Journal of Chemical Technology and Metallurgy, 61, 6, 2026, 981-994

IN SILICO INVESTIGATION FOR MODELING ISATIN DERIVATIVES AS ANTI-COLORECTAL CANCER AGENTS ASSISTED BY MOLECULAR DOCKING

Anfal R. Mahmoud, Rabah Ali Khalil

Department of Chemistry, College of Science, University of Mosul
Mosul 41002, Iraq, anfal.rm.albrhawi@uomosul.edu.iq (A.R.M.);
rabahalikhalil@uomosul.edu.iq (R.A.K.).

Received 10 August 2026
Accepted 20 September 2026 *DOI: 10.59957/jctm.v61.i6.2026.5*

ABSTRACT

The present paper introduces an in silico study using quantitative structure-activity relationship (QSAR) for the possibility of developing anti-colorectal cancer from isatin derivatives. The theoretical treatments were carried out using quantum mechanical calculations for 29 compounds, with 25 as the training set and the rest as the test set. The results indicate that the estimated properties as descriptors are not adequate for building the QSAR model. It was found that the outcome parameters from the docking study are quite effective to be used as descriptors in building the model. Two QSAR models were developed for estimating the cytotoxic activity data IC_{50} regarding the colon cancer (HCT-116) cell line utilizing the binding energy that resulted from the docking study as a descriptor. It was concluded that the significance of this research extends beyond the prediction of new compounds and the reduction of trial-and-error approaches in chemistry through introducing the resulting binding energy from the docking study as a descriptor. In the long term, these computational strategies may be integrated with artificial intelligence to suggest the most suitable compounds for specific diseases. Such advancement could represent a qualitative leap in drug discovery by reducing economic costs, minimizing environmental impact, decreasing human effort, and significantly shortening the required development time.

Keywords: *QSAR, in silico, isatin, colorectal cancer, docking.*

INTRODUCTION

Colorectal cancer (CRC) is the third most frequently diagnosed malignant tumor and is considered the second leading cause of global cancer-related fatalities [1]. The incidence and mortality rates for CRC vary significantly across different countries, influenced by factors such as gender, age, and ethnicity [2]. This disease imposes a substantial burden on global health due to its complications, mortality rates, treatment side effects, healthcare resource utilization, and associated medical expenses [3]. All of these factors highlight the urgent need for innovative therapeutic approaches by scientists.

One of the updated methods employed in these investigations is the *in silico* approach, which is considered very important due to its great advantages that have played an essential role in drug discovery. Among these advantages is reducing the problem of trial and error, particularly in terms of time consumption, economic cost and environmental aspects. In other words, theoretical study could lead to selected molecules that could be used as an efficient agent for treating the specific biological target, as well as for understanding the factors that influence the treatment of that case. In this approach, a quantitative structure-activity relationship (QSAR) is employed for building suitable predictive models using the theoretically calculated descriptors derived from quantum mechanical methods. These models will then be exploited in predicting new, more efficient substituents that can be used for therapy, noting that the factor of toxicity is already taken into account in this study. In general, such an *in silico* strategy gives a successful outcome when applied to isatin-

based antibiotics [4], pesticides [5] and even corrosion inhibitors [6]. In this study, a theoretical investigation was conducted using previously published data on the anti-colorectal cancer activity of 29 isatin derivatives reported by Kumar et al. [7]. The molecular framework of these compounds is illustrated in Fig. 1. Isatin is an indole-derived heterocyclic compound that is found naturally in mammalian systems and various biological sources and has attracted considerable attention as a versatile pharmacophore in anticancer drug discovery. Its significance is exemplified by the development of clinically approved anticancer agents such as Sunitinib and Nintedanib [8, 9]. The isatin scaffold possesses a significant amount of structural flexibility and inherent biological activity, which makes it very amenable to chemical modifications, and derivatives with enhanced binding affinity and selectivity for the hydrophobic binding pocket of murine double minute 2 (MDM2) can be designed [10].

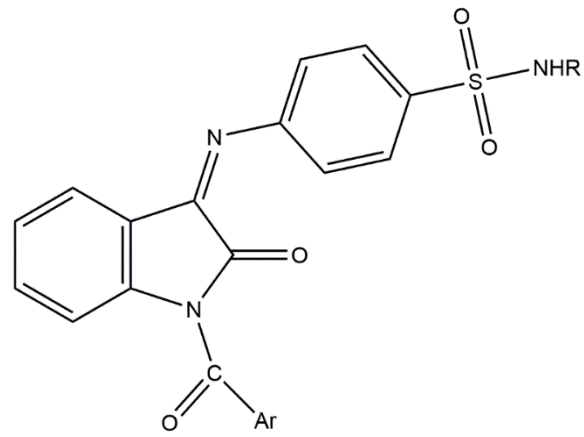


Fig. 1. The skeleton of isatin derivatives.

The present study is anticipated to theoretically explore the potential of isatin derivatives as inhibitors of MDM2 using advanced in silico approaches. Special emphasis is placed on the search for compounds disrupting the binding of MDM2 with the tumor suppressor protein P53 - the "guardian of the genome" in the human colorectal cancer (HCT-116) cells. Such computational studies are important not only for the development of a predictive mathematical model but also for gaining valuable insights into the molecular factors controlling cancer cell proliferation, provided that the resulting model is based on sound physicochemical grounds. To construct a QSAR model with physically meaningful interpretation, the limitations of the conventional statistical methods should be eliminated. Unlike rigorous mathematical approaches, statistical analyses may generate apparently acceptable relationships without adequately considering the validity and physicochemical relevance of the input descriptors. Consequently, integrating mathematically and physically meaningful parameters into QSAR modelling is crucial for enhancing the reliability, interpretability, and predictive performance of the developed model [11].

It should be noted that the previous studies concerning the presented topic have validated the effectiveness of various small molecules, including certain natural products and synthetic derivatives in obstructing the P53-MDM2 interaction [12 - 15]. In addition, some isatin-based compounds have shown potential as activators of P53 [16 - 20]. According to the literature, no MDM2 inhibitors have yet been approved by the Food and Drug Administration (FDA) [21]. In a more specific aspect, three studies have utilized computational approaches for deriving suitable models for isatin derivatives as anti-colon cancer elucidate the mechanism of binding and conformational changes induced by small molecule inhibitors within the MDM2 binding pocket [22 - 24]. A more negative binding energy in molecular docking studies often signifies a more stable and favourable interaction between a ligand and the MDM2-P53 complex, thereby indicating stronger inhibitory potential [25]. This work presents an innovative *in silico* methodology that combines molecular docking results with QSAR modeling to construct a predictive model with improved physicochemical relevance. The integration of docking-derived descriptors is expected to strengthen the theoretical foundation of the model, thereby enhancing its reliability, interpretability, and predictive capability.

