Available online at www.sciencedirect.com

SCIENCE DIRECT®

Polymer 46 (2005) 9443–9451

polymer

www.elsevier.com/locate/polymer

QSPR studies of polyvinyls by density functional theory

Xinliang Yu, Xueye Wang*, Jinwei Gao, Xiaobing Li, Hanlu Wang

College of Chemistry, Xiangtan University, Xiangtan, Hunan 411105, People's Republic of China

Received 25 May 2005; received in revised form 10 July 2005; accepted 12 July 2005
Available online 8 August 2005

Abstract

Density functional theory (DFT) calculations are carried out for polyvinyls repeating units at the B3LYP/6-31G(d) level, and the calculated results of E_T, E_{int}, C_v, S, Q_{ii}, μ, α and q^- are used to predict V (298 K), P_s, F_d, R_{LL}, χ, H_{vsum}, U_R and U_H. Multiple linear stepwise regression analysis is used to generate eight more physically meaningful quantitative structure–property relationship (QSPR) models having correlation coefficient R of 0.996 for V (298 K), 0.998 for P_s, 0.997 for F_d, 0.997 for R_{LL}, 0.997 for χ, 0.992 for H_{vsum}, 0.992 for U_R and 0.991 for U_H, and the conclusions are in consistence with theoretical analysis. Investigated results indicate QSPR models given here are easy to apply and have good predictive capability.

Keywords: Polyvinyls; QSPRs; Density functional theory

1. Introduction

Quantitative structure–property relationship (QSPR) models of polymers are the theoretical basis for polymeric molecular designs and material designs, and have the ability to survey a list of possible candidates and exclude ones that do not fall into the desired property range for the application. In particular, no time is wasted on synthesis and testing of new materials that are deemed inappropriate by QSPR models. With recent progress in computational hardware and the development of molecular quantum mechanical calculations, the exponential growth in the number of papers dealing with QSPR studies for polymers clearly demonstrates the rapid progress in this area. Hamerton et al. [1] have used molecular modeling and molecular orbital to find molecular descriptors that could be used to derive an empirical equation to describe the glass transition temperature (T_g) of two related classes of poly(arylene ether)s and succeeded in predicting the thermal characteristics of another polymer of the same type using the equation. Katritzky at el. [2] have developed a QSPR model having a coefficient of determination of 0.946 to predict T_gs for a diverse set of 88 polymers with comprehensive descriptor for structural and statistical analysis (CODESSA) program. Cao and Lin [3] have tested the same set of 88 polymers against a correlation involving their own chosen descriptors in an attempt to derive a more physically meaningful QSPR with a coefficient of determination of $R^2 = 0.9056$. Faulon et al. [4] have constructed QSPRs, which could be used to efficiently predict transport properties in amorphous polymeric material from molecular dynamics simulation since only bulk modulus and/or cohesive energy need to be determined from a simulation. Kholodovych et al. [5] have presented a surrogate model for the prediction of cellular response to the surface of biodegradable polymers. The prediction of their model, when tested against experimental results, has shown a high degree of accuracy that was sufficient for rational design of polymeric materials for biomedical applications. Therefore, their models can provide direct guidance to the synthetic chemist in biomaterial design. Zhang et al. [6] have built a QSPR model, based on the structural analysis of polymers, to predict the refractive index of linear polymers. It must be noted that Morrill et al. [7] apply the AM1 method within AMPAC and CODEESSA to calculate molecular properties to derived descriptors, which have strongly predictive nature and is a testimony to the role that semi empirical methods play in studying such large molecular systems. In addition, many authors have studied QSPRs for polymers

---

* Corresponding author. Tel.: +86 732 829 2206; fax: +86 732 829 2477.
E-mail address: wxueye@xtu.edu.cn (X. Wang).

0032-3861/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.polymer.2005.07.039

using artificial neural networks (ANNs) and obtained valuable results [8–12]. The purpose of this article is to study QSPRs for Polyvinyls using DFT B3LYP method. 8 QSPR models, based on quantum chemical descriptors, are obtained to predict V (298 K), P_s, F_d, R_{LL}, χ, H_{vsum}, U_R and U_H of polyvinyls. Comparison of our results with other existing QSPR models was also carried out.

2. Data set

Table 1 shows the data set for 39 polymers with structure –(C^1H_2–$C^2R^3R^4$)–, which are taken from Ref. [13]. These proprieties are the molar volume V (298 K), the molar parachor P_s, the dispersion component F_d of the molar attraction constant, the molar refraction R_{LL}, the molar diamagnetic susceptibility χ, the molar viscosity–temperature function H_{vsum}, the molar Rao function U_R and the molar Hartmann function U_H. The molar volume V (298 K) is defined as the space occupied by one mole of polymeric repeat units at the room temperature ($T=298$ K). The molar parachor P_s is used to calculate the surface tension with the equation: $\gamma(T) \approx [P_s/V(T)]^4$. The mole attraction constant F is used to express the cohesive energy: $E_{coh} \approx F^2/V(298\ \mathrm{K})$. The cohesive energy E_{coh} can be divided into the dispersion, polar and hydrogen bonding components. Just like E_{coh} itself, the mole attraction constant F also has three components, relate to the dispersion (F_d), polar (F_p) and hydrogen bonding (F_h). The molar refraction R_{LL} is used to estimate the refractive index n with equation: $n=\sqrt{(V+2R_{LL})/(V-R_{LL})}$. The molar diamagnetic susceptibility χ is used to calculate the magnetic susceptibility. The molar viscosity–temperature function H_{vsum} is useful in estimating the activation energy

Table 1

Experimental data of properties for polymers [13]

