

医薬資源の自動合成を志向した 機械学習による反応条件最適化

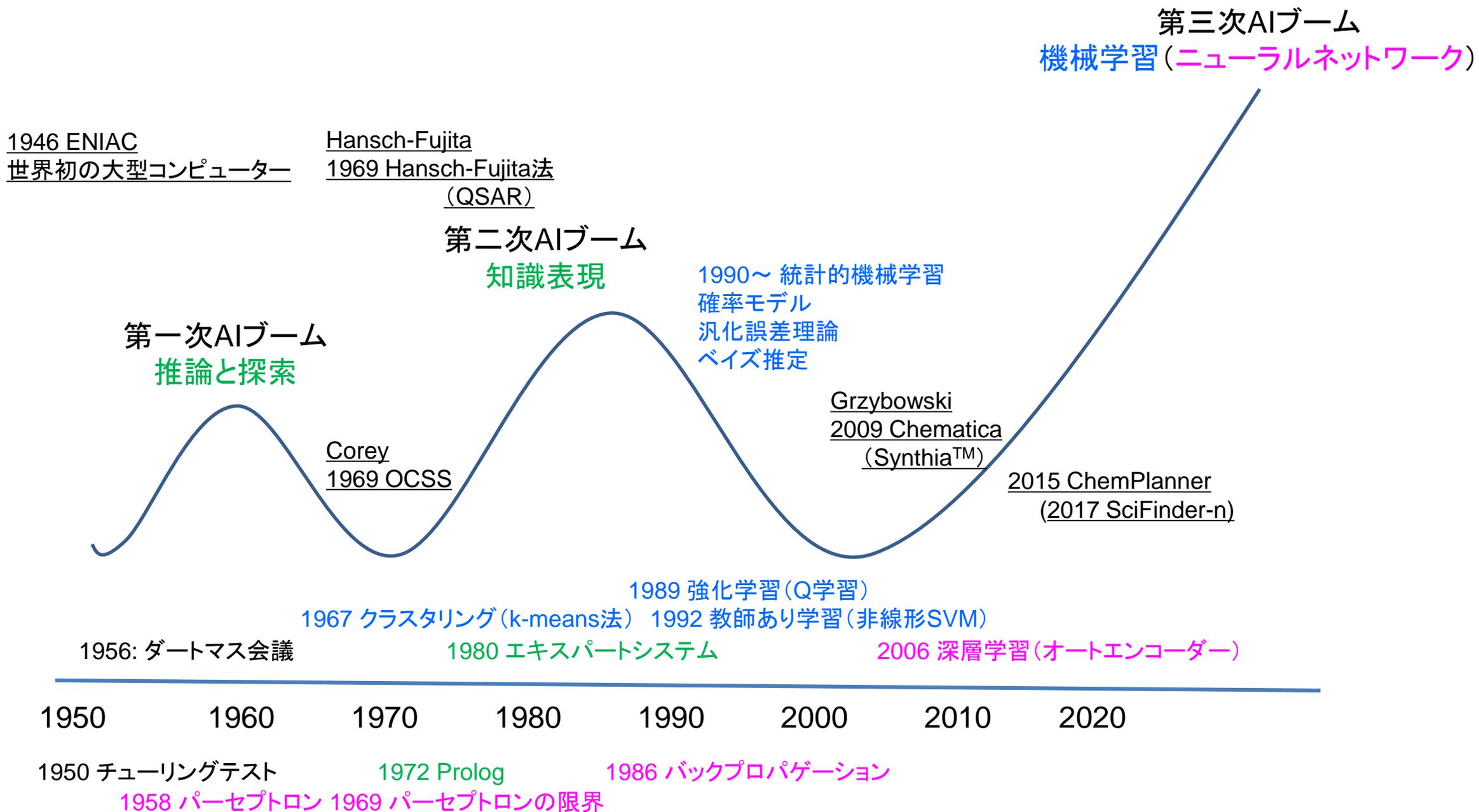
Machine-learning-assisted screening for
automated synthesis of pharmaceutical resources

滝澤 忍

大阪大学 産業科学研究所

分子システム創成化学研究分野 & 産業科学AIセンター

化学分野でのAI



化学分野でのAI・機械学習

1. 逆合成経路解析

1. OCSS, 2. LHASA, 3. SYNCHEM, 4. SYNGEN, 5. WODCA, 6. CHIRON, 7. SESAM, 8. SECS, 9. SST, 10. SYNSUP-MB, 11. KOSP, 12. LILITH, 13. TRESOR, 14. Chemetica (Synthia™), 15. ChemPlanner (SciFinder-n)

2. 新規反応探索

Grzybowski (Tactical Combinations (TCs)), Maeda (Global Reaction Route Mapping (GRRM))

3. 触媒設計

Yada & Sato, Denmark, Yamaguchi

4. 反応条件最適化: 収率・エナンチオ選択性の向上

Sigman & Phipps, Adams & Doyle, Takeda & Mase
Takizawa & Washio

5. 創薬、材料創製etc



学術変革研究A 2021年9月発足

機械学習

教師あり学習: あらかじめ問題に正解を与える

【回帰】

ある基準の確率分布に基づいて、結果を予測する手法

例: いくつかの反応条件とその収率から最高収率を与える反応条件を予測する

【分類】

ある基準に基づいてデータを分類することにより、結果を予測する手法

例: 犬、猫、鳥の写真から犬の写真だけを選ぶ

教師なし学習: あらかじめ問題に正解を与えない

【クラスタリング】

類似性を基に与えられたデータをグループ化

例: 色の類似性からレモンとバナナを同一グループとする

【次元削減】

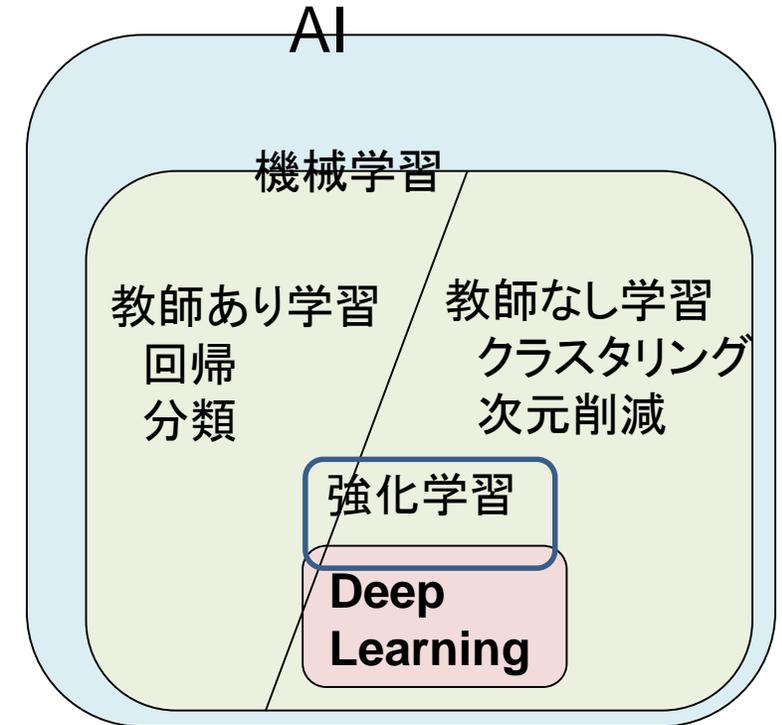
主成分分析 (PCA: Principal Component Analy)

例: 身長と体重から体格を予測

例: 顧客満足度の調査

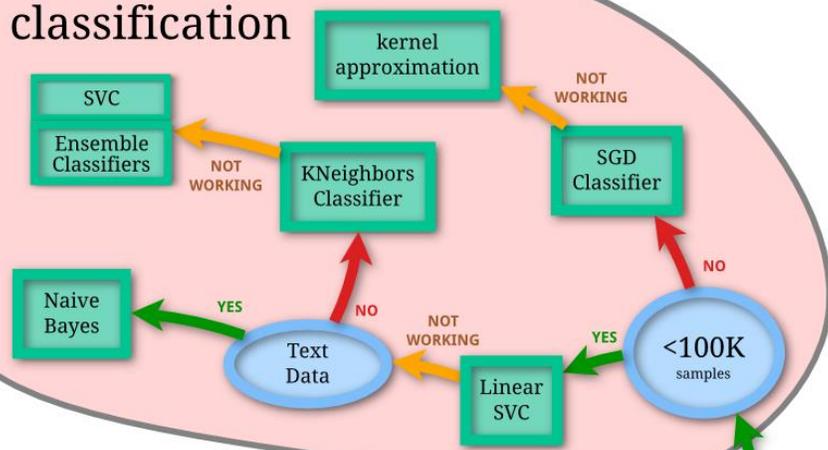
強化学習: 正解はなく目的の報酬を最大化する行動を学習

例: 将棋AI、囲碁AI (深層強化学習)



