DESIGN, SCREENING AND EXPLORING NOVEL METAL-THIOSEMICARBAZONE COMPLEXES TARGETING HT-29 CELL LINE USING QSPR-BASED *IN SILICO* MODELING

NGUYEN MINH QUANG[1,*], PHAM VAN TAT[2]
[1]*Faculty of Chemical Engineering, Industrial University of Ho Chi Minh City*
[2]*Institute of Pharmaceutical Education and Research, Binh Duong University*
** Corresponding author: nguyenminhquang@iuh.edu.vn*
DOIs: https://www.doi.org/10.46242/jstiuh.v79i1.5804

Abstract. *In silico* models were used to calculate the stability constants ($\log\beta_{11}$) of metal ions and thiosemicarbazone in complex solutions. The 0-3D molecular descriptors, physicochemical, and quantum parameters were produced using the semi-empirical quantum computation PM7 and PM7/sparkle as well as the molecular geometric structure. Multivariate linear regression (MLR) and artificial neural networks (ANN) were used to build the quantitative structure and property relationship (QSPR). The constructed $QSPR_{MLR}$ model involves descriptors such as: *k0, core-core repulsion, xvp10, ^{5}C, and xch5*. The statistical values of the model were pointed out: R^2_{train} = 0.847, R^2_{adj} = 0.834, Q^2_{LOO} = 0.764, SE = 1.371, and F-stat = 66.20. The neural network $QSPR_{ANN}$ I(5)-HL(8)-O(1) model was also developed with the statistical values: R^2_{train} = 0.976, Q^2_{test} = 0.956, and $Q^2_{validation}$ = 0.926. A series of new thiosemicarbazone derivatives and complexes of this ligand and metal ions were designed using these models. These complexes were screened using the Applicability Domains (AD) and Outliers technique. The ability prediction of models applied complexes on the test data set agrees with those from the experimental literature. Besides, the results of fifteen new complexes were selected because they were within the predictive application domain. Next, nine of fifteen new complexes overcame drug-likeness analysis of Lipinski and Veber rules. Finally, nine typical complexes simulated the docking on protein for anti-colorectal cancer (code: 6GUE-PDB). The results screened out six novel complexes considered potential inhibitors to the HT-29 cell line in supporting colorectal cancer treatment.

Keywords. Metal-thiosemicarbazone complexes, quantitative structure-property relationship model, colorectal cancer, docking.

1. INTRODUCTION

Thiosemicarbazone (TIO) ligands and related metal-TIO complexes have several real-world uses. Derivatives of TIO are widely used analytical agents [1] and have shown great biological activity [2]. These compounds and their metal ion complexes are recognized for their medicinal properties, which include antibacterial, antifungal, antimalarial, antitumor, and antiviral effects [2]. Additionally, one of the most important parameters for complexes is the stability constant, which indicates their stability in solutions. It gauges how strongly ligands and metal ions bind, causing different complexes to form. Furthermore, the stability constant is essential for elucidating phenomena like response mechanisms and the many characteristics of biological systems.

By knowing the stability constant, we can calculate the equilibrium concentrations of the components in a solution. Changes in the structure of the complexes can be predicted using initial concentrations of metal ions and ligands. In recently, metal-TIO complexes' the stability constants of have been estimated using UV/Vis spectrophotometry alongside computational techniques [3]. Furthermore, the links between structural descriptors and their attributes are used in theoretical methods to predict stability constants [4]. Quantum mechanics has been used in this context to identify several complicated descriptors involving metal ions and thiosemicarbazone ligands [5].

In recent years, computers have emerged as valuable tools for intense calculations across various fields of chemistry, including inorganic, analytical, organic, physical chemistry, material simulation, and data mining; specially, computer-aided molecular design has expedited the discovery process of material properties and has helped minimize the traditional trial-and-error methods [6-7]. Additionally, *in silico* models, such as conformational search techniques and quantitative structure-property relationships (QSPR), have greatly aided new molecule research and discovery. Consequently, multivariate analysis methods have become valuable and accessible tools for supporting empirical and theoretical models, allowing for assessing different characteristics within chemical systems through multivariable linear relationships.

Colorectal cancer (CRC) is the third most commonly diagnosed cancer worldwide and represents the second leading cause of cancer-related mortality [8]. By 2040, 3.2 million CRC cases will likely be diagnosed annually, up from around 1.9 million new cases and 935,000 deaths in 2020. CRC is prevalent in both men and women, ranking third in morbidity and second in mortality among cancer types. It accounts for 10% of cancer cases in men and 9.2% in women worldwide [8]. While treatments such as Bevacizumab, Camptosar, and Cetuximab are used to manage CRC, their effectiveness can vary significantly from patient to patient. Additionally, these drugs can have different side effects, highlighting the need for novel therapies that offer improved efficacy and fewer adverse effects. *In silico* drug design methods present a faster and more cost-effective lead identification and optimization approach. Therefore, this study aimed to design novel thiosemicarbazone ligands and their complexes using *in silico* design techniques for CRC therapy.
This study present building the quantitative structure and properties correlations between stability constants of metal-TIO complexes and the descriptors. Molecular mechanics, connectivity computation, and the semi-empirical quantum mechanics methods with new version of PM7 and PM7/sparkle [9] are used to compute the structure descriptors. Structural descriptors and least squares methods are used to create the multivariate linear models ($QSPR_{MLR}$). Using a multilayer perceptron (MLP) algorithm and the error back-propagation approach, the artificial neural network models ($QSPR_{ANN}$) are developed with the input variables descriptors of the best-chosen $QSPR_{MLR}$ model. As the models are being constructed, the stability constant $\log\beta_{11}$ of the metal-thiosemicarbazone complexes in the test set that the QSPR models produce is verified and contrasted with those found in experimental data published in the literature. The results obtained from the models were utilized to design a series of novel ligands and complexes. These metal-thiosemicarbazone complexes were examined for their potential anti-colon cancer effects using drug screening techniques, including Lipinski [10] and Veber [11] rules and molecular docking methods. Ultimately, this study produced promising results, indicating that the complexes could serve as potential therapeutics for colorectal cancer.

2. METHODOLOGY

2.1. Selection of complexes

A substitution process is a metal ion (M) and a thiosemicarbazone ligand (L) producing a complex in an aqueous solution. This equilibrium phenomenon showcases the dynamic interactions that underpin coordination chemistry and highlights the versatility and significance of such ligand-metal combinations in various chemical applications.

$$p\ \mathrm{M} + q\ \mathrm{L} \rightleftharpoons \mathrm{M}_p\mathrm{L}_q \qquad (1)$$

When the complex is formed from the reagents, the overall or stability constant, represented by the symbol β_{pq}, is the constant. For the production of M_pL_q, when p = 1 and q = 1, the stability constant provided by the complex ML is formed in a single step, as indicated by the following stability constant β_{11}.

$$\beta_{11} = \frac{[\mathrm{ML}]}{[\mathrm{M}][\mathrm{L}]} \qquad (2)$$

Figure 1: (a) Thiosemicarbazone ligand; b) Metal-thiosemicarbazone complex

2.2. Data set and descriptors computations

Literature [12-22] provided the $\log\beta_{11}$ values of complexes between metal ions and the ligand thiosemicarbazone (Table 1). On the MoPac2016 system, the metal ion and ligand thiosemicarbazone complexes were reconstructed and optimized using quantum mechanics [9]. The semi-empirical quantum technique was used to derive the quantum descriptors for lanthanides using the updated PM7 and PM7/sparkle. The QSARIS system was used to compute the 0-3D descriptors. $QSPR_{MLR}$ models were

created using the forward regression and back-elimination method on the Regress tool [23] added-in Excel software. The Matlab2016 [24] platform's multilayer training method was used to build the artificial neural network model $QSPR_{ANN}$. The leave-one-out approach (LOO) was used to cross-validate the predictability of QSPR models using the statistic Q^2_{LOO}.

2.3. QSPR modeling

Multivariate linear regression

Multivariate linear regression (MLR) models can predict the values of dependent quantitative or qualitative variables through a linear combination of explanatory quantitative and qualitative factors. This method overcomes the drawbacks of ordinary least squares regression regarding the ratio of variables to data. MLR is often referred to simply as linear regression. The following is the expression for the regression model with k explanatory variables in this context [6,23]:

$$Y = b_0 + \sum_{i=1}^{k} b_i . X_i + \varepsilon \qquad (3)$$

where X_i is an independent variable, k is the number of variables in the equation, ε is an error, b_0 is the model's intercept, b_i is a slope connected to X_i, and Y is the dependent variable.

