Correlation and Prediction of the Refractive Indices of Polymers by QSPR

Alan R. Katritzky,*,§ Sulev Sild,§,‡ and Mati Karelson*,‡

Center for Heterocyclic Compounds, University of Florida, Gainesville, Florida 32611-7200, and Institute of Chemical Physics, University of Tartu, 2 Jakobi Street, EE 2400 Tartu, Estonia

A general QSPR model ($R^2 = 0.940$, $s = 0.018$) was developed for the prediction of the refractive index for a diverse set of amorphous homopolymers with the CODESSA program. The five descriptors, involved in the model, are calculated from the structure of the repeating unit of the polymer. The average prediction error by this model is 0.9%.

INTRODUCTION

The refractive index n is a basic optical property of polymers that is directly related to other optical, electrical, and magnetic properties. The refractive index is also widely used in material science. The specific refractive index increment (dn/dc) is an important parameter in light scattering measurements of dilute polymer solutions, which can be used for the determination of molecular weight, size, and shape.[1] Importantly, the refractive index can indicate the potential of a polymer for a specific purpose. A satisfactory quantitative structure−property relationship (QSPR) that would allow quantitative prediction of the refractive index of as yet unsynthesized polymers would clearly be of significant utility. In principle, combining the QSPR method with pattern recognition techniques should make possible the theoretical prediction of structures with desired property values.

Theoretical methods for calculating the refractive indices of polymers generally utilize equations formulated by (i) Lorentz and Lorentz (eq 1) and (ii) Gladstone and Dale (eq 2). Both approaches require the availability (or theoretical estimation) of molar refraction and molecular volume data. A good summary of early attempts to estimate the molar refraction of polymers using group contributions was provided by Krevelen.[1] In a recent review, Askadskii[2] proposed several semiempirical equations for the calculation of various physical properties of polymers and copolymers (with accuracy usually within 3−5%). The calculation of refractive index is based on eq 1, where the molecular refraction (R_{LL}) is calculated as a sum of corresponding atom and bond contributions, and the volume (V) is estimated as a van der Waals volume of the compound divided by the average coefficient of molecular packing.

$$R_{LL} = \frac{n^2 - 1}{n^2 + 2} \cdot V \qquad (1)$$

$$R_{GD} = (n - 1) \cdot V \qquad (2)$$

The main advantage of using the group contributions method is its simplicity. Prediction with reasonable accuracy can be easily performed provided all the necessary increments are known from the experimental data for every structural element. However, interactions between functional groups can introduce significant errors in predicted refractive index values. Agrawal and Jenekhe[3] demonstrated that the refractive index of π-conjugated polymers predicted by existing group contribution methods can have deviations from experimental values as high as 22%. The source of these discrepancies is believed to be large optical dispersion and π-electron delocalization effects in conjugated polymers. To overcome this problem, Yang and Jenekhe[4] developed new Lorentz and Lorentz molar refraction group contributions for 24 functional groups commonly found in conjugated polymers. They successfully used these new R_{LL} data to calculate the refractive indices of 33 conjugated polymers (with an average error of 0.9%).[4]

Some of the shortcomings and limitations of group contribution methods can be avoided by using the theoretical QSPR approach. The quantum-chemical descriptors used in this approach encode information about the electronic structure of the molecule and thus implicitly account for the cooperative effects between functional groups, charge redistribution, and possible hydrogen bonding in the polymer. The only previously published QSPR relationship for the prediction of refractive index was developed by Bicerano ($R^2 = 0.955$) for a set of 183 polymers, with 10 descriptors involved.[5] These descriptors included three different topological indices, the total number of rotational degrees of freedom (both of the polymer backbone and the side groups), and several constitutional descriptors such as the number of fluorine atoms, the number of chlorine atoms bonded to an aromatic ring, the number of sulfur atoms, and the number of hydrogen bonding moieties, etc. Alternative topological descriptors for polymers have been developed in the framework of the topological extrapolation method (TEM) by Mekenyan *et al.*[6] and used to calculate the refractive index within a homologous series of polymers.

