iDMT

Innovation Centre in Digital Molecular Technologies

Annual Report

2022

Yusuf Hamied Department of Chemistry along Lensfield Road, Cambridge
Nathan Pitt, © University of Cambridge

Produced by
Innovation Centre in Digital Molecular Technologies
Yusuf Hamied Department of Chemistry
University of Cambridge
Lensfield Road
Cambridge, CB2 1EW

iDMT@ch.cam.ac.uk
www.idmt.online

Foreword

by Professor Alexei Lapkin, Director of the iDMT

Founding the *i*DMT has been a long-term ambition of Prof Matthew Gaunt and myself, so it is with the greatest pleasure that I write my first introduction to our annual report.

The unique vision of the *i*DMT is to pose challenging scientific and technical questions that require innovation in academic research and industrial capabilities. Jointly the University, large industry partners and small and medium-sized enterprises (SMEs) will significantly speed-up the development of tools and services for revolutionising chemistry through digitalisation.

Establishing the Centre

The Grand Opening of the *i*DMT was held on Friday 26th November 2021. This hybrid event was composed of an opening ceremony, virtual lab tours and presentation contributions from: SMEs who have received support from the *i*DMT, University of Cambridge academic groups, and the *i*DMT's industrial partners. A recording of the event can be viewed on the iDMT's website.

One of the main focuses of the *i*DMT within the first year was on developing a new facility, which integrates high-throughput synthesis, analytics, chemical informatics, machine learning, robotics, and reaction engineering. Refurbishment works of the *i*DMT office and lab space were completed in December 2021, and commissioning of state-of-the-art equipment is well underway. The lab spaces have been fitted with gloveboxes, UPLC-MS systems, a liquid handling high throughput robot and a solid handling robot. Our custom designed hyperspectral instrument is expected before the end of the year.

First Year Highlights

The first *i*DMT-SME projects are well underway, spanning a range of activities across the digital chemistry space, including equipment vendors, both new and well-established chemical manufacturers, and companies who develop AI and machine learning tools. Starting from page 9, you will find case studies and snapshot introductions to some of the projects currently taking place in the centre.

A key element of the *i*DMT's success and longevity is the involvement of world-leading academics from the University of Cambridge who develop innovative ideas and methods for potential training and support of SMEs. To find out more about some of the research taking place in the centre, start on page 17 to read scientific synopses, prepared by our researchers and groups.

Looking Ahead

In the coming year, the *i*DMT team plans to continue building research and SME project momentum, and strengthening networking ties, to position the *i*DMT as a national resource. Further ahead, by mid-2023 the *i*DMT's current funding period will end and the *i*DMT team is currently establishing the next funding model. The *i*DMT is committed to continue leading science-based innovation, and operational development, until 2030 and beyond.

International Landscape

The World Economic Forum (2017) projected that the introduction of digital technologies will unlock $100 trillion for business and society. The positive impact in the UK from the introduction of digital technologies is estimated to be £435 billion by 2027 (Accenture report, 2017). Introduction of digital technologies, specifically in pharmaceuticals, is estimated to bring an additional £22.4 billion to the industry through cost savings, accident avoidance, carbon emissions decreases, and a reduction in the cost of clinical trials. Of this, £12 billion will come from new business models, enabled by digital technologies.

> The collaboration with *i*DMT allows us to tackle some of the grand challenges of automating chemical synthesis.
>
> *Dr Clive Green, Executive Director AstraZeneca*

Innovation in chemistry supports 90% of advances in other sectors including sport, health, food, buildings, electronics, and fashion. However, chemistry itself has not evolved, and access to novel functional molecules and materials continues to be a major bottleneck in many chemistry-using industries. This leads to staggering expenses for novel molecule development, excessive cost for the personalisation of medicines, and extremely long commercialisation times for new materials. Digital technologies have already revolutionised many areas of life, and the *i*DMT aims to make this revolution accessible to the chemicals manufacturing sector in the UK.

The revolution requires a transformational change in reaction discovery, process optimisation and chemical manufacturing. This involves a drastic increase in the throughput of chemical discovery and process development, achievable through automation of largely routine procedures, and adoption of artificial intelligence. The introduction of digital manufacturing in the chemical industry is best supported through highly innovative and flexible SMEs. Therefore, the *i*DMT aims to create an industrial base for implementing Industry 4.0 by supporting both new and existing SMEs in digital discovery and manufacturing of functional molecules.

> Digitalisation creates opportunities for completely new business models.
>
> *Professor Alexei Lapkin, Director of the iDMT*

About the *i*DMT

Hosted by the University of Cambridge's Yusuf Hamied Department of Chemistry, the *i*DMT is a project co-funded by the European Regional Development Fund (ERDF), the University of Cambridge, AstraZeneca and Shionogi. It has been established to support the transition of small and medium-sized enterprises (SMEs) in England towards the digital paradigm of R&D and manufacturing.

The *i*DMT offers a platform for the rapid development and testing of new products and services for the emerging digital economy, specifically in the sectors of chemical synthesis of molecules and materials. It includes two dedicated, newly refurbished laboratories, space for communication and computational facilities. The laboratories are equipped with robotic and automation equipment for chemical synthesis, as well as high-throughput analytical instruments. Visit our website or YouTube channel for virtual lab tours.

The objective of this centre is the development of a sustainable pipeline of new products and services for the digital transformation of chemical synthesis. For SMEs it is the opportunity to upskill in AI and other tools of digital molecular technologies, and to de-risk and tailor their new products for the emerging markets within Industry 4.0.

4
PARTNERS
co-funded the *i*DMT: the European Regional Development Fund (ERDF), the University of Cambridge, AstraZeneca and Shionogi.

£1.8m
INVESTMENT
in newly refurbished labs with state-of-the-art equipment for high throughput synthesis and bespoke robotics for automated chemistry.

LOCATION
hosted by the Yusuf Hamied Department of Chemistry, University of Cambridge in the centre of Cambridge city.

Visit the iDMT's website:

11
CURRENT SME COLLABORATIONS
currently approved by the *i*DMT Management Board.

HIGH IMPACT
areas of specialism are: AI for chemical synthesis, robotic equipment for chemistry, new algorithms for chemical process development.

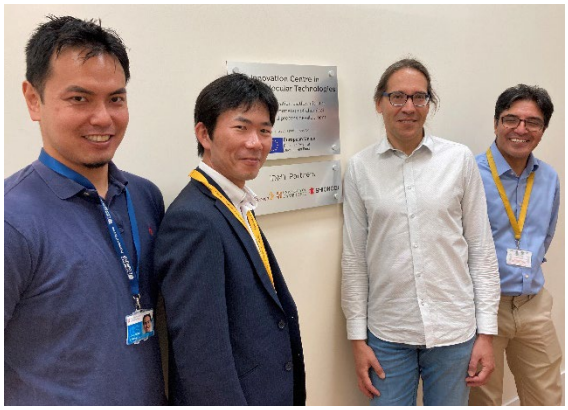
7
ACADEMIC LEADERS
contributing their knowledge and time to SME-driven projects for transforming the digital future of chemistry.

Browse the iDMT's YouTube videos:

Gabriella Bocchetti, © University of Cambridge

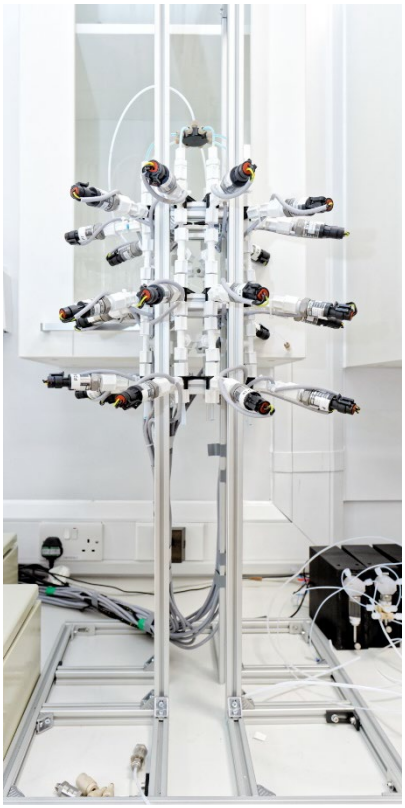
Gabriella Bocchetti, © University of Cambridge

Gabriella Bocchetti, © University of Cambridge



© Shionogi

Nathan Pitt, © University of Cambridge



Gabriella Bocchetti, © University of Cambridge

The *i*DMT management structure involves three boards, each with their own roles and responsibilities. Day-to-day management of the *i*DMT occurs through the Executive Board, which is overseen by the Management Board and provided with guidance from an External Advisory Board. The *i*DMT is staffed by both academics, researchers, and PhD students from the University of Cambridge as well as researchers from the industrial delivery partners.

