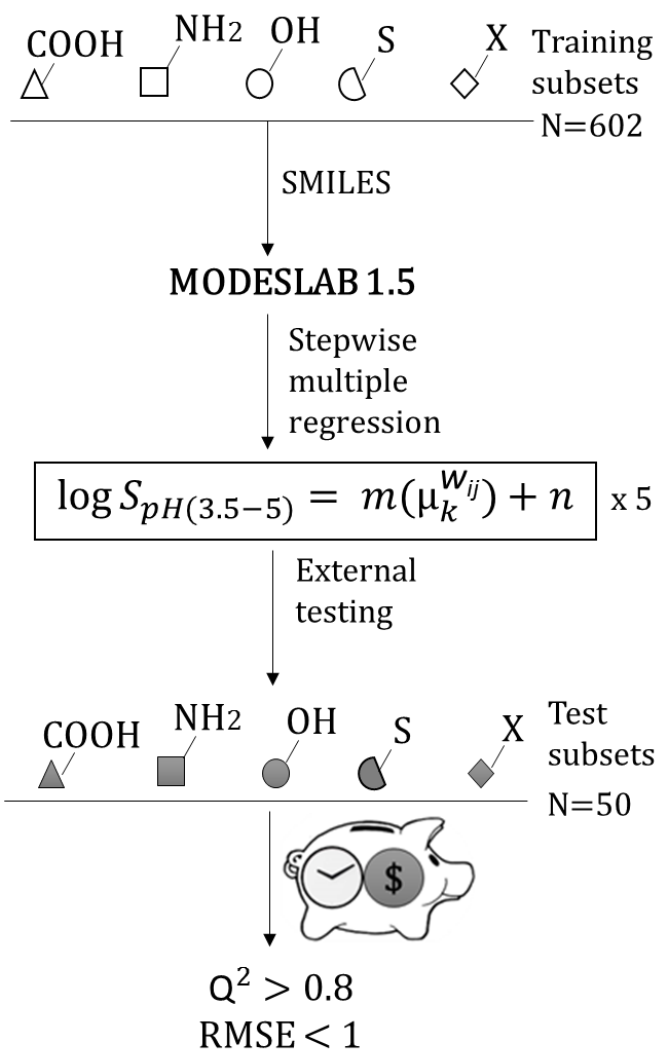
MOL2NET, International Conference Series on Multidisciplinary Sciences

Models to Predict pH–dependent Aqueous Solubility of Chemically Diverse Druglike Compounds

Gretter León Sánchez [a], Laura Machín Galarza (laura@ifal.uh.cu) [a], Juan Carlos Polo Vega [a], Luis Alberto Torres Gómez [a]

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

Graphical Abstract



Abstract.

The aqueous solubility plays a key role in the biological processes that involved ionizable molecules like drugs, particularly, in the design and development of pharmaceutical formulations oriented to oral administration. Taking into account that experimental determination of aqueous solubility presents several limitations, like the elevated inverted time and the consumption of considerable quantities of sample, the use of structure-activity relationship (QSAR) studies has been suggested. These are intended to defined the function able to predict specific properties of a compound, using the information contained in their molecular descriptors. Therefore, it is possible to examine a great quantity of molecules in a minor time and with less resources. Then, the present work was oriented to obtain mathematical models for the prediction of aqueous solubility in pH range from 3.5 to 5.0. After processing a training serie of 602 compounds divided into five chemically different sub-series, five mathematical models were defined. In general, these models demonstrated a good prediction capacity after external testing. Four of them exhibited standard error of estimate inferior or close to the logarithmic unit. Also, the external prediction coefficients were superior to 0.8. Finally, the results obtained suggest these models for the design and development of new oral medicaments.

Introduction

The aqueous solubility (S) plays a key role in the biological processes where drugs are involved, due to dissolution rate and permeability of molecules depend on this physicochemical parameter.[1] In particular, the therapeutic efficacy of oral medications depends on drug bioavailability, which considers the extent and rate of drug absorption from the pharmaceutical preparation to the action site.[2] Then, knowing the aqueous drug solubility is of major interest. In the other hand, the facility or difficulty of drugs to become ionized should be consider during experimental determination of S.[3] In this sense, the final pH and the pKa of molecules are very importance,[4] however they are frequently ignored.

