Turkish Computational and Theoretical Chemistry

Turkish Comp Theo Chem (TC&TC)

Volume(Issue): 7(1) – Year: 2023 – Pages: 51-61

e-ISSN: 2602-3237

https://doi.org/10.33435/tcandtc.1093987

Received: 28.03.2022 ***Accepted:*** 29.06.2022 *Research Article*

Cytotoxic activity of new thioglycosides of the arylamino-1,3,4-thiadiazole and glycosyl-1,3,4-thiadiazole-triazole conjugates against Escherichia coli: QSAR study

Hassan BADAOUI[a, 1], Yassine KOUBI[a], Younsse MOUKHLISS[a], Khalile EL KHATABI[a], Mohmed ELBOUHI[a], Mohamed OUABANE[a], M'barek CHOUKRAD[a, 2], Mohammed Aziz JANA[a], Abdelouahid SBAI[a], Tahar LAKHLIFI[a], Mohammed BOUACHRINE[a,b]

[a]Molecular Chemistry and Natural Substances Laboratory (MCNSL),
Department of Chemistry, Faculty of Science. Moulay Ismail University, BP 298 Meknes, Morocco.
[b]Higher School of Technology (HST), Sultan Moulay Sliman University, BP 23000 Beni-Mellal, Morocco.

Abstract: The main objective of our study is the modeling of the cytotoxic activity of a series of 17 compounds derived from new thioglycosides of the arylamino-1, 3,4-thiadiazole and glycosyl-1, 3,4-thiadiazole-triazole conjugates against Escherichia coli using 2D-QSAR and some statistical tools (Principal Component Analysis (PCA), Multiple Linear Regression (MLR), This study consists of three main steps: Data set selection, molecular descriptor generation, construction and validation of predictive models for the studied activity. To build and validate our QSAR model, the dataset was divided into two sets: 17 molecules constitute the training set and the remaining 4 molecules constitute the test set. The division of the data set was done by random selection. The training set and the test set were validated separately using internal and external tests, such as the y-randomization and Golbraikh and Trouphsa model validation criteria. The model is validated with the high values of R2, R2test and Q2cv (R2= 0.731, R2adj =0.641 and Q2cv=0.51, R2test =0.666, MSE=0.003).

Keywords: 2D-QSAR, E. coli, RLM, y-randomization.

1. Introduction

The thiadiazole ring system, presented as a structural bioisostere of pyrimidine and oxadiazole, has shown numerous biological activities, including anticancer, antifungal, antiviral, antidiabetic, analgesic, antiepileptic, antibacterial and anti-inflammatory activities [2-3]. The increased lipid solubility conferred by the sulfur atom of the thiadiazole ring in addition to the mesoionic nature allowed compounds incorporating this motif to interact with biological targets with distinct affinities after penetrating through cell membranes. Recent studies have shown the ability of 1,3,4-thiadiazole compounds to inhibit a variety of molecular targets, including kinases that enable their activity against different cancer cell lines, presenting them as a promising structure for antitumor drug discovery. [1] On the other hand, nucleoside analogues have been recognized as active drug candidates for the treatment of various forms of cancer. A number of nucleoside analogues such as cytarabine, azacitidine, capecitabine, floxuridine and decitabine are indeed ratified with antitumor activity [4].

The main objective of our study is to model the cytotoxic activity of a series of 17 compounds derived from new thioglycosides of the arylamino-1, 3,4-thiadiazole and glycosyl-1, 3,4-thiadiazole-triazole conjugates against Escherichia coli using some statistical tools (Principal Component

[1] Corresponding Authors

e-mail: hassanbadaoui5050@gmail.com

[2] Corresponding Authors

e-mail: M'barekchoukrad@gmail.com

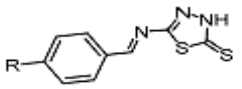
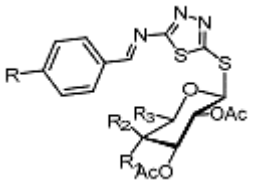
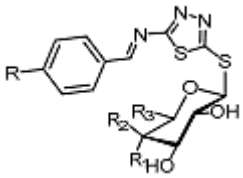
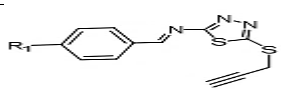
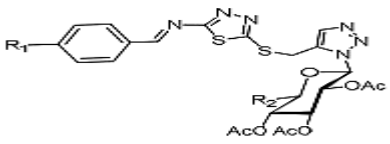
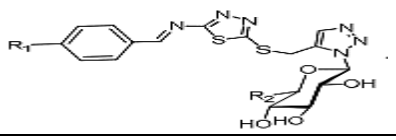
Analysis (PCA), Multiple Linear Regression (MLR). This study includes three main steps: selection of the data set, generation of molecular descriptors, construction and validation of predictive models of the studied activity

2. Computational Method

2.1. Data sources

In the present study, we selected 17 novel thioglycoside derivatives of arylamino-1, 3,4-thiadiazole and glycosyl-1,3,4-thiadiazole-triazole conjugates (Table 1) with activity values reported in the literature. Activity expressed as IC50 is defined as the concentration required to decrease the initial rate of effect of CVB3 with a percentage of 50%. For modeling purposes, the activity was expressed as pIC50 such that: $pIC50 = -\log_{10}(IC50)$. To build and validate our QSAR model, the data set was divided into two sets: 17 molecules constitute the training set and the remaining 4 molecules constitute the test set. The division of the data set was done by random selection.

