

2021.05.26

Cypher Society 5.0 研究支援プログラム 成果発表

機械学習を利用したマイクロフロー電解合成の効率的最適化

佐藤英祐

岡山大学学術研究院自然科学学域

合成プロセス化学研究室(菅誠治 教授) 助教

自己紹介

佐藤英祐

岡山大学学術研究院自然科学学域 助教

合成プロセス化学研究室 (菅誠治 教授)

専門：有機合成化学

博士課程：天然有機化合物の全合成と生物活性

ポスドク：陽極酸化を用いた天然有機化合物の全合成

現在：有機電解合成・電解マイクロフロー→ここに機械学習を取り入れたい

略歴：

2018年3月 慶応義塾大学大学院理工学研究科修了

博士(理学) 末永聖武教授

2017年4月 日本学術振興会特別研究員DC2

2018年4月 日本学術振興会特別研究員PD (学位取得に伴う資格変更)

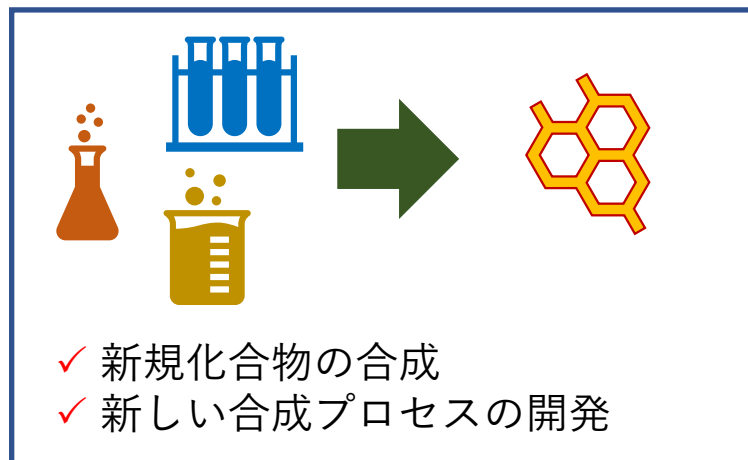
2018年7月 Alexander von Humboldt財団博士研究員

Johannes Gutenberg-University Mainz (Prof. Dr. Till Opatz)

2020年4月 岡山大学大学院自然科学研究科 助教

2021年4月 現職 (組織改編による名称変更)

有機合成化学と機械学習



有機合成化学×人工知能(特に機械学習)

