

## **Original Research Article**

### **Isatin analogues as anti-bacterial agents; QSAR, Docking**

#### **Abstract:**

Computational chemistry is a unique method in drug discovery which reduces cost. In this study 86 molecules containing isatin core were subjected to quantitative structure-activity relationship analysis and docking study to find the structure requirements for ligand binding. The structures were sketched and optimized in Hyperchem. The structural invariants used in this study were those obtained from whole molecular structures: by both hyperchem and dragon software (16 types of descriptors). Four chemometrics methods including MLR, FA-MLR, PCR and GA-PLS were employed to make the connection between structural parameters and anticancer effects. MLR analysis explained the positive effect of the number of urea derivatives, thiourea, amide, thioamide, hydrazone, thiocarbohydrazone, nBnz with the halogen substitution on 5 position of isatin ring on the antimicrobial activity. It also shows nArCN, nPyridines have negative effects on the antimicrobial activity of studied compound. The FA-MLR describes the effect of 3D-MORSE and Galvez Topological charge descriptors and on antimicrobial activity of the studied compounds. The quality of PCRA equation is better than those derived from FA-MLR. A comparison between the different statistical methods employed revealed that GA-PLS represented superior results and it could explain and predict 73% and 68% of variances in the  $-\text{LogMIC}$  data, respectively. Comparison between QSAR and docking analysis revealed that by decreasing in a number of ring and lipophilicity (also  $\text{Logp}$ ) for the design of new compounds can have better activity. Substitutions such as urea, thiourea, thiocarbohydrazone, benzhydrazide as on isatin ring, can cause better interaction with receptor.

**Keywords:** Isatin, QSAR, Docking, antibacterial

## Introduction:

Fatty acid synthesis in bacteri (FabI) was confirmed to be the only enoyl-acyl carrier protein (ACP) reductase required for the synthesis of fatty acids. FabI is required for the elongation of both long-chain saturated and unsaturated fatty acids in *Escherichia coli*. Thus, the activity of this enzyme plays a determinant role in completing cycles of fatty acid biosynthesis in *E. coli*. Indole compounds can inhibit FabI [1]. One of the best promising new indole compounds having many interesting activity profiles is isatin and isatin derivatives. The isatin (1*H*-indole-2,3-dione) moiety is responsible for a wide spectrum of biological property such as antibacterial, antifungal, antiviral, anticancer, anticonvulsant, anti-HIV and antiparkinsonian activity in many synthetically versatile molecules [2–9]. Among these properties antibacterial activities (against *Escherichia coli*) of this moiety was of our interest to study the quantitative structure-activity relationships of a series of 86 isatin derivatives reported in literature. Synthesis and evaluation of the biological activity of novel compound usually time consuming and take large amounts of money. Today, the use of computational methods for designing newly biologically active compounds have opened a new window to modern drug discovery study. Computational methods can accelerate the procedure of discovering new drugs by designing new compounds and predicting potency or activity of them. Quantitative structure-activity relationship (QSAR) studies provide pharmaceutical chemists valuable information that is useful for drug design and prediction of drug activity [10-14]. QSAR studies, as one of the most important areas in chemometrics, give information that is useful for molecular design and medicinal chemistry [8-12]. QSAR models are mathematical equations constructing a relationship between the chemical scaffold and biological property. These models have another ability, which is providing a deeper knowledge of molecule design.

Linear and nonlinear QSAR models are mathematical equations that display us enough information about the mechanism of biological activity of compounds by constructing a relationship between chemical structures and biological activities. The first step in constructing QSAR models is the proper representation of the structural and physicochemical features of chemical compounds [15-18]. These features named molecular descriptors that represent variation in the structural features property of the molecules by numerical and have a high effect on the biological property of

compound [19-22]. Molecular descriptors have been classified into different categories such as physicochemical, constitutional, geometrical, topological, and quantum chemical descriptors. Dragon and hyperchem are two famous computational software provide us more than 4000 of these descriptors [23,24].

Different QSAR methods including multiple linear regression (MLR), partial least squares combined with genetic algorithm for variable selection (GA-PLS), factor analysis–MLR (FA-MLR), principal component regression analysis (PCR) were used to make connections between structural descriptors and antibacterial activity of studied compounds [25-28]. An important approach of the researchers in the modification of the isatin moiety has been to establish a comprehensive structure–activity relationship (SAR), for this class of antibacterial agents (against *Escherichia coli*).

The study shows that a series of compounds didn't evaluate for QSAR studies. Our Different QSAR analysis establishes mathematical relationship between biological activities and computable parameters such as chemical, topological, physicochemical, stereochemical or geometrical and so on indices [29].

The molecular docking studies help us to understand the different interactions between the ligands and enzyme active sites (FabI) in detail and also help to design novel potent structure. Molecular docking simulation technique was also performed on eighty six compounds to show the details molecular binding models for these compounds interacting with the key active site of protein.

## 2. Methods

### 2.1. Data set

The biological data used in this study were antimicrobial activity against *E. coli*, (in terms of  $-\log$  MIC), of a set of 86 isatin derivatives [30-35]. Firstly, outlier data was removed from the data set according to principle component analysis, it was shown in Figure 1. The data set was classified into calibration and prediction set by kenardston algorithm of the 16 prediction molecules from the spaces of the calculated descriptors. The structural features and biological activity of these compounds are listed in Table 1.

[Table 1. near here], [Figure 1 near here]

### 2.2. Descriptor generation

The structural features of the studied compounds are listed in Table 1. The two-dimensional structures of molecules were drawn by Hyperchem 8.0 software (Hypercube Inc.) to calculate whole molecular structure-based descriptors. The final geometries were obtained with semi-empirical AM1 calculations in Hyperchem program. The molecular structures were optimized using the Polak-Ribiere algorithm until the root mean square gradient was  $0.01 \text{ kcal mol}^{-1}$  [23]. Some physicochemical parameters including molecular volume (V), molecular surface area (SA), hydrophobicity (Log P), hydration energy (HE) and molecular polarizability (MP) were calculated using Hyperchem Software. To calculate some molecular descriptors including topological, constitutional and functional group descriptors the optimized molecules were transferred into the Dragon package, developed by the Milano chemometrics and QSAR Group [24]. The calculated descriptors from whole molecular structures are briefly described in Table 2.

[Table 2. near here]

### 3.2. Data screening & model building

The selected descriptors from each class and the experimental data were analyzed by the stepwise regression SPSS (version 22.0) software. The calculated descriptors were collected in a data matrix whose number of rows and columns were the number of molecules and descriptors, respectively. Multiple linear regressions (MLR) and partial

least squares (PLS) were used to derive the QSAR equations and feature selection was performed by the use of genetic algorithm (GA). MLR with factor analysis as the data pre-processing step for variable selection (FA-MLR) and principal component regression analysis (PCRA) methods were also used to explore the QSAR equations.

The resulted models were validated by leave-one out cross-validation procedure (using MATLAB software) to check their predictability and robustness.

A key step in QSAR modelling is evaluating model's stability and prediction ability. We used cross-validation and external test set for these proposes. Cross-validation has different variants such as leave-one-out (LOO), leave-group-out (LGO) and v-fold. It was shown previously that LOO can lead to chance and overfitted models whereas LGO is more sensitive to chance variables [36]. Therefore, we used LGO for model-validation utilizing correlation coefficient and root mean square error of cross-validation ( $q^2$  and  $RMSECV$ , respectively) as the scoring function. In addition, an external test set composed of 6 molecules was also used. The molecules in this set did not have a contribution in the model step and thus their predicted values can give a final prediction power of the models as measured by correlation coefficient, root mean square errors of prediction, the relative error of prediction ( $R^2_p$ ,  $RMSEP$  and  $REP$ , respectively).

The PLS regression method used in this study was the NIPALS-based algorithm existed in the chemometrics toolbox of MATLAB software (version 12 Math work Inc.). Leave-one-out cross-validation procedure was used to obtain the optimum number of factors based on the Haaland and Thomas F-ratio criterion [37].

### **3. Results and discussion**

#### **3.1. MLR analysis**

In the first step, separate stepwise selection-based MLR analyses were performed using different types of descriptors, and then, an MLR equation was obtained utilizing the pool of all calculated descriptors. The resulted QSAR models from different types of descriptors for the compounds (86 molecules as calibration and 16 molecules as prediction sets) are listed in Table 3.

[Table 3. near here]

The equation E1 of Table 3 shows among chemical descriptors, the negative effect of log p of the molecules on the antimicrobial activity. This equation shows the hydrophilic molecules shows better antimicrobial effect. The second equation of Table 3 demonstrated the effect of constitutional descriptors on the activity of these compounds. It also explains the negative effects nCIC (number of rings), and nR10 (number of 10-membered rings) on activity (such as molecule series 20-39, 49, 55, 65 have intermediate activity and inactive compounds).

