ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Computational Study of Quinolone's Antibacterial Activity Using QSAR Approach.

Mohamed Aissaoui[a]*, Sofiane Ali Rachedi[a], Ahmed Mostefa Mokhnane[b], Habiba Louzim[b], and Abdelhafid Djerourou[a]

[a]Organic Synthesis and Biocatalysis Laboratory, Faculty of Science, Chemistry Department, P.B.: 12, 23000, Annaba, Algeria.

[b]Chemistry Therapeutic Group, Organic Synthesis and Biocatalysis Laboratory, Faculty of Medicine, Pharmacy Department, P.B.: 12, 23000, Annaba, Algeria.

ABSTRACT

The urinary tract infection is a common cause of consultation and hospitalization in Urology. Quinolones are used in second line treatment of these infections. The resistance developed by the bacteria against these molecules is evolving. To evaluate this resistance, we have processed the files of 1506 patients hospitalized at Urology Service of the University Hospital center of Annaba, Algeria. In an attempt to achieve a better understanding of the parameters controlling the bacterial properties against gram negative *E. coli* and linking chemical activities with molecular structures and compositions, a QSAR study was undertaken. In the present study, 85 quinolone derivatives (Data set), were evaluated as antibacterial agents, expressed by the minimal inhibition concentration (MIC) of these compounds against *E. coli*. A linear QSAR model was developed using Multiple Linear Regression technique, while Genetic algorithm was adopted for selecting the most appropriate descriptors. The predictive ability of the Model proposed was experimentally validated using the external validation set and the Y-randomization technique. The QSAR model developed in this study will be helpful to design new potent quinolone derivatives.

Keywords: QSAR, Quinolone Derivatives, Minimum Inhibitory Concentration (MIC), *E. coli*, Genetic Algorithm, Multiple Linear Regression.

Corresponding author

ISSN: 0975-8585

INTRODUCTION

It's been over seventy (70) years that antibiotics have been one of the greatest medical advances making them indispensable. [1, 2] Nowadays, bacterial resistance to antibiotics is both a present reality and a threat for the future.[3] Its involvement in morbidity and mortality makes the therapeutic choices more complex, affecting seriously the quality of medical care. It reduces our therapeutic range without giving the pharmaceutical researches the time to respond to the new needs with new more active products. Therefore, control of bacterial resistance to antibiotics is a major health issue for our country. Although actions have been taken for several years in both the monitoring of resistance and the prevention of the transmission of resistant bacteria in health facilities in order to promote better use of antibiotics. This work started with an observation in Urology Service of the University Hospital Center at Annaba (Algeria) of bacterial resistance to quinolones.

This study was carried out from 1 January 2011 to 31 December 2013, including 1506 patients. The microbiological exam of urine cultures showed that about 26% of patients are positive as shown in (Figure 1). *E. coli* is the most incriminated germ being responsible for about 3⁄4 of the positive results of urinalysis, other bacteria were also isolated, as shown in Figure 1.

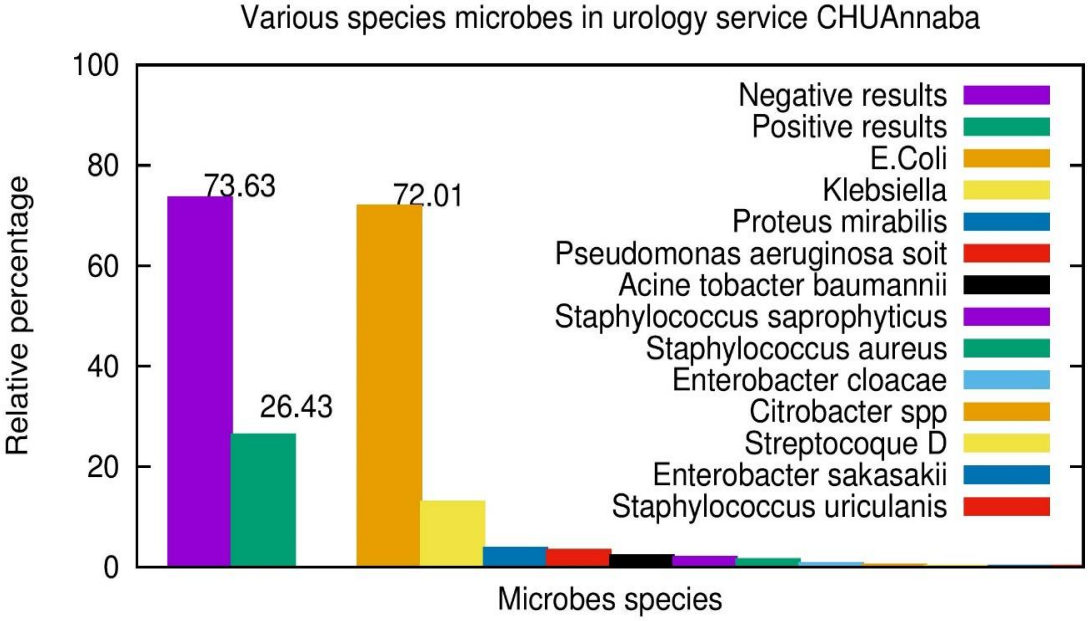

Figure 1 The percentage of microbiological exam of urine.

As for the resistance to quinolones a worrying development was noted for Ofloxacin®: passing from 23% in 2011 to 30% in 2013. Another quinolone: The Levofloxacin® seems more active with a resistance rate of 10 % in 2011, slightly increased to 13 % in 2013. [3, 4] Hence, conclusions were drawn; the bacterial resistance developed in this service is increasing, mainly for *E. coli.* [5, 6] Quinolones of different generations, however, remain effective in these urinary tract infections without confirming how long this effectiveness is affirmed, Ofloxacin® and Levofloxacin® are stereoisomers with different resistance levels, suggest that even small structural variations affect the antibiotic activity of quinolones.[7, 8].

Based on these data, and remaining in the same vein, the QSAR study we conducted is interested to quinolones following the establishment of a relationship between their structure and their activities.[9] The concept of structure-activity relationship (QSAR) used for several years in the designing new drugs and predicting their therapeutics activities even before their synthesis. This prediction although not total, has the advantage of saving a part of the initial testing; thus, the structure activity relationship has become an integral part in the modeling of biological activity. From this finding, recent works involving toxicologists, chemists and computer scientist have shown that compliance with a number of simple rules is possible to obtain comparable results to those obtained during the development of therapeutics molecules.

ISSN: 0975-8585

This article presents at first the attempts to set out the basic rules that must be respected in this approach. Then, an application example is given with the limits of the prediction obtained when using a QSAR. This example treats the minimum inhibitory concentration (MIC) of a set of quinolone molecules against the most isolated germ in the study previously mentioned *E. coli.*

MATERIALS AND METHODS

QSAR Study

Select data set and biological data

In an attempt to achieve a better understanding of the parameters controlling the antibacterial properties against the gram negative (*E. coli*), a QSAR study was realised [10–14].

The present QSAR study is limited to the properties exhibited against *E. coli* due to the incredible increase in the resistance of these bacteria.[15, 16] The data set contains 85 quinolone derivatives with an antibacterial activity against *E. coli*. Their molecular structure and activity in vitro are listed in figures 2 and table 2 respectively.

