JIE XU[1),*], BIAO CHEN[2)], HAO LIANG[3)], WEILIN XU[1)], WEIGANG CUI[1)]

Prediction of cohesion energy of polymers using topological indices

Summary — The quantitative structure-property relationship (QSPR) study of the cohesion energy (E_{coh}) values for a training set of 60 polymers was carried out using only topological indices as input parameters. A general four-parameter correlation was obtained to predict E_{coh} values by stepwise multilinear regression analysis (MLRA), with squared correlation coefficient (R^2) equal to 0.9775 and standard error of estimation (s) 2455 J/mol. The mean relative error (MRE) of E_{coh} prediction was 5.82 %. The stability of the proposed model was validated using Leave-One-Out cross-validation, randomization experiments and external test set. The model requires only topological indices for the predictions and its advantage is relative ease in calculation of descriptors which makes it easy to apply.
Key words: cohesion energy, prediction, QSPR, polymers, topological index, multilinear regression analysis.

PROGNOZOWANIE ENERGII KOHEZJI POLIMERÓW Z ZASTOSOWANIEM WSKAŹNIKÓW TOPOLOGICZNYCH
Streszczenie — Dotyczącą energii kohezji (E_{coh}) ilościową zależność struktura-właściwości (QSPR — *quantitative structure-property relationship*) wykorzystano do określenia wartości E_{coh} wzorcowej serii 60 polimerów z zastosowaniem w charakterze parametrów wejściowych wyłącznie wskaźników topologicznych. Uzyskano ogólną czteroparametrową zależność [równanie (6)] pozwalającą na prognozowanie E_{coh} metodą krokowej wieloliniowej analizy regresji (MLRA — *stepwise multilinear regression analysis*) z kwadratem współczynnika korelacji R^2 = 0,9975 i ze standardowym błędem oznaczania s = 2455 J/mol (tabele 1 i 2, rys. 1). Średni błąd względny (MRE) tego oznaczania wynosił 5,82 % (rys. 2). Trwałość omawianego modelu oceniono wykorzystując sposób walidacji krzyżowej w wariancie *Leave-One-Out* (LOO) (rys. 3, tabela 3) oraz testu prognozowania E_{coh} serii polimerowych próbek zewnętrznych (tabela 5, rys. 4). Określono również istotność poszczególnych parametrów występujących w wyprowadzonej 4-parametrowej zależności (tabela 4). Pozwala ona na wiarygodne prognozowanie wartości E_{coh} na podstawie wyłącznie wskaźników topologicznych a jej dodatkową zaletą użytkową jest możliwość uniknięcia skomplikowanych obliczeń.
Słowa kluczowe: energia kohezji, polimery, wskaźnik topologiczny, prognozowanie, QSPR, wieloliniowa analiza regresji.

COHESION ENERGY — SIGNIFICANCE IN POLYMERS' CHARACTERIZATION AND SOME METHODS OF ITS ESTIMATION

The cohesion energy (E_{coh}) of a material is an increase in the internal energy per mol of the material when all intermolecular forces are removed [1]. The cohesion energy density (e_{coh}), which is the energy required to break all intermolecular physical links in a unit volume of the material, is given by the following equation [2]:

$$e_{coh} = \frac{E_{coh}}{V} \qquad (1)$$

where: V — *molar volume of the polymer*.

E_{coh} is a very fundamental property of a polymer and plays a major role in the prediction of many other physical properties. The most important use of E_{coh} is in calculation Hildebrand solubility parameter (δ), which quantifies the interaction between polymer and solvent and is defined as follows:

$$\delta = \sqrt{e_{coh}} \qquad (2)$$

E_{coh} is also useful in correlation or prediction of many other important properties of a polymer, such as its glass transition temperature [3, 4], surface tension, dielectric constant, mechanical properties, permeability to small molecules [5, 6], energies of mixing of polymer blends [7, 8], and surface free energy [9].

For liquids of low molecular weight, the value of E_{coh} can be easily calculated from the molar heat of evaporation or from the vapor pressure as a function of the temperature. However, the experimental determination of E_{coh} for polymers is not as straightforward, since such materials do not evaporate. The experimental values of E_{coh}, which are usually deduced indirectly from dissolution and/or swelling measure-

ments at room temperature, vary in a wide range for many polymers [3].

Several semi-empirical group-contribution methods based on data for small molecules as well as dissolution and swelling data for polymers have been well-established to predict the E_{coh} values of polymers [1]. Group-contribution methods can sometimes give the prediction with reasonable accuracy, but a serious limitation is that these methods are only applicable for polymers containing chemical structural groups previously investigated. The E_{coh} values of polymers also can be calculated from molecular mechanics modeling of bulk polymers [10].

