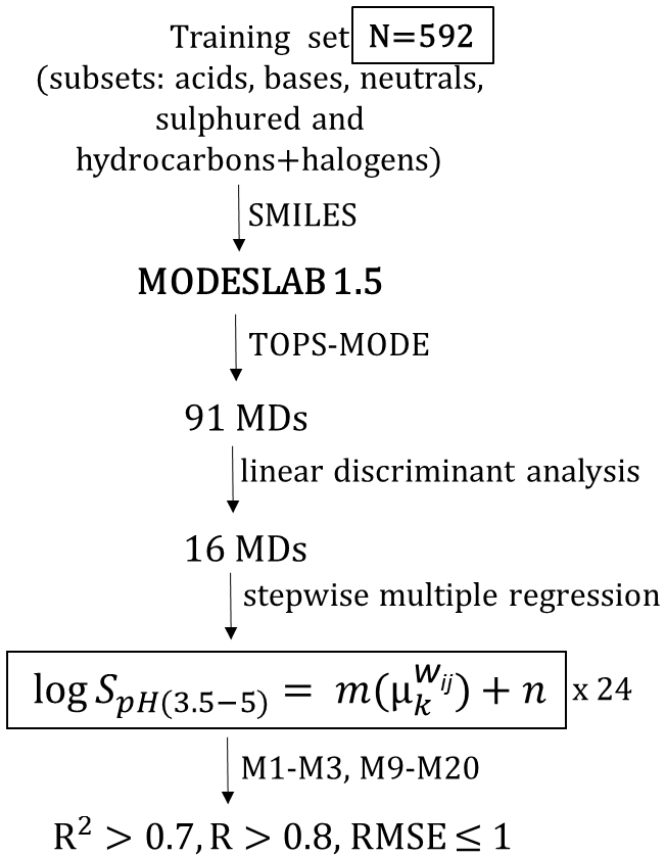
MOL2NET, International Conference Series on Multidisciplinary Sciences

Modeling aqueous solubility of druglike organic compounds in strongly acidic medium

Yeiny Cabrera Martínez [a], *Juan Carlos Polo Vega* [a],
Laura Machín Galarza (laura@ifal.uh.cu) [a], *Luis Alberto Torres Gómez* [a]

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

Graphical Abstract



MDs: molecular descriptors

Abstract.

Experimental determination of aqueous solubility is affected by long time invested and consumption of considerable amounts of sample. To solve these problems, QSPR studies have been applied, in order to establish quantitative relationships between the structure and some property of the molecule of interest, by means of a function capable of predicting a certain quality of a compound. Considering that solubility of orally-administered drugs is influenced by the strongly acid pH of the gastrointestinal tract, obtaining predictive models based on pH is of great interest, which was the purpose of the present study. Then, some computer programs were used: ACDLabs (construction of the training series), MODESLAB (calculation of molecular descriptors), IBM SPSS Statistics (data reduction) and BuildQSAR (obtaining and optimization of predictive models). Finally, 24 mathematical models (M) for the prediction of the aqueous solubility of organic compounds of pharmaceutical interest were defined, by ranges of pH (1-1.3, 1.4-1.7, 1.8-2.1 and 2.2-2.5) and by group of chemical behavior. The relative simplicity along with correlation coefficients and standard errors of estimate close to the unity, suggest the external evaluation of models M3-M5 and M11-M22 and their subsequent use as part of the design and development of orally-administered medications.

Introduction

Aqueous solubility (S) is a physical-chemical property that characterizes each compound and significantly influences the biological behavior of drugs, including their absorption and bioavailability. The relationship between pH and S was evidenced in the Henderson-Hasselbach equation [1]. In this sense, oral administration medications (syrups, elixirs, suspensions, tablets, capsules) must surpass the acidic pH of the stomach (1-3 units), which is capable to modify the S of the Active Pharmaceutical Ingredient (API). Therefore, knowing the S of APIs at strongly acidic pH becomes essential in the design of medications that are intended to be orally administered.

