Journal of Physical and Theoretical Chemistry
of Islamic Azad University of Iran, 18 (1) 1-19: Spring 2021
(J. Phys. Theor. Chem. IAU Iran)
ISSN 1735-2126

Monte Carlo and QSAR Study on Biological Activity of Several Platinum (IV) Anti Cancer Drugs

R. SayyadikordAbadi[1*], A. Fallah Shojaei[1], A. Alizadehdakhel[1], L. Mohammadinargesi[2], G. Ghasemi[1]

[1] Depatment of Chemistry and Chemical Engineering, Rasht Branch, Islamic Azad University, Rasht, Iran
[2] Department of Chemistry, Faculty of Sciences, University of Guilan, P.O.BOX 1914, Rasht, Iran

Received April 2021; Accepted June 2021

ABSTRACT

QSAR investigations of some platinum (IV) derivatives were conducted using multiple linear regression (MLR) and artificial neural network (ANN) as modelling tools, along with simulated annealing (SA) and genetic algorithm (GA) optimization algorithms. In addition, CORAL software was used to correlate the biological activity to the structural parameters of the drugs. The obtained results from different approaches were compared and GA-ANN combination showed the best performance according to its correlation coefficient (R^2) and mean sum square errors (RMSE). From the GA-ANN method, it was revealed that MTAS8e, ESpm05d, BElv3, MWC09, ESpm14u, BEHe2, RDF125e, and S3K are the most important descriptors. From Monte Carlo simulations, it was found that the presence of double bond, present of Platinum, number of chlorine connected to Pt, branching in molecular skeleton and presence of N and O atoms are the most important molecular features affecting the biological activity of the drug. It was concluded that simultaneous utilization of QSAR and Monte Carlo method can lead to a more comprehensive understanding of the relation between physico-chemical, structural or theoretical molecular descriptors of drugs to their biological activities.

Keywords: Platinum (IV) Antitumor Drugs ,QSAR, Genetic Algorithm, Monte Carlo method

1. INTRODUCTION

Platinum compounds occupy a leading position among drugs for cancer chemotherapy [1]. Platinum (IV) compounds with an inert higher oxidation state and a reactive lower oxidation state can be potentially hypoxia-selective agents most of which do not react significantly with proteins in the blood stream, [2] though they are still highly effective anticancer agents [3]. The mechanisms of drug effects strongly depend on their molecular structure [4,5].

Platinum (IV) complexes undergo ligand substitution reactions much more

*Corresponding author: sayyadi_04@yahoo.com, sayyadi@iaurasht.ac.ir

slowly than their Pt(II) counterparts, and are therefore considered as prodrugs; Pt(IV) species can be activated by reduction to their Pt(II) counterparts, which is followed by hydrolysis and electrophilic attack on DNA. A sensible choice of the axial ligands is thus the key for being able to modulate the lipophilic character of these Pt(IV) complexes (basically, their ability to enter the tumor cells by passive diffusion) and their redox properties (the ability to undergo reduction under the hypoxic conditions typical of tumor tissues) [52-54].

Modelling and optimization approaches that relate the descriptors (constitutional, geometrical, topological, quantum chemical, etc.) to the biological activity of drugs are named QSAR [56,57]. Multiple Linear Regression (MLR), Artificial Neural Networks (ANN), Simulated Annealing algorithm (SA), [6] Genetic Algorithm (GA) [7], and Partial Least Squares (PLS), are the most common mathematical methods that have been utilized to describe the quantitative relationship between the molecular descriptors of the drugs and their properties[8,9] .

CORAL has been proposed as a competent software for the QSAR studies. It uses Monte Carlo method to find the most important simplified molecular input-line entry system (SMILES)-based descriptors and calculate their correlation weights to predict an endpoint (e.g., -log(IC_{50})). SMILES are lines of symbols, representing the molecular structure [10].

B3LYP/ANL2DZ is a valid model that is used for compounds with intermediate metals with “d” orbitals [44, 45]. Density functional theory (DFT) based computations are the main tools in contemporary computational chemistry. Although B3LYP underestimates reaction barrier heights, yields too low bond dissociation enthalpies, and fails to bind van der Waals systems, it is very promising in a number of areas of chemistry, such as geometry, stability of complexes, molecular pectroscopic properties and electronic structures and bonding characters [46-49].

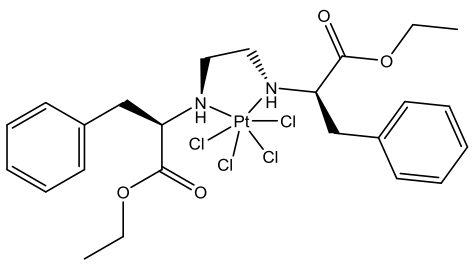
It should be noted that structural optimization is not the main focus of the present work. Nonetheless, QSAR study is the main part and novelty of this research. In many QSAR studies, the structural optimization of complexes is not performed by Gaussian [50, 51] and only a preliminary optimization is performed using “Hyper” software. However, in this study, in addition to the preliminary optimization using Hyper, optimization with Gaussian has also been performed. Most of the studies reported in the literature were limited to establishment of simple structure–activity relationships for small sets of Pt(IV) complexes [54,55]. This study considers a larger number and new series of Pt(IV) complexes. In the current study, MLR and ANN modelling tools coupled with SA and GA optimization techniques and Monte Carlo method were used to find the best set of descriptors that correlate the half maximal inhibitory concentration of 34 platinum (IV) complexes.

2. COMPUTATIONAL METHODS

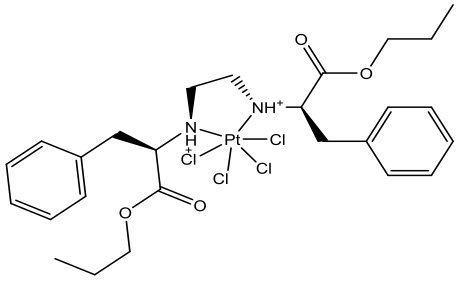
2.1. SELECTION OF DESCRIPTORS USING LINEAR REGRESSION

Geometrical optimizations of platinum (IV) complexes were carried out with B3lyp/lanl2dz at the Gaussian 03W [11, 12] (Figure 1).

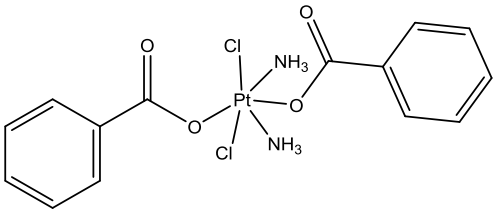
Dragon program [13] was used for calculation of 3226 molecular descriptors for each of the 34 Platinum (IV) compounds [14] which were categorized into topological, geometrical, MoRSE [15], RDF [16], GETAWAY[17], auto-correlations [15] and WHIM [18].