EXPERIMENTAL

In this study, the cytotoxic activity data IC_{50} regarding the colon cancer (HCT-116) cell line in μg mL^{-1} unit were obtained from already published research by Kumar et al., the data set comprised twenty-nine isatin derivatives whose general molecular scaffold presented

in Fig. 1 [7]. These data were divided into a training set comprising twenty-five compounds, and the remaining four were designated as a test set. The training dataset was utilized to develop and conduct internal validation of the model, whereas the test dataset was reserved for external validation to guarantee an impartial evaluation of performance.

The theoretical treatments were started by drawing chemical structures of the isatin derivatives using two-dimensional ChemDraw Professional 17.0 software. Each molecular structure underwent systematic energy minimization with Chem 3D 17.0 software. To achieve the global minimum in optimized time using an ordinary PC, the process was started by energy minimization using molecular mechanics calculations MM2, followed by MMFF94 methods to attain an RMS gradient below 0.1 kcal mol^{-1} [26]. The minimization continued with semi-empirical calculations incorporating AM1 and PM3 methods to achieve a negative heat of formation and a positive frequency sign. Then, the ab initio method of Hartree-Fock (HF) with a 6-31G basis set was employed for further minimization. For density functional theory (DFT) calculations, energy minimizations progressed using a hybrid DFT approach that included Becke's three-variable exchange potential and the Lee-Yang-Parr functional (B3LYP) at the 6-31G basis set until an RMS gradient minimum of 0.1 was achieved [4]. Descriptor estimation was conducted using Gaussian 09w software as the software default, employing DFT, HF, and PM3 methods for closed-shell molecular orbitals based on the chosen descriptor type. These calculated descriptors include constitutional, topological, geometrical, electronic, and thermodynamic properties.

For the docking study, phenylalanine and leucine amino acids were optimized using the same steps as above to reach global energy minimization. Then each isatin derivative was grouped with an amino acid, followed by rigorous geometry optimization of the grouped structure to achieve the global energy minimum using the HF method with a basis set of 6-21G. Subsequently, intermolecular hydrogen bond distances were measured, and the binding energy of the resulting complex was calculated. This parameter was utilized as a novel molecular descriptor for the subsequent development of a robust QSAR model. Statistical analysis for QSAR was performed with Minitab software release 18.

RESULTS AND DISCUSSION

QSAR was performed on 29 isatin derivatives to investigate the relationship between molecular structure and the inhibitory effects on the MDM2 protein. Table 2 shows some of the theoretically calculated descriptors for the isatin derivatives using quantum mechanics methods. The descriptors chosen for statistical analyses were based on their interdependencies, as determined by the constructed correlation matrix that provides useful information on the degree of correlation of different pairs of descriptors. The filtration process by the correlation matrix is based on the combination of descriptors, as shown in Table 3. For example, as observed in Table 3, there is a direct correlation between the ionization potential (Ip) and the energy of the highest occupied molecular orbital (HOMO), as the value of the correlation coefficient is close to or equal to unity, which means one of them must be filtered.

Where Ip is ionization potential, log S is the compound's aqueous solubility, and E bin is the energy of binding between the isatin derivative and leucine amino acid. It should be noted that the theoretically calculated descriptors can be divided into two main groups, including electronic and field effects. The first effect includes the properties that belong to the valence electrons, such as the frontier molecular orbitals of HOMO, LUMO, band gap, and so on, which are responsible for chemical reactivity, such as the rate of chemical reaction. In contrast, the field effect refers to the physical forces that are dependent not only on the charge distribution but also on the topography of molecules, as well as the types of atoms and functional groups they contain. The last effect is responsible for molecular interactions, especially the physical attractions between drugs and living tissues. Over 100 theoretical descriptors were estimated for isatin derivatives, including those obtained for curve-fitting treatments by taking their mathematical functions, such as natural logarithm (ln), exponential (exp), and so on, to improve the statistical parameters of the constructed model. Noting that considerable attention must be taken when dealing with statistical issues, as it is not an exact method like mathematics. In other words, the statistical method could treat any entered data with no regard to the reality of these data, in contrast to that of mathematics. Moreover, the number of observations and model terms play a major

Table 1. The chemical structure of isatin derivatives with their HCT-116 values of IC_{50} according to Fig. 1, as taken from reference [7].

