Arabian Journal of Chemical and Environmental Research

Vol. 06 Issue 1 (2019) 57–69

Antimicrobial evaluation against Escherichia coli (MTCC 1652) by using 1,4-disubstituted 1,2,3-triazole and derivatives: QSAR study

Y. KOUBI[1], Y. MOUKHLISS[1], A. SBAI[1], Y. El MASAOUDY[1], H. MAGHAT[1], R. EL-MERNISSI[1], M. AZIZ AJANA[1], M. BOUACHRINE[1,2,] and T. LAKHLIFI[1*]

[1]Molecular Chemistry and Natural Substances Laboratory (MCNSL),
Department of Chemistry, Faculty of Science. Moulay IsmailUniversity, BP 298Meknes, Morocco.
[2]Higher School of Technology (EST Khenifra), Sultan Moulay Sliman University, BP 23000 Beni-Mellal, Morocco.

Received 18 Dec 2018, Revised 26 Dec 2018, Accepted 26 Jan 2019

Abstract

In the present work, we carry out the antimicrobial evaluation against E. coli (MTTC1652), with a study between the activity and 2D-QSAR structure. The study made by a set consisting of 21 chemical structures of the 1,3-disubstituted triazole derivatives, each chemical structure is characterized by topological and electronic descriptors. The software ChemOffice and ChemSketch use for the computations of a topological descriptor, and for the computation of electronic descriptors with the functional theory of density whose hybrid base and the basic set are, successively: 3-Lee-Yang-Paar, 6-31G(d,p), we use the software Gaussian09. The data of 21 chemical compounds and divides into two sets at random, one is the learning set made up of 17 chemical compounds and there is the test set made up of 4 chemical compounds. The analysis performed in this work done by principal component analysis (PCA) and multilinear regression (RLM). The set of tests to be performed by changing the compounds of the two sets and the following were calculated the predicted pMIC values for the test set. The training and testing set validated separately with internal and external tests, such as y-randomization and Golbraikh and Tropsha validation model criteria. The model is validated with the high values of R^2, R^2_{test} and Q^2cv values (R^2= 0.80, R^2_{adj} =0.75 and Q^2cv=0.68, R^2_{test} =0.93, MSE=0.024).

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

*Corresponding author.
E-mail address: tahar.lakhlifi@yahoo.fr

ISSN : 2458-6544

1. Introduction

Escherichia coli was first isolated by Dr. Theodor Escherichia [1]. E. coli classes in the bacterial family of enterobacteria and living in the gastrointestinal tract of animals and humans [2]. Usually, it lives with hosts in association and rarely causes disease. However, it has caused uncomplicated urinary tract infections in humans and animals [3]. And frequently in women due to the proximity of the urethra and anus [4]. The appearance of multi-resistance between bacterial and microbial infection has seen the discovery of new powerful drugs of the treatment of infection resulting from infections by medical researchers. For that, the theoretical study between the target and the chemical structure and a means of discovering a new drug having biological activity, antibacterial against E. coli, certain works were used the QSAR study [5-6].

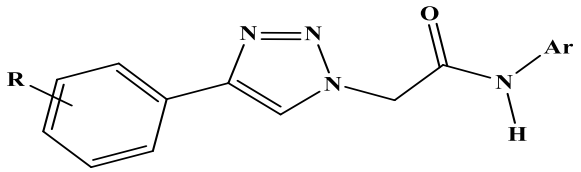
The evolution of the computer tool makes it possible to obtain a physical-chemical description of the chemical compounds in the theoretical study in a very precise manner. This descriptor uses variables to find molecules linking biological activity to the multidimensional chemical structure which does not allow high predictability of the biological of new organic molecules [5-7]. In this present study, we have sought to develop a model that links biological activity with the chemical structure of a series of 21 compounds derived from 1,3-disubstituted triazole by using 15 descriptors to build a 2D-QSAR model. Also, the statistical consistency of the advanced model was estimated on the base of their liaison capability for the traineeship set, as their predictive force for an exterior test est. Different groups of molecular descriptors were calculated to forecast the studied property of the 21 triazole derivatives using multiple linear regression (MLR). We, therefore, suggest quantitative models, and we try to explain the property of the studied compounds from the 2D-QSAR study [7]. The triazole derivatives the subject of this study was synthesized by a click reaction between the terminal alkynes and the N-substituted 2-azido-acetamide, all of the newly synthesized triazole (Table 1), were found to exhibit inhibitory activity against E. coli (MTCC 1652) [8].

2. Material and methods

2.1. Data sources

The data of the QSAR study of this work is collected from the literature [8], however of study consists of a 21 compounds drive of 1,4-disubstituted triazole which has present by their minimum inhibitory concentration (pMIC), Table 1 shows the 21 compounds with their pMIC (mol/m^3). For correct validation of the QSAR model, we have separated all 21 chemical compounds into two groups. The training group made up of 17 molecules to validate the QSAR model and 4 molecules the model test obtained there, the last group is called the test set, and the isolation of the two groups is done randomly.

