

A general QSPR model for the prediction of θ (lower critical solution temperature) in polymer solutions with topological indices

Jie Xu ^{a,*}, Li Liu ^b, Weilin Xu ^a, Sanpin Zhao ^a, Danying Zuo ^a

^a Hubei New Textile Material & Application Key Laboratory, Wuhan University of Science & Engineering, Wuhan 430073, PR China

^b College of Electronics & Information Engineering, Wuhan University of Science & Engineering, Wuhan 430073, PR China

Received 29 August 2006; received in revised form 2 January 2007; accepted 12 January 2007

Available online 18 January 2007

Abstract

The quantitative structure–property relationship (QSPR) study was performed between topological indices representing the molecular structures and θ (LCST, lower critical solution temperature) with a database of 169 data points including 12 polymers and 67 solvents. A linear 10-parameter model was developed to predict θ (LCST) with squared correlation coefficient (R^2) of 0.8874 and standard error of estimation (SEE) of 24.57 for the training set of 112 systems. The mean relative error (MRE) for the calculation of θ (LCST) was 3.49%. The stability of the proposed model was validated using Leave-One-Out cross-validation and randomization experiments. The model is generally predictive and requires only topological indices during the predictions with the advantage of the relative ease in calculating descriptors which makes it easy to apply. A satisfactory MRE of 4.95% for a test set of 57 systems indicates that the model would be very useful to obtain reliable estimates of θ (LCST) in polymer solutions before they are actually synthesized.

© 2007 Elsevier Inc. All rights reserved.

Keywords: θ (Lower critical solution temperature); QSPR; Polymer solutions; Topological index; Multilinear regression analysis

1. Introduction

The phase behavior of polymers is an important property since most polymeric materials are processed in solutions [1]. Partially miscible polymer solutions often exhibit two solubility boundaries [2], the upper critical solution temperature (UCST) and the lower critical solution temperature (LCST). Both the UCST and LCST depend on the molar mass and pressure; however, the θ temperatures, which are the critical solution temperatures at infinite chain length, are not affected by polymer molar mass. θ (LCST), the LCST at infinite chain length, is an important property of polymer solutions because it can serve as an upper temperature limit for polymer processing. Experimental θ (LCST) data have been available for a lot of systems; however, a relatively accurate prediction method is still required when experimental data are lacking.

There have been numerous attempts to correlate and estimate LCSTs, including θ (LCST), by various approaches.

According to the view of Melagraki et al. [3], these methods can be roughly divided into three groups. The first group of models is those methods that have a solid theoretical background but that require vapor–liquid or liquid–liquid experimental data to adjust the unknown parameters, resulting in limited predictive ability [4–8]. The second group of models is empirical correlations relating θ (LCST) to other physiochemical properties, such as density, critical properties, or solubility parameters [2,9–11]. These correlations are simple and have reasonable accuracy, but they cannot be applied to systems for which the required physicochemical properties are not available. An alternative approach to estimate θ (LCST) is quantitative structure–property relationship (QSPR) based on 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 [12]. 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

* Corresponding author. Tel.: +86 27 87426559; fax: +86 27 87426559.

E-mail address: xujie0@ustc.edu (J. Xu).

diverse physicochemical properties of polymers and polymer solutions [1,3,13–16]. Liu and Zhong [17] have developed a QSPR model to predict the θ (LCST) in polymer solutions by using the connectivity indices of the polymer and solvent concerned, with R^2 of 0.77 and mean relative error (MRE) of 5.44%. Recently, Melagraki et al. [3] obtained a nine-parameter QSPR model for θ (LCST) in polymer solutions by multilinear regression, with R^2 of 0.8860 and MRE of 3.73%.

There exist more than a thousand of theoretical descriptors available in the literature to represent the molecular structure, and one usually faces the problem of selecting those which are the most representatives for the property under consideration. Topological indices, the commonly used molecular descriptors, 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, whose 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, also known as graph theoretical 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.

QSPR models produced with only topological indices have advantages over models produced with other descriptors due to the relative ease in calculating descriptors, thus lowering computational cost and time [18,19]. Moreover, because topological indices can be calculated solely from the molecular structure, the models based on them are predictive. There have been numerous models developed with only topological indices [20–24]. Some of the recent studies are outlined here. Garcia-Domenech and Julian-Ortiz [18] 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 of 0.962 was obtained; the glass transition temperatures were predicted through a model with R^2 of 0.894, consisting of 10 indices. Xu and Chen [24] have developed a QSPR model with R^2 of 0.9304 to predict the glass transition temperatures of organic light-emitting diode materials, with five topological indices involved.

The goal of the present work is to obtain an exclusively topological index-based QSPR model, which is expected to predict the θ (LCST) values in polymer solutions with better accuracy and relative ease in calculating. Also, various model validation methods are used to test the predictive potential of the model produced through stepwise multilinear regression analysis (MLRA).

2. Materials and methods

A total set of 169 experimental θ (LCST) data, including 12 polymers and 67 solvents (as shown in Tables 1 and 2), was taken from the article by Liu and Zhong [17]. Among them, 112 systems were chosen as the training set and the other 57 systems

were used as the test set, which was the same with Ref. [17]. The polymers studied in this work are polyethylene (PE), polypropylene (PP), polybutene-1 (PB1), polyisobutene (PIB), polypentene-1 (PP1), poly(4-methylpentene-1) (P4MP1), poly(*cis*-1,4-butadiene) (PBD), polystyrene (PS), poly(α -methylstyrene) (PMS), poly(*p*-chlorostyrene) (PPCS), poly(dimethyl siloxane) (PDMS), and poly(isotactic methyl methacrylate) (PMMA).

