Journal of Physical and Theoretical Chemistry
of Islamic Azad University of Iran, 18 (2) 23-44: Summer 2021
(J. Phys. Theor. Chem. IAU Iran)
ISSN 1735-2126

Quantitative structure–property relationship models to Predict thermodynamic properties of Imidazole Derivatives using molecular descriptor and genetic algorithm-multiple linear regressions

Shiva Moshayedi, Fatemeh Shafiei, * Tahereh Momeni Isfahani

Department of Chemistry, Arak Branch, Islamic Azad University, Arak, Iran

Received March 2021; Accepted May2021

ABSTRACT

Imidazole is a compound with a wide range of biological activities and imidazole derivatives are the basis of several groups of drugs. In this study the relationship between molecular descriptors and the thermal energy (E_{th} kJ/mol), and heat capacity (Cv J/mol) of imidazole derivatives is studied. The chemical structures of 85 Imidazole derivatives were optimized at HF/6-311G* level with Gaussian 98 software. Molecular descriptors were calculated for selected compounds by using the Dragon software. The Genetic algorithm- multiple linear regression (GA-MLR) and backward methods were used to select the suitable descriptors and also for predicting the thermodynamic properties of imidazole derivatives. The obtained models were evaluated by statistical parameters, such as correlation coefficient (R^2_{adj}), Fisher ratio (F), Root Mean Square Error (RMSE), Durbin-Watson statistic (D) and significance (Sig). The predictive powers of the GA- MLR models are studied using leave-one-out (LOO) cross-validation and external test set. The predictive ability of the GA-MLR models with two-three selected molecular descriptors was found to be satisfactory. The developed QSPR models can be used to predict the property of compounds not yet synthesized.

Keywords: QSPR; imidazole derivatives; leave-one-out (LOO) cross-validation; genetic algorithm- multiple linear regressions

INTRODUCTION

Quantitative Structure–Property Relationships (QSPRs) and Quantitative structure-activity relationship (QSAR) are useable method to predict and estimate real thermodynamic properties and biological activity of compounds approved on knowledge of molecular structure without having to synthesize and purify the substance.

Molecular descriptors play a fundamental role in creating models for variety of sciences such as health research, chemistry, quality control and toxicology. A molecular descriptor is defined as the final result of a logical and mathematical

*Corresponding author: f-shafiei@iau-arak.ac.ir

method, which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment.

Many different chemometrics methods, such as multiple linear regression (MLR), partial least squares regression (PLS), different types of artificial neural networks (ANN), genetic algorithms (GAs), and support vector machine (SVM) can be employed to derive correlation models between the molecular structures and properties.

Imidazole is an organic compound with the formula $C_3H_4N_2$. This aromatic heterocyclic is classified as an alkaloid. Imidazole ring system is present in important biological building blocks such as histidine, and the related hormone histamine. It has a wide range of pharmacological activity including; antifungal [1] antimycobacterial [2] antiprotozoal [3] analgesic[4] anticancer [5] angiotensin antagonist [6] antidepressant [7] antihistaminic [8] antihypertensive [9] and as heme oxygenase inhibitor [10-12].

Heterocyclic compounds containing nitrogen are widely used in the field of energetic materials as well as commonly used as corrosion inhibitors to protect the metals [13].

QSAR models have been improved to determine the biological activity (pIC50) of imidazole derivatives using the physicochemical parameters such as density (D), surface tension (St), index of refraction (Ior), balaban index (J) and partition coefficient (Log P) [14].

2D and 3D-QSAR analysis has been used on a series of imidazole derivatives to predict Heme Oxygenase Inhibitor, by the Partial Least Square regression (PLS) method [15].

Linear and non-linear 2D and 3D QSAR models have been developed to predict the activity of Imidazole against Alzheimer's disease using Stepwise Multiple Linear Regression (S-MLR) and neural networks [16].

QSAR models have been used for predicting the anticancer activity of 192 imidazole containing Farnesyl transe ferase inhibitors (FTIs) using genetic algorithms-partial least squares (GA-PLS), SVM, MLR and ANN methods [17].

QSAR models have been improved to determine the molecular surface area (MSA), molar volume (MV), molar refractivity (MR), lipophilicity (log P) and hydration energy (HE) of some imidazole derivatives using different electronic, topologic, functional groups and physicochemical descriptors and multiple linear regression[18].

QSPR and QSAR methods have been used to study the quantitative relationship between the toxicity and structure of 43 kinds of imidazolium ionic liquids [19].

Few studies have been conducted on the relationship between the chemical structure and physical properties of imidazole derivatives. In this study the relationship between the structure of imidazole derivatives and its thermodynamic properties have been researched using the Genetic algorithm- multiple linear regression (GA-MLR) methods and molecular descriptors.

MATERIALS AND MATHEMATICAL METHODS

The imidazoles discussed in this study consist of 83 derivatives and their chemical structures are shown in Table 1. These data were randomly divided into 2 groups: training and test sets consisting of 61(75%), 22(25%) data point, respectively. The chemical structures of molecules were drawn by Gauss View 05 program and then they were optimized with Gaussian09 using ab initio 6-311G* basis set method. Based on the output files created by

Gaussian, thermodynamic properties, such as the thermal energy (E_{th} kJ/mol) and heat capacity (Cv J/molK) of the imidazole derivatives were calculated (see Table 2).

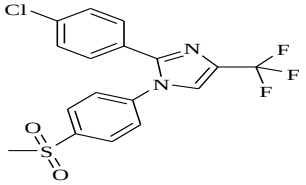
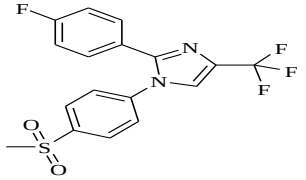
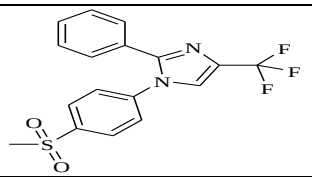
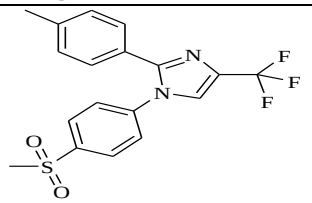
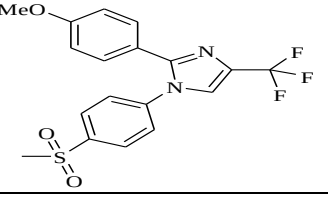
A set of different categories descriptors including Connectivity, 2D matrix-based, 2D autocorrelations, constitutional, geometrical, GETAWAY and topological descriptors was calculated for selected compound by using the Dragon software for Windows Version 5.4- 2006 package [20]. More than 1531 molecular descriptors calculated by the Dragon software for each molecule.

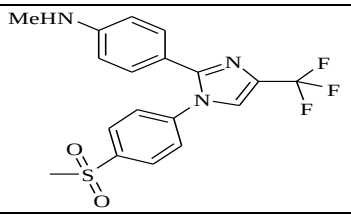
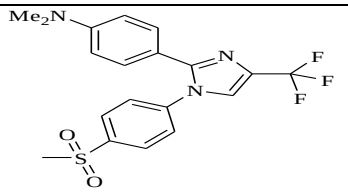
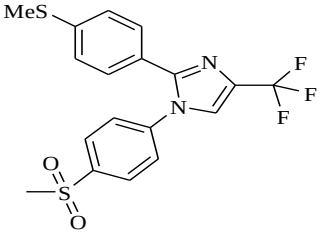
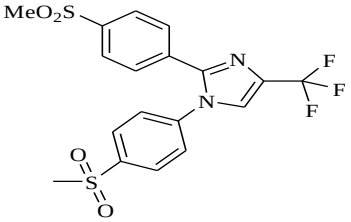
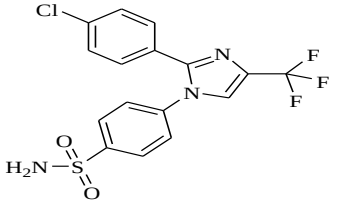
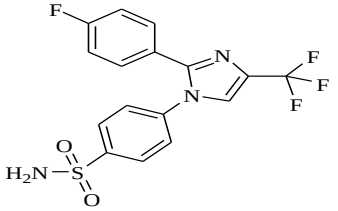
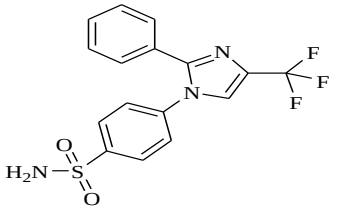
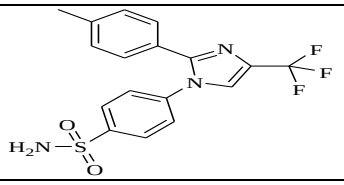
The genetic algorithm - multiple linear regression (GA-MLR) method and the SPSS program Version 21 were employed to give the QSPR model.