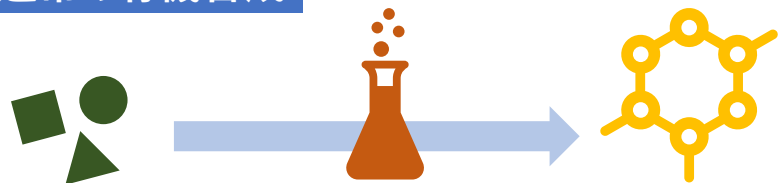
「データ駆動型有機合成化学」

あくまで実験化学者の立場から、
有機合成化学をさらに発展させる手段として
機械学習を効率的に利用していきたい。

従来型の網羅的な条件探索、化合物探索から脱却
⇒ 効率的な反応条件探索、膨大な数のパラメータの合理的最適化

有機合成化学と機械学習

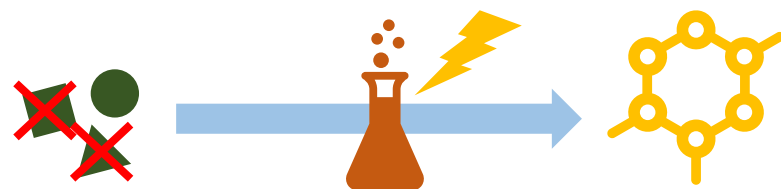
通常の有機合成



フラスコに出発原料と試薬を入れ混合
反応停止、後処理、単離精製、構造解析

- × 一回の実験に時間がかかる
- × 大量の廃棄物が生じてしまう

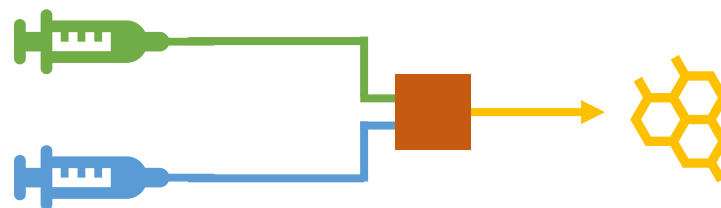
電解有機合成



電気の力を駆動力とする化学変換

- ✓ 廃棄物の少ない環境低負荷な手法
- × 電極表面上のみで化学反応が進行する

マイクロフロー合成



狭い合成装置(マイクロリアクタ)で試薬を混合
断続的な溶液供給により大量合成が簡単

マイクロフロー電解有機合成

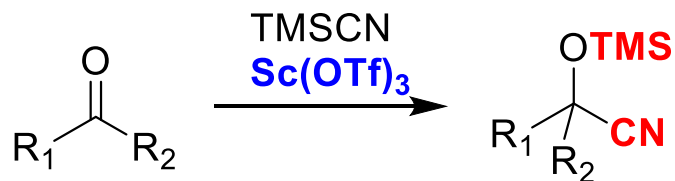
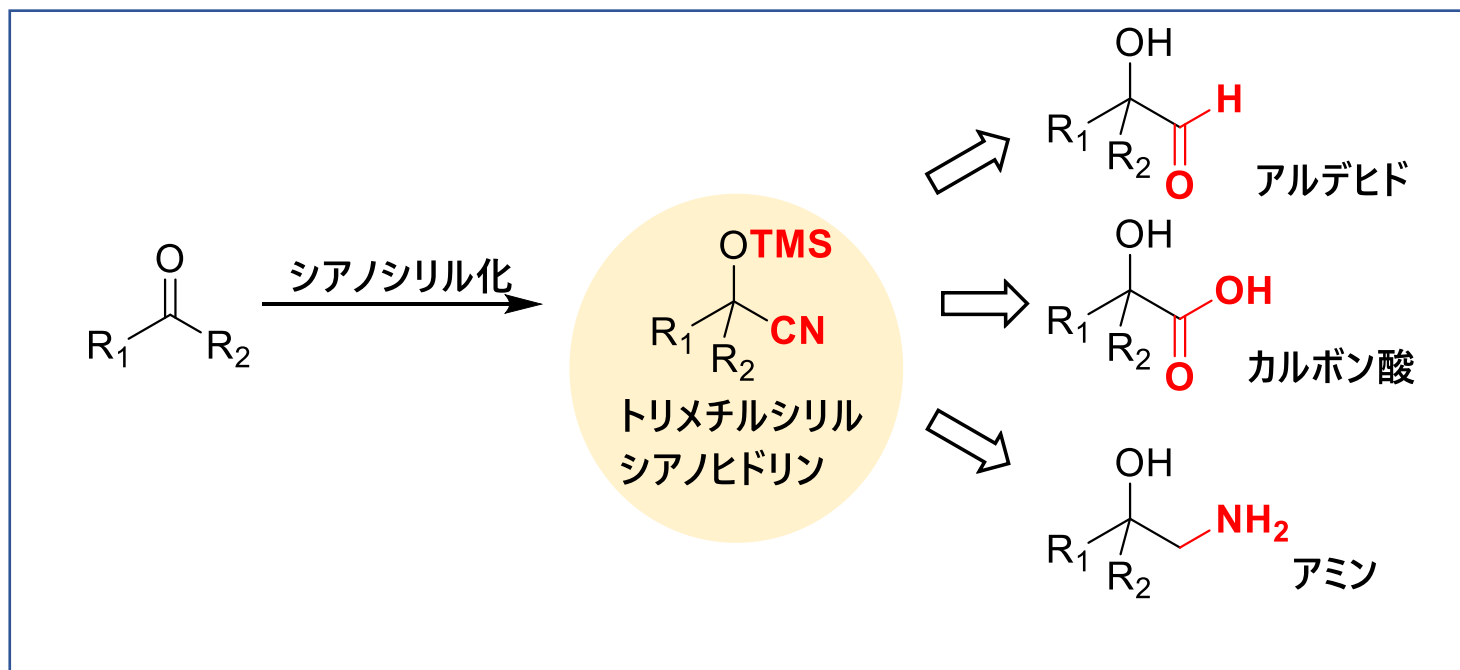