Artificial neural network

A neural network comprises hidden nodes, which are derived inputs. These hidden nodes are made up of nonlinear functions obtained from the initial inputs. A neural network can have several levels of hidden nodes, with several hidden nodes in each layer. Activation functions are applied at the nodes inside the hidden layers. An activation function transforms a linear combination of the input variables. The function used in the output layer is usually a logistic transformation for nominal or ordinal outcomes or a linear combination of continuous outcomes [25]. Its primary benefit is the neural network model's ability to model various response surfaces effectively. In addition to being highly adaptable models, neural networks often overfit data. The work selected the standard feed-forward neural network, specifically the multi-layer perceptron (MLP), for training using the error back-propagation approach. The MLP was chosen because it is the most basic and widely used ANN architecture, suitable for a variety of applications. There are also many other ANN designs tailored to different tasks [26].

Table 1: 66 stability constants ($\log\beta_{11,exp}$) values of experimental metal-thiosemicarbazone

Ligand				Metal ions	$\log\beta_{11,exp}$ (ref.)	Ligand				Metal ions	$\log\beta_{11,exp}$ (ref.)
R_1	R_2	R_3	R_4			R_1	R_2	R_3	R_4		
H	$-C_5H_4N$	H	H	Ni^{2+}	5.630 [12]	$-CH_3$	$-C_5H_4N$	H	H	Pr^{3+}	7.330 [18]
H	$-C_5H_4N$	H	H	Mn^{2+}	4.320 [12]	$-CH_3$	$-C_5H_4N$	H	H	Nd^{3+}	7.590 [18]
H	$-C_8H_7$	H	H	Cd^{2+}	5.544 [13]	$-CH_3$	$-C_5H_4N$	H	H	Gd^{3+}	7.630 [18]
H	$-C_6H_4OH$	H	H	V^{5+}	5.322 [14]	$-CH_3$	$-C_5H_4N$	H	H	Sm^{3+}	7.790 [18]
H	$-C_5H_4N$	H	H	Co^{2+}	5.360 [15]	$-CH_3$	$-C_5H_4N$	H	H	Tb^{3+}	8.060 [18]
H	$-C_5H_4N$	H	H	Zn^{2+}	5.230 [15]	$-CH_3$	$-C_5H_4N$	H	H	Dy^{3+}	8.080 [18]
H	$-C_4H_3O$	H	H	Co^{2+}	5.099 [16]	$-CH_3$	$-C_5H_4N$	H	H	Ho^{3+}	8.150 [18]
H	$-C_8H_7$	H	H	Mo^{6+}	6.551 [17]	$-CH_3$	$-C_5H_4N$	H	H	La^{3+}	7.140 [18]
$-CH_3$	$-C_5H_4N$	H	H	La^{3+}	7.600 [18]	$-CH_3$	$-C_5H_4N$	H	H	Pr^{3+}	7.400 [18]
$-CH_3$	$-C_5H_4N$	H	H	Pr^{3+}	7.760 [18]	$-CH_3$	$-C_5H_4N$	H	H	Nd^{3+}	7.740 [18]
$-CH_3$	$-C_5H_4N$	H	H	Nd^{3+}	7.950 [18]	$-CH_3$	$-C_5H_4N$	H	H	Gd^{3+}	7.840 [18]
$-CH_3$	$-C_5H_4N$	H	H	Gd^{3+}	8.160 [18]	$-CH_3$	$-C_5H_4N$	H	H	Sm^{3+}	7.950 [18]
$-CH_3$	$-C_5H_4N$	H	H	Sm^{3+}	8.260 [18]	$-CH_3$	$-C_5H_4N$	H	H	Tb^{3+}	8.060 [18]
$-CH_3$	$-C_5H_4N$	H	H	Tb^{3+}	8.340 [18]	$-CH_3$	$-C_5H_4N$	H	H	Dy^{3+}	8.240 [18]
$-CH_3$	$-C_5H_4N$	H	H	Dy^{3+}	8.490 [18]	$-CH_3$	$-C_5H_4N$	H	H	Ho^{3+}	8.260 [18]
$-CH_3$	$-C_5H_4N$	H	H	Ho^{3+}	8.640 [18]	H	$-C_6H_5$	H	H	Ag^{+}	15.50 [19]
$-CH_3$	$-C_5H_4N$	H	H	La^{3+}	6.820 [18]	H	$-C_5H_4N$	H	H	Ag^{+}	14.00 [19]
$-CH_3$	$-C_5H_4N$	H	H	Pr^{3+}	7.050 [18]	H	$-C_6H_4OH$	H	H	Ag^{+}	15.60 [19]
$-CH_3$	$-C_5H_4N$	H	H	Nd^{3+}	7.380 [18]	H	$-C_6H_5$	H	H	Cu^{2+}	17.70 [20]

$-CH_3$	$-C_5H_4N$	H	H	Gd^{3+}	7.510 [18]	H	$-C_5H_4N$	H	H	Cu^{2+}	20.40 [20]
$-CH_3$	$-C_5H_4N$	H	H	Sm^{3+}	7.600 [18]	$-CH_3$	$-C_2H_4NO$	H	H	Cu^{2+}	19.10 [20]
$-CH_3$	$-C_5H_4N$	H	H	Tb^{3+}	7.860 [18]	$-CH_3$	$-C_6H_4OH$	H	H	Mg^{2+}	3.300 [21]
$-CH_3$	$-C_5H_4N$	H	H	Dy^{3+}	7.880 [18]	$-CH_3$	$-C_6H_4OH$	H	H	Mg^{2}	3.030 [21]
$-CH_3$	$-C_5H_4N$	H	H	Ho^{3+}	7.950 [18]	$-CH_3$	$-C_6H_4OH$	H	H	Mg^{2}	2.920 [21]
$-CH_3$	$-C_5H_4N$	H	H	La^{3+}	6.910 [18]	$-CH_3$	$-C_6H_4OH$	H	H	Cd^{2+}	5.590 [21]
$-CH_3$	$-C_5H_4N$	H	H	Pr^{3+}	7.120 [18]	$-CH_3$	$-C_6H_4OH$	H	H	Cd^{2+}	4.830 [21]
$-CH_3$	$-C_5H_4N$	H	H	Nd^{3+}	7.540 [18]	$-CH_3$	$-C_6H_4OH$	H	H	Cd^{2+}	4.740 [21]
$-CH_3$	$-C_5H_4N$	H	H	Gd^{3+}	7.570 [18]	H	$-C_6H_4NH_2$	H	H	Co^{2+}	11.95 [22]
$-CH_3$	$-C_5H_4N$	H	H	Sm^{3+}	7.660 [18]	H	$-C_6H_4NH_2$	H	H	Co^{2+}	9.870 [22]
$-CH_3$	$-C_5H_4N$	H	H	Tb^{3+}	7.810 [18]	H	$-C_6H_4NH_2$	H	H	Mn^{2+}	12.14 [22]
$-CH_3$	$-C_5H_4N$	H	H	Dy^{3+}	7.930 [18]	H	$-C_6H_4NH_2$	H	H	Mn^{2+}	9.990 [22]
$-CH_3$	$-C_5H_4N$	H	H	Ho^{3+}	8.020 [18]	H	$-C_6H_4NH_2$	H	H	Zn^{2+}	11.32 [22]
$-CH_3$	$-C_5H_4N$	H	H	La^{3+}	7.020 [18]	H	$-C_6H_4NH_2$	H	H	Zn^{2+}	8.740 [22]

The investigation employed a MLP type with a single hidden layer as our artificial neural network (ANN) architecture. The model consists of an input layer, a hidden layer, and an output layer. The models were trained the network using a standard feed-forward approach and an error back-propagation learning technique. Additionally, various transfer functions were utilized for training the ANN models, including logistic, exponential, and hyperbolic tangent functions [26].

The output generated by the network is compared to the actual values obtained from experimental findings after the trained artificial neural network (ANN) model until the mean square error (MSE_{ANN}) is minimized. The average squared error between the target outputs (t) and the network's outputs (o) is MSE_{ANN}. This is represented as follows [24]:

$$MSE_{\mathrm{ANN}} = \frac{1}{n}\sum_{1}^{n}\left(t_i - o_i\right)^2 \tag{4}$$

2.4. Validation of models

Validation is a critical step in QSPR modelling, as it determines the reliability and predictive power of the models developed. This process evaluates how effectively a QSAR model can predict the activity of new, unseen compounds. Without appropriate validation, a model may demonstrate strong performance on the data used for its creation but may not generalize well to new data, resulting in unreliable predictions [6].

QSPR validation typically involves two main categories: internal and external validation. Internal validation assesses the model's stability and predictive ability using only the data used to build the model. Meanwhile, external validation is considered the most reliable way to assess a QSPR model's true predictive power. It involves testing the model on a dataset not used to build or select the model [6].