Executive Board

The *i*DMT's Executive Board is responsible for operational activities of the *i*DMT, including the review of ongoing issues related to recruitment, expenditure, delivery of targets, communication activities, interaction with SME partners, liaising with funders, and review of procedures. The Executive Board is comprised of the Directors, the Project Manager, and the Administrator.

Professor Alexei Lapkin is the *i*DMT Director and leads the Lapkin Lab (Sustainable Reaction Engineering) which is developing clean, intensive processes for the manufacture of molecules, formulations, and functional materials. Professor Lapkin leads the group's work on methods of modelling chemical processes, starting from molecular modelling methods and extending to multi-objective process optimisation and life cycle assessment. Lapkin Lab have pioneered methods of machine learning for automated process optimisation and are exploring methods of Big Data for chemical reaction networks. The group works with many industry sectors, from inorganic materials and formulations to platform chemicals and pharmaceuticals, and have an international network of collaborations. Professor Lapkin is also the founder of several small companies, co-director of the SynTech Centre for Doctoral Training and sits on various boards including an advisory board position for the Global Entrepreneurship Centre (GEC, Dusseldorf).

Professor Matthew Gaunt is the *i*DMT co-Director and leads the Gaunt Group which is focussed on the development of new catalytic methods for small and biomacromolecule synthesis and functionalisation. The Gaunt Group also develop and apply high throughput experimentation strategies to accelerate synthesis and generate chemical data sets for machine learning applications towards predictive chemistry. The group has extensive connections with global and local pharmaceutical industry and biotech SMEs. Professor Gaunt is also the Yusuf Hamied 1702 Chair of Chemistry, Chair of the Synthetic Chemistry Research Interest Group and Director of the SynTech Centre for Doctoral Training.

Dr Celeste van den Bosch is the Project Manager of the *i*DMT. Dr van den Bosch has an MSci in Chemistry and a PhD in Materials Science & Engineering. She manages the centre, aligning the centre's work with milestones, strategic goals, financial targets and outputs through detailed planning. Dr van den Bosch is the first point of contact for those who want to know more about the centre, including SMEs who are interested in receiving support from the *i*DMT. She also maintains the relationships between the core *i*DMT team and the wider network, including funders and delivery partners. Dr van den Bosch is

passionate about the centre's purpose of supporting innovation in chemistry by bringing together academia, industry, and small businesses.

Kerstin Enright is the Administrator of the *i*DMT.

> "Digitalisation will result in a radical increase in the throughput of chemical discovery, guiding synthetic chemists towards successful solutions more efficiently, and freeing up their time to develop new ideas.
>
> *Professor Matthew Gaunt, iDMT co-Director*

Management Board

The Management Board is the decision-making body of *i*DMT, responsible for its strategic direction and operational oversight. The Management Board is comprised of the *i*DMT Directors, the *i*DMT Project Manager and representatives from each Delivery Partner: AstraZeneca and Shionogi.

Dr Clive Green is the Executive Director and Head of Compound Synthesis and Management, Discovery Sciences for AstraZeneca R&D. Dr Green leads the strategic development of automated compound synthesis at AstraZeneca. He is also an expert in medicinal chemistry and currently leads the compound management capability to support all stages of drug discovery with process and distribution of small molecules.

Dr Michael Kossenjans is the Director and Head of iLAB, Discovery Sciences for AstraZeneca R&D. Dr Kossenjans leads the design, development, and deployment of the automated compound synthesis capability at AstraZeneca. He is also an expert in medicinal chemistry and has established a prototype automated chemistry platform supporting drug discovery projects with small molecule libraries.

Dr Kenji Yamawaki is a representative from Shionogi Co., & Ltd involved in *i*DMT activities. Dr Yamawaki is also an Associate Corporate Officer and a Vice President of the Laboratory for Medicinal Chemistry Research for Shionogi. The development of new synthetic technologies for lab automation, including synthetic design by machine learning to improve throughput activities for drug discovery, forms a part of the Shionogi medicinal chemistry team's future strategy.

Dr Makoto Kawai is in charge of the *i*DMT project in Shionogi. Dr Kawai is also the director of Shionogi's Drug Discovery 3 programme, leading the development of new technologies for medicinal chemistry, including lab automation and cheminformatics, to accelerate drug discovery. Dr Kawai is keenly interested in the development of new technologies, and the growth opportunities for Shionogi members through working with the *i*DMT.

University of Cambridge Academic Team

A key element of the *i*DMT's success is the involvement of world-leading academics who are supported in their research to develop innovative ideas and methods for potential training and support of SMEs. The core research of the academic team is the main magnet for SMEs to work with the *i*DMT and learn from the ongoing research.

Dr Sebastian Ahnert is a University Assistant Professor in the Department of Chemical Engineering and Biotechnology, and a Senior Research Fellow at the Alan Turing Institute in London. His research interests lie in the intersection of theoretical physics, biology, mathematics, and computer science. He is particularly interested in using algorithmic descriptions of structures and functional systems to quantify and classify complexity.

Professor Jonathan Goodman is a Professor of Chemistry, leading the Goodman Group which uses computational organic chemistry to understand molecular phenomena and to gain knowledge from molecular data. Their work includes method development for analysing reactions and property prediction from chemical structures. The group also works with IUPAC and other international bodies to help communicate chemistry effectively.

Dr Laura Torrente Murciano is a Reader in the Department of Chemical Engineering and Biotechnology where she leads the Catalysis and Process Integration group. She has expertise on the development of manufacturing technologies for nanomaterials combining aspects of reaction engineering, fluid dynamics and in-situ characterisation. Her group is pioneering automation strategies for self-regulated synthesis of materials.

Dr Alpha Lee is a Royal Society University Research Fellow and leads the Lee Group. The Lee Group develops predictive models for molecular design, chemical reaction prediction, and aims to integrate physical simulations into machine learning models. Dr Lee's specific expertise is developing machine learning methods to accelerate the design-make-test cycle in drug design and materials discovery, combining science with statistical models to learn from big datasets.

Dr Miguel Lobato is an Associate Professor in Machine Learning and is one of the leaders of the Cambridge machine learning group. His research expertise is in probabilistic machine learning and its applications to molecular design. His group works on deep generative models of molecules and their combination with Bayesian optimization techniques for data-efficient automated molecular design.

External Advisory Board

The role of the External Advisory Board is to calibrate the activities of the *i*DMT against international competition, national objectives, and local priorities. The External Advisory Board is comprised of individuals with knowledge complementary to the Management Board, providing a mixture of experiences in the areas of UK-based SMEs, innovation support, and small-molecule discovery pharmaceuticals.

Professor Andy Neely OBE is the Pro-Vice-Chancellor for Enterprise and Business Relations at the University of Cambridge and former Head of the Institute for Manufacturing (IfM). He was the Founding Director of the Centre for Digital Built Britain and the Cambridge Service Alliance. Professor Neely is widely recognised for his work on the servitisation of manufacturing, as well as his work on performance measurement and management. He received an OBE for services to Research and to University/Industry Collaboration in 2020.

Neelam Patel is the CEO of MedCity. MedCity was established in 2014 by the Mayor of London in partnership with London's Academic Health Science Centres as the cluster organisation for world-leading health and life sciences in the region. Patel has extensive leadership experience in the biopharmaceutical industry and public sector including National Institute for Health Research (NIHR) service improvement and strategy. Developing optimal innovation pathways for SMEs is a focus of her role at MedCity, in addition to driving MedCity's strategy which includes promoting SME growth by enabling research collaborations, access to investment and infrastructure. Patel also sits on the board of the Health Research Authority, the NICE steering group for the Evidence Standards Framework for digital health technologies and is on the advisory board of the NIHR research design service.

> "The development and application of digital technologies including artificial intelligence and robotics has the potential to significantly accelerate progress in different fields of chemistry.
>
> *Alleyn Plowright, Wren Therapeutics Ltd*

Dr Alleyn Plowright is the Head of Translational Science and Pre-Clinical Development at Wren Therapeutics Ltd. Wren Therapeutics is a spin-off company from the University of Cambridge (UK) and Lund University (Sweden) dedicated to eliminating protein misfolding diseases by discovering and developing new therapeutics. Dr Plowright previously held appointments as the Head of Integrated Drug Discovery for Sanofi (Germany) and was a Project Leader in Medicinal Chemistry at AstraZeneca. He has extensive experience in driving innovative approaches to drug discovery.