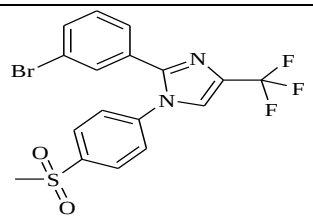
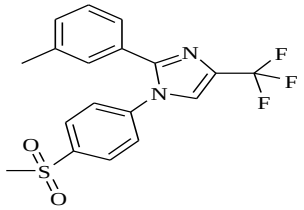
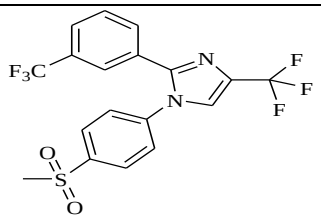
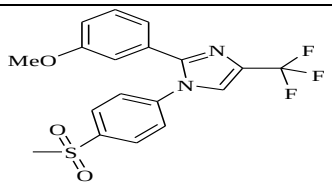
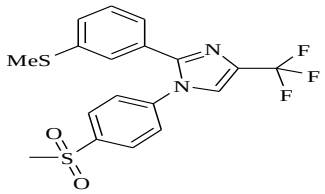
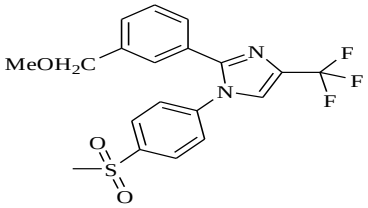
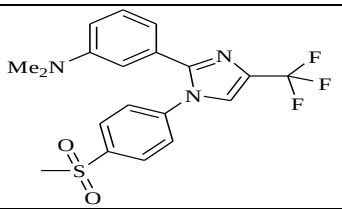
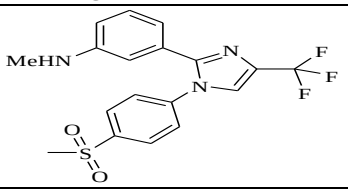
The most significant molecular descriptors among the selected descriptors were identified using the genetic algorithm method.

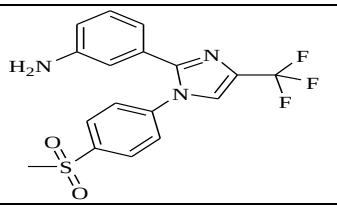
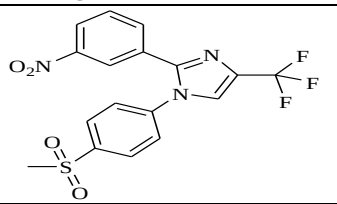
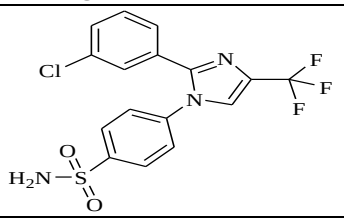
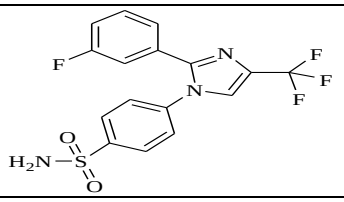
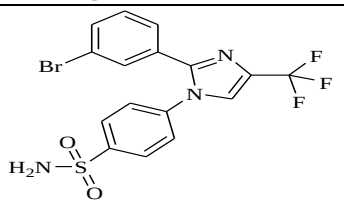
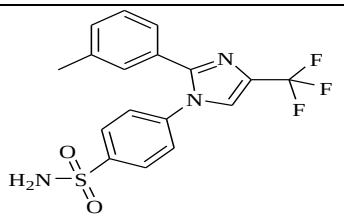
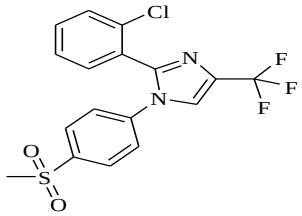
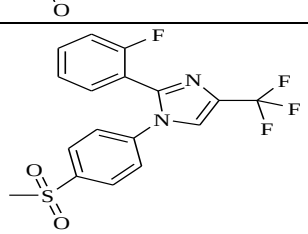
The genetic algorithm is implemented in MATLAB (2017a) software and backward stepwise regression was used to minimize obtained descriptors and build robust QSPR models.

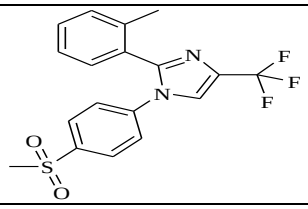
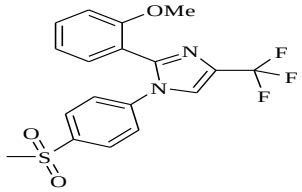
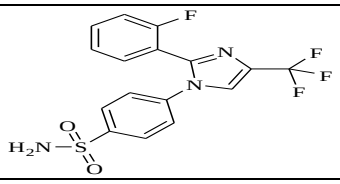
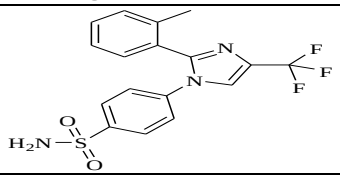
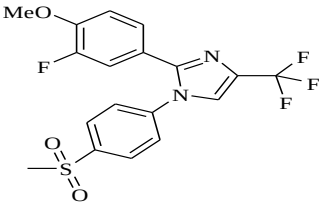
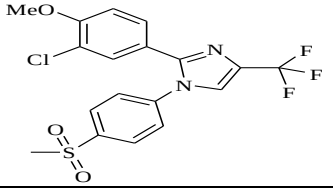
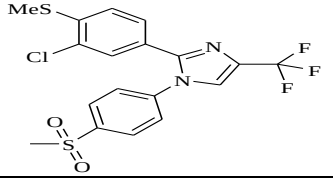
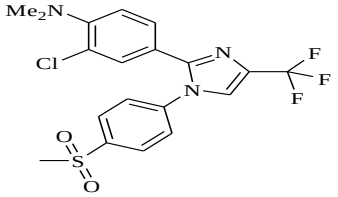
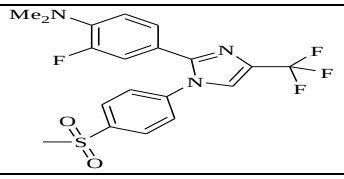
Table 1. Iupac Name and chemical structure of 83 imidazole derivatives used in present study

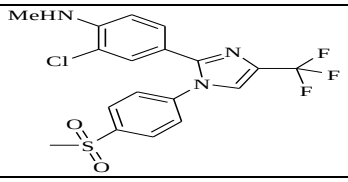
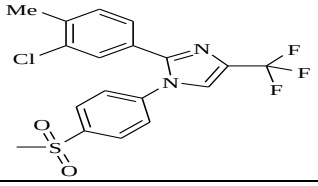
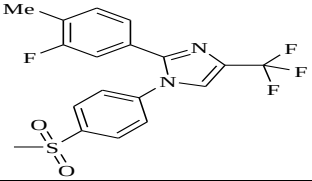
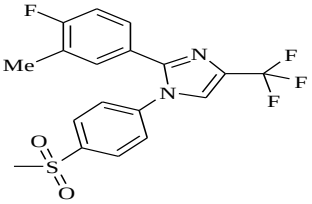
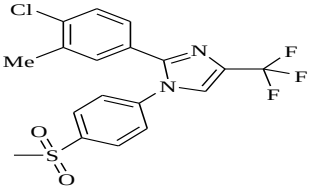
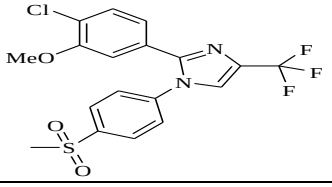
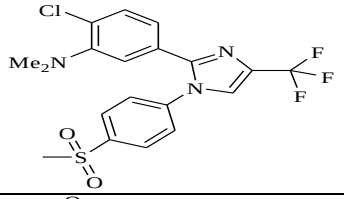
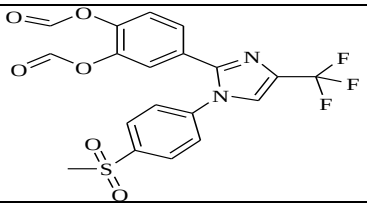
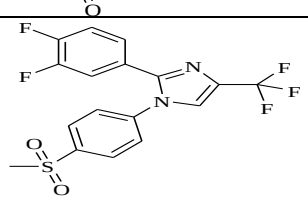
No.	IUPAC name	structure
1	2-(4-Chlorophenyl)- 4-(trifluoromethyl)-1-(4-methylsulfonylphenyl) imidazole	
2	4-(trifluoromethyl)-2-(4-fluorophenyl)-1-(4-methylsulfonylphenyl)imidazole	
3	4-(trifluoromethyl)-1-(4-methylsulfonylphenyl)-2-phenylimidazole	
4	4-(trifluoromethyl)-2-(4-methylphenyl)-1-(4-methylsulfonylphenyl)imidazole	
5	4-(trifluoromethyl)-2-(4-methoxyphenyl)-1-(4-methylsulfonylphenyl)imidazole	