Additionally, the average absolute values of the relative error (MARE,%), which are utilized in the work to evaluate the overall error of the QSPR models, are computed using the following formula [27].

$$MARE,\% = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|\log\beta_{11,\mathrm{exp}} - \log\beta_{11,\mathrm{cal}}\right|}{\log\beta_{11,\mathrm{exp}}}100,\% \tag{5}$$

n is the number of test substances; $\beta_{11,exp}$ and $\beta_{11,cal}$ are the experimental and calculated stability constants.

2.5. Development of new ligands and complexes

In order to develop new thiosemicarbazone, the study chose carbazole and phenothiazine derivatives and added these groups to the R_4 position of the structure. Hydrogen atoms occupy the remaining places (R_1, R_2, and R_3). Ions metal such as Ag^+, Cu^{2+}, Cd^{2+}, Ni^{2+}, and Zn^{2+} are common metal ions forming complexes with newly developed ligands. These derivatives were selected due to their comparable uses to the original derivatives, particularly their antiviral and antibacterial qualities [28-29]. This is a long-term line of inquiry into thiosemicarbazone derivative activity. Furthermore, practical synthesis of these compounds has been carried out, and complete applications in numerous sectors have been documented [29-30].

Forty four of the twenty two phenothiazine and twenty two carbazole compounds that were created in a fruitful investigation formed 220 complexes with the five metal ions that were previously discussed. The

application domain (AD) and outliers was then tested by carefully screening the new complexes and embedding them in the data space of the original training set [6,31].

2.6. ADMET analysis and Molecular docking

Drug-likeness is the intricate balancing act between structural and molecular characteristics that influence a molecule's pharmacological activity. Drug-likeness evaluation identifies prospective candidates for drug development from among new metal-thiosemicarbazone compounds with anti-colorectal cancer activity above the QSPR results. Lipinski and Veber criteria predict drug-likeness based on several characteristics, including molecular weight (MW<500 Da), high lipophilicity (MLogP<5), fewer than five hydrogen bond donors (HBD<5), and fewer than ten hydrogen bond acceptors (HBA<10). Molecular refraction (40<MR<130); no molecular weight (160<MW<480); a range of lipophilicity (-0.4<WLOG<5.6); and hefty atomic number (20<atoms<70).

In the meantime, the interactions between proteins (enzymes) and ligands are predicted using a molecular simulation technique called docking. We computed the binding energy (E_bonding, kcal/mol) and the root mean square deviation (RMSD, Å) to assess the novel derivatives' capacity to bind to the active sites. These calculations were performed by docking the derivatives against colon cancer cells, which is recognized as a signalling agent of inflammatory cytokines in colorectal cancer. The procedure was adequate if the RMSD value was less than 1.5 Å and the binding energy was less than -6.5 kcal.mol^{-1} [32]. Additionally, the interactions between amino acids and their corresponding binding energies were used to evaluate the ability of the binders to bind effectively. The docking technique consists of the following steps: (1) preparation of the ligand structure, (2) preparation of the protein, (3) identification of the binding site, and (4) docking.

3. RESULTS AND DISCUSSION

3.1. QSPR$_{MLR}$ models

The library of experimental complexes consisting of metal ions and ligands, which included quantum parameters and 0-3D molecular descriptors, was utilized to develop the QSPR$_{MLR}$ model. Figure 1 illustrates a typical complex structure, while Table 1 presents the stability constant values logβ_{11}. The dataset was divided into a training set and a test set, with 20% of the data allocated for testing to construct the linear regression model. The model's quality was evaluated using statistical parameters such as R^2_{train}, R^2_{adj}, Q^2_{LOO}, SE, and F_{stat} (Fisher's value). Table 1 shows the statistical values along with the QSPR$_{MLR}$ models.

The important descriptors of the linear model were selected based on changes in statistical parameters. The QSPR$_{MLR}$ models was established using both back-elimination and forward techniques, with a critical value set at 0.05. As a result, the number of descriptors (k) varied from 1 to 6. As the structural parameters changed, the values of values R^2_{train}, Q^2_{LOO}, and SE also fluctuated (see Figure 2a).

The results presented in Table 2 and Figure 2a indicate that the regression parameters also show a positive increase as the number of variables (k) increases from 1 to 6. Specifically, when the number of variables in the model is k of 2, the R^2_{train} value meets our expectations; however, the Q^2_{LOO} does not satisfy the requirements of a valid regression model. Continuing to add variables to the model, guided by the adjusted R^2_{adj} value, the R^2_{train} value increases positively. However, when the value of k rises from 5 to 6, we observe only a marginal increase in the R^2_{train} value. The unexpected decrease in the Q^2_{LOO} value is more concerning, which negatively impacts the model's predictive ability. Consequently, we have chosen the model with five variables as the most suitable within the context of this dataset.

Table 2: The results of built QSPR$_{MLR}$ models (k of 1 to 6)

k	Descriptors	SE	R^2_{train}	R^2_{adj}	Q^2_{LOO}	F_{stat}
1	x_1	2.575	0.423	0.414	0.338	46.893
2	x_1/x_2	2.050	0.640	0.629	0.570	56.021
3	$x_1/x_2/x_3$	1.652	0.770	0.759	0.700	69.207
4	$x_1/x_2/x_3/x_4$	1.567	0.796	0.783	0.687	59.664
5	$\mathbf{x_1/x_2/x_3/x_4/x_5}$	**1.371**	**0.847**	**0.834**	**0.764**	**66.198**
6	$x_1/x_2/x_3/x_4/x_5/x_6$	1.320	0.860	0.846	0.756	60.564

Notation of molecular descriptors

k0	x_1	^{5}C	x_4
Core-core repulsion	x_2	xch5	x_5
xvp10	x_3	LUMO	x_6

As a result, the $QSPR_{MLR}$ model wit k of 5 was selected and the results of the $QSPR_{MLR}$ model is shown in R^2_{train} = 0.847; SE = 1.371; F_{stat} = 66.198 and Q^2_{LOO} = 0.764. The equation of the MLR model is as follows

$$\log\beta_{11} = 31.455 - 4.201 \cdot x_1 + 0.004 \cdot x_2 + 20.867 \cdot x_3 - 10.432 \cdot x_4 - 38.042 \cdot x_5 \quad (6)$$

As such, the training dataset is well-structured, and the constructed $QSPR_{MLR}$ model is statistically significant. The cross-validation method demonstrates that the model effectively predicts the stability constant values. The criteria of statistics are utilized for assessing the significance of the coefficients in the $QSPR_{MLR}$ models, as shown in Table 3.

Furthermore, the average percentage, MP_mx_k, is the contribution of each independent variable in the selected QSPR models (with k of 4 to 6) and is determined according to formula (7) [28].

$$MPx_k, \% = \frac{1}{n}\sum_{j=1}^{n}\left(100.\left|b_{m,i}x_{m,i}\right| \Big/ \sum_{i=1}^{k}\left|b_{m,k}x_{m,k}\right|\right) = \frac{1}{n}\sum_{j=1}^{n}\left(100.\left|b_{m,i}x_{m,i}\right| \Big/ C_{total}\right) \quad (7)$$

here n = 66 is the number of compounds; m is the number of compounds used to determine the P_mx_k value.

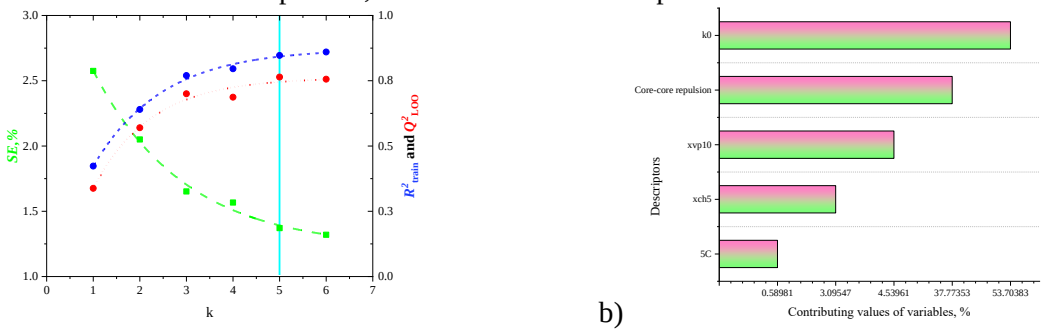


Figure 2: (a) Change of regression parameters according to variable k; (b) Contribution of variables in $QSPR_{MLR}$ model (k of 5)

The regression coefficients along with the contributions of MP_mx_k and GMP_mx_k in the $QSPR_{MLR}$ models for k values ranging from 4 to 6, are presented in Table 3. The significant contributions of the variables are organized in order of their GMP_mx_k values: *k0>core-core repulsion>xvp10>xch5>^{5}C.* According to the results in Table 3, the important contribution of each descriptor GMP_mx_k, % in the $QSPR_{MLR}$ model (6) or rather the contribution of the stability constant of the complexes. The k0 parameter strongly influences the stability constant of complexes. The k0 parameter (x_1) is called Kappa-zero index [33] based on atom classes *(58,98%).* Then, core-core repulsion (x_2) refers to the forces between the atomic cores, which include nuclei and core electrons. In semi-empirical methods, this concept is used to model the repulsive forces that arise from the interaction of atomic cores or electron clouds at short distances (37,77%). The xvp10 (x_3) and xch5 (x_5) are the Chi valence, and chain terms [33], for both simple and valence indices, are defined up to the maximum order present in the molecule. The xvp10 descriptor is called Chi valence path 10 [32], it is the valence 10th-order path chi index (4,54%). The xch5 descriptor is called Chi chain 5, the simple 5th-order chain chi index (3,09%). The weakest descriptor affecting MLR equation is ^{5}C (0,59%). ^{5}C (x_4) is a partial charge of C-atom in 5 sites of complex structure.