The QSPR method has already been applied in the framework of the CODESSA program[7] to predict successfully various physical properties of low molecular weight compounds; early examples were summarized in our review,[8] for later examples see refs 9−11. This approach was extended to calculate appropriate descriptors for the repeating

§ Center for Heterocyclic Compounds, University of Florida.
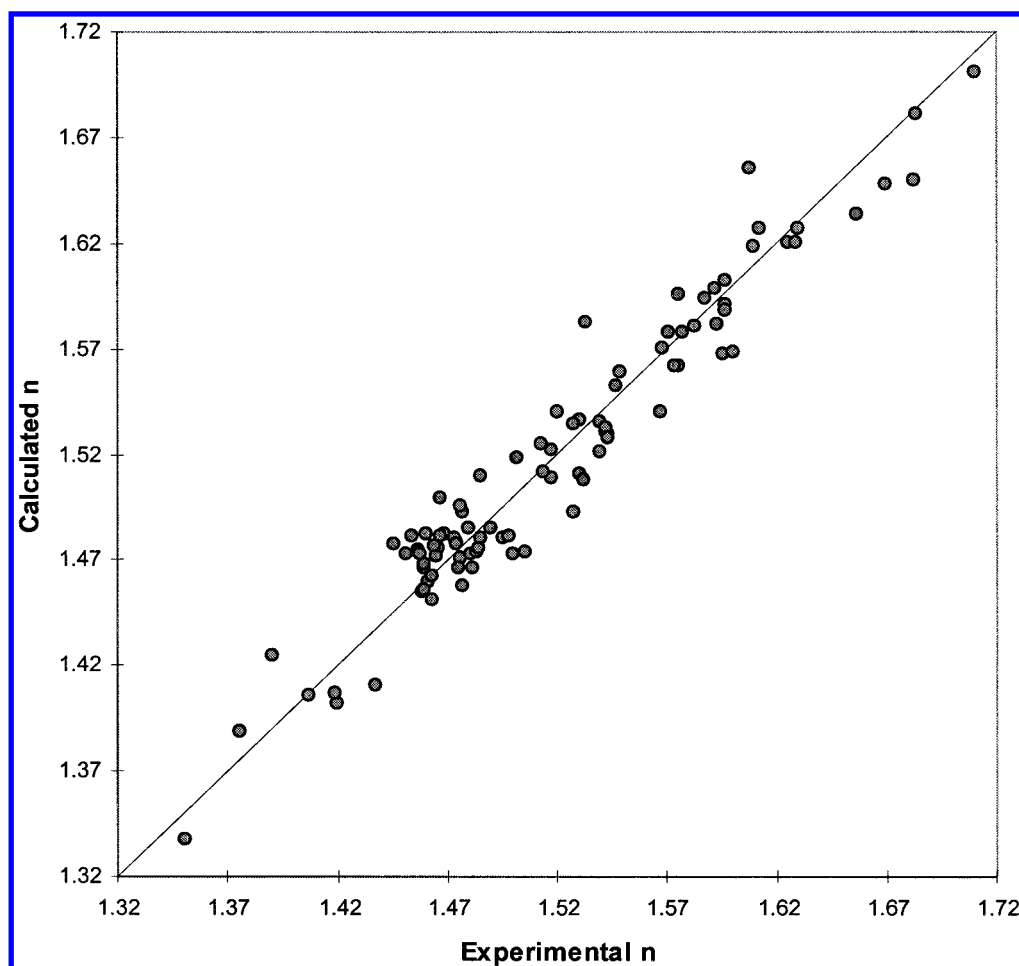‡ University of Tartu.

Figure 1. The plot of the best five parameter correlation for refractive index.

units of polymers which were subsequently used to develop correlations for the glass transition temperatures of polymers.[12,13] For a set of 22 relatively low molecular weight fluorinated polymers and copolymers, the glass transition temperatures were correlated with four descriptors ($R^2 = 0.928$).[12] Glass transition temperatures for a structurally diverse data set of 88 high molecular weight homopolymers were described by a five descriptors model ($R^2 = 0.946$).[13]

The refractive indices of a diverse set of 125 common low molecular weight organic compounds were successfully correlated by the CODESSA approach in a general QSPR relationship ($R^2 = 0.945$).[14] Five descriptors were involved in this model: HOMO−LUMO energy gap, quantum-chemically (AM1 method) calculated lowest (absolute value) electron-nucleus attraction energy for a carbon atom, total charge-weighted partial positively charged surface area, surface area of hydrogen donor atoms, and gravitation index (calculated over all bonds). In the present study, a new QSPR relationship is developed for the refractive indices of a diverse set of polymers. The descriptors selected for this polymer data set are then compared with the descriptors selected in our previous study[14] for correlation of the refractive indices of low molecular weight compounds.