Highlights

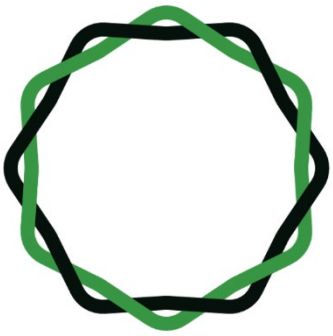
from iDMT-supported SMEs

vapourtec
precision flow chemistry

Vapourtec, page 13

CDI Chemical Data Intelligence

Chemical Data Intelligence, page 14



Green Rose Chemistry, page 15

Memgraph, page 16

DeepMatter, page 11

STOLI chem

Stoli Chem, page 15

ATMPS

ATMPS, page 16

Accelerated Materials, page 10

rutterdesign, page 12

Thomas Swan

Thomas Swan, page 16

Case Study:

Accelerated Materials Ltd

Accelerated Materials Ltd. is a University of Cambridge start-up focussing on the manufacture of nanomaterials by wet synthesis and offers consulting in the area of materials development and scale-up of nanomaterial manufacturing. This engineering consultancy specialises in providing businesses and investors with the knowledge and technology to accelerate materials innovation and commercialization.

> With the *i*DMT's capabilities, Accelerated Materials has been able to provide services for multiple clients in pilot trials. We have also been successful in using the space as a platform for marketing and increasing brand awareness.
>
> *Dr Nicholas Jose, Accelerated Materials founder*

Accelerated Materials provides companies with the abilities to rapidly generate new materials, processes and products using a unique technology platform for scale-up. This platform consists of several inventive features:

- advanced high shear microreaction technology for precision tuning of particle formation at the nanoscale,
- a workflow implementing machine learning methods for efficient multi-objective optimization of processes, and
- a cost-effective, passive manifolding strategy that mitigates the issue of process variability during number-up.

The Accelerated Materials platform is widely applicable with the capacity to be used for a wide range of materials and chemicals used in the pharmaceutical industry and beyond, including anisotropic nanocrystals for APIs, layered double hydroxides, and quantum dots for bioimaging.

The partnership of Accelerated Materials with the *i*DMT aims to create a new commercial platform for product development and optimisation, integrating the core process technology of Accelerated Materials with the algorithms and automation expertise of the *i*DMT. The goal of the project was to develop this system for a range of products, and in particular this collaborative project is considering:

(1) metal organic framework nanoparticles, used as precursors for high capacity, high selectivity adsorbents, and
(2) anisotropic zinc oxide nanomaterials which have applications as antimicrobial additives ranging from skin care to exterior building coatings.

Recently Accelerated Materials has completed a demo of its product to Shionogi, with the support of the *i*DMT, and currently a whitepaper is being prepared for publication. Accelerated Materials has also received interest in their nanoparticle product and its software product and is working on two developments to complete these projects within 2022.

Case Study:

DeepMatter Group Plc

At DeepMatter we are building a single platform that allows reaction data to be consumed from multiple data sources (internal, external, public, or private) to provide a cleansed, harmonized, and categorized repository for the exploitation of reaction data through API's, search tools and ML/AI learning. We combine existing proven software with new components to provide a scalable, extensible and performant platform for the chemistry community, underpinned by proven cheminformatics solutions and tools.

Our Digital Glassware (DG) technology provides chemists with the ability to capture entire experiment time course data to provide outcome data, real time guidance and notifications. Using proprietary algorithms, we can find previously unknown correlations in the data to provide a better understanding of chemical reactions and deliver insight to predict outcomes and develop better, safer and more sustainable routes for synthesis.

InfoChem, part of the DeepMatter Group, has more than 30 years' experience in the development and integration of sophisticated software tools for the storage and handling of structure and reaction information. Our SPRESI hand curated database of organic reactions underpins the chemical knowledge in our reaction prediction tools ICSynth and ICfrp and is complemented with additional data sources. We provide comprehensive tools for the cleaning, labelling, and categorizing of chemical reaction data sets. These can be used to facilitate the incorporation of further data sources into our reaction prediction tools ensuring that you can maximise the value from your existing IP.

> "The expertise we can garner by engaging with these influential institutions will be invaluable in helping us to innovate our platform even further.
>
> *Mark Warne,*
> *CEO of DeepMatter Group*

We are pleased to be working with Prof. Lapkin and the team at the *i*DMT to demonstrate the capabilities of our Digital Glassware technology in synthesis and automation. This collaboration enables us to access facilities and personnel that will help guide us on our journey to provide chemists with the best tools that they require to work more efficiently in the future. A key element of this is access to the integrated workflows within *i*DMT and access to a range of hardware. We are interfacing with the Vapourtec flow chemistry platform to enable the capture of key reaction parameters into the DG platform and facilitating cloud-based storage and access. This, combined with the chemistry details and experimental outcomes will enable scientists to efficiently optimise their reactions and maximise the use of the reaction condition data provided by this flow synthesis platform. DeepMatter is grateful for the support it is receiving from Vapourtec to enable the utilisation of their recently updated API to facilitate the integration.

Case Study:

rutterdesign Ltd

rutterdesign is a one-man consultancy business which helps shape and deliver programmes that identify the wider enterprise changes needed for success in manufacture. The original rutterdesign-*i*DMT project remit was to explore the evolution of new business models using AI to model and map decision networks from macro-level to the individual, with an aim to support decision making processes and help clarify complex manufacturing investment decisions.

After in-depth discussions on the concept of the proposed approach with experts in supply chain optimisation at the University of Cambridge, this project has evolved into a potential large collaboration with multi-nationals as end-users and SMEs as providers, and will include rutterdesign, as well as multiple other *i*DMT SME partners and multinational corporations.

Current chemical supply chains have evolved under the pressures of costs (labour and regulatory avoidance) and business opportunities presented by global trade agreements. Today the drivers have changed radically under new pressures of resilience in the face of climate change impact, and the societal demand for sustainable manufacturing. Specifically, recent catastrophic weather events and the COVID pandemic clearly exposed vulnerability of the existing supply chains.

The identified problem is of a system-level scale: today manufacture of molecules in the UK is largely impossible due to the lack of traditional basic raw materials, loss of traditional manufacturing capabilities, and price competition on the global markets with the low labour cost / low regulatory regimes in Southeast Asia and China.

The project will deliver a unique solution that addresses the challenge as an overall system. The key output of the collaboration will be the clear business strategy for implementing UK-based manufacturing of strategically important molecules. The data-supported decision making will be grounded in network optimisation tools, where the network elements comprise continuously updated information on:

(i) manufacturing capabilities,
(ii) large data sets of properties of molecules and chemical transformations, and
(iii) price information.

The project will also deliver a new service offer for a consulting business, a new tech product for chemical manufacturers and demonstration of UK-based manufacturing of a functional strategically important product.

Case Study:

Vapourtec Ltd

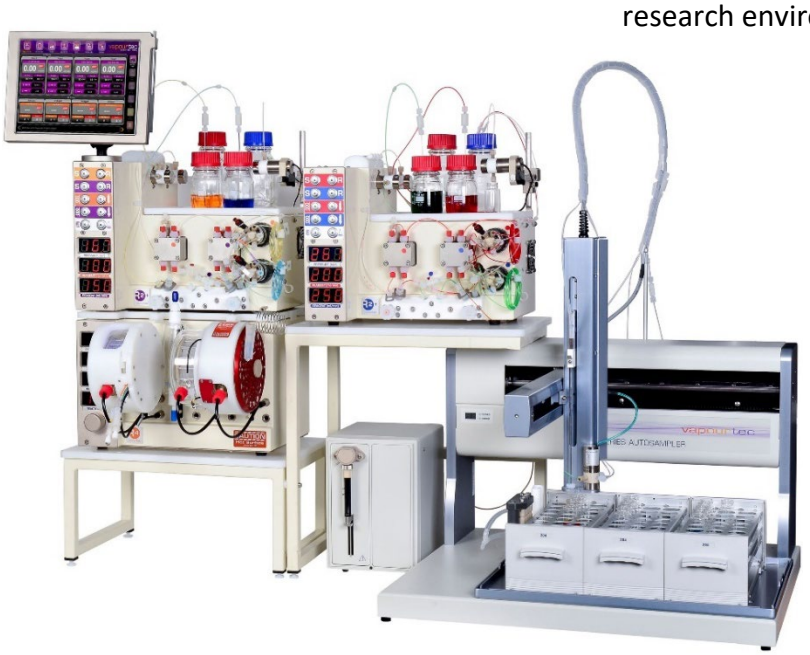
vapourtec
precision flow chemistry

Vapourtec Ltd is a technology company located close to Cambridge. It designs and manufactures flow chemistry laboratory equipment. Their systems include features that have been designed for precision and repeatability, faster reaction optimisation, and instant reproducible scaleup. A key focus of Vapourtec is designing systems for easy integration with upstream and downstream equipment.