【分類】

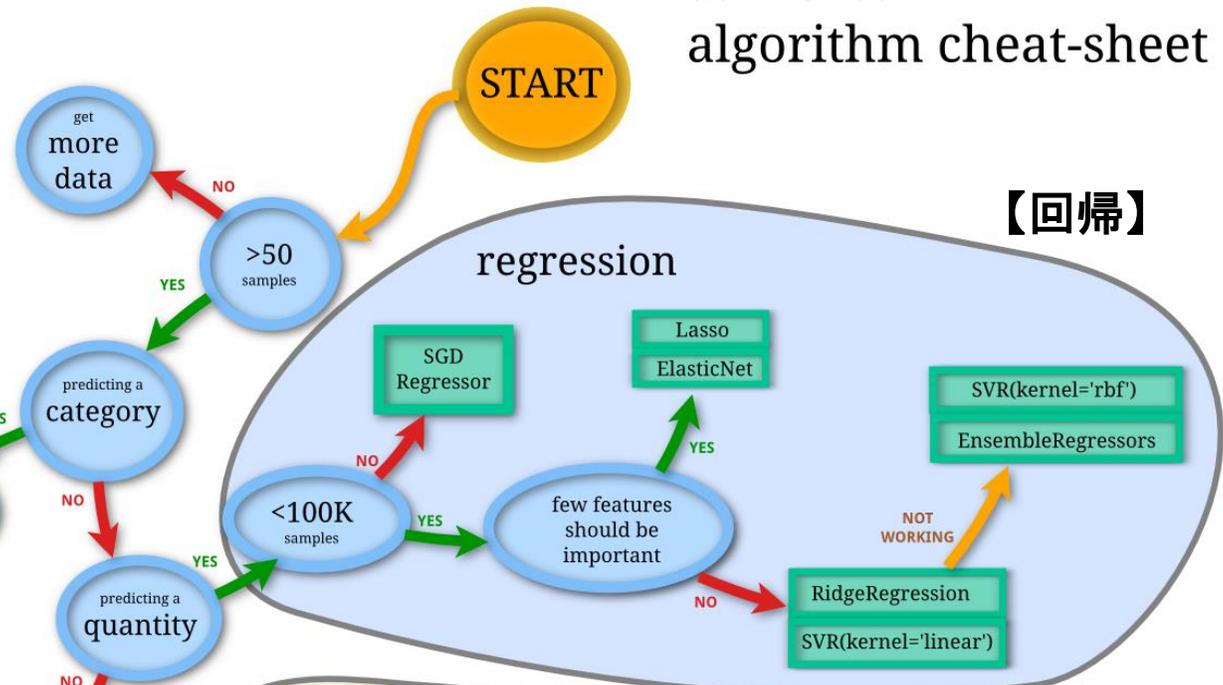
classification



scikit-learn
algorithm cheat-sheet

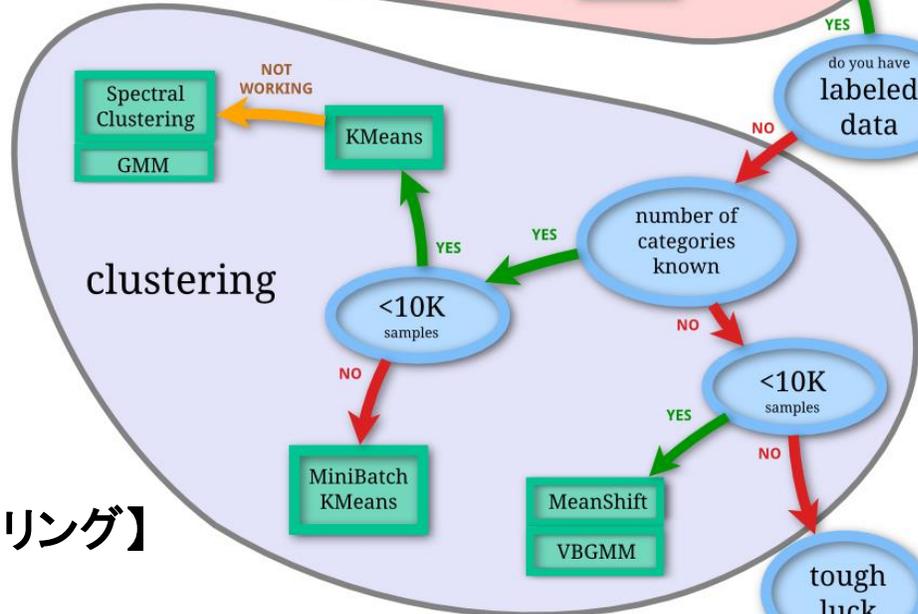
【回帰】

regression



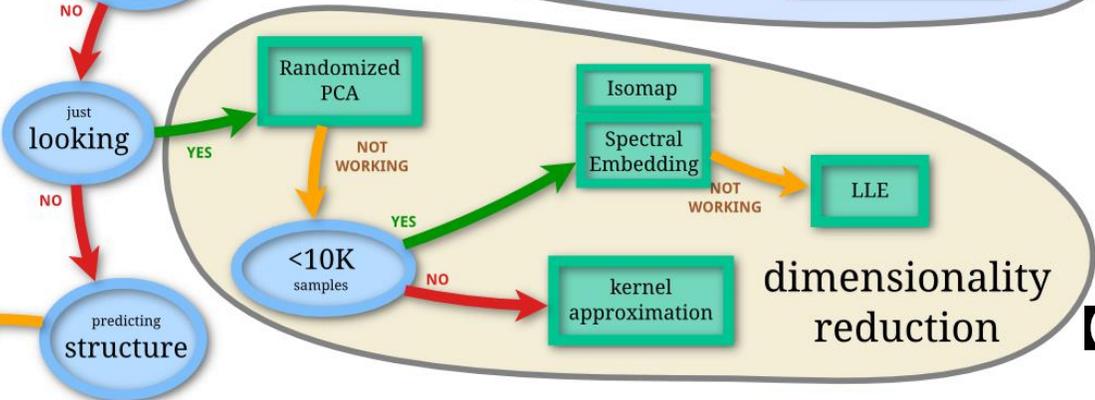
【クラスタリング】

clustering



【次元削減】

dimensionality reduction



Scikit-learn: Python(プログラミング言語)で使用できる

フロー合成と電解合成

フロー合成 (20世紀後半～)

1990年代後半に工業プロセスの安全性と効率向上を目的に開発
スケールアップ合成が比較的容易

2010年頃から医薬原料などファインケミカルズの自動合成ツールとして注目



UNIQSIS社製
FlowSyn

電解合成 (19世紀前半～)

触媒や試薬を必要としない電気化学的酸化・還元が可能

2010年代半ばから複雑な天然物合成にも利用(電位など電解条件による選択性の発現)

近年、フロー電解合成研究も活発に展開



IKA社製
ElectraSyn 2.0

フロー合成・電解合成の特徴

- ①分子拡散・混合や熱移動を精密に制御可能
- ②個々の操作が実験者の技術に依存しにくい(高い再現性)
- ③反応温度・基質当量・溶液の混合速度などのパラメータを容易に変更可能
- ④コンピュータ制御による自動化、信頼性の高いデータ集積化が容易

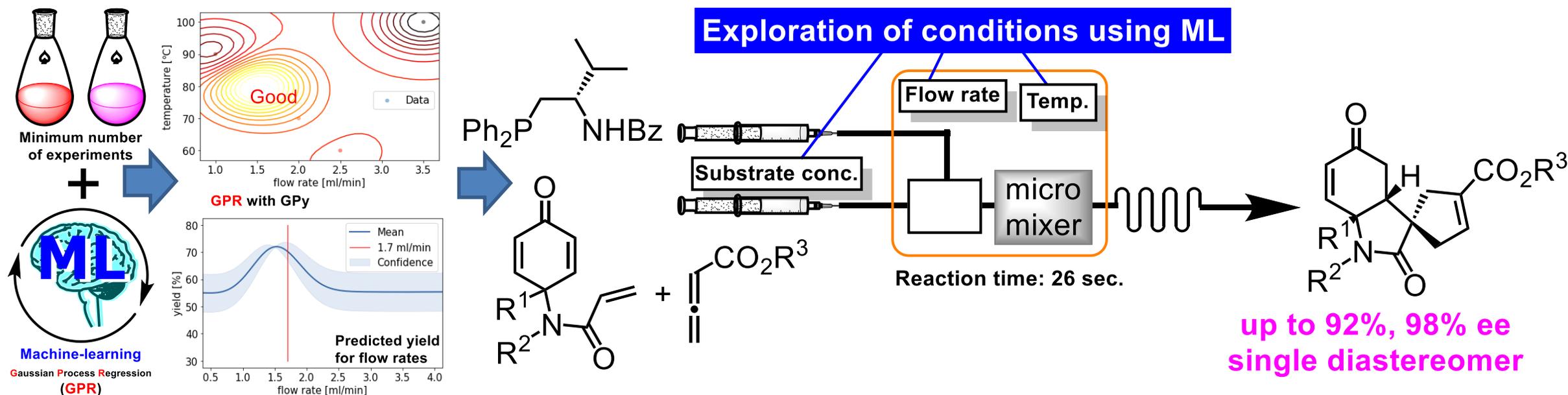
機械学習との
高い親和性

最少実験データ(≦15)による機械学習合成→大幅コスト削減

1. 最少実験データとガウス過程回帰によるフロー反応条件(2次元パラメータ)の迅速最適化 -予測収率の可視化-

不斉フロー合成: *Chem. Commun.* **2020**, *56*, 1259 (Highlighted in *Synfacts*, *EurekAlert!*)

フロー電解合成: *J. Org. Chem.* **2021**, *86*, 16035; *Org. Process Res. Dev.* **2023**, in press, DOI: 10.1021/acs.oprd.2c00267 (菅誠治教授@岡山大学との共同研究)



2. ベイズ最適化による電解ケチミン合成反応条件(5次元パラメータ)の迅速最適化

Green Chem. **2021**, *23*, 5825

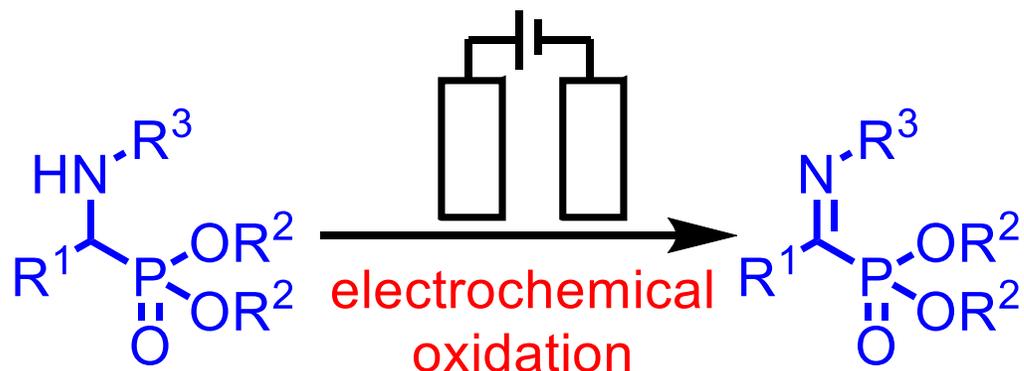
3. 並列ベイズ最適化によるビアリール化合物のフロー合成反応条件(ミキサー選択を含む6次元パラメータ)の迅速最適化

Commun. Chem. **2022**, *5*, 148

最少実験データ(≦15)による機械学習合成→大幅コスト削減

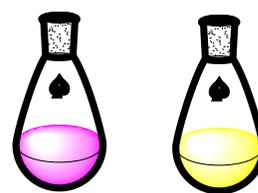
1. 最少実験データとガウス過程回帰によるフロー反応条件(2次元パラメータ)の迅速最適化 -予測収率の可視化-
不斉フロー合成: *Chem. Commun.* **2020**, 56, 1259 (Highlighted in *Synfacts*, *EurekaAlert!*)
フロー電解合成: *J. Org. Chem.* **2021**, 86, 16035; *Org. Process Res. Dev.* **2023**, in press,
DOI: 10.1021/acs.oprd.2c00267 (菅誠治教授@岡山大学との共同研究)

Electrolysis of α -ketiminophosphonate



1. metal free
2. external oxidant free
3. mild conditions

ML-assisted simultaneous multiparameter screening



minimum number of experiments
collecting dataset
evaluating ML estimation

+



Machine Learning
(Bayesian optimization)

1. efficient and rapid screening
2. save energy, time, and labor

2. ベイズ最適化による電解ケチミン合成反応条件 (5次元パラメータ) の迅速最適化

Green Chem. **2021**, 23, 5825

3. 並列ベイズ最適化によるビアリアル化合物のフロー合成反応条件(ミキサー選択を含む6次元パラメータ)の迅速最適化

Commun. Chem. **2022**, 5, 148

機械学習を活用する電解酸化反応の最適反応条件探索

本研究

ベイズ最適化を利用して、5パラメータ（出発物質の濃度、電解質の濃度、反応時間、反応温度、電流値）を同時に探索・最適化

ベイズ最適化

- ・次に行うべき（最大値を取り得る）実験点を探索
- ・局所解に陥りにくい（活用と探索）
- ・人間には予測困難な多次元パラメータに応用可能

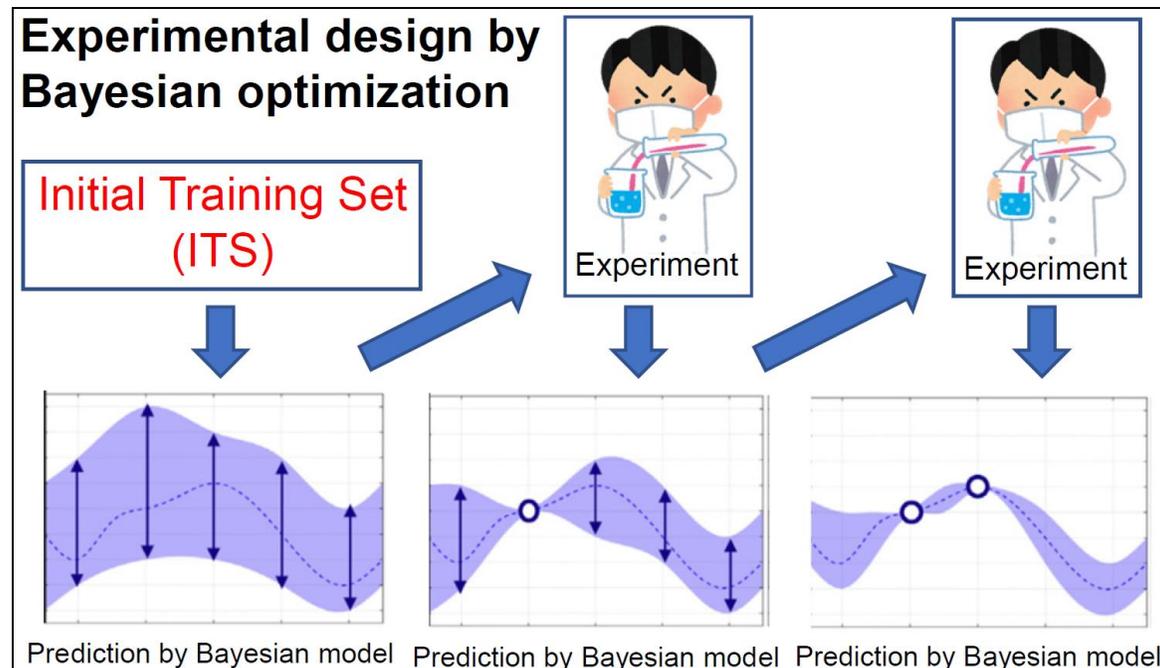
収率に大きく影響を与える
複数のパラメータを検討



必要な実験回数を大幅に削減

ベイズ最適化

適切なパラメータ値を設定



ベイズ最適化の流れ

- ① 数例の実験データ(ITS)を収集
- ② ガウス過程回帰による平均と分散の算出
- ③ ②で得られた値を利用して探索点を求める
- ④ 探索点を評価し、ガウス過程を更新する