No.	Polymers	V (298 K)	P_s	F_d	R_{LL}	χ	H_{vsum}	U_R	U_H
1	Polyethylene	33.1	78.0	540	9.298	22.70	840	1760	1350
2	Poly(vinyl alcohol)	35.0	98.0	560	10.723	27.85	2208	1930	1545
3	Poly(vinyl fluoride)	–	86.6	570	–	26.95	1542	1830	1445
4	Poly(vinyl chloride)	45.1	115.2	800	14.288	38.85	2750	2480	2125
5	Poly(vinyl bromide)	–	128.9	900	–	47.85	4969	2610	2125
6	Poly(vinylidene fluoride)	40.0	95.2	640	9.025	31.55	2239	1930	1575
7	Poly(vinylidene chloride)	58.4	152.4	1100	19.275	51.35	4580	3230	2710
8	Poly(vinyl acetate)	72.4	182.0	1160	–	48.85	2930	3980	3225
9	Poly(vinyl ethyl ether)	76.7	176.0	1140	20.199	51.20	2380	4010	3300
10	Poly(*N*-vinyl pyrrolidone)	88.9	246.9	1660	–	67.74	3985	5272	4181
11	Poly(vinyl propionate)	98.1	221.0	1430	24.864	60.20	3350	4860	3900
12	Poly(vinyl *n*-butyl ether)	108.1	254.0	1680	29.497	73.90	3220	5770	4650
13	Poly(vinyl *sec*-butyl ether)	108.9	254.0	1640	29.456	74.70	3440	5860	4800
14	Poly(vinyl methyl ether)	–	137.0	870	15.496	39.85	1960	3130	2625
15	Poly(vinylide bromide)	–	179.8	1300	–	69.35	9112	3490	2710
16	Poly(vinyl methyl ketone)	62.6	–	–	18.696	–	–	3630	2925
17	Poly(vinyl *iso*-butyl ether)	107.7	–	–	29.495	–	–	–	–
18	Poly(vinyl *n*-pentyl ether)	124.4	–	–	34.146	–	–	–	–
19	Poly(vinyl *n*-octyl ether)	171.0	–	–	48.093	–	–	–	–
20	Poly(vinyl *n*-decyl ether)	208.7	–	–	57.391	–	–	–	–
21	Poly(vinyl *n*-dodecyl ether)	238.1	–	–	66.689	–	–	–	–
22	Poly(*N*-vinyl carbazole)	161.0	427.7	3196	–	126.30	8565	10261	8138
23	Poly(vinyl isopropyl ether)	93.2	–	–	–	–	–	–	–
24	Poly(*o*-vinyl pyridine)	–	–	–	–	–	–	5655	4485
25	Poly(vinyl benzoate)	–	315.9	2170	40.081	87.35	5470	7230	5745
26	Poly(vinyl cyclohexane)	116.0	283.9	1970	34.951	85.85	2728	6380	5045
27	Poly(β-vinyl naphthalene)	140.0	–	–	–	–	–	–	–
28	Poly(α-vinyl naphthalene	137.7	360.0	2590	–	103.97	6452	8623	6839
29	Poly(vinyl *n*-butyl sulfide)	118.6	–	–	–	–	–	5920	4790
30	Poly(vinyl pivalate)	–	298.8	1930	–	84.85	4550	6830	5370
31	Poly(vinyl chloroacetate)	83.1	–	–	25.265	–	–	–	–
32	Poly(vinyl cyclopentane)	–	–	1700	–	–	–	5480	4346
33	Poly(vinyl sulfuric acid)	–	–	1042	–	–	4396	3210	2670
34	Poly(vinyl 2-ethylhexyl ether)	172.9	–	–	48.055	–	–	–	–
35	Poly(5-vinyl 2-methyl-pyridine)	–	–	–	–	–	–	6584	5222
36	Poly(*p*-vinyl pyridine)	–	241.5	–	–	69.01	3985	5655	4485
37	Poly(vinyl *n*-hexyl ether)	138.6	–	–	38.795	–	–	–	–
38	Poly(vinyl trimethylsilane)	–	–	1624		–	–	5640	4515
39	Poly(vinyl methyl sulfide)	–	–	–	–	–	–	3280	2765

The data units are cm^3/mol for V, $(cm^3/mol)(10^{-5}\ N/cm)^{1/4}$ for P_s, $J^{0.5}\ cm^{1.5}/mol$ for F_d, cm^3/mol for R_{LL}, $10^{-6}\ cm^3/mol$ for χ, $g\ J^{1/3}/mol^{-4/3}$ for H_{vsum}, $cm^{10/3}/(s^{1/3}\ mol)$ for U_R and U_H.

($E_{V(\infty)}$) for viscous flow at zero shear rate in the limit of $T \to \infty$.The molar Rao function U_R is useful in predicting the bulk modulus B and the thermal conductivity. The molar Hartmann function U_H can be used to estimate the shear modulus G.

3. Quantum chemical descriptors

It is impossible to calculate descriptors directly for entire molecule because all the polymers possess high molecular weights, but molecular descriptors calculated directly from polymeric repeating unit structure can be used on the study of QSPRs for polymers (Fig. 1). Katritzky et al. [2,14] have chosen models consisting of repeating units end-capped by hydrogen, as small, yet representative structures, to calculate the descriptors. Cao et al. [3] and Zhang et al. [6] also have adopted this method to derive the descriptors and succeeded in developing QSPRs for polymers. Therefore, we focus on the following model to calculate molecular quantum chemical descriptors.

