Turkish Computational and Theoretical Chemistry

Turkish Comp Theo Chem (TC&TC)

Volume(Issue): 6(1) – Year: 2022 – Pages: 1-8

e-ISSN: 2602-3237

https://doi.org/10.33435/tcandtc.894168

Received: 10.03.2021 ***Accepted***: 10.02.2022 *Research Article*

A Developed QSPR Model for the Melting Points of Isatin Derivatives

Rabah Ali Khalil [1]**; Shayma'a Hashim Abdulrahman**
Department of Chemistry, College of Science, University of Mosul, Mosul, Iraq.

Abstract: This paper suggests a developed quantitative structure property relationship (QSPR) model for coping the melting point (M.P) which is considered as the main and important physical property of solid state. The development was based on the decreasing in number of descriptors in order to be statistically intensive with excellent values of statistical parameters. The model was applied successfully to the already published data of M.P for 32 biologically active molecules derived from 4-(1-aryl-2-oxo-1,2-dihydro-indol-3-ylideneamino)-N-substituted benzene sulfonamides. The calculations of descriptors were carried out using density functional theory (DFT) with bases set of 6-311G (d, P). A statistically intensive QSPR model contains only three descriptors with physical meaning has been introduced. Two of them are belonging to the direct theoretical calculations but the third was considered as three dimensional correcting term of which depending on the chemical structure of the substituent. The theoretically calculated descriptors were the total connectivity (TC) and the average charge on the aryl group (AQArH) as both depending on the packing of molecules and responsible on M.P. The last descriptor was suggested as a correction term with respect to the packing of molecules of which depending on their three dimensional chemical structure which only taking the values of -1, 0 and 1. A relatively excellent statistical parameters for the developed model were obtained with square regression coefficient (r^2), cross-validation (q^2) and root mean squared error (RMSE) are equal to 0.925, 0.903 and 15.26°C, respectively. It was concluded that the developed model gives more confidence results in addition to physical significance which can be considered as a helpful tool for understanding the factors affecting the melting point.

Keywords: *QSPR, computational chemistry, isatin derivatives, melting point, DFT*

1. Introduction

There is no doubt in the existence of essential needs for changing from experimental methods to that of theoretical using computer simulations. Such a movement requires a crucial development in simulations methods such as quantitative structure activity relationship (QSAR) and quantitative structure property relationship (QSPR). It is not surprising to state that the employment of QSAR as *in silico* method in some important aspects such as ecology and toxicology will be more popular than that of related experimental ways (*in vitro* and *in vivo*). It should be noted that QSAR and QSPR are not separated topic as having a link between each other. For example, there is a relationship between melting point and solubility which is the latter could be considered as an important factor for drug delivery [1,2].

On the other side, the prediction of melting points (M.P) using theoretical investigations is considered as unreachable endeavor for computational chemists. This because there is no clear factor can be responsible for the melting point. In other words, the melting point can be considered as a balance of many factors that belonging to molecular interactions and molecular topography [3]. Therefore, serious efforts for obtaining an excellent model for prediction the melting points of chemicals can be considered as a good task for theoretical chemists. The operative QSPR model must be significance from both statistical and physical meaning points of views. In other words,

[1] Corresponding Authors
e-mail: rakhalil64@yahoo.com

the effective model should build up by a minimum number of descriptors and having excellent statistical parameters. Indeed, the decrease in the number of descriptors increases the confidence of the suggested model. Furthermore, the selected descriptors must give physical meaning that should be helpful in understanding the factors affecting the melting point or any another measured property. It should be noted that anyone whom dealing with statistical issue must bears in mind that the statistics not like mathematics as the former can gives results with no regards about the reality of input data. Thus, our present task is to develop a reliable QSPR model based on computational chemistry that could help in understanding the factors affecting the melting point which may then helping in pharmaceutical matters [1].

According to the literature, there are considerable efforts concerning theoretical prediction of melting points [3-13]. The common feature that one could realized from these published works is that the building up of models are based on the term of trial and error for obtaining the best statistical parameters regardless to the physical meaning of the suggested descriptors. In addition, the use of the numbers of atoms such as C, H, N, and O as descriptors for building up the suggested model does not give any general sense as there is many different compounds having same chemical formula [10].