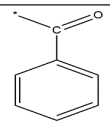
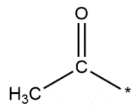
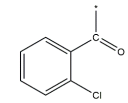
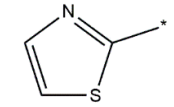
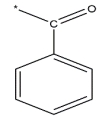
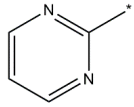
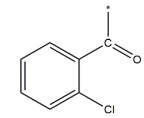
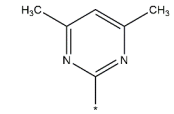
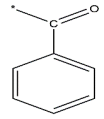
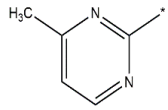
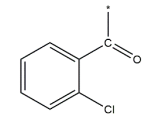
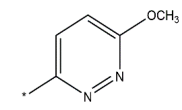
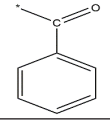
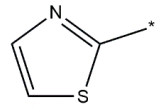
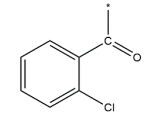
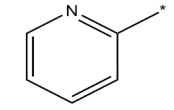
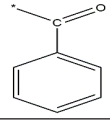
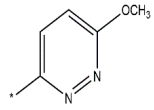
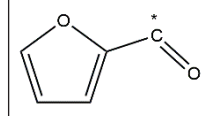
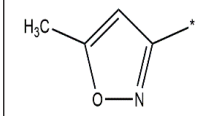
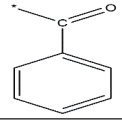
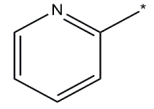
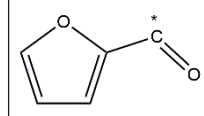
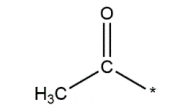
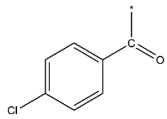
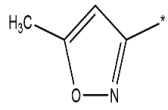
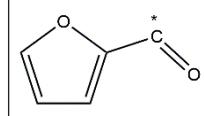
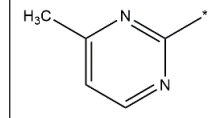
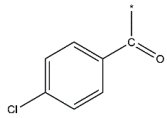
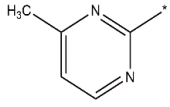
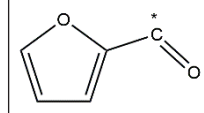
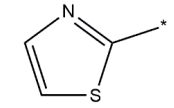
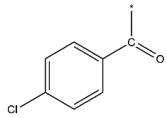
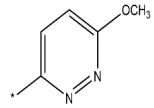
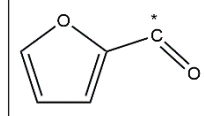
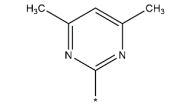
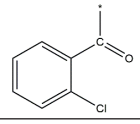
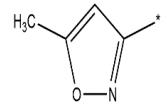
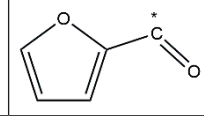
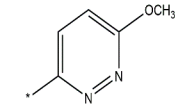
No	R1 (Aryl group)	R2 (Alkyl group)	IC_{50}, μg mL^{-1}	No	R1 (Aryl group)	R2 (Alkyl group)	IC_{50}, μg mL^{-1}
Training set							
1			70	15			200
2			70	16			210
3			45	17			210
4			40	18			480
5			120	19			70
6			50	20			400
7			35	22			260
8			150	23			200
10			120	24			120
11			280	25			210
12			400	26			500
13			500	27			70
14			300				

Table 1. The chemical structure of isatin derivatives with their HCT-116 values of IC_{50} according to Fig. 1, as taken from reference [7] - *continued.*

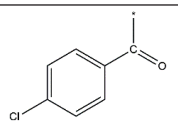
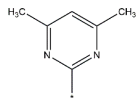
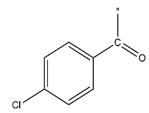
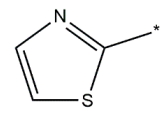
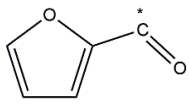
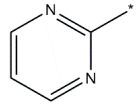
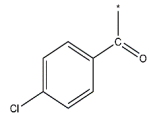
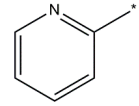
No	R1 (Aryl group)	R2 (Alkyl group)	IC_{50}, µg mL^{-1}	No	R1 (Aryl group)	R2 (Alkyl group)	IC_{50}, µg mL^{-1}
Training set							
Test set							
9			50	28			90
21			110	29			80

Table 2. Example of some of the calculated descriptors of twenty-nine isatin derivatives.

N.	HOMO, eV	LUMO, eV	E-Gap, eV	Hardness	µ*	log S, log units	Balaban index	Dipole, Debye
1	-0.244	-0.1130	0.131	0.065	-0.178	-5.586	1359822	7.587
2	-0.240	-0.1134	0.127	0.064	-0.177	-6.375	1816973	7.376
3	-0.238	-0.1124	0.126	0.063	-0.175	-6.648	2083132	7.343
4	-0.235	-0.1116	0.124	0.062	-0.173	-6.553	1572317	2.952
5	-0.236	-0.1122	0.124	0.058	-0.177	-7.169	2372081	8.305
6	-0.235	-0.1188	0.117	0.059	-0.1703	-7.111	2414561	6.199
7	-0.230	-0.1112	0.118	0.061	-0.192	-7.768	1816973	13.242
8	-0.253	-0.1307	0.122	0.062	-0.178	-7.337	2080770	7.975
9	-0.240	-0.1157	0.124	0.061	-0.176	-7.609	237720	9.619
10	-0.237	-0.1155	0.122	0.052	-0.173	-7.857	2697801	8.423
11	-0.225	-0.1207	0.105	0.060	-0.179	-7.752	2743970	8.098
12	-0.239	-0.1184	0.120	0.065	-0.180	-6.259	2054269	6.819
13	-0.245	-0.1144	0.130	0.063	-0.178	-7.048	1555851	8.707
14	-0.242	-0.1148	0.127	0.060	-0.179	-7.321	2054706	8.896
15	-0.239	-0.1184	0.120	0.055	-0.170	-7.226	2348338	6.045
16	-0.225	-0.1148	0.110	0.062	-0.175	-7.593	1784218	8.902
17	-0.237	-0.1136	0.124	0.058	-0.181	-7.841	2666457	8.015
18	-0.239	-0.1231	0.116	0.059	-0.172	-7.784	2712561	7.769
19	-0.231	-0.1126	0.118	0.060	-0.178	-6.011	2054706	11.667
20	-0.238	-0.1173	0.121	0.061	-0.181	-4.520	1585232	7.339
21	-0.242	-0.1208	0.121	0.064	-0.177	-5.308	1169971	6.209
22	-0.241	-0.1133	0.127	0.061	-0.178	-5.580	1585627	5.996
23	-0.239	-0.1169	0.122	0.055	-0.168	-5.485	1824777	3.764
24	-0.223	-0.1133	0.110	0.061	-0.178	-5.853	1366340	6.194
25	-0.238	-0.1174	0.121	0.058	-0.179	-6.101	2085016	7.296
26	-0.237	-0.1210	0.116	0.059	-0.170	-6.043	2124021	5.259
27	-0.229	-0.1110	0.118	0.062	-0.182	-6.275	1585627	6.483
28	-0.243	-0.1197	0.124	0.054	-0.170	-7.242	1579399	6.557
29	-0.224	-0.1166	0.107	0.058	-0.179	-7.800	1808479	6.563

**µ is chemical potential.*

Table 3. An example of the correlation matrix between some theoretically calculated properties and experimental IC_{50}, μg mL^{-1}.