In the other hand, the experimental determination of S depends on traditional disadvantageous procedures, that employ long times and high quantities of samples, as well as qualified personal [1]. Among the alternative solutions to these inconveniences, there are the quantitative structure-property relationship (QSPR) studies, which are oriented to identify a function capable to predict a determined property [2]. Having into account that QSPR solutions can accelerate the development of new pharmaceutical products [3], this work was aimed to quantitative relate the chemical structure of a series of organic compounds, of pharmaceutical interest, with their S in strongly acidic medium.

Materials and Methods

Training set and solubility source. The sets were assembled from ACD/Labs version 10.04,[4] commercial software that was also employed to obtain the S values (in logarithmic scale) of compounds in pH interval from 3.5 to 5. According to a qualitative classification criterion that consider the chemical functional groups and the IUPAC priority, the training set was divided into five chemical groups.

Descriptors. The TOPS–MODE approach (in MODESLAB software, version 1.5 [5]) was used to generate the molecular descriptors (MDs). The mathematical details of the method were widely described [6–8]. The selection of the parameters hydrophobicity (hyd), atomic weight (ato), van der Waals ratio (van), polarizability (pol), dipole moment (dip) and link distance (std) to weight the matrix was based on the influence of them in S behavior of ionizable molecules. Therefore, the descriptors calculated can allow the adequate codification of the molecular structure.

Statistical tools. Group–specific and general QSPR models were developed considering the pool of calculated descriptors after linear discriminant analysis. The statistical processing was carried out by using the stepwise multiple regression through IBM SPSS Statistics, used to reduce data, and BuildQSAR, usted to obtain and optimize predictive models.

Validation of models. The determination coefficient (R^2), correlation coefficient (R) and the root mean square error (RMSE) were considered for identification of statistical quality models.

Results and Discussion

After initial processing of the training set, atypical cases were eliminated, then original set was reduced to 592 compounds. Subsequent, they were grouped in five subsets: (1) acids, (2) bases, (3) neutrals, (4) sulphured and (5) hydrocarbons of 126, 127, 123, 88 and 128 molecules, respectively. Later, molecular descriptors based on TOPS-MODE approach were calculated. As a result, a matrix containing the spectral moments from μ_1 to μ_{15} was obtained per weight, in addition to the μ_0 (number of atoms), leading to a total of 91 MDs for each compound. The use of a large number of variables in QSPR studies makes difficult to explain the functions. Therefore, the linear discriminant analysis technique

was developed. From the initial set of independent variables, only 16 (µ0, µstd1, µstd3, µstd7, µdip3, µdip4, µhyd1, µhyd3, µhyd4, µhyd13, µpol1, µpol4, µpol12, µvan9, µvan12 and µato3) were included in the discriminant functions, representing a significant reduction. These MDs should be the variables with the greatest capacity to discriminate, from the chemical point of view, the belonging group of each compound (five sub-series of interest) from training set.

In order to obtain prediction models of aqueous solubility (in terms of log S) in strongly acidic pH ranges (1-1.3; 1.4-1.7; 1.8-2.1; 2.2 - 2.5) the multiple linear regression analysis was developed. As a result, specific (considering chemical classification) and general (not considering chemical classification) mathematical models were obtained. In total, 20 specific models (one for each pH range for each sub-set, M1-M20) and four general models (one for each pH range, M21-M24) were obtained (See Table 1).

In general, the models obtained included no more than two DM, in correspondence with what literature recommended, in order to minimize the risks of independent interchangeable correlation [9]. Also, the coefficients changed according to the pH in the same sub-series. This demonstrates the importance of considering the influence of pH on the prediction of S of compounds of pharmaceutical interest and, therefore, the utility of obtaining prediction models as a function of pH. However, the coefficients of the DM included in the models M17-M20 showed very little variation depending on the pH. This behavior may be related to hydrocarbons characteristics (halogenated or not), whose pKa values determine that they are not so susceptible to ionization in the pH range under study, with respect to the other sub-sets of chemical classification considered.

Except for compounds with basic characteristics (sub-set II), all models include at least one DM related to hydrophobicity, which demonstrates the importance of weighting this molecular property for the prediction of S in strongly acidic medium.