The collaboration between Vapourtec and the *i*DMT supports this goal, with a project focussed on the development of a next generation API. This project is building a "requirements specification" for an API interface that will allow both high level and low-level system control, whilst also increasing the transparency of the system to facilitate a more connected laboratory, with seamless integration of instruments. Specifically, the project will also research and define a generic method for integration of a flow chemistry platform to HPLC system including dilution control.

> "Through working with the *i*DMT we plan to develop a user-friendly platform, enabling our customers to easily configure our Vapourtec systems in a complex set up, such as those required for AI driven self-optimisation platforms.
>
> *Duncan Guthrie, Director of Vapourtec*

Currently the *i*DMT is using the Vapourtec API which was developed with the input from the *i*DMT team on the required functionality. Within the *i*DMT, work is also ongoing to install a digital twin of the Vapourtec instrument which will demonstrate the potential use cases of Vapourtec instruments within a fully digitised research environment.



Case Study:

Chemical Data Intelligence Ltd

Chemical Data Intelligence (CDI) was established as a spinout from Cambridge Centre for Advanced Research and Education in Singapore (CARES Ltd) and recently incorporated in England as CDI Ltd. CDI is a consultancy and software development firm based on the vision of creating a more sustainable and smarter chemical industry.

> *i*DMT provides an invaluable platform for SMEs like us, which enables synergistic interactions with world-class industrial players and academic professionals. Through this collaboration, we will understand more about industrial requirements and know how to solve real-world issues using cutting-edge technologies, which accelerates the development of the product version of our software tools.
>
> *Dr Zhen Guo,*
> *Director of CDI*

CDI has developed an initial series of software tools for optimisation of synthetic routes. Through interacting with the *i*DMT, CDI will benefit from access to big pharma researchers to ensure end-customer requirements are met, and to the *i*DMT academic knowledge base to contribute to the optimisation algorithms, as well as to other SME partners to provide holistic solutions for industrial challenges. This collaboration will support CDI in implementing new features and improving robustness of the synthesis development tools.

The CDI-*i*DMT project has accelerated the testing of an initial demo-tool that CDI has developed for optimising synthetic chemistry routes. CDI demonstrated their initial product to their lead customer in June 2022 and jointly, with the *i*DMT and the customer, have developed a roadmap to further test and optimise the product for their end users' requirements.

Snapshot SME Introductions:

Green Rose Chemistry Ltd is a consultancy that provides the green chemistry expertise companies need to develop sustainable products and processes for a renewable, circular bioeconomy. One of their specialties is solvent substitution, specifically introducing sustainable solvents into industrial applications including cleaning, extraction, and formulation. They are working with the *i*DMT to develop a tool that accurately predicts polymer dissolution with no lab work involved, speeding development of biobased and biodegradable polymers. Through working with the *i*DMT, Green Rose Chemistry aims to enhance their consultancy offering by creating a new service and entering the market of polymer development.

GREEN ROSE CHEMISTRY

Stoli Chem is a Warwick University spinout which designs and manufactures continuous flow processes with an aim to reduce manufacturing costs of fine and speciality chemicals. Their purpose in connecting with the *i*DMT is to learn about integrating their reactor designs into fully automated setups and expanding the applicability of their process technology. The Stoli Chem- *i*DMT project is studying a slurry hydrolysis as a flow chemistry testcase to understand the limits of the current reactor design and how these can be improved.

STOLI chem

If you are (a representative of) an SME based in England who is interested in finding out more about how the *i*DMT can support your project, we invite you to get in touch with us. Contact the *i*DMT by sending an email (idmt-info@ch.cam.ac.uk). We want to hear from you at any stage, from idea inception to market delivery of a new product or service!

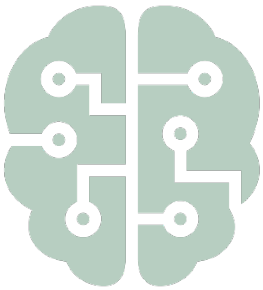
Memgraph Ltd is a graph streaming platform that provides instantaneous insights for better decision-making. It was founded in 2016 to address the limitations of current graph database systems by developing a state-of-the-art high performance and scalable graph database. Memgraph developed a technology that enables software developers to build graph-based tools without the need for delving into graph mathematics. The Memgraph-*i*DMT project supports a collaborative effort between Memgraph and CDI which will bring together Memgraph's software and CDI's graph tools for the specific domain of early-stage chemical development, with expert knowledge support from *i*DMT staff.

ATMPS Ltd is a company focused on vein-to-vein tracking solution for the personalised medicine sector and has developed a patented software platform called Hataali for ensuring regulatory compliance for treatments known as advanced therapy medicinal products. The project with *i*DMT aims to access a new market for their blockchain technology within chemical digital labs and digital manufacturing, allowing easy traceability of all assets and leading to opportunities for optimisation. The pilot project currently underway is developing a blockchain based chemical inventory optimisation and maintenance product for the *i*DMT lab and one of the academic departments of the University of Cambridge as a demonstration for future digital labs and factories.

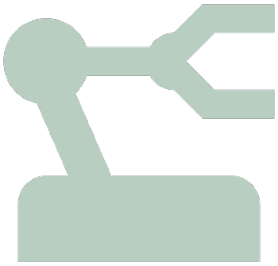
Thomas Swan & Co. Ltd is a well-established, independent manufacturer of performance and fine chemicals. Through a joint project with the *i*DMT, Thomas Swan expects to develop further deep learning skills and internally demonstrate the benefits of digitalisation. Recently, Thomas Swan and *i*DMT have been working on identification of a specific non-proprietary case study, and setting up the preparatory work, including training of Thomas Swan employees in machine learning and coding, and creating the scientific backing for the project within *i*DMT with the work on ML-based structure-property relationships predictions and on predictive toxicology.

Scientific Synopses

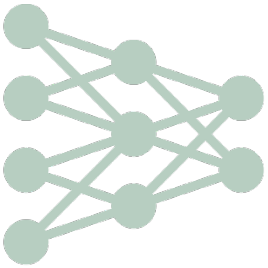
from iDMT-associated researchers and groups



Artificial intelligence in molecular technologies



Robotic equipment and Automation



Algorithms for digital process development

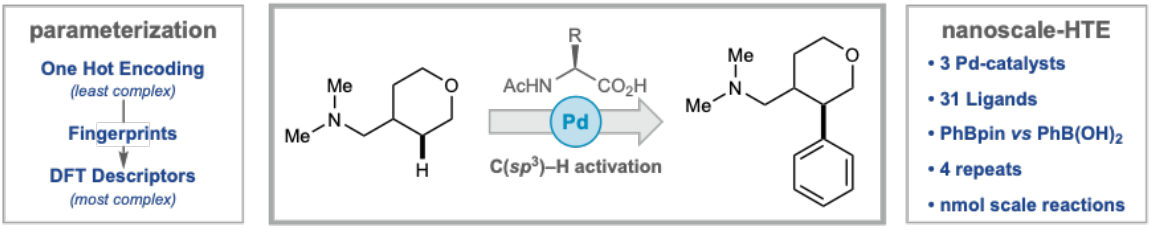
The Effect of Chemical Representation on Active Machine Learning Towards Closed-Loop Optimization

