Bulgarian Chemical Communications, Volume 48, Number 4 (pp. 736 – 745) 2016

QSAR analysis of N[1]-substituted 1,2,4-triazoles against *Escherichia coli*

V. Dimova[1]*, I. Jordanov[1], L. Dimitrov[2]

[1]*Faculty of Technology and Metallurgy, University Ss. Cyril and Methodius Rudjer Boskovic 16, 1000 Skopje, Republic of Macedonia*

[2]*Institute of Agriculture, University Ss. Cyril and Methodius, 16ta Makedonska brigada 3a, 1000 Skopje, Republic of Macedonia*

Received October 2, 2015; Revised March18, 2016

QSAR analysis of a series of previously synthesized N[1]-substituted 1,2,4-triazole derivatives tested for growth inhibitory activity with respect to *Escherichia coli*, was performed using the computer-assisted multiple regression procedure. Using the Hansch and Free Willson approach the activity contribution of the aminomethyl/aminoethyl unit and aromatic/heteroaromatic ring was determined from the correlation equation. In accordance with the statistical parameters (R=0.8729; R^2_{adj}=0.6430; Sd=0.2983; Q^2=0.7548 and PRESS/SSY=0.2452), the biparametric model which involves R and L, is selected as the best biparametric model, for determing the activity of the chosen triazole derivatives against *E. coli*. Spreading the investigated system: the subset B (aminomethyl unit replaced with the aminoethyl group) and subset C (aromatic replaced with a heteroaromatic ring), statistically significant QSAR models were not obtained.

Keywords: quantitative structure - activity relationships, 1,2,4-triazole, antibacterial activity, Escherichia coli,

INTRODUCTION

The 1,2,4-Triazole system is a structural element of many drugs that have antimycotic activity such as fluconazole, itraconazole, voriconazole [1-3]. Because of the synthetic utility and broad range of pharmacological effects, the 1,2,4-triazole nucleus is an important five member ring, and the interest in the synthesis and microbiology of this pharmacophore continues to be fuelled by its analgetic, antiasthmatic, diuretic, antihypertensive, antibacterial, antifungal and anti-inflammatory properties [4-10].

One of the methods of preparing derivatives of 1,2,4-triazole is the Mannich reaction (aminomethylation), a well know process [11,12]. N-hydroxymethyl derivatives of heterocycles such as: benzotriazoles and benzimidazoles under the influence of amines, can also give corresponding Mannich bases [12-14]. It is also known that some aminomethyl heterocycles, that possess biological and corrosion-inhibition activity can be used as additives in greasy oils as well as photopolymerizing paints to improve adhesion [15,16].

Quantitative structure activity relationships (*QSARs*) are estimation methods developed and used in order to predict certain effects or properties of chemical substances which are primarily based on the structure of the substances. They have been developed on the basis of experimental data on model substances. Today the investigation of the *QSAR* of the substances is an important issue in modern chemistry, biochemistry, medicinal chemistry, as well as in drug discovery [17-20].

This information is composed of mathematical equations relating the chemical structure of the compounds to a wide variety of their physical, chemical, biological and technological properties. Once a correlation between the structure and activity/property is found, any number of compounds, including those not yet synthesized, can be readily screened on the computer in order to select the structure with the desired properties. Then, it is possible to select the most promising compounds to synthesize and test in the laboratory.

In the above mentioned contex and in the continuation of the studies of new antimicrobial agents which possess heterocyclic rings in their structure, such as the 1,2,4-triazole, during the last ten years,some new 1,2,4-triazole derivatives [21-24] were synthesized and investigated.

A group of 18 N[1]-aryl/heteroarylamino-/methyl/ethyl-1,2,4-triazole derivatives was synthesized by condensation of the hydroxymethyl derivative of 1,2,4-triazole and the appropriate aromatic/heteroaromatic amines and by the reaction of 1,2,4-triazole, acetaldehyde and a few aromatic/heteroaromatic amines. All the synthesized compounds were screened for their antibacterial activities with respect to *Escherichia coli* [24].

The objective of this investigation was to study the usefulness of QSAR in the prediction of the antibacterial activity of the investigated triazole derivatives with respect to *Escherichia coli*. Multiple linear regression (MLR) models have been

* To whom all correspondence should be sent:
E-mail: vdimova@tmf.ukim.edu.mk

developed as a mathematical equation which correlate the chemical structure to the activity.

EXPERIMENTAL

Materials and methods

All N^1-aryl/heteroarylamino/methyl/ethyl-1,2,4-triazole derivatives (1-18), (Table1), used in this study were previously synthesized and reported elsewhere [24].

Antibacterial Investigation

All 1,2,4-triazole derivatives were tested for their *in vitro* growth inhibitory activity against *Escherichia coli*. The antimicrobial activity of the investigated compounds were tested by the filter paper disc method [25], using standard conditions in a Mueller-Hinton agar medium, inoculated with 0.5 mL of the 24 h liquid cultures containing 10^7 microorganisms/mL.

Stock solutions of the compounds were prepared in DMSO, as an inert medium in 3 concentrations: 1 mg/mL; 5 mg/mL and 10 mg/mL DMSO (Figure 1). Filter paper discs (5 mm diameter) saturated with each compound solution were placed on the indicated agar mediums. The incubation time was 24 h at 37^0C. A control disc using DMSO without any test compound was included. There was no inhibitory activity in those disks. The diameter of the zone inhibition (mm) was measured. The MIC values of the triazoles tested were obtained as µg/mL (Table 1). Every test was done in triplicate to confirm the findings.