Table 1. Structure and biological activity of the studied compounds.

N°	Molecules	R	R_1	R_2	R_3	IC_{50}
2		OCH_3				12,2
3		$N(CH_3)_2$				15,6
4		OCH_3	H	OA_C	$CH2OA_C$	20,7
5		OCH_3	OA_C	H	H	22,5
6		$N(CH_3)_2$	H	OA_C	CH_2OA_C	19,2
7		$N(CH_3)_2$	OA_C	H	H	19,7
8		OCH_3	H	OH	CH_2OH	------
9		OCH_3	OH	H	H	------
10		$N(CH_3)_2$	H	OH	CH_2OH	14,4
11		$N(CH_3)_2$	OH	H	H	22,2
12		OCH_3				17,6
13		$N(CH_3)_2$				23,5
14			OCH_3	CH_2OA_C		15,3
15			OCH3	H		15,9
16			$N(CH_3)_2$	CH_2OA_C		21
17			$N(CH_3)_2$	H		18,2
18			OCH_3	CH_2OH		18,3
19			OCH_3	H		
20			$N(CH_3)_2$	CH_2OH		12,1
21			$N(CH_3)_2$	H		19,2

2.2. Molecular descriptors

Several descriptors are calculated using Chemoffice 2016, ACD / ChemSketch [5], and MarvinSketch [6] to predict the correlation between the descriptors and the activity of the molecules studied (Tables 2).

2.3. Statistical analysis

2.3.1. Principal component analysis

The statistical method used to apply the quantitative relationship between the chemical descriptor and the biological activity using a mathematical model, are the principal component analysis ACP [7], using XLSTAT 16 [8]. The objective of the use

ACP is for analyzing the relationship between the quantitative descriptor of novel thioglycoside derivatives of arylamino-1, 3,4-thiadiazole and glycosyl-1,3,4-thiadiazole-triazole conjugates, and also used to understand the distribution of the compounds [9], the result of ACP present in Table 5.

2.3.2. Multiple Linear Regression (MLR)

Multiple linear regression (MLR) is one of the most transparent modeling methods due to its ease of use and ease of interpretation. It is based on the fact that there is a linear relationship between a dependent variable Y (the activity pIC50) and a series of independent variables Xi (the descriptors) [10] according to the following relation:

$$y = a_0 + \sum_{i=1}^{n} a_i x_i$$

Where: y is the dependent variable;
x_i is the independent variables;
a_0 is the constant of the model equation;
a_i are the coefficients of descriptors in the model equation

Internal validation is the set of tests applied to the group of learning molecules. The solidity if this set is characterized by the values of the parameters R2, R2adj , MSE, and F, which are successive : the coefficient of determination and the coefficient of the determination adjusts, the average of the square errors of the model, and Fisher test ¬[11]. There are other internal validation procedures to reinforce the reliability of the model obtained, on the one hand, cross-validation LOO (leave-one-out), which consists in extracting a number k (k=1) molecule from the set initial of N molecule and build a model with (N-1) molecule, remaining using the chosen descriptor, this process and then repeated to remove and predict the values of all the molecules in the learning series [12]. On the other hand, y-randomization consists of randomly mixing the properties by contribution to the experimental activity for the learning series which we use the same descriptor. The parameters of the randomization must be to have poor quality.

2.4. Internal Validation

Table 2. Software used to calculate the various descriptors

ACD / ChemSketch	Molar Volume (MV); Molecular Weight (MW); Parachor (Pc); Refractive Index (n); Surface Tension (γ); density(d) and Polarizability (α)
MarvinSketch	Partition Coefficient (LogP) ; Number of H-bond Acceptors (NHA); Number of H-bond donors (NHD); Van der Waals volume (VDWV); Refractivity (R)
Chemoffice2016	Balaban Index (J); Weiner Index (W);

2.5. External validation

Internal validation is a standard compulsory part of QSAR modeling according to some researchers [13].This model consists of predicting the activity of a series of molecules commonly called test series which do not appear in the model development series. This validation is characterized by the parameters R_{test}, Q^2_{cv} (test). Recently, several studies, have demonstrated insufficient parameter R_{test} , Q^2_{cv} (test) to control the predictive capacity of QSAR model. Therefore, other parameters must be checked for this. These parameters are called "external validation criteria" or are often called "Tropsha criteria''.

The domain applicability is a mandatory condition for measuring the use of the QSAR model according to the OECD validation principal [14]. We cannot say a model valid if and only the model and capable of creating to prediction of new compounds can say exists as credible and not an extrapolation of the model.

The domain applicability is presented by the Williams graph (residual= f (leverage values)), the leverage value for each compound is calculated by the following method:

$$h_i = X_i^t (X^t X)^{-1}$$

And Xi is the descriptor vector of compound examined X, and is the descriptive matrix derived from the values of the descriptors of the learning set. According to the Williams graph, the domain of applicability exists inside a square bounded by the value ±x which is the value of the standard deviation. The levier value (h_i) of each compound must be lower than the threshold value h* for a

model to be meaningful and a strong significant and a strong predictability.

3. Results and discussion

3.1. Molecular descriptors

Thanks to the software (MarvinSketch, ACD/ChemSketch), Chemoffice 2016 and the 15 descriptors; we have calculated the different descriptors mentioned above, the results are gathered on the following (table 3)

The polarizability and molar refractivity increase relatively with the size and molecular weight of the studied thiadiazolyl-1, 2, 3-triazole and thiadiazole derivatives (Table 3). Then, these results are in agreement with the Lorentz-Lorenz formula.

