

Calculation of Hildebrand Solubility Parameters of Some Polymers Using QSPR Methods Based on LS-SVM Technique and Theoretical Molecular Descriptors

Nasser Goudarzi*, M. Arab Chamjangali and A.H. Amin

Faculty of Chemistry, Shahrood University of Technology, Box 316, Shahrood, Iran

Abstract In this work, some chemometrics methods are applied for the modeling and prediction of the Hildebrand solubility parameter of some polymers. A genetic algorithm (GA) method is designed for the selection of variables to construct two models using the multiple linear regression (MLR) and least square-support vector machine (LS-SVM) methods in order to predict the Hildebrand solubility parameter. The MLR method is used to build a linear relationship between the molecular descriptors and the Hildebrand solubility parameter for these compounds. Then the LS-SVM method is utilized to construct the non-linear quantitative structure-activity relationship (QSAR) models. The results obtained using the LS-SVM method are then compared with those obtained for the MLR method; it was revealed that the LS-SVM model was much better than the MLR one. The root-mean-square errors of the training set and the test set for the LS-SVM model were 0.2912 and 0.2427, and the correlation coefficients were 0.9662 and 0.9518, respectively. This paper provides a new and effective method for predicting the Hildebrand solubility parameter for some polymers, and also reveals that the LS-SVM method can be used as a powerful chemometrics tool for the quantitative structure-property relationship (QSPR) studies.

Keywords: Hildebrand solubility parameter; Least square-support vector machine (LS-SVM); Quantitative structure-property relationship (QSPR); Multiple linear regression (MLR); Genetic algorithm (GA).

INTRODUCTION

The Hildebrand solubility parameter is the square root of the cohesive energy density. The cohesive energy density is the amount of energy required to completely remove a unit volume of molecules from their neighbors to infinite separation (an ideal gas), which is equal to the heat of vaporization divided by molar volume. For a material to dissolve, these same interactions need to be overcome as the molecules are separated from each other and surrounded by the solvent. Hildebrand has suggested the square root of the cohesive energy density as a numerical value indicating the solvency behavior of a solvent known as the “Hildebrand solubility parameter”. Materials with similar solubility parameters are able to interact with each other, resulting in solvation, miscibility or swelling^[1, 2].

Division of the Hildebrand parameter into the three-component Hansen parameters (dispersion force, polar force, and hydrogen bonding force) considerably increases the accuracy with which non-ionic molecular interactions can be predicted and described. Hansen parameters can be used to interpret not only the solubility behavior but also the mechanical properties of polymers, pigment binder relationships and activity of surfactants and emulsifiers. Being a three-component system, however, place limitations on the ease with which this information can be practically applied. Translating this three-component data into a 2D graph (by ignoring one of the components) solves this problem, although a certain amount of accuracy is sacrificed at the same time.

* Corresponding author: Nasser Goudarzi, E-mail: goudarzi@shahroodut.ac.ir

Received July 15, 2013; Revised November 19, 2013; Accepted November 24, 2013

doi: 10.1007/s10118-014-1423-z

What is needed is a simple, planar graph on which polymer solubility areas can be drawn in their entirety in two dimensions. A triangular graph meeting these qualifications has been introduced by Teas in 1968 using a set of fractional parameters mathematically derived from the three-component Hansen parameters. Due to its clarity and ease of use, the Teas graph has found increasing application among conservators for problem-solving, documentation and analysis, and it is an excellent vehicle for teaching the practical solubility theory. In order to plot all the three parameters on a single planar graph (using Teas graph), a certain departure must be made using the established solubility theory. Construction of the Teas graph is based upon the hypothetical assumption that all materials have the same Hildebrand value. According to this assumption, solubility behavior is determined not by differences in the total Hildebrand value but by the relative amounts of the three-component forces (dispersion force, polar force, and hydrogen bonding force) that contribute to the total Hildebrand value. This allows us to speak in terms of percentages rather than unrelated sums.