Another alternative approach to estimate E_{coh} is quantitative structure-property relationship (QSPR) on the basis of descriptors derived solely from the molecular structure to fit experimental data. The QSPR approach is based on the assumption that the variation of the behavior of the compounds, as expressed by any measured properties, can be correlated with changes in molecular features of the compounds termed descriptors [11]. The advantage of this approach lies in the fact that it requires only the knowledge of the chemical structure and is not dependent on any experimental properties. The QSPR has been successfully applied to the correlation of many diverse physicochemical properties of polymers and polymer solutions [12—17]. Bicerano [3] has developed two successful models for the prediction of E_{coh} of two data sets with 16 and 19 descriptors involved, respectively. Bortolotti *et al.* [9] correlated the E_{coh} values of polymers with their connectivity indices and constitutional descriptors, and got a 12-parameter correlation with square correlation coefficient (R^2) equal to 0.9948 and standard error of estimation (*s*) equal to 2834 J/mol. There are too many descriptors involved in these models though the predictions are of good accuracy. Yu *et al.* [18] have built a four-parameter model using quantum chemical descriptors obtained by density functional theory, with R^2 equal to 0.9722 and *s* equal to 2636 J/mol. However, no external data set polymers were used to validate this model.

Topological indices have been widely used in the correlation of physicochemical properties of organic compounds. In chemical graph theory, molecular structures are normally represented as hydrogen-depleted graphs, which vertices and edges act as atoms and covalent bonds, respectively. Chemical structural formulas can be then assimilated to undirected and finite multigraphs with labeled vertices, commonly known as molecular graphs. Topological indices are descriptors that characterize molecular graphs and contain a large amount of information about the molecule, including the numbers of hydrogen and non-hydrogen atoms bonded to each non-hydrogen atom, the details of the electronic structure of each atom, and the molecular structural features [19, 20].

QSPR models produced with only topological indices have advantages over the models produced with other descriptors due to the relative ease in calculating of descriptors, thus lowering computational cost and time [21, 22]. Moreover, because topological indices can be calculated solely from the molecular structure, the models based on them let predict the parameters required.

There have been numerous models developed with only topological indices [19, 20, 23—26]. Some of the recent studies concerning polymers are outlined here. García-Domenech and Julián-Ortiz [21] correlated the refractive indices and the glass transition temperatures of polymers with their topological indices: for the prediction of refractive indices, a 10-parameter correlation with R^2 equal to 0.962 was obtained; the glass transition temperatures were predicted through a model with R^2 = 0.894, consisting of 10 indices. Xu *et al.* [27] have developed a QSPR model with R^2 = 0.8874 to predict θ (lower critical solution temperature) in polymer solutions, with ten topological indices involved.

The goal of this work was to obtain a QSPR model on the basis of topological indices exclusively, which is expected to predict the E_{coh} values of polymers with better accuracy and relative ease in calculation.

MATERIALS AND METHOD

A total of 75 polymers with extensive structural diversity (Table 1 and 5) were selected as the data set, among which, 60 polymers were randomly chosen as the training set, and the other 15 polymers were used as the test set. The experimental E_{coh} data were taken from a book published by Bicerano [3]. The E_{coh} values ranged from 9883 to 73 620 J/mol. The polymers chosen in the data set were polyolefins, polyacrylates, polymethacrylates, polystyrenes, poly(vinyl ethers), and some others. It is impossible to calculate indices directly for the entire molecules because all polymers possess high values and wide distributions of molecular weights. So, the molecular indices calculated from their repeating unit structures end-capped with two hydrogen atoms could only be used in the QSPR studies for polymers [12, 28, 29]. The Dragon software [30] was applied to calculate a set of 199 topological indices for each polymer. Most of these indices are reviewed in the textbook by Todeschini and Consonni [31].

To develop QSPR models, stepwise multilinear regression analysis (MLRA) [32] was applied to the training set. Step-by-step variables are added to the equation, and a new regression is performed. If the new variable contributes significantly to the regression equation, the variable is retained; otherwise, the variable is excluded, hence preventing over-fitting. *F*-to-enter and *F*-to-remove were 4 and 3, respectively.

