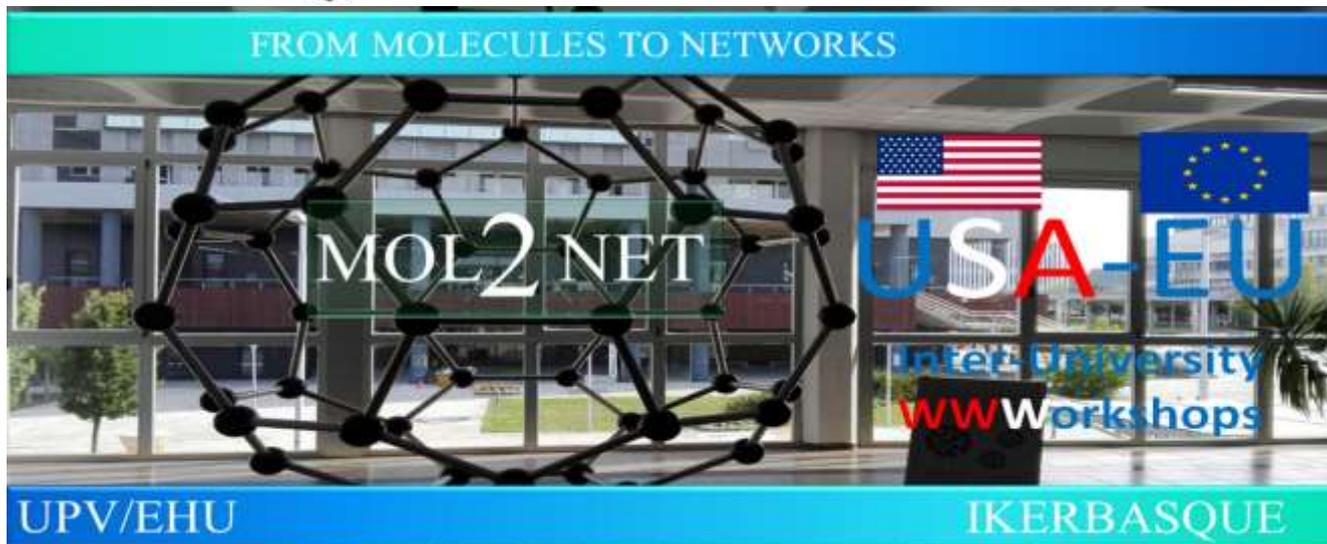




MOL2NET, International Conference Series on Multidisciplinary Sciences

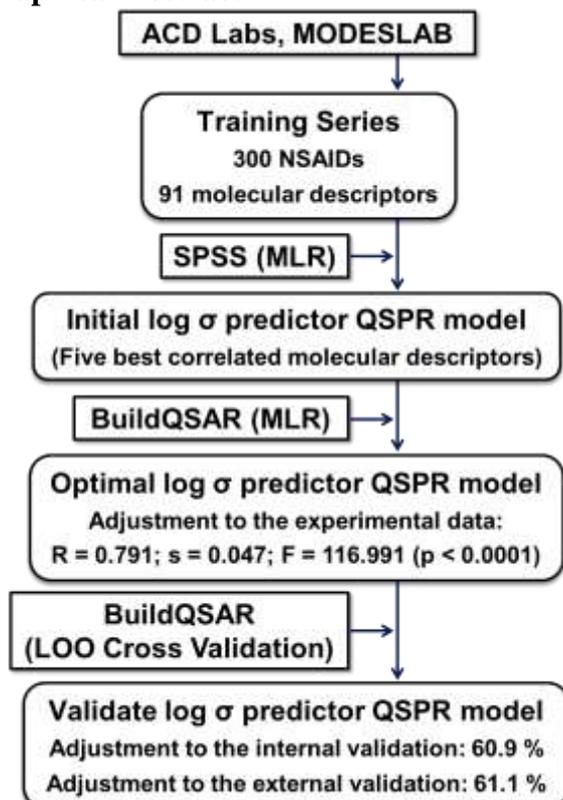


A QSPR model for the prediction of the surface tension of NSAIDs

Eliani Muñoz Alcuria ^a, Juan Carlos Polo Vega ^a, Luis A. Torres Gómez ^a, Laura Machín Galarza ^a

^a Department of Pharmacy, Institute of Pharmacy and Food Sciences, University of Havana, Cuba.

.Graphical Abstract



Abstract.

