated correlation coefficient $R_{cv}^2 = 0.890$. Although the derived equation is of restricted validity due to the limited size and variety of the training set, this result encourages the application of quantitative structure–property relationships to a wider selection of polymer properties and to other classes of polymers, including biopolymers.

REFERENCES AND NOTES

- (1) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.
- (2) Hansch, C.; Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; John Wiley & Sons: New York, 1979.
- (3) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure-Activity Analysis*; Research Studies Press: Letchworth, 1986.
- (4) Hansch, C.; Kim, D.; Leo, A. J.; Novellino, E.; Silipo, C.; Vittoria, A. *Critical Reviews in Toxicology*; McClellan, R. O., Ed.; CRC Press: Florida, 1989; Vol. 19, p 185.
- (5) Gupta, S. P. Quantitative Structure–Activity Relationship Studies on Local Anesthetics. *Chem. Rev.* **1991**, 91, 1109–1119.
- (6) Leo, A. J. Calculating $\log P_{oct}$ from Structures. *Chem. Rev.* **1993**, 93, 1281–1306.
- (7) Gupta, S. P. Quantitative Structure–Activity Relationship Studies on Anticancer Drugs. *Chem. Rev.* **1994**, 94, 1507–1551.
- (8) Kowalski, B. R. *Chemometrics*; Reidel: Dordrecht, 1984.
- (9) Brown, S. D.; Barker, T. Q.; Larivee, R. J.; Monfre, S. L.; Wilk, H. R. *Chemometrics. Anal. Chem.* **1988**, 60, 252R–273R.
- (10) Brown, S. D. *Chemometrics. Anal. Chem.* **1990**, 62, 84R–101R.
- (11) Brereton, R. A. *Chemometrics: Applications of Mathematics and Statistics to Laboratory Systems*; E. Horwood: New York, 1990.
- (12) Brown, S. D.; Bear, R. S., Jr.; Blank, T. B. *Chemometrics. Anal. Chem.* **1992**, 64, 22R–49R.
- (13) Meloun, M.; Militky, J.; Forina, M. *Chemometrics for Analytical Chemistry*; Ellis Horwood: New York, 1992; Vol. 1.
- (14) Brown, S. D.; Blank, T. B.; Sum, S. T.; Weyer, L. G. *Chemometrics. Anal. Chem.* **1994**, 66, 315R–359R.
- (15) Katritzky, A. R.; Ignatchenko, E. S.; Barcock, R. A.; Lobanov, V. S.; Karelson, M. Prediction of Gas Chromatographic Retention Times and Response Factors Using a General Quantitative Structure–Property Relationship Treatment. *Anal. Chem.* **1994**, 66, 1799–1807.
- (16) Katritzky, A. R.; Lobanov, V.; Karelson, M. *CODESSA, Reference Manual*; University of Florida: 1994.
- (17) Askadskii, A. A. *Polym. Sci., Ser. B* **1995**, 37, 66–88.
- (18) Hopfinger, A. J.; Koehler, M. G.; Pearlstein, R. A.; Tripathy, S. K. *J. Polym. Sci., Part B: Polym. Phys.* **1988**, 26, 2007–2028.
- (19) Koehler, M. G.; Hopfinger, A. J. *Polymer* **1989**, 30, 116–126.
- (20) Vaz, R. J. *Makromol. Chem., Macromol. Symp.* **1993**, 65, 261–268.
- (21) Weyland, H. G.; Hoftyzer, P. J.; Van Krevelen, D. W. *Polymer* **1970**, 11, 79–87.
- (22) Lee, W. A. *J. Polym. Sci., Part A-2* **1970**, 3, 555–570.
- (23) DiBenedetto, A. T.; DiLandro, L. *J. Polym. Sci., Part B: Polym. Phys.* **1989**, 27, 1405–1417.
- (24) Wiff, D. R.; Altieri, M. S.; Goldfarb, I. J. *J. Polym. Sci.: Polym. Phys.* **1985**, 23, 1165–1176.
- (25) Rohrbaugh, R. H.; Jurs, P. C. Descriptions of Molecular Shape Applied in Studies of Structure/Activity and Structure/Property Relationships. *Anal. Chim. Acta* **1987**, 199, 99–109.
- (26) (a) Bonchev, D.; Balaban, A. T.; Mekenyan, O. Generalization of the Graph Center Concept, and Derived Topological Centric Indexes. *J. Chem. Inf. Comput. Sci.* **1980**, 20, 106–113. (b) Balaban, A. T. Highly Discriminating Distance-Based Topological Index. *Chem. Phys. Lett.* **1982**, 89, 399–404. (c) Bonchev, D. *Information-Theoretic Indices for Characterization of Chemical Structures*; Research Studies Press: Chichester, 1983. (d) Stankevich, M. I.; Stankevich, I. V.; Zefirov, N. S. *Topological Indices in Organic Chemistry. Russ. Chem. Rev.* **1988**, 57, 191–208.
- (27) Zefirov, N. S.; Kirpichenok, M. A.; Izmailov, F. F.; Trofimov, M. I. Scheme for the Calculation of the Electronegativities of Atoms in a Molecule in the Framework of Sanderson's Principle. *Dokl. Akad. Nauk SSSR* **1987**, 296, 883–887.
- (28) Kirpichenok, M. A.; Zefirov, N. S. Electronegativity and Molecular Geometry. I. General Principles of the Method and Analysis of the Effect of Short-Range Electrostatic Interactions of Bond Lengths in Organic Molecules. *Zh. Org. Khim.* **1987**, 23, 673–691.
- (29) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity – A Rapid Access to Atomic Charges. *Tetrahedron* **1980**, 36, 3219–3288.
- (30) Stanton, D. T.; Jurs, P. C. Development and Use of Charged Partial Surface Area Structural Descriptors in Computer-Assisted Quantitative Structure–Property Relationship Studies. *Anal. Chem.* **1990**, 62, 2323–2329.
- (31) Stanton, D. T.; Egolf, L. M.; Jurs, P. C.; Hicks, M. G. Computer-Assisted Prediction of Normal Boiling Points of Pyrans and Pyrroles. *J. Chem. Inf. Comput. Sci.* **1992**, 32, 306–316.
- (32) Klopman, G.; Iroff, L. D. Calculation of Partition Coefficients by the Charge Density Method. *J. Comput. Chem.* **1981**, 2, 157.
- (33) Buydens, L.; Massart, D. L.; Geerlings, P. Prediction of Gas Chromatographic Retention Indexes with Topological, Physicochemical, and Quantum Chemical Parameters. *Anal. Chem.* **1983**, 55, 738–744.

