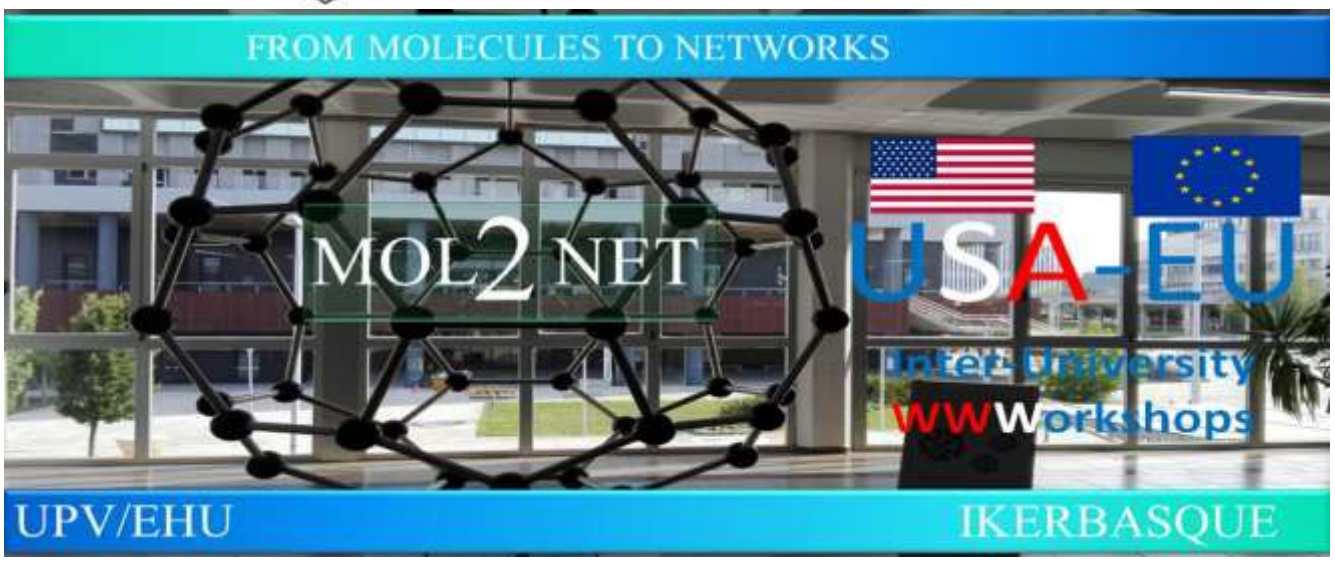
MOL2NET, 2021, 7, ISSN: 2624-5078
https://mol2net-07.sciforum.net/