Semi empirical quantum chemical methods (such as AM1) include empirical parameters, which come from a special system. Thus these methods is not universality for others system. DFT method does not include any empirical parameters and thus belongs to the family of ab initio methods. In the DFT method, the exact exchange term used in the Hartree–Fock method is replaced by a more general expression: The exchange correlation functional. The simplicity of the DFT stems from the fact that it uses a functional of electron density to model exchange and correlation. Further, DFT energy includes, besides the exchange contributions, also some portion of the correlation energy. In addition, DFT requires shorter computer times when compared with conventional ab initio methods with inclusion of electron correlation. Care should be taken, however, when applying DFT to molecular clusters, where the dominant part of the stabilization energy comes from the dispersion energy [15]. The reason is that DFT method does not include the dispersion energy. In this paper, the calculated models are repeating units of polyvinyls. The dispersion energy is very smaller than the total energy. Therefore, we adopt DFT method to optimize and calculate the models by the GAUSSIAN 03 program, at the B3LYP/6-31G(d) level.

We calculate two types of quantum chemical descriptors. One is thermodynamic parameters, such as the total energy E_T, the thermal energy E_{int}, the heat capacity at constant volume C_v and the entropy S. E_T is the contributions from translational (E_t), vibrational (E_v), rotational (E_r), electronic (E_e) movement, as well includes zero-point corrections. E_{int}, C_v and S are values of a system at the condition $T = 298.150$ K and $P = 1.00000$ atm, and all are the contributions from translational, vibrational and rotational movement of a molecular. The eight properties studied in this paper are extensive properties. An extensive property depends upon the size of the system. Therefore, we choose extensive descriptors E, E_{int}, C_v and S, to correlate with these extensive properties.

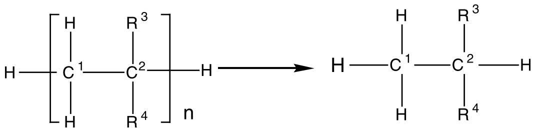

Fig. 1. The calculated models of polymers.

The other is electric parameters, including the quadrupole moment Q_{ii}, the dipole moment μ, the average polarizability of the molecule α and the net charge of the most negative atom q^-. According to Ref. [16], the elements Q_{ij} of the quadrupole moment tensor are defined as

$$Q_{ij} = \int \sigma(r) r_i r_j \mathrm{d}r \quad (i,j = x,y,z) \tag{1}$$

where $\sigma(r)$ is the charge density distribution and r_i and r_j are the component of the distance vector (x, y, z) from the molecular center. We define Q_{ii} with following equation

$$Q_{ii} = \frac{(Q_{xx} + Q_{yy} + Q_{zz})}{3} \tag{2}$$

Q_{ii} can reflect the degree of the asymmetry of charge spherical distribution in the molecular. α is defined as

$$\alpha = \frac{(\alpha_{xx} + \alpha_{yy} + \alpha_{zz})}{3} \tag{3}$$

where α_{xx}, α_{yy} and α_{zz} reflect electric perturbation in the x-, y-, and z-coordinates. α increases with the size of the species either as a result of an increase with the number of electrons or by the expansion of the molecular radius. μ is the total dipole moment and a measure of the asymmetry in molecular charge distribution. q^- can reflect the force of electrostatic interactions and also be used for the descriptor of the molecular polarity. Since the intermolecular forces energy are larger than the bond energy for a polymer, it is very important to analyze the factors affecting the intermolecular forces, which include van der Waals forces (electrostatic force, induction force, dispersion force) and hydrogen bond forces. The electrostatic force can be expressed in μ, Q_{ii}, etc. Both induction force and dispersion force are depending on α [17]. q^- can enhance intermolecular forces between the backbones of polymers as the hydrogen bond can [3]. Therefore, we chose these four electric parameters to develop QSPRs.

4. Results and discussion

By carrying out the correlation between these properties (Table 1) and eight descriptors (Table 2) using stepwise regression analysis, respectively, the regression analysis