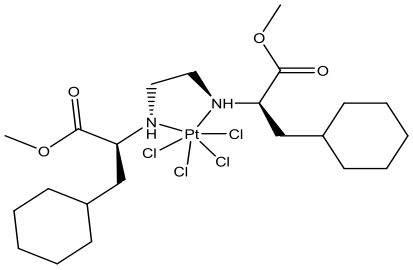
1- IC_{50} : 5.04



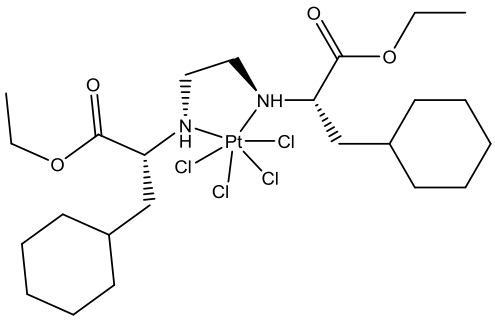
2- IC_{50} : 6.08



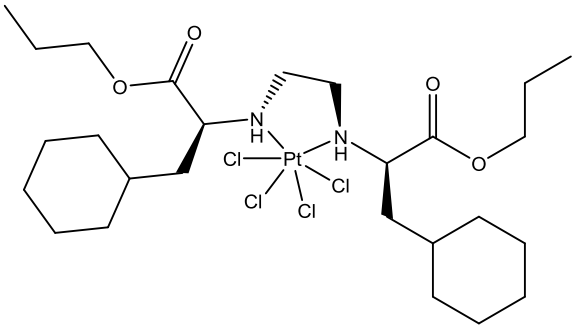
3- IC_{50} : 0.45



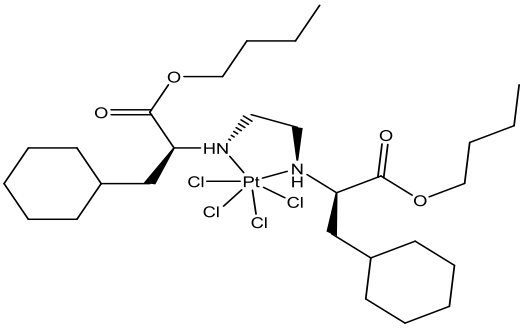
4- IC_{50} : 2.5



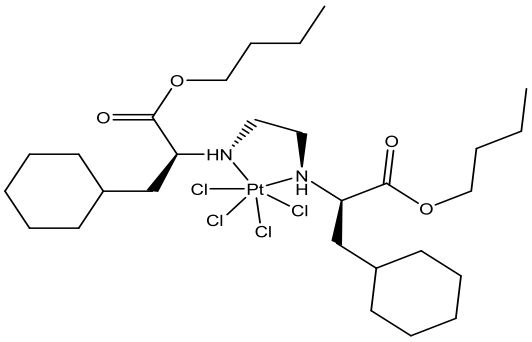
5- IC_{50} : 1.9



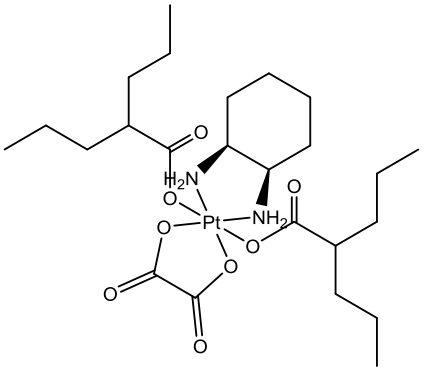
6- IC_{50} : 2.2



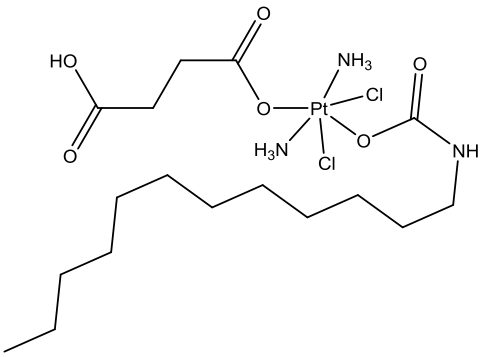
7- IC_{50}: 7.4



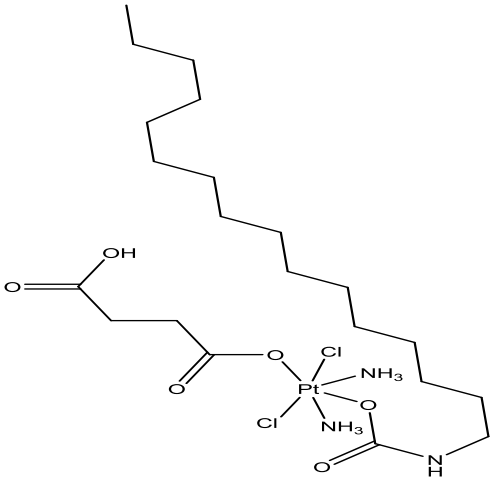
8- IC_{50} : 5.4



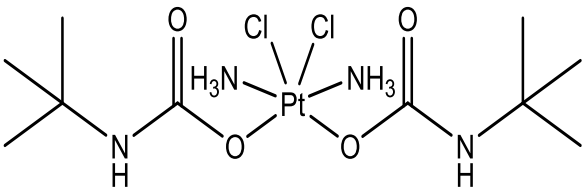
9- IC_{50} : 1.3



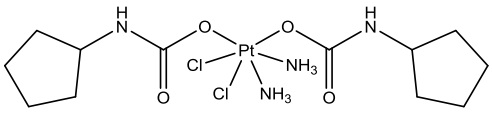
10- IC_{50} : 0.30



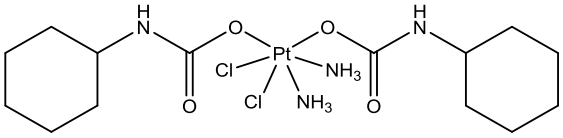
11- IC_{50} : 0.16



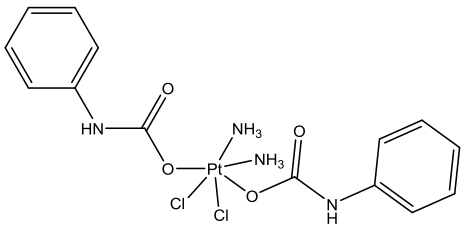
12- IC_{50} : 1



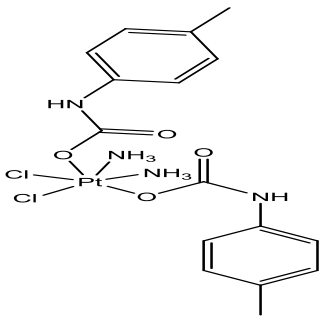
13- IC_{50} : 0.6



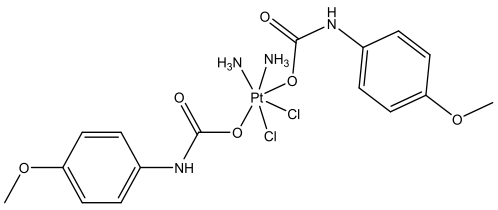
14- IC_{50}: 6.7



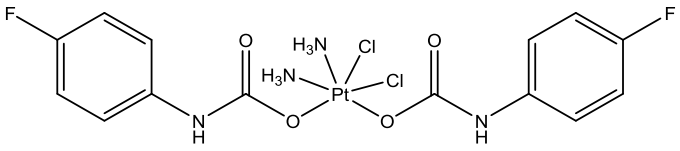
15- IC_{50} : 6.7



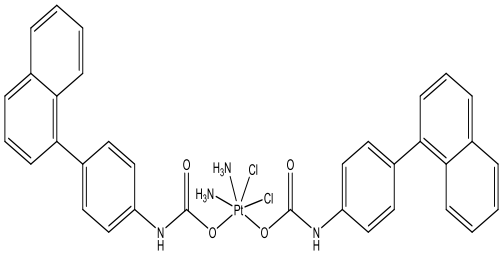
16- IC_{50} : 3



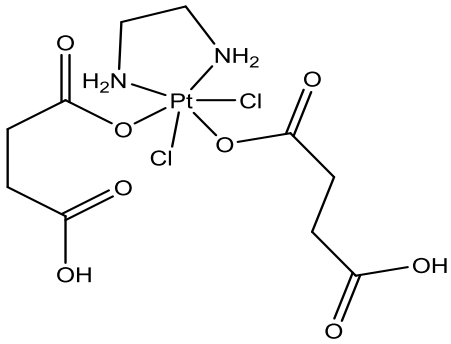
17- IC_{50} : 5.3



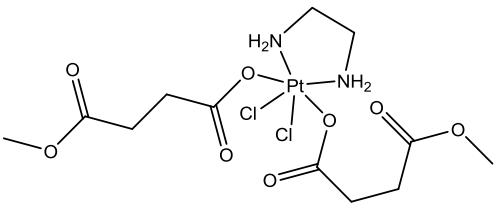
18- IC_{50} : 3.7



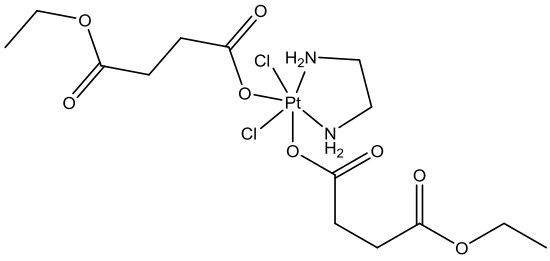
19- IC_{50} : 4.3



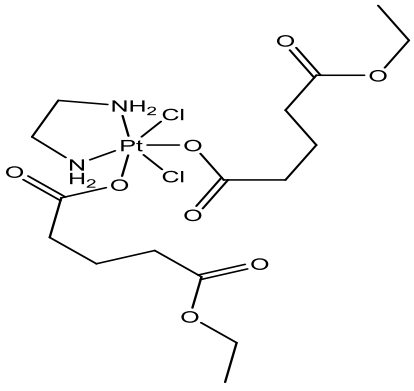
20- IC_{50} : 5.5



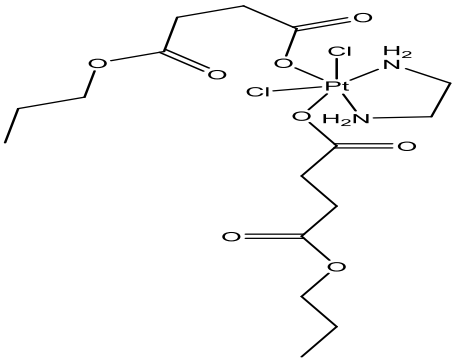
21- IC_{50} : 0.68



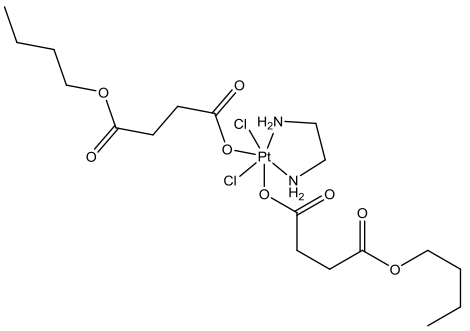
22- IC_{50} : 0.34



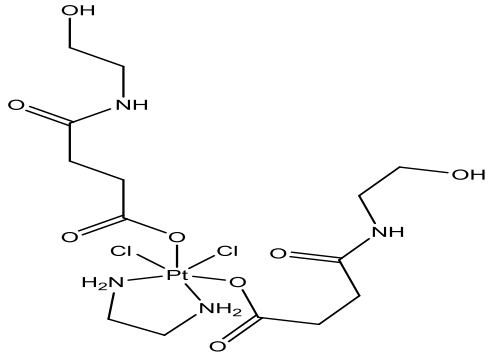
23- IC_{50} : 1.1



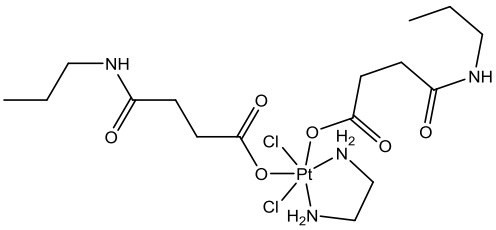
24- IC_{50} : 0.068



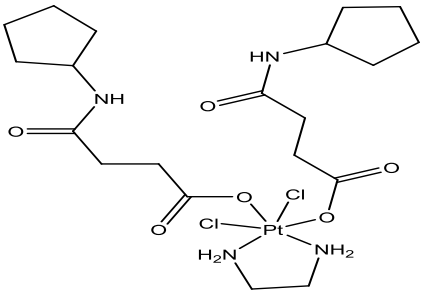
25- IC_{50} : 0.018



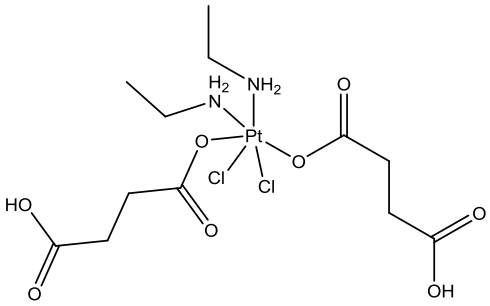
26- IC_{50} : 24



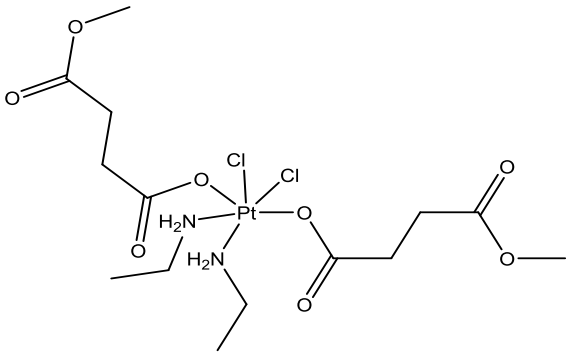
27- IC_{50} : 2.3



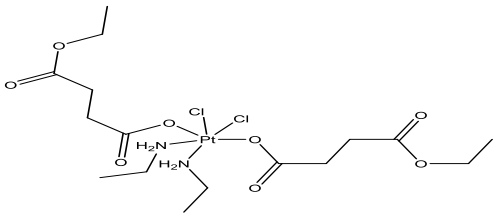
28- IC_{50} : 1.9



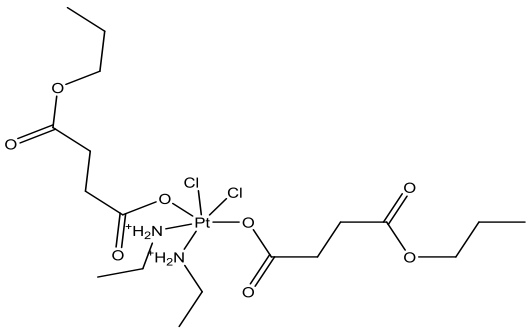
29- IC_{50} : 5.6



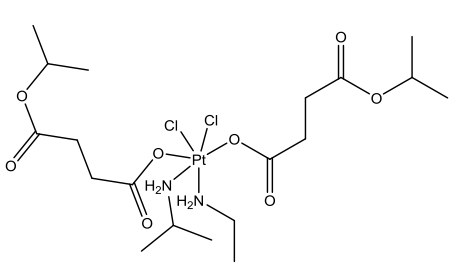
30- IC_{50} : 0.16



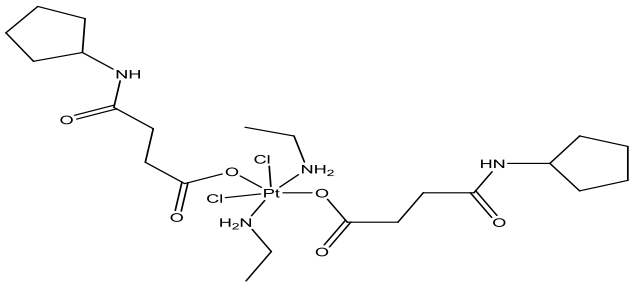
31- IC_{50} : 0.061



32- IC_{50} : 0.014



33- IC_{50} : 0.0094



34- IC_{50} : 0.75

Fig. 1. Structure of the platinum complexes used to build QSAR models with B3lyp/lanl2dz

2.2. Primary Selection of Descriptors Using Linear Regression

Three refining steps were performed to reduce the number of descriptors using SPSS [19]. In the first stage, the descriptors that had the same value for at least 70% of the Pt (IV) compounds in the dataset were removed and then the descriptors with correlation coefficient of less than 0.25 with the dependent variable - log IC50 (empirical negative logarithm of half maximal inhibitory concentration) were considered redundant and subsequently removed [20]. Following these two steps, the number of descriptors was reduced to 1408. The selected descriptors were then further screened as described in the following sections.

2.3. Final Selection of Descriptors Using Linear Regression

