CAMBRIDGE CARES
CAMBRIDGE CENTRE FOR ADVANCED RESEARCH AND EDUCATION IN SINGAPORE LTD.

Pharma Innovation Programme Singapore (PIPS)
A*STAR

New Processes Development in Pharmaceutical R&D and Manufacture

Alexei Lapkin

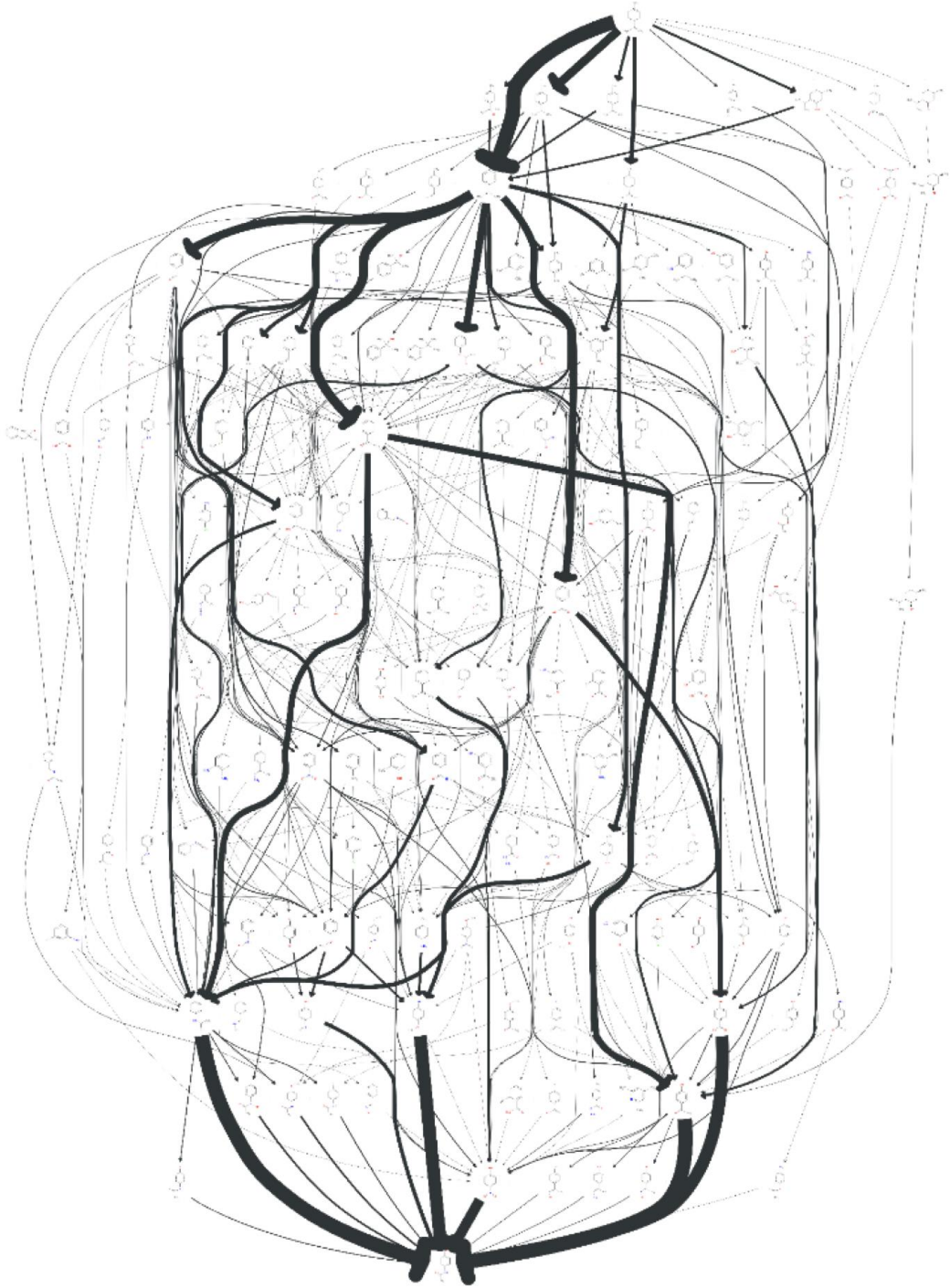
CARES 10-year anniversary
1/12/2023

Sustainable Reaction Engineering

UNIVERSITY OF CAMBRIDGE

Generic Challenges of Process Development in Pharma

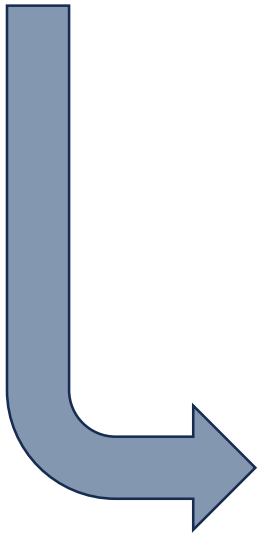
- Synthesis of an active pharmaceutical ingredient (API) is typically a multi-step sequence of synthesis and purification tasks.
- There are different options of how to build the required molecule (different synthetic paths).
- There is a large number of combinations of possible reagents and reaction conditions.



An example of possible chemical routes from limonene (a bio-waste based starting molecule) to paracetamol. In total there are 458 routes involving 132 chemical species. Each reaction (arrow) will have a set of optimal reaction conditions and many of the reactions must be followed by product separation (not shown).

Generic Challenges of Process Development in Pharma

- How to efficiently develop a single step (reaction or separation)?
- How to efficiently design the complete sequence of steps?



- Do we know all the unknowns about our reaction system?
- Could we predict / calculate reaction outcome (what is being formed) and reaction conditions?
- Is it possible to generate required data using minimum time and quantities of reactants/reagents?
- Can we calculate or efficiently measure all required physical properties to evaluate separation strategy computationally?
- How to efficiently enumerate all possible options for a multi-step synthesis?
- What is the most effective way of optimizing a multi-step synthesis?

PIPS projects hosted by CARES

Pharma Innovation Programme Singapore

The Pharma Innovation Programme Singapore (PIPS) is an industry-led platform which aims to synergistically and strategically bring together public sector research capabilities and domain expertise of the pharmaceutical industry to enhance the productivity and operational efficiency within Singapore's pharmaceutical sector through leveraging novel manufacturing technologies and data analytics.

CARES PIPS-1 Projects (completed in 2022/2023)

C4: Development of Multi-Step Processes in Pharma
C12: Data-2-Knowledge in the Digital Manufacture of Pharmaceuticals
Pfizer-specific:

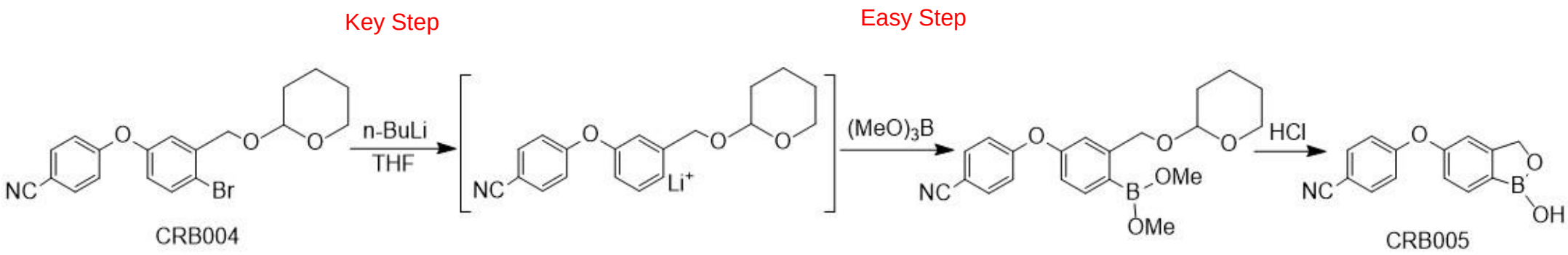
CARES PIPS-2 Projects (started in 2023)

T1: From Digital Twins to Real Time AI-supported Plant Operations
T2: Automated Evaluation of Environmental Impacts of Pharma Manufacturing Processes

Efficient Development of a Single Reaction Step

Li-Halogen exchange is a frequently used synthetic strategy despite well-known practical difficulties:

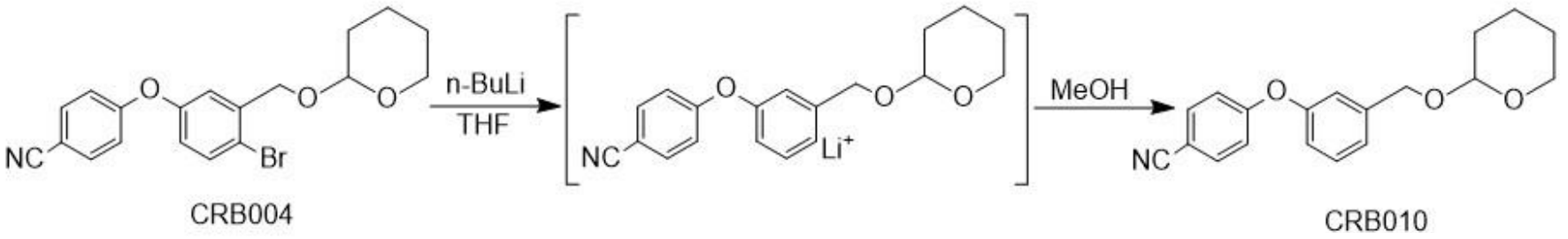
- very high exotherm and reaction rate (requires cryogenic conditions);
- low solubilities – solids formation;
- high sensitivity to traces of water and oxygen;
- low tolerance to functional groups – formation of impurities.



Could we propose a robust procedure based on Li-halogen exchange and used at scale?

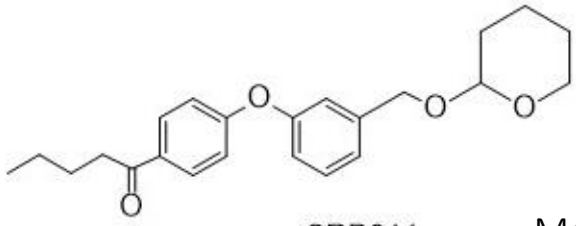
Dr Dogancan Karan

Efficient Development of a Single Reaction Step



CRB004

CRB010

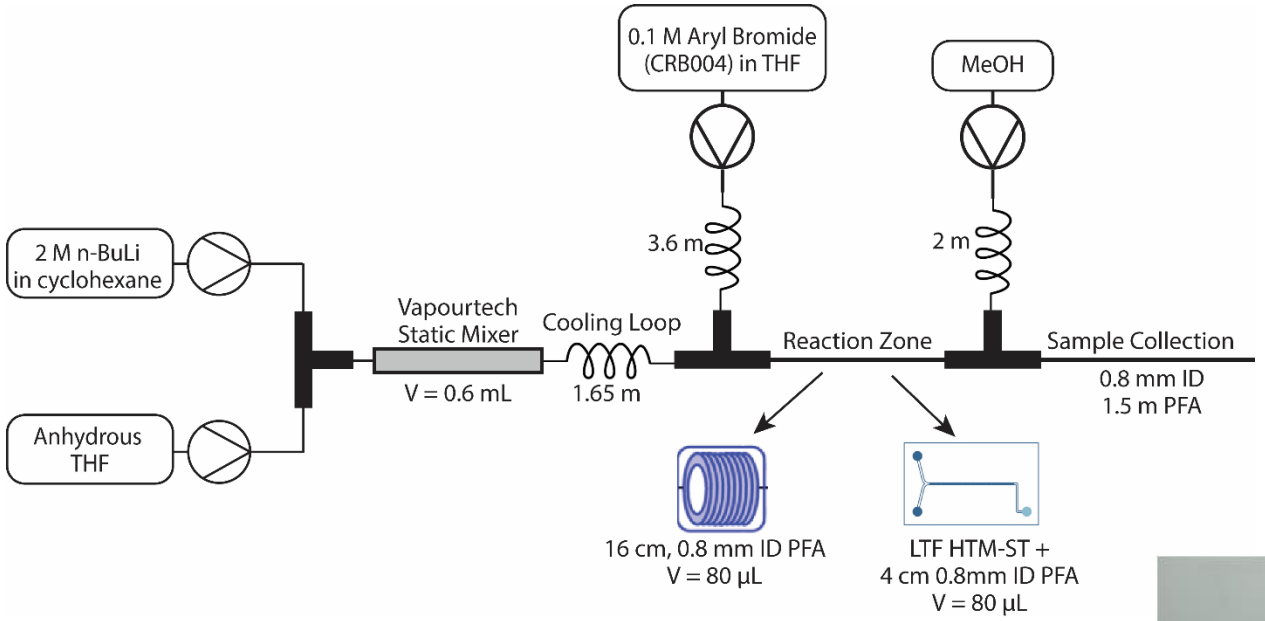


CRB011 Main Impurity

Simplified reaction – focus on the most complex step of Li-halogen exchange.

