Journal of Chemical Information and Modeling 2016 vol.56 N11, pages 2140-2148

Automatized assessment of protective group reactivity: A step toward big reaction data analysis

Lin A., Madzhidov T., Klimchuk O., Nugmanov R., Antipin I., Varnek A.
Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

© 2016 American Chemical Society.We report a new method to assess protective groups (PGs) reactivity as a function of reaction conditions (catalyst, solvent) using raw reaction data. It is based on an intuitive similarity principle for chemical reactions: similar reactions proceed under similar conditions. Technically, reaction similarity can be assessed using the Condensed Graph of Reaction (CGR) approach representing an ensemble of reactants and products as a single molecular graph, i.e., as a pseudomolecule for which molecular descriptors or fingerprints can be calculated. CGR-based in-house tools were used to process data for 142,111 catalytic hydrogenation reactions extracted from the Reaxys database. Our results reveal some contradictions with famous Greene's Reactivity Charts based on manual expert analysis. Models developed in this study show high accuracy (ca. 90%) for predicting optimal experimental conditions of protective group deprotection.

http://dx.doi.org/10.1021/acs.jcim.6b00319