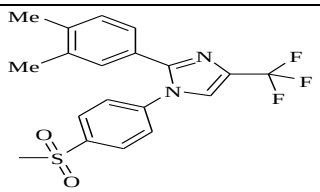
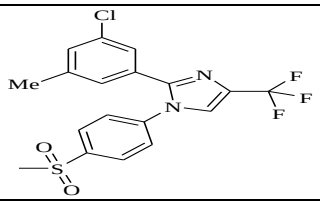
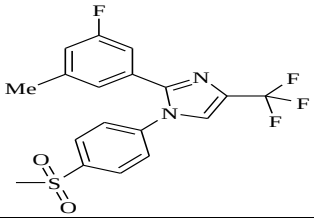
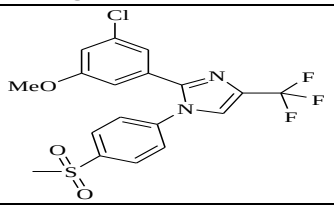
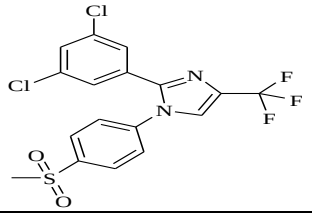
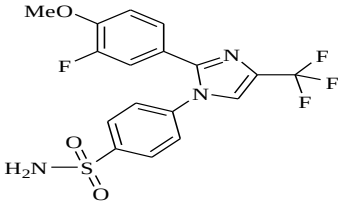
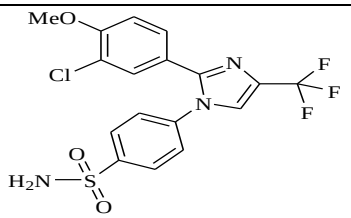
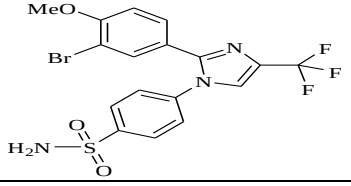
No.	IUPAC name	structure
6	4-(trifluoromethyl)-2-(4-N-methylaniline)-1-(4-methylsulfonylphenyl)imidazole	
7	4-(trifluoromethyl)-2-(4-N,N-dimethylaniline)-1-(4-methylsulfonylphenyl)imidazole	
8	4-(trifluoromethyl)-1-(4-methylsulfonylphenyl)-2-(4-methylthiophenyl)imidazole	
9	4-(trifluoromethyl)-1,2-bis (4-(methylsulfonylphenyl)imidazole	
10	4-(2-(4-chlorophenyl)-4-(trifluoromethyl)-1*H*-imidazol-1-yl)benzenesulfonamide	
11	4-(4-(trifluoromethyl)-2-(4-fluorophenyl)-1*H*-imidazol-1-yl)benzenesulfonamide	
12	4-(4-(trifluoromethyl)-2-phenyl-1*H*-imidazol-1-yl)benzenesulfonamide	
13	4-(4-(trifluoromethyl)-2-(4-methylphenyl)-1*H*-imidazol-1-yl)benzenesulfonamide	

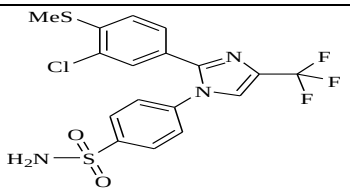
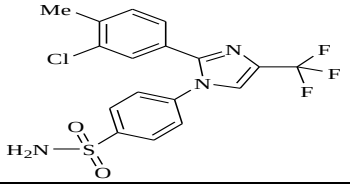
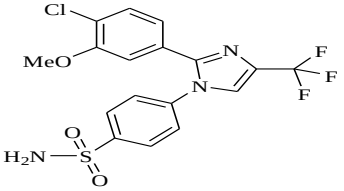
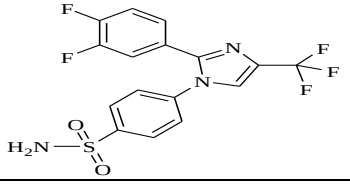
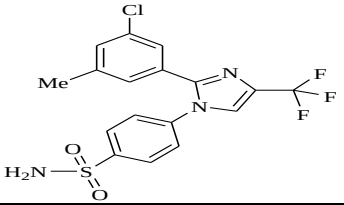
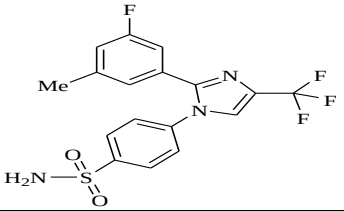
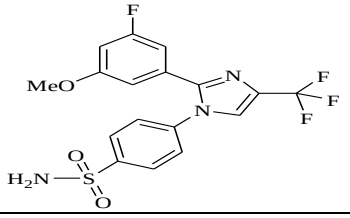
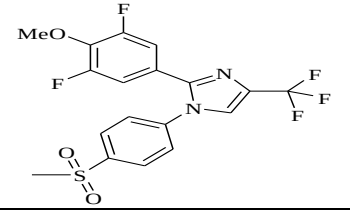
No.	IUPAC name	structure
14	2-(3-bromophenyl)-4-(trifluoromethyl)-1-(4-methylsulfonylphenyl)imidazole	
15	4-(trifluoromethyl)-2-(3-methylphenyl)-1-(4-methylsulfonylphenyl)imidazole	
16	4-(trifluoromethyl)-2-(3-trifluoromethylphenyl)-1-(4-methylsulfonylphenyl)imidazole	
17	4-(trifluoromethyl)-2-(3-methoxyphenyl)-1-(4-methylsulfonylphenyl)imidazole	
18	4-(trifluoromethyl)-1-(4-methylsulfonylphenyl)-2-(3-methylthiophenyl)imidazole	
19	4-(trifluoromethyl)-2-(3-methoxymethylphenyl)-1-(4-methylsulfonylphenyl)imidazole	
20	4-(trifluoromethyl)-2-(3-N,N-dimethylaniline)-1-(4-methylsulfonylphenyl)imidazole	
21	4-(trifluoromethyl)-2-(3-N-methylaniline)-1-(4-methylsulfonylphenyl)imidazole	

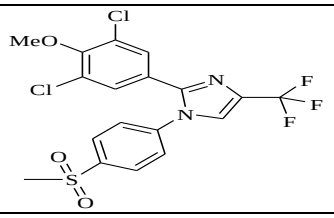
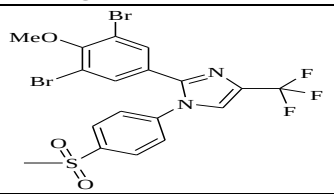
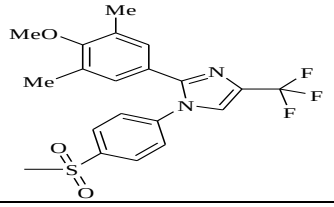
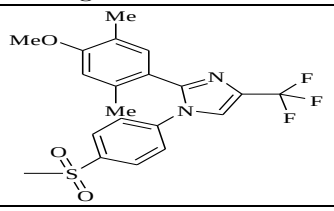
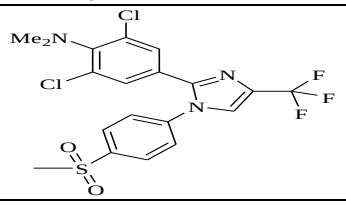
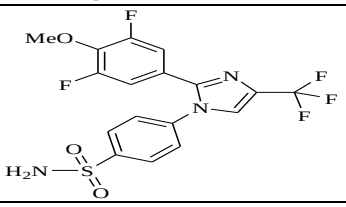
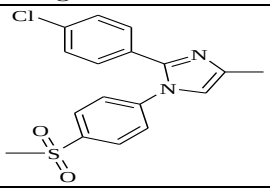
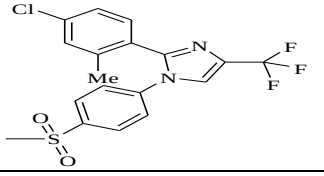
No.	IUPAC name	structure
22	2-(3-aminophenyl)-4-(trifluoromethyl)-1-(4-methylsulfonylphenyl) imidazole	
23	4-(trifluoromethyl)-1-(4-methylsulfonylphenyl)-2-(3-nitrophenyl)imidazole	
24	4-(2-(3-chlorophenyl)-4-(trifluoromethyl)-1*H*-imidazol-1-yl)benzenesulfonamide	
25	4-(4-(trifluoromethyl)-2-(3-fluorophenyl)-1*H*-imidazol-1-yl)benzenesulfonamide	
26	4-(2-(3-bromophenyl)-4-(trifluoromethyl) -1*H*-imidazol-1-yl) benzenesulfonamide	
27	4-(4-(trifluoromethyl)-2-(3-methylphenyl)-1*H*-imidazol-1-yl)benzenesulfonamide	
28	2-(2-chlorophenyl)- 4-(trifluoromethyl)-1-(4-methylsulfonylphenyl) imidazole	
29	4-(trifluoromethyl)-2-(2-fluorophenyl)-1-(4-methylsulfonylphenyl)imidazole	