Table 2
Quantum chemical descriptors for polymers

No.	E_T	E_{int}	C_v	S	Q_{ii}	μ	α	q^-
1	−79.8304172	49.390	9.995	56.564	−14.8853	0.0000	22.922	−0.432886
2	−155.0338053	53.080	13.387	64.386	−18.9995	1.5623	26.394	−0.613228
3	−179.0564234	45.395	11.530	62.395	−18.7139	1.7179	23.250	−0.470982
4	−539.4262724	44.649	12.413	65.654	−26.1104	2.2701	31.436	−0.435643
5	−2650.9347582	44.301	12.881	68.484	−30.9478	2.1738	38.099	−0.438196
6	−278.1249231	34.767	13.240	66.386	−23.2619	2.1116	26.919	−0.528470
7	−999.0165315	39.347	15.804	72.511	−38.0945	2.2729	41.068	−0.430556
8	−307.5978522	79.081	24.409	84.235	−37.5420	4.9613	48.124	−0.491285
9	−233.6634595	90.549	22.643	79.629	−32.1344	1.0250	48.498	−0.467381
10	−365.2602387	110.814	29.644	89.172	−49.4005	3.7102	69.913	−0.444441
11	−347.0221322	98.073	29.068	92.262	−42.0187	1.8352	58.895	−0.464483
12	−312.2906337	128.148	32.324	95.229	−45.3354	1.0006	70.155	−0.475837
13	−312.2925054	127.823	33.116	94.639	−45.4648	1.0028	69.347	−0.487260
14	−194.3443061	71.814	17.736	71.990	−25.7032	1.1453	37.491	−0.490140
15	−5222.0343369	38.757	16.717	78.111	−48.5852	2.0005	54.688	−0.435758
16	−232.4663240	74.479	19.488	73.740	−32.4567	2.8307	43.917	−0.455009
17	−312.2915224	127.889	32.819	93.762	−45.2702	1.0042	69.204	−0.472939
18	−351.6044314	146.963	37.148	102.694	−52.0378	0.9452	81.126	−0.475982
19	−469.5455456	203.335	51.641	125.756	−72.0328	0.9939	114.308	−0.476197
20	−548.1729253	240.266	59.370	135.840	−85.4039	0.9905	136.564	−0.476155
21	−626.8004483	278.573	70.911	155.084	−98.8229	0.9924	158.866	−0.476204
22	−596.0986390	154.192	47.837	106.697	−83.3121	1.8874	152.602	−0.623371
23	−272.9065010	109.367	28.310	88.182	−39.0738	1.5392	57.830	−0.529383
24	−326.9115327	94.686	23.158	77.444	−46.2467	1.6980	73.626	−0.450420
25	−499.4475483	114.683	37.144	99.549	−60.2600	1.8148	96.369	−0.494372
26	−314.5043750	148.496	32.949	88.624	−53.2023	0.0172	81.756	−0.448398
27	−464.5217458	133.794	37.576	92.128	−67.5162	0.5642	124.825	−0.454383
28	−464.5209841	134.601	39.373	95.508	−67.7182	0.4207	122.212	−0.453627
29	−635.2700256	126.471	34.390	100.094	−52.7242	1.6846	83.539	−0.453634
30	−425.6491260	135.102	40.289	105.089	−55.4266	1.9133	79.454	−0.488248
31	−767.2938960	74.479	26.705	93.240	−47.9605	2.1090	57.057	−0.462173
32	−275.1854730	129.319	28.853	87.510	−46.4936	0.0417	71.594	−0.446000
33	−780.0062176	78.418	32.276	94.338	−48.9198	1.3325	59.617	−0.714710
34	−510.0618742	231.911	51.466	121.819	−79.5074	0.9794	125.333	−0.477772
35	−366.2375647	113.705	30.940	89.827	−53.2940	2.0405	87.026	−0.519765
36	−326.9179046	96.003	26.742	84.490	−48.26703	2.6659	73.633	−0.436578
37	−390.9181549	165.797	41.941	109.676	−58.6595	0.9952	92.153	−0.476187
38	−488.5008189	118.083	37.209	97.928	−48.2456	0.0607	74.036	−0.658183
39	−517.3282203	69.903	19.919	77.108	−33.1509	1.7129	49.294	−0.583014

The parameters units are a.u. for E_T, α and q^-; 4.19×10^3 J/mol for E_{int}; 4.19 J/(mol K) for C_v, and S; Debye Å for Q_{ii}; Debye for μ.

results (Tables 3 and 4) and plots of calculated values versus experimental values (Fig. 2) have been obtained.

Q_{ii} is a negative value because Q_{xx}, Q_{yy}, and Q_{zz} are negative ones, which indicates the negative charge distribution is farther removed from the calculated model center of the nuclear charges and the model is characterized by the quadrupolarity $\{-\ +\ -\}$ in all three independent directions. The larger the descriptors E_{int}, C_v, S and α, and

Table 3
Results of stepwise regression analysis

Regression equation	No.	Statistical analysis					
		N	R	s	F	ME (%)	R_{LOO}
$V\,(298\ \text{K})=2.129+3.407C_v-2.668\mu$	1	26	0.996	4.8	1494	5.218	0.995
$P_s=-82.408+1.842\alpha+1.905S+0.239E_{int}$	2	21	0.998	6.2	1621	2.189	0.996
$F_d=310.858+19.193\alpha+1168.492q^-+5.978S$	3	23	0.997	59	973	8.224	0.992
$R_{LL}=0.469+0.412\alpha$	4	21	0.997	1.306	2942	4.589	0.996
$\chi=1.627+0.061E_{int}-1.418Q_{ii}-2.234\mu$	5	21	0.997	2.183	1013	3.692	0.996
$H_{vsum}=-1229.704-0.905E_T-34.878E_{int}+44.453S+63.585\alpha$	6	22	0.992	292	264	9.426	0.987
$U_R=1188.506+14.542E_{int}+2608.088q^-+55.610\alpha$	7	29	0.992	289	492	6.377	0.988
$U_H=924.186+11.766E_{int}+1858.584q^-+43.136\alpha$	8	29	0.991	239	443	6.960	0.986

N is the number of sample; s is the standard error; R_{LOO} is the correlation coefficients between the experimental values and the predicted values by leave-one-out (LOO) method; ME is the average predicted error.

Table 4
The characteristics of parameters appearing in regression equations

No	Descriptor	Std. error	*t*-test	Sig.	VIF
1	(Constant)	3.076	0.692	0.496	–
	C_v	0.067	51.061	0.000	1.104
	μ	0.919	−2.904	0.008	1.104
2	(Constant)	13.111	−6.285	0.000	–
	α	0.087	21.215	0.000	4.453
	S	0.238	8.002	0.000	6.586
	E_{int}	0.080	2.979	0.008	5.610
3	(Constant)	112.247	2.769	0.012	–
	α	0.748	25.650	0.000	3.671
	q^-	169.073	6.911	0.000	1.129
	S	1.672	3.576	0.002	3.899
4	(Constant)	0.610	0.769	0.451	–
	α	0.008	54.239	0.000	1.000
5	(Constant)	1.549	1.050	0.308	–
	Q_{ii}	0.049	−28.925	0.000	3.070
	μ	0.464	−4.816	0.000	1.165
	E_{int}	0.021	2.850	0.011	3.216
6	(Constant)	562.818	−2.185	0.043	–
	E_T	0.071	−12.693	0.000	1.618
	E_{int}	4.490	−7.767	0.000	7.890
	α	4.317	14.727	0.000	4.915
	S	10.396	4.276	0.001	5.790
7	(Constant)	384.549	3.091	0.005	–
	α	3.471	16.022	0.000	3.537
	E_{int}	2.746	5.296	0.000	3.474
	q^-	749.724	3.479	0.002	1.034
8	(Constant)	317.900	2.907	0.008	–
	α	2.869	15.034	0.000	3.537
	E_{int}	2.270	5.183	0.000	3.474
	q^-	619.783	2.999	0.006	1.034

the more negative E_T and Q_{ii} stand for a larger atomic numbers in the calculated model.

