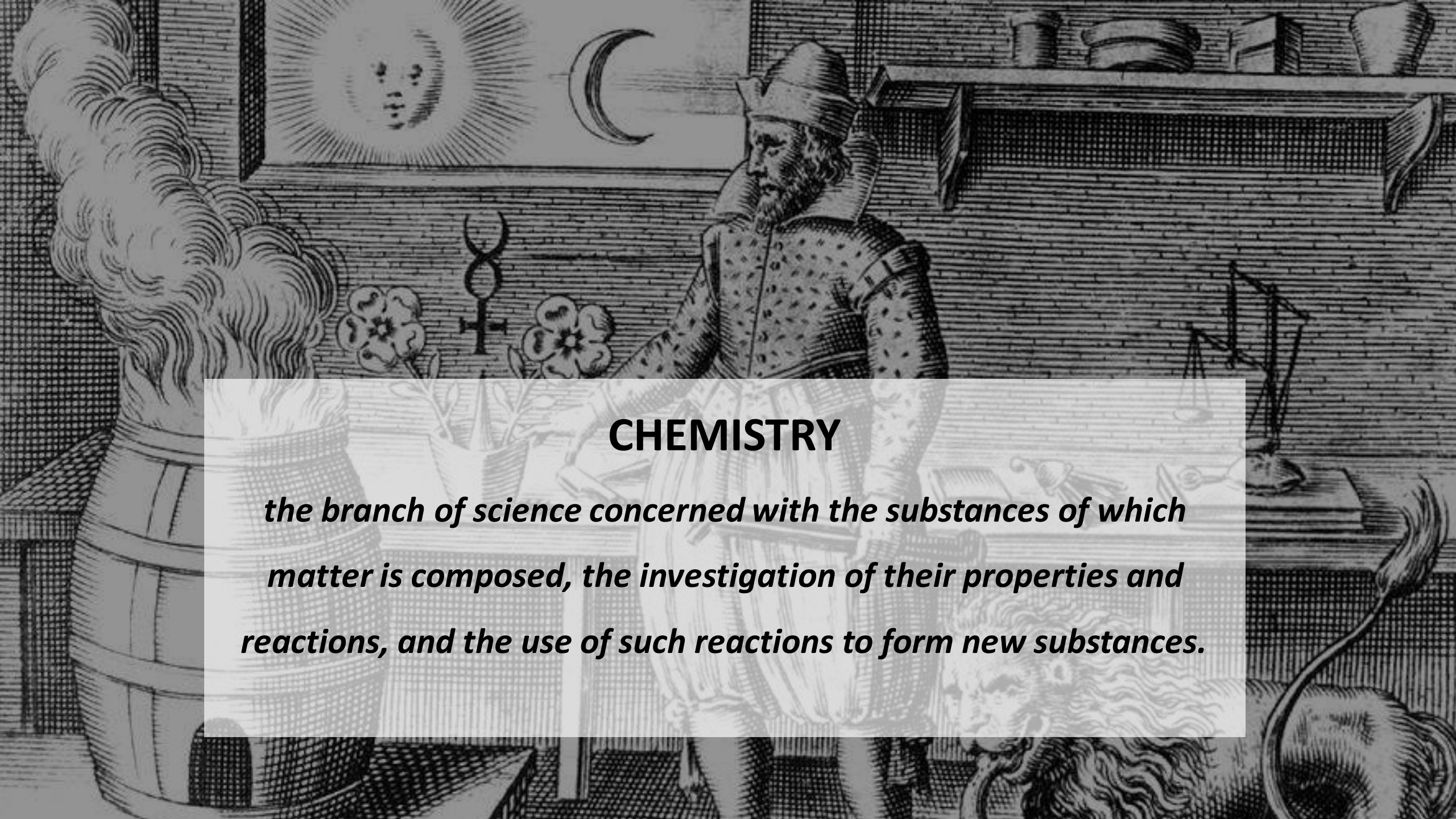
IMPERIAL

Pivoting synthetic chemistry towards a data-led subject

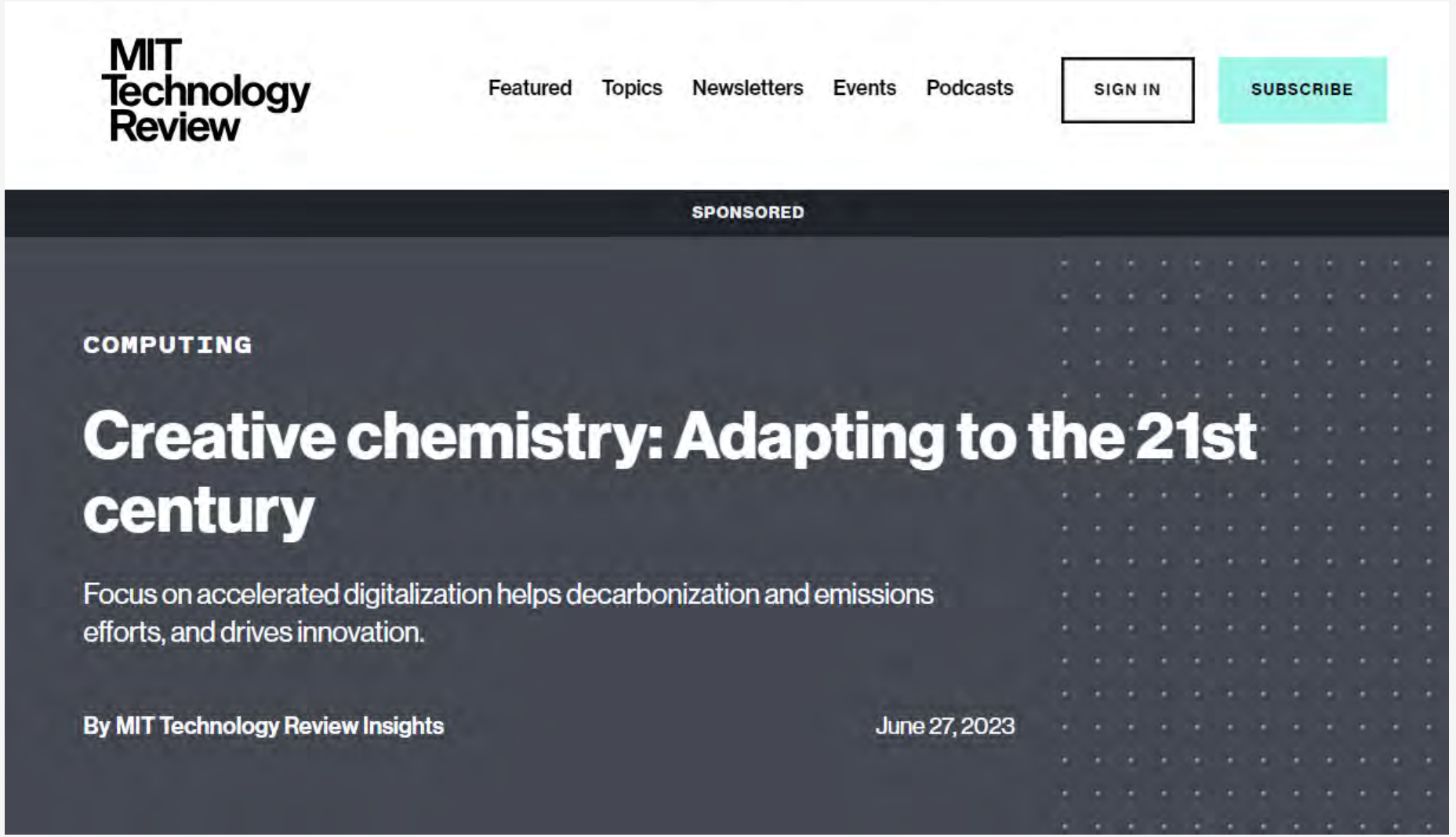
Professor King Kuok (Mimi) Hii

Professor of Catalysis
Director of Imperial Centre for Rapid Online Analysis of Reactions (ROAR)



CHEMISTRY

the branch of science concerned with the substances of which matter is composed, the investigation of their properties and reactions, and the use of such reactions to form new substances.

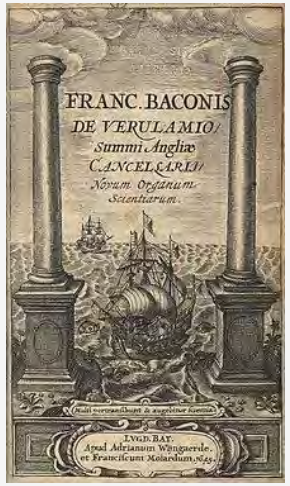


“AI, digital twins, robotics, and lab automation are accelerating research productivity, improving operational efficiency, and enhancing safety. AI, machine learning (ML) and digitization of research are enabling high-throughput research, allowing scientists to build on past efforts efficiently, identify promising ideas, and forecast product performance.”

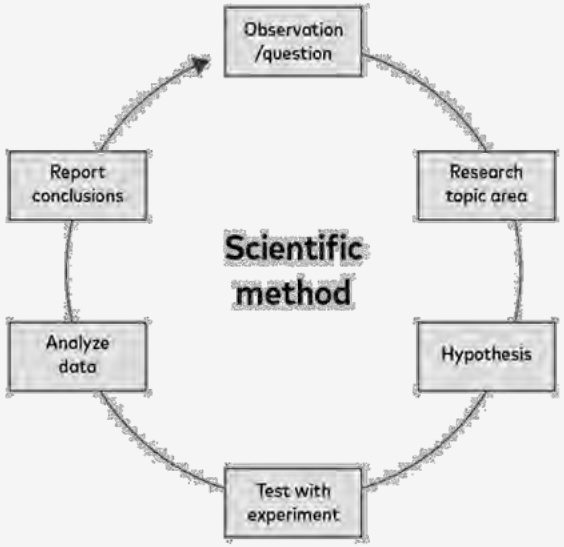
The Scientific Method

- Acquiring knowledge through *observation and experimentation*, rather than *guessing and authority (Aristotelian)*.

Francis Bacon, 1617-1621

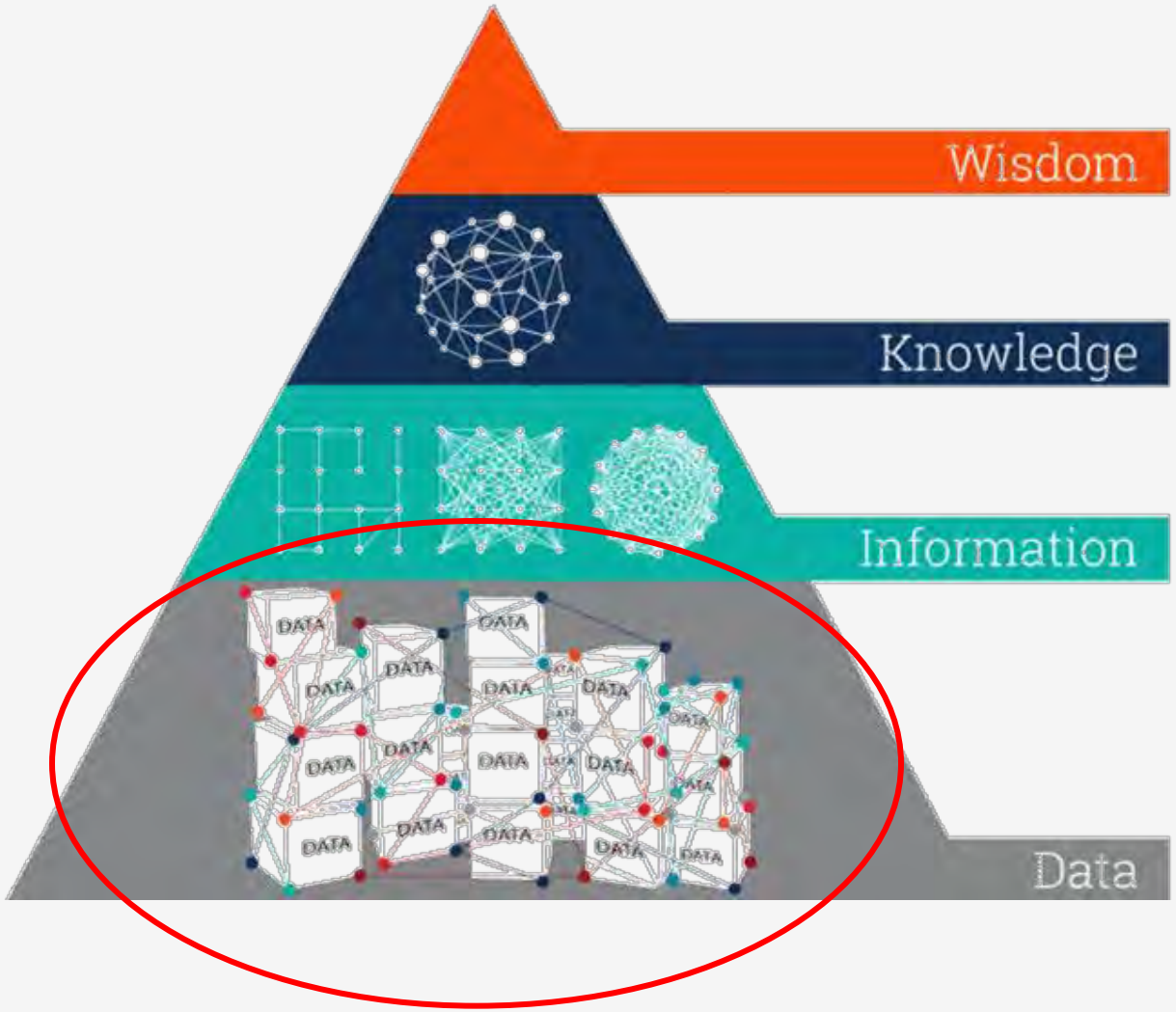


Novum Organum (1620)



Galileo Before the Holy Office in the Vatican
Joseph-Nicolas Robert-Fleury

The DIKW Pyramid:
Acquiring wisdom through data



Allows us to *predict* based on *understanding* –'gut feeling' and 'intuition'

Requires learning, understand relationships, detect patterns

Organise, classify – assign 'meaning' to data

Unorganised, unstructured, unprocessed, noisy.

https://towardsdatascience.com/rootstrap-dikw-model-32cef9ae6dfb

Data scarcity: Looming crisis for the development of AI

PIECES
FOR DEVELOPERS
JUNE 17, 2024

**Data Scarcity: Wh
Wall?**

Martha J. Lindeman, Ph.D.

IBM Products Solutions Consulting Support Think TechXchange 2024

Think Artificial intelligence Cloud Security Sustainability Blog Videos Reports Events More

Explore IBM's AI hallucinations solution → Sign up for AI updates →

- What are AI hallucinations?
- Implications of AI hallucination
- Preventing AI hallucinations
- AI hallucination applications
- Products
- Resources
- Take the next step

What are AI hallucinations?

