# *Inverse design of chemical reagents for fast and sensitive detection of pesticides in the environment using mass spectrometry*

Henrik Hupatz, Berit Olofsson, Anneli Kruve

henrik.hupatz@mmk.su.se
*Stockholm University Center of Circular and Sustainable Systems (SUCCeSS), Svante Arrhenius väg 16C, 114 16 Stockholm, Sweden*

## BACKGROUND

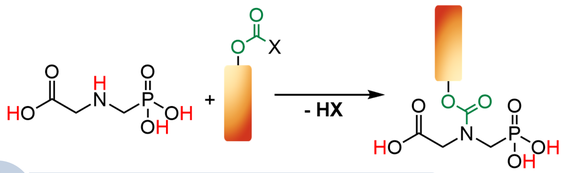
- **Accumulation** of polar **pesticides** in the environment is **threatening** water quality
- A **fast** and **sensitive analysis** is required (Limit of quantification: 0.1 µg/l)
- **Mass spectrometry** (MS) coupled with **chromatography** is a sensitive analytical method for a wide range of analytes
- **Derivatization** can facilitate the analysis of challenging **highly polar** small molecules

## AIM

1. Development of a **machine learning** model for the **inverse molecular design** of easily **accessible** derivatizing reagents for **glyphosate** with **high sensitivity** and improved **sustainability** of chemical analysis.
2. **Verify** predicted properties of reagents by **synthesis** and **practical analysis** of **water samples**.

## PRELIMINARY RESULTS

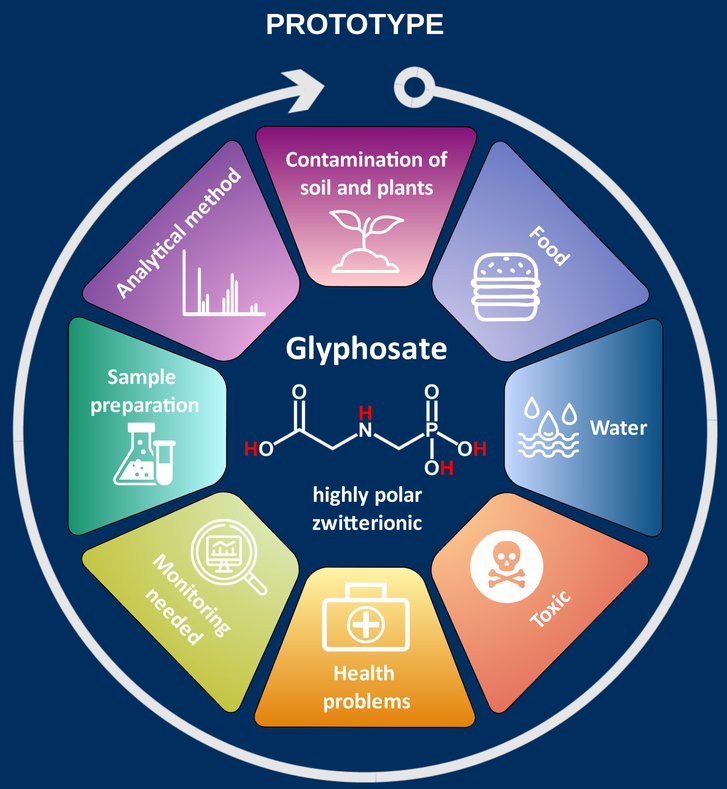
- **Sensitivity** of an analyte in **electrospray ionization** (ESI-MS) is expressed with the **ionization efficiency** (*IE*)
- Our **machine learning** algorithm **predicts log*IE*** for given structures based on **molecular fingerprints** or **PaDEL descriptors**
- **Monte Carlo tree search** (MCTS) generates structures (orange) with **improved** range for **log*IE***, **log*p*** and synthetic accessibility (**SAScore**) compared to **established reagents**