Multiple Linear Regression

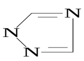
The mathematical foundation of the quantitative structure – activity relationship is based on the principle of polylinearity. Multiple linear regression is a common method used in QSAR studies. The QSAR equations were obtained by the multilinear forms:

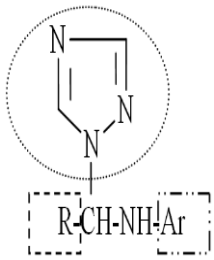
$$\log 1/c = a_0 + a_1D_1 + a_2D_2 + a_3D_3 + \ldots + a_nD_n, \quad (1)$$

where D_1, D_2, D_3 and D_n are descriptors, n is the number of desriptors. The intercept (a_0) and regression coefficient of the descriptors were determined using the least squares method.

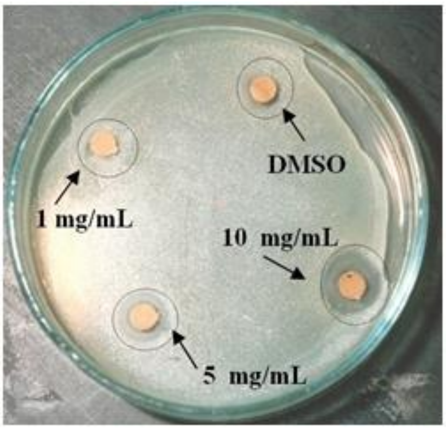
The MVA (multi variable analysis) approach in *QSAR* analysis has been most widely and effectively used for theoretical drug design due to various physicochemical (electronic, steric and hydrophobic) parameters and structural indicator parameters used together (*Hansch* and *Free Willson* approach) [26,27].

Table 1. Calculated $\log 1/c_{MIC}$ values; logP values and matrix of the Free Willson approach for N^1-aryl/heteroarylaminomethyl/ethyl-1,2,4-triazole derivatives (1-18)

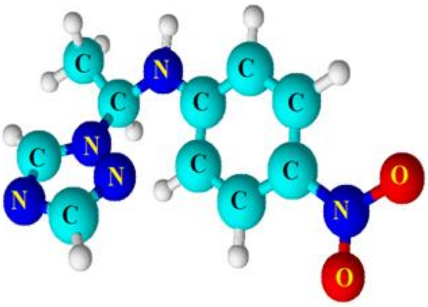
Comp.	R	Ar	log 1/c_{MIC}	logP[a]	I_H	$I_{=CH-}$
1	H	$-C_6H_4-COOC_2H_5$ (*p*)	5.3914	1.2981	1	1
2	H	$-C_6H_4-COOH$(*p*)	4.3388	0.4579	1	1
3	H	$-C_6H_4-COOH$(*o*)	4.3388	1.4690	1	1
4	H	$-C_6H_4-Cl$(*p*)	5.3194	1.2210	1	1
5	H	$-C_6H_4-Br$(*p*)	5.4034	1.4665	1	1
6	H	$-C_6H_4-CH_3$ (*p*)	5.2747	1.1238	1	1
7	H	$-C_6H_4-C_6H_5$ (*p*)	5.3985	2.3405	1	1
8	H	$-C_6N_4-NH-CH_2-$ 	4.7328	-1.6900	1	1
9	CH_3	$-C_6H_4-COOC_2H_5$ (*p*)	5.4156	1.7157	0	1
10	CH_3	$-C_6H_4-NO_2$ (*p*)	4.3678	1.3895	0	1
11	H	2-Pyridyl-	5.2435	0.0301	1	0
12	H	4-Methyl-2- pyridyl -	5.2769	0.5774	1	0
13	H	6- Methyl-2- pyridyl -	5.2769	0.5774	1	0
14	H	5-Chloro-2-pyridyl-	-	0.6746	1	0
15	H	2-Pyrimidyl-	-	-1.4763	1	0
16	H	1,2,4-Triazole-4-yl	4.5189	-2.1800	1	0
17	CH_3	6-Methyl-2-pyridyl-	-	0.9950	0	0
18	CH_3	2-Thiazolyl-	-	0.7441	0	0



[a]*Ref. [28].*



a



b

Fig. 1. a) Filter paper discs saturated with compound (10) solution (1 mg/mL; 5 mg/mL and 10 mg/mL DMSO) and b) optimized geometry of the compound (10)

The assumption can be formulated as given in the following equation (*Hansch* approach):

$$\log 1/c = A_1 x + A_2 y + A_3 z + b, \qquad (2)$$

where *x*, *y* and *z* are the molecular properties and log1/c are the desired biological activities.

From the values of linear slopes A_1, A_2, A_3 we can see the correlation of the particular molecular properties with the activity of the investigated compounds. Applying the same chosen descriptors in *Free Willson* analysis, the activity contributions of either methyl- or substituted heterocyclic ring systems were determined from the correlation equation:

$$\log 1/c = \Sigma a_i I_i + \Sigma b_i x_i + b \qquad (3)$$

where I_i is the structural indicator parameter.

Descriptors

The variables used as descriptors in the analysis are electronic, steric and structural parameters (Tables 1 and 2). The physicochemical parameters taken into consideration in the *QSAR* study are: σ, π, MR, F, R, L and logP (Table 2). For each compound the partition coefficient logP has been calculated (Table 1) [28].

