

## [P36] Generative Topographic Mapping for chemical reactions analysis: Evolution of the reaction space between 1976 and 2016

Bort William<sup>1</sup>, Bonachera Fanny<sup>1</sup>, Madzhidov Timur<sup>2</sup>, Varnek Alexandre<sup>1</sup>

[ 1 ] Univ Strasbourg, CNRS, Lab Chemoinformat, UMR 7140, 4 Rue Blaise Pascal, F-67081 Strasbourg, France

[ 2 ] Kazan Fed Univ, Butlerov Inst Chem, Lab Chemoinformat & Mol Modeling, Kremlevskaya 18, Kazan, Russia

One of the main successes of modern chemistry is the development of synthetic procedures for rapid and mass production of chemical substances. In this work the evolution of chemical reactions used in the synthesis is studied. The approach is based on application of data visualization techniques such as Generative Topographic Mapping<sup>1</sup> (GTM) for exploring time evolution of chemical space of reactions.

The dataset of more than 3 million reactions extracted from patents<sup>2</sup> published in 1976-2016 was used. Reactions were encoded by Condensed Graphs of Reactions<sup>3</sup> approach for which ISIDA fragment descriptors were calculated. Since the initial dataset was very large, 1% portion of data (called frameset) was used for map preparation, and the rest of the data was projected on it forming all-embracing map of reaction space. Random and reaction center-based strategies for frame set selection were tried and compared.

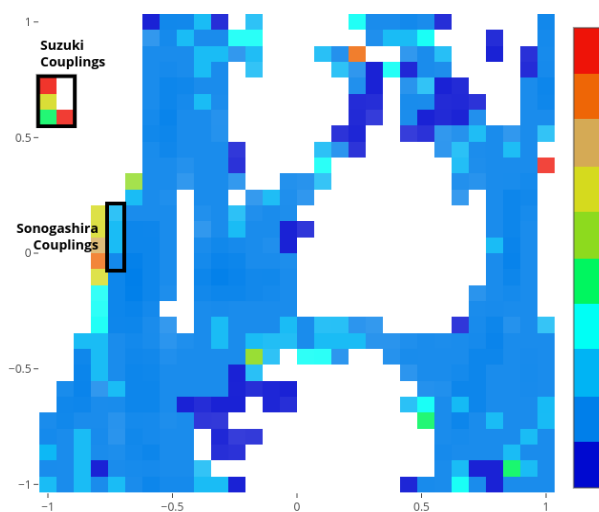


Figure 1 : Map of 1992 (RED) versus all other years previous (BLUE) for Augmented Atoms descriptors with an equally distributed frameset.

Each zone on the maps is associated with a reaction type and the coloring used allow trends to be visualized.

Local tendencies, such as the rise of popularity in Suzuki couplings in industrial synthesis since the 1990's (figure 1), have been identified. A 10 year-gap separates the invention of the Suzuki reaction from its use in the industry while Sonogashira couplings took less than 5 years to transition. This could be explained by historical and/or chemical arguments

The study of specific reactions behavior through GTM maps gives an insight into the transition from academic knowledge to industrial processes and helps understand the factors that determine the time needed for an adaption to the industry.

Á

TM acknowledges Russian Science

Foundation (grant 14-43-00024) for the support.

### Bibliography :

[1] C. M. Bishop, M. Svensén. *Neural Computations*. 10(1) : 215-234, 1998.

[2] D. M. Lowe. *Extraction of chemical structures and reactions from the literature*. PhD thesis, University of Cambridge, 2012.

[3] F. Hoonakker, N. Lachiche, A. Varnek, A. Wagner. *Internation Journal of Artificial Intelligence Tools*. 20(2):253-270, 2011.