When manipulation of reagents involved some risk and determination of a property/activity results complex, an interesting alternative are quantitative structure property relationship (QSPR) studies. These methodologies are oriented to find the best functions that predict a compound specific property through its molecular descriptors. Also, QSPR studies analyze a great quantity of molecules in less time and with minus resources. Therefore, they can be recommended to accelerated the development of pharmaceutical products.[5]

The quality of mathematical models depends on several factors like the descriptors, whose determination can be based on different theories.[6] In this sense, the TOPS–MODE approach can be used in order to calculated the spectral moments of the adjacency matrix between edges of the molecular graph with suppressed hydrogens[7–11]. Therefore, the aims of this work were to generate and evaluate mathematical functions oriented to predict the aqueous solubility of ionizable molecules considering their chemical behavior and pH range close to acidity.

Materials and Methods

Training set and solubility source. According to a qualitative classification criterion that consider the chemical functional groups and the IUPAC priority, the training set was divided in five groups: (1) acids, (2) bases, (3) neutrals, (4) sulphured and (5) hydrocarbons of 126, 131, 123, 89 and 133 molecules, respectively. The sets were assembled from ACD/Labs version 10.04,[12] commercial software employed to obtain the S values (in logarithmic scale) of compounds in pH interval from 3.5 to 5.

Descriptors. The SMILES derived from ACD/Labs were introduced in the software package MODESLAB version 1.5,[13] which includes the TOPS–MODE approach used to generate the molecular descriptors. The mathematical details of the method have been largely reported [14–16]. 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 molecular descriptors for each compound. The selection of these parameters to weight the matrix was based on the influence of them in solubility behavior of ionizable molecules. Therefore, the descriptors calculated allowed the adequate codification of the molecular structure.

Statistical tools. Group–specific QSPR models were developed considering the whole pool of calculated descriptors. The statistical processing was carried out by using the stepwise multiple regression, where the independent variables are individually deleted from the model in order to obtain the best one.

pH–dependent aqueous solubility prediction of a test set. Following the same classification criterion, a test set of 50 compounds was divided in five chemically different groups of ten molecules each. These sub-series were used to evaluate the prediction power of QSPR models outside the training set.

Validation of models. The determination coefficient (R^2), the prediction coefficient for external test (Q^2) and the external root mean square error ($RMSE_{ext}$) determined the accuracy of the models.

Results and Discussion

After processing the training sub-sets (acids, bases, neutrals, sulphured and hydrocarbons + halogens) five mathematical predictive models (1-5) were obtained (See Table 1). The inclusion of less than ten predictive variables (five, seven, four, five and five, respectively) in each model highlights the economy of descriptors achieved, which is desirable for a better interpretation of results.

Table 1. Models to predict pH-dependent aqueous solubility

#	Equations	*Sub-serie*
1	$logS_{pH:3.5-5} = -0{,}1379\,(\mu_3^{dip}) + 0{,}0382\,(\mu_4^{dip}) - 0{,}0021\,(\mu_5^{hyd}) - 0{,}1342(\mu_2^{pol}) + 0{,}0319(\mu_1^{ato}) + 0{,}5412$	*acids*
2	$logS_{pH:3.5-5} = -2{,}8777\,(\mu_0) - 0{,}1278\,(\mu_3^{pol}) + 1{,}7464(\mu_1^{pol}) + 1{,}3816(\mu_1^{std}) + 0{,}2227(\mu_3^{hyd}) + 0{,}0000\,(\mu_5^{ato}) - 1{,}7845(\mu_1^{hyd}) + 1{,}3329$	*bases*
3	$logS_{pH:3.5-5} = -0{,}0164\,(\mu_1^{ato}) - 0{,}4123\,(\mu_1^{hyd}) + 0{,}0000\,(\mu_6^{ato}) + 0{,}0032\,(\mu_3^{dip}) + 1{,}0371$	*neutrals*
4	$logS_{pH:3.5-5} = -0{,}6955\,(\mu_2^{pol}) + 0{,}1299\,(\mu_3^{pol}) + 1{,}6418\,(\mu_0) - 0{,}3889\,(\mu_1^{van}) - 0{,}0000\,(\mu_5^{ato}) + 0{,}8201$	*sulphured*
5	$logS_{pH:3.5-5} = -1{,}3805\,(\mu_1^{std}) + 1{,}6610\,(\mu_0) + 0{,}1359\,(\mu_2^{pol}) - 0{,}4642\,(\mu_1^{pol}) - 0{,}0000\,(\mu_6^{ato}) - 0{,}5212$	*hydrocarbons + halogens*

The statistical parameters obtained for the five QSPR models after external validation are shown in table 2.