The effect of topological group counts parameter on antimicrobial activity of the studied compounds has been described by equation E<sub>3</sub> of Table 3. It shows that among topological descriptors spanning tree number (STN) has a negative effect on cytotoxic activity of the compounds.

The equation E<sub>4</sub> of Table 3 was found by using Mol-Walk descriptors (E<sub>4</sub>), which explains the negative effect of PIPCO9 of studied compounds on the activity of the compounds. The equation E<sub>5</sub>-E<sub>16</sub> and E<sub>16</sub> of Table 3 demonstrated the effect of positive and negative effects of BCUT, Galvz topological Charge indices, 2D autocorrelations, Charge, Burden eigenvalues, RDF, 3D MoRSE, WHIM, GETAWAY and charge descriptors on the anti-cancer activity of these compounds.

The MLR equation of Table 3 obtained from the pool of functional groups descriptors, E<sub>17</sub>, explained the positive effect of the number of urea derivatives (nCONN), thiourea (nCSNN) (such as molecules of 1-9, 69-86), amide(nCONH2), thioamide (nRCONHR), hydrazone (nC=N-N), thiocarbohydrazone (such as molecules of 1-9) on the antimicrobial activity. It also indicates the positive effect of nBnz (number of benzene rings) with the halogen substitution on 5 position of isatin ring F > Cl > Br (nArX) (molecules series 14-19). This equation also shows nArCN (Aromatic nitrile such as compounds 22, 25, 28, 31, 34, 37, 39), nPyridines (pyridine derivatives) have negative effects on the antimicrobial activity. Maybe there isn't any electron withdrawing group in the receptor site, thus for design of new compounds it's better to don't use these substitution on the backbone of compounds. The negative sign of this group proposed that a decrease in the number of these descriptors resulted in an activity enhancement. This equation, which has a high statistical quality ( $R^2 = 0.63$ ,  $Q^2 = 0.59$ ).

The statistical parameters of prediction, listed in Table 4, indicate the suitability of the proposed QSAR model based on MLR analysis of molecular descriptors. The

correlation coefficient of prediction is 0.62, which means that the resulted QSAR model could predict 62% of variances in the antimicrobial activity data. It has root mean square error of 0.21.

[Table 4. near here]

### 3.2. GA-PLS model

Multicollinearity is a real problem in MLR analysis. This problem in the descriptors is omitted by PLS analysis. In fact, in PLS analysis, the descriptors data matrix is decomposed to orthogonal matrices with an inner relationship between the dependent and independent variables. This modeling method coincides with noisy data better than MLR, because a minimal number of latent variables are used for modeling in PLS. In GA-PLS analysis a variable selection method is used to find the more convenient set of descriptors because redundant variables degrade the performance of PLS analysis, similar to other regression methods. In the present study, GA was used as variable selection method. The data set ( $n = 86$ ) was divided into two groups: calibration set ( $n = 70$ ) and prediction set ( $n = 16$ ). Given 70 calibration samples; cross-validation procedure was used to find the optimum number of latent variables for each PLS model. In this work, in each run of GA-PLS method a large number of acceptable models were created. GA produces a population of acceptable models in each run. In this work, many different GA-PLS runs were conducted using different initial set of populations (50-250) and therefore a large number of acceptable models were created. The most convenient GA-PLS model that resulted in the best fitness contained 8 descriptors including, one constitutional descriptor (nCIC), one 3D MoRSE descriptors (MOR30M) parameter, two WHIM descriptor (P1P, E1U) and four functional descriptors (nRCONHR, nCONN, nArCN, nPyridines). The majority of these descriptors are functional indices All of them being those obtained by different MLR-based QSAR models. The PLS estimate of the regression coefficients are shown in Figure 2.

This model not only has a high cross-validation statistics, but also represents a high ability for modeling external test samples. It could explain and predict about 75% of variances in the antimicrobial activity (against *E. coli*) of the studied molecules. There is a close agreement between the experimental and predicted values of antimicrobial activity data.

To measure the significance of the 8 selected PLS descriptors in the protein tyrosine kinase inhibitory activity; In order to investigate the relative importance of the variable appeared in the final model obtained by GA-PLS method, variable important in projection (VIP) was employed [38]. VIP values reflect the importance of terms in PLS model. According to Erikson *et al.* X-variables (predictor variables) could be classified according to their relevance in explaining y (predicted variable), so that  $VIP > 1.0$  and  $VIP < 0.8$  mean highly or less influential, respectively, and  $0.8 < VIP < 1.0$  means moderately influential. The VIP analysis of PLS equation is shown in Figure 3. As it is observed, nRCONHR, P1P and E1U indices represent the most significant contribution in the resulted QSAR model. In addition, parameters such as nArCN and MOR30M have been found to be moderately influential parameters.

[Figure 2. Near here], [Figure 3. Near here]

### 3.3. FA-MLR and PCRA

FA-MLR was performed on the dataset. Factor analysis (FA) was used to reduce the number of variables and to detect structure in the relationships between them. This data-processing step is applied to identify the important predictor variables and to avoid collinearities among them [39]. Principle component regression analysis, PCRA, was tried for the dataset along with FA-MLR. With PCRA collinearities among **X** variables are not a disturbing factor and the number of variables included in the analysis may exceed the number of observations [40]. In this method, factor scores, as obtained from FA, are used as the predictor variables [39]. In PCRA, all descriptors are assumed to be important while the aim of factor analysis is to identify relevant descriptors.

Table 5 shows the nine factor loadings of the variables (after VARIMAX rotation) for the compounds tested for cytotoxic activity. As it is observed, about 81.2% of variances in the original data matrix could be explained by the selected nine factors.

Based on the procedure explained in the experimental section, the following two-parametric equation was derived (Table 6).

$$Y = 5.766(\pm 0.547) - 0.031(\pm 0.012) \text{MOR30m} - 2.336(\pm 0.721) \text{JGI5}$$

$$R^2 = 0.64 \quad S.E = 0.24 \quad F = 14.69 \quad Q^2 = 0.59 \quad RMS_{cv} = 0.12$$



This equation could explain about 59% of the variance and predict 64% of the variance in pMIC data. It has a root mean square error of 0.12. This equation describes the effect of 3D-MORSE and Galvez Topological charge descriptors (MOR30m and JGI5) and on antimicrobial activity of the studied compounds.

When factor scores were used as the predictor parameters in a multiple regression equation using forward selection method (PCRA), the following equation was obtained (Table 7):

$$Y=4.285(\pm 0.57)-0.258(\pm 0.031)F1-.189(\pm 0.021)F7+0.124(\pm 0.027)F6+0.123(0.057)F8$$
$$R^2 = 0.73 \quad S.E. = 0.34 \quad F = 15.54 \quad Q^2 = 0.68 \quad RMScv = 0.15$$

This equation could explain and predict 68% and 73% of the variances in pMIC data, respectively. The root mean square error of PCRA analysis was 0.15. Since factor scores are used instead of selected descriptors, and any factor-score contains information from different descriptors, loss of information is thus avoided and the quality of PCRA equation is better than those derived from FA-MLR. Whilst the data of this analysis show acceptable prediction, we see that the predicted values of some molecules are near to each other.

As it is observed from Table 5, in the case of each factor, the loading values for some descriptors are much higher than those of the others. These high values for each factor indicate that this factor contains higher information about which descriptors. It should be noted that all factors have information from all descriptors but the contribution of descriptor in different factors are not equal. For example, factors 1 and 2 have higher Constitutional, Charge, WHIM, Atom-center, Connectivity, Functional, MORSE and GETAWAY whereas information about RDF, MORSE, burden eigenvalues 2Dautocorrelations and functional descriptors are highly incorporated in factor 3 and 4. Factor score 5, 6, 7 and 8 signify the importance of functional chemical and Atom-center descriptors.

[Table 5 near here], [Table 6 near here], [Table 7 near here]

### 3.4. Robustness and applicability domain of the models

Leverage is one of standard methods for this purpose. Warning leverage ( $h^*$ ) is another criterion for interpretation of the results. The warning leverage is, generally, fixed at  $3k/n$ , where  $n$  is the number of training compounds and  $k$  is the number of model parameters. A leverage greater than warning leverage  $h^*$  means that the predicted response is the result of substantial extrapolation of the model and therefore may not be reliable [41]. The calculated leverage values of the test set samples for different models and the warning leverage, as the threshold value for accepted prediction, are listed in Table 8. As seen, the leverages of all test samples are lower than  $h^*$  for all models. This means that all predicted values are acceptable.

