Seminar

The Application of Bayesian Optimization for Chemical Reaction

Reporter: Kunhua Li

Supervisor: Prof. Junliang Zhang

A. Prof. Zhiming Li

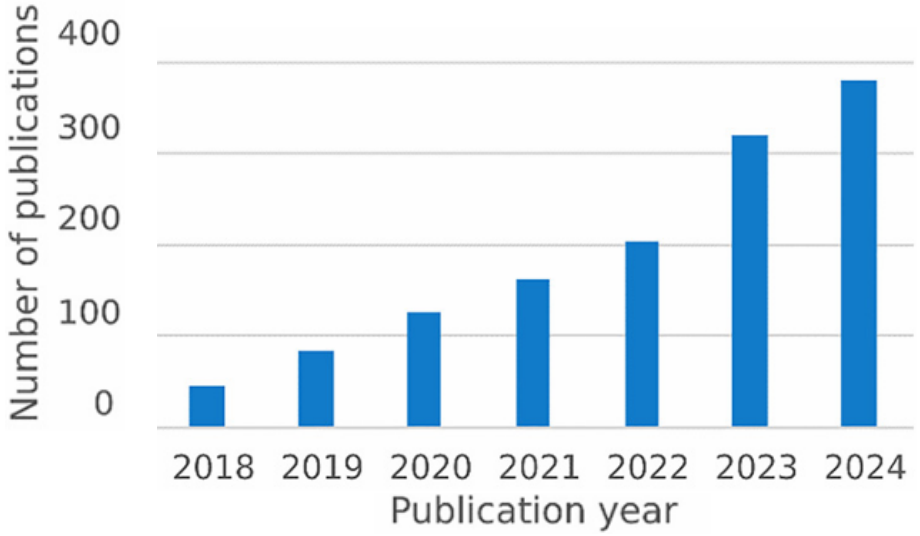
Date: 2026-04-24

Content

Content

Brief history of Bayesian Optimization

Concept Formulation 1760s - 1970s	Algorithm Implemented 1990s - 2010s	Chemical Application 2010s - Today
• 1763 Bayes' Theorem	• 1998 Algorithm Implemented	• 2018 Schweidtmann
• 1975 BO's framework	• 2012 Applied in AI	• 2021 Doyle



Lagarias, J. C. et al. *SIAM J. Optim.* **1998**, *9*, 112–147.
Criminisi, A. et al. *Found. Trends Comput. Graph. Vis.* **2012**, *7*, 81–227.
Schweidtmann, A. M. et al. *Chem. Eng. J.* **2018**, *352*, 277-282..
Shields, B. J. et al. *Nature* **2021,** *590*, 89-96
Li, S. W. et al. *Angew. Chem. Int. Ed. Engl.* **2026,** e4418883.

Workflow of Bayesian Optimization

Problem Definition
- Parameters: T, conc., etc
- Target: yield, ee, etc

Featurization
- One-hot Encoding
- Descriptors
- Fingerprints

Initialization
- 5-10 Experiments
- Random Distributed

Bayesian optimization modeling
- Surrogate Model
- Acquisition Function

Execution & Loop
- Conduct Experiments
- Data Feedback
- Model Update

Solvent
MeOH
DMF
MeCN
THF

→ **One-hot Encoding**

[1, 0, 0, 0]
[0, 1, 0, 0]
[0, 0, 1, 0]
[0, 0, 0, 1]

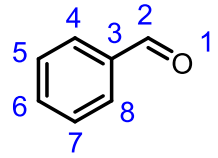
Descriptors
Atomic charge
Buried volume…

BO Model

Fingerprint

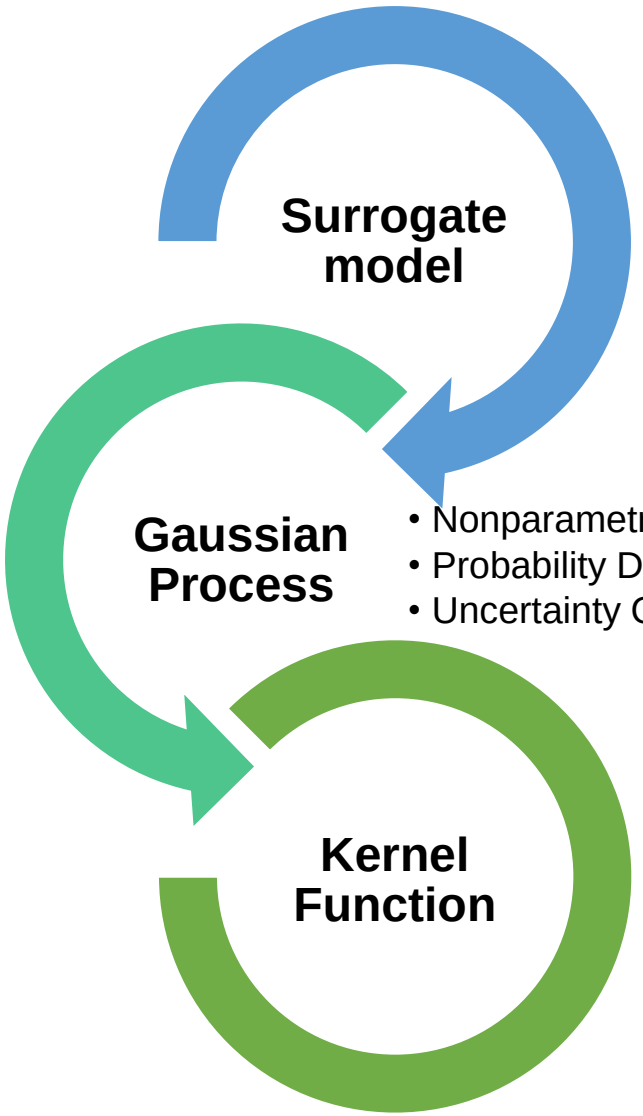
	1	2	3	4	5	6	7	8
1	0	2	0	0	0	0	0	0
2	2	1	0	0	0	0	0	0
3	0	1	0	2	0	0	0	1
4	0	0	2	0	1	0	0	0
5	0	0	0	1	0	2	0	0
6	0	0	0	0	2	0	0	0
7	0	0	0	0	0	0	0	2
8	0	0	1	0	0	0	2	0

Chemical Structure



Doyle, A. G. et al. *Nature* **2021,** *590*, 89-96.