This relationship shows that molar refractivity and polarizability increase with volume and molecular weight. For example, compound 16 carry bulky substituents that include (N-dimethylamine) and (methoxyacetate) have large values of polarizability (67.17Å3) and molar refractivity (169.44Å3), respectively. In contrast, compound 2 is the small molecule in the studied series of thiadiazole at the small value of polarizability (27.36Å3) and molar refractivity (69.03Å3).

Compound 18 has the lowest partition coefficient (Log P) (0.61), This compound is the most hydrophilic.

Compounds 12 and 13 which have the highest values of partition coefficient (3.85) and (3.59) respectively, they show significant abilities to bind to proteins. [15]

The hydrophobic groups in the structures of thiadiazole derivatives induce a decrease in the hydration energy.

3.2. Principal component analysis (PCA)

The method of Principal Component Analysis (PCA) is quite necessary to reduce the number of descriptors to even precede the RLM in order to be able to eliminate any type of redundancy, [16] we carried out a PCA study on our studied system, trying to find correlations between the selected descriptors, the main results drawn by the PCA is contained in the following table:

Table 3. Calculated values of the different descriptors

N°	pIC_{50}	MW	MR	MV	Pc	N	ɣ	D	Ae	W	IB	LogP	HLB	HD	HA	VVDW
1	**4,91**	**251,33**	**69,03**	**179,50**	**480**	**1,70**	**51,10**	**1,39**	**27,36**	**511**	**91761**	**2,98**	**12,4**	**1**	**3**	**201,49**
2	**4,81**	**264,37**	**76,02**	**199,00**	**526**	**1,69**	**48,70**	**1,32**	**30,13**	**605**	**122226**	**3,25**	**11,72**	**1**	**3**	**221,58**
3	**4,68**	**581,62**	**141,11**	**399,50**	**1074**	**1,62**	**52,30**	**1,45**	**55,94**	**6999**	**6682412**	**2,42**	**15,3**	**0**	**9**	**483,27**
4	**4,65**	**509,55**	**125,22**	**348,60**	**941**	**1,64**	**53,20**	**1,46**	**49,64**	**4937**	**3517495**	**2,61**	**14,25**	**0**	**8**	**421,06**
5	**4,72**	**594,66**	**148,10**	**419,00**	**1121**	**1,62**	**51,20**	**1,41**	**58,71**	**7530**	**7521740**	**2,69**	**14,57**	**0**	**14**	**504,21**
6	**4,71**	**522,59**	**132,21**	**368,10**	**987**	**1,64**	**51,80**	**1,41**	**52,41**	**5364**	**4026941**	**2,88**	**15,33**	**0**	**8**	**440,98**
7	**4,84**	**413,47**	**99,42**	**248,90**	**719**	**1,73**	**69,70**	**1,66**	**39,41**	**2431**	**1055746**	**0,93**	**14,7**	**4**	**9**	**356,79**
8	**4,65**	**396,48**	**100,95**	**255,10**	**721**	**1,72**	**63,80**	**1,55**	**40,02**	**2019**	**727815**	**1,56**	**13,67**	**3**	**8**	**331,24**
9	**4,75**	**302,42**	**90,15**	**252,30**	**655**	**1,63**	**45,40**	**1,19**	**35,73**	**1002**	**277357**	**3,85**	**8,63**	**0**	**4**	**260,83**
10	**4,63**	**289,38**	**83,16**	**232,80**	**609**	**1,63**	**46,80**	**1,24**	**32,96**	**869**	**217578**	**3,59**	**9,3**	**0**	**4**	**240,73**
11	**4,82**	**664,71**	**162,45**	**439,10**	**1209**	**1,66**	**57,50**	**1,51**	**64,40**	**8119**	**6926367**	**1,16**	**32,52**	**2**	**12**	**550,72**
12	**4,80**	**592,64**	**146,57**	**388,30**	**1076**	**1,68**	**59,00**	**1,52**	**58,10**	**6201**	**4216160**	**1,34**	**31,5**	**2**	**11**	**488,14**
13	**4,68**	**677,75**	**169,44**	**458,60**	**1255**	**1,66**	**56,20**	**1,47**	**67,17**	**8725**	**7762254**	**1,42**	**31,78**	**2**	**12**	**571,09**
14	**4,74**	**605,69**	**153,56**	**407,70**	**1122**	**1,68**	**57,40**	**1,48**	**60,87**	**6708**	**4785033**	**1,61**	**30,76**	**2**	**11**	**509,91**
15	**4,74**	**496,56**	**120,76**	**288,60**	**854**	**1,78**	**76,60**	**1,72**	**47,87**	**3825**	**1799016**	**0,61**	**32,71**	**6**	**12**	**404,19**
16	**4,92**	**509,60**	**127,75**	**308,10**	**900**	**1,77**	**72,80**	**1,65**	**50,64**	**4192**	**2088391**	**0,75**	**15,95**	**4**	**11**	**423,97**
17	**4,72**	**479,58**	**122,29**	**294,80**	**855**	**1,77**	**71,00**	**1,62**	**48,48**	**3609**	**1601018**	**1,38**	**14,63**	**3**	**10**	**398,28**

Table 4: Selected models and statistical parameters of correlations between molecular descriptors and biological activity pIC5