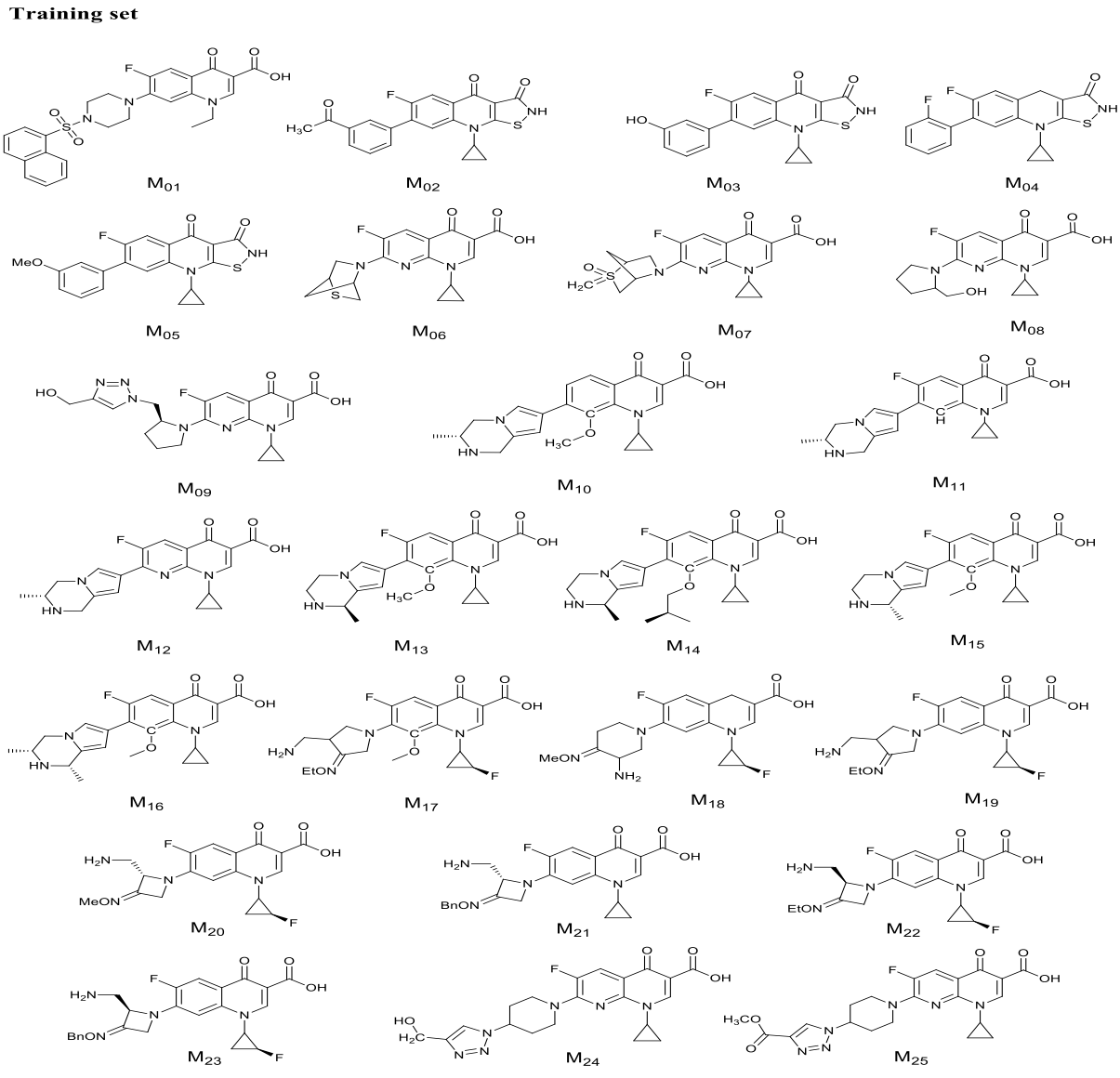

Figure 2: Training Set and Test Set Molecules

ISSN: 0975-8585

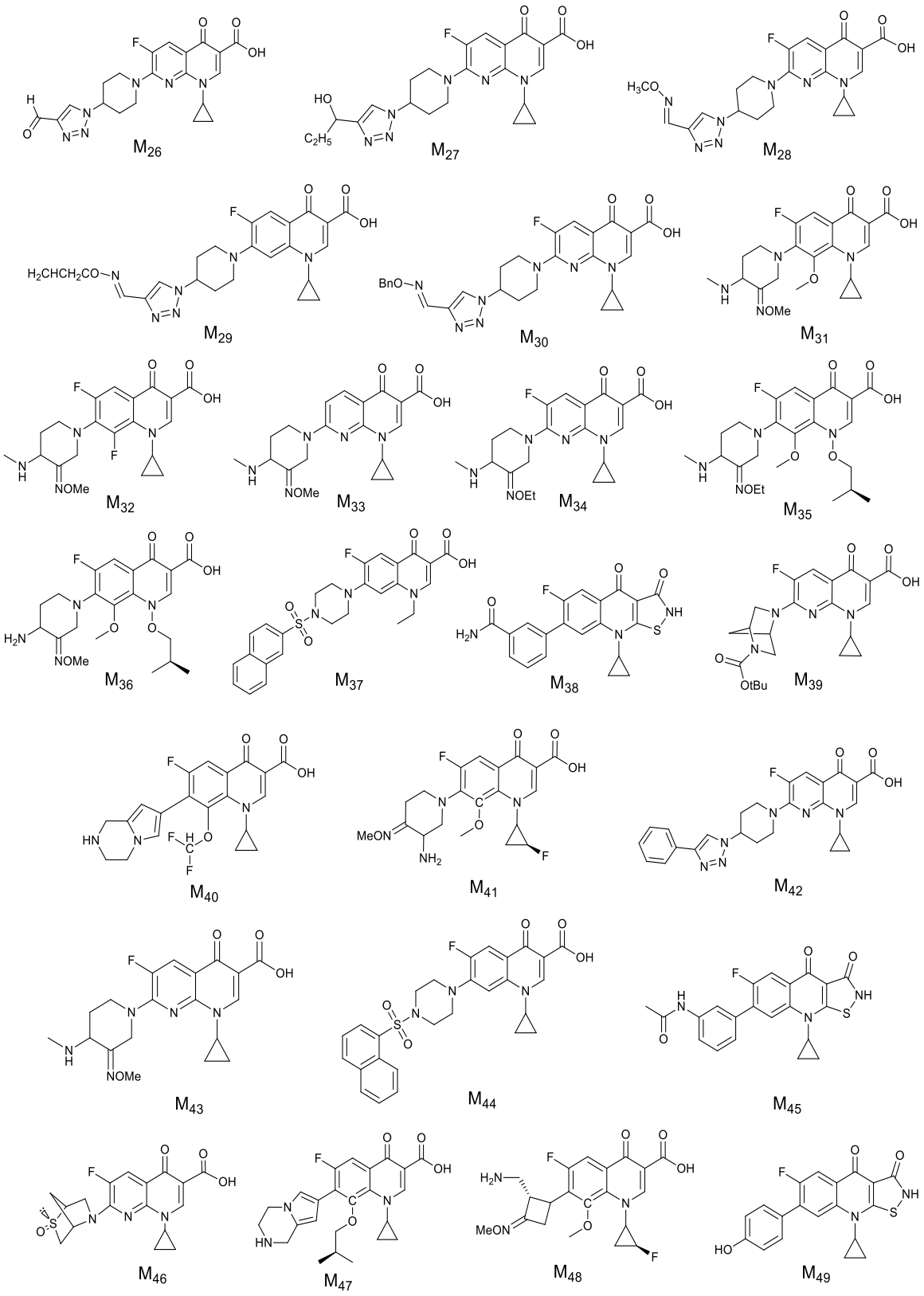

Figure 2: Training Set and Test Set Molecules continued.

ISSN: 0975-8585

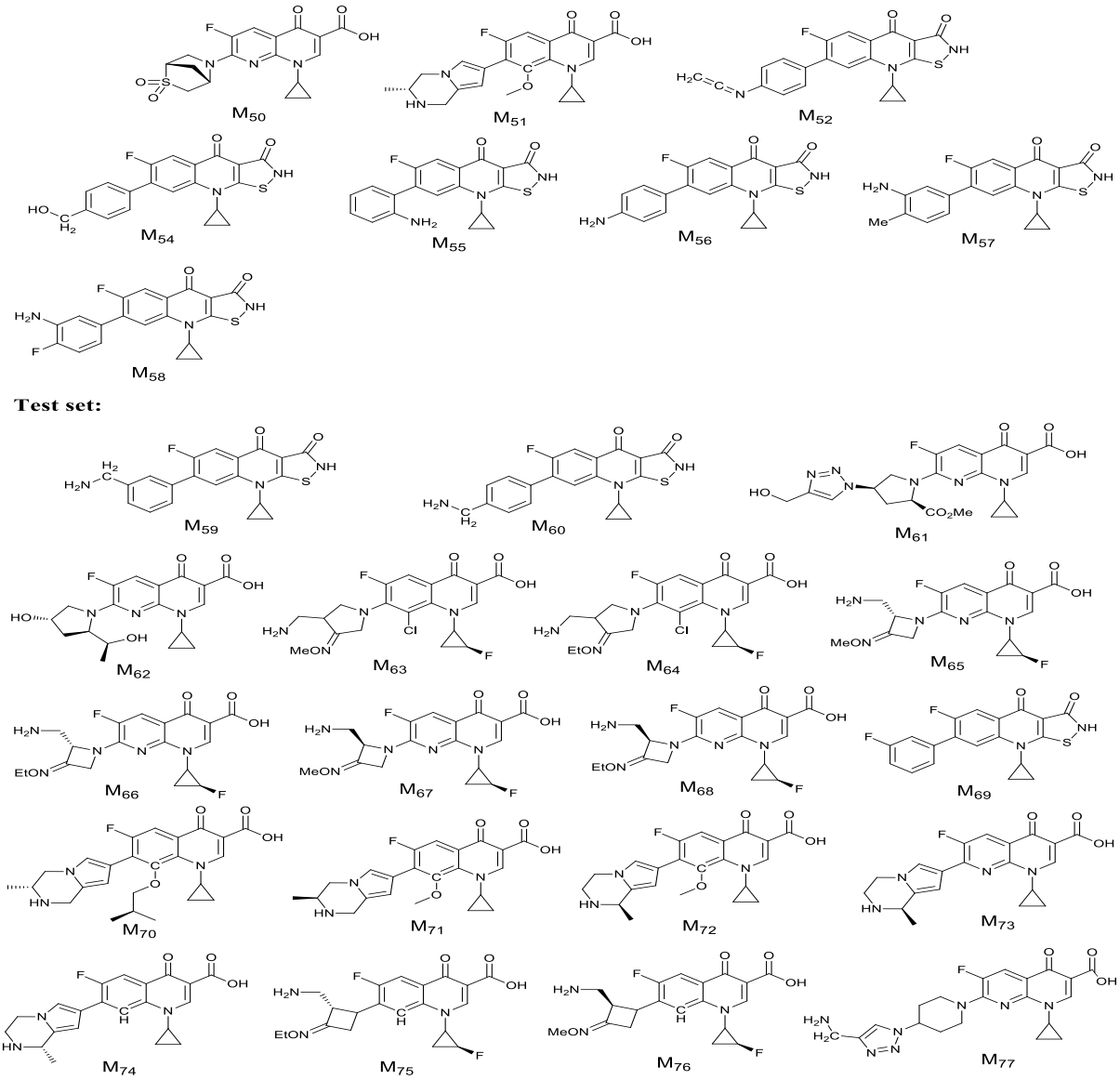

Figure 2: Training Set and Test Set Molecules continued.

