

Poster presentation

Open Access

Topological pharmacophores: pros & cons of QSARs based on 2D pharmacophore fingerprints

D Horvath

Address: UMR 8576 CNRS – University of Lille, 59655 Villeneuve d'Ascq, France
from 3rd German Conference on Chemoinformatics
Goslar, Germany. 11-13 November 2007

Published: 26 March 2008

Chemistry Central Journal 2008, 2(Suppl 1):P3 doi:10.1186/1752-153X-2-S1-P3

This abstract is available from: <http://www.journal.chemistrycentral.com/content/2/S1/P3>

© 2008 Horvath

Topological (2D) Fuzzy Pharmacophore Triplets (2D-FPT), using the number of interposed bonds to measure separation between the atoms representing pharmacophore types, were employed to establish and validate Quantitative Structure-Activity Relationships (QSAR). Thirteen data sets for which state-of-the-art QSAR models were reported in literature were revisited in order to benchmark 2D-FPT biological activity-explanatory propensities. Linear and non-linear QSAR models were constructed for each compound series (following the original author's splitting into training/validation subsets) with three different 2D-FPT versions, using the Genetic Algorithm (GA)-based Stochastic QSAR Sampler (SQS) to pick relevant triplets and fit their coefficients.

Beyond the benchmarking studies, in which 2D-FPT models validated better than, or as well as the best among reported literature equations, including 3D overlay-dependent computer-intensive techniques, this work is concerned with the in-depth analysis of the 'topological pharmacophores' defined by the sets of triplets entering the models. Do these really pinpoint the actual functional groups involved in the non-covalent binding of the ligand to the active site? May the use of topological distances replace the actual 3D distances in the bioactive conformer? On hand of an in-depth study of thrombin inhibitor models, it was shown that triplets entering a model are often at risk of having been selected because of correlation artefacts caused by the low diversity of the training/validation set. The initial set of models were, robust validation scores notwithstanding, either unable to recognize any of the structurally different actives they were challenged with, or else tended to predict any molecule outside the training set as active. Fortunately, while CoMFA restricts training molecule diversity in order to get meaningful overlays, 2D-FPT models can (and should!) be

trained on sets of compounds of arbitrary diversity – addition of assumed inactive random compounds to the structurally homogeneous thrombin inhibitor series produced models that successfully extrapolate to molecules of different topology.