METHODOLOGY

The refractive index data for 95 essentially amorphous polymers, measured at room temperature (298 K), were taken from a published compilation (Table 1).[5] The polymers chosen for the data set cover a wide range of refractive index values and represent a diverse set of chemical structures. The majority of the polymers fall into the classes of homochain polymers (only carbon atoms in the main chain) and polyoxides, but several polyamides and polycarbonates were also included. The data set contained large subsets of polyethylenes, polyacrylates, polymethacrylates, polystyrenes, polyethers, and polyoxides. The entire set was characterized by a high degree of structural variety; the functionalities represented in the side chains include halides, cyanides, carboxylates, acetates, amides, ethers, alcohols, hydrocarbon chains, aromatic, and nonaromatic rings.

For high molecular weight polymers it is at best extremely difficult to calculate descriptors directly. Instead, we used the repeating unit end-capped by hydrogen atoms as the small representative model structure. All polymer chains were assumed to be terminated by a hydrogen atom.

The three-dimensional structure of the repeating unit for each polymer was drawn and preoptimized using the PC-MODEL program.[15] The preoptimized structures were then fully optimized with the semiempirical AM1 method[16] using the MOPAC 6.0 program[17] to obtain the necessary quantum-chemical descriptors for the further calculations. More than 800 constitutional, topological,[18] geometrical, and quantum chemical[19] descriptors were calculated for the repeating unit from the results of the semiempirical calculations using the CODESSA (COmprehensive DEscriptors for Structural and Statistical Analysis)[7] program.

