Chemistry Central Journal

MUVing SAFEly: Scale for Assessing Figures of Effectiveness (SAFE) in virtual screening using Maximum Unbiased Validation (MUV) datasets

S Rohrer* and K Baumann

Address: Institut für Pharmazeutische Chemie, Technische Universität Braunschweig, Beethovenstr. 55, 38106 Braunschweig, Germany
* Corresponding author

This abstract is available from: http://www.journal.chemistrycentral.com/content/3/S1/P17

Refined nearest neighbour analysis was recently introduced for the analysis of bias introduced by benchmark datasets into the validation of virtual screening techniques [1]. It is based on spatial statistics [2] and provides a mathematical framework for the nonparametric analysis of mapped point patterns.

Here, the application of refined nearest neighbour analysis to the design of benchmark datasets for virtual screening based on PubChem bioactivity data is presented. Datasets of compounds active against pharmaceutically relevant targets were purged from unselective hits by a data cantered workflow. Topological optimization using experimental design strategies monitored by refined nearest neighbour analysis functions was applied to generate corresponding datasets of actives and decoys that are as unbiased as possible with regard to analogue bias and artificial enrichment.

These datasets provide a tool for Maximum Unbiased Validation (MUV) of virtual screening methods. The datasets and a software package implementing the MUV design workflow are freely available at http://www.pharmchem.tu-bs.de/lehre/baumann/MUV.html.

Using a large number of designed sub-samples from established VS benchmark datasets, a general regression model was developed that allows the prediction of the estimated performance of VS methods. Based on this model and the MUV datasets, SAFE (Scale for Assessing Figures of Effectiveness) an open, standardized, platform for the benchmarking of virtual screening methods was developed. SAFE will allow developers to benchmark novel methods against established ones and facilitate the selection of the best VS tool for researchers on prospective virtual screening projects.

References

1. Rohrer SG, Baumann K: **Impact of Benchmark Data Set Topology on the Validation of Virtual Screening Methods: Exploration and Quantification by Spatial Statistics.** *J Chem Inf Model* 2008, **48:**704-718.
2. Fortin MJ, Dale MRT: *Spatial Analysis: A Guide for Ecologists Cambridge University Press: Cambridge, UK*; 2005.