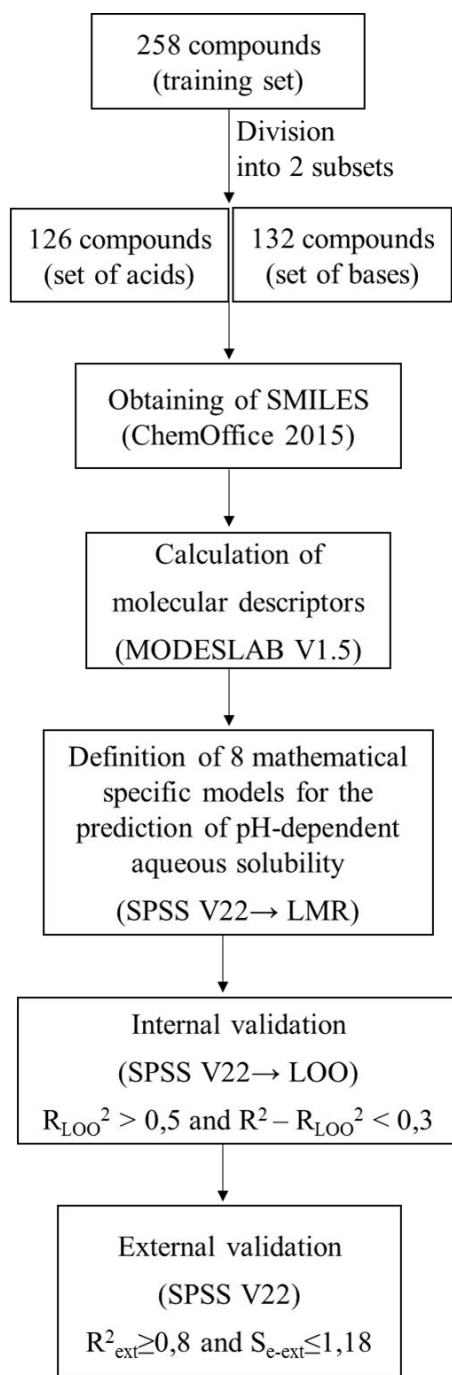


Prediction of pH-dependent aqueous solubility of druglike molecules of different chemical behavior

<Laura Machín Galarza> (laura@ifal.uh.cu)^a, <Luis Alberto Torres Gómez> (luistg@ifal.uh.cu)^a
^a <Department of Pharmacy. Institute of Pharmacy and Foods. University of the Havana>

Graphical Abstract



Abstract.

Aqueous solubility is an important factor that influence the biological processes where participate ionized molecules like drugs. The experimental determination of this property have several limitations like the elevated inverted time, as well as the consumption of considerable quantities of product. Among the requirements of the drugs of parenteral use, the pH of the final formulation must be close to neutrality. Small changes in this factor can have an important effect in molecule's solubility and, therefore, in its concentration in the final preparation. The fact that a lot of drugs has a pKa value around 7,4 determines that the selection of the pH, that guarantee an adequate commitment between dosage and the exigences of parenteral medicaments, remains like a real challenge during the design and development of this type of products. The structure-property relationship studies are oriented to find the function capable to predict a particular property of a compound, using the information contained in their descriptors. This strategy allows us to analyze a great quantity of molecules in a minor term of time and with less resources. Eight specific mathematical models were defined in the present work in order to predict the aqueous solubility by pH range and chemical group. In general, these models demonstrated a good prediction capacity when they faced a test set. The standard error of estimate for the specific functions were inferior or close to the logarithmic unit. Also, the prediction coefficients were between 0,6 and 0,9. The results suggest the employment of them in the design and development of medicaments of parenteral use.