Variables	pIC_{50}	MW	MR	MV	Pc	n	ɣ	d	αe	IW	IB	LogP	HLB	HD	HA	Vvw
pIC_{50}	**1**															
MW	-0,096	**1**														
MR	-0,340	**0,602**	**1**													
MV	-0,154	**0,960**	**0,584**	**1**												
Pc	-0,145	**0,987**	**0,602**	**0,992**	**1**											
N	0,146	-0,040	0,011	-0,309	-0,186	**1**										
ɣ	0,004	0,338	0,257	0,069	0,197	**0,888**	**1**									
D	0,113	0,485	0,311	0,222	0,342	**0,804**	**0,962**	**1**								
Ae	-0,138	**0,992**	**0,611**	**0,981**	**0,997**	-0,121	0,251	0,389	**1**							
IW	-0,056	**0,979**	**0,568**	**0,984**	**0,988**	-0,200	0,168	0,339	**0,983**	**1**						
IB	-0,031	**0,904**	0,480	**0,955**	**0,935**	-0,366	-0,015	0,174	**0,915**	**0,969**	**1**					
LogP	-0,124	**-0,611**	-0,294	-0,372	-0,484	**-0,698**	**-0,902**	**-0,955**	-0,529	-0,468	-0,296	**1**				
HLB	0,164	**0,763**	0,350	**0,653**	**0,710**	0,242	0,463	**0,563**	**0,739**	**0,702**	**0,568**	**-0,723**	**1**			
HD	0,108	0,211	0,101	-0,057	0,068	**0,894**	**0,947**	**0,897**	0,123	0,044	-0,131	**-0,867**	0,517	**1**		
HA	-0,146	**0,922**	**0,638**	**0,823**	**0,877**	0,177	0,550	**0,657**	**0,889**	**0,874**	**0,799**	**-0,722**	**0,689**	0,422	**1**	
Vvw	-0,106	**0,998**	**0,612**	**0,964**	**0,989**	-0,063	0,323	0,470	**0,992**	**0,980**	**0,909**	**-0,595**	**0,737**	0,191	**0,921**	**1**

The results of the PCA analysis are used to select the input data for the MLR. Thus, initially, we eliminated all descriptors with small (non-significant, $r \leq 0.3$) correlations to the dependent variable (pIC50). To reduce the redundancy existing in our data matrix, descriptors that are highly correlated ($r \geq 0.9$) and have the small value of the correlation coefficient with respect to the dependent variable were excluded (Table 4).

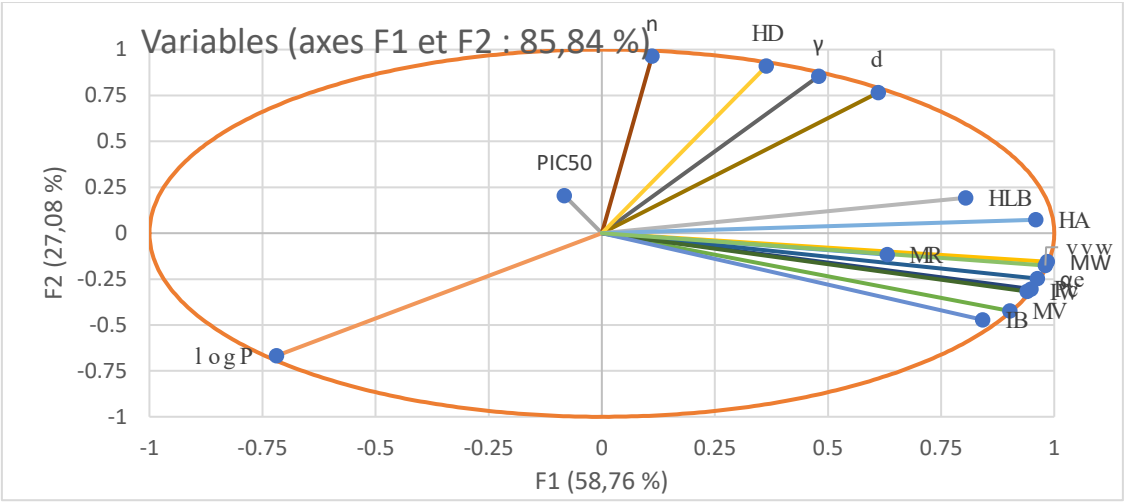


Figure 1. The correlation circle is the projection of the cloud of variables on the principal components plane

Figure 1 represents these descriptors in a correlation circle, the first two components F1 and F2 contributing respectively 58.76% and 27.08% to the total variance, the total information is estimated at a percentage of 85.84%.

3.3. Multiple linear regression (MLR)

Multiple Linear Regression has been used to study the effect in different descriptors on activity. Like simple linear regression, multiple linear regression seeks to approximate an overly complex relationship in general, by a simple mathematical function. [20]

It is based on the assumption that there is a linear relationship between the activity and several descriptors, so we have carried out a statistical study by this method. The mathematical equation obtained is the following:

$$pIC_{50}=0.82+1.22*10^{-4}\,MW+2.45*n-3.57*10^{-3}\,\gamma$$

The equation obtained from the RLM is an expression of activity as a function of 3 descriptors MW, n, and γ. We find that these descriptors are related to activity by coefficients of positive and negative sign, Indeed:

The pIC50 activity increases with increasing values of MW, n (both descriptors and their coefficients are positive).

The pIC50 activity increases with decreasing value of γ (sign (-) of the coefficient) taking into account the positive value of the surface tension.

