Expert system for predicting reaction conditions: The Michael reaction case

Marcou G., Aires De Sousa J., Latino D., De Luca A., Horvath D., Rietsch V., Varnek A.

Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

© 2015 American Chemical Society. A generic chemical transformation may often be achieved under various synthetic conditions. However, for any specific reagents, only one or a few among the reported synthetic protocols may be successful. For example, Michael β-addition reactions may proceed under different choices of solvent (e.g., hydrophobic, aprotic polar, protic) and catalyst (e.g., Brønsted acid, Lewis acid, Lewis base, etc.). Chemoinformatics methods could be efficiently used to establish a relationship between the reagent structures and the required reaction conditions, which would allow synthetic chemists to waste less time and resources in trying out various protocols in search for the appropriate one. In order to address this problem, a number of 2-classes classification models have been built on a set of 198 Michael reactions retrieved from literature. Trained models discriminate between processes that are compatible and respectively processes not feasible under a specific reaction condition option (feasible or not with a Lewis acid catalyst, feasible or not in hydrophobic solvent, etc.). Eight distinct models were built to decide the compatibility of a Michael addition process with each considered reaction condition option, while a ninth model was aimed to predict whether the assumed Michael addition is feasible at all. Different machine-learning methods (Support Vector Machine, Naive Bayes, and Random Forest) in combination with different types of descriptors (ISIDA fragments issued from Condensed Graphs of Reactions, MOLMAP, Electronic Effect Descriptors, and Chemistry Development Kit computed descriptors) have been used. Models have good predictive performance in 3-fold cross-validation done three times: balanced accuracy varies from 0.7 to 1. Developed models are available for the users at http://infochim.u-strasbg.fr/webserv/VSEngine.html. Eventually, these were challenged to predict feasibility conditions for ~50 novel Michael reactions from the eNovalys database (originally from patent literature).