Table 1. In vitro antimicrobial evaluation against E. coli of 1,4-disubstituted-1,2,3-triazole.



	Compound	Ar	R	pMIC
Training set	**4a**	C_6H_5	H	7.0467
	4b	4-$CH_3OC_6H_4$	H	7.0909
	4c	4-$NO_2C_6H_4$	H	7.4122
	4d	4-FC_6H_4	H	6.6757
	4e	4-ClC_6H_4	H	7.0974
	4f	4-BrC_6H_4	H	7.4190
	4g	α-Naphthyl	4-CH_3	7.1101
	4h	C_6H_5	4-CH_3	7.4306
	4i	4-$CH_3OC_6H_4$	4-CH_3	7.0936
	4l	4-ClC_6H_4	4-CH_3	6.8356
	4m	4-BrC_6H_4	4-CH_3	7.3767
	4n	α-Naphthyl	4-CH_3	7.4168
	4r	C_6H_5	3-F	7.4225
	4p	4-$CH_3OC_6H_4$	3-F	7.4775
	4q	4-$NO_2C_6H_4$	3-F	7.7569
	4r	4-FC_6H_4	3-F	7.0680
	4u	α-Naphthyl	3-F	7.4424
Test set	**4j**	4-$NO_2C_6H_4$	4-CH_3	6.8153
	4k	4-FC_6H_4	4-CH_3	7.4723
	4s	4-ClC_6H_4	3-F	7.7375
	4t	4-BrC_6H_4	3-F	7.7011

2.2. Molecular descriptors

In the QSAR study, there are numerous molecular descriptors to use as a variable in the QSAR model, and these obtained variables allow us to predict the biological activity of the compounds not yet formed. In this study we use to predict of the descriptor, on the one hand, the electronic descriptor that was obtained by theoretical calculates, with the theory of functional density (DFT) [9], with hybrid function B3LYP [10], and the 6-31G(d,p) [11], using Gaussian 09W software [12] are as follows: the total energy (Et), highest occupied molecular orbital energy (E_{HOMO}), lowest unoccupied molecular orbital energy (E_{LUMO}), the absolute electronegativity (χ), the dipole moment (DM), the softness index (S), the nucleophilic index (N), reactivity index (ω) and the absolute hardness (n) [13].

The n, χ, and ω were determined using the following equations:

$$S = \frac{1}{2\mu} \quad ; \quad \omega = \frac{\mu^2}{2\eta} \; ; \; \chi = \frac{-\varepsilon_{HOMO}+\varepsilon_{LUMO}}{2} \; ; \; \eta = (\varepsilon_{LUMO} - \varepsilon_{HOMO})$$

Else by using ChemOffice.16 and ChemSketch.2020 Software [14], we chose the following topological descriptor:

partition coefficient (LogP), molar volume (VM), polarizability (αe), parachor (Pr), molar weight (MW), and molar reactivity (MR), all topological ana electronic descriptors presented in Table 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 [15], using XLSTAT 16 [16]. The objective of the use ACP is for analyzing the relationship between the quantitative descriptor of substituted 1,3-triazole and also used to understand the distribution of the compounds [17], the result of ACP present in Table 3.

2.3.1. Principal component analysis

Multiple Linear regression (MLR) analysis is the conventional and standard approach for multivariate data analysis [18]. It is based on the ordinary least square regression (OLS) method. MLR is a method used for modeling the linear relationship between a dependent variable y (pMIC) and independent variable x (descriptors) [19]. MLR estimates the values of regression coefficients (R^2) by stratifying the least-squares adequate method. The model makes a relationship in the form of a straight line (linear) tat better present in the form:

$$y = a_1 * x_1 + a_2 * x_2 + a_3 * x_3 \dots\dots\dots\dots\dots + b$$

a_i : regression coefficients;

b: regression coefficients of the intersection;

2.4. Internal validation

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 R^2, R^2_{adj}, 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 [20]. 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) molecules, remaining using the chosen descriptor, this process and then repeated to remove and predict the values of all the molecules in the learning series [21]. 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.

2.5. External validation

Internal validation is a standard compulsory part of QSAR modeling according to some researchers [22]. 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 [23-24], have demonstrated insufficient parameter R_{test}, Q^2_{cv} (test) to control the predictive capacity of the QSAR model. Therefore, other parameters must be checked for this. These parameters are called “external validation criteria” or are often called “Tropsha criteria’’[25-22].

The domain applicability is a mandatory condition for measuring the use of the QSAR model according to the OECD validation principal [26]. We cannot say a model valid if and only the model and capable of creating to the prediction of new compounds can say exists as credible and not an extrapolation of the model [27].