電解合成とフロー合成の組み合わせ

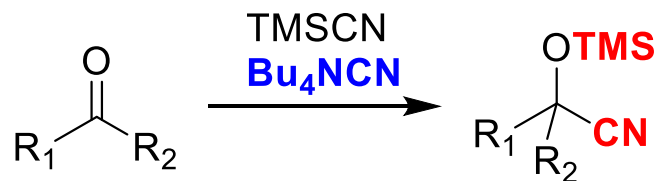
- ✓ 電解合成の弱点を克服
- × パラメータの種類が膨大

➡ 機械学習を用いた合理的な最適化

カルボニル化合物のシアノシリル化反応



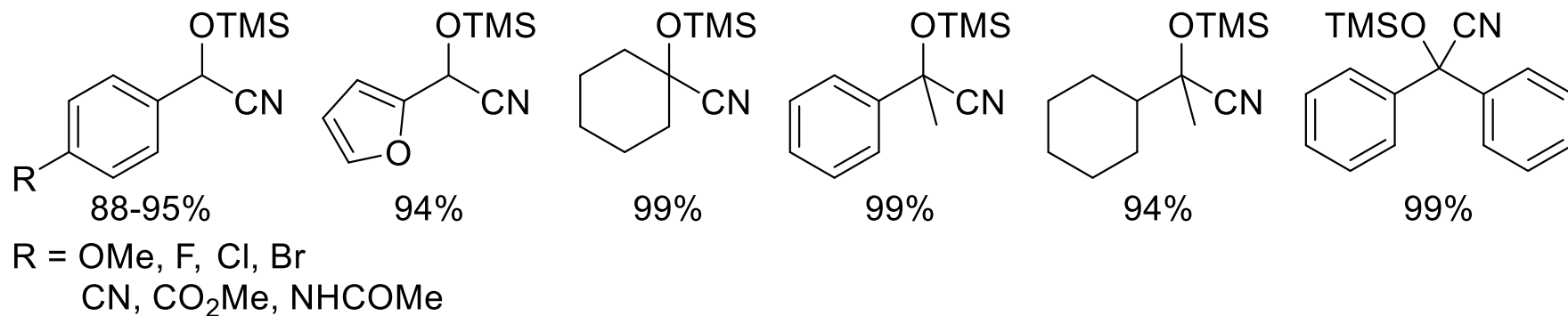
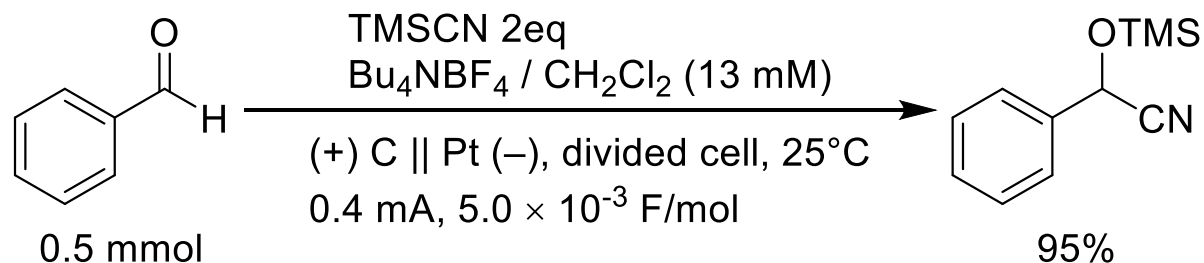
Lee, S. *et al. Green Chem.* **2009**, *11*, 946-948.



Plumet, J. *et al. Tetrahedron* **2004**, *60*, 10521-10524.

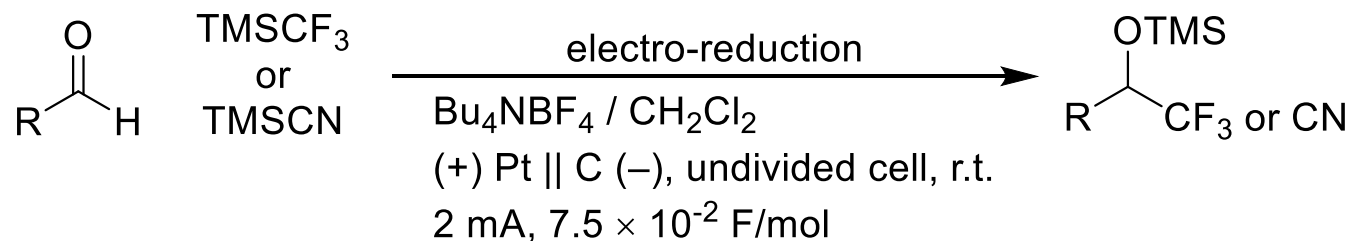
Evans, D. A. *et al. Tetrahedron Lett.* **1973**, *49*, 4929-4932.