N=17 R=**0,854**, $\mathbf{R^2 = 0,731}$, $\mathbf{R^2_{adj} = 0,666}$, **F=8,152**, **MSE= 0,003** , **P< 0, 0001**

The QSAR model must consider an $R^2 > 0.5$ for it to be valid. The values **R = 0.854** ($\mathbf{R^2}$ = **0.731**); allowed us to strongly indicate the correlation between the different descriptors used and the biological activity pIC50

This model presents significant correlation coefficient $\mathbf{R^2}$ value. The $\mathbf{R^2}$ value is equal to **0.731**, for this model. It indicates that this model can be successfully applied to predict the minimum inhibitor concentration of pIC50 bacteria.

In order to test the predictive power validity of the selected RLM models (final Eq), the LOO (leave-one-out) cross-validation technique was used. The developed models were validated by calculating the following statistical parameters: R^2_{adj}, goodness of fit, cross-validation correlation coefficient $Q^2_{CV(LOO)}$: $Q^2_{CV(LOO)}$=**0, 51** R^2_{adj} =**0,641**

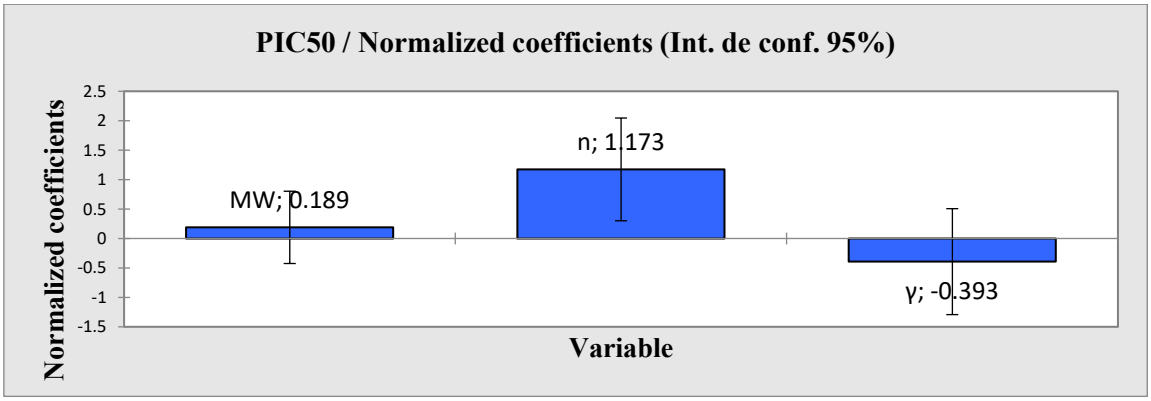


Figure 2: Box plot of standardized coefficients as a function of variables

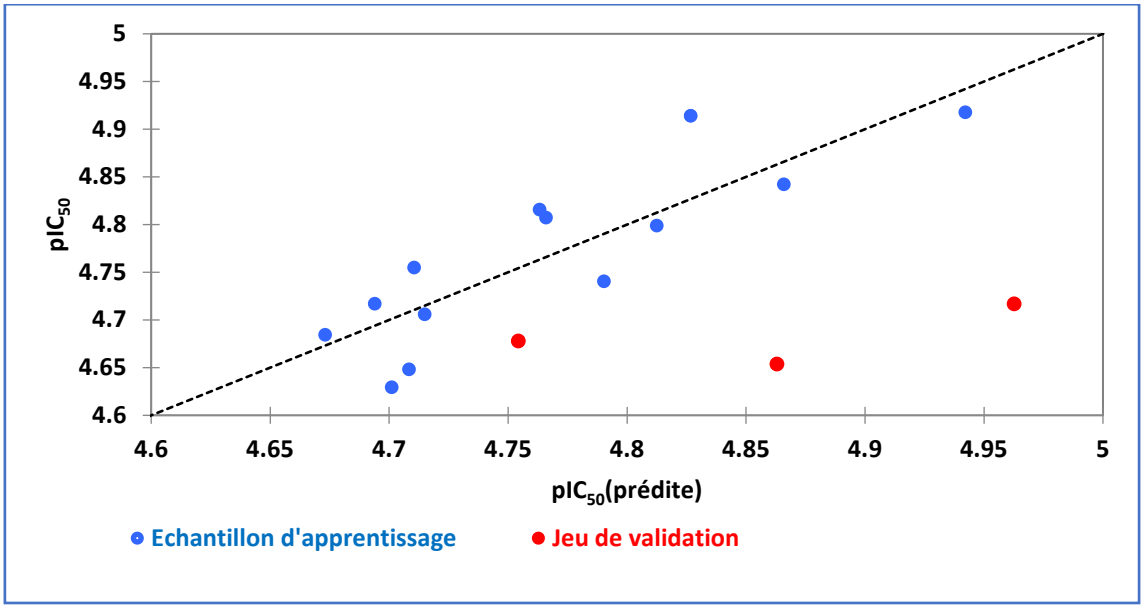


Figure 3. Correlation between the values of predicted and observed activities calculated.

The refractive index **n** and **MW** contribute positively while the surface tension contributes negatively.