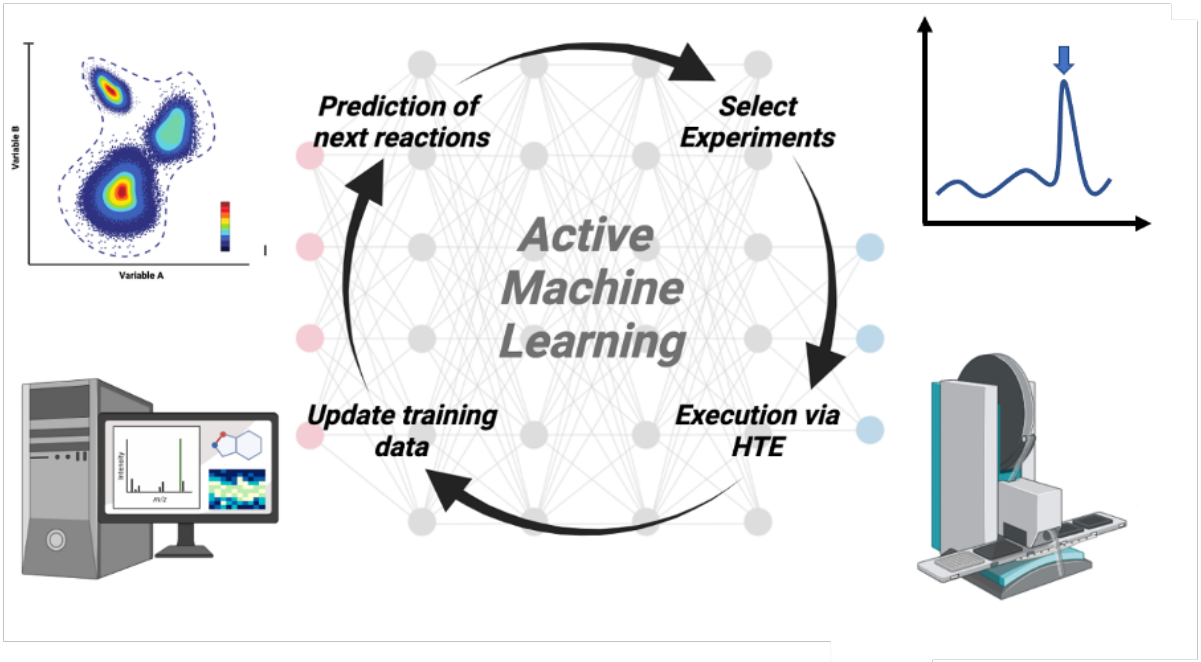
Alexander Pomberger

Multivariate chemical reaction optimization involving catalytic systems is a non-trivial task due to the high number of tuneable parameters and discrete choices. Active Machine Learning (ML) represents a powerful strategy for automating reaction optimization. However, the translation of chemical reaction conditions into a machine-readable format requires the identification of highly informative features which accurately capture the factors which determine reaction success. Herein, we compare the efficacy of different calculated chemical descriptors for a high throughput experimentation generated dataset to determine the impact on a supervised ML model when predicting reaction yield. Then, the effect of featurization and size of the initial dataset within a closed-loop reaction optimization was examined. Finally, the balance between descriptor complexity and dataset size was considered. Ultimately, tailored descriptors did not outperform simple generic representations, however, a larger initial dataset accelerated reaction optimization.[1]

References

1. Pomberger, A.; Pedrina McCarthy, A. A.; Khan, A.; Sung, S.; Taylor, C. J.; Gaunt, M. J.; Colwell, L.; Walz, D.; Lapkin, A. A., Reaction Chemistry & Engineering 2022, 7, 1368-1379





Exploring molecular context: reaction predictions studies

Henrique Magri Marçon

During the last year we have been studying how to teach computers what influences regioselectivity in chemical transformations using the concept of molecular context. In chemistry it is typical to have prediction models with excellent accuracy, but they come at a price that they work on a limited part of the chemical space. On the other hand, models that explore a wider chemical space usually lack the selectivity or precision presented by local models. Our goal within this project is to create a workflow that will help scientists to identify which variables drive the reaction towards one product or the other based on records already available in the literature and databases. Therefore, with the knowledge of the factors that drive reactions we will be able to make more informed decisions on synthetic planning and acquire the high accuracy from local models with the comprehension to use it in the vast chemical space.

Currently, we have selected reactions for case studies to develop and showcase future interpretable machine learning models, from which we expect to extract first principles knowledge from chemical databases and literature. Two main datasets have been curated from the literature: aromatic substitution and *N*-alkylation of indazoles. The former was selected to guide the development of explainable models that comprehend the chemical structure and properties of starting materials. The latter is based on reactions that are highly influenced by experimental conditions. Because the work will be based on the curated dataset, it is essential to have a proper data cleaning, for example duplicate removal and management of invalid entries. The next steps in these projects include data engineering to encode the molecules and their environment in a computer-readable way. Finally, we will be able to develop a machine learning model that can show us where a reaction is going to happen and the reasons why. Therefore, it is possible to understand the reaction paths and actively design chemical synthesis.

Additionally, we are investigating how the modelling of side reactions can help improve reaction predictions and the influence of pH on transformations involving boronic acids. A model on the stability of boronic acids against the pH they are in has been developed and will soon be published. In the next period we expect to collect experimental data, in semi-automated fashion, to validate boronic acid decomposition model and use it to help designing a cleaner path for the main transformation.

Reaction impurity prediction using a data mining approach

Adarsh Arun, Zhen Guo, Simon Sung, Alexei Lapkin

1 Introduction

Early-stage knowledge of impurities and conditions under which they may form is crucial for the rapid design of robust, scalable, and sustainable synthesis pathways for target molecules. This is especially relevant for the pharmaceutical industry where im- purity tolerance is low to maximize the safety and efficacy of the final drug product. Conventional methods for impurity profiling rely on chemical intuition and analytical techniques which can be time consuming, iterative, and expensive. Prediction of impurities a priori is an attractive option to alleviate some of these drawbacks. Notably, given the development of large chemical reaction databases such as Reaxys® (>148 million substances and >54 million reactions), leveraging this available chemical data to make informed predictions is more viable now than ever before.

Existing data-driven prediction methodologies[1,2] are numerous, with some leveraging large chemical reaction networks containing molecules as nodes connected via edges to reactions. However, they all have at least one of the following drawbacks: I) They exclusively focus on main product prediction and ignore impurities and II) They are often black-box and difficult to interpret if wrong predictions are suggested. Therefore, the goal of this work is to propose an automated impurity prediction workflow that is interpretable and transparent. In the context of chemical reaction networks, this task can be reframed as a type of link prediction where impurity information is included via additional edges.

2 Methods

Figure 1 illustrates the proposed automated workflow for impurity prediction, which is based on data mining a subset of 17 million reactions in Reaxys®. The workflow, which was implemented in Python using RDKit, consists of fourteen steps split across four main modules: I. Data mining, II. Data processing, III. Impurity prediction and IV. Impurity ranking.

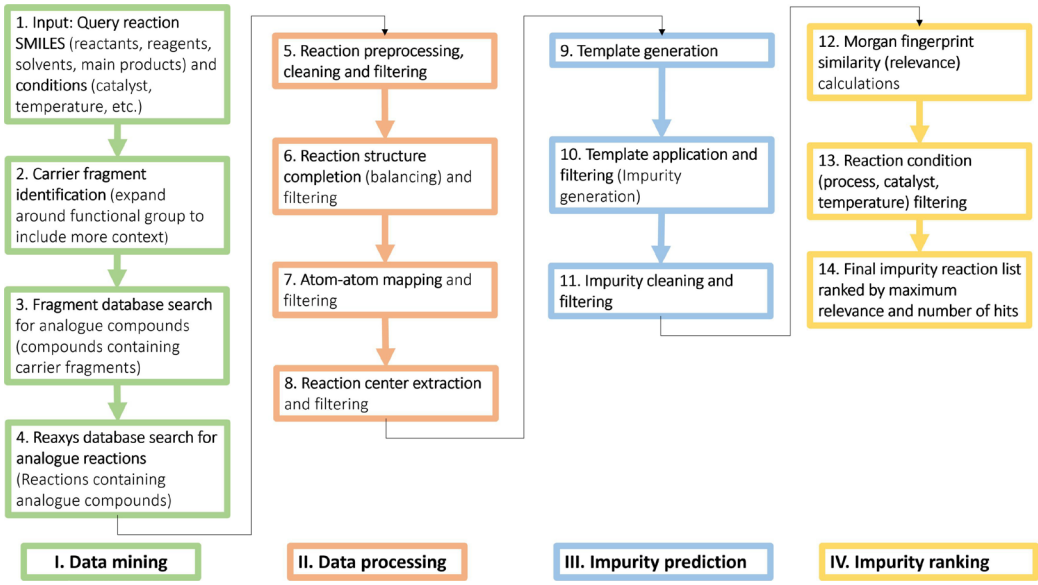


Figure 1 - An illustration of the developed workflow for automated impurity prediction based on data mining.