In the present work, we have selected the melting point of a set of 32 drug like molecules derived from 4-(1-aryl-2-oxo-1,2-dihydro-indol-3-ylideneamino)-N-substituted benzene sulfonamides [14]. In relative terms, these compounds give a good opportunity for QSPR study as they possessing random substituents as could offer interpretation for the factors affecting the melting point. For instance, random substituents having different groups of aromatic and aliphatic in contrast to that of successive change in chemical structure as increasing the hydrocarbonic chain length and so on [11].

2. Computational Method

2.1. Experimental data

The experimental M.P of a set of thirty two derivatives of isatin (Figure 1) were taken from Kumar et al. [14] as illustrated in Table 1.

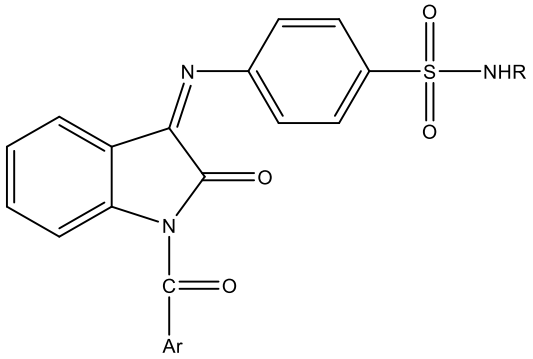
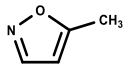
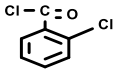
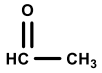
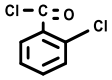
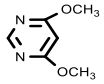
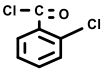
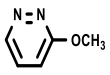
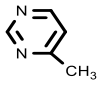
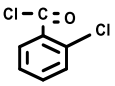
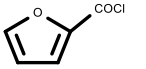
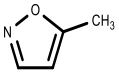
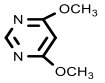
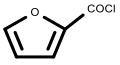
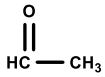
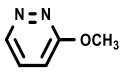
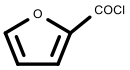
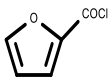
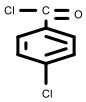
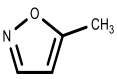
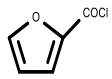
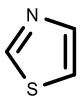
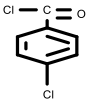
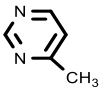
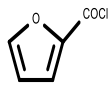
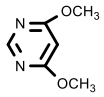


Figure 1. The skeleton of istatin derivatives.

2.2. Computational methodology

The chemical structures of each istain derivative together with their models were drawn, using two dimensional Chemdraw ultra version 11.0. Then, the molecular structure was transferred for organized energy minimization using Chem 3D-ChemBioOffice software version 16.0.0.82 (level: Ultra). The molecules must reach the global minima through energy minimization process using both molecular and quantum mechanics methods. The molecular mechanics calculations were started using MM2 followed by MMFF94 methods in order to catch a value of root mean square (RMS) gradient of less than 0.1 kcal/mol [3,15,16]. Then the optimization process must be continued using semi-empirical procedures which started using Austin Model number 1 (AM1), followed by the more advanced Parameterized Model number 3 (PM3) in order to get a negative sign of heat of formation and definitely a positive sign of frequency. The energy minimization process was continued using density functional theory (DFT) calculations B3LYP level with 6-311G (d,p) basis set till reaching the minimum RMS gradient of 0.1 kcal/mol [3]. The evaluations of quantum mechanical properties in order to employ them as descriptors in the developed model were made by Gaussian 03w software using, Hartree–Fock ab initio (HF), PM3 and DFT methods (closed-shell MOs). In other words, the selection of the method was depending on the kind of properties as each of these methods having a special characteristics in calculating definite theoretical descriptors. Statistical analysis is quite essential for QSPR investigations which was completed using professional software, Minitab software release 14.1. The root mean squared error (RMSE) was computed according to the following relation [5].

Table 1. The structure of 32 isatin derivatives with their melting points (M.P ºC) as taken from Kumar et al. [14] with the average melting points (M.P ºC (Av.)) that used for computations.