[Table.8 near here]

### 3.5. Molecular Docking Studies

The docking study was performed using the AutoDock 4.2. All the eighty six isatin derivatives were docked into the active site of the enzymes FabI (PDB:1lx6). All the docking protocols were done on validated structures, with RMSD values below 2 Å. The conformation with the lowest ones was considered as the best docking result. Docking binding energies of these active compounds were summarized in Table 9. Docking analysis showed that Compounds 1-9 with thiocarbazone moiety, was good inhibitor for FabI, because of good interaction between enzyme and cofactor. With suitable orientation of thiocarbazone group, hydrogen and hydrophobic bounds can occur. An electron rich group such as  $\text{NH}_2$  substitution on phenyl ring that increase electron charge can create better interaction with receptor and has low binding energy. Also compounds have benzhydrazide substitution hydrogen binding interaction between tyrosine 156 and benzhydrazide. Halogen and methyl substitution on isatin ring of this series (1-9) can cause better interaction with receptor. Benzyl amide, methyl and halogen groups on isatin ring of compound 10-19 showed good docking score. But Halogen and methyl groups at C-5 isatin ring of series (40-43) can cause bad interaction with receptor. Urea moiety can show good interaction between tyrosine 156 and coenzyme. Our results indicated Benzamide group show good interaction but OH group has bad interaction with receptors. Compounds 47 and 48 with isonicotine amide group on isatin ring showed good interaction with receptor. The interaction modes of 2, 14 and 27, 40 those with the best docking score are shown in Figure 4.

[Table 9 near here], [Figure 4 near here]

#### 4. Conclusions

Quantitative relationships between molecular structure and antibacterial activity of isatin derivatives were discovered by four chemometrics methods: MLR, GA-PLS, PCR and FA-MLR. MLR analysis explained the positive effect of the number of urea derivatives (nCONN), thio urea (nCSNN), amide (nCONH<sub>2</sub>), thioamide (nRCONHR), hydrazone (nC=N-N), thiocarbohydrazone on the antimicrobial activity. It also indicate the positive effect of nBnz (number of benzene rings) with the halogen substitution on 5 position of isatin ring F > Cl > Br (nArX). This equation also shows nArCN, nPyridines (pyridine derivatives) have negative effects on the antimicrobial activity of studied compound. The FA-MLR describes the effect of 3D-MORSE and Galvez Topological charge descriptors (MOR30m and JGI5) and on antimicrobial activity of the studied compounds. The quality of PCRA equation is better than those derived from FA-MLR. Whilst the data of this analysis show acceptable prediction, we see that the predicted values of some molecules are near to each other. Factors 1 and 2 have higher Constitutional, Charge, WHIM, Atom-center, Connectivity, Functional, MORSE and GETAWAY whereas information about RDF, MORSE, burden eigenvalues 2Dautocorrelations and functional descriptors are highly incorporated in factor 3 and 4. Factor score 5, 6, 7 and 8 signify the importance of functional chemical and Atom-center descriptors. A comparison between the different statistical methods employed revealed that GA-PLS represented superior results and it could explain and predict 73% and 68% of variances in the -LogMIC data, respectively. Comparison between QSAR and docking analysis revealed that by decreasing in number of ring and lipophilicity (also logP) for design of new compounds can have better activity. Substitutions such as urea, thiourea, thiocarbohydrazone, benzhydrazide as on isatin ring, can cause better interaction with receptor.

## References

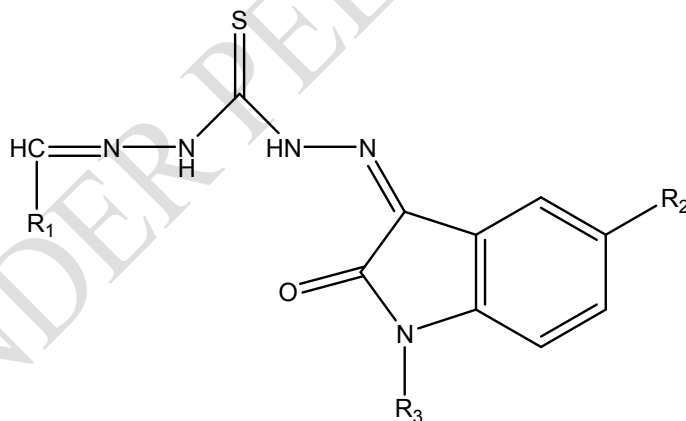
1. Heath RJ, Rubin JR, Holland DR, Zhang E, Snow ME, Rock CO. Mechanism of triclosan inhibition of bacterial fatty acid synthesis, *J Biol Chem.* 1999,16; 274(16): 11110-4.
2. S.N. Pandeya, S. Smitha, M. Jyoti, S.K. Sridhar, *Acta Pharm.* 55, (2005) 27-46.
3. V.M. Sharma, P.Prasanna, V.A. Seshu, B. Renuka, V.L. Rao, G.S. Kumar, C.P. Narasimhulu, P.A. Babu, R.C. Puranik, D. Subramanyam, A. Venkateswarlu, S. Rajagopal, K.B.S. Kumar, C.S. Rao, N.V.S. R. Mamidi, D.S. Deevi, R. Ajaykumar, R. Rajagopalan, *Bioorg. Med. Chem. Lett.* 12, (2002) 2303-2307.
4. M. Verma, P.S. Nath, K. Nandsingh, J.P. Stables, *Acta Pharm.* 54 (2004) 49-54.
5. S.N. Pandeya, D. Sriram, G. Nath, E. De Clercq, *Eur. J. Med. Chem.* 35 (2000)249-255.
6. N. Igosheva, C. Lorz, E. O'Connor, V. Glover, H. Mehmet, *Neurochem. Int.* 47 (2005) 216-224
7. M.J. Moon, S.K. Lee, J.-W. Lee, W.K. Song, S.W. Kim, J.I. Kim, C. Cho, S.J. Choi, Y.-C. Kim, *Bioorg. Med. Chem.* 14, (2006) 237-246.
8. A.H. Abadi, S.M. Abou-Seri,; D.E. Abdel-Rahman, C. Klein, O. Lozach, L. Meijer, *Eur. J. Med. Chem.* 41, (2006) 296-305.
9. A. Gursoy, N. Karali, *Eur. J. Med. Chem.* 38, (2003) 633-643.
10. H. Schmid, Multivariate prediction for QSAR, *Chemom. Intell. Lab. Syst.* 37 (1997) 125-134.
11. C. Hansch, A. Kurup, R. Garg, H. Gao, Chem-bioinformatics and QSAR: A review of QSAR lacking positive hydrophobic terms, *Chem. Rev.* 101(2001) 619-672.
12. S. Wold, J. Trygg, A. Berglund, H. Antti, Some recent developments in PLS modeling, *Chemom. Intell. Lab. Syst.* 58 (2001) 131-150.
13. Sabet R.; Fassihi A.; Hemmateenejad B.; Saghaie L.; Miri R.; Gholami M.; Computer-aided drug design of novel antibacterial 3-hydroxypyridine-4-ones: application of QSAR methods based on the MOLMAP approach. *Journal of Computer-Aided Molecular Design.* 2012, 26,349-361.

14. Sabet, R.; Fassihi, A.; Moeinifard, B., QSAR study of PEST Derivatives as Potent HIV-Reverse Transcriptase Inhibitors. *J. Mol. Graph & Model.* 2009, 28, 146-155.
15. Razihi Sabet, Soghra Khabnadideh, Dara Fathi, Leila Emami, QSAR and docking study of isatin analogues As cytotoxic agent, *Journal of Pharmaceutical Research International. Soc.* 27(5): 1-22, 2019.
16. J. Wang, L. Zhang, G. Yang, C.G. Zhan, Quantitative structure-activity relationship for cyclic imide derivatives of protoporphyrinogen oxidase inhibitors: A study of quantum chemical descriptors from density functional theory, *J. Chem. Inf. Comput. Sci.* 44 (2004) 2099-2105.
17. C. Hansch, D. Hoekman, H. Gao, Comparative QSAR: Toward a deeper understanding of chemobiological interactions, *Chem. Rev.* 96 (1996) 1045-1075.
18. R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors.* Wiley-VCH, Weinheim, 2000.
19. D. Horvath, B. Mao, Neighborhood behavior. Fuzzy molecular descriptors and their influence on the relationship between structural similarity and property similarity, *QSAR Comb. Sci.* 22 (2003) 498-509.
20. S. Putta, J. Eksterowicz, C. Lemmen, R. Stanton, A novel subshape molecular descriptor, *J. Chem. Inf. Comput. Sci.* 43 (2003) 1623-1635.
21. S. Gupta, M. Singh, A.K. Madan, Superpendent index: A novel topological descriptor for predicting biological activity. *J. Chem. Inf. Comput. Sci.* 39 (1999) 272-277.
22. V. Consonni, R. Todeschini, M. Pavan, Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 2. Application of the novel 3D molecular descriptors to QSAR/QSPR studies, *J. Chem. Inf. Comput. Sci.* 42(2002) 693-705.
23. HyperChem, Release 8.0 for Windows, Molecular Modeling System: HyperCube.
24. Todeschini, R. Milano Chemometrics and QSAR Group. <http://michem.disat.unimib.it/>.
25. Fassihi, A.; Sabet, R., QSAR Study of p56<sup>lck</sup> Protein Tyrosine Kinase Inhibitory Activity of Flavonoid Derivatives Using MLR and GA-PLS. *Int. J. Mol. Sci.* 2008, 9, 1876-1892.

