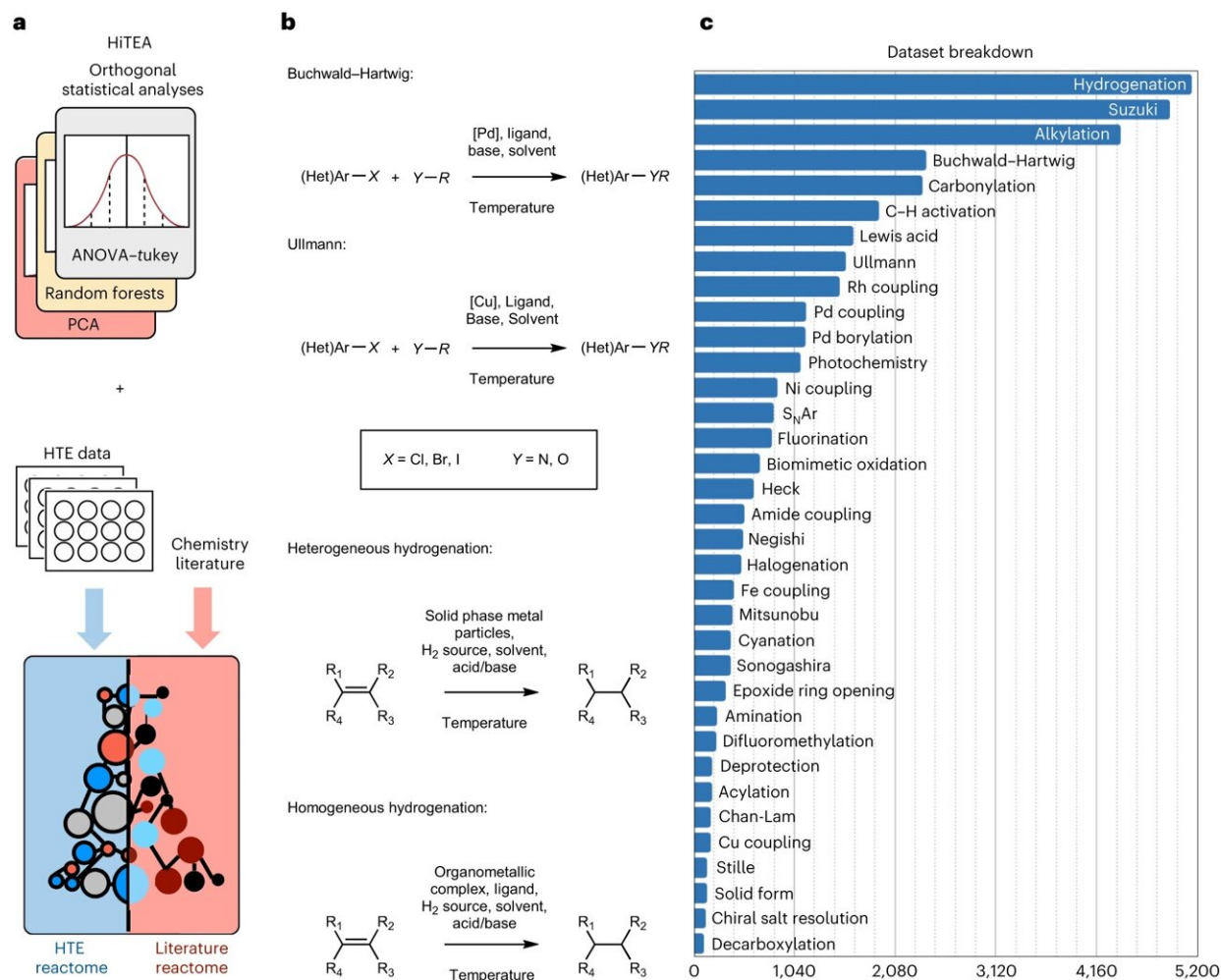


Accelerating how new drugs are made with machine learning

January 15 2024



Overview of the HTE dataset and framework. **a**, Overview of HiTEA and its analysis is shown. Comparison of the literature reactome with HiTEA reactome will reveal support for our mechanistic conclusions (agreement of reactomes) or reveal areas of bias/unusual chemical phenomenon (disagreement of reactomes).

b, Abstracted representations of the four reaction classes analyzed by HiTEA in this publication are shown. c, Breakdown of the HTE dataset by reaction class is shown. Credit: *Nature Chemistry* (2024). DOI: 10.1038/s41557-023-01393-w

Researchers have developed a platform that combines automated experiments with AI to predict how chemicals will react with one another, which could accelerate the design process for new drugs.