②から④を繰り返し、最大値を取り得るパラメータを探索

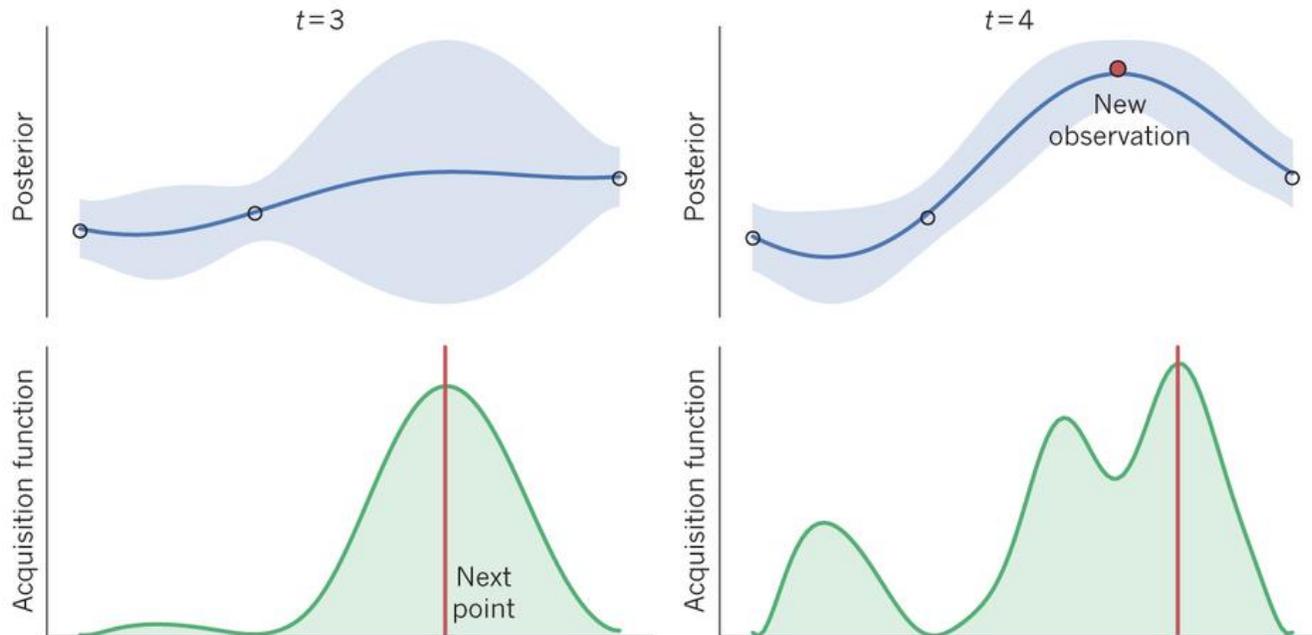
ベイズ最適化のイメージ

ベイズ最適化

ベイズ最適化は、不確かさを言いながら次に探索すべき点を決定するブラックボックス最適化アルゴリズム。

下の緑色の図で示す獲得関数が最大になる点を探索する。

真の関数を近似する獲得関数(acquisition function)の構成に確率過程、特にガウス過程を用いるのがベイズ最適化。期待値と分散から計算される獲得関数($f(x)$ 由来の近似関数)を利用することで、期待値が最大を示す近傍のパラメータ値だけでなく(=活用, exploitation)、分散が大きく最大値になり得るパラメータ値も探すことができる(=探索, exploration)



獲得関数

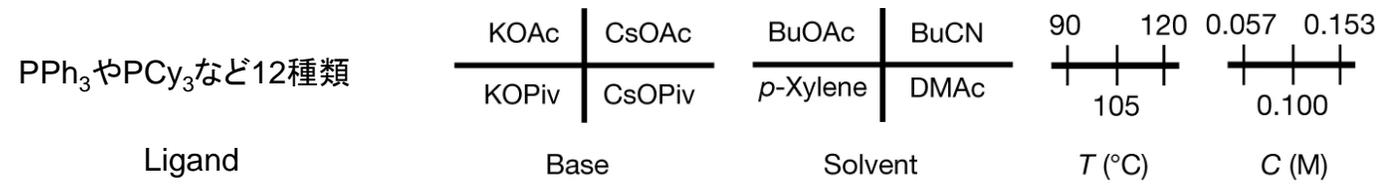
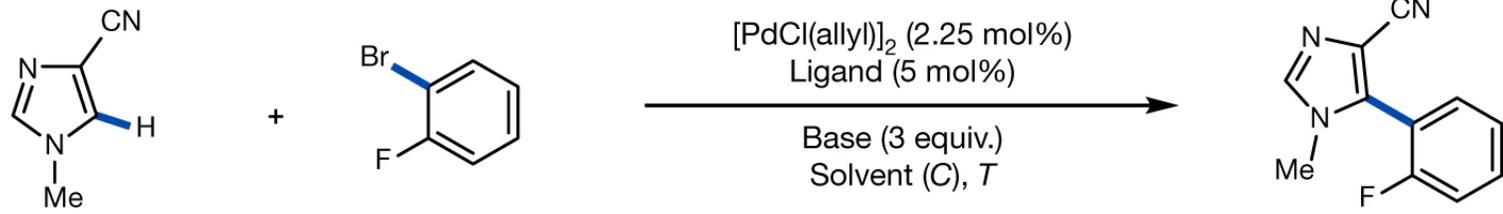
- Probability of Improvement (PI)
改善確率を指標 \rightarrow 活用重視
- Expected Improvement (EI)
改善確率と改善量の期待値を指標
- Lower Confidence Bound (LCB)
探索と活用のトレードオフ

$$LCB = \mu - \beta \times \sigma$$

期待値: μ , 標準偏差: σ

β : 探索と活用の割合をコントロールする係数

ベイズ最適化による合成条件探索が世界的に注目されつつある



Doyle, A. G. *et al.*, *Nature*, **2021**, 590, 89; *J. Am. Chem. Soc.* **2022**, 144, 19999

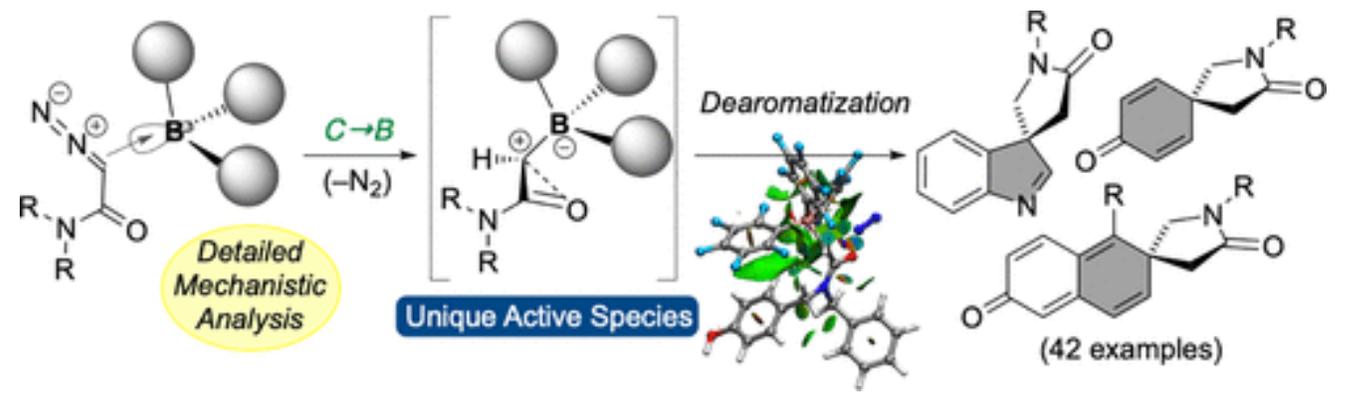
Micro-flow approach

Bayesian optimization

0.30 M
Me₂NBn (pKaH = 8.9)
20 °C
5.1 s

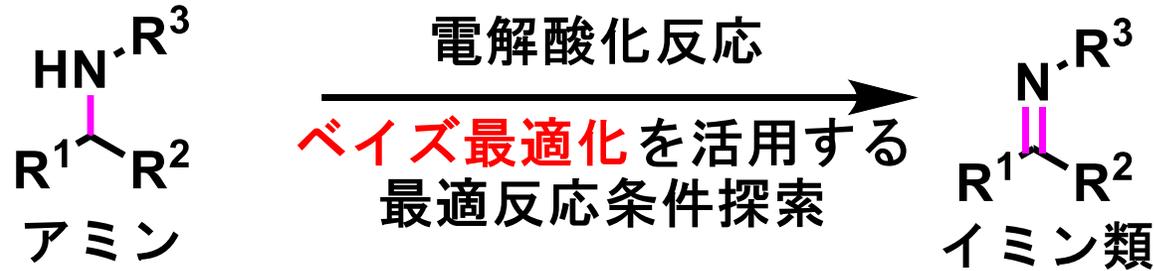
next conditions are ...

Fuse, S. *et al.*, *Chem. Methods* **2021**, 1, 484



Harada, S.; Nemoto, T. *et al.*, *ACS Catal.* **2023**, 13, 147

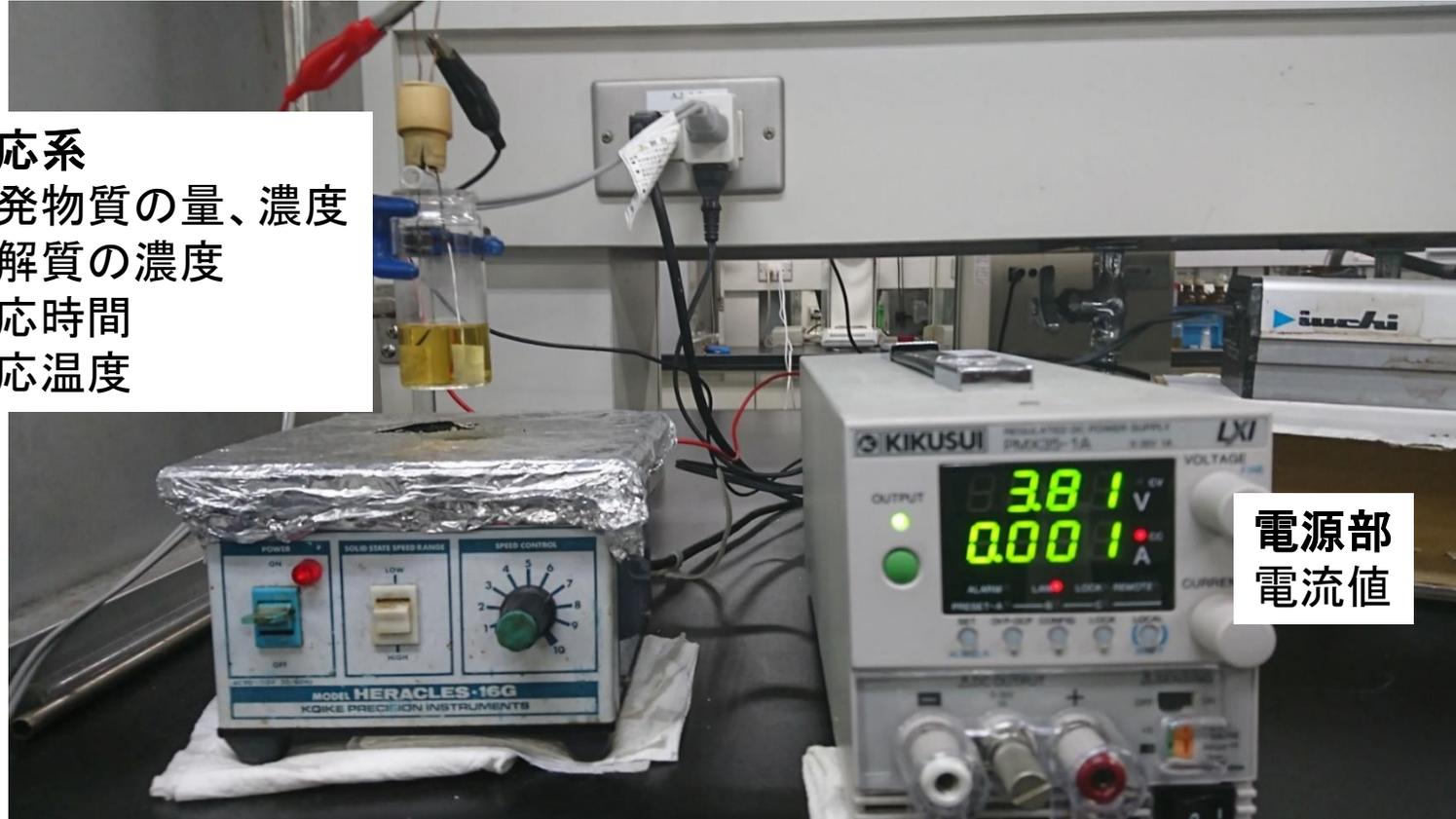
電解酸化反応のパラメータ設定



バイズ最適化を活用する
最適反応条件探索

Green Chem. 2021, 23, 5825

反応系
出発物質の量、濃度
電解質の濃度
反応時間
反応温度



電源部
電流値

人力による網羅的スクリーニング（従来法）：膨大な実験データが必要
機械学習による条件探索（本研究）：最少限の実験データで達成可能

Experimental design by Bayesian optimization

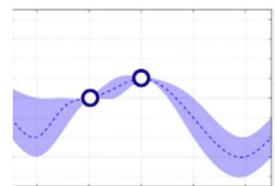
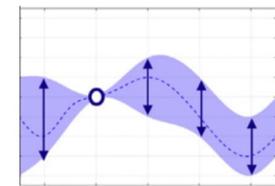
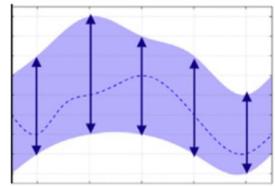
Initial Training Set (ITS)