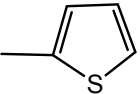
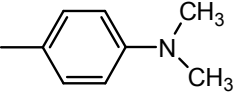
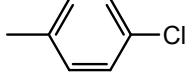
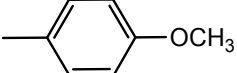
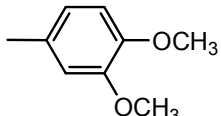
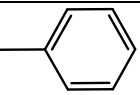
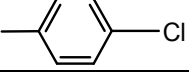
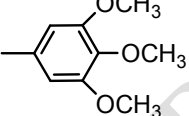
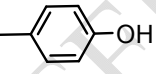
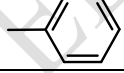
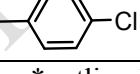
26. Sabet, R.; Fassihi, A., QSAR Study of Antimicrobial 3-Hydroxypyridin-4-one and 3-Hydroxypyran-4-one Derivatives Using different Chemometric Tools. *Int. J. Mol. Sci.* 2008, 9, 2407-2423.
27. Fassihi, A.; Abedi, D.; Saghaie, L.; Sabet, R.; Fazeli, H.; Bostaki, Gh.; Deilami, O.; Sadinpour, H., Synthesis, Antimicrobial Evaluation and QSAR Study of Some 3-hydroxypyridine-4- one and 3-hydroxypyran-4-one Derivatives. *Eur. J. Med. Chem.* 2009, 44, 2145-2157.
28. V. Consonni, R. Todeschini, M. Pavan, J. Chem. Inf. Comput. Sci. 42 (2002) 693-705.
29. Thanh ND, Giang NTK, Quyen TH, Huong DT, Toan VN. Synthesis and evaluation of in vivo antioxidant, in vitro antibacterial, MRSA and antifungal activity of novel substituted isatin N-(2, 3, 4, 6-tetra-O-acetyl- $\beta$ -d-glucopyranosyl) thiosemicarbazones. *Eur J Med Chem.* 2016;123:532-43.
30. Tehrani KHME, Hashemi M, Hassan M, Kobarfard F, Mohebbi S. Synthesis and antibacterial activity of Schiff bases of 5-substituted isatins. *Chin Chem Lett.* 2016;27(2):221-5.
31. Sridhar SK, Saravanan M, Ramesh A. Synthesis and antibacterial screening of hydrazones, Schiff and Mannich bases of isatin derivatives. *Eur J Med Chem.* 2001;36(7):615-25.
32. Šekularac GM, Nikolić JB, Petrović P, Bugarski B, Đurović B, Drmanić SŽ. Synthesis, antimicrobial and antioxidative activity of some new isatin derivatives. *Journal of the Serbian Chemical Society.* 2014;79(11):1347-54.
33. Prakash CR, Raja S. Synthesis, characterization and in vitro antimicrobial activity of some novel 5-substituted Schiff and Mannich base of isatin derivatives. *Journal of Saudi Chemical Society.* 2013;17(3):337-44.
34. Lian Z-M, Sun J, Zhu H-L. Design, synthesis and antibacterial activity of isatin derivatives as FtsZ inhibitors. *J Mol Struct.* 2016;1117:8-16.
35. Kiran G, Maneshwar T, Rajeshwar Y, Sarangapani M. Microwave-assisted synthesis, characterization, antimicrobial and antioxidant activity of some new isatin derivatives. *J Chem.* 2012;2013.
36. Leardi, R. Genetic Algorithms in Chemometrics and Chemistry: A Review. *J. Chemometrics.* 2001, 15, 559-569.

37. Sabet R.; Fassihi A.; Saghaie L., Octanol-water partition coefficients determination and QSPR study of some 3-hydroxy pyridine-4-one derivatives, *Journal of Pharmaceutical Research International*. 2018 .22(4), 1-15.
38. Olah, M.; Bologa, C.; Oprea, T.I. An Automated PLS Search for Biologically Relevant QSAR Descriptors. *J. Comput. Aided Mol. Des.* 2004, 18, 437-449.
39. R. Franke, A. Gruska, Chemometrics Methods in molecular design, in: H. van Waterbeemd, (Ed.), *Methods and Principles in Medicinal Chemistry*, VCH, Weinheim, 1995, Vol. 2, pp. 113–119.
40. H. Kubinyi, The quantitative analysis of structure-activity relationships, in: M.E. Wolff, (Ed.), *Burger's Medicinal Chemistry and Drug Discovery*, 5<sup>th</sup> Ed.; Wiley, New York, 1995, Vol. 1, pp. 506-509.
41. Brereton R. *Chemometrics Data Analysis for the Laboratory and Chemical Plant*. Wiley. 2004:47–54.

**Table 1.** Chemical structure of isatin derivatives used in this study



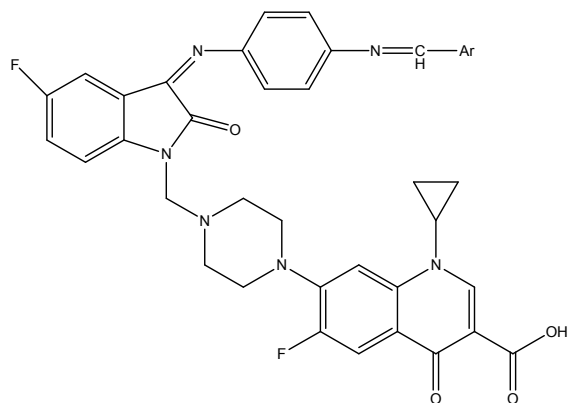
Compound	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	PMIC
1		H	H	4.00
2		H	H	4.60

3**		H	H	3.70
4		H	H	4.00
5		H	H	4.60
6		H	H	3.70
7**		H	H	4.60
8		CH3	H	4.00
9		CH3	H	4.30
*		CH3	H	
*		H	CH3	
*		H	CH3	
*		H	CH3	

\*outlier data

\*\*test set

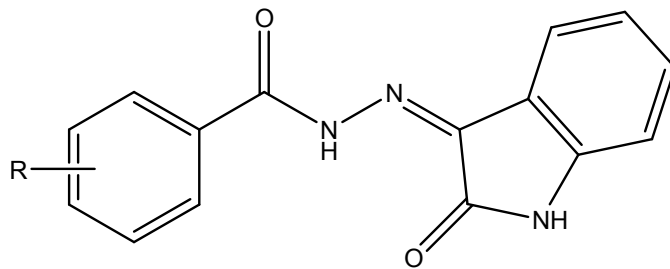




Compound	Ar	PMIC
*		
*		
10**		4.81
*		
*		
*		
11**		4.51
12		4.51
13		4.51

\*outlier data

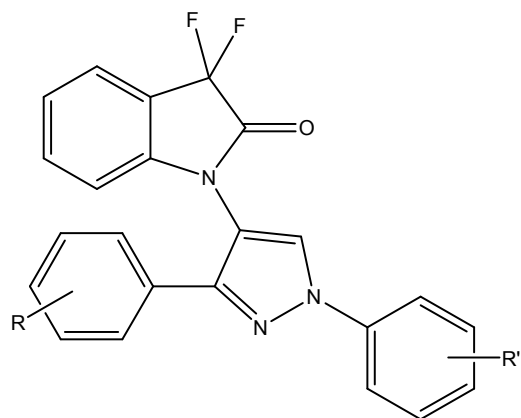
\*\*test set



Compound	R	PMIC
14	H	5.18
15	2-F	6.74
16	3-Cl	5.95
17	3-CH3	5.31
18	2-Br	5.61
19**	3-F	6.17

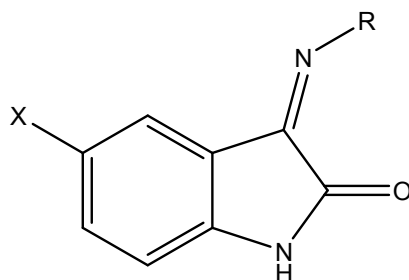
\*\*test set

UNDER PEER REVIEW

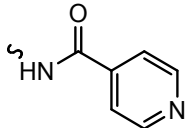
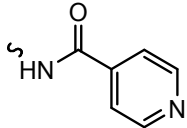