No	Ar	R.	M.P ºC	M.P ºC (Av.)	No	Ar	R.	M.P ºC	M.P ºC (Av.)
1			246-248	247	17			213-215	214
2			88-90	89	18			124-126	125
3			252-254	253	19			208-210	209
4			268-270	269	20			220-222	221
5			213-215	214	21			108-110	109
6			124-126	125	22			112-114	113
7			208-210	209	23			100-102	101
8			220-222	221	24			195-197	196
9			203-205	204	25			148-150	149
10			112-114	113	26			180-182	181
11			180-182	181	27			223-225	224
12			223-225	224	28			228-230	229
13			246-248	247	29			117-119	118
14			120-122	121	30			120-122	121
15			252-254	253	31			148-150	149
16			268-270	269	32			228-230	229

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1,n} y_{exp}^{i} - y_{cal}^{i}}$$

where y_{exp}^{i} and y_{cal}^{i} are experimental and calculated M.P values, n is the number of observations.

3. Results and discussion

In general, the calculated theoretical properties that could be used as descriptors for building up the developed model may be depend on the demanded purpose from the selected property of the suggested model. For example, estimating QSAR and QSPR (quantitative structure activity relationship) models are depend on the properties that belonging to the physical characteristics (i.e. electronic field) and topography. On the other side, building up of models dealing with chemical reactivity such as reaction kinetics are depending on the electronic properties such as HOMO, LUMO, softness and so on. Actually, the correlation matrix can be considered as the major filtration of the suitable properties for building up the developed model. Moreover, the correlation matrix does two important issues at once; the first is for distinguishing the most appropriate descriptors and the second is for giving an idea about the presence of linear combination between any descriptor with each other. The value of the correlation coefficient (r) can be considered as the best factor for the filtration process of correlation matrix. Therefore, descriptors of not good value of r in correlation with M.P are eliminated as well as those who having linear combination with others.

Table 2 lists the refined descriptors by correlation matrix in addition to new adopted three dimensional (3DI) descriptor of which it depends on the shape of 3D chemical structure of molecules. These descriptors are include zero-point energy (ZPE), Connolly accessible area (CAA), Connolly molecular area (CMA), Connolly solvent excluded volume (CSEV), molecular volume (MV), total connectivity (TC), average charge of RH groups (AQRH), average charge of R groups (AQR), average charge of Aryl hydrogen groups (AQArH), average charge of aryl groups (AQA) and three dimensional correction term (3DI). The last developed descriptor can be considered as a correction term with respect to the packing of molecules in optimized three dimensional chemical structure model which having values of -1, 0 and 1. The statistical treatments were carried out with regard to physical significance yielded the following QSPR model for coping the melting point of 32 observations of isatin derivatives.

M.P = 212 - 3598578 TC - 2732 AQArH + 55.5 3DI

Hence, a model of only three descriptors with relatively excellent statistical parameters of square regression coefficient (r2), cross-validation (q2), root mean squared error (RMSE) and standard error are equals to 0.925, 0.903, 15.26oC and 16.3oC, respectively. The r2 is also known as coefficient of determination or as goodness of fit which gives the strength of the linear relationship as its value indicates the closeness ratio of the points to the linear line. While q2 is termed goodness of prediction as it value indicates the strength of the model from the prediction point of view. RMSE is termed prediction errors which considered as the standard deviation of residuals between estimated and experimental values. Interestingly, this developed QSPR model was built up with respect to the physical meaning of which can be considered as a helpful tool in understanding the factors affecting the melting point of these substances. Thus, the constant term has a value of 212 as lying within the values of melting points which could reflect the reliability of the suggested model. The first descriptor of the total connectivity (TC) representing the packing of intermolecular alignment as could take a good role with M.P [17]. While the average charge on the aryl group (AQArH) of isatin was the second descriptor which also plays a good role in molecular interactions as the main responsible for M.P. It is not surprising to state that the melting point does depend on the strength of charge, but also it depends on the charge delocalization as the melting point increases with the increase in charge strength and charge localization [3]. The third developed parameter (3DI) is considered as a correcting term which taking only 1, 0 or -1 value of which depending on the optimized three dimensional chemical structure with respect to the skeleton of isatin (Figure 1). In other words, the suggested descriptor (3DI) is belonging to the stereochemical structure of the derivatives. The corrected value of 3DI (1, 0 or -1) can be selected according to the angle between the