These QSPR models show good correlations between these properties and these descriptors, and all the correlation coefficients R are above 0.990 (Table 3). According to the statistical theory, if there is $\mathrm{VIF}_j \geq 10$ in a model, the descriptor j is strongly correlated with the others and it is not significant to explain for that model, which is not reliable. In this paper (Table 4), $\mathrm{VIF}_j < 10$, therefore, these descriptors are weakly correlated with each other and these QSPR models can be regarded as the optimal regression equations. According to the *P*-test (Table 4), all the descriptors appearing in correlation equations were significant descriptors. We predicted these properties with the leave-one-out (LOO) method for the same experimental data set in Table 1 to test the reliability of these QSPR models. The plots of predicted values versus experimental values are shown in Fig. 3. In fact Fig. 3 is very similar to Fig. 2, i.e. predicted values are very close to calculated values. We can draw a conclusion that these QSPR models are reliable. The correlation coefficient R_{LOO} (between the experimental value and predicted value) and the average predicted error *ME* are listed in Table 3.

In order to make a comparison between present wok and previous work, Table 5 shows the standard errors obtained from this paper and Sun et al.'s work [18,19]. Sun et al. [18] obtained two kinds of calculated models (Models 1A and 2A) on these properties except the molar refraction R_{LL} by artificial neural works for linear chain polymers. After this work, Sun et al. [19] further built two models (Models 1F and 2F) on these properties for the same polymeric data set by fuzzy set theory. The group average method is used to calculate the descriptors for Models 1A and 1F, the connectivity indexes method is used for Models 2A and 2F.

Regression Eq. (1) shows a good correlation between the molar volume V (298 K) and the descriptors, C_v and μ. The former is the more significant descriptor according to the *t*-test (Table 4) and increases V (298 K); While the descriptor μ deceases V (298 K). This phenomenon can be explained by that the larger the descriptor C_v indicates a larger substitute side (Fig. 1), thus increase V (298 K); In addition, the presence of a large μ could enhance intermolecular force, make molecules closer to each other, and decrease V (298 K). Therefore, V (298 K) is positive correlation with C_v, but negative with μ. Regression Eq. (1) gives a correlation coefficient R_{LOO} with 0.995. The average predicted error ME is 5.218%, which reveals the quality of regression Eq. (1) for prediction of V (298 K). Table 5 shows that the standard error ($s = 4.8\ \mathrm{cm^3/mol}$) obtained from regression Eq. (1) is smaller than that from previous every models for V (298 K) [18,19]. The standard error obtained from Model 1F is $15.1\ \mathrm{cm^3/mol}$, which is about three times as larger as that obtained from regression Eq. (1). While the s of Model 2F is 11 times larger than that of ours.

Regression Eq. (2) is of good statistical quality with $R = 0.998$. By the *t*-test the most significant descriptor appearing in regression Eq. (2) is the descriptor α, and the second and the last significant descriptor are S and E_{int}, respectively (Table 4). The descriptors all are increase P_s. The reason is that a larger α, stands for a larger atomic numbers in the calculated model, makes an increase of V; A larger S or E_{int} indicates a larger substitute side, causes an increase of V, too. While P_s can be expressed in the term V with following

Table 5
The standard errors comparison of the present paper with previous work

Property	V (298 K)	P_s	F_d	R_{LL}	χ	H_{vsum}	U_R	U_H
The present paper	4.8	6.2	59	1.306	2.183	292	289	239
Model 1A [18]	18.9	25	228	–	7.45	649	729	568
Model 2A [18]	40.5	37	235	–	5.36	719	593	674
Model 1F [19]	15.1	23	274	–	9	511	570	549
Model 2F [19]	58.4	66	491	–	24	1729	1240	1428

