ICC
Iranian Chemical Communication
Original Research Article
Payame Noor University http://icc.journals.pnu.ac.ir

The use of topological indices to predict thermodynamic properties of amino acid derivatives

Afsaneh Safari, Fatemeh Shafiei*

Department of Chemistry, Arak Branch, Islamic Azad University, P.O. BOX 38135-567, Arak, Iran

Received: 11 October 2018, Accepted: 24 October 2018, Published: 1 July 2019

Abstract

In the present investigation the applicability of various topological indices are tested for the QSPR study on 80 amino acid derivatives. Relationship between the Randic' (^1X), Balaban (J), Szeged (Sz), Harary (H), Wiener (W), Hyper-Wiener (WW) and Wiener Polarity (Wp) indices to the thermodynamic properties such as thermal energy E_{th} (J/mol) and heat capacity (C_V J/mol. K) of amino acids is represented. The thermodynamic properties are taken from HF level using the ab initio 6-31G basis sets from the program package Gaussian 98. We have used Multiple Linear Regression (MLR) techniques and followed back ward regression analysis for obtaining properties. By analyzing the correlation between the indices in the made models, the most suitable indicators for modeling properties were determined. The predictive powers of the models were discussed using leave-one-out (LOO) cross-validation. The obtained results show that combining of the two descriptors (J, ^1X) could be used successfully for modeling and predicting the heat capacity (C_V), and thermal energy (E_{th}) of amino acid derivatives.

Keywords: Amino acids; QSPR; MLR method; topological indices; validation.

Introduction

Amino acids (AAs) are building blocks of proteins. AAs are organic compounds with an amino group (-NH_2) and an acid carboxy group (-COOH) [1]. There are over 500 amino acids found in nature, yet, of these, the human genetic code includes information for only 20 amino acids. Every protein in the human body is made up of some linked combination of these amino acids.

In theoretical chemistry, topological indices have been used to develop molecular graph descriptors in order to express chemical structures in the numerical form [2,3]. Graph theory has been successfully applied in developing novel topological indices to predict some thermodynamic properties of compounds [4]. It is one of the most important tools in the fields of Quantitative Structure-Property Relationship (QSPR) and Quantitative Structure-Activity Relationship (QSAR) [5-8]. QSPR/QSAR are mathematical models designed for predicting the properties of a wide range of chemical compounds based on the correlation between these properties and molecular descriptors as topological indices [9-14]. Topological indices (TI) are the digital values associated to chemical constitution for

*Corresponding author: Fatemeh Shafiei
Tel: +98 (918) 3617386, Fax: +98 (86) 13670017
E-mail: f-shafiei@iau-arak.ac.ir

correlation of chemical structure with various physical properties, chemical reactivity or biological activity and useful mathematical methods for finding good relationship between several data of the properties in these materials [15–17]. The use of these methods for making good correlations between several data properties of chemicals is important. Graph-theoretical topological indices are taken into consideration attention because they can be obtained directly without any trial achieved of the molecular structure [18]. For this reason, these indices in QSPR/QSAR studies which are means of a simple and clear design are molecules. Topological indices such as molecular connectivity index of Randic' [19] and the Wiener [20], Balaban [21], Hosoya [22] indices have received great attention due to their applications in chemistry and drug research. Relationship between topological indices and thermodynamic properties such as heat capacity, standard Gibbs energy of formation, thermal energy and entropy of the monocarboxylic acids, alkanes [23,24], alcohols [25], aldehydes and ketones [26,27] has been searched. QSAR studies of the natural and unnatural amino acids were developed using partial least squares (PLS) regression and a novel 3D amino acid descriptor [28,29]. Structural topology scale (ST-scale) as a novel amino acid descriptor was applied for the study of quantitative sequence-activity models (QSAMs) of 167 amino acids using PLS regression [30]. Quantitative relationships studies between structure and physical properties using topological and quantum-chemical molecular descriptors were employed to derive descriptors for the 20 natural amino acids [31]. To determine the three dimensional structure of protein, 10 orthogonal factors were obtained by using 188 properties of the natural amino acids [32]. In the present study, the multiple linear regression (MLR) techniques and back ward methods are estimated for modeling the thermal energy (E_{th}) and heat capacity (C_V) of 80 amino acids. The main aim of this study is to provide reliable QSPR models for predicting thermodynamic properties of 80 natural and unnatural amino acid derivatives.

Materials and methods

The data set consists of 80 compounds, 20 natural amino acid and 60 unnatural of the amino acids have been selected. A complete list of the compound names are listed in Table 1. All thermodynamic data of the present investigation were obtained from the quantum mechanics methodology with Hartree- Fock (HF) level using the ab initio 6-31G basis sets. The quantum chemistry data of the amino acid derivatives are listed in Table 1.

Topological indices

Mathematical descriptors have been widely used in structure-property-activity studies. The descriptors used for the QSPR analyses were calculated with the chemicalize database program [33]. The topological indices (molecular descriptors) were Wiener index, W [20,34], Hyper-Wiener index, WW [35,36], Wiener polarity index, Wp [37], Randić index, ^1X [19, 38], Balaban index, J [21,39], Harary number, H [40], Szeged index, Sz [41,42]. Nowadays, hundreds of topological indices, suitable to describe different properties, are reported. Seven topological indices tested in the present study are recorded in Table 2.