indoline group with other main part of isatin from the aliphatic side with respect to the direction of the benzene ring plane in contrast to that of indoline as explained in Figure 2. Indeed, the type of angle reflects the change on molecular structure which plays a good role in the molecular packing as responsible for the melting point. Thus, if the orientation of the plane of benzene ring differs from that of indoline ring it takes 1 or -1 depending on the angle between them. The somewhat right angle takes a value of one (U shape), but obtuse angle takes a value of -1. On the other side, if the angle is relatively very obtuse and at the same time the orientation of the plane of benzene ring is close to that of indoline the correction term will take zero value. It was found that the ratio of substituents of their correction term having a value of zero is equal 31.25% for compounds 9, 11, 12, 17, 19, 20, 22, 24, 25 and 26 as their model only has two descriptors. On the other hand, derivatives with correction term of both 1 and -1 values having same percentage of 37.5%. The increasing of melting point due to the presence U shape for those having correction term equal to one may be related the effect of this shape in packing molecules in more tidiness and then increases the melting point in contrast to others. It should be noted that the terms of 3D as well as 4D were mentioned in the literature for QSAR modeling [18-20].

Table 2 Values of selected descriptors used in the regression analysis in addition to adopted 3D-correction term one.

No	ZPE	CAA	CMA	CSEV	MV	TC $x10^7$	AQRH $x\ 10^3$	AQR $x\ 10^3$	AQArH $x\ 10^3$	AQA $x\ 10^3$	3DI
1	260.5	699.0	385.6	333.8	3356	23	-1.2	-86.4	6.57	-106.7	1
2	241.7	691.1	384.9	341.9	2959	114	24.4	-191.4	9.10	-155.3	-1
3	254.9	730.2	401.4	357.3	3432	20	-4.4	-87.9	6.52	-105.1	1
4	276.3	733.5	405.8	360.2	39145	16	18.4	-125.4	8.51	-155.9	1
5	239.9	726.1	410.2	360.2	3762	28	26.3	-175.6	10.70	-154.8	1
6	295.5	761.4	431.8	350.8	3544	7	12.3	-132.5	10.65	-151.0	-1
7	272.9	749.2	427.6	349.3	3167	12	20.1	-67.6	12.96	-105.0	1
8	263.0	756.5	472.3	341.2	3899	20	25.5	-109.6	6.95	-106.8	1
9	254.1	725.2	399.7	338.6	3977	19	25.8	-118.3	9.39	-116.9	0
10	270.0	762.2	416.7	363.7	3502	13	18.2	-127.1	7.82	-116.9	-1
11	295.6	767.4	424.1	351.8	3881	5	12.9	-132.9	7.59	-117.3	0
12	272.5	766.8	444.5	354.2	3979	9	18.9	-106.8	3.61	-126.7	0
13	273.3	701.9	388.3	340.7	3958	19	26.3	-116.0	0.22	-124.4	1
14	238.1	693.6	387.6	352.6	3619	93	25.8	-198.6	0.07	-126.2	-1
15	251.3	732.8	402.8	370.4	3446	16	22.0	-72.5	-0.39	-126.2	1
16	269.9	738.2	407.9	369.9	3319	13	18.0	-129.1	-0.33	-126.6	1
17	233.1	728.4	409.1	373.2	3876	23	24.6	-55.1	-0.15	-126.2	0
18	295.2	763.1	433.4	359.7	3776	5	9.6	-138.2	-1.28	-123.1	-1
19	268.5	751.2	428.6	360.2	3172	9	18.6	-109.3	0.81	-125.0	0
20	258.8	758.8	474.1	342.6	3772	16	24.2	-104.1	3.73	-116.9	0
21	244.7	659.8	388.9	358.3	3172	33	25.6	-120.0	12.99	-114.7	-1
22	224.4	667.7	363.9	312.7	2945	161	24.5	-191.4	13.03	-113.0	0
23	237.8	706.3	388.1	334.4	3274.3	28	22.2	-72.5	12.94	-113.0	-1
24	256.0	692.8	400.9	381.7	3619.9	23	18.6	-125.1	9.09	-121.2	0
25	223.2	704.1	385.3	334.0	2947.5	40	24.7	-53.6	13.08	-113.2	0
26	281.7	741.5	398.7	340.8	3948.2	9	13.3	-132.5	13.26	-112.9	0
27	259.1	699.4	431.6	362.4	3834.2	16	18.8	-112.4	12.50	-114.0	1
28	245.5	716.7	395.6	340.7	2797	28	25.3	-92.7	11.93	-114.9	1
29	238.0	718.7	399.3	348.0	3406.4	93	28.0	-183.6	7.542	-117.5	-1
30	251.5	752.8	416.8	362.7	3474.8	16	22.9	-72.0	7.25	-117.9	-1
31	237.0	751.8	417.0	365.2	3478	23	25.1	-53.5	7.90	-116.9	-1
32	259.5	765.4	404.3	350.0	3134	1	24.4	-104.0	7.50	-116.9	1