The *Hammett* constant σ is an electronic substituent constant reflecting the electron-donating or electron-withdrawing properties of a substituent. It thus serves as a quantitative measure of the change in the electronic density of the reaction center caused by the introduced substituent. Positive σ values are defined to represent the residues that unfold an electron-withdrawing effect stronger thanthe hydrogen present in the unsubstituted derivative, negative σ values vice versa. It is possible to express the electronic effect of a substituent in terms of non-resonance and resonance capability. The parameter σ_m describes the inductive effect on the electron density present in the reaction center, hence emphasizing the non-resonance capability. The mesomeric effect on the electron density is characterized by σ_p, putting an emphasis on the resonance capability. As resonance conjugation cannot be performed from the *meta*-position, σ_R contributes indirectly to σ_m. A distinction between the non-resonance and resonance capability can also be made by employing the parameters F and R, which are the field (F) and resonance (R) constants calculated by *Swain* and *Lupton* from the *Hammett* σ_m and σ_p values. *Swain* and *Lupton* claim the descriptors F and R "to be more accurately defined and more physically significant independent variables for correlating or predicting substituent effects" than any other pair of *Hammett* values. F and R are supposed to be independent of the performed reaction, of the solvent, and of the temperature. *Hansch* attests the field and resonance constants a remarkable orthogonality and thus theavoidance of the common problem of multi-collinearity. The empirical weighting factors *f* and *r* are independent of the substituent but different for each substituent constant (i.e., σ_m, σ_p):

$$\sigma = f\,F + r\,R \qquad (4)$$

where σ is the *Hammett* constant, *f* the field weighting factor, F the field constant, *r* the resonance weighting factor and R is the resonance constant.

A parameter readily available for each substituent is the molar refractivity (MR), described as the molar volume of the substituent, corrected by the refractive index. The refractive index being a part of the definition hints a contribution of the polarizability of the substituent to the MR value which must be regarded as an electronic contribution:

$$MR = \frac{M(n^2 - 1)}{d(n^2 + 2)} \qquad (5)$$

where MR is the molar refractivity of the substituent, M is the molecular weight of the substituent, dis the density of the substituent, and n is the refractive index, measured at 20 °C.

A purely geometrical definition of the size of a substituent is provided by *Verloop* who thereby wanted to overcome the problem of asymmetry of the substituents. In order to be able to describe the substituents deviating from a spherical shape, a set of five parameters was developed. The length parameter L and the four width parameters B1-4, represent the four rectangular directions perpendicular to the axis describing the length.

The partition coefficient, referred to as logP, represents a parameter expressing the hydrophobicity of the molecule as a whole. The logP value of a compound is commonly determined by measuring the distribution behavior in a biphasic system consisting of 1-octanol, representing the lipid phase, and water:

$$P = \frac{c_{(\text{lipid phase})}}{c_{(\text{water phase})}} \qquad (6)$$

where P is the partition coefficient of the compound, $c_{(\text{lipid phase})}$ is the concentration of the compound present in the lipid phase, and $c_{(\text{water phase})}$ is the concentration of the compound present in the water phase.

The *Hansch* constant π describes the contribution of a substituent to the lipophilicity of a compound. The π values are usually derived from the benzene solute system, that is, by partitioning substituted benzene derivatives between 1-octanol and water. Subtraction of the logP value of the unsubstituted benzene from the logP value of the substituted compound yields the π value of the appropriate substituent. Positive π values represent an amplification of the lipophilic character caused by the substituent; negative π values that symbolize an increase in hydrophilicity:

$$\pi_x = \log P_{R\text{-}X} - \log P_{R\text{-}H} \qquad (7)$$

where π_X is the Hansch constant characteristic of the substituent X, $P_{R\text{-}X}$ is the partition coefficient of the X-substituted benzene, and $P_{R\text{-}H}$ is the Partition coefficient of the unsubstituted benzene.

Applying the *Free Willson* analysis, in the first step, the structural variable indicator I_H expresses the replacement of the hydrogen atom with a methyl group in an aminomethyl unit. I_H is defined as 1 for the N^1-aryl/heteroarylaminomethyl-1,2,4-triazoles (1-8, 11-16), and 0 for N^1-aryl/heteroarylaminoethyl-1,2,4-triazole derivatives (9,10,17,18). In the second step, the other indicator $I_{=CH-}$ is defined as 1 for compounds with =CH- in six membered ring (1-10) and 0 for compounds with –N- replacement (11-18) (Table 1).

Statistical analysis

The statistical evaluation of the data was performed using the STATISTICA program package [31]. To test the quality of the regression models, the following statistical parameters were used:

- ***Correlation coefficient (R)*** - measures the degree to which the dependent variable is linearly related to the explanatory variables.
- ***Standard deviation of the estimate (Sd)*** - defined as the square root of the variance represents the most commonly used measure of the spread. Sd expresses the degree of deviation of the calculated biological activity from the experimentally determined biological activity.
- ***Fisher test for singnificance of the equation (F-test)*** mirrors the ratio of the variance, which is a measure of the spread of data, that is explained by the established regression equation to the variance not explained, taking the degrees of freedom into account. The F-distribution can be regarded as a measure of significance of the established regression equation as a whole.
- ***Adjusted R^2 (R^2_{adj})*** an adjusted version of R**:**

$$R^2_{adj} = 1 - R^2((n-1)/(n-p-1)) \qquad (8)$$

where *n* is the number of compounds and *p* is the number of independent parameters.

- ***Predictive residual error Sum of Squares (PRESS)*** and ***Sum of squares of deviation of the experimental values from their mean (SSY)***:

$$PRESS = \Sigma(Y_{pred} - Y_{exp})^2 \quad \text{and}$$

$$SSY = \Sigma\,(Y_{exp} - Y_{mean})^2 \qquad (9)$$

where Y_{pred}-predicted, Y_{exp}-experimental and Y_{mean}-mean are the values of the target property; in our case the (log1/C) values respectively. PRESS appears to be an important cross-validation parameter accounting for a good estimate of the real predictive error of the model. Its value less than SSY (PRESS<<SSY) indicates that the model predicts better than chance and can be consided statistically significant. If the sum of the squared deviations of the calculated values from the observed values (PRESS) is larger than the sum of the squared deviations of the observed values from the mean experimental value (SSY), Q^2 adopts a negative value. This implies that the proposed regression equation does not provide reasonable predictions.