Compound	R	R'	PMIC
20	H	H	3.60
21	H	4-CH <sub>3</sub>	3.60
22	H	4-CN	3.70
23	4-OCH <sub>3</sub>	H	3.60
24	4-OCH <sub>3</sub>	4-CH <sub>3</sub>	3.70
25	4-OCH <sub>3</sub>	4-CN	3.82
26	4-Cl	H	3.60
27	4-Cl	4-CH <sub>3</sub>	3.70
28	4-Cl	4-CN	3.60
29	4-CH <sub>3</sub>	H	3.70
30	4-CH <sub>3</sub>	4-CH <sub>3</sub>	3.60
31	4-CH <sub>3</sub>	4-CN	3.70
32	4-Br	H	3.30
33	4-Br	4-CH <sub>3</sub>	3.60
34	4-Br	4-CN	3.70
35	4-F	H	3.82
36	4-F	4-CH <sub>3</sub>	3.82

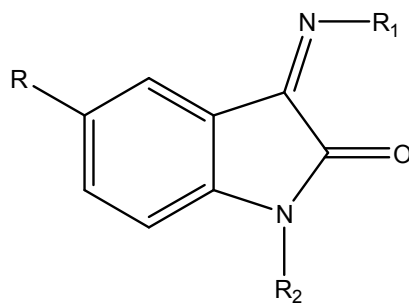
37	4-F	4-CN	4.10
38	2-Cl,4-Cl	H	3.60
39**	2-Cl,4-Cl	4-CN	3.82



Compound	R	X	PMIC
40		H	4.90
41**		F	4.90
42		Cl	4.60
43		CH <sub>3</sub>	4.60
44		H	4.90
45**		F	4.30
46		F	4.90

47	 <chem>NC(=O)c1ccncc1F</chem>	F	4.60
48	 <chem>NC(=O)c1ccncc1Cl</chem>	Cl	4.90

UNDER PEER REVIEW



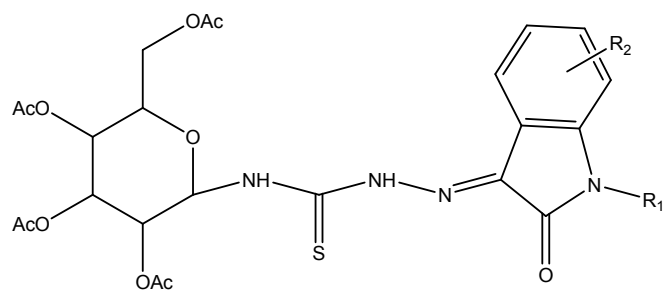
Compound	R	R1	R2	PMIC
49	H	1-naphthyl	—	3.78
50	H	4-Chloro phenyl	—	4.92
51	H	4-Bromo phenyl	—	3.78
52	H	4-Methyl phenyl	—	3.94
53	H	Phenyl hydrazino	—	3.88
54	H	Thiosemicarbazino	—	4.17
55	CH3	1-Naphthyl	—	3.85
56	CH3	4-Chloro phenyl	—	3.86
57	NO2	1-Naphthyl	—	4.39
58	NO2	4-Bromo phenyl	—	4.18
59	NO2	4-Chloro phenyl	—	3.78
60	CH3	4-Methoxy phenyl	—	3.58
61	CH3	Thiosenicarbazino	—	4.40
62	Cl	4-chloro phenyl	—	4.01
63	Cl	4-Methyl phenyl	H	4.85

64	Cl	Thiosemicarbazino	H	4.00
65	Br	1-Naphtyl	H	3.78
66	Br	4-Methoxy phenyl	H	3.79
67	H	4-Methyl phenyl	CH2- N(C6H5)2	3.89
68	Cl	4-Bromo phenyl	CH2- N(C6H5)2	3.79
*	Br	4-Methoxy phenyl	CH2- N(C6H5)2	

\*outlier data

\*\* test set

UNDER PEER REVIEW



Compound	R1	R2	PMIC
*	H	H	
*	H	5'-NO <sub>2</sub>	
69**	H	5'-Cl,7'-NO <sub>2</sub>	4.60
70	H	5'-Br,7'-NO <sub>2</sub>	4.60
71**	H	5',7'-diBr	4.30
72**	H	5'-F	4.30
73**	H	5'-Cl	4.30
74	H	5'-Br	4.30
75**	H	5'-I	4.30
76**	H	5'-Me	4.60
77	H	7'-Me	4.60
78	H	5'-iPr	4.30
79	Me	H	4.90
80	Et	H	4.90
81	n-Pr	H	4.90
82**	n-Bu	H	4.90



83**	i-Bu	H	4.90
84	Allyl	H	4.00
85	Bn	H	4.00
86	Phen	H	4.00

\*outlier data/ \*\*test set

UNDER PEER REVIEW

**Table 2.** Brief description of some descriptors used in this study

Descriptor type	Molecular Description
<b>Chemical</b>	LogP (Octanol-water partition coefficient), Hydration Energy (HE), Polarizability (Pol), Molar refractivity (MR), Molecular volume (V), Molecular surface area (SA).
<b>Constitutional</b>	mean atomic vander Waals volume (MV), no. of atoms, no. of non-H atoms, no. of bonds, no. of heteroatoms, no. of multiple bonds (nBM), no. of aromatic bonds, no. of functional groups (hydroxyl, amine, aldehyde, carbonyl, nitro, nitroso, etc.), no. of rings, no. of circuits, no of H-bond donors, no of H-bond acceptors, no. of Nitrogen atoms (NN), chemical composition, sum of Kier-Hall electrotopological states (Ss), mean atomic polarizability (Mp), number of rotatable bonds (RBN), mean atomic Sanderson electronegativity (Me), number of Chlorine atoms (NCl), number of 9-membered rings (NR09), etc.
<b>Topological</b>	Molecular size index, molecular connectivity indices (X1A, X4A, X2v, X1Av, X2Av, X3Av, X4Av), information content index (IC), Sum of topological distances between F..F (T(F..F)), Ratio of multiple path count to path counts (PCR), Mean information content vertex degree magnitude (IVDM), Eigenvalue sum of Z weighted distance matrix (SEigZ), reciprocal hyper-detour index (Rww), Eigenvalue coefficient sum from adjacency matrix (VEA1), radial centric information index, 2D petijean shape index (PJI2), mean information index on atomic composition(AAC), Kier symmetry index(S0K), mean information content on the distance degree equality (IDDE), structural information content (neighborhood symmetry of 3-order) (SIC3), Randic-type eigenvector-based index from adjacency matrix (VRA1), sum of topological distances between N..N (T(N..N)), sum of topological distances between O..O(T(O..O)),etc.
<b>Geometrical</b>	3D-Balaban index (J3D), span R (SPAN), length-to-breadth ratio by WHIM (L/BW), sum of geometrical distances between N..N (G(N..N)), sum of geometrical distances between N..O (G(N..O)), sum of geometrical distances between O..O (G(O..O)), ect.
<b>Walk-Mol</b>	molecular walk count of order 08 (MWC08), self-returning walk count of order 05 (SRW05), total walk count (TWC), etc.
<b>Burden matrix</b>	highest eigenvalue n. 1 of Burden matrix / weighted by atomic masses (BEHM1), highest eigenvalue n. 7 of Burden matrix / weighted by atomic masses (BEHM7), lowest eigenvalue n. 1 of Burden matrix / weighted by atomic masses (BELM1), highest eigenvalue n. 1 of Burden matrix / weighted by atomic van der Waals volumes (BELV1), highest eigenvalue n. 2 of Burden matrix / weighted by atomic Sanderson electronegativities (BEHE2), etc.
<b>Galvez</b>	topological charge index of order 1 (GGI1), topological charge index of order 6 (GGI6),topological charge index of order 7 (GGI7), global topological charge index (JGT), etc.
<b>2D autocorrelation</b>	Broto-Moreau autocorrelation of a topological structure - lag 7 / weighted by atomic Sanderson electronegativities (ATS7E), Moran autocorrelation -lag 4 / weighted by atomic Sanderson electronegativities (MATS4E), Broto-Moreau autocorrelation of a topological structure - lag 3 / weighted by atomic Sanderson electronegativities (ATS3E), Broto-Moreau autocorrelation of a topological structure - lag 3 / weighted by atomic van der Waals volumes (ATS3V), etc.
<b>Charge</b>	maximum positive charge (QPOS), partial charge weighted topological electronic charge (PCWTE), etc.
<b>Aromaticity</b>	Harmonic Oscillator Model of Aromaticity index,RCI;Jug RC index HOMA