- (34) Cartier, A.; Rivail, J.-L. Electronic Descriptors in Quantitative Structure–Activity Relationships. *Chemometrics and Intelligent Laboratory Systems* **1987**, *1*, 335–347.
- (35) Bodor, N.; Gabanyi, Z.; Wong, C.-K. A New Method for the Estimation of Partition Coefficient. *J. Am. Chem. Soc.* **1989**, *111*, 3783–3786.
- (36) Tuppurainen, K.; Lötjönen, S.; Laatikainen, R.; Vartiainen, T.; Maran, U.; Strandberg, M.; Tamm, T. About the Mutagenicity of Chlorine-Substituted Furanones and Halopropenals. A QSAR Study Using Molecular Orbital Indices. *Mutation Res.* **1991**, *247*, 97–102.
- (37) Clare, B. W. Frontier Orbital Energies in Quantitative Structure–Activity Relationships: A Comparison of Quantum Chemical Methods. *Theoret. Chim. Acta* **1994**, *87*, 415–430.
- (38) Stewart, J. J. P. *MOPAC 6.0*; QCPE No 455, 1990.
- (39) McQuarrie, D. A. *Statistical Thermodynamics*; Harper & Row Publ.: New York, 1973.
- (40) Akhiezer, A. I.; Peletminskii, S. V. *Methods of Statistical Physics*; Pergamon Press: Oxford, 1981.
- (41) Wiener, H. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, *69*, 17–20.
- (42) Randić, M. On Characterization of Molecular Branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609–6615.
- (43) Brédas, J. L.; Chance, R. R. *Conjugated Polymeric Materials: Opportunities in Electronics, Optoelectronics, and Molecular Electronics*, NATO-ARW; Kluwer: Dordrecht, 1990; Vol. 182.
- (44) André, J.-M.; Delhalle, J.; Brédas, J.-L. *Quantum Chemistry Aided Design of Organic Polymers*; World Scientific: London, 1991; Vol. 2.
- (45) Myers, R. H. *Classical and Modern Regression with Applications*; Duxbury Press: Boston, 1986.
- (46) Murugan, R.; Grendze, M. P.; Toomey, J. E., Jr., Katritzky, A. R.; Karelson, M.; Lobanov, V.; Rachwal, P. Predicting Physical Properties from Molecular Structure. *CHEMTECH* **1994**, *24*, 17–23.
- (47) Rogers, D.; Hopfinger, A. J. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854.
- (48) Luke, B. T. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1279.

CI950156W