Table 3: Contribution of variables in $QSPR_{MLR}$ models with k of 4 to 6

Regression coefficients	$QSPR_{MLR}$			MP_mx_k, %			GMP_mx_k, %
	k = 4	k = 5	k = 6	k = 4	k = 5	k = 6	
b_0	10.347	31.455	34.003	-	-	-	-
b_1	-3.333	-4.201	-4.258	50.760	54.848	55.503	53.704
b_2	0.005	0.004	0.004	45.684	34.855	32.781	37.774
b_3	10.566	20.867	21.489	3.086	5.195	5.338	4.540
b_4	-6.495	-10.432	-10.307	0.470	0.654	0.646	0.590
b_5	-	-38.042	-41.475	-	4.448	4.838	3.095
b_6	-	-	-0.956	-	-	0.893	0.298

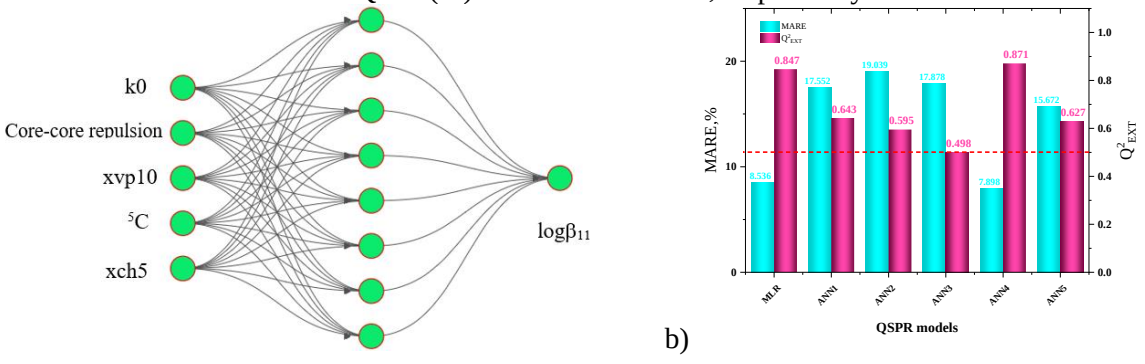
3.2. $QSPR_{ANN}$ models

In addition to MLR model, the ANN model is also developed with the neural network technique on the Matlab tool [24] using the variables of MLR model as input layer of ANN. The MLP of ANN model is the I(5)-HL(*m*)-O(1) architecture. The error back-propagation algorithm is used to train the network. The exponential, logistic, and tangent transfer function sets on each node of the layer neural network. The construction of ANN models involve two stages. In the first stage, a training dataset (Table 1) consisting of 66 experimental constant values is utilized to determine the m value of the hidden layer. The *m* values of screening are presented in Table 4. In the next stage, the work utilize the survey results from the first stage to select the appropriate ANN models for the external evaluation dataset, which includes 16 experimental stability constants values (Table 5). The network development process integrates the external evaluation of the MLR and ANN models by examining the Q^2_{EXT} and MARE(%) values. In this case, the optimal network is chosen if the Q^2_{EXT} value exceeds 0.5 and is closer to 1.0 [34], indicating better performance. Additionally, a lower MARE value is preferred, indicating improved accuracy.

Table 4: The finding results of m neuron for ANN models in the first stage

Code	$QSAR_{ANN}$	R^2_{train}	Q^2_{test}	Q^2_{CV}	Training error	Test error	Validation error	Transfer function
ANN1	I(5)-HL(3)-O(1)	0.897	0.880	0.896	1.285	3.387	1.234	Exponential
ANN2	I(5)-HL(3)-O(1)	0.916	0.904	0.944	0.862	5.772	0.769	Logistic
ANN3	I(5)-HL(11)-O(1)	0.928	0.777	0.904	0.702	6.807	0.830	Exponential
ANN4	**I(5)-HL(8)-O(1)**	**0.926**	**0.872**	**0.915**	**0.724**	**5.412**	**0.665**	**Exponential**
ANN5	I(5)-HL(3)-O(1)	0.915	0.894	0.910	0.922	5.064	0.524	Tanh

Figure 3b displays the models' external validation results. Several ANN models meet the specified criteria; however, the ANN4 model is selected because it achieved the highest Q^2_{EXT} value and the lowest MARE value. Additionally, the results of the ANN4 model are even better than the MLR model when comparing the results of the two models with Q^2_{EXT} (%) of 0.8771and 0.874, respectively.



a) b)

Figure 3: (a) $QSPR_{ANN4}$ I(5)-HL(8)-O(1) model; (b) Q^2_{EXT} and MARE (%) values between QSPR models

3.3. Predictability of QSPR models

An independent training dataset, the external evaluation dataset, is used to assess the prediction abilities, construct full MLR models, and find appropriate ANN models. Besides, the ability prediction of the MLR and ANN models was scrupulously validated using the phasing-each-case method. Table 5 provides specifics about this dataset's data and the predictions made by the built models.

Table 5: The external validation on the data set of 16 experimental stability constants values

n	Ligand				Metal	$\log\beta_{11,exp}$	$\log\beta_{11,cal}$	
	R_1	R_2	R_3	R_4	Ions	(Ref.)	$QSPR_{MLR}$	$QSPR_{ANN4}$
1	H	$-C_6H_3(OH)(OCH_3)$	H	H	Mn^{2+}	10.550 [35]	10.097	10.714
2	$-CH_3$	$-CH=N-NHC_6H_5$	H	H	Cu^{2+}	11.840 [36]	10.034	10.924
3	$-CH_3$	$-CH=N-NHC_6H_5$	H	H	Cu^{2+}	11.530 [36]	10.034	10.924
4	$-CH_3$	$-CH=N-NHC_6H_5$	H	$-CH_3$	Cu^{2+}	12.300 [36]	12.596	13.669

5	$-CH_3$	$-CH=N-NHC_6H_5$	H	$-CH_3$	Co^{2+}	10.590 [36]	11.202	11.242
6	$-CH_3$	$-CH=N-NHC_6H_5$	H	$-CH_3$	Cu^{2+}	12.140 [36]	12.596	11.669
7	$-CH_3$	$-CH=N-NHC_6H_5$	H	$-CH_3$	Cu^{2+}	19.100 [20]	16.751	17.289
8	H	$-C_{10}H_6OH$	H	H	Pb^{2+}	6.570 [37]	6.292	6.502
9	-	$-C_9H_7NO$	H	H	Cu^{2+}	8.714 [38]	9.510	9.537
10	-	$-C_9H_7NO$	H	H	Ni^{2+}	8.500 [38]	9.695	7.218
11	-	$-C_9H_7NO$	H	H	Co^{2+}	7.251 [38]	6.091	7.537
12	H	$-C_6H_4NH_2$	H	H	Mn^{2+}	9.990 [22]	10.358	9.808
13	H	$-C_6H_4NH_2$	H	H	Zn^{2+}	8.740 [22]	9.245	9.065
14	H	$-C_6H_4NO_2$	H	H	Pr^{3+}	11.040 [39]	10.701	9.526
15	H	$-C_6H_4NO_2$	H	H	Cd^{2+}	10.950 [40]	8.974	8.768
16	H	$-C_6H_4NO_2$	H	H	Cr^{2+}	10.150 [40]	10.741	8.883
						MARE, %	8.536	7.898

According to Table 5, the MARE values for models $QSPR_{MLR}$ and $QSPR_{ANN}$ I(5)-HL(8)-O(1) are 7.898% and 8.536%, respectively. These results indicate that the $QSPR_{ANN}$ model demonstrates greater predictive accuracy than the $QSPR_{MLR}$ model and the $\log\beta_{11}$ values generated by the $QSPR_{ANN}$ model are closer to the experimental values.