The goodness of the correlation was tested by the R^2, the adjusted R^2, the *F* ratio values, the standard error of estimate *s* and the significance level value *p*. The adjusted R^2 value was calculated using the following equation:

$$R_{adj}^2 = 1 - \left[\left(\frac{n-1}{n-m-1} \right) R^2 \right] \qquad (3)$$

where: n — *number of members of the data set*, m — *number of indices involved in the correlation.*

The adjusted R^2 is a better measure of the proportion of variance in the data explained by the correlation than R^2 (especially for the correlations developed with use of small data sets) because R^2 is somewhat sensitive to changes of n and m. In particular, in small samples, if m is large relative to n, there is a tendency for R^2 to be artificially high, *i.e.*, for the correlation to fit the data very well. In the extreme case if $n = (m + 1)$ the correlation will fit the data exactly, *i.e.*, $R^2 = 1$ [33]. The adjusted R^2 corrects the artificiality introduced when m approaches n through the use of a penalty function which scales the result.

After the generation of the correlation equation, a variance inflation factor (VIF) was calculated to test if multicollinearities existed among the indices in the model, which was defined as [34]:

$$VIF = \frac{1}{1 - R_j^2} \qquad (4)$$

where: R_j^2 — *coefficient of determination between the j-th coefficient regressed against all the other indices in the model.*

This is an arbitrary statistic in that the null hypothesis is not being accepted or rejected. Because of this, there is no distinct cutoff value for VIF. A common procedure is to set the value at 5.0, with any number higher indicating "serious" multicollinearity [35].

The predictive ability of the selected equation was measured through the percentage of mean relative error (MRE), defined as:

$$MRE = \frac{100}{n} \sum_i \left| \frac{E_{coh}^{exp} - E_{coh}^{pred}}{E_{coh}^{exp}} \right| \qquad (5)$$

The reliability of the final QSPR model was further validated internally using Leave-One-Out (LOO) cross-validation. Randomization experiments were also performed to prove the possible existence of fortuitous correlations. To do this, the dependent variable was randomly scrambled and used in the experiment. Models were then investigated with all members in the index pool to find the best models. The s and R values found using random dependent variables should be very poor if the original model did accurately represent the relationship between chemical structure and E_{coh}. Finally, the model was validated using the external test set.

RESULTS AND DISCUSSION

Stepwise MLRA was used to select the indices for the best model and the final correlation obtained for the training set contained four indices, with $R^2 = 0.9775$ and $R^2_{adj} = 0.9759$. The correlation equation for E_{coh} is the following:

$$E_{coh} = 46754.6 - 37057.9\ \chi0A + 7177.3\ \chi0sol + {} - 2129.8\ Jhetm - 6311.8\ CIC1$$

$$(6)$$

($N = 60$, $R = 0.9887$, $R^2 = 0.9775$, $R^2_{adj} = 0.9759$, $R^2_{CV} = 0.9747$, $s = 2455$, $F = 597.9$, $p < 0.000000$)

where: $\chi0A$ — *average connectivity index χ-0* [20]; $\chi0sol$ — *solvation connectivity index χ-0* [20]; *Jhetm* — *Balaban-type index from mass weighted distance matrix* [36]; *CIC1* — *complementary information content (neighborhood symmetry of 1-order)* [37].

The E_{coh} values predicted from Eq. (6) are shown in Table 1 and Fig. 1, while the distributions of relative errors (RE) for the prediction are given in Fig. 2. MRE for the prediction is 5.82 %. The min/max values of RE are 0.08 %/26.31 %. Among the 60 polymers, there are 41 samples having RE less than 5.82 % and only five present RE greater than 15 %. The absolute mean error between the experimental and predicted E_{coh} values is 1845 J/mol, which reveals the quality of the selected model for the prediction of E_{coh}. The statistical characteristics of the best four indices in Eq. (6) are shown in Table 2. The p values are quite small for each index, which shows that each is needed for accounting the variance and that the model is not overfit. The final ratio of training set poly-
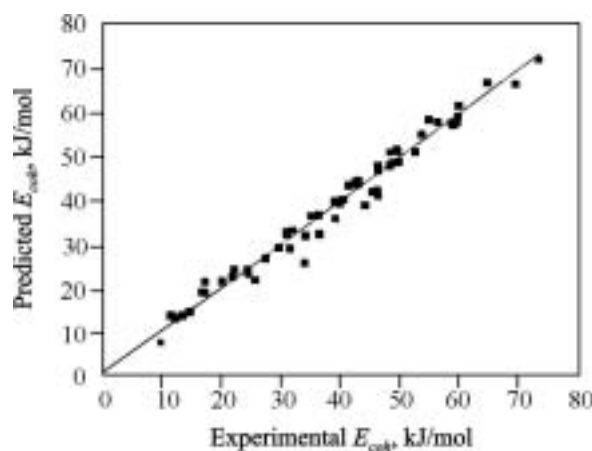