For models M1-M3, M13-M16 and M9-M12 values of the correlation (R) and determination (R^2) coefficients were obtained above 0.80 and 0.70 respectively (See table I). These results show the significant linear correlation of the DM included in these models with respect to S and that, together, these DMs explain more than 70% of the variability of the S depending on the chemical structure, suggesting a good adjustment to experimental data. Although models M17-M20 meet the recommended statistical quality criteria [10], they show a more discrete linear adjustment ($R \leq 0.80$ and $R2 \leq 0.6$) in comparison with the previous models.

The models M5-M8 exhibited the values of R and R^2 farther than 1 (See table I), which may be due to that compounds with basic characteristics easily ionize in strongly acidic medium, making them very soluble molecules. Therefore, their S in this pH range will depend on the modifications (ionization) that these molecules undergo, which are not contemplated in the SMILES of the compounds included in the training series, which were obtained from neutral form.

Consequently, the DM used (graph-theoretical type) were not sufficient to explain the S behavior of the bases in strongly acidic medium. In fact, the only DM that contributed significantly to the S of this group of compounds was µ0, which corresponds to the number of atoms in the molecule.

Table 1. Models to predict pH-dependent aqueous solubility

pH	Sub sets	**Specific predictive models**	R	R^2
1,0-1,3	I (acids)	M1: $logS = -0{,}4503(\mu hyd1) - 0{,}0078(\mu hyd4) + 0{,}536$	0,86	0,74
1,4-1,7		M2: $logS = -0{,}4290(\mu hyd1) - 0{,}0072(\mu hyd4) + 0{,}423$	0,85	0,72
1,8-2,1		M3: $logS = -0{,}4458(\mu hyd1) - 0{,}0070(\mu hyd4) + 0{,}432$	0,80	0,71
2,2-2,5		M4: $logS = -0{,}0086(\mu hyd4) + 0{,}1402$	0,71	0,50
1-1,3	II (bases)	M5: $\log S = -0{,}1702(\mu 0) + 2{,}8160$	0,71	0,51
1,4-1,7		M6: $logS = -0{,}1669(\mu 0) + 2{,}6668$	0,70	0,49
1,8-2,1		M7: $logS = -0{,}1653(\mu 0) + 2{,}5599$	0,69	0,48
2,2-2,5		M8: $logS = -0{,}1514(\mu 0) + 2{,}1223$	0,59	0,35
1,0-1,3	III (neutrals)	M9: $logS = -0{,}1447(\mu 0) - 0{,}4302(\mu hyd1) + 0{,}6258$	0,92	0,84
1,4-1,7		M10: $logS = -0{,}1480(\mu 0) - 0{,}4269(\mu hyd1) + 0{,}6428$	0,92	0,85
1,8-2,1		M11: $logS = -0{,}0941(\mu 0) - 0{,}4820(\mu hyd1) + 0{,}546$	0,90	0,81
2,2-2,5		M12: $logS = -0{,}1510(\mu 0) - 0{,}4410(\mu hyd1) + 0{,}670$	0,92	0,85
1,0-1,3	IV (sulphured)	M13: $logS = -0{,}7563(hyd1) - 0{,}719$	0,89	0,78
1,4-1,7		M14: $logS = -0{,}7341(hyd1) - 0{,}8056$	0,88	0,78
1,8-2,1		M15: $logS = -0{,}0393(\mu std1) - 0{,}6539(hyd1) - 0{,}367$	0,88	0,77
2,2-2,5		M16: $logS = -0{,}0425(\mu std1) - 0{,}6427(hyd1) - 0{,}374$	0,87	0,76
1,0-1,3	V (hydrocarbons and halogens)	M17: $logS = -0{,}1237(\mu std1) - 0{,}3911(hyd1) - 1{,}108$	0,79	0,62
1,4-1,7		M18: $logS = -0{,}1236(\mu std1) - 0{,}3910(hyd1) - 1{,}108$	0,79	0,62
1,8-2,1		M19: $logS = -0{,}1231(\mu std1) - 0{,}3755(hyd1) - 1{,}145$	0,79	0,63
2,2-2,5		M20: $logS = -0{,}1231(\mu std1) - 0{,}3755(hyd1) - 1{,}145$	0,79	0,63
		General predictive models		
1,0-1,3		M21: $logS = -0{,}0241(\mu std1) - 0{,}8888(\mu hyd1) + 0{,}0471$	0,83	0,68
1,4-1,7		M22: $logS = -0{,}0286(\mu std1) - 0{,}8714(\mu hyd1) + 0{,}0591$	0,83	0,69
1,8-2,1		M23: $logS = -0{,}0306(\mu std1) - 0{,}8552(\mu hyd1) + 0{,}0355$	0,83	0,68
2,2-2,5		M24: $logS = -0{,}0068(\mu dip3) - 0{,}8392(\mu hyd1) + 0{,}1943$	0,75	0,56