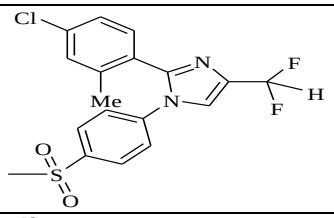
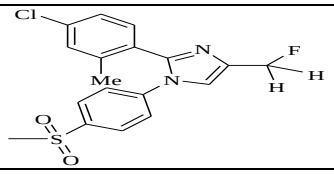
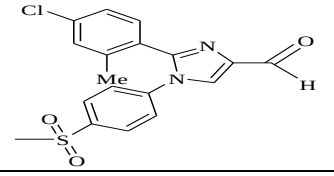
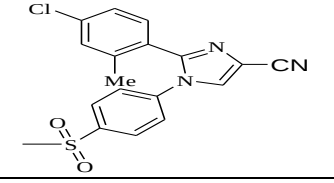
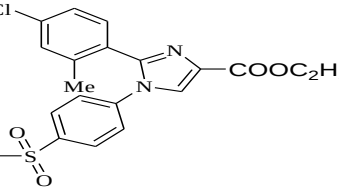
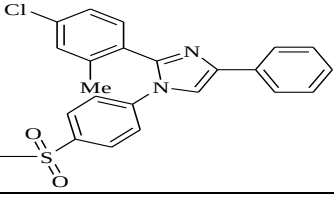
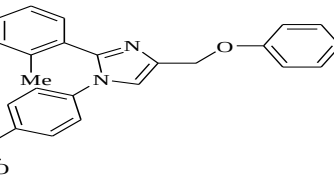
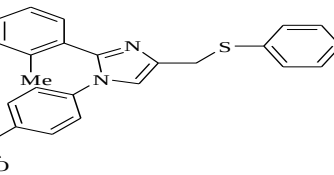
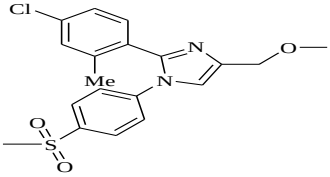
No.	IUPAC name	structure
30	4-(trifluoromethyl)-2-(2-methylphenyl)-1-(4-methylsulfonylphenyl)imidazole	
31	4-(trifluoromethyl)-2-(2-methoxyphenyl)-1-(4-methylsulfonylphenyl)imidazole	
32	4-(4-(trifluoromethyl)-2-(2-fluorophenyl)-1*H*-imidazol-1-yl)benzenesulfonamide	
33	4-(4-(trifluoromethyl)-2-(2-methylphenyl)-1*H*-imidazol-1-yl)benzenesulfonamide	
34	2-(3-fluoro-4-methoxyphenyl)- 4-(trifluoromethyl)-1-(4-methylsulfonylphenyl)imidazole	
35	2-(3-chloro-4-methoxyphenyl)- 4-(trifluoromethyl)-1-(4-methylsulfonylphenyl)imidazole	
36	2-(3-chloro-4-methylthiophenyl)- 4-(trifluoromethyl)-1-(4-methylsulfonylphenyl)imidazole	
37	2-(3-chloro-4-N,N-dimethylaniline)-49methylsulfonylphenyl)imidazole	
38	4-(trifluoromethyl)-2-(3-fluoro-4-N,N-dimethylaniline)-1-(4-methylsulfonylphenyl)imidazole	

No.	IUPAC name	structure
39	2-(3-chloro-4-N-methylaniline)-4-(trifluoromethyl)-1-(4-methylsulfonylphenyl)imidazole	
40	2-(3-chloro-4-methylphenyl)-4-(trifluoromethyl)-1-(4-methylsulfonylphenyl)imidazole	
41	4-(trifluoromethyl)-2-(3-fluoro-4-methylphenyl)-1-(4-methylsulfonylphenyl)imidazole	
42	4-(trifluoromethyl)-2-(4-fluoro-3-methylphenyl)-1-(4-methylsulfonylphenyl)imidazole	
43	2-(4-chloro-3-methylphenyl)-4-(trifluoromethyl)-1-(4-methylsulfonylphenyl)imidazole	
44	2-(4-chloro-3-methoxyphenyl)-4-(trifluoromethyl)-1-(4-methylsulfonylphenyl)imidazole	
45	2-(4-chloro-3-N,N-dimethylaniline)-4-(trifluoromethyl)-1-(4-methylsulfonylphenyl)imidazole	
46	2-formyloxy-4-[4-trifluoromethyl-1-(4-methylsulfonylphenyl)imidazol-2-yl]-phenyl formate	
47	4-(trifluoromethyl)-2-(3,4-difluorophenyl)-1-(4-methylsulfonylphenyl)imidazole	

No.	IUPAC name	structure
48	4-(trifluoromethyl)-2-(3,4-dimethylphenyl)-1-(4-methylsulfonylphenyl)imidazole	
49	2-(3-chloro-5-methylphenyl)-4-(trifluoromethyl)-1-(4-methylsulfonylphenyl)-1*H*-imidazole	
50	4-(trifluoromethyl)-2-(3-fluoro-5-methylphenyl)-1-(4-(methylsulfonyl)phenyl)imidazole	
51	2-(3-chloro-5-methoxyphenyl)-4-(trifluoromethyl)-1-(4-methylsulfonylphenyl)imidazole	
52	2-(3,5-dichlorophenyl)-4-(trifluoromethyl)-1-(4-methylsulfonylphenyl)-imidazole	
53	4-(4-(trifluoromethyl)-4-(2-(3-fluoro-4-methoxyphenyl)-1*H*-imidazol-1-yl)benzenesulfonamide	
54	4-(2-(3-chloro-4-methoxyphenyl)-4-(trifluoromethyl)-1*H*-imidazol-1-yl)benzenesulfonamide	
55	4-(2-(3-bromo-4-methoxyphenyl)-4-(trifluoromethyl)-1*H*-imidazol-1-yl)benzenesulfonamide	

No.	IUPAC name	structure
56	4-(2-(3-chloro-4-(methylthio)phenyl)-4-(trifluoromethyl)-1*H*-imidazol-1-yl)benzenesulfonamide	
57	4-(2-(3-chloro-4-methylphenyl)-4-(trifluoromethyl) -1*H*-imidazol-1-yl)benzenesulfonamide	
58	4-(2-(4-chloro-3-methoxyphenyl)-4-(trifluoromethyl)-1*H*-imidazol-1-yl)benzenesulfonamide	
59	4-(4-(trifluoromethyl)-2-(3,4-difluorophenyl)- 1*H*-imidazol-1-yl)benzenesulfonamide	
60	4-(2-(3-chloro-5-methylphenyl)-4-(trifluoromethyl)-1*H*-imidazol-1-yl)benzenesulfonamide	
61	4-(4-(trifluoromethyl)-2-(3-fluoro-5-methylphenyl)- 1*H*-imidazol-1-yl)benzenesulfonamide	
62	4-(4-(trifluoromethyl)-2-(3-fluoro-5-methoxyphenyl)- 1*H*-imidazol-1-yl)benzenesulfonamide	
63	4-(trifluoromethyl)-2-(3,5-difluoro-4-methoxyphenyl)-1-(4-methylsulfonylphenyl)imidazole	

No.	IUPAC name	structure
64	2-(3,5-dichloro-4-methoxyphenyl)- 4-(trifluoromethyl)-1-(4-methylsulfonylphenyl)imidazole	
65	2-(3,5-dibromo-4-methoxyphenyl)- 4-(trifluoromethyl)-1-(4-methylsulfonylphenyl)imidazole	
66	4-(trifluoromethyl)-2-(4-methoxy-3,5-dimethylphenyl)1-(4-methylsulfonylphenyl)imidazole	
67	4-(trifluoromethyl)-2-(4-methoxy-2,5-dimethylphenyl)-1-(4-methylsulfonylphenyl)- 1*H*-imidazole	
68	2-(3,5-dichloro-4-N,N-dimethylaniline)-4-(trifluoromethyl)-1-(4-methylsulfonylphenyl)imidazole	
69	4-(4-(trifluoromethyl)-2-(3,5-difluoro-4-methoxyphenyl)-1*H*-imidazol-1-yl) benzenesulfonamide	
70	2-(4-chlorophenyl)-4-methyl-1-(4-methylsulfonylphenyl)imidazole	
71	2-(4-chloro-2-methylphenyl)-4-(trifluoromethyl)-1-(4-methylsulfonylphenyl)imidazole	

No.	IUPAC name	structure
72	2-(4-chloro-2-methylphenyl)-4- (difluoromethyl)-1-(4-methylsulfonylphenyl)imidazole	
73	2-(4-chloro-2-methylphenyl)-4-(fluoromethyl)-1-(4-methylsulfonylphenyl)imidazole	
74	2-(4-chloro-2-methylphenyl)-1-(4-methylsulfonylphenyl)imidazole-4-carbaldehyde	
75	2-(4-chloro-2-methylphenyl)-1-(4-methylsulfonylphenyl)imidazole-4-carbonitrile	
76	Ethyl-2-(4-chloro-2-methylphenyl)-1-(4-methylsulfonylphenyl)imidazole-4-carboxylate	
77	2-(4-chloro-2-methylphenyl)-1-(4-methylsulfonylphenyl)-4-phenylimidazole	
78	2-(4-chloro-2-methylphenyl)-4-((4-chlorophenoxy) methyl)-1-(4-methylsulfonylphenyl)imidazole	
79	2-(4-chloro-2-methylphenyl)-4-((4-chlorophenylthio)methyl)-1-(4-methylsulfonylphenyl)imidazole	
80	2-(4-chloro-2-methylphenyl)-4-(methoxymethyl)-1-(4-methylsulfonylphenyl)imidazole	