Table 2. Performance on the test set of the QSPR models.

Group	eq	Q^2	R^2 - Q^2	$RMSE_{ext}$
Acids	1	0,862	0,213	0,874
Bases	2	0,811	0,094	0,703
Neutrals	3	0,950	0,057	0,345
Sulphured	4	0,805	0,023	1,035
Hydrocarbons + halogens	5	0,984	0,106	2,321

R^2: determination coefficient, Q^2: prediction coefficient for external test, $RMSE_{ext}$: standard error for external test.

The five models accomplished the external validation criteria, due to $Q^2 > 0,8$ and $R^2 - Q^2 < 0,3$. According to the Q^2 obtained, we can consider that the models obtained are capable to explain around the 80% of the variability of aqueous solubility outside the training molecules. Also, the $RMSE_{ext}$ inferior or close to the logarithmic unit obtained for models 1-4 indicates an adequate prediction capacity of them. The higher value achieved after testing the fifth model may be associated to the structural variability of the correspondent group, which includes aromatic and aliphatic compounds of saturated and/or unsaturated chemical nature.

Conclusions

Four of the mathematical predictive models defined exhibited a good predictive capacity. Therefore, the use of them is suggested in the prediction of aqueous solubility of druglike molecules in a range close to acidity.

References

1. Alelyunas, Y. W.; Empfield, J. R.; McCarthy, D.; Spreen, R. C.; Bui, K.; Pelosi-Kilby, L.; Shen, C. Experimental solubility profiling of marketed CNS drugs, exploring solubility limit of CNS discovery candidate. Bioorg. Med. Chem. Lett. 2010, 20, 7312–7316.
2. Morgan P, Van Der Graaf PH, Arrowsmith J, Feltner DE, Drummond KS y col. Can the flow of medicines be improved? Fundamental pharmacokinetic and pharmacological principles toward improving Phase II survival. Drug Discov Today. 2012, 17, 419-24.
3. Comer, J. E. A. High-throughput Measurement of log D and pKa. In Drug Bioavailability; Van de Waterbeemd, H., Lennernas, H., Artursson, P., Eds.; VCH: New York, 2003, 21–45.
4. Li, P., Tabibi, S. E., Yalkowsky, S. H. Solubilization of ionized and un-ionized flavopiridol by ethanol and polysorbate 20. J. Pharm. Sci. 1999, 88,507–509.
5. Podunavac-Kuzmanović, S. O.; Cvetković, D. D.; Barna, D. J. QSAR Analysis of 2-Amino or 2-Methyl-1-Substituted Benzimidazoles Against Pseudomonas aeruginosa. Int. J. Mol. Sci. 2009, 10, 1670–1682.
6. Katritzky, A. R.; Lobanov, V. S.; Karelson, M. QSPR - the correlation and quantitative prediction of chemical and physical properties from structure. Chem. Soc. Rev. 1995, 24, 279–287.
7. Cabrera, M. A.; Bermejo, M.; Ramos, L.; Grau, R.; Pérez, M.; González, H. A topological sub-structural approach for predicting human intestinal absorption of drugs. Eur. J. Med. Chem. 2004, 39, 905–916.
8. Pérez, M.; Morales, A.; González, H. A TOPS-MODE approach to predict permeability coefficients. Polymer, 2004, 45, 2073–2079.
9. Molina, E.; Sobarzo-Sanchez, E.; Speck-Planche, A.; Matos, M. J.; Uriarte, E.; Santana, L.; Yáñez, M.; Orallo, F. Monoamino oxidase a: an interesting pharmacological target for the development of multi-target QSAR. Mini. Rev. Med. Chem. 2012, 12, 947–958.
10. Morales, A.; Pérez, M.; Rieumont, J. TOPS-MODE approach to predict mutagenicity in dental monomers. Polymer, 2004, 45, 2045–2050.
11. Rojas-Vargas, J. A.; Speck-Planche, A.; García-López, A.; Acevedo-Martínez, J. Modelación del índice de retención en iminas usando descriptores TOPS-MODE. Rev. Cub. Quím. 2012, 24, 272–283.
12. ACD/Labs software, version 10.04; ACD/Labs 1994-2007. Available at http://www.acdlabs.com/ilab/.
13. MODESLAB software, version 1.5; Estrada, E. and Gutiérrez, Y.; 2002-2004. Available at http://www.modeslab.
14. 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.
15. 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.
16. 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.