The contribution of the refractive index is more important than those of **MW** and surface tension γ

We have represented on **figure 3**, a correlation between the experimental activity and the predicted activity according to the obtained model

3.3. y-randomization

The randomization test for the learning set constitutes the QSAR model and consists in randomly mixing 100 times the variables of the equation obtained by the statistical analysis RLM with the experimental values of the biological activity pIC, and the favorable result of this test and that if of the bad values of Rrand, R2rand and Q2cv(LOO)RAND. The computer tool used in this study hppt: //dtclab.webs.com/software-tools **[17].** Tables (4,5) show that the values of all the 100 tests i slow and also the average value of the model validation parameters by randomization. **[18].** these results and show that the model was chosen is not developed by chance.d

Table 5. Values of y-randomization test results

Rand	R	R^2	Q^2	rand	R	R^2	Q^2	Rand	R	R^2	Q^2	rand	R	R^2	Q^2
1	0,774	0,600	-2,822	26	0,471	0,222	-2,306	51	0,659	0,434	-2,862	76	0,785	0,616	-1,492
2	0,621	0,386	-3,052	27	0,658	0,433	-1,635	52	0,819	0,671	-0,484	77	0,667	0,445	-1,496
3	0,606	0,368	-2,920	28	0,788	0,621	-0,678	53	0,495	0,245	-2,724	78	0,768	0,590	-1,107
4	0,722	0,521	-1,226	29	0,680	0,462	-1,519	54	0,578	0,335	-2,163	79	0,855	0,732	-1,048
5	0,717	0,514	-1,181	30	0,828	0,685	-2,050	55	0,544	0,296	-3,367	80	0,745	0,555	-1,701
6	0,550	0,303	-3,237	31	0,346	0,120	-4,599	56	0,791	0,626	-1,192	81	0,740	0,548	-0,515
7	0,502	0,252	-3,729	32	0,707	0,500	-1,304	57	0,301	0,090	-3,784	82	0,541	0,292	-1,964
8	0,740	0,548	-1,839	33	0,711	0,505	-1,723	58	0,461	0,212	-2,690	83	0,706	0,498	-2,066
9	0,774	0,598	-1,513	34	0,442	0,196	-1,921	59	0,650	0,422	-2,222	84	0,577	0,333	-4,091
10	0,838	0,702	-1,163	35	0,903	0,816	-0,354	60	0,816	0,666	-0,025	85	0,695	0,483	-1,312
11	0,461	0,212	-2,847	36	0,511	0,261	-3,100	61	0,817	0,667	-0,471	86	0,628	0,395	-1,631
12	0,658	0,433	-0,981	37	0,802	0,643	-1,571	62	0,708	0,501	-1,217	87	0,521	0,272	-2,560

13	0,742	0,551	-0,609	38	0,522	0,273	-3,131	63	0,805	0,648	-0,228	88	0,659	0,435	-3,062
14	0,514	0,264	-3,600	39	0,536	0,287	-3,047	64	0,725	0,525	-0,957	89	0,526	0,277	-2,313
15	0,664	0,441	-2,237	40	0,553	0,306	-2,753	65	0,670	0,449	-1,625	90	0,700	0,490	-2,170
16	0,698	0,487	-0,696	41	0,708	0,501	-1,333	66	0,720	0,518	-1,815	91	0,682	0,466	-2,132
17	0,814	0,662	-0,823	42	0,898	0,806	0,232	67	0,749	0,560	-0,870	92	0,692	0,478	-1,210
18	0,745	0,554	-1,853	43	0,654	0,428	-1,333	68	0,563	0,316	-1,919	93	0,898	0,807	-0,037
19	0,742	0,551	-0,923	44	0,616	0,380	-2,742	69	0,806	0,649	-0,678	94	0,527	0,278	-2,548
20	0,736	0,542	-1,225	45	0,737	0,543	-2,434	70	0,859	0,739	-0,505	95	0,645	0,416	-2,546
21	0,830	0,689	-1,047	46	0,723	0,523	-1,215	71	0,591	0,349	-2,132	96	0,714	0,509	-1,375
22	0,704	0,495	-3,369	47	0,527	0,278	-1,660	72	0,539	0,291	-3,121	97	0,801	0,642	-0,193
23	0,834	0,696	-1,102	48	0,894	0,799	-0,726	73	0,784	0,614	-0,719	98	0,671	0,451	-2,216
24	0,814	0,662	-0,666	49	0,807	0,651	-1,199	74	0,668	0,446	-2,338	99	0,696	0,484	-1,565
25	0,621	0,386	-4,362	50	0,585	0,342	-3,514	75	0,755	0,570	-1,691	100	0,590	0,349	-3,718

Table 6. Random Models Parameters

Average R:	0,479
Average R^2:	0,477
Average Q^2_{cv}:	-1,845
cR_p^2:	0,571

Table 7. Comparison of the statistical parameter and Golbraikh and Tropsha's criteria