It is impossible to calculate indices directly for the entire molecules because all polymers possess high molecular weights and have wide distribution of molecular weights. And the molecular indices calculated from their repeating unit structures end-capped with two hydrogen atoms could be used in the QSPR studies for polymers [13,25,26]. The Dragon software [27] was used to calculate a set of 199 topological indices for each polymer and solvent. Most of these indices are reviewed in the recent textbook by Todeschini and Consonni [28].

Objective feature selection was done to remove those indices that provide minimal or redundant information. Constant values and descriptors found to be correlated pair-wise were excluded in a pre-reduction step (one of any two indices with a correlation greater than 0.96 was removed to reduce redundant information), thus 147 indices for each system underwent subsequent variable selection for the modeling.

To develop QSPR models, stepwise MLRA [29] 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 (SEE) and *p*-values all corresponding to a 95% confidence level. The adjusted R^2 value was calculated using the following formula:

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

where *n* is the number of members of the data set and *m* is the 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 correlations developed using small data sets) because R^2 is somewhat sensitive to changes in *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., $R^2 = 1$ [30]. The adjusted R^2 corrects for the artificiality introduced when *m* approaches *n* through the use of a penalty function which scales the result.

The predictive ability of the selected equations was measured through the percentage of MRE, defined as:

$$\text{MRE} = \frac{100}{n} \sum_i \left| \frac{\theta_i^{\text{exp}} - \theta_i^{\text{calc}}}{\theta_i^{\text{exp}}} \right| \quad (2)$$

Table 1

Experimental, calculated and cross-validated values of θ (LCST) for the training set of 112 polymer solution systems

Polymer	Solvent	Exp. θ (LCST) (K)	Calc. θ (LCST) (K)	Relative error (%)	CV θ (LCST) (K)	Relative error (CV, %)
PE	<i>n</i> -Hexane	411	439.07	6.83	441.65	7.46
PE	<i>n</i> -Heptane	459	476.39	3.79	477.85	4.11
PE	<i>n</i> -Nonane	531	533.26	0.43	533.40	0.46
PE	<i>n</i> -Decane	557.55	556.52	0.19	556.39	0.21
PE	<i>n</i> -Dodecane	605.55	595.80	1.62	594.65	1.81
PE	<i>n</i> -Tridecane	631.55	613.22	2.91	610.55	3.33
PE	2,2-Dimethyl pentane	399	433.08	8.55	436.84	9.49
PE	2,2,3-Trimethyl butane	444	460.31	3.68	462.91	4.26
PE	2,3-Dimethyl pentane	463	471.77	1.90	472.48	2.05
PE	3-Ethyl pentane	471	489.33	3.90	490.68	4.18
PE	2,2,4,4-Tetramethyl pentane	513	502.43	2.07	498.28	2.88
PE	2,3,4-Trimethyl hexane	545	539.98	0.93	539.40	1.03
PE	Methyl cyclopentane	488	511.19	4.76	514.41	5.42
PE	Methyl cyclohexane	537	538.50	0.28	538.64	0.31
PE	1-Octanol	621	513.83	17.26	497.15	19.95
PE	<i>n</i> -Pentyl acetate	528	530.69	0.51	531.21	0.61
PP	<i>n</i> -Pentane	422	431.67	2.30	432.37	2.46
PP	<i>n</i> -Hexane	470	478.62	1.84	479.08	1.94
PP	<i>n</i> -Octane	542	546.57	0.85	546.77	0.88
PP	<i>n</i> -Nonane	571	572.81	0.32	572.91	0.34
PP	2,2-Dimethyl butane	441	435.61	1.23	435.19	1.32
PP	2,3-Dimethyl butane	465	467.48	0.54	467.68	0.58
PP	2,2-Dimethyl pentane	489	472.63	3.35	471.22	3.64
PP	2,2,3-Trimethyl butane	511	499.86	2.19	498.47	2.46
PP	2,3-Dimethyl pentane	513	511.32	0.33	511.19	0.36
PP	3-Ethyl pentane	520	528.88	1.71	529.42	1.82
PP	2,2,4,4-Tetramethyl pentane	548	541.98	1.10	539.63	1.53
PP	2,3,4-Trimethyl hexane	585	579.53	0.94	578.82	1.06
PP	Methyl cyclopentane	518	550.74	6.33	555.22	7.19
PP	Methyl cyclohexane	564	578.05	2.50	579.77	2.80
PP	Diethyl ether	420	407.86	2.90	405.75	3.40
PB1	<i>n</i> -Pentane	421	428.06	1.68	428.37	1.75
PB1	<i>n</i> -Heptane	509	512.33	0.66	512.41	0.67
PB1	<i>n</i> -Octane	540	542.96	0.55	543.04	0.57
PB1	2-Methyl butane	416	399.34	4.01	398.45	4.23
PB1	2,2-Dimethyl butane	444	432.00	2.71	431.37	2.85
PB1	2,2,3-Trimethyl butane	507	496.25	2.13	495.18	2.34
PB1	2,5-Dimethyl hexane	519	482.41	7.06	474.01	8.67
PB1	3-Ethyl pentane	523	525.27	0.44	525.32	0.45
PB1	3,4-Dimethyl hexane	559	546.39	2.26	545.48	2.42
PIB	Cyclopentane	461	458.71	0.50	458.47	0.55
PIB	Cyclohexane	516	515.38	0.13	515.32	0.14
PIB	Cycloheptane	572	567.54	0.78	567.06	0.87
PIB	Cyclooctane	637	610.39	4.18	606.52	4.79
PIB	<i>n</i> -Heptane	442	457.88	3.60	458.27	3.68
PIB	<i>n</i> -Octane	477	488.50	2.42	488.81	2.48
PIB	<i>n</i> -Dodecane	582	577.29	0.81	576.80	0.90
PIB	2-Methyl butane	318	344.88	8.46	346.65	9.01
PIB	2-Methyl pentane	376	390.91	3.97	391.67	4.17
PIB	3-Methyl pentane	405	412.67	1.90	412.97	1.97
PIB	3-Methyl hexane	446	441.66	0.98	441.48	1.02
PIB	3-Ethyl pentane	458	470.82	2.80	471.37	2.92
PIB	2,3-Dimethyl pentane	451	453.25	0.50	453.33	0.52
PIB	2,4-Dimethyl pentane	403	420.72	4.40	420.02	4.23
PIB	2,2,3-Trimethyl butane	445	441.80	0.72	441.39	0.82
PIB	Ethyl cyclopentane	524	533.18	1.76	534.13	1.94
PIB	3-Methyl heptane	478	452.57	5.33	450.82	5.69
PIB	2,2-Dimethyl hexane	454	428.22	5.68	426.70	6.02
PIB	2,5-Dimethyl hexane	446	427.96	4.05	426.76	4.32
PIB	3,4-Dimethyl hexane	497	491.94	1.02	491.51	1.11
PIB	<i>n</i> -Propyl cyclopentane	547	550.35	0.62	550.74	0.69
PP1	<i>n</i> -Pentane	433	445.90	2.98	446.56	3.14
PP1	<i>n</i> -Hexane	482	492.86	2.26	493.29	2.35
PP1	<i>n</i> -Octane	556	560.80	0.87	560.98	0.90