REFRACTIVE INDICES OF POLYMERS BY QSPR

J. Chem. Inf. Comput. Sci., Vol. 38, No. 6, 1998 **1173**

Table 1. Experimental and Calculated Refractive Index Values

compound	representative structure	exp. n	calcd n	Δn
poly(ethylene)	HCH_2CH_2H	1.4760	1.4714	0.0046
poly(acrylic acid)	$HCH_2CH(COOH)H$	1.5270	1.4927	0.0343
poly(methyl acrylate)	$HCH_2CH(COOMe)H$	1.4790	1.4852	0.0062
poly(ethyl acrylate)	$HCH_2CH(COOEt)H$	1.4685	1.4830	−0.0145
poly(vinyl alcohol)	$HCH_2CH(OH)H$	1.5000	1.4731	0.0269
poly(vinyl chloride)	$HCH_2CH(Cl)H$	1.539	1.5353	0.0037
poly(acrylonitrile)	$HCH_2CH(CN)H$	1.5200	1.5405	−0.0205
poly(vinyl acetate)	$HCH_2CH(OCOMe)H$	1.4670	1.5001	−0.0331
poly(styrene)	$HCH_2CH(C_6H_5)H$	1.5920	1.5988	−0.0068
poly(2-chlorostyrene)	$HCH_2CH(C_6H_4Cl)H$	1.6098	1.6184	−0.0086
poly(2-methylstyrene)	$HCH_2CH(C_6H_4Me)H$	1.5874	1.5946	−0.0072
poly(propylene)	$HCH_2CH(Me)H$	1.4735	1.4791	−0.0056
poly(ethoxyethylene)	$HCH_2CH(OEt)H$	1.4540	1.4818	−0.0278
poly(n-butyl acrylate)	$HCH_2CH(COOC_4H_9)H$	1.4660	1.4760	−0.0100
poly(vinyl hexyl ether)	$HCH_2CH(OC_6H_{13})H$	1.4591	1.4670	−0.0079
poly(1,1-dimethylethylene)	$HCH_2C(Me)_2H$	1.5050	1.4738	0.0312
poly(methyl methacrylate)	$HCH_2C(Me)(COOMe)H$	1.4893	1.4852	0.0041
poly(ethyl methacrylate)	$HCH_2C(Me)(COOEt)H$	1.4850	1.4811	0.0039
poly(isopropyl methacrylate)	$HCH_2C(Me)(COOCH(Me)_2)H$	1.4728	1.4804	−0.0076
poly(2-chloroethyl methacrylate)	$HCH_2C(Me)(COOC_2H_4Cl)H$	1.5170	1.5093	0.0077
poly(phenyl methacrylate)	$HCH_2C(Me)(COOC_6H_5)H$	1.5706	1.5779	−0.0073
poly(tetrafluoroethylene)	HCF_2CF_2H	1.3500	1.3379	0.0121
poly(chlorotrifluoroethylene)	$HCFClCF_2H$	1.3900	1.4249	−0.0349
poly(oxymethylene)	$HOCH_2H$	1.4800	1.4730	0.0070
poly(oxyethylene)	$HOCH_2CH_2H$	1.4563	1.4752	−0.0189
poly(ε-caprolactam)	$H(CH_2)_5C(O)NHH$	1.5300	1.5112	0.0188
poly(ethylene terephthalate)	$H(CH_2)_2OC(O)C_6H_4COOH$	1.5750	1.5621	0.0129
poly(vinyl n-octyl ether)	$HCH_2CH(OC_8H_{17})H$	1.4613	1.4598	0.0015
poly(vinyl n-decyl ether)	$HCH_2CH(OC_{10}H_{21})H$	1.4628	1.4516	0.0112
poly(vinyl n-pentyl ether)	$HCH_2CH(OC_5H_{11})H$	1.4590	1.4689	−0.0099
poly(vinyl 2-ethylhexyl ether)	$HCH_2CH(OCH_2CH (Et)(C_4H_9))H$	1.4626	1.4630	−0.0004
poly(vinyl n-butyl ether)	$HCH_2CH(OC_4H_9)H$	1.4563	1.4746	−0.0183
poly(vinyl isobutyl ether)	$HCH_2CH(OCH_2CH(Me)_2)H$	1.4507	1.4736	−0.0229
poly(vinyl sec-butyl ether)	$HCH_2CH(OCH(Me)(Et))H$	1.4740	1.4783	−0.0043
poly(isobutyl methacrylate)	$HCH_2C(Me)(COOCH_2CH(Me)_2)H$	1.4770	1.4927	−0.0157
poly(n-hexyl methacrylate)	$HCH_2C(Me)(COOC_6H_{13})H$	1.4813	1.4664	0.0149
poly(n-butyl methacrylate)	$HCH_2C(Me)(COOC_4H_9)H$	1.4830	1.4740	0.0090
poly(4-methyl-1-pentene)	$HCH_2C(CH_2CH(Me)_2)H$	1.4650	1.4724	−0.0074
poly(vinyl chloroacetate)	$HCH_2CH(OC(O)CH_2Cl)H$	1.5130	1.5251	−0.0121
poly(n-propyl methacrylate)	$HCH_2C(Me)(COOC_3H_7)H$	1.4840	1.4757	0.0083
poly[oxy(2,6-dimethyl-1,4-phenylene)]	$HOC_6H_2(Me)_2H$	1.5750	1.5961	−0.0211
poly(p-xylylene)	$HCH=CHC_6H_4H$	1.6690	1.6486	0.0204
poly(vinyl butyral)	$HCH_2CH(OC(O)C_3H_7)H$	1.4850	1.5104	−0.0254
poly(vinyl benzoate)	$HCH_2CH(OC(O)C_6H_5)H$	1.5775	1.5786	−0.0011
poly(N-vinylpyrrolidone)	$HCH_2CH(NC_4OH_6)H$	1.5300	1.5361	−0.0061
poly[oxy(methylphenylsilylene)]	$HOSi(Me)(C_6H_5)H$	1.5330	1.5827	−0.0497
poly(vinylidene fluoride)	HCH_2CF_2H	1.4200	1.4023	0.0177
poly(trifluoroethyl acrylate)	$HCH_2CH(COOCH_2CF_3)H$	1.4070	1.4061	0.0009
poly(2,2,2-trifluoro-1-methylethyl methacrylate)	$HCH_2CH(Me)(COOCH(Me)CF_3)H$	1.4185	1.4069	0.0116
poly(trifluoroethyl methacrylate)	$HCH_2C(Me)(COOCH_2CF_3)H$	1.4370	1.4104	0.0266
poly(N-methyl methacrylamide)	$HCH_2C(Me)(CONMe)H$	1.5398	1.5211	0.0187
poly(N-vinylcarbazole)	$HCH_2CH(NC_{12}H_8)H$	1.6830	1.6816	0.0014
poly(α-vinylnaphthalene)	$HCH_2CH(C_{10}H_9)H$	1.6818	1.6500	0.0318
poly(styrene sulfide)	$HSCH_2CH(C_6H_5)H$	1.6568	1.6337	0.0231
poly(pentabromophenyl methacrylate)	$HCH_2C(Me)(C_6Br_5)H$	1.7100	1.7009	0.0091
poly(phenyl α-bromoacrylate)	$HCH_2C(Br)(COOC_6H_5)H$	1.6120	1.6271	−0.0151
poly(2,6-dichlorostyrene)	$HCH_2CH(C_6H_3Cl_2)H$	1.6248	1.6206	0.0042
poly(chloro-p-xylylene)	$HCH=CHC_6H_4ClH$	1.6290	1.6208	0.0082
poly(β-naphthyl methacrylate)	$HCH_2C(Me)(COOC_{10}H_9)H$	1.6298	1.6274	0.0024
poly(sec-butyl α-bromoacrylate)	$HCH_2C(Br)(COOCH(Me)(Et))H$	1.5420	1.5311	0.0109
poly(2-bromoethyl ethacrylate)	$HCH_2C(Et)(COOC_2H_4Br)H$	1.5426	1.5300	0.0126
poly(methyl α-bromoacrylate)	$HCH_2C(Br)(COOMe)H$	1.5672	1.5400	0.0272
poly(ethylmercaptyl methacrylate)	$HCH_2C(Me)(COSEt)H$	1.5470	1.5525	−0.0055
poly(benzyl methacrylate)	$HCH_2C(Me)(COOCH_2C_6H_5)H$	1.5679	1.5702	−0.0023
poly[oxy(methyl-n-hexylsilylene)]	$HOSi(Me)(C_6H_{13})H$	1.4450	1.4779	0.0249
poly(propylene oxide)	$HOCH(Me)CH_2H$	1.4570	1.4736	−0.0166
poly(3-butoxypropylene oxide)	$HOCH(CH_2OC_4H_9)CH_2H$	1.4580	1.4548	0.0032
poly(3-hexoxypropylene oxide)	$HOCH(CH_2OC_6H_{13})CH_2H$	1.4590	1.4562	0.0028
poly(4-fluoro-2-trifluoromethylstyrene)	$HCH_2CH(C_6H_3F(CF_3))H$	1.4600	1.4825	−0.0225
poly(propylene sulfide)	$HSCH(Me)CH_2H$	1.5960	1.5674	0.0286
poly(p-bromophenyl methacrylate)	$HCH_2C(Me)(COOC_6H_4Br)H$	1.5964	1.6024	−0.0060
poly(vinylidene chloride)	HCH_2CCl_2H	1.6000	1.5688	0.0312
poly(pentachlorophenyl methacrylate)	$HCH_2C(Me)(COOC_6Cl_5)H$	1.6080	1.6555	−0.0475

