Moroccan Journal of Chemistry
ISSN: 2351-812X
http://revues.imist.ma/?journal=morjchem&page=login
Hammoudan & al. / Mor. J. Chem. 10 N°3 (2022) 405-416
DOI: https://doi.org/10.48317/IMIST.PRSM/morjchem-v10i3.33059

QSAR study of a series of peptidomimetic derivatives towards MERS-CoV inhibitors

***I. Hammoudan*[(a,*)], *S. Chtita*[(b)], *M. Bakhouch*[(c)], *D. Riffi Temsamani*[(a)].**

[(a)] *Laboratory Physical Chemistry, Faculty of Sciences of Tetouan, University Abdelmalek Essaadi, Tetouan, Morocco.*
[(b)] *Laboratory of Analytical and Molecular Chemistry, Faculty of Sciences Ben M'Sik, Hassan II University of Casablanca, B.P. 7955, Casablanca, Morocco.*
[(c)] *Laboratory of Bioorganic Chemistry, Department of Chemistry, Faculty of Sciences, Chouaïb Doukkali University, P.O. Box. 24, 24000, El Jadida, Morocco*

***** *Correspondence:*
imad_2005_05@hotmail.com

Received 12 April 2022
Revised 10 Jun 2022
Accepted 15 Jun 2022

Abstract

In this study, we report the quantitative structure activity relationships (QSAR) investigation to determine the relationship between the anti-MERS-CoV activity and a set of chemical descriptors computed using ChemSketch, MarvinSketch and ChemOffice software. Herein, the principal components analysis (PCA), multiple linear regression (MLR) and multiple non-linear regression (MNLR) methods were used with the intention to obtain a reliable QSAR model with good predictive capacity. The original data set of 43 peptidomimetic compounds was randomly divided into training and test set of 35 and 8 compounds, respectively. The values obtained by MLR and MNLR for the determination coefficient are 0.777 and 0.813, respectively. The predictive ability of the MLR model was assessed by external validation using the eight compounds of the test set with predicted determination coefficients R^2_{test} of 0.655.

Keywords: *MERS-CoV, QSAR, ACP, MLR, NLMR.*

1. Introduction

Middle East Respiratory Syndrome (MERS) is an emerging infectious disease, discovered for the first time in September 2012 in Saudi Arabia [1,2]. This disease has spread throughout many countries of the Middle East, such as Egypt, Oman, Qatar and Saudi Arabia, which they have recorded the largest percentage of positive cases (> 85%) [3-5]. In 2015, among 27 countries affected by MERS-CoV, South Korea has registered the largest confirmed MERS cases (186 cases) with 38 deaths. Thus, this country became the top of the list with high rate of positive MERS-CoV cases outside the Saudi Arabia Kingdom. The unusual and the scary thing of the MERS-CoV outbreak is the high rate of death, which is estimated about 35% [6]. The transmission of the virus from an animal reservoir to human was associated with dromedary camels along with human-to-human transmission are the two known modes of infection [7-8]. The virus appears to cause more severe disease in older people, people with weakened immune systems, and those with chronic diseases such as renal disease, cancer, chronic lung disease, and diabetes. More than two hundred cases were reported in 2019 and there is currently no specific vaccine or treatment for MERS-CoV. Available treatment is supportive and depends on the patient's clinical condition. The 3C-like protease ($3CL^{pro}$) is essential for viral replication and thus represents a potential target for antiviral drug development. Nowadays, very few data are available on MERS-CoV $3CL^{pro}$ inhibition by small molecules; HKU4-CoV $3CL^{pro}$ shares high sequence identity (81%) with the MERS-CoV enzyme and thus represents a potential surrogate model for anti-MERS drug discovery [9]. In this study, quantitative structure-activity relationship (QSAR) was used as basic tools in drug design. Whereas, QSAR attempts to explore the relationship between molecular descriptors that describe the unique physicochemical properties of a set of compounds of interest with their respective biological activity [10]. It defines a chemical structure through a variety of molecular descriptors, such as constitutional, topological, thermodynamic, geometrical descriptors,…etc. Recently, the development of new cheminformatics software allows to calculate thousands of molecular descriptors [11]. The QSAR method involves recognition that molecules (peptide, protein, …etc.) are really a three- dimensional distribution of properties. The most important of these properties are steric (eg shape and volume), electronic (eg electric charge and electrostatic potential), and what are termed 'lipophilic' properties (how polar or non – polar the sections of the molecular are, usually exemplified by the log of octanol-water partition coefficient, log P). However, molecules 'look' different when viewed in electrostatic, of lipophile 'space' (Figure 1) [12].

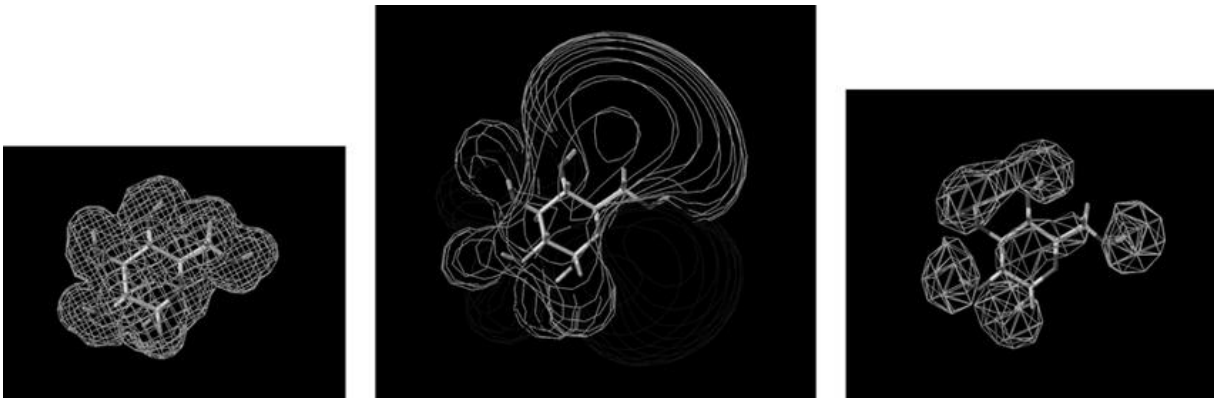


Figure1. A small organic molecule (glucopyranose) viewed in steric (left), electronic (centre) and lipophilic (right) space.