A stepwise multiple linear regression procedure based on the forward-selection and backward-elimination techniques was used for the rejection of descriptors in the linear models. The Multiple Linear Regression (MLR) model maps independent variables X to a dependent variable (response) Y using the following equation:

$$Y=W_1X_1+W_2X_2+\ldots+W_pX_p \quad (1)$$

Where, W_i is the coefficient of the regression [21]. An ideal model is one that has low standard deviation, high correlation coefficient (R^2), minimum number of independent variables, high predictive power, and a high F-statistic value [22].

2.4. Combination Methods

The 1408 descriptors chosen in the previous primary linear selection were implemented under additional screening using QSAR methods including MLR-MLR, MLR-PLS1, MLR-PCR, GA-MLR, SA-ANN, GA-ANN.

Principal Component Analysis (PCA) [23] is an unsupervised parametric method that reduces and classifies the number of variables by extracting those with a higher percentage of variance in the data (called principal components, PCs) without significant loss of information. PCA is essentially a coordinate transformation. The original data are plotted on an X-axis and a Y-axis. For two-dimensional data, PCA seeks to rotate these two axes so that the new axis X' lies along the direction of maximum variation in the data. PCA requires that the axes be perpendicular, so in two dimensions the choice of X' will determine Y'. The transformed data is obtained by reading the x and y values off this new set of axes, X' and Y'. For more than two dimensions, the first axis is in the direction of most variation; the second, in direction of the next-most variation; and so on. Principal components regression (PCR) is a regression technique based on PCA.

The basic idea behind PCR is to calculate the principal components and then use some of these components as predictors in a linear regression model fitted using the typical least squares procedure. Partial Least Squares (PLS) is similar to PCR, but works in one step. It also determines latent variables that are linear combinations of the original variables, but the criterion applied is maximal covariance between the response variable and the explanatory variables. In the MLR-PCR, MLR-PLS1 and MLR-MLR models, the best descriptors were selected and then used as inputs in unscramble software. The GA-ANN and SA-ANN approaches are described in Fig. 2.

In each run via ANN, the eight descriptors selected by the optimization method (GA or SA) were used as the inputs and corresponding values of (-logIC50) were defined as the target values (Fig. 3).

The neural network included eight neurons in the input layer and one neuron in the output layer. Equation (2) describes the calculation of each neuron's output from the hidden layer:

$$n_i = b_i + \sum_{j=1}^{m} (Win(j,i) * In_j), i = 1,2,..., N' \quad (2)$$

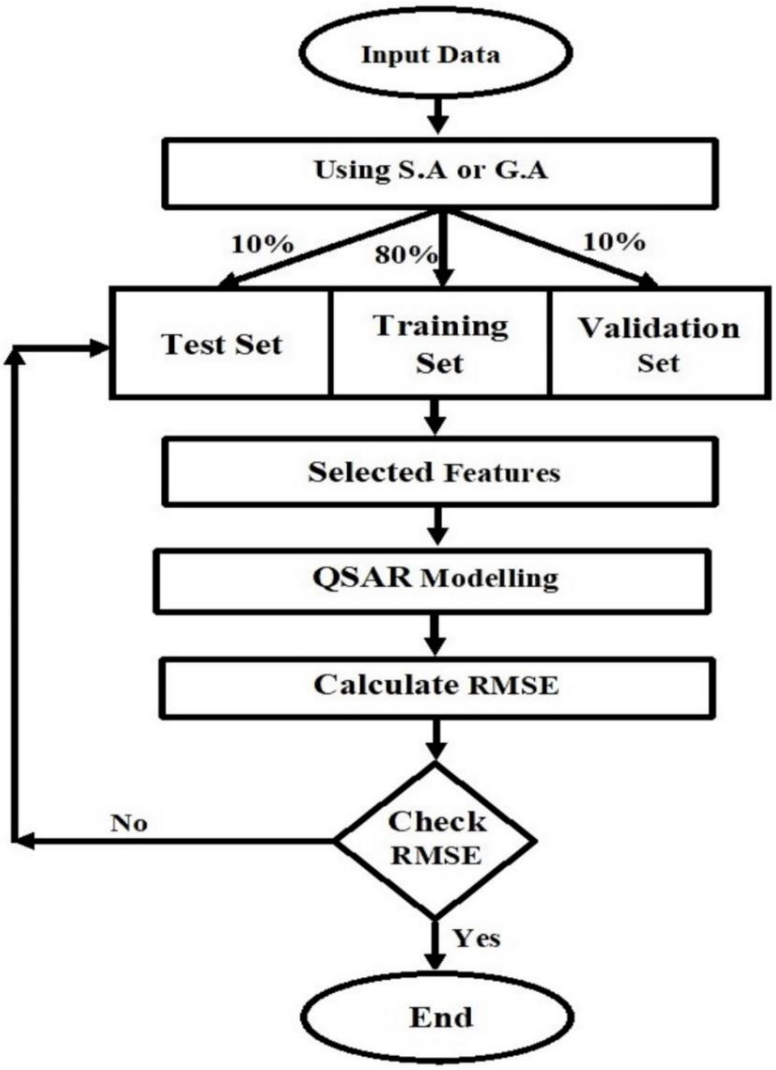


Fig. 2. The employed procedure for finding optimum descriptors of the nonlinear models in Pt(IV) complexes.

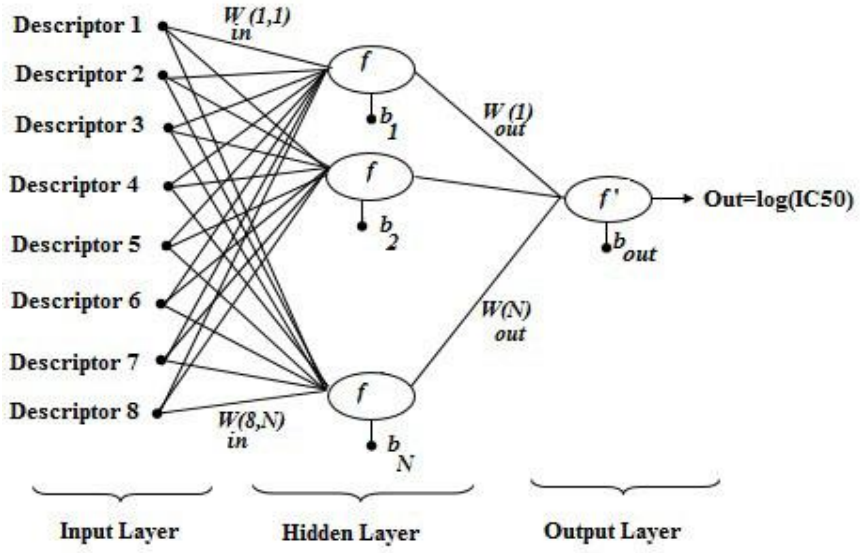


Fig. 3. The neural network computational model.

Where, b is the bias that is added to the weighted inputs. N' and m represent the number of neurons in the hidden layer and the number of inputs, respectively. The transfer function of hidden layer (f) acts on n_i to give the outputs of each neuron in the hidden layer:

$$a_i = f(n_i) \tag{3}$$

The network output can be given by the following equation:

$$out = f'(b_{out} + \sum(W_{out}(i) * a_i)) \qquad i = 1,2,...,N' \tag{4}$$

Where, f' is the transfer function for the output layer [18]. Three neurons were used in the hidden layer and 80%, 10% and 10% of data sets in were randomly chosen as training, validation and test sets, respectively. The networks were trained using the training set members via Levenberg-Marquardt Algorithm [43], while logarithmic sigmoid and linear transfer function were used as the hidden and output transfer functions, respectively. The logarithmic sigmoid transfer function is defined as follows:

$$\log sigm = \frac{1}{1+e^{-x}} \tag{5}$$

Genetic Algorithm (GA) works on a method based on the biological evolution. This approach modifies a population of individual solutions in a repetitive mechanism [24]. SA is an algorithm that utilizes random movements in the domain in order to evade choosing local optimum values rather than the global optimal [6].

The 1408 descriptors were fed to the ANN as data series including eight inputs (chosen by the optimization algorithm; S.A or G.A) and one output. This screening approach was performed for the gas phase. High correlation coefficient (R^2) and low RMSE are characteristics of an ideal model. The RMSE is defined as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - y_o)^2}{n}} \tag{6}$$

Where, y_i is the desired output, y_a is the predicted value by method, and n is the number of molecules in our data set. Matlab. 2014a was used for modelling and optimization calculations

2. 5. Momte Carlo Method

CORAL [25] software was used for calculation of descriptor correlation weight (DCW) of the 34 platinum (IV) compounds with a hybrid optimization scheme including hydrogen-suppressed molecular graph (HSG) and SMILES representation of molecular structures. The data was randomly divided into training, calibration, and test subsets in three splits**.** Modelling using CORAL software was carried out for thresholds [26] of 1 up to 5 and 100 epochs (i.e., an overall number of 1500 runs were performed). The SMILES-based and Graph - based optimal descriptors are achieved using the following equations [16]:

$$DCW(T, Nepoch)^{SMILES} = \alpha \sum CW(Sk) + \beta \sum CW(SSk) + \gamma \sum CW(SSSk) + x.CW(NOSP) + y.CW(HALO) + z.CW(BOND) \quad (7)$$

$$DCW(T, Nepoch)^{Graph} = \sum CWAk + \alpha \sum CW(^{0}ECk) + \beta \sum CW(^{1}ECk) + \gamma \sum CW(^{2}ECk) + \delta \sum CW(^{3}ECk) \quad (8)$$

Where, S_k, SS_k and SSS_k is one, two, and three component SMILES attributes. NOSP (nitrogen, oxygen, sulfur, and phosphorus) and HALO (fluorine, chlorine, and bromine) shows the presence or absence of chemical elements. Also "BOND" denotes double (=), triple (#), or stereo chemical bonds (@ or @@). A_k in equation (2) indicates the occurrence of the C, N, O atoms in the HSG and HFG molecular graphs. The α, β, γ, and δ coefficients and combinations of their values are used to define various versions of the graph-based optimal descriptor and can be 1 or 0. The hybrid objective function for finding the optimal descriptors is defined as:

$$DCW(T, Nepoch)^{Hybrid} = DCW(T, Nepoch)^{SMILES} + DCW(T, Nepoch)^{Graph}$$

3. RESULTS AND DISCUSSION

3.1. Linear and non- Linear Combination Methods

All the studied platinum complexes [34-40] are presented in Figure 1. The calculated Pt–N, Pt-O, and Pt-Cl bond lengths using B3lyp/lanl2dz are 2.0495-2.1577, 2.0251-2.0585, and 2.4006-2.4676 A˚ respectively, which are in accordance with their experimental values [227].

The statistical parameters of all the QSAR approaches are shown in Table 1. The presented results in this table show that as expected, the efficiency of nonlinear methods were better that the linear ones. In Addition, MLR-MLR method was the best approach among the linear methods, while the GA-ANN combination method exhibited the best performance in comparison to the other non-linear approaches.

In order to compare the accuracy of the employed approaches with previous researches, the reported correlation coefficients of some similar studies have been presented in table 2, which proves the good performance of the developed nonlinear approaches in the present work.

Table 1. Statistical data of different QSAR models in gas phase

QSAR Model	Predicted		Train	
	R^2	RSME	R^2	RSME
MLR-MLR	0.942	0.204		
MLR-PLS1	0.906	0.258		
MLR-PCR	0.903	0.262		
SA-ANN	0.9685	0.0227	0.9647	0.1647
GA-MLR	0.9870	0.0095	0.9849	0.1062
GA-ANN	0.9879	0.0093	0.9887	0.0925

Table 2. Reported correlation coefficients from previous similar QSAR studies

Compounds	QSAR model	R^2	Ref.
Pt(IV) complexes with 3,5-dimethyl-5-(4-pyridyl) hydantoin and 5-methyl-5-(4-pyridyl) hydantoin and halogen ligands	PLS	0.778	[30]
Several cis-platinum complexes	Leave-One-Out cross-validation method.	0.795	[31]
Octahedral platinum (IV) complexes	MLR and PCA	0.9	[32]
Several Pt(IV) complexes	MLR and GA – Variable Subset Selection	0.8	[33]

Definitions of the selected descriptors using GA-ANN are given Table 3. In this table, BELv3, BEHe2 are Burden eigenvalues descriptors that the B matrix defines as the electronegativity of the atoms and the number of atoms, bond order between two atoms. MWC09, S3k are walk and path counts and topological that these are often sensitive to the electronic characteristics of molecules and size, shape, branching, and 2D-frequency fingerprints (fragment descriptors are representations of local atomic environments) descriptors, respectively. MATS8e is 2D-autocorrelation. The 2D-autocorrelation descriptors that describe the considered property are distributed along the topological structure. RDF125e is RDF descriptors. RDF descriptors are independent of the number of atoms, i.e. the size of a molecule, and these descriptors provide further valuable information, for instance, about bond distances, ring types planar and non-plane systems and atom types. ESPm05d, ESPm14u are edge adjacency indices, which are new topographic indices used in molecular graphs. Molecules as weighted graphs were used for calculation of the novel index, in which the elements of edges set were substituted by the bond orders between connected atoms in the molecule [15] The RMSE of the predicted set in GA-ANN model was 0.0093, which is acceptable in comparison to previous works [28,29].