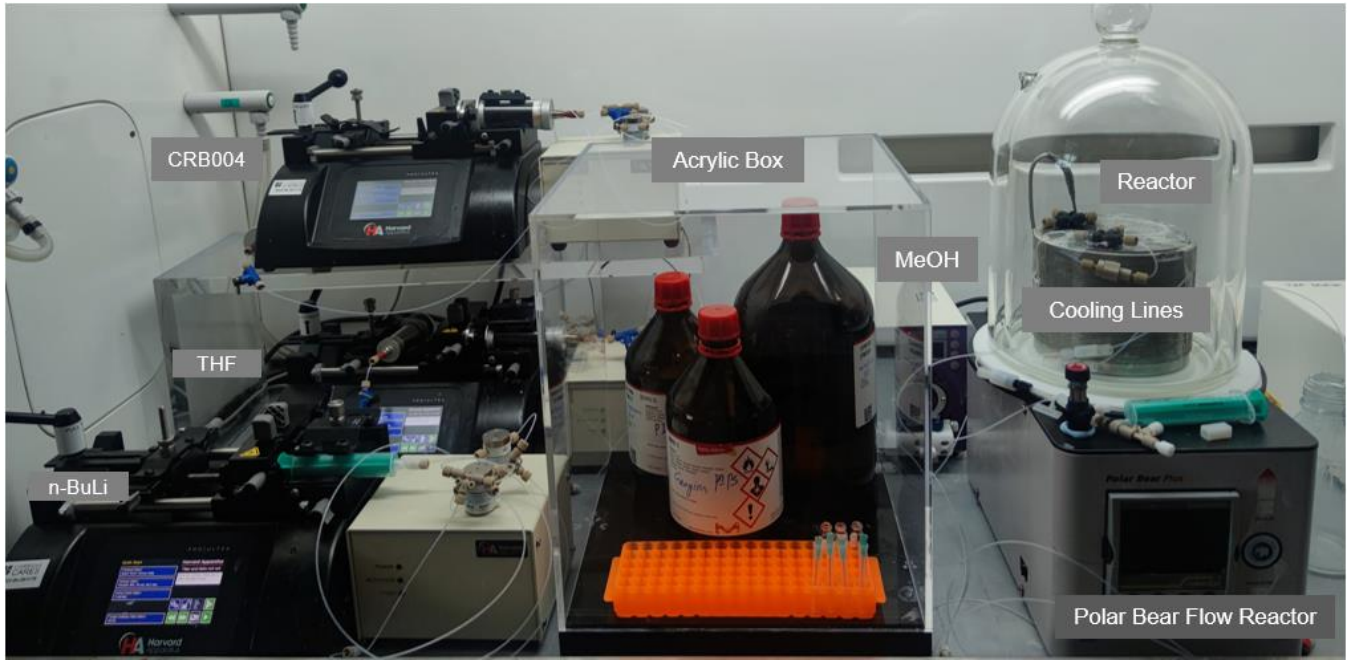
Drs Dogancan Karan & Guoying Chen

Flow Chemistry for ‘Difficult’ Synthetic Procedures



Benefits of flow technology:

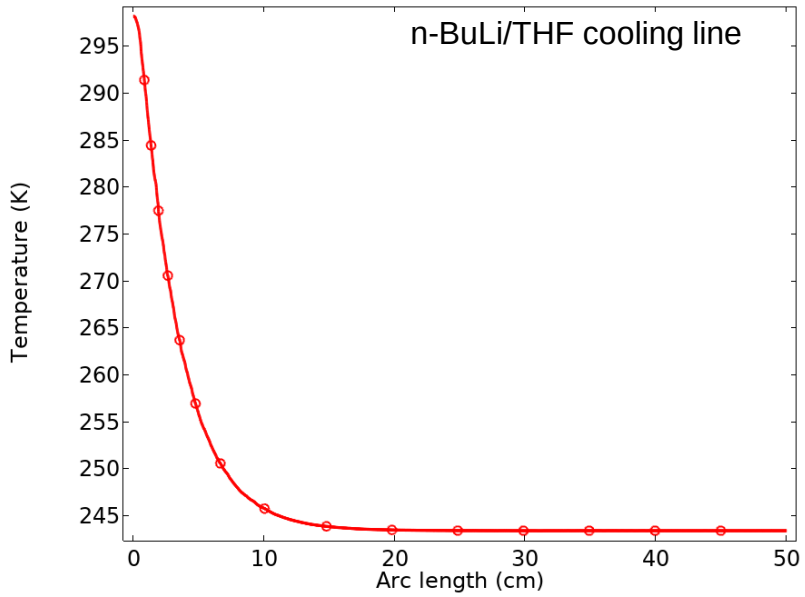
- Many options for mixing arrangement in flow.
- Excellent temperature control – high exchange area.
- Enclosed system for control of environment.
- Low hazard – small reaction volumes.
- High space-time-yield: no need for scale-up.



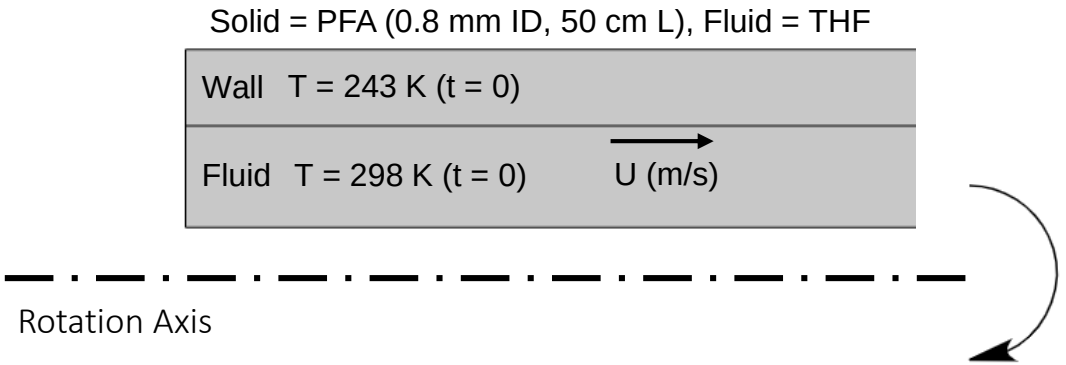
Drs Dogancan Karan & Guoying Chen

Engineering a Good Experiment

- CFD simulations to ensure that cooling tube is long enough to reach the desired temperature.
- Scaled down, axisymmetric 2D geometry to save computational time.
- Highest flow speed (worst case scenario) is simulated.



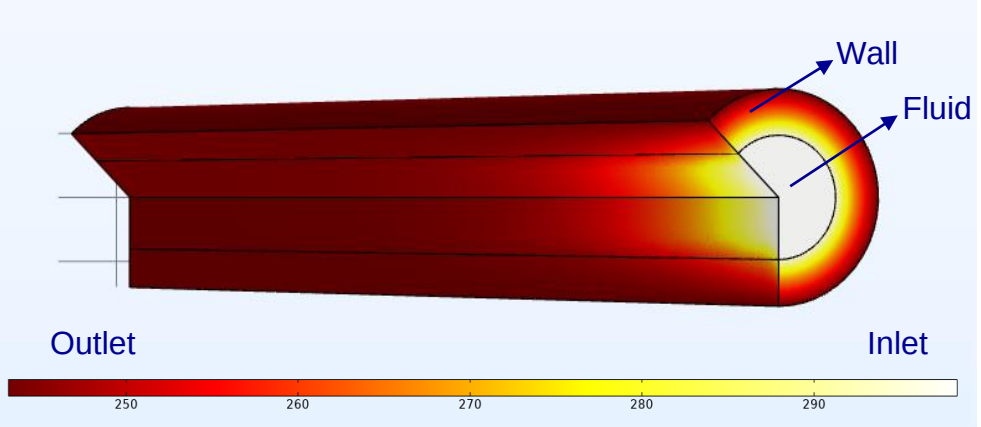
3 – 3.2 m tube length is enough to reach target temperature



$$\rho(\boldsymbol{u}\cdot\nabla)\boldsymbol{u} = \nabla\cdot[-p\boldsymbol{I}+\boldsymbol{K}]+\boldsymbol{F} \quad \rho\nabla\cdot\boldsymbol{u}=0$$ NSE for velocity field

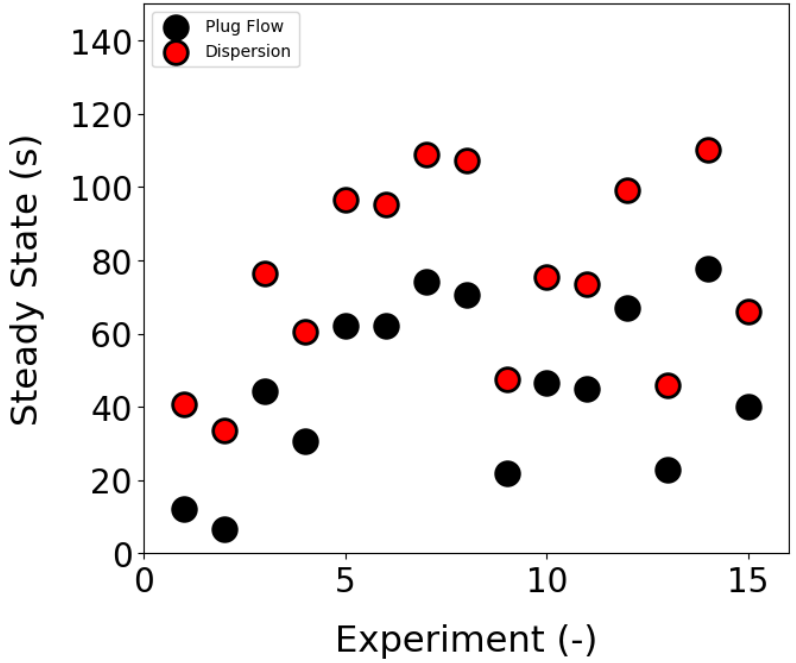
$$\rho C_p\boldsymbol{u}\cdot\nabla T + \nabla\cdot\boldsymbol{q} = Q+Q_{ted} \quad \boldsymbol{q} = -\mathrm{k}\nabla\mathrm{T}$$ Heat transport at solid side

$$\rho C_p\boldsymbol{u}\cdot\nabla T + \nabla\cdot\boldsymbol{q} = Q+Q_p+Q_{vd} \quad \boldsymbol{q} = -\mathrm{k}\nabla\mathrm{T}$$ Heat transport at liquid side