MOL2NET, International Conference Series on Multidisciplinary Sciences

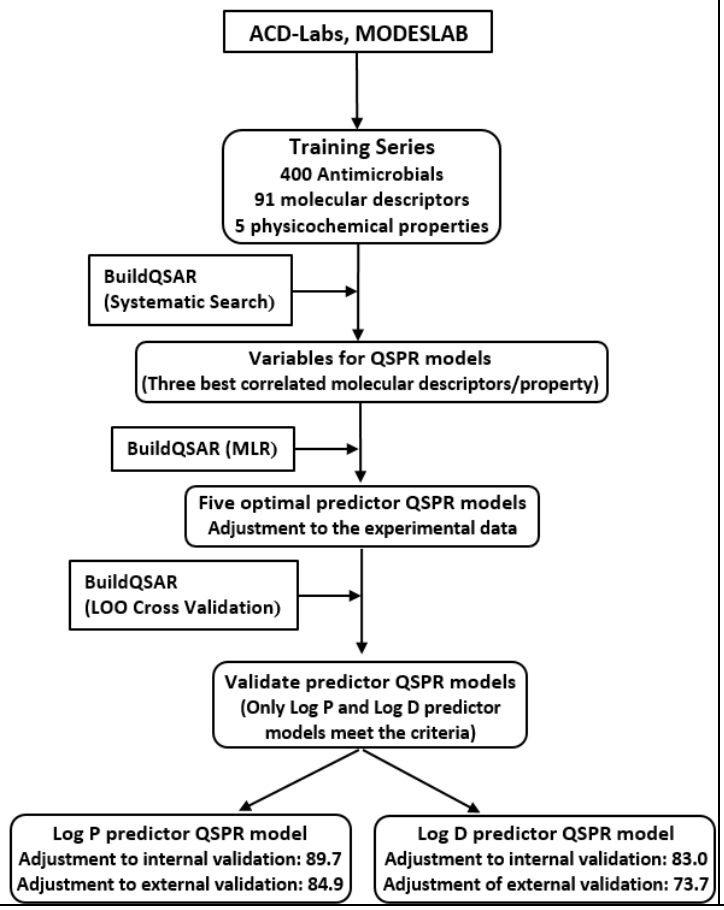


Obtaining QSPR models for the prediction of physicochemical properties of topical antimicrobials

Tran Le Quan [a], Juan Carlos Polo Vega [a], Luis A. Torres Gómez [a]

[a] Department of Pharmacy, Institute of Pharmacy and Food Sciences, University of Havana, Cuba

Graphical Abstract



Abstract

The traditional form of development and investigation off the antimicrobial has been resulting inefficient according to the delay of the new candidates discovery in the last years. Several limitations have been demonstrated, such as the long time invested, the expensive experimental trials or the errors in the manipulation of the researcher. To solve this problem, the application of computational methods in the design of drugs raised as a promised alternative. Specifically, the QSPR studies are oriented to determine the functions that capable to predict a particular property of a compound, using the information contained in their molecular descriptors. This strategy allowed analyzing a great quantity of molecules in a minor time and with less resources. Five specific models were defined in the present work in order to predict the interested physicochemical properties (aqueous solubility (S), partition coefficient (P), distribution constant (D), acid dissociation constant (K_a) and superficial tension (σ)) for the external use only of a series of 400 antimicrobial compounds, with simplified representations, physicochemical properties and

	molecular descriptors were obtained through ACD-Labs and MODESLAB software. After an exhausted validation, the specific models of Log P and Log D demonstrated a better prediction capacity with the standard errors of estimate for the specific functions were inferior or close to the logarithmic unit. These results suggested their employment in the design and development of antimicrobials for topical use.

INTRODUCTION

Throughout history, antimicrobials have become an essential tool for humans in the battle against diseases caused by microorganisms because of numerous benefits such as the simplicity of production and stograge condition, the convenience and quickness of their use and a broad spectrum of action with minimal toxicity. However, multiple factors such as the execssive and incorrect use of antimicrobials or the lack of a complete education on drugs rational use, all this factors have caused serious consequences and among them, the phenomenon called antibiotic resistance, recognized widely as one of the main problems related to the use of medicines.[1,2]

Traditional research methods have demosntrated their inefficacy by not being able to give an adequate response to this situation, thus it is pertinent to establish a new series of methods which allow to boost the research and development rate in this sector. In this sense, computational methods emerge as a promising tool for molecular design of new candidates of pharmaceutical interest, allowing to analyze a large number of previously selected compounds in a reduced time with minimized resources access. Technological evolution has given place to the born of powerful computer software of outstanding utilities in the construction of predictive models of convinent interpretation and relative simplicity based on the quantitative dependence between property of interest and molecular structure (QSPR).[3-5]

In the particular case of antimicrobials for topical use, the following physicochemical mesurements compose strong influence in their pharmacological performance, which are solubility (S), partition coefficient (P), distribution constant (D), acid dissociation constant (K_a) and the superficial tension (σ). A profound knowledge of the relation between these properties and chemical structure of interest is essential not only to successfully develop a new pharmaceutical candidate, but also to enhance the behavior of already existed molecules.[6,7]

Based on the potential advantages, this work aims to obtain general QSPR models for the prediction of five mentioned physicochemical properties of antimicrobials for topical use.