	IC_{50}	HOMO, eV	LUMO, eV	E-Gap, eV	Ip	log S
HOMO, eV	0.062					
LUMO, eV	0.159	0.411				
E-Gap, eV	0.041	-0.787	0.239			
Ip	-0.062	-1.000	-0.411	0.787		
log S, log unit	0.101	-0.135	0.254	0.316	0.135	
E bin	-0.068	0.160	-0.340	-0.401	-0.160	-0.838

role in the reliability of the outcome results. For example, statistical parameters such as the square correlation coefficient (r^2) and standard error (SE) are significantly improved by increasing the number of model descriptors and decreasing the number of observations, and vice versa. Hence, a constructed model with relatively few descriptors is quite close to that of a mathematical model from the reliability and physical meaning point of view. However, the attempts at statistical treatments for the obtained descriptors cannot give a reasonable model for the HCT-116 dependent variable due to not good statistical parameters. Such a surprising phenomenon may be related to the relative complexity of this type of cancer, which is medically considered a wild case. Consequently, considerable attention should be directed towards this issue for obtaining insights that may contribute to a better understanding of therapeutic strategies for colorectal cancer. Therefore, the present study proposes the incorporation of molecular docking outcomes of isatin derivatives with the selected biological target into the construction of a QSAR model. In this investigation, the selected target is the murine double minute 2 (MDM2) protein, which plays a critical role in the proliferation of human colorectal cancer cells, particularly the HCT-116 cell line [27].

The interaction between MDM2 and P53 represents a key regulatory pathway in colorectal cancer progression. Under pathological conditions, overexpression of MDM2 suppresses the activity of the tumor suppressor protein P53, commonly referred to as the "guardian of the genome." Since p53 is encoded by the TP53 gene and serves as a fundamental regulator of genomic stability, apoptosis, and cell-cycle control, its inactivation is strongly associated with cancer development. In many colorectal cancers, P53 activity is impaired due to the excessive expression of its negative regulator, MDM2, an E3 ubiquitin-protein ligase [28, 29]. Accordingly, disrupting the P53–MDM2 interaction has emerged as a promising pharmacological strategy for colorectal cancer treatment [25]. The inhibition of MDM2 can restore the biological function of P53, which will promote DNA repair, induce cell-cycle arrest, and trigger apoptosis in cancer cells [30, 31]. Furthermore, P53 interacts predominantly with the hydrophobic pocket of MDM2, as shown by alanine-scanning mutagenesis studies for the key residues Phe19, Trp23, and Leu26 [32]. This observation indicates that small molecule inhibitors, including Nutlin analogs, can mimic the α-helical structure of P53 and compete for the binding cavity of MDM2 in vitro and in solution. This leads to a disruption of the inhibitory interaction between MDM2 and P53, resulting in the reactivation and release of P53 protein [27]. The design of leucine-isatin derived inhibitors is based on a strategic approach to the structural topology of the MDM2 binding domain that contains a characteristic V-shaped hydrophobic cleft. Small-molecule ligands that block the interaction between P53 and MDM2 must include specific hydrophobic pharmacophores (usually aromatic rings) mimicking the three key residues of P53 (Phe19, Trp23 and Leu26), and bind their respective sub-pockets. The most important anchor point for high binding affinity is the interaction in the central hydrophobic cavity, normally filled with the Trp23 indole ring.

On the other hand, due to the huge size of the protein, the docking process is usually done using molecular mechanics rather than quantum mechanics. But there is no doubt that the accuracy of quantum mechanics is relatively too high, as the mechanics depend on classical physics laws of Newton and Hook, in addition to considering the atoms as only nuclei without electrons. Therefore, the docking procedure will be more sufficient if it can be done using quantum mechanics if the key (amino acid) is known, particularly in the database of

GenBank. Both literature and GenBank suggest that the leucine and phenylalanine amino acids are the keys of P53 [33]. Thus, molecular docking integrated with quantum mechanics was applied to the isatin derivatives with leucine and phenylalanine residues, subsequently. The group of isatin derivatives with leucine was optimized to their global energy minima. Noting that, the docking study was carried out using the ab initio HF method, as the DFT is too expensive, which takes a very long time. Then the distances between the atoms involved in hydrogen bonding, as illustrated in Table 4, were measured according to the final optimized state, as clearly shown in the supplementary materials. The maximum intermolecular distance is equal to 2.555 Å, indicating the significant interaction between leucine and isitin derivatives. In other words, the intermolecular distance should be lower than 3.5 Å, as the stability of the resulting group is directly affected by it, as shorter bonds generally indicate more robust interactions [34]. However, no reasonable correlation was found between the intermolecular distance and HCT-116 for building the QSAR model.

The binding energy (E_{bin}) of the complex that was formed by the interaction of isatin with the amino acid was estimated according to the model presented in Eq. (1).

$$E_{bin} = E_{comp} - (E_{isa} + E_{am}) \quad (1)$$

Table 4. Close contact distance between isatin derivatives and leucine amino acid.

No.	Isatin d. + Leucine	Close contact, Å	No.	Isatin d. + Leucine	Close contact, Å
1	O9-H71 H37-O57	1.614 1.709	16	H55-O66	2.452
2	O8-H62 H36-O60	2.110 2.385	17	H56-O64 O38-H66	2.175 2.204
3	H50-O63	2.508	18	H41-O61 O8-H70	1.889 2.410
4	O26-H60	2.500	19	H49-O58 H41-O59 O11-H62	2.223 2.373 2.436
5	H53-N58	2.411	20	O21-H68 H45-O54 H41-O55	1.877 2.335 2.367
6	H40-N55 O8-H63 H41-O61	1.925 2.115 2.407	21	O8-H71 H35-O57	1.678 2.213
7	O27-H75 H37-N55 H42-O62	1.824 2.280 2.323	22	O9-H74 H39-O60	1.699 2.067
8	H53-N57	2.354	23	O8-H69 H38-O55	1.631 1.689
9	H46-O66	2.437	24	O8-H77 H37-O63	1.638 2.202
10	O9-H78 H42-O64	1.781 2.261	25	O17-H72	2.555
11	N12-H75 H41-N55 H37-O61	1.980 1.773 2.242	26	H49-O58	2.281
12	O9-H71 H38-O57	1.624 1.718	27	H38-O57 O9-H71	1.716 1.626
13	O8-H68	2.469	28	H40-O59 O8-H62	1.857 2.387
14	H55-O63	2.276	29	H41-N55 O8-H64 H42-O61	1.919 1.972 2.454
15	H39-O59 N13-H62	2.303 2.414			

where E_{comp} is the obtained complex minimized energy from the docking study, E_{isa} is the minimizing energy that has already been measured for isatin derivatives, and E_{am} is the amino acid minimization energy.