Fig. 1. Plot of the predicted vs. experimental E_{coh} values of polymers for the training set (see Table 1)
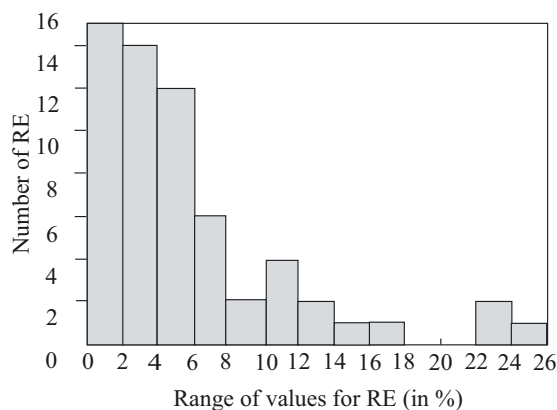

Fig. 2. Distributions of relative errors (RE) for the training set

T a b l e 1. **Prediction of cohesion energy (E_{coh}) of polymers (the training set)[*]**

No.	Polymer	E_{coh}, J/mol experimental	E_{coh}, J/mol predicted	RE, %	E_{coh}, J/mol (CV)[*]	RE (CV), %
1	Poly(1,2-butadiene)	17 299	19 456.1	12.47	19 580.8	13.20
2	Poly(1-butene)	19 092	18 621.0	2.47	18 575.4	2.71
3	Poly(2-ethylbutyl methacrylate)	59 130	57 431.8	2.88	57 240.8	3.20
4	Poly(3,4-dichlorostyrene)	60 062	61 839.5	2.96	62 030.1	3.28
5	Poly(4-methyl-1-pentene)	27 549	27 202.1	1.26	27 176.1	1.36
6	Poly(α-methyl styrene)	42 616	44 420.7	4.24	44 509.8	4.45
7	Poly(α-vinyl naphthalene)	58 796	57 968.7	1.41	57 905.0	1.52
8	Poly(benzyl methacrylate)	64 919	67 017.3	3.24	67 255.6	3.60
9	Poly(cyclohexyl methacrylate)	59 978	59 484.2	0.83	59 450.3	0.88
10	Poly(dimethyl siloxane)	16 870	19 525.2	15.74	18 914.1	12.12
11	Poly(ethyl α-chloroacrylate)	46 495	47 072.7	1.25	47 108.4	1.32
12	Poly(ethyl acrylate)	36 619	32 695.4	10.72	32 593.4	11.00
13	Poly(ethyl methacrylate)	40 039	39 706.3	0.84	39 695.9	0.86
14	Poly(isobutyl methacrylate)	48 496	48 220.5	0.57	48 201.5	0.61
15	Poly(isoprene)	21 863	23 068.1	5.52	23 130.3	5.80
16	Poly(methyl α-chloroacrylate)	41 554	43 715.7	5.21	43 907.1	5.67
17	Poly(methyl acrylate)	31 678	29 336.4	7.4	29 217.8	7.77
18	Poly(methyl ethacrylate)	40 637	40 598.7	0.10	40 597.7	0.10
19	Poly(methyl methacrylate)	35 097	36 726.5	4.65	36 772.2	4.78
20	Poly(n-butyl acrylate)	46 502	41 354.1	11.08	41 263.2	11.27
21	Poly(n-butyl methacrylate)	49 921	49 136.1	1.58	49 102.1	1.65
22	Poly(n-hexyl methacrylate)	59 804	57 918.0	3.16	57 755.0	3.43
23	Poly(n-octyl methacrylate)	69 686	66 711.0	4.27	66 211.3	4.99
24	Poly(N-vinyl carbazole)	73 620	72 369.3	1.70	72 347.2	1.73
25	Poly(N-vinyl pyrrolidone)	45 488	42 316.0	6.98	42 093.1	7.47
26	Poly(o-chloro styrene)	49 712	51 593.0	3.79	51 692.2	3.99
27	Poly(o-methyl styrene)	43 256	44 672.4	3.28	44 739.1	3.43
28	Poly(p-bromo styrene)	53 851	55 288.2	2.67	55 371.9	2.83
29	Poly(p-chloro styrene)	49 546	51 814.5	4.58	51 938.8	4.83
30	Poly(p-methoxy styrene)	48 405	51 183.4	5.74	51 298.6	5.98
31	Poly(p-methyl styrene)	43 090	44 853.4	4.10	44 936.8	4.29
32	Poly(p-t-butyl styrene)	55 060	58 716.7	6.65	58 986.8	7.14
33	Poly(p-vinyl pyridine)	46 371	42 646.4	8.04	42 294.5	8.80
34	Poly(sec-butyl methacrylate)	48 872	48 909.2	0.08	48 916.2	0.10
35	Poly(styrene sulfide)	52 747	51 366.0	2.62	51 293.9	2.76
36	Poly(t-butyl methacrylate)	46 427	48 154.3	3.73	48 351.4	4.15
37	Poly(vinyl acetate)	31 136	32 695.4	5.01	32 736.3	5.14
38	Poly(vinyl benzoate)	56 558	58 108.9	2.75	58 209.6	2.93
39	Poly(vinyl bromide)	24 536	24 584.9	0.20	24 588.5	0.22
40	Poly(vinyl chloride)	20 231	21 937.6	8.44	22 096.3	9.22
41	Poly(vinyl cyclohexane)	39 197	36 148.5	7.78	35 627.5	9.11
42	Poly(vinyl cyclopentane)	34 256	32 332.3	5.62	31 978.8	6.65
43	Poly(vinyl ethyl ether)	22 240	24 559.0	10.43	24 678.9	10.97
44	Poly(vinyl fluoride)	11 622	14 254.3	22.65	14 556.4	25.25
45	Poly(vinyl methyl ether)	17 299	21 849.8	26.31	22 070.9	27.59
46	Poly(vinyl n-butyl ether)	32 122	33 534.3	4.40	33 604.4	4.62
47	Poly(vinyl pivalate)	43 107	43 962.7	1.99	44 017.8	2.12
48	Poly(vinyl propionate)	36 453	36 956.3	1.39	36 969.1	1.42
49	Poly(vinyl sec-butyl ether)	31 073	33 184.9	6.80	33 272.1	7.08
50	Poly(vinyl trimethylsilane)	25 746	22 346.3	13.21	23 067.6	10.41
51	Poly(vinylidene chloride)	29 784	29 585.1	0.67	29 564.2	0.74
52	Poly(vinylidene fluoride)	12 566	13 512.9	7.54	13 594.8	8.19
53	Polyacrylamide	44 296	39 278.0	11.33	38 664.5	12.72
54	Polyacrylonitrile	34 157	26 194.8	23.32	25 559.9	25.17
55	Polychlorotrifluoroethylene	24 706	23 754.5	3.86	23 627.6	4.37
56	Polyethylene	9883	8104.9	18.00	8135.7	17.69
57	Polyisobutylene	16 870	17 230.8	2.14	17 271.6	2.39
58	Polypropylene	13 774	14 358.2	4.25	14 439.5	4.84
59	Polystyrene	39 197	40 156.8	2.45	40 250.8	2.69
60	Polytrifluoroethylene	14 837	15 053.0	1.46	15 097.9	1.76
MRE				5.82		6.10