Experiment



Experiment



Prediction by Bayesian model

Prediction by Bayesian model

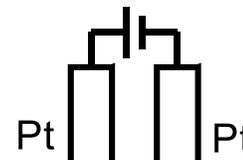
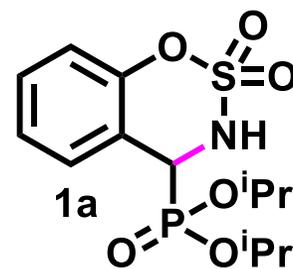
Prediction by Bayesian model

Initial Training Set

Training Set

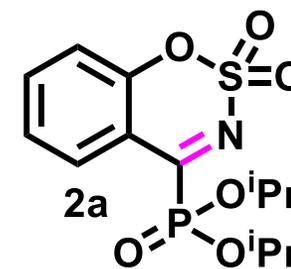
Training Set

ベイズ最適化を活用する最適反応条件探索



電流値 (mA)

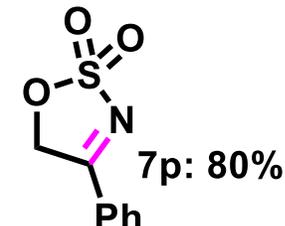
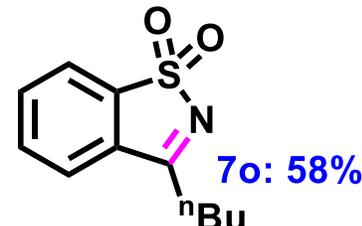
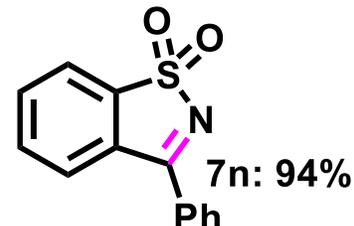
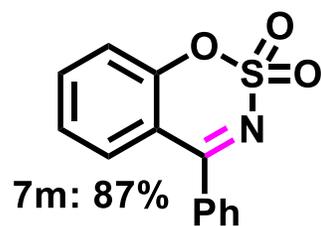
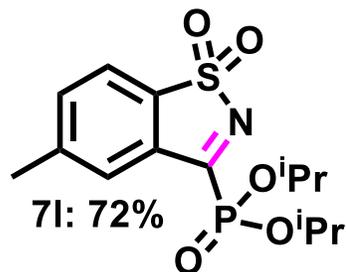
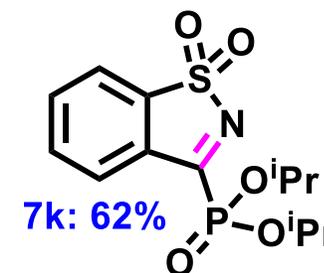
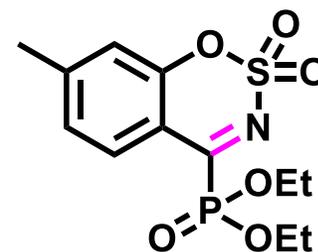
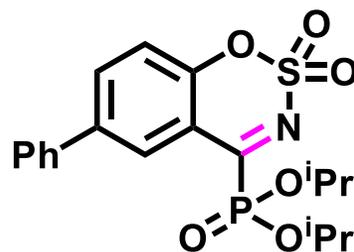
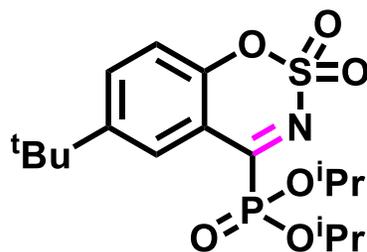
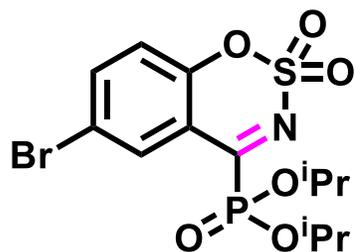
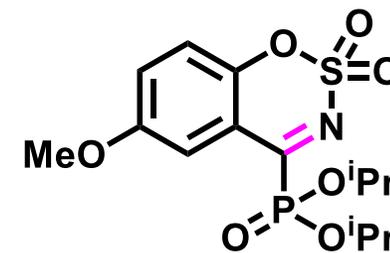
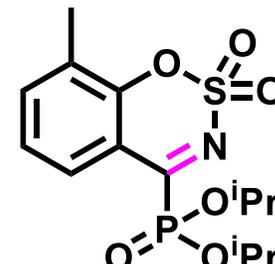
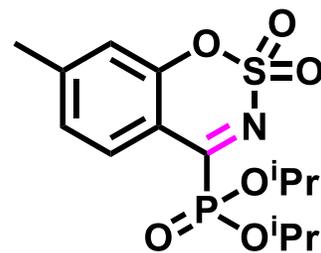
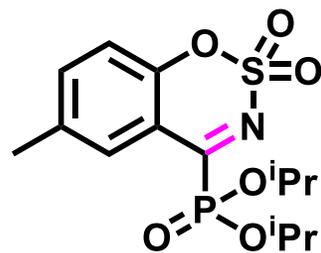
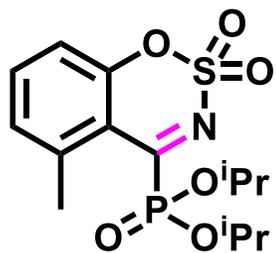
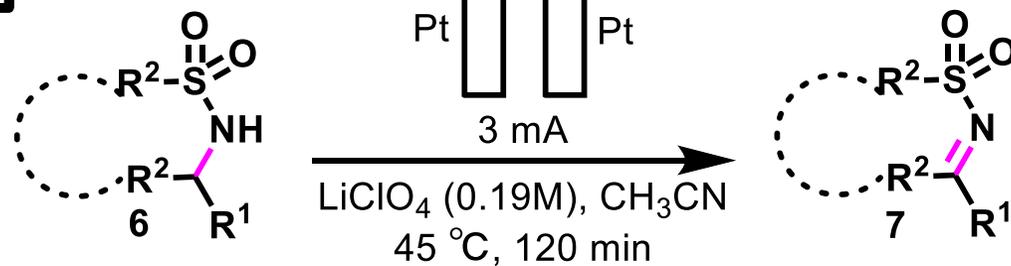
電解質 (M), CH₃CN
反応温度, 反応時間



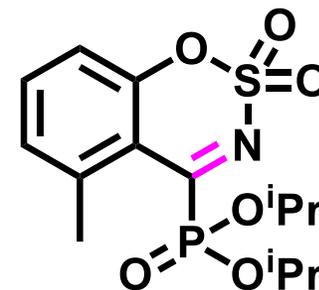
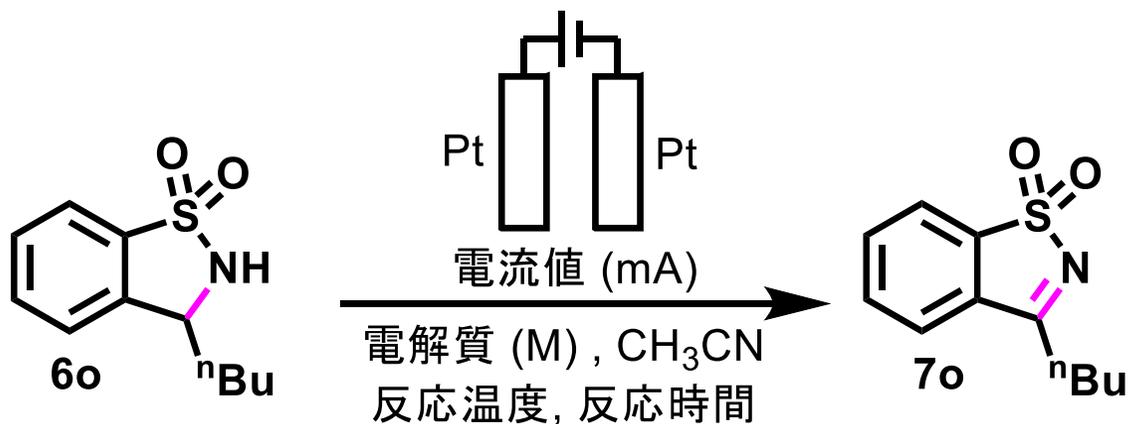
entry	電流値 (mA)	1a (mM)	電解質 (M)	温度 (°C)	時間 (min)	2aの収率% (NMR)
1	1	10	0.05	60	180	8
2	2	20	0.2	25	60	16
3	3	10	0.1	40	120	65
4	4	15	0.1	40	60	38
5	5	5	0.05	25	120	26

電解質 : LiClO₄ (過塩素酸リチウム)

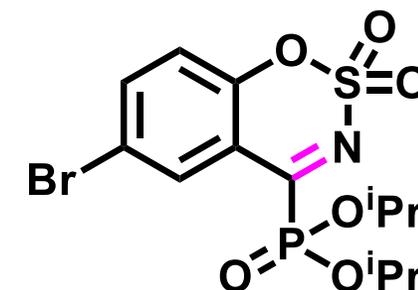
基質一般性の検討 (電解反応)



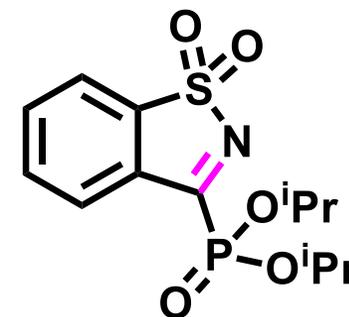
ベイズ最適化を活用する最適反応条件の再探索



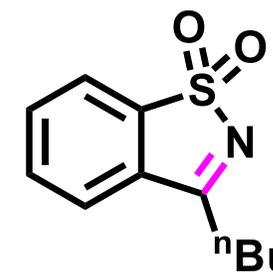
7b: **86%** (61%)



7g: **75%** (56%)



7k: **71%** (62%)



7o: **66%** (58%)

entry	電流値 (mA)	6o (mM)	電解質 (M)	温度 (°C)	時間 (min)	7oの収率% (NMR)
1	1	10	0.05	60	180	46
2	2	20	0.2	25	60	22
3	3	10	0.1	40	120	62
4	4	15	0.1	40	60	53
5	5	5	0.05	25	120	14
6	4	10	0.12	60	180	29
7	3	11.7	0.22	40	120	67 (66%単離収率)
前回の条件	3	10.4	0.19	45	120	58

電解質 : LiClO₄ (過塩素酸リチウム)

社会実装研究: ベイズ最適化の化成品合成への応用



協力: 旭化学工業(株)

論文投稿中

毎日新聞朝刊15面

(2022年11月10日)

化学工業日報4面

(2023年4月28日)



SM

(2 grams scale)

concentration in toluene (M)

1) NaSH aq. (eq.) and concentration in H₂O (M)
30 °C, 10 min.
2) Sulfur (eq.)
95 °C, 1 h, then 50 °C

3) TBAB (8.5 mol%)
95 °C, 24 h



spiro-dithiolane

2023年度より商業生産化

Entry	NaSH aq. (eq.)	concentration in H ₂ O (M)	Sulfur (eq.)	SM concentration in toluene (M)	Yield (GC %)
1	6.5	14.8	2.4	1.19	65
2	4.5	7.4	2.4	2.78	51
3	5.5	14.8	3.0	0.76	39
4	5.5	7.4	1.2	1.19	76
5	6.5	9.8	2.0	2.78	66
6	4.5	9.8	1.2	0.76	52