Interfaces are crucially important in pharmaceuticals, biotechnology and biomedicine. There is a growing need for specific interfacial consideration that is using routinely to solve pharmaceutical problems. In order to meet manufacturing challenges and develop new better performing pharmaceutical products with improved qualities, knowledge of surface tension (σ) is of utmost importance. The experimental determination of this property has several limitations, such as the high time invested and the consumption of considerable amounts of sample. In the recent years, constant increase in the performance of hardware and software transformed quantitative structure property relationship (QSPR) into powerful and widely used model for the prediction of many biological, toxicological and physicochemical properties in the field of medicinal chemistry.

The aim of the present work was to find a QSPR model for prediction of surface tension of non-steroidal anti-inflammatory drugs (NSAIDs). To do this, a training series, consisting of 300 compounds, was constructed. By the ACD-Labs and MODESLAB, the simplified representation, surface tension value and molecular descriptors of each compound in the series were obtained.

An initial mathematical model of $\log \sigma$, obtained using the Multiple Lineal Regression method (MLR) of SPSS, was optimized and validate through BuildQSAR program. The final model showed a good predictive power, results which suggest their use as part of the design and development of NSAIDs.

Introduction

The normal inflammatory response is an acute process that resolves after removal of the initial stimulus. Diseases of inflammation and immunity can occur either when the normal inflammatory response progresses to chronic inflammation, from an inappropriate response to a long-term stimulus, or because the causative agent is not eliminated. Non-steroidal anti-inflammatory drugs, known as NSAIDs, are effective in reducing inflammation, as well as pain and fever, inhibiting the action of cyclooxygenases, enzymes that participate in the biosynthesis of the prostaglandins that cause these symptoms, but present numerous adverse effects. One of the challenges of current science is the search for new leading compounds that have anti-inflammatory activity and lack the characteristic adverse effects of these drugs.¹⁻³

Computer-aided drug design is an alternative for the synthesis and evaluation of new candidates, without the high cost and time consuming experimental trials that characterize traditional methods. In the recent years, constant increase in the performance of hardware and software transformed quantitative structure property relationship (QSPR) into powerful and widely used model for the prediction of many biological, toxicological and physicochemical properties in the field of medicinal chemistry.⁴⁻⁸

The role of interfaces are crucially important in pharmaceuticals, biotechnology and biomedicine. Due to this bigger interest, there is a growing need for specific interfacial consideration that is using routinely to solve pharmaceutical problems. In order to meet manufacturing challenges and develop new better performing pharmaceutical products with improved qualities, knowledge of surface tension (σ) is of utmost importance.⁹

Based on these premises, the present work seeks to obtain QSPR models from which it is possible efficiently predict the surface tension values of new NSAID candidates.

Materials and Methods

Construction of training series. The training series used included 300 compounds of interest, representative of the anti-inflammatory, analgesic, antipyretic or combined actions that NSAIDs may present.^{2, 3, 10}

The ACD-Labs 10 computer program was used to represent the simplified molecular structures (SMILES) of each of the compounds in the training series and to obtain the experimental values of surface tension. The primary data were transformed to their logarithmic values ($\log \sigma$) in order to achieve a better fit of the corresponding predictive models.

From the SMILES codes, a set of molecular descriptors that weight the structural properties related to the surface activity of the molecules was calculated, using the TOPS-MODE approach of the MODESLAB software: bond distance (Std), dipole moment (Dip), hydrophobicity (Hyd), polarizability (Pol), Van der Waals radius (Van) and atomic weight (Ato). As a result, was obtained a matrix with the spectral moments from μ_0 to μ_{15} for each descriptor.¹¹

Construction of QSPR predictive model. The Linear Regression menu of the IBM SPSS Statistics 26 program was used to select, from the 91 molecular descriptors calculated for each compound, those with the greatest capacity to structure as independent variables, an initial mathematical model for predicting surface tension. The Multiple Linear Regression analysis (MLR) offered by the BuildQSAR software was used to optimize the initial mathematical model. The optimization process included the elimination of outliers, the analysis of the significance of the slopes and compliance with the orthogonality principle. The following minimum requirements for statistical quality were considered: multiple correlation coefficient 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 ($F \gg 1$ with $p < 0.05$).⁵⁻⁸

Validation of QSPR predictive model. For the internal and external validation of the obtained model, the LOO (Leave-one-out) method was used. External validation was carried out with a test series that included 30 new compounds with anti-inflammatory activity similar to that of the compounds in the training series. As criteria whose satisfaction ensures obtaining a predictive and reliable QSPR model, it was considered that the determination coefficient R^2_{pred} , should be greater than 0.6 and be analogous to the cross-correlation coefficient Q^2 . In addition, that the difference between R^2 and R^2_{pred} , must be less than 0.3 and that the standard error of the Spread estimate, is less than unity and less than the experimental error.¹²⁻¹⁴

Results and Discussion

The initial predictive model of $\log \sigma$ was obtained by applying the stepwise method of the Linear Regression menu of the SPSS software. The characteristics of this model, called Model 1, are summarized in tables I, II and III.