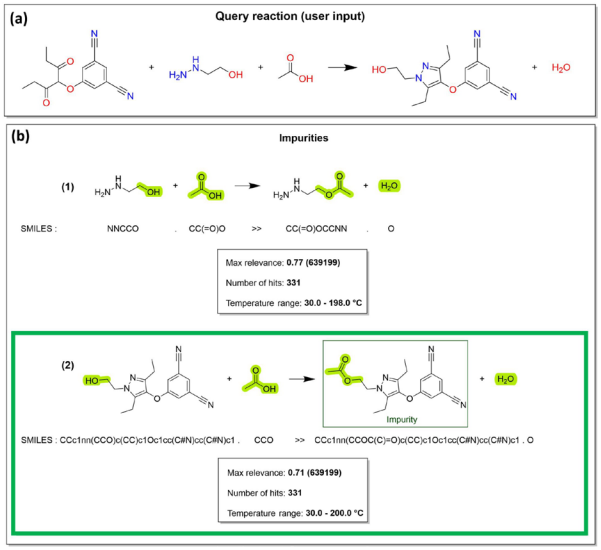


Figure 2 - Summary of results for lersivirine case study: (a) Query reaction; (b) Top two ranked impurity reactions based on maximum relevance. Each reaction is captioned with a box containing the max relevance score with associated Reaxys® ID in brackets, number of hits and temperature range. The literature-identified impurity reaction is ranked second and boxed in green as is the literature identified impurity. Relevant functional group fragments are highlighted in green.

The rationale behind this approach lies in the understanding that impurity prediction entails assessing how functional groups present in reactants, reagents, solvents, and main products (henceforth referred to as *query species*) may react with each other under specific conditions. This assessment can be performed by directly mining a database of *analogue reactions*, here defined as reactions involving functional group fragments derived from the query species. The analogue reactions are cleaned, stoichiometrically balanced and atom mapped. Finally, extracting reaction templates from the analogue reactions, and applying them to relevant query species can suggest potential impurities and transformations of interest. These are then ranked based on *maximum relevance* (Highest average morgan fingerprint similarity between analogue reactants and associated query species, across all analogue reactions suggesting the impurity) and *number of hits* (Total number of analogue reactions suggesting the impurity).

3 Results

To validate the workflow, three proof-of-concept case studies based on active pharmaceutical ingredients (paracetamol, agomelatine and lersivirine) were conducted. In all cases, the correct impurities were suggested within the top two outcomes. Figure 2 illustrates the results of the workflow when applied to the synthesis of lersivirine, a drug for HIV treatment.[3] As shown in Figure 2b, the correct impurity due to an esterification of lersivirine with ethanoic acid was ranked second (boxed in green) with a maximum relevance score of 0.71 (Reaxys® ID 639199) and 331 hits. As additional validation, this impurity was also detected in a lab environment in Singapore.

It is important to note that at all stages, suggested impurities can be traced back to the originating reaction template and Reaxys® analogue reaction in literature, allowing for closer inspection and user validation.

4 Conclusion

This work could be useful as a benchmark for more sophisticated algorithms or models since it is interpretable, as opposed to purely black-box solutions, and illustrates the potential of chemical data in impurity prediction. In the long run, applying this workflow to the entire Reaxys database will allow for an enriched chemical reaction network with impurity information that can facilitate rapid decision making and synthesis route planning.[4]

References

1. Coley CW, Jin W, Rogers L, Jamison TF, Jaakkola TS, Green WH, et al. A graph-convolutional neural network model for the prediction of chemical reactivity. *Chem Sci*. 2019; *10*(2): 370–377.
2. Jacob PM, Lapkin AA. Prediction of chemical reactions using statistical models of chemical knowledge. *ChemRxiv*.
3. Codina A, Derrick A, Joyce R, Susanne F. *Process for the preparation of lersivirine*. WO2013050873A1 2012.
4. Arun, A., Guo, Z., Sung, S., & Lapkin, A. (2022). Reaction impurity prediction using a data mining approach. *ChemRxiv*. doi:10.26434/chemrxiv-2022-0btmt. This content is a preprint and has not been peer-reviewed.

Carbonyl Alkylative Amination as a new method for alkylamine synthesis

Joseph Phelps

The CAA array project has focused on developing a high throughput microscale array platform with a second-generation version of the carbonyl alkylative amination reaction developed in our laboratory. There were several challenges associated with translating the reaction into high throughput: 1) stirring a heterogenous mixture, 2) containing each reaction well inside an inert environment in the presence of corrosive chemicals, 3) purifying the products produced and 4) analysing the reaction mixtures. To overcome the first two challenges, we created a customised reaction vessel which can be placed on commercial stirring platforms. To overcome the purification challenge we relied on reactive powders for the removal of unwanted chemical species. Finally, we used automated microscale nuclear resonance spectroscopy to provide a rapid but detailed analysis method. The platform was then used to optimise a difficult reaction, taking a several week process down to less than a week and can now be used to make a diverse range of pharmaceutically relevant alkyl amines.

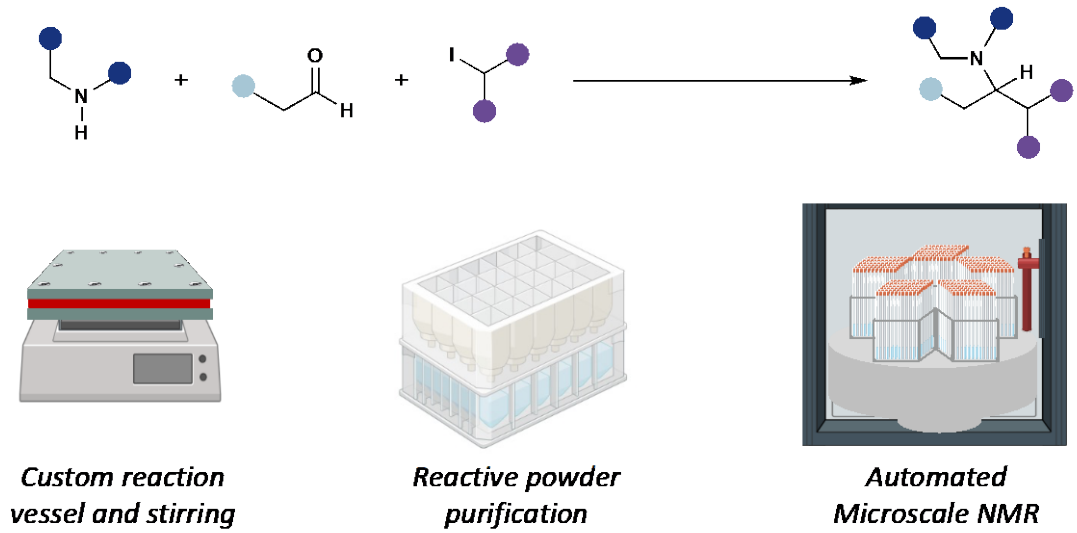


Figure 1 – The carbonyl alkylative amination

Development of Automated Chemical Transformations in Air Using High Throughput Micellar Catalysis

Hanna Maliszewska

Currently, we are witnessing a transformation of chemical synthesis from one at a time reaction to high throughput, automated preparation of tens or hundreds of compounds in a single experiment. This change requires simultaneous development of chemical transformations that could be used reliably across a wide area of chemical space. Such reactions need to provide an accessible methodology to form the most important types of chemical bonds and compound classes. For these high throughput methods to be generally adopted, they should be experimentally straightforward and take into account sustainability aspects. Micellar catalysis has been brought forward as an attractive method for high throughput synthesis applications thanks to its known effects of reaction facilitation and acceleration, mild conditions and green character (exclusion of use of toxic organic solvents). In short, under these conditions the reaction between organic molecules takes place within micelles formed by surfactant molecules in aqueous media.

This project investigates micellar catalysis methods in the context of miniaturised, parallel synthesis. The micellar reaction conditions were first adapted to small scale high throughput format (where routinely 24, 96, 384 or 1536 reactions are performed within one experiment in low microlitre or nanolitre volumes). This means that a standard batch reaction scale has to be decreased by tens to hundreds of times. Working with such large volume of reactions necessitates the use of laboratory automation such as liquid handling robots. Miniaturised, semi-automated micellar workflow under standard laboratory environment has been established. As a proof of concept, the project scope encompassed chemical reactions that are commonly used in pharmaceutical industry, such as Suzuki-Miyaura cross-coupling. This high throughput method was used both for multifactor optimisation of reaction conditions (see Figure 1) and preparation of small libraries of targets using medicinally relevant building blocks in a 96 well microtiter plate format. The work on extended range of chemical transformations that can be used with this method is ongoing.