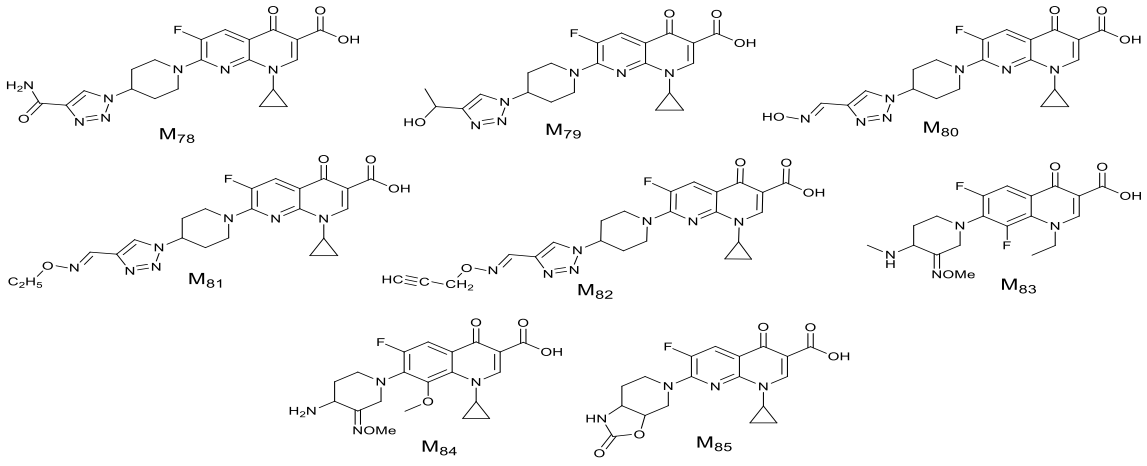

Figure 2: Training Set and Test Set Molecules continued.

All molecules were selected from different papers in literature,[17–22] The reason of this selection is to obtain a structural diversity for a better understanding of the observed bio-properties and to determine the most important structural parameters controlling the antibacterial activity. The biological activity modeled in this study was converted into -Log MIC where MIC is the minimal inhibition concentration against the organism shown.[23].

To reduce the number of descriptors. The objective selection is used to select a subset of descriptors that are best in encoding the property of quinolone derivatives using only the independent variables (descriptors). The reduction of variables was performed by the Dragon software using objective selection to remove descriptors which had identical or zero values for greater than 90% of the data set, one descriptors in any pair of descriptors whose pairwise correlation exceeding 0.9 was also eliminated.[32, 33]

Calculation

Geometric optimization

The structures of 85 quinolones derivatives was sketched using *ChemBioDraw* ultra 12.0 software [24] and was exported to *Gaussview* 5.0.9 and *Gaussian* 09 software [25]. The three dimensional structures of all molecules were generated, and their geometries were optimized preliminary with semi-empirical method AM1 [26], then using the quantum chemical DFT method(Density Functional Theory) included in Gaussian 09 software.[27, 28] After optimization the (x,y,z) atomic coordinates of the minimal energy conformation for each molecule can be determined.

Molecular descriptors

The optimized structures were transferred into *Dragon* 5.0 software[29] (developed by Milano chemometrics and QSPR group) to calculate 4485 descriptors in 29 di*ff*erent blocs.[30, 31]

QSAR model development

Objective feature selection of descriptors

Initially, over 4485 descriptors were calculated for each molecule in the data set, a feature selection was used to reduce the number of descriptors. The objective selection is used to select a subset of descriptors that are best in encoding the property of quinolone derivatives using only the independent variables (descriptors). The reduction of variables was performed by the Dragon software using objective selection to remove descriptors which had identical or zero values for greater than 90% of the data set, one descriptors in any pair of descriptors whose pairwise correlation exceeding 0.9 was also eliminated.[32, 33]

Subjective feature selection of descriptors

After elimination with objective selection, we obtained only 2255 descriptors, the number was reduced an- other time by subjective feature selection,[34–36] In this stage, descriptors selection is based on MLR analysis in combination with GA.[37, 38] The general aim behind the MLR procedure is to build a multiple linear regression model from a set of independent variables (descriptors) by entering and removing predictor in a stepwise manner until there is no justifiable reason to enter or remove any more (until no significant variables varies). All molecular descriptors are used to build QSAR model by MLR analysis, as implemented in XLStat software. [39] In this study, we used the stepping criteria: $\alpha = 0.15$ to enter and $\alpha = 0.15$ to remove in XLStat software. [39]. GA is a search heuristic method that belongs to the larger class of the *Evolution Algorithms* (EA) which generate solutions to optimization problems using techniques inspired by natural evolution, such as inheritance, crossover, mutation, and selection. [40] The GA simulation conditions were: 10000 iterations, equation length change from 3 to 8 descriptors per model for map the set of quinolone derivatives to activity, number of crossovers was 5000, smoothness factor was 1, mutation probability was 0.5, and initial number of equations generated was 500. GA was executed multiple times (10-15 times) till an optimized solution is found.

A major decision to develop successive QSAR model is knowing when to stop adding descriptors to the model during GA-MLR procedure. A simple technique to control the model expansion is called "Breaking Point" in the improvement of the statistical quality of the model, by analyzing the plot of the number of descriptors involved in the models obtained (in our study from 3 to 8 descriptors) versus the squared correlation of coefficient R^2 and of cross validation correlation coefficient Q^2_{cv} . The model corresponding to the breaking point is considered as the optimum model.[41].

Model validation and evaluation

Cross validation test

The first technique applied for the validation of the proposed model was based on leave- one-out algorithm. The "Leave one out" of the proposed model was based on leave-one-out algorithm. The "Leave one out" (Loo) cross validation was used to evaluate the predictivity of the final QSAR equation. This step is necessary, because a high value of the square of correlation coefficient R^2 indicates the best fit of the data, but does not contain information about the ability to predict the dependent variable or no included data in the training set. From the Loo cross-validation procedure, the square of cross-validation coefficient Q^2 is obtained, which is used as a criterion to evaluate both the robustness and the predictive ability of the generated model. According to Tropsha et *al*, a QSAR model is considered predictive if the following conditions are satisfied.[42, 43]

$$Q^2_{cv-Loo} > 0.5 \qquad (1)$$

$$R^2 > 0.6 \qquad (2)$$

$$R^2 - \frac{R_0^2}{R^2} < 0.1 \qquad (3)$$

$$0.85 < K < 1.15 \qquad (4)$$

The Mathematical definition of R^2, R_0^2, R'^2_0, K and K' are based on regression for the test set of the observed activity against predicted activity and vice versa. According to Roy et *al*[44], the di*ff*erence between values of R_0^2 and R'^2_0 must be studied and given importance. They suggested the following modified R^2 from:

$$R_m = R^2\left(1 - \left|\sqrt[2]{R^2 - R_0^2}\right|\right) \qquad (5)$$

Y-Randomization test for the MLR model

The model was further validated by applying the Y-randomization test,[45] in which random MLR models are generated by randomly shuffling. The dependent variables while keeping the independent variables as they are. The new QSAR models are expected to have significant low R^2 and Q^2 values for several trials, which confirm that the developed QSAR models are robust. Another parameter, CR^2 is also calculated, which should be more than 0.5 for passing this test with:

$$CR_p^2 = R\sqrt{(R^2 - (average\,R_r)^2)} \qquad (6)$$

Where: average *Rr* = average *R* of random models.

Euclidean based applicability domain (AD)

Applicability domain (AD) is the physicochemical structure or biological space, knowledge or information on which the training set of the model has been developed. The resulting model can be reliably applicable for only those compounds which are inside this domain.[46, 47] It is based on mean distance scores calculated by the Euclidean distance norms. A simple way of defining the range of a QSAR model is according

to the leverage of a compound. The leverage *hi* of a compound can be used to measure outlyingness in the X-space and measures its influence on the model. Observations that are outliers in the X-space are known as high leverage points to distinguish them from observations that are outliers in the response variable (those with large standardized residuals). In matrix terms, the leverage of a compound in the original variable space is defined as:

$$h_i = X_i^T (X^T X)^{-1}) X_i) \ldots (i = 1, .. n) \qquad (7)$$

Where *Xi* is the descriptor row-vector of the query compound, and X the matrix of k model descriptor values for *n* training set compounds. The superscript T refers to the transpose of the matrix/vector. The warning leverage *h** is defined as follows:

$$h^* = 3h = \frac{3(\sum_i hi)}{n} = \frac{3p'}{n} \ldots (i = 1, N, n) \qquad (8)$$

Where *n* is the number of training compounds and *p'* is the number of model adjustable parameters.[40].