Table I. Summary of Model 1. Source: SPSS 26

Model	R	R ²	R ² adjusted	Standard error of the estimate (S)
1	0.743	0.552	0.544	0.05902

Predictors: (Constant), $\mu(\text{Hyd})^1$, $\mu(\text{Pol})^5$, $\mu(\text{Van})^7$, $\mu(\text{Hyd})^3$, $\mu(\text{Dip})^3$

Dependent variable: $\log \sigma$

Table II. ANOVA test for Model 1. Source: SPSS 26

ANOVA						
Model	Sum of squares	f. g.	Quadratic mean	F	Sig.	
1	Regression	1.261	5	0.252	72.385	0.000
	Residue	1.024	294	0.003		
	Total	2.285	299			

Table III. Model 1 coefficients. Source: SPSS 26

Model	Non-standardized coefficients		Standardized coefficients	T	Sig.	
	B	Standard error	Beta			
1	(Constant)	1.694	0.012		142.690	0.000
	$\mu(\text{Hyd})^1$	-0.055	0.004	-0.956	-12.368	0.000
	$\mu(\text{Pol})^5$	3.652E-5	0.000	0.729	7.749	0.000
	$\mu(\text{Van})^7$	-2.037E-6	0.000	-1.134	-8.607	0.000
	$\mu(\text{Hyd})^3$	0.003	0.001	0.836	4.007	0.000
	$\mu(\text{Dip})^3$	0.000	0.000	0.232	2.622	0.009

As can be seen in the previous tables, Model 1 is relatively simple, it is a function of only five predictor variables: $\mu(\text{Hyd})^1$, $\mu(\text{Pol})^5$, $\mu(\text{Van})^7$, $\mu(\text{Hyd})^3$ and $\mu(\text{Dip})^3$, a considerable reduction from the 91 molecular descriptors initially considered. Table III indicates that the five descriptors show significant slopes ($p < 0.05$), which shows their influence on the variation of the surface tension of the compounds of the training series.

Although this model satisfies the minimum requirements for statistical quality, it presents a moderate fit to the experimental data ($R = 0.743$; $R^2 = 0.552$), for which reason it was optimized using the BuildQSAR software.

Fifteen atypical cases were identified, corresponding to compounds whose $\log \sigma$ values differed by more than two standard deviations from the mean of the series. The elimination of these compounds led to the obtaining of Model 2, whose characteristics were summarized below.

Table IV. Summary of Model 2. Source: BuildQSAR

Model	R	R ²	Standard error of the estimate (S)	F (p)
2	0.8010	0.6413	0.046	99.7746 (0.0001)

Table V. Model 2 Coefficients. Source: BuildQSAR

	Coefficients	Dev. Std.	95%Conf	t-ratio	P	Commentary
Constant	1.6827	0.0098	0.0196	172.074	0.0000	Significant
Dip ³	0.0004	0.0001	0.0002	5.0263	0.0000	Significant
Hyd ¹	-0.0511	0.0038	0.0077	-13.34429	0.0086	Significant
Hyd ³	0.0019	0.0006	0.0011	3.4598	0.0006	Significant
Pol ⁵	0.0000	0.0000	0.0000	10.4914	0.0000	Significant
Van ⁷	0.0000	0.0000	0.0000	-10.6246	0.0000	Significant

Table VI. Model 2 Correlation Matrix. Source: BuildQSAR*

	Dip ³	Hyd ¹	Hyd ³	Pol ⁵	Van ⁷
Dip ³	1	0.210	0.802	0.707	0.867
Hyd ¹	0.210	1	38	0.595	0.333
Hyd ³	0.8022	0.638	1	0.906	0.896
Pol ⁵	0.707	0.595	0.906	1	0.773
Van ⁷	0.8677	0.333	0.896	0.773	1

* Shows the Pearson correlation coefficients between independent variables.

