Kazius *Journal of Cheminformatics* 2010, **2**(Suppl 1):P51
http://www.jcheminf.com/content/2/S1/P51

Journal of Cheminformatics

POSTER PRESENTATION Open Access

Get the best from substructure mining

Jeroen Kazius

From 5th German Conference on Cheminformatics: 23. CIC-Workshop
Goslar, Germany. 8-10 November 2009

The chemical information that is present in a set of compounds is rarely fully exploited. This is mostly because no descriptor set can capture all biologically important features. As a result, valuable chemical knowledge can thus stay hidden from hypothesis-based drug design. The simplest form of a structure-activity relationship (SAR) is a substructure that predisposes compounds towards reduced or increased biological activity. Such simple patterns should not be missed during drug design.

The aim of substructure mining is to present those substructures that are most likely related to biological activity. This method thus provides rapid access to a substantial repertoire of chemical descriptors that otherwise remains hidden: *substructures*. In short, substructure mining consists of a focused, but exhaustive, series of substructure searches.

This poster describes how AweSuM, the new *Awe*some *Su*bstructure *M*ining tool from Curios-IT, was employed to learn the most interesting substructures. The poster also discusses the value of enriching the data with 2D pharmacophore information prior to mining. An enriched, detailed SAR analysis produced a scaffold that summarises the chemical content of datasets better than any standard substructure. The pharmacophore that AweSuM extracted shows predictive power and agrees with published chemical knowledge. These results demonstrate that useful SAR knowledge can be extracted from the vast space of substructure descriptors. In this way, AweSuM reveals key substructures (e.g., pharmacophores or toxicophores), which can often be predictive for biological activities.

Published: 4 May 2010

doi:10.1186/1758-2946-2-S1-P51
Cite this article as: Kazius: Get the best from substructure mining. *Journal of Cheminformatics* 2010 **2**(Suppl 1):P51.

Curios-IT, Anna van Burenhof 63, 2316GP Leiden, The Netherlands