[*] RE — relative error, CV — coefficient of variantion, MRE — mean relative error.

mers and number of indices is 1:15, what further supports this conclusion. The VIF values for each index are well below the general cutoff of 5.0, indicating that the indices are weakly correlated each to other and the QSPR model can be regarded as an optimal regression equation.

T a b l e 2. Indices involved in the best correlation derived for E_{coh} [Equation (6)]

Index	χ	$D\chi$	t-test	p-level	VIF
Constant	46 754.6	7068.2	6.615	0.000 000	
$\chi 0A$	-37 057.9	7848.8	-4.721	0.000 017	2.474
$\chi 0sol$	7177.3	258.8	27.737	0.000 000	3.307
$Jhetm$	-2129.8	413.7	-5.149	0.000 004	1.335
$CIC1$	-6311.8	648.4	-9.735	0.000 000	1.718

The LOO cross-validation was used to test the stability of the model obtained and the suitability of its application to unknown polymers. The results are shown in Table 1 and Fig. 3. The squared correlation coefficient and MRE for the cross-validation are 0.9747 and 6.10 %, respectively. Thus, the selected model is reliable in prediction of E_{coh} values of polymers. These results are acceptable, thus the introduction of additional descriptors, such as electronic or quantum mechanical, which afford more information but are sophisticated to calculate, is unnecessary.

The model was validated further by applying the randomization test. It consists of repetition of the calculation procedure several times after scrambling the dependent variables randomly. If all models obtained by the randomization test have relatively high values for both R^2 and R^2_{CV} statistics, this is due to a chance correlation and implies that the current modeling method cannot lead to an acceptable model using the available data set. Several randomizations of the dependent variables were performed and the results are shown in Table 3. The low R^2 and R^2_{CV} values indicate that the good results
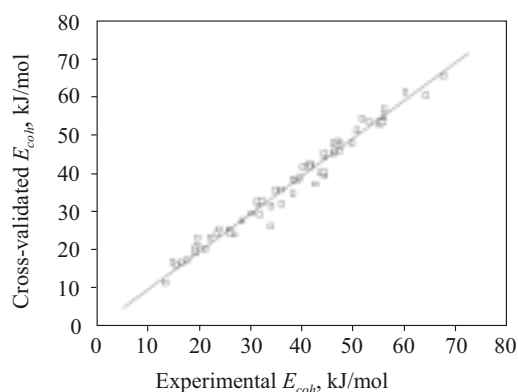
of the original model are not due to a chance correlation or structural dependency of the data set.