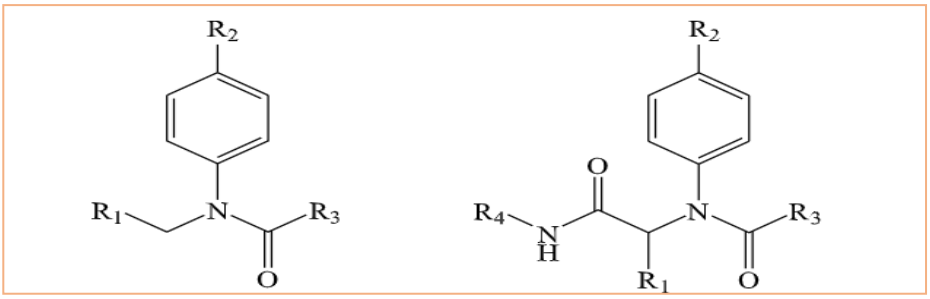
This study aims to build QSAR models, which explain the relationship between anti-MERS-CoV activity and the structure of 43 peptidomimetic derivatives based on physicochemical descriptors using many statistical methods such

as the principal components analysis (PCA), the multiple linear regression (MLR) Y-randomization test and the multiples non-linear regression (MNLR).

2. Materials and methods

2.1. *Data Set*

A series of 43 peptidomimetic derivatives (Figure 2) [13], were studied on a logarithmic scale **[pIC_{50} = - log (IC50)],** as shown in (Table 1). The compounds of this series were drawn using the Chem3D available in the ChemOffice, ACD/ChemSketch and MarvinSketch software as shown in (Table 1). The studied compounds were randomly divided into two set, a training one containing 35 compounds to build a QSAR model, and a test one containing 8 compounds to check the predictive power of the built model. Two definitions of IC_{50} are considered: relative and absolute. The relative IC_{50} is the parameter c in the 4-parameter logistic model and is the concentration corresponding to a response midway between the estimates of the lower and upper plateaus. The absolute IC_{50} is the response corresponding to the 50% control (the mean of the 0% and 100% assay controls) [14].

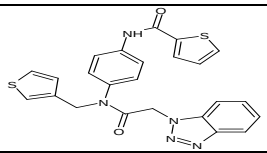
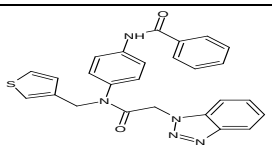
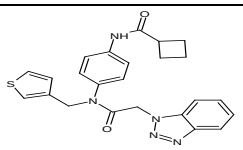
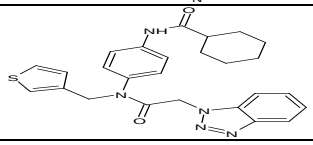
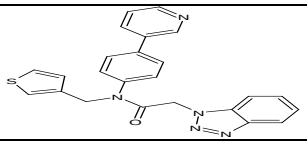
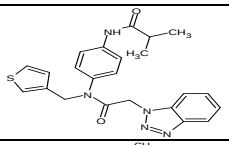
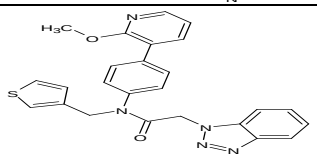
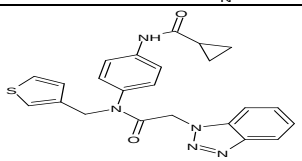
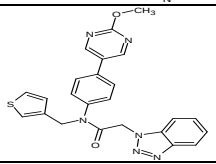
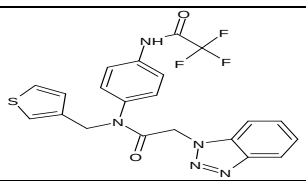
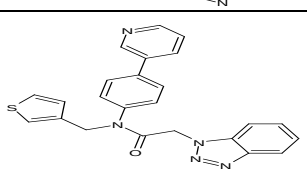
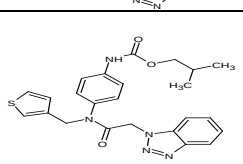