ITS

Prediction with ITS (entries 1 to 6)

experimental results

7 5.8 8.1 1.4 1.22 **78**

Champion data in company (none-BO screening): 2 grams scale

6.0 11.9 3.0 1.71 68

Entry	NaSH aq. (eq.)	concentration in H ₂ O (M)	Sulfur (eq.)	SM concentration in toluene (M)	Yield (GC %)
1	5.5	5.3	2.1	4.54	67
2	2.7	7.0	3.3	2.16	31
3	6.9	18.4	2.7	1.42	68
4	8.3	4.3	0.9	0.84	48
5	4.1	10.0	1.5	1.06	66

ITS

Prediction with ITS (entries 1 to 5)

experimental results

6 6.7 16.2 2.7 1.50 70

Prediction with ITS (entries 1 to 6)

experimental results

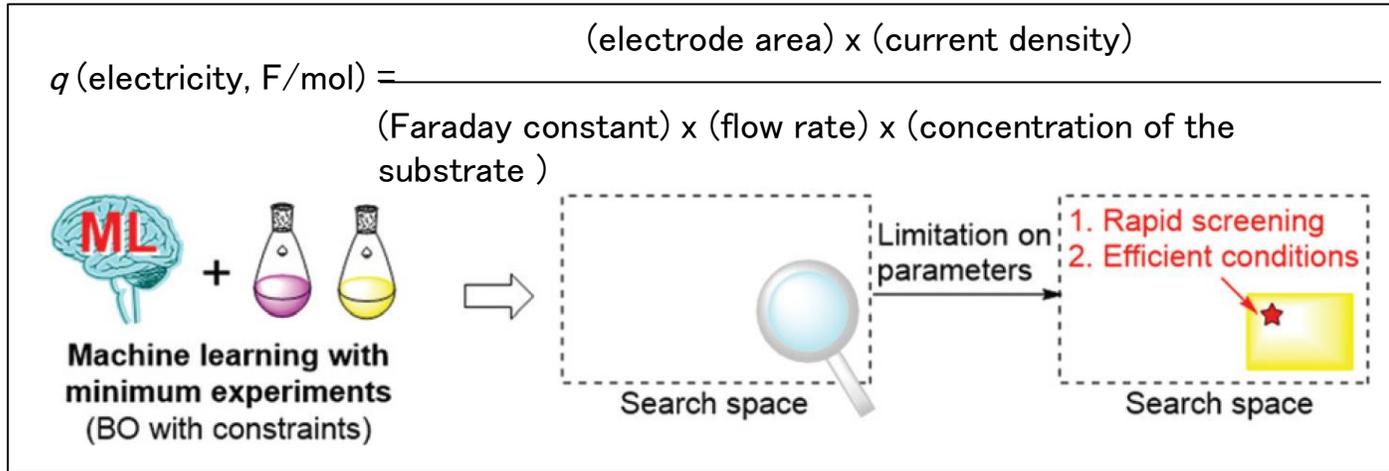
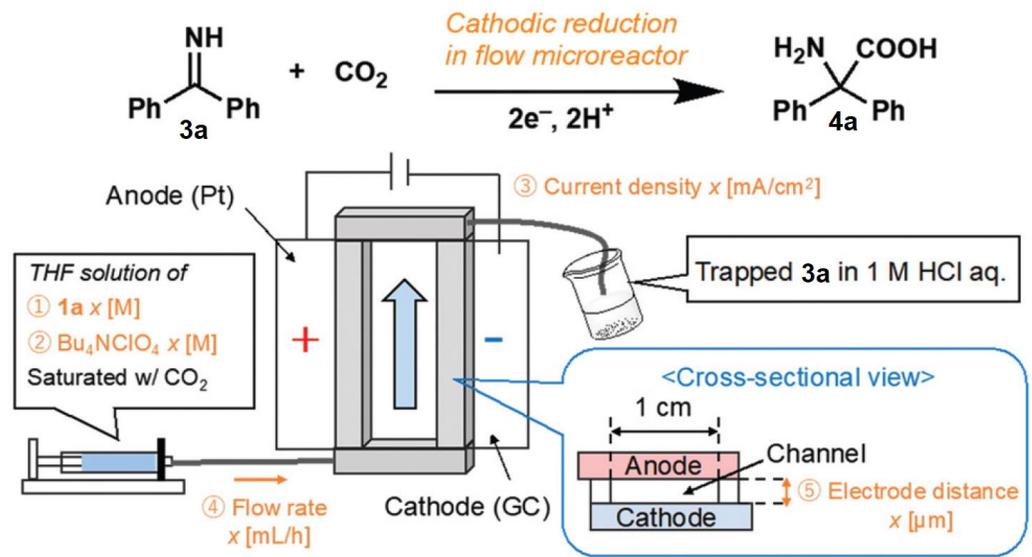
7 6.5 29.8 1.3 4.54 **89 (87)**

ITS prepared by random sampling

20 grams scale

ITS prepared by Latin hypercube sampling (LHS)

制約付きベイズ最適化によるアミノ酸フロー電解合成の反応条件最適化



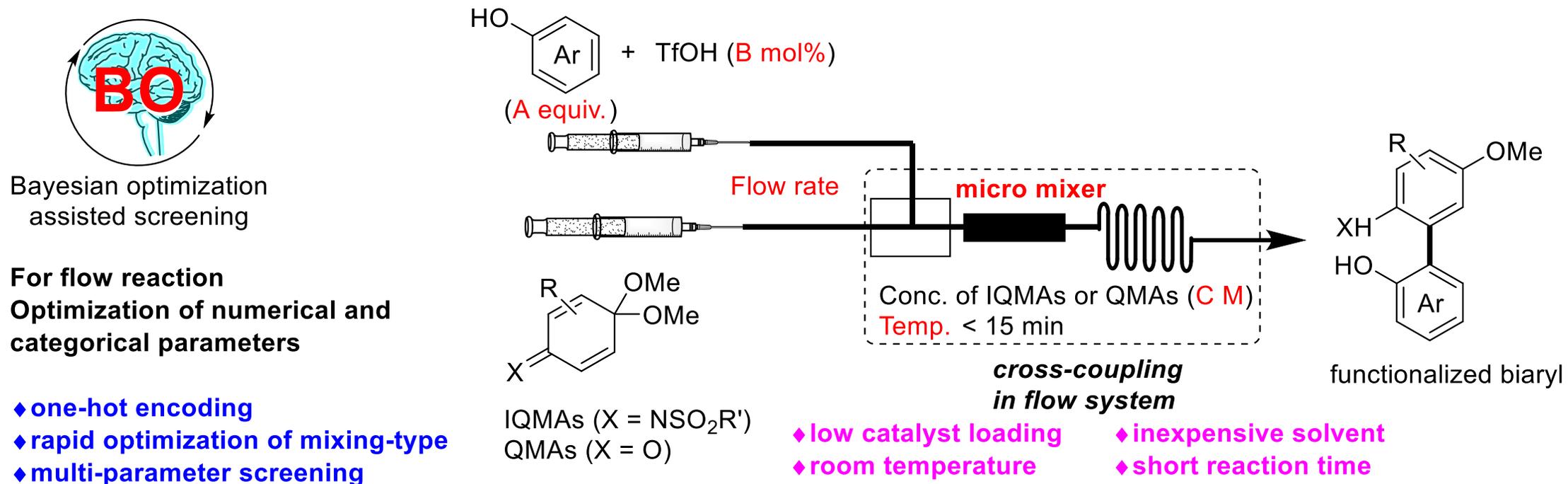
Entry	[1a] / M	[Bu ₄ NClO ₄] / M	Current density / mA cm ⁻²	Flow rate / mL h ⁻¹	Electrode distance / μm	Charge Passed (q) / F mol ⁻¹	Yield of 4a (%)	Current efficiency (%)
Training data	1	0.08	0.1	25	15	2.33	28	24
	2	0.08	0.15	40	25	2.24	31	28
	3	0.04	0.1	10	10	2.80	88	28
	4	0.04	0.05	25	25	2.80	39	33
	5	0.12	0.15	15	5	2.80	46	63
2.0 < q < 3.0	6A	0.049	0.1	11.1	8.9	2.85	90	52
	7A	0.062	0.075	12.2	7.4	2.98	78	57
	8A	0.045	0.12	10.1	9.2	2.73	78	58
	9A	0.043	0.077	11.9	11.1	2.79	81	24
2.0 < q < 2.1	6B	0.04	0.084	10.6	14.6	2.03	85	84
	7B	0.071	0.053	10.6	8.3	2.01	71	71
	8B	0.06	0.15	11.7	10.8	2.02	77	76
	9B	0.042	0.094	12.2	16	2.03	88	87



最少実験データ(≦15)による機械学習合成→大幅コスト削減

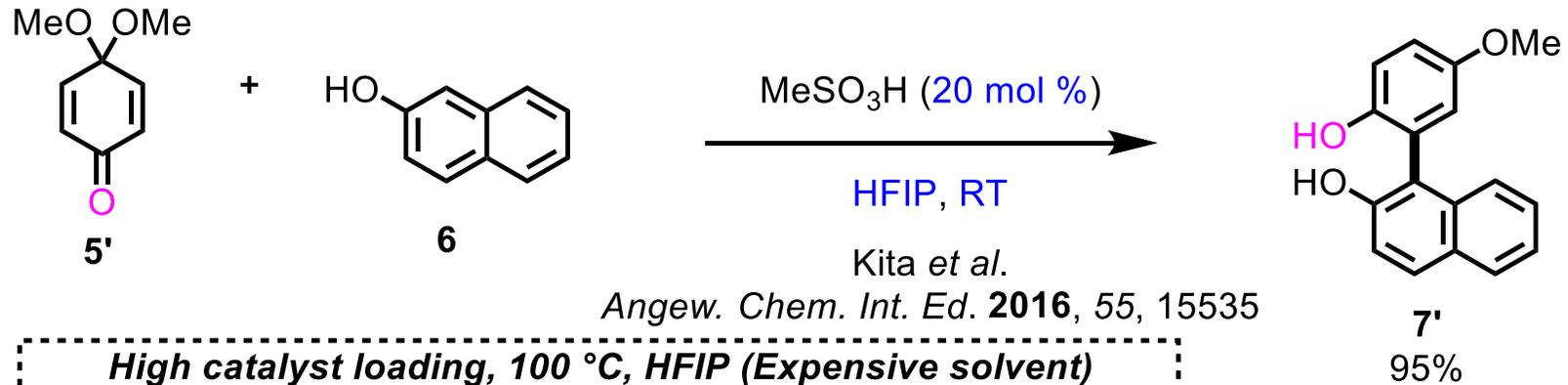
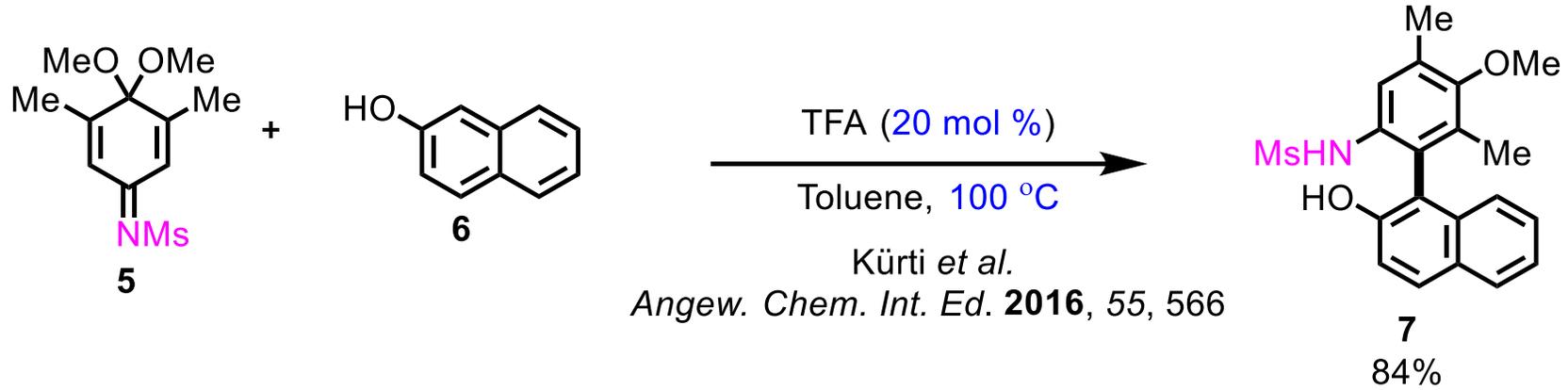
1. 最少実験データとガウス過程回帰によるフロー反応条件(2次元パラメータ)の迅速最適化 -予測収率の可視化-
不斉フロー合成: *Chem. Commun.* **2020**, 56, 1259 (Highlighted in *Synfacts*, *EurekaAlert!*)
フロー電解合成: *J. Org. Chem.* **2021**, 86, 16035 (菅誠治教授@岡山大学との共同研究)