Table 2. Physicochemical parameters of the triazole derivatives studied [a]Ref. [29]; [b]Ref. [30]

Substituent	σ^a	π^a	MR^a	F^a	R^a	L^a
p-$COOC_2H_5$	0.45	0.51	17.47	0.33	0.15	5.96
p-COOH	0.45	-0.32	6.93	0.33	0.15	3.91
o-COOH	1.2^b	-0.32	6.93	0.33	0.15	3.91
p-Cl	0.23	0.71	6.03	0.41	-0.15	3.52
p-Br	0.23	0.86	8.88	0.44	-0.17	3.83
p-CH_3	-0.17	0.56	5.65	-0.04	-0.13	2.87
p-C_6H_5	-0.01	1.96	25.36	0.08	-0.08	6.28
p-NO_2	0.78	-0.28	7.36	0.67	0.16	3.44

- ***Cross-validation squared correlation coefficient (Q^2)*** is widely adopted to quantitatively express the predictive power of a correlation and can by calculated by equation:

$$Q^2 = 1 - PRESS/SSY \quad (10)$$

- ***Quality factor (Q)*** can be calculated by equations**:**

$$Q = R/Sd \quad (11)$$

A higher value of *Q* indicates a better prediction of the model.

- ***Uncertainty of Prediction (S_{PRESS})*** and ***Predictive Square Error (PSE)***can by calculated by equations:

$$S_{PRESS} = \sqrt{\frac{PRESS}{(N-p-1)}}$$

$$PSE = \sqrt{\frac{PRESS}{N}} \quad (12)$$

The lower values of S_{PRESS} and *PSE* indicate a better model.

- ***Variance inflation factor (VIF)*** is defined as:

$$VIF=1/(1-R_i^2) \quad (13)$$

where R_i is the multiple correlation coefficient of the *i*-th independent variable on all of the other independent variables.

RESULTS AND DISCUSSION

The series of 18 substiuted 1,2,4-triazole derivatives may be organized in 3 subsets:

Subset A - structurally identical compounds (1-8);

Subset B - compounds with structural changes (*aminomethyl* unit replaced with the *aminoethyl* group) (1-10) and

Subset C - compounds with ring changes (the aromatic ring has been replaced with a heteroaromatic one) (1-18).

In this work QSAR analyses were made between selected physicochemical properties and experimentally obtained values for antimicrobial activities with respect to *Escherichia coli*, applying the general *Hansch* equation for **subset A**, *Free-Wilson* approach for **subset B** and extend *Free-Wilson* equation for **subset C**.

It is well known that there are three important phases in any QSAR study:

i. *development of the models;*
ii. *statistical validation of the obtained models and*
iii. *utility of the developed models.*

In the first step for the development of QSAR models, the selected 1,2,4-triazoles were evaluated for *in vitro* antibacterial activity against *Escherichia coli*. After applying the filter paper disc method, the compounds 14, 15, 17 and 18 do not inhibit the growth of the test strain [24]. In the second step, efforts were focused on developing the QSAR models of compounds with antibacterial activity. Inhibitory activity data determined as µg/mL were first transformed to the negative logarithms of molar MICs ($log1/c_{MIC}$), (Tab. 2) which were used as a dependent variable in the QSAR study.

In accordance with the calculated values, $log1/C_{MIC}$ is lowest for *orto/para* COOH substituted 1,2,4-triazoles (**2** and **3**). Following the sequence of antibacterial activity the values observed were:

2 = 3 < 10 < 16 < 8 < 11 < 6 < 12 = 13 < 4 < 1 < 7 < 5 < 9

Since the previous work descriptors such as: surface tension, molar refraction, molar volumen, parachor, index of refractivity, density and polarizability were used [23], in the present study different electronic, steric and structural descriptors (σ, π, MR, F, R, L, logP) and structural variable indicators (I_H and $I_{=CH-}$) (Tables 1 and 2), were used as an independent variable and were correlated with antibacterial activity ($log1/c_{MIC}$).

One-variable model. The relatively good monoparametric model was obtained only for π indicating the importance of the descriptor in contribution to the inhibitory activity (**Model 1**; Figure 2a).

Model 1 $\log 1/c_{MIC} = 4.7747(\pm0.1569) + 0.5156(\pm0.1714)\pi$

$R=0.8025$ $R^2_{adj}=0.5779$ $F=9.0471$
$Sd=0.3263$ $p<0.2983$

Furthermore, the data shows that some of the chosen descriptors such as σ (R=0.7235) and R (R=0.6926) correlate median with the activity. A statistically unreliable model was obtained for the following descriptors: logP, MR, F and L (R<0.4).

From this it was concluded that no single variable model is capable of good modelling of activity and that the refereed descriptors should be combined to obtain a statistically significant multiparametric model for modelling the activity.

Two-variable models. In bivariate correlation analysis, by applying a stepwise multiple linear regression method, 21 models were obtained. Among them a few best models were selected for the futher discussion (**Models 2-9**). The selection was based on the statistical quality of the models (R; Sd; F-test; R^2_{adj}; p-level).