Introduction

The facility/difficulty of ionization of a particular drug have to be consider during the experimental determination of its solubility, due to in the majority of the cases the drugs behave as acids or bases weekly ionizable. In this sense, the pKa of the molecule and the pH of the solution are factors of major interest¹; whose dependence with solubility is shown in the pH-solubility profile of each compound. However, the experimental determination of aqueous solubility of ionizable molecules accompanies several limitations, like the high time invested, as well as the elevated consumption of product². In the other hand, the parenteral preparations have to satisfy some requirements in order to reach an adequate adaptation to physiological conditions of blood and tissues. Among these requirements, the pH of the final formulation have to be close to neutrality³. But, small changes in this factor can have an important effect in molecule's solubility, as well as in its concentration in the final preparation.⁴ To choose the pH, that accomplish an adequate commitment between the drug dose and the exigences of this type of preparations, is one of the main challenges to face in the design and development of parenteral products; taking into account that a lot of drugs have a pKa value close to 7,4. An interesting alternative is the combined employment of experimental and computational methods. Specifically, the *Quantitative structure-property relationship* (QSPR) studies are oriented to find the best function that predict a determined quality of a compound, using the information that contain its molecular descriptors. In this sense, the MODESLAB approach is used in order to calculated the spectral moments of the adjacency matrix between edges of the molecular graph with suppressed hydrogens.⁵ The main objective of the present work was to predict the aqueous solubility of druglike molecules considering four ranges of pH close to neutrality and two different chemical behaviors.

Materials and Methods

A training set and a prediction one of 258 and 46 compounds were designed respectively. Both sets were divided into two subgroups attending to their chemical behavior: (i) the acid one and (ii) the basic one. The spectral moments of each compound were calculated with the MODESLAB software by weighting the following molecular graphs: bond distance (Std), dipole moment (Dip), hydrophobicity (Hyd), polarization (Pol), atomic radius of van der Waals (Van) and atomic weight (Ato). As a result, a matrix containing the spectral moments from μ_0 to μ_{15} was obtained for each molecule. The logS values were obtained from the ACD/Labs program. The statistical software IBM SPSS version 22 for Windows was used for the definition of mathematical predictive models, using multiple linear regression.

Results and Discussion

After the processing of the training set eight mathematical predictive models (M1-8) were obtained, four for each chemical group. Each function predicts the logS of compounds in a specific range: (i) 6,5-6,7; (ii) 6,8-7; (iii) 7,1-7,3 and (iv) 7,4-7,5. The functions that correspond to the acid group included five descriptors, whereas those that describe the basic one included seven descriptors.

$\log S_{pH:6,5-6,7} = ,048 (\mu_4^{dip}) - ,157 (\mu_3^{dip}) - ,167 (\mu_2^{pol}) + ,043 (\mu_1^{ato}) - ,005 (\mu_5^{hyd}) + 1,149$	(M1, acid group)
$\log S_{pH:6,8-7} = ,048 (\mu_4^{dip}) - ,153 (\mu_3^{dip}) - ,169 (\mu_2^{pol}) + ,044 (\mu_1^{ato}) - ,006 (\mu_5^{hyd}) + 1,237$	(M2, acid group)
$\log S_{pH:7,1-7,3} = ,047 (\mu_4^{dip}) - ,150 (\mu_3^{dip}) - ,169 (\mu_2^{pol}) + ,044 (\mu_1^{ato}) - ,006 (\mu_5^{hyd}) + 1,313$	(M3, acid group)
$\log S_{pH:7,4-7,5} = ,046 (\mu_4^{dip}) - ,144 (\mu_3^{dip}) - ,166 (\mu_2^{pol}) + ,044 (\mu_1^{ato}) - ,006 (\mu_5^{hyd}) + 1,337$	(M4, acid group)

$\log S_{\text{pH}:6,5-6,7} = -2,221 (\mu_0) - ,113 (\mu_3^{\text{pol}}) + 1,44 (\mu_1^{\text{pol}}) + 1,107 (\mu_1^{\text{std}}) + ,177 (\mu_3^{\text{hyd}}) + 5,664E - 10 (\mu_5^{\text{ato}}) - 1,361 (\mu_1^{\text{hyd}}) + ,751$	(M5, basic group)
$\log S_{\text{pH}:6,8-7} = -2,177 (\mu_0) - ,111 (\mu_3^{\text{pol}}) + 1,411 (\mu_1^{\text{pol}}) + 1,082 (\mu_1^{\text{std}}) + ,173 (\mu_3^{\text{hyd}}) + 5,571E - 10 (\mu_5^{\text{ato}}) - 1,374 (\mu_1^{\text{hyd}}) + ,778$	(M6, basic group)
$\log S_{\text{pH}:7,1-7,3} = -2,128 (\mu_0) - ,109 (\mu_3^{\text{pol}}) + 1,38 (\mu_1^{\text{pol}}) + 1,054 (\mu_1^{\text{std}}) + ,169 (\mu_3^{\text{hyd}}) + 5,469E - 10 (\mu_5^{\text{ato}}) - 1,361 (\mu_1^{\text{hyd}}) + ,809$	(M7, basic group)
$\log S_{\text{pH}:7,4-7,5} = -2,081 (\mu_0) - ,107 (\mu_3^{\text{pol}}) + 1,35 (\mu_1^{\text{pol}}) + 1,026 (\mu_1^{\text{std}}) + ,165 (\mu_3^{\text{hyd}}) + 5,355E - 10 (\mu_5^{\text{ato}}) - 1,351 (\mu_1^{\text{hyd}}) + ,835$	(M8, basic group)