Table 1 (Continued)

Polymer	Solvent	Exp. θ (LCST) (K)	Calc. θ (LCST) (K)	Relative error (%)	CV θ (LCST) (K)	Relative error (CV, %)
PP1	2-Methyl butane	422	417.18	1.15	416.92	1.21
PP1	2,4-Dimethyl pentane	493	493.02	0.01	493.02	0.01
PP1	2,2-Dimethyl pentane	502	486.87	3.02	485.84	3.22
PP1	2,3-Dimethyl pentane	529	525.55	0.66	525.34	0.70
PP1	3-Methyl heptane	537	530.08	1.29	529.74	1.36
P4MP1	<i>n</i> -Butane	388	371.20	4.33	369.16	4.86
P4MP1	<i>n</i> -Pentane	441	446.32	1.21	446.65	1.28
P4MP1	<i>n</i> -Heptane	522	530.60	1.65	531.07	1.74
P4MP1	<i>n</i> -Octane	553	561.22	1.49	561.72	1.58
P4MP1	2-Methyl butane	431	417.60	3.11	416.73	3.32
P4MP1	2,2-Dimethyl butane	462	450.27	2.54	449.48	2.72
P4MP1	2,2-Dimethyl pentane	499	487.29	2.35	486.25	2.56
P4MP1	2,2,3-Trimethyl butane	521	514.52	1.25	513.71	1.40
P4MP1	Cyclopentane	505	531.44	5.24	534.71	5.89
PBD	<i>n</i>-Hexane	373	423.82	13.63	425.17	13.99
PBD	2,2,3-Trimethyl butane	414	445.06	7.51	448.59	8.36
PBD	<i>n</i>-Octane	390	491.77	26.10	494.41	26.78
PBD	Ethyl propyl ketone	510	509.13	0.18	509.34	0.14
PBD	Diethyl ketone	481	471.40	2.00	470.93	2.10
PS	Cyclopentane	427	404.53	5.27	400.79	6.14
PS	Cyclohexane	486	461.19	5.11	457.87	5.79
PS	Methyl cyclopentane	417	438.49	5.16	441.55	5.89
PS	Methyl cyclohexane	480	465.80	2.96	463.98	3.34
PS	Benzene	523	550.27	5.22	554.74	6.07
PS	Toluene	550	593.63	7.94	596.12	8.39
PS	Ethyl acetate	412	392.04	4.85	390.12	5.32
PS	<i>n</i> -Propyl acetate	451	428.43	5.01	426.25	5.49
PS	<i>i</i> -Propyl acetate	380	378.09	0.51	378.18	0.49
PS	Ethyl <i>n</i> -butyrate	471	456.63	3.06	454.91	3.42
PS	Isobutyl acetate	445	400.86	9.92	401.68	9.74
PS	<i>sec</i> -Butyl acetate	442	425.88	3.65	423.89	4.10
PMS	Cyclopentane	417.6	394.14	5.62	390.05	6.60
PMS	Cyclohexane	456	450.80	1.15	450.07	1.31
PMS	Butyl chloride	412	386.15	6.28	384.55	6.67
PMS	<i>n</i> -Butyl acetate	446.9	437.70	2.06	436.74	2.28
PMS	<i>n</i> -Hexyl acetate	500.9	452.00	9.77	446.51	10.86
PPCS	<i>Isobutyl acetate</i>	348.5	402.60	15.53	403.18	15.69
PPCS	<i>Ethyl carbitol</i>	300.8	371.38	23.47	378.49	25.83
PPCS	<i>n</i> -Butyl carbitol	323.1	336.74	4.23	336.28	4.08
PDMS	<i>n</i> -Pentane	453	439.04	3.09	438.05	3.31
PDMS	<i>n</i> -Heptane	528	523.31	0.89	523.05	0.94
PDMS	<i>n</i> -Octane	553	553.94	0.17	553.97	0.18
PDMS	<i>n</i> -Dodecane	643	642.73	0.05	642.68	0.06
PDMS	<i>n</i> -Cetane	708	705.85	0.31	705.08	0.42
PMMA	Methyl acetate	441	426.93	3.20	424.38	3.78
PMMA	Ethyl acetate	478	463.51	3.04	460.81	3.60
PMMA	Ethyl propyl ketone	511	523.16	2.38	525.31	2.80
PMMA	Diethyl ketone	497	485.42	2.33	483.77	2.67
MRE				3.49		3.80