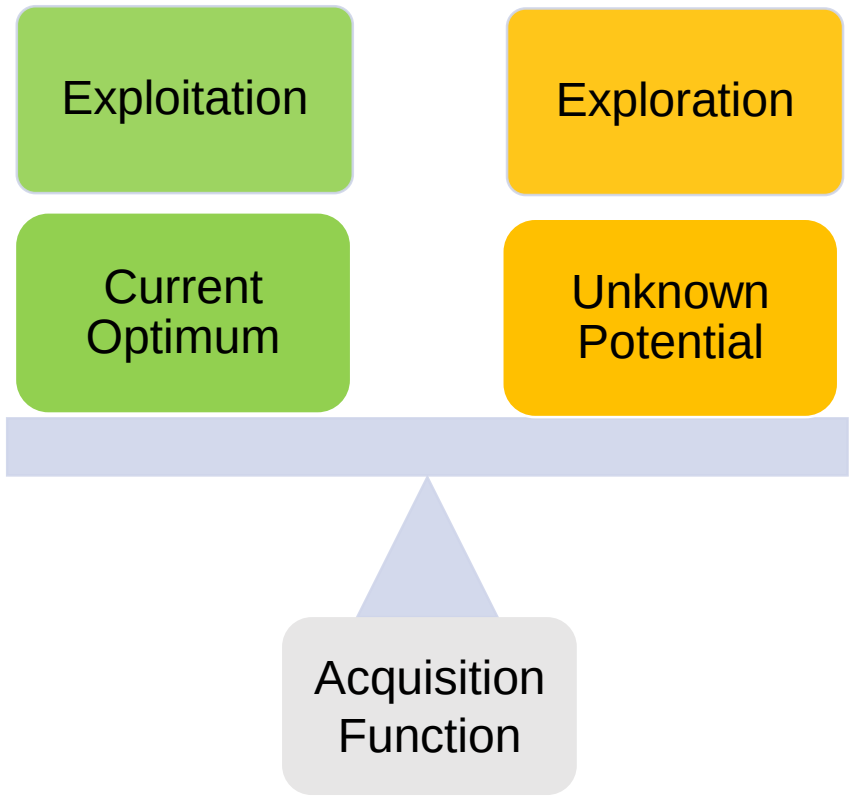
Principles of Bayesian Optimization



- Mathematical Approximation
- “Virtual Experiments"

- Nonparametric Characteristics
- Probability Distribution
- Uncertainty Quantification

- Similarity Measure
- Control Function
- Control Uncertainty

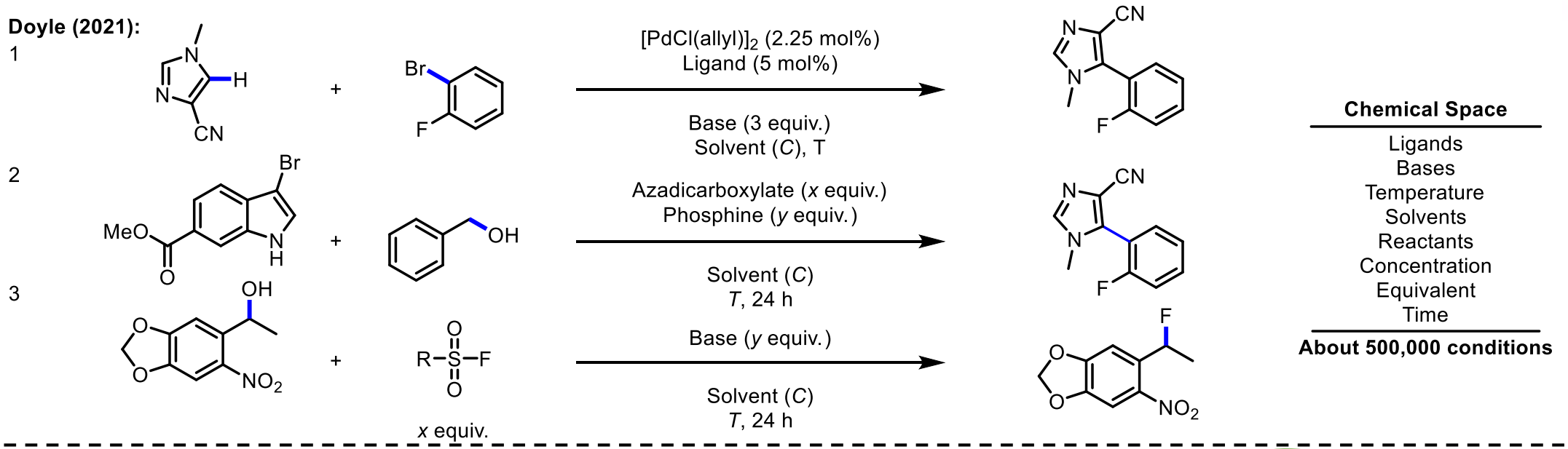


Wagner, F. L. et al. *Chem Sci* **2025,** *16,* 22447-22455.
Hong X. et al. *Angew. Chem. Int. Ed. Engl.* **2026,** e4418883.