Engineering a Good Experiment

- Sequence of events to collect a data point causes dispersion in multiple locations in the system
- Steady state time needs to be estimated from the theory of Residence Time Distribution (RTD)
- Dispersion regimes are identified from Ud_t/D_{AB} vs L/d_t plot



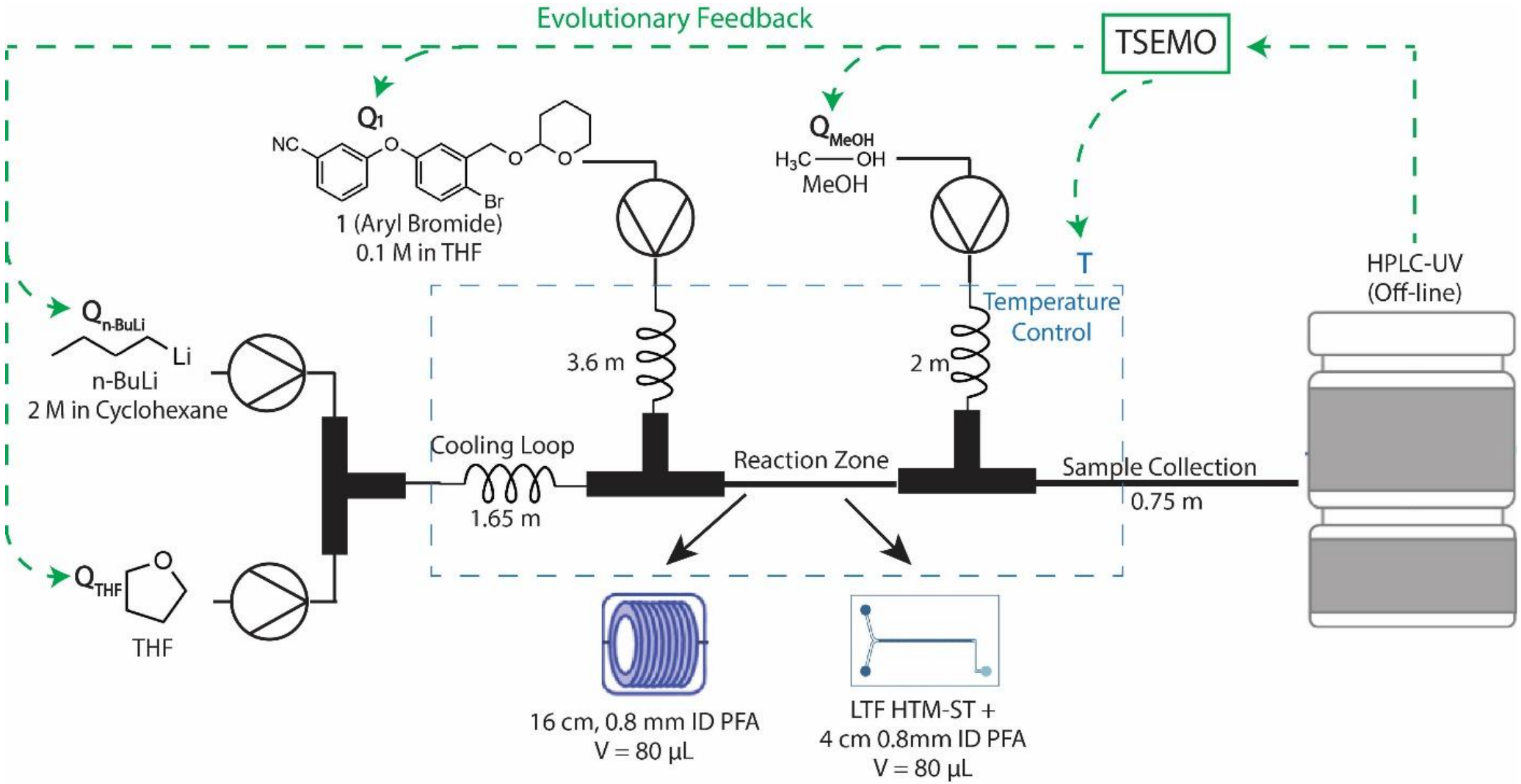
- Taylor dispersion model is used for n-BuLi+THF and sample collection part
- Plug flow is assumed in reaction zone due to short τ
- Predicted steady state with dispersion model is 50 – 100 % higher than plug flow steady state time for 15 randomly generated samples

- Fluctuation in steady state samples (avg of 10 samples)

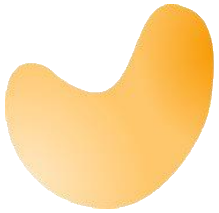
Sample I - 47.3 (± 3.8) % Yield, 19.5 (± 1.1) % Impurity

Sample II – 95.5 (± 0.5) % Yield, 4.7 (± 0.5) % Impurity

Reaction Development & Optimisation using Automated Experiments



An open-source solution



flab

https://pypi.org/project/flab/

About

- A coding *framework* to simplify automation
  - Providing a robust architecture and tools for implementing routines
- Open-source *Python*
  - Zero capital cost, and native migration of many packages for ML & data-science
- Designed for experimentalists with novice coding experience
  - Providing a gradual transition into coding expertise
- Agile principles
  - Making modification and sharing a feature

Automation Workflow

1. Choose the devices you need, and load them into the automation framework

2. Program a task (or multiple tasks) in python with high-level commands
   - i.e. “ device1.start()” or “experiment.start(parameters)”

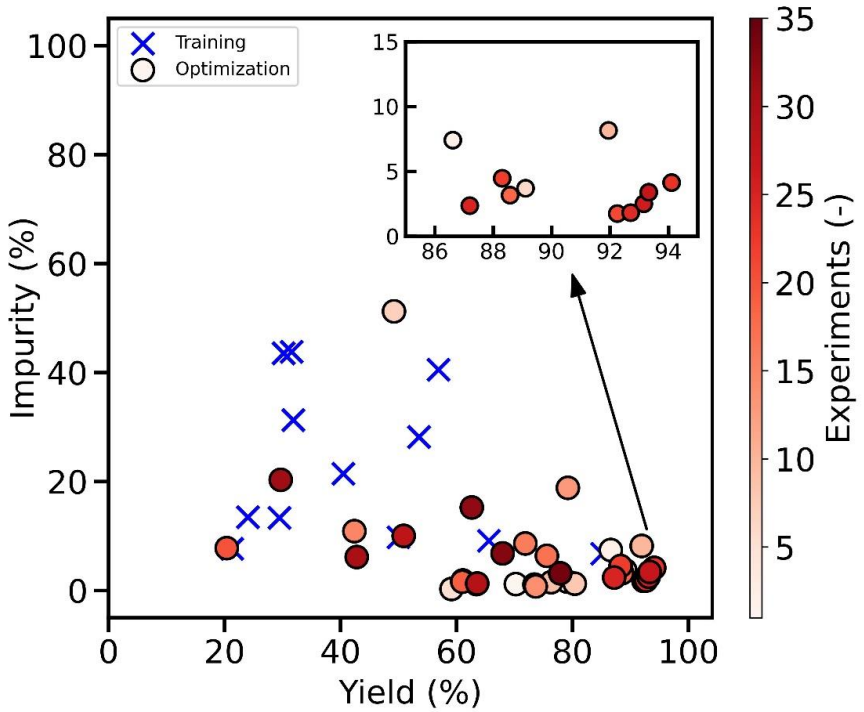
3. Construct AI “loops”

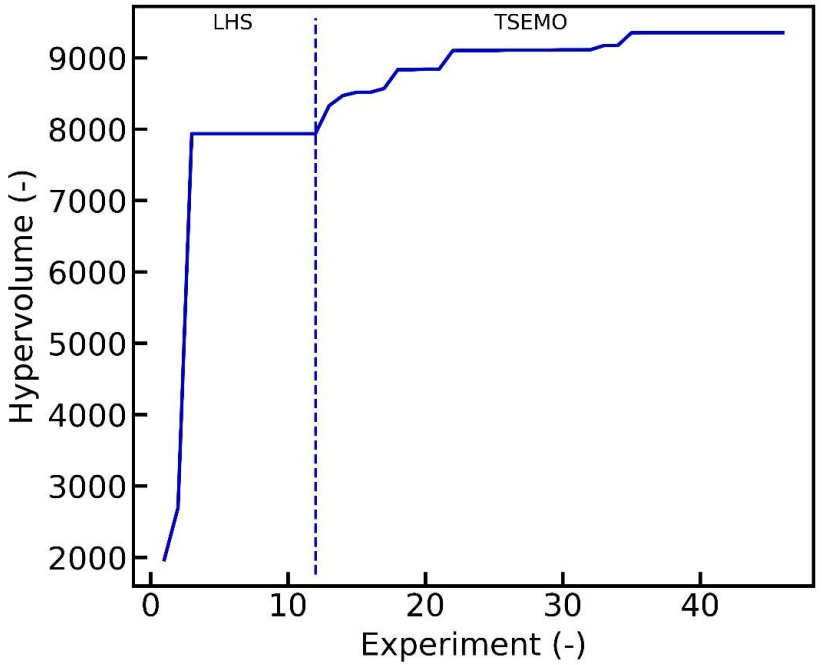
4. Run and stop task(s) as you need during an experiment

5. Increase in complexity as necessary

Generating Process Knowledge through ML-driven DoE

Multi-objective Bayesian Optimization: Yield (max) vs Impurity (min) as competing objectives





- TSEMO designed 34 experiments to establish the pareto plot
- Optimum operation range for max. yield
- τ = 0.185-0.230 s, T = -(30-23) ºC, n-BuLi = 1.03 – 1.08 Eq.