Statistical analysis

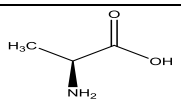
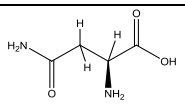
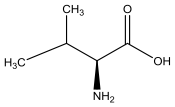
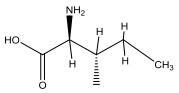
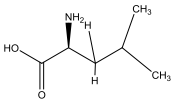
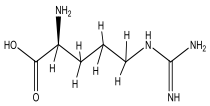
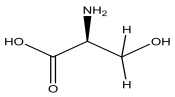
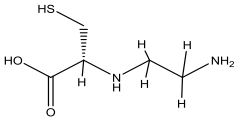
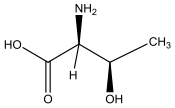
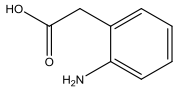
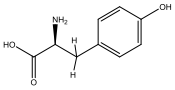
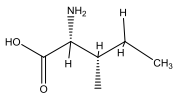
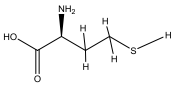
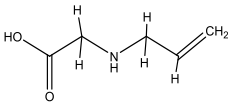
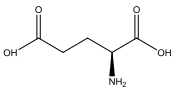
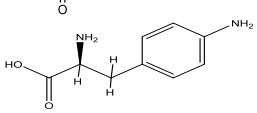
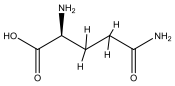
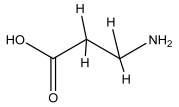
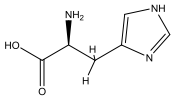
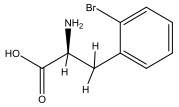
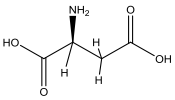
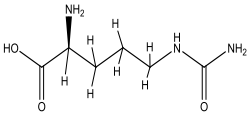
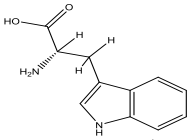
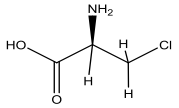
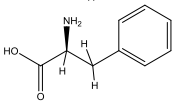
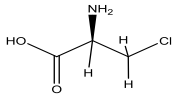
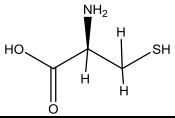
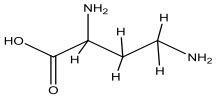
In the present study, Structure- Property models (MLR models) are generated using the multilinear regression

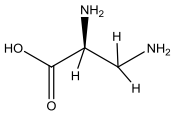
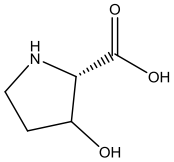
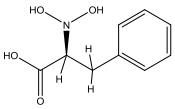
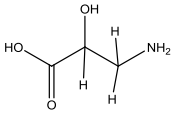
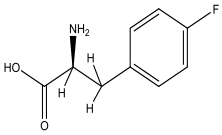
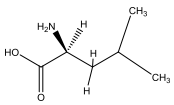
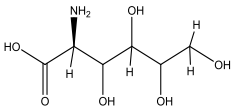
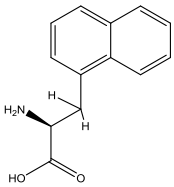
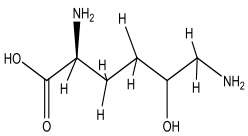
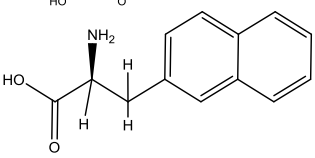
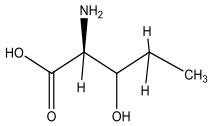
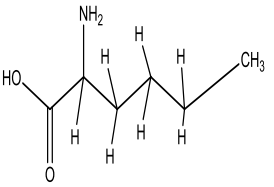
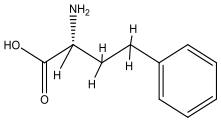
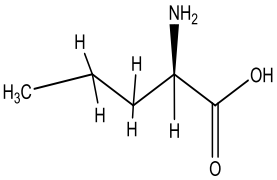
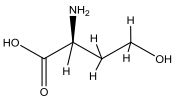
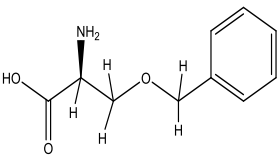
procedure of SPSS version 20 and backward stepwise regression was used to construct the QSPR models.