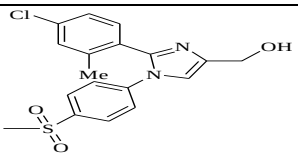
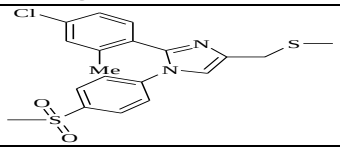
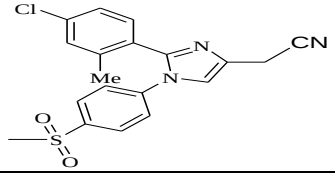
No.	IUPAC name	structure
81	(2-(4-chloro-2-methylphenyl)-1-(4-methylsulfonylphenyl)imidazol-4-yl)methanol	
82	2-(4-chloro-2-methylphenyl)-1-(4-(methylsulfonyl)phenyl)-4-(methylthiomethyl)imidazole	
83	2-(2-(4-chloro-2-methylphenyl)-1-(4-methylsulfonylphenyl)imidazol-4-yl)acetonitrile	

Table 3. The observed, predicted and residual values for training and test sets of 83 imidazole derivatives

No.	Cv J/mol.K	Predicted (Cv)	Residual (Cv)	E_{th} kJ/mol	Predicted(E_{th})	Residual(E_{th})
1	326.967	316.845	10.121	779.111	803.587	-24.476
2	370.56	356.655	13.905	906.271	893.615	12.656
3	307.286	345.049	-37.763	865.691	866.951	-1.261
4	328.666	330.284	-1.6180	776.061	787.808	-11.747
5	331.942	333.202	-1.260	784.374	784.933	-0.559
6	349.393	346.810	2.582	1182.465	1138.485	43.980
7	427.559	417.527	10.032	807.96	802.082	5.878
8	333.18	340.203	-7.023	880.083	900.121	-20.038
9	340.954	324.195	16.758	877.494	876.340	1.154
10	364.564	347.108	17.455	1174.507	1106.046	68.461
11	435.764	429.896	5.868	912.79	896.991	15.799
12	341.749	340.659	1.090	807.499	806.992	0.507
13	333.532	347.607	-14.075	873.485	898.482	-24.997
14	369.192	359.440	9.751	1074.372	1016.990	57.382
15	364.577	374.247	-9.670	929.99	934.266	-4.276
16	354.167	356.603	-2.436	811.316	810.102	1.214
17	328.624	329.416	-0.792	890.1	880.403	9.696
18	345.368	366.340	-20.972	936.065	942.840	-6.775
19	346.577	354.792	-8.215	912.773	915.413	-2.640
20	360.251	359.217	1.0337	965.081	944.061	21.019
21	341.552	329.985	11.566	865.36	835.482	29.877
22	354.142	350.483	3.659	1037.933	1021.759	16.173
23	378.455	398.342	-19.887	1048.858	998.295	50.563
24	354.142	354.107	0.0354	930.082	910.064	20.018
25	355.381	353.119	2.262	857.641	876.541	-18.900
26	350.561	348.006	2.554	950.174	954.606	-4.432

No.	Cv J/mol.K	Predicted (Cv)	Residual (Cv)	E_{th} kJ/mol	Predicted(E_{th})	Residual(E_{th})
27	363.23	383.111	-19.882	921.262	917.981	3.281
28	326.766	328.248	-1.4820	779.195	809.615	-30.420
29	369.853	374.133	-4.280	872.883	898.659	-25.776
30	393.083	405.640	-12.557	946.638	947.146	-0.508
31	339.657	330.833	8.824	880.623	887.888	-7.265
32	338.778	338.838	-0.060	758.952	669.534	89.418
33	332.335	342.165	-9.830	775.174	802.642	-27.468
34	366.05	376.160	-10.110	909.903	912.655	-2.752
35	393.083	405.640	-12.557	946.638	947.146	-0.508
36	327.812	332.455	-4.643	778.743	811.742	-33.000
37	369.05	373.342	-4.292	906.882	907.798	-0.916
38	352.72	344.612	8.108	859.912	891.451	-31.539
39	316.356	321.952	-5.596	798.629	795.899	2.7298
40	379.25	379.768	-0.518	897.623	865.250	32.373
41	364.309	362.3605	1.948	877.402	877.836	-0.434
42	340.988	350.661	-9.673	880.021	884.520	-4.499
43	395.083	393.244	1.839	1023.729	1021.796	1.933
44	378.819	379.785	-0.966	888.874	926.567	-37.693
45	333.285	332.202	1.083	808.048	816.442	-8.394
46	391.049	391.376	-0.327	1028.239	1055.154	-26.915
47	386.865	379.171	7.694	811.319	826.868	-15.549
48	328.816	333.113	-4.297	942.702	953.818	-11.116
49	375.744	372.112	3.632	882.15	882.022	0.128
50	388.342	390.70	-2.361	882.15	876.411	5.7389
51	333.875	335.473	-1.598	808.232	815.924	-7.692
52	357.326	356.869	0.456	889.803	908.660	-18.857
53	403.471	387.662	15.809	1092.907	1118.080	-25.173
54	333.49	341.933	-8.443	910.320	910.997	-0.677
55	353.113	360.080	-6.967	893.267	894.639	-1.372
56	402.852	389.319	13.533	1093.400	1137.404	-44.004
57	367.067	336.635	30.432	893.481	887.028	6.453
58	352.908	351.750	1.1581	999.430	999.718	-0.288
59	412.417	400.954	11.463	855.930	874.968	-19.038
60	354.414	352.819	1.595	929.543	919.283	100.26
61	357.209	358.591	-1.382	889.715	903.164	-13.449
62	330.549	333.633	-3.084	865.691	867.167	-1.476
63	363.121	349.867	13.254	866.833	888.803	-21.970
64	371.025	377.306	-6.281	905.451	920.739	-15.288
65	307.286	334.525	-27.239	865.691	786.789	78.902
66	373.916	364.634	9.281	1013.114	1016.903	-3.789

No.	Cv J/mol.K	Predicted (Cv)	Residual (Cv)	E_{th} kJ/mol	Predicted(E_{th})	Residual(E_{th})
67	395.150	381.801	13.349	1032.105	1027.559	4.545
68	357.929	356.232	1.697	889.205	891.939	-2.734
69	378.74	369.747	8.993	1048.314	987.835	60.478
70	395.522	377.426	18.096	929.471	951.116	-21.645
71	347.418	352.041	-4.623	910.765	939.315	-28.550
72	361.272	360.912	0.360	964.358	952.394	11.964
73	340.603	343.250	-2.647	761.186	764.123	-2.937
74	338.26	345.017	-6.757	930.053	924.736	5.317
75	343.728	340.567	3.161	879.991	880.264	-0.273
76	365.226	363.996	1.230	994.679	995.240	-0.561
77	334.812	340.945	-6.133	896.866	895.166	1.700
78	345.368	334.513	10.855	842.745	836.988	5.757
79	358.037	359.656	-1.619	889.305	881.713	7.592
80	330.549	333.633	-3.084	866.833	888.803	-21.9705
81	331.227	327.348	3.879	775.877	787.041	-11.1643
82	345.368	351.719	-6.351	890.100	904.863	-14.763
83	387.802	374.858	12.944	1081.171	1101.763	-20.592

RESULTS AND DISCUSSION

QSPR models and statistical coefficients

In this study QSPR models for predicting the thermodynamic properties of 83 imidazole derivatives were presented, using the GA-MLR method and molecular descriptors.

The thermal energy (E_{th}) and heat capacity (Cv) was used as dependent variables and the molecular descriptors were used as independent variables.

The best multiple linear regression models were selected by using statistical factors such as correlation coefficient (R), squared multiple correlation coefficient (R^2), adjusted correlation coefficient (R^2_{adj}), root mean square error (RMSE), Fisher ratio (F), Durbin-Watson value (D) and significance (Sig)) [21-23].

QSPR models for the thermal energy (E_{th} / kJ mol^{-1})

Several linear QSPR models for predicting the thermal energy including five-eleven descriptors are established. Table **3** shows the statistical parameters of models for the entropy of pyridines. The value of R^2, the value of RMSE, the value of Fisher function (F) and a small number of descriptors are good measures of the quality of QSPR models. The best linear model for the thermal energy includes five molecular descriptors namely: Ss, Mor01e, Mor14v, Se and Har. The regressions parameters of the best model consist of five molecular descriptors are presented in Equation (1):

$$E_{th}= 620.275-10.46\ (Har) + 12.092\ (mor014v) +0.650\ (mor01e) -1.742\ (Ss) + 8.854(se) \quad (1)$$

N=61, R=0.998, R^2=0.995, R^2_{adj}=0.995, RMSE= 2.64, DW=2.066, F=2.342×10^3

QSPR models for the heat capacity (Cv)

The best linear model for the heat capacity contains six descriptors, namely, Mor16p, Mor14v, Vm, HATS7v, QW and GMTI descriptors (see Table **4**). The model and its statistical coefficients are presented below:

Cv=313.583+0.329 (Qw) -0.056(GMTIv) + 21.555(Mor14v)16.285 (Mor16p)+0.351(Vm) - 214.280 (HATS7v) (2)

N=61, R=0.984, R^2=0.968, R^2_{adj}=0.965, RMSE= 2.12, F=273.825, DW=2.269, Sig=0.000

We studied the relationship between the molecular descriptors and the thermodynamic properties of 85 imidazole derivations. In this study, to find the best model for predicting the mentioned properties, we will use the following sections.

