

## Estimation of the Accuracy of ADMET Predictions and Secure Sharing of Information Are Two Sides of the Same Coin

Igor V. Tetko

Institute for Bioinformatics, GSF, Neuherberg, Germany and Institute of Bioorganic & Petrochemistry, Murmanskaya 1, 02094 Kiev, Ukraine,  
<http://www.vcclab.org>

This article discusses importance of correct estimation of the applicability domain of ADMET models and methods to predict physico-chemical properties in the early stages of drug development. The methods are classified into two main groups, namely, methods based on the analysis of similarity of molecules (“molecular-based similarities”) and methods based on the analysis of calculated properties (“property-based similarities”). Using the example of octanol-water distribution coefficients I will exemplify consistency of estimated and calculated accuracy of the ALOGPS program to predict *in house* and publicly available datasets using examples from our previous publications.(Tetko, 2005; Tetko et al., 2006)

In the second part of the talk I will discuss a relation of the accuracy of ADMET predictions and problem of secure sharing of information. Indeed, if one can accurately estimate predictions for compounds structurally dissimilar from those used in the training set, the compounds with highly accurate predictions (“surrogate compounds”) will contain the same information as the training ones. Therefore the surrogate compounds can be shared together with their predicted to develop new predictive models. Moreover, one can *a priori* estimate/select the security level of the data sharing using surrogate compounds. An example of secure sharing of information for lipophilicity prediction is exemplified and discussed using examples from our previous work.(Tetko et al., 2005)

**Acknowledgements** I thank Tudor Oprea, Ruben Abagyan, Gennadiy I. Poda, Pierre Bruneau, Douglas C. Rohrer and Hans-Werner Mewes whose contribution and continuous support made this study possible. This work was partially supported with Virtual Computational Chemistry Laboratory INTAS INFO 00-0363 grant.

**Tetko IV** (2005) Encoding molecular structures as ranks of models: A new secure way for sharing chemical data and development of ADME/T models. Abstracts of Papers of the American Chemical Society **229**: U602-U602

**Tetko IV, Abagyan R, Oprea TI** (2005) Surrogate data - a secure way to share corporate data. J Comput Aided Mol Des **19**: 749-764

**Tetko IV, Bruneau P, Rohrer D, Poda GI** (2006) Can we estimate accuracy of ADMET predictions? Drug Discovery Today **in press**.