R^2: determination coefficient, R: correlation coefficient.

The ANOVA test for regression was highly significant ($p < 0.0001$) for all the models obtained. This gives greater reliability to them, since it results in the existence of a significant linear relationship between S and the DMs considered. This fact was complemented with the t test of significance of the slopes, which in all cases was significant ($p = 0.000$), which indicates that coefficients of DMs contribute in a relevant way to the prediction of log S. Except for models M5-M8, the standard error values of the estimate obtained for the models were lower than the logarithmic unit, which should be between 0.5 and 0.7 logarithmic units, according to recommendations for this regression parameter [11]. Therefore, the specific models that meet the requirements to be validated and subsequently used in the prediction of the S of new compounds in strongly acidic medium are M1-M3 and M9-M20.

The results obtained for the general models (M21-M24) in each pH range for the entire training set showed a more discrete adjustment with respect to specific ones. This behavior is associated with the greater diversity and quantity of compounds included in the training sub-set, as a result of the consideration of the five chemical classification groups. Then, it is not recommended to continue with the validation and subsequent use of the general models derived from this work for the correct prediction of S of compounds of pharmaceutical interest.

Conclusions

Fifty of the mathematical predictive models defined exhibited a good predictive capacity. Therefore, the validation of them is suggested, in order to be used in predicting the S of new compounds in strongly acidic medium.

References

1. Bhattachar SN, Deschenes LA, Wesley JA. Solubility: it'snot just for physical chemists. Drug Disc Today. 2006; 11: 1012-8.
2. Santos-Filho OA, Hopfinger AJ, Cherkasov A, De-Alencastro RB. The receptor-dependent QSAR paradigm: anoverview of the currentstate of the art. MedChem. 2009; 5: 359-66.
3. Duchowicz P, Castro EA. QSPR Studieson Aqueous Solubilities of Drug-LikeCompounds, Int J Mol Sci. 2009; 10: 2558-77.
4. ACD/Labs software, version 10.04; ACD/Labs 1994-2007. Available at http://www.acdlabs.com/ilab/.
5. MODESLAB software, version 1.5; Estrada, E. and Gutiérrez, Y.; 2002-2004. Available at http://www.modeslab.
6. Estrada, E. Spectral Moments of the Edge Adjacency Matrix in Molecular Graphs. 1. Definition and Applications to the Prediction of Physical Properties of Alkanes. J. Chem. Inf. Comput. Sci. 1996, 36, 844–849.
7. Estrada, E. Spectral Moments of the Edge-Adjacency Matrix of Molecular Graphs. 2. Molecules Containing Heteroatoms and QSAR Applications. J. Chem. Inf. Comput. Sci. 1997, 37, 320–328.
8. Estrada, E. Spectral Moments of the Edge Adjacency Matrix in Molecular Graphs. 3. Molecules Containing Cycles J. Chem. Inf. Comput. Sci. 1998, 38, 23–27.
9. Topliss JG. Utilization of OperationalSchemes for AnalogSynthesis in Drug Design. J MedChem. 1972; 15:1006-11.
10. Veerasamy R, Rajak H, Jain A, Sivadasan S, Varghese CP,et al. Validation of QSAR Models-Strategies and Importance. International Journal of DrugDesign and Discovery. 2011; 2: 511-19.
11. Cherkasov A, Muratov EN, Fourches D, Varnek A, Baskin II,et al. QSAR Modeling: Where have youbeen? Where are yougoing to? J MedChem. 2014; 57: 4977-10.