Hansen parameters are additive components of the total Hildebrand value:

$$\partial_t^2 = \partial_d^2 + \partial_p^2 + \partial_h^2 \quad (1)$$

where ∂_t^2 , ∂_d^2 , ∂_p^2 and ∂_h^2 are total Hildebrand parameter, dispersion component, polar component and hydrogen bonding component respectively. In other words, if all the three Hansen values (squared) are added together, their sum is equal to the Hildebrand value for that liquid (squared). Teas parameters, called fractional parameters are mathematically derived from Hansen values and indicate the percent contribution that each Hansen parameter contributes to the whole Hildebrand value:

$$f_d = \frac{\partial_d}{\partial_d + \partial_p + \partial_h} \quad f_p = \frac{\partial_p}{\partial_d + \partial_p + \partial_h} \quad f_h = \frac{\partial_h}{\partial_d + \partial_p + \partial_h} \quad (2)$$

In other words, if all the three fractional parameters are added together, the sum will always be the same (100).

$$f_d + f_p + f_h = 100 \quad (3)$$

The solubility parameter, δ , is very useful in determining the fundamental properties of materials, especially in the coatings^[3] and pharmaceutical industries^[4]. The values of polymeric δ can be obtained from the experimental approach indirectly, group contributions theory or QSPR models. The first approach is to deduce this property from the results of various experiments, so that its value is subject to great uncertainty, and spans a wide range^[5]. The second approach fails to account for the presence of neighboring groups or conformational influences^[6]. There are various methods used to determine the Hildebrand solubility parameter^[7, 8]. Recently, some molecular modeling methods based on widely spread quantitative structure-property/activity relationships (QSPR/QSAR) techniques have found their place as an important tool for the chemical engineers and chemists^[9–11]. Also, QSPR approach has been widely used in the study of polymeric properties such as the glass transition temperature (T_g)^[12–14] and refractive index^[15].

The ultimate role of the different formulations of the QSPR theory is to suggest mathematical models for estimating the relevant endpoints of interest, especially when these cannot be experimentally determined for some reason. These studies simply rely on the assumption that the structure of a compound determines the physico-chemical properties it manifests. The molecular structure is therefore translated into the so-called molecular descriptors through mathematical formulae obtained from several theories such as the chemical graph theory, information theory and quantum mechanics theory^[16, 17]. There exist more than a thousand of theoretical descriptors available in the literature, and one usually faces the problem of selecting those which are the most representatives for the property under consideration. There are some other reports about the quantitative structure-property/activity relationship studies^[18–33].

The aim of the present article is to propose and validate a novel, more rapid approach for the development of robust QSPR models for the Hildebrand solubility parameter (δ) as based on the large space of theoretically calculated molecular descriptors. In this work, two linear and non-linear methods including the MLR and

LS-SVM were used for predicting the Hildebrand solubility parameter (δ) values for some polymeric compounds, and the results of these methods were compared. Also the genetic algorithm (GA) method was used for the variables selection. According to the literature search, to the best of our knowledge, this is the first report published involving prediction of δ of polymeric compounds using the MLR and LS-SVM methods.

MATERIALS AND METHODS

Data Set

The experimental Hildebrand solubility parameter (δ) data for some polymers were taken from Ref. [1]. These data contain in total namely 97 solubility parameter values. Also the Hildebrand solubility parameters for these polymers calculated by the MLR and LS-SVM methods and experimental data of these compounds were obtained under the same instrumental conditions.

Molecular Descriptors

It is impossible to calculate descriptors directly for an entire molecule because all polymers have a wide distribution of molecular weights, and possess high molecular weights. However, the molecular descriptors calculated directly from their repeating unit structures can be used in the QSPR studies of the polymers^[7, 12–14]. A variety of different types of descriptors can be used.