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 most predictive models. The SEEs and correlation coefficients found using random dependent variables should be very poor if the original model did accurately represent the relationship between chemical structure and θ (LCST).

3. Results and discussion

The number of indices in the final QSPR model was determined on the basis of the training set size (112 systems) and on the basis of the obtained correlation coefficient, F ratio values, adjusted correlation coefficient, and p -values for the selected indices. The results during stepwise MLRA were shown in Table 3, including the selected indices, R , and SEE values of each model. The best seven-parameter equation exhibited the R and SEE comparable to those of models containing more indices; however, not all indices or constants

Table 2
Experimental and predicted values of θ (LCST) for the test set of 57 polymer solution systems

Polymer	Solvent	Exp. θ (LCST) (K)	Calc. θ (LCST) (K)	Relative error (%)
PE	<i>n</i> -Pentane	353	392.12	11.09
PE	<i>n</i> -Octane	496	507.02	2.23
PE	<i>n</i> -Undecane	581.75	552.79	4.98
PE	2,4-Dimethyl pentane	395	439.24	11.20
PE	2,2,4-Trimethyl pentane	495	448.23	9.45
PE	3,4-Dimethyl hexane	515	510.45	0.89
PE	Cyclopentane	472	477.23	1.11
PE	Cyclohexane	518	533.89	3.07
PE	<i>n</i> -Butyl acetate	490	520.79	6.29
PP	<i>n</i> -Heptane	511	515.94	0.97
PP	2-Methyl butane	413	402.95	2.44
PP	2,4-Dimethyl pentane	481	478.79	0.46
PP	2,2,4-Trimethyl pentane	510	487.78	4.36
PP	3,4-Dimethyl hexane	553	550.00	0.55
PP	Cyclopentane	495	516.78	4.40
PP	Cyclohexane	540	573.44	6.20
PB1	<i>n</i> -Hexane	464	475.01	2.38
PB1	<i>n</i> -Nonane	564	569.20	0.93
PB1	2,4-Dimethyl pentane	480	475.18	1.01
PB1	2,3-Dimethyl pentane	517	507.71	1.80
PB1	Cyclopentane	498	513.17	3.05
PIB	<i>n</i> -Pentane	344	373.60	8.61
PIB	<i>n</i> -Hexane	402	420.56	4.62
PIB	Methyl cyclopentane	478	492.67	3.07
PIB	2-Methyl hexane	426	414.55	2.69
PIB	2,2-Dimethyl pentane	404	414.57	2.62
PIB	3,3-Dimethyl pentane	451	453.57	0.57
PIB	Methyl cyclohexane	526	519.98	1.15
PIB	2-Methyl heptane	466	434.69	6.72
PIB	2,4-Dimethyl hexane	458	450.04	1.74
PIB	2,2,4-Trimethyl pentane	435	429.72	1.22
PIB	<i>n</i> -Decane	535	538.01	0.57
PP1	<i>n</i> -Heptane	522	530.18	1.57
PP1	2,2-Dimethyl butane	457	449.85	1.57
PP1	2,2,4-Trimethyl pentane	527	502.02	4.75
PP1	Cyclopentane	502	531.02	5.79
P4MP1	<i>n</i> -Hexane	487	493.28	1.29
P4MP1	<i>n</i> -Nonane	579	587.47	1.47
P4MP1	2,4-Dimethyl pentane	499	493.44	1.12
P4MP1	3-Ethyl pentane	532	543.54	2.17
PBD	2-Methyl hexane	370	417.82	12.93
PBD	2,2,4-Trimethyl pentane	388	432.98	11.60
PBD	Propylene oxide	414	359.47	13.18
PS	<i>tert</i> -Butyl acetate	329	394.02	19.77
PS	Diethyl malonate	578	551.61	4.57
PS	<i>i</i> -Amyl acetate	493	422.21	14.36
PS	Methyl acetate	384.5	355.45	7.56
PS	Methyl ethyl ketone	420	353.20	15.91
PMS	Methyl cyclohexane	431	455.41	5.67
PMS	<i>n</i> -Pentyl acetate	475.8	447.60	5.93
PPCS	<i>tert</i> -Butyl acetate	338.4	395.77	16.96
PDMS	<i>n</i> -Hexane	493	486.00	1.42
PDMS	<i>n</i> -Decane	603	603.44	0.08
PMMA	Butyl chloride	454	468.02	3.09
PMMA	Dipropyl ketone	522	544.97	4.41
PMMA	Methyl ethyl ketone	464	424.67	8.48
PMMA	Tetrahydrofuran	519.5	497.64	4.21
MRE				4.95