Table 1 (Continued)

compound	representative structure	exp. n	calcd n	Δn
poly(N-benzyl methacrylamide)	$HCH_2C(Me)(CONHCH_2C_6H_5)$	1.5965	1.5918	0.0047
poly(trifluorovinyl acetate)	$HCF_2CF(OC(O)Me)H$	1.3750	1.3891	−0.0141
poly($tert$-butyl-methacrylate)	$HCH_2C(Me)(COOC(Me)_3)H$	1.4638	1.4773	−0.0135
poly(vinyl methyl ether)	$HCH_2CH(OMe)H$	1.4670	1.4816	−0.0146
poly(3,3,5-trimethylcyclohexyl methacrylate)	$HCH_2C(Me)(COOC_9H_{17})H$	1.4850	1.4810	0.0040
poly(3-methylcyclohexyl methacrylate)	$HCH_2C(Me)(COOC_7H_{13})H$	1.4947	1.4804	0.0143
poly(4-methylcyclohexyl methacrylate)	$HCH_2C(Me)(COOC_7H_{13})H$	1.4975	1.4815	0.0160
poly(ethyl α-chloroacrylate)	$HCH_2C(Cl)(COOEt)H$	1.5020	1.5184	−0.0164
poly(N-methylmethacrylamide)	$HCH_2C(Me)(CONMe)H$	1.5135	1.5119	0.0016
poly(methyl α-chloroacrylate)	$HCH_2C(Cl)(COOMe)H$	1.5170	1.5222	−0.0052
poly(1,3-dichloropropyl methacrylate)	$HCH_2C(Me)(COOC_3H_5Cl_2)H$	1.5270	1.5343	−0.0073
poly(cyclohexyl α-bromoacrylate)	$HCH_2C(Br)(COOC_6H_{11})H$	1.5420	1.5331	0.0089
poly(1-phenylethyl methacrylate)	$HCH_2C(Me)(COOCH(C_6H_4)Me)H$	1.5487	1.5596	−0.0109
poly(2,3-dibromopropyl methacrylate)	$HCH_2C(Me)(COOC_3H_5Br_2)H$	1.5739	1.5618	0.0121
poly(o-chlorobenzyl methacrylate)	$HCH_2C(Me)(COOCH_2C_6H_4Cl)H$	1.5823	1.5810	0.0013
poly(o-methoxystyrene)	$HCH_2CH(C_6H_4OMe)H$	1.5932	1.5821	0.0111
poly(p-methoxystyrene)	$HCH_2CH(C_6H_4OMe)H$	1.5967	1.5881	0.0086
poly(ethylene succinate)	$HCH_2CH(OC(O)C_2H_4COOH)H$	1.4744	1.4670	0.0074
poly(vinyl formate)	$HCH_2CH(OC(O)H)H$	1.4757	1.4962	−0.0205
poly(2-fluoroethyl methacrylate)	$HCH_2C(Me)(COOC_2H_4F)H$	1.4768	1.4582	0.0186
poly(cyclohexyl α-chloroacrylate)	$HCH_2C(Cl)(COOC_6H_{11})H$	1.5320	1.5085	0.0235
poly(2-bromoethyl methacrylate)	$HCH_2C(Me)(COOC_2H_4Br)H$	1.5426	1.5280	0.0146