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 a correlation between the independent variables or should not have happened multicollinearity.

Test multicollinearity is as a basis the variance inflation factor (VIF) value of multicollinearity test results using SPSS [24- 27]. The VIF and Pearson correlation coefficient (PCC) was detected for the selected descriptors. If the VIF value lies between 1 and 10, there is no multicollinearity; if VIF<1 or >10, there is multicollinearity [28]. The VIF values calculated as follows:

$$\mathrm{VIF} = \frac{\mathbf{1}}{\mathbf{1-R^2}} \quad (3)$$

In Equation (3), R^2 is the square correlation coefficient of multiple regressions between variables in the studied model.

Table 3. Statistical parameters of the models calculated with the SPSS software for E_{th} (kJ.mol^{-1})

Model	Independent Variable	R	R^2	R^2_{adj}	RMSE	F	Sig
1	MW, Mor28p, Ss, H4p, Rww, Mor01e, QW, Mor14v, Se, D, Har	0.998	0.996	0.995	2.61	1.110×10^3	0.000
2	MW, Mor28p, Ss, H4p, Mor01e, QW, Mor14v, Se, D, Har	0.998	0.996	0.995	2.62	1.212×10^3	0.000
3	MW, Mor28p, Ss, H4p, Mor01e, QW, Mor14v, Se, Har	0.998	0.996	0.995	2.61	1.360×10^3	0.000
4	MW, Mor28p, Ss, H4p, Mor01e, Mor14v, Se, Har	0.998	0.996	0.995	2.61	1.527×10^3	0.000
5	MW, Ss, H4p, Mor01e, Mor14v, Se, Har	0.998	0.996	0.995	2.64	1.686×10^3	0.000
6	MW, Ss, Mor01e, Mor14v, Se, Har	0.998	0.995	0.995	2.64	1.952×10^3	0.000
7	**Ss**(struct usb_descriptor), **Mor01e** (signal 01 / weighted by Sanderson electronegativity), **Mor14v (**signal 14 / weighted by van der Waals volume), **Se (**sum of atomic Sanderson electronegativities), **Har**(Wi_H22D matrix-based descriptors Reciprocal squared)	0.998	0.995	0.995	2.64	2.342×10^3	0.000

Table 4. Statistical parameters of the models calculated with the SPSS software for Cv (J.mol^{-1}.K^{-1})

Model	Independent Variable	R	R^2	R^2_{adj}	RMSE	F	Sig
1	Mor16p, RDF095e, RDF100e, RDF070u, Mor14v, Vm, H5m, HATS7v, RDF095u, RDF070e, QW, GMTI	0.986	0.972	0.965	2.13	137.111	0.000
2	Mor16p, RDF095e, RDF100e, RDF070u, Mor14v, Vm, H5m, HATS7v, RDF070e, QW, GMTI	0.965	0.971	0.965	2.13	150.012	0.000
3	Mor16p, RDF100e, RDF070u, Mor14v, Vm, H5m, HATS7v, RDF070e, QW, GMTI	0.965	0.971	0.965	2.12	165.368	0.000
4	Mor16p, RDF100e, RDF070u, Mor14v, Vm, HATS7v, RDF070e, QW, GMTI	0.965	0.970	0.965	2.12	184.568	0.000
5	Mor16p, RDF100e, Mor14v, Vm, HATS7v, RDF070e, QW, GMTI	0.965	0.970	0.965	2.12	207.445	0.000
6	Mor16p, RDF100e, Mor14v, Vm, HATS7v, QW, GMTI	0.965	0.969	0.965	2.12	238.508	0.000
7	**Mor16p** (signal 16 / weighted by polarizability), **Mor14v (**signal 14 / weighted by van der Waals volume), **Vm**(V total size index / weighted by mass), **HATS7v** (leverage-weighted autocorrelation of lag 7 / weighted by van der Waals volume), **QW** QW_L 2D matrix-based descriptors Laplace matrix (L)), **GMTI** (Gutman MTI by valence vertex degrees)	0.965	0.968	0.995	2.12	273.825	0.000

VIF study of models

To study the validation of linearity between the molecular descriptors in built models, we used SPSS program to obtain the PCC, and collinearity statistics.

The suitable linear model for QSPR study of the thermal energy (Equation**1**) includes six molecular descriptors (Ss, Mor01e, Mor14v, Se and Har). The results of the correlation between these descriptors are listed in Table 5.

These results can be expressed the Pearson correlation between Se and Mor01e is close to unity, and VIF for some descriptors such as Mor01e, and Har are bigger than 10, therefore there is linearity between these descriptors. After removing Se from this model, and then in the next step Har and Ss from this model, we corrected Equation 1 as follows:

E_{th}= 503.907 + 177.283 (Mor14v) + 0.577 (Mor01e) (4)

N=61, R=0.995, R^2=0.913, R^2_{adj}=0.910, DW=2.01, RMSE=5.42, F=303.002, Sig=0.000

In Equation (**2)**, the Pearson correlation between Qw and GMTIv descriptors is close to unit. Also, the VIF value for these descriptors are bigger than 10(see Table **6**), therefore there is linearity between these descriptors. After removing GMTIv and then in the next step Qw from model, we obtained the best model of three descriptors. The regression parameters of this model are provided in Equation (**5**).

Cv=33.053-18.802(Mor16p) + 0.537(Vm) - 344.665(HATS7v) (5)

N=61, R=0.918, R^2=0.844, R^2_{adj}=0.835, RMSE= 3.13, F=102.407, DW=1.821, Sig=0.000

Validation

The success of any QSAR/ QSPR models depends on the accuracy of the input data, selection of appropriate descriptors, statistical tools and validation of the developed model. In this section, for verification, the validity of the regression models and the predictive ability and statistical significance of the QSPR models, squared cross-validation coefficient for leave-one-out (Q^2LOO) was used [29-32]. The Q^2LOO value (Eq. 6) computed from 20 % of randomly chosen data was found to be positive and smaller than one.

Table 5. Correlation between the molecular descriptors (Eq. 1)

Pearson Correlation for E_{th}						**Collinearity Statistical**			**Corrected model**
Descriptor	Ss	Mor01e	Mor14v	Se	Har	VIF(1)	VIF(2)	VIF(3)	VIF(4)
Ss	1.000	-0.135	0.055	0.303	-0.603	5.100	4.632	2.623	----
Mor01e		1.000	0.23	-0.944	0.545	60.441	2.925	1.267	1.126
Mor14v			1.000	-0.377	0.554	3.409	6.613	2.340	1.126
Se				1.000	-0.776	119.284	----	----	----
Har					1.000	37.082	14.748	----	----

Table 6. Correlation between the molecular descriptors (Eq. 2)

Pearson Correlation for Cv							**Collinearity Statistical**		**Corrected model**
Descriptor	Mor16p	Mor14v	Vm	HATs7v	Qw	GMTIv	VIF(1)	VIF(2)	VIF(3)
Mor16p	1.000	-0.467	0.231	-0.366	0.267	0.267	1.660	2.585	1.518
Mor14v		1.000	-0.485	0.303	-0.340	0.405	3.093	1.531	----
Vm			1.000	0.146	-0.587	-0.658	6.011	3.406	1.219
HATs7v				1.000	0.367	-0.341	1.822	1.610	1.403
Qw					1.000	-0.994	421.351	5.167	----
GMTIv						1.000	487.873	----	----

$$Q^2 = 1 - \frac{\sum(Y_i - \hat{Y}_{i|i})^2}{\sum(Y_i - \bar{Y})^2} = 1 - \frac{PRESS}{TSS} \quad Q^2 \leq 1 \qquad (6)$$

In Equation (6), the notation i|i indicates that the quantity is predicted by a model estimated when the i-th sample was left out from the training set.

The Q^2_{LOO} values of the thermal energy and heat capacity (Eqs. 4 and 5) of the imidazole derivatives were calculated 0.923and 0.825 respectively.

In purpose to create and test model(external validation), which was used data set of 83compounds and randomly separated into a training set of 61 compounds (75%), that was applied to test the made model a prediction set of 22 compounds (25%).

Statistical factors such as correlation coefficient (R), squared multiple correlation coefficient (R^2), adjusted correlation coefficient (R^2_{adj}), Fisher ratio (F) and root mean square error (RMSE), of these models for training and external validation sets which that was used to create and test model are listed in Table 7. These parameters show potential of the GA-MLR models for prediction of E_{th} and Cv values of imidazole derivatives and this method could simulate the complicated linear relationship between the studied properties and molecular descriptors.