Interestingly, the statistical treatments showed a significant improvement in constructing the QSAR model due to the presence of the binding energy descriptor. It is not surprising to find that as the intermolecular interactions have a direct proportional relationship with binding force, as the binding energy decreases, the attraction forces are increased. Indeed, such a combination between QSAR and molecular docking based on quantum mechanics may be considered an interesting approach for developing theoretical chemistry in pathological treatments. Hence, the suggested QSAR model for estimating HCT-116 for 25 compounds as a training set is presented in Eq. (2), with a square correlation coefficient (r^2) and standard error equal to 0.844 and 63.65 μg mL^{-1}, respectively.

$$HCT\text{-}116 = -1956 + 134.9\ \text{Ebin leu} + 48056\ \exp(\text{LogS}) + + 369.1ZO \quad (2)$$

where exp (log S) is the exponential of the solubility of compounds in aqueous media, Ebin leu is the binding energy between isatin derivatives and leucine, and ZO is the zero-one correction term for only 4 specific unusual compounds.

The obtained model (2) is considered quite strong from a statistical point of view as it contains only two descriptors with 25 observations. Noting that, the term of ZO just takes the value of one for only four molecules and zero for the rest, which means only two descriptors for the rest of the compounds. According to the Topliss and Costello rule, the ratio of training set observations to the number of descriptors is about 10:1, which indicates the significance of this model, as the ratio should be at least 5:1 using regression analysis [5]. In addition, this model gives a physical meaning, as the sign of the binding energy term is positive, which is parallel to the inverse relationship between the strength of binding and the value of energy. The term log S, which belongs to the solubility of isatin derivatives in water, plays an important role in drug delivery to the infected cells. The third descriptor was the zero-one correction term (ZO) that is constructed based on the chemical structure of a molecule, specifically impacted by its substituents. This descriptor was already adopted by previous research [4], which only takes the values of 0 and 1. ZO can support the physical relevance of the suggested model, considering potential issues associated with statistical analyses. The ZO descriptor depends on the chemical framework of the derivatives, as it identifies structural effects based on the residuals from predictions made using simple regression with the descriptor. As the four included compounds show a significant negative effect against the disease, with a ZO value coefficient of 396.1, they should be excluded.

Table 5 presents the theoretically estimated values of descriptors that belong to the model above (Eq. (2)) together with the values of residuals between the experimental and predicted values for both the training and test sets. Fig. 2 shows the relationship between the predicted and experimental HCT-116 values for the 29

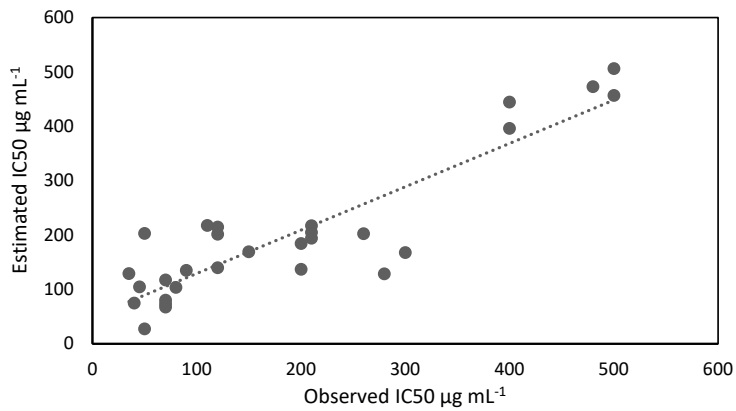


Fig. 2. Estimated versus observed IC_{50} values for the isatin derivatives in the training and test sets, obtained using Model (2). The model exhibited $r^2 = 0.802$, $q^2 = 0.791$, and a standard error (SE) of 68.67.

Table 5. The values of descriptors that are used in model (2) and the values of residual and estimated IC_{50}, µg mL^{-1}according to the suggested model.

No.	IC_{50}, µg mL^{-1}	E. isatin der., Ha*	exp (log S), log unit	E. bin Leu., Ha**	ZO	Residual, µg mL^{-1}	Estimated, µg mL^{-1}
Training set							
1	70	-1824.64	0.00375	13.6676	0	0.4930	69.507
2	70	-1935.08	0.00170	14.4903	0	-5.870	75.870
3	45	-1974.4	0.00130	14.8180	0	-51.471	96.471
4	40	-2239.79	0.00143	14.5507	0	-30.717	70.717
5	120	-2049.53	0.00077	15.2660	0	-5.952	125.952
6	50	-1919.06	0.00082	14.4170	0	22.038	27.962
7	35	-2415.68	0.00042	15.3120	0	-80.995	115.995
8	150	-2433.97	0.00065	15.5250	0	-1.176	151.176
10	120	-2509.11	0.00039	15.9568	0	-70.358	190.358
11	280	-2415.67	0.00043	15.3040	0	164.639	115.361
12	400	-2284.21	0.00191	14.3805	1	-48.778	448.778
13	500	-2394.65	0.00087	15.2100	1	-0.260	500.260
14	300	-2433.97	0.00066	15.5130	0	149.762	150.238
15	200	-2699.36	0.00073	15.2600	0	76.654	123.346
16	210	-2473.28	0.00050	15.8450	0	27.630	182.370
17	210	-2509.1	0.00039	15.9727	0	17.486	192.514
18	480	-2378.63	0.00042	15.1250	1	9.7870	470.213
19	70	-1953.86	0.00245	14.5000	0	-40.095	110.095
20	400	-1822.40	0.01089	13.5620	0	26.940	373.060
22	260	-1972.15	0.00377	14.6605	0	72.557	187.443
23	200	-2237.55	0.00415	14.3940	0	27.279	172.721
24	120	-2011.46	0.00287	14.9760	0	-64.784	184.784
25	210	-2047.29	0.00224	15.1450	0	33.237	176.763
26	500	-1916.81	0.00237	14.3090	1	39.250	460.750
27	70	-2284.22	0.0018826	14.3760	0	-0.329	70.329
$r = 0.919$, $r^2 = 0.844$, SE = 60.82							
Test set							
9	50	-2473.29	0.000496	16.5829	0	-152.898	202.898
21	110	-1932.84	0.004954	15.1069	0	-107.48	217.4797
28	90	-2699.37	0.000716	15.9919	0	-45.0935	135.0935
29	80	-2378.63	0.00041	15.8629	0	-23.6608	103.6608
SE = 136.97							

** The global energy minimum of isatin derivatives in Hartree units.*
*** The global energy minimum of leucine, which is equal to -441.546 Hartrees.*

isatin derivatives in both the training and test sets. The plot reveals a strong linear correlation, with r^2 equal to 0.802.