2. ベイズ最適化による電解ケチミン合成反応条件(5次元パラメータ)の迅速最適化 *Green Chem.* **2021**, 23, 5825

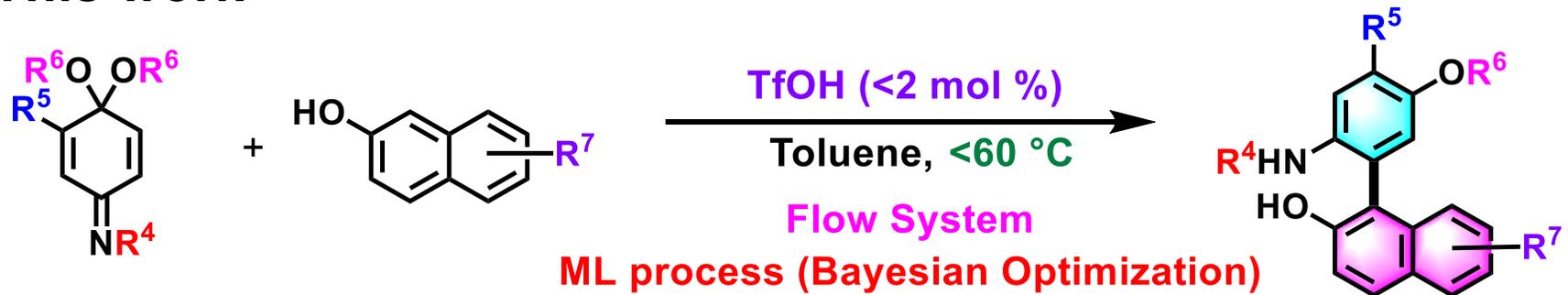


3. 並列ベイズ最適化によるビアリール化合物のフロー合成反応条件(ミキサー選択を含む6次元パラメータ)の迅速最適化
Commun. Chem. **2022**, 5, 148

有機酸触媒を用いるキノンアセタール原料からのビアリール合成

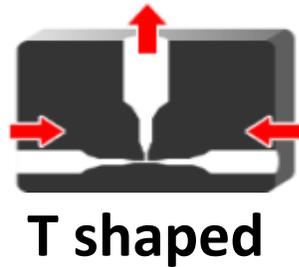


This work

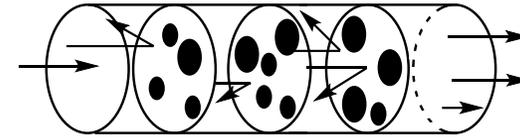


ミキサー (カテゴリカルパラメータ)を含む6次元パラメータの迅速最適化

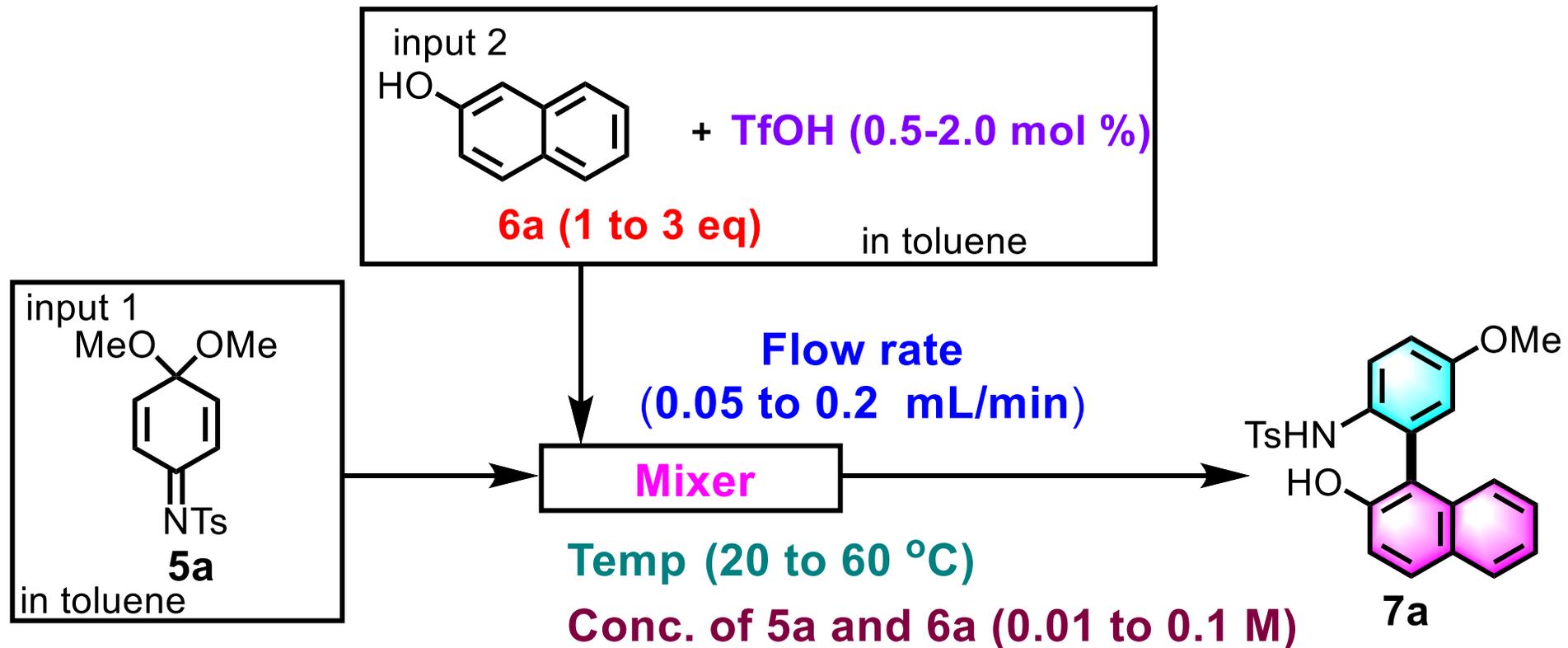
Categorical Parameters: Types of micro mixers



β -type

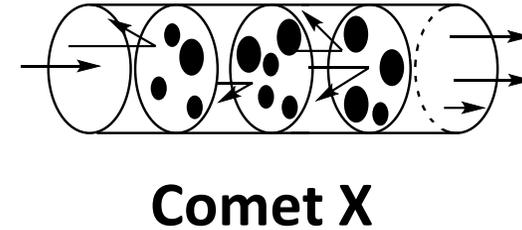
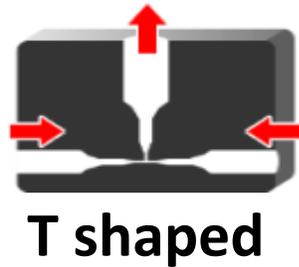


Comet X



ミキサー (カテゴリカルパラメータ)を含む6次元パラメータの迅速最適化

Categorical Parameters: Types of micro mixers



Mixer type	T shaped	β -type	Comet X
OrdinalEncoder	1	2	3
OneHotEncoder	[1.0.0]	[0.1.0]	[0.0.1]
BinaryEncoder	[1.0]	[0.1]	[1.1]

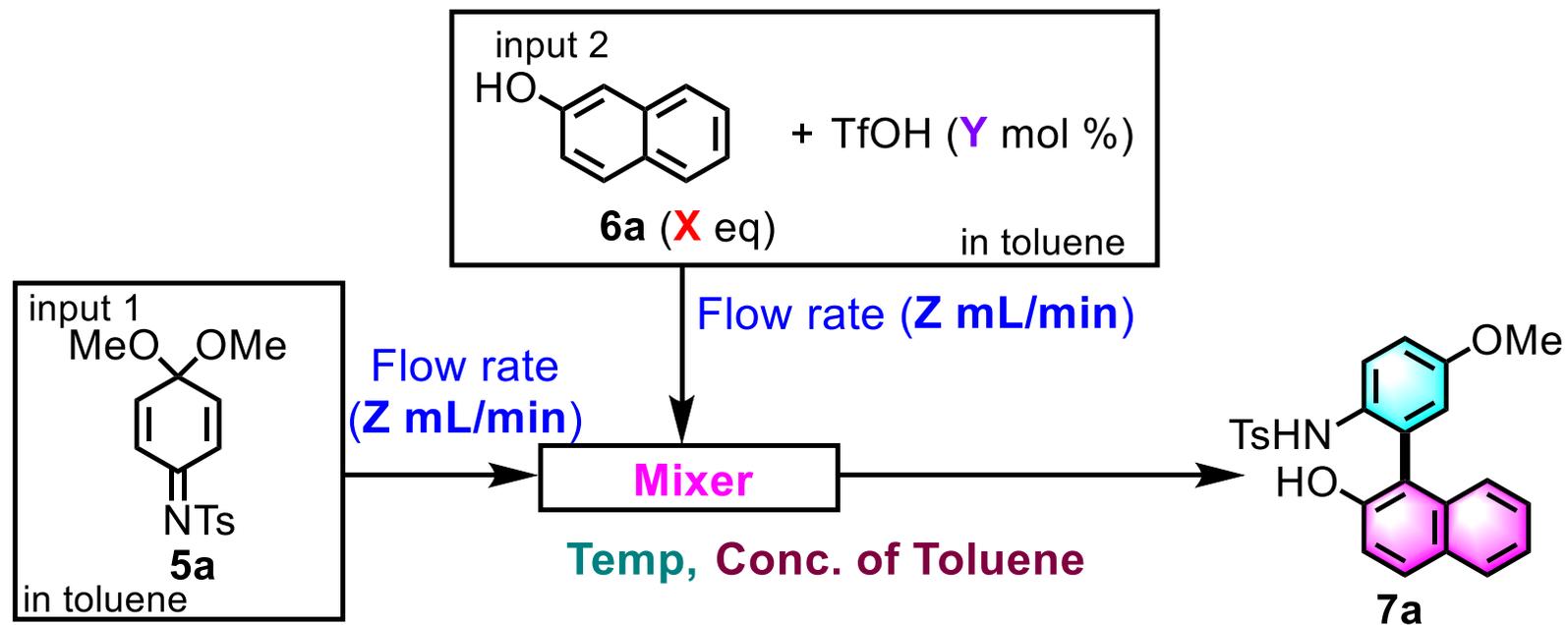
Single EI (一つずつ実験・評価)

➔ Parallel LCB (三つずつ実験・評価) 時間の節約

Acquisition Function

- Probability of Improvement (PI)
改善確率を指標 ➔ 活用重視
- Expected Improvement (EI)
改善確率と改善量を指標
- Lower Confidence Bound (LCB)
探索と活用のトレードオフ

カテゴリカルパラメータを含むベイズ最適化 : Initial Training Set

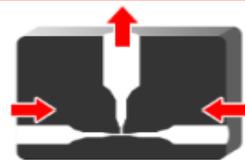


Entry	Mixer	X Equiv of 6a	Temp. (°C)	Conc. (M)	Flow rate (Z mLmin ⁻¹)	TfOH (Y mol %)	Yield 7a (%)
1)	Comet X	2	60	0.05	0.05	1	68
2)	Comet X	3	40	0.01	0.2	0.5	73
3)	β -type	1	60	0.01	0.1	2	42
4)	β -type	3	20	0.1	0.1	0.5	28
5)	T-shaped	1	40	0.05	0.05	2	55
6)	T-shaped	2	20	0.1	0.2	1	75

カテゴリカルパラメータを含むベイズ最適化

BO was applied for initial training data set

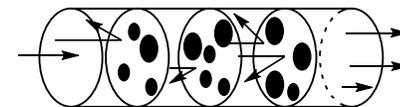
Entry	Mixer	X Equiv of 6a	Temp. (°C)	Conc. (M)	Flow rate (Z mLmin ⁻¹)	TfOH (Y mol %)	Yield 7a (%)
1)	T-shaped	2.2	65	0.039	0.15	11	68
2)	Comet X	1.3	45	0.15	0.21	2.4	46
3)	β-type	1	60	0.04	0.1	7	17