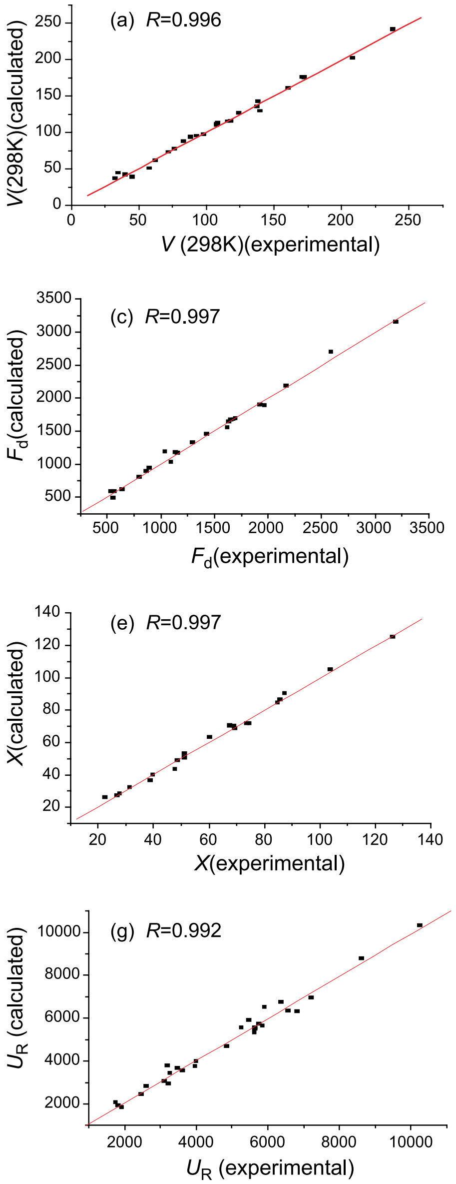

Fig. 2. Plots of calculated values versus experimental values, (a): V (298 K); (b): P_s; (c): F_d; (d): R_{LL}; (e): χ; (f): H_{vsum}; (g): U_R; (h): U_H.

equation

$$P_s = V\gamma^{1/4} \tag{4}$$

where γ is the surface tension and $\gamma^{1/4}$ is almost a constant. Therefore, P_s and V are consistent in variable trend, i.e. a larger α, S or E_{int} also indicates a larger P_s. Thus P_s is positive correlation with E_{int}, S and α. Seen from Table 3, the correlation coefficient R_{LOO} for the P_s is 0.996 and the average predicted error is 2.189%. These confirm that the predictive ability of regression Eq. (2) is excellent. The standard error is 6.2 (cm^3/mol) (10^{-5} N/cm)$^{1/4}$, which is about 10% of standard error value obtained from Model 2F, and also very smaller than that obtained from the other models (Table 5).

Regression Eq. (3) shows a high correlation coefficient with R of 0.997. As seen in Table 4, the most important descriptor used in regression Eq. (3) is α, the others descriptors are q^- and S. The polar component of cohesive energy E_p is positive correlation with α. While the molar attraction constant F is often expressed in terms of the cohesive energy E and V at room temperature.

$$F = (EV(298\ \text{K}))^{1/2} \tag{5}$$

Therefore, a larger α indicates a larger E, and results in a larger F_d. On the other hand, a larger S expresses a larger V (298 K), increase F and F_d according to Eq. (5). So F_d is positive correlation with α and S. A more negative q^- denotes a larger intermolecular force, makes molecules closer to each other, and causes a smaller V (298 K). But a larger intermolecular force also indicts a larger E. In the end, the former (the molar volume V (298 K)) may

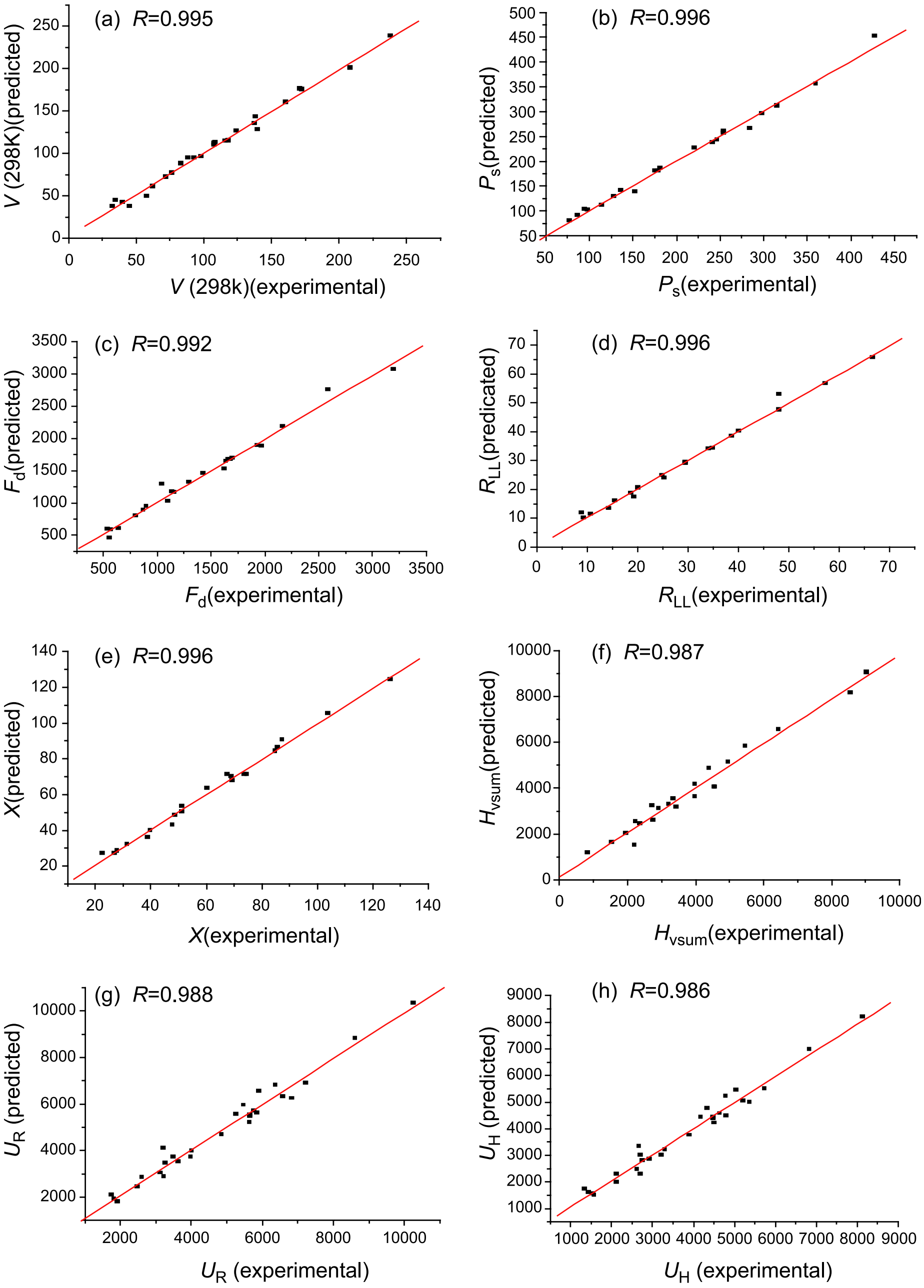

Fig. 3. Plots of predicted values versus experimental values. (a): V (298 K); (b): P_s; (c): F_d; (d): R_{LL}; (e): χ; (f): H_{vsum}; (g): U_R; (h): U_H.