The linear fit of Model 2 is superior to that of Model 1 (See Tables I and IV) and the slopes of the five predictor variables that comprise it are significant (See Table V). However, in table VI it can be seen that the variable Hyd³ presents a high correlation with the other variables (correlation coefficients ≈ 1), which suggests that its elimination may contribute to optimizing Model 2. Removal of the variable Hyd³ led to Model 3 described below.

Table VII. Summary of Model 3. Source: BuildQSAR

Model	R	R ²	Standard error of the estimate (S)	F (p)
2	0.791	0.6257	0.047	116.991 (0.0001)

Table VIII. Model 3 coefficients. Source: BuildQSAR

	Coefficients	Dev. Std.	95%Conf	t-ratio	P	Commentary
Constant	1.6688	0.0087	0.0174	192.2609	0.0000	Significant
Dip ³	0.0006	0.0001	0.0002	6.8470	0.0000	Significant
Hyd ¹	-0.0386	0.0024	0.0048	-16.1123	0.0000	Significant
Pol ⁵	0.0000	0.0000	0.0000	14.8871	0.0000	Significant
Van ⁷	0.0000	0.0000	0.0000	-11.3945	0.0000	Significant

Table IX. Model 3 correlation matrix. Source: BuildQSAR*

	Dip ³	Hyd ¹	Pol ⁵	Van ⁷
Dip ³	1	0.197	0.706	0.868
Hyd ¹	0.197	1	0.559	0.316
Pol ⁵	0.706	0.559	1	0.773
Van ⁷	0.868	0.316	0.773	1

* Shows the Pearson correlation coefficients between independent variables.

As the regression parameters indicate, the statistical quality of Model 3 is superior to that of Model 1 and similar to that of Model 2 (See Tables I, IV and VII). Furthermore, Model 3 is simpler and easier to interpret than the previous models, since it is only a function of four predictor variables, whose slopes were significant (See Table VIII). The scatter plot in Figure 1 shows a good fit between the observed (experimental) values and those calculated from Model 3.

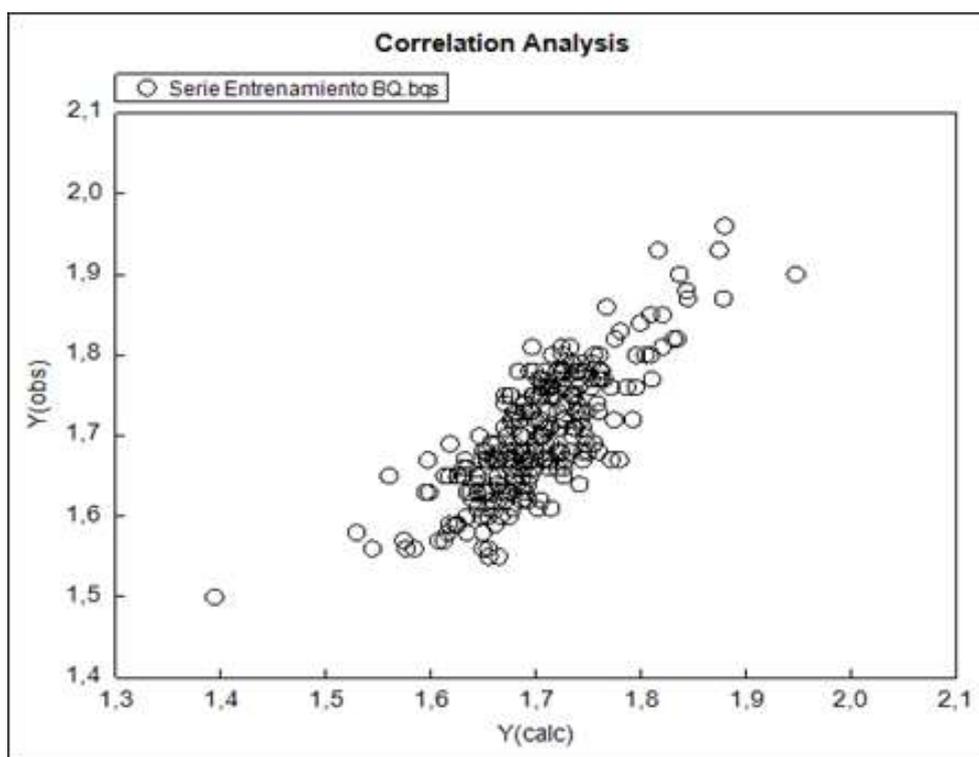


Figure 1. Scatter plot of Model 3. Source: BuildQSAR

Table IX indicates that in Model 3 relatively high correlations persist between the variables Pol⁵ and Van⁷ and the rest of the independent variables. However, the elimination of these variables causes a significant decrease in the statistical quality of the model. Consequently, Model 3 was considered the optimal as a linear expression of the relationship between log σ and the structure of the compounds that make up the training series.