First, the 2D structures of the molecules were drawn by the Hyperchem 7 software^[34]. These were pre-optimized with the molecular mechanics force field (MM+) and final geometries were obtained using the semi-empirical AM1 method in the Hyperchem program. It's mentioned that, molecular mechanics expresses the total energy as a sum of Taylor series expansions for stretches for every pair of bonded atoms, and adds additional potential energy terms coming from bending, torsional energy, van der Waals energy, electrostatics and cross terms. By separating out the van der Waals and electrostatic terms, molecular mechanics attempts to make the remaining constants more transferrable among molecules than they would be in a spectroscopic force field. All calculations were carried out at the restricted Hartree-Fock level with no configuration interaction. The molecular structures were optimized using the Polak-Ribiere algorithm until the root mean square gradient was $0.0042 \text{ kJ}\cdot\text{mol}^{-1}$. The resulting geometry was transferred into the Dragon program package, which was developed by the Milano chemometrics and QSPR group^[35], to calculate about 1479 descriptors in the constitutional, topological, geometrical, charge, GETAWAY (geometry, topology and atoms-weighted assembly), WHIM (weighted holistic invariant molecular descriptors), 3D-MoRSE (3D-molecular representation of structure based on electron diffraction), molecular walk count, BCUT, 2D-autocorrelation, aromaticity index, randic molecular profile, radial distribution function, functional group and atom-centered fragment classes. Initially, the 1479 descriptors were analyzed for the existence of constant or near-constant variables. The detected ones were then removed leaving 695 descriptors, as too many of them included zero values and did not have any information on structures. Secondly, correlation among descriptors and the Hildebrand solubility parameter of the molecules was examined, and collinear descriptors (*i.e.* correlation coefficient between descriptors greater than 0.9) were detected. Descriptors containing a high percentage ($> 90\%$) of identical values for all molecules were discarded to decrease the redundancy in the descriptor data matrix. Among the collinear descriptors, the one presenting the highest correlation with the activity to be predicted was retained and others were removed from the data matrix.

RESULTS AND DISCUSSION

Polymers play important roles in human life and industry. In fact, our body is made of lot of polymers, *e.g.* proteins, enzymes, *etc.* Other naturally occurring polymers like wood, rubber, leather and silk have been serving the humankind for many centuries. Modern scientific tools revolutionized the processing of polymers giving available synthetic polymers like useful plastics, rubbers and fiber materials. As with other engineering materials (metals and ceramics), the properties of polymers are related with their constituent structural elements and their arrangement. Therefore, to calculate the properties of these compounds is very important. On the other hand, the

QSAR/QSPR studies help us to avoid consuming time and money. In this work, we paid attention to make a model with high precision and accuracy to predict the δ values of 97 polymers.

The task of variable selection is to find a subset of variables from n variables. The number of possible combinations is equal to 2^n , where n is the number of variables. Conventional variable selection methods like stepwise MLR are based upon a single solution or a few solutions. To overcome this problem, a GA designed for the selection of variables was used^[36]. A GA, based on evolutionary principles, is a stochastic optimization method^[37–40]. What all methods based on GAs have in common is that they optimize not only a single solution but also the population of solutions. Thus the first step is to generate the initial population of chromosomes. In this population, each chromosome of length n (the number of features) consists of zeros and ones indicating the selected features. Implementing genetic competition requires the definition of a fitness function for the chromosome population. In the next step, reproduction takes place. The next population is the result of genetic manipulation of the chromosome in the current population through recombination (cross-over) and mutation basis on their fitness scores. This process going from the current population to the next one is called one generation or one cycle. Generation of new populations is repeated until a satisfactory solution is identified. In this work, the GA parameters were cross-validation random subset: number of subsets, 5; iteration, 100; population size, 64; initial terms (%), 30; maximum generations, 100; convergence (%) 50, mutation rate, 0.005; and cross-over, double. The selected descriptors using GA were Hb, alk, nk, Qii, Ein, and QH. These descriptors were used as input for MLR and LS-SVM to construct reliable models in order to predict the Hildebrand solubility parameter for the polymers studied. As it can be seen in Table 1, there is no significant correlation between the descriptors that were used in this study.

Table 1. Correlation matrix for six selected descriptors

	Hb	alk	nk	Qii	Ein	QH
Hb	1					
alk	0.0043	1				
nk	0.1132	0.0102	1			
Qii	0.1076	0.1438	0.0036	1		
Ein	0.1197	0.0514	0.0601	0.6937	1	
QH	0.3635	0.0188	0.0166	0.0499	0.1098	1

LS-SVM was performed with radial basis function (RBF) as a kernel function. Thus γ (the relative weight of the regression error) and σ (the kernel parameter of the RBF kernel) had to be optimized. The optimal parameters were found by an intensive grid search method. The result of this grid search was an error-surface spanned by the model parameters. A robust model was obtained by selecting the parameters that gave the lowest error in a smooth area. In order to find the optimized combination of the parameters γ and σ , a process of leave-one-out cross validation of the whole training set was performed. The σ parameter of the RBF kernel in the style of σ^2 and the γ parameter were tuned simultaneously in a grid 20×20 ranging from 0.01 to 10^9 and 0.01 to 10^2 , respectively. In this case, the optimized values for σ^2 and γ were 121 and 6, respectively. The predicted results of the optimal LS-SVM model and also the experimental and calculated δ values by MLR are shown in Table 2.