AI hallucination is a phenomenon wherein a large language model (LLM)—often a generative AI chatbot or computer vision tool—perceives patterns or objects that are nonexistent or imperceptible to human observers, creating outputs that are nonsensical or altogether inaccurate.

ogy and Medicine
mber 2024, 108734

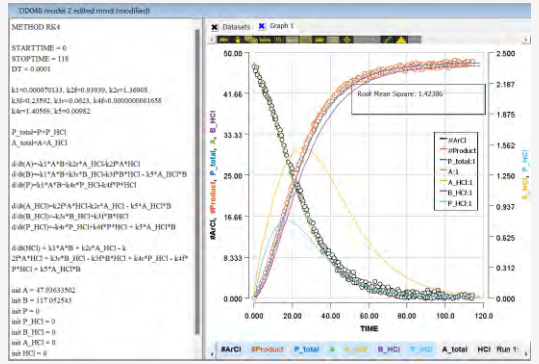
ddress data scarcity
e-based drug
ensive review

, Abul Kalam Azad [d] ,

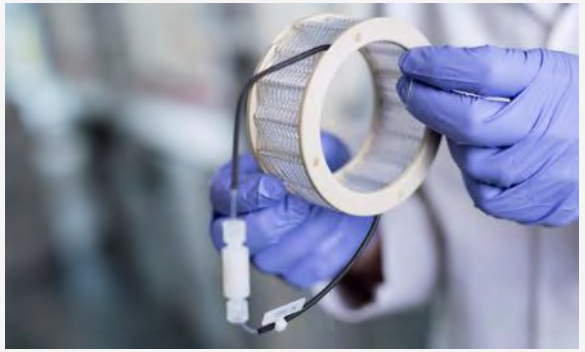
Get rights and content ↗

High Throughput Experimentation

Reaction Kinetics & In-situ Studies



Automated Batch & Flow Reactors



Analysis of Reactions

https://www.imperial.ac.uk/rapid-online-analysis-of-reactions

ROAR
Imperial College
London

rEaCt

EPSRC Centre for Doctoral Training in Next Generation Synthesis & Reaction Technology

ROAR

Imperial College London

ACS Publications C&EN CAS Hi, Mimi! Logout

c&en CHEMICAL & ENGINEERING NEWS TOPICS MAGAZINE COLLECTIONS VIDEOS JOBS

SYNTHESIS

Automation for the people: Training a new generation of chemists in data-driven synthesis

ROAR and other UK initiatives aim to democratize access to high-throughput equipment and industrial know-how

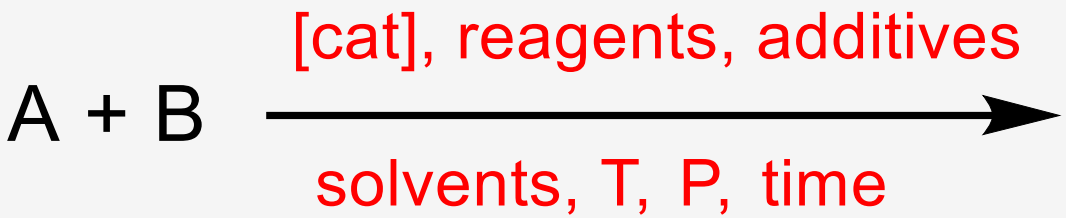
by Mark Peplow

OCTOBER 27, 2019 | APPEARED IN VOLUME 97, ISSUE 42

The ‘classical’ approach to synthesis

Optimise for product yield (single-objective)

The ‘data-centric’ approach to synthesis

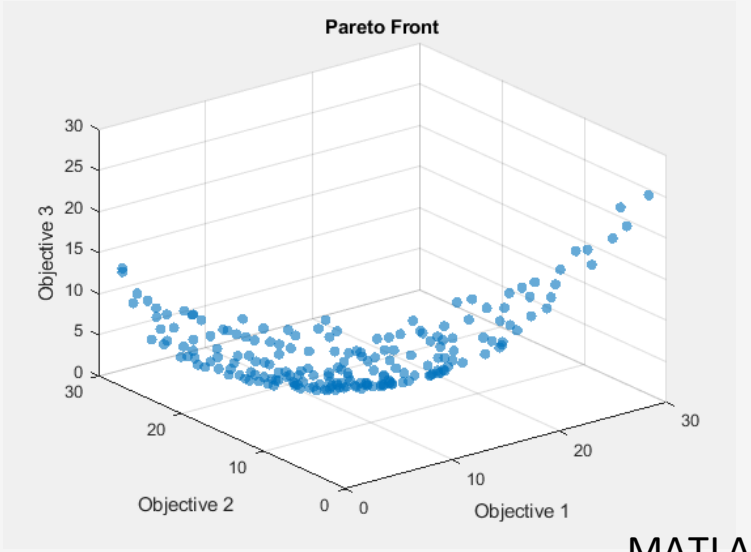


Optimise for product yield (single-objective)

Not necessarily compatible with a **‘process-driven’** approach…

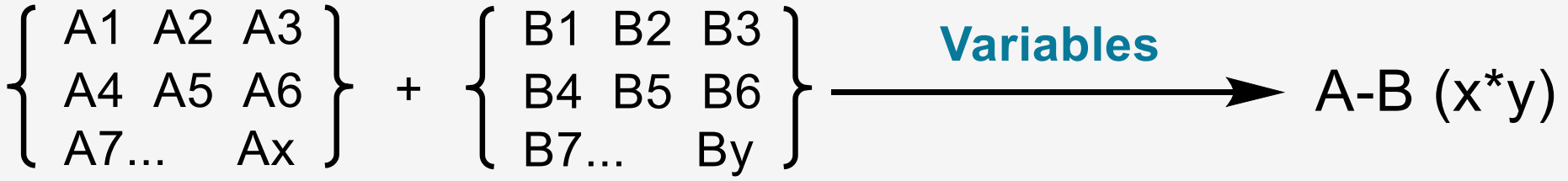
- Optimised: Productivity, efficiency
- Reliable: Quality control
- Sustainable: Waste/energy management
- H&S: Hazard management
- Cost: Availability of feedstock, reagents

Multi-objective

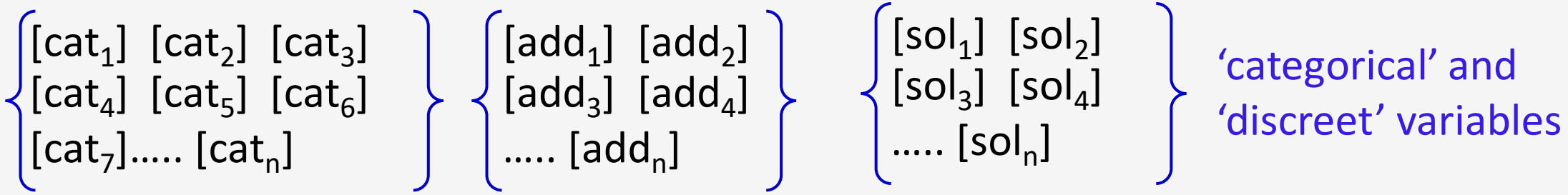


MATLAB

The ‘data-driven’ approach



‘categorical’ variables



‘categorical’ and
‘discreet’ variables

Temperature, pressure, time: ‘continuous’ variables.

To understand and quantify the relationships between the variables ➡ **KNOWLEDGE**

‘Digital Chemistry’ workflow

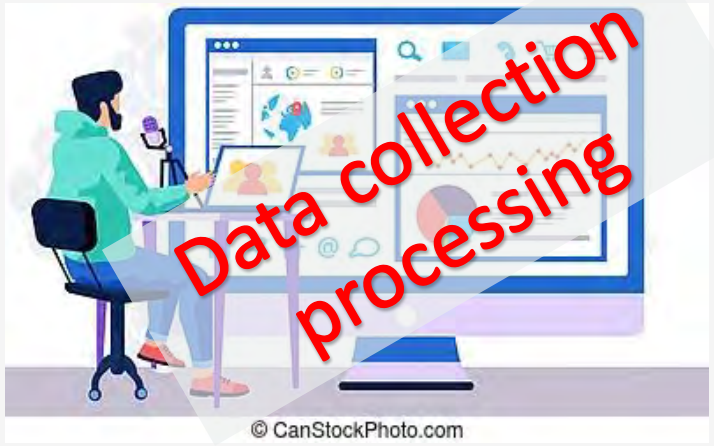
Hypothesis

Design

Experiment

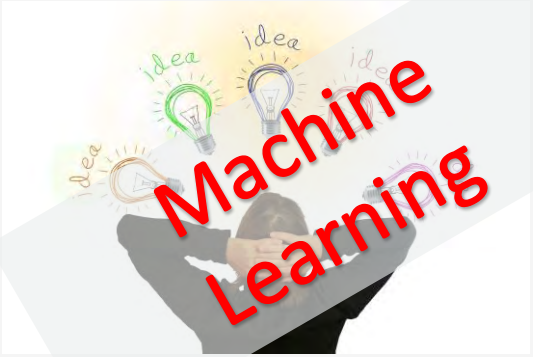
Record

Data collection processing



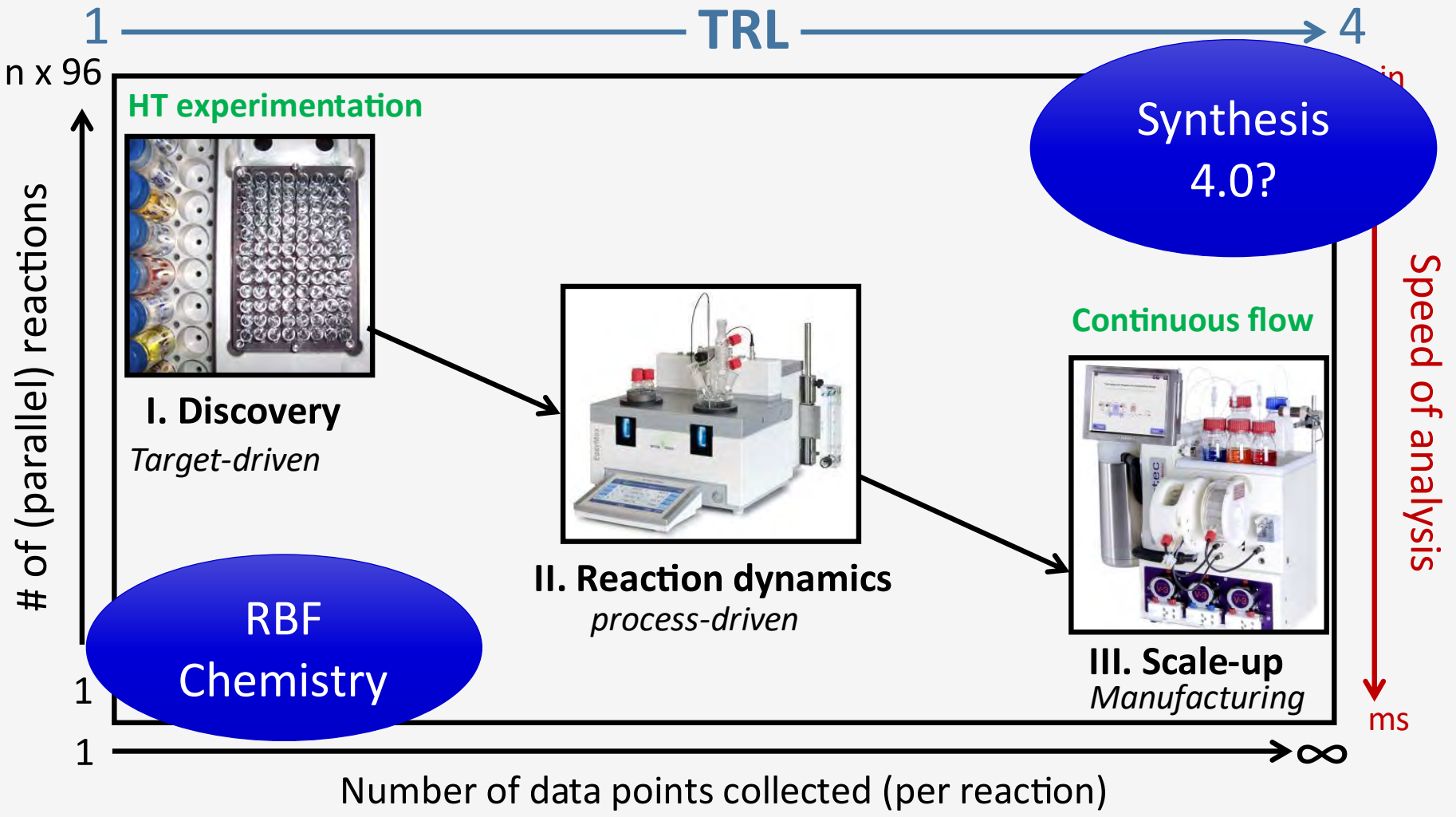
Analyse

Machine Learning



Understanding

Current state-of-art for ‘data-rich experimentation’



Matt Takle (left) and Linden Schrecker (right)

rawpixel.com

Collaborators:

Klaus Hellgardt (ICL, Chemical Engineering)

Christian Holze (BASF)

Joachim Dickhaut (BASF)

Andy Wieja (BASF)

BASF
We create chemistry

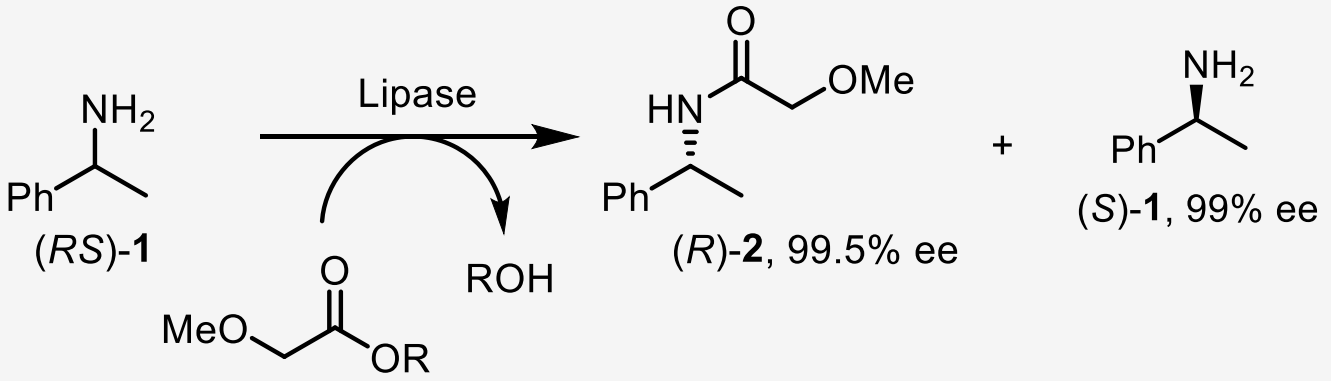
rEaCt
EPSRC Centre for Doctoral Training in Next Generation Synthesis & Reaction Technology

ROAR
Imperial College London

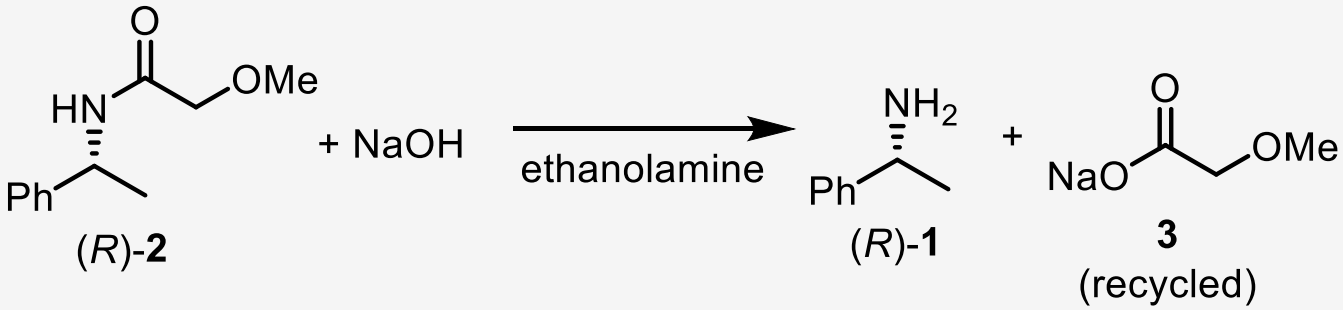
UKRI
UK Research and Innovation

Enzymatic Kinetic Resolution (ChiPros®)

1. lipase-catalyzed kinetic resolution



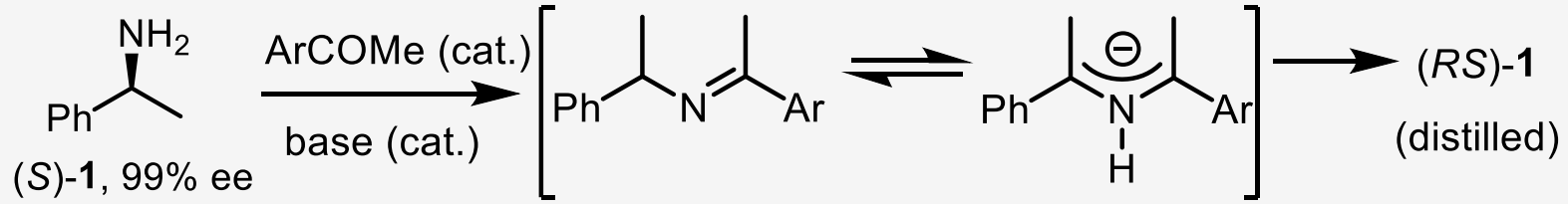
2. Recovery of *R*-enantiomer



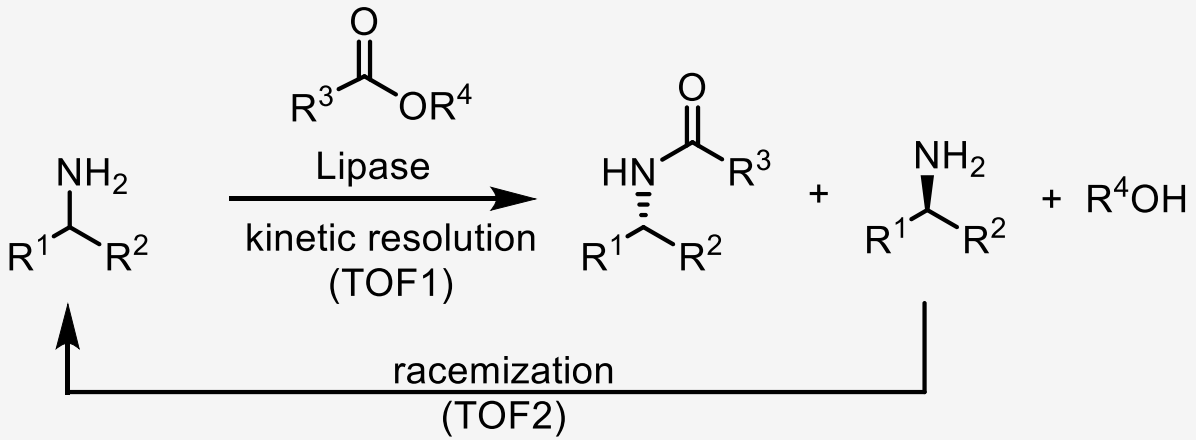
Hieber, G.; Ditrich, K. Introducing chipros: Biocatalytic production of chiral intermediates on a comme
Chim. Oggi **2001**, *19*, 16-20.