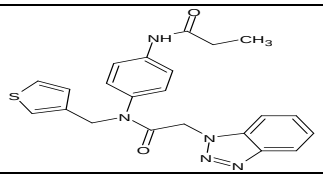
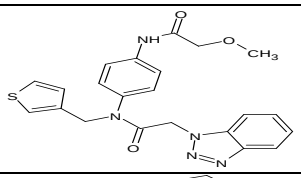
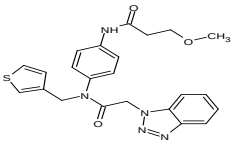
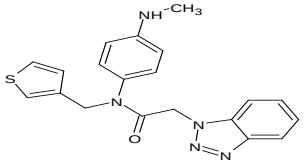
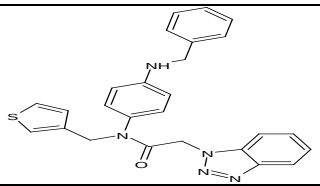
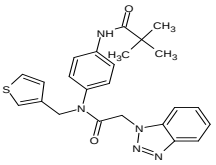
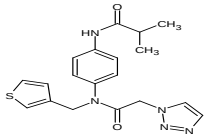
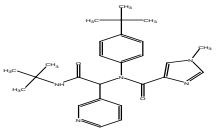
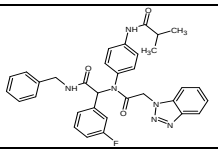
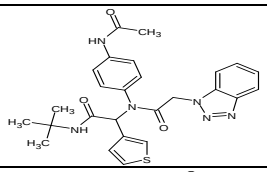
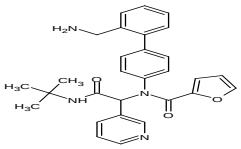
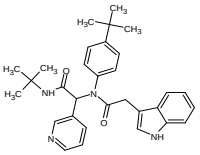
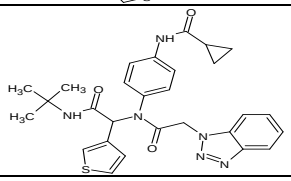
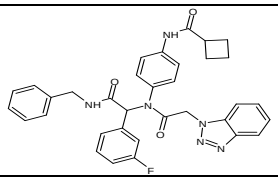
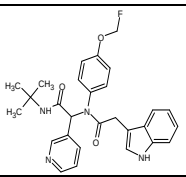
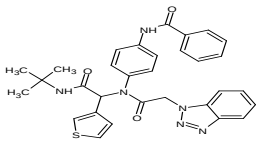
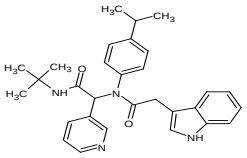
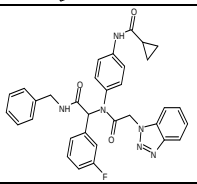
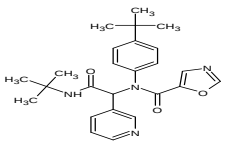
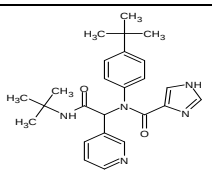
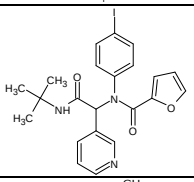
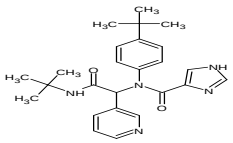
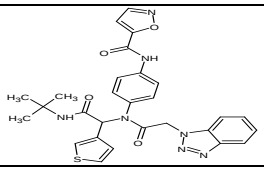
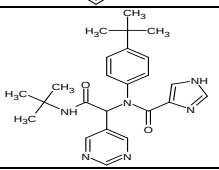
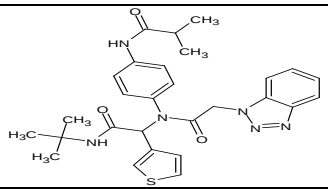
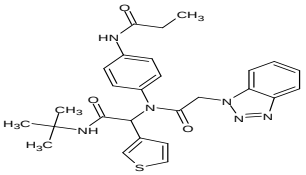
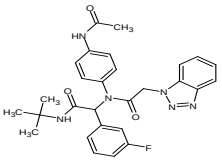
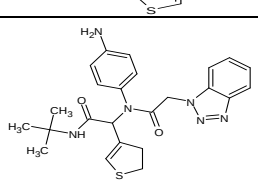
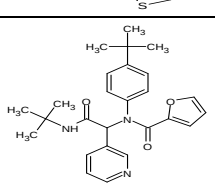
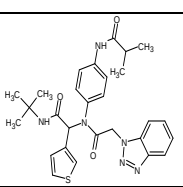
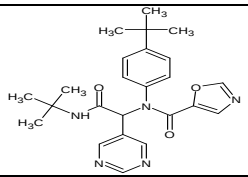
Peptidomimetic Backbone 1 (Class A) **Molecules 1-19**

Peptidomimetic Backbone 2 (Class B) **Molecules 20-43**

Figure2. The general structure of the two peptidomimetic inhibitors.

Table 1: ***Chemical structures and activity experimental of 43 peptidomimetic compounds.***

N°	Structure	pIC_{50} (M)	N°	Structure	pIC_{50} (M)	N°	Structure	pIC_{50} (M)
1		6.48	**2**		6.39	**3**		5.92
4		5.92	**5**		5.82	**6**		5.80
7		5.77	**8**		5.77	**9**		5.70
10		5.66	**11**		5.62	**12**		5.55

13		5.51	**14**		5.51	**15**		5.43
16		5.32	**17**		5.28	**18**		5.06
19		4.80	**20**		5.89	**21**		5.82
22		5.74	**23**		5.66	**24**		5.66
25		5.57	**26**		5.47	**27**		5.41
28		5.38	**29**		5.16	**30**		5.15
31		5.07	**32**		5.03	**33**		5.02
34		4.95	**35**		4.83	**36**		4.81
37		4.76	**38**		4.74	**39**		4.73
40		4.66	**41**		4.45	**42**		4.28
43		4.56						