The statistical parameters obtained for the eight QSPR models after internal validation, using leave-one-out (LOO) procedure, as well as external validation are shown in table 1.

Table 1. Statistical performance of the QSPR models.

Group	Model	pH range	R	R ²	S _e	R ² _{LOO}	S _{e-LOO}	R ² _{ext}	S _{e-ext}
acid	1	6,5 – 6,7	,899	,807	,951	,768	1,03	,804	,927
	2	6,8 – 7	,906	,821	,939	,782	1,02	,809	,972
	3	7,1 – 7,3	,911	,829	,928	,790	1,01	,827	,889
	4	7,4 – 7,5	,913	,834	,912	,794	1,00	,804	1,18
basic	5	6,5 – 6,7	,852	,725	1,00	,672	1,06	,845	,51
	6	6,8 – 7	,859	,738	,991	,687	1,04	,831	,529
	7	7,1 – 7,3	,866	,750	,973	,701	1,02	,813	,572
	8	7,4 – 7,5	,872	,761	,956	,713	1,01	,799	,625

R: correlation coefficient, R²: determination coefficient, S_e: standard error of estimate, R²_{LOO}: cross correlation coefficient, S_{e-LOO}: standard cross correlation coefficient, R²_{ext}: prediction coefficient for external test, S_{e,pred}: standard error for external test.

The contribution of descriptors to solubility in the mathematical models that correspond to the same group is different for each pH range, which demonstrates the importance of considering this factor in the prediction of the solubility of ionized drug compounds. For the functions that described the solubility behavior of the acid and basic groups, the R and R² values were superior to 0,85 and to 0,72, respectively; which suggest that the molecular descriptors involved are capable to explain more than the 85% and 72% of pH-dependent solubility's variability, as well as a good adjustment of the predicted values to the experimental ones. Also, the eight models accomplished the cross validation criteria, due to R²_{LOO} > 0,5 and R² – R²_{LOO} < 0,3. In the other hand, according to the R²_{ext} obtained, we can consider that the models obtained are capable to explain around the 80% of the variability of the aqueous solubility of drug compounds that were not considered for the construction of the models. Also, the S_{e-ext} inferior to or close to the logarithmic unit indicates an adequate prediction capacity of them.

Conclusions

The eight mathematical predictive models presented in this study showed R^2_{ext} values superior to and close to 0,8; as well as $S_{e-\text{ext}}$ values inferior to or close to the logarithmic unit. The results presented suggest the employment of these models in the prediction of aqueous solubility of druglike molecules, as a part of the design and development of parenteral preparations.

References

1. Li, P. et al. (1999). Solubilization of ionized and un-ionized flavopiridol by ethanol and polysorbate 20. *J. Pharm. Sci.*, **88**: 507–509.
2. Duchowicz, P. & Castro, E. A. (2009). QSPR Studies on Aqueous Solubilities of Drug-Like Compounds, *Int. J. Mol. Sci.*, **10**: 2558-2577.
3. The Department of Health. General monographs. Parenteral preparations. British Pharmacopeia (BP) Volume III. (2009).
4. Vila, J. L. (2005). Formas farmacéuticas Vol. II. 157-249.
5. Estrada, E. & Gutiérrez, Y. (2002). MODESLAB for Windows User's Manual, University of Santiago de Compostela, Santiago de Compostela.