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 [28]. 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 [27]. 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 strong predictability.

3. Results and discussion

3.1. Molecular descriptors

To determine a quantitative relationship between the structure and activity 2D-QSAR study was performed 21 amides linked 1,4-disubstituted 1,2,3-triazoles to antimicrobial E. coli (MTCC 1652) [8]. The values of the 15 descriptors are shown in Table 2.

The matrix result of the analysis in the principal component has regrouped the sum of data 15 descriptors associated with the 21 components of triazole (Table 3). The variance of the three main axes F1, F2, F3, is 38.057%, 33.268%, and 15.304% and adequate to describe the maximum information of the matrix that is estimated to be equal (86.99%).

Table 2. values of the obtain descriptors of the studied 1,4-disubstituted triazoles.

Compounds	pMIC	E_t	DM	E_{HOMO}	E_{LUMO}	χ	η	S(I)	ω	N	MW	MR	MV	Pc	αe	LogP
4a	7.05	-912.37	3.77	-0.22	-0.03	-0.09	0.19	2.68	0.023	0.21	280.32	82.60	226.80	606.70	32.74	2.97
4b	7.09	-1026.89	5.46	-0.21	-0.02	-0.10	0.19	2.61	0.024	0.22	308.33	88.42	248.50	657.00	35.05	2.81
4c	7.41	-1116.87	2.31	-0.23	-0.10	-0.06	0.13	3.99	0.016	0.21	323.31	88.26	232.10	652.20	34.99	2.91
4d	6.68	-1011.60	3.25	-0.22	-0.04	-0.09	0.18	2.73	0.023	0.21	296.30	82.47	229.70	606.90	32.69	3.11
4e	7.10	-1371.96	4.23	-0.23	-0.03	-0.10	0.19	2.62	0.024	0.21	312.75	87.20	236.10	635.60	34.57	3.57
4f	7.42	-3483.48	4.16	-0.22	-0.03	-0.09	0.19	2.65	0.024	0.21	404.46	123.59	330.40	883.30	48.99	3.96
4g	7.11	-1066.01	4.23	-0.22	-0.06	-0.08	0.16	3.12	0.020	0.21	322.36	92.84	263.70	688.10	36.80	3.32
4h	7.43	-951.70	2.58	-0.22	-0.03	-0.10	0.19	2.59	0.024	0.21	337.33	92.69	247.30	683.30	36.74	3.42
4i	7.09	-1066.22	2.16	-0.20	-0.03	-0.09	0.17	2.87	0.022	0.23	310.33	86.90	244.90	638.00	34.45	3.62
4j	6.82	-1156.20	8.17	-0.23	-0.08	-0.07	0.15	3.43	0.018	0.20	326.78	91.63	251.30	666.70	36.32	4.08
4k	7.47	-1050.93	3.84	-0.22	-0.03	-0.10	0.19	2.62	0.024	0.21	371.23	94.58	254.60	681.40	37.49	4.25
4l	6.84	-1411.29	2.46	-0.21	-0.04	-0.09	0.17	2.91	0.021	0.22	418.49	128.02	345.60	914.40	50.75	6.12
4m	7.38	-3522.80	4.51	-0.22	-0.03	-0.10	0.19	2.63	0.024	0.21	296.30	82.47	229.70	606.90	32.69	3.11
4n	7.42	-11105.3	3.75	-0.21	-0.06	-0.08	0.15	3.23	0.019	0.22	326.33	88.29	251.40	657.20	35.00	2.95
4o	7.42	-1011.61	1.04	-0.22	-0.04	-0.09	0.19	2.70	0.023	0.21	330.74	87.07	239.00	635.80	34.52	3.71
4p	7.48	-1126.13	1.88	-0.20	-0.03	-0.09	0.17	2.91	0.021	0.23	375.20	90.03	242.30	650.50	35.69	3.88
4q	7.76	-1216.10	6.50	-0.23	-0.10	-0.07	0.14	3.67	0.017	0.20	357.20	90.16	239.40	650.30	35.74	3.73
4r	7.07	-1110.83	5.01	-0.23	-0.03	-0.10	0.19	2.62	0.024	0.21	292.34	87.03	242.00	637.80	34.50	3.48
4s	7.74	-1471.13	5.00	-0.23	-0.04	-0.09	0.18	2.70	0.023	0.20	341.30	88.13	235.00	652.40	34.94	3.05
4t	7.70	-3582.70	4.96	-0.23	-0.04	-0.09	0.18	2.71	0.023	0.20	314.29	82.34	232.60	607.00	32.64	3.25
4u	7.44	-1165.25	5.69	-0.22	-0.06	-0.08	0.16	3.04	0.021	0.21	422.45	123.46	333.30	883.40	48.94	5.74