Recycling of (unwanted) enantiomer (ChiPros®)

3. Racemization of (unwanted) *S*-enantiomer

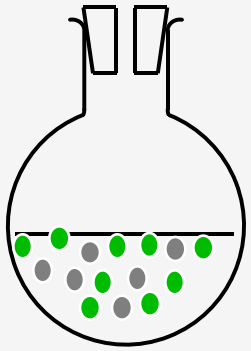


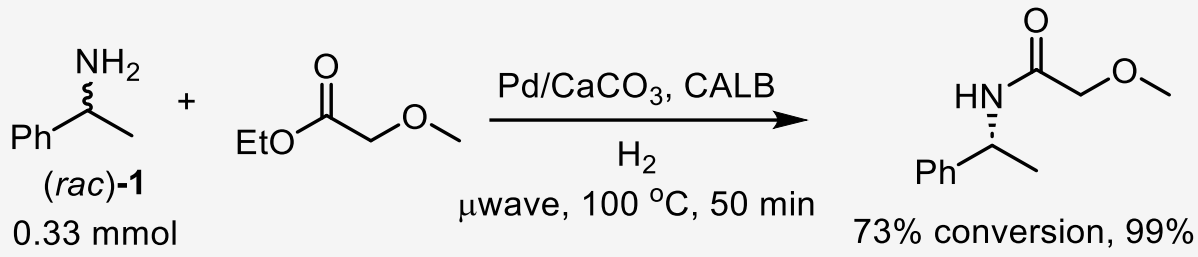
Alternative: chemoenzymatic kinetic resolution



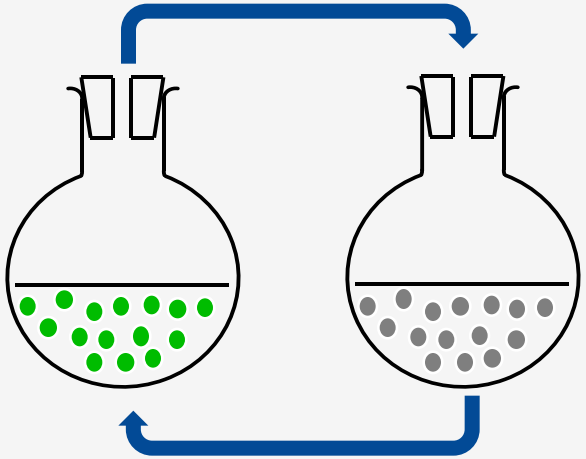
Hieber, G.; Ditrich, K. Introducing chipros: Biocatalytic production of chiral intermediates on a comm scale. *Chim. Oggi* **2001**, *19*, 16-20.

CE-DKR: One vs two ‘pots’





A. N. Parvulescu, E. Van der Eycken, P. A. Jacobs, D. E. De Vos, *J. Catal.* **2008**, *255*, 206.



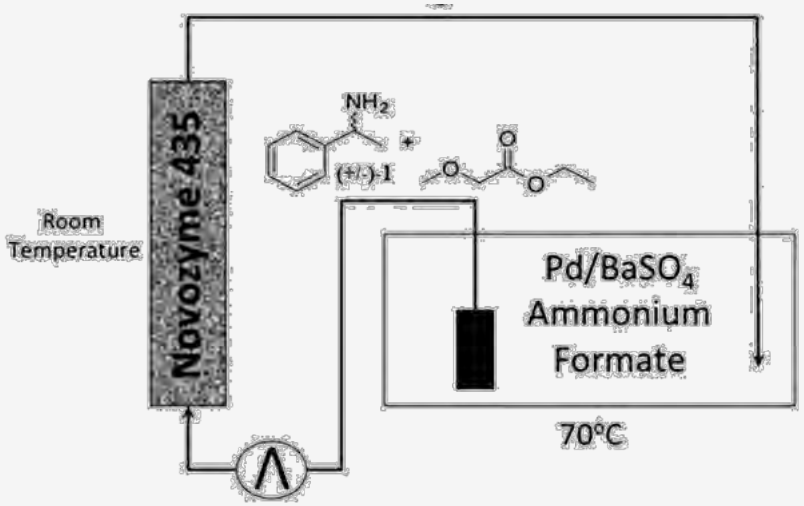


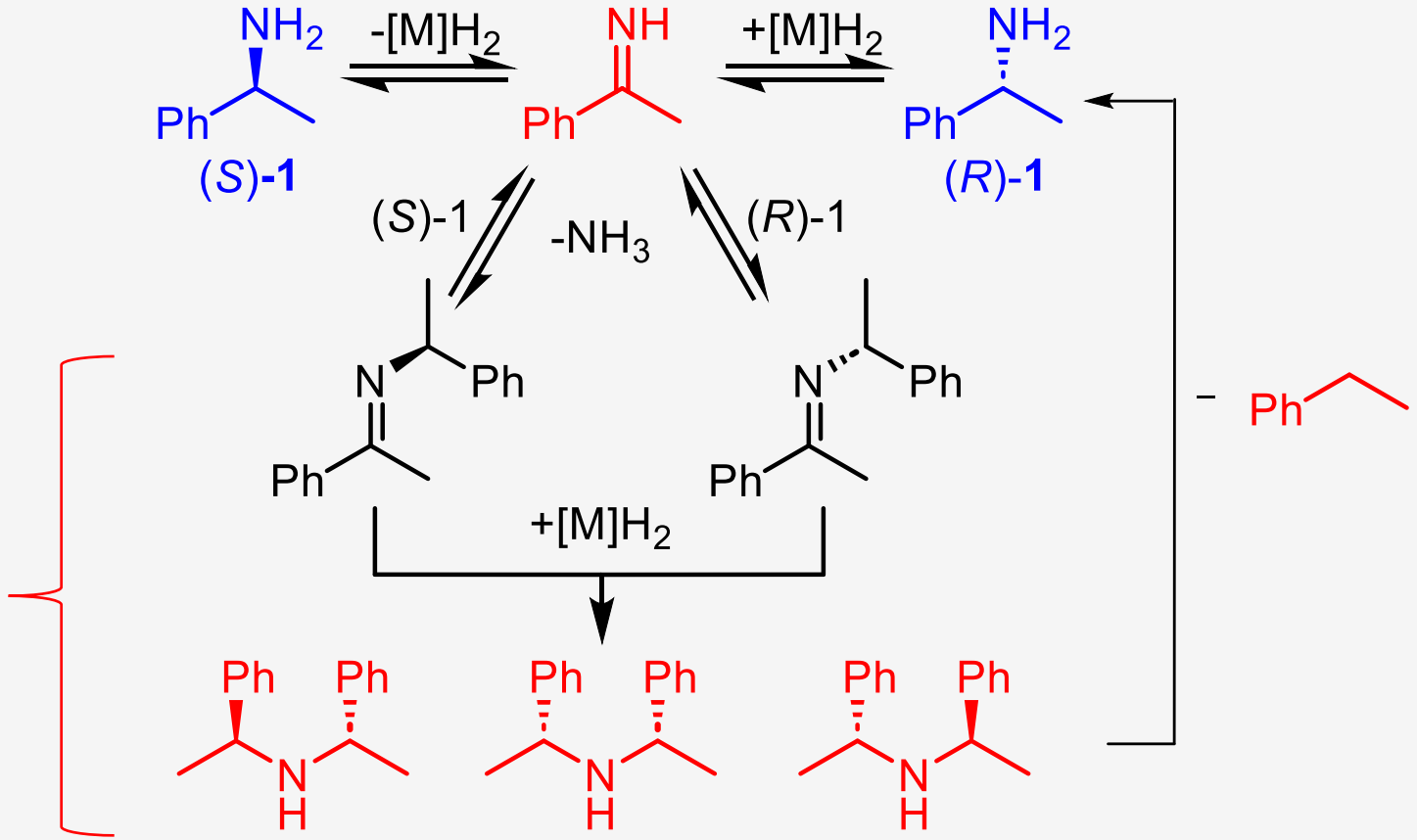
Table 5 Dynamic kinetic resolution of (+/–)-α-methylbenzylamine catalyzed by Novozyme 435, $Pd/BaSO_4$ and ammonium formate as green hydrogen source, under semi-continuous flow conditions

Entry	Reaction time (h)	2 (%)	ee (%)
1	2.5	53	99
2	4.5	64	97
3	7	68	96
4	8	76	96
5	10	77	95

de Miranda, A. S.; de Souza, R. O. M. A.; Miranda, L. S. M., *RSC Adv*. **2014**, *4*, 13620.

Catalyst deactivates thermally or poisoning by additives

Selectivity issues

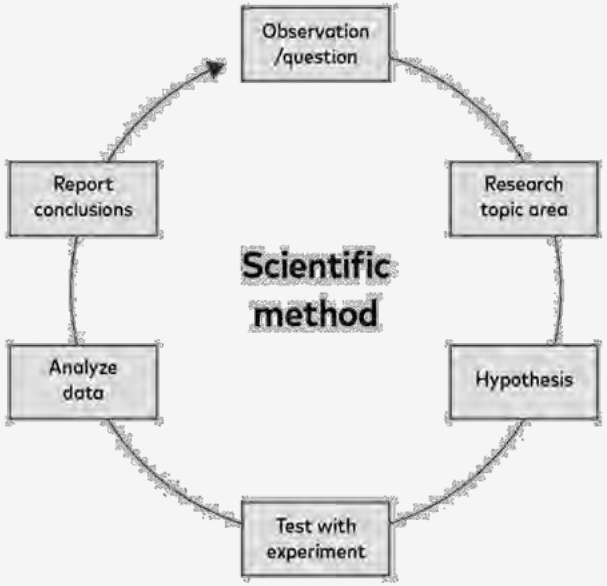


Suppressed by addition of H_2 or base

Research topic area: Scalable CE-DKR of chiral amines

Observation/question:

Lack of an effective amine racemization catalyst that is also compatible with enzyme-catalyzed kinetic resolution



Hypothesis:

- Compatibility issue: Compartmentalize the two catalysts
- Focus on amine racemization using heterogeneous catalysts
- Selectivity: Utilize residence time control to eradicate the need for additives

➡ **FLOW**

Development of the Flash Thermal Racemization (FTR) methodology: Catalyst screening

Image: ROAR

Parallel screening:
[M] catalysts, (H_2 pressure), T

	8		9			10				11
	Fe	Ru	Co	Rh	Ir	Ni	Pd	$Pd(OH)_2$	Pt	Cu	Au
Al_2O_3		89/86		100/96			51*/87		94/78		
TiO_2	100/100		100/100	76*	95/78	100/97	25*/63		73/40	100/100	100/100
C		72/57					39*/5	-			100/93

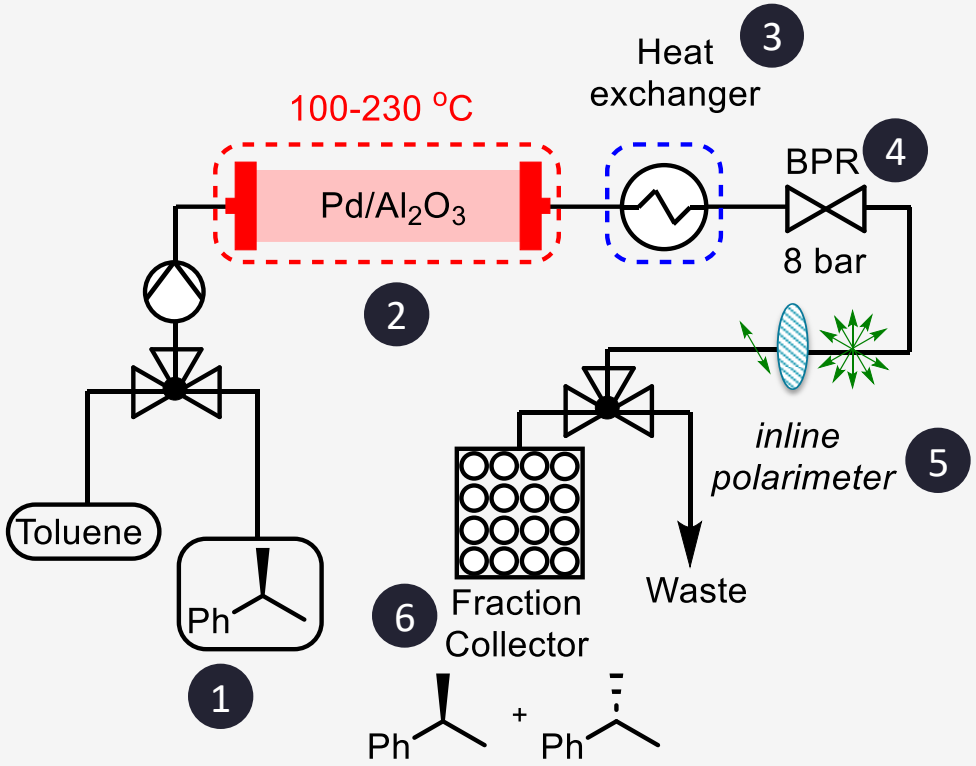
145 °C, H_2 (5 bar), 1 h. *100 °C, H_2 (5 bar), 1 h