IC_{50} = Half maximal inhibitory concentration; M = mol/L

2.2. Calculation of descriptors

Molecular descriptors play a fundamental role in QSAR study, they formally are the numerical representation of a molecular structure. Molecular descriptors can be classified using different criteria. Among them, there are two main categories, experimental and theoretical descriptors [15]. The purpose of molecular descriptors is to numerically describe the structure of a molecule in order to establish a relationship between the descriptors and biological activity [16]. Today, several programs have been developed to calculate various molecular descriptors, such as MOPAC [17], CDK [18], MOLGEN [19], DRAGON …etc [20]. Therefore, this study relied on a group of programs, namely: ChemSketch [21], MarvinSketch [22] and ChemOffice [23]. ChemSketch was used to calculate several chemical recipes, as this program is not only limited to the calculation of constitutional descriptors, but also enables us to identify several other types of descriptors (such as physicochemical, quantum, Geometric, …etc.). The descriptors calculated by this program are as follows: Formula weight (FW), percentage of C atoms (C%), percentage of H atoms (H%), percentage of N atoms (N%), percentage of O atoms (O%) , percentage of S atoms (S%), Molar Refractivity (MR), Molar volume (MV), Parachor (Pa), Index of Refraction (IR), Surface Tension (ST), Density (D), Polarizability (Po), RDBE, Nominal Mass (NM). All these descriptions are shown in (Table S1). As for MarvinSketch program, it has a wide range of functions to enable the fast and accurate drawing of chemical compounds, reactions, Markush structures and query molecules. Furthermore, MarvinSketch has built-in structure and valence checkers to provide guidance, and integrated property calculators to pull live results-upon your request [22]. With this program, the following descriptors were calculated: Atom Count (AC), Partition coefficient octanol-water (Log P), HLB, Deriding Energy (DE), MMFF94 Energy (ME), Van Der Waals Volume (ME), Polar Surface Area 2D (PSA), Van Der Waals Surface Area 3D (VDWSA), Refractivity (R). These descriptors are presented in (Table S2). ChemOffice Professional is a scientifically intelligent, integrated suite of personal productivity tools that enables scientists and researchers to capture, store, retrieve, analyze and share data and information on compounds, reactions, materials and their properties [24]. By ChemOffice, the following recipes were calculated: Henry's Law Constant (HLC), Mol Refractivity (MR), Exact Mass (EM), Mass (M), Mol Weight (MW), Number of H-bond Acceptors (NHA), Number of H-bond Donors (NHD), Mol Refractivity (MR), Partition Coefficient (PC), Balaban Index (BI), Cluster Count (CC), Molecular Topological Index (MTI), Num Rotatable Bonds (NRB), Radius (R), Shape Attribute (SA), Sum of Degrees (SD), Sum of Valence Degrees (SVD), Topological Diameter (TD), Total Connectivity (TC), Total Valence Connectivity (TVC), Wiener Index (WI). As shown in (Table S3).

2.3. Statistical analysis

XLSTAT version 2020 [25], was used to accomplish both principal component analysis (PCA), multiple linear regression (MLR) and multiple non-linear regression (MNLR), with the aim of deriving a mathematical relationship between inhibitory activity and a set of molecular descriptors. In another meaning, MLR and MNLR depend on the assumption that there is a relationship that combines both the inhibitory activity and a series of molecular descriptors.

3. Validation of the QSAR Model

The predictive power of the QSAR model is verified by the results obtained by two basic principles: internal verification and external verification.

3.1. Internal validation

In the leave-one-out (LOO) method of the cross-validation (CV), the process of removing a molecule, and creating and validating the model against the individual molecules is performed for the entire training set. Once complete, the mean is taken of all the Q^2 values and reported. The data utilized in obtaining Q^2 is an augmented training set of the

compounds (data points) used to determine R^2 [26]. The Cross-validated squared correlation coefficient, R^2_{CV} (or Q^2) was calculated using the following expression:

$$Q^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{PRESS}{TSS}$$

Where yi represent the observed activity of the training set compounds, $\hat{y}_i$ is the predicted activity of the training set compounds, and $\bar{y}$ corresponds to the mean observed activity of the training set compounds and n number of objects. PRESS is the predictive residual sum of the squares.

Also calculated was the adjusted R^2 (R_{adj}^2) which is a modification of R^2 that adjust the number of explanatory terms in a model. Unlike R^2 in which addition of descriptors to the developed QSAR model increases its value, the value of R_{adj}^2 increases only if the new term improves the model more than what would be expected by chance [27,28]. Hence, R_{adj}^2 overcomes the draw backs associated with the value of R^2 that was calculated using the expression:

$$R^2_{adj} = \frac{(n-1)R^2 - p}{n-p-1}$$

Where p is the number of predictor variables used in the model development.

3.2. Y-Randomization test

Y-Randomization is a tool used in validation of QSAR models, whereby the performance of the original model in data description (R^2) is compared to that of models built for permuted (randomly shuffled) response, based on the original descriptor pool and the original model building procedure [29]. In other words, in this test, random MLR models are generated by randomly shuffling the dependent variable while keeping the independent variables as it is. The new QSAR models are expected to have significantly low R^2 and Q^2 values for several trials, which confirm that the developed QSAR models are robust. Another parameter, CR_P^2 is also calculated which should be more then 0.5 for passing this test [30].

$$CR_P^2 = R \times (R^2 - (Average\ R_r)^2)^{1/2}$$

where, Average Rr = average 'R' of random models

3.3. External Validation

Several authors have suggested that the only way to estimate the true predictive power of a QSAR model is to compare the predicted and observed activities of an (sufficiently large) external test set of compounds that were not used in the model development [31-35]. The final validation of the predictive power of the developed model is carried out by external verification. The quality of the QSAR model is mostly determined by its ability to make predictions for things not included in the training groups. To calculate the predictive R^2 (R_{pred}^2) value is based on the following expression:

$$R^2_{pred} = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2}$$

Here, $y_i$and $\bar{y}_i$ are the observed and predicted activity data for the test set compounds while $\bar{y}$ indicates the mean observed activity of the training set molecules. The models were subjected to external validation criteria according to the proposed test by Golbraikh and Tropsha [36-37], The model QSAR can be considered predictive, if it meets the following conditions:

$R^2_{pred} > 0{,}6$ (1)

$Q^2 > 0{,}5$ (2)

$\frac{r^2 - r_0^2}{r^2} < 0{,}1$; $\frac{r^2 - r_0'^2}{r^2} < 0{,}1$ (3)

$0{,}85 \leq K \leq 1{,}15 \quad Or \quad 0{,}85 \leq K' \leq 1{,}15$ (4)

Where $Q^2$is the square of the LOO cross-validation (CV) coefficient, r^2 is the regression coefficient for the test set exclusively, r_0^2 observed activities; $r_0'^2$predicted activities; slopes k and k' of the regression lines through the origin.

The external predictive power of QSAR models is verified using the test set, by observing metrics the r_m^2 measurements, as shown below [38].

$$r_m^2 = r^2 \left(1 - \sqrt{r^2 - r_0^2}\right) \quad (5)$$

$$r_m'^2 = r^2 \left(1 - \sqrt{r^2 - r_0'^2}\right) \quad (6)$$

Where r^2 is squared correlation coefficient between observed and predicted values, r_0^2 is squared correlation coefficient between observed and predicted values with intercept value set to zero. In addition, the compounds with and without the intercept respectively and $r_0'^2$ bears the same meaning but uses reversed axes.

4. Results and Discussion

4. 1. Principal Components Analysis (PCA)

During this phase, 45 descriptors were used, which is the sum of the number of recipes that were previously calculated by ChemSketch, MarvinSketch and ChemOffice. The matrix of correlation coefficients derived from the original data set is a starting point of PCA. This matrix is studied and analyzed to extract important information from a multivariate spreadsheet and to express this information as a set of few new variables called the main components. Therefore, PCA is a very important stage before linear regression is accomplished, as PCA reduces the number of descriptors while ensuring a minimum level of information loss. The descriptors that remain after the PCA for the rest of this study are: FW ; C% ; H% ; N% ; O% ; S% ; IR ; ST ; D ;RDBE ; LogP ; HLB ;DE ; ME ; PSA ; R; HLC ; NHA ; NHD ; MR; PC ; NRB ; SVD ; TD ; TC and TVC.

4. 2. Multiple Linear Regression (MLR)

Multiple Linear regression (MLR) is one of the most basic modeling methods known to QSAR, when twenty-six descriptors remaining after completion of PCA are built by building an input file for gradual selection based on MLR analysis. A comparison of the statistical parameters R^2, R_{test}^2, R_{adj}^2, MCE, RMCE and F-value was used to obtain, the best model whose equation is generally written as follows:

$$y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots\ldots\ldots + \beta_K X_K$$

Where β_0 is the model constant, X_1; ...; X_K are molecular descriptors with their corresponding coefficients β_1; ...; β_K (for molecular descriptors 1 through k). These coefficients can be obtained through the use of estimators like least-squares method which minimizes the sum of squared residuals. The best equation was obtained with eight variables using the multiple linear regression (MLR) method, and write this equation as follows:

pIC_{50} = -10.37 – 0.02 **FW** – 0.22 **N%** + 12.38 **IR** + 0.48 **LogP** + 4.05 10^{-02} **PSA** +0.29 **NHA** – 7.25 10^{-02} **TC** – 6.36 10^{-02} **TVC**

R^2 = 0.777; R_{test}^2 = 0.655; R_{adj}^2 = 0.709; **MSE = 0.079; RMSE = 0.281**;

F = 11.332; Pr <0.0001.

Where **R^2** is the coefficient of determination; **R_{test}^2** is the coefficient of determination of the external test; **R_{adj}^2** is the adjusted coefficient of determination; **MSE** is the means of the square errors of the model; **RMSE** is root mean square error, **F** the coefficient of Fischer (Fisher statistics F) and **Pr** is the significance level .This equation consists of 8 descriptors: Formula weight (FW), percentage of N atoms (N%), Index of Refraction (IR), Partition coefficient octanol-water (Log P), Polar Surface Area 2D (PSA), Number of H-bond Acceptors (NHA), Total Connectivity (TC), Total Valence Connectivity (TVC). High values for each of the coefficient of determination, the coefficient of determination of the external test, and the adjusted coefficient of determination, that exceed 0.6, in addition to the low value of mean squared errors and root mean square error, all confirm that the model has reliable predictive power. On

the other hand, The Fisher F-test associated with the Pr value indicates that we will take less than 0.01% of the risk assuming the null hypothesis is false and the regression equation is statistically significant. Based on these results, we conclude that the model contains a large amount of information. For this MLR model the expected activity values pIC_{50} are calculated along with the observed activity values, as shown in Table 2. The correlations between the expected activities are illustrated and observed in Figure 3. Therefore, only the proposed descriptors are used in the MLR equation to perform multiples non-linear regression (MNLR).