Table. 3. Values of correlation matrix obtained through ACP

Variables	E_T	DM	E_{HOMO}	E_{LUMO}	χ	η	S(I)	ω	N	MW	MR	MV	Pc	αe	LogP
E_T	**1**	-0.02	-0.17	0.04	-0.12	0.12	-0.09	0.12	-0.17	0.02	0.02	-0.03	0.007	0.02	0.20
DM	-0.02	**1**	-0.56	-0.38	0.19	-0.19	0.18	-0.19	-0.56	-0.03	0.02	0.03	0.017	0.02	0.02
E_{HOMO}	-0.17	-0.56	**1**	0.49	-0.15	0.15	-0.20	0.15	**1.00**	0.03	0.03	0.10	0.048	0.03	0.04
E_{LUMO}	0.04	-0.38	0.49	**1**	**-0.93**	**0.93**	**-0.94**	**0.93**	0.49	-0.12	-0.04	0.08	-0.039	-0.04	-0.05
χ	-0.12	0.19	-0.15	**-0.93**	**1**	**-1.00**	**0.99**	**-1.00**	-0.15	0.16	0.06	0.03	0.064	0.06	0.08
η	0.12	-0.19	0.15	**0.93**	**-1.00**	**1**	**-0.99**	**1.00**	0.15	-0.16	-0.06	-0.03	-0.064	-0.06	-0.08
S(I)	-0.09	0.18	-0.20	**-0.94**	**0.99**	**-0.99**	**1**	**-0.99**	-0.20	0.12	0.02	-0.01	0.029	0.02	0.03
ω	0.12	-0.19	0.15	**0.93**	**-1.00**	**1.00**	**-0.99**	**1**	0.15	-0.16	-0.06	-0.03	-0.064	-0.0	-0.08
N	-0.17	-0.56	**1.00**	0.49	-0.15	0.15	-0.20	0.15	**1**	0.03	0.03	0.10	0.048	0.03	0.04
MW	0.02	-0.03	0.03	-0.12	0.16	-0.16	0.12	-0.16	0.03	**1**	**0.88**	**0.84**	**0.87**	**0.88**	**0.81**
MR	0.02	0.02	0.03	-0.04	0.06	-0.06	0.02	-0.06	0.03	**0.88**	**1**	**0.98**	**0.99**	**1.00**	**0.83**
MV	-0.03	0.03	0.10	0.01	0.03	-0.03	-0.01	-0.03	0.10	**0.84**	**0.98**	**1**	**0.99**	**0.98**	**0.82**
Pc	0.01	0.01	0.04	-0.03	0.06	-0.06	0.03	-0.06	0.04	**0.87**	**0.99**	**0.99**	**1**	**0.99**	**0.81**
αe	0.02	0.02	0.03	-0.04	0.06	-0.06	0.02	-0.06	0.03	**0.88**	**1.00**	**0.98**	**0.99**	**1**	**0.83**
LogP	0.20	0.02	0.05	-0.05	0.08	-0.08	0.03	-0.08	0.04	**0.81**	**0.83**	**0.82**	**0.81**	**0.83**	**1**

The purpose of this analysis is to reduce the number of variables (descriptor) according to PCA **[17]**. The correlation condition between the variables is defined by the value of **r**: if r greater than 0.5 we have

collinearity between the variables if r less than 0.5 we have a non-collinear relationship between the variables [29].

In this study the correlation is perfectly correlated between, (MR, αe), (ω, η), (ω, E_{LUMO}) and (N, E_{HOMO}) with a correlation value equal to 1, in these variables are redundant. ω and χ relate to a perfectly negative correlation r = - 1, and a large observed value of collinearity between the variables (S, ω) with r = 0.99 and (Pc, αe) r = 0.98 at the end by eliminating the following variable, E_{LUMO}, αe, χ, MR, S and Pc.

3.2. Multiple linear regression

After several tests, we have developed a model that links the descriptor of chemical structure with the biological activity pMIC, but the best combination obtained by RLM admits a linear relation with only three descriptors MW, LogP, and E_{HOMO} by contribution to the other descriptor manipulated in the study. The linearity equation of this model and presented in the form:

$$\mathbf{pMIC = 2.56+1.10^{-2}MW-0.53LogP-13.92*E_{HOMO}}$$

the model chosen is valid with the following parameters:

$$\mathbf{N=21 \quad R^2 = 0.80 \quad , \quad R^2_{adj} =0.75 \quad , \quad F=17.281 \quad , \quad MSE= 0.024 \quad , \quad P< 0.0001}$$

Based on these values, the pMIC equation obtained is statistically significant.