Table X shows the statistical results obtained from the internal validation of Model 3, carried out using the LOO method.

Table X. Internal validation statistics for Model 3 (LOO method). Source: BuildQSAR

Property	Q^2	$R^2 - Q^2$	S_{mode}	S_{press}	S_{dep}
$\log \sigma$	0.609	0.017	0.047	0.048	0.048

As seen in Table X, Model 3 satisfies the statistical excellence requirements assumed for internal validation: $Q^2 > 0.5$; $(R^2 - Q^2) < 0.3$ and S_{dep} and S_{press} values are similar to those obtained for the models built with the intact training series, S_{mode} .

For external validation of the $\log \sigma$ predictive model, the test series of 30 NSAIDs compounds obtained from the library of the same ACD-Labs 10 software was used. The statistical parameters resulting from this procedure was summarized in Table XI.

Table XI. External validation statistics for Model 3 (LOO method). Source: BuildQSAR

Property	R^2_{pred}	$R^2 - R^2_{pred}$	S_{pred}	S_{mode}
$\log \sigma$	0.6110	0.0147	0.0399	0.0470

The good predictive capacity of a function is associated with a difference between R^2 and R^2_{pred} not greater than 0.3 while the S_{pred} values must be less than the experimental error S_{mode} . The values obtained after external validation indicate that Model 3 satisfies these requirements and explains more than 60% of the variability of the predicted property in the compounds of the test series, therefore it exhibits a good predictive capacity concerning surface tension.

Conclusions

In summary, according to the results obtained, it is possible to conclude that the optimal QSPR model of the surface tension of NSAIDs (Model 3) meets the criteria of statistical excellence, which guarantees the reliability of its use as a predictive tool for this important property in the development of new NSAIDs.

References

1. Carda M. Inflamación: síntesis de antiinflamatorios. Máster en Química Aplicada y Farmacología. Universidad Jaime I. 2011.
2. David E. et al. Principles of Pharmacology: The Pathophysiologic Basic of Drug Therapy, 3ra Edition. Chapter 41: Principles of Inflammation and the Immune System. 2012; 736-74.
3. Rang H.P, Dale M.M, Ritter J.M, Flower R.J, Henderson G. Antiinflamatorios e inmunodepresores. Churchill Livingstone. Rang y Dale, Farmacología. 9ª ed. Barcelona: Elsevier. 2019; 349-356.
4. Lokendra O, Rachana S, Mukta B. Modern drug design with advancement in QSAR: A review, International Journal of Research in BioSciences. 2013; 2(1): 1-12.
5. Patel H, Noolvi M, Sharma P, Jaiswal V, Bansal S. Quantitative structure - activity relationship (QSAR) studies as strategic approach in drug discovery. Med. Chem. Res. 2014.
6. Polishchuk PG, Kuzmin VA, Artemenko AG, Muratov EN. Universal Approach for Structural Interpretation of QSAR/QSPR Models. Mol. Inf. 2013; 32: 843 – 853.
7. Ojha K, Sharma R, Bhawsar M. Modern drug design with advancement in QSAR: A review. International Journal of Research in BioSciences. 2013; 2(1): 1-12. <http://www.ijrbs.in>.

8. Torres LA. Relationship studies structure property of pharmaceutical interest. QSPR methods applied to pharmaceutical analysis. Spanish Academic Editorial. 2016.
9. Fathi-Azarbayjani A, Jouyban A, Chan S. Impact of Surface Tension in Pharmaceutical Sciences. J Pharm Pharmaceut Sci. 2009; 12 (2): 218-228.
10. 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>.
11. Dearden JC. The use of topological indices in QSAR and QSPR modeling. Roy, K, ed. Advances in QSAR Modeling, Challenges and Advances in Computational Chemistry and Physics. Springer International Publishing. 2017.
12. 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 Information and Modeling. 2012; 52(8): 2044–58.
13. Tropsha A. Best Practices for QSAR Model Development, Validation and Exploitation. Molecular Informatics. 2010; 29(6–7): 476–488.
14. Veerasamy R, Rajak H, Jain A, Sivadasan S, Varghese CP. Validation of QSAR Models - Strategies and Importance. International Journal of Drug Design and Discovery. 2011; 2(3): 511-19