Table 3
Results of stepwise MLRA for best correlation model

Best index subset	<i>R</i>	SEE
X3v (Sol)	0.621	54.99
X3v (Sol), X4v (Sol)	0.674	52.09
X3v (Sol), X4v (Sol), TI1 (Poly)	0.717	49.39
X3v (Sol), X4v (Sol), TI1 (Poly), IVDM (Sol)	0.746	47.40
X3v (Sol), X4v (Sol), TI1 (Poly), IVDM (Sol), TIC2 (Sol)	0.811	41.80
X3v (Sol), X4v (Sol), TI1 (Poly), IVDM (Sol), TIC2 (Sol), Jhete (Poly)	0.867	35.74
X3v (Sol), TI1 (Poly), IVDM (Sol), TIC2 (Sol), Jhete (Poly)	0.865	35.82
X3v (Sol), TI1 (Poly), IVDM (Sol), TIC2 (Sol), Jhete (Poly), IC2 (Sol)	0.894	32.22
X3v (Sol), TI1 (Poly), IVDM (Sol), TIC2 (Sol), Jhete (Poly), IC2 (Sol), BIC2 (Poly)	0.924	27.56
X3v (Sol), X4v (Sol), TI1 (Poly), IVDM (Sol), TIC2 (Sol), Jhete (Poly), IC2 (Sol), BIC2 (Poly)	0.929	26.75
X3v (Sol), X4v (Sol), TI1 (Poly), IVDM (Sol), TIC2 (Sol), Jhete (Poly), IC2 (Sol), BIC2 (Poly), Jhetv (Sol)	0.935	25.93
X3v (Sol), X4v (Sol), TI1 (Poly), IVDM (Sol), TIC2 (Sol), Jhete (Poly), IC2 (Sol), BIC2 (Poly), Jhetv (Sol), IVDE (Sol)	0.942	24.57

involved in the best seven-parameter equation were highly significant according to the *t*-test. The similar observation was also found in the equations containing eight or nine indices. Thus, the final correlation obtained for the training set contained 10 indices, with $R^2 = 0.8874$ and $R^2_{adj} = 0.8762$. The correlation equation for θ (LCST) is the following:

$$\begin{aligned} \theta (\text{LCST}) = & -166.975 + 53.901X3v(\text{Sol}) + 36.305X4v(\text{Sol}) \\ & - 26.060TI1(\text{Poly}) + 218.370IVDM(\text{Sol}) \\ & - 6.420TIC2(\text{Sol}) - 89.700Jhete(\text{Poly}) \\ & + 136.480IC2(\text{Sol}) + 148.520BIC2(\text{Poly}) \\ & + 35.922Jhetv(\text{Sol}) - 44.559IVDE(\text{Sol}) \\ (n = 112, R = 0.9420, R^2 = 0.8874, R^2_{adj} = 0.8762, R^2_{CV} = 0.8658, SEE = 24.57, \\ SEE(CV) = 25.71, F = 79.58, p < 0.000000). \end{aligned} \quad (3)$$

Here, X3v is the valence connectivity index chi-3 [21]; X4v is the valence connectivity index chi-4 [21]; TI1 is the first Mohar index [31]; IVDM is the mean information content on the vertex degree magnitude [32]; TIC2 is the total information content index (neighborhood symmetry of two-order) [33];

Jhete is the Balaban-type index from electronegativity weighted distance matrix [34]; IC2 is the information content index (neighborhood symmetry of two-order) [33]; BIC2 is the bond information content (neighborhood symmetry of two-order) [33]; Jhetv is the Balaban-type index from van der Waals weighted distance matrix [34]; IVDE is the mean information content on the vertex degree equality [32]; Sol and Poly corresponds to solvent and polymer, respectively.

The predicted θ (LCST) for the training and test sets are presented in Tables 1 and 2. Figs. 1 and 2 plot the experimental versus predicted values for the training and test sets. The distributions of relative errors (REs) for the training and test sets are also given in Fig. 3. Among the 112 training systems, only five systems have REs greater than 10%, (see the bold and italic values in Table 1) two have REs greater than 20% (26.10% for PBD-*n*-octane, and 23.47% for PPCS-ethyl carbitol) and 80 present REs less than 4.0%. Among the 57 test systems, 9 have REs greater than 10%, and 30 present REs less than 4.0%. The MREs for the training and test sets are 3.49 and 4.95%, respectively. These values are clearly improved compared to the model described in the article by Liu and Zhong [17], where the corresponding MREs were 5.44 and 5.63% for the training and test sets, respectively. It is noteworthy that the experimental data used in this work were