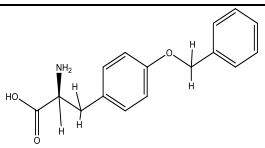
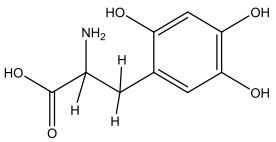
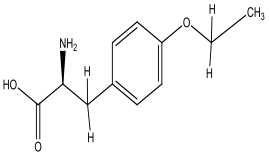
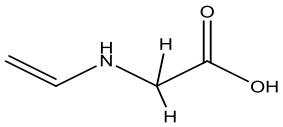
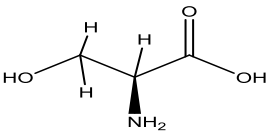
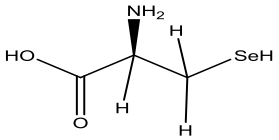
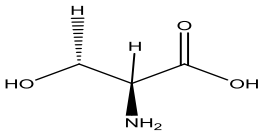
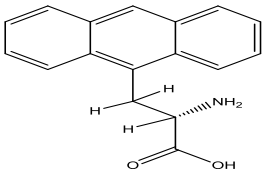
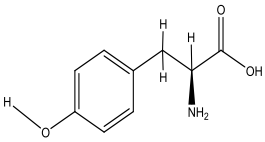
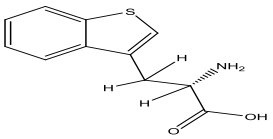
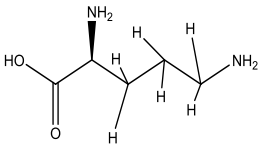
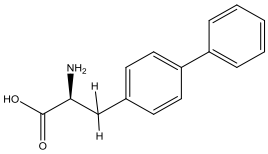
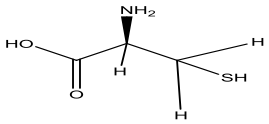
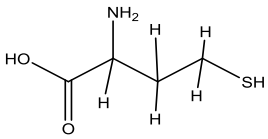
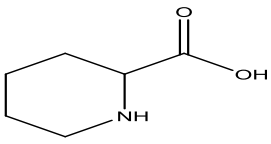
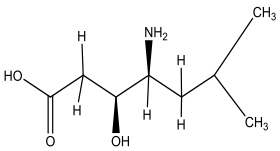
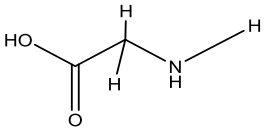
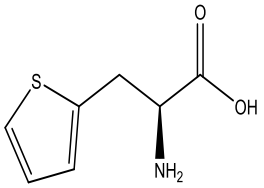
Regression analyses

The thermal energy (E_{th} kJ/mol) and heat capacity (C_v J/mol. K) are used as the dependent variable and topological indices as the independent variables.

Table 1. Amino acid derivatives used in present study

No.	Compounds	Structure	No.	Compounds	Structure
1	Alanine		**15**	Asparagine	
2	Valine		**16**	Isoleucine	
3	Leucine		**17**	Arginine	
4	Serine		**18**	N-(2-Aminoethyl)cysteine	
5	Threonine		**19**	2-Aminophenylacetate	
6	Tyrosine		**20**	D-allo-Isoleucine	
7	Methionine		**21**	Allylglycine	
8	Glutamat		**22**	4-Amino-L-phenylalanine	
9	Glutamine		**23**	β-Alanine	
10	Histidine		**24**	2-Bromophenylalanine	
11	Aspartic acid		**25**	3-Cyclohexylalanine	
12	Tryptophan		**26**	L-Citrulline	
13	Phenylalanine		**27**	3-Chloro-L-alanine	
14	Cysteine		**28**	2,4-Diaminobutanoic acid	

No.	Compounds	Structure	No.	Compounds	Structure
29	DL-3-Aminoalanine		**37**	L-Hydroxyproline	
30	dihydroxyphenylalanine		**38**	Isoserine	
31	4-Fluorophenylalanine		**39**	2-Methyl-L-leucine	
32	3,4,5,6-Tetrahydroxynorleucine		**40**	l-3-(1-naphthyl)alanine	
33	5-Hydroxylysine		**41**	3-(2-Naphthyl)alanine	
34	3-Hydroxynorvaline		**42**	DL-Norleucine	
35	(2R)-2-Amino-4-phenylbutanoic acid		**43**	L-Norvaline	
36	Homoserine		**44**	O-Benzylserine	

No.	Compounds	Structure	No.	Compounds	Structure
45	O-Benzyltyrosine		**54**	L-6-Hydroxydopa	
46	O-Ethyltyrosine		**55**	vinylglycine	
47	O-Methylserine		**56**	L-selenocysteine	
48	O-Methylthreonine		**57**	3-(9-Anthryl)alanine	
49	O-Methyltyrosine		**58**	3-(1-Benzothiophen-3-yl)-L-alanine	
50	L-Ornithine		**59**	(2S)-2-Amino-3-(4-biphenylyl)propanoic acid	
51	Penicillamine		**60**	DL-Homocysteine	
52	Pipecolic acid		**61**	Statine	
53	Sarcosine		**62**	3-(2-Thienyl)-L-alanine	