A one-way ANOVA assessed the differences between the experimental and predicted $\log\beta_{11}$ values obtained from the $QSPR_{MLR}$ and $QSPR_{ANN}$ models. The analysis showed that the discrepancies between the experimental and calculated values of stability constants $\log\beta_{11}$ for the $QSPR_{MLR}$ model and the $QSPR_{ANN}$ model I(5)-HL(8)-O(1) were insignificant ($F = 0.1087 < F_{0.05} = 3.2043$). Therefore, the predictability of both QSPR models aligns well with the experimental data.

3.4. Design and screening of new TIO and metal-TIO

The study designed hundreds of new thiosemicarbazone derivatives, which were screened for predicting new $\log\beta_{11}$ values through Cook's distance value (D-Cook) [31]. QSPR models will predict new derivatives within the AD of the training dataset with |D-Cook| values less than 1, and derivatives located in the outliers of the dataset with |D-Cook| values greater than 1 will be discarded [31]. As a result, 12 new ligands (as Figure 4) and 15 metal-TIO complexes met the evaluation criteria of AD and outliers, and the constructed $QSPR_{MLR}$ and $QSPR_{ANN}$ models predicted their $\log\beta_{11,new}$ values.

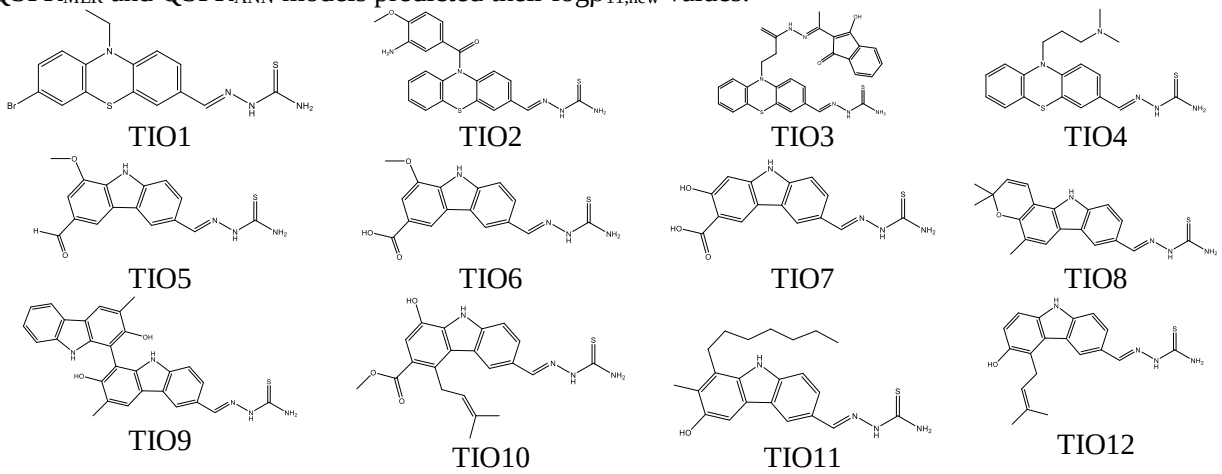


Figure 4: Structures of 12 new ligands

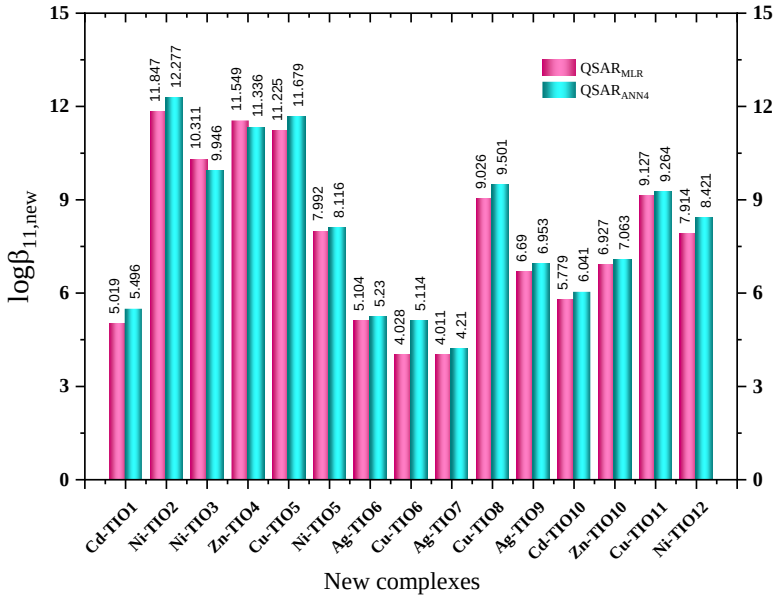


Figure 5: The log$\beta_{11,new}$ predicted values of 15 new complexes from built models

Thus, the predicted results of 15 new complexes with their corresponding log$\beta_{11,new}$ are presented in Figure 5. The predicted results show that the prediction values of the two models are very close to each other, meaning that this result will also be very close to the experimental results and can be further applied in some fields because the models, when built, have confirmed this.

Continue using the ANOVA approach to assess the forecasting outcomes of the two models, $QSAR_{MLR}$ and $QSAR_{ANN4}$. The variance analysis results demonstrate no significant difference ($F = 0.0509 < F_{0.05} = 4.1709$). As a result, both models can make accurate and consistent predictions.

3.5. ADMET analysis and Molecular docking

This research aims to identify potential complexes for use in the pharmaceutical field, necessitating the evaluation of new derivatives for drug-likeness properties. As noted, these new derivatives were assessed simultaneously using the Lipinski and Veber rules via the SwissADME online platform [41]. The results indicated that nine of fifteen compounds met the criteria of both rules, and the findings are fully detailed in Table 6.

The results of the drug-likeness screening from Table 6 show that six derivatives do not satisfy both of the selected rules, namely Cd-TIO1, Ni-TIO2, Ni-TIO3, Ag-TIO7, Ag-TIO9 and Cd-TIO10. These derivatives have violated one of the two mentioned rules, so the work will eliminate the new complexes in the next step in the pathway to find potential derivatives as colorectal cancer inhibitors. Thus, nine derivatives that pass these two rules will perform binding to cancer cell proteins. In the next stage, the study uses the docking technique to evaluate the binding ability of the protein with the code 6GUE-PDB. The docking results with E_bonding (kcal.mol^{-1}) and RMSD (Å) values are presented in Figure 6.

Table 6: Results of ADMET analysis using Lipinski and Veber rules and docking values

No.	Code	Lipinski rules				Veber rules		Drug-likeness
		MW*	MLOGP	HBD	HBA	RB	TPSA	
1	Cd-TIO1	518.75	3.96	1	1	2	104.58	No
2	Ni-TIO2	507.23	3.19	2	3	4	156.79	No
3	Ni-TIO3	614.34	3.19	3	5	7	183.34	No
4	Zn-TIO4	449.92	3.22	1	2	5	107.82	Yes
5	Cu-TIO5	388.91	1.24	2	3	3	118.13	Yes
6	Ni-TIO5	384.06	1.24	2	3	3	118.13	Yes
7	Ag-TIO6	449.23	1.52	3	4	3	138.36	Yes
8	Cu-TIO6	404.91	1.52	3	4	3	138.36	Yes

9	Ag-TIO7	435.21	1.27	4	4	2	149.36	No
10	Cu-TIO8	427.00	2.99	2	2	1	101.06	Yes
11	Ag-TIO9	600.44	3.41	5	3	2	148.08	No
12	Cd-TIO10	521.89	2.61	3	4	5	138.36	No
13	Zn-TIO10	474.86	2.61	3	4	5	138.36	Yes
14	Cu-TIO11	459.09	3.51	3	2	7	112.06	Yes
15	Ni-TIO12	410.14	2.76	3	2	3	112.06	Yes

As mentioned in the methodology section, the docking results were evaluated based on RMSD (<2.0 Å) and E_bonding energy values (<-6.5 kcal.mol^{-1}). The results of nine new derivatives on the 5GUE-PDB code showed that six novel derivatives have less than -6.5 kcal binding energy values.mol-1, namely Zn-TIO4 (-7.4643), Ni-TIO5 (-6.5564), Cu-TIO8 (-7.0769), Zn-TIO10 (-7.0753) Cu-TIO11 (-7.5159) and Ni-TIO12 (-6.5052) and all six metal-TIO complexes also had RMSD values less than 1.5 Å. The docking results of the six novel derivatives with amino acids in the protein structure are shown in Figure 7.