MATERIALS AND METHODS

Construcción of training series. The training series used 400 compounds organized into ten families according to their specific action, with recognized antimicrobial activity. In each group, 40 representative compunos from each family were included.[1,8]

The ACD-Labs software was used to obtain the simplified molecular structures, in form of SMILES code, of each compund of the training series and the experimental values of the evaluated physicochemical properties.

From the derived SMILES CODES, using the TOPS-MODE approach of the MODESLAB software, a set of molecular descriptors (DM) that weight the structural properties related to the modeled physicochemical properties was calculated: bond distance (Std), dipole moment (Dip), hydrophobicity (Hyd), polarizability (Pol), Van der Waals radius (Van) and atomic weight (Ato). As a result, a matrix was formed with the spectral moments from μ0 to μ15 by each graph.[9,10]

Construction of QSPR predictive model. The Systematic Search method of the BuildQSAR Software was engaged to select, from the 91 molecular descriptors calculated for each compound, those with the greatest capacity to structure as independent variables of an efficient QSPR model. The Multiple Linear Regression analysis (MLR) offered by the BuildQSAR software was used to optimize the initial QSPR models. By eliminating the atipical compounds, analyzing the significance of the slopes and compliance with the orthogonality principle, the final predictive QSPR model was considered if following statistical measurements are satisfied: multiple correlation R ($R > 0.6$), coefficient of determination R^2 ($R^2 > 0.5$), standard error of the estimate s ($s < 1$) and coefficient F of the test ANOVA ($F >> 1$ with $p < 0.05$).[5,11-13]

Validation off QSPR predictive model. For the internal and exxternal validation of the obtained model, the LOO (Leave-one-out). In the internal validation, to evaluate the robustness and predictive power of a QSPR model, following conditions of statistical excellence need to be satisfied: coefficient of cross-correlation Q^2 ($Q^2 > 0.5$) and the residual predictive summary of the standard squared deviation S_{press} ($S_{press} < 0.3$). The external evaluation was carried out with a test that included 40 new compounds with antimicrobial activity similar to that compounds in the training series. As criteria whose satisfaction ensures a predictive and reliable QSPR model, following requirements for statistical quality were considered: predictive determination coefficient R^2_{pred} should be greater than 0.6 and the difference with R^2 must be less than 0.3; the standard error of the s_{pred} is less than unity and less than the experimental error.[14-16]

RESULTS AND DISCUSSION

The predictive capacity off a QSPR model depends immensely on the characteristics of the compounds of the training series. 400 antimicrobial compounds included in the training series represent ten pharmacological groups which corresponde to the polyfunctionality that distinguishes the molecules of interest. Table I shows the detailed classification.

Table I. Antimicrobial families represented in the training series.

Antimicrobial families	
Antibacterial	Bactericide
Antifungal	Fungicide
Antimalarial	Anti-infective
Antihelmintic	Pesticide
Antineoplastic	Antiseptic

The calculations was caried out using the MODESLAB software generated 91 molecular descriptors as independent variables for each compound in the training series, with the corresponding parameters related to the estimated properties: bond distance (Std), dipole moment (Dip), hydrophobicity (Hyd), polarizability (Pol), Van der Waals radius (Van) and atomic weight (Ato).

The inclusion of a large number of independent variables in a QSPR function may hamper its explanatory power. For this reason, it is recommended to use an adequadate number of descriptors, with high statistical quality and relatively easy to interpret. The BuildQSAR software performed the selection of variables using the Systematic Search method, to sort out the three best molecular descriptors to include as independent variables for each model, according to the four considered esential criteria for the evaluation of the candidate: (i) multiple correlation R ($R > 0.6$); (ii) standard error of the estimate s ($s < 1$); (iii) coefficient F of the test ANOVA ($F >> 1$ with $p < 0.05$).

