## **ORAL PRESENTATION**



**Open Access** 

## Statistical modeling of value distributions of similarity coefficients in virtual screening and its application to predicting fingerprint search performance

Martin Vogt<sup>\*</sup>, Jürgen Bajorath

*From* 8th German Conference on Chemoinformatics: 26 CIC-Workshop Goslar, Germany. 11-13 November 2012

Similarity searching using fingerprints is a popular ligandbased virtual screening approach. The Tanimoto coefficient (Tc) is the most widely used measure for quantifying fingerprint similarity. In general, it is very difficult to assess the significance of the similarity of two molecules solely based on their calculated Tc values. In the literature, Tc cut-off values are frequently intuitively chosen as similarity criteria for virtual screening. This can be very problematic because the distribution of similarity scores largely depends on the specific type of fingerprint that is used and the reference compound for which the fingerprint is calculated. In order to rationalize similarity value considerations, a statistical approach named the conditional correlated Bernoulli model is presented that models similarity scores based on the statistical distribution of fingerprint features in large compound databases. Fingerprint features are modeled as dependent Bernoulli variables and conditional distributions of Tanimoto similarity values of database compounds are determined with respect to given reference compounds. The model makes it possible to estimate the position of a compound in a database ranking only based on its Tc value relative to the reference. This rank estimation of molecules enables the quantitative comparison of similarity values of different fingerprint types. Moreover, it can be utilized to rapidly assess the potential of fingerprints to identify new active molecules in a database search given a set of known reference molecules [1].

\* Correspondence: martin.vogt@bit.uni-bonn.de

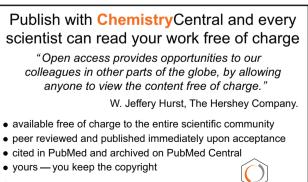


## Reference

 Vogt M, Bajorath J: Introduction of the Conditional Correlated Bernoulli Model of Similarity Value Distributions and its Application to the Prospective Prediction of Fingerprint Search Performance. J Chem Inf Model 2011, 51:2496-2506.

## doi:10.1186/1758-2946-5-S1-O5

**Cite this article as:** Vogt and Bajorath: **Statistical modeling of value** distributions of similarity coefficients in virtual screening and its application to predicting fingerprint search performance. *Journal of Cheminformatics* 2013 **5**(Suppl 1):O5.



**Chemistry**Central

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



© 2013 Vogt and Bajorath; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Department of Life Science Informatics, B-IT, LIMES Program Unit Chemical Biology and Medicinal Chemistry, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, 53113, Germany