New method to perform similarity search in chemical reactions domain.

Frank Hoonakker[1,2], Alexander Varnek[1*]

[1] *Laboratoire d'Infochimie, UMR 7177 CNRS, Université Louis Pasteur, 4, rue B. Pascal, Strasbourg 67000, France, e-mail : varnek@chimie.u-strasbg.fr*

[2] *NOVALYST DISCOVERY, 23 rue du Loess BP 20, 67037 Strasbourg CEDEX 02,France*

Abstract :

Similarity search methods are widely used in computer-aided design of new compounds possessing desirable properties. The basic assumption is the *similar property principle* which states that similar chemical structures possess similar physicochemical properties. While this approach is well developed for individual compounds, it is still very little applied for the chemical reactions because of the difficulties to present simultaneously reactants and products of a given reaction. This problem might be solved if a chemical reaction is represented as a Condensed Graph of Reaction (CGR) involving both standard bond types (simple, double, etc.) and dynamical bonds (simple transformed to double, etc.). Our method implies fragmentation of CGR into atom/bond sequences involving dynamical bonds. Two types of fingerprints are suggested. In the first one each bit corresponds to the presence or absence of a particular fragment, whereas in the second one the occurrences of fragments are taken into account. A number of tests performed on the data set of about 200.000 reactions demonstrate the performance of reaction similarity search over traditional sub-structural search technique.