

Poster presentation

Open Access

Data integration and knowledge transfer: application to the tissue: air partition coefficients

C Gaudin^{1,2}, G Marcou¹, P Vayer², I Tetko³, I Baskin⁴ and A Varnek^{*1}

Address: ¹Laboratoire d'Infochimie (UMR 7177 CNRS), Strasbourg, France, ²Technologie Servier, Orléans, France, ³GSF-Institute for Bioinformatics, Neuherberg, Germany and ⁴Department of Chemistry, Moscow State University Moscow, Russia

* Corresponding author

from 4th German Conference on Chemoinformatics
Goslar, Germany. 9–11 November 2008

Published: 5 June 2009

Chemistry Central Journal 2009, 3(Suppl 1):P30 doi:10.1186/1752-153X-3-S1-P30

This abstract is available from: <http://www.journal.chemistrycentral.com/content/3/S1/P30>

© 2009 Gaudin et al; licensee BioMed Central Ltd.

Conventional QSAR/QSPR models are built only for one target property without exploiting any a priori knowledge stored in datasets of related properties. Here, individual models are not viewed as separate entities but as nodes in the network of interrelated models. Such interrelated models can be built in parallel by means of multitask learning (MTL), or sequentially using feature nets (FN). MTL and FN are kinds of data integration, as opposed to traditional single-task learning (STL), in which all models are built separately. We apply this strategy to model Human blood:air, human and rat tissue:air partition coefficients of organic compounds using diverse and relatively small datasets.