Almost all complexes except Ni-TIO5 show interactions of sulfur atoms with amino acids in the structure. The S atom of the Zn-TIO4 complex interacts with the amino acid TYR199 with the H-donor interaction mode and the corresponding interaction energy and distance values are -0.6 kcal.mol^{-1} and 3.79 Å, the S atom of the Cu-TIO8 complex interacts with the amino acid ASP68 and LYS300 with the H-donor and H-acceptor bonds (E = -1.0 and -5.7 kcal.mol^{-1}; D = 3.85 and 3.27 Å), the S atom of the Zn-TIO10 complex interacts with the amino acid ASN3 with the H-acceptor mode (E = -1.9 kcal.mol^{-1}; D = 3.84 Å), the S atom of the Cu-TIO11 complex interacts with the amino acid LYS300 with the H-acceptor bond (E = -6.7 kcal.mol^{-1}; D = 3.31 Å), and the S atom of the Ni-TIO12 complex interacts with the amino acid GLU2 and ASN73 with two modes of the H-donor and H-acceptor (E = -1.3 and -0.5 kcal.mol^{-1}; D = 3.51 and 4.09 Å). In particular, these interactions of complexes such as Cu-TIO8 and Cu-TIO11 are robust and stable on amino acid LYS300 with corresponding energy values of -5.7 and -6.7 kcal.mol^{-1}. The interaction mode of the receptors and ligands is H-acceptor. In addition, the expected interactions in the complex structure were also present, namely the metallic interaction of metal atoms with amino acids. This occurred in three of the six complexes: the Zn-TIO4 complex with amino acid LYS202 (E = -1.4 kcal.mol^{-1}; D = 2.22 Å), Cu-TIO11 with amino acid ACE-1 (E = -0.6 kcal.mol^{-1}; D = 2.79 Å), and Ni-TIO12 with amino acid GLU2 (E = -1.5 kcal.mol^{-1}; D = 2.52 Å).

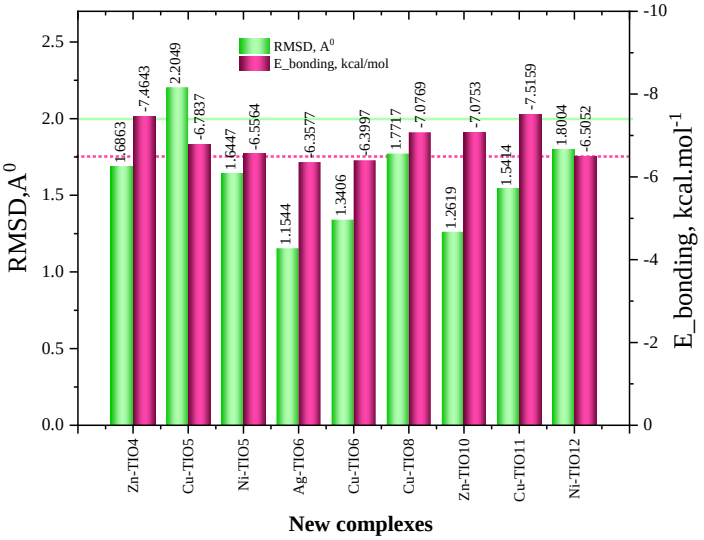


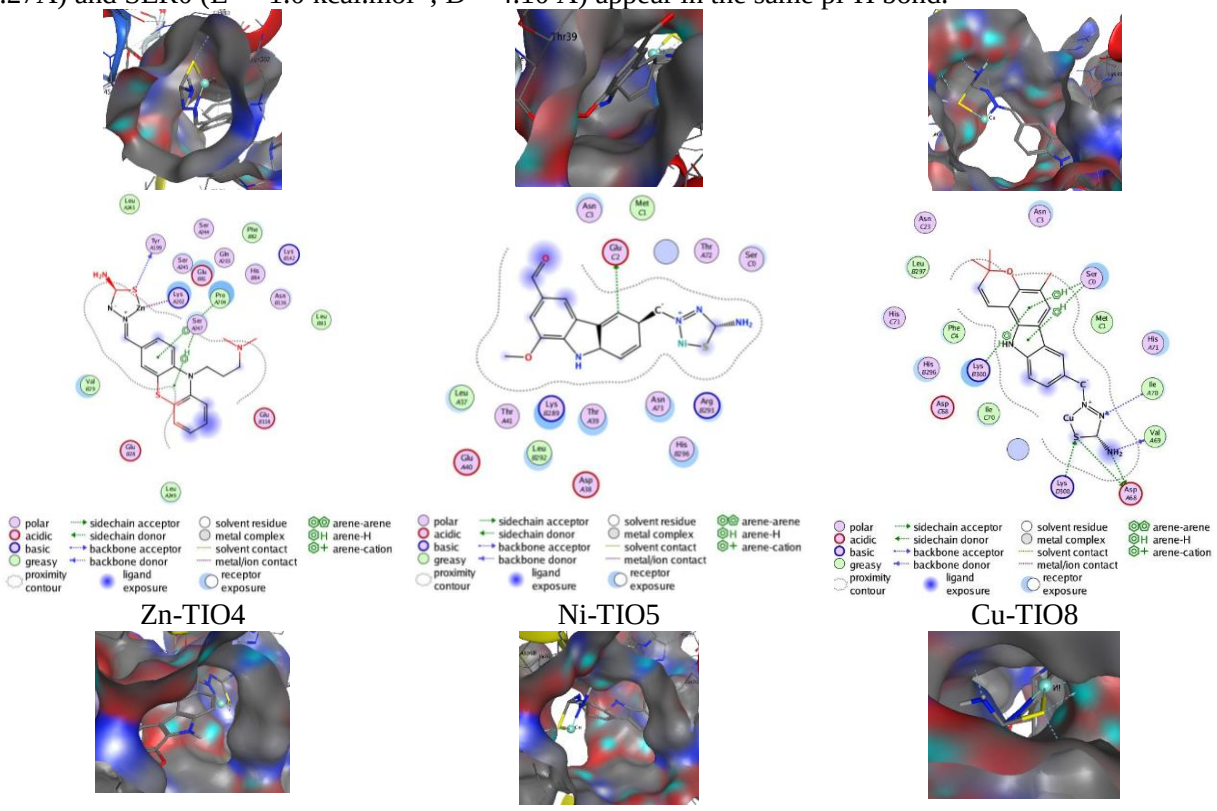
Figure 6: RMSD (Å) and E_bonding (kcal.mol^{-1}) values of nine derivatives docked on 5GUE

Table 7: Interactions of six novel complexes from the docking results

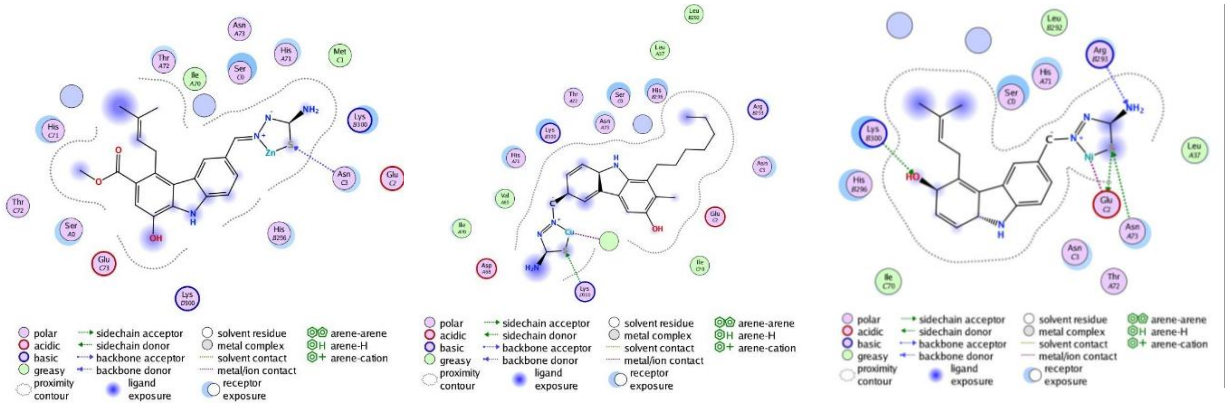
Complexes	Ligand	Receptor	Interaction	Distance D, Å	Energy E, kcal.mol^{-1}
Zn-TIO4	S18	TYR199	H-donor	3.79	-0.6
	Zn28	LYS202	metal	2.22	-1.4

	6-ring	PRO204	pi-H	4.16	-0.5
	6-ring	SER247	pi-H	4.37	-0.9
Ni-TIO5	C11	GLU2	H-donor	3.17	-0.5
Cu-TIO8	S23	ASP68	H-donor	3.85	-1.0
	N27	ASP68; VAL69	H-donor	3.01; 2.92	-1.0; -1.1
	S23	LYS300	H-acceptor	3.27	-5.7
	N26	ILE70	H-acceptor	3.54	-0.7
	6-ring	LYS300; SER0; SER 0;	pi-H	3.59; 4.27; 4.10	-0.5; -0.5; -1.0
Zn-TIO10	S26	ASN3	H-acceptor	3.84	-1.9
Cu-TIO11	S27	LYS300	H-acceptor	3.31	-6.7
	Cu32	ACE-1	metal	2.79	-0.6
Ni-TIO12	S24	GLU2	H-donor;	3.51	-1.3
		ASN73	H-acceptor	4.09	-0.5
	O17	LYS300	H-acceptor	3.19	-0.5
	N28	ARG293	H-acceptor	3.40	-0.5
	Ni25	GLU2	metal	2.52	-1.5