T a b l e 3. Results of randomization test

Iteration	R^2	R^2_{CV}	Iteration	R^2	R^2_{CV}
1	0.1529	0.0000	11	0.0776	0.0000
2	0.1038	0.0000	12	0.1296	0.0000
3	0.2338	0.0000	13	0.0912	0.0000
4	0.1958	0.0000	14	0.2877	0.0000
5	0.2922	0.0000	15	0.1480	0.0000
6	0.2836	0.0000	16	0.1731	0.0000
7	0.1587	0.0000	17	0.2869	0.0000
8	0.3501	0.0000	18	0.1536	0.0000
9	0.2064	0.0000	19	0.3995	0.0000
10	0.1931	0.0000	20	0.3251	0.0000

The statistical parameters obtained exclusively using topological indices herein are comparable in quality to those obtained by Yu *et al.* [18] using quantum chemical descriptors from more complicated density functional theory calculation (R^2 = 0.9775 *vs.* R^2 = 0.9722, R^2_{CV} = 0.9747 *vs.* 0.9643, F = 597.9 *vs.* F = 480.6, s = 2455 *vs.* s = 2636, MRE = 5.82 % *vs.* 5.93 %).

The results from Bicerano's model [3] are slightly better than those of the present model. However, this is not surprising because Bicerano's model consists of 16 topological and constitutional descriptors while the present model comprises only four descriptors. On the other hand, an improvement in the results by introducing more descriptors into the correlation equation should be carefully considered because overfitting and chance correlations may in part be due to such an approach [28].

The topological characteristics of the present model indicate that dispersion interactions and the extent of branching of the molecules affect the cohesion energy of polymers. To test the dependency of E_{coh} on each index, R^2, R^2_{adj}, R^2_{CV} and s were calculated with one-index linear regression for the entire set of 60 polymers (Table 4).

T a b l e 4. Summary of one-parameter models for 60 training data set

Index	E_{coh}, prediction equation	R^2	R^2_{adj}	R^2_{CV}	s
$\chi 0sol$	E_{coh} = -1747.8 + 6745.3 $\chi 0sol$	0.9185	0.9171	0.9133	4551
$\chi 0A$	E_{coh} = 19 647 409 − 194 835 $\chi 0A$	0.6232	0.6167	0.5979	9787
$CIC1$	E_{coh} = 14 612.5 + 10 542.3 $CIC1$	0.1856	0.1716	0.1348	14 389
$Jhetm$	E_{coh} = 35 255.0 + 1068.8 $Jhetm$	0.0036	-0.0135	-0.0930	15 916

Fig. 3. Plot of the cross-validated (LOO) vs. experimental Ecoh values of polymers for the training set

The first important index is the 0-order solvation connectivity index $\chi 0sol$ [20] which gave the best overall fit

with the R^2 value equal to 0.9185 for the data set. The $\chi 0sol$ is calculated by Eq. (7):

$$\chi 0sol = \frac{1}{2} \sum_{i=1}^{n} \frac{L_i}{\delta_i^{1/2}} \qquad (7)$$

where: L_i — *principal quantum number of the i-th atom, δ_i — corresponding vertex degree, n — total number of atoms in the molecule.*

This index can be considered as entropy of solvation [38] and somehow indicates the dispersion interactions occurring in the polymers. The large contribution of this index in the E_{coh} values of polymers is in agreement with the contribution that one would expect for the dispersion interactions of polymers, since E_{coh} can be formally divided into the contributions of the dispersion, polar and hydrogen bonding components.

The second important index is $\chi 0A$ [20] which correlates with E_{coh} of $R^2 = 0.6232$. $\chi 0A$ is defined by Eq. (8),

$$\chi 0A = \frac{\sum_{i=1}^{n} \frac{1}{\delta_i^{1/2}}}{B} \qquad (8)$$

where: *B — number of bonds in the molecule.*

The 1-order complementary information content *CIC*1 [37] is defined by Eq. (9),

$$CIC1 = \log_2 n - IC1 \qquad (9)$$

$$IC1 = \sum_{i=1}^{1} \frac{n_i}{n} \log_2 \frac{n_i}{n} \qquad (10)$$

where: n_i — *number of atoms in the i-th class, IC1 — 1-order information content itself defined by Eq. (10).*

The *CIC*1 index describes the atomic connectivity in the molecule, while $\chi 0A$ and *CIC*1 encode the size and the atomic constitution of the polymer. These parameters directly affect the intermolecular interaction [39]. *Jhetm* shows only a small contribution to the cohesion energy of polymers.