To ensure the normality of the distrution of variables, provide stability to the regressores and reduce the atypical observations, logarithmic transformation of solubility variable (Log Sol) and inversion of acid dissociation constant variable (pK_a^{-1}) were triggered.

Table II shows the statistical parameters.

Table II. Summary of the selection of independent variables. Source: BuildQSAR.

	X1	X2	X3	R	s	F
Log Sol	**μ(Hyd)1**	**μ(Pol)4**	**μ(Pol)15**	0.612	1.475	79.211
Log P	**μ(Std)1**	**μ(Hyd)1**	**μ(Van)14**	0.623	2.66	49.931
Log D	**μ0**	**μ(Hyd)1**	**μ(Van)13**	0.708	1.911	132.91
pK_a^{-1}	**μ(Dip)1**	**μ(Dip)15**	**μ(Hyd)1**	0.551	1.557	57.63
σ	**μ(Hyd)1**	**μ(Pol)4**	**μ(Van)14**	0.599	13.53	73.872

In order to build the predction models for the physicochemical properties of interest, the Multiple Linear Regression analysis (MLR) offered by the BuildQSAR software was proceeded to perform the predictive mathematical functions from the obtained independent variables. Five QSPR models corresponding to the five properties of interest: solubility, partition constant, distribution constant, acid dissociation constant and the superficial tension.

As an example, the construction of the QSPR model of partition coefficient (Log P) was highlighted by a brief qualitative analysis of its optimization.

Analysis of predictive QSPR model of the partition constant (Log P)

Initial modelo (M1):

log P = 0.0116 (±0.0217) μ(Std)1 + 0.9061 (±0.1307) μ(Hyd)1 + 0.00001 (±0.00001) μ(Van)14 - 0.0377 (±0.6307)
Regression statistics: R = 0.621; R^2= 0.381; s = 2.658; F = 82.955 (p<0.0001)

According to the values of regression statistics, the model M1 fits moderately to the experimental data with the value of R, greater than 0.6, which indicates the ability to explain more than 60% of the training series. Nevertheless, value of R^2 and s do not satisfy the established criteria reflecting a poor predictive capacity and thus this model is susceptible to optimization.

By excluding the atypical observations (outiers), the model M2 is obtained.

log P = -0.0052 (±0.0074) μ(Std)1 + 0.9743 (±0.0419) μ(Hyd)1 + 0.00001 (±0.00001) μ(Van)14 - 0.1579 (±0.1978)
Regression statistics: R = 0.95; R^2= 0.901; s = 0.695; F = 912.81 (p<0.0001)

The elimination of the outliers leads to an considerable increase of the value of R and F (0.95 and 912.81 respectively), whereas the standard error s go down sharply (0.695), attaining the very good fit of this model to the experimental data. On the other hand, due to the simplicity of model M2 whose represented by only three independent variables, its interpretation and applicability were facilitated as well.

The MLR analysis includes the t-test of the significance of the slopes and the results demonstrate the significant constribution of the selected variables to the variation of the partition coefficient.

Table III. Coefficients of model M2 and t-test of the significance of the slopes. Source: BuildQSAR/MLR.

	Coeficiente	St. Desv.	95% Conf	t-ratio	p	Comentario
Constant	0.1579	0.0989	0.1978	1.5963	0.1115	No Significant
μ(Std)1	-0.0052	0.0037	0.0074	-1.4172	0.1575	No Significant
μ(Hyd)1	0.9743	0.0210	0.0419	46.4866	0.0000	Significant
μ(Van)14	0.00001	0.0000	0.0000	8.3453	0.0000	Significant

The values of p indicate that only the variation of μ(Hyd)1 and μ(Van)14 contributes significantly to the variation of the partition coefficient of the compounds in the training series. Considering that the new model is composed only by μ(Hyd)1 and μ(Van)14 as independent variables, after eliminating the atypical variables, the model M3 is obtained.