One interaction of note in these complexes is the 6-ring. The 6-ring of these ligands appears in the Zn-TIO4 and Cu-TIO8 complexes. For the Zn-TIO4 complex, the two 6-ring of phenothiazine structures appear in two interactions with the amino acids PRO204 (E = -0.5 kcal.mol^{-1}; D = 4.16 Å) and SER247 (E = -0.6 kcal.mol^{-1}; D = 4.37 Å), and these interactions occur in the same pi-H type. For the Cu-TIO8 complex, the interactions of the 6-ring with the two amino acids LYS300 (E = -0.5 and -0.5 kcal.mol^{-1}; D = 3.59 and 4.27Å) and SER0 (E = -1.0 kcal.mol^{-1}; D = 4.10 Å) appear in the same pi-H bond.



Zn-TIO4 Ni-TIO5 Cu-TIO8



Zn-TIO10 Cu-TIO11 Ni-TIO12

Figure 7: Docking results of six potential complexes derivatives with 5GUE-PDB

One atom of the thiosemicarbazone structure that is expected to have an affinity for the receptors is the N atom. The docking results in Figure 7 and Table 7 show that the Cu-TIO8 and Ni-TIO12 complexes show interactions of this atom. The N atom at the 27 site of the Cu-TIO8 complex shows interactions with the amino acids ASP68 (E = -1.0 kcal.mol-1; D = 3.01 Å) and VAL69 (E = -1.1 kcal.mol^{-1}; D = 2.92 Å) in the same interaction of the H-donor bond. In addition, the N atom at the 26 site interacts with the amino acid ILE70 (E = -0.7 kcal.mol^{-1}; D = 3.54 Å) with the H-acceptor interaction mode. The second complex shows this interaction, which is the Ni-TIO12 complex. The interaction occurs between the N atom at the 28 site and the amino acid ARG 293 (E = -0.5 kcal.mol^{-1}; D = 3.40 Å) with the H-acceptor bond.
Among the potential complexes, the Ni-TIO5 complex appears to have the poorest interaction; the second image of Figure 7 records that this complex's interaction with amino acids is very little and not outstanding. Finally, the study successfully built $QSAR_{MLR}$ and $QSAR_{ANN}$ models to validate newly designed metal-TIO derivatives' anti-cancer activity. the work selected six potential derivatives, Zn-TIO4, Ni-TIO5, Cu-TIO8, Zn-TIO10, Cu-TIO11, and Ni-TIO12, that can inhibit colorectal cancer cells.

4. CONCLUSION

Using experimental $\log\beta_{11}$ values, this study successfully created the $QSPR_{MLR}$ and $QSPR_{ANN}$ models to predict the complexation of metal-thiosemicarbazone complexes. The experimental data set and structural descriptions from PM7 and PM7/sparkle semi-empirical quantum computations were used to construct QSPR models. The multivariate linear regression approach was used to construct the QSPRMLR models, and the 5-descriptors $QSPR_{MLR}$ model successfully produced the best $QSPR_{ANN}$ model I(5)-HL(8)-O(1). The QSPR models were built to satisfy the statistical requirements. R^2_{train}, Q^2_{LOO}, Q^2_{Ext}, SE, MARE (%), and the single-factor ANOVA approach were among the statistical indicators used to assess the QSPR model-building process thoroughly. The $\log\beta_{11}$ values have been projected for the fifteen new complexes developed between specific metal ions and thiosemicarbazones containing phenothiazine and carbazole derivatives. The newly developed complexes were further evaluated using appropriate techniques for their potential use in treating anti-colorectal cancer. An ADMET analysis revealed that nine of fifteen complexes met the Lipinski and Veber rules. The docking technique on the protein with the 6GUE-PDB code for cancer yielded promising results, including six potential complexes. By anticipating and experimentally guiding the synthesis of thiosemicarbazone derivatives with strong complexing capacity with metal ions, the work's results open up a new and exciting avenue for investigation in the fields of analysis and the environment, especially pharmacology.

REFERENCES

[1] R. B. Singh, B. S. Garg, and R. P. Singh, Analytical applications of thiosemicarbazones and semicarbazones: A review, *Talanta*, vol. 25, no. 11-12, pp. 619-632, 1987.
[2] G. Pelosi, Thiosemicarbazone metal complexes: from structure to activity, *Open Crystal. J.*, vol. 3, pp. 16-28, 2010.

[3] R. Chaudhary and Shelly, Synthesis, spectral and pharmacological study of Cu(II), Ni(II) and Co(II) Coordination Complexes, *Res. J. chem. Sci*., vol. 1, no. 5, pp. 1-5, 2011.
[4] N. M. Quang, T. X. Mau, N. T. A. Nhung, T. N. M. An, and P. V. Tat, Novel QSPR modeling of stability constants of metal-thiosemicarbazone complexes by hybrid multivariate technique: GA-MLR, GA-SVR and GA-ANN, *J. Mol. Struct*., vol. 1195, pp. 95-109, 2019.
[5] M. Ante and N. Raos, Estimation of stability constants of mixed copper(II) chelates using valence connectivity index of the 3rd order derived from two molecular graph representations, *Acta Chim. Slov*., vol. 56, pp. 373-378, 2009.
[6] R. Kunal, K. Supratik, and D.R. Narayan, *A Primer on QSAR/QSPR Modeling: Fundamental Concepts*. Springer: London, 2015.
[7] A. Speck-Planche, V. V. Kleandrova, F. Luan, M. Natália, and D. S. Cordeiro, Rational drug design for anti-cancer chemotherapy: multi-target QSAR models for the *in silico* discovery of anti-colorectal cancer agents, *Bioorg. Med. Chem*., vol. 20, no. 15, pp. 4848-4855, 2012.
[8] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA Cancer J. Clin*., vol. 71, pp. 209-249, 2021.
[9] J. J. P. Stewart, Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters, *J. Mol. Model*., vol. 19, pp. 1-32, 2013.
[10] C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, Experimental and omputational approaches to estimate solubility and permeability in drug discovery and development settings, *Adv. Drug Deliv. Rev*., vol. 23, pp. 3-25, 1997.
[11] D. F. Veber, S. R. Johnson, H. Y. Cheng, B. R. Smith, K. W. Ward, and K. D. Kopple, A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases, *J. Med. Chem*., vol. 45, pp. 2615-2623, 2002.
[12] D. N. Kenie and A. Satyanarayana, Protolitic equilibria and stability constants of Mn(II) and Ni(II) complexes of 3-formylpyridine thiosemicarbazone in sodium dodecyl sulphate (SDS)-water mixture, *Sci. Technol. Arts Res. J*., vol. 4, no. 1, pp. 74-79, 2015.
[13] D. G. Krishna and K. Devi, Determination of cadmium(II) in presence of micellar medium using cinnamaldehyde thiosemicarbazone by spectrophotometry, *Int. J. Green Chem. Biopro*., vol. 5, no. 2, pp. 28-30, 2015.
[14] N. S. R. Reddy and D. V. Reddy, Spectrophotometric determination of Vanaditun(V) with salicylaldehyde thiosemicarbazone, *J. Indian Inst. Sci.*, vol. 64B, pp. 133-136, 1983.
[15] D. N. Kenie and A. Satyanarayana, Solution equilibrium study of the complexation of Co(II) and Zn(II) with nicotinaldehyde thiosemicarbazone, *Sci. Technol. Arts Res. J.*, vol. 4, no. 3, pp. 145-149, 2015.
[16] V. Veeranna, V. S. Rao, V. V. Laxmi, and T. R. Varalakshmi, Simultaneous second order derivative spectrophotometric determination of cadmium and cobalt using Furfuraldehyde thiosemicarbazone (FFTSC), *Res. J. Pharm. Technol*., vol. 6, no. 5, pp. 577-584, 2013.
[17] D. G. Krishna and G. V. K. Mohan, A Facile Synthesis, Characterization of cinnamaldehyde thiosemicarbazone and determination of Molybdenum(VI) by spectrophotometry in presence of micellar medium, *Indian J. Appl. Res*., vol. 3, no. 8, pp. 7-8, 2013.
[18] B. S. Garg, S. R. Singh, S. B. Basnet, and R. P. Singh, Potentiometric studies on the complexation equilibria between La(III), Pr(III), Nd(III), Gd(III), Sm(III), Tb(III), Dy(III), Ho(III) and 2-acetylpyridinethiosemicarbazone (2-APT), *Polyhedron*, vol. 7, no. 2, pp. 147-150, 1988.
[19] M. A. Jiménez and M. D. Luque De Castro, Potentiometric study of silver(I)-thiosemicarbazonates, *Microchem. J*., vol. 25, pp. 301-308, 1980.
[20] T. Atalay and E. Ozkan, Thermodynamic studies of some complexes of 4'-morpholinoacetophenone thiosemicarbazone, *Thermochim Acta*, vol. 237, pp. 369-374, 1994.
[21] B. S. Garg, S. Ghosh, V. K. Jain, and P. K. Singh, Evaluation of thermodynamic parameters of bivalent metal complexes of 2-hydroxyacetophenonethiosemicarbazone (2-HATS), *Thermochima Acta*., vol. 157, pp. 365-368, 1990.
[22] S. S. Sawhney and S. K. Chandel, Solution chemistry of Cu(II)-, Co(II)-, Ni(II)-, Mn(II)- and Zn(II)-p-aminobenzaldehyde thiosemicarbazone systems, *Thermochim Acta*, vol. 71, pp. 209-214, 1983.
[23] D. D. Steppen, J. Werner, and P. R. Yeater, E*ssential Regression and Experimental Design for Chemists and Engineers*, Free Software Package, 1998.
[24] *Matlab R2016a 9.0.0.341360*, MathWorks, MA, 2016.
[25] R. Rojas, *Neural Networks*. Springer-Verlag: Berlin, 1996.
[26] M. Hagan, H. Demuth, and M. Beale, *Neural Network Design*, PWS Publishing: Boston, 1996.
[27] P. V. Tat, *Development of QSAR and QSPR,* Ha Noi Publisher of Natural sciences and Technique: Ha Noi, 2009.
[28] S. W. Priyanka, R. C. Anuruddha, G. R. Kunal, T. G. Pooja, A. R. Chabukswar, K. G. Raut, and P. T. Giri, A review on carbazole and its derivatives as anticancer agents from 2013 to 2024, *Chirality*, vol. 37, no. 2, pp. e70021, 2025.