Table 2: ***Experimental and predicted inhibitory activities (pIC50) and residual values, according to different methods (MLR, MNLR) obtained from the model.***

	N°	pIC_{50} exp	MLR		MNLR	
			pIC_{50} pred	Resid	pIC_{50} pred	Resid
Training set	**1**	6.480	6.148	0.332	6.148	0.332
	2	6.390	6.024	0.366	5.914	0.476
	3	5.920	5.751	0.169	5.684	0.236
	4	5.920	5.917	0.003	5.929	-0.009
	6	5.800	5.816	-0.016	5.814	-0.014
	7	5.770	5.851	-0.081	6.040	-0.270
	8	5.720	6.090	-0.370	5.957	-0.237
	9	5.700	5.981	-0.281	6.027	-0.327
	10	5.660	5.578	0.082	5.518	0.142
	12	5.550	5.482	0.068	5.558	-0.008
	13	5.510	5.750	-0.240	5.761	-0.251
	15	5.430	5.286	0.144	5.301	0.129
	16	5.320	5.183	0.137	5.186	0.134
	17	5.280	5.173	0.107	5.369	-0.089
	18	5.060	5.211	-0.151	5.191	-0.131
	19	4.800	4.853	-0.053	4.806	-0.006
	21	5.820	5.507	0.313	5.469	0.351
	22	5.740	5.191	0.549	5.277	0.463
	23	5.660	5.572	0.088	5.746	-0.086
	24	5.660	6.010	-0.350	5.848	-0.188
	25	5.570	5.246	0.324	5.450	0.120
	26	5.470	5.587	-0.117	5.442	0.028
	28	5.380	5.310	0.070	5.430	-0.050
	29	5.160	5.235	-0.075	5.209	-0.049
	30	5.150	5.417	-0.267	5.460	-0.310
	31	5.070	5.024	0.046	5.013	0.057
	32	5.030	4.718	0.312	4.768	0.262
	33	5.020	4.985	0.035	4.953	0.067
	34	4.950	4.842	0.108	4.906	0.044
	36	4.810	4.650	0.160	4.640	0.170
	37	4.760	4.996	-0.236	4.898	-0.138
	38	4.740	5.054	-0.314	5.084	-0.344
	40	4.660	4.723	-0.063	4.622	0.038
	42	4.280	4.502	-0.222	4.380	-0.100
	43	4.250	4.828	-0.578	4.692	-0.442
Test set	**5**	5.820	4.855	0.965	4.776	1.044
	11	5.620	4.974	0.646	4.862	0.758
	14	5.510	5.219	0.291	5.113	0.397
	20	5.890	4.183	1.707	4.260	1.630
	27	5.410	5.392	0.018	5.645	-0.235
	35	4.830	5.838	-1.008	6.113	-1.283

39	4.730	5.275	-0.545	5.334	-0.604
41	4.450	5.326	-0.876	5.285	-0.835

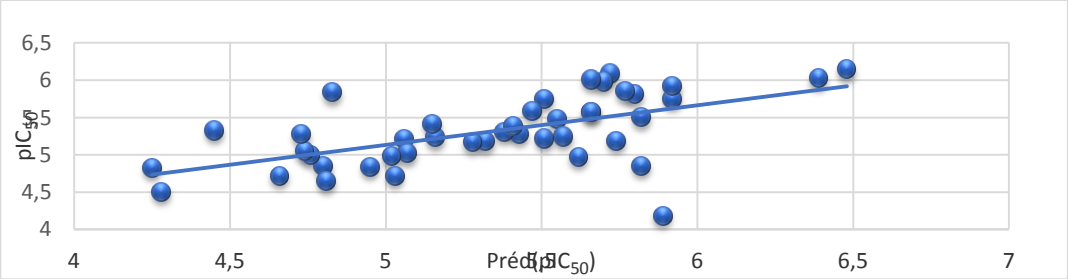


Figure 3. Graphical representation of predicted and observed inhibitory activities (pIC50) values calculated by MLR.

4.3. Y-Randomization

Y-Randomization test was applied to verify the validity and robustness of the model. Where it is assumed that he obtained the basic LOO (Leave-one-out) statistics of the random models worse than the original model, and thus it is confirmed that the model obtained is not by chance. The various results are shown in (Table 3) below. Where R is the correlation coefficient for the Y-randomization and $CR_P^2$is the coefficient of Y- randomization.

Table 3: ***Various values obtained after testing of Y-Randomization.***

Random Models Parameters	**Value**	**Model Original**	**Value**
R	0.467	R	0.882
R^2	0.231	R^2	0.777
Q^2	-0.436	Q^2	0.611
CR_P^2	0.659 > 0.6		

4. 2. Multiples Non-Linear Regression (MNLR)

The MNLR model method has also been used to improve the relationship between structure and activity. The MNLR shows the nonlinear relationship between input and output data, where we applied this technique to the data matrix composed of descriptors proposed by the best MLR model corresponding to 43 molecules. To have a preprogrammed function in the XLSTAT, it is written as:

$$y = a_0 + \sum_{i=1}^{n} a_i x_i + b_i {x_i}^2$$

Where y is the dependent variable, x_i are the independent variables, i is the number of explanatory variables, a_0 is the constant of the model equation, a_i and b_i are the descriptor coefficients in the model equation.

The resulting equation is as follows:

pIC_{50} = -60.94 – 7.50 10^{-03} **FW** – 0.21 **N%** + 69.41 **IR** + 1.08 **LogP** + 1.02 10^{-02} **PSA** + 0.80 **NHA** – 2.59 10^{-02} **TC** + 7.38 10^{-02} **TVC** – 9.49 10^{-06} **FW^2** – 3.91 10^{-04} **$N\%^2$** – 17.43 **IR^2** – 7.91 10^{-02} **$(LogP)^2$** + 1.55 10^{-04} **PSA^2** – 4.05 10^{-02} **NHA^2** – 3.91 10^{-03} **TC^2** – 1.49 10^{-02} **TVC^2**

R^2 = 0.813; MSE = 0.096; RMSE = 0.309

For a clearer understanding of the MNLR model we compared it with the MLR model. various values are shown in (Table 4). Through these results, we can easily notice the increase in the value of, coefficient of determination R^2, mean squared error MSE and root mean square error RMSE, of the nonlinear model MNLR, compared to the linear model MLR [39]. Based on all these results obtained by MLR, Y-Randomization and MNLR, we can conclude that

this model has a good predictive power. QSAR investigations receive more attention of several researchers to compute and explain the Covid activity [40-42].