The applicability domain (AD) of QSAR model is defined from the Williams plot. The plot of leverage values versus standardized residuals was used to give a graphical detection of both the response outliers (Y outliers) and the structurally influential compounds (X outliers). In this plot, the two horizontal lines indicate the limit of normal values for Y outliers (i.e., samples with standardized residuals greater than 3.0 standard deviation units, 3.0 s); the vertical straight lines indicate the limits of normal values for X outliers (i.e., samples with leverage values greater than the threshold value, h > h∗). For a sample in the external test set, whose leverage value is greater than *h**. It's prediction is the result of a substantial extrapolation of the model.[48] Conversely, when the leverage value of a compound is lower than the critical value, the probability of accordance between predicted and experimental values is as high as that for the compounds in the training set.

RESULT AND DISCUSSION

During model selection or formulation process, the task is to design, or more often to select an algorithm to mathematically describe the relations of descriptors and biological activity.

Selection of relevantdescriptors

In order to select the predominant descriptors that will affect the MIC of these compounds, correlation analysis was performed with statistical software XLStat software. [39], taking every calculated descriptor as independent variable and -log MIC as a dependent variable, using the GA technique as method of selection. To select the most important descriptors and the optimal number in the model, we performed consecutively several equations with different number of descriptors from 3 to 8 variables. To define the optimum model, we must pass through an important step, that ensures the over parameterization of the model and prevents to some extent the chance correlations between descriptors. As mentioned above, this procedure is based on the break point rule, the change in the slope as shown in the plot of R^2 and Q^2_{cv-Loo} versus number of descriptors added [49]. It was indicated that the maximum improvement of R^2 and Q^2_{cv-Loo} was at five descriptors as shown in figure3.

Table 1 Correlation coefficients and VIF values among the five variables.

VARIABLES (DESCRIPTORS)	DP	RDF055U	RCI	MOR17I	CATS2D 04 AL	VIF
DP	1.000	0.384	0.213	-0.302	0.588	1.72
RDF055U		1.000	0.341	-0.354	0.256	1.36
RCI			1.000	-0.396	0.157	1.28
MOR17I				1.000	-0.273	1.37
CATS2D 04 AL					1.000	1.58

ISSN: 0975-8585

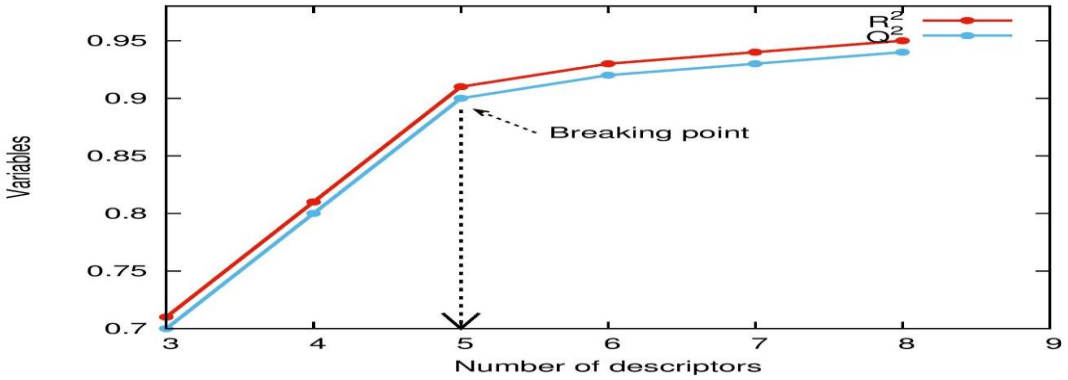

Figure 3 The optimum number of descriptors for the MLR-GA model.

The multi-collinearity test between the 5 descriptors in the equation was investigated by the VIF parameter, where VIF is the variable inflation factor. The values for these parameters are given in table1.

All descriptors have variance of VIF values between 1.28 and 1.72, indicating that the collinearity is not a problem for these data, so the obtained model have statistical significance, and the descriptors were found to be reasonably orthogonal. The same picture emerges if we examine the correlation coefficient matrix using XLStat software. As a result of the collinearity, low correlation coefficients (table 1) were detected among the value of the independent variables of the model represented by equation 9.

QSAR modeling

The aforementioned stepwise MLR technique coupled with GA, was used to establish the QSAR model. The structures of 85 quinolone derivatives was drawn (figure7, 8) and their antibacterial activity against *E. coli* were listed in table 2. Before starting the actual calculation for the model, we divided the data set into training and test set, using the Euclidean Based Kennard-Stone algorithm. The training set and the test set consisted 58 and 17 compounds respectively. On the MLR-GA method, we perform several equations from 3 to 8 variables. As noted above, the powerful QSAR model contains 5 descriptors shown with their high statistical parameters:

$$MIC = 5.723 - 8.60\mathbf{Dp} - 0.03385\mathbf{RDF055u} - 6.808\mathbf{RCI} - 1.022\mathbf{Mor17i} - 0.2049\mathbf{CATS2D\ 04\ AL} \quad (9)$$

$$N = 58;\ R^2 = 0.91;\ S = 0.2743;\ F = 120.41;\ P < 10^{-4};\ Q^2_{cv-Loo} = 0.9031;\ S_{cv-Loo} = 0.2886$$

Where
- **N** is the number of compounds (training set),
- **S** is the standard deviation of the regression,
- R^2 is the squared correlation coefficient,
- **F** is the Fischer ratio,
- S_{cv-Loo} is the cross validation standard deviation,
- Q^2_{cv-Loo} is deviation and square of the correlation coefficient.

Table 2 The experimental and predicted value of –log MIC with the selected descriptors.

No.	-Log MIC Exp (mol/ml)	-Log MIC Pred (mol/ml)	Dp	RDF055u	RCI	Mor17i	CATS2D 04 AL	Leverage (hi)
				Training Set				
1	0.60	0.74	0.380	28.397	-0.091	-1.066	12.000	0.25067135
2	3.50	3.55	0.345	10.716	-0.086	-1.566	5.000	0.25962742
3	3.79	4.10	0.314	9.572	-0.113	-1.228	3.000	0.06509144
4	3.47	3.58	0.295	8.198	-0.019	-1.134	3.000	0.09549577

