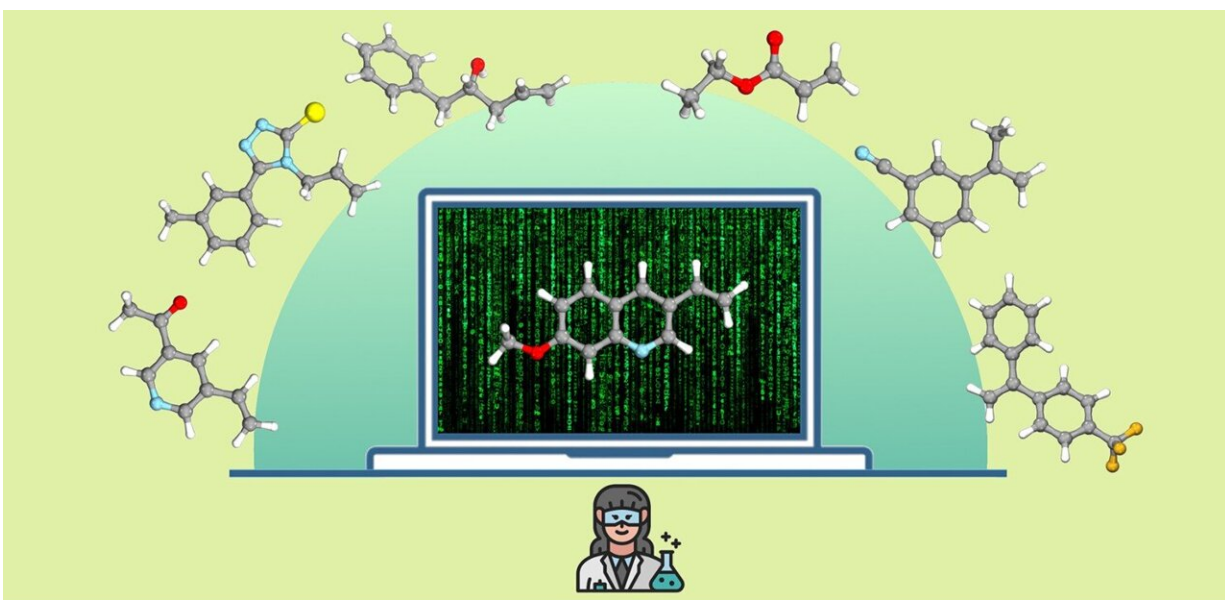# New strategy for assessing the applicability of reactions

April 10 2024

The team led by Prof Frank Glorius transferred all approved active pharmaceutical ingredients into a digital code. Credit: Uni MS – Debanjan Rana/ Glorius Group

Chemists often develop and optimize new chemical reactions using so-called model systems, i.e., simple, easily accessible substrates. They then use up to around 100 other substrates as examples to show that the reaction works. This demonstration of versatile applicability is called "scope" in technical jargon.

However, a subjective selection of substrates often results in a distorted picture of the range of applications of the newly developed reaction. It is often unclear whether it can be used to synthesize a desired product. To address this problem, a team led by chemist Prof Frank Glorius from the University of Münster (Germany) is proposing a computer-aided, bias-free method for selecting the model substrates to evaluate new chemical reactions.

The selection of substrates is based on the complexity and structural properties of real pharmaceutical compounds. "Our method aims to improve the quality and information content of chemical reaction data in the future and to close knowledge gaps," explains Glorius.

A deeper understanding of new reactions lowers the barriers to their application in both an academic and industrial context. The availability of high-quality, unbiased data also significantly facilitates the use of machine learning and paves the way for a more comprehensive use of the data. The work has been published in the journal *ACS Central Science*.

According to the team's authors, attempts to standardize and objectify the development and evaluation of chemical reactions are still quite new and relatively uncommon. "We would like to initiate a 'rethinking process' with our publication. Instead of doing as many experiments as possible, which are often biased or have a predictable outcome, the focus should be on obtaining the best possible data about new chemical reactions," says first author Debanjan Rana.

Other scientists have also tried to evaluate chemical reactions on the basis of "better" selected substrates. However, this work was limited to special cases—either to firmly selected structures with pharmaceutical relevance or to structures specially tailored for a single reaction, which have to be calculated and selected in a complex process.

In contrast to the previous work, the method presented by the Münster team takes the entire structure of a molecule into account, which makes it universally applicable to any chemical reaction.

Niklas Hölter, one of the paper's authors in Münster, explains the thought process behind the study: "Scope is of central importance in all publications on chemical synthesis. However, chemists are often biased in their choice of substrate compounds to test.

"For example, they choose substrates that are structurally very simple, very similar to the model substrate or simply just available in the laboratory ('selection bias'). They often don't mention unsuccessful reactions at all in their publication in order to paint a better picture ('reporting bias')."

When synthesizing new chemical compounds, such as active ingredients or materials, chemists have to select the most suitable method for producing the target compound from a large number of known chemical reactions and methods. To do this, they consider several factors, such as the yield of the desired product as well as environmental and safety aspects. The development of new, versatile chemical reactions therefore continues to be a focus of current chemical research.

The method developed by the team at the University of Münster used molecular fingerprints to transfer all approved active pharmaceutical ingredients into a digital code. Using unsupervised machine-learning and clustering methods, they created a model that divides this "space" of active pharmaceutical ingredients into chemically meaningful regions based on molecular structures.

To evaluate a new chemical reaction, thousands of potential test substrates can be projected into the same space using the machine-learning model. A test substrate is automatically selected from the center

of each of the previously identified regions to cover the entire space without bias.