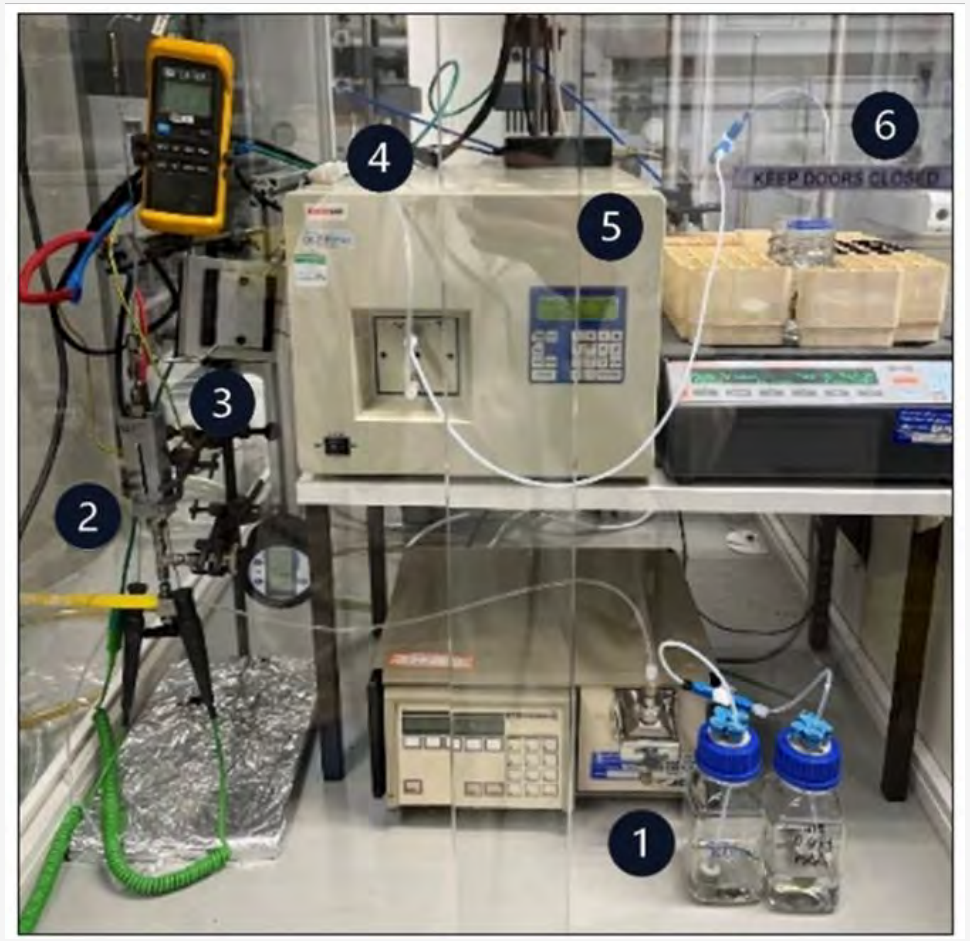
	8		9			10				11
	Fe	Ru	Co	Rh	Ir	Ni	Pd	$Pd(OH)_2$	Pt	Cu	Au
Al_2O_3		100/95		99/99			5/23		28/68		
TiO_2	100/100		100/100	90/95	100/92	100/95	7/15		100/76	100/100	100/94
C		100/98					1/9	31/38	14/30		100/90

145 °C, without H_2, 1 h

M. J. Takle, PhD thesis, 2023

Development of the Flash Thermal Racemization (FTR) methodology: DoE Optimization





Three inputs ('factors'):

Temperature (100-230 °C)

Flow rate (1-7 mL min^{-1})

Concentration (20.6-82.5 mM)

⟹

Two outputs ('responses'):

e.e. (chiral HPLC)

selectivity (GC)

M. J. Takle, B. J. Deadman, K. Hellgardt, J. Dickhaut, A. Weija, K. K. Hii, *ACS Catal.* **2023**, *13*, 10541.

DoE optimization (3 factors, 2 responses)

Initial custom design (18 experiments)

Entry	Whole Plots	Temperature °C	Flow rate mL min^{-1}	Concentration mM	e.e.[a] / %	Selectivity[b] %
1	1	100	1	20.6	100	94
2	1	165	4	20.6	10	48
3	1	230	1	20.6	15	6
4	2	100	7	20.6	100	94
5	2	165	4	20.6	7	54
6	2	230	7	20.6	8	40
7	3	100	1	51.6	100	96
8	3	165	4	51.6	10	65
9	4	230	4	51.6	3	25
10	4	165	7	51.6	37	77
11	5	100	7	82.5	100	97
12	5	230	1	82.5	8	16
13	6	100	1	82.5	100	98
14	6	230	7	82.5	5	12
15	7	165	4	51.6	8	76
16	7	230	7	51.6	3	56
17	8	100	4	51.6	100	97
18	8	165	1	51.6	9	44

Augmented Design (6 experiments)

19	9	145	6.4	51.6	23	70
20	9	135	5	51.6	28	73
21	9	110	5.8	51.6	100	88
22	9	150	2.4	51.6	8	61
23	9	125	1.7	51.6	19	72
24	9	115	3.2	51.6	90	79

Validation (3 experiments)

25	10	125	5.5	51.6	90 (70)	100 (88)
26	10	165	4	51.6	5 (8)	65 (63)
27	10	200	2	51.6	<5 (0)	26 (36)

M. J. Takle, B. J. Deadman, K. Hellgardt, J. Dickhaut, A. Weija, K. K. Hii, *ACS Catal.* **2023**, *13*, 10541.

DoE optimization

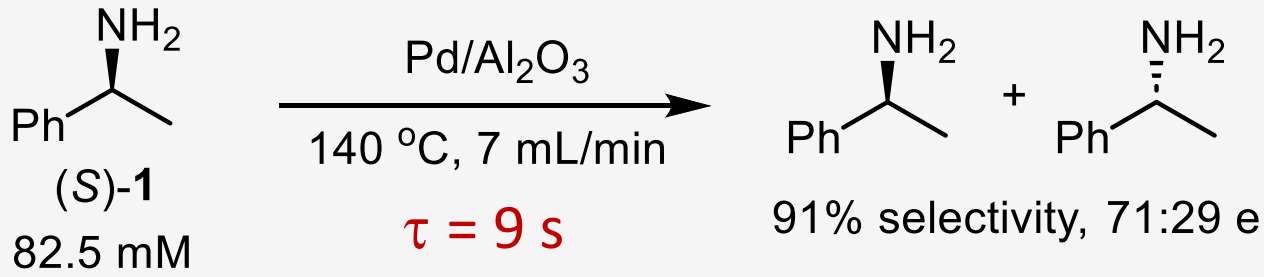
Validation (3 experiments)

25	10	125	5.5	51.6	90 (70)	100 (88)
26	10	165	4	51.6	5 (8)	65 (63)
27	10	200	2	51.6	<5 (0)	26 (36)

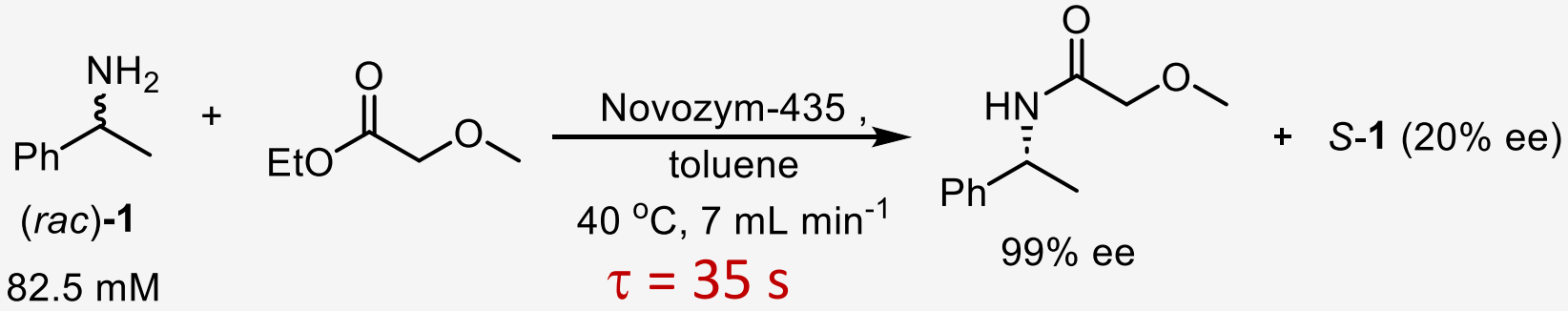
$$(1)\ e.e._{S-(1)}(\%) = 7.82 - 54.8\left(\frac{T-165}{65}\right) + 9.23\left(\frac{F-4}{3}\right) + 55.7\left(\frac{T-165}{65}\right)^2 - 13.2\left(\frac{T-165}{65}\right)\left(\frac{F-4}{3}\right)$$

$$(2)\ Selectivity_{S-(1)}(\%) = 61.3 - 40.0\left(\frac{T-165}{65}\right) + 10.7\left(\frac{F-4}{3}\right) + 3.03\left(\frac{S.C-0.0516}{0.0310}\right) - 11.0\left(\frac{T-165}{65}\right)\left(\frac{S.C-0.0516}{0.0310}\right)$$

Predicted optimized conditions

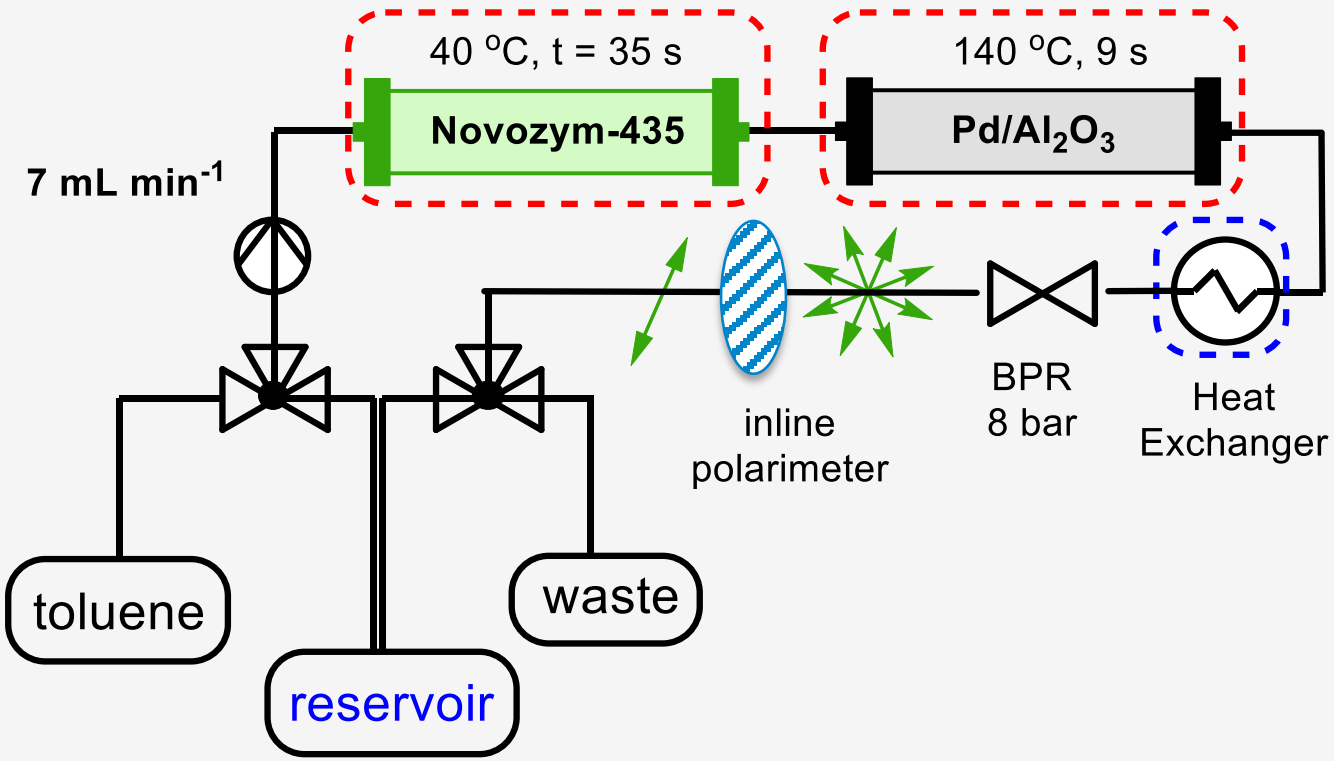


Kinetic resolution:



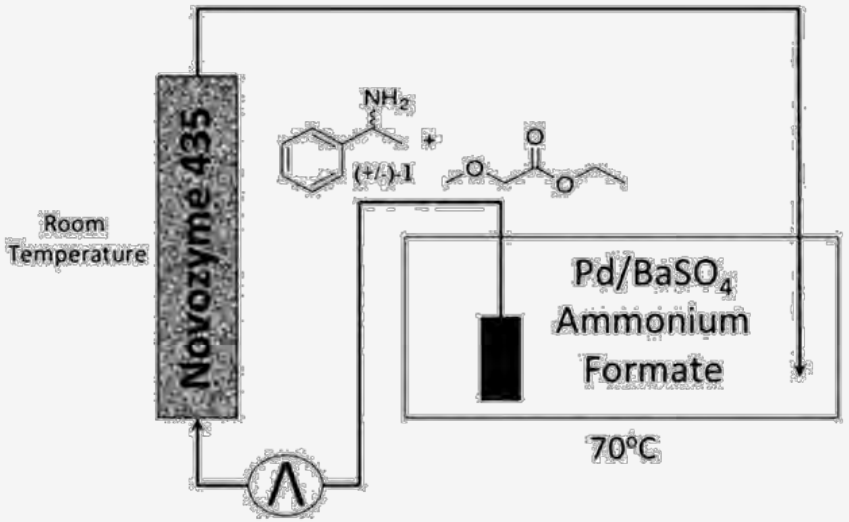
M. J. Takle, B. J. Deadman, K. Hellgardt, J. Dickhaut, A. Weija, K. K. Hii, *ACS Catal.* **2023**, *13*, 10541.

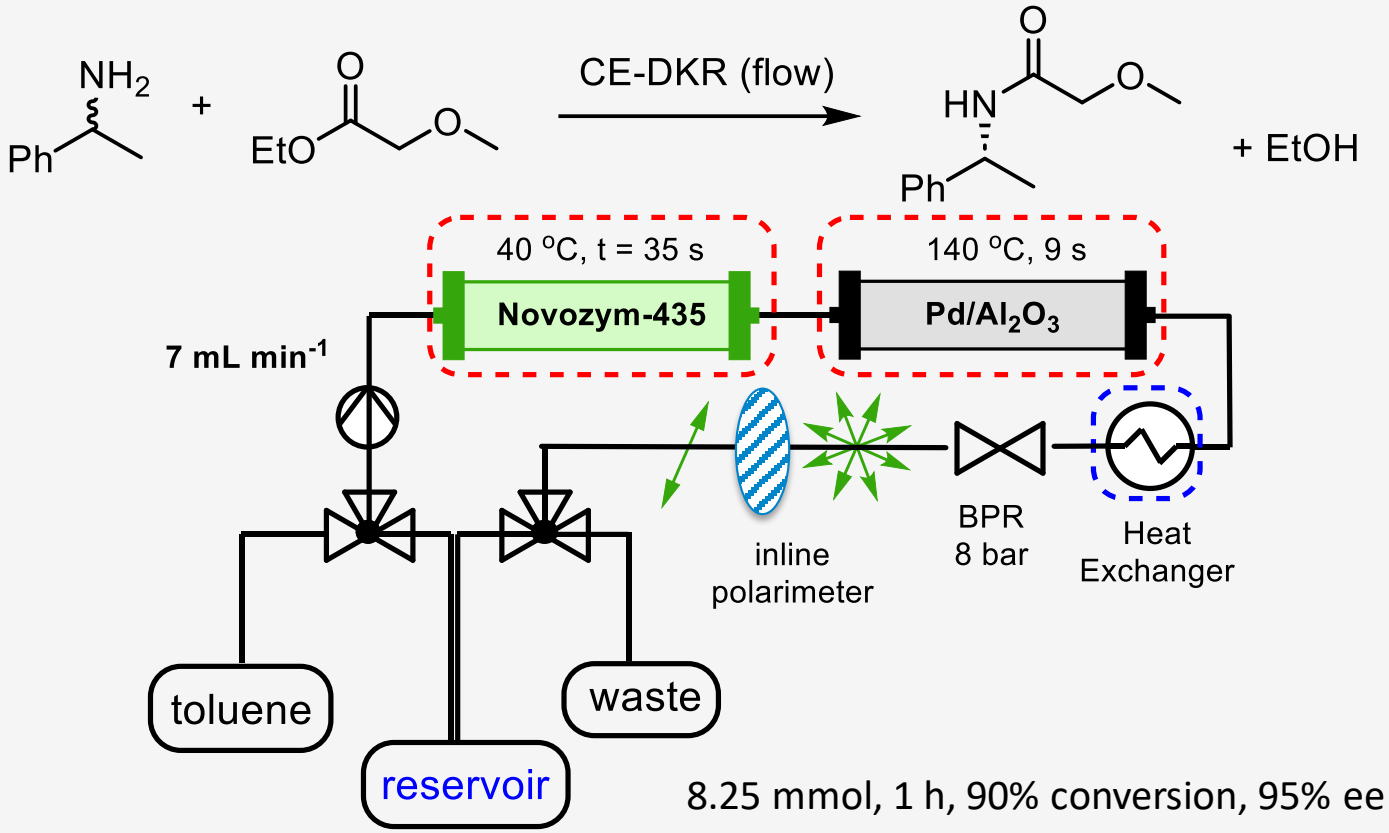
CE-DKR-FTR Reactor Design



M. J. Takle, B. J. Deadman, K. Hellgardt, J. Dickhaut, A. Weija, K. K. Hii, *ACS Catal.* **2023**, *13*, 10541.

Comparison (batch-flow): 1-Phenylethylamine





3.2 mmol scale, 10 h, 77% conversion, 95% ee

de Miranda *et al*, *RSC Adv*. **2014**, *4*, 13620-13625.