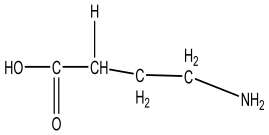
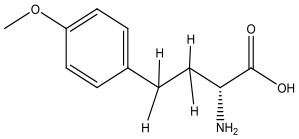
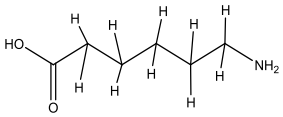
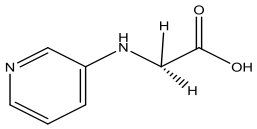
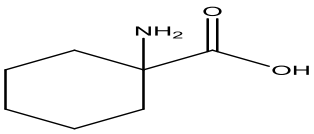
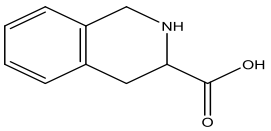
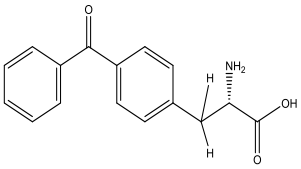
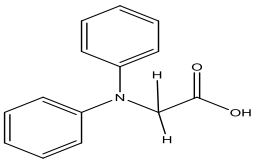
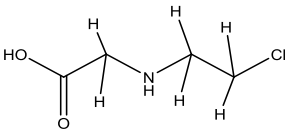
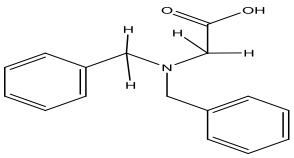
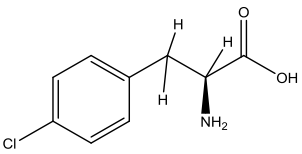
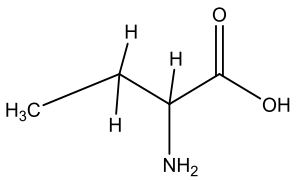
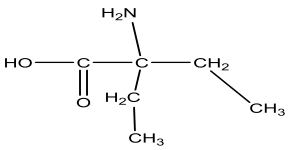
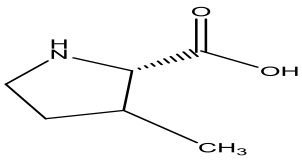
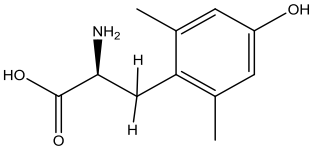
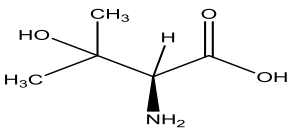
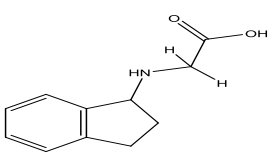
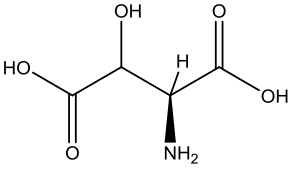
No.	Compounds	Structure	No.	Compounds	Structure
63	γ-Aminobutyric acid		**72**	(2R)-2-Amino-4-(4-methoxyphenyl)butanoic acid	
64	Aminocaproic acid		**73**	N-3-Pyridinylalanine	
65	1-Aminocyclohexanecarboxylic acid		**74**	Tetrahydroisoquinoline-3-carboxylic acid	
66	4-Benzoyl-L-phenylalanine		**75**	N,N-Diphenylglycine	
67	N-(2-Chloroethyl)glycine		**76**	N,N-Dibenzylglycine	
68	4-Chlorophenylalanine		**77**	2-Aminobutanoic acid	
69	2-Amino-2-ethylbutanoic		**78**	3-Methylproline	
70	2,6-Dimethyltyrosine		**79**	3-Hydroxyvaline	
71	1-Indanylglycine		**80**	β-hydroxyaspartate	

Table 2. Amino acids derivatives and their topological indices

No	^1X	J	H	W	WW	W_p	S_z	No	^1X	J	H	W	WW	W_p	S_z
1	29	4	47	29	9.33	2.99	2.64	41	46	6	83	46	12	3.14	3.18
2	18	2	28	18	6.67	2.54	2.27	42	70	7	143	70	14.73	3.17	3.68
3	65	8	122	65	15.17	3.46	4.09	43	460	15	1137	361	35.56	2.02	7.02
4	96	8	206	96	17.97	3.38	4.04	44	1438	25	4052	988	61.33	1.59	9.65
5	65	8	122	65	15.17	3.46	3.55	45	582	18	1342	420	40.23	2.05	7.13
6	376	15	735	268	32.99	2.05	6.09	46	70	7	143	70	14.73	3.17	3.68
7	102	8	235	102	17.55	3.16	4.18	47	92	10	188	92	18.23	3.58	4.09
8	136	9	330	136	20.87	3.3	5.07	48	472	17	997	337	36.64	2.22	6.63
9	182	10	397	165	25.6	1.98	5.2	49	102	8	235	102	17.55	3.16	4.18
10	96	8	206	96	17.97	3.38	4.04	50	86	10	164	86	18.83	3.88	3.85
11	518	20	1018	369	42.96	1.76	7.72	51	103	8	173	87	19.32	2.05	4.2
12	293	13	542	212	29.17	2.2	6.24	52	142	9	176	88	19.15	1.98	4.3
13	46	6	83	46	12	3.14	3.18	53	32	3	58	32	9	2.63	2.77
14	143	9	368	143	20.45	3.13	4.63	54	524	21	1035	374	41.99	2.31	6.91
15	92	10	188	92	18.23	3.58	4.09	55	46	6	83	46	12	3.14	3.18
16	247	11	739	247	26.92	3.2	5.54	56	46	6	83	46	12	3.14	3.18
17	134	10	317	134	20.83	3.4	4.72	57	793	21	2975	762	51.06	2.03	9.06
18	234	12	362	158	25.98	2.08	5.2	58	518	20	1018	369	42.96	1.76	7.18
19	122	12	259	122	21.6	3.8	4.63	59	1061	24	2414	683	54.21	1.61	8.66
20	70	7	143	70	14.73	3.17	3.86	60	70	7	143	70	14.73	3.17	3.68
21	376	6	83	46	12	3.14	3.18	61	220	12	576	220	28.04	3.68	5.43
22	375	15	735	268	32.99	2.05	6.09	62	182	10	397	165	25.6	1.98	5.2
23	32	3	58	32	9	2.63	2.77	3	52	4	108	52	11.48	2.68	3.27
24	352	16	659	256	33.4	2.15	6.11	64	114	6	297	114	16.84	2.75	4.27
25	247	11	739	247	26.92	3.2	5.54	65	174	13	220	111	23.23	2.48	4.65
26	46	6	83	46	12	3.14	3.18	66	842	25	1426	504	52.23	1.98	8.08
27	70	7	143	70	14.73	3.17	3.68	67	1366	28	3464	916	63.24	1.73	9.58
28	46	6	83	46	12	3.14	3.18	68	79	5	185	79	14.1	2.72	3.77
29	451	18	893	321	37.31	2.33	6.5	69	376	15	735	268	32.99	2.05	6.09
30	376	15	735	268	32.99	2.05	6.09	70	86	12	136	86	18.75	3.92	4.06
31	254	18	640	254	32.79	4.15	5.91	71	488	18	858	313	38.38	1.81	6.77
32	180	11	469	180	24.1	3.37	5.07	72	578	18	1390	428	39.95	1.99	7.13
33	92	10	188	92	18.23	3.58	4.09	73	289	15	431	188	31.37	1.79	5.77
34	70	7	143	70	14.73	3.17	3.68	74	293	13	542	212	29.17	2.2	5.7
35	103	8	173	87	19.32	2.05	4.2	75	462	18	580	238	35.22	1.64	6.27
36	736	15	735	268	32.99	2.05	6.09	76	762	22	1437	510	51.4	1.95	8.27