T shaped

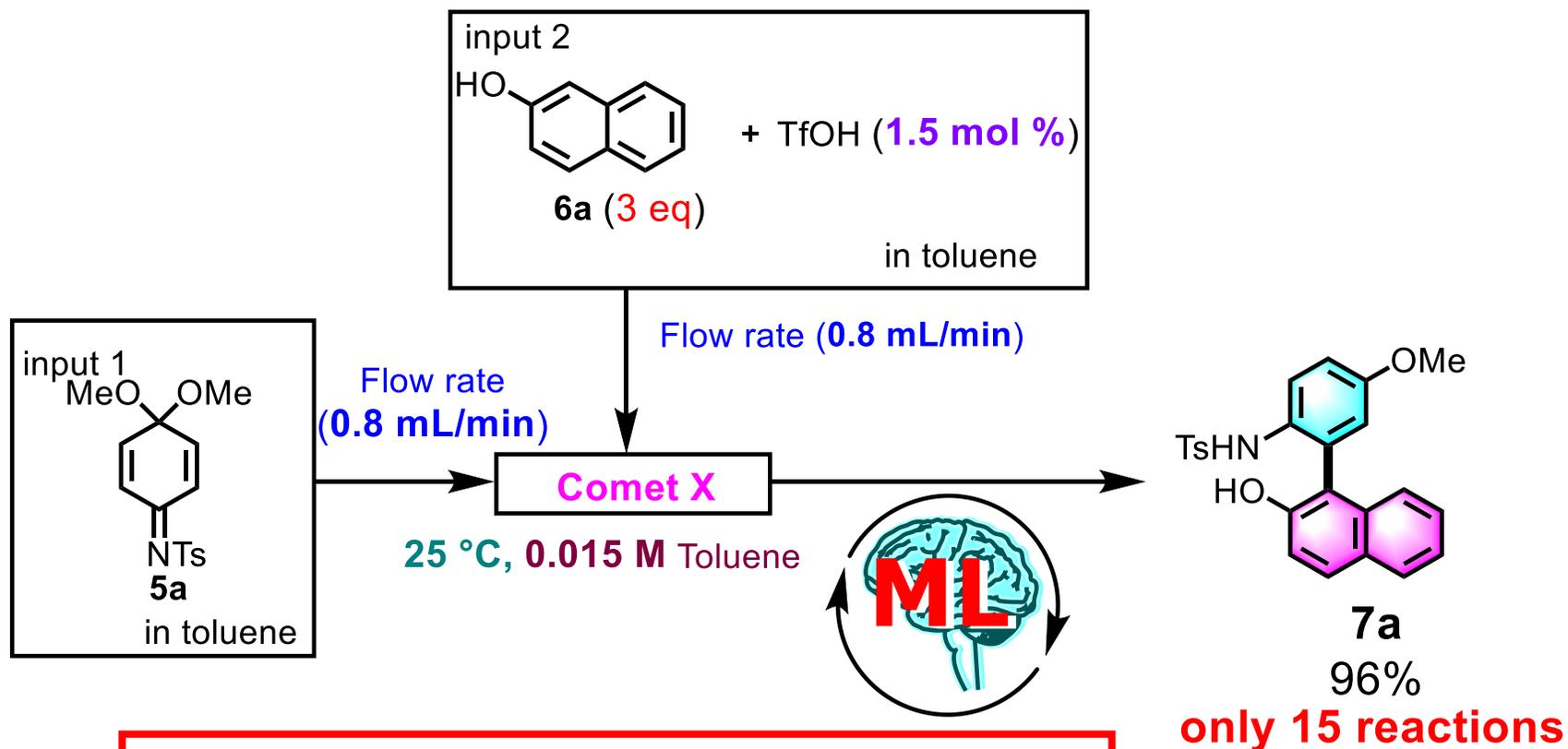


β-type



Comet X

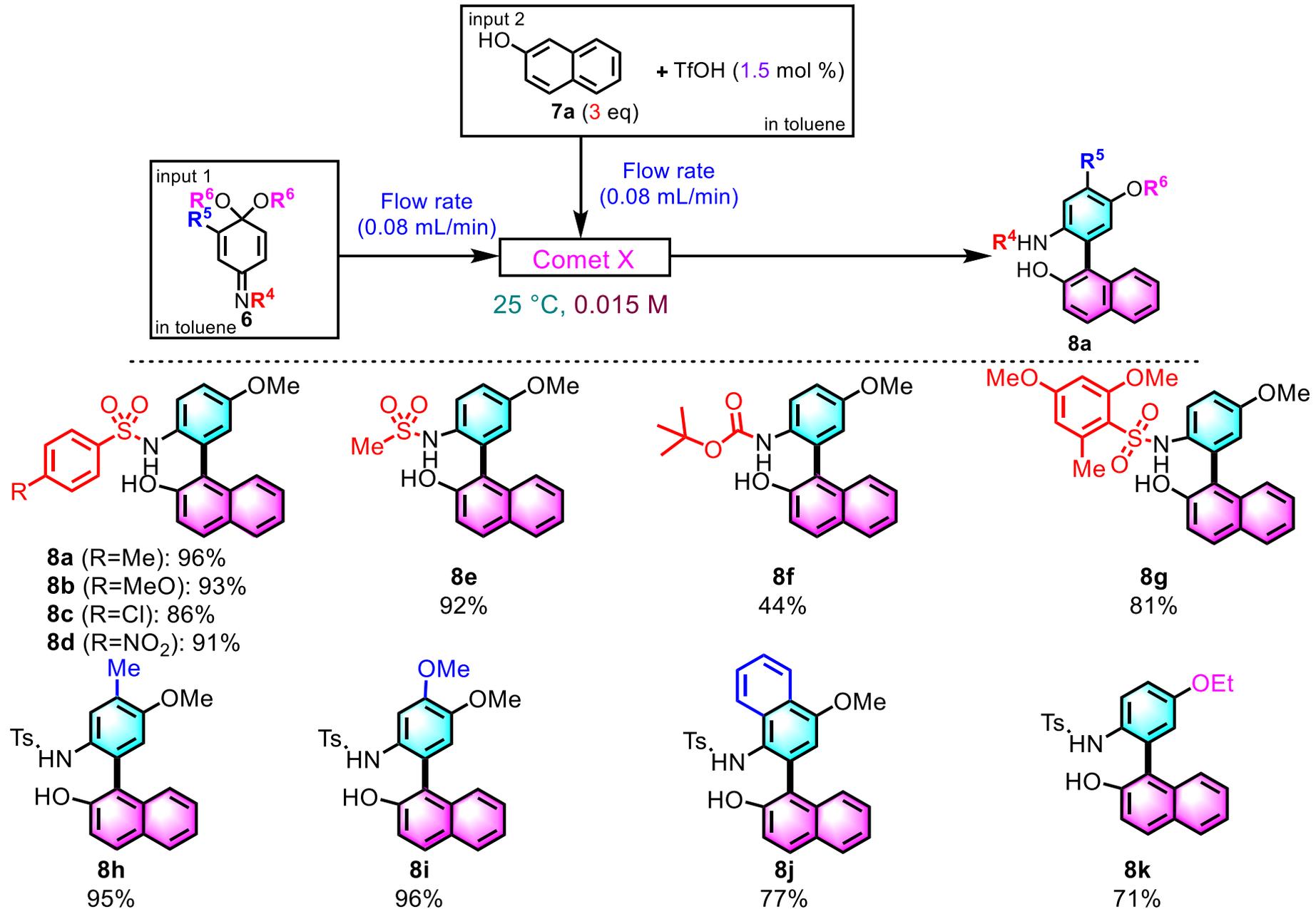
カテゴリカルパラメータを含むベイズ最適化の結果



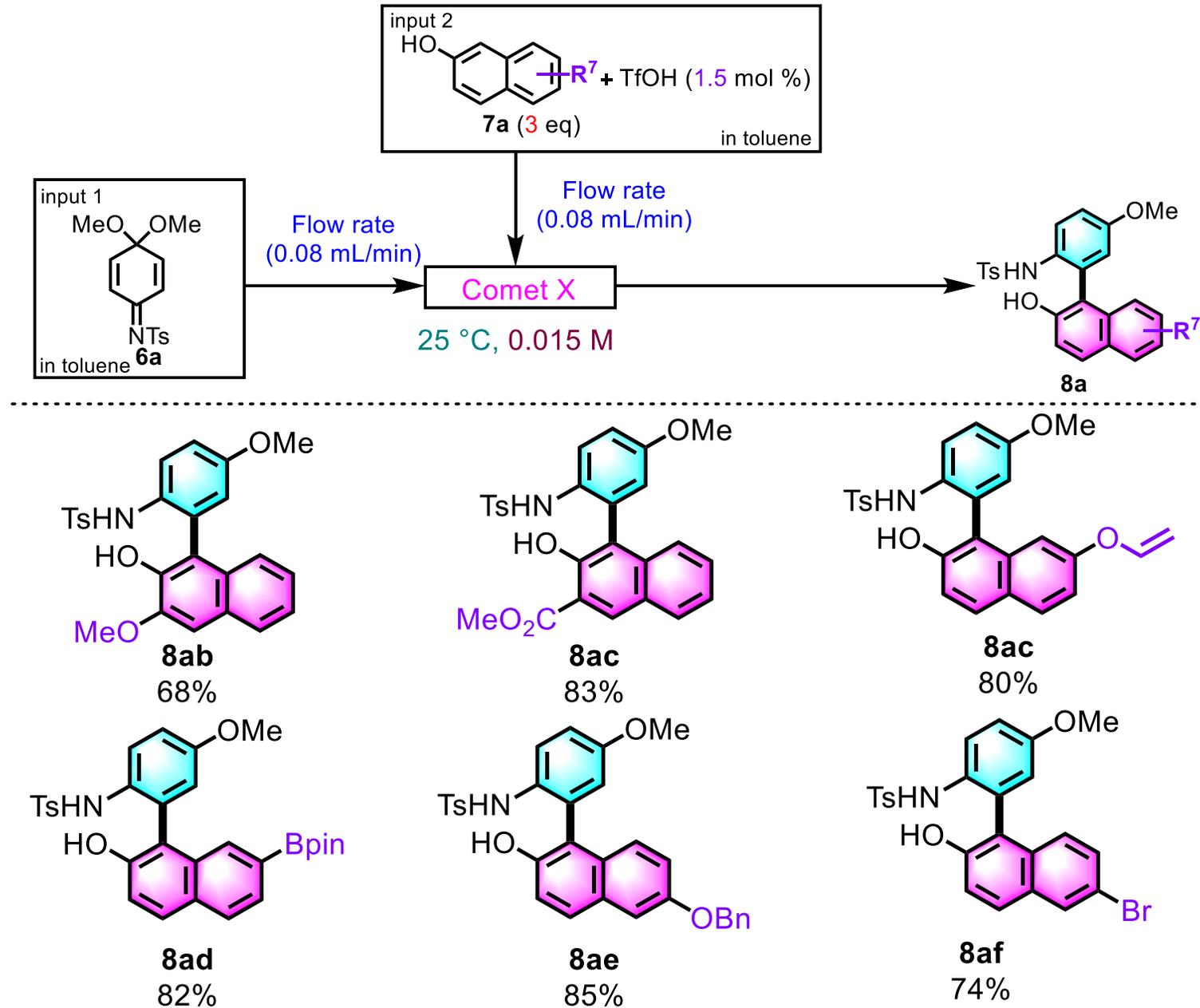
ML Predicted results:

- Mixer = **Comet X**
- Flow rate = **0.8 mL/min**
- Catalyst loading = **1.5 mol %**
- Temperature = **25 °C**
- Concentration = **0.015 M**
- Amount of reactant = **3.0 eq**