The predicted values of –logIC50 using the GA-ANN are plotted against the observed values in Figure 4, which indicates a very strong agreement.

The changes of the negative logarithm half maximal inhibitory concentration according to the chosen (i.e. the most effective) descriptors are depicted in

Table 3. Definition of the selected descriptors using GA-ANN Method in gas phase

Description	Definition	Type
BElv3	Lowest eigenvalue n.3 of Burden matrix/weighted by atomic van der Waals volumes	Burden eigenvalues
MWC09	Molecular walk count of order 09	Walk and path counts
MATS8e	Moran autocorrelation- lag8/weighted by atomic Sanderson electronegativities	2D autocorrelations
ESpm05d	Spectral moment 05 from edge adj. matrix weighted by dipole moments	edge adjacency indices
RDF125e	Radial Distribution Function-12.5/weighted by atomic Sanderson electronegativities	RDF descriptors
ESpm14u	Spectral moment 14 from edge adj. matrix	edge adjacency indices
BEHe2	Highest eigenvalue n.2 of Burden matrix/weighted by atomic Sanderson electronegativities	Burden eigenvalues
S3k	3-path kier alpha-modified shape index	Topological descriptors

Figure 5. The figure shows that the descriptors have different effects and ESpm05d is the most important one. Furthermore, the optimum values/ranges of these descriptors can be extracted as given in Table 4. This information can be employed for designing new anti-cancer platinum (IV) drugs. In other words, the results suggest that BElv3, MWC09, MATS8e, BEHe2 and ESpm05d be in their maximum values, while keeping S3k and ESpm14u minimized and RDF125e in the range of 0 to 34.02 in new designs for this class of drugs.

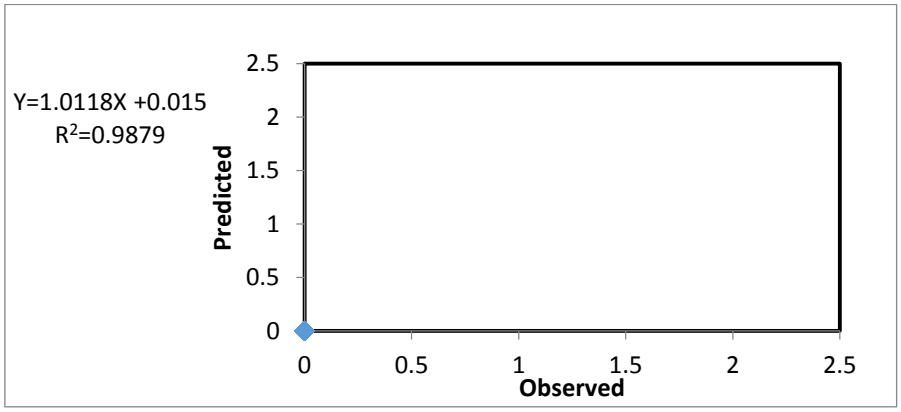
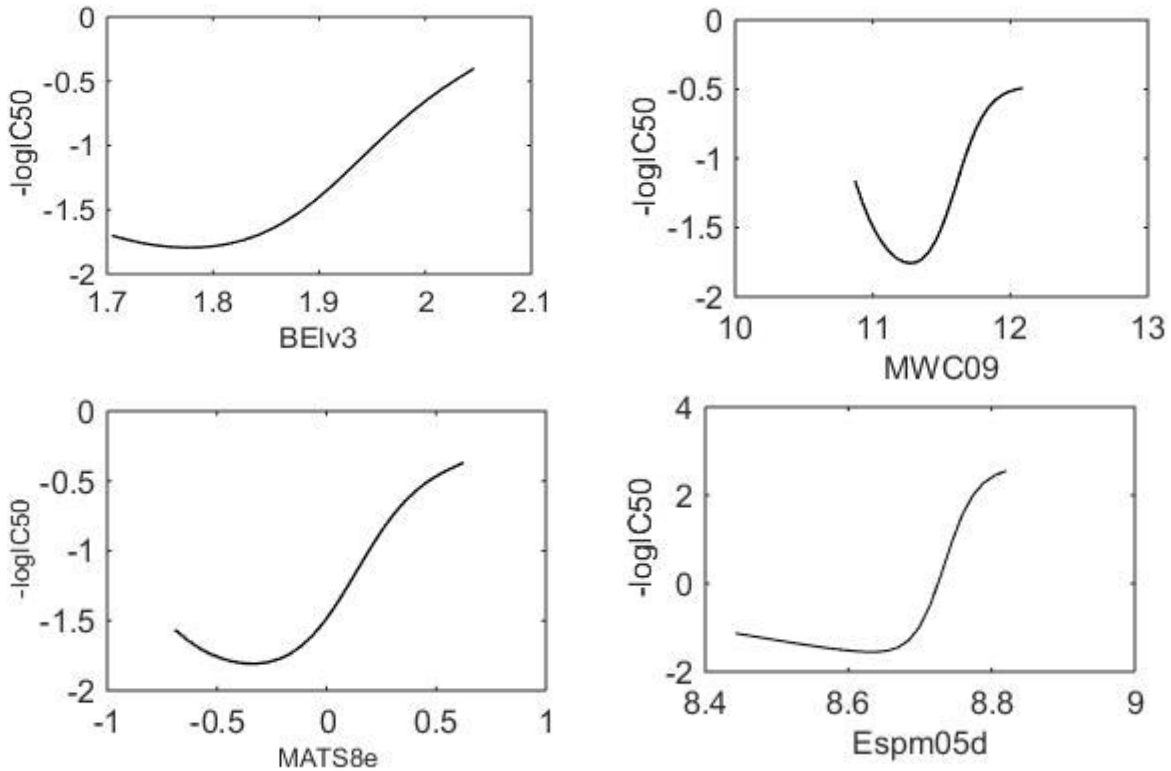


Fig. 4. Predicted versus observed values of –log (IC50) of platinum complexes using GA-ANN approach.



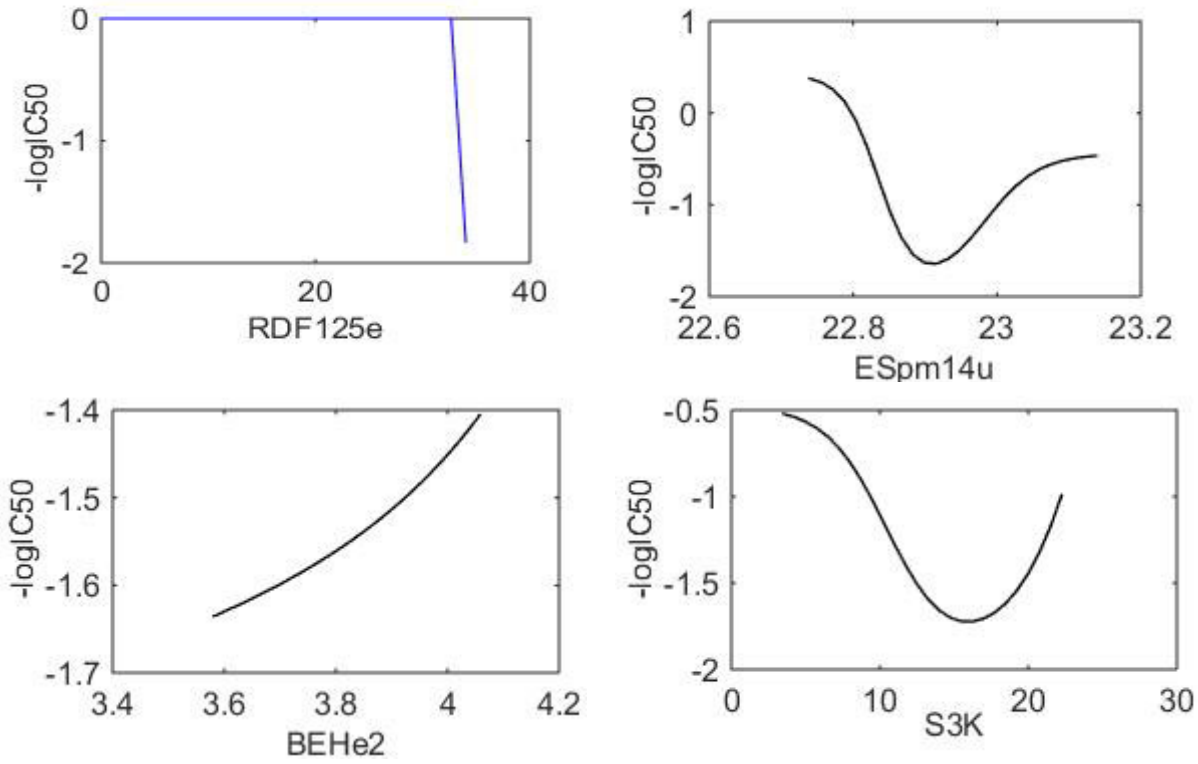


Fig. 5. Changes of –log IC50 against the chosen descriptors.