Optimized model (M3):

log P = 0.9606 (±0.0392) μ(Hyd)1 + 0.00001 (±0.00001) μ(Van)14 - 0.0601 (±0.1429)
Regression statistics: R = 0.94; R^2= 0.898; s = 0.703; F = 1335.01 (p<0.0001)

The value of R, R^2 and s decrease slightly in respect to the previous model, but the value of F is increased considerably (1335.01), which implies that the this simplified model with only two molecular descriptors as independent variables, reaches an comparable adjustment to the experimental data with the model M2, which further facilitates its interpretation and applicability.

Table IV. Partial correlation coefficients between μ(Hyd)1 and μ(Van)14. Fuente: BuildQSAR/MLR

DM	μ(Hyd)1	μ(Van)14
μ(Hyd)1	1	0.142 (p>0.05)

The model M3 fulfills the compliance of the principle of orthogonality or independence between μ(Hyd)1 and μ(Van)14, as has been asumed in the MLR analysis.

Finally, Figure 1 indicates a good linear correlation between the experimental and calculated values of Log P, thus ensuring the reliability of the model M3.

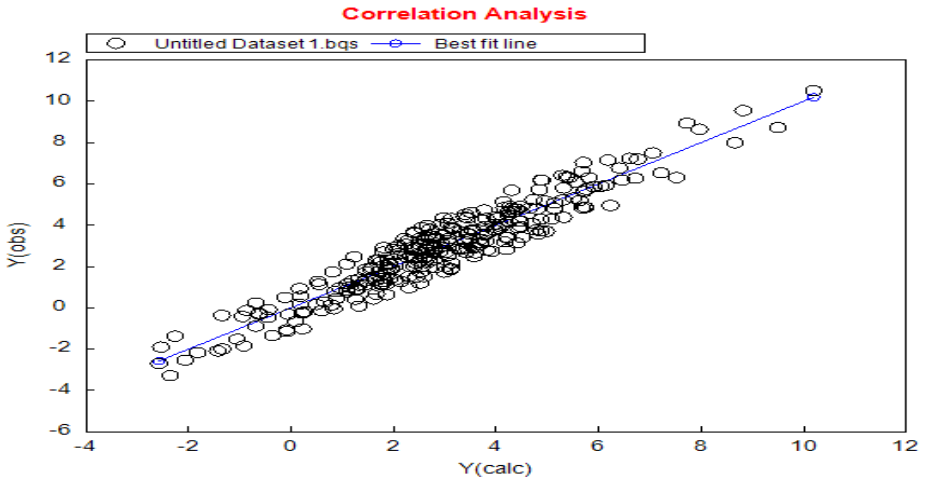


Figure 1. Linear correlation between the observed and calculated values of Log P from the model M3. Source: BuildQSAR/MLR

For these reasons, the model M3 is optimal for validation of its prediction of the dependency between the proposed property (partition coefficient) and the structure of the antimicrobials.

Following the same methodology, the predictive QSPR models of the remaining physicochemical properties of interest were obtained with summarized information in the table V. The models of Log Sol, Log P and Log D show better statistical quality with the error estandar of the estimate is less than 1, since all the parameters meet the established criteria, than the models of pK_a^{-1} and σ. But the rest of the statistical parameters indicate an acceptable fit to the experimental data for this type of model.

Table V. QSPR models for each selected physicochemical property and statistical parameters of MLR. Source: BuildQSAR/MLR.