Table 4: ***Some of the main statistical indicators for MLR model and MNLR model.***

Model	R^2	MSE	RMSE
MLR	0.777	0.079	0.281
MNLR	0.813	0.096	0.309

5. Conclusion

In this study, 45 descriptors were computed and selected for 43 peptidomimetic compounds using a ChemSketch, MarvinSketch and ChemOffice programs. These descriptors are subject to a statistical method PCA with the aim of analyzing and visualizing this set of data, and thus grouping these data into only a few new variables called basic components, these new variables correspond to a linear set of the original variables. The MLR technique was used in order to find linear equations that combine many of the descriptors and the pIC_{50} values. Indeed, this equation was found consisting of eight descriptors between the QSAR random models, and the results confirm that, where the model was validated internally (Q^2 = 0.611), and the same was verified externally (R^2_{test} = 0.655), In addition to the values of the main statistical terms for this model such as R^2 = 0.777, R^2_{adj} = 0.709 and CR_P^2 = 0.658. MNLR is another modeling method that has been used in this study, as it contributes to building the QSAR model by raising and stimulating the statistical factors of the equation of the model MLR. Using this method, we managed to get good results for both R^2 = 0.813, MSE = 0.096 and RMSE = 0.309. Finally, based on the obtained results the proposed model has sufficient capacity to predict the anti-MERS-CoV activity.

References

[1] Z.A. Memish, A.I. Zumla, R.F. Al-Hakeem, A.A. AlRabeeah, G.M. Stephens Family cluster of Middle East respiratory syndrome coronavirus infections. *New Engl. J. Med.*, 368 (2013) 2487-2494

[2]A.M. Zaki, S. vanBoheemen, T.M. Bestebroer, A.D. Osterhaus, R.A. Fouchier Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med*, 367 (2012) 1814-1820

[3] A. Bermingham, M.A. Chand, C.S. Brown, E. Aarons, C. Tong, C. Langrish, et al.Severe respiratory illness caused by a novel coronavirus, in a patient transferred to the United Kingdom from the Middle East, September 2012. *Euro Surveill*, 17 (2012) 20290

[4] S. Sridhar, P. Brouqui, P. Parola, P. GautretImported cases of Middle East respiratory syndrome: an update. *Travel Med Infect Dis*, 13 (2015) 106-109

[5] R. Bialek, D. Allen, F. Alvarado-Ramy, R. Arthur, A. Balajee, D. Bell, et al.First confirmed cases of Middle East respiratory syndrome coronavirus (MERS-CoV) infection in the United States, updated information on the epidemiology of MERS-CoV infection, and guidance for the public, clinicians, and public health authorities—May 2014. *MMWR Morb Mortal Wkly Rep*, 63 (2014) 431-436

[6] World Health organization WHO, Middle East respiratory syndrome coronavirus (MERS-CoV), (2019). www.who.int/news-room/fact-sheets/detail/middle-east-respiratory-syndrome-coronavirus-(mers-cov)

[7] B. L. Haagmans, et al. Middle East respiratory syndrome coronavirus in dromedary camels: an outbreak investigation. *Lancet Infect. Dis.* 14 (2014) 140–145

[8] J. S. Sabir, et al. Co-circulation of three camel coronavirus species and recombination of MERS-CoVs in Saudi Arabia. *Science* 351 (2016) 81–84.