37	46	6	83	46	12	3.14	3.18	77	1044	22	2498	756	58.38	1.61	9.24
38	724	23	1213	434	47.43	1.69	7.68	78	46	6	83	46	12	3.14	3.18
39	772	22	1377	458	46.72	1.59	7.66	79	97	9	159	84	19.5	2.13	4.22
40	102	8	235	102	17.55	3.16	4.18	80	97	10	164	86	18.83	3.88	3.85

The models are assessed with R value (correlation coefficient), the R^2 (squared multiple correlation coefficient), the R^2_{adj} (adjusted correlation coefficient), the RMSE value (root of the mean square of errors), the F value (Fischer statistic), the D value (Durbin-Watson) and the Sig (significant).

Results

Multiple linear regression analysis has been carried out to derive the best QSPR model and strongest correlations are identified by the Back ward step wise regression routine implemented in SPSS is used to develop the linear model for the prediction of the thermo-physical properties.

QSPR models

The best linear model for Cv and E_{th} contains three topological descriptors, namely, Randić (^1X), Balaban (J), Hyper-Wiener (WW), Wiener polarity (Wp), Wiener (W) and Szeged (Sz) indices. The regression parameters of the best three descriptors correlation model are gathered in equation (1, 2).

C_v= 28.472+11.471(^1X) +8.946(J) +0.234(W) +3.439(Wp)-0.144(Sz) (1)

N=80, R=0.973, R^2 =0.946, R^2_{adj} = 0.943, RMSE = 9.566, F= 261.317, Sig=0.000 DW=1.783

E_{th} = -246.159 + 150.104(^1X)+ 66.416(J)-1.834 (W) +0.355(WW) (2)

N=80, R=0.934, R^2 =0.0.873, R^2_{adj} = 0.867, RMSE= 51.932, F=129.218, Sig=0.000 DW=1.872

Results and discussion

We studied the relationship between topological indices to the thermal energy (E_{th}) and heat capacity (C_v) of 80 amino acids derivatives. In this study, to find the best model for predicting the mentioned properties, we will use the following sections.

The Durbin-Watson statistic

The Durbin-Watson statistic ranges in value from 0 to 4. A value near 2 indicates non-autocorrelation; a value toward 0 indicates positive autocorrelation; a value toward 4 indicates negative autocorrelation. Therefore the value of Durbin-Watson statistic is close to 2 if the errors are uncorrelated. In our all models, the value of Durbin-Watson statistic is close to 2(See eqs.1, 2) and hence the errors are uncorrelated.

Multicollinearity

Multicollinearity in regression is a condition that occurs when some predictor variables in the model are correlated with other predictor variables. Good regression model should not exist in correlation between the independent variables or should not happen multicollinearity. Test multicollinearity is as a basis the variance inflation factor (VIF) value of multicollinearity test results using SPSS. If the VIF value lies between1-10, then there is no multicollinearity, and if the VIF<1 or >10, then there is multicollinearity. In all our final models, the multicollinearity has existed, because the values of correlations between independent

variables are near to one and VIFs value lies are not use of between 1-10.

Validation

Predictive power of the MLR models for a QSAR/QSPR analyses can be conveniently estimated using statistical parameters. A good QSPR model should have both suitable relativity and good predictability. In the constructed model internal validation is usually done by leave-one-out (LOO). We studied the validation of linearity between the molecular descriptors in the models 1 and 2. We obtained by SPSS the Pearson coefficient correlation and collinearity statistics as shown in Tables 3 and 4. For model 1 the Pearson correlation (Sz, W) and (Wp, ^{1}X) are near one, and VIF (Sz), (W), (^{1}X), and VIF (Wp)>10(See Table 3). After removed (W) from this model, we corrected models 1 as follows:

C_v= -22.418+ 26.498 (^{1}X) + 17.460 (J) (3)

N=80, R=0.964, R^2=0.930, R^2_{adj} = 0.928, RMSE= 10.698, F=513.412, Sig=0.000 DW=1.795, Q^2_{LOO} =0.824

For model 2 the Pearson correlation (WW, W), (WW, ^{1}X) and (^{1}X, W) are near one, and VIF (WW), (^{1}X) and VIF (W)>10(See Table 4). After removed Wiener index from this model, we corrected model 2 as follow:

E_{th} = -93.740+ 87.647 (^{1}X) +65.787(J) (4)

N=80, R=0.921, R^2 =0.848, R^2_{adj} =0.845, RMSE= 56.049, F=215.544, Sig=0.000 DW=1.967, Q^2_{LOO} =0.864

Table 3. Correlation between the molecular descriptors (model 1)