Model 2 $\log 1/c_{MIC} = 4.9839(\pm0.2837) - 0.3745(\pm0.4194)\sigma + 0.3710(\pm0.2383)\pi$

$R=0.8385$ $R^2_{adj}=0.5548$ $F=4.7391$
$Sd=0.3331$ $p<0.0881$

Model 3 $\log 1/c_{MIC} = 4.8484(\pm0.4007) - 0.7481(\pm0.3259)\sigma + 0.3526(\pm0.2575)\log P$
$R=0.8219$ $R^2_{adj}=0.5132$ $F=4.1639$
$Sd=0.3483$ $p<0.1052$

Model 4 $\log 1/c_{MIC} = 5.1377(\pm0.4425) + 0.7216(\pm0.2925)\pi - 0.3579(\pm0.4067)\log P$
$R=0.8377$ $R^2_{adj}=0.5527$ $F=4.7069$
$Sd=0.3339$ $p<0.0889$

Model 5 $\log 1/c_{MIC} = 4.9288(\pm0.2491) + 0.6698(\pm0.2593)\pi - 0.0218(\pm0.0268)MR$
$R=0.8335$ $R^2_{adj}=0.5422$ $F=4.5533$
$Sd=0.3378$ $p<0.0931$

Model 6 $\log 1/c_{MIC} = 4.8340(\pm0.1755) + 0.3907(\pm0.2288)\pi - 0.9911(\pm1.1590)R$

$R=0.8361$ $R^2_{adj}=0.5486$ $F=4.6460$
$Sd=0.3354$ $p<0.0905$

Model 7 $\log 1/c_{MIC} = 5.0401(\pm0.5498) + 0.5747(\pm0.2192)\pi - 0.0688(\pm0.1355)L$

$R=0.8158$ $R^2_{adj}=0.4984$ $F=3.9809$
$Sd=0.3536$ $p<0.1118$

Model 8 $\log 1/c_{MIC} = 4.6927(\pm0.2371) + 0.0314(\pm0.0181)MR - 2.3438(\pm0.8880)R$
$R=0.8384$ $R^2_{adj}=0.5545$ $F=4.7343$
$Sd=0.3332$ $p<0.0882$

Model 9 $\log 1/c_{MIC} = 4.0750(\pm0.4575) - 2.7899(\pm0.8309)R + 0.2208(\pm0.1014)L$
$R=0.8729$ $R^2_{adj}=0.6430$ $F=6.4038$
$Sd=0.2983$ $p<0.0566$

Three-variable models. In the next step an attempt was made for finding the triparametic correlation analysis, involving some of the parameters. Only **Model 10** and **11** gave relatively statistically good results.

Model 10 $\log 1/c_{MIC} = 4.8897(\pm0.4245) - 0.2822(\pm0.4814)\log P + 0.0474(\pm0.0337)MR - 2.7389(\pm1.1823)R$
$R=0.8565$ $R^2_{adj}=0.46708$ $F=2.7529$
$Sd=0.3654$ $p<0.2138$

Model 11 $\log 1/c_{MIC} = 4.6032(\pm0.4202) + 0.0332(\pm0.0216)MR + 0.2578(\pm0.9266)F - 2.4082(\pm1.0385)R$
$R=0.8429$ $R^2_{adj}=0.4097$ $F=2.4541$
$Sd=0.3790$ $p<0.2401$

Models 10 and **11** indicate that logP and R negatively contribute to the biological activities, opposite to MR and F.

Model with structural variable indicators. The aminoalkyl linker between the triazole and substituted aromatic core was also investigated (Subset B). The extension of the alkyl chain (methyl) by one carbon (ethyl) led to the statistically realable regression **model 12**:

Model 12 $\log 1/c_{MIC} = 4.8251(\pm0.4245) + 0.5799(\pm0.1748)\pi - 0.0866(\pm0.2889)I_H$
$R=0.8095$ $R^2_{adj}=0.5403$ $F=5.7024$ $Sd=0.3467$ $p<0.0410$

where I_H is a structural indicator parameter representing $-CH_2-$ group as 1 and CH_3-CH- group as 0.

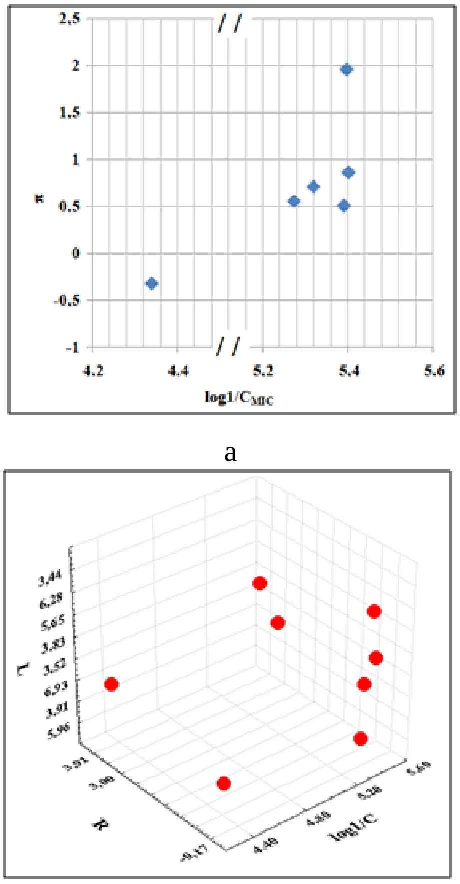
Model with heteroaromatic unit. Further modification of the investigated triazole set whenthe aromatic was replaced with a heteroaromatic core (subset C), led to a reduction in activity. After applying an extend *Free – Wilson* equation for such a structured group of compoundsthe following model was obtained:

Model 13 $\log 1/c_{MIC}=5.1300(\pm0.2110) + 0.2051(\pm0.1057)\log P - 0.3532(\pm0.2849) I_{=CH-}$
$R=0.5101$ $R^2_{adj}=0.1257$ $F=1.9345$ $Sd=0.4189$ $p<0.1905$

A good correlation was not obtained, indicating that another type of heteroaromatic core must be chosen to be a part of the system investigated.