dominate the variable trend of F in Eq. (5) and results in a smaller F and F_d. This is the reason why q^- increases F_d. From Table 3, it is shown that regression Eq. (3) possesses admissible predictive ability with ME = 8.224%. The standard error is 59 $J^{0.5}$ $cm^{1.5}$/mol, which is about 169 $J^{0.5}$ $cm^{1.5}$/mol smaller than the s of Model 1A [18] (Table 5). Model 1A possesses the smallest s of 228 $cm^{1.5}$/mol for F_d among the models in Table 5.

According to regression Eq. (4), α increases R_{LL}, which can be understood from the Lorenz–Lorentz equation

$$R_{LL} = \frac{N_A\alpha}{3\varepsilon_0} \qquad (6)$$

where N_A is a Avogadro constant, ε_0 is permittivity of the free space. So a larger α denotes a larger R_{LL}. Regression Eq. (4) shows a fairly high correlation coefficient with R = 0.997, in spit of a single descriptor as input. The standard error is 1.306 cm^3/mol. The average error (ME) of prediction is 4.589%, which is satisfactory.

Any materials are characterized by diamagnetic susceptibility. Polymers usually belong to the class of diamagnetic materials, and the diamagnetic susceptibility χ can be expressed as

$$\chi = -\frac{\mu_0 ZNe^2}{6m}\bar{r}^2 M \qquad (7)$$

where M is the molecule weight of the repeat unit, N is the atomic number in per unit volume, Z is the electronic number in per atom, μ_0 is the space permeability, m is the electronic mass, $\bar{r}^2$ indicates the charge distribution in a atom and '-' expresses the induced magnetization is diamagnetic. Regression Eq. (5) shows a high correlation coefficient with R of 0.997 between χ and the descriptors, Q_{ii}, μ and E_{int}. The most significant descriptor is Q_{ii}, according to the t-test (Table 4). This descriptor decreases χ, which can be understood as that a more negative Q_{ii} suggests there are more atomics in substitute R (Fig. 1), and results in a larger χ according to Eq. (7). The second most significant descriptor is μ. The descriptor μ decrease χ, since μ affects the distance of atoms and the electric Larmor precession. The last descriptor is E_{int}, which increases χ. This can be explained from Eq. (7). Regression Eq. (5) gives a good enough prediction power with ME = 3.692%. The standard error is 2.183×10^{-6} cm^3/mol, which is quite satisfactory in contrast to previous work [18,19] (Table 5). Model 2A has the smallest standard error ($s=5.36\times10^{-6}$ cm^3/mol), which is still 3.177×10^{-6} cm^3/mol larger than the value from regression Eq. (5).

Regression Eq. (6) shows a good correlation between the molar viscosity–temperature function H_{vsum} and the four descriptors α, E_{T}, E_{int} and S. The molar viscosity–temperature function H_{v} can be expressed as a sum of two terms

$$H_{\mathrm{v}} = H_{\mathrm{vsum}} + H_{\mathrm{vstr}} \tag{8}$$

H_{vsum} is a simple additive quantity estimated as a sum of the group contributions made by the structural units in the repeat unit. H_{vstr} is a structural term related to side chains, which are a special type of side group. There is a equation for H_{v}

$$H_{\mathrm{v}} = ME_{\mathrm{V}(\infty)}^{1/3} \tag{9}$$

where $E_{\mathrm{V}(\infty)}$ is the activation energy for viscous flow at zero shear rate in the limit of $T\rightarrow\infty$. For polymers a larger α, S or a more negative E_{T} would cause a stronger intermolecular force, enhance stiffness of the polymeric chains, and result in a larger $E_{\mathrm{V}(\infty)}$. Thus there are a larger H_{v} and a larger H_{vsum} by Eq. (9). In addition, a larger E_{int} denotes a smaller $E_{\mathrm{V}(\infty)}$, thus results in a smaller H_{v} and a smaller H_{vsum}, according to Eq. (9). Table 4 shows the four descriptors all are significant descriptors from the t-test. Seen in regression Eq. (6), the average predicted error is 9.426%, which shows the predictive ability of this model is generally acceptable. The standard error is 292 g J$^{1/3}$/mol$^{-4/3}$. Table 5 shows Model 2F [19] and Model 1F [19] have the largest and smallest standard error with $s=1729$ g J$^{1/3}$/mol$^{-4/3}$ and 511 g J$^{1/3}$/mol$^{-4/3}$ for H_{vsum}, respectively, which are far larger than the value in present paper.