	Parameter	Equation	Model score	Threshold
Fitting criteria	R^2	$R^2=1-\frac{\sum(Y_{obs}-Y_{calc})^2}{\sum(Y_{obs}-\overline{Y_{obs}})^2}$	0,731	>0,60
	R^2_{adj}	$R^2_{adj}=\frac{(N-1)R^2-p}{N-P-1}$	0,641	>0,60
	MSE	$MSE=\frac{\sum(Y_{obs}-Y_{calc})^2}{N}$	0,003	A low value
	F	$F=\frac{\sum(Y_{calc}-\overline{Y_{calc}})^2-N-P-1}{\sum(Y_{obs}-Y_{calc})^2\ P}$	8,152	A high value
Internal validation	Q^2_{adj}	$Q^2_{adj}=1-\frac{\sum(Y_{obs}-Y_{calc})^2}{\sum(Y_{obs}-\overline{Y_{obs}})^2}$	0,518	>0,50
	R_{rand}	Average of the 100 $R_{rand(i)}$	0,479	<R
	R^2_{rand}	Average of the 100 $R^2_{rand(i)}$	0,477	$< R^2$
	$Q^2_{cvLoo(rand)}$	Average of the 100 $Q^2_{cvLoo(rand)(I)}$	-1,845	$<Q^2_c$
	cR_p^2	$cR^2_p=\sqrt{(R^2-(AverageR_{rand})^2)}$	0,571	>0,50
Extwrnal validation	R^2_{test}	$1-\frac{\sum(Y_{obs(test)}-Y_{(test)})^2}{\sum(Y_{obs(test)}-\overline{Y_{obs(test)}})^2}$	0,816	>0,50
	$\overline{r_m}\ test$	$\frac{\lvert r_m^2-r_m'^2\rvert}{2}$	0,660	>0,50
	$\Delta\ r^2_{test}$	$\lvert r_m^2-r_m'^2\rvert$	-0,007	<0,20
	$\Delta\ r^2_{0\ test}$	$\lvert r_0^2-r_0'^2\rvert$	0.084	<0,30
	$(r^2-r^2_0)/\ r^2$	$\frac{\lvert r^2-r_0^2\rvert}{r^2}$	0,077	<0,10

	$(r^2 - r'^2_0)/ r^2$	$\frac{\lvert r^2 - r_0'^2 \rvert}{r^2}$	0,02	<0,10
	K	$\frac{\sum Y_{obs} Y_{calc}}{\sum Y_{calc}^2}$	0.956	O,85≤K≤1,15
	K’	$\frac{\sum Y_{obs} Y_{calc}}{\sum Y_{obs}^2}$	1.045	O,85≤ K’≤1,15

3.4. External validation

For external validation we calculated the correlation and determination coefficient, R test = 0,816, R2test =0,666 and also the validation parameters of Golbraikh **[19],** the values obtained in Table 6 show the reliability of model chosen.

The result of the calculate leverage parameters hi presents in the form of Williams diagram (fig. 2) the diagram and presents the residual standard value as a function of hi with h* = 0.857 and x = ±2

We observe from fig. 4 that all the leverage values of each learning compound (blue), and test (red) are inside the domain of applicability, which means the solidity and reliability of model chosen by statistical analysis

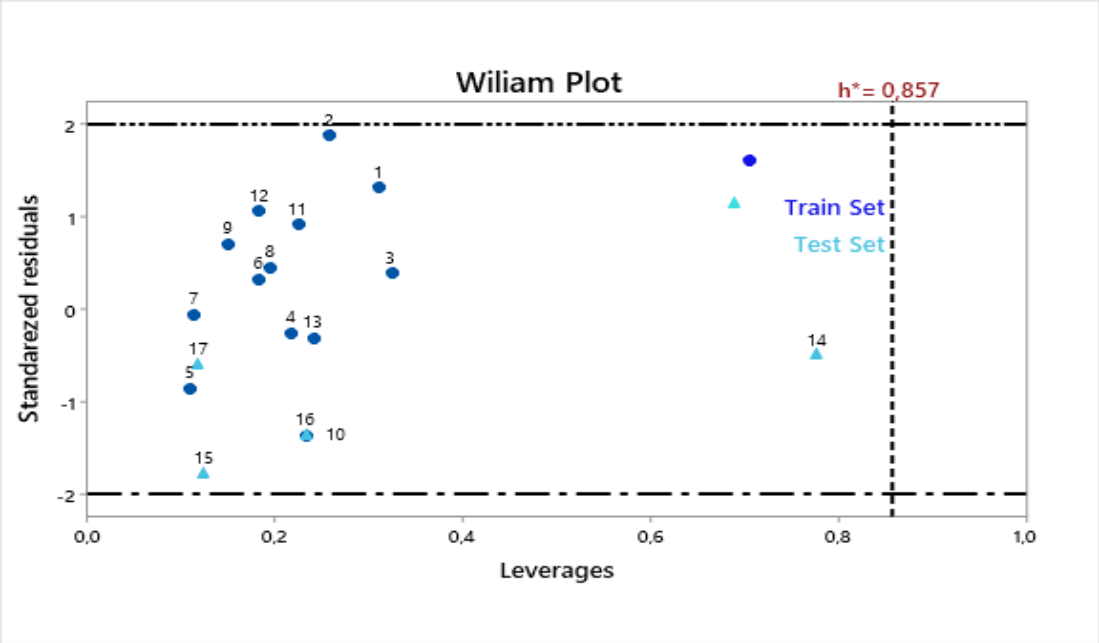


Figure 4. Williams plot of the normalized residual compared to the leverage for the MLR model.

4. Conclusions

This work is a study of the structural relationship of a series of analogues of 1,3,4-Thiadiazole Thio glycosides and 1,2,3-Triazolyl-1,3,4-Thiadiazole N-glycosides which is based on the QSAR method, consists in finding a correlation between the biological activity measured for a set of compounds of the series and certain descriptors calculated with the help of the MarvinSketch and ChemSketch software. The validity of the model was established by determining the appropriate statistical parameters. The implemented model was used to predict the minimum inhibitory concentration of 1, 3,4-Thiadiazole Thio glycosides and 1, 2,3-Triazolyl-1, 3,4-Thiadiazole N-glycosides derivatives with close agreement between experimental and predicted values. This model presents significant correlation coefficient R^2 value. The R2 values is equal to 0.731 for the MLR applied to predict the minimum inhibitory concentration of pIC50 bacteria, allow us to conclude that the model works well.