Content

Applications in Laboratory

Doyle (2021):

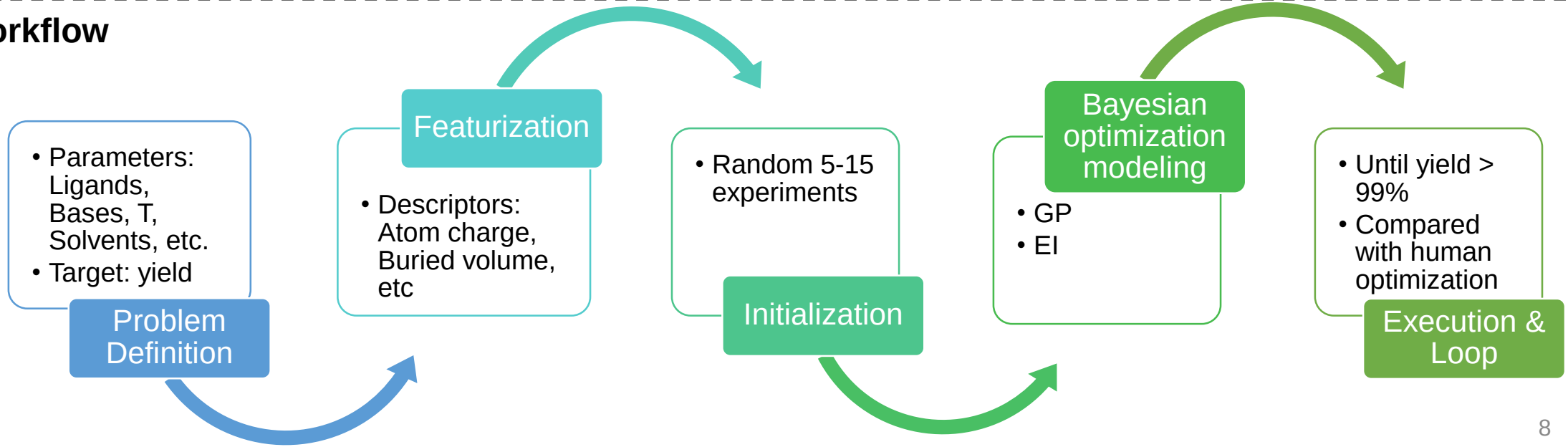


Chemical Space

Ligands
Bases
Temperature
Solvents
Reactants
Concentration
Equivalent
Time

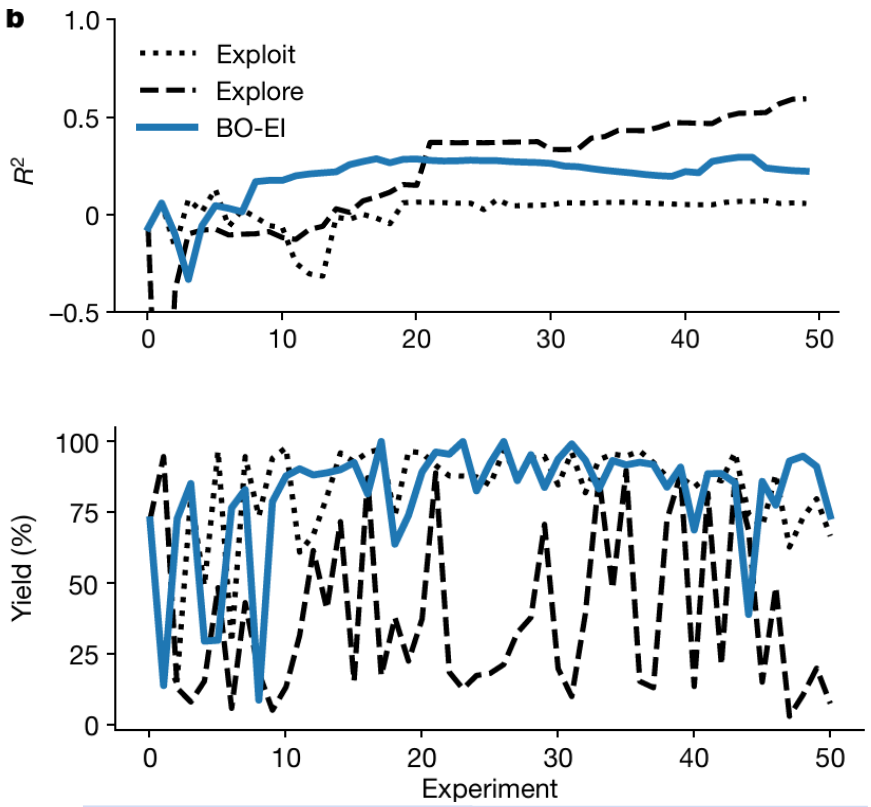
About 500,000 conditions

Workflow

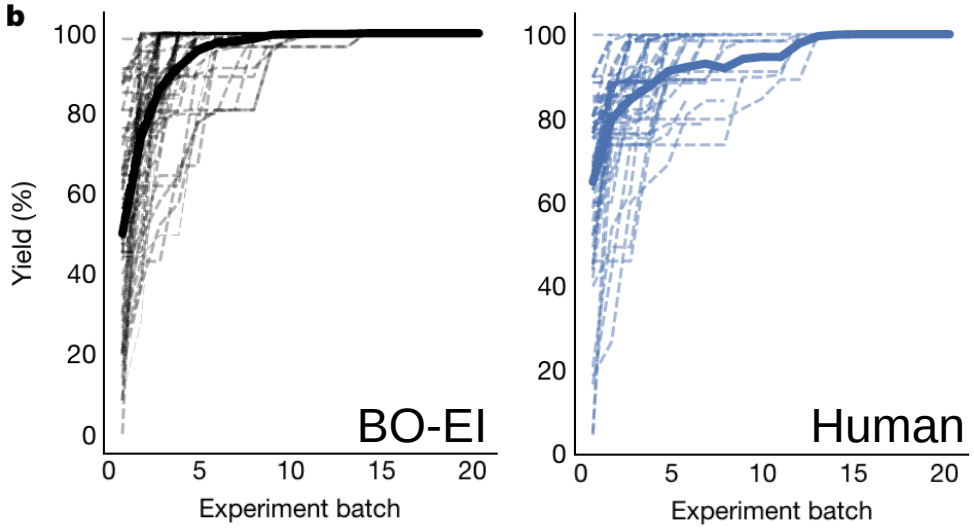


Doyle, A. G. et al. *Nature* **2021,** *590*, 89-96.