Table 2. The experimental and calculated Hildebrand solubility parameters

No.	Compound	Exp. δ (kJ/cm ³) ^{0.5}	LS-SVM	MLR
1	polyethylene	0.0732	0.0717	0.0720
2	poly(vinyl alcohol)	0.1297	0.1298	0.1305
3 ^t	poly(vinyl chloride)	0.0887	0.0852	0.0883
4 ^v	poly(vinyl bromide)	0.0883	0.0868	0.0895
5	poly(vinyl acetate)	0.0879	0.0873	0.0858
6	poly(vinyl ethyl ether)	0.0728	0.0745	0.0803
7 ^t	poly(<i>N</i> -vinyl pyrrolidone)	0.0933	0.0934	0.0933
8 ^v	poly(vinyl propionate)	0.0853	0.0858	0.0828

Table 2. Continued

No.	Compound	Exp. δ (kJ/cm ³) ^{0.5}	LS-SVM	MLR
9	poly(vinyl <i>n</i> -butyl ether)	0.0728	0.0776	0.0782
10	poly(<i>p</i> - <i>t</i> -butyl styrene)	0.0766	0.0797	0.0778
11 ^t	poly(methyl acrylate)	0.0895	0.0860	0.0841
12 ^v	polypropylene	0.0703	0.0717	0.0707
13	poly(vinyl cyclohexane)	0.0786	0.0792	0.0757
14	poly(<i>p</i> -vinyl pyridine)	0.0937	0.0942	0.0933
15 ^t	poly(vinylidene chloride)	0.0954	0.0980	0.0958
16 ^v	poly(vinylidene bromide)	0.0954	0.0964	0.0983
17	poly(ethyl acrylate)	0.0858	0.0816	0.0828
18	poly(<i>n</i> -butyl acrylate)	0.0824	0.0801	0.0807
19 ^t	poly(<i>n</i> -butyl methacrylate)	0.0799	0.0815	0.0799
20 ^v	poly(methyl α -cyanoacrylate)	0.1083	0.1079	0.1113
21	poly(sec-butyl methacrylate)	0.0786	0.0815	0.0800
22	poly(isobutyl methacrylate)	0.0786	0.0817	0.0795
23 ^t	polyacrylonitrile	0.1150	0.1119	0.1088
24 ^v	polystyrene	0.0841	0.0827	0.0824
25	poly(<i>p</i> -methyl styrene)	0.0812	0.0820	0.0816
26	poly(<i>p</i> -chloro styrene)	0.0887	0.0883	0.0874
27 ^t	poly(<i>o</i> -chloro styrene)	0.0887	0.0892	0.0895
28 ^v	poly(<i>p</i> -bromo styrene)	0.0891	0.0893	0.0887
29	poly(methyl methacrylate)	0.0845	0.0816	0.0828
30	polymethacrylonitrile	0.1025	0.1034	0.1063
31 ^t	poly(ethyl methacrylate)	0.0820	0.0809	0.0816
32 ^v	poly(α -methyl styrene)	0.0807	0.0824	0.0812
33	poly(<i>n</i> -propyl acrylate)	0.0837	0.0808	0.0820
34	poly(cyclohexyl methacrylate)	0.0828	0.0818	0.0807
35 ^t	poly(<i>n</i> -hexyl methacrylate)	0.0786	0.0780	0.0778
36 ^v	poly(benzyl methacrylate)	0.0866	0.0866	0.0841
37	poly(<i>n</i> -octyl methacrylate)	0.0778	0.0778	0.0757
38	poly(<i>N</i> -vinyl carbazole)	0.0904	0.0906	0.0950
39 ^t	poly(α -vinyl naphthalene)	0.0874	0.0847	0.0837
40	polyisobutylene	0.0669	0.0675	0.0736
41 ^v	poly(1-butene)	0.0715	0.0717	0.0694
42	poly(4-methyl-1-pentene)	0.0703	0.0717	0.0674
43 ^t	poly(1, 2-butadiene)	0.0720	0.0717	0.0715
44 ^v	poly(ethyl α -chloroacrylate)	0.0900	0.0865	0.0895
45	poly(<i>p</i> -fluoro styrene)	0.0837	0.0857	0.0849
46	poly(isobutyl acrylate)	0.0812	0.0808	0.0820
47 ^t	poly(<i>t</i> -butyl methacrylate)	0.0766	0.0815	0.0799
48 ^v	poly(2-ethoxyethyl methacrylate)	0.0803	0.0820	0.0803
49	Poly(vinyl sec-butyl ether)	0.0711	0.0736	0.0782
50	poly(acrylic acid)	0.1075	0.1078	0.1088
51 ^t	polyacrylamide	0.1176	0.1171	0.1130
52 ^v	poly(vinyl butyrate)	0.0767	0.0779	0.0821
53	poly(vinyl phenyl ether)	0.0845	0.0839	0.0085
54	poly(vinyl methyl sulfide)	0.0817	0.0828	0.0858
55 ^t	poly(vinyl methyl ether)	0.0822	0.0790	0.0816
56 ^v	poly(<i>m</i> -methyl styrene)	0.0809	0.0821	0.0819
57	poly(methoxy styrene)	0.0845	0.0828	0.0828
58	poly(1-methyl vinyl ethyl ether)	0.0807	0.0785	0.0794
59 ^t	poly(1-ethyl vinyl ethyl ether)	0.0804	0.0775	0.0787
60 ^v	poly(1-phenyl vinyl ethyl ether)	0.0838	0.0863	0.0812