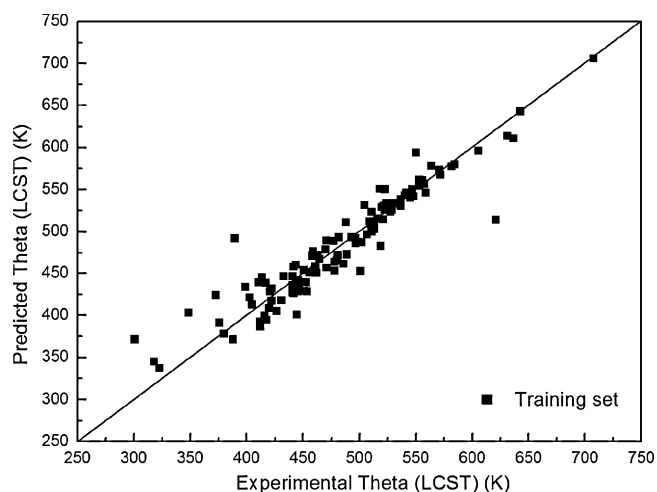


Fig. 1. Plot of predicted vs. experimental values of theta for the training set with 10 indices involved.

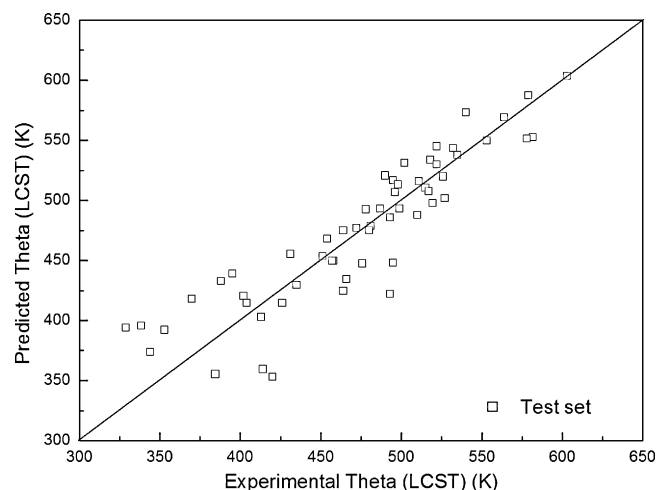


Fig. 2. Plot of predicted vs. experimental values of theta for the test set.

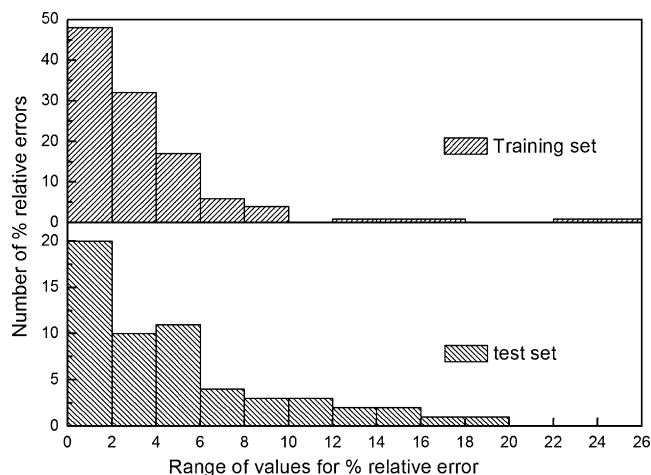


Fig. 3. Distributions of relative errors for the training and test set.

Table 4

Indices involved in the best 10-parameter correlation derived for θ (LCST)

Index	X	DX	t-test	p-Level
Constant	-166.975	42.487	-3.930	0.000156
X3v (Sol)	53.901	7.193	7.493	0.000000
X4v (Sol)	36.305	8.475	4.284	0.000042
TI1 (Poly)	-26.060	1.828	-14.256	0.000000
IVDM (Sol)	218.370	14.086	15.502	0.000000
TIC2 (Sol)	-6.420	0.461	-13.931	0.000000
Jhete (Poly)	-89.700	7.326	-12.244	0.000000
IC2 (Sol)	136.480	15.954	8.555	0.000000
BIC2 (Poly)	148.520	19.639	7.562	0.000000
Jhetv (Sol)	35.922	7.921	4.535	0.000016
IVDE (Sol)	-44.559	12.575	-3.543	0.000600

taken from various sources and have different measurement accuracies and systemic errors due to different measuring methods. As a result, the experimental uncertainties in θ (LCST) can be quite large, depending on polymer samples used, the measurement methods and techniques. Keeping these factors in mind, it can be considered that the present model works well for the representation of θ (LCST) in polymer solutions.

All the indices in Eq. (3) are highly significant as Table 4 shows the values of p and the standard errors for each index. The correlation matrix for the indices in Eq. (3) is given in Table 5, which shows that the 10 indices are independent of each other and could be used to develop a QSPR model. The

Table 5

Correlation matrix for the selected indices

	X3v (Sol)	X4v (Sol)	TI1 (Poly)	IVDM (Sol)	TIC2 (Sol)	Jhete (Poly)	IC2 (Sol)	BIC2 (Poly)	Jhetv (Sol)	IVDE (Sol)
X3v (Sol)	1									
X4v (Sol)	0.658	1								
TI1 (Poly)	0.160	0.245	1							
IVDM (Sol)	0.542	0.613	0.340	1						
TIC2 (Sol)	0.441	0.334	0.198	0.739	1					
Jhete (Poly)	-0.392	-0.294	-0.419	-0.147	-0.214	1				
IC2 (Sol)	-0.174	-0.194	0.097	0.241	0.670	0.101	1			
BIC2 (Poly)	-0.343	-0.279	-0.276	-0.138	-0.158	0.706	0.073	1		
Jhetv (Sol)	0.385	-0.030	-0.128	0.045	0.184	-0.327	-0.124	-0.268	1	
IVDE (Sol)	-0.226	-0.383	-0.088	-0.088	0.381	0.055	0.752	0.044	0.279	1