	aromaticity indices,HOMT;HOMA total (trial) , etc.
<b>Randic</b>	DP0;molecular profile, SP0;shape profile; <b>SHP</b> ;average shape profile index , etc.
<b>RDF</b>	Radial Distribution Function - 7.0 / unweighted(RDF070U),Radial Distribution Function - 13.5 / unweighted(RDF135U),Radial Distribution Function - 1.0 / weighted by atomic masses(RDF010M),Radial Distribution Function - 3.0 / weighted by atomic masses(RDF030M),Radial Distribution Function - 4.5 / weighted by atomic masses(RDF045M),Radial Distribution Function - 12.5 / weighted by atomic masses(RDF125M),Radial Distribution Function - 2.0 / weighted by atomic van der Waals volumes(RDF020V),Radial Distribution Function - 8.5 / weighted by atomic van der Waals volumes(RDF085V),Radial Distribution Function - 1.0 / weighted by atomic Sanderson electronegativities(RDF010E), etc.
<b>3D-MoRSE</b>	3D-MoRSE - signal 01 / unweighted (MOR01U)(01U,02U,...,32U), 3D-MoRSE - signal 01 / weighted by atomic van der Waals volumes (MOR01V)(01V,02V,...,32V), ect.
<b>WHIM</b>	1st component symmetry directional WHIM index / weighted by atomic polarizabilities (G1P), 2st component symmetry directional WHIM index / weighted by atomic electrotopological states (G2S), D total accessibility index / weighted by atomic van der Waals volumes (DV), etc.
<b>GETAWAY</b>	H autocorrelation of lag 1 / lag2/ lag3 weighted by atomic Sanderson electronegativities (H1E,H2E,H3E), total information content on the leverage equality (ITH), R maximal autocorrelation of lag 3 / lag4 unweighted (R3U+,R4U+), R maximal autocorrelation of lag 6 / weighted by atomic masses (R6M+), R maximal autocorrelation of lag 5 / weighted by atomic van der Waals volumes (R5V+), R maximal autocorrelation of lag 1 / lag 4 weighted by atomic Sanderson electronegativities (R1E+), R maximal autocorrelation of lag 3 / weighted by atomic polarizabilities (R3P+), etc.
<b>Functional</b>	number of total secondary C(sp3) (NCS), number of ring tertiary C(sp3) (NCRHR), number of secondary C(sp2) (n=CHR), number of tertiary amines (aliphatic) (NNR2), number of N hydrazines (aromatic) (nN-NPH), number of nitriles (aliphatic) (NCN), number of phenols (NOHPH), number of ethers (aromatic) (NRORPH), number of sulfures (NRSR), etc.
<b>Atom-Centred</b>	CHR3 (C-003), CR4 (C-004), X--CR..X (C-034), Ar-C(=X)-R (C-039), R-C(=X)-X / R-C#X / X=C=X (C-040), X--CH..X (C-042), H attached to C1(sp3) / C0(sp2) (H-047), RCO-N< / >N-X=X (N-072),R2S / RS-SR (S-107), etc.
<b>connectivity indices</b>	X0(connectivity index chi-0), connectivity index chi-1(x1), average connectivity index chi-0(XOA)
<b>information indices</b>	Uindex(Balaban U index), IC0(information content index), TIC0(total information content index)
<b>edge adjacency indices</b>	EEig01x(Eigenvalue 01),EEig01r(Eigenvalue 01 from edge)
<b>eigenvalue-based indices</b>	Eig1v(Leading eigenvalue from van der Waals weighted distance Eigenvalue sum from mass weighted distance matrix),SEigm <b>matriceigenvalue</b> -based indices

**Table 3.** The results of MLR analysis with different types of descriptors.

Equation	Descriptors	(+) effect	(-) effect	R <sup>2</sup>	F	Q <sup>2</sup>	SE
E1	chemical	--	Logp	0.72	5.3	0.67	0.21
E2	constitutional	--	NCIC,nR10	0.58	18.52	0.52	0.32
E3	topological	--	STN	0.214	18.53	0.17	0.43
E4	Walk and path counts	--	PIPC09	0.283	26.81	0.21	0.53
E5	Connectivity indices	X4A	--	0.63	18.99	0.58	0.34
E6	Information indices	BIC5	--	0.38	19.685	0.32	0.36
E7	2D autocorrelation	GATS1M	MATS3E	0.64	10.417	0.58	0.17
E8	Edge adjacency indices	--	EEIG03X	0.231	20.45	0.18	0.55
E9	Burden eigenvalues	--	BEHm1	0.63	28.562	0.58	0.28
E10	Topological charge indices	--	JGI5	0.68	7.50	0.54	0.34
E11	Eigenvalue-based indices	--	LP1	0.25	23.358	0.17	0.54
E12	Geometrical descriptors	--	G(N..F),DISP V	0.46	8.704	0.39	0.41
E13	RDF descriptors	RDF030M	RDF020M	0.65	10.05	0.59	0.23
E14	3D MoRSE descriptors	--	MOR30M,MOR24U	0.69	17.67	0.62	0.38
E15	WHIM descriptors	E1U,G1M	E3U,P1P	0.61	15.25	0.56	0.46
E16	GETAWAY descriptors	R4M	R2V,HOU,HTM	0.59	10.88	0.53	0.39
E17	Fuctional group counts	nBnz, nCONN, nRCONHR, nArX, nC=N-N	nArCN, nPyridines	0.63	15.29	0.59	0.34
E18	Atom-centred fragments	C-039,C-034	--	0.56	13.51	0.49	0.24
E19	Charge descriptors	QMEAN, QPOS		0.58	13.52	0.52	0.54

**Table 4.** Statistical parameters for testing prediction ability of the MLR, GA-PLS, PCR, and FA-MLR models

<b>Model</b>	<b>R<sup>2</sup></b>	<b>R<sup>2</sup><sub>LOOCV</sub></b>	<b>RMSE<sub>cv</sub></b>	<b>R<sup>2</sup><sub>p</sub></b>	<b>RMSE<sub>p</sub></b>
<b>MLR</b>	0.67	0.62	0.21	0.74	0.18
<b>GA-PLS</b>	0.81	0.75	0.19	0.87	0.23
<b>PCR</b>	0.73	0.68	0.16	0.78	0.22
<b>FA-MLR</b>	0.64	0.59	0.15	0.70	0.14

R<sup>2</sup>: Regression Coefficient for Calibration set

R<sup>2</sup><sub>LOOCV</sub>: Regression Coefficient for Leave One Out Cross Validation

RMSE<sub>cv</sub>: Root Mean Square Error of cross validation

R<sup>2</sup><sub>p</sub>: Regression Coefficient for prediction set

RMSE<sub>p</sub>: Root Mean Square Error of prediction set

UNDER PEER REVIEW

**Table 5.** Numerical values of factor loading numbers 1–9 for descriptors after VARIMAX rotation

	Component									extraction
	F1	F2	F3	F4	F5	F6	F7	F8	F9	
LOGP	.190	.152	.038	-.175	.002	.091	<b>.724</b>	-.083	-.123	.646
NCIC	<b>.897</b>	.188	-.017	-.184	-.060	-.107	.013	.137	-.030	.909
NR10	<b>.774</b>	.073	-.091	-.021	.210	.398	.064	.254	-.063	.887
X4A	<b>-.655</b>	-.587	-.004	-.101	-.173	-.085	.006	.283	.105	.912
BIC5	-.304	-.314	.059	.025	-.444	.461	-.225	.373	-.035	.795
MATS3E	.456	.541	-.085	-.100	.145	-.301	.304	.094	-.286	.812
GATS1M	-.293	.306	<b>-.664</b>	.155	-.050	.184	-.002	.004	-.106	.691
BEHM1	-.006	.183	<b>.851</b>	.034	-.011	-.074	-.017	.001	-.147	.785
JG13	.004	-.125	-.106	.094	.009	<b>.932</b>	.078	-.006	-.017	.910
JG15	.325	.150	-.009	.105	<b>.818</b>	.107	-.040	-.074	.049	.830
L/Bw	-.319	<b>-.787</b>	.151	-.090	-.008	-.197	.129	.042	.059	.813
DISPV	-.059	-.107	.286	-.194	.547	-.357	.240	-.047	-.179	.653
RDF020M	-.176	.316	<b>.653</b>	.314	-.214	-.022	.245	-.027	-.144	.783
RDF030M	.019	<b>.893</b>	.188	-.021	.007	.026	.182	.003	-.023	.869
RDF025P	.485	<b>.724</b>	-.104	-.190	.170	-.219	.190	.045	-.078	.927
MOR24U	.230	-.099	-.006	<b>-.719</b>	.283	-.156	.092	.312	.046	.792
MOR30M	<b>.631</b>	.100	-.125	.102	-.282	-.142	.221	-.394	.014	.738
MOR32V	-.002	.245	.067	<b>.757</b>	.268	.076	-.113	.104	.011	.739
E1U	<b>-.736</b>	-.493	-.028	-.006	-.211	-.136	.002	.148	.011	.872
E3U	.128	.543	.089	.457	-.215	-.196	.215	-.259	.278	.803
G1M	-.073	<b>-.677</b>	.092	-.079	-.148	.237	-.312	-.159	.293	.764
P1P	-.013	<b>-.898</b>	-.035	-.285	.090	.011	.048	.159	-.065	.930
H0U	.150	.257	-.114	.322	.354	-.172	.537	-.259	.226	.767
HTM	.020	<b>.747</b>	.581	.092	.172	-.120	.073	.039	.015	.956
R4M	.079	<b>.709</b>	.444	.218	.029	.062	.052	-.032	.259	.829
R2V	<b>.773</b>	.204	.144	.154	.031	-.063	.225	.018	.401	.900
nCb	<b>.728</b>	-.199	.158	-.150	.180	-.237	-.069	-.010	-.096	.720
NRCONHR	-.592	-.406	.136	.290	-.107	-.205	-.411	.066	.006	.845
NCONN	<b>-.708</b>	.282	-.109	.211	.162	.125	.290	.300	-.136	.872
NARCN	<b>.617</b>	.033	.028	.203	.179	.036	.107	.112	-.012	.480
nC=N-N	<b>-.687</b>	.035	-.107	-.158	.158	-.006	.139	.540	.080	.851
NArX	.140	-.046	<b>.652</b>	-.024	.229	.173	-.350	-.160	.260	.746
nPyridines	-.062	-.069	-.037	-.007	.000	-.009	-.084	-.025	.876	.784
C-034	.008	-.037	-.064	-.052	-.151	.010	-.119	<b>.839</b>	-.051	.752
C-039	<b>-.868</b>	-.138	-.051	.082	-.260	-.087	-.092	.070	.016	.870
QPOS	-.367	<b>.881</b>	-.057	.111	.027	-.161	.082	.029	-.038	.962
QMEAN	<b>-.732</b>	.348	-.053	.222	.085	.161	-.304	.123	.048	.853
<b>% variance</b>	21.769	19.668	7.805	5.894	5.753	5.520	5.368	5.277	4.153	81.208