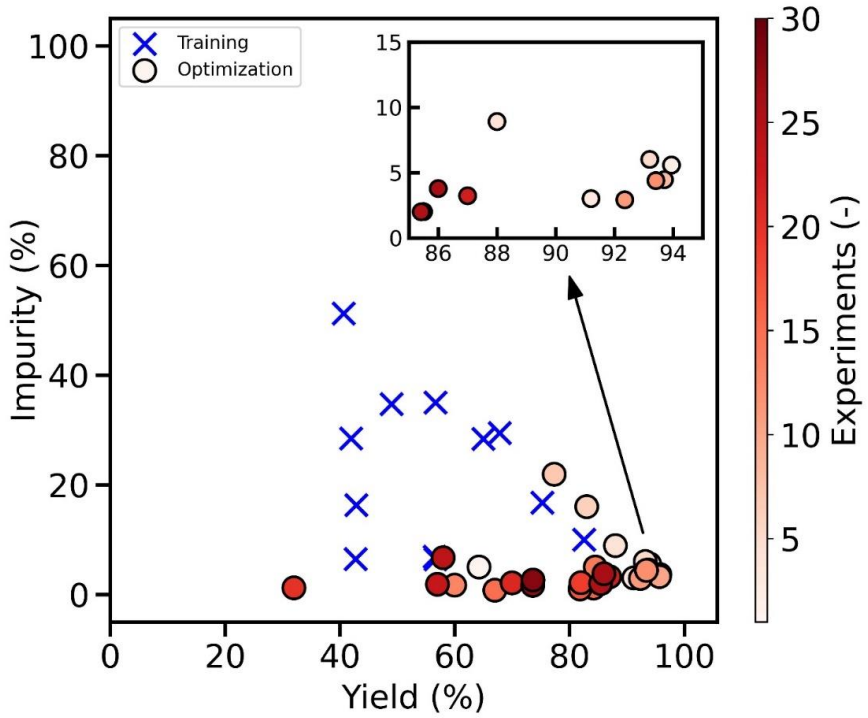
Top 3 best conditions for max yield

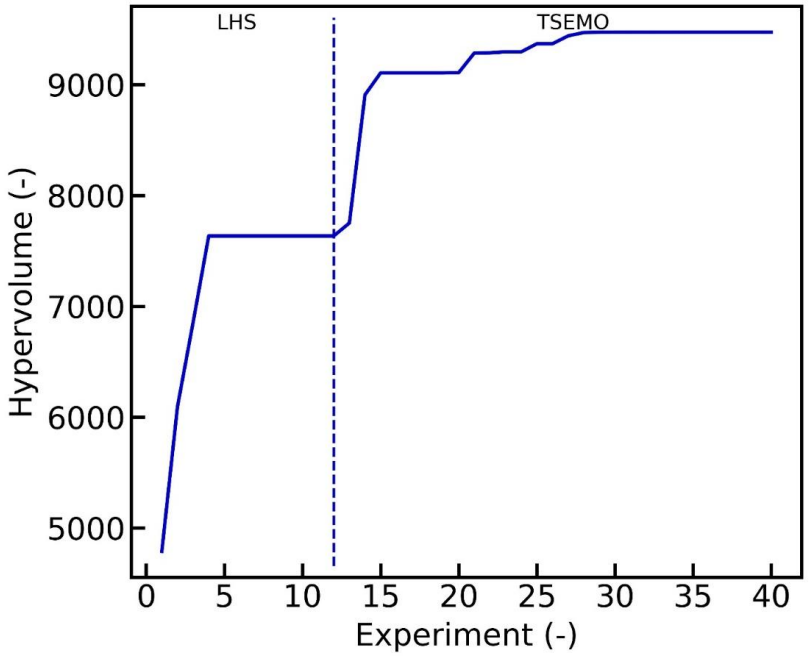
τ (s)	T (°C)	n-BuLi (Eq)	Yield (%)	Imp (%)
0.191	-23.41	1.079	94.1	4.1
0.23	-30	1.047	93.3	3.4
0.185	-26	1.037	93.1	2.5

Reactor setup = 16 cm, 0.8mm ID PFA tubing

Generating Process Knowledge through ML-driven DoE

Multi-objective Bayesian Optimization: Yield (max) vs Impurity (min) as competing objectives





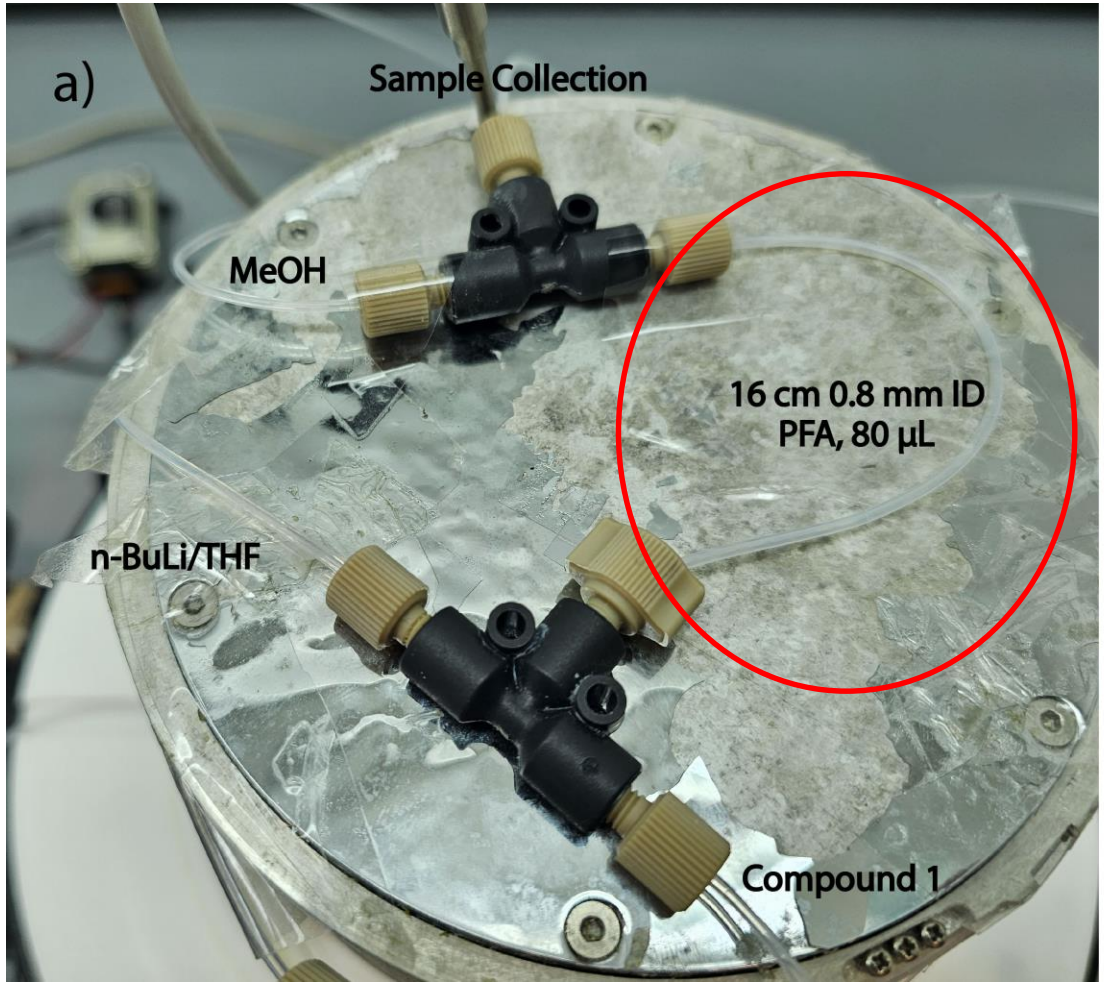
- TSEMO designed 28 experiments to establish the pareto plot
- Optimum operation range for max. yield
- τ = 0.185-0.266 s, T = -(30-23) ºC, n-BuLi = 1.0 – 1.005 Eq.

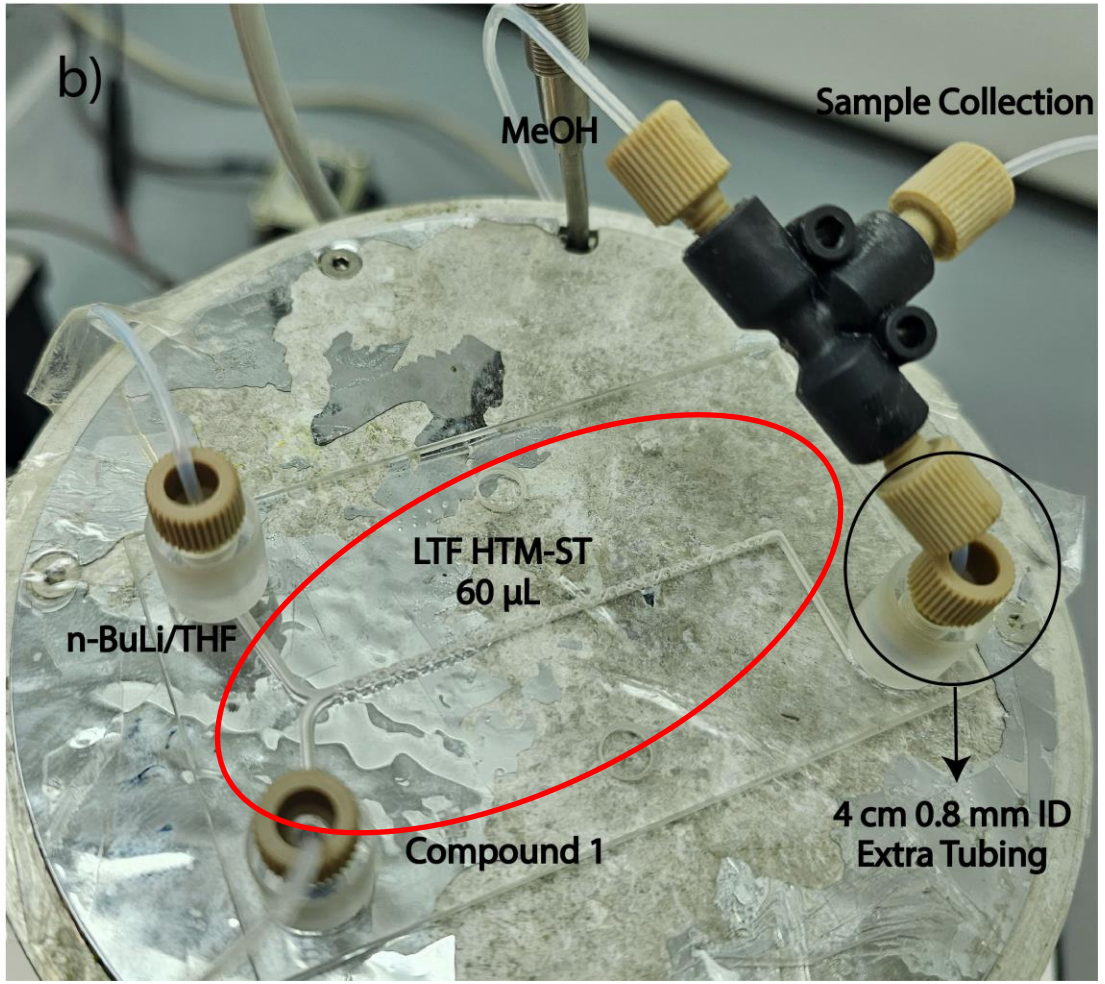
Top 3 best conditions for max yield

τ (s)	T (°C)	n-BuLi (Eq)	Yield (%)	Imp (%)
0.24	-28.6	1.003	95.7	3.6
0.266	-23.5	1.005	95.6	3.3
0.19	-23.4	1.06	93.9	5.57