Table 4. Optimum value/range descriptors in GA-ANN method

Descriptor	range	Optimum value/range
BElv3	1.706 to 2.045	2.045
MWC09	10.875 to 12.081	12.081
MATS8e	-0.687 to 0.620	0.620
ESpm05d	8.440 to 8.820	8.820
RDF125e	0.000 to 36.100	0.000 to 34.021
ESpm14u	22.740 to 23.138	22.740
BEHe2	3.581 to 4.058	4.058
S3k	3.495 to 22.240	3.495

The -logIC50 value grows via increasing BElv3, MATS8e, BEHe2 and ESpm05d Physico-chemical descriptors, thereby, the half maximal inhibitory concentration (IC50) value is reduced. Noting that the aforementioned descriptors are the most effective descriptors among the eight ones in the gas phase, van der Waals volumes, Sanderson electronegativities and dipole moment should be maximized in designing new drugs.

3.2. Result of the Monte Carlo Method

The statistical parameters of the models obtained using molecular graphs (HSG and EC0) and SMILES are shown in Table 5. Performance of the models were compared with each other by the criterion of the predictability in test set (Rm^2), which should be larger than 0.5 [42], correlation coefficient (R^2) in each set, cross-validated correlation coefficient (Q^2) and standard error of estimation (s).

The difference between R^2m and R'^2m

values (ΔRmTEST) was used as another criterion in this issue. The depicted results in table 6 discloses that for all of the three splits, threshold of two and probe 3 gives the best results.

The good correlation between calculated values of –log IC50 using Monte Carlo method and its empirical values, observed in figure 6 approves the appropriateness of the developed model. The variation of correlation coefficient (test set) with respect to the threshold and the number of epochs are plotted in figure 7. This figure confirms that 2 and 60 are the most appropriate values for threshold and number of epochs, respectively.

The distribution of SMILES notations in the train, calibration and test sets are reported in table 6.

Molecular features are sorted according to their correlation weights and are given in table 8. Molecular feature with negative correlation weights are omitted due to their inverse effect on the -logIC50 value. The higher the correlation weigh of a molecular feature, the lower the value of IC50, therefore, the feature is more significant. Definitions of the molecular features are given in table 9. According to table 9, the structural descriptors including presence of double bond, present of Platinum, number of chlorine connected to Pt, branching in molecular skeleton and presence of N and O atoms increase the –logIC50 values (i.e. reduce the half maximal inhibitory concentration (IC50) value).

Table 5. The split models in Monte Carlo Method

Split 1: (T=2)
-logIC50 = -4.7971488 (± 0.1423026) + 0.0953002 (± 0.0032195) * DCW(2,100)
n=18, R^2= 0.8441, Q^2=0.7923, s=0.314 (training set)
n=10, R^2= 0.9193, Q^2=0.8737, s=**0.**579 (calibration set)
n=6, R^2= 0.9172, Q^2=0.8053, s=**0.**805 (test set), R^2m TEST= 0.5507
Spit 2: (T=1)
-logIC50 = -3.2027780 (± 0.0682975) + 0.0910515 (± 0.0019751) * DCW(1,100)
n=16, R^2= 0.9009, Q^2=0.8745, s=0.290 (training set)
n=11, R^2=0.9375, Q^2=0.8840, s=0.468 (calibration set)
n=76, R^2=0.7044, Q^2= 0.4966, s=0.558, R^2m TEST=0.5596
Spit 3: (T=4)
-logIC50 = -3.3819207 (± 0.0763026) + 0.1301079 (± 0.0032454) * DCW(4,100)
n=16, R^2=0.8236 , Q^2=0.7834, s=0.345 (training set)
n=10, R^2=0.9754 , Q^2=0.9663, s=1.38 (calibration set)
n=8, R^2=0.579 , Q^2=0.4367 s=1.82 (test set), R^2m TEST= 0.558

Table 6. Statistical data calculated with both HSG and SMILES for three random splits into test set. Best models are indicated in bold

Threshold	**R^2 test Probe 1**	**R^2 test Probe 2**	**R^2 test Probe 3**	**R^2 test Average**	**Dispersion**
SPLIT 1					
1	0.9198	0.9095	0.9203	0.9165	0.005
2	0.9027	0.9104	**0.9251**	0.9127	0.0093
3	0.9222	0.9128	0.9085	0.9145	0.0057
4	0.9159	0.9163	0.9227	0.9183	0.0031
5	0.9095	0.9118	0.9115	0.9109	0.001
SPLIT2					
1	0.7232	0.7267	0.7242	0.7247	0.0015
2	0.7222	0.7208	**0.7231**	0.7220	0.0009

Threshold	R^2 test Probe 1	R^2 test Probe 2	R^2 test Probe 3	R^2 test Average	Dispersion
3	0.7198	0.7184	0.7213	0.7198	0.0012
4	0.7177	0.7203	0.7188	0.7189	0.0011
5	0.7205	0.7198	0.7199	0.7201	0.0003
SPLIT3					
1	0.6703	0.6814	0.6883	0.68	0.0074
2	**0.6187**	0.6622	0.65	0.6436	0.0183
3	0.6112	0.6033	0.6236	0.6127	0.0084
4	0.7098	0.6869	0.6870	0.6945	0.0108
5	0.6795	0.6720	0.6696	0.6737	0.0043

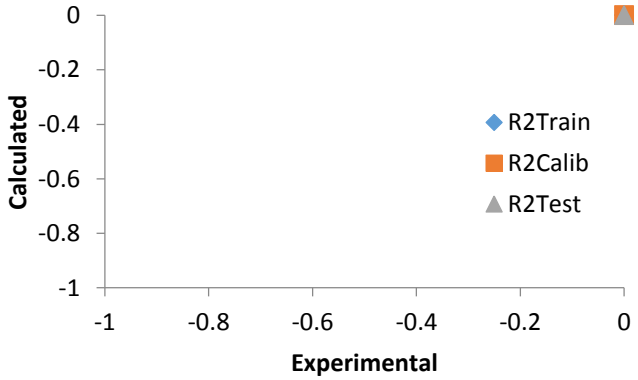


Fig. 6. Correlation between experimental and predicted values of –logIC50 using the Monte Carlo Method.

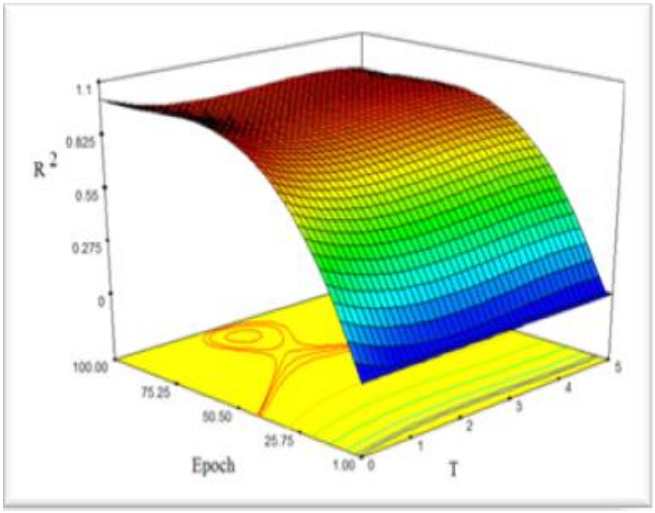


Fig. 7. 3-D surface plot of R^2 according to the threshold and the number of epochs.

Table 7 .SMILES notations 34 compound of Pt(IV) and their corresponding sets

Compound	SMILES	Set
1	CCOC(=O)C(CC1=CC=CC=C1)N|2CCN(C(CC3=CC=CC=C3)C(=O)OCC)|[Pt]|2(Cl)(Cl)(Cl)Cl	Train
2	CCCOC(=O)C(CC1=CC=CC=C1)N|2CCN(C(CC3=CC=CC=C3)C(=O)OCCC)|[Pt]|2(Cl)(Cl)(Cl)C	Train
6	CCCOC(=O)C(CC1CCCCC1)N|2CCN(C(CC3CCCCC3)C(=O)OCCC)|[Pt]|2(Cl)(Cl)(Cl)Cl	Train
9	CCCC(CCC)C(=O)O[Pt]|12(|NC3CCCCC3N|1)(OC(=O)C(CCC)CCC)OC(=O)C(=O)O2	Train
11	N|[Pt](|N)(Cl)(Cl)(OC(=O)CCC(O)=O)OC(=O)NCCCCCCCCCCCCCCCCC	Train
13	N|[Pt](|N)(Cl)(Cl)(OC(=O)NC1CCCC1)OC(=O)NC2CCCC2	Train
15	N|[Pt](|N)(Cl)(Cl)(OC(=O)NC1=CC=CC=C1)OC(=O)NC2=CC=CC=C2	Train