References

[1] « Synthesis, biological evaluation and 2D-QSAR analysis of benzoxazoles as antimicrobial agents », Eur. J. Med. Chem.,

vol. 44, no 2, p. 501-510, févr. 2009, doi: 10.1016/j.ejmech.2008.04.001.

[2] Hu, Y.; Li, C.Y.; Wang, X.M.; Yang, Y.H.; Zhu, H.L. 1,3,4-Thiadiazole: Synthesis, Reactions, and Applications in Medicinal, Agricultural, and Materials Chemistry. Chem. Rev. 2014, 114, 5572–5610

[3] El-Sayed, W.A.; Metwally, M.A.; Nada, D.S. Abdel-Rahman A.A.-H. Synthesis and Antimicrobial Activity of New Substituted 5-(Pyridine-3-yl)-1,3,4-thiadiazoles and Their Sugar Derivatives. J. Heterocycl. Chem. 2013, 50, 194–201.

[4] Jordheim, L.P.; Durantel, D.; Zoulim, F.; Dumontet, C. Advances in the development of nucleoside and nucleotide analogues for cancer and viral diseases. Nat. Rev. Drug. Disc. 2013, 12, 447–464

[5] "Chemical Structure Drawing Software | ACD/ChemSketch." https://www.acdlabs.com

[6] "ChemAxon - Software Solutions and Services for Chemistry & Biology." https://chemaxon.com

[7] Samir Chtita, Majdouline Larif, Mounir Ghamali, Mohammed Bouachrine, Tahar Lakhlifi, Quantitative structure–activity relationship studies of dibenzo[a , d]cycloalkenimine derivatives for non-competitive antagonists of N -methyl- d -aspartate based on density functional theory with electronic and topological descriptors . J. Taibah Univ. Sci. 9: 143–154 (2015). https://doi.org/10.1016/j.jtusci.2014.10.006

[8] XLSTAT | Logiciel statistique pour Excel. https://www.xlstat.com/fr/.

[9] M. Ghamali, Samir Chtita, Rachid Hmamouchi, Azeddine Adad, Mohammed Bouachrine, Tahar Lakhlifi, The inhibitory activity of aldose reductase of flavonoid compounds: Combining DFT and QSAR calculations. J. Taibah Univ. Sci. 10: 534–542 (2016). https://doi.org/10.1016/j.jtusci.2015.09.006

[10] Bok-Hee Jung and SoonGohn Kim " A Festival Satisfaction Evaluation Method using Multiple

[11] M. Ghamali, et al. QSAR analysis of the toxicity of phenols and thiophenols using MLR and ANN. J. Taibah Univ. Sci. 11: 1–10 (2017).

[12] S. Chtita, M. Ghamali, M. Larif, A. Adad, R. Hmammouchi, M. Bouachrine and T. Lakhlifi, Prediction of biological activity of imidazo [1, 2-a] pyrazine derivatives by combining DFT and QSAR results, International Journal of Innovative Research in Science, Engineering and Technology, 2(12): 7951-7962 (2013).

[13] Partha Pratim Roy, Somnath Paul, Indrani Mitra, Kunal Roy, On Two Novel Parameters for Validation of Predictive QSAR Models. Molecules 14: 1660–1701 (2009). https://doi.org/10.3390/molecules14051660

[14] L. Zhang, H. Zhu, T. I. Oprea, A. Golbraikh, A. Tropsha, QSAR modeling of the blood-brain barrier permeability for diverse organic compounds. Pharm. Res. 25: 1902–1914 (2008). https://doi.org/10.1007/s11095-008-9609-0

[15] K. Roy, I. Mitra, S. Kar, P. Kumar Ojha, R. N. Das, H. Kabir, Comparative studies on some metrics for external validation of QSPR models. J. Chem. Inf. Model. 52: 396–408 (2012). https://doi.org/10.1021/ci200520g

[16] A. Sankar, « Principal Component Analysis Part 1: The Different Formulations. », Medium, 29 septembre 2021. https://towardsdatascience.com/principal-component-analysis-part-1-the-different-formulations-6508f63a5553 (consulté le 26 juin 2022).

[17] Validation of (Q)SAR Models – OECD https://www.oecd.org/chemicalsafety/riskassessment/validationofqsarmodels.html.

[18] S. Chtita, A. Aouidate, A. Belhassan, A. Ousaa, A. Idrissi Taourati, B. Elidrissi, M. Ghamali, M. Bouachrine and Tahar Lakhlifi, QSAR study of N -substituted oseltamivir derivatives as potent avian influenza virus H5N1 inhibitors using quantum chemical descriptors and statistical methods. New J. Chem. 44: 1747–1760 (2020). https://doi.org/10.1039/C9NJ04909F

[19] A. Aouidate, A. Ghaleb, M. Ghamali, S. Chtita, A. Ousaa, M. Choukrad, A. Sbai, M. Bouachrine, and T. Lakhlifi, QSAR study and rustic ligand-based virtual screening in a search for aminooxadiazole derivatives as PIM1 inhibitors. Chem. Cent. J. 12: 1–12 (2018). https://dx.doi.org/10.1186%2Fs13065-018-0401-x

[20] Bok-Hee Jung and SoonGohn Kim " A Festival Satisfaction Evaluation Method using Multiple Regression Analysis "International Journal of Software Engineering and Its Applications

Vol.8, No.4 (2014), pp.187-196
http://dx.doi.org/10.14257/ijseia.2014.8.4.20