[29] G. Sudeshna and K. Parimal, Multiple non-psychiatric effects of phenothiazines: A review, *Eur. J. Pharmacol*., vol. 648, no. 1-3, pp. 6-14, 2010.
[30] I. J. Al-Busaidi, A. Haque, N. K. Al Rasbi, and M. S. Khan, Phenothiazine-based derivatives for optoelectronic applications: A review, *Synth. Met*., vol. 257, no. 116189, 2019.
[31] OECD, *Guidance document on the validation of (quantitative) structure-activity relationships models*. Organisation for Economic Co-operation and Development: France, 2007.
[32] S. Podlewska and A. J. Bojarski, *Chapter 3 - Post-processing of Docking Results: Tools and Strategies*, Molecular Docking for Computer-Aided Drug Design, Academic Press, USA, 2021.
[33] R. Todeschini and V. Consonni, *Handbook of Molecular Descriptors*. Wiley-VCH: Weinheim, Germany, 2000.
[34] A. Golbraikh and A. Tropsha, Beware of q2!, *J. Mol. Graph. Model*, vol. 20, no. 4, pp. 269-76, 2002.
[35] D. K. Singh, P. K. Jha, R. K. Jha, P. M. Mishra, A. Jha, S. K. Jha, and R. P. Bharti, Equilibrium studies of transition metal complexes with tridentate ligands containing N, O, S as donor atoms, *Asian J. Chem*., vol. 21, no. 7, pp. 5055-5060, 2009.
[36] A. T. A. El-Karim and A. A. El-Sherif, Potentiometric, equilibrium studies and thermodynamics of novel thiosemicarbazones and their bivalent transition metal(II) complexes, *J. Mol. Liq*., vol. 219, pp. 914-922, 2016.
[37] Sahadev, R. K. Sharma, and S. K. Sindhwani, Thermal studies on the chelation behaviour of biologically active 2-hydroxy-1-naphthaldehyde thiosemicarbazone (HNATS) towards bivalent metal ions: a potentiometric study, *Thermochim. Acta*, vol. 202, pp. 291-299, 1992.
[38] K. Sarkar and B. S. Garg, Determination of thermodynamic parameters and stability constants of the complexes of p-MITSC with transition metal ions, *Thermochim Acta*, vol. 113, pp. 7-14, 1987.
[39] S. S. Sawhney and S. K. Chandel, Stability and thermodynamics of La(III)-, Pr(III)-, Nd(III)-, Gd(III)- and Eu(III)-p-nitrobenzaldehyde thiosemicarbazone systems, *Thermochim Acta*, vol. 72, pp. 381-385, 1984.
[40] S. S. Sawhney and R. M. Sati, pH-metric studies on Cd(II)-, Pb(II)-, Al(III)-, Cr(III)- and Fe(III)-p-nitrobenzaldehyde thiosemicarbazone systems, *Thermochim Acta*, vol. 66, pp. 351-355, 1988.
[41] SwissADME: a free web tool to evaluate pharmacokinetics, *Sci. Rep*., vol. 7, no. 42717, 2017.

THIẾT KẾ, SÀNG LỌC VÀ KHÁM PHÁ CÁC PHỨC CHẤT MỚI TÁC ĐỘNG VÀO DÒNG TẾ BÀO HT-29 BẰNG CÁCH SỬ DỤNG MÔ HÌNH *IN SILICO* QSPR

NGUYEN MINH QUANG[1*], PHAM VAN TAT[2]
[1] Khoa Công nghệ Hóa học, Trường Đại học Công nghiệp Thành phố Hồ Chí Minh
[2] Viện Đào tạo và Nghiên cứu Dược phẩm, Trường Đại học Bình Dương
** Tác giả liên hệ: nguyenminhquang@iuh.edu.vn*

Tóm tắt. Sự mô hình hóa theo phương pháp *in silico* đã được áp dụng nhằm phát triển mới các phưc chất giữa các ion kim loại và dẫn xuất thiosemicarbazone thông qua việc tính toán hằng số bền ($\log\beta_{11}$) của các phức chất. Các mô tả phân tử 0-3D, các tham số lý hóa và lượng tử đã được tính toán bằng cách sử dụng phép tính lượng tử bán thực nghiệm theo phương pháp mới PM7 và PM7/sparkle. Kỹ thuật hồi quy tuyến tính đa biến (MLR) và mạng thần kinh nhân tạo (ANN) đã được sử dụng để xây dựng các mô hình quan hệ định lượng cấu trúc-tính chất (QSPR). Kết quả, mô hình $QSPR_{MLR}$ được xây dựng bao gồm các mô tả như: *k0, core-core repulsion, xvp10,* ^5C và *xch5*. Các giá trị thống kê của mô hình đã được xác định: R^2_{train} = 0,847; R^2_{adj} = 0,834; Q^2_{LOO} = 0,764; SE = 1,371 và F-stat = 66,20. Mô hình mạng thần kinh $QSPR_{ANN}$ I(5)-HL(8)-O(1) cũng được phát triển với các giá trị thống kê: R^2_{train} = 0,976; Q^2_{test} = 0,956 và $Q^2_{validation}$ = 0,926. Một loạt các dẫn xuất thiosemicacbazone mới và các phức chất của các phối tử này và các ion kim loại đã được thiết kế bằng cách sử dụng các mô hình đã được xây dựng. Các phức chất này đã được sàng lọc bằng kỹ thuật miền ứng dụng và giá trị ngoại biên (AD and Outliers). Khả năng dự đoán của các mô hình áp dụng lên các phức chất trên tập dữ liệu thử nghiệm phù hợp với các kết quả từ nghiên cứu thực nghiệm. Hơn nữa, kết quả của mười lăm phức chất mới đã được chọn vì các phức chất này nằm trong phạm vi ứng dụng dự đoán. Tiếp theo, chín trong số mười lăm phức chất mới đã vượt qua kết quả phân tích tính giống dược theo hai quy tắc Lipinski và Veber. Cuối cùng, chín trong số mười lăm phức chất đã được mô phỏng docking lên protein chống ung thư trực tràng (mã: 6GUE-PDB). Kết quả đã sàng lọc ra sáu phức chất mới được coi là chất ức chế tiềm năng đối với dòng tế bào HT-29 trong việc hỗ trợ điều trị ung thư trực tràng.

Từ khóa: Docking, mô hình định lượng về cấu trúc-tính chất, Phức chất kim loại-thiosemicarbazone, ung thư đại trực tràng.

Ngày nhận bài: 05/8/2025
Ngày nhận đăng: 21/10/2025