Occurrence of colinearity

At this step, it is worth examining the occurrence or otherwise of colinearity in the proposed models. The best candidates for this purpose would obviously be the two-parametric models. We can do this in two different ways: (*i*) by examining the correlation matrices for used descriptors and (*ii*) by statistical evaluation - calculation of the additional statistical parameters: PRESS/SSY; Q^2; PSE; SPRESS and Q.



a

b

Fig. 2. a) Linear correlation between log $1/c_{MIC}$ and π (**model 1**) b) 3D Scatterplot of correlation between log $1/c_{MIC}$, R and L (**model 9**)

Correlation matrix. It was important for further analysis to find the correlation matrix for used descriptors and their correlation with the activity (Table 3). The results show mutual correlation between some of the used descriptors. So, if a combination of them is present in the regression expression, then the model may suffer from the defect due to collinearly. Also, it may result in a change in signs of the coefficients, a change in the values of the previous coefficient, a change of a significant variable into an insignificant one or an increase in the standard error of the estimate in the addition of an additional parameter to the model.

In accordance with Table 3, **models 2, 4** and **5** were excluded from further statistical analysis, although those models had relatively well correlated coefficients and a standard deviation (R>0.83; Sd<0.34). In the statistically best regression **model 9**, descriptor L has the positive effect on the log$1/c_{MIC}$ value, but its influence is not significant compared with the influence of the descriptor R (Figure 2b). The descriptor R receives a relatively large negative coefficient (-2.7899), indicating that this descriptor leads to a lower log $1/c_{MIC}$ value.

Taking into consideration the above mentioned and preliminary conclusions of the statistical evaluation of the quality of all the models (R, R^2_{adj}; F; Sd, p), only **models 9** and **10** can be used as relatively statistically significant.

Finally, in order to confirm these findings, the antimicrobial activity with respect to *Escherichia coli* (log $1/c_{MIC}$) as predicted by **models 9** and **10** was compared with the corresponding observed values reported in Table 1. Within experimental error, the values agree well.

Also, the predictive correlation coefficient (R_{pre}^2) has been calculated, (Table 4). The obtained predictive correlation coefficient (R_{pre}>0.8) confirms our conclusion. The values R_{pre} are found >0.8, respectively, for the **models 9** and **10**. Correlation between the observed log $1/c_{MIC}$ and predicted log$1/c_{MIC}$ values and for all active compounds, calculated by: *(i)* **model 9** and *(ii)* **model 10**, are presented in Figure 3.

Table 3. Correlation matrix for the chosen electronic, steric and hydrophobic parameters.

	σ	**π**	**logP**	**MR**	**F**	**R**	**L**
σ	**1.0000**	*-0.7226*	-0.1338	-0.2845	0.6140	*0.7557*	-0.1180
π		**1.0000**	*0.7255*	*0.7325*	-0.5044	-0.6925	0.5748
logP			**1.0000**	*0.7202*	-0.1879	-0.2990	0.5669
MR				**1.0000**	-0.3256	-0.0191	*0.9540*
F					**1.0000**	0.3934	-0.2317
R						**1.0000**	0.1589
L							**1.0000**

Table 4. Calculated predictive correlation coefficient (R_{pre}) for **models 9** and **10**.

Model	Correlation: log1/C_{model} with log1/C_{exp}	R_{pre}^2	Sd
9	log1/ $c_{model\ 9}$= 1.3505 +0.7334 log1/$c_{exp.}$	0.8565	0.2417
10	log1/ $c_{model\ 10}$=1.4669 +0.7105 log1/$c_{exp.}$	0.8429	0.2482

Table 5. Cross-validation parameters (Q^2, PRESS/SSY, S_{PRESS}, PSE and Q) calculated for **models 9** and **10**.

Model	Parameters	Q^2	PRESS/SSY	*S_{PRESS}*	*PSE*	*Q*
9	R; L	0.7548	0.2452	0.271	0.214	2.926
10	logP; MR; R	0.7335	0.2665	0.283	0.223	2.328

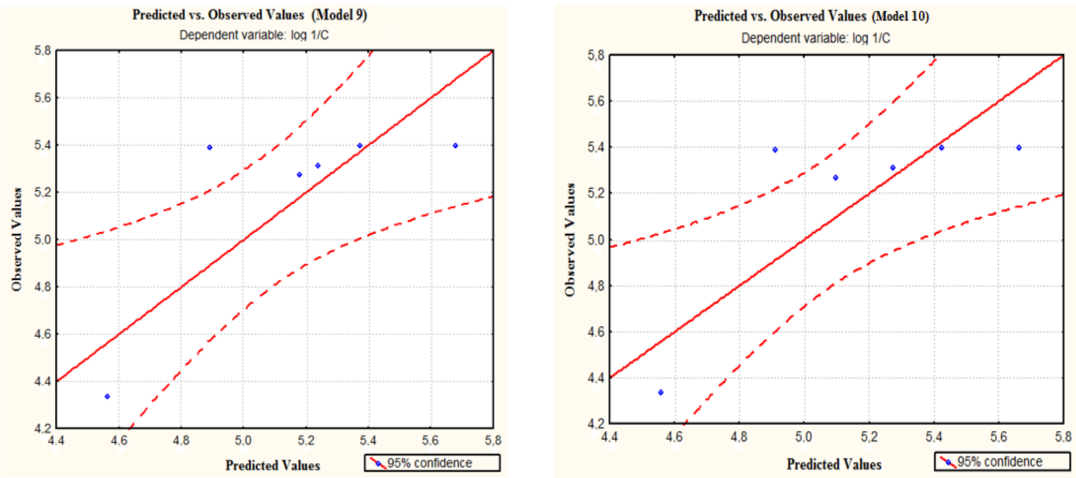


Fig. 3. Plot between the observed and predicted log 1/c_{MIC} values for a) **model 9** and b) **model 10**

Validation

Performing the multiple linear regression of a dependent variable (y; log 1/c_{MIC}) offers the possibility of choosing a large number of explanatory variables (x) and thus raises the question of significance in an acute form. Statistical quantities need to be calculated in order to assess the success of the correlation. Validation is a crucial aspect of any QSAR analysis and a cross-validation methodology was undergoing for deciding the predictive power of the proposed models. This is needed because a model with good statistics may not have a good predictive potential. The predictivity of each model was measured by several cross-validation parameters: Q^2, PRESS/SSY; *S_{PRESS}*; *PSE* and *Q*.