Table 2. Continued

No.	Compound	Exp. δ (kJ/cm ³) ^{0.5}	LS-SVM	MLR
61	poly(1,1-diphenyl ethylene)	0.0834	0.0838	0.0833
62	poly(allyl 4, tolyl ether)	0.0838	0.0827	0.0812
63 ^t	poly(allyl cyanide)	0.1065	0.1019	0.1012
64 ^v	poly(vinyl ethyl ketone)	0.0926	0.0889	0.0818
65	poly(vinyl phenyl sulfide)	0.0848	0.0867	0.0873
66	poly(vinyl propyl ether)	0.0807	0.0783	0.0793
67 ^t	poly(vinyl isopropyl ether)	0.0807	0.0785	0.0794
68 ^v	poly(vinyl isoamyl ether)	0.0800	0.0782	0.0772
69	poly(4-hydroxystyrene)	0.1027	0.1029	0.1007
70	poly(divinyl ether)	0.0792	0.0807	0.0831
71 ^t	poly(vinyl-1-phenyl methyl ether)	0.0844	0.0832	0.0802
72 ^v	poly(vinyl-1-methyl phenyl ether)	0.0844	0.0833	0.0826
73	poly(vinyl-1-phenyl phenyl ether)	0.0858	0.0845	0.0847
74	poly(nitro styrene)	0.0950	0.0933	0.0987
75 ^t	poly(2-nitro styrene)	0.0929	0.0933	0.0981
76 ^v	poly(benzyl acrylate)	0.0811	0.0840	0.0855
77	poly(<i>o</i> -methyl styrene)	0.0809	0.0821	0.0820
78	poly(vinyl isobutyl ether)	0.0804	0.0793	0.0782
79 ^t	poly(2-ethyl hexyl acrylate)	0.0772	0.0779	0.0770
80 ^v	poly(allyl isocyanide)	0.1065	0.1019	0.1004
81	poly(isopropyl methacrylate)	0.0770	0.0799	0.0810
82	poly(allyl phenyl ether)	0.0845	0.0834	0.0821
83 ^t	poly(allyl methyl ether)	0.0813	0.0787	0.0804
84 ^v	poly(allyl ethyl ether)	0.0807	0.0783	0.0793
85	poly(allyl propyl ether)	0.0804	0.0779	0.0777
86	poly(allyl isopropyl ether)	0.0804	0.0774	0.0788
87 ^t	poly(diallyl ether)	0.0788	0.0797	0.0797
88	poly(allyl 2, tolyl ether)	0.0838	0.0821	0.0826
89	poly(allyl 3, tolyl ether)	0.0838	0.0826	0.0815
90	poly(allyl acetate)	0.0764	0.0776	0.0838
91 ^t	poly(allyl acetonitrile)	0.1012	0.1013	0.0982
92	poly(propyl methacrylate)	0.0769	0.0802	0.0804
93	poly(3-chloropropyl methacrylate)	0.0820	0.0829	0.0880
94	poly(cyano styrene)	0.0935	0.0935	0.0982
95 ^t	poly(4-acetoxy styrene)	0.0908	0.0834	0.0858
96	poly(vinyl methyl ketone)	0.0959	0.0942	0.0845
97	poly(vinyl-1-amyl methyl ether)	0.0871	0.0849	0.0765