Compound	SMILES	Set
16	N|[Pt](|N)(Cl)(Cl)(OC(=O)NC1=CC=C(C)C=C1)OC(=O)NC2=CC=C(C)C=C2	Train
17	N|[Pt](|N)(Cl)(Cl)(OC(=O)NC1=CC=C(OC)C=C1)OC(=O)NC2=CC=C(OC)C=C2	Train
18	N|[Pt](|N)(Cl)(Cl)(OC(=O)NC1=CC=C(F)C=C1)OC(=O)NC2=CC=C(F)C=C2	Train
19	N|[Pt](|N)(Cl)(Cl)(OC(=O)NC1=CC=C(C=C1)C2=C3C=CC=CC3=CC=C2)OC(=O)NC4=CC=C(C=C4)C5=C6C=CC=CC6=CC=C5~:	Train
21	COC(=O)CCC(=O)O[Pt]|1(|NCCN|1)(Cl)(Cl)OC(=O)CCC(=O)OC	Train
23	+:CCOC(=O)CCCC(=O)O[Pt]|1(|NCCN|1)(Cl)(Cl)OC(=O)CCCC(=O)OCC	Train
25	CCCCOC(=O)CCC(=O)O[Pt]|1(|NCCN|1)(Cl)(Cl)OC(=O)CCC(=O)OCCCC	Train
27	CCCNC(=O)CCC(=O)O[Pt]|1(|NCCN|1)(Cl)(Cl)OC(=O)CCC(=O)NCCC	Train
28	N|1CCN|[Pt]|1(Cl)(Cl)(OC(=O)CCC(=O)NC2CCCC2)OC(=O)CCC(=O)NC3CCCC3	Train
32	CCN|[Pt](|NCC)(Cl)(Cl)(OC(=O)CCC(=O)OCCC)OC(=O)CCC(=O)OCCC	Train
34	+:CCN|[Pt](|NCC)(Cl)(Cl)(OC(=O)CCC(=O)NC1CCCC1)OC(=O)CCC(=O)NC2CCCC2	Train
3	N|[Pt](|N)(Cl)(Cl)(OC(=O)C1=CC=CC=C1)OC(=O)C2=CC=CC=C2	Calib
4	COC(=O)C(CC1CCCCC1)N|2CCN(C(CC3CCCCC3)C(=O)OC)|[Pt]|2(Cl)(Cl)(Cl)Cl	Calib
7	CCCCOC(=O)C(CC1CCCCC1)N|2CCN(C(CC3CCCCC3)C(=O)OCCCC)|[Pt]|2(Cl)(Cl)(Cl)Cl	Calib
12	N|[Pt](|N)(Cl)(Cl)(OC(=O)NC(C)(C)C)OC(=O)NC(C)(C)C	Calib
14	N|[Pt](|N)(Cl)(Cl)(OC(=O)NC1CCCCC1)OC(=O)NC2CCCCC2	Calib
22	CCOC(=O)CCC(=O)O[Pt]|1(|NCCN|1)(Cl)(Cl)OC(=O)CCC(=O)OCC	Calib
29	CCN|[Pt](|NCC)(Cl)(Cl)(OC(=O)CCC(O)=O)OC(=O)CCC(O)=O	Calib
30	CCN|[Pt](|NCC)(Cl)(Cl)(OC(=O)CCC(=O)OC)OC(=O)CCC(=O)OC	Calib
31	CCN|[Pt](|NCC)(Cl)(Cl)(OC(=O)CCC(=O)OCC)OC(=O)CCC(=O)OCC	Calib
33	CCN|[Pt](|NC(C)C)(Cl)(Cl)(OC(=O)CCC(=O)OC(C)C)OC(=O)CCC(=O)OC(C)C	Calib
5	CCOC(=O)C(CC1CCCCC1)N|2CCN(C(CC3CCCCC3)C(=O)OCC)|[Pt]|2(Cl)(Cl)(Cl)Cl	Test
8	CCCC(CCC)C(=O)O[Pt]|12(|NC3CCCCC3N|1)(O)OC(=O)C(=O)O2	Test
10	N|[Pt](|N)(Cl)(Cl)(OC(=O)CCC(O)=O)OC(=O)NCCCCCCCCCCCC	Test
20	N|1CCN|[Pt]|1(Cl)(Cl)(OC(=O)CCC(O)=O)OC(=O)CCC(O)=O	Test
24	CCCOC(=O)CCC(=O)O[Pt]|1(|NCCN|1)(Cl)(Cl)OC(=O)CCC(=O)OCCC	Test
26	N|1CCN|[Pt]|1(Cl)(Cl)(OC(=O)CCC(=O)NCCO)OC(=O)CCC(=O)NCCO	Test

Table 8. SMILES attributes with positive correlation weights for split 1

SMILES attributes	CWs	SMILES attributes	CWs
(...(.....	1.09868	[...Pt.....	4.97087
=...(.....	4.30333	...N.....	7.87514
C...(.....	1.59722	...[.....	1.91161
C...C.....	1.13491	EC0-Cl..1..	0.31632
BOND10000000	5.16467	Cl..(.....	3.75072
EC0-C...2...	0.51325	Cl.....	1.16651
EC0-C...3..	1.77541	...(.....	3.31111
EC0-O...2...	3.02902	EC0-C...1..	0.41414
EC0-Pt..6…	5.20013	C...2.....	3.21252
N...C.....	1.53331	[...(.....	6.83496
O...=.....	5.36745	=...1.....	2.37866
O...C.....	1.34930	EC0-N...1..	4.74585
NOSP11000000	5.30952	C...3.....	0.12818
Pt.....	5.32754	...1.....	3.60455
[.....	3.93797	=...3.....	0.55311

Table 9. Definition of the promoter of A_k

Attribute Ak	Comment
HALO00000000	Absence of F, Cl, Br
C...C.....	Presence of carbon – carbon bonds (sp3)
C…(…C…	SP3 Carbon atoms with branching
++++O---B2==	Presence of oxygen and double bonds
C…=…….	SP2 Carbon atom
(.....	Branching in molecular skeleton
O.....	Presence of oxygen

Attribute Ak	Comment
1.....	Presence of rings
++++N---B2==	Presence of nitrogen and double bond
=	Double bond
@	Stereo specific bond
#	Triplet bond

4. CONCLUSION

In this study, six different approaches were employed to study the structure-activity relationships of 34 Platinum (IV) anticancer drugs. The GA-ANN showed the best performance among the considered approaches. The results proved that BElv3, MWC09, MATS8e, ESpm05d, RDF125e, ESpm14u, BEHe2 and S3k descriptors in the gas phase were the most significant descriptors in defining the biological activity of Platinum (IV) compounds. Furthermore, the optimum values/ranges of the chosen descriptors were presented. The obtained results can be employed for designing new anti-cancer drugs. Additionally, Monte Carlo method was employed to find out the quality of the effects of structural descriptors on the biological activity of the studied drugs. It can be said that simultaneous use of Monte Carlo and GA-ANN methods gives deeper and more comprehensive knowledge of the effect of molecular and structural descriptors on the activity of drugs and provides better insights for designing new drugs.

In GA-ANN method the BElv3, MWC09, MATS8e, ESpm05d, ESpm14u and BEHe2 Physico-chemical descriptors were found to have the most important role on the drug activity. In Monte Carlo method, the structural descriptors including presence of double bond, present of Platinum, presence of N and O are important. Precise setting of these Physico-chemical and structural descriptors can reduce the half maximal inhibitory concentration (IC50).

5. ACKNOWLEDGEMENT

The support for this study provided by the Islamic Azad University of Rasht and University of Guilan are gratefully acknowledged.

6. REFERENCES

[1] L.R. Kelland, Nature Rev., 7 (2007); 573-584.
[2] R. Dolman, T.W. Hambley, G.B. Deacon., J. Inorg. Biochem. 88 (2002); 260-267.
[3] H. R. Mellor, S. Snelling, M. D. Hall, S. Modok, M. JaVar, T.W.Hambley, R. Callaghan, Biochem. Pharmacol. 70 (2005); 1137-1146.
[4] R. Sayyadi kord Abadi, A. Alizadehdakhel and S. Tajadodi Paskiabei, J. Korean Chem. Soc. 60 (2016); 225.
[5] R. Sayyadi kord Abadi, A. Alizadehdakhel, S. Dorani Shiraz, Russ. J. Physic. Chem. B, 11 (2017; 307.
[6] V.O.Černý., J. Optimiz. Theory. App. 45(1985); 41-51.
[7] L.M. Schmitt., Theor. Comput. Sci., 259(2001); 1-61.
[8] D. Bertsimas, J. Tsitsiklis, Statistical Science, 8(1983); 10-15.
[9] N.A. Meanwell , O.B. Wallace, H. Fang., H. Wang, M.Deshpande, T. Wang,Z. Yin., L. Zadjura, D.L.Tweedie, S.Yeola, F. Zhao, S.Ranadive, B.A. Robinson , Y.F.Gong, H.G. Wang , T.P. Spicer, W.S. Blair, P.Y. Shi., R.J. Colonno.P.F. Lin., Bioorg. Med. Chem. Lett. 19 (2009); 1977-1981.
[10] Toropova, A.P.; Toropov, A. A; Benfenati, E.; Gini, G.; Leszczynska, D.; Leszczynski, J. J. Comput. Chem, 32 (2011); 2727.
[11] E. B. DeMelo, M .M. Ferreira., Eur, J.Med.Chem. 44 (2009); 3577-3583.
[12] M. J. Frisch, G.W. Trucks, H.B. Schlegel,

G.E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr.J.E. Peralta, F. Ogliaro, M. Bearpark, J. J Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J. Dannenberg, S. Dapprich, A.D. Daniels, Ö. Farkas,.; J. B. ForesmanJ.V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian 09 (Gaussian, Inc., Wallingford CT, 2009).