8.25 mmol, 1 h, 90% conversion, 95% ee

Takle et al, *ACS Catal.* **2023**, *13*, 10541.

STY = 14.5 µmol mL^{-1} h^{-1}

0.030 µmol µmol(Pd)$^{-1}$ mL^{-1} h^{-1}

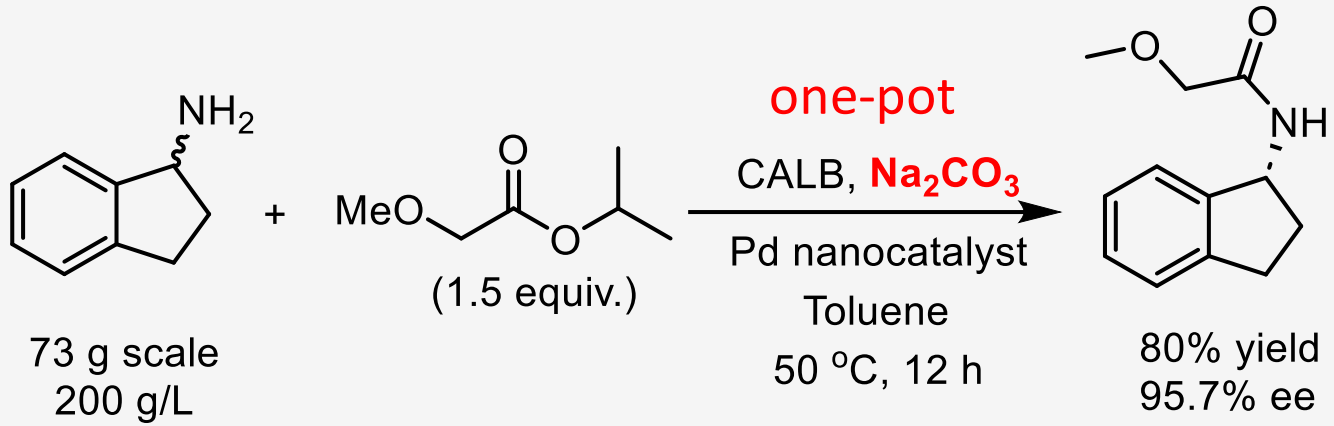
STY = 1465 µmol mL^{-1} h^{-1}

15.5 µmol µmol(Pd)$^{-1}$ mL^{-1} h^{-1}

>500x more productive

M. J. Takle, B. J. Deadman, K. Hellgardt, J. Dickhaut, A. Weija, K. K. Hii, *ACS Catal.* **2023**, *13*, 10541.

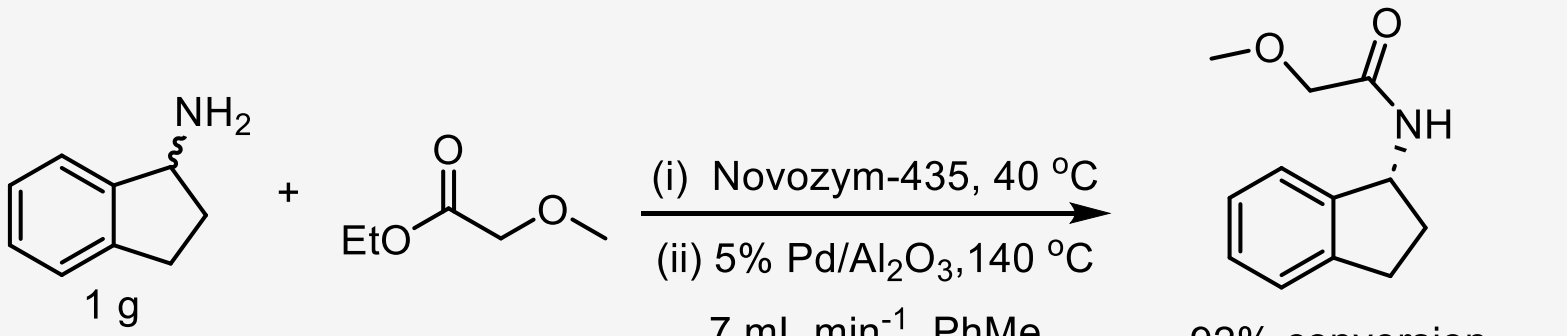
Comparison (batch-flow) Rasagiline (treatment of Parkinson's Disease)



73 g scale
200 g/L

STY = 109.6 µmol mL^{-1} h^{-1}
0.0008 µmol µmol$(Pd)^{-1}$ mL^{-1} h^{-1}

Ma, G.; Xu, Z.; Zhang, P.; Liu, J.; Hao, X.; Ouyang, J.; Liang, P.; You, S.; Jia, X., *Org. Process Res. Dev*. **2014**, *18*, 1169-1174



1 g

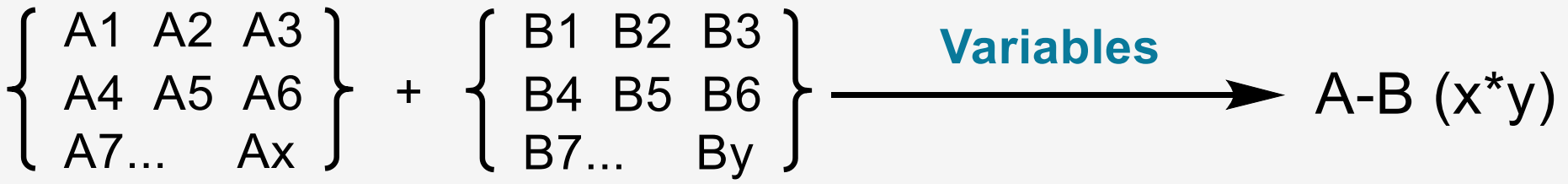
This work:

STY = 1362 µmol mL^{-1} h^{-1}
14.5 µmol µmol$(Pd)^{-1}$ mL^{-1} h^{-1}
>18,000 times more productive

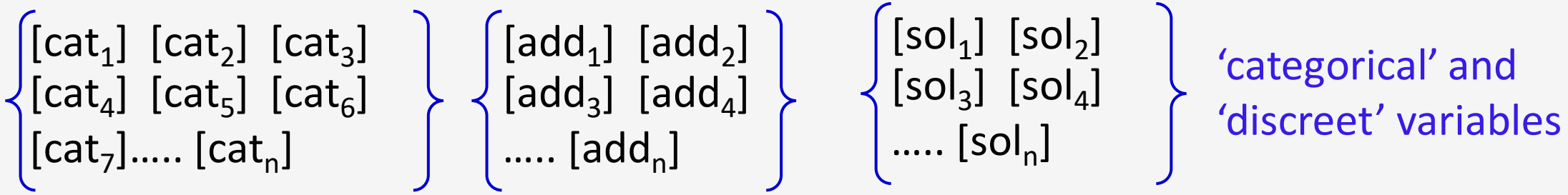
Bonus: No base/H-donor needed

M. J. Takle, B. J. Deadman, K. Hellgardt, J. Dickhaut, A. Weija, K. K. Hii, *ACS Catal.* **2023**, *13*, 10541.

The ‘data-driven’ approach



‘categorical’ variables

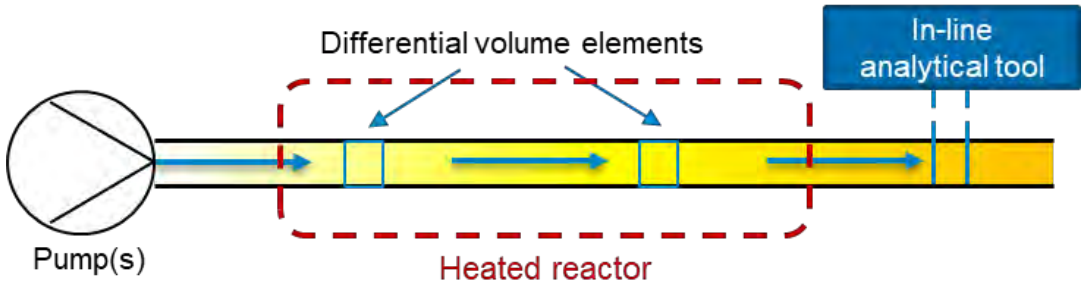


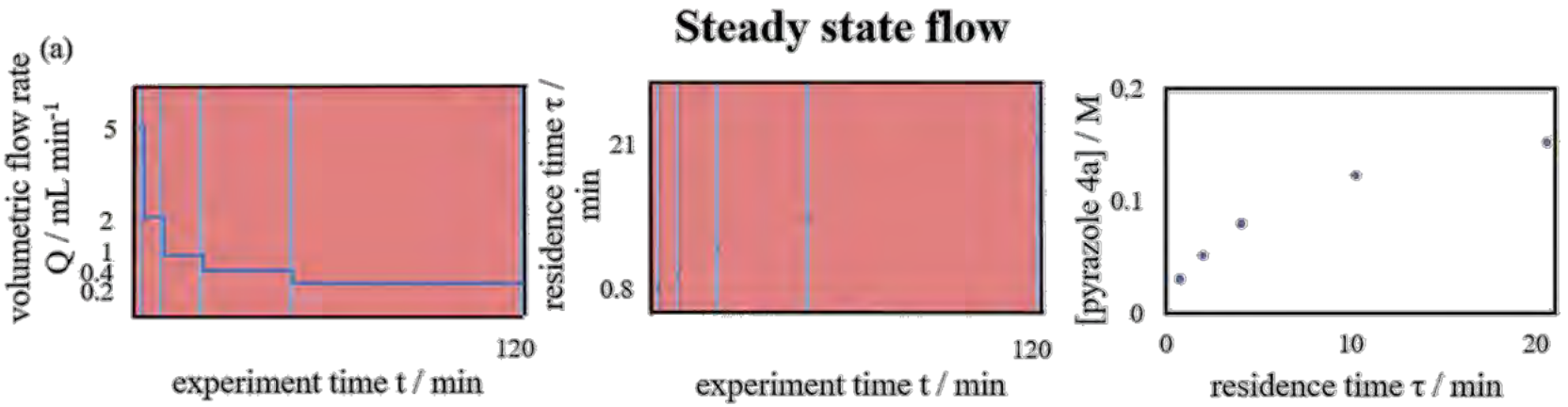
‘categorical’ and
‘discreet’ variables

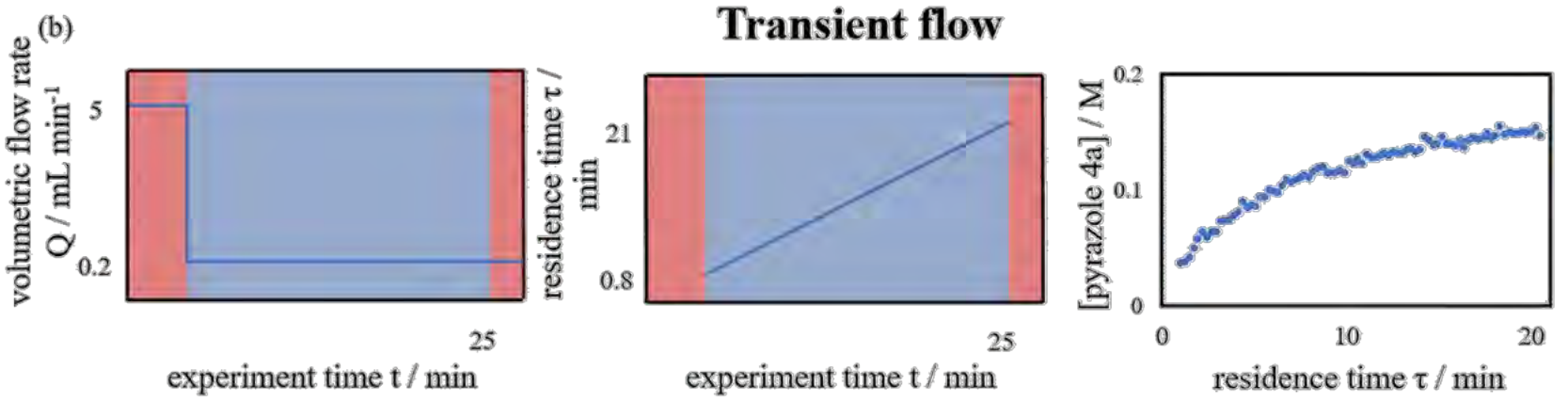
Temperature, pressure, time: ‘continuous’ variables.

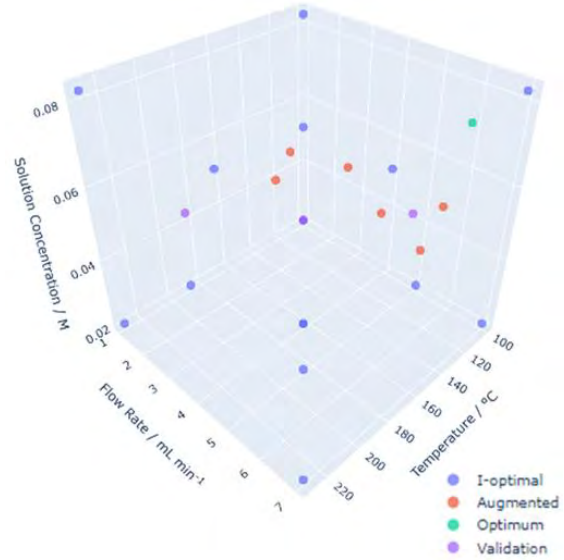
To understand and quantify the relationships between the variables ➡ **KNOWLEDGE**

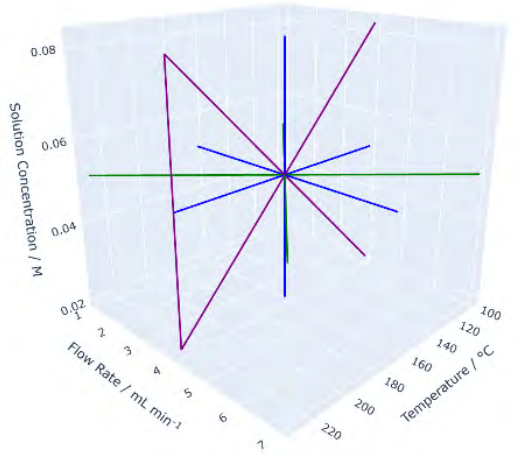
Steady state vs transient flow





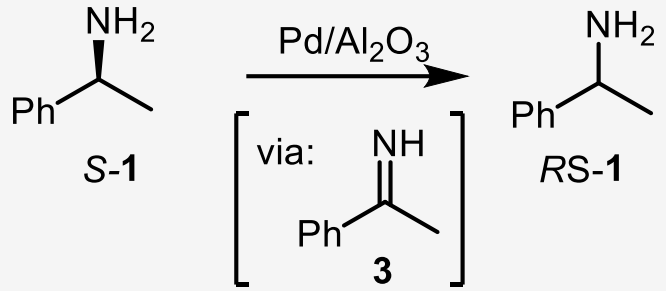




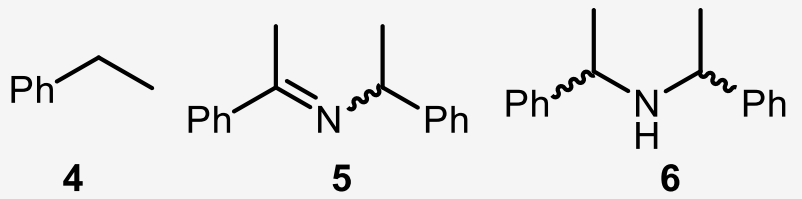


L. Schrecker, J. Dickhaut, C. Holtze, P. Staehle, M. Vranceanu, K. Hellgardt, K. K. Hii, *React. Chem. Eng.* **2023**, *8*, 41.