t = test set and v = validation set

Figure 1 shows a plot of the LS-SVM predicted values of the training and test sets against the experimental values of the solubility parameter for the molecules included in the data set. The residuals of the LS-SVM-calculated values of the training and test sets of the solubility parameter were plotted against the experimental values in Fig. 2. The propagation of residuals at both sides of the zero line indicates that no systematic error exists in the development of the LS-SVM model.

For evaluation of the prediction ability of the LS-SVM model, we performed some statistical tests. The correlation coefficients (R^2) for the training, validation, and test sets were 0.9662, 0.9533, and 0.9518, respectively. Also the root mean square error of prediction (RMSEP) for the training, validation, and test sets were 0.4510, 0.5007, and 0.6669, respectively. The relative standard error of prediction (RSEP) for the training, validation, and test sets were 2.2295, 2.4345, and 3.1305, respectively. The statistical test shows that this LS-SVM model is achieved in modeling. Thus the prediction results indicate that LS-SVM is a powerful method in the QSAR study of the class of solubility parameter.

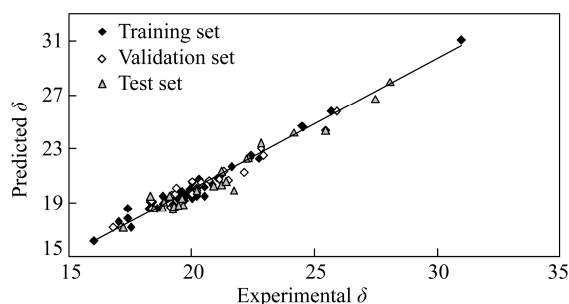


Fig. 1 Plot of the calculated Hildebrand solubility parameter against the experimental values

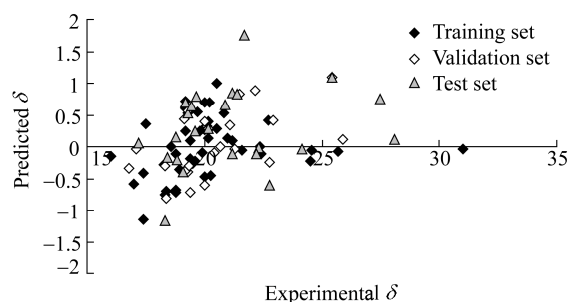


Fig. 2 Plot of the residuals versus experimental values of the Hildebrand solubility parameter

CONCLUSIONS

The results obtained from the training set, validation set, and test set show that the QSPR model can provide satisfactory prediction of the solubility parameter for polymers. Since these descriptors can be calculated from the repeating unit structure, and have clear physical meanings, the model proposed in this paper is predictive and very easy to apply. The results obtained reveal the superiority of LS-SMV over the MLR model. This is due to the ability of the LS-SMV model to allow for flexible mapping of the selected features by manipulating their functional dependence implicitly, unlike regression analysis. Descriptors appearing in these QSPR models provide information related to different molecular properties which can participate in the physico-chemical process that affects the solubility parameter of the polymers.