The statistical parameters of the square of the correlation coefficient (r^2) and cross-validation (q^2) for both the training set and the test set were equal to 0.802 and 0.791, respectively. The interesting value of q^2, which is positive and close to unity, indicates significant confidence of the developed models [35]. However, the main aim of this work is to attempt to find a suitable

model that can be used for predicting new, more effective anti-colorectal cancer agents, as previously made with anti-bacterial and pesticides [4, 5]. But as one can see here, the residuals of the test set are not that good, in addition to the fact that the correction term is not very distinguishable, as it only covers the compounds with low effect towards HCT-116. Thus, the overall results of this work indicate that isatin derivatives are unsuitable for use as an anti-colon cancer agent.

In an analogous manner to leucine, phenylalanine is employed to study molecular docking as another preferred amino acid as a key of the P53 protein. The obtained distances of close contact that belong to hydrogen bonds are shown in Table 6. Interestingly, the obtained model, as presented in Eq. (3), is so close to Eq. (2), which supports the present achievement of employing the binding energy in the QSAR model.

$$HCT - 116 = -2078 + 48301\exp(LogS) + + 136.1Ebind\ phen + 369.8\ ZO \quad (3)$$

Table 6. Close contact distances between isatin derivatives and phenylalanine.

No	Isatin d. + Phen.	Close contact, Å	No.	Isatin d. + Phen.	Close contact, Å
1	H37-O60 O8-O62	1.820 2.028	16	H50-O69 O9-H77	2.330 2.472
2	H47-N54 O35-H66	2.455 2.296	17	H40-O67 H56-N58	2.422 2.480
3	O9-H69	2.356	18	H50-N55 O8-H71 H51-O64	2.353 2.358 2.420
4	O34-H63	2.210	19	O8-H73 O29-H72 H35-O61 H40-O62	1.815 2.422 2.190 2.392
5	H42-N58	1.801	20	O8-H59 H36-O57 O8-H68	2.030 1.817 2.473
6	H37-O64	2.909	21	N12-H72 H39-O61 H35-N51	1.957 1.854 2.448
7	O8-H75 O27-H67	2.406 2.488	22	H41-O64 H45-N54	2.469 2.596
8	H46-O66 O8-H76	2.311 2.473	23	O29-H68 O26-H67	2.550 2.535
9	H40-O69 O8-H80	2.310 2.356	24	H51-O66 O8-H67	2.466 2.456
10	H46-O67 H54-N58	2.389 2.382	25	O9-H76 H40-O64	1.661 2.210
11	H41-O64 H37-O64 O8-H71	1.861 2.370 2.455	26	O9-H71	2.517
12	H38-O60 O9-H67	1.787 2.319	27	O9-H72 H38-O60	1.617 1.718
13	O8-H67	2.463	28	O8-H63 H40-O61	2.038 1.748
14	H43-O66 H45-O66	2.414 2.500	29	O8-H76 H41-O64	1.607 1.738
15	H47-O61 O8-H71	2.425 2.410			

where Ebind phen is the binding energy between isatin derivatives and phenylalanine.

Table **7** presents the theoretically estimated descriptor values for model (3), alongside the residual values between the experimental and predicted activities for both the training and test sets. Fig. 3 illustrates the relationship between the predicted and experimental HCT-116 values for the 29 isatin derivatives. The plot reveals a strong linear correlation, with an r^2 value of 0.804.

According to the literature, the model developed in the present study offers notable advantages from both statistical and physicochemical perspectives compared with previously reported models [22 - 24]. In contrast

Table 7. The values of descriptors that are used in model (3) and the values of residual and estimated IC_{50}, µg mL^{-1}according to the suggested model.

No.	IC_{50}, µg mL^{-1}	ex (log S), log unit	Ebin phen., Ha*	ZO	Residual, µg mL^{-1}	Estimated, µg mL^{-1}
Training set						
1	70	0.00375	14.4299	0	19.0920	50.9080
2	70	0.00170	15.2503	0	-6.6150	76.6150
3	45	0.00130	15.5283	0	-55.237	100.237
4	40	0.00143	15.3106	0	-87.192	127.192
5	120	0.00077	16.0029	0	-9.3520	129.352
6	50	0.00082	15.1839	0	23.134	26.8660
7	35	0.00042	16.0779	0	-48.334	83.3340
8	150	0.00065	16.2659	0	42.456	107.544
10	120	0.00039	16.7257	0	-17.284	137.284
11	280	0.00043	16.0664	0	141.26	138.742
12	400	0.00191	15.1454	1	-39.840	439.840
13	500	0.00087	15.9429	1	-10.034	510.034
14	300	0.00066	16.2769	0	122.610	177.389
15	200	0.00073	16.0149	0	29.424	170.576
16	210	0.00050	16.5909	0	-6.3120	216.312
17	210	0.00039	16.7243	0	-2.4490	212.449
18	480	0.00042	15.8889	1	1.1730	478.827
19	70	0.00245	15.2359	0	-56.482	126.482
20	400	0.01089	14.3009	0	42.019	357.981
22	260	0.00377	15.4414	0	60.793	199.207
23	200	0.00415	15.1899	0	5.648	194.352
24	120	0.00287	15.7489	0	-83.025	203.025
25	210	0.00224	15.8719	0	20.516	189.484
26	500	0.00237	15.0579	1	48.700	451.300
27	70	0.00188	15.1289	0	50.080	19.9200
$r = 0.920$, $r^2 = 0.846$, SE = 60.49						
Test set						
9	50	0.00049	16.5829	0	-152.881	202.881
21	110	0.00495	15.1069	0	-107.339	217.339
28	90	0.00072	15.9919	0	-43.069	133.069
29	80	0.00041	15.8629	0	-20.730	100.730
SE = 115.85						