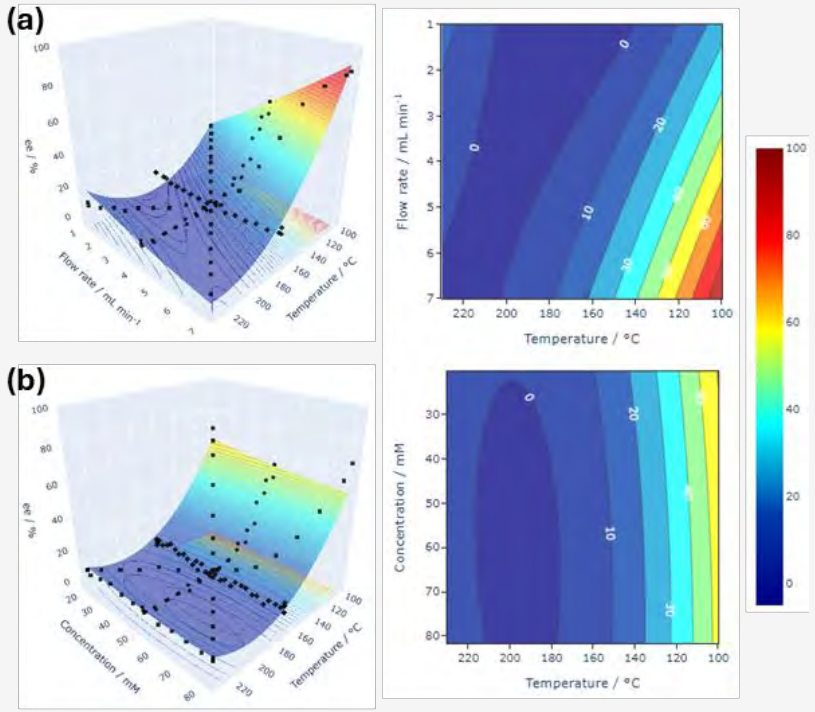
Development of the Flash Thermal Racemization (FTR) methodology: Interdependent reaction parameters



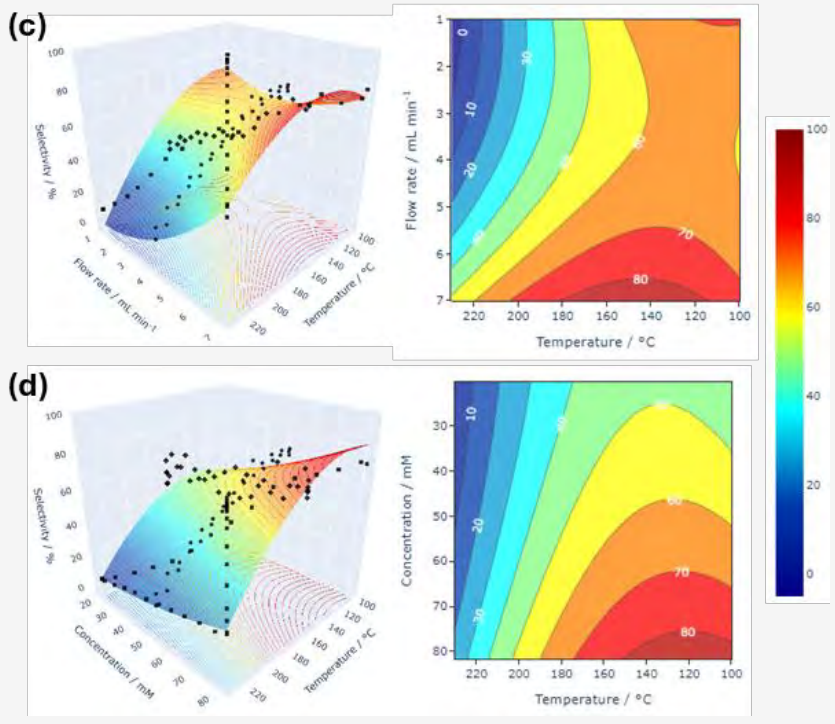
side products:



e.e. / %

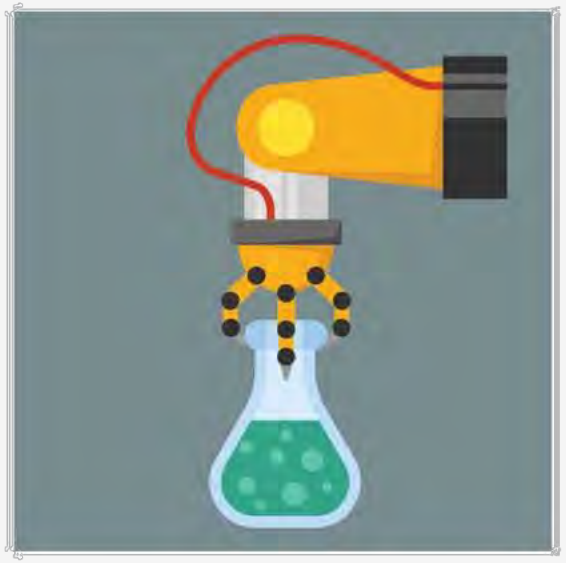


selectivity / %



SUMMARY

- The scientific method depends on experimentation fn observation.
- Chemistry needs to evolve and become much more data-led.
- Generation of accurate predictive models (‘AI’) can only be achieved with good quality data.
- Lab of the future: Fusion of automated ‘autonomous’ hardware and software: ELN, LIMS, connectivity.
- To understand complexity, chemists need to understand statistical tools.



Extra slides

Will robots kill chemistry?

Image: Shutterstock

https://cen.acs.org/careers/employment/robots-kill-chemistry/97/i15

Transition from target-led to data-led synthesis

In the lab:

- Understand limitations of RBF
- Translate RBF → automated platforms
- Qualitative → quantitative
- Utilisation of reaction technologies (μwave, flow, electro/photo/mechano-chem)

At the computer:

- How to handle/interpret large amount & different types of data
- Planning experiments (ELN, software)
- Reaction kinetics, modelling
- Statistical analysis/ML algorithms

Learning new skills

Change in mindset

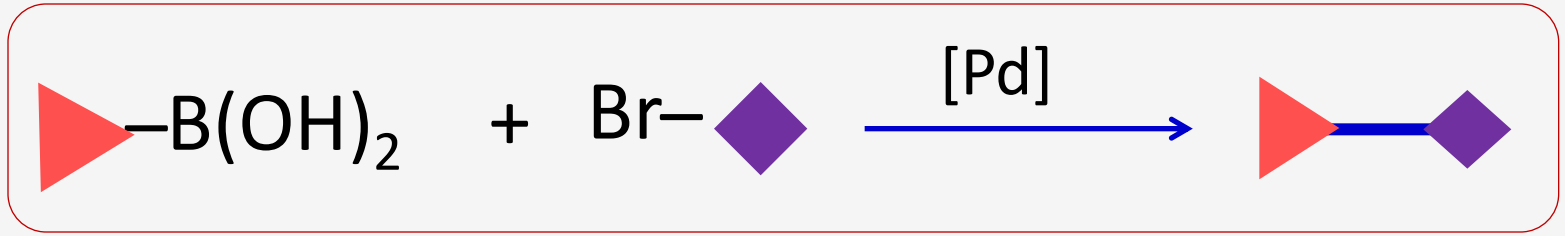
Holistic and agile: finding the right solution as demanded by the chemistry.

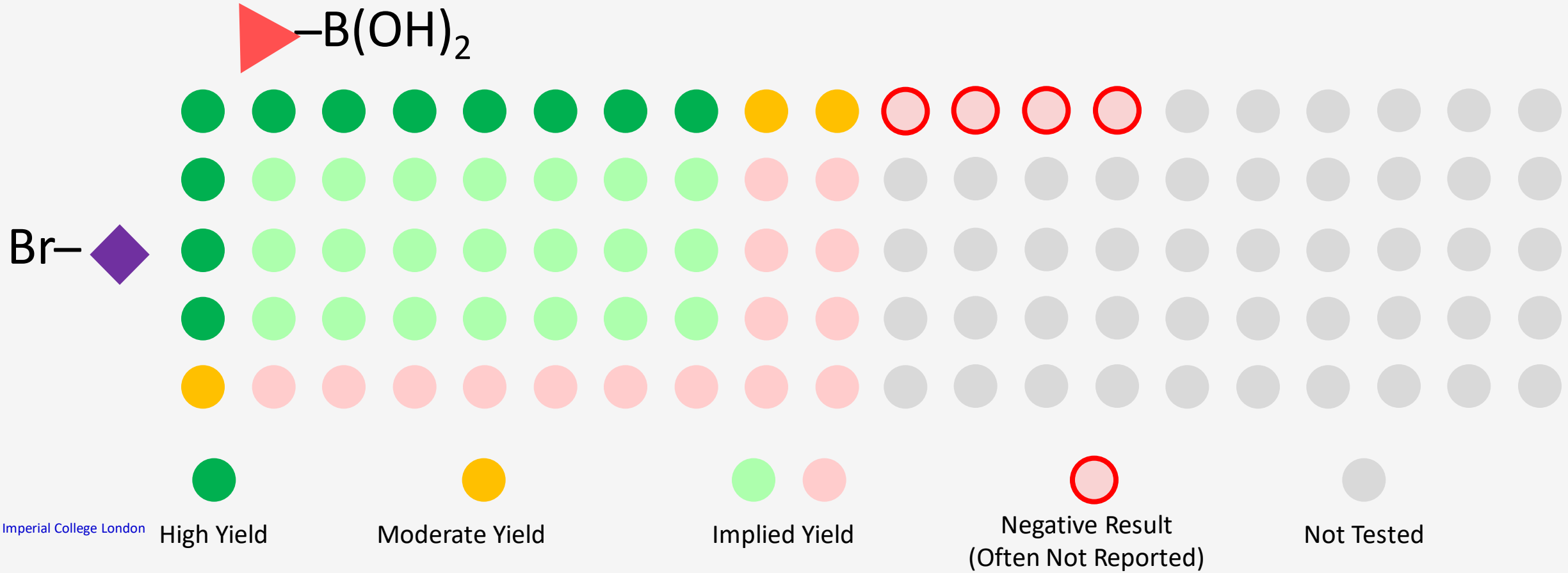
Digital Chemistry: Challenges

Infrastructure

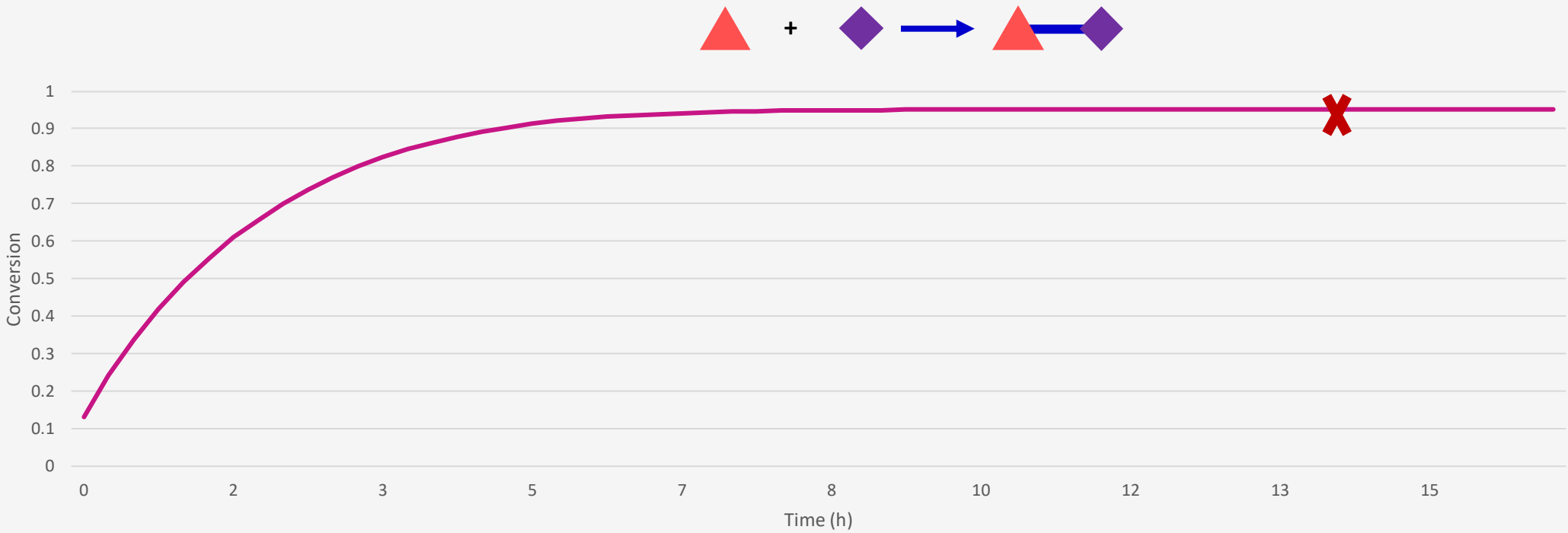
Training/expertise

Problems with existing data:
Exhibit #1: Limited substrate scope

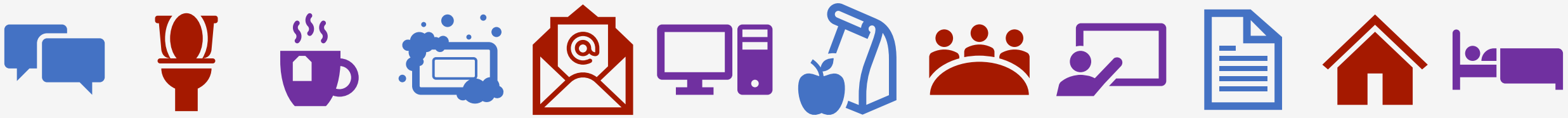




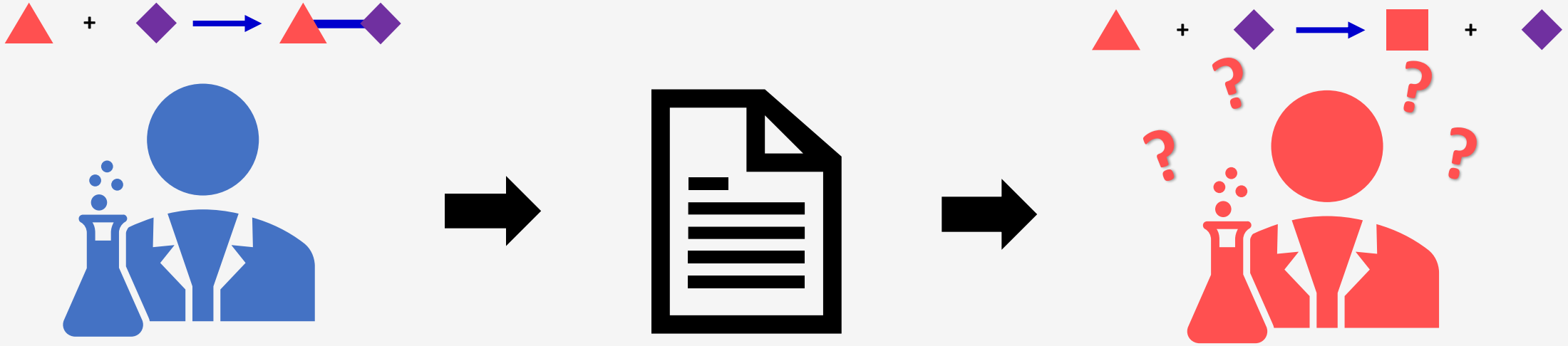
Problems with existing data:
Exhibit #2: Single timepoints results