5	3.48	3.43	0.322	9.096	-0.093	-1.157	5.000	0.05317436
6	3.46	3.34	0.325	11.623	-0.014	-1.901	6.000	0.27592132
7	2.60	2.71	0.327	17.098	-0.017	-1.863	8.000	0.27394688
8	2.24	1.88	0.310	18.826	0.003	-0.895	7.000	0.16156481
9	1.13	1.24	0.329	28.471	0.023	-1.284	9.000	0.17257736
10	2.63	2.93	0.320	24.247	-0.028	-1.979	7.000	0.07082532
11	3.50	3.24	0.309	12.863	0.058	-1.790	4.000	0.14068250
12	3.18	3.07	0.309	14.710	0.047	-1.811	5.000	0.11444626
13	2.92	2.83	0.342	22.960	0.018	-1.731	4.000	0.05801818
14	2.92	2.98	0.348	38.413	0.036	-2.561	4.000	0.21743082
15	2.32	2.58	0.332	24.797	0.018	-1.469	4.000	0.06486185
16	2.63	2.67	0.344	30.443	-0.016	-1.615	4.000	0.06958903
17	2.05	2.10	0.361	31.690	-0.034	-0.917	3.000	0.20257029
18	3.21	3.54	0.289	22.881	-0.095	-1.025	3.000	0.10438099
19	3.85	3.32	0.321	21.233	-0.066	-1.216	3.000	0.05038511
20	4.69	4.48	0.285	20.024	-0.148	-1.460	3.000	0.13707945
21	2.64	2.98	0.330	31.095	-0.115	-1.768	7.000	0.14241815
22	4.70	4.22	0.284	19.092	-0.135	-1.253	3.000	0.10910278
23	3.27	2.73	0.348	21.720	-0.115	-1.360	7.000	0.06447378
24	2.93	2.83	0.325	15.535	-0.026	-1.256	5.000	0.03874262
25	2.66	2.89	0.368	15.313	-0.054	-1.482	5.000	0.06484982
26	3.23	3.02	0.339	14.457	-0.035	-1.464	5.000	0.03153311
27	2.66	2.40	0.361	16.897	-0.006	-1.911	8.000	0.06316960
28	2.36	2.70	0.377	15.577	-0.031	-1.533	5.000	0.28750275
29	2.08	2.01	0.405	17.715	-0.014	-1.474	6.000	0.15183708
30	1.82	1.66	0.414	19.303	-0.017	-1.846	9.000	0.13406368
31	3.54	3.16	0.365	30.427	-0.070	-1.906	4.000	0.11804362
32	4.15	3.87	0.360	18.795	-0.113	-1.689	3.000	0.13063692
33	2.92	2.97	0.342	19.481	-0.108	-1.923	9.000	0.12224275
34	2.01	2.40	0.383	21.149	-0.055	-1.716	7.000	0.25903428
35	1.47	1.48	0.369	38.565	0.031	-1.847	7.000	0.12157641
36	1.75	1.61	0.348	28.655	0.065	-1.692	7.000	0.10470795
37	0.91	0.87	0.425	27.258	-0.052	-1.190	9.000	0.20343004
38	3.50	3.73	0.296	9.319	-0.100	-1.186	5.000	0.06889501
39	2.05	2.26	0.402	28.763	-0.013	-2.673	9.000	0.18387913
40	3.22	3.20	0.307	21.240	-0.036	-1.990	7.000	0.22526157
41	3.85	3.25	0.325	23.785	-0.061	-1.300	3.000	0.05582371
42	1.17	1.35	0.376	18.490	-0.025	-1.541	11.00	0.12495147
43	3.51	3.14	0.357	19.955	-0.071	-2.075	7.000	0.07228965
44	1.21	1.04	0.375	27.245	-0.067	-1.439	12.00	0.18091071
45	3.51	3.75	0.349	11.730	-0.122	-1.589	5.000	0.08634628
46	2.92	2.92	0.335	16.078	-0.017	-2.103	8.000	0.09257504
47	2.64	2.60	0.328	36.111	0.067	-2.152	4.000	0.17339414
48	3.23	3.77	0.300	30.370	-0.106	-1.519	3.000	0.14764938
49	4.08	4.31	0.306	8.642	-0.101	-1.414	3.000	0.06714957
50	3.50	3.50	0.323	16.047	-0.027	-2.104	6.000	0.07766224
51	2.21	2.58	0.334	22.232	-0.051	-1.613	4.000	0.09129070
52	4.29	3.87	0.295	9.758	-0.083	-1.446	5.000	0.06670504
53	3.82	3.61	0.332	10.493	-0.091	-1.476	5.000	0.04814234
54	3.82	4.05	0.307	9.992	-0.075	-1.383	3.000	0.05245198
55	3.46	3.57	0.310	10.480	-0.091	-0.852	3.000	0.07675033
56	3.78	3.98	0.284	8.658	-0.094	-0.946	3.000	0.08288749

57	3.48	3.64	0.311	9.843	-0.058	-1.122	3.000	0.05990887
58	3.48	3.50	0.304	9.503	-0.101	-0.635	3.000	0.10936951
				Test Set				
59	4.4	4.04	0.324	11.910	-0.102	-1.404	3.000	0.16794792
60	3.82	4.08	0.284	9.576	-0.068	-1.250	3.000	0.22267215
61	0.97	1.17	0.347	28.622	-0.054	-1.181	7.000	0.27527671
62	1.67	2.04	0.357	18.471	-0.064	-0.796	6.000	0.27688470
63	3.54	3.05	0.345	19.473	-0.074	-0.841	2.000	0.27251481
64	2.96	2.85	0.344	22.528	-0.062	-0.820	2.000	0.27840659
65	3.81	3.38	0.287	19.446	-0.036	-1.338	4.000	0.13920913
66	3.51	3.29	0.288	19.427	-0.019	-1.362	4.000	0.15194296
67	3.19	3.17	0.297	21.033	-0.036	-1.269	4.000	0.11544441
68	2.92	2.60	0.325	25.584	-0.019	-1.202	4.000	0.12669077
69	3.77	4.06	0.309	9.101	-0.101	-1.208	3.000	0.1753176
70	2.96	2.56	0.361	38.159	0.069	-2.476	4.000	0.25828249
71	2.95	2.63	0.335	21.783	0.051	-1.661	4.000	0.19939862
72	2.89	3.31	0.342	14.502	0.021	-1.948	4.000	0.24221746
73	2.88	3.29	0.316	16.515	0.008	-1.883	5.000	0.13524025
74	3.18	3.38	0.314	16.040	0.021	-1.825	4.000	0.17348944
75	4.13	4.38	0.286	22.151	-0.135	-1.534	3.000	0.30026026
76	3.89	4.26	0.282	18.356	-0.148	-1.166	3.000	0.30520997
77	2.43	2.69	0.332	20.012	-0.018	-1.375	5.000	0.04948037
78	2.24	2.51	0.342	16.272	-0.014	-1.190	5.000	0.08904159
79	2.64	2.24	0.350	15.798	0.019	-1.598	7.000	0.20680778
80	2.34	2.88	0.330	15.452	-0.024	-1.357	5.000	0.06155180
81	2.06	2.68	0.395	17.214	-0.036	-1.685	5.000	0.28856971
82	2.07	2.16	0.343	17.514	-0.015	-1.090	6.000	0.13723289
83	2.91	3.32	0.356	18.897	-0.124	-1.051	3.000	0.24274444
84	2.52	2.81	0.349	24.323	-0.032	-1.485	4.000	0.09370465
85	2.59	2.55	0.324	13.671	0.026	-1.659	7.000	0.23920959

Table 3 The definition of descriptors based on the model with their bloc kind.

DESCRIPTORS SYMBOL	DESCRIPTION	BLOC
DP	D total accessibility index / weighted by polarizability	WHIM descriptors
RDF055U	Radial Distribution Function - 055 / unweight	RDF descriptors
RCI	ring complexity index	Ring descriptors
MOR17I	signal 17 / weighted by ionization potential	3D-MoRSE descriptors
CATS2D 04 AL	CATS2D ACCEPTOR-LIPOPHILIC AT LAG 04	CATS 2D

The definition of descriptors based on the model used in the present study was designed as fellow in the table 3.

The values of descriptors for all the training set as well as the test set are summarized in table 2. As can be seen from the statistical parameters of the above equation, a considerable improvement was achieved by combining five descriptors, equation 9 can explain the 91% of the experimental variance of the dependent variable MIC. The MLR equation is given by value of regression ($R^2 = 0.91$) which explain around 91%of the variance of data, the model presents the greatest Fisher value ($F = 120.41$) and the lowest standard deviation for the data ($S = 0.2743$) which confirm the former selection and indicate the ability of predictive power of this QSAR model. The present QSAR model developed with MLR coupled with GA as method of selection for variable, confirms that the prediction of the MIC against *E. coli* bacteria is statistically very significant. The validity and predictability of the QSAR model for anti-bacterial activity were cross validated by correlation coefficient value ($R^2 = 0.903$) and standard deviation of cross validation value($S_{cv-Loo} = 0.2886$), obtained by the leave on out (Loo) method. The high value of the determination

coefficient of leave one out cross-validation for the obtained model , and small standard deviation cross validation are proving the predictive power of this approach and the stability of the model. The model was further validated by applying the Y-randomization test, several random shuffles of the Y vector were performed and the low R^2 and Q^2_{cv-Loo} values that were obtained show that the good results in our model are not due to a chance correlation or structural dependency of the training set. The results of the Y-randomization test are presented in table 4.

Table 4 R^2 and Q^2_{cv-Loo} values after several Y-randomizations.

Model	R	R^2	Q^2_{cv-Loo}
Original model	0.959	0.920	0.903
Random 1	0.386	0.149	-0.075
Random 2	0.227	0.052	-0.236
Random 3	0.296	0.088	-0.106
Random 4	0.278	0.077	-0.117
Random 5	0.222	0.049	-0.225
Random 6	0.168	0.028	-0.205
Random 7	0.260	0.067	-0.131
Random 8	0.318	0.101	-0.103
Random 9	0.233	0.054	-0.190
Random 10	0.255	0.065	-0.155

As mentioned in section 1.4. We verified the predictive ability using Golbraikh et *al* criteria (equations 1, 2, 3, 4).

$$Q^2_{cv-Loo} = 0.9031, \ passed \quad \text{(threshold value } Q^2 > 0.5)$$

$$R^2 = 0.719, \quad passed \ (\text{threshold value R2} > 0.6)$$

$$\frac{R^2 - R_0^2}{R^2} < 0.00051, \quad passed \ (\text{threshold value} \ \frac{R^2 - R_0^2}{R^2} < 0.1)$$

$$K = 1.035, \quad passed \ (\text{threshold value } 0.85 < K < 1.15$$

Equation 9 represents our best performing QSAR model, Figure 4 shows the corresponding scatter plot of the estimated versus experimental activity values for quinolone derivatives, this figure indicates that there is a significant correlation between experimental and predicted values of log MIC for training set and test set.