The one-parameter model with $\chi 0sol$ as the sole index has an R^2 value equal to 0.9185 (*vs.* 0.9775), which means that introduction of the three additional indices does not
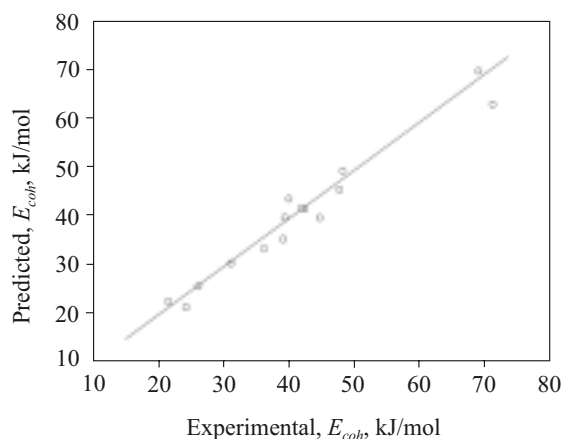

*Fig. 4. Plot of the predicted vs. experimental E*coh *values of polymers for the test set*

increase R^2 too much. However, the *s* value for this one-parameter model is 4551 J/mol, *i.e.* 85.4 % greater than that obtained with the four-parameter model (2455 J/mol). Thus the three additional indices should be introduced.

T a b l e 5. Prediction of the cohesive energy E_{coh} of polymers (the test set)

No.	Polymer	E_{coh} (expt.) J/mol	E_{coh} (pred.) J/mol	RE %
61	Poly(1,4-butadiene)	17 972	19 456.2	8.26
62	Poly(acrylic acid)	35 329	32 420.4	8.23
63	Poly(chloro-*p*-xylylene)	49 546	51 462.4	3.87
64	Poly(ethylene sulfide)	23 433	23 153.2	1.19
65	Poly(ethylene terephthalate)	74 087	75 908.3	2.46
66	Poly(maleic anhydride)	39 831	44 607.6	11.99
67	Poly(methacrylic acid)	38 748	34 732.0	10.36
68	Poly(*N*-phenyl maleimide)	76 653	67 757.8	11.60
69	Poly(oxy-1,1-dichloro-methyltrimethylene)	48 954	46 894.3	4.21
70	Poly(*p*-phenylene)	29 314	28 794.5	1.77
71	Poly(*p*-xylene)	39 031	40 032.7	2.57
72	Poly(thiocarbonyl fluoride)	21 175	18 050.7	14.75
73	Poly(α,α,α′,α′-tetrafluoro--*p*-xylene)	45 353	39 966.7	11.88
74	Poly(ε-caprolactone)	42 067	42 404.9	0.80
75	Poly[thio(*p*-phenylene)]	42 698	42 098.4	1.40
MRE				6.36

The results of prediction for the test set from Eq. (6) for 14 polymers are shown in Table 5 and Fig. 4. MRE is 6.36 %, which confirms the validity of the proposed model.

CONCLUSIONS

A general QSPR model with good statistical parameters ($R^2 = 0.9775$ and MRE = 5.82 %) has been obtained to predict the cohesion energy values of polymers with four topological indices. This model is very useful because topological indices can be calculated easily as long as the molecular structure of the compound concerned is known. The topological characteristics of the present model also show the importance of the dispersion interactions and the extent of branching of the molecules to the cohesion energy of polymers.

REFERENCES

1. van Krevelen D. W.: "Properties of polymers: correlation with chemical structure", 2nd Edition, Elsevier, Amsterdam 1976.

2. Niemela S., Leppanen J., Sundholm F.: *Polymer* 1996, **37**, 4155.

3. Bicerano J.: "Prediction of polymer properties", 2nd Edition, Marcel Dekker Inc., New York 1996.

4. Liu H., Uhlherr A., Bannister M. K.: *Polymer* 2004, **45**, 2051.

5. Hirayama Y., Yoshinaga T., Kusuki Y., Ninomiya K., Sakakibara T., Tamari T.: *J. Memb. Sci.* 1996, **111**, 183.

6. Tokarski J. S., Hopfinger A. J., Hobbs J. D., Ford D. M., Faulon J.-L. M.: *Comput. Theor. Polym. Sci.* 1997, **7**, 199.