Model	QSPR model and regression statistics
M3	**log P** = 0.9606 (± 0.0392) μ(Hyd)1 + 0.00001 (± 0.00001) μ(Van)14 - 0.0601 (± 0.1429) R = 0.94; R^2 = 0.898; s = 0.703; F = 1335.01 (p<0.0001)
M4	**log Sol** = -0.3605 (± 0.0351) μ(Hyd)1 - 0.0019 (± 0.0001) μ(Pol)4 + 0.00001 (± 0.00001) μ(Pol)15 + 2.3697 (± 0.1577) R = 0.953; R^2 = 0.906; s = 0.483; F = 761.82 (p<0.0001)
M5	**log D** = 0.9029 (± 0.0512) μ(Hyd)1 + 0.00001 (± 0.00001) μ(Van)13 – 0.0096 (± 0.1830) R = 0.91; R^2 = 0.832; s = 0.857; F = 673.904 (p<0.0001)
M6	**pK_a^{-1}** = -0.0477 (± 0.0139) μ(Dip)1 - 0.5083 (± 0.0523) μ(Hyd)1 + 1.241 (± 0.2369) R = 0.758; R^2 = 0.572; s = 1.054; F = 223.69 (p<0.0001)
M7	**σ** = -4.3321 (± 0.4202) μ(Hyd)1 + 0.0185 (± 0.0017) μ(Pol)4 - 0.00001 (± 0.00001) μ(Van)14 + 41.7459 (± 1.8972) R = 0.808; R^2 = 0.649; s = 6.728; F = 200.25 (p<0.0001)

According to the information in table V, all the five QSPR models can be subjected to the validation procedure for their further application in the research and development of new topical antimicrobials.

Internal validation of the QSPR models

Table VI shows the statistical results obtanied from the internal validation of the obtained QSPR models:

Table VI. Internal validation statistical parameters obtained by using LOO method. Source: BuildQSAR/LOO

Model	Property	Q^2	R^2-Q^2	S_{mode}	S_{press}	S_{dep}
M3	**P**	0.897	0.001	0.703	0.711	0.709
M4	**Sol**	0.905	0.001	0.483	0.49	0.487
M5	**D**	0.83	0.002	0.857	0.867	0.863
M6	**pK_a**	0.566	0.006	1.054	1.064	1.061
M7	**σ**	0.644	0.005	6.728	6.821	6.789

As can be seen, both the parameters Q^2 and (R^2 - Q^2) of five models meet the validation criteria demonstrating an appropiate level of stability when internal compounds are excluded for the construction of predictive models. However the value of S_{dep} and S_{press} of the models of M6 (pK_a^{-1}) and M7 (σ), although they are similar to those of the obtained models, is higher than the logarithmic unit and thus not satified the established criteria for the standard error of the estimate. Therefore, this measure of internal consistency is not enough to suggest the use of these functions as useful prediction tools.

External validation of the QSPR models

For the external validation of the predictive capacity, an external series was built from 40 compounds obtained from the library of the same ACD-Labs software. Table VII shows the stastitical results of the external validation.

Table VII. External validation statistical parameters obtained by using LOO method. Source: BuildQSAR/LOO

Model	Property	R^2_{pred}	R^2-R^2_{pred}	S_{pred}	S_{mode}
M3	**P**	0.849	0.052	0.646	0.703
M4	**Sol**	0.695	0.201	0.864	0.483
M5	**D**	0.737	0.095	0.893	0.857
M6	**pK_a**	0.594	-0.022	1.035	1.054
M7	**σ**	0.243	0.396	8.426	6.728

The model M4, although does not exhibit such a low value of R^2_{pred} but presents a higher value of s_{pred} than the experimental s_{mode}. On the other hand, the models M6 and M7 do not satisfy any of the statistical criteria of external validation. For this reason these three models M4, M6, M7 are not suitable for the prediction of corresponding physicochemical properties. Nevertheless, these results could be used as references for other studies of this area of research by using other types of molecular descriptor or criteria for selection and optimization of variables.

