STATISTICS FOR QSAR MODELS VALIDATION

Sorana D. Bolboacă[1], Carmen E. Stoenoiu[2], Lorentz Jäntschi[3]

[1] "Iuliu Haţieganu" University of Medicine and Pharmacy
13 Emil Isac, 400023, ROMANIA
sbolboaca@umfcluj.ro

[2] Technical University of Cluj-Napoca,
15 C-tin Daicoviciu, 400020, ROMANIA
carmen@j.academicdirect.org

[3] Technical University of Cluj-Napoca,
15 C-tin Daicoviciu, 400020, ROMANIA
lori@j.academicdirect.org

Key Words and Phrases: MDF-SAR (Molecular Descriptors Family on the Structure-Activity Relationship); Models Assessment and Validation; Statistical Methods

AMS Subject Classification: 62J05, 62H20, 62J99

General validation principles for quantitative structure-activity relationship (qSAR) models in the context of chemical regulation were developed [1] due to the importance and implication in of these methods in drug design. A brief analysis of different techniques used in validation of multiple linear regression models [2] is reviewed. The hierarchical steps in models validation are highlighted and a validation technique is proposed. The following statistical approaches are considered: correlation analysis (Pearson, Spearman, Kendall and Gamma coefficients as parameters and associated significance levels [3]), regression analysis (leave-one-out cross-validation and determination coefficients), and other inferential statistics (cross correlation coefficients, training vs. test experiment, correlated correlations analysis). The proposed statistical validation technique is exemplified on a qSAR model obtained by applying the molecular descriptors family on the structure-activity relationship approach [4].

Acknowledgements

References

[1] Gramatica P. QSAR Comb Sci 2007;26(5):694-701.
[2] Bolboacă SD, Jäntschi L. Env Chem Lett. 10.1007/s10311-007-0119-9.
[3] Bolboaca SD, Jäntschi L. Leonardo J Sci 2006;9:17-200.
[4] Jäntschi L. Leonardo Elect J Pract Techol 2005;6:76-98.