Predicting how [molecules](#) will react is vital for the discovery and manufacture of new pharmaceuticals, but historically this has been a trial-and-error process, and the reactions often fail. To predict how molecules will react, chemists usually simulate electrons and atoms in simplified models, a process that is computationally expensive and often inaccurate.

Now, researchers from the University of Cambridge have developed a data-driven approach, inspired by genomics, where automated experiments are combined with machine learning to understand [chemical reactivity](#), greatly speeding up the process. They've called their approach, which was validated on a dataset of more than 39,000 pharmaceutically relevant reactions, the chemical "reactome."

Their [results](#), reported in the journal *Nature Chemistry*, are the product of a collaboration between Cambridge and Pfizer.

"The reactome could change the way we think about [organic chemistry](#)," said Dr. Emma King-Smith from Cambridge's Cavendish Laboratory, the paper's first author. "A deeper understanding of the chemistry could enable us to make pharmaceuticals and so many other useful products much faster. But more fundamentally, the understanding we hope to generate will be beneficial to anyone who works with molecules."

The reactome approach picks out relevant correlations between reactants, reagents, and performance of the reaction from the data, and points out gaps in the data itself. The data is generated from very fast, or high throughput, automated experiments.

"High throughput chemistry has been a game-changer, but we believed there was a way to uncover a deeper understanding of chemical reactions than what can be observed from the initial results of a high throughput experiment," said King-Smith.

"Our approach uncovers the hidden relationships between reaction components and outcomes," said Dr. Alpha Lee, who led the research. "The dataset we trained the model on is massive—it will help bring the process of chemical discovery from trial-and-error to the age of big data."

In a related [paper](#), published in *Nature Communications*, the team developed a machine learning approach that enables chemists to introduce precise transformations to pre-specified molecule regions, enabling faster drug design.

The approach allows chemists to tweak [complex molecules](#)—like a last-minute design change—without having to make them from scratch. Making a molecule in the lab is typically a multi-step process, like building a house. If chemists want to vary the core of a molecule, the conventional way is to rebuild the molecule, like knocking the house down and rebuilding from scratch. However, core variations are important to medicine design.

A class of reactions known as late-stage functionalization reactions attempts to directly introduce chemical transformations to the core, avoiding the need to start from scratch. However, it is challenging to make late-stage functionalization selective and controlled—there are

typically many regions of the molecules that can react, and it is difficult to predict the outcome.

"Late-stage functionalizations can yield unpredictable results and current methods of modeling, including our own expert intuition, isn't perfect," said King-Smith. "A more predictive model would give us the opportunity for better screening."

The researchers developed a machine learning model that predicts where a molecule would react, and how the site of reaction vary as a function of different reaction conditions. This enables chemists to find ways to precisely tweak the core of a molecule.

"We pretrained the model on a large body of spectroscopic data—effectively teaching the model general chemistry—before fine-tuning it to predict these intricate transformations," said King-Smith. This approach allowed the team to overcome the limitation of low data: there are relatively few late-stage functionalization reactions reported in the scientific literature. The team experimentally validated the [model](#) on a diverse set of drug-like molecules and was able to accurately predict the sites of reactivity under different conditions.

"The application of machine learning to chemistry is often throttled by the problem that the amount of data is small compared to the vastness of chemical space," said Lee. "Our approach—designing models that learn from large datasets that are similar but not the same as the problem we are trying to solve—resolve this fundamental low-data challenge and could unlock advances beyond late stage functionalization."

More information: Emma King-Smith et al, Probing the chemical 'reactome' with high-throughput experimentation data, *Nature Chemistry* (2024). [DOI: 10.1038/s41557-023-01393-w](https://doi.org/10.1038/s41557-023-01393-w)

Predictive Minisci Late Stage Functionalization with Transfer Learning,
Nature Communications (2024). DOI: [10.1038/s41467-023-42145-1](https://doi.org/10.1038/s41467-023-42145-1).
www.nature.com/articles/s41467-023-42145-1

Provided by University of Cambridge

Citation: Accelerating how new drugs are made with machine learning (2024, January 15)
retrieved 28 April 2024 from <https://phys.org/news/2024-01-drugs-machine.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.