カルボニル化合物のシアノシリル化反応 (Batch反応系)



分離型電解セル

Yuan's work

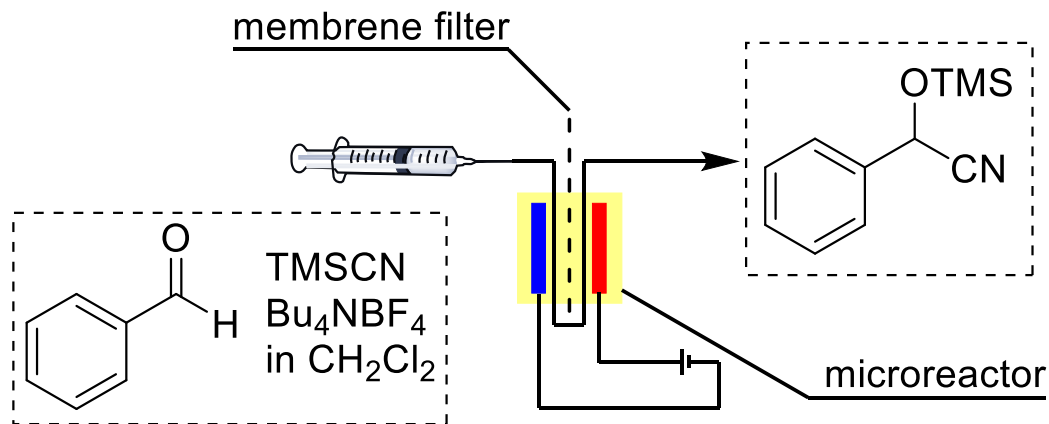


Chem. Commun. **2020**, 56, 2435.

カルボニル化合物のシアノシリル化反応(Flow反応系)