7. Lee S., Lee J. G., Lee H., Mumby S. J.: *Polymer* 1999, **40**, 5137.

8. Moolman F. S., Meunier M., Labuschagne P. W., Truter P.-A.: *Polymer* 2005, **46**, 6192.

9. Bortolotti M., Brugnara M., Volpe C. D., Maniglio D., Siboni S.: *J. Coll. Interf. Sci.* 2006, **296**, 292.

10. Eichinger B. E., Rigby D., Stein J.: *Polymer* 2002, **43**, 599.

11. Yao X. J., Wang Y. W., Zhang X. Y., Zhang R. S., Liu M. C., Hu Z. D., Fan B. T.: *Chemom. Intell. Lab. Syst.* 2002, **62**, 217.

12. Xu J., Chen B., Zhang Q., Guo B.: *Polymer* 2004, **45**, 8651.

13. Afantitis A., Melagraki G., Sarimveis H., Koutentis P. A., Markopoulos J., Igglessi-Markopoulou O.: *Polymer* 2006, **47**, 3240.

14. Melagraki G., Afantitis A., Sarimveis H., Koutentis P. A., Markopoulos J., Igglessi-Markopoulou O.: *J. Mol. Model.* 2007, **13**, 55.

15. Yu X., Wang X., Li X., Gao J., Wang H.: *Macromol. Theory Simul.* 2006, **15**, 94.

16. Yu X., Wang X., Wang H., Li X., Gao J.: *QSAR Comb. Sci.* 2006, **25**, 156.

17. Gao J., Wang X., Yu X., Li X., Wang H.: *J. Mol. Model.* 2006, **12**, 521.

18. Yu X., Wang X., Li X., Gao J., Wang H.: *J. Polym. Sci. Part B: Polym. Phys.* 2006, **44**, 409.

19. Kier L. B., Hall L. H.: "Molecular Connectivity in Chemistry and Drug Research", Academic Press, New York 1976.

20. Kier L. B., Hall L. H.: "Molecular Connectivity in Structure Activity Analysis", Wiley, New York 1986.

21. Garcia-Domenech R., Julian-Ortiz J. V. D.: *J. Phys. Chem. B* 2002, **106**, 1501.

22. Mattioni B. E., Jurs P. C.: *J. Chem. Inf. Comput. Sci.* 2002, **42**, 232.

23. Devillers J., Balaban A. T.: "Topological Indices and Related Descriptors in QSAR and QSPR", Gordon and Breach, The Netherlands 1999.

24. Pogliani L.: *Chem. Rev.* 2000, **100**, 3827.

25. Zhong C., Yang C., Li Q.: *Ind. Eng. Chem. Res.* 2002, **41**, 2826.

26. Xu J., Chen B.: *J. Mol. Model.* 2005, **12**, 24.

27. Xu J., Liu L., Xu W., Zhao S., Zuo D.: *J. Mol. Graph. Model.* 2007, **26**, 352.

28. Katritzky A., Sild S., Karelson M.: *J. Chem. Inf. Comput. Sci.* 1998, **38**, 1171.

29. Cao C., Lin Y.: *J. Chem. Inf. Comput. Sci.* 2003, **43**, 643.

30. Todeschini R., Consonni V., Mauri A., Pavan M., DRAGON, Version 5.1, TALETE srl, Milan 2004.

31. Todeschini R., Consonni V.: "Handbook of Molecular Descriptors", Wiley-VCH, Weinheim 2000.

32. Jurs P. C.: "Computer Software Applications in Chemistry", Wiley, New York 1996.

33. Jobson J. D.: "Applied Multivariate Data Analysis", Vol. 1: "Regression and Experimental Design", Springer-Verlag, New York 1991.

34. Peterangelo S. C., Seybold P. G.: *Int. J. Quantum Chem.* 2004, **96**, 1.

35. Holder A. J., Kilway K. V., Code J. E., Giese G. J., Travis D. M., Fleckenstein J. E., Marzluf K. R., Clevenger R. C., Vastlik H. L., Eick J. D., Chappelow C. C.: *Macromol. Theory Simul.* 2005, **14**, 117.

36. Barysz M., Jashari G., Lahh R. S., Srivastava A. K., Trinajstic N.: in "Chemical Applications of Topology and Graph Theory" (Ed. King R. B.), Elsevier, Amsterdam 1983, p. 222.

37. Magnuson V. R., Harriss D. K., Basak S. C.: in "Studies in Physical and Theoretical Chemistry" (Ed. King R. B.), Elsevier, Amsterdam 1983, p. 178.

38. Jalali-Heravi M., Kyani A.: *J. Chem. Inf. Comput. Sci.* 2004, **44**, 1328.

39. Tamm K., Burk P.: *J. Mol. Model.* 2006, **12**, 417.