

ORAL PRESENTATION

Open Access

A unified approach to the applicability domain problem of QSAR models

Dragos Horvath*, Gilles Marcou, Alexandre Varnek

From 5th German Conference on Cheminformatics: 23. CIC-Workshop
Goslar, Germany. 8-10 November 2009

The present work proposes a unified conceptual framework to describe and quantify the important issue of the Applicability Domains (AD) of Quantitative Structure-Activity Relationships (QSARs). AD models are conceived as meta-models designed to associate an untrustworthiness score to any molecule *M* subject to property prediction by a QSAR model. Untrustworthiness scores or "AD metrics" are an expression of the relationship between *M* (represented by its descriptors in chemical space) and the space zones populated by the training molecules at the basis of model μ . Scores integrating some of the classical AD criteria (similarity-based, box-based) were considered in addition to newly invented terms, such as the dissimilarity to outlier-free training sets and the correlation breakdown count.

A loose correlation is expected to exist between this untrustworthiness and the error affecting the predicted property. While high untrustworthiness does not preclude correct predictions, inaccurate predictions at low untrustworthiness must be imperatively avoided. This kind of relationship is characteristic for the Neighborhood Behavior (NB) problem: dissimilar molecule pairs may or may not display similar properties, but similar molecule pairs with different properties are explicitly "forbidden". Therefore, statistical tools developed to tackle this latter aspect were applied, and lead to a unified AD metric benchmarking scheme.

A first use of untrustworthiness scores resides in prioritization of predictions, without need to specify a hard AD border. Moreover, if a significant set of external compounds is available, the formalism allows optimal AD borderlines to be fitted. Eventually, consensus

AD definitions were built by means of a nonparametric mixing scheme of two AD metrics of comparable quality, and shown to outperform their respective parents.

Published: 4 May 2010

doi:10.1186/1758-2946-2-S1-O6

Cite this article as: Horvath et al.: A unified approach to the applicability domain problem of QSAR models. *Journal of Cheminformatics* 2010 2(Suppl 1):O6.

Laboratoire d'InfoChimie, Université de Strasbourg - CNRS, Institut de Chimie, 4 rue Blaise Pascal, 67000 Strasbourg, France

Publish with **ChemistryCentral** and every scientist can read your work free of charge

"Open access provides opportunities to our colleagues in other parts of the globe, by allowing anyone to view the content free of charge."

W. Jeffery Hurst, The Hershey Company.

- available free of charge to the entire scientific community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
<http://www.chemistrycentral.com/manuscript/>



ChemistryCentral