Manipulation of the model obtained by the statistical study of the RLM is cross-validation and shows their solidity according to the values of the cross-validation LOO Q^2_{cv}= 0.68, and this value is greater than 0.5. The model and also valid by the best metric value of $\overline{r_m^2}(cvLoo)$=0.578 $\overline{\Delta r_m^2}(cvLoo)$ =0.001

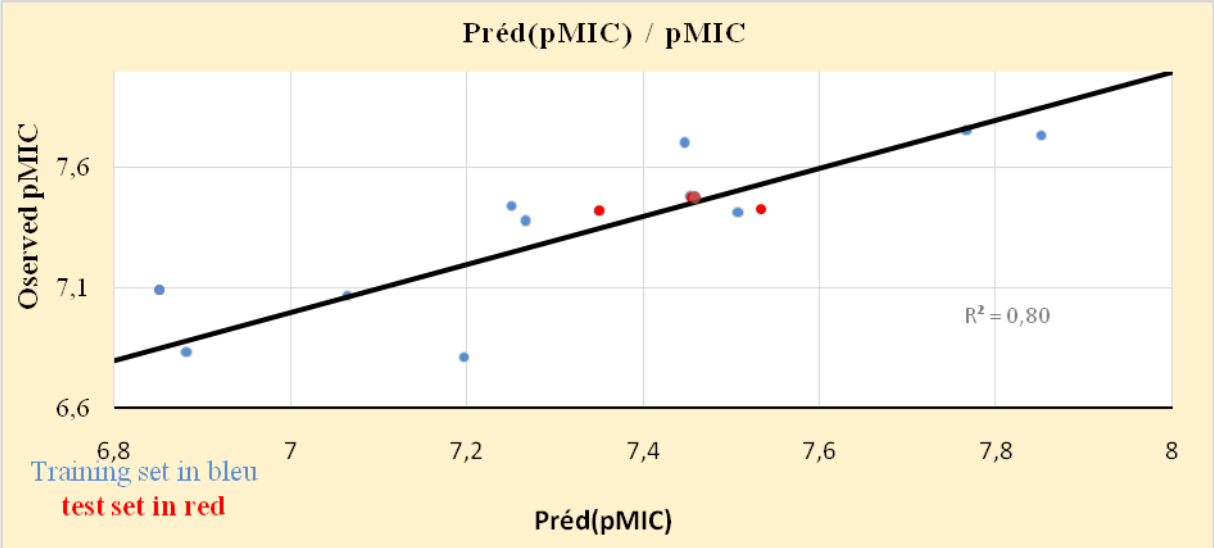


Fig 1. Correlation between the values of predicted and observed activities calculated.

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 pMIC, and the favorable result of this test and that if of the bad values of R_{rand}, R^2_{rand} and $Q^2_{cv(LOO)RAND}$. The computer tool used in this study https: //dtclab.webs.com/software-tools **[27]**. **Tables 4 and 5** show that the values of all the 100 tests i and also the average value of the model validation parameters by randomization. These results and show that the model was chosen is not developed by chance.