Table 2. Best Five Parameter Correlation for Refractive Index[a]

X	ΔX	t-test	descriptor
1.000			intercept
0.118E-01	0.154E-02	−7.682	HOMO−LUMO energy gap
0.574E-03	0.362E-04	15.881	AM1 heat of formation
0.167E-01	0.513E-03	32.556	max nuclear repulsion for a C−H bond
0.477E-03	0.480E-04	9.939	partial negative surface area [Zefirov's PC]
−0.260	0.298E-01	−8.740	relative number of F atoms

[a] ($R^2 = 0.940$, $F = 282.13$, and $s^2 = 0.000\ 313$).

The QSPR models were developed using both the *heuristic* and the *best multilinear regression analysis* methods available in the framework of the CODESSA program.[9] In both cases, a preselection of descriptors was carried out, by removing descriptors having an essentially constant value for all structures. The final QSPR model was selected on the basis of the highest correlation coefficient (R^2), the lowest standard error, and the relevance of involved descriptors to refractive index as a physical phenomena. Another important criteria for the model selection was the intercept value, since the refractive index in a vacuum is unity. Assuming that all the descriptors involved in the QSPR model have zero values in a vacuum, the intercept of the respective (multi)-linear relationship should be determined by the refractive index of a vacuum. Therefore, we also used a modified *best multilinear regression analysis* program that fixed the intercept value to one during regression analysis. The stability of every potential model was tested against the cross-validated correlation coefficient (R^2_{CV}). The R^2_{CV} describes the stability of a regression model obtained by focusing on the sensitivity of the model to the elimination of any single data point.

RESULTS AND DISCUSSION

The final QSPR model with a correlation coefficient of 0.940 was developed from a preselected pool of more than 655 CODESSA descriptors. The model consisted of four quantum-chemical descriptors and one constitutional descriptor as follows: (i) HOMO−LUMO energy gap, (ii) AM1 calculated heat of formation, (iii) maximum nuclear repulsion

for a C−H bond, (iv) partial negative surface area (PNSA) calculated from Zefirov's partial charges, and (v) the relative number of F atoms (for details, see Table 2).

The HOMO−LUMO energy gap is defined as the energy difference between highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO). The refractive index and the HOMO−LUMO energy gap are both related to the polarizability of the molecule. A small difference between HOMO and LUMO energies usually means that the molecule is easily polarized. This particular descriptor was also involved in our previous QSPR treatment of low molecular organic compounds. Herve et al. showed that an empirical relationship exists between the refractive index and the energy gap in semi-conductors.[20]

The AM1 calculated heat of formation reflects the thermodynamic stability of the polymer. Its emergence in the correlation equation for the refractive index is possibly connected with the "looseness" of the electrons in the molecule that is interacting with the electromagnetic radiation. The positive value of the corresponding regression coefficient (cf. Table 2) indicates that the compounds that are less stable (higher heats of formation) possess higher refractive indices. Apparently, the electronic distribution in these molecules is, on average, more flexible to interact with light.