**The global energy minimum of phenylalanine, which is equal to -554.630 Hartrees.*

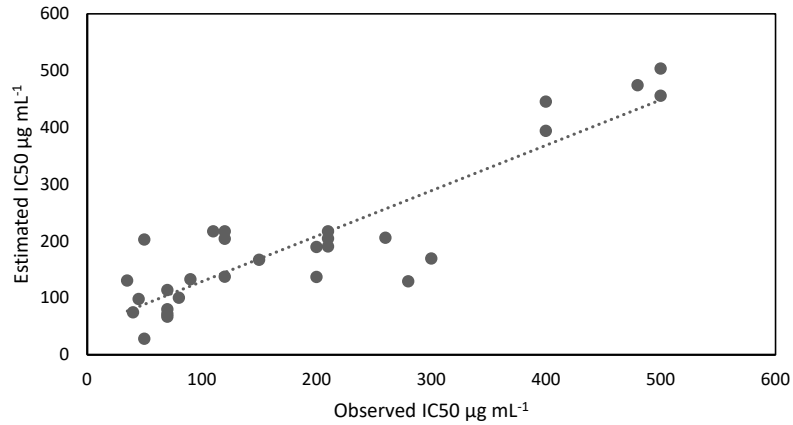


Fig. 3. Estimated versus observed IC_{50} values for the isatin derivatives in the training and test sets, obtained using Model (3). The model exhibited $r^2 = 0.804$, $q^2 = 0.800$, and a standard error (SE) of 67.04.

to many published QSAR models based mainly on descriptive correlations, the proposed model includes descriptors of evident physicochemical meaning. Such a feature not only improves the interpretability of the model but also provides useful insights into molecular factors that may influence the development and treatment of colorectal cancer.

CONCLUSIONS

The incorporation of quantum mechanics in molecular docking and the use of the resulting binding energies as descriptors in QSAR modelling represents a significant advancement over traditional classical mechanics. The integrated computational strategy combines the high accuracy of quantum mechanics calculations for electronic levels with the predictive ability of QSAR analysis, providing a robust and reliable platform for the identification and design of novel molecule inhibitors with better potential for cancer therapy.

In summary, the present study demonstrates that modern computational approaches can be of great value not only for the discovery of new compounds but also for reducing the use of traditional trial-and-error procedures in chemistry. In addition, the future integration of these approaches with artificial intelligence could allow for the identification of highly effective disease-specific compounds with greater precision and efficiency. Such advances are expected to revolutionize the drug discovery process by significantly cutting down the time, cost, environmental impact, and human labor involved.

Authors' contributions

A.R.M. carried out the theoretical and statistical work, performed data collection, and contributed to the initial drafting of the manuscript; R.A.K. supervised the research project, designed the theoretical and statistical methodology, and critically reviewed and finalized the manuscript.

REFERENCES

1. R.L. Siegel, N.S. Wagle, A. Cercek, R. A. Smith, A. Jemal, Colorectal cancer statistics, CA Cancer J. Clin., 73, 3, 2023, 233-254.
2. M. Díaz-Gay, W. Dos Santos, S. Moody, M. Kazachkova, A. Abbasi, C. D. Steele, R. Vangara, L. B. Alexandrov, Geographic and age variations in mutational processes in colorectal cancer, Nature 643, 8070, 2025, 230-240.
3. F. Baidoun, K. Elshiwy, Y. Elkeraie, Z. Merjaneh, G. Khoudari, M. T. Sarmini, M. Gad, M. Al-Husseini, & A. Saad, Colorectal Cancer Epidemiology: Recent Trends and Impact on Outcomes, Curr. Drug Targets, 22, 9, 2021, 998-1009.
4. R.A. Khalil, S.H. Abdulrahman, Newly developed statistically intensive QSAR models for biological activity of isatin derivatives, Studia Universitatis Babeș-Bolyai Chemia, 67, 1, 2022, 139-152.
5. R.A. Khalil, A.H.A. Zarari, In silico treatment for