Table 4. 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	0.222	0.049	-0.345	**26**	0.409	0.167	-0.177	**51**	0.373	0.139	-0.363	**76**	0.077	0.006	-0.407
2	0.301	0.091	-0.303	**27**	0.361	0.130	-0.294	**52**	0.137	0.019	-0.439	**77**	0.247	0.061	-0.257
3	0.323	0.104	-0.336	**28**	0.485	0.235	-0.258	**53**	0.351	0.123	-0.399	**78**	0.282	0.079	-0.344
4	0.199	0.040	-0.412	**29**	0.474	0.224	-0.041	**54**	0.271	0.073	-0.483	**79**	0.358	0.128	-0.283
5	0.368	0.136	-0.355	**30**	0.307	0.094	-0.344	**55**	0.068	0.005	-0.480	**80**	0.181	0.033	-0.680
6	0.356	0.126	-0.208	**31**	0.183	0.034	-0.340	**56**	0.347	0.120	-0.401	**81**	0.425	0.181	-0.198
7	0.323	0.104	-0.425	**32**	0.404	0.163	-0.340	**57**	0.309	0.095	-0.269	**82**	0.293	0.086	-0.328
8	0.496	0.246	-0.194	**33**	0.111	0.012	-0.399	**58**	0.370	0.137	-0.381	**83**	0.382	0.146	-0.397
9	0.382	0.146	-0.236	**34**	0.156	0.024	-0.420	**59**	0.347	0.120	-0.270	**84**	0.501	0.251	-0.032
10	0.380	0.145	-0.260	**35**	0.290	0.084	-0.321	**60**	0.163	0.027	-0.474	**85**	0.663	0.440	0.084
11	0.265	0.070	-0.424	**36**	0.090	0.008	-0.330	**61**	0.411	0.169	-0.269	**86**	0.092	0.009	-0.303
12	0.185	0.034	-0.495	**37**	0.643	0.414	0.099	**62**	0.312	0.097	-0.419	**87**	0.488	0.238	-0.155
13	0.270	0.073	-0.396	**38**	0.278	0.077	-0.371	**63**	0.480	0.230	-0.203	**88**	0.306	0.094	-0.306
14	0.422	0.178	-0.246	**39**	0.478	0.229	-0.128	**64**	0.643	0.413	0.086	**89**	0.360	0.129	-0.254
15	0.337	0.114	-0.494	**40**	0.448	0.200	-0.368	**65**	0.459	0.211	-0.126	**90**	0.257	0.066	-0.452
16	0.365	0.133	-0.382	**41**	0.151	0.023	-0.294	**66**	0.344	0.118	-0.314	**91**	0.258	0.066	-0.328
17	0.240	0.058	-0.497	**42**	0.502	0.252	-0.025	**67**	0.274	0.075	-0.354	**92**	0.344	0.119	-0.238
18	0.336	0.113	-0.241	**43**	0.114	0.013	-0.452	**68**	0.184	0.034	-0.462	**93**	0.369	0.136	-0.479
19	0.424	0.180	-0.332	**44**	0.311	0.096	-0.326	**69**	0.407	0.165	-0.201	**94**	0.207	0.043	-0.448
20	0.186	0.035	-0.441	**45**	0.421	0.178	-0.234	**70**	0.192	0.037	-0.401	**95**	0.482	0.233	-0.155
21	0.484	0.234	-0.197	**46**	0.304	0.092	-0.300	**71**	0.415	0.172	-0.293	**96**	0.477	0.228	-0.137
22	0.553	0.306	0.063	**47**	0.605	0.366	0.007	**72**	0.313	0.098	-0.508	**97**	0.337	0.113	-0.436
23	0.254	0.065	-0.296	**48**	0.247	0.061	-0.483	**73**	0.623	0.388	-0.033	**98**	0.287	0.083	-0.339
24	0.445	0.198	-0.131	**49**	0.341	0.117	-0.411	**74**	0.385	0.148	-0.214	**99**	0.528	0.279	-0.008
25	0.307	0.094	-0.393	**50**	0.475	0.225	-0.113	**75**	0.327	0.107	-0.319	**100**	0.386	0.149	-0.416

Table 5. Random Models Parameters

Average R:	0.34102
Average R^2:	0.13306
Average Q^2_{cv}:	-0.3243
cR_p^2:	0.72034

3.4. External validation

For external validation, we calculated the correlation and determination coefficient, R_{test}= 0.96, R^2_{test}=0.93, and also the validation parameters of Golbarikh and Tropsha **[25]**, the values obtained in **Table 6** show the reliability of model chosen.

Table 6. Comparison of the statistical parameter and Golbraikh and Tropsha criteria

	Parameter	Equation	Model score	Threshold
Fitting criteria	R^2	$R^2=1-\frac{\sum(Y_{obs}-Y_{calc})^2}{\sum(Y_{obs}-\overline{Y_{obs}})^2}$	0,800	>0,60
	R^2_{adj}	$R^2_{adj}=\frac{(N-1)R^2-p}{N-P-1}$	0,753	>0,60
	MSE	$MSE=\frac{\sum(Y_{obs}-Y_{calc})^2}{N}$	0,024	A low value
	F	$F=\frac{\sum(Y_{calc}-\overline{Y_{calc}})^2-N-P-1}{\sum(Y_{obs}-Y_{calc})^2\ P}$	17,281	A high value
Internal validation	Q^2_{adj}	$Q^2_{adj}=1-\frac{\sum(Y_{obs}-Y_{calc})^2}{\sum(Y_{obs}-\overline{Y_{obs}})^2}$	0,682	>0,50
	R_{rand}	Average of the 100 $R_{rand(i)}$	0,341019	<R
	R^2_{rand}	Average of the 100 $R^2_{rand(i)}$	0,133061	< R^2
	$Q^2_{cvLoo(rand)}$	Average of the 100 $Q^2_{cvLoo(rand)(I)}$	-0,3243	<Q^2_c
	cR_p^2	$cR^2_p=\sqrt{(R^2-(AverageR_{rand})^2)}$	0,720336	>0,50
	$\overline{r^2_m}(cvLoo)$	$\frac{\lvert r_m^2-r_m'^2\rvert}{2}$	0,578	>0,50
	$\Delta\overline{r^2_m}(cvLoo\ (cvLoo)$	$\lvert r_m^2-r_m'^2\rvert$	0,001	<0,20
External validation	R^2_{test}	$1-\frac{\sum(Y_{obs(test)}-Y_{(test)})^2}{\sum(Y_{obs(test)}-\overline{Y_{obs(test)}})^2}$	0,936	>0,50
	$\overline{r_m}\ test$	$\frac{\lvert r_m^2-r_m'^2\rvert}{2}$	0,741	>0,50
	$\Delta\ r^2_{test}$	$\lvert r_m^2-r_m'^2\rvert$	-0,11	<0,20
	$\Delta\ r^2_{0\ test}$	$\lvert r_0^2-r_0'^2\rvert$	0,096	<0,30
	$(r^2-r^2_0)/\ r^2$	$\frac{\lvert r^2-r_0^2\rvert}{r^2}$	-0,002	<0,10
	$(r^2-r'^2_0)/\ r^2$	$\frac{\lvert r^2-r_0'^2\rvert}{r^2}$	0,07	<0,10
	K	$\frac{\sum Y_{obs}Y_{calc}}{\sum Y_{calc}^2}$	1,007	0,85≤K≤1,15
	K’	$\frac{\sum Y_{obs}Y_{calc}}{\sum Y_{obs}^2}$	0,992	0,85≤ K’≤1,15