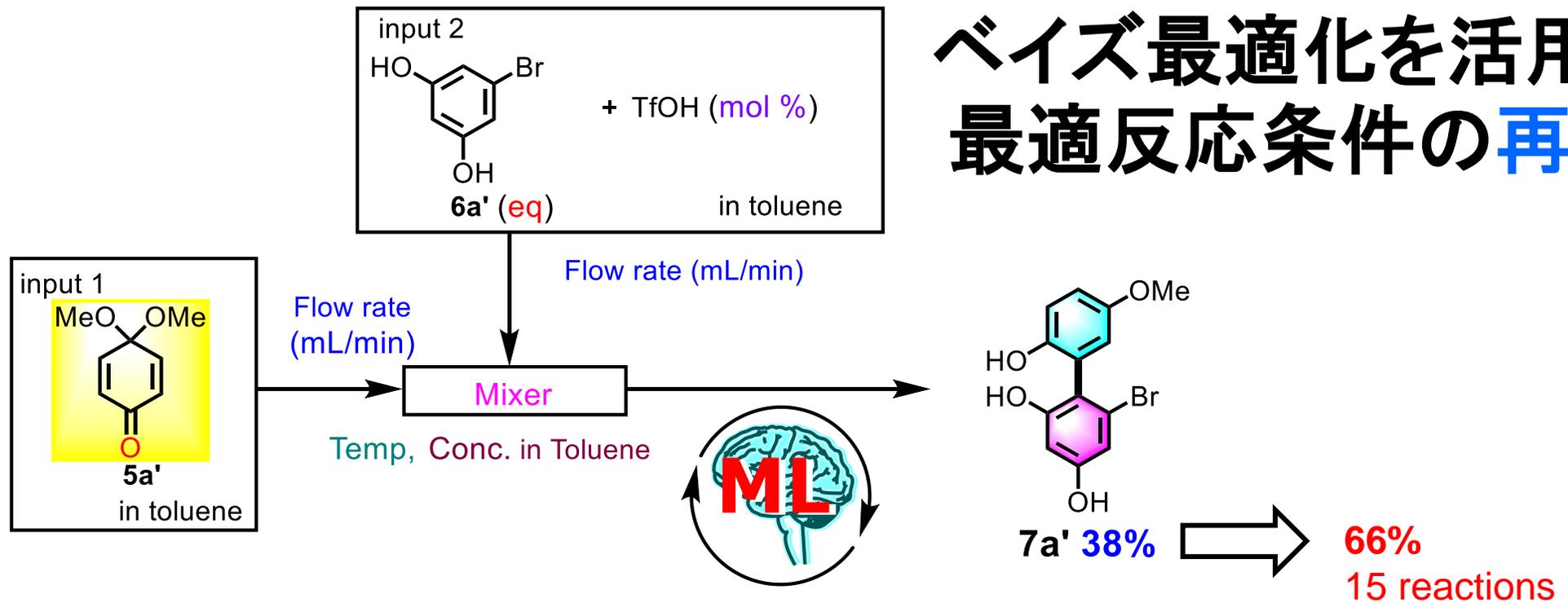
基質一般性①



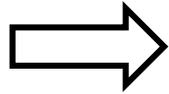
基質一般性②



ベイズ最適化を活用する 最適反応条件の再探索

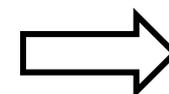


Flow rate:
0.8 mL/min



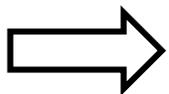
Flow rate:
0.068 mL/min

Mixer:
Comet X



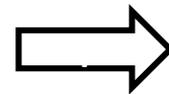
Mixer:
β-type

Stoichiometry:
3 eq.



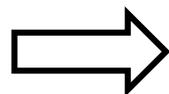
Stoichiometry:
3.2 eq.

Concentration:
0.015M



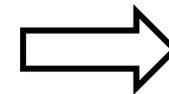
Concentration:
0.044M

Temperature:
25 °C

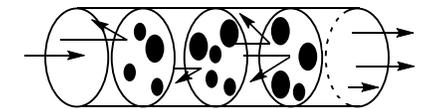


Temperature:
30 °C

TfOH:
1.5 mol %



TfOH:
0.35 mol %



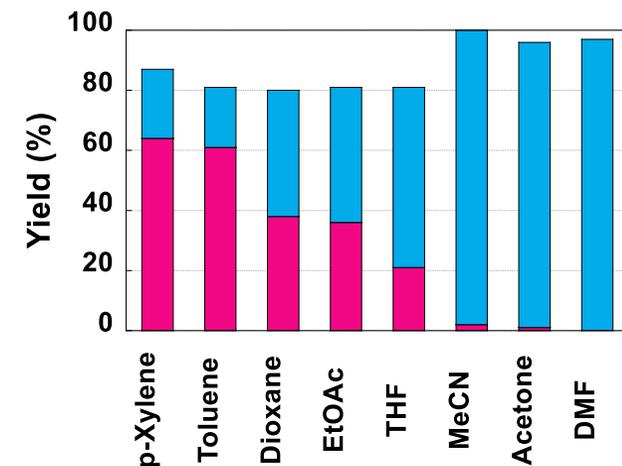
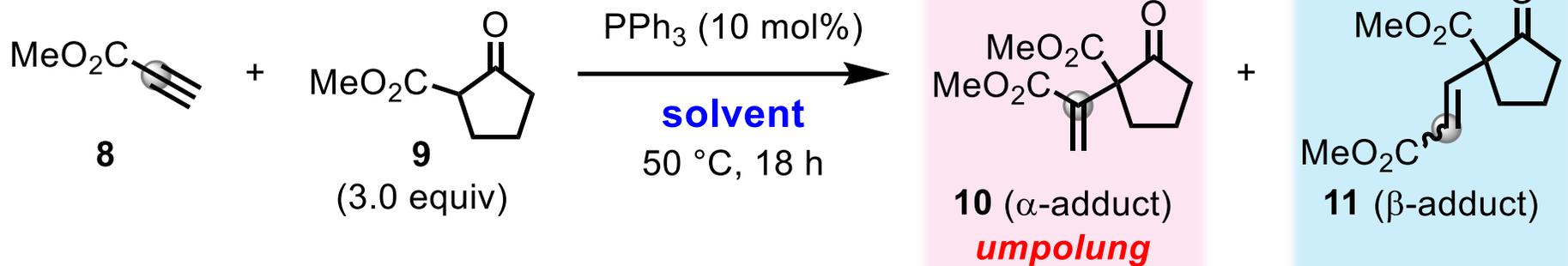
Comet X



β-type

今後の展望

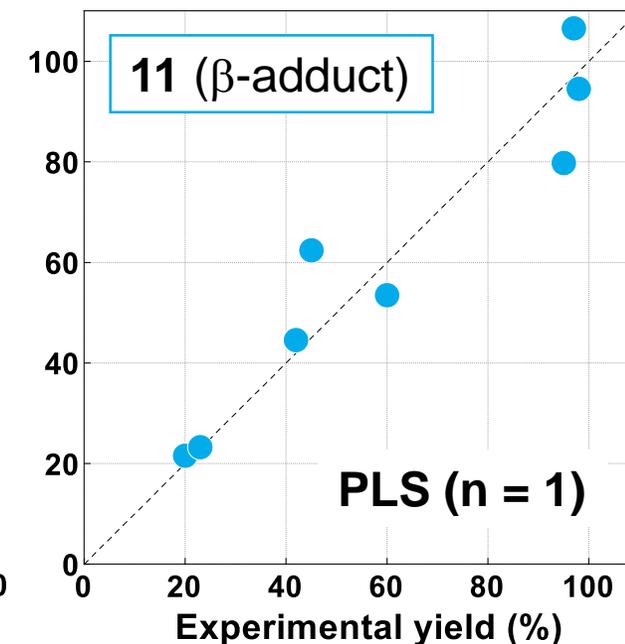
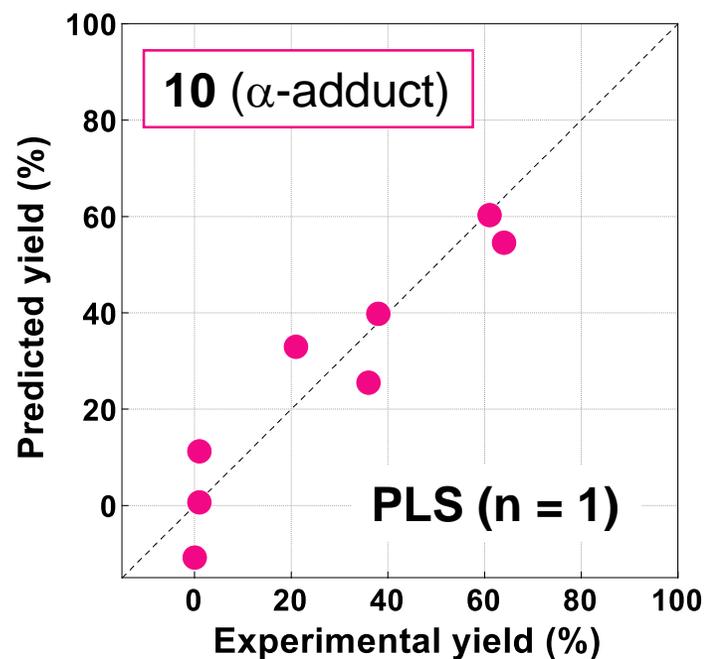
極性転換型触媒反応における溶媒効果説明モデルの構築



説明変数

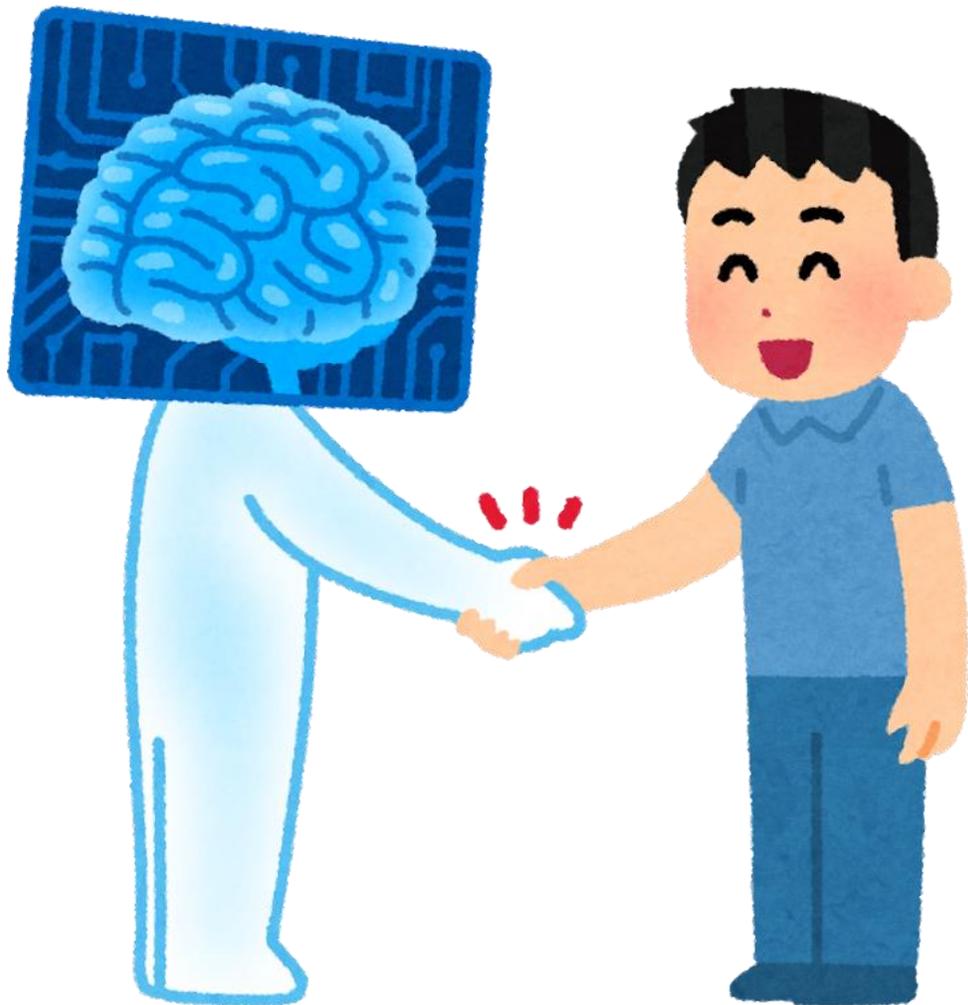
① ϵ	誘電率	④ δ_D	分散力
② DN	ルイス塩基性	⑤ δ_P	分極
③ AN	ルイス酸性	⑥ δ_H	水素結合

PLS, LASSO, Ridge, Random Forest



- ✓ 溶媒効果の支配因子予測
- ✓ 未実施溶媒の収率予測
- ✓ 混合溶媒系の提案
- ✓ 反応機構や選択性の考察

最少実験データ(≦15)による機械学習⇒大幅なコスト削減



医薬資源の自動合成を志向した 機械学習による反応条件最適化

Machine-learning-assisted screening for
automated synthesis of pharmaceutical resources

滝澤 忍

大阪大学 産業科学研究所

分子システム創成化学研究分野 & 産業科学AIセンター

応用分野は、ライフサイエンス・創薬化学・精密有機合成反応プロセスです。
ファインケミカル関連企業様との連携・共同研究を希望しております！