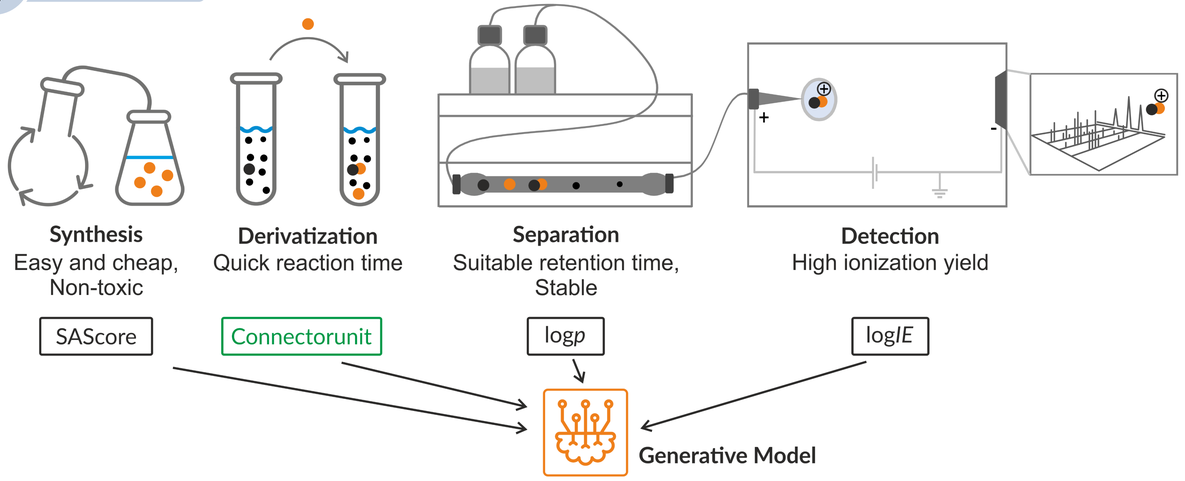

## SUGGESTIONS APPRECIATED

**Neural network** based **property-guided** generative models might yield higher **quality** structures and cover a wider **chemical space**. Which machine learning methods might be well suited?

# Can **predicted ionization efficiencies** facilitate the **inverse design** of **derivatizing** reagents for the **sensitive detection** of polar pesticides by **mass spectrometry?**

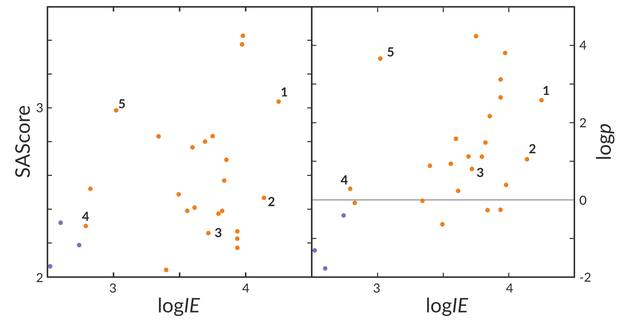


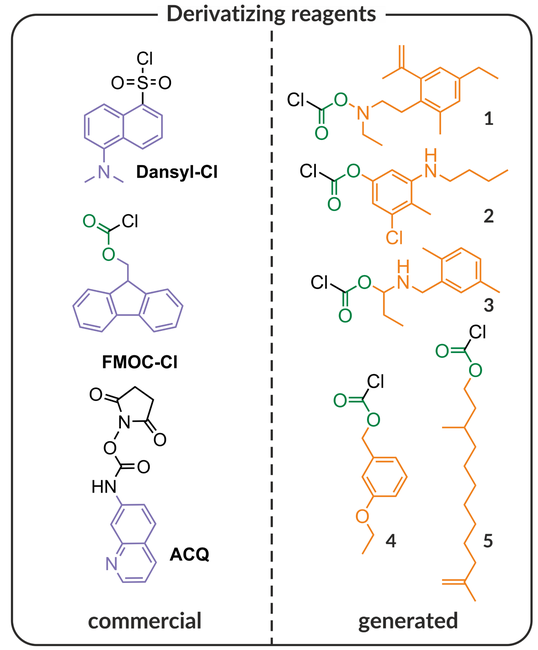

## METHODS




## MONTE CARLO TREE SEARCH

- **Molecular graphs generated** by MCTS from **carbon seed** to molecular weight of **400** with MCTS **depth of 2** and **width of 12**
- **Evaluation function:** *Ev* = log*IE* – SAScore
- For each MCTS the structure with highest *Ev* was picked




## GENERATIVE MODEL

- SMILES or molecular graphs?
- Which architecture? Variational autoencoders or recurrent neural networks?
- How to train the model efficiently? Reinforcement learning.

*For references, additional information and a poster copy:*

Kruve Lab
Stockholm University
SUCCeSS