REFERENCES

- Burke, J., Part 2. "Hildebrand Solubility Parameter". The American Institute for Conservation, Washington, 1984, p. 121
- Vandenburg, H., *Analyst*, 1999, 124: 1707
- Lindvig, T., Michelsen, M.L. and Kontogeorgis, G.M., *Fluid Phase Equilib.*, 2002, 203: 247
- Bustamante, P., Navarro-Lupión, J. and Escalera, B., *Eur. J. Pharm. Sci.*, 2005, 24: 229
- Bicerano, J., "Prediction of Polymer Properties". Second Edition, Marcel Dekker, New York, 1996, p. 126
- Delgado, E.J., *Fluid Phase Equilib.*, 2002, 199: 101
- Eguiazabal, J.I., Fernández-Berridi, M.J., Iruin, J.J. and Elorza, J.M., *Polym. Bull.*, 1985, 13: 463
- King, J.W., *Food Sci. Technol.*, 1995, 28: 190
- Goudarzi, N. and Goodarzi, M., *Mol. Phys.*, 2008, 106: 2525
- Goodarzi, M. and Freitas, M.P., *QSAR Comb. Sci.*, 2008, 27: 1092
- Fatemi, M.H. and Goudarzi, N., *Electrophoresis*, 2005, 26: 2968
- Afantitis, A., Melagraki, G., Makridima, K., Alexandridis, A., Sarimveis, H. and Markopoulou, O.I., *J. Mol. Struct., (Theochem)* 2005, 716: 193
- Mattioni, B.E., and Jurs, P., *J. Chem. Inf. Comput. Sci.*, 2002, 42: 232
- GarcQa-Domenech, R. and de Julia'n-Ortiz, J.V., *J. Phys. Chem. B*, 2002, 106: 1501
- Katritzky, A.R., Sild, S. and Karelson, M., *J. Chem. Inf. Comput. Sci.*, 1998, 38: 1171
- Trinajstić, N. "Chemical Graph Theory". CRC Press, Boca Raton, FL, 1992, p. 216
- Katritzky, A.R., Lobanov, V.S. and Karelson, M., *Chem. Rev. Soc.*, 1995, 24: 279
- Goudarzi, N. and Goodarzi, M., *Mol. Phys.*, 2009, 107: 1739
- Goudarzi, N. and Goodarzi, M., *Mol. Phys.*, 2009, 107: 1615
- Elmi, Z., Faez, K., Goodarzi, M. and Goudarzi, N., *Mol. Phys.* 2009, 107: 1787
- Goudarzi, N. and Goodarzi, M., Araujo, M.C.U. and Galvao, R.K.H., *J. Agric. Food Chem.*, 2009, 57: 7153

- 22 Goudarzi, N., Fatemi, M.H. and Samadi, A., *Spect. Lett.*, 2009, 42: 186
- 23 Goodarzi, M. and Freitas, M.P., *QSAR Comb. Sci.*, 2008, 27: 1092
- 24 Goodarzi, M. and Freitas, M.P., *J. Phys. Chem. A*, 2008, 112: 11263
- 25 Goodarzi, M. and Freitas, M.P., *Chemometr. Intell. Lab. Syst.*, 2008, 96: 59
- 26 Golmohammadi, H. and Fatemi, M.H., 2005, 26: 3438
- 27 Goodarzi, M., Freitas, M.P. and Jensen, R., *J. Chem. Inf. and Mod.*, 2009, 49: 824
- 28 Goodarzi, M., Duchowicz, P.R., Wu, C.H., Fernández, F.M. and Castro, E.A., *J. Chem. Inf. and Mod.*, 2009, 49: 1475
- 29 Goudarzi, N., Goodarzi, M. and Arab Chamjangali, M., *J. Environ. Chem. Ecotoxicol.*, 2010, 2: 47
- 30 Goudarzi, N. and Kalhor, P., *Anal. Chem. Lett.*, 2012, 2: 13
- 31 Goudarzi, N. and Goodarzi, M., *Anal. Methods*, 2010, 2: 758
- 32 Goudarzi, N., Goodarzi, M. and Chen, T., *Med. Chem. Res.*, 2012, 21: 437
- 33 Hosseini, J., Nekoei, M., Mohammadhosseini, M. and Goudarzi, N., *J. App. Res. Chem.*, 2011, 5: 5
- 34 HyperChem Release 7, HyperCube, Inc., <http://www.hyper.com>
- 35 Todeschini, R., Milano Chemometrics and QSPR Group, <http://www.disat.unimib.it/vhml>
- 36 Siedlecki, W. and Sklansky, J., *Pat. Recog. Let.*, 1989, 10: 335
- 37 Abbass, H.A., Sarker, R.A. and Newton, C.S., "Data mining: A Heuristic Approach". University of New South Wales, Australia, 2002, p. 341
- 38 Jia, Y., Liu, L., Mu, Y. and An, L., *Acta Mater.*, 2004, 52: 4153
- 39 Chen, J.S. and Hou, J. L., "A Combination Genetic Algorithm with Applications on Portfolio Optimization". National Central University, Jungli, Taiwan, 320
- 40 Rosenberg, A.N., Pritchard, J.K., Weber, J.L., Cann, H.M., Kidd, K.K., Zhivotovsky, L.A. and Feldman, M.W., *Science*, 2002, 298: 2381