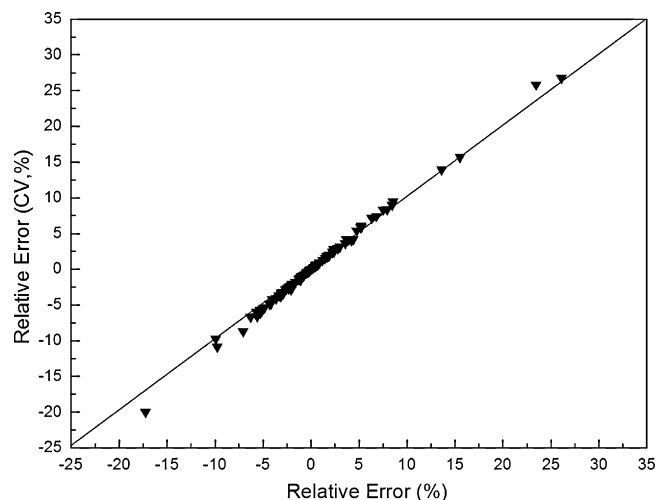


Fig. 4. Relative error obtained with the QSPR model vs. RE obtained by cross-validation for theta.

LOO cross-validation was used to test the stability of the obtained model and the results are shown in Table 1 and Fig. 4. The MRE for the cross-validation is 3.80%. The discrepancies between the REs obtained with the QSPR model and by cross-validation are small for most of the studied systems, which indicates that the model is reliable to predict the θ (LCST) in polymer solutions.

The same model size and algorithm that produced the QSPR model for the standard experiment were also tested with the randomized dependent variables. The most predictive model with SEE of 55.7 ($R^2 = 0.404$, $R^2_{\text{adj}} = 0.363$) was obtained with the randomized dependent variables. The values of SEE and R^2 indicate that a poor correlation was found between structure and θ (LCST), which proves the validity of the real model.

The results obtained exclusively using topological indices herein are comparable in quality to those obtained by Melagraki et al. [3] for the θ (LCST) in polymer solutions using connectivity indices and quantum-chemical descriptors ($R^2 = 0.8874$ versus $R^2 = 0.8860$, $R^2_{\text{CV}} = 0.8658$ versus $R^2_{\text{CV}} = 0.8546$, $F = 79.58$ versus $F = 88.04$, $\text{SEE} = 24.57$ versus $\text{SEE} = 25.51$, $\text{MRE} = 3.49\%$ versus $\text{MRE} = 3.73\%$). This confirms the potency of topological indices as a useful tool for the structural characterization and prediction of θ (LCST) in polymer solutions.

4. Conclusions

This work shows that θ (LCST) in polymer solutions can be estimated with a general QSPR model ($R^2 = 0.8874$ and $MRE = 3.49\%$) using only topological indices as input parameters. The proposed model gives reasonable accuracy and it is predictive because topological indices can be calculated easily as long as the molecular structure of the compound concerned is known. Therefore, reliable predictions of θ (LCST) in polymer solutions can be obtained before they are actually synthesized. This work also demonstrates that topological indices are useful for the structural characterization of polymers.

Acknowledgements

Special thanks were given to Prof. Todeschini and other members in Milano Chemometrics and QSAR Group for providing Dragon package to use in this research.

