BIOTRANS
LA ROCHELLE 2023

N°109 /
TOPIC(s) : Artifical intelligence / computational methods / Industrial biocatalysis

# Fueling the Digital Revolution in Biocatalysis with Language Models

## AUTHORS
Teodoro LAINO / IBM RESEARCH EUROPE, SAEUMERSTRASSE 4, RUESCHLIKON

## PURPOSE OF THE ABSTRACT

Artificial Intelligence (AI) and Data are driving one of the most notable revolutions in organic chemistry. The use of experimental data, extracted from public documents or collected using ELNs, emerged as one of the most effective, scalable approaches for capturing human knowledge, modelling and improving chemical processes. Machine learning tasks demonstrated high quality and ease of use in problems such as predicting chemical reactions [1-2], retrosynthetic routes [3], digitizing chemical literature [4], predicting detailed experimental procedures [5], designing new fingerprints [6] and yield predictions [7]. In this talk, I'll briefly introduce the use of AI as the cornerstone of the accelerated discovery method, in which intelligent workflows are used to collect and synthesize known information, augment the known data with rich simulation, predict potential solutions with the desired attributes based on data-driven models.

The rest of the presentation will focus on recent applications of language models to the field of enzymatic catalysis [8], including the development of retrosynthetic strategies as a mean to promote sustainability and green chemistry at scale [9].

[1] IBM Research Europe, Chem. Sci., 2018, 9, 6091-6098
[2] IBM Research Europe, ACS Cent. Sci. 2019, 5, 9, 1572-1583
[3] IBM Research Europe, Chem. Sci., 2020, 11, 3316-3325
[4] IBM Research Europe, Nat. Comm., 2020, 11, 3601
[5] IBM Research Europe, Nat. Comm., 2021, 12, 2573
[6] IBM Research Europe, Nat. Mach. Intel., 2021, 3, 144?152
[7] IBM Research Europe, Mach. Learn.: Sci. Technol., 2021, 2, 015016
[8] IBM Research Europe, Nat Comm., 2022, 13, 964
[9] https://rxn.res.ibm.com

## FIGURES

FIGURE 1

FIGURE 2

## KEYWORDS

## BIBLIOGRAPHY