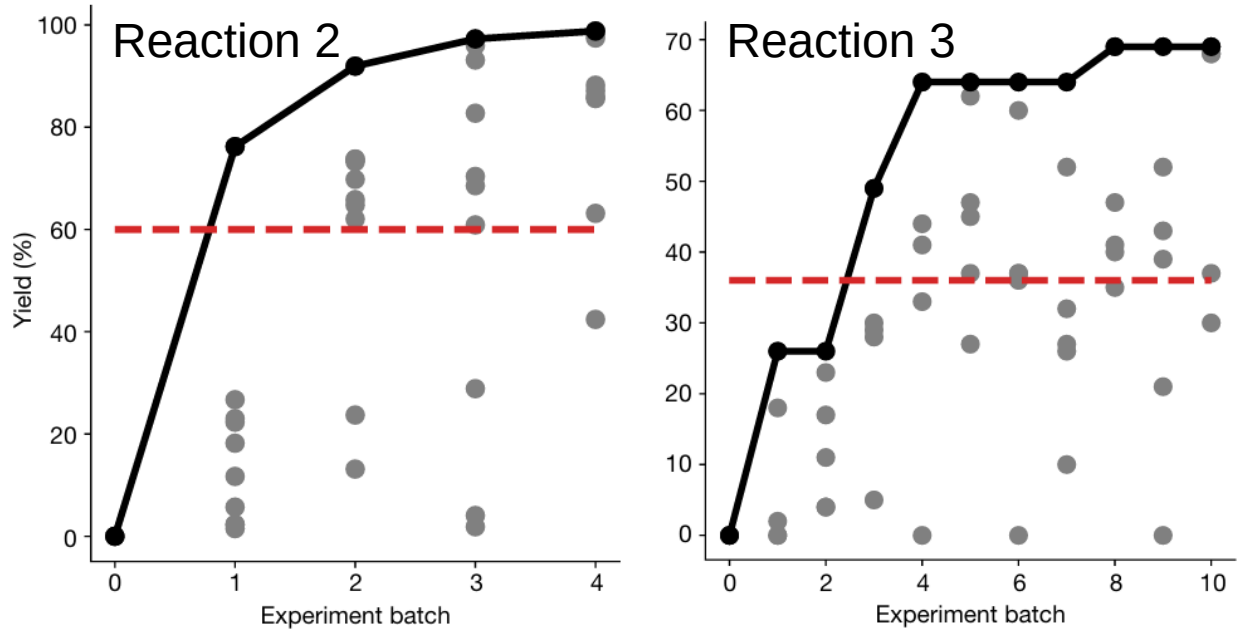
Applications in Laboratory



R^2 score	Yield score
Exploration best	Exploitation best
BO-EI	
Stable and excellent fit with better long-term performance	



BO-EI
• Initial lower but quickly improvement
Human
• Empirical bias • Individual differences



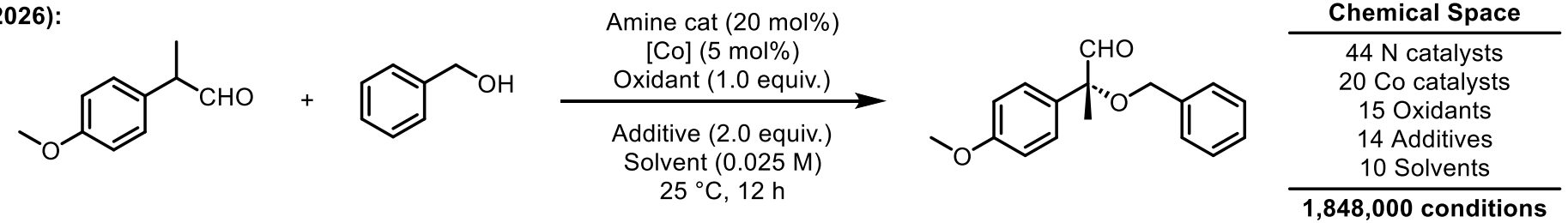
40-50 experiments

Yield: 0% to 70%+

Doyle, A. G. et al. *Nature* **2021,** *590*, 89-96.

Applications in Laboratory

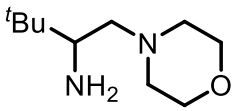
Luo (2026):



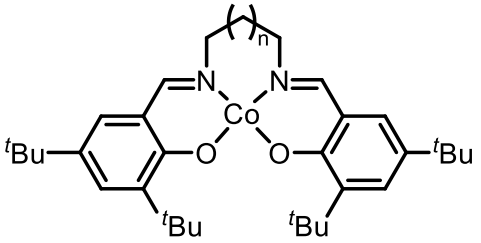
Chemical Space
44 N catalysts
20 Co catalysts
15 Oxidants
14 Additives
10 Solvents
1,848,000 conditions