The Durbin-Watson statistics

In this study, we also focus on the Durbin-Watson statistics (DW) and unstandardized predicted and residual values. If the DW value lies between 2 and 0, then there is non-autocorrelation. The value of Durbin-Watson statistics in our models is close to 2 (see Eqs.4 and 5) and hence the errors are uncorrelated [33-35].

The residual is the difference between the observed value and the predicted value. The residual values of the thermal energy and heat capacity expressed by Equations 4 and 5 are shown in Table 2.These values show a relatively random pattern (see Figs.1 and 2). This relatively random pattern shows that a linear model provides a decent fit to the data.

Figures 3, 4 present the linear correlation between the observed and predicted values of mentioned properties that were obtained using Equations 3 and 4.

Interpretation of the best descriptors in QSPR models

The aforementioned results and discussion lead us to conclude that the three descriptors have been classified into WHIM (Weighted Holistic Invariant Molecular descriptors), 3D-MoRSE (3D-Molecule Representation of Structures based on Electron diffraction descriptors), and GETWAY (Geometric, Topological and Atomic Weighted Assembly descriptors) indices can be used successfully for modeling and predicting the heat capacity (Cv /J $mol^{-1}K^{-1}$) of the studied imidazole derivatives (see Table **8**).

WHIM descriptors are 3D structural descriptors obtained from the (*x*,*y*,*z*)-atomic coordinates of a molecular conformation of a chemical, and are used successfully in QSAR/QSPR modelling. They are built in so as to capture relevant 3D information regarding different features of molecular structure: size, shape, symmetry and atom distribution. Different weights are used to obtain particular information for each set of descriptors[36-38].

GETAWAY descriptor calculated from the leverage matrix obtained by the centered atomic coordinates [39].This descriptor could be used for satisfactory prediction of the thermal energy.

3D-MoRSE is a very flexible 3D structure encoding framework for chemoinformatics and QSAR/QSPR

purposes due to the range of scattering parameter values and variety of weighting schemes used [40-43].

The thermal energy can be better modeled using a combination of the two descriptors (Mor14v, and Mor01e). These descriptors have been classified into 3D-MoRSE descriptor and weighted by atomic van der Waals volume and Sanderson electronegativity.

Table 7. Statistical parameters obtained by the GA- MLR model for the thermal energy and heat capacity for training and test sets (Eqs. 4, 5)

Data set	Property	N	R	R^2	R^2_{adj}	RMSE	DW	F	sig
training	E_{th}	61	0.995	0.913	0.910	5.42	2.041	303.002	0.000
test	E_{th}	22	0.910	0.828	0.810	5.34	1.966	45.789	0.000
training	Cv	61	0.918	0.844	0.835	3.13	1.821	102.407	0.000
test	Cv	22	0.902	0.813	0.769	4.041	2.008	18.459	0.000

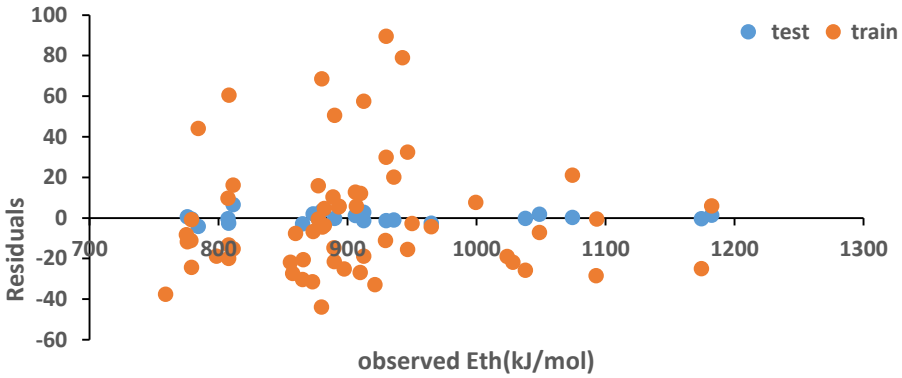


Fig. 1. Plot of residuals against the observed values of the thermal energy (E_{th}) of 83 imidazole derivatives for training and test sets using GA-MLR method.

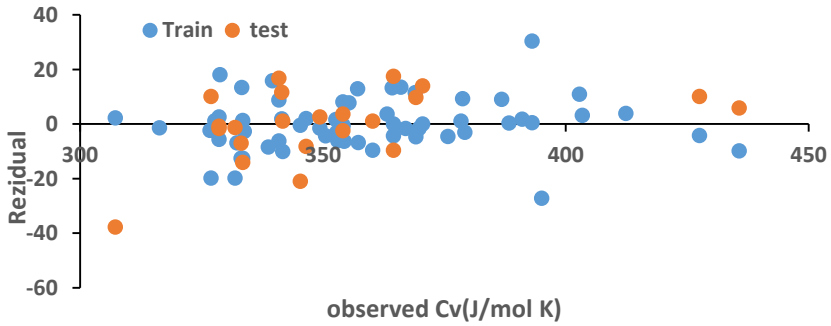


Fig. 2. Plot of residuals against the observed values of the heat capacity (Cv) of 83 imidazole derivatives for training and test sets using GA-MLR method.

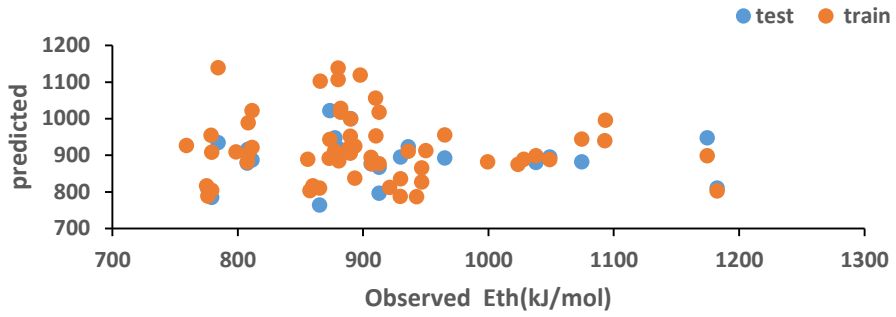


Fig. 3. Comparison between predicted and observed values of the thermal energy calculated by the GA-MLR method.

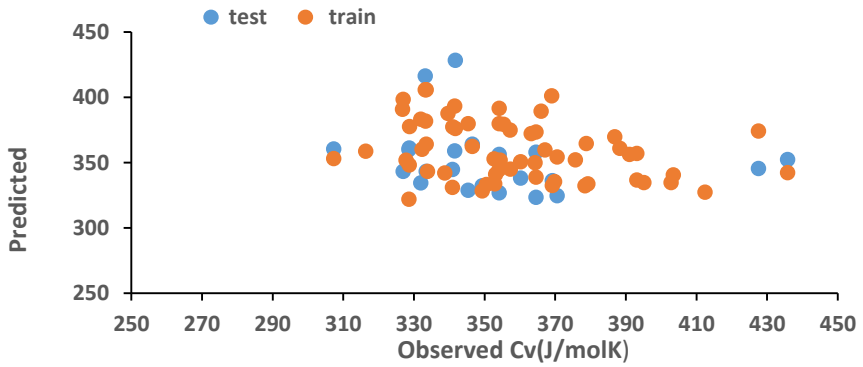


Fig. 4. Comparison between predicted and observed values of the heat capacity calculated by the GA-MLR method.

Table 8. List of the best selected molecular descriptors that appear in the final models

Property	Name descriptor	description	Block descriptor
Cv	Vm	V total size index / weighted by mass	WHIM
	HATS7v	leverage-weighted autocorrelation of lag 7 / weighted by van der Waals volume	GETAWAY
E_{th}	Mor16p	signal 16 / weighted by polarizability	3D-MoRSE
	Mor14v	signal 14 / weighted by van der Waals volume	3D-MoRSE
	Mor01e	01 / weighted by Sanderson electronegativity	3D-MoRSE

CONCLUSION

QSPR studies are mathematical correlations between molecular property and molecular descriptors. In this study, QSPR models were developed to predict the thermal energy (E_{th}/ kJ mol^{-1}) and heat capacity (Cv/ J $mol^{-1}K^{-1}$) of some imidazole derivatives. The genetic

algorithm (GA) technique and multiple linear regression (MLR) method were applied for modeling and predicting properties which are used in present study. Molecular descriptors calculated by the DRAGON software. The suitable descriptors were selected with the aid of the genetic algorithm (GA) technique and SPSS program. To evaluate the robustness and predictive ability of the constructed models, leave-one-out cross-validation and internal and external validation methods were implemented.

The results and discussion lead us to conclude that combining the three descriptors (Vm, Mor16p and HATS7v) can be used successfully for modeling and predicting the heat capacity, and the thermal energy can be better modeled using two descriptors (Mor14v, and Mor01e). These descriptors were classified in 3D-MoRSE, WHIM and GETAWAY indices.

REFERENCES

[1] S. Mghazli, A. Jaouad, M. Mansour, D. Villemin and D. Cherqaoui, Chemosph. 43(2001) 385.