	Pearson correlations					**Collinearity Statistical**		**Corrected model**
	Sz	J	Wp	^{1}X	W	Tolerance	VIF	VIF
Sz	1	0.622	0.678	0.617	-0.945	0.024	42.376	-
J		1	-0.568	0.639	-0.605	0.287	3.489	1.992
Wp			1	-0.846	-0.643	0.010	97.175	-
^{1}X				1	-0.743	0.041	24.607	1.992
W					1	0.011	89.146	-

Table 4. Correlation between the molecular descriptors (model 2)

Pearson correlations					**Collinearity Statistical**		**Corrected model**
	WW	J	^{1}X	W	Tolerance	VIF	VIF
WW	1	0.017	-0.836	-0.983	0.027	148.05	-
J		1	-0.209	-0.032	0.531	1.883	1.869
^{1}X			1	-0.914	0.003	37.321	1.869
W				1	0.007	293.564	-

We have computed Q^2 (Eq.5) by 50% of data, randomly, that are positive and less than one.

$$\mathbf{Q^2 = 1 - \frac{\sum(Y_i - \hat{Y}_{i|i})^2}{\sum(Y_i - \bar{Y})^2} Q^2 \leq 1} \qquad (\mathbf{5})$$

Where the notation i|i indicates that the response is predicted by a model estimated when the i-th sample was left out from the data set.

Regular residuals

The residual is the difference between the observed value of the dependent variable (y) and the predicted (calculated) value (ŷ). Comparison between predicted (calculated) and observed values of C_v and E_{th} of respect amino acids which is shown in Tables 5, 6. Figures 1, 2 shows the linear correlation between the observed and the predicted heat capacity and the obtained thermal energy values using equations 3 and 4 respectively.

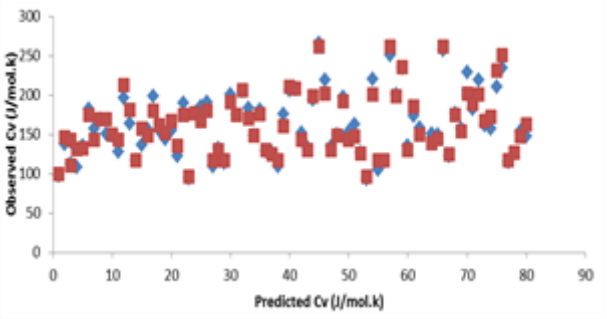


Figure 1. Comparison between the predicted and observed values of the heat capacity (model 1) by MLR method

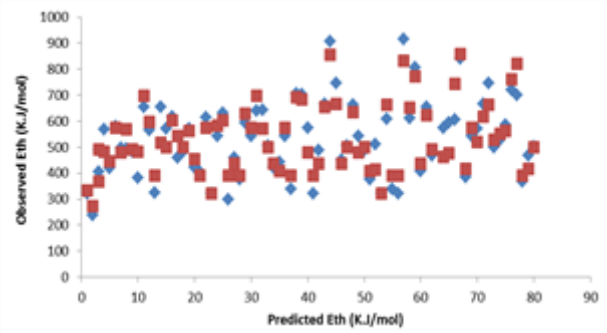


Figure 2. Comparison between the predicted and observed values of the thermal energy (model 2) by MLR method

Table 5. Comparison between predicted (calculated) and observed values of the heat capacity (C_V) of respect amino acid derivatives

Comp.No	Observed Cv(J/mol.K)	Predicted Cv(J/mol.K)	Residual	Comp.No	Observed Cv(J/mol.K)	Predicted Cv(J/mol.K)	Residual
1	97.66	99.74	-2.09	41	207.38	208.32	-0.95
2	137.64	146.37	-8.73	42	151.62	143.52	8.10
3	141.45	143.65	-2.20	43	133.42	130.44	2.98
4	108.28	130.98	-22.70	44	194.09	198.87	-4.78
5	135.52	132.06	3.45	45	265.98	261.05	4.93
6	181.91	174.75	7.16	46	219.11	202.31	16.80
7	157.78	143.52	14.26	47	134.32	130.44	3.87

8	168.88	169.55	-0.67	48	149.78	148.47	1.32
9	151.18	169.55	-18.37	49	197.18	192.03	5.15
10	148.54	149.95	-1.41	50	152.84	143.52	9.32
11	127.92	143.65	-15.73	51	162.41	147.35	15.06
12	195.91	212.88	-16.97	52	126.98	126.10	0.89
13	163.76	181.34	-17.59	53	93.04	96.90	-3.86
14	117.33	116.67	0.66	54	220.04	201.02	19.02
15	136.38	157.70	-21.32	55	104.65	116.67	-12.03
16	155.85	148.47	7.38	56	115.72	116.67	-0.95
17	198.99	180.26	18.74	57	251.02	261.49	-10.47
18	155.59	162.02	-6.43	58	200.41	198.57	1.84
19	144.16	151.69	-7.53	59	234.27	235.17	-0.90
20	155.41	166.62	-11.21	60	135.95	130.44	5.50
21	122.19	135.21	-13.02	61	173.91	185.72	-11.81
22	189.84	174.75	15.09	62	158.63	149.95	8.68
23	94.14	96.90	-2.76	3	113.01	111.03	1.98
24	177.26	177.03	0.23	64	150.47	138.75	11.72
25	185.91	167.04	18.88	65	149.22	144.10	5.12
26	190.97	180.26	10.71	66	256.88	261.64	-4.77
27	109.77	116.67	-6.90	67	122.5	124.97	-2.47
28	132.79	130.44	2.35	68	176.81	174.75	2.06
29	114.04	116.67	-2.63	69	157.13	153.61	3.52
30	201.26	190.50	10.76	70	228.87	201.72	27.15
31	174.97	174.75	0.21	71	182.65	188.58	-5.93
32	204.74	206.65	-1.90	72	219.73	201.26	18.47
33	183.64	170.77	12.87	73	161.08	167.04	-5.95
34	148.91	148.47	0.44	74	157.73	172.36	-14.63
35	181.01	175.75	5.26	75	210.93	230.77	-19.84
36	129.42	130.44	-1.03	76	234.41	250.54	-16.13
37	127.11	124.67	2.45	77	114.52	116.67	-2.16
38	109.11	116.67	-7.56	78	125.39	126.60	-1.20
39	175.49	160.69	14.80	79	154.05	147.35	6.70
40	206.58	210.60	-4.02	80	147.99	162.29	-14.30

