Cepos InSilico SAR System - CeSAR

Ute Seidel

Cepos InSilico GmbH, Computer-Chemie-Centrum, Nägelsbachstr. 25, 91052 Erlangen; http://www.ceposinsilico.com

Cepos InSilico SAR is a web-based system designed to find linear and non-linear relationships between data. It can predict the physical, chemical and biological behaviors of compounds based on measured or calculated information in order to develop predictive models. The software is especially designed as an easy-to-use tool for both beginners and experts. In addition, the Cepos InSilico SAR system includes an integrated database for data storage. In order to guarantee the optimum accuracy of the predictive models, partial least squares (PLS), [1,2] Bagged Stepwise Multiple Linear Regression using the descriptor-pool-size corrected F-value, [3] Random Forests and Artificial Neural Networks (ANNs) [4] are all available.

An important feature of the SAR-system is its ability to predict likely model errors and to describe the applicability domain exactly.

The system is designed primarily to use the surface-based descriptors and surface-integral models available in ParaSurf™. [5] These proved powerful and robust models, as will be shown using several examples.

[1] Wold, H., *Estimation of Principal Components and Related Models by Iterative Least Squares*, in Multivariate Analysis, ed. P. R. Krishnaiah, New York: Academic Press, **1966**, Seiten 391-420.
[2] Leach , A. R., Molecular Modelling: Principles and Applications. 2nd ed. **2001**, Harlow: Prentice-Hall.
[3] Kramer, C.; Tautermann, C. S.; Livingstone, D. J.; Salt, D. W.;Whitley, D. C.; Beck, B.; Clark, T. Sharpening the toolbox of computational chemistry: A new approximation of critical F-values for multiple linear regression. J. Chem. Inf. Model. **2009**, 49, 28–34.
[4] Zupan, J.; Gasteiger, J. *Neural Networks in Chemistry and Drug Design*, Wiley-VCH, Weinheim, **1999**, 2nd edition.
[5] Clark, T. ParaSurf'11™, **2011**, Users' Manual, www.ceposinsilico.de/products/parasurf.htm