Regression Eq. (7) displays a good correlation coefficient with $R=0.992$ between U_{R} and the descriptors, α, E_{int} and q^-. These descriptors all are significant descriptors by the t-test (Table 4). U_{R} correlates with the bulk modulus B

$$\frac{B}{\rho} = \left(\frac{U_{\mathrm{R}}}{V}\right)^6 \tag{10}$$

where ρ is the density of the polymer, V is the mole volume. While there is $B\approx 8.04\, E_{\mathrm{sub1}}/V$, E_{sub1} is the lattice energy of molecular crystals. Eq. (10) can be expressed as

$$U_{\mathrm{R}} \approx 1.4\frac{E_{\mathrm{sub1}}^{1/6}V^{7/6}}{M^{1/6}} \tag{11}$$

For polyvinyls (Fig. 1) a larger α or E_{int} indicts a larger $V^{7/6}$ and $M^{1/6}$. But $V^{7/6}$ increases faster than $M^{1/6}$, which indicts that U_{R} will increase, too. So U_{R} is positive correlation with α and E_{int}. As stated above, a more negative q^- denotes a smaller V, and decreases U_{R}. Regression Eq. (7) shows a satisfactory average predicted error with ME of 6.377%. As seen in Table 5, a very small standard error ($s=289$ cm$^{10/3}$/(s$^{1/3}$ mol)) is obtained from regression Eq. (7). While the values obtain from Models 1A, 2A, 1F and 2F in literatures [18,19] are 440, 304, 281 and 951 cm$^{10/3}$/(s$^{1/3}$ mol) larger that the value obtained from in this paper, respectively, (Table 5).

According to Table 3, regression Eq. (8) is similar to regression Eq. (7), since U_{H} can be expressed in the term U_{R}

$$U_{\mathrm{H}} \approx 0.8017U_{\mathrm{R}} \tag{12}$$

but the predictive ability of regression Eq. (8) ($ME=$ 6.960%) is slightly lower than that of regression Eq. (7) ($ME=6.377\%$). Regression Eq. (8) also shows a rather small standard error ($s=239$ cm$^{10/3}$/(s$^{1/3}$ mol)) in contrast with standard errors in literatures [18,19] (Table 5).

It should be pointed out that the correlation coefficients obtained from Bicerano's models [13] are slightly high than these reported in this work. But this is not surprising in view of the fact that the numbers of descriptors involved in Bicerano's models are 2–7.5 times as large as in our equations. For example, Bicerano's model for V (298 K) consisted of 15 topological and constitutional descriptors, our model for V (298 K) is quite distinct as it comprises only two descriptors. On the other hand, improvement of results by increasing the number of descriptors in the correlation equation should be considered with care, since over fitting and chance correlations may in part be due to such an approach [14].

5. Conclusion

Eight QSPR models are obtained using multiple linear stepwise regression and eight descriptors, E_{T}, E_{int}, C_{v}, S, Q_{ii}, μ, α and q^-, to predict the properties V (298 K), P_{s}, F_{d} R_{LL}, χ, H_{vsum}, U_{R} and U_{H} of polyvinyls. The produced QSPR models are proved to be accurate and reliable. Since these quantum chemical descriptors have clear physical meanings, express all of the electronic and geometric properties of molecules and can be obtained quickly and accurately by calculation, these QSPR models are valuable for predicting the properties and providing theoretical guidance for polymeric molecular designs and material designs.

Acknowledgements

The authors wish to acknowledge the financial supports from the Scientific Research Fund of Hunan Provincial Education Department, the Natural Science Foundation of Hunan Province and the Prominent Mid-youth Science and Technology Foundation of Hunan Province (Grant No. 04JJ1010) for the research work.

References

[1] Hamerton I, Howlin BJ, Larwood V. J Mol Graphics 1995;13:14–17.
[2] Katritzky AR, Sild S, Lobanov V, Karlson M. J Chem Inf Comput Sci 1998;38(2):300–4.
[3] Cao CZ, Lin YB. J Chem Inf Comput Sci 2003;43(2):643–50.
[4] Tokarski JS, Hopfinger AJ, Hobbs JD, Ford DM, Faulon JM. Comput Theor Polym Sci 1997;7:199–214.
[5] Kholodovych V, Smith JR, Knight D, Abramson S, Kohn J, Welsh WJ. Polymer 2004;45:7367–79.
[6] Xu J, Chen B, Zhang QJ, Guo B. Polymer 2004;45:8651–9.
[7] Morrill JA, Jensen RE, Madison PH, Chabalowski CF. J Chem Inf Comput Sci 2004;44:912–20.
[8] Joyce SJ, Osguthorpe DJ, Padgett JA, Price GJ. J Chem Soc 1995; 91(16):2491–6.
[9] Sumpter BG, Noid DW. Macromol Theor Simul 1994;3:363–78.
[10] Ulmer II CW, Smith DA, Sumpter BG, Noid DI. Comput Theor Polym Sci 1998;8:311–21.
[11] Mattioni BE, Jurs P. J Chem Inf Comput Sci 2002;42:232–40.
[12] Sumpter BG, Noid DW. Ann Rev Mater Sci 1996;26:233–77.
[13] Bicerano J. Prediction of polymer properties. 2nd ed. New York: Marcel Dekker; 1996.
[14] Katritzky AR, Sild S, Karelson M. J Chem Inf Comput Sci 1998;38: 1171–6.
[15] Hobza P, Sÿponer J. Chem Rev 1999;99(11):3247–76.
[16] Glaser R, Wu Z, Lewis M. J Mol Struct 2000;556:131–41.
[17] Staikova M, Wania F, Donaldson DJ. Atmos Environ 2004;38: 213–25.
[18] Sun H, Tang YW, Wu GS, Zhang FS, Chen XQ. Comput Appl Chem 2003;20:210–6.
[19] Sun H, Tang YW, Wu GS, Zhang FS, Chen XQ. Comput Appl Chem 2003;20:381–5.