

Similarity based assessment of Model Applicability Domain and quantitative evaluation of the reliability of the prediction

P. Japertas^a, R. Didziapetris^a, A. Sazonovas^{a,b} and A. Petrauskas^a

^a Pharma Algorithms, Inc., A.Mickeviciaus g. 29, LT-08117 Vilnius, Lithuania

^b Faculty of Chemistry, Vilnius University, Naugarduko g. 24, LT-03225 Vilnius, Lithuania
presenting author e-mail: pranas@pharma-algorithms.com

Every model, no matter what data, descriptors or modeling techniques were used building it, has certain applicability domain, beyond which the quality of predictions becomes highly questionable.

“The applicability domain of QSAR model is the response and chemical structure space in which the model makes predictions with a given reliability”¹

Assessment of this domain is critical and is therefore one of the main requirements for the method recognition by regulatory authorities and initiatives such as REACH:

“According to the OECD Principles for (Q)SAR validation, a (Q)SAR model that is proposed for regulatory use should be associated with five types of information: 1) a defined endpoint; 2) an unambiguous algorithm; 3) a defined domain of applicability; 4) appropriate measures of goodness-of-fit, robustness and predictivity; and 5) a mechanistic interpretation, if possible.”²

In this presentation the theoretical background of the novel methodology is outlined. It includes the determination of the test compound similarity to the training set of the model as well as the analysis of the experimental data for the most similar compounds found in it. The similarity is defined dynamically from the dataset of the model under consideration, i.e. it is different for each individual property or activity. For each compound that the model is applied to, all this information is summarized in the Reliability Index value which serves as a quantitative measure of the quality of every prediction. Apart from the identification whether a particular test compound falls within the Applicability Domain of the model (i.e. the usability of the values predicted for that molecule), these developed indices can also be useful in experimental planning and prioritization of measurements.

References

1. Netzeva, T.I., et al., Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. *Alternatives to Laboratory Animals*, 2005 **33**, 155-173.
2. Worth, A.P., et al. The Characterisation of (Quantitative) Structure-Activity Relationships: Preliminary Guidance. EUR 21866 EN, 2005.
(http://ecb.jrc.it/DOCUMENTS/QSAR/QSAR_characterisation_EUR_21866_EN.pdf
accessed on 2008.04.30)