The maximum nuclear repulsion for a C−H bond is the maximal nuclear repulsion energy (E_{nn}) between a pair of bonded carbon and hydrogen nuclei. This nuclear repulsion energy is calculated by eq 3, where Z is the nuclear (core)

Table 3. Descriptor Coefficients Calculated for Three Subsets

sets I, II	sets II, III	sets I, III	descriptor
1.000	1.000	1.000	intercept
−0.120E-01	−0.118E-01	−0.127E-01	HOMO−LUMO energy gap
0.586E-03	0.570E-03	0.548E-03	AM1 heat of formation
0.168E-01	0.165E-01	0.170E-01	max nuclear repulsion for a C−H bond
0.458E-03	0.484E-03	0.431E-03	partial negative surface area [Zefirov's PC]
−0.2320	−0.255	−0.2471	relative number of F atoms

charge and the R is the distance between the carbon and hydrogen atoms. This descriptor depends on the reciprocal of the C−H bond length and thus possibly encodes the information about the hybridization of the carbon atoms, since the carbon−hydrogen bond length varies depending whether the carbon atom is in the sp^3, sp^2, or sp hybridization state.

$$E_{nm}(CH) = \frac{Z_C Z_H}{R_{CH}} \qquad (3)$$

The partial negative surface area (PNSA) is an electrostatical descriptor calculated from the Zefirov's partial charges and is defined as a sum over the surface areas of the negatively charged atoms. This descriptor encodes information about the charge distribution in the repeating unit. The PNSA is dependent on the size of the repeating unit; the squared correlation coefficient of 0.735 shows a moderate intercorrelation between PNSA and the molecular weight of repeating unit. Thus PNSA also describes molecular size related bulk properties of repeating units linked into the polymer chain.

The relative number of F atoms is defined as a ratio between the number of fluorine atoms and the total number of atoms in the repeating unit. This descriptor is required due to the extraordinary chemical nature of the fluorine. Fluorine containing polymers have usually very low refractive index values, and the negative slope for the relative number of fluorine atoms is in good agreement with this trend. The use of quantum-chemical descriptors alone appears to overestimate the refractive index for this set of polymers.

The model as described above shows a standard error of 0.018. The average prediction error is 0.9%, and the highest prediction error is 3.2%. The cross-validated correlation coefficient ($R_{CV}^2 = 0.934$) shows the stability of the model. An alternative method for cross-validation was also used to test the stability of the model. The data set of experimental refractive indices was divided into three subsets according to their magnitude. When two of the subsets were combined and the QSPR model recalculated using the same descriptors but newly optimized regression coefficients, the predicted refractive indices for the third subset gave a squared correlation coefficient of 0.906. We applied similar procedures to calculate the squared correlation coefficients (0.959 and 0.951) for the other two subsets. The average correlation coefficient for the three subsets was 0.938, and the descriptor coefficients were essentially constant (see Table 3).

A comparison between the QSPR model developed in the present paper for polymers with the model previously found[14] for low molecular weight organic compounds shows that the HUMO−LUMO energy gap is a common descriptor for both data sets. Several of the other descriptors describe similar types of physicochemical interactions. Thus, both QSPR models include electrostatic descriptors which describe the charge distribution in the molecule or repeating unit of the polymer (partial negative surface area for low molecular weight organic compounds, partial positive surface area, and hydrogen donor dependent surface area for polymers). Also, the lowest E−N attraction for a C atom (for low molecular weight organic compounds) and the strongest nuclear repulsion for a C−H bond (for polymers) are both descriptors that can be related to the hybridization of the carbon atoms. The differences in the descriptors selected for the low and high molecular weight models may in part be done to the variation of physical interactions in different media, e.g., solid phase versus liquid phase.

Bicerano's QSPR model consisted of 10 topological and constitutional descriptors;[5] our QSPR model is quite distinct as it comprises four general quantum-chemical descriptors, augmented by one constitutional descriptor. Bicerano's model implies that the refractive index for a vacuum should be 1.885. The comparison of statistical parameters shows better statistical quality in Bicerano's model ($R^2 = 0.955$ versus $R^2 = 0.940$, $s = 0.0157$ versus $s = 0.0177$), but this is not surprising in view of the fact that the number of descriptors involved in this correlation equation is twice as large (10 instead of five) as in our equation. We have also attempted to correlate topological and constitutional descriptors with the refractive index and verified that results comparable with Bicerano's QSPR model[5] can be reproduced if a sufficient number of topological and constitutional descriptors is used. On the other hand, improvement of results by increasing the number of descriptors in the correlation equation should be considered with care, since overfitting and chance correlations may in part be due to such an approach.