The immense structural diversity of the training series, where the majority are ionizable compounds, makes it difficult to estimate the aqueous solubility by a general predictive QSPR model. Beside of the close proportionality between the aqueous solubility and the acid dissociation constant, the influence of the pH-dependence of these properties is reflected in the different contribution of the ionizable degree of each compound, which prove the poor predictive capacity of models M4 and M6. Abnormality of the distribution levels of polarity also posiblely could be a cause which leads to a statitical dissatisfaction of predictive power of model M7. All this explanations suggest to perform for each anitimicrobial family, a specific predictive QSPR model of such high-variable properties.

Notwithstanding the poor statistical results of M4, M6 and M7, the models M3 (Log P) and M5 (Log D) are those that meet all the criteria of statistical excellence corresponding to the partition coeffcient and distribution constant.

CONCLUSIONES

In summary, it is possible to guarantee the use of the optimal QSPR models of the partition coeficient (M3) and the distribution constant (M5) as a reliable predictive tool for this important properties in the development of new antimicrobial candidate for topical use.

REFERENCES

1. Lesher J, McConnell-Woody C. Fármacos antimicrobianos. En: Bolognia JL, Jorizzo JL, Rapini RP, editores. Dermatología. 1ª ed. Madrid: Elsevier. 2004: 2007-31.
2. Kukso F. Para 2050 la resistencia a los antibióticos será la principal causa de muerte. Scientific American – Español Revista. 2016.
3. J.C. Escalona, R. Carrasco y J. A. Padrón. Introducción al diseño racional de farmacos. Ciudad de La Habana : Editorial Universitaria, 2008: 36.
4. Nantasenamat C, Isarankura-Na-Ayudhya C, Prachayasittikul V. Advances in computational methods to predict the biological activity of compounds. Expert Opinion on Drug Discovery. 2010; 5(7): 633–54.
5. Polishchuk PG, Kuzmin VA, Artemenko AG, Muratov EN. Universal Approach for Structural Interpretation of QSAR/QSPR Models. Mol. Inf. 2013; 32: 843 – 853.
6. J.I.Porras-Luque. Topical antimicrobial agents in dermatology. 2017; 98(1): 29-39.
7. Torres LA. Relationship studies structure property of pharmaceutical interest. QSPR methodsapplied to pharmaceutical analysis. Spanish Academic Editorial. 2016.

8. Tam LM, Torres LA, Polo JC. Machin L. A QSPR model for the prediction of the partition coefficient of organic compounds of pharmaceutical interest. MOL2NET, 2019, 5, ISSN: 2624-5078. http://sciforum.net.
9. Garcia D, Torres LA, Polo JC, Machín L. Obtaining a computer-assisted QSAR model for the prediction of anti-inflammatory activity. MOL2NET, 2019, 5, ISSN: 2624-5078. http://sciforum.net.
10. Dearden JC. The use of topological indices in QSAR and QSPR modeling. Roy, K, ed. Advances inQSAR Modeling, Challenges and Advances in Computational Chemistry and Physics. Springer International Publishing. 2017.
11. Torres LA, Polo JC, Guevara YC, Mutsauri TB. Multivariate classification of a series of organic compounds of pharmaceutical interest using MODESLAB methodology. MOL2NET, 2017, 3, doi:10.3390/mol2net-03. http://sciforum.net.
12. Polo JC. Apuntes sobre bioestadística aplicada a las ciencias farmacéuticas. Vol I. Universidad de La Habana, Instituto de Farmacia y Alimentos; 2017: 295-306.
13. Paul L, Jenny C. Systematic searching: practical ideas for improving results. Facet Publishing. 2019.
14. Chirico N. Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection. Journal of Chemical Informationand Modeling. 2012; 52(8): 2044–58.
15. Veerasamy R, Rajak H, Jain A, Sivadasan S, Varghese CP. Validation of QSAR Models - Strategiesand Importance. International Journal of Drug Design and Discovery. 2011; 2(3): 511-19
16. Machín L. Predicción de la solubilidad acuosa pH-dependiente de compuesto de interés farmacéutico con comportamiento químico diferente. Instituto de Farmacia y Alimentos, UH. 2018