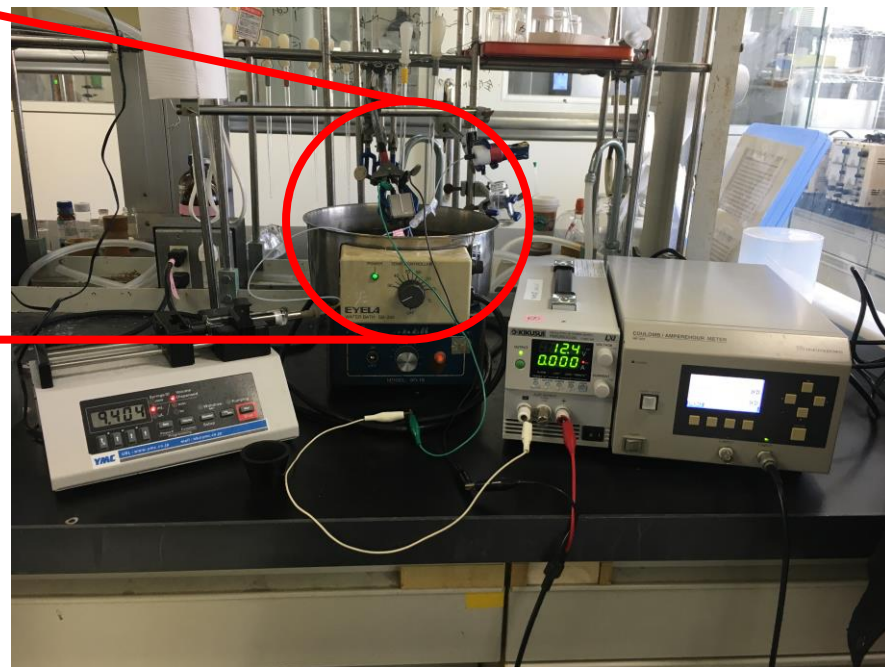
分離型電解セル



マイクロフローセル

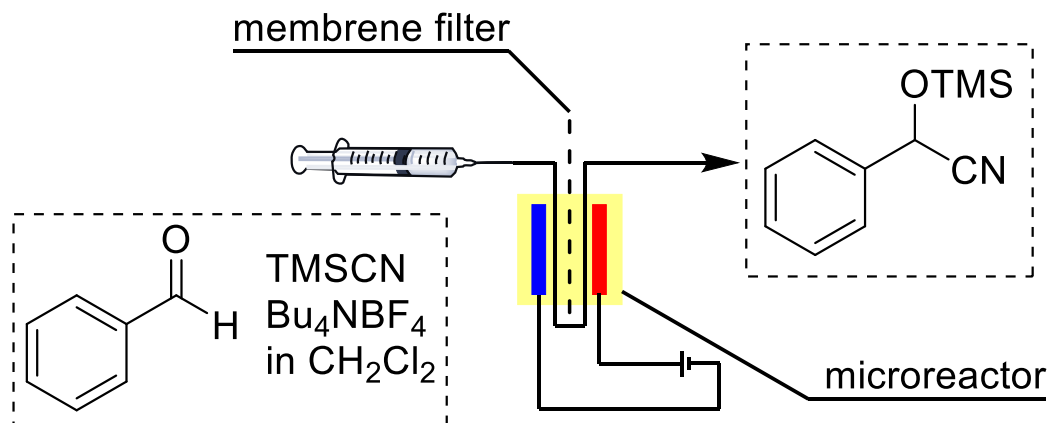


反応装置拡大

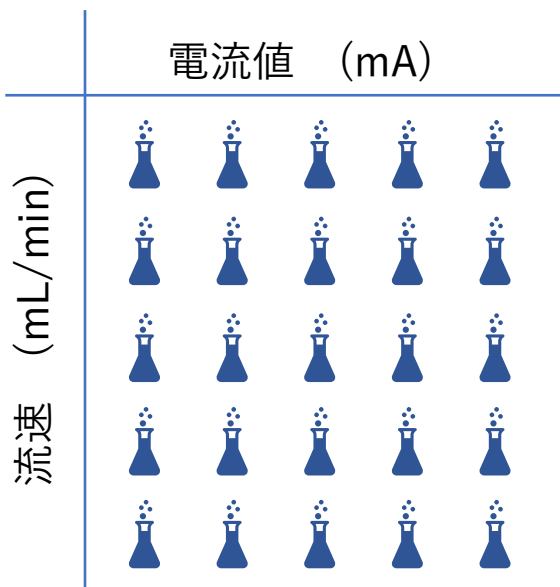


反応系の全体図

ガウス過程回帰による反応条件の最適化

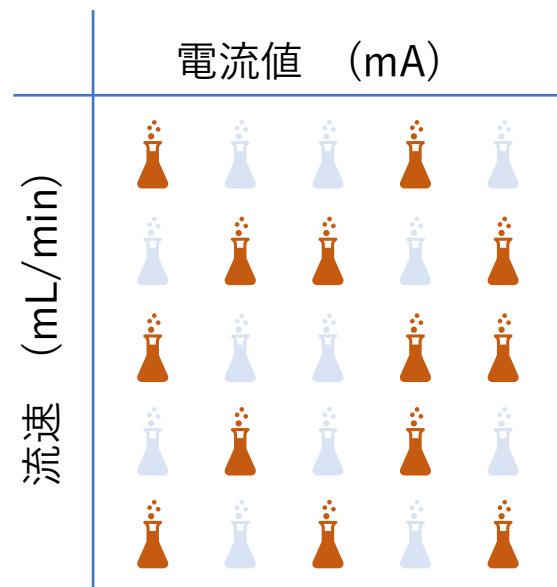


「流速」と「電流値」に関する最適な条件探索



網羅的なスクリーニング

機械学習の活用



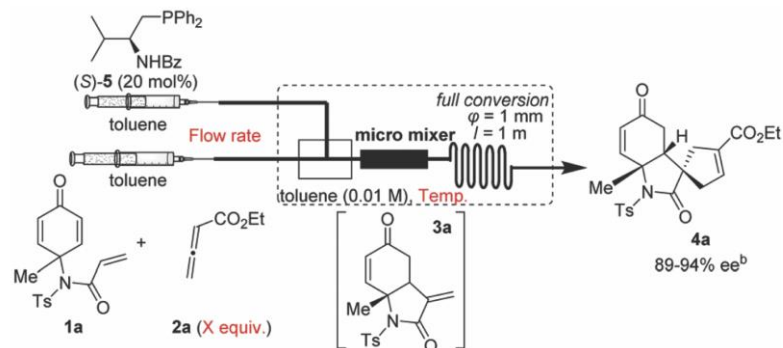
少ない実験量での最適条件探索

ガウス過程回帰

- ✓ 入力変数 x から出力変数 y への関数 $y=f(x)$ を推定するモデルの一つ
- ✓ ベイズ推定を用いることで、推定される関数が分布として得られる
- ✓ データ量が少ない系にも適用することができる