**Table 6.** The results of FA-MLR analysis with different types of descriptors

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	R <sup>2</sup>	F	Q <sup>2</sup>	SE
	B	Std. Error	Beta						
(Constant)	5.766	.547		10.537	.000	0.64	14.691	0.59	0.24
MOR30m	-0.031	.012	-.303	-2.619	.000				
JGI5	-2.336	0.721	-.442	-3.240	.002				

UNDER PEER REVIEW

**Table 7.** The results of PCR analysis

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	R <sup>2</sup>	F	Q <sup>2</sup>	SE
	B	Std. Error	Beta						
(Constant)	4.285	.057		75.368	.000	0.73	17.16	0.68	0.34
F1	-.258	.031	-.409	-4.510	.000				
F7	-.189	.021	-.300	-3.305	.001				
F6	.124	.027	.197	2.176	.033				
F8	.123	.057	.196	2.159	.034				

UNDER PEER REVIEW



**Table 8.** Leverage ( $h$ ) of the external test set molecules for different models. The last row ( $h^*$ ) is the warning leverage.

Molecule .no	MLR	GA-PLS	PCR	FA-MLR
3	0.117418	0.084802	0.026426	0.028202
7	0.058532	0.067963	0.03644	0.034729
10	0.087443	0.157524	0.131804	0.035335
11	0.071099	0.093302	0.092915	0.021066
19	0.054337	0.08314	0.03296	0.037432
39	0.081619	0.077263	0.136844	0.040156
41	0.097168	0.134119	0.13121	0.056011
45	0.158855	0.144921	0.152167	0.036003
69	0.045048	0.101806	0.06149	0.068055
71	0.109807	0.13409	0.023009	0.022631
72	0.102708	0.297308	0.041009	0.060281
73	0.105906	0.198805	0.022111	0.063121
75	0.087529	0.127991	0.018691	0.025659
76	0.04769	0.084609	0.021734	0.045611
82	0.081846	0.058078	0.022526	0.016686
83	0.077447	0.07017	0.016547	0.014426
$h^*$	0.214286	0.342857	0.191429	0.108571

**Table9.** Binding interaction of some studied compounds in active site of enzyme

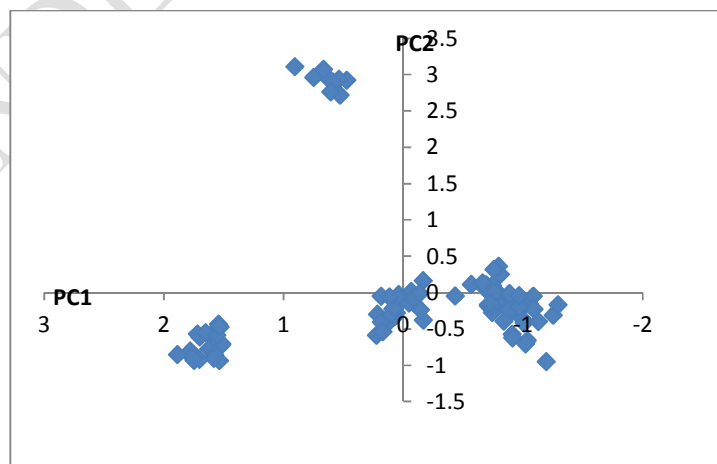
Compounds	$\Delta G$ kcal/ mol	$K_i$ ( $\mu M$ )	Atom of the ligand	Receptor/Coenzyme	Interaction	Distance( $\text{\AA}$ )
1	-8.46	631.45	C=O (isatine) C=O (isatine)	Tyr 156 2'-OH (ribose)NAD <sup>+</sup>	Hydrogen bond Hydrogen bond	1.80 1.85
2	-9.01	248.47	C=O (isatine) C=O (isatine) Isatine Indole Indole	Tyr 156 2'-OH (ribose)NAD <sup>+</sup> NAD <sup>+</sup> -Pyridine Phe 94 Ala 95	Hydrogen bond Hydrogen bond Pi-Pi H-Pi H-pi	1.82 1.83
3	-8.58	512.32	C=O (isatine) C=O (isatine) Isatine	Tyr 156 2'-OH (ribose)NAD <sup>+</sup> NAD <sup>+</sup> -Pyridine	Hydrogen bond Hydrogen bond Pi-Pi	1.81 1.89
4	-8.86	322.58	C=O (isatine) C=O (isatine) C=S Isatine Indole	Tyr 156 2'-OH (ribose)NAD <sup>+</sup> NH <sub>2</sub> - NAD <sup>+</sup> Pyridine- NAD <sup>+</sup> Phe 94	Hydrogen bond Hydrogen bond Hydrogen bond Pi-Pi H-Pi	1.77 1.94 2.66
5	-8.67	443.90	C=O (isatine) C=O (isatine) Isatine Indole	Tyr 156 2'-OH (ribose)NAD <sup>+</sup> NAD <sup>+</sup> -Pyridine Phe 94	Hydrogen bond Hydrogen bond Pi-Pi H-Pi	1.78 1.89
6	-8.24	912.17	C=O (isatine) C=O (isatine) Isatine	Tyr 156 2'-OH (ribose)NAD <sup>+</sup> NAD <sup>+</sup> -Pyridine	Hydrogen bond Hydrogen bond Pi-Pi	1.85 1.88
7	-8.51	577.86	C=O (isatine) C=O (isatine) Isatine	Tyr 156 2'-OH (ribose)NAD <sup>+</sup> NAD <sup>+</sup> -Pyridine	Hydrogen bond Hydrogen bond Pi-Pi	1.95 2.09
8	-8.46	630.94	C=O (isatine) C=O (isatine)	Tyr 156 2'-OH (ribose)NAD <sup>+</sup>	Hydrogen bond Hydrogen bond	1.89 1.95
9	-8.47	614.00	C=O (isatine) NH-thiocarboydrazone N-thiocarboydrazone Isatine	2'-OH (ribose)NAD <sup>+</sup> Met 159 Tyr 156 Ala 95	H-Pi Hydrogen bond Hydrogen bond H-Pi	1.87 2.35
14	-7.44	3.52	C=O(benzohydrazid e) C=O(benzohydrazid	Tyr 156 2'-OH- ribose(NAD <sup>+</sup> )	Hydrogen bond Hydrogen	2.22 2.17 2.66