CONCLUSION

It is evident that the QSPR approach can be applied to develop successful QSPR models for polymers. The five-parameter QSPR model, proposed in present study, can predict the refractive index values of structurally diverse polymers with a significant degree of confidence (the average prediction error is 0.9%). The model employs only theoretical descriptors calculated from structure of repeating units and is thus applicable to not yet synthesized polymers. Therefore, this QSPR model should be useful in development of new polymeric materials.

ACKNOWLEDGMENT

REFERENCES AND NOTES

(1) Van Krevelen, D. W. In *Properties of Polymers: Correlation with Chemical Structure*; Elsevier: Amsterdam, 1972; Chapter 11, p 195.

(2) Askadskii, A. A. Structure−Property Relationships in Polymers: A Quantitative Analysis. *Polym. Sci., Ser. B.* **1995**, *37*, 66−88.

(3) Agrawal, A. K.; Jenekhe, S. A. Thin-Film Processing and Optical Properties of Conjugated Rigid-Rod Polyquinolines for Nonlinear Optical Applications. *Chem. Mater.* **1992**, *4*, 95−104.

(4) Yang, C.-J.; Jenekhe, S. A. Group Contribution to Molar Refraction and Refractive Index of Conjugated Polymers. *Chem. Mater.* **1995**, *7*, 1276−1285.

(5) Bicerano, J. In *Prediction of Polymer Properties*, 2nd ed.; Marcel Dekker: New York, 1996.

(6) Mekenyan, O.; Dimitrov, S.; Bonchev, D. Graph-Theoretical Approach to the Calculation of Physico-Chemical Properties of Polymers. *Eur. Polym. J.* **1983**, *19*, 1185−1193.

(7) Katritzky, A. R.; Lobanov, V. S.; Karelson, M. *CODESSA, Reference Manual,* University of Florida, 1994.

(8) Katritzky, A. R.; Lobanov, V. S.; Karelson, M. QSPR: The Correlation and Quantitative Prediction of Chemical and Physical Properties from Structure. *Chem. Soc. Rev.* **1995**, 279−287.

(9) Katritzky, A. R.; Mu, L.; Lobanov, V. S.; Karelson, M.Correlation of Boiling Points with Molecular Structure. 1. A Training Set of 298 Diverse Organics and a Test Set of 9 Simple Inorganics. *J. Phys. Chem.* **1996**, *100*, 10400−10407.

(10) Katritzky, A. R.; Maran, U.; Karelson, M.; Lobanov, V. S. Prediction of Melting Points for the Substituted Benzenes: A QSPR Approach. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 913−919.

(11) Huibers, P. D. T.; Lobanov, V. S.; Katritzky, A. R.; Shah, D. O.; Karelson, M. Prediction of Critical Micelle Concentration Using a Quantitative Structure−Property Relationship Approach. *J. Colloid Interface Sci.* **1997**, *187*, 113−120.

(12) Katritzky, 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.

(13) Katritzky, A. R.; Sild, S.; Lobanov, V.; Karelson, M. QSPR Correlation of Glass Transition Temperatures of High Molecular Weight Polymers. *J. Chem. Inf. Comput. Sci.* **1997**, accepted.

(14) Katritzky, A. R.;Sild, S.; Karelson, M. A General QSPR Treatment of the Refractive Index of Organic Compounds. *J. Chem. Inf. Comput. Sci.* **1998**, submitted.

(15) *PCMODEL User Manual*; Serena Software: 1992.

(16) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.

(17) Stewart, J. J. P. *MOPAC 6.0*; 1989; QCPE No. 455.

(18) Kier, L. B.; Hall, L. H. In *Molecular Connectivity in Structure−Activity Analysis*; Research Studies Press Ltd.: Letchworth, 1986.

(19) Karelson, M.; Lobanov, V. S.; Katritzky, A. R. Quantum-Chemical Descriptors in QSAR/QSPR Studies. *Chem. Rev.* **1996**, *96*, 1027−1043.

(20) Hervé, P.; Vandamme, L. K. J. General Relation Between Refractive Index and Energy Gap in Semiconductors. *Infrared Phys. Technol.* **1994**, *35*, 609−615.

CI980087W