COMPUTATIONAL QSPR MODELING OF LIPOPHILICITY FOR ORGANIC DRUG-LIKE MOLECULES

Pushkarova Ya. M., Zaitseva G. M., Probyta A. A.
Bogomolets National Medical University, Ukraine
yaroslava.pushkarova@gmail.com

Lipophilicity is a key physicochemical property influencing the absorption, distribution and stability of drug-like molecules in biological systems. Accurate prediction of lipophilicity is essential for drug design, optimization of bioactive compounds and biotechnological applications.

Computational approaches, such as Quantitative Structure–Property Relationship (QSPR) modeling, allow estimation of lipophilicity at the molecular level prior to experimental testing, thereby reducing time and resources in the development process. Integrating in silico QSPR predictions into the early stages of molecular design provides a powerful tool for optimizing the physicochemical and biological properties of organic drug-like molecules, supporting more efficient and rational strategies in pharmaceutical and industrial biotechnology.

The modeling dataset included 76 organic drug-like molecules, which were characterized using nine electronic, physicochemical and topological descriptors. A detailed correlation analysis and multiple linear regression were performed to identify the most informative descriptors and evaluate their influence on lipophilicity. Based on this analysis, three descriptors – hydrogen bond acidity, polar surface area and molecular weight – were selected for inclusion in the predictive model. The compounds were randomly divided into a training set (85%) and a test set (15%). The training data were used to build and optimize the model coefficients, while the test data were employed to assess its predictive performance on unseen compounds. All computations and statistical evaluations were carried out using Matlab R2024b.

The developed three-parameter QSPR model demonstrated high predictive accuracy. High values of the coefficient of determination (0.7428) and the correlation coefficient (0.8618) between experimental and predicted lipophilicity values indicate that the model can reliably predict lipophilicity for new molecules.
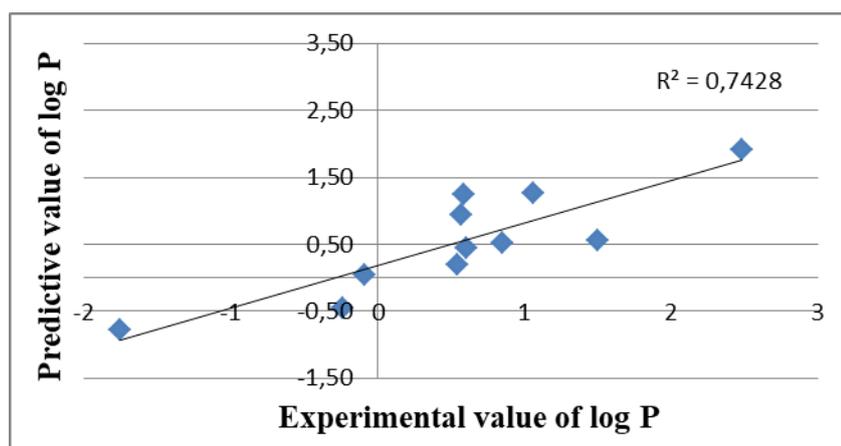

Figure 1. Dependence of predicted lipophilicity values on experimental lipophilicity values of the test set drug-like molecules for the three-parameter QSPR model

This model can be effectively used for preliminary screening of compounds during early stages of drug development, facilitating the identification of promising drug-like molecules and supporting rational design strategies.