cat-1: R = ^{t}Bu
cat-2: R = Ad

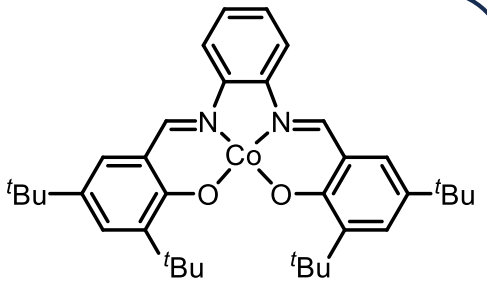
F-1



cat-3



[Co]-1: n = 0
[Co]-2: n = 1



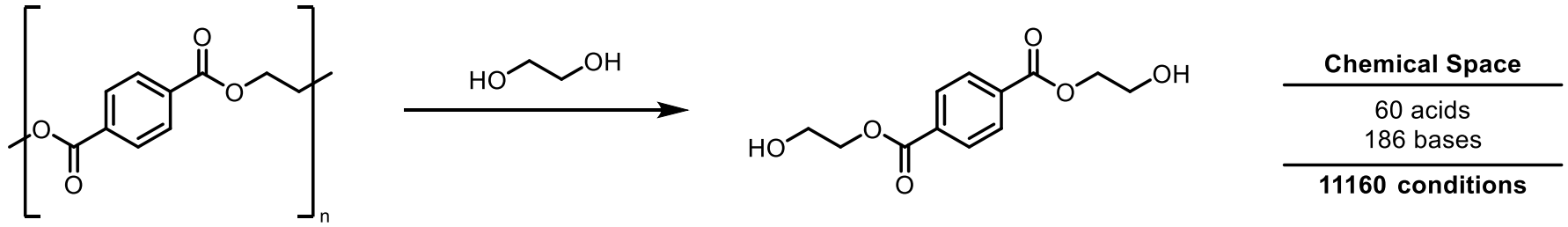
[Co]-3

Entry	HO (Human Optimization)	BO-1 (Using CVT to select initial points)	BO-2 (Using HO Optimal Condition as initial points)	BO-3 (Another set of initial points)
Optimal Condition	cat-1•TFA (20 mol%) [Co]-1 (5 mol%) F-1 (1.0 equiv.) KHCO3 (2 equiv.) DCM (0.025 M)	cat-1•HNTF2 (20 mol%) [Co]-2 (5 mol%) NCS (1.0 equiv.) KHCO3 (2 equiv.) DCM (0.025 M)	cat-2•TFA (20 mol%) [Co]-1 (5 mol%) F-1 (1.0 equiv.) KHCO3 (2 equiv.) DCM (0.025 M)	cat-3•HOTf (20 mol%) [Co]-3 (5 mol%) F-1 (1.0 equiv.) no additive DCE (0.025 M)
Total Experiments	200+	83	15	106
Yield/ee (%)	50/88	57/90	54/88	50/80

Luo S. et al. *Nat. Syn.* **2026,** *5 (3)*, 455-465.

Applications in Laboratory

Jiang (2026):



Chemical Space

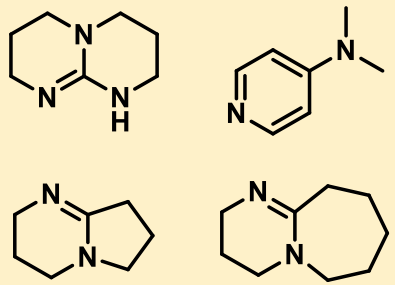
60 acids
186 bases

11160 conditions

Stage 0
Literature-guided screening as baseline

Acid
- $Zn(OAc)_2$
- $Zn(TFA)_2$
- $Zn(OTf)_2$

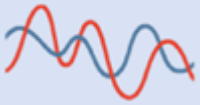
Base:



Stage 1
Chemical space exploration by BO

Lewis acid-base pairs

Knowledge-encoded chemical space



Bayesian optimization

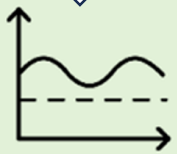
Stage 2
Insight distillation and hypothesis generation by LLM

Encode chemical features

Extract key insight

Generate hypotheses

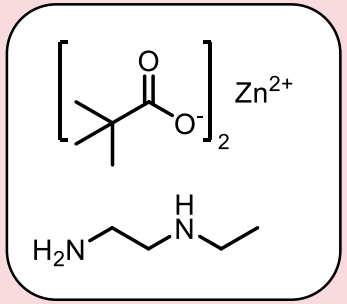
Provide recommendation



Inject domain preferences

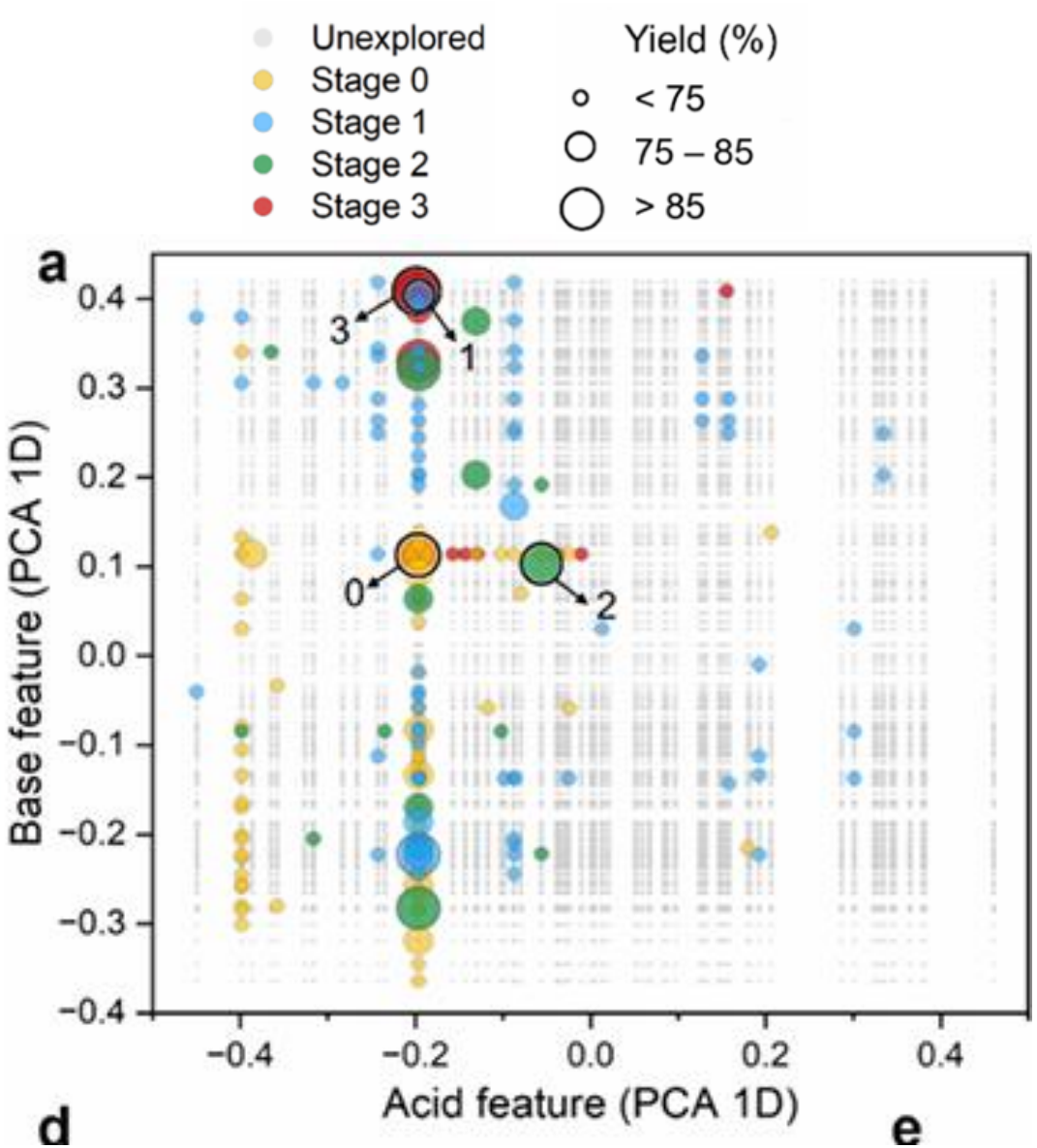
Stage 3
Exploration of out-of-sample chemical space