Reported reaction time often influenced by other activities



Problems with existing data:
Exhibit #3: Method interpretation



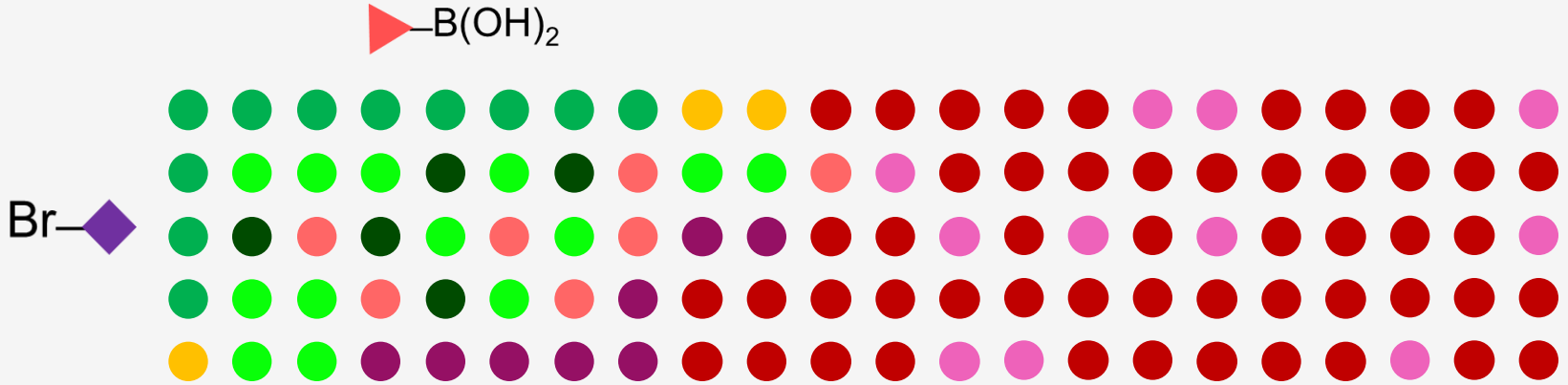
Example Procedure

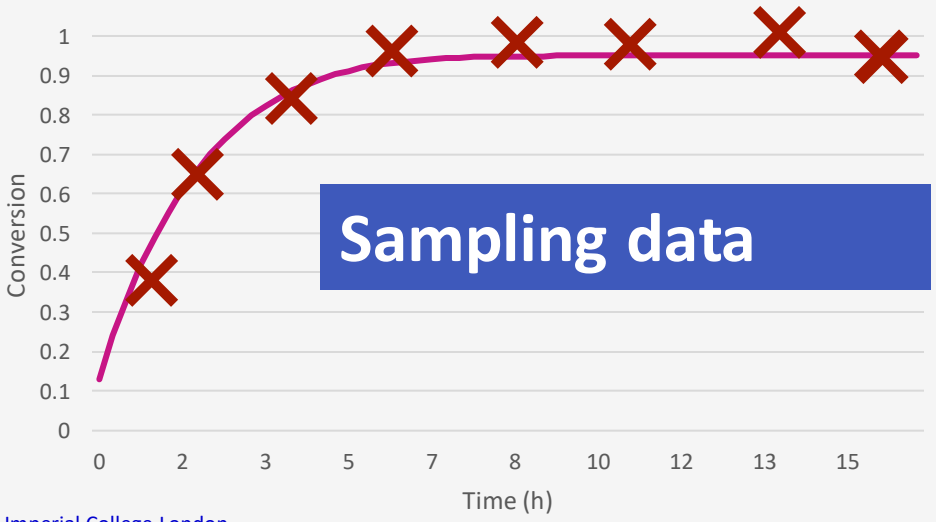
A dry 100 mL round bottom flask was purged with argon. The appropriate indole or NHheterocycle (10 mmol) was added to the round bottom with CuI (190 mg, 10 mol%) and K_3PO_4 (5.5 g, 2.6 equiv). Then 2-bromopyridine (1.8 mL, 1.5 equiv) and DMEDA (0.22 mL, 20 mol%) was added in toluene (12.5 mL). The reaction was stirred at 110 °C for 24 h. The reaction was cooled to room temperature and filtered over a silica pad. The product was isolated by flash column chromatography, typically 20% -> 60% DCM/pentane w/ 1% TEA. In instances where N-pyridyl indole was difficult to separate from remaining bromopyridine the product was triturated or recrystallized from pentane or the reaction was run with 1.05-1.5 equivalents of indole relative to bromopyridine.

J. Am. Chem. Soc., **2020**, DOI: 10.1021/jacs.0c03246

How Can We Improve Reaction Data?

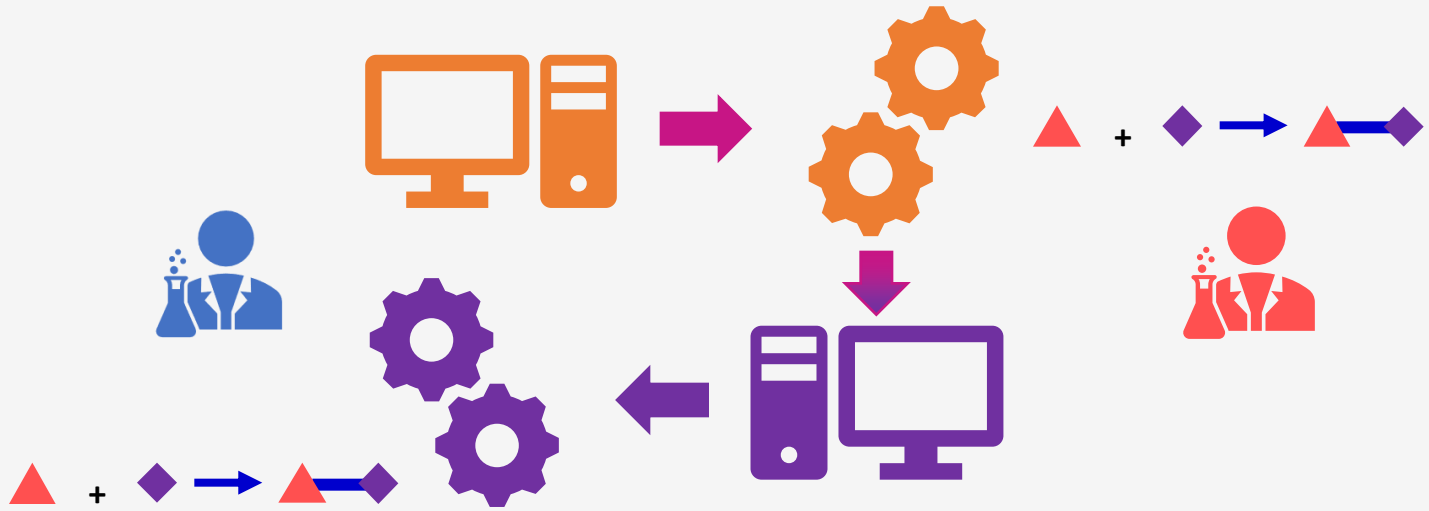
Comprehensive substrate scope



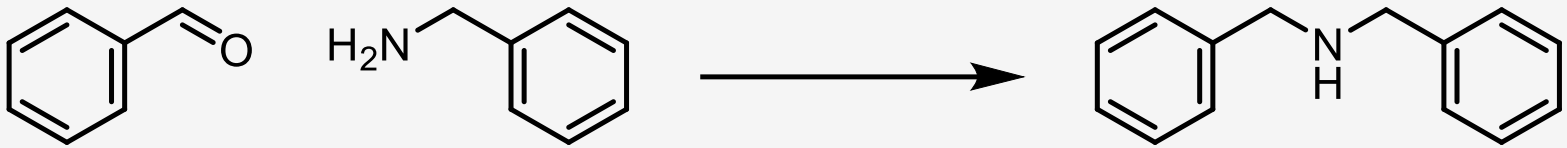


Sampling data

Interoperable procedures



Case study: Reductive amination reactions



Option 1: (i) $NaBH(OAc)_3$, AcOH; (ii) aqeuous workup.

Option 2: (i) H_2, $[cat]_{het}$; (ii) (filter,) evaporate.

Literature (heterogeneous catalysis), chronological order:

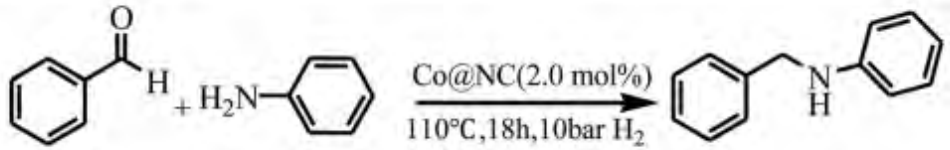
1. Pd/C, MeOH-$CHCl_3$, 'room temperature and pressure', 2 h, 96%: Xin *et al*, *Tetrahedron* **2008**, *64*, 11783.
2. Rh NP, EtOH, 30 bar, 80 °C, 12 h, 99%, Huang *et al*, *RSC Adv*. **2015**, *5*, 56936.
3. Co/N-doped C, toluene, 10 bar, rt →110 °C, 18 h, 99%, Mao et al, *RSC Adv*. **2016**, *6*, 94068.
4. Rh NP, glycerol/heptane, 40 bar, 80 °C, 3 h, 98% (3% imine), Serrano-Maldonado *et al*, *Eur. J. Inorg. Chem.* **2020**, 2506.

Table 1 Screening of t

Entry	Catalyst
1	Rh@CN
2	Rh@CN
3	Rh@CN
4	**Rh@CN**
5	Rh@CN
6	Rh@CN
7	Rh/C
8	Pd/C
9	Ru@CN
10	Rh@CN

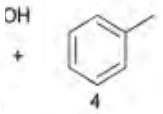
[1] Rh NP, EtOH, 3(

Table 1 Co@NC-catalyzed reductive amination of benzaldehyde and aniline ove Co catalysts[a]



Entry	Catalyst	Conversion[b] (%)	Yield[b] (%)
1	Co@NC (600-2 h)	79	75
2	Co@NC (700-2 h)	88	84
3	Co@NC (800-2 h)	100	98
4	Co@NC (800-3 h)	98	96
5	Co/C (800-2 h)	7	6
6	IL/C (800-2 h)	Trace	0
7	Co–IL/C	10	9
8[c]	Co@NC (800-2 h)	96	0
9[d]	Pd/C	100	0[e]

[a] Reaction conditions: 1.0 mmol aniline; 1.5 mmol benzaldehyde; Co 2.0 mol% added; toluene (2.0 mL); at 110 °C; 10 bar H_2; 18 h. [b] Determined by GC. [c] Without H_2. [d] 3.0 mol% Pd (Pd/C) was added, H_2 balloon. [e] Toluene was detected as product from benzaldehyde. In case of lower yields, the imine was detected as the by-product. IL = [MCNIm]Cl, C = active carbon.

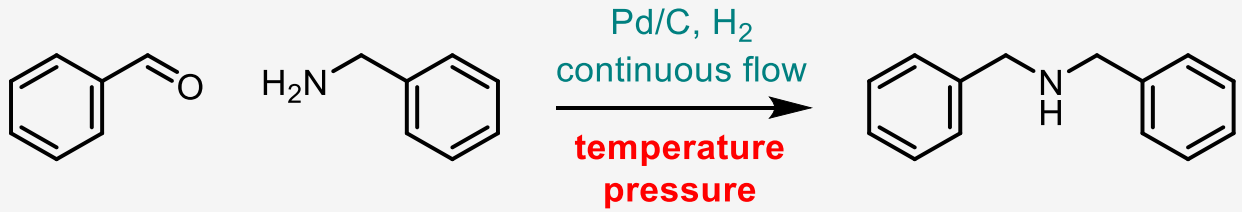


Selectivity/%[b]			
1	2	3	4
89	11	0	0
[illegible]	[illegible]	[illegible]	[illegible]
95	5	0	0
[illegible]	[illegible]	[illegible]	[illegible]
0	0	0	0
[illegible]	[illegible]	[illegible]	[illegible]
58	42	0	0
0	82		18
40	60	0	0
75	25	0	0

"With commercial Pd/C catalyst, benzaldehyde was reduced to toluene directly, and no N-benzylaniline was detected (Table 1, entry 9)."

[2] Co/N-doped C, toluene, 10 bar, rt →110 °C, 18 h Mao et al, *RSC Adv*. **2016**, *6*, 94068

Automated workflow (Batch reactions)



Reaction conditions: benzaldehyde 4.5 mmol (0.3 M), amine 9 mmol (0.6 M), 50 mg Pd/C, 15 mL methanol, **10 bar H_2**.

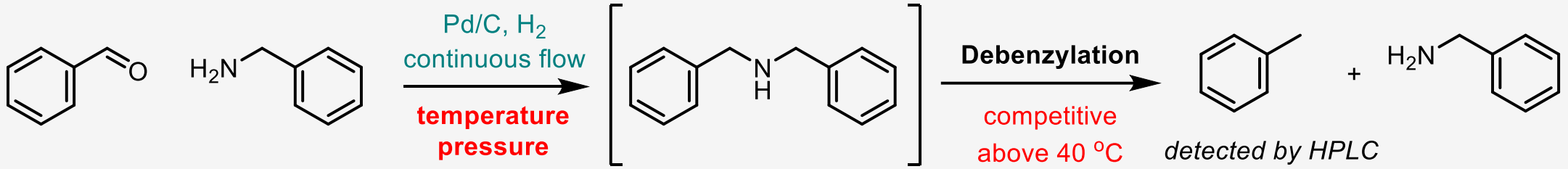
Optimization Sampling Reactor (OSR)

- 8 // reactors
- Independent T &P control
- Gas uptake measurement
- Autosampling under pressure

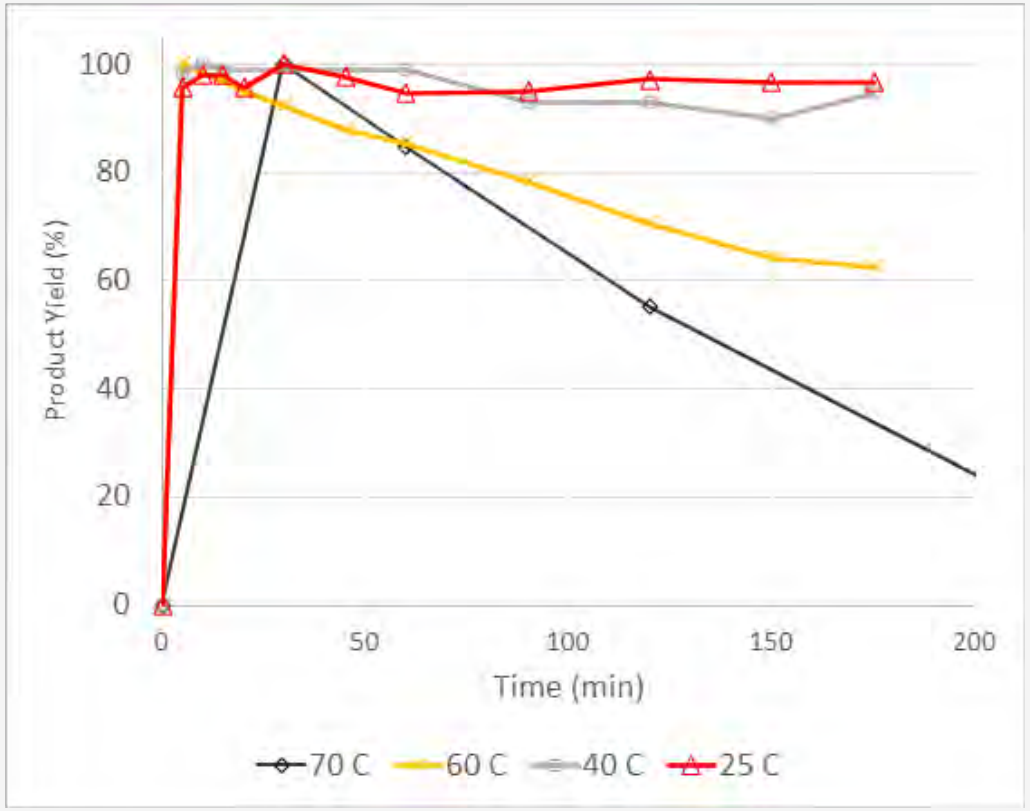
4 temperatures: 25, 40, 60, 70 °C

Reactions monitored (GC) over 3 h

Paola Ferrini, ROAR, *unpublished results.*



Reaction conditions: benzaldehyde 4.5 mmol (0.3 M), amine 9 mmol (0.6 M), 50 mg Pd/C, 15 mL methanol, **10 bar H_2**.