The values of correction term may also related to the type of substituents, i.e. the presence of acetyl group in aliphatic part of isatin takes -1 value of 3DI as compounds 2, 14 and 29 except for compound 22 which has 3DI equal to zero due to the absence of benzene ring in the aromatic part as substituted by furan. Actually, the presence of oxygen atom in furan ring may increases the melting point by enhancing the electrostatic forces that responsible for molecular interactions between molecules. In the similar manner the presence of pyridine also increases the melting point (3DI =1) as compounds 8, 28 and 32 except that for compound 20 (3DI = 0) which may be related to presence of chlorine in ortho position of benzene of the aromatic part of isatin. In general, the substituent effect on M.P can be summarized by both inductive and resonance effects as both could influence the intermolecular interactions of which the main responsible for melting point [3].

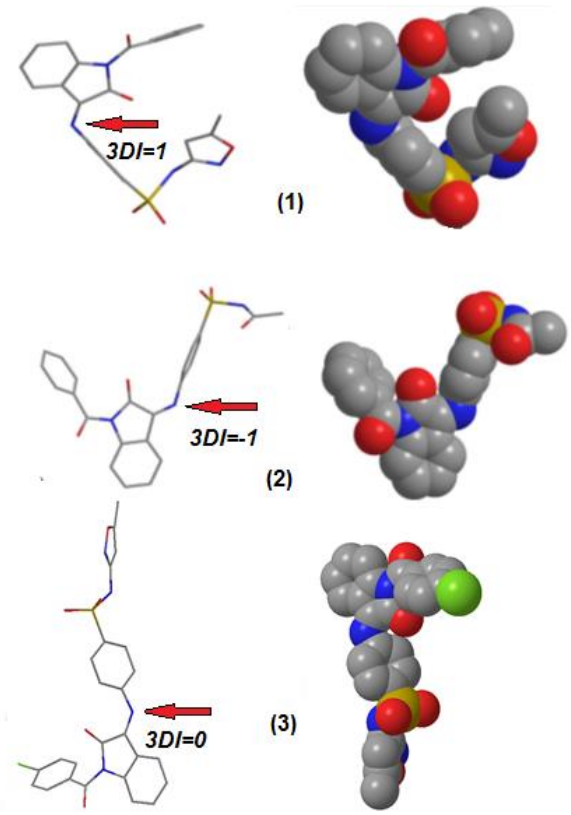


Figure 2. Illustrates the depending of 3DI parameter on the stereochemical of isatin derivatives as shown by the arrows for compounds 1, 2 and 9 of which they take values of 1, -1 and 0 respectively.

Table 3 Estimated ($M.P_{est}$) and residuals values for the melting point according to the suggested model.