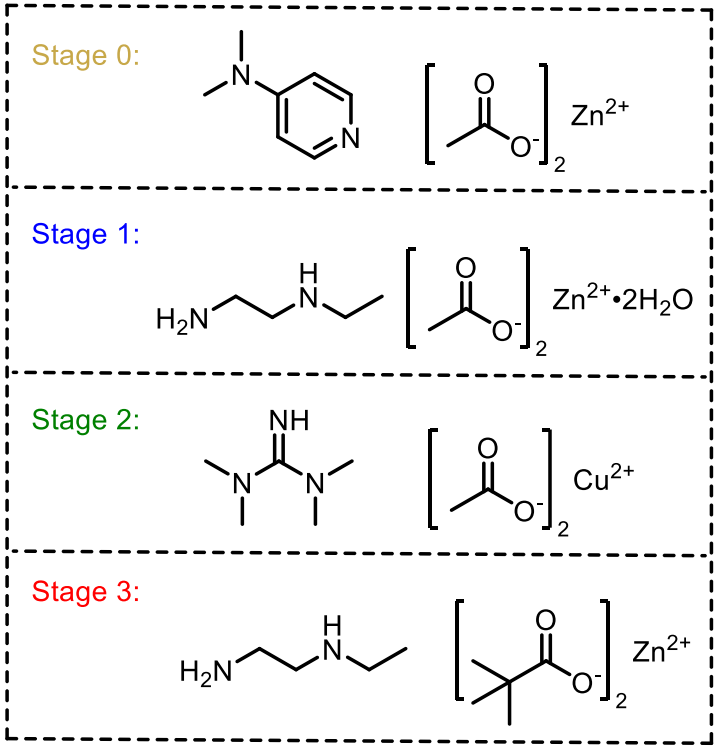
Human experts design

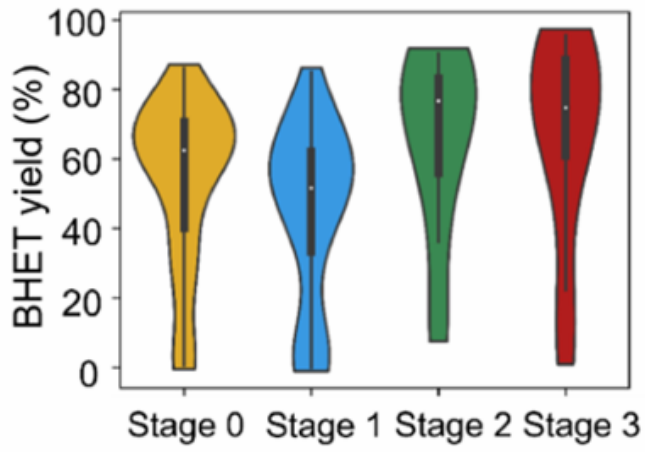


Jiang J. et al. *J. Am. Chem. Soc.* **2026,** *148*, 4635-4644.

Applications in Laboratory







Baseline
86% yield

Expand the explored chemical space

Yield up to 90%

Global optimal catalyst
95% yield

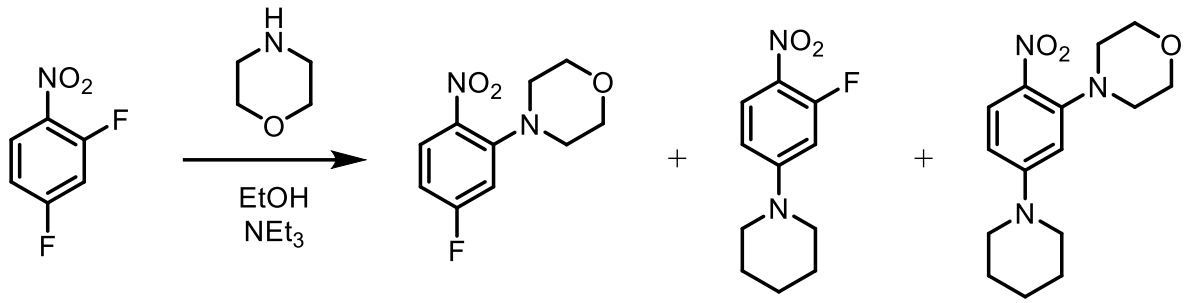
AI-Automation-Human cooperative paradigm

Jiang J. et al. *J. Am. Chem. Soc.* **2026,** *148*, 4635-4644.