The result of the calculate leverage parameters hi presents in the form of Williams diagram (**Figure 2**) the diagram and presents the residual standard value as a function of hi with h* = 0.71 and x = ±3

We observe from **figure 2** 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

$$h^* = \frac{3(k+1)}{n}$$

And k=3 n = 17

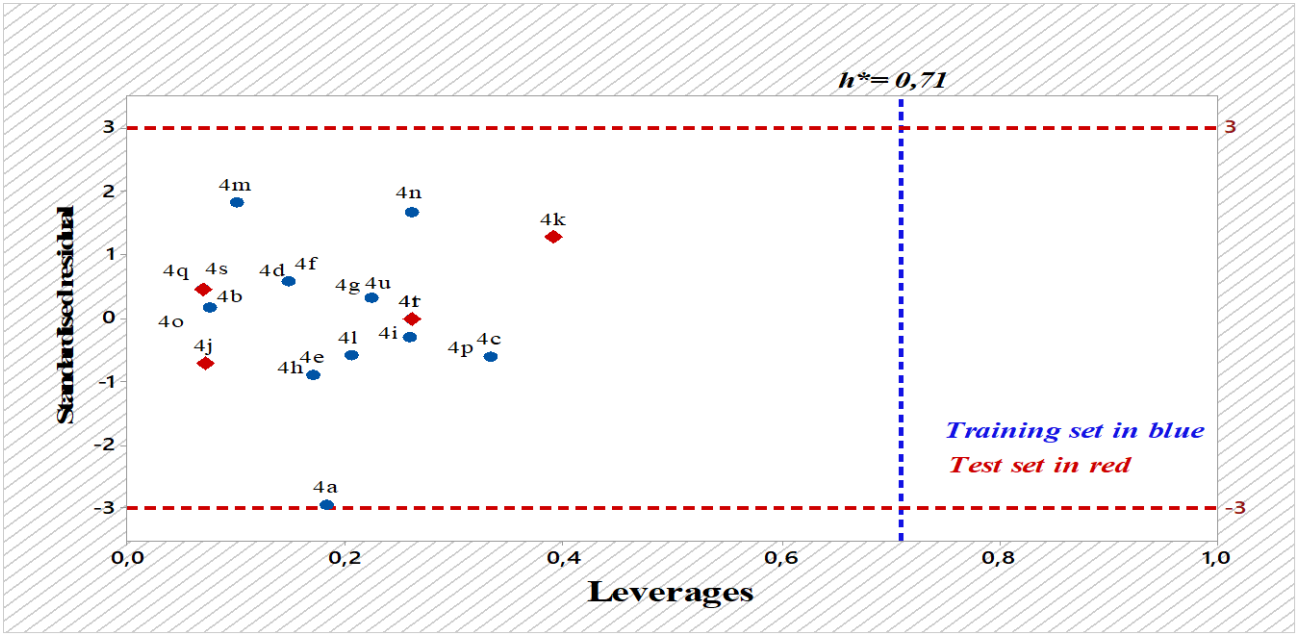


Figure 2. Williams plot of the normalized residual compared to the leverage for the MLR model

Conclusion

In this 2D-QSAR study of 21 triazole derivatives, we have studied the predictability of a mathematical model for biological activity pMIC. The model is shown here reliability and solidity values of R^2 and R^2_{adj}. So, the model is manipulated to be a set of tests. First, the LOO cross-validation test which does not give a value of Q^2_{cv} = 0.68 sufficient and the Y-randomization test for also shows that the model was not chosen random according to the values of R_{rand}, R2rand, $Q^2_{cv\ Loo(rand)}$, rand, and cR^2_p and secondly the parameter of Golbraikh and Trouphsa For still ensures the model obtained and finally the domain of applicability to verify that the model and capable of predicting antibacterial biological activity E. coli (MTCC 1652) for molecules not yet formed.

References

[1]. S. E. Woodward, coli: A brief overview. *British Journal of Nursing* 24: 158–159 (2015).