			e) C=O(isatine)	Gly 93	bond Hydrogen bond	
15	-7.94	1.52	C=O(benzohydrazid e) C=O(benzohydrazid e) C=O(isatine)	Tyr 156 2'-OH- ribose(NAD <sup>+</sup> ) Gly 93	Hydrogen bond Hydrogen bond Hydrogen bond	1.88 1.52 2.02
16	-7.87	1.66	C=O(benzohydrazid e) C=O(benzohydrazid e) C=O(isatine) Phenyl	Tyr 156 2'-OH- ribose(NAD <sup>+</sup> ) Gly 93 Tyr 146	Hydrogen bond Hydrogen bond Hydrogen bond H-Pi	2.19 2.03 2.36
17	-7.79	1.95	C=O(benzohydrazid e) C=O(benzohydrazid e) C=O(isatine) Phenyl	Tyr 156 2'-OH- ribose(NAD <sup>+</sup> ) Gly 93 Tyr 146	Hydrogen bond Hydrogen bond Hydrogen bond H-Pi	1.97 2.12 2.92
18	-7.81	1.88	C=O(benzohydrazid e) C=O(benzohydrazid e) C=O(isatine)	Tyr 156 2'-OH- ribose(NAD <sup>+</sup> ) Gly 93	Hydrogen bond Hydrogen bond Hydrogen bond	1.98 1.63 3.15
19	-7.86	1.72	C=O(benzohydrazid e) C=O(benzohydrazid e) C=O(isatine)	Tyr 156 2'-OH- ribose(NAD <sup>+</sup> ) Gly 93	Hydrogen bond Hydrogen bond Hydrogen bond	1.79 2.23 2.46
20	-6.57	15.30	Pyrazole	Tyr 156	H-Pi	
21	-6.42	19.80	Pyrazole	Tyr 156	H-Pi	
22	-6.21	28.20				
23	-6.28	25.03				
24	-6.30	26.91				
25	-6.79	10.57	Pyrazole	Tyr 156	H-Pi	
26	-6.33	27.55				
28	-6.25	26.00				
29	-6.46	18.25				
30	-6.03	38.14				

31	-6.37	21.35				
32	-6.35	22.18				
33	-6.12	32.63				
34	-6.30	24.20				
35						
36	-6.26	26.17				
37	-6.65	13.38				
38	-6.65	13.38	2-Cl	Leu100	Hydrogen bond	2.90
39	-6.68	12.61	2-Cl	Leu100	Hydrogen bond	2.83
40	-8.47	618.68 nano	C=O (isatine) N-H (isatine) C=O (imidazolidine) C=O (imidazolidine)	NH <sub>2</sub> ( NAD <sup>+</sup> ) Phosphate ( NAD <sup>+</sup> ) Tyr 156 2'-OH (ribose)NAD <sup>+</sup>	Hydrogen bond Hydrogen bond Hydrogen bond Hydrogen bond	2.18 2.33 1.81 1.83
41	-8.30	828.33 nano	C=O (isatine) C=O (isatine) Isatine	Tyr 156 2'-OH (ribose)NAD <sup>+</sup> NAD <sup>+</sup> -Pyridine	Hydrogen bond Hydrogen bond Pi-Pi	1.70 1.76
42	-8.08	1.19	C=O (isatine) C=O (isatine) Isatine	Tyr 156 2'-OH (ribose)NAD <sup>+</sup> NAD <sup>+</sup> -Pyridine	Hydrogen bond Hydrogen bond Pi-Pi	1.81 1.89
43	-7.93	1.53	C=O (isatine) C=O (isatine) Isatine	Tyr 156 2'-OH (ribose)NAD <sup>+</sup> NAD <sup>+</sup> -Pyridine	Hydrogen bond Hydrogen bond Pi-Pi	1.77 1.94
44	-8.39	709.17 nano	C=O (isatine) C=O (isatine) NH <sub>2</sub> (urea) NH <sub>2</sub> (urea)	Tyr 156 2'-OH (ribose)NAD <sup>+</sup> Phosphate ( NAD <sup>+</sup> ) Phosphate ( NAD <sup>+</sup> )	Hydrogen bond Hydrogen bond Hydrogen bond Hydrogen bond	1.91 1.64 2.81 3.74
46	-8.24	912.17	C=O (isatine) C=O (benzamide) C=O (benzamide)	Gly 93 Tyr 156 2'-OH (ribose)NAD <sup>+</sup>	Hydrogen bond Hydrogen bond Hydrogen bond	2.44 1.86 2.30
47	-8.33	778.55 nano	C=O (isatine) C=O (isatine) C=O (nicotinamide)	Tyr 156 2'-OH (ribose)NAD <sup>+</sup> Gly 93	Hydrogen bond Hydrogen bond	1.81 1.81 2.10

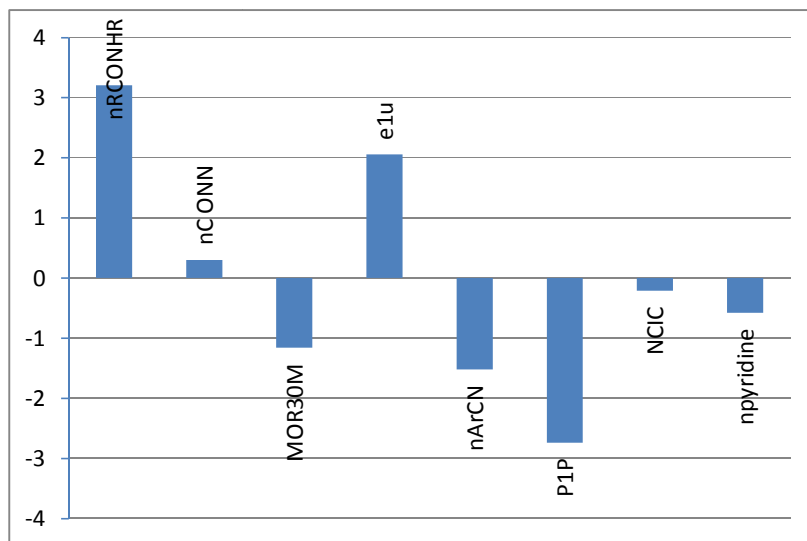
					Hydrogen bond	
48	-8.22	943.86 nano	C=O (isatine) C=O (isatine) C=O (nicotinamide)	Tyr 156 2'-OH (ribose)NAD <sup>+</sup> Gly 93	Hydrogen bond Hydrogen bond Hydrogen bond	1.97 1.63 2.80
49	-8.57	519.34	C=O (isatine) C=O (isatine)	Tyr 156 2'-OH (ribose)NAD <sup>+</sup>	Hydrogen bond Hydrogen bond	1.83 1.68
50	-8.47	618.68	C=O (isatine) C=O (isatine)	Tyr 156 2'-OH (ribose)NAD <sup>+</sup>	Hydrogen bond Hydrogen bond	2.01 1.76
52	-7.58	2.79 (μM)	C=O (isatine) C=O (isatine)	Tyr 156 2'-OH (ribose)NAD <sup>+</sup>	Hydrogen bond Hydrogen bond	1.81 1.85
53	-7.80	1.90 (μM)	C=O (isatine) C=O (isatine) Isatine	Tyr 156 2'-OH (ribose)NAD <sup>+</sup> NAD <sup>+</sup>	Hydrogen bond Hydrogen bond Pi-Pi	1.87 2.05

**Figure 1.** Outlier data by Principle Component Analysis before QSAR analysis



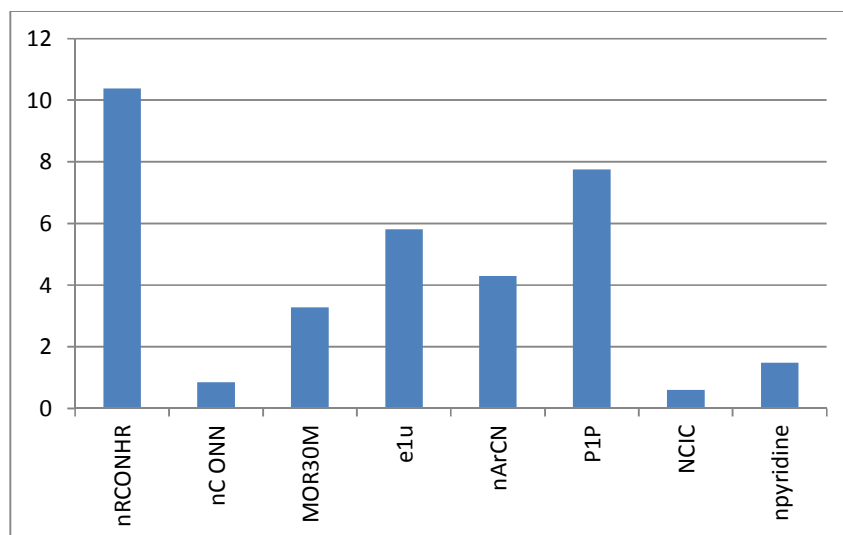
UNDER PEER REVIEW

**Figure 2.** PLS regression coefficients for the variables used in GA-PLS model.



UNDER PEER REVIEW

**Figure 3.** Plot of variables important in projection (VIP) for the descriptors used in GA-PLS model.



UNDER PEER REVIEW



**Figure 4.** The docked configuration of 2 (A), 14(B), 27(C) and 40 (D) in the binding site of FabI