Reactor setup = LTF HTM-ST + 4 cm, 0.8 mm ID PFA tubing

Tubular reactor vs Microfluidic chip reactor





Tubular reactor vs Microfluidic chip reactor

Comparison between tubular system and chip system

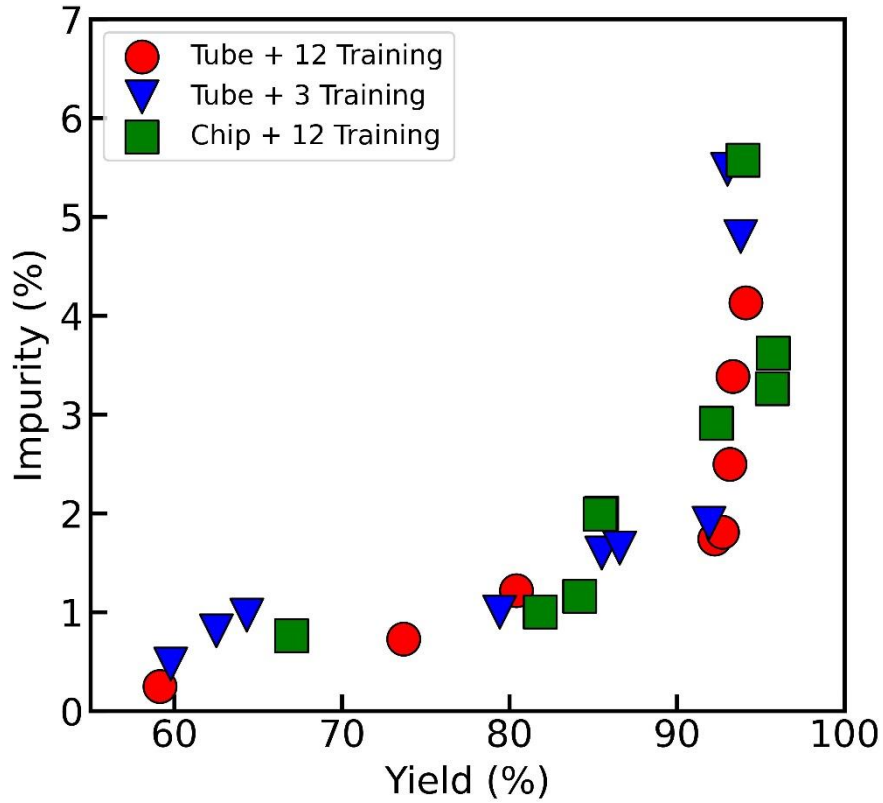
Best performing condition of different optimization campaigns

Optimization	τ (s)	T (°C)	n-BuLi (Eq)	Yield (%)	Imp (%)
Tube+12 Training	0.191	-23.41	1.079	94.1	4.1
Tube+3 Training	0.185	-16.3	1.065	93.8	4.8
Chip+12 Training	0.24	-28.6	1.003	95.7	3.6

- Chip system can achieve higher yield and lower impurity due to better mixing
- Tubular system requires more n-BuLi to achieve the max yield

	τ (s)	T (°C)	n-BuLi (Eq)	Yield (%)	Imp (%)
Tube	0.185	-21	1.1	91.9	8
Chip	0.19	-23.4	1.06	93.9	5.57

- Tubular system requires more n-BuLi to compromise the yield for impurity



Consistent Pareto efficient points for two optimization campaigns with tubular system

Slightly better pareto efficient points for chip system due to improve mixing

Efficient data generation in an automated laboratory

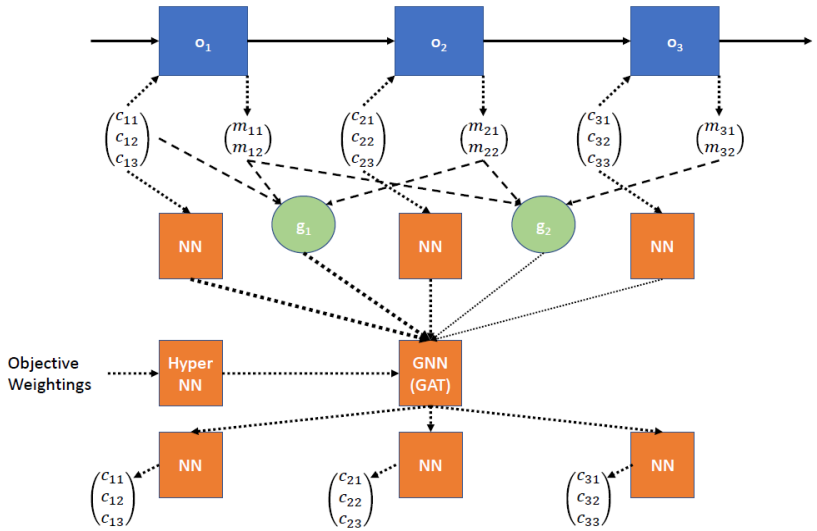
C4 PIPS-1 Project: Development of Multi-Step Processes in Pharma

Project Objectives

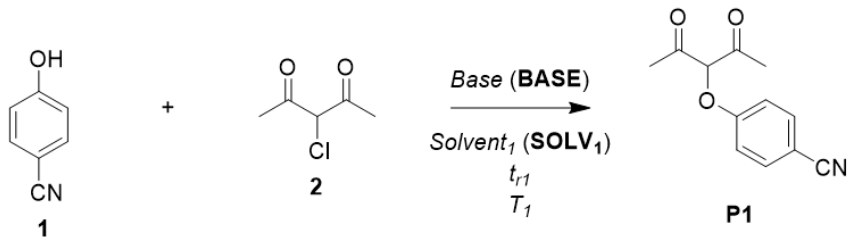
- To develop a generic methodology for accelerated development of multi-step reaction-separation processes in the manufacture of pharmaceuticals.
- To demonstrate the methodology on specific industry-defined case studies.
- To identify critical technological challenges for implementation.

Drs Mohammed Jeraal, Simon Sung, Magda Barecka

Industrial Case Study

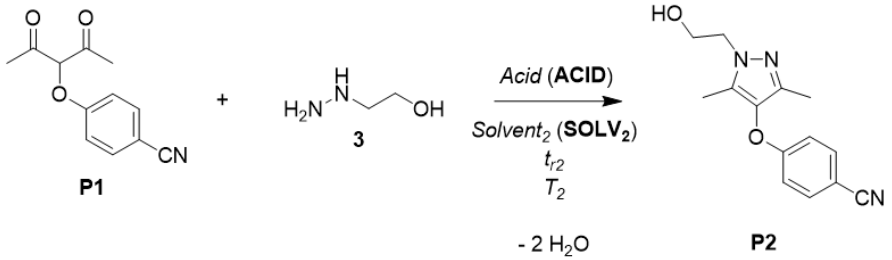


Step 1



Entry	Continuous variables	Discrete variables
1	Reaction temperature, T_1	Type of **BASE**
2	Reaction time, t_{r1}	Type of **SOLV$_1$**
3	Equivalents of **6**	
4	Equivalents of **BASE**	

Step 2



Entry	Continuous variables	Discrete variables
1	Reaction temperature, T_2	Type of **ACID**
2	Reaction time, t_{r2}	Type of **SOLV$_2$**
3	Equivalents of **7**	
4	Equivalents of **ACID**	

SOLV$_1$ Options

- Methanol
- Chloroform
- 1,3-Dimethyl-2-imidazolidinone
- Tetrahydrofuran
- N, N'-Dimethylpropyleneurea

BASE Options

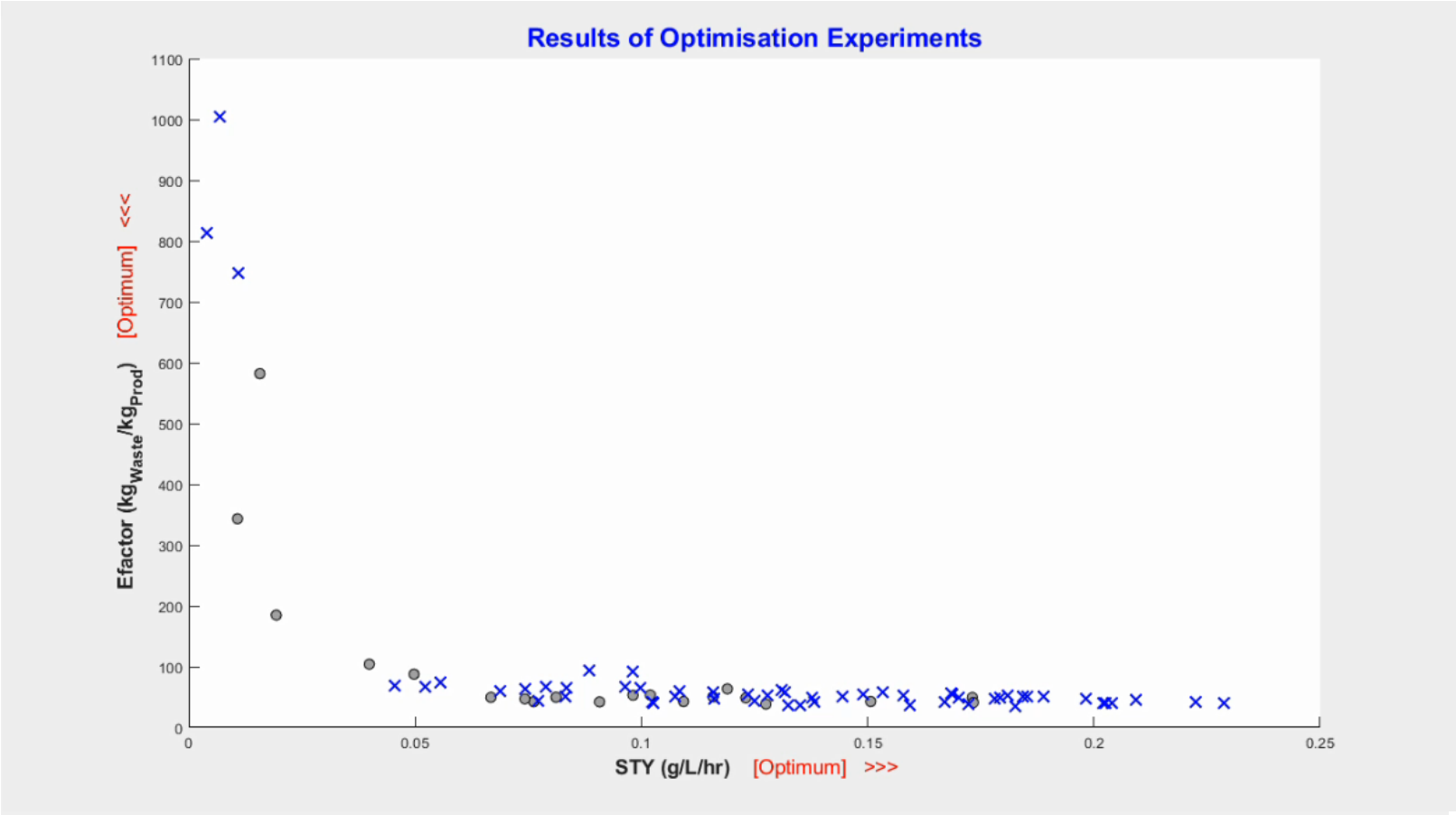
- 1,8-Diazabicycloundec-7-ene
- 1,5-Diazabicyclo[4.3.0]non-5-ene
- N-Methylmorpholine
- N,N-Diisopropylethylamine
- Triethylamine
- 2,6-Lutidine
- Pyridine

ACID Options

- Acetic Acid
- Cyclobutanecarboxylic acid
- Cyclopentanecarboxylic acid
- Difluoroacetic acid
- Formic acid
- Isobutyric acid
- Isovaleric acid
- Lactic acid
- Pentanoic acid
- Propionic acid

Machine-learning based design of experiments: TSEMO

Optimisation of STY (g/L/h) and E-Factor (Reaction Mass/Product Mass)