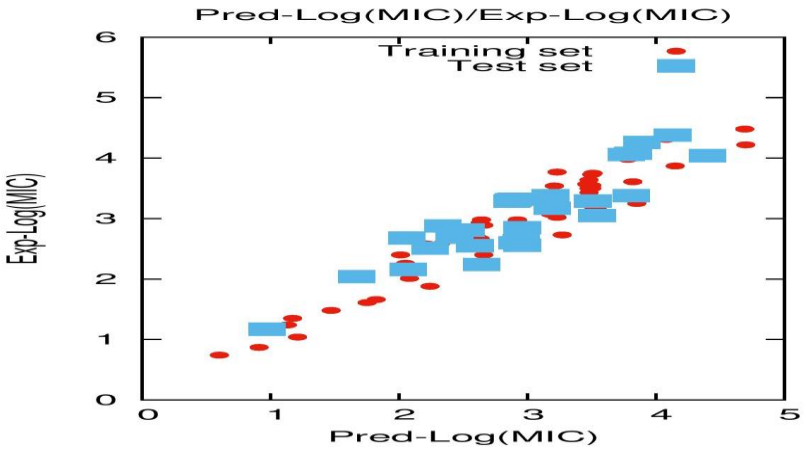

Figure 4 Predicted versus the experimental values of –log MIC For the training set and validation set.

ISSN: 0975-8585

The results illustrate that the MLR technique combined with GA as variable selection procedure are adequate to generate an efficient QSAR model. The residual of the predicted values of MIC against the experimental values for the present model is shown in the Figure 5.

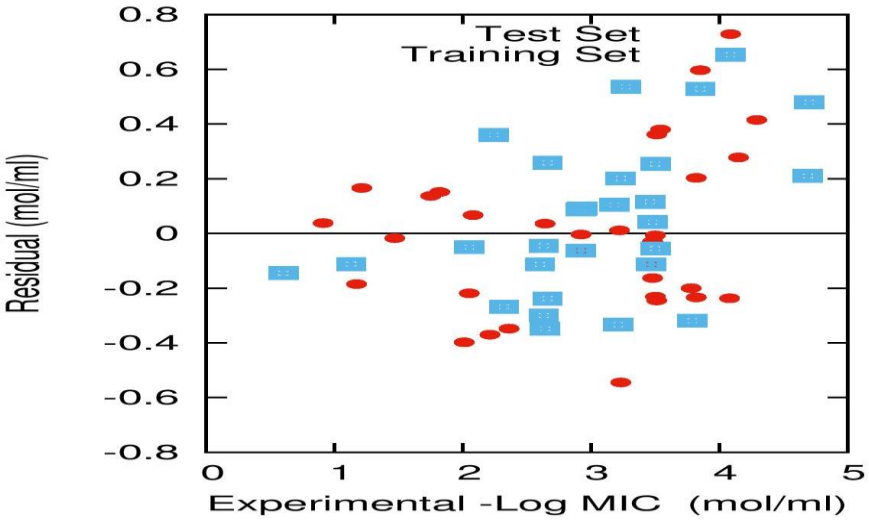

Figure 5 Plot of the residuals for calculated values of –log MIC from the GA-MLR model versus their experimental values for the training and validation sets.

As most of the calculated residuals are distributed on two sides of zero line, a conclusion maybe drawn that there is no systematic error in the development of the present model. To see the importance of each descriptor for the prediction of antibacterial activity, the relative contributions of 5 descriptors to the MLR-GA model were determined and are plotted in Figure 6. Interpreting a QSAR model in terms of the specific contribution of substituents and other molecular features to the modeled activity is always a difficult task. Dp and RCI are the most important variables in this equation (figure 6), because they present the highest contribution in our model. This two descriptors have negative influence on the studied property, the reason why the inhibitory activity of the quinolone derivatives increases.

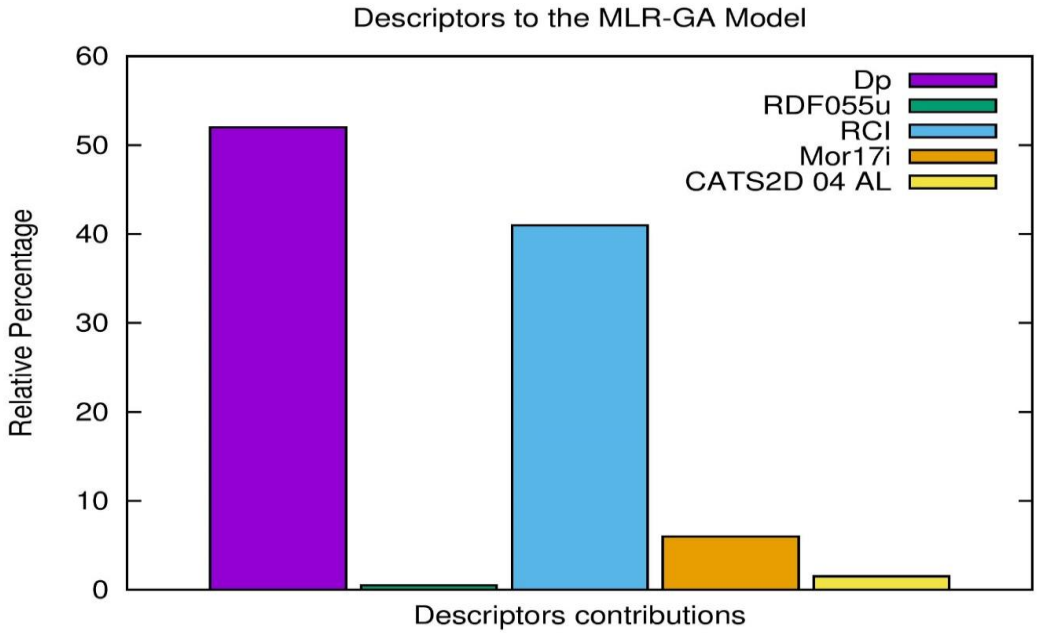

Figure 6 Histogram depicting the relative contributions of the five descriptors to the MLR-GA model.

ISSN: 0975-8585

We made a comparison between compounds N: 26 and 30, 20 and 48. As we can see, the difference between compounds 30 and 26 is in the radical at position 7 of quinolone derivatives, whereas the radical in molecule 26 contains 4 atoms of nitrogen and one of oxygen, and for molecule 30, the radical contains 5 atoms of nitrogen and one of oxygen (increase of polarizability), this difference leads to increase Dp values from 0.339 to 0.414 and consequently decrease the MIC value as shown in figure 7.

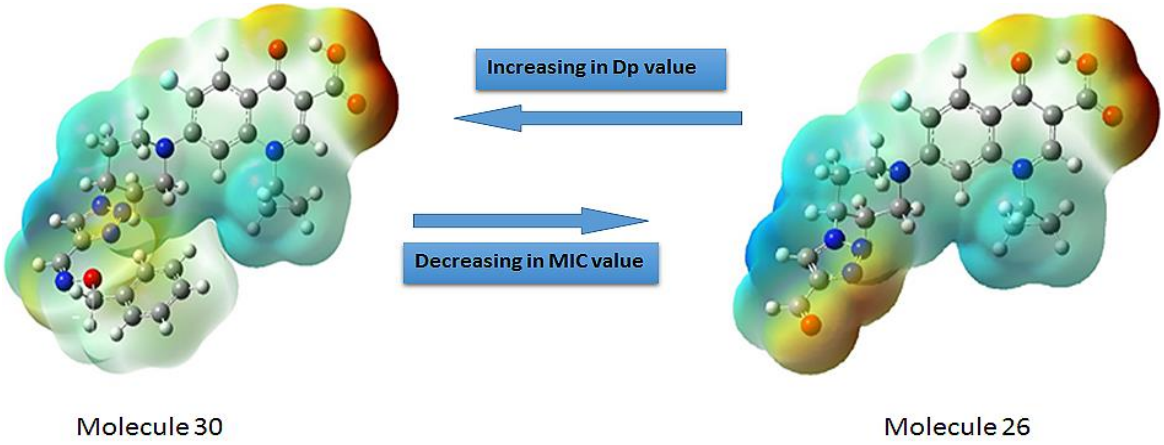

Figure 7 Molecular Electrostatic potentials (MEP) comparison mapped on the electron density surface calculated by the DFT/B3LYP method between compounds N30 and N26.

On the other hand, as shown in equation 9, the inhibition of quinolone derivatives was dependent on the steric parameter (ring complexity index). Indeed, structure analog of 2-benzoylacrylic in quinolone molecule consists a big conjugated system contained 12 of π electrons. Carbon atoms and Fluorine, Nitrogen and Oxygen heteroatoms provide alone pair electrons. Previous study stating that the electron transfer may occur when quinolones interact with DNA or protein.[50] It is clear that which substitution position and what kind of substituent may affect the mechanism of action of pharmacophore (2-benzoylacrylic)with DNA. That's why another comparison was made between compounds 48 and 20; the only difference between these two compounds is the kind of substituent in the pharmacophore of quinolone derivatives at position 8. In spite of this, apparently small difference leads to increase of RCI values from −0.148 to −0.106 and consequently decreases the value of MIC as shown in figure 8.