prediction of new effective pesticides based on acetylcholinesterase inhibition, Doklady Chemistry, 518, 1-2, 2024, 154-165.
6. M.K. Qasim, R.A. Khalil, Employment of in silico in prediction of furan derivatives as efficient corrosion inhibitors for mild steel in acidic media, Chemistry Journal of Moldova, 20, 2, 2025, 84-93.
7. M. Kumar, K. Ramasamy, V. Mani, R. K. Mishra, A. B. A. 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 Journal of Chemistry, 7, 4, 2014, 396-408.
8. R.E. Ferraz de Paiva, E.G. Vieira, D. Rodrigues da Silva, C. A. Wegermann, A. M. Costa Ferreira, Anticancer compounds based on isatin-derivatives: strategies to ameliorate selectivity and efficiency, Front. Mol. Biosci., 7, 2021, 627272.
9. Varun, Sonam, R. Kakkar, Isatin and its derivatives: a survey of recent syntheses, reactions, and applications, Medchemcomm 10, 3, 2019, 351-368.
10. O. Fedorova, A. Daks, V. Petrova, A. Petukhov, L. Lezina, O. Shuvalov, P. Davidovich, D. Kriger, E. Lomert, D. Tentler, V. Kartsev, B. Uyanik, V. Tribulovich, O. Demidov, G. Melino, N. A. Barlev, Novel isatin-derived molecules activate p53 via interference with Mdm2 to promote apoptosis, Cell Cycle, 17,15, 2018, 1917-1930.
11. Y.A. Shahab, R.A. Khalil, A new approach to NMR chemical shift additivity parameters using simultaneous linear equation method, Spectrochim. Acta A Mol. Biomol. Spectrosc., 65, 2, 2006, 265-270.
12. N. Estrada-Ortiz, C.G. Neochoritis, A. Dömling, How to Design a Successful p53-MDM2/X Interaction Inhibitor: A Thorough Overview Based on Crystal Structures, ChemMedChem, 11, 8, 2016, 757-772.
13. A. Beloglazkina, N. Zyk, A. Majouga, E. Beloglazkina, Recent small-molecule inhibitors of the p53–MDM2 protein-protein interaction, molecules, 25, 5, 2020, 1211.
14. Y.A. Ivanenkov, S. V. Vasilevski, E.K. Beloglazkina, M. E. Kukushkin, A. E. Machulkin, M. S. Veselov, N. V. Chufarova, E. S. Chernyaginab, A. S. Vanzcool, N. V. Zyk, D. A. Skvortsov, A. A. Khutornenko, A. L. Rusanov, A. G. Tonevitsky, O. A. Dontsova, A. G. Majouga, Design, synthesis and biological evaluation of novel potent MDM2/p53 small-molecule inhibitors, Bioorg. Med. Chem. Lett., 25, 2, 2015, 404-409.
15. C.J. Gohil, M.N. Noolvi, Design and in silico study of the novel small molecular mdm2 inhibitors, Biointerface Res. Appl. Chem., 11, 1, 2021, 8052-8064.
16. P. Davidovich, V. Aksenova, V. Petrova, D. Tentler, D. Orlova, S. Smirnov, V. Gurzhiy, A. L. Okorokov, A. Garabadzhiu, G. Melino, N. Barlev, V. Tribulovich, Discovery of novel isatin-based p53 inducers, ACS Med. Chem. Lett., 6, 8, 2015, 856-860.
17. X. Xie, S.-S. Xiong, X. Li, H. Huang, F.-B. Wu, P.-F. Shen, C. Peng, G. He, B. Han, Design and organocatalytic synthesis of spirooxindole–cyclopentene–isoxazole hybrids as novel MDM2–p53 inhibitors, Organic Chemistry Frontiers, 8, 8, 2021, 1836-1843.
18. M.A.Vaali-Mohammed, M.H. Abdulla, S. Matou-Nasri, W. M. Eldehna, M. Meeramaideen, E. B. Elkaeed, M. El-Watidy, N. S. Alhassan, K. Alkhaya, O. Al Obeed, The anticancer effects of the pro-apoptotic benzofuran-isatin conjugate (5a) are associated with p53 upregulation and enhancement of conventional chemotherapeutic drug efficiency in colorectal cancer cell lines, Front. Pharmacol., 13, 2022, 923398.
19. Y.A. Ivanenkov, M.E. Kukushkin, A.A. Beloglazkina, R.R. Shafikov, A.A. Barashkin, A.A. Ayginin, M.S. Serebryakova, A.G. Majouga, D.A. Skvortsov, V.A. Tafeenko, E.K. Beloglazkina, Synthesis and biological evaluation of novel dispiro-indolinones with anticancer activity, Molecules, 28, 3, 2023, 1325.
20. D. Khylyuk, S. Holota, N. Finiuk, R. Stoika, T. Rumynska, R. Lesyk, Design and optimization of spiro-isatin-thiazolidinone hybrids with promising anticancer activity, Pharmaceuticals, 18, 10, 2025, 1502.
21. V. Khanna, G. Eslami, R. Reyes, R. Diep, S. Fernandez-Pol, H. Stehr, C.J. Suarez, H. Pinto, J.M. Ford, T.Y. Zhang, C.T. Chen, MDM2 inhibition is associated with the emergence of TP53-altered clonal hematopoiesis, NPJ Precis. Oncol., 9, 1, 2025, 34.
22. A.S. Girgis, J. Stawinski, N.S.M. Ismail, H. Farag, Synthesis and QSAR study of novel cytotoxic spiro[3H-indole-3,2′(1′H)-pyrrolo[3,4-c]pyrrole]-2,3′,5′(1H,2′aH,4′H)-triones, Eur. J. Med. Chem.,

47, 2012, 312-322.
23. K. Tabti, A. Sbai, L. Elmchichi, H. Maghat, T. Lakhlifi, M. Bouachrine, Computational investigation of pyrrolidin derivatives as novel GPX4/MDM2–p53 inhibitors using 2D/3D-QSAR, Molecular docking, Molecular dynamics simulations and MM-GBSA free energy, 33, 4, 2022, 1019-1039.
24. S. Kaur, J. Kaur, B.A. Zarger, N. Islam, N. Mir, Quantitative structure-activity relationship and ADME prediction studies on series of spirooxindoles derivatives for anti-cancer activity against colon cancer cell line HCT-116, Heliyon, 10, 16, 2024, e35897.
25. S.P.C. Kurmi, S. Thapa, D. Karati, Molecular docking and pharmacokinetic evaluations of curcumin-based scaffolds as MDM2-p53 inhibitors, Discover Chemistry, 2, 1, 2025, 53.
26. R.A. Khalil, S.H. Abdulrahman, A developed QSPR model for the melting points of isatin derivatives, Turkish Computational and Theoretical Chemistry, 6, 1, 2022, 1-8.
27. H. Zhu, H. Gao, Y. Ji, Q. Zhou, Z. Du, L. Tian, Y. Jiang, K. Yao, Z. Zhou, Targeting p53–MDM2 interaction by small-molecule inhibitors: learning from MDM2 inhibitors in clinical trials, J. Hematol. Oncol., 15, 1, 2022, 91.
28. E.R. Kastenhuber, S.W. Lowe, Putting p53 in Context, Cell, 170, 6, 2017, 1062-1078.
29. C.W. Chee, N.H. Zamakshshari, V.S. Lee, I. Abdullah, R. Othman, Y.K. Lee, N.M. Hashim, N.N. Rashid, Morindone from Morinda citrifolia as a potential antiproliferative agent against colorectal cancer cell lines, PLoS One, 17, 7, 2022, e0270970.
30. N.J. Basha, R.M. Mohan, Insight on heterocycles as p53-MDM2 protein-protein interaction inhibitors: molecular mechanism for p53 activation, ChemistrySelect, 9, 10, 2024, e202304525.
31. N.A. Ghafoor, A. Yildiz, Targeting MDM2-p53 axis through drug repurposing for cancer therapy: A multidisciplinary approach, ACS Omega, 8, 38, 2023, 34583-34596.
32. C.A. Menéndez, S.R. Accordino, D.C. Gerbino, G.A. Appignanesi, hydrogen bond dynamic propensity studies for protein binding and drug design, PLoS One, 11, 10, 2016, e0165767.
33. P.H. Kussie, S. Gorina, V. Marechal, B. Elenbaas, J. Moreau, A.J. Levine, N.P. Pavletich, Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain, Science, 274, 5289, 1996, 948-953.
34. M. Mohanty, P.S. Mohanty, Molecular docking in organic, inorganic, and hybrid systems: a tutorial review, Monatshefte Für Chemie - Chemical Monthly, 154, 7, 2023, 683-707.
35. A. Tropsha, Best practices for QSAR model development, validation, and exploitation, Mol. Inform., 29, 6 -7, 2010, 476-488.