No	$M.P_{est}$ ºC	Residual ºC	No	$M.P_{est}$ ºC	Residual ºC
1	241.70	5.30	17	204.57	9.43
2	91.22	-2.22	18	158.53	-33.53
3	242.97	10.03	19	206.86	2.14
4	238.85	30.15	20	196.41	24.59
5	228.54	-14.54	21	109.69	-0.69
6	125.48	-0.48	22	119.06	-6.06
7	228.43	-19.43	23	111.40	-10.40
8	241.79	-20.79	24	179.31	16.69
9	180.02	23.98	25	162.30	-13.30
10	130.80	-17.80	26	172.85	8.15
11	189.79	-8.79	27	227.95	-3.95
12	199.22	24.78	28	225.17	3.83
13	260.57	-13.57	29	102.98	15.02
14	123.39	-2.39	30	131.27	-10.27
15	263.16	-10.16	31	127.06	21.94
16	264.09	4.91	32	241.60	-12.60

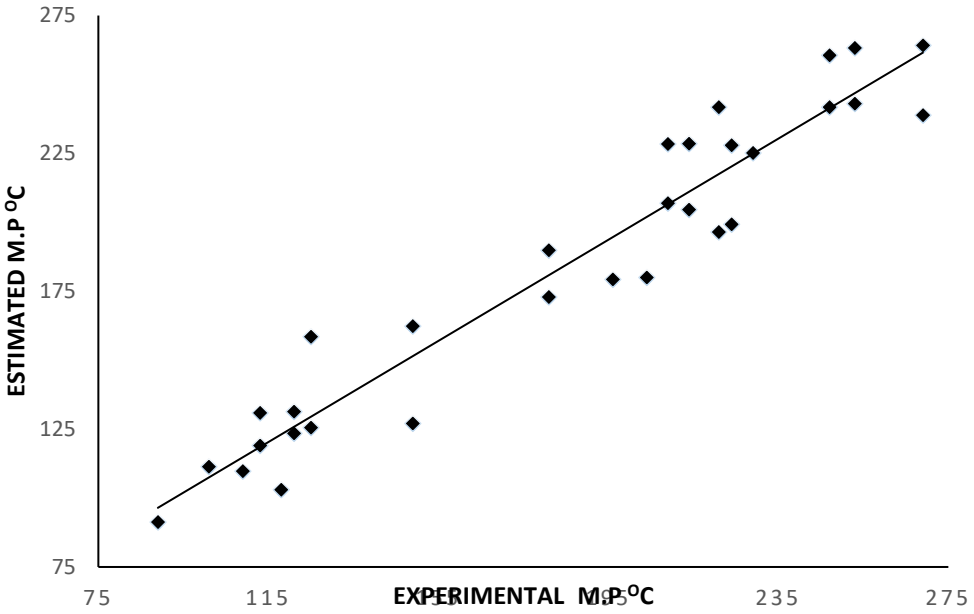


Figure 3. The plot of estimated M.P values versus observed M.P values according to the developed model with r^2 equals to 0.925.

Table 3 lists the predicted and residual values of M.P resulted from the application of the suggested model, while Figure 3 shows the linear relationship between estimated and observed M.P. with r2 equal to 0.925 indicating the reliability of the developed QSPR model.

4. Conclusion

Theoretical investigations for the melting point may be considered as the most important task for theoretical chemists due to the significant importance of this physical property. A statically intensive QSPR model with only three descriptors has been developed for 32 biologically active molecules of isatin derivatives. Two of these descriptors are belonging to direct theoretical calculations but the third was a suggested correcting term which is depend on three dimensional optimized chemical structure of molecule and could only take the value of -1, 0 or 1. The results indicate that the developed model is more reliable from statistical point of view as having values of r^2, q^2 and RMSE equal to 0.925, 0.903 and 15.26°C, respectively, in addition to physical significance which can be considered as a helpful tool for understanding the factors affecting the melting point.

References

[1] Y. Ran, S.H. Yalkowsky, Prediction of drug solubility by the general solubility equation (GSE), J Chem Inf Comput Sci 41 (2001) 354–357.

[2] W.L. Jorgensen, E. M. Duffy, Prediction of drug solubility from structure. Adv Drug Deliv Rev 45 (2002) 355-366.

[3] R.A. Khalil, A.Y. Hamed, Theoretical investigation using DFT for predicting the factors affecting the melting point of series of alkylammoniumformates ionic liquids, Arab J Phys Chem 2 (2015) 56-63.