References

- [1] H. Liu, C. Zhong, Modeling of the θ (lower critical solution temperature) in polymer solutions using molecular connectivity indices, *Eur. Polym. J.* 41 (2005) 139–147.
- [2] A.R. Imre, Y.C. Bae, B.H. Chang, T. Kraska, Semiempirical method for the prediction of the theta (lower critical solution temperature) in polymer solutions, *Ind. Eng. Chem. Res.* 43 (2004) 237–242.
- [3] G. Melagraki, A. Afantitis, H. Sarimveis, P.A. Koutentis, J. Markopoulos, O. Igglessi-Markopoulou, A novel QSPR model for predicting θ (lower critical solution temperature) in polymer solutions using molecular descriptors, *J. Mol. Model.* 13 (1) (2007) 55–64.
- [4] C. Zhong, W. Wang, H. Lu, Open-cell model equation of state for liquids. 1. P – V – T behavior for liquids and liquid polymers, *Macromolecules* 27 (1994) 660–664.
- [5] C. Zhong, W. Wang, H. Lu, Open-cell model equation of state for liquids. 2. Polymer solutions and blends, *Macromolecules* 28 (1995) 7737–7743.
- [6] B.H. Chang, Y.C. Bae, Liquid–liquid equilibria of binary polymer solutions with specific interactions, *Polymer* 39 (1998) 6449–6454.
- [7] G. Bogdanic, J. Vidal, A segmental interaction model for liquid–liquid equilibrium calculations for polymer solutions, *Fluid Phase Equilib.* 173 (2000) 241–252.
- [8] G.D. Pappa, E.C. Voutsas, D.P. Tassios, Liquid–liquid phase equilibrium in polymer–solvent systems: correlation and prediction of the polymer molecular weight and the pressure effect, *Ind. Eng. Chem. Res.* 40 (2001) 4654–4663.
- [9] A. Vetere, Empirical method to correlate and to predict the vapor–liquid equilibrium and liquid–liquid equilibrium of binary amorphous polymer solutions, *Ind. Eng. Chem. Res.* 37 (1998) 2864–2872.
- [10] A. Vetere, An empirical method to predict the liquid–liquid equilibria of binary polymer systems, *Ind. Eng. Chem. Res.* 37 (1998) 4463–4469.
- [11] F. Wang, S. Saeki, T. Yamaguchi, Absolute prediction of upper and lower critical solution temperatures in polymer/solvent systems based on corresponding state theory, *Polymer* 40 (1999) 2779–2785.
- [12] X.J. Yao, Y.W. Wang, X.Y. Zhang, R.S. Zhang, M.C. Liu, Z.D. Hu, B.T. Fan, Radial basis function neural network-based QSPR for the prediction of critical temperature, *Chemom. Intell. Lab. Syst.* 62 (2002) 217–225.
- [13] J. Xu, B. Chen, Q. Zhang, B. Guo, Prediction of refractive indices of linear polymers by a four-descriptor QSPR model, *Polymer* 45 (2004) 8651–8659.
- [14] A. Afantitis, G. Melagraki, H. Sarimveis, P.A. Koutentis, J. Markopoulos, O. Igglessi-Markopoulou, Prediction of intrinsic viscosity in polymer–solvent combinations using a QSPR model, *Polymer* 47 (2006) 3240–3248.
- [15] X. Yu, X. Wang, H. Wang, A. Liu, C. Zhang, Prediction of the glass transition temperatures of styrenic copolymers using a QSPR based on the DFT method, *J. Mol. Struct. (Theochem)* 766 (2006) 113–117.
- [16] X. Yu, X. Wang, J. Gao, X. Li, H. Wang, QSPR studies of polyvinyls by density functional theory, *Polymer* 46 (2005) 9443–9451.
- [17] H. Liu, C. Zhong, General correlation for the prediction of theta (lower critical solution temperature) in polymer solutions, *Ind. Eng. Chem. Res.* 44 (2005) 634–638.
- [18] R. Garcia-Domenech, J.V.d. Julian-Ortiz, Prediction of indices of refraction and glass transition temperatures of linear polymers by using graph theoretical indices, *J. Phys. Chem. B* 106 (2002) 1501–1507.
- [19] B.E. Mattioni, P.C. Jurs, Prediction of glass transition temperatures from monomer and repeat unit structure using computational neural networks, *J. Chem. Inf. Comput. Sci.* 42 (2002) 232–240.
- [20] L.B. Kier, L.H. Hall, *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York, 1976.
- [21] L.B. Kier, L.H. Hall, *Molecular Connectivity in Structure Activity Analysis*, Wiley, New York, 1986.
- [22] L. Pogliani, From molecular connectivity indices to semiempirical connectivity terms: recent trends in graph theoretical descriptors, *Chem. Rev.* 100 (2000) 3827–3858.
- [23] C. Zhong, C. Yang, Q. Li, Correlation of Henry's constants of nonpolar and polar solutes in molten polymers using connectivity indices, *Ind. Eng. Chem. Res.* 41 (2002) 2826–2833.
- [24] J. Xu, B. Chen, Prediction of glass transition temperatures of OLED materials using topological indices, *J. Mol. Model.* 12 (2005) 24–33.
- [25] A. Katritzky, S. Sild, M. Karelson, Correlation and prediction of the refractive indices of polymer by QSPR, *J. Chem. Inf. Comput. Sci.* 38 (1998) 1171–1176.
- [26] C. Cao, Y. Lin, Correlation between the glass transition temperatures and repeating unit structure for high molecular weight polymers, *J. Chem. Inf. Comput. Sci.* 43 (2003) 643–650.
- [27] R. Todeschini, V. Consonni, A. Mauri, M. Pavan, DRAGON, Version 5.1, TALETE srl, Milan, 2004.
- [28] R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley-VCH, Weinheim, 2000.
- [29] P.C. Jurs, *Computer Software Applications in Chemistry*, Wiley, New York, 1996.
- [30] J.D. Jobson, *Applied Multivariate Data Analysis*, vol. 1: Regression and Experimental Design, Springer-Verlag, New York, 1991.
- [31] B. Mohar, in: A. Graovac (Ed.), *MATH/CHEM/COMP 1988*, Elsevier, Amsterdam, 1989, pp. 1–8.
- [32] D. Bonchev, *Information Theoretic Indices for Characterization of Chemical Structures*, RSP-Wiley, Chichester, 1983.
- [33] V.R. Magnuson, D.K. Harriss, S.C. Basak, in: R.B. King (Ed.), *Studies in Physical and Theoretical Chemistry*, Elsevier, Amsterdam, 1983, pp. 178–191.
- [34] M. Barysz, G. Jashari, R.S. Lahh, A.K. Srivastava, N. Trinajstić, in: R.B. King (Ed.), *Chemical Applications of Topology and Graph Theory*, Elsevier, Amsterdam, 1983, pp. 222–230.