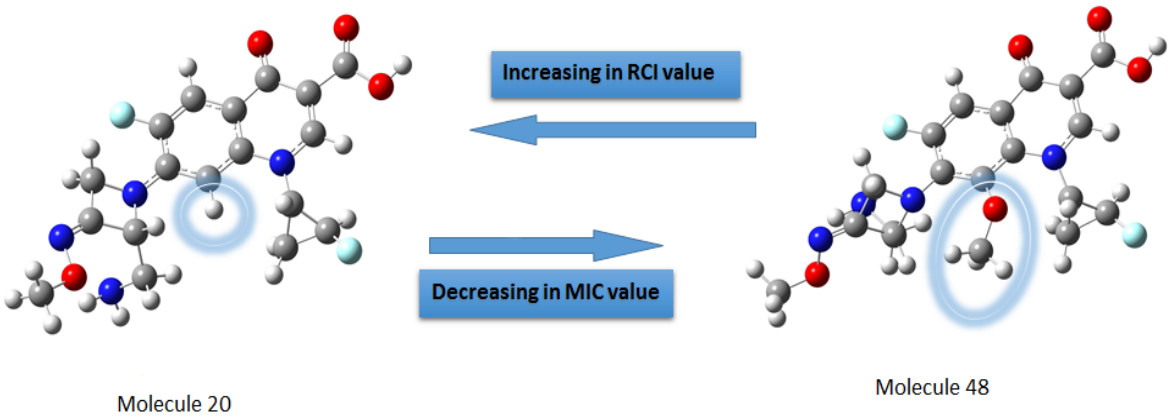

Figure 8 Structure comparison between compounds N20 and N48.

Applicability domain

The prime overall goal of QSAR research is to develop models that provide accurate predictions for as many chemical structures as possible in the universe, particularly for those that have not been tested or for which reliable experimental data is still not available. To this end, however, QSAR model must always be

verified for their applicability with regard to chemical domain, in order to produce predicted data that can be considered reliable only for too structurally dissimilar chemicals. The applicability domain of the model was analyzed using a Williams plot (Figure 9), where the vertical line is the critical leverage value (h∗=0, 31), and the horizontal lines are the cut off value for Y space. From this plot, the applicability domain is established inside a squared area within 3 standard deviations and a leverage threshold *h∗* leverage= 0.31. For making predictions: training data must be considered reliable only for those compounds that fall within this AD on which the model was constructed. It can be seen from Figure 9 that the majority of compounds in the data set are inside this area. Apparently, no influential or outlier data was detected. Therefore, the model presented by Equation 9 displays the best statistical parameters, good prediction, and applicability.

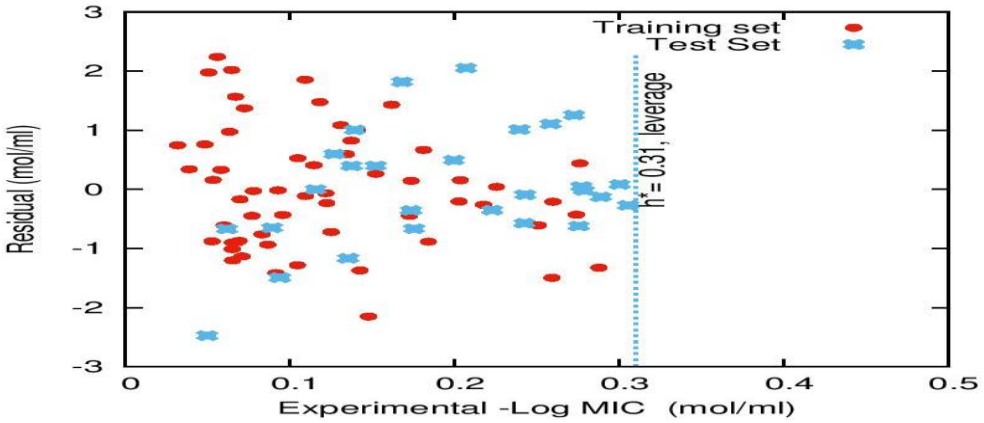

Figure 9 Projection of the training set and the validation set in the Williams plot.

1. CONCLUSION

The aim of the present work was to develop a QSAR study and to predict the minimal inhibition concentration of quinolone derivatives to gram negative (*E. coli*). Minimal inhibition concentration for a set of 85 quinolone derivatives was modeled with success by Multiple Linear Regression analysis, using Genetic Algorithms as variable selection method. The five selected descriptors showed that the polarizability properties and the structure of the molecule play a main role in the inhibition of quinolone derivatives. The proposed model has good stability, robustness and predictivity when verified by internal validation (cross validation by Leave One Out and Y randomization) and external validation. The chemical applicability domain of the studied model served as a valuable tool to filter out dissimilar and outlier compounds.

ACKNOWLEDGEMENTS

This work was supported financially by The General Directorate for Scientific Research and Technological Development (DG-RSDT), Algerian Ministry of Scientific Research, Organic Synthesis and Biocatalysis Laboratory (FNR 2000).

REFERENCES

[1] Alexander Fleming. The discovery of penicillin. *British Medi- cal Bulletin*, 2(1):4–5, 1944.

[2] AW Bauer, WMM Kirby, J C&turck Sherris, and Marvin Turck. Antibiotic susceptibility testing by a standardized single disk method. *American journal of clinical pathology*, 45(4):493, 1966.

[3] JACKSON E Fowler Jr and THOMAS A Stamey. Studies of in- troital colonization in women with recurrent urinary infections. vii. the role of bacterial adherence. *The Journal of urology*, 117 (4):472–476, 1977.

[4] G Ka¨llenius, SB Svenson, H Hultberg, R Mo¨llby, I Helin, B Cedergren, and J Winberg. Occurrence of p-fimbriated es- cherichia coli in urinary tract infections. *The Lancet*, 318(8260): 1369–1372, 1981.

[5] Jochen Kuhlmann, Axel Dalho*ff*, and H-J Zeiler. *Quinolone an- tibacterials*, volume 127. Springer Science & Business Media, 2012.

[6] Hubert Ph Endtz, Gijs J Ruijs, Bert van Klingeren, Wim H Jansen, Tanny van der Reyden, and R Peter Mouton. Quinolone resistance in campylobacter isolated from man and poultry following the introduction of fluoroquinolones in veterinary medicine. *Journal of antimicrobial chemotherapy*, 27(2):199– 208, 1991.

[7] Lester A Mitscher. Bacterial topoisomerase inhibitors: quinolone and pyridone antibacterial agents. *Chemical reviews*, 105(2):559–592, 2005.

[8] Munekazu Yamakuchi, Masanori Nakata, Ko-ichi Kawahara, Isao Kitajima, and Ikuro Maruyama. New quinolones, ofloxacin and levofloxacin, inhibit telomerase activity in transitional cell carcinoma cell lines. *Cancer letters*, 119(2):213–219, 1997.

[9] Corwin Hansch. Quantitative structure-activity relationships in drug design. *Drug design*, 1:271–342, 1971.

[10] Mohammed I El-Gamal, Said M Bayomi, Saadia M El-Ashry, Shehta A Said, A-M Alaa, and Naglaa I Abdel-Aziz. Synthe- sis and anti-inflammatory activity of novel (substituted) ben- zylidene acetone oxime ether derivatives: molecular modeling study. *European journal of medicinal chemistry*, 45(4):1403– 1414, 2010.

[11] [A-M Alaa, Kamal EH ElTahir, and Yousif A Asiri. Synthesis, anti-inflammatory activity and cox-1/cox-2 inhibition of novel substituted cyclic imides. part 1: Molecular docking study. *Eu- ropean journal of medicinal chemistry*, 46(5):1648–1655, 2011.

[12] A-A Magda, Naglaa I Abdel-Aziz, A-M Alaa, Adel S El-Azab, Yousif A Asiri, and Kamal EH ElTahir. Design, synthesis, and biological evaluation of substituted hydrazone and pyrazole derivatives as selective cox-2 inhibitors: molecular docking study. *Bioorganic* & *medicinal chemistry*, 19(11):3416–3424, 2011.

[13] Adel S El-Azab, Mohamed A Al-Omar, A-M Alaa, Naglaa I Abdel-Aziz, A-A Magda, Abdulaziz M Aleisa, Mohamed M Sayed-Ahmed, and Sami G Abdel-Hamide. Design, synthesis and biological evaluation of novel quinazoline derivatives as po- tential antitumor agents: molecular docking study. *European journal of medicinal chemistry*, 45(9):4188–4198, 2010.

[14] Fatmah AM Al-Omary, Laila A Abou-Zeid, Mahmoud N Nagi, El-Sayed E Habib, A-M Alaa, Adel S El-Azab, Sami G Abdel- Hamide, Mohamed A Al-Omar, Abdulrahman M Al-Obaid, and Hussein I El-Subbagh. Non-classical antifolates. part 2: syn- thesis, biological evaluation, and molecular modeling study of some new 2, 6-substituted-quinazolin-4-ones. *Bioorganic* & *medicinal chemistry*, 18(8):2849–2863, 2010.