大阪大学産業科学技術研究所 (共同研究)

笹井宏明 教授、滝澤忍 准教授、近藤健 助教(現在は茨城大学)



“Exploration of flow reaction conditions using machine-learning for enantioselective organocatalyzed Rauhut–Currier and [3+2] annulation sequence”

Chem. Commun. **2020**, 56, 1259

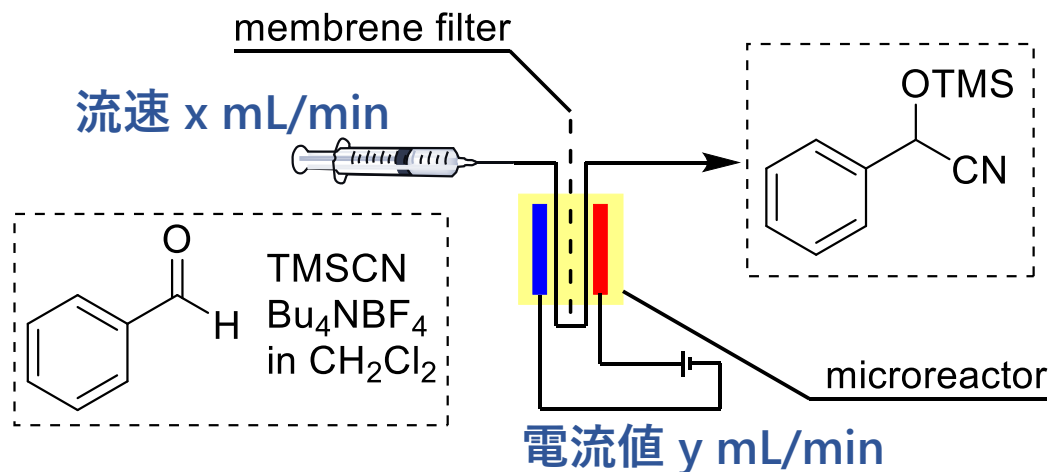
人工知能 (Artificial Intelligence)

機械学習 (Machine Learning)

深層学習 (Deep Learning)

判断や予測を行うモデルを
機械的に学習する

シアノシリル化反応の最大収率条件探索



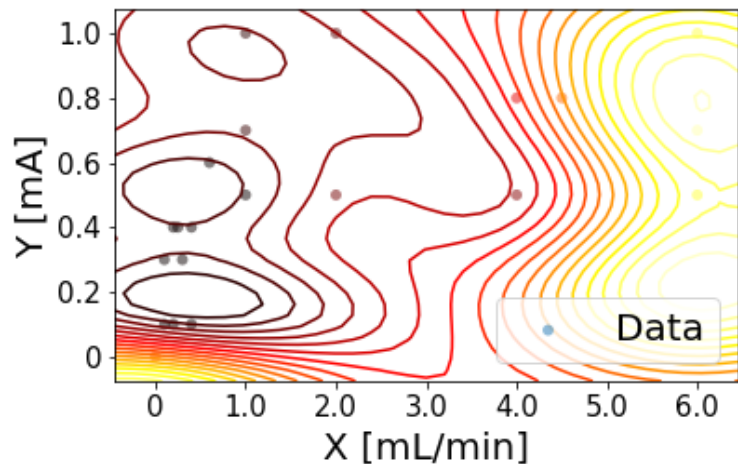
channel	flow_rate_ml_min	current_mA	yield of 1	yield of 2'	yield of 2
A	0	0	20	80	60
A	0.1	0.3	0	99	100
A	0.1	0.1	0	98	100
A	0.2	0.1	0	99	100
A	0.2	0.4	0	96	100
A	0.25	0.4	0	100	100
A	0.3	0.3	0	100	100
A	0.4	0.1	0	98	100
A	0.4	0.4	0	98	100
A	0.6	0.6	0	104	100
A	1	0.5	0	97	100
A	1	1	4	96	96
A	1	0.7	5	93	95
A	2	0.5	12	86	88
A	2	1	9	91	91
A	4	1	13	86	87
A	4	0.5	12	75	86
A	4	0.8	13	86	75
A	4.5	0.8	13	86	60
A	6	0.5	65	41	39
A	6	1	64	40	38
A	6	0.7	71	30	30
A	6	0.2	80	33	24

