

POSTER PRESENTATION

Open Access

Introducing fuzziness into maximum common substructures for meaningful cluster characterisation

Christian Herhaus

From 9th German Conference on Chemoinformatics
Fulda, Germany. 10-12 November 2013

Arranging similar structures in clusters is one of the typical tasks of modern Cheminformatics with high impact in HTS follow-up, generation of structure activity relationships (SAR) and selection of starting points for compound optimisation. Methods for cluster generation are as diverse as the structures which they are applied to [1], may they be e.g. similarity- or substructure-based. Typically, medicinal chemists tend to orientate themselves in structure subsets like clusters with the help of substructures, so-called “scaffolds”, which intuitively characterise the structural relationships between the molecules of the subset. In the case of substructure-based clustering, well established methods are existing for the generation of Maximum Common Substructures (MCS) which are present in all members of the structure population or a defined proportion thereof [2]. But in the case of similarity-based clusters, such MCS may either not be existing for the required dataset proportion or the common substructure may be so small that it is no longer representative and therefore meaningless.

The approach presented here allows the generation of MCS also for similarity-based clusters with a given inherent structural diversity. It does so by generating an MCS of reduced graphs in a first step, followed by mapping atom and bond indexes of this reduced MCS onto the full structures and aggregation of atom and bond information for each indexed atom/bond. In a final step, query features of the MDL SDF format (atom lists, query bonds) are utilized to map aggregated element and bond information onto the reduced MCS. As a result, “fuzziness” in atom and bond information is

added to the MCS which, although still being fully database-searchable, is more meaningful for the characterisation of clusters as it can cover larger parts of the full structures than a conventional MCS could do. The approach was implemented in Pipeline Pilot™ for proof of concept but is general enough to be transferred to other technical platforms as well.

Published: 11 March 2014

References

1. Downs GM, Barnard JM: **Clustering Methods and Their Uses in Computational Chemistry**. In *Reviews in Computational Chemistry*. Volume 78. Chichester: Wiley and Sons; 2003:1-40.
2. Ehrlich HC, Rarey M: **Maximum common subgraph isomorphism algorithms and their applications in molecular science: a review**. *WIREs Comput Mol Sci* 2011, **1**:68-79.

doi:10.1186/1758-2946-6-S1-P17

Cite this article as: Herhaus: Introducing fuzziness into maximum common substructures for meaningful cluster characterisation. *Journal of Cheminformatics* 2014 **6**(Suppl 1):P17.

Publish with **ChemistryCentral** and every scientist can read your work free of charge

“Open access provides opportunities to our colleagues in other parts of the globe, by allowing anyone to view the content free of charge.”

W. Jeffery Hurst, The Hershey Company.

- available free of charge to the entire scientific community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

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


ChemistryCentral

Correspondence: christian.herhaus@merckgroup.com
Global Computational Chemistry, Merck Serono, Darmstadt, 64293, Germany

 **ChemistryCentral**

© 2014 Herhaus; licensee Chemistry 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. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.