[13] https://gaussian.com/glossary/g09/

[14] Dragon 3.0 Evaluation Version. Available online: http://www.disat.unimib.it/chm

[15] R. Todeschini, Milano chemometrics, QSAR Group, http://www.disat.unimib.it/chem.

[16] R. Todeschini, V. Consonni, Handbook of Molecular Descriptors (Wiley-VCH), 2000.

[17] A. M. Veselinović, A. A. Toropov, A. P. Toropova, I. Damnjanović, G. M. Nikolić, Scientific Journal of the Faculty of Medicine in Niš, 31(2014);95-103.

[18] V.Consonni, R.Todeschini, M. Pavan. P, Gramatica., J. Chem. Inf. Comput .Sci. 2002, 42, 693-705.

[19] P.Gramatica, V. Consonni. R. Todeschini, Chemosphere, 41(2000); 63-777.

[20] SPSS, Version 19, available at http://www.spssscience.com, (2010).

[21] M.H. Fatemi, S. Gharaghani, Bioorg. Med. Chem. 15(2007); 7746-7754.

[22] N. Kenneth, J.Am. Stat. Assoc, 72 (1997); 865-866.

[23] M. Jalali-Heravi, M.F. Parastar, J. Chem. Inf. Comput. Sci. 40 (2000); 147.

[24] Y-L. Xie, J.-H. Kaliva, Analytica Chemica Acta, 348 (1997); 19-27

[25] L.M. Schmitt, Theor. Comput. Sci., 259 (2001);1-61.

[26] (http://www.insilico.eu/coral)

[27] S. H. Sadat Hayatshahi, P. Abdolmaleki, M. Ghiasi, S. Safarian, FEBS Lett 581 (2007); 506-514.

[28] H. Varbanov; etal, Eur. J. Med. Chem. 46 (2011); 5456-5464.

[29] T. Asadollahi, S. Dadfarnia, A.M. Haji Shabani, J.B. Ghasemi. MATCH Commun. Math. Comput. Chem. 71(2014); 287-304.

[30] A.M.Veselinović, JB. Milosavljević, AA. Toropov, G. M. Nikolić, Eur. J. Pharm. Sci. 48 (2013), 532-41.

[31] A. Bakalova, H. Varbanov, R. Buyukliev, G. Momekov, D. Ivanov, I. Doytchinov., Arch. Pharm. Chem. Life Sci. 11 (2011); 209–216 .

[32] P. Sarmah, R.C. Deka, J Comput. Aided. Mol. Des. 23 (2009); 343–354.

[33] H. P. Varbanov, M. A. Jakupec, A. Roller, F. Jensen, M. Galanski, B. K. Keppler, J. Med. Chem. 56 (2013); 330−344.

[34] P. Gramatica, E. Papa, M. Luini., E. Monti, M. B. Gariboldi, M. Ravera, E. Gabano, L. Gaviglio, D. Osella, J. Biol. Inorg. Chem. 15 (2010); 1157–1169.

[35] D. Dimitrijevic, et al., Inorganica Chimica Acta, 402 (2013); 83-89.

[36] S. L. Yoong, Biomaterials, 35 (2014); 748-759.

[37] L. E. Mihajlovic, Int. J. Electrochem. Sci, 8 (2013); 8433-8441.

[38] V. Novohradsky, et al., J. Inorg. Biochem., 140 (2014); 72-79.

[39] Y-R. Zheng, et al, J. Am. Chem. Soc, 136 (2014); 8790-8798.

[40] J. J. Wilson, etal., Inorg. Chem., 50 (2011), 3103-3115.

[41] M. R. Reithofer, et al., J. Inorg. Biochem. 105 (2011); 46-51.

[42] A. Golbraikh, A.Tropsha, A.; J. Mol. Graph. Model. 20 (2002), 269-276.

[43] K. Levenberg, Quarterly of Applied Mathematics, 2 (1944), 164-168.

[44] İ. Yılmaz, N. Acar-Selçuk, S. J. Coles, F. Pekdemir, A. Şengül. J. Mol. Struct. 1223 (2021); 129271.
[45] A. M. .Fathi, H. S.Mandour, El. Hassane Anouar. *J. Mol. Struct.* 1224 (2021); 129263.
[46] S. Baskaran, M.M Krishnan, R. Kumar, J. Molecul. Struct, 1224 (2021); 129236.
[47] A. Abkari, I. Chaabane, K. Guidara, Physica E: Low-dimensional Systems and Nanostructures, 81 (2016); 136-144.
[48] I. Yılmaz, N. Acar-Selçuki, A. Şengül, J. Molecul. Struct., 1223 (2021), 129271.
[49] R. E. Hag, M. M Abdusalam, C, Aclian, H. Kayi, S. Özalp-Yaman., Polyhedron., 170 (2019), 25-33.
[50] G. P. Rosa., A. Palmeira., D. I. S. P. Resende, I. F. Almeida, A. Kane-Pagès, M. C. Barreto., E. Sousa, M. M. M. Pinto., Bioorganic & Medicinal Chemistry. 29 (2021); 115873.
[51] E. Pourbasheer., R. Aalizadeh, M. R. Ganjali., Arabian Journal of Chemistry, 12 (2015); 2141-2149.
[52] MD. Hall, RA. Alderden, M. Zhang, PJ. Beale, Z. Cai, B. Lai, APJ. Stamp, TW. Hambley, J. Struct. Biol, 155 (2006), 38–44.
[53] MD. Hall MD, HR. Mellor, R. Callaghan, TW. Hambley, J. Med. Chem. 50 (2007); 3403–3411.
[54] A. M. Montan, C. Batalla, Curr. Med. Chem, 16 (2009); 2235–2260
[55] M. D. Hall, S. Amjadi, M. Zhang, P.J. Beale, T.W. Hambley, J. Inorg.Biochem. 98 (2004);1614–1624.
[56] S. Hosseini, M.R. Gholami, M. Haghgu., J. Phys. Theor. Chem. IAU Iran ., 13 (2016), 171-177.
[57] L. Mahdavi, J. Phys. Theor. Chem. IAU Iran., 14 (2017), 103-110.

مجله شیمی فیزیک و شیمی نظری
دانشگاه آزاد اسلامی واحد علوم و تحقیقات
جلد ۱۸، شماره ۱، بهار ۱۴۰۰
ISSN ۱۷۳۵-۲۱۲۶

مونت کارلو و مطالعه QSAR فعالیت بیولوژیکی چند داروی ضد سرطان پلاتین (IV)

ربابه صیادی کردابادی[۱*]، عبدالله فلاح شجاع[۲]، اصغر علیزاده داخل[۱]، لیلا محمدی نرگسی[۲]، قاسم قاسمی[۱]

[۱] گروه شیمی و مهندسی شیمی، واحد رشت، دانشگاه آزاد اسلامی، رشت، ایران
[۲] گروه شیمی، دانشکده علوم، دانشگاه گیلان، P.O.BOX 1914، رشت، ایران

چکیده

مطالعه QSAR برخی از مشتقات پلاتین (IV) با استفاده از رگرسیون خطی چندگانه (MLR) و شبکه عصبی مصنوعی (ANN) به عنوان ابزارهای مدل سازی، همراه با الگوریتم‌های بهینه‌سازی شبیه سازی (SA) و الگوریتم ژنتیک (GA) انجام شد. علاوه بر این، از نرم‌افزار CORAL برای ارتباط فعالیت بیولوژیکی با پارامترهای ساختاری داروها استفاده شد. نتایج بدست آمده از روش‌های مختلف مقایسه شد و روش ترکیبی GA-ANN با توجه به ضریب همبستگی (R^2) و میانگین خطای مجموع مربعات (RMSE) بهترین عملکرد را نشان داد. از روش GA-ANN مشخص شد که MTAS8e، ESpm05d، BElv3، WC09، ESpm14u، BEHe2، RDF125e و S3K مهمترین توصیف کننده‌ها هستند. از شبیه سازی مونت کارلو ، مشخص شد که وجود پیوند دوگانه، پلاتین، تعداد کلر متصل به Pt، شاخه فرعی در مولکول و وجود اتم‌های N و O مهمترین ویژگی‌های مولکولی موثر بر فعالیت بیولوژیکی دارو هستند. نتیجه‌گیری شد که استفاده همزمان از روش QSAR و مونت کارلو می‌تواند به درک جامع‌تری از رابطه بین توصیف کننده‌های فیزیکی – شیمیایی، ساختاری یا تئوری داروها با فعالیت‌های بیولوژیکی آنها منجر شود.

کلید واژه‌ها: داروهای ضد سرطان پلاتین (IV)، QSAR، الگوریتم ژنتیک، روش مونت کارلو

[*] مسئول مکاتبات: Sayyadi@iaurasht.ac.ir , Sayyadi_04@yahoo.com