Content

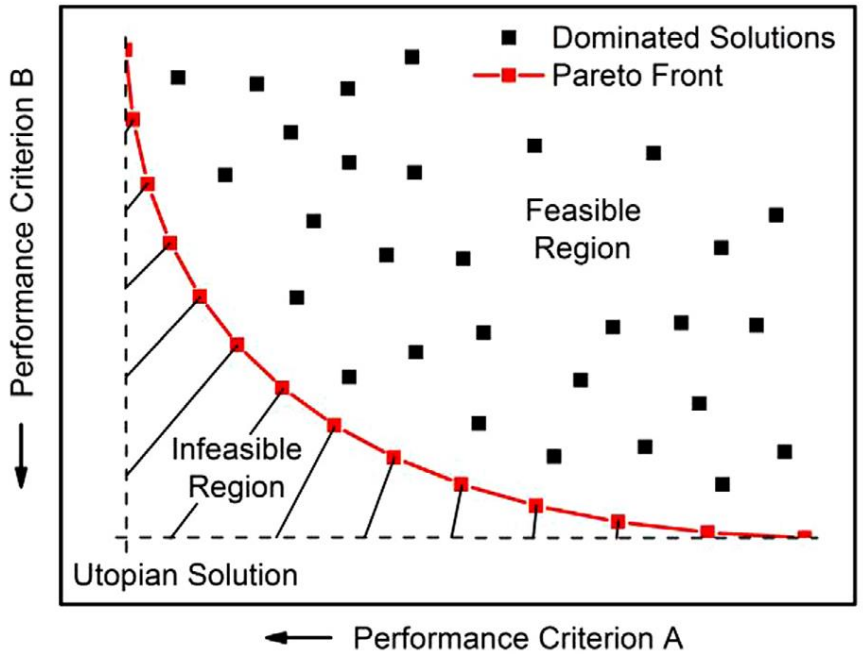
Applications in Industry

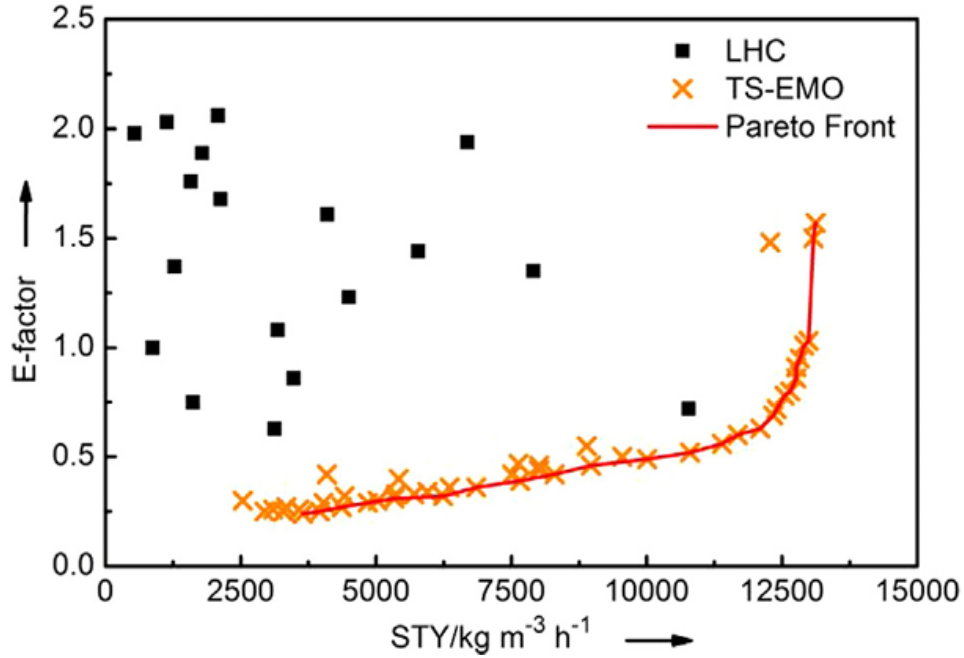
Schweidtman (2018):



Chemical Space

Flow rate
Equivalents of reagent
Temperature
Concentration of reactant





Pareto Front

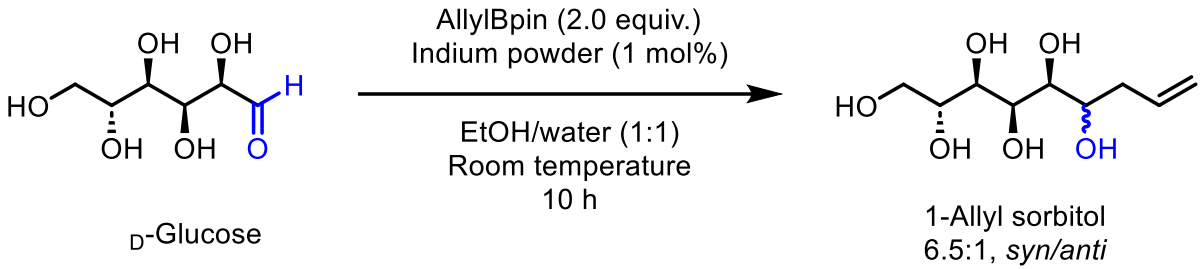
the best trade-off between two competing parameters

One night + about 50 data ⇨ Pareto Front

Schweidtmann, A. M. et al. *Chem. Eng. J.* **2018**, *352*, 277-282.