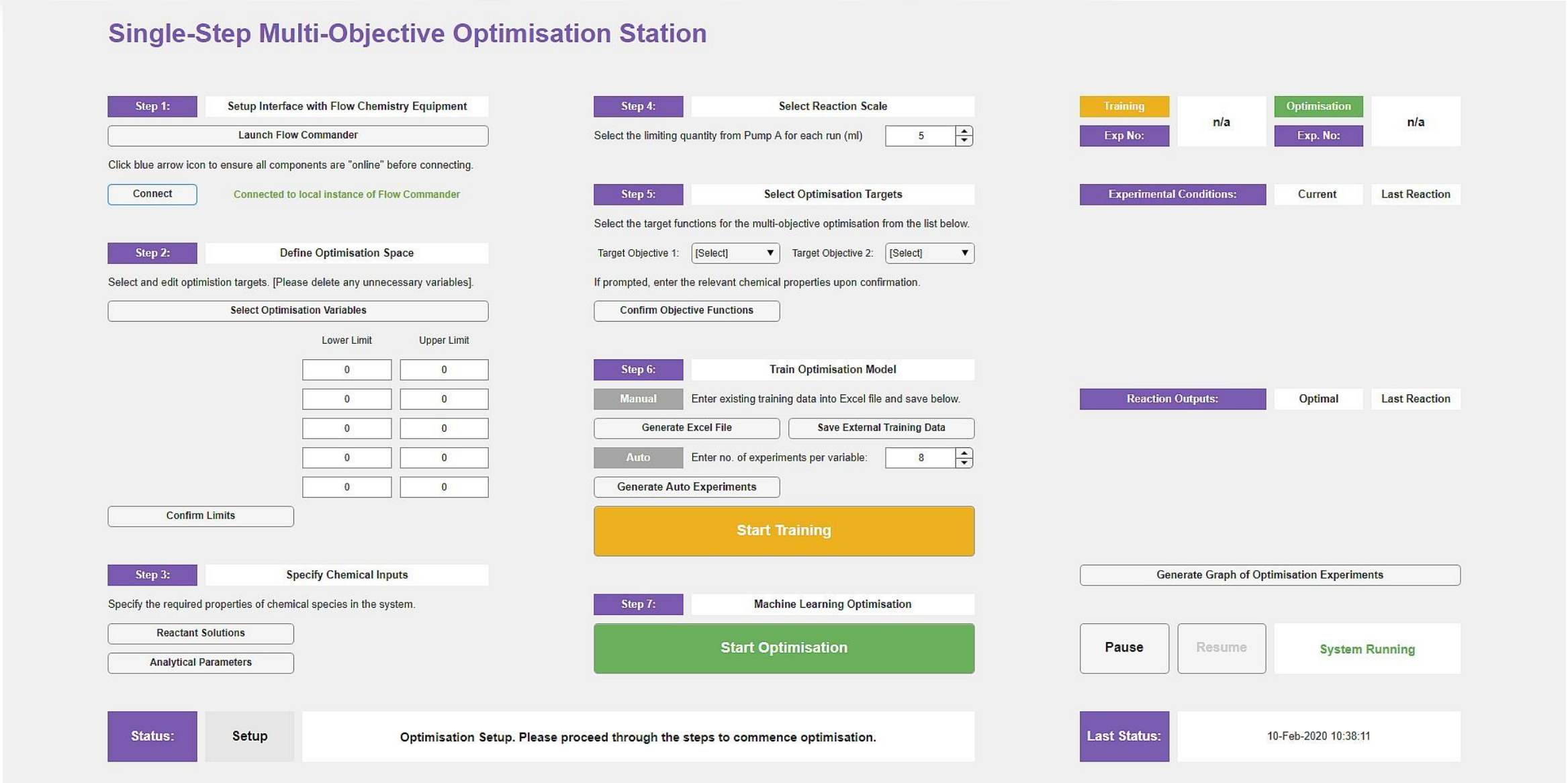


- Optimisation space same as original
- Same training experiments as original optimisation
- System performed 74 optimisation experiments.

Drs Mohammed Jeraal, Simon Sung, Magda Barecka

M.I. Jeraal, S. Sung, A.A. Lapkin, A Machine Learning-Enabled Autonomous Flow Chemistry Platform for Process Optimization of Multiple Reaction Metrics. Chemistry–Methods, 2 (2021) 71-77

User interface for working with automated experiments

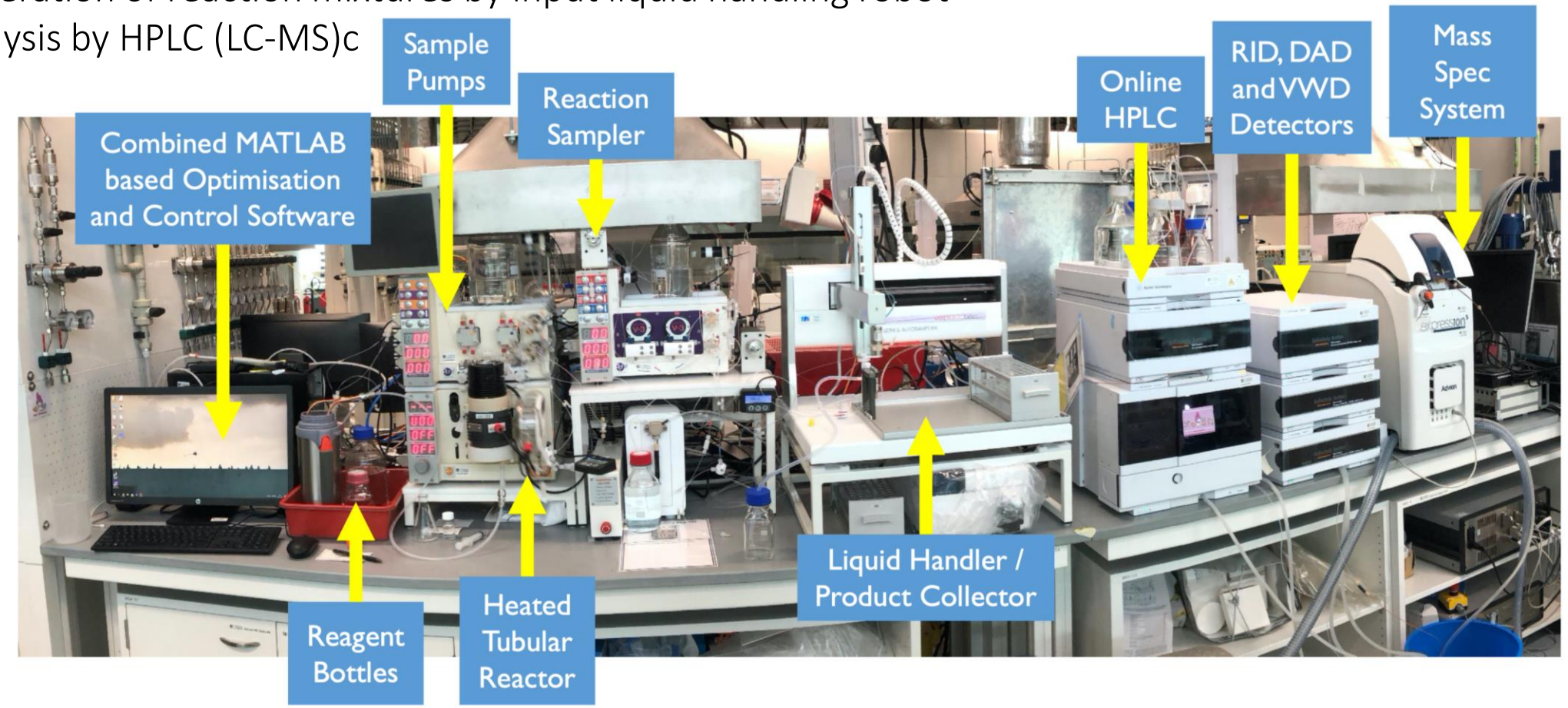


Drs Mohammed Jeraal, Simon Sung, Magda Barecka

A chemical robot = an automated Vapourtec flow reactor + Liquid Handler + HPLC

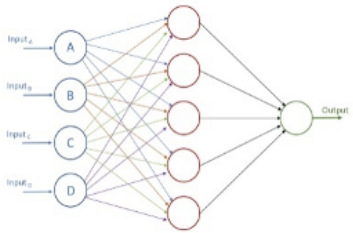
Components:

- Flow chemistry experiment – precise control of conditions, broad range and rapid change of operating conditions
- Reactions in slugs to conserve material and increase data throughput
- Generation of reaction mixtures by input liquid handling robot
- Analysis by HPLC (LC-MS)c



*Nomadic Evolutionary Multiobjective Optimisation (**NEMO**)*

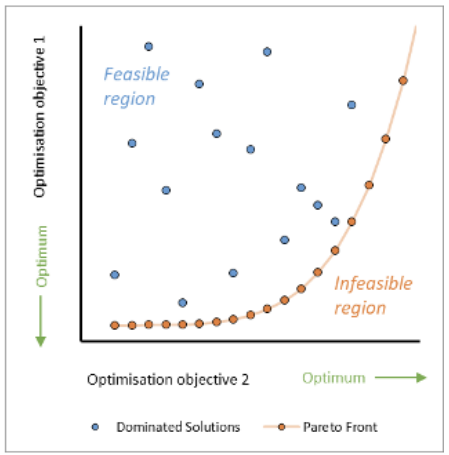
1. (Re-)Fit black-box models, e.g., Neural Network



Fit experimental data to a model for each black-box objective

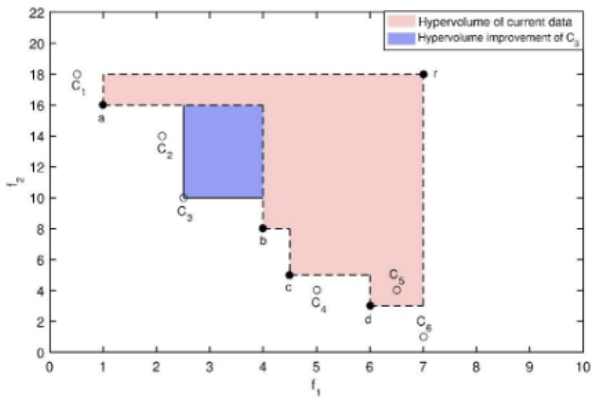
2. Generates new points using Latin Hypercube Sampling (LHS)

Use the models to predict the black-box outcomes of possible non-dominated solutions in NSGA2 (Evolutionary Genetic Algorithm)



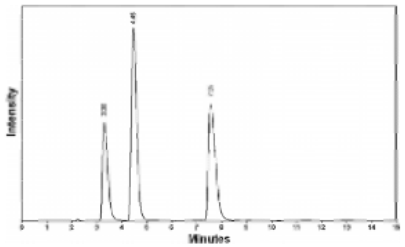
3. Finds combination of points with highest EHVI

Calculate which predicted non-dominated solution gives the best hypervolume improvement and most filled pareto front