[4] I.V. Tetko, Y. Sushko, S. Novotarskyi, L. Patiny, I. Kondratov, A. E. Petrenko, L. Charochkina, A. M. Asiri, How accurately can we predict the melting points of drug-like compounds? J Chem Info Model 54 (2014) 3320-3329.

[5] I.V. Tetko, D. M. Lowe, A.J. Williams, The development of models to predict melting and pyrolysis point data associated with several hundred thousand compounds mined from PATENTS, J Cheminform 8 (2016) 186-203.

[6] C.A.S. Bergström, U. Norinder, K. Luthman, P. Artursson, Molecular descriptors influencing melting point and their role in classification of solid drugs, J Chem Inf Comput. Sci 43 (2003) 1177-1185.

[7] H. Modarresi, J.C. Dearden, H. Modarress, QSPR correlation of melting point for drug compounds based on different sources of molecular descriptors, J Chem Info Model 46 (2006) 930-936.

[8] A. R. Katritzky, U. Maran, M. Karelson, V.S. Lobanov, Prediction of melting points for the substituted benzenes: A QSPR approach. J Chem Inf Comput. Sci 37 (1997) 913-919.

[9] B. Johnson-Restrepo, L. Pacheco-Londoño, J. Olivero-Verbel, Molecular Parameters Responsible for the Melting Point of 1,2,3-Diazaborine Compounds, J Chem Inf Comput. Sci 43 (2003) 1513-1519.

[10] R.F. Alamdari, M.H. Keshavarz, A simple method to predict melting points of non-aromatic energetic compounds. Fluid Phase Equilibria 292 (2010) 1–6.

[11] A. Guendouzi, S.M. Mekelleche, Prediction of the melting points of fatty acids from computed molecular descriptors: A quantitative structure–property relationship study, Chem Phys Lipids 165 (2012) 1-6.

[12] J.A. Morrill, E.F.C. Byrd, Development of quantitative structure property relationships for predicting the melting point of energetic materials, J Mol Graph Model 62 (2015) 190–201.

[13] D. Wang, Y. Yuan, S. Duan, R. Liu, S. Gu, S. Zhao, L, Liu, J. Xu, QSPR study on melting point of carbocyclic nitroaromatic compounds by multiple linear regression and artificial neural network, Chemom Intell Lab Syst 143 (2015) 7-15.

[14] M. Kumar, K. Ramasamy, V. Mani, R. K. Mishra, A. Abdul Majeed, E. De Clercq, B. Narasimhan. Synthesis, antimicrobial, anticancer, antiviral evaluation and QSAR studies of 4-(1-aryl-2-oxo-1,2-dihydro-indol-3-ylideneamino)-N-substituted benzene sulfonamides, Arabian J Chem 7 (2014) 396-408.

[15] R.A. Khalil, A.A. Zarari, Theoretical investigations for the behavior of hydrotropes in aqueous solution, J Turk Chem Soc A 2 (2015) 42-52.

[16] R.A. Khalil, A.A. Zarari, Theoretical estimation of the critical packing parameter of amphiphilic self-assembled aggregates, Appl Surf Sci 318 (2014) 85-59.

[17] M. Črepnjak, N. Tratnik, P.Ž. Pleteršek, Predicting melting points of hydrocarbons by the Graovac-Pisanski index, Fullerenes, Nanotubes Carbon Nanostructures 26 (2018) 1080-1101.

[18] B. Tüzün, E. Saripinar, Molecular docking and 4D-QSAR model of methanone derivatives by electron conformational-genetic algorithm method, J Iran Chem Soc 17(2020) 985-1000.

[19] B.Tuzun, S.C. Yavuz, N.Sabanci, E. Saripinar, 4D-QSAR Study of some pyrazole pyridine carboxylic acid derivatives by electron conformational-genetic algorithm method, Current Computer-Aided Drug Design 14 (2018) 370- 384.

[20] B. Tuzun, S.C. Yavuz, E. Saripinar, 4D-QSAr analysis and pharmacophore modeling: propoxy methylphenyl oxasiazole derivatives by electron conformatitional-genetic algorithm method, J. Physical Theoretical chem 14 (2017) 149-164.