[2]. J. B. Kaper, J. P. Nataro, & H. L. T. Mobley, Pathogenic Escherichia coli. *Nature Reviews Microbiology*. 2 123–140 (2004).

[3]. Tarun Madappa, Chi Hiong U Go, Escherichia coli (E coli) Infections: Background, Pathophysiology, (2019) *Epidemiology.* https://emedicine.medscape.com/article/217485-overview.

[4]. R. Ben-Ami, J. Rodríguez-Baño, H. Arslan, J. D D Pitout, C. Quentin, E. S Calbo, O. K Azap, C. Arpin, A. Pascual, D.M Livermore, J. Garau, Y. Carmeli, A Multinational Survey of Risk Factors for Infection with Extended-Spectrum β-Lactamase–Producing Enterobacteriaceae in Nonhospitalized Patients. *Clin. Infect. Dis.* 49: 682–690 (2009). https://doi.org/10.1086/604713

[5]. D. A. Montero, F. Del Canto, J. C. Salazar, S. Céspedes, L. Cádiz, M. Arenas-Salinas, J. Reyes, Á. Oñate & R. M. Vidal, Immunization of mice with chimeric antigens displaying selected epitopes confers protection against intestinal colonization and renal damage caused by Shiga toxin-producing Escherichia coli. *NPJ Vaccines* 5: 20 (2020). https://doi.org/10.1038/s41541-020-0168-7

[6]. X. Lin, X. Li, X. Lin, A review on applications of computational methods in drug screening and design. *Molecules* 25: 1–17 (2020).

[7]. A. Cherkasov, E. N. Muratov, D. Fourches, A. Varnek, I. I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y. C. Martin, R. Todeschini, V. Consonni, V. E. Kuz'min, R. Cramer, R. Benigni, C. Yang, J. Rathman, L. Terfloth, J. Gasteiger, A. Richard, and A. Tropsha, QSAR modeling: Where have you been? Where are you going to? *Journal of Medicinal Chemistry* vol. 57: 4977–5010 (2014). https://dx.doi.org/10.1021%2Fjm4004285

[8]. C. P. Kaushik, A. Pahwa, D. Singh, K. Kumar, & R. Luxmi, Efficient synthesis, antitubercular and antimicrobial evaluation of 1,4-disubstituted 1,2,3-triazoles with amide functionality. *Monatshefte fur Chemie* 150: 1127–1136 (2019).

[9]. A. D. Becke, Perspective: Fifty years of density-functional theory in chemical physics. *J. Chem. Phys.* 140: 18A301 (2014).

[10]. O. Salomon, M. Reiher & B. A. Hess, Assertion and validation of the performance of the B3LYP* functional for the first transition metal row and the G2 test set. *J. Chem. Phys.* 117: 4729–4737 (2002).

[11]. M. E. Rădulescu-Grad, A. Visa, M. Silviu Mile, R. Ioan Lazău, S. Popa, S. Funar-Timofei, Synthesis, spectral characterization, and theoretical investigations of a new azo-stilbene dye for acrylic resins. *J. Mol. Struct.* 1217: 128380 (2020). https://doi.org/10.1016/j.molstruc.2020.128380

[12]. Gaussian.com | Expanding the limits of computational chemistry. https://gaussian.com/.

[13]. Anouar Zerouali & Houria Nacer, L'Electrophylicite, un descripteur prometteur pour la prediction

de la toxicite. Université Oum El Bouaghi, (2019). http://hdl.handle.net/123456789/8944

[14]. PerkinElmer | For The Better. https://www.perkinelmer.com/fr.

[15]. 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

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

[17]. 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

[18]. A. Adebanji, M. Asamoah-Boaheng, & O. Osei-Tutu, Asymptotic performance of the quadratic discriminant function to skewed training samples. *Springerplus* 5: (2016).

[19]. K. Ennouri, *et al.* Multiple linear regression and artificial neural networks for delta-endotoxin and protease yields modelling of Bacillus thuringiensis. *3 Biotech* 7: (2017).

]20]. 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).

[21]. 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).

[22]. 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

[23]. 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

[24]. 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

[25]. A. Golbraikh, & A. Tropsha, Beware of q2! *J. Mol. Graph. Model.* 20: 269–276 (2002). https://doi.org/10.1016/s1093-3263(01)00123-1

[26]. Validation of (Q)SAR Models - OECD. https://www.oecd.org/chemicalsafety/risk-assessment/validationofqsarmodels.html.

[27]. 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

[28]. 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

[29]. Pavel Polishchuk, Interpretation of Quantitative Structure-Activity Relationship Models: Past, Present, and Future. *Journal of Chemical Information and Modeling* 57: 2618–2639 (2017). https://doi.org/10.1021/acs.jcim.7b00274