[2] P. O. Miranda, L. L. Gundersen, Arch. Pharm. Chem. Life Sci. 343(2010) 40.

[3] T. H. Lean and D. Vengadasalam, Br. J. Vener. Dis.49 (1973) 69.

[4] R. Palin, J. K. Clark, L. Evans, A.K. Houghton, P.S. Jones and A. Prosser, Bioorg. Med. Chem. 16 (2008) 2829.

[5] B. G. Wachall, M. Hector, Y. Zhuang and R. W. Hartmann, Biorg. Med. Chem.7 (1999) 1913.

[6] S. E.Yoo, S. K. Kim, S. H. Lee, K. Y. Yi and D. W. Lee, Bioorg. Med. Chem. 7 (1999) 2971.

[7] F. Hadizadeh, H. Hosseinzadeh, V. S. M. Shariaty, M. Seifi and S. Kazemi, Iranian J. of Pharm. Res. 7 (2008) 29.

[8] M. Mor, F. Bordi, C. Silva, S. Rivara, V. Zuliani, F. Vacondio, M. Rivara, E. Barocelli, S.Bertoni, V. Ballabeni and F. Magnanini, Biorg. Med. Chem.12 (2004) 663.

[9] D. P. Clough, R. Hatton, S. J. Pettinger, G. M. Samuels and A. Shaw, Brit. J. Pharmacol. 62 (1978) 385P.

[10] M. N. Rahman, J. Z. Vlahakis, D. Vukomanovic, W. A. Szarek, K. Nakatsu and Z. Jia , J. Med. Chem. 2009; 52: 4946–4950.

[11] G. Roman, M. N. Rahman, D. Vukomanovic, Z. Jia, K. Nakatsu and W. Szarek, Chem. Biol. Drug. Des. 75 (2010) 68.

[12] B. Gómez, N. Likhanova, M. Dominguez Aguilar, O. Olivares, J. Hallen and J. Martínez-Magadán, J. Phys. Chem. A. 109 (2005) 8950.

[13] R. Kanojia and G. Singh, Surf. Eng. 21 (2005) 180.

[14] A. K. Srivastava and N. Shukla, J. Saudi. Chem. Soc. 17 (2013) 321.

[15] G. Darshana, and M. Parul, Current Research in Pharmaceutical Sciences. 10 (2020)19.

[16] O. H. A. Al-Attraqchi, K. N. Venugopala, Curr. Comput. Aided. Drug .Des. 16 (2020) 682.

[17] A. Shayanfar, S. Ghasemi, S. Soltani, K. Asadpour-Zeynali, R. J. Doerksen and A. Jouyban, Med. Chem. 9 (2013) 434.

[18] A. Khalafi-Nezhad, M. N. Soltani Rad, H. Mohabatkar, Z. Asrari and B. Hemmateenejad, Bioorg. Med. Chem. 13 (2005) 1931.

[19] J. H. Zhao, Y. S. Zhao, H. Z. Zhang and X. P. Zhang, Chinese. 34 (2013) 1882.

[20] DRAGON Software for calculation of topological indices. <http:// www.disatunimib.it>

[21] M. V. Diudea, QSPR/QSAR studies for molecular descriptors; Ed Nova Science Hunting don: New York, 2000.

[22] B. L. Podlgar and D. M. Ferguson, Drug Des. Discov. 17 (2000) 4.
[23] F. Hadizadeh, S. Vahdani and M. Jafarpour, Iran J. Basic. Med. Sci. 16 (2013) 910.
[24] E. Pourbasheer, S. Ahmadpour, R. Zare-Dorabei and M. M. Nekoei, Arab. J. Chem. 10 (2017) 33.
[25] C. B. García, J. García, M. M. López Martín and R. Salmerón, J. Appl. Stat. 42 (2015) 648.
[26] [26] D.J. Dupuis and M. P. Victoria-Feser, Ann. Appl. Stat. 7 (2013) 319.
[27] S. Velilla, Am .Stat. 1 (2018**)** 1.
[28] M. Golberg and H. A. Cho, Introduction to regression analysis, WIT press, 2004.
[29] S. Ajmani, K. Jadhav and S. A. Kulkarni, J. Chem. Inf. Model. 46 (2006) 24.
[30] M. Zhao and D. Wei, Exploring the ligand-protein networks in traditional chinese medicine: current databases, methods and applications. In Advance in Structural Bioinformatics, Springer, Dordrecht, 2015.
[31] R. D. Cramer III, J. D. Bunce and D.E. Patterson, Quant. Struct. Act. Relat. 7(1988) 18.
[32] D. G. Kleinbaum, Applied regression analysis and other multivariable methods, Australia; Belmont, CA: Brooks/Cole, 2008, 906.
[33] H. Hisamatsu and K. Maekawa, J. Econom. 61 (1994) 367.
[34] S. Chatterjee and J. S. Simonoff, Handbook of regression analysis, John Wiley & Sons, 2013**.**
[35] N. R. Hateka, Tests for Detecting Autocorrelation. Principles of Econometrics: An Introduction (Using R). SAGE Publications, 2010, 379–382.
[36] V. Consonni, D. Ballabio and R. Todeschini, J. Chemom. 24 (2010) 194.
[37] R. Todeschini, M. Lasagni and E. Marengo, J. Chemom. 8 (1994) 263.
[38] R. Todeschini and P. Gramatica, Quant. Struct.–Act. Relat. 16 (1997) 113.
[39] A. O. Aptula, N. G. Jeliazkova, T. W. Schultz and M. T. D. Cronin, QSAR. Comb. Sci. 24 (2005) 385.
[40] P. Gramatica, QSAR. Comb. Sci. 4 (2006) 327.
[41] O. Devinyak, D. Havrylyuk and R. Lesyk, J. Mol. Graph. Model. 54 (2014) 194.
[42] J. H. Schuur, P. Selzer and J. Gasteiger, J. Chem. Inf. Comput. Sci. 36 (1996**)** 334.
[43] L. Saíz-Urra, Y. Pérez-Castillo, M. Pérez González, R. Molina Ruiz, M. Cordeiro, J.E.
[44] Rodríguez-Borges and X. García-Mera, QSAR & Comb. Sci. 28 (2009**)** 98.

مجله شیمی فیزیک و شیمی نظری
دانشگاه آزاد اسلامی واحد علوم و تحقیقات
جلد ۱۸، شماره ۲، تابستان ۱۴۰۰
ISSN ۱۷۳۵-۲۱۲۶

مدل‌های ارتباط کمی ساختار- خاصیت برای پیش‌بینی خواص ترمودینامیکی مشتقات ایمیدازول با استفاده از توصیف کننده‌های مولکولی و الگوریتم ژنتیک- رگرسیون چند متغیره خطی

شیوا مشیدی، فاطمه شفیعی*، طاهره مومنی اصفهانی

گروه شیمی، واحد اراک، دانشگاه آزاد اسلامی، اراک، ایران

چکیده

ایمیدازول ترکیبی با طیف گسترده ای از فعالیت های بیولوژیکی است و مشتقات آن اساس چندین گروه از داروها می‌باشند. در این مطالعه رابطه بین توصیف کننده‌های مولکولی و انرژی گرمایی (E_{th} kJ/mol) و ظرفیت گرمایی (Cv J/mol) مشتقات ایمیدازول مورد بررسی قرار گرفته است. ساختارهای شیمیایی ۸۵ مشتق ایمیدازول در سطح هارتری فاک (HF) و مجموعه پایه 6-311G* با نرم افزار گوسین۹۸ بهینه سازی شده است. توصیف کننده های مولکولی برای ترکیبات منتخب با استفاده از نرم افزار دراگون محاسبه شد. روش‌های الگوریتم ژنتیک- رگرسیون خطی چندمتغیره (GA-MLR) و برگشتی برای انتخاب توصیف کننده های مناسب و همچنین پیش‌بینی خواص ترمودینامیکی مشتقات ایمیدازول استفاده گردید. مدلهای بدست آمده با استفاده از ضرایب آماری مانند ضریب همبستگی تعدیل شده (R^2adj)، ضریب فیشر (F)، جذر میانگین مربع خطا (RMSE)، آماره دوربین واتسون (D) و سطح معناداری (Sig) مورد ارزیابی قرار گرفت. قدرت پیش بینی مدل‌های GA-MLR با استفاده از اعتبار سنجی متقابل یکی بیرون (LOO) واعتبار سنجی خارجی مجموعه آزمون مورد مطالعه قرار گرفت. توانایی پیش‌بینی مدل‌های GA-MLR با دو- سه توصیف کننده مولکولی انتخاب شده رضایت بخش بود. از مدل‌های توسعه یافته QSPR می‌توان برای پیش بینی خصوصیات ترکیباتی که هنوز سنتز نشده اند استفاده کرد..

کلید واژه‌ها: ارتباط کمی ساختار- خاصیت؛ مشتقات ایمیدازول؛ اعتبارسنجی متقابل یکی بیرون؛ الگوریتم ژنتیک - رگرسیون چند متغیره خطی

* مسئول مکاتبات: f-shafiei@iau-arak.ac.ir