流速と電流値を少しずつ変更しながら
合計で22のデータポイントの取得



これらを用いたガウス過程回帰による
反応性予測モデルの構築

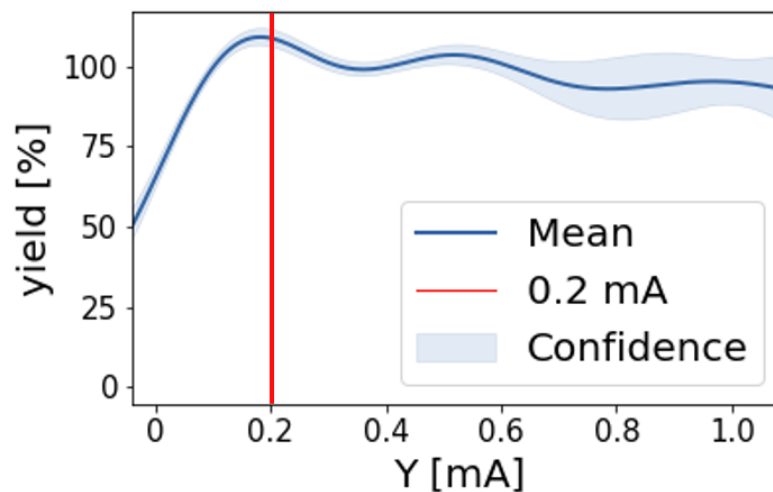
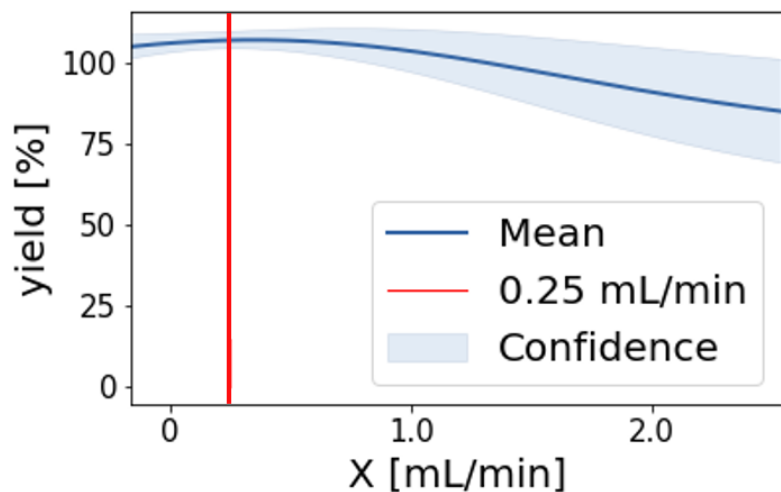
シアノシリル化反応の最大収率条件探索



