Imperial College London

Digital Chemistry: A new era for digital molecular design and fabrication

Imperial Business Partners - Technical Digest Series

Professor Sophia Yaliraki
Professor of Theoretical Chemistry
Faculty of Natural Sciences, Department of Chemistry
Institute of Digital Molecular Design and Fabrication (DigiFab)
Imperial College London

As state-of-the-art labs with automated facilities increasingly generate a wealth of new and unique data in both academic and industrial settings, a new era in digital molecular design and fabrication in synthesis of small molecules, materials, vaccines and pharmaceuticals is moving away from slow, labour-intensive methods to highly automated and data-driven approaches.

The practice of Chemistry in academia and industry is being revolutionised by advances in automation, artificial intelligence, and data-based methods. Computational modelling can now allow predictive insights into the behaviour of complex molecules and systems, whilst parallel experiments can be monitored in real-time to give spectroscopic data readouts that allow rapid analysis of kinetics and yield optimisation. Artificial intelligence and machine learning are transforming our ability to predict reaction outcomes, plan complex multistep syntheses and understand small-molecule-macromolecule interactions to name just a few advances.

Automated closed-loops experiments aim to start from the design of experiments, using automated platforms optimising a set of reactions with data capture and on-the-fly data analysis, to lead to optimal results or new chemicals discovery with precise properties. The harnessing of the digital approaches in the chemical sector will lead to faster, more efficient and reproducible synthesis with a sustainable and more environmentally friendly approach. With scaling up factored from the start of the process, smart and resource efficient manufacturing will bring renewed benefits to society and the economy.

The acquisition of high quality data will allow for optimisation of the route design to achieve the synthesis of a molecular target in an expedient manner with minimal waste. In the longer term, the data acquired will allow for more accurate prediction of reaction outcomes and, ultimately, to the prediction of unknown reactions or access to unknown chemical space of materials with predesigned properties in diverse contexts ranging from the environment to health.

The need for data to inform automation before, during and after the process is at the heart of this digital approach. However, a major challenge is the curation, integration and sharing of old and new data in an open manner.

As the context of existing data is important, standardisation of captured data is essential. Only by addressing this oversight will we be able to reap the potential of Digital Chemistry through stewardship, management, sharing and integration of different types of data (previously generated, as well as new and emerging forms). This may be achieved by establishing (inter)-national standards, software development, architecture and data engineering, and use of good practice including reproducibility, design of experiments, workflows and Findable, Accessible, Interoperable and Reusable, (FAIR) data principles.

Another challenge comes from automated platforms that are universal and not just optimised for a particular system or set of reactions. This is especially so in heterogeneous or non-equilibrium systems. The automation of non-standard characterisation techniques is still on-going and although there are analytical techniques currently commercially available, there is a need for more specialised approaches that can keep up with the throughput of synthesis possible on platforms, for example, high-throughput screening of porosity, thermal characteristics, and viscosity. Finally, there is the issue of purification and isolation of products after high-throughput synthesis and whether it is possible to make synthetic routes more 'process-friendly' from the start to make isolation simpler.

The link from automated to fully autonomous system in a universal way is still missing and in the foreseeable future the human interaction will remain important in setting the goals and questions, interpreting results and identifying further avenues of investigation, to name just a few points. Life-long digital skills will be required in the chemical community as well as team work with data engineers, data scientists, roboticists and process engineers. Rather than eliminating jobs, this approach will create new jobs, where molecular scientists are conversant and work across and with other disciplines.

Digitisation has been the top agenda for the Chemical, Life Sciences, Agrochemical, Personal Care and Materials Industries, affecting every aspect of the value chain ranging from large corporations through to SME and start-ups leading innovation in Industry 4.0 and 5.0. COVID-19 has been accelerating the acceptance of digital technologies and is leading to a re-examination of supply chains across these industries. This has been evident for example in the design of the scaling up of vaccine production from the start of the process.

Chemical companies such as BASF and BAYER are adopting digitisation in their plants and processes and universities are supporting multidisiplinary Institutes in collaboration with industry. An example Institute is The Materials Innovation Factory at the University of Liverpool, which is bringing together materials chemistry experts together with cutting edge robotics and high-performance computing to accelerate materials discovery, and the Machine Learning for Pharmaceutical Discovery and Synthesis Consortium (MLPDS) at MIT. SMEs such as Deepmatter develop hardware and software for the digitisation of chemistry. Finally, new Centres for Doctoral Training (CDT) in this area such as Cambridge's SynTech which aims to connect chemistry, machine learning and robotics, and Imperial's rEaCT which brings together automation, data analytics and process optimisation, aim to fill the skills gap that will train the next generation of chemists in this area.

The aim to achieve autonomous, smart molecular fabrication by harvesting the data generated by rapid, parallel and online analyses of reaction processes – to produce new molecules and materials with designed properties – is in the 10-year goals of many industries essential for health, the environment and energy.