4. Run experiments

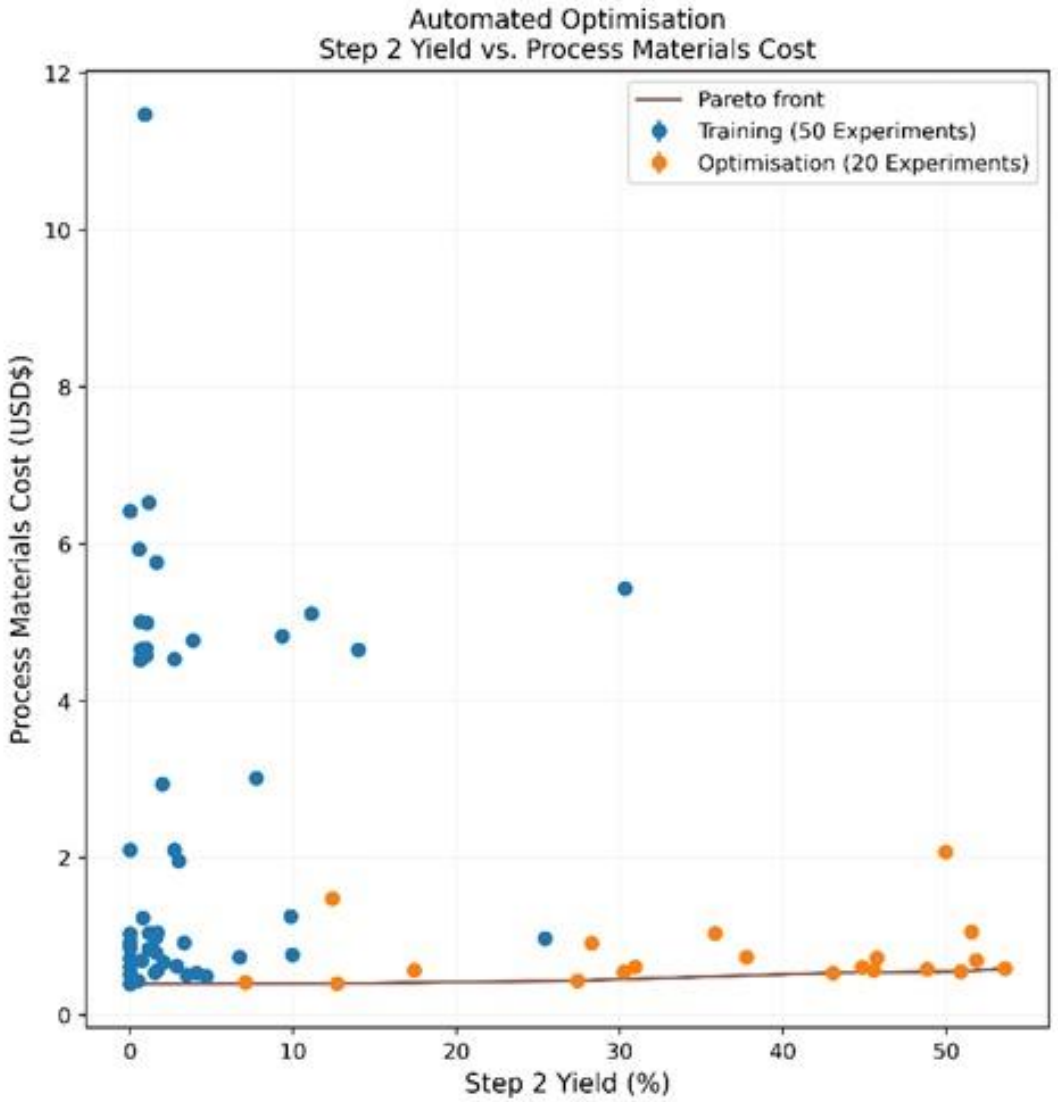
Run the experiment using the proposed best reaction conditions, determine the outcome, and add the new data point to the complete data set



Use SciPy minimize function of the identified points to refine the locations

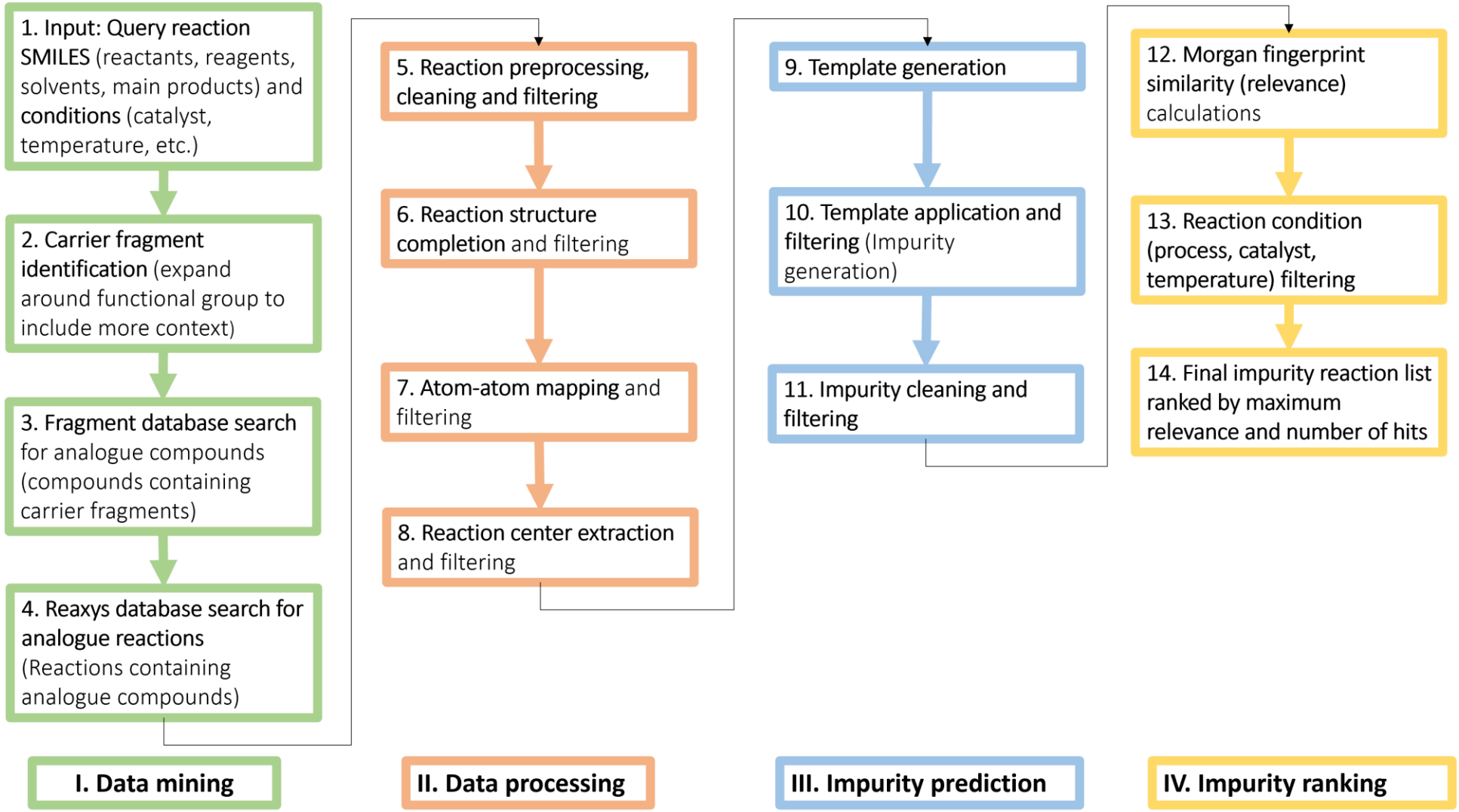
Mohammed Jeraal, Simon Sung

Advance in algorithmic optimisation of reactions



NEMO improves search for optimal conditions in multi-objective optimisation

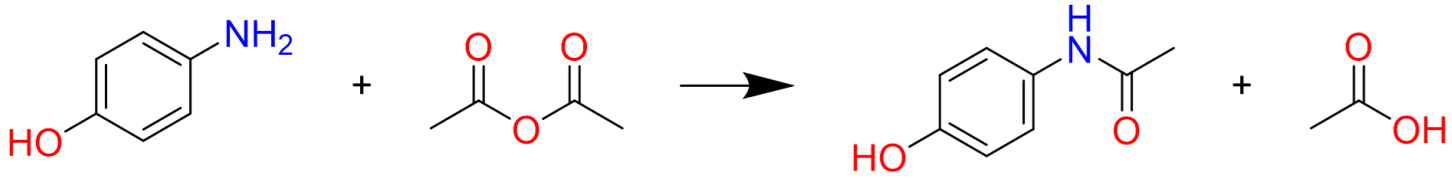
Prediction of reaction impurities in chemical reactions



Adarsh Arun

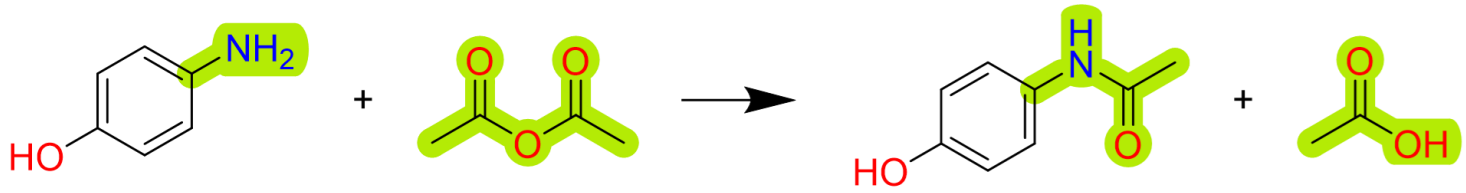
Prediction of reaction impurities in chemical reactions