Table 6. Comparison between predicted and observed values of the thermal energy (E_{th}) of respect amino acid derivatives

Comp.No	Observed E_{th}(J/mol.k)	Predicted E_{th}(J/mol)	Residual	Comp.No	Observed E_{th}(J/mol.k)	Predicted E_{th}(J/mol)	Residual
1	320.65	334.35	-13.70	41	321.37	391.55	-70.17
2	238.01	272.32	-34.31	42	489.12	437.34	51.78
3	487.14	492.36	-5.21	43	665.2	654.43	10.77
4	567.81	482.71	85.10	44	906.1	856.65	49.45
5	419.01	445.03	-26.02	45	746.35	666.04	80.30
6	579.05	574.89	4.16	46	450.27	437.34	12.93
7	495.66	480.51	15.15	47	504.48	500.25	4.23
8	494.52	567.72	-73.21	48	663.89	633.40	30.49
9	481.37	492.28	-10.91	49	542.36	480.51	61.85
10	383.17	482.71	-99.54	50	487.71	498.95	-11.24
11	654.54	698.68	-44.14	51	377.68	409.24	-31.56
12	564.66	597.91	-33.24	52	511.62	413.40	98.22
13	323.92	391.55	-67.63	53	321.4	322.06	-0.66
14	654.42	517.98	136.44	54	608.03	663.87	-55.83
15	571.39	500.25	71.14	55	338.66	391.55	-52.89
16	618.58	602.34	16.24	56	322.82	391.55	-68.72
17	459.71	543.63	-83.91	57	917.04	833.89	83.16
18	481.34	498.86	-17.51	58	610.71	651.35	-40.63
19	571.85	562.05	9.80	59	806.91	771.20	35.71
20	422.69	453.12	-30.43	60	408.29	437.34	-29.05
21	405.11	391.55	13.56	61	654.93	624.28	30.65
22	614.51	574.89	39.62	62	468.35	492.28	-23.93
23	323.09	322.06	1.03	3	406.26	369.17	37.09
24	542.44	583.22	-40.78	64	573.61	461.42	112.19
25	630.62	602.34	28.28	65	594.6	476.97	117.63
26	298.8	391.55	-92.74	66	604.76	744.70	-139.95
27	458.96	437.34	21.61	67	841.62	859.73	-18.11
28	375.32	391.55	-16.23	68	384.46	415.63	-31.17
29	593.47	629.25	-35.78	69	542.5	574.89	-32.39
30	543.9	574.89	-30.99	70	570.44	519.99	50.45
31	641.07	697.27	-56.20	71	667.36	618.70	48.66
32	643.41	572.33	71.08	72	745.55	662.10	83.45
33	505.02	500.25	4.77	73	498.93	529.74	-30.81
34	422.71	437.34	-14.63	74	528.5	550.58	-22.08

35	442.32	409.24	33.09	75	586.96	563.69	23.27
36	541.6	574.89	-33.29	76	719.77	759.38	-39.62
37	339.48	391.55	-52.07	77	699.81	822.03	-122.22
38	705.55	690.57	14.98	78	366.96	391.55	-24.58
39	704.38	682.23	22.15	79	468.39	416.25	52.13
40	573.05	480.51	92.54	80	501.88	498.95	2.93

Conclusion

At the present study, QSPR mathematical models for the prediction of the heat capacity (C_v) and thermal energy (E_{th}) of 80 amino acids derivatives using MLR method based on topological descriptors calculated from molecular structure alone have been developed. MLR model is proved to be a useful tool in the prediction of C_v and E_{th}. Cross-validation as the evaluation technique has been designed to evaluate the quality and predictive ability of the MLR model. The obtained results showed that the best model for predicting the heat capacity and thermal energy contains two parameters having to be optimized: the Randic (^1X) and Balaban (J) indices.

Acknowledgements

The authors would like to thank Islamic Azad University of Arak for their support on this work.