Applications in Industry

Denmark (2025):



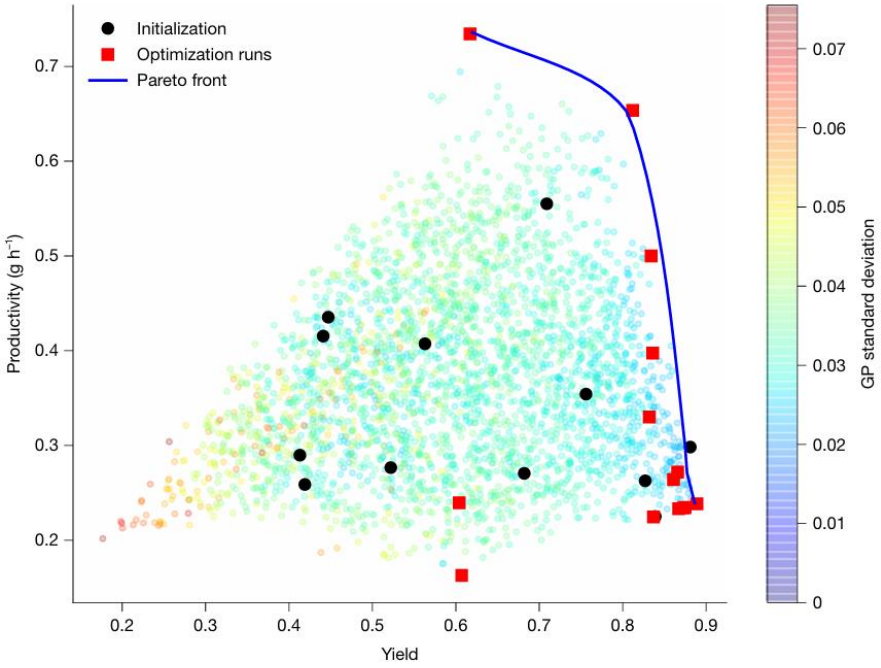
D-Glucose

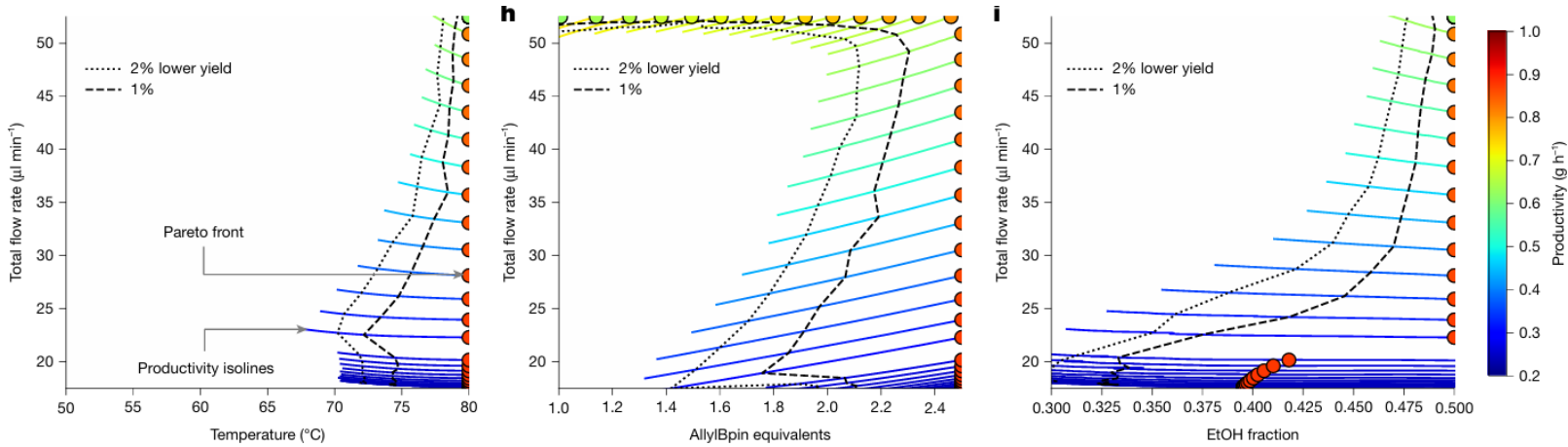
1-Allyl sorbitol
6.5:1, *syn/anti*

Chemical Space

Flow rate
Equivalents of allylBpin
Temperature
Concentration of EtOH

$$P(X)=m(X)\cdot Y(X)$$





Robust

Widely operate

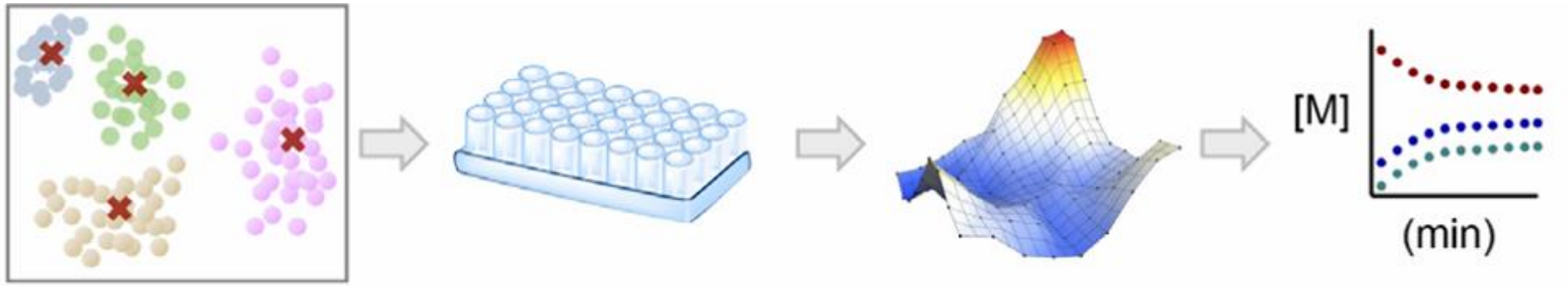
Amenable to industrial

Denmark, T. et al. *Nature* **2025,** *640*, 94-99.

Content

Conclusion & Outlook

Training set design | Data collection | Bayesian optimization | Execution & Loop



Current Achievements

- Mechanism and efficiency
- A wide range of application scenarios
- Integrated with other technologies

Future Outlook

- Deeply integrate with multi-modal AI
- Fully autonomous "Self-Driving" laboratories
- Conquering high-dimensional chemical spaces

Sin, J. W. et al. *Nat Commun* **2025**, *16*, 6464.

Thanks for attention!