[15] W Goettsch, W Van Pelt, N Nagelkerke, MGR Hendrix, AGM Buiting, PL Petit, LJM Sabbe, AJA Van Griethuysen, and AJ De Neeling. Increasing resistance to fluoroquinolones in es- cherichia coli from urinary tract infections in the netherlands. *Journal of Antimicrobial Chemotherapy*, 46(2):223–228, 2000.

[16] TW Maddox, PD Clegg, NJ Williams, and GL Pinchbeck. An- timicrobial resistance in bacteria from horses: Epidemiology of antimicrobial resistance. *Equine veterinary journal*, 47(6):756– 765, 2015.

[17] Jason A Wiles, Qiuping Wang, Edlaine Lucien, Akihiro Hashimoto, Yongsheng Song, Jijun Cheng, Christopher W Mar- lor, Yangsi Ou, Steven D Podos, Jane A Thanassi, et al. Isoth- iazoloquinolones containing functionalized aromatic hydrocar- bons at the 7-position: synthesis and in vitro activity of a series of potent antibacterial agents with diminished cytotoxicity in human cells. *Bioorganic* & *medicinal chemistry letters*, 16 (5):1272–1276, 2006.

[18] Xiaoguang Huang, Dongliang Chen, Ning Wu, Aiqin Zhang, Zhenhua Jia, and Xingshu Li. The synthesis and biological evaluation of a novel series of c7 non-basic substituted fluo- roquinolones as antibacterial agents. *Bioorganic* & *medicinal chemistry letters*, 19(15):4130–4133, 2009.

[19] Bin Zhu, Brett A Marinelli, Raul Goldschmidt, Barbara D Foleno, Jamese J Hilliard, Karen Bush, and Mark J Ma- cielag. Synthesis and antibacterial activity of 7-(1, 2, 3, 4- tetrahydropyrrolo [1, 2-a] pyrazin-7-yl) quinolones. *Bioorganic* & *medicinal chemistry letters*, 19(17):4933–4936, 2009.

[20] Xiaoguang Huang, Aiqin Zhang, Dongliang Chen, Zhenhua Jia, and Xingshu Li. 4-substituted 4-(1h-1, 2, 3-triazol-1-yl) piperi- dine: Novel c7 moieties of fluoroquinolones as antibacterial agents. *Bioorganic & medicinal chemistry letters*, 20(9):2859–2863, 2010.

[21] Kai Lv, Jinwei Wu, Jian Wang, Mingliang Liu, Zengquan Wei, Jue Cao, Yexin Sun, and Huiyuan Guo. Synthesis and in vitro antibacterial activity of quinolone/naphthyridone

[22] derivatives containing 3-alkoxyimino-4-(methyl) aminopiperi- dine sca*ff*olds. *Bioorganic & medicinal chemistry letters*, 23 (6):1754–1759, 2013.

[23] Hongmin Liu, Ju Huang, Jiayang Wang, Minghua Wang, Min- gliang Liu, Bin Wang, Huiyuan Guo, and Yu Lu. Synthe- sis, antimycobacterial and antibacterial evaluation of l-[(1r, 2s)- 2-fluorocyclopropyl] fluoroquinolone derivatives containing an oxime functional moiety. *European journal of medicinal chem- istry*, 86:628–638, 2014.

[24] Jennifer M Andrews. Determination of minimum inhibitory concentrations. *Journal of antimicrobial Chemotherapy*, 48 (suppl 1):5–16, 2001.

[25] ChemBioDraw Ultra. 12.0. *Cambridge, USA: Cambridge Soft Corporation*, 1986.

[26] MJ Frisch, GW Trucks, HB Schlegel, GE Scuseria, MA Robb, JR Cheesemann, VG Zakrzewski, JA Montgomery Jr, RE Strat- mann, JC Burant, et al. Gaussian software, 2003.

[27] Michael JS Dewar, Eve G Zoebisch, Eamonn F Healy, and James JP Stewart. Development and use of quantum mechani- cal molecular models. 76. am1: a new general purpose quantum mechanical molecular model. *Journal of the American Chemi- cal Society*, 107(13):3902–3909, 1985.

[28] Chengteh Lee, Weitao Yang, and Robert G Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Physical review B*, 37(2):785, 1988.

[29] Warren J. Hehre. *Ab initio molecular orbital theory*. Wiley- Interscience, 1986.

[30] Andrea Mauri, Viviana Consonni, Manuela Pavan, and Roberto Todeschini. Dragon software: An easy approach to molecular descriptor calculations. *Match*, 56(2):237–248, 2006.

[31] R Todeschini, V Consonni, A Mauri, and M Pavan. Dragon- software for the calculation of molecular descriptors. *Web ver- sion*, 3, 2003.

[32] R Todeschini, V Consonni, and R Mannhold. Methods and prin- ciples in medicinal chemistry. *Kubinyi H, Timmerman H (Se- ries eds) Handbook of molecular descriptors. Wiley-VCH, Wein- heim*, 2000.

[33] Mar´ıa Jimena Mart´ınez, Ignacio Ponzoni, Mo´nica F D´ıaz, Gus- tavo E Vazquez, and Axel J Soto. Visual analytics in cheminfor- matics: user-supervised descriptor selection for qsar methods. *Journal of cheminformatics*, 7(1):1–17, 2015.

[34] Brian E Mattioni and Peter C Jurs. Prediction of dihydrofolate reductase inhibition and selectivity using computational neural networks and linear discriminant analysis. *Journal of Molecular Graphics and Modelling*, 21(5):391–419, 2003.

[35] Subhash C Basak. Mathematical descriptors for the predic- tion of property, bioactivity, and toxicity of chemicals from their structure: A chemical-cum-biochemical approach. *Current computer-aided drug design*, 9(4):449–462, 2013.

[36] Linnan He and Peter C Jurs. Assessing the reliability of a qsar model's predictions. *Journal of Molecular Graphics and Mod- elling*, 23(6):503–523, 2005.

[37] Rajarshi Guha, Jon R Serra, and Peter C Jurs. Generation of qsar sets with a self-organizing map. *Journal of Molecular Graphics and Modelling*, 23(1):1–14, 2004.

[38] Leona S Aiken, Stephen G West, and Steven C Pitts. Multiple linear regression. *Handbook of psychology*, 2003.

[39] Raimund Mannhold, Hugo Kubinyi, Hendrik Timmerman, and David E Clark. *Evolutionary algorithms in molecular design*, volume 8. John Wiley & Sons, 2008.

[40] Addinsoft. Xlstat 2015, data analysis and statistics software for microsoft excel, 2010.

[41] Peter J Rousseeuw and Annick M Leroy. *Robust regression and outlier detection*, volume 589. John Wiley & Sons, 2005.

[42] Alan R Katritzky, Liliana M Pacureanu, Svetoslav Slavov, Dim- itar A Dobchev, and Mati Karelson. Qsar study of antiplatelet agents. *Bioorganic* & *medicinal chemistry*, 14(22):7490–7500, 2006.

[43] Alexander Golbraikh and Alexander Tropsha. Predictive qsar modeling based on diversity sampling of experimental datasets for the training and test set selection. *Molecular diversity*, 5(4): 231–243, 2000.

[44] Alexander Golbraikh and Alexander Tropsha. Beware of q 2! *Journal of Molecular Graphics and Modelling*, 20(4):269–276, 2002.

[45] Kunal Roy, Pratim Chakraborty, Indrani Mitra, Probir Kumar Ojha, Supratik Kar, and Rudra Narayan Das. Some case studies on application of ”rm2” metrics for judging quality of quanti- tative structure–activity relationship predictions: emphasis on scaling of response data. *Journal of computational chemistry*, 34(12):1071–1082, 2013.

[46] Liying Zhang, Hao Zhu, Tudor I Oprea, Alexander Golbraikh, and Alexander Tropsha. Qsar modeling of the blood–brain bar- rier permeability for diverse organic compounds. *Pharmaceuti- cal research*, 25(8):1902–1914, 2008.

[47] E Papa, JC Dearden, and P Gramatica. Linear qsar regres- sion models for the prediction of bioconcentration factors by physicochemical properties and structural theoretical molecular descriptors. *Chemosphere*, 67(2):351–358, 2007.

[48] Paola Gramatica, Elisa Giani, and Ester Papa. Statistical exter- nal validation and consensus modeling: A qspr case study for k oc prediction. *Journal of Molecular Graphics and Modelling*, 25(6):755–766, 2007.

[49] Lennart Eriksson, Joanna Jaworska, Andrew P Worth, Mark TD Cronin, Robert M McDowell, and Paola Gramatica. Methods for reliability and uncertainty assessment and for applicability evaluations of classification-and regression-based qsars. *Envi- ronmental health perspectives*, 111(10):1361, 2003.

[50] Lee Phillips. *gnuplot Cookbook*. Packt Publishing Ltd, 2012.

[51] Jianying Hu, Wanfeng Wang, Zhou Zhu, Hong Chang, Feng Pan, and Binle Lin. Quantitative structure-activity relationship model for prediction of genotoxic potential for quinolone an- tibacterials. *Environmental science* & *technology*, 41(13):4806– 4812, 2007.