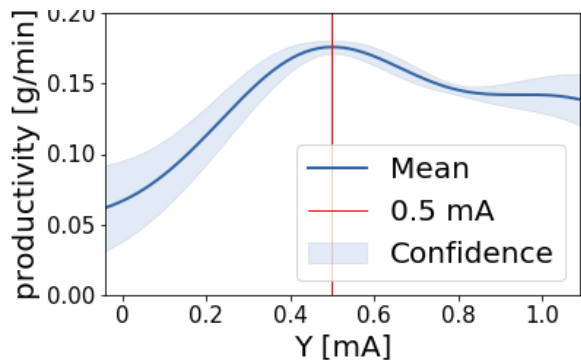
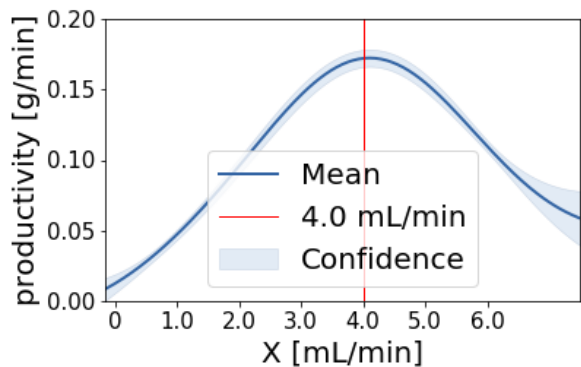
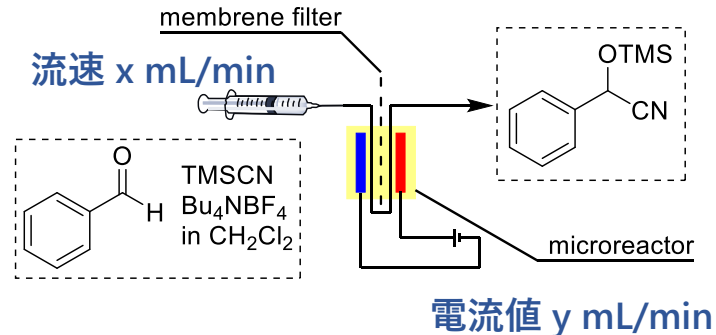
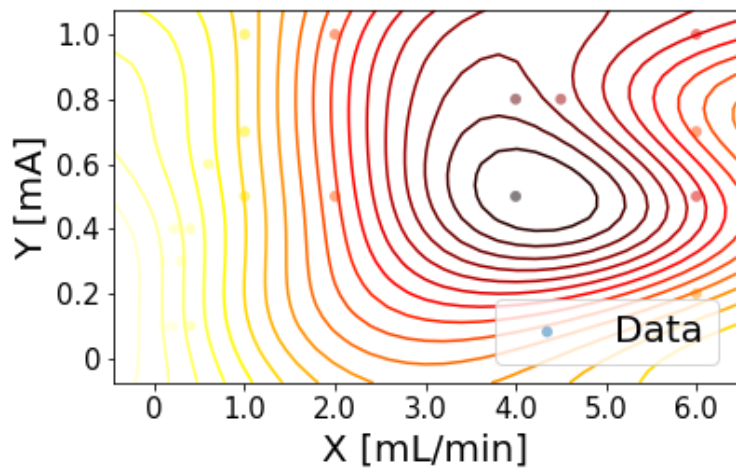
等高線図の色の濃い部分が「高収率」を示す反応条件



それぞれ下の断面図より
0.25 mL/min, 0.2 mAの時に
最高収率を予測した



シアノシリル化反応の最大生産性条件探索



単位時間あたりにどれだけの化合物を生産できるか？
「収率」だけでなく「流速」も重要なパラメーター

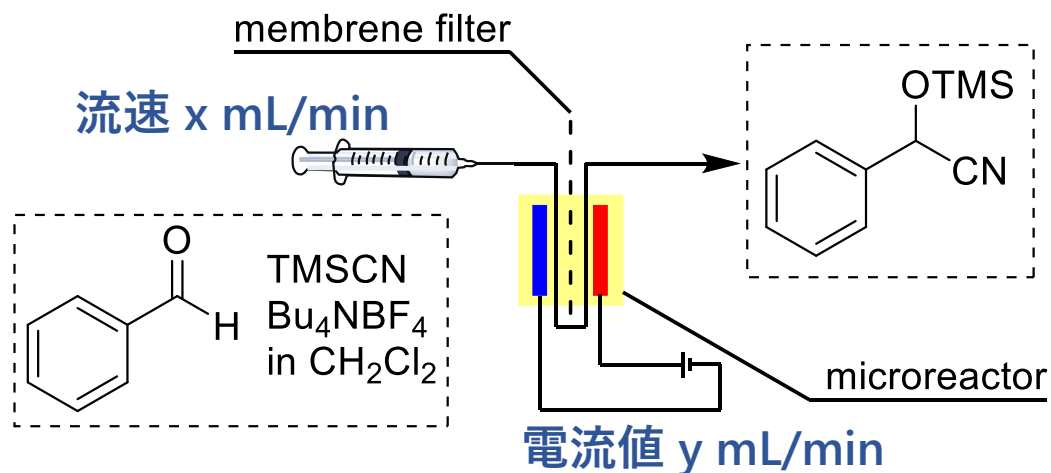


「収率×流速」を新たな「生産性」とした
目的変数とした最適化モデルの構築



