LABORATORY FOR CHEMICAL TECHNOLOGY
Technologiepark 125 9052 Gent, Belgium

Coach	**Supervisor(s)**	**Funding**
Maarten Dobbelaere	Kevin Van Geem Chris Stevens	FWO European Commission

Predicting Solvents for Chemical Reactions with Machine Learning

Aim

The aim of this master thesis is to apply machine learning techniques to predict, for a given chemical reaction, which solvent is best suited so that more sustainable pharmaceutical production processes can be designed.

Justification

Bringing a new drug to the market takes on average 15 years, costs over 1 billion euro, and has an enormous environmental impact. After medical experts define the necessary target of a drug, potential drug candidates are synthesized, and passed to later stages of drug testing (such as preclinical and clinical phases), so that – in an ideal case – one drug will be successful. In reality, the success rate of drug development is strongly limited by the number of proposed drug candidates that can effectively be synthesized. Therefore, researchers have developed new machine learning-based tools that can predict for a desired small molecule, how it can be synthesized from commercially available materials. While the accuracy of computer-aided synthesis planning software is reasonably good, a large human effort is needed to determine at which reaction conditions these reactions can be performed. One of the key reaction conditions is the choice of a good solvent. There are several ways to select a solvent today. The first method is using experience from other reactions and hence trying out a solvent that worked in other cases. Another method uses solubility parameters, such as the ones developed by Kamlet and Taft, Hansen, or Abraham. All of these methods are limited to solvents that are already used. With the very restricting legislation in Europe, that requires researchers to select sustainable solvents, it is crucial to develop new green solvents. Machine learning is a method that learns from data and is excellently suited to predict new solvents, that would never be selected with the current existing selection tools.

Program

- Literature survey on solvent selection methods and on molecular design via machine learning
- Curation of a large chemical database with solvation properties
- Development/extension and application of existing machine learning models for solvation properties (python)
- Sanity check: evaluation of machine learning results and comparison with solvation chemistry rules