References

[1] H.D. Jakube, H. Jeschkeit, Eine Einführung, Akademie-Verlag, Berlin., **1982**, 26, 838-839.
[2] N. Ahmadinejad , M. Talebi Tari, *J. Chem. Method*., **2019**, *3*, 55-66.
[3] M. Nabati , M. alsadat Kermanian, H. Mohammadnejad-Mehrabani, H. Rahbar Kafshboran, M. Mehmannavaz, S. Sarshar, *J. Chem. Method*., **2018**, *2*, 128-140.
[4] A.T. Balaban, T.S. Balaban, *J. Math. Chem.*, **1991**, *8*, 383-397.
[5] P.J. Hansen, P.C. Jurs, *J. Chem. edu.*, **1988**, *65*, 574-580.
[6] G. Rucker, C. Rucker, *J. Chem. Inf. Comput. Sci.*, **1999**, *39*, 788-802.
[7] M. Shahpar, S. Esmaeilpoor, *A. J. Green Chem.*, **2017**, *1*, 116-129.
[8] Y. Boukarai, F. Khalil, M. Bouachrine, *J. Chem. Method*., **2017**, *1*, 173-193.
[9] M.P. Gonzales, A.M. Helguera, M.A. Cabrera, *Bioorg. Med. Chem.*, 2005, *13*, 1775-1781.
[10] M. Randic', S.C. Basak, *SAR. QSAR. Environ. Res.*, **2000**, *11*, 1-23.
[11] O. Ivanciuc, T. Ivanciuc, D. Cabrol-Bass, A.T. Balaban, *J. Chem. Inf. Comput. Sci.*, **2000**, *40*, 631-643.
[12] M. Nabati, *J. Chem. Method*., **2017**, *1*, 121-135.
[13] A.R. Katritzky, V.S. Lobanov, M. Karelson, *Chem. Soc. Rev.*, **1995**, *24*, 279-287.
[14] A. R.Katritzky, U. Maran, V.S. Lobanov, M. Karelson, *J. Chem. Inf. Comput. Sci.*, **2000**, *40*, 1-18.
[15] Y.P. Du, Y.Z. Liang, B.Y. Li, C.J. Xu, *J. Chem. Inf. Comput. Sci.*, **2002**, *42*, 1128-1138.
[16] Z. Slanina, F. Uhlik, S.L. Lee, E. Osawa, *MATCH Commun. Math. Comput. Chem.*, **2001**, *44*, 335-348.
[17] Yi. Gao, M. Farahani, W. Nazeer, *J. Chem. Method*., **2018**, *2*, 39-46.
[18] A.T. Balaban, *J. Chem. Inf. Comput. Sci.*, **1995**, *35*, 339-350.
[19] M. Randić, *J. Math. Chem.*, **1991**, *7*, 155-168.
[20] H. Wiener, *J. American. Chem. Soc.*, **1947**, *69*, 17-20.
[21] A.T. Balaban, *J. Chem. Inf. Comput. Sci.*, **1985**, *25*, 334-343.

[22] A.R. Nizami, T. Farman, *J. Appl. Comput. Math.*, **2018**, *7*, 1-5.
[23] F. Shafiei, *Iranian J. Math. Chem*., **2015**, *6*, 15-28.
[24] D. Bonchev, *J. Chem. Inf. Comput. Sci*., **2000**, *40*, 934-941.
[25] S. Liu, H. Liu, Z. Xia, C. Cao, Z. Li, *J. Chem. Inf. Comput. Sci*., **1999**, *39*, 951-957.
[26] O. Ivanciuc, T. Ivanciuc, D.J. Klein, W.A. Seitz, A.T. Balaban, *J. Chem. Inf. Comput. Sci.*, **2001**, *41*, 536-549.
[27] A. Alaghebandi, F. Shafiei, *Iranian J. Math. Chem.*, **2016**, *7*, 235-251.
[28] Elizabeth R. Collantes, and William J. Dunn III, *J. Med. Chem.*, **1995**, *38*, 2705-2713.
[29] R. Beigzadeh, *J. Chem. Method*., **2019**, *3*, 67-82.
[30] Li. Yang, Mao. Shu, Kaiwang. Ma, Hu. Mei, Yongjun. Jiang, Zhiliang. Li, *Amino Acids*., **2010**, *38*, 805–816.
[31] S. Sahoo, M. Kuanar, S. Patel, B.K. Mishra, *Indian. J. chem*., **2014**, *53A*, 1324-1331.
[32] A. Kidera, Y. konishi, M. Oka, T. Ooi, H. Scheraga, *J. Protein Chem.*, **1985**, *4*, 23–55.
[33] Web search engine developed by ChemAxon; software available at http:// WWW. Chemicalize. Org.
[34] M. Randić, *Acta. Chim. Slov*., **2002**, *49*, 483–496.
[35] B. Zhou, I. Gutman, *Chem. Phys. Lett.*, **2004**, *394*, 93-95.
[36] D.J. Klein, W. Yan, Y.N. Yeh, *INT. J. QUANTUM. CHEM*., **2006**, *106*, 1756-1761.
[37] M. Liu, B. Liu, *MATCH Commun. Math. Comput. Chem.*, **2011**, *66*, 293-304.
[38] M. Randić, X. Guo, T. Oxley, H. Krishnapriyan, *J. Chem. Inf. Comput. Sci*., **1994**, *34*, 361-367.
[39] A.T. Balaban, *Chem. Phys. Lett.*, **1982**, *89*, 399-404.
[40] K.C. Das, B. Zhou, N. Trinajstić, *J. Math. Chem.*, **2009**, *46*, 1369-1376.
[41] I. Gutman, S. Klavžar, *J.Chem. Inf. Comput. Sci.*, **1995**, *35*, 1011-1014.
[42] P.V. Khadikar, N.V. Deshpande, P.P. Kale, A. Dobrynin, I. Gutman, G. Dömötör, *J. Chem. Inf. Comput. Sci.*, **1995**, *35*, 547-550.

How to cite this manuscript: Afsaneh Safari, Fatemeh Shafiei. The use of topological indices to predict thermodynamic properties of amino acid derivatives. *Iranian Chemical Communication*, 2019, 7(3), 310-323.