(a) **Query reaction (user input)**



SMILES : Nc1ccc(O)cc1 . CC(=O)OC(C)=O >> CC(=O)Nc1ccc(O)cc1 . CC(=O)O

(b) **Main product**

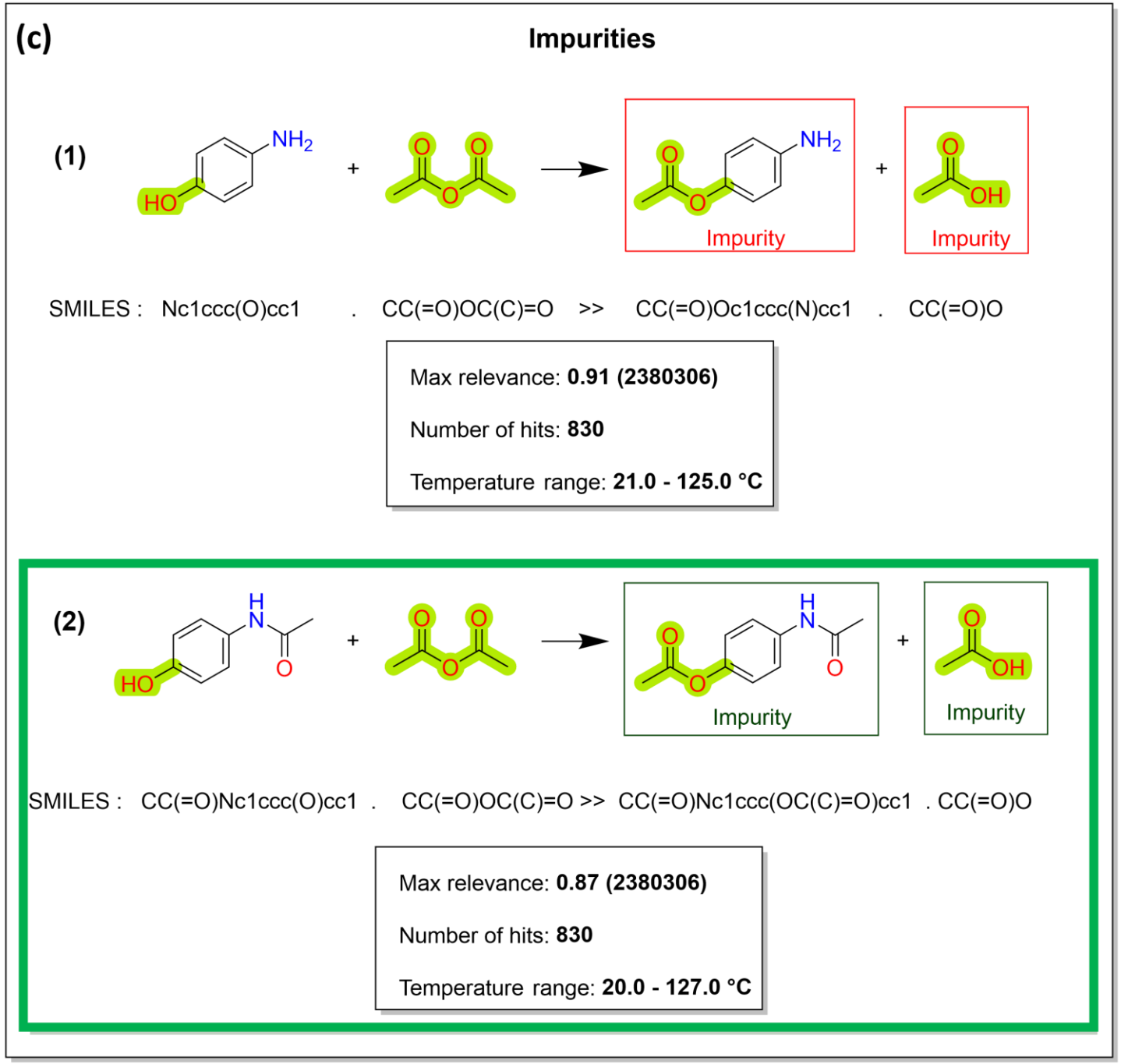


SMILES : Nc1ccc(O)cc1 . CC(=O)OC(C)=O >> CC(=O)Nc1ccc(O)cc1 . CC(=O)O

Max relevance: **1.0 (570100)**

Number of hits: **12180**

Temperature range: **1.0 - 64.0 °C**



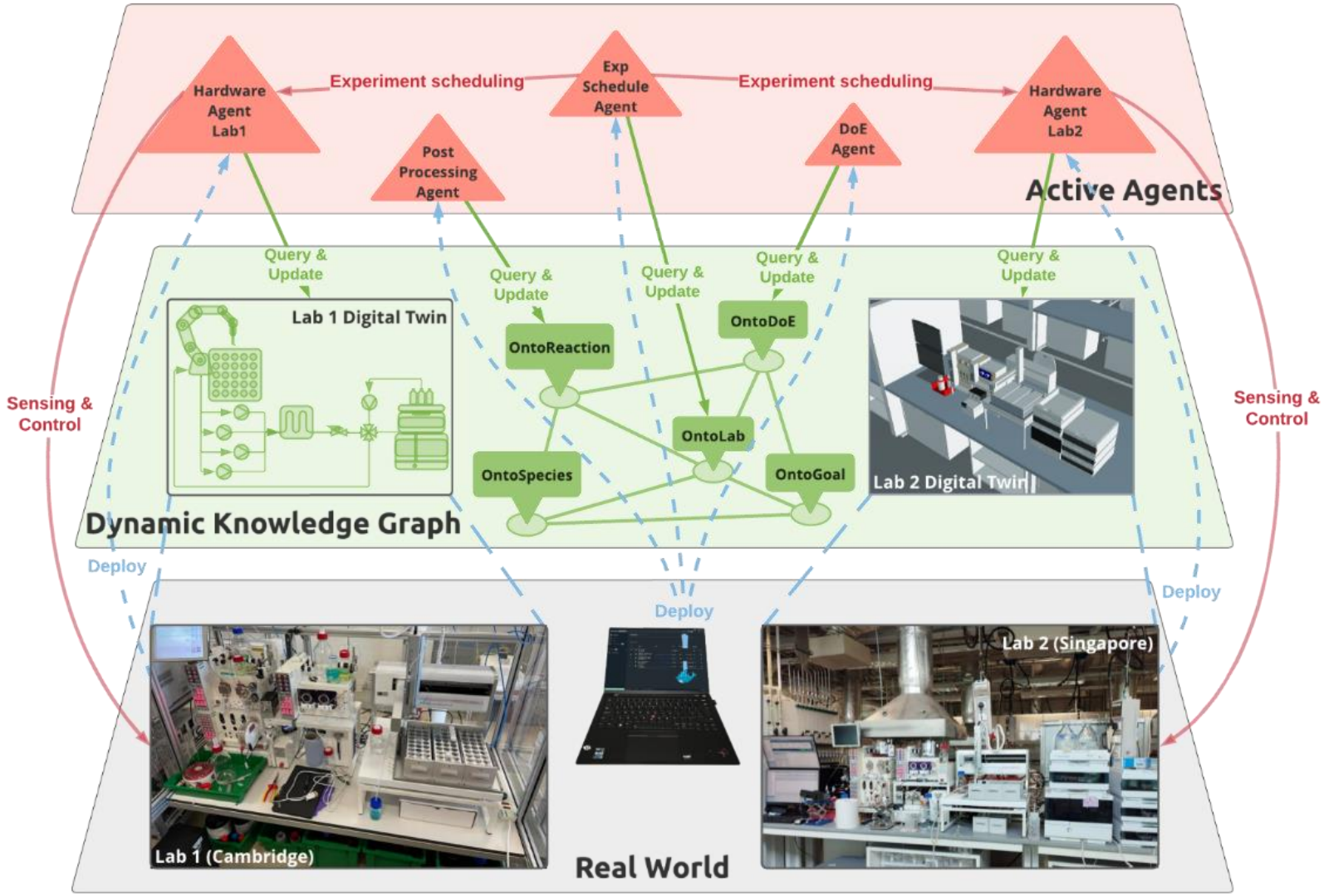
A. Arun, Z. Guo, S. Sung, A.A. Lapkin, Reaction Impurity Prediction using a Data Mining Approach. Chemistry Methods, 3 (2023) e2022000062.

Ongoing PIPS projects hosted by CARES

T1: From Digital Twins to Real Time AI-supported Plant Operation

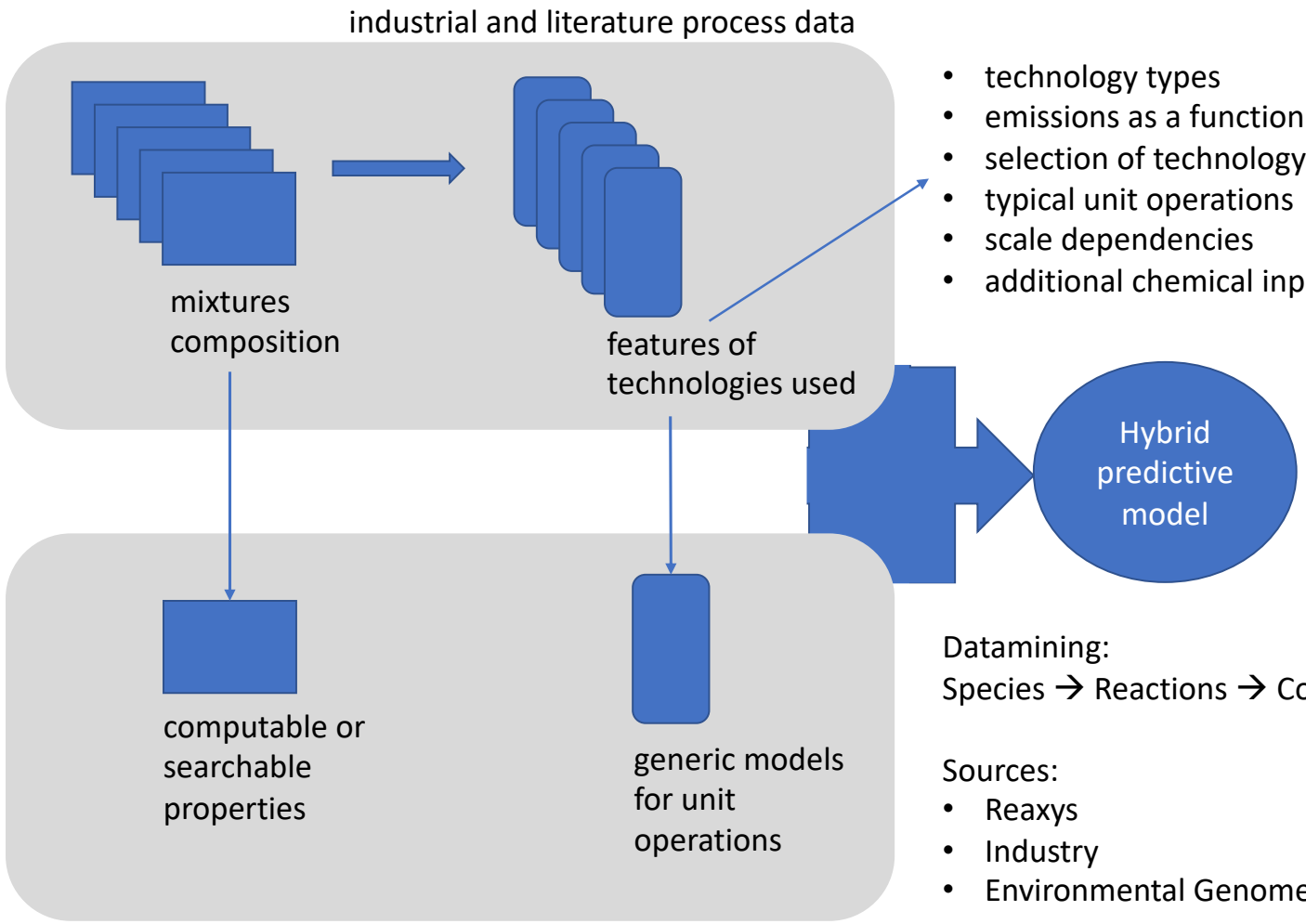
CARES:
Alexei Lapkin (PI)
Markus Kraft

I2R A*STAR:
Lianlian Jiang



Ongoing PIPS projects hosted by CARES

T2: Automated Evaluation of Environmental Impacts of Pharma Manufacturing Processes



- technology types
- emissions as a function of inputs
- selection of technology as a function of composition
- typical unit operations
- scale dependencies
- additional chemical inputs

Hybrid predictive model

Datamining:
Species → Reactions → Conditions → Inventories

Sources:
- Reaxys
- Industry
- Environmental Genome

Acknowledgements

- **Public Funding**

  UKRI, ERDF, NRF Singapore, PIPS(ingapore)

- **Industrial Funding**

  BASF, Pfizer, GSK, MSD, Syngenta, Astex, Elsevier, UCB Pharma, CDI Pte Ltd

- **Academic Collaborations**

  Markus Kraft, Leroy Cronin, Jonathan Goodman, Gabor Csanyi, Pietro Lio, Lucy Colwell, Matthew Gaunt, Ning YAN, Hua Chun ZENG, Saif Khan, Wen LIU, Tej Choksi, Samir Mushrif

Early career researchers

Dr Zhen Guo
Dr Simon Sung
Dr Mohammed Jeraal
Dr Dogancan Karan
Dr Buoying Chen
Dr Magda Barecka
Dr Shuyuan Zhang
Dr Jiyizhe Zhang
Kobi Felton
Adarsh Arun
Jiaru Bai (MK)

CDI Chemical Data Intelligence

Spin-out company offering consultancy and developing a new product
(cdi-sg.com)

SynTech Automated Chemical Synthesis

Centre for Doctoral Training @ University of Cambridge
(syntechcdt.com)

iDMT Innovation Centre in Digital Molecular Technologies

Innovation Centre to work with SMEs hosted by the University of Cambridge
(idmt.online)

Sustainable Reaction Engineering