

Poster presentation

Two-step hierarchical assignments on molecular graphs

A Jahn*, N Fechner, G Hinselmann and A Zell

Address: University of Tübingen, Sand 1, 72076 Tübingen, Germany

* 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):P13 doi:10.1186/1752-153X-3-S1-P13

This abstract is available from: <http://www.journal.chemistrycentral.com/content/3/S1/P13>

© 2009 Jahn et al; licensee BioMed Central Ltd.

Measures for the similarity of molecules are of interest for several *in silico* based tasks like virtual screening or de novo structure design. The Optimal Assignment Kernel (OAK) [1] is a successful similarity measure, although it is not a valid kernel, since the function is not positive definite [2]. Careful investigations of the assignment on the atom level disclose that the optimal assignment with the Hungarian algorithm may result in topological errors. These errors are mappings of atoms from chemical substructures like ring systems to atoms of the other molecule, which belong to different substructures or even can be scattered among the molecule. This yields an overall higher similarity score but is problematic from a chemical point of view. To avoid these topological errors we developed a two-step hierarchical assignment method and compared it with the OAK.

As a pre-processing step, our method separates the molecules in disjunctive molecular fragments like aromatic systems, rings and conjugated environments. The first assignment step maps these fragments of the molecules to corresponding fragments and guarantees a substructure preserving mapping at the atom level in the second assignment step. A special penalty function penalizes mappings between atoms from different substructures, which were not mapped in the first step. This function also adjusts the similarity score of mappings from atoms included in fragments to atoms, which were not part of a fragment. These modifications reduce the probability of topological errors and produce assignments with a reasonable mapping between molecular substructures.

Virtual screening results of the OAK and the hierarchical assignment approach on several datasets from the directory of useful decoys (DUD) [3] showed that the hierarchical assignment achieved better BEDROC scores. [4] This performance gain is the result of the penalty of topological errors and ensures an improved distinction between biologically active and inactive compounds.

References

1. Fröhlich H, et al.: *QSAR Comb Sci* 2006, **25**:317-326.
2. Vert J-P: *Technical Report HAL-00218278* 2008.
3. Huang N, et al.: *J Med Chem* 2006, **49**:6789-6801.
4. Truchon J-F, Bayly Cl: *J Chem Inf Model* 2007, **41**:488-508.