For a reasonable QSAR model, PRESS/SSY should be smaller than 0.4 [32]. In our case the ratio PRESS/SSY<0.27 indicates that the proposed models are reliable QSPR models (Tab. 5). Good Q^2, *S_{PRESS}* and *PSE* values, were also obtained: **(**Q^2>0.73; *S_{PRESS}* and *PSE*<0.3)(Tab. 5),confirming the assumption that these models can be used as a tool for predicting the inhibition of *E. coli* [33]. But, the lowest value for *S_{PRESS}* and *PSE*, for **model 9**, compared to other models, supports its highest predictive potential. An additional requirement should be fulfilled, the difference between R^2 and Q^2 should not be more than 0.3 [33].

The predictive power, determined by the Pogliani Q parameter [34] for **model 9** (Q=2.926) confirms that this model has excellent statistics as well as excellent predictive power, compared with the other model (Tab. 5). In our study, since the previous data are ofa similar ranking, Q is taken as proof of the high predictive ability of the QSAR **model 9**.

We also calculated the variance inflation factor (VIF) for each of the parameters in the selected models, as a measure of multicollinearity [35]. A VIF 10 or more (no upper limit is defined) for large data sets indicates a collinearity problem. For small data sets, even VIFs of five or more (here no upper limit is defined as well) can signify collinearity. The variables with a high VIF are candidates for exclusion from the model. First, we calculated VIF values for **model 9** and **10** (as statistically the best models according to the preliminary statistical data) all VIF values are<4 (VIF=1.102 ÷3.269), meaning that these models are free from defects caused by colinearity (Tab.6).

Table 6. VIF values for the two- and three-variabel models.

Models	Descriptors	Variance inflation factor (VIF)
9	R	1.102
	L	1.102
10	logP	3.269
	MR	2.902
	R	1.487

In the statistically best regression **model 9**, the descriptor L has a positive effect on the log1/c_{MIC} value, but his influence is not significant (7.33%) compared with the influence of descriptor R (92.67%), (Figure 2b). The descriptor R receives a relatively large negative coefficient (-2.7899), indicating that this descriptor leads to a lower log 1/c_{MIC} value. In this model, the coefficients of R and L are much higher than their standard deviation, which is another confirmation for the statistical significance of **model 9**.

The statistical evaluation of the data used to test the quality of the obtained models indicated that **model 9** is statistically significant (R=0.8729; R^2_{adj}=0.6430; F=6.4038; Sd=0.2983 p<0.0566; R_{ped}=0.8565; PRESS/SSY=0.2452; Q^2=0.7548; S_{PRESS}=0.271; *PSE*=0.214; Q=2.926), when all the parameters are summarized.

The MLR method can be useful when a relatively small number of descriptors are used. In this case, for the active compounds, the obtained QSAR models with two descriptors can be good in order to avoid a high chance of spurious correlations. Therefore, we selected **model 9** as the best statistically biparametric models, for determing the activity of the chosen triazole derivatives against *E. coli*.

As a result of spreading the investigated system (Subsets B and C) the results showed that statistically significant QSAR models were not obtained. That means that maybe another heterocyclic nucleus (beside those used in this study) may lead to developing a better *QSAR* system.

CONCLUSION

Spurred by the need of new antimicrobial agents and the fact that many effective drugs, insecticides and fungicides possess heterocyclic systems in their structure, such as triazole core, some new 1,2,4-triazole derivatives were synthesized. Form the results and discussion presented in this work, a conclusion can be made that part of the investigated N^1-substituted 1,2,4-triazole derivatives are effective *in vitro* against the tested strain *Escherichia coli*. The inhibitor activity of triazoles derivatives were modeled using multi linear models based on the chosen physico - chemical descriptors and structure variable indicators. Analysis of this limited set of substituted 1,2,4-triazole molecules allowed us to build a model of their antimicrobial activity against *E. coli* in which logP, MR, R and L are important factors.

The obtained biparameter models showed a relatively good correlation and predictive ability, in comparison with the monoparametric models. The validity of the models have been established by the determination of suitable statistical parameters such as: R; R^2_{adj}; Sd; F-test; Q^2; PRESS/SSY, S_{PRESS}, *PSE* and Q.

The structural modification which was made of the basic set (Subset A) didn't show any upgrading of the QSAR models.This result is a good base for expanding the 1,2,4-triazole set with new compounds which will have improved characteristics. In consequence this will help the medical and agriculture chemist in their prediction of an increasing activity and thus the synthesis of new triazoles exhibiting better activities than those reported in this paper.