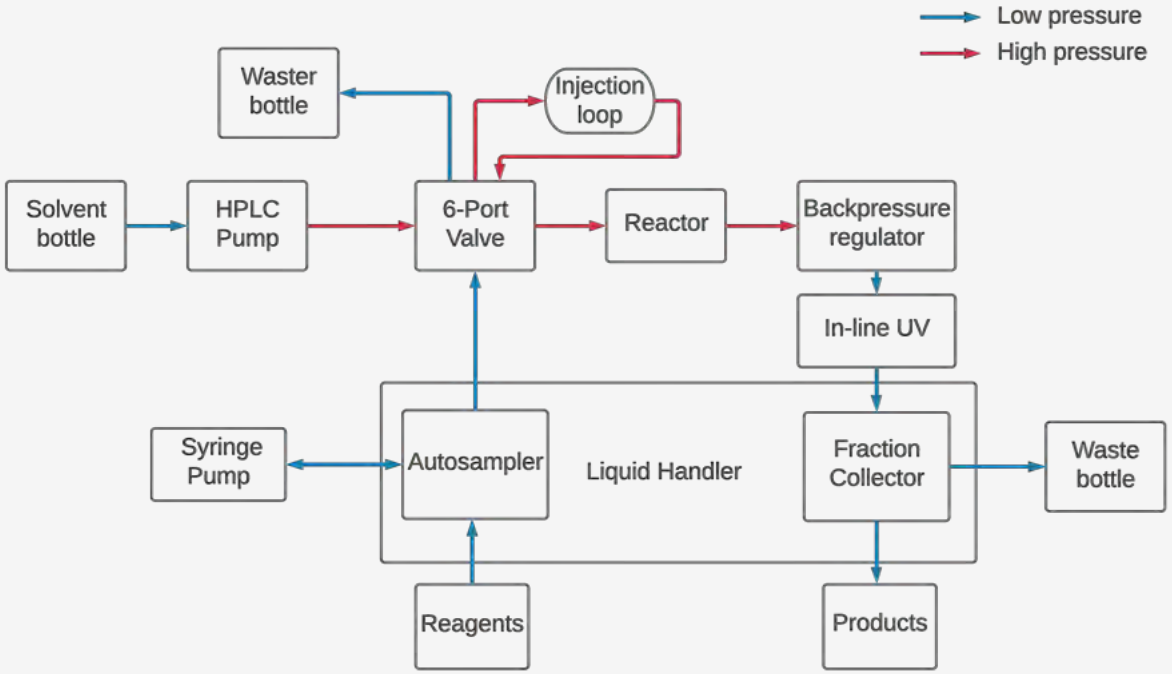
Conclusions:

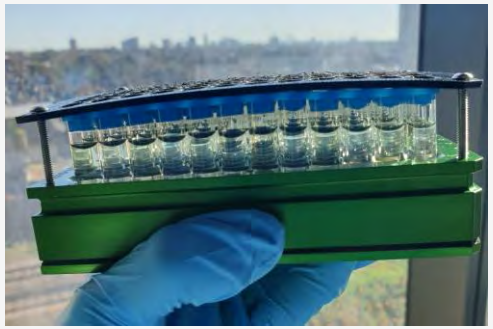
1. Pd/C is far more active than Rh or Co;
2. Toluene is generated from the product, <u>not</u> the precursor.

Improve selectivity in flow (τ control)

Paola Ferrini, ROAR, *unpublished results*.

Automated HT workflow (in flow)



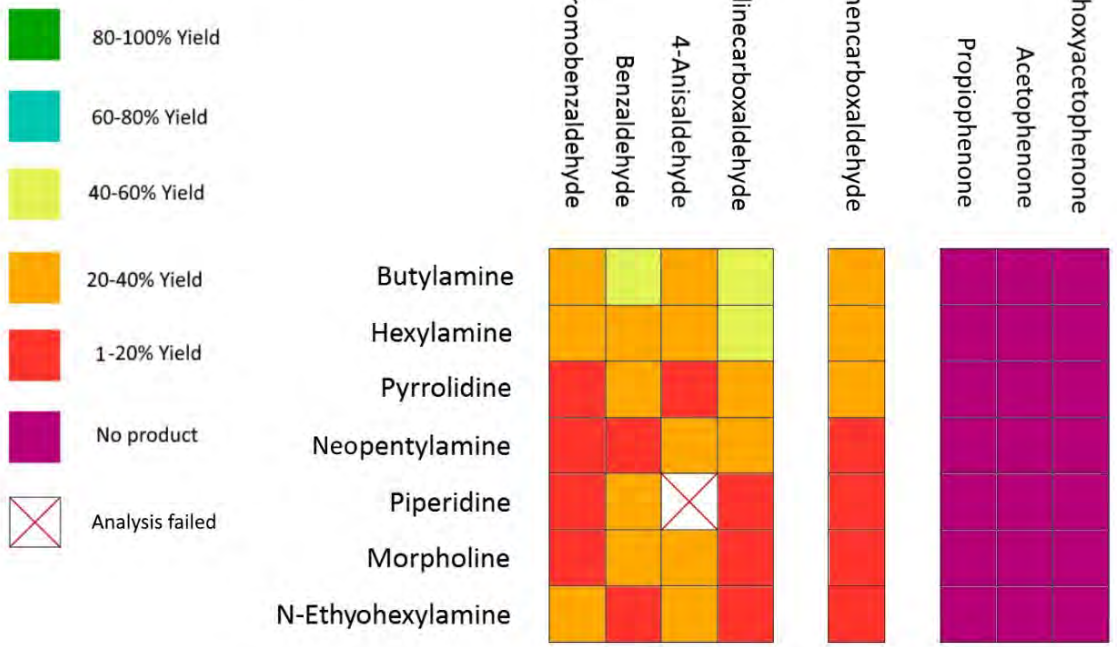


Imperial College London

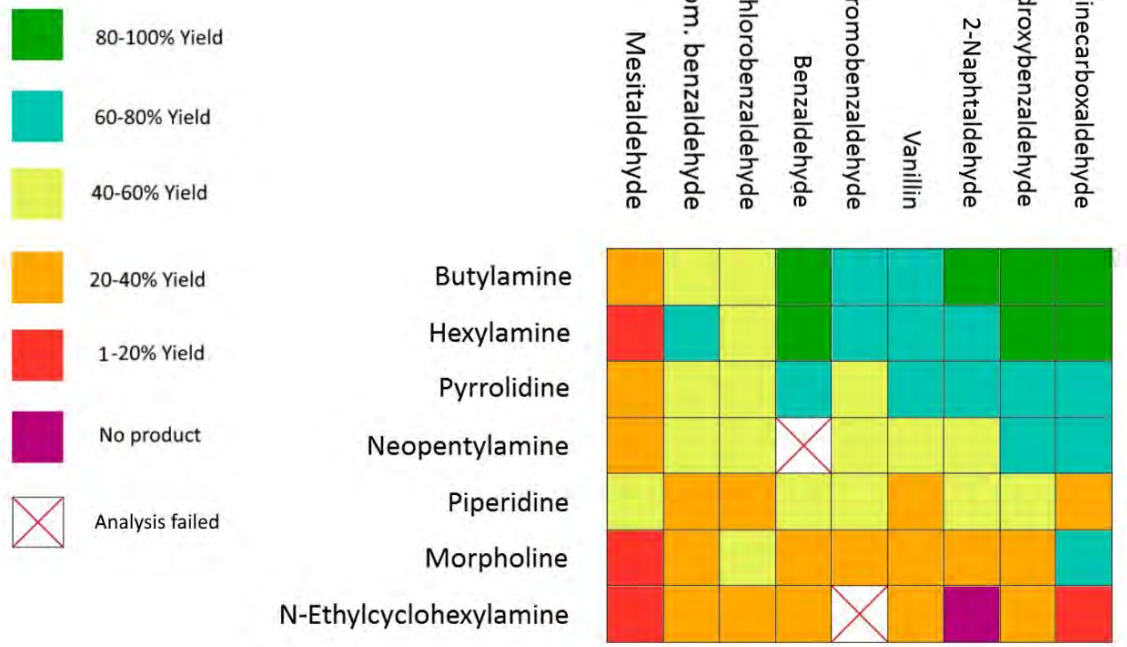
Marvin Alberts, MSci thesis

Representative results

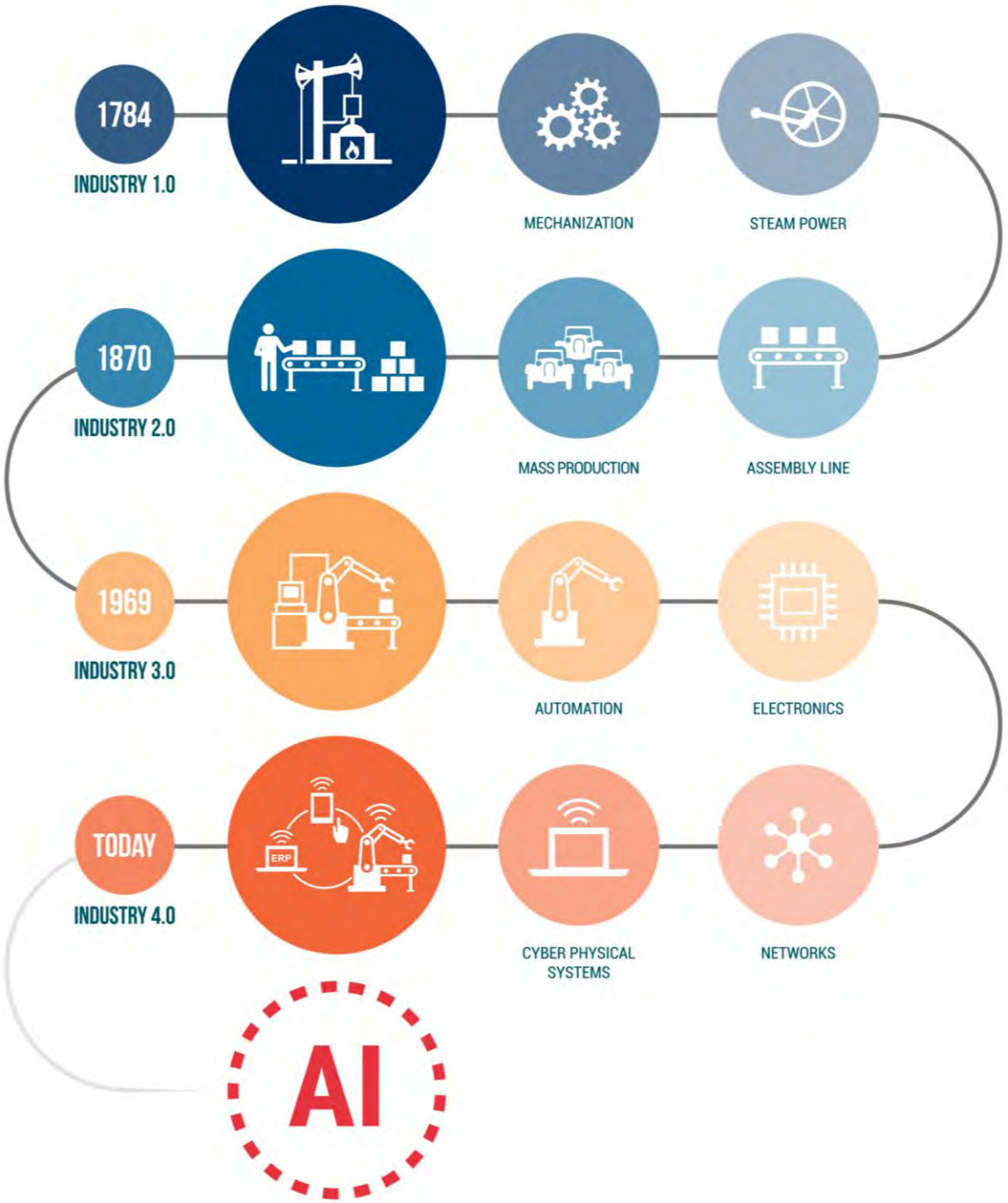
PdEnCat30



Pd/Al_2O_3



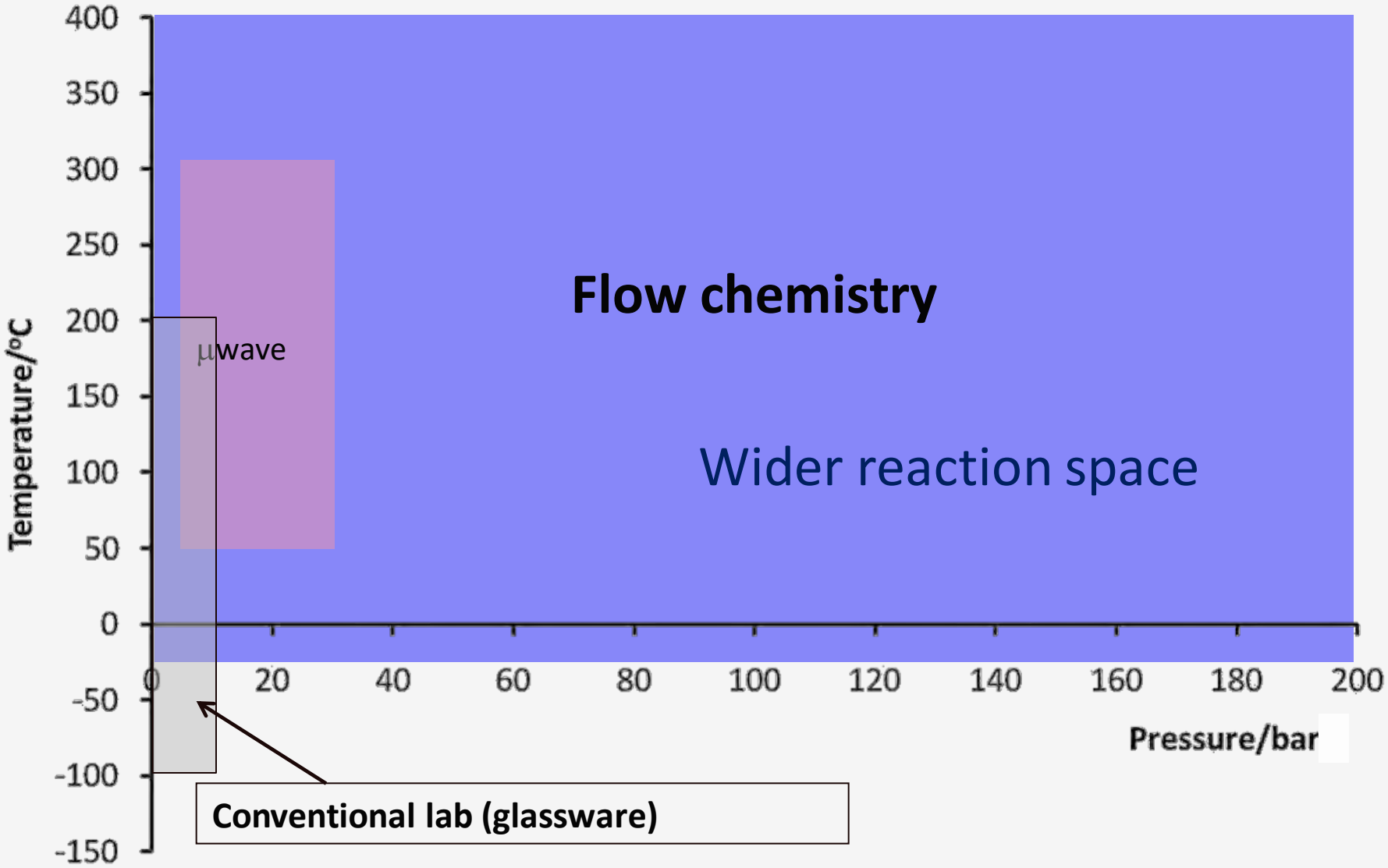
Marvin Alberts, MSci thesis (2021)



AT THE CUSP OF THE 5TH INDUSTRIAL REVOLUTION

Recent rapid adoption and application of artificial intelligence algorithms — triggered by access to big data and better hardware-processing capabilities — are changing the face of blue and white collar jobs.

CBINSIGHTS


Flow chemistry
Wider reaction space
μwave
Conventional lab (glassware)
Temperature/°C
Pressure/bar
400
350
300
250
200
150
100
50
0
-50
-100
-150
0
20
40
60
80
100
120
140
160
180
200