[9] A. Abuhammad, R.A. Al-Aqtash, B.J. Anson, A.D. Mesecar, M.O. Taha, Computational modeling of the bat HKU4 coronavirus 3CLpro inhibitors as a tool for the development of antivirals against the emerging Middle East respiratory syndrome (MERS) coronavirus. *J. Mol. Recognit.* 30 (2017) e2644.
[10] C. Nantasenamat, C. Isarankura-Na-Ayudhya, T. Naenna & V. Prachayasittikul, A Practical Overview of Quantitative Structure-Activity Relationship, *J. Excli.* 8 (2009) 74-88. http://dx.doi.org/10.17877/DE290R-690
[11] Khan, A. U. Descriptors and their selection methods in QSAR analysis: paradigm for drug design. *Drug Discovery Today*. 21(8) (2016) 1291–302.
[12] D.A. Winker. The role of quantitative structure-activity relationships (QSAR) in biomolecular discovery, *J. Henry Stewart Publications*, 73-86 (2002).
[13] S.E. St. John, S. Tomar, S.R. Stauffer, A.D. Mesecar, Targeting zoonotic viruses: Structure-based inhibition of the 3C-like protease from bat coronavirus HKU4—The likely reservoir host to the human coronavirus that causes Middle East Respiratory Syndrome (MERS), *Bioorganic & Medicinal Chemistry*, 23 (2015) 6036-6048.
[14] J. L. Sebaugh, Guidelines for accurate EC50/IC50 estimation, *OFC.J.PSI*, 10, 2 (2011) 128-134.
[15] A. Mauri, V. Consonni, R. Todeschini, Molecular Descriptors, *Handbook of Computational Chemistry* pp 1-29 (2016).
[16] M. Dehmer, K. Varmuza, and D. Bonchev, Eds., 2012 "Statistical Modelling of Molecular Descriptors in QSAR/QSPR: DEHMER: MOL. DESCRIPTOR O-BK. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA".
[17] J. Baker. An Algorithm for the Location of Transition States. *J. Comp. Chem.*, 7 (1986) 385.
[18] Willighagen, Egon L, Mayfield, John W, Alvarsson, Jonathan, Berg, Arvid, Carlsson, Lars, Jeliazkova, Nina, Kuhn, Stefan, Pluskal, Tomás, Rojas-Chertó, Miquel, Spjuth, Ola, Torrance, Gilleain, Evelo, Chris T, Guha, Rajarshi, Steinbeck, Christoph: The Chemistry Development Kit (CDK) v2.0: atom typing, depiction, molecular formulas, and substructure searching. In: Journal of cheminformatics, 9 (1), pp. 33, 2017.
[19] A. Kerber, MOLGEN, a Generator for Structural Formulas. MATCH Commun. *Math. Comput. Chem.* 80 (2018) 733-744
[20] R. Todeschini, Milano Chemometrics and QSPR Group, Talete srl, Milan, Italy. (http://www.talete.mi.it/products/dragon_description.htm).
[21] ChemSketch on ACD/Labs 2019.2.0, http://www.acdlabs.com
[22] Marvin 20.9.0 software, Chem Axon, 2020, http://www.chemaxon.com
[23] ChemOffice Or ChemDraw 16.0, http:// www.chemistry1science.blogspot.com
[24] PerkinElmer – ChemOffice, ADDITIVE Soft -und Hardware für Technik und wissenschaft GmbH, https://www.additive-net.de/en/software/produkte/perkinelmer/chemoffice
[25] XLSTAT software 2020, http://www.xlstat.com
[26] R. Veerasamy, H. Rajak, A. Jain, S.Sivadasan , C. Varghese, R.Agrawal. Validation of QSAR Models - Strategies and Importance. *International Journal of Drug Design and Discovery*, 2 (3) (2011) 511-519
[27] Muhammad U, Uzairu A, Arthur DE, Review on: quantitative structure activity relationship (QSAR) modeling, *J. Anal. Pharm. Res.* 7(2) (2018) 240–242.
[28] K Roy, I Mitra, S Kar, et al. Comparative Studies on some metrics for external validation of QSAR model. *J. Chem. Inf. Model.* 52(2) (2012) 396–408.
[29] C. Rücker, G. Rücker, M. Meringer, y-Randomization and Its Variants in QSPR/QSAR, *J. Chem. Inf. Model.* 47, 6 (2007) 2345–2357
[30] Pravin Ambure (ambure.pharmait@gmail.com) of Drug Theoretics & Cheminformatics (DTC) Laboratory, Jadavpur University (2013).

[31] H. Kubinyi, F.A. Hamprecht, T. J. Mietzner, Three-dimensional quantitative similarity-activity relationships (3D QSiAR) from SEAL similarity matrices, *Med. Chem.* 41 (1998) 2553-2564. https://doi.org/10.1021/jm970732a
[32] R. Guha, P.C. Jurs, Determining the Validity of a QSAR Model − A Classification Approach, *J. Chem. Inf. Model.* 45 (2005) 65-73. https://doi.org/10.1021/ci0497511
[33] E. Novellino, C. Fattorusso, G. Greco, Use of comparative molecular field analysis and cluster analysis in series design, *Pharm. Acta Helv.* 70 (1995) 149-154.
[34] U. J. Norinder, Single and domain mode variable selection in 3D QSAR applications, *Chemom.* 10 (1996) 95-105. https://doi.org/10.1002/(SICI)1099-128X(199603)10:2%3C95::AID-CEM407%3E3.0.CO;2-M
[35] N.S. Zefirov, V.A. Palyulin, QSAR for Boiling Points of "Small" Sulfides. Are the "High-Quality Structure-Property-Activity Regressions" the Real High Quality QSAR Models? *J. Chem. Inf. Comput. Sci.* 41 (2001) 1022-1027. https://doi.org/10.1021/ci0001637
[36] A. Golbraikh, A. Tropsha, Beware of q2! *J. Mol. Graphics Mod.* 20 (2002) 269-276. https://doi.org/10.1016/s1093-3263(01)00123-1
[37] A. Tropsha, Best practices for QSAR model development, validation, and exploitation. *Mol. Inform.* 29 (2010) 476–488.
[38] S. Chtita, A. Aouidate, A. Belhassan, A. Ousaa, A. Idrissi Taourati, B. Elidrissi, M. Ghamali, M. Bouachrine and T. Lakhlifib, QSAR study of N-substituted oseltamivir derivatives as potent avian influenza virus H5N1 inhibitors using quantum chemical descriptorsand statistical methods, *J. Chem*, 44 (2020) 1747.
[39] I. Hammoudan, S. Matchi, M. Bakhouch, S. Belaidi, et S. Chtita, « QSAR Modelling of Peptidomimetic Derivatives towards HKU4-CoV 3CLpro Inhibitors against MERS-CoV », *Chemistry*, 3, no 1 (2021) 391-401, doi: 10.3390/chemistry3010029.
[40] H. Saputra, C. Nur Albar, D. Sulistiyo Soegoto, Bibliometric Analysis of Computational Chemistry Research and Its Correlation with Covid-19 Pandemic, *Mor J. Chem.*, 10, No 1 (2022) 37-49, https://doi.org/10.48317/IMIST.PRSM/morjchem-v10i1.31723