REFERENCES

1. J. K. Sahu, S. Ganguly, A. Kaushik, *J App Pharm Sci.*,**4,** 81 (2014).
2. N. Chaudharya, R. Dubeyb, H. Panwar, *Der Pharma Chemica*, **6**, 115 (2014).
3. Z. Li, Y. Cao, P. Zhan, C. Pannecouque, J. Balzarini, E. D. Clercq, X. Liu, *Letters in Drug Design & Discovery*, **10,** 27 (2013).
4. K. Rakesh, M. Shahar Yar, B. Srivastava, A. K. Rai, *Der Pharma Chemica*, **6**, 137 (2014).
5. N. S. Dighe, R. B. Saudagar, D. A. Jain, *Bulgarian Chemical Communications*, **46**, 85 (2014).
6. G. R. Kokil, P. V. Rewatkarb, S. Gosainc, S. Aggarwalc, A. Vermac, A. Kalrac, S. Thareja, *Letters in Drug Design & Discovery*,**7**, 46 (2010).
7. A. Martin, R. Martin, *Int. J. Life Sc. Bt & Pharm. Res*., **3**, 323 (2014).
8. V. N. R Desabattina, R. G. P. Aluru, S. Y. Narasimha, R. R. Dharmapuri, R. L. Rao. *J. App. Pharm*., **6**, 1 (2014).
9. R. Kumar, M. Shahar Yar, S. Chaturvedi, A.Srivastava, *Int. J. Pharm Tech Res.*, **5,** 1844 (2013).
10. A. Reddy, S. G. Kini, M. Mubeen, *Der Pharma Chemica*, **5**, 259 (2013).
11. B. Andrews, A. Mansur, *Der Pharma Chemica*, **6**, 162 (2014).
12. B. Anil Reddy, *E-Journal of Chemistry*, **7**, 222 (2010).
13. R. Katritzky, S. Rachwal, B. Rachwal, *J. Chem. Soc. Perkin Trans. I.*,**1**, 805 (1987).
14. F. Collino, S. Volpe, *Boll. Chim. Farm*., **121**, 328 (1982).

15. J. Cruz, E. Gracia-Ochoa, M. Castro, *J. Electrochem. Soc*.,**150**, B26 (2003) .
16. M. A. Quraishi, D. Jamal, *J. Am. Oil Chem. Soc*.,**77,** 1107 (2000).
17. S. S. Panda, S. C. Jain, *Med. Chem. Res*., **23**, 848 (2014).
18. F. Ding, J. Guo, W. Song, W. Hu, Z. Li, *Chemistry and Ecology*, **27**, 359 (2011).
19. O. Adebimpe, R. C. Dash, M. E. S. Soliman, *Letters in Drug Design & Discovery*, **11**, 618 (2014).
20. M. Polyakova, L. Mei Jin, K. Ho Row, *Bull. Korean Chem. Soc*., **27**, 211 (2006).
21. V. Dimova, K. Colanceska Ragenovic, V. Kakurinov, *Int. J. Mol. Sci.*, **7**, 119 (2006).
22. V. Dimova, N. Perisic-Janjic, *Organic Chemistry, An Indian Journal*, **3**, 51 (2007).
23. V. Dimova, N. Perisic-Janjic, *Maced. J Chem. Chem. Eng*., **28**, 79 (2009).
24. V. Dimova, PhD Thesis, University Ss. Cyril and Methodius, Skopje, Macedonia, 2006.
25. S. Rollas, N. Kalyoncuoglu, D. Sür-Altiner,Y. Yegenoglu, *Pharmazie*, **48**, 308 (1993).
26. C. Hansch, A. Leo, Exploring QSAR: Fundamentals and Applications in Chemistry and Biology, American Chemical Society. Washington DC (1995).
27. M. Karelson, V. Lobanov, A. Katritzky, *Chem. Rev*., **96**, 1027 (1996).
28. http://www.molinspiration.com.
29. R. Todeschini, V. Consonni, Handbook of molecular descriptors, WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, 2000.
30. T. Sotomatsu, M. Shigemura, *Bull. Chem. Soc. Jpn*., **65**, 3157 (1992).
31. STATISTICA program package http://www.statsoft.com
32. A. Thakur, *ARKIVOC*, (**xiv**). 49 (2005).
33. R. Veerasamy, H. Rajak, A. Jain, S. Sivadasan, C. P. Varghese, R. K. Agrawal, *International Journal of Drug Design and Discovery*, **2**, 511 (2011).
34. L. Pogliani, *J. Phys. Chem.* **100**, 18065 (1996).
35. J. Singh, B. Shaik, S. Singh, V. K. Agrawal, P. V. Khadikar, O. Deeb, C. T. Supuran, *Chem. Biol. Drug Des.*, **71**, 244 (2008).

QSAR АНАЛИЗ НА N[1]-ЗАМЕСТЕНИ 1,2,4-ТРИАЗОЛИ СРЕЩУ *Escherichia coli*

В. Димова[1 *], И. Йорданов[1], Л. Димитров[2]

[1]*Факултет по технология и металургия, Университет „Св. Св. Кирил и Методий“, ул. Руджер Бошкович 16, 1000 Скопие, Република Македония*

[2]*Институт по земеделие, Университет „Св. Св. Кирил и Методий“, ул. 16-та Македонска бригада 3а, 1000 Скопие, Република Македония*

Получена на 12 Декември 2014 г.; ревизирана на 26 февруари 2016 г.

(Резюме)

Бе направен QSAR анализ на серия от предварително синтезирани N[1]-заместени 1,2,4-триазолови производни тествани за инхибиторна активност по отношение на растеж на *Escherichia coli*, като се използва компютеризирана схема за множествена регресия. Използвайки подхода на Hansch и Free – Willson, приноса за активността на аминометил / аминоетил заместител и ароматен / хетероароматният пръстен се определя от получените корелационни уравнения. В съответствие със статистическите параметри ($R = 0{,}8729$; $R^2_{adj} = 0.6430$; $Sd = 0.2983$; $Q^2 = 0{,}7548$ и $PRESS / SSY = 0{,}2452$), двупараметричен модел, който включва R и L, е избран като най-добър, за определяне на активността на избраните триазоловите производни срещу *E. colli*. Разширяването на изследваната система: подмножеството B (аминометилова група заменена с аминоетилова група) и подмножеството C (ароматен пръстен заменен с хетероароматен пръстен), не води до получаване на статистически значими QSAR модели.