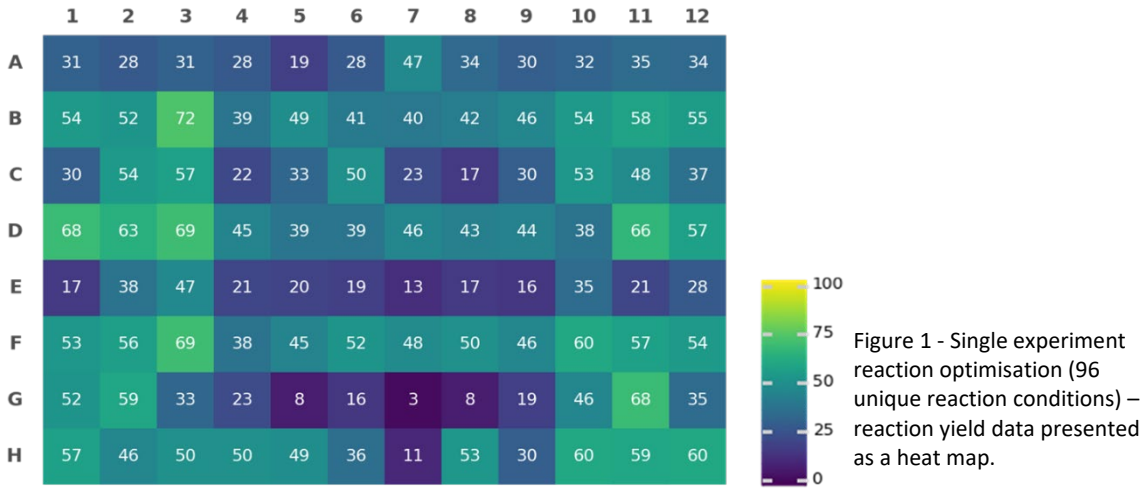


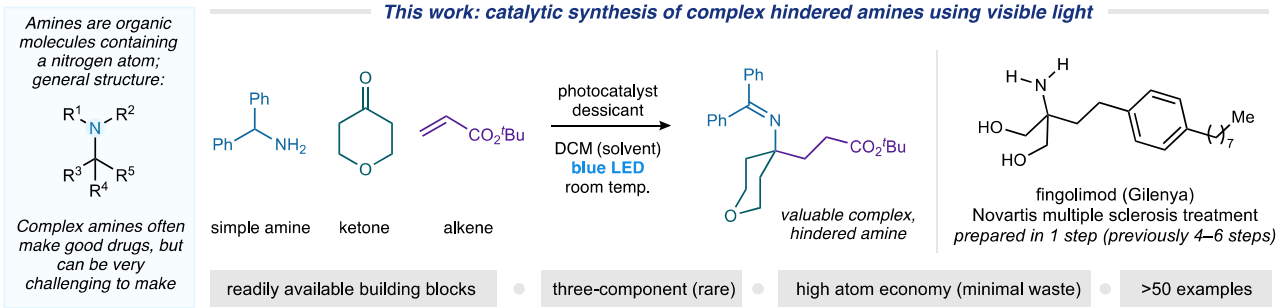
Figure 1 - Single experiment reaction optimisation (96 unique reaction conditions) – reaction yield data presented as a heat map.

Photocatalytic synthesis of alkylamines

Gaunt Group

The development of efficient new methods for the facile preparation of complex biologically active molecules is an important goal in contemporary organic chemistry. Amines are a particularly interesting type of molecule due to their profound therapeutic potential – indeed, over 40% of pharmaceutical agents belong to the amine class. Hindered amines – amines with three bonds next to the nitrogen atom – represent a subclass of amine that remains particularly challenging to synthesise, despite having proven application in a therapeutic context; examples of hindered-amine-containing therapeutic agents include fingolimod (multiple sclerosis), Elayta (Alzheimer's disease), CCT128930 (AKT inhibitor), and virantmycin (antiviral).

In this work, we developed a catalytic method for the synthesis of hindered amines, using an iridium-based catalyst activated by visible light. The key to the method lies in the ability of the catalyst to absorb energy from blue light and transfer it, with high selectivity, to a single 'receiver' molecule in the reaction mixture. This input of energy sets off a sequence of bond formations which knit together the amine, ketone and alkene starting materials, at which point the catalyst returns to remove the remaining energy from the molecule. This delivers the desired product and regenerates the active form of the catalyst, which is then ready to absorb more light and react with more starting material. Our design stands out in the fact that it requires very few chemical reagents to affect the process — only a catalyst and a desiccant are required, in addition to the starting materials and the solvent — and in that it combines three starting materials rather than two. In addition to preparing over 50 novel and structurally interesting hindered amines, we were able to prepare Novartis' multibillion dollar multiple sclerosis treatment, fingolimod, in a single step, in contrast to other approaches which typically require 4–6 steps.

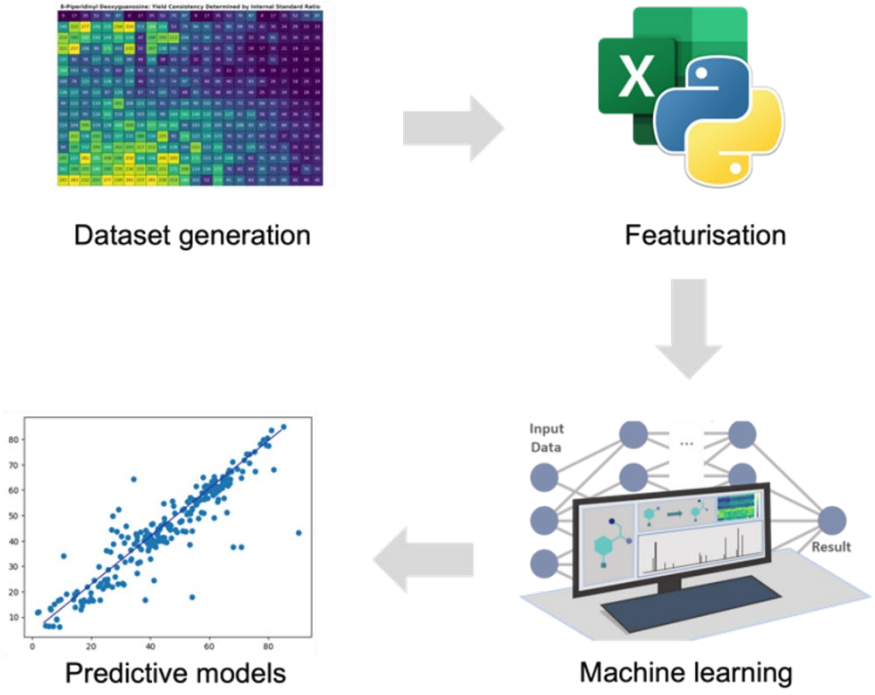


Using HTE to generate datasets for machine learning models

Abigail Barker-Mountford

Yield prediction is a challenging problem for expert synthetic chemists as reaction outcomes are influenced by many parameters that cannot always be rationalized by using chemical intuition alone, however, could be unravelled by machine learning algorithms if given enough data. By using high-throughput experimentation (HTE), rapid large-scale data generation is possible; thousands of reactions can be carried out on a nanomolar scale utilising automated liquid handling platforms in a fraction of the time it would take to carry out these reactions by traditional methods. The merger of machine learning models (ML) with HTE presents a wealth of opportunities to predict yield ahead of time, uncover hidden facets of reactivity and learn more about reaction mechanism.

In this work, HTE was used to generate a library of yield data for the simple, yet robust SNAr reaction. This reaction lends itself well to the automated platform due to the air stable nature of the reagents and well understood reaction mechanism. The reactions were carried out in well-plates, sampled, and analysed to give a yield measurement for each experiment. This data was then combined with both structural and physico-chemical descriptors to train a series of ML models that were able to accurately predict reaction yield ahead of time. This work highlighted the utility of automated liquid handling robots for both the accuracy and speed of data generation, and the paradigm shifting potential of HTE for future reaction optimisation and discovery.

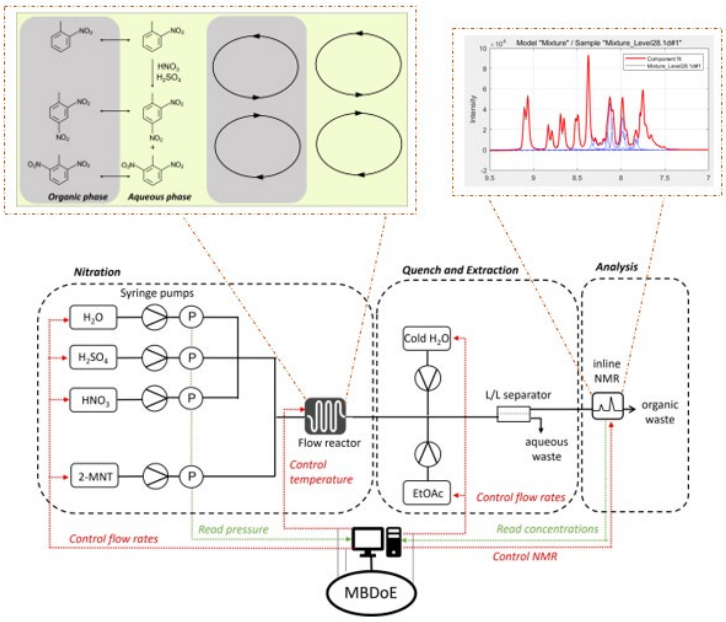


Development of multi-phase processes using rapid data generation and hybrid mechanistic-ML models

Anna Katsarou

Batch reactors, although widely used in chemical reaction engineering, present several limitations in terms of process intensification and scale-up.[1] Conventional stirred-tank reactors are adversely affected by inefficient mass and heat transfer as well as safety concerns in cases of highly exothermic reactions. Benefiting from their smaller reactor volumes, continuous flow microreactors can alleviate these challenges and can be particularly useful in multi-phase reaction systems: where both chemical kinetics and mass transfer are controlling the outcome of the process. In this project we aim to enable reliable and routine scale-up of continuous flow liquid/liquid multi-phase processes utilising efficient data generation techniques.

Aromatics nitration is a reaction of great interest since nitroaromatics are encountered in pharmaceuticals, dyes, explosive, pesticides, etc., as intermediates or final products.[2] The nitration of 2-nitrotoluene was chosen as a model system for measuring reaction kinetics and characterizing liquid/liquid mass transfer. This fast biphasic reaction involves an aqueous phase, composed of the nitrating agent (nitric and sulfuric acids), and an organic phase, composed of the organic substrate and the nitrated products.



We opted to create an automated continuous flow platform to enable the closed-loop development of reliable and robust kinetic models. Syringe pumps are used to deliver accurate and constant flows free from pulsation, active and passive micromixers are utilized for microfluidic mixing, water/ethyl acetate (EtOAc) streams are combined with the reaction stream to quench the reaction and simultaneously extract the organic products into the organic phase, and membrane separation is finally employed for inline phase separation. The organic phase passes through a benchtop NMR for real-time acquisition of data and a chemometric model, based on an indirect hard model approach,[3] is integrated to deconvolute the overlapping peaks of nitrated compounds and quantify them. To further increase the robustness of the platform and generate high data density, we perform non-steady state experiments by applying temperature and flow ramps. This automated platform is coupled with a Model-Based Design of Experiments (MBDoE) algorithm to design experiments with the objective of identifying optimal model parameters for a pre-selected kinetic expression from open academic literature. This allows us to obtain mechanistic insight into the reaction chemistry in a more efficient manner than conventional experimental strategies that rely on steady-state experimentation, offline analysis as well as factorial design for experimental design, i.e. reduced experimentation time and less reagent consumption.

References

1. Kockmann, N.; Gottsponer, M.; Roberge, D. M., Scale-up concept of single-channel microreactors from process development to industrial production. Chemical Engineering Journal 2011, 167 (2-3), 718-726. 2. Weeranoppanant, N., Enabling tools for continuous-flow biphasic liquid–liquid reaction. Reaction Chemistry & Engineering 2019, 4 (2), 235-243. 3. Kriesten, E.; Alsmeyer, F.; Bardow, A.; Marquardt, W., Fully automated indirect hard modelling of mixture spectra. Chemometrics and Intelligent Laboratory Systems 2008, 91 (2), 181-193.

PUBLISHED ARTICLE:

Transition to sustainable chemistry through digitalization

P. Fantke, C. Cinquemani, P. Yaseneva, J. De Mello, H. Schwabe, B. Ebeling, and A. Lapkin

Modern chemistry is the backbone of our society, but it is also a major contributor to global environmental pollution and the ongoing climate crisis. The transition toward a sustainable future requires a radical transformation of how chemistry is designed, developed, and used. This represents a “break it or make it” challenge for the chemical industry with significant technology lock-in and high entry barriers to radical innovations. We propose that urgently required systemic changes in chemical industry, research and development (R&D), chemicals assessment and management, and education to advance sustainable chemistry are attainable through increased and more rapid adoption of digitalization and new digital tools. This will enable flexible data exchange, increased transparency of information flows along cross-country chemical, material, and product life cycles, and chemistries that are safe and sustainable by design, addressing the complexity of chemicals-environment-health interactions and lowering the costs of entry into chemical R&D and manufacture, and new, more sustainable and collaborative business models.

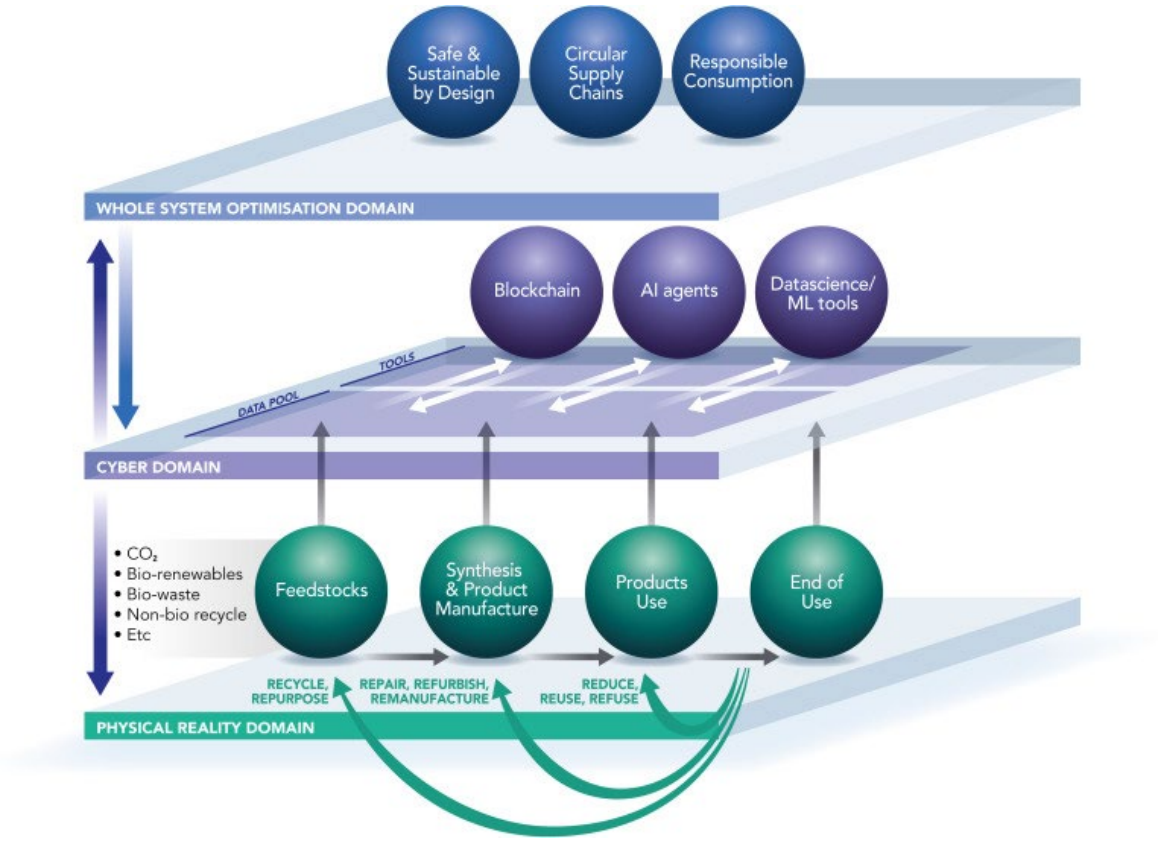


Figure 1 - Data points on the value chain of the chemical industry from raw materials to finished products and interfaces for digitalization and big data approaches to boost safer and more sustainable chemistries

Read the full article:
10.1016/j.chempr.2021.09.012

Notes

Notes

Innovation Centre in Digital Molecular Technologies

An open innovation platform for the digital transformation of chemical synthesis and process development

*i*DMT Partners:

University of Cambridge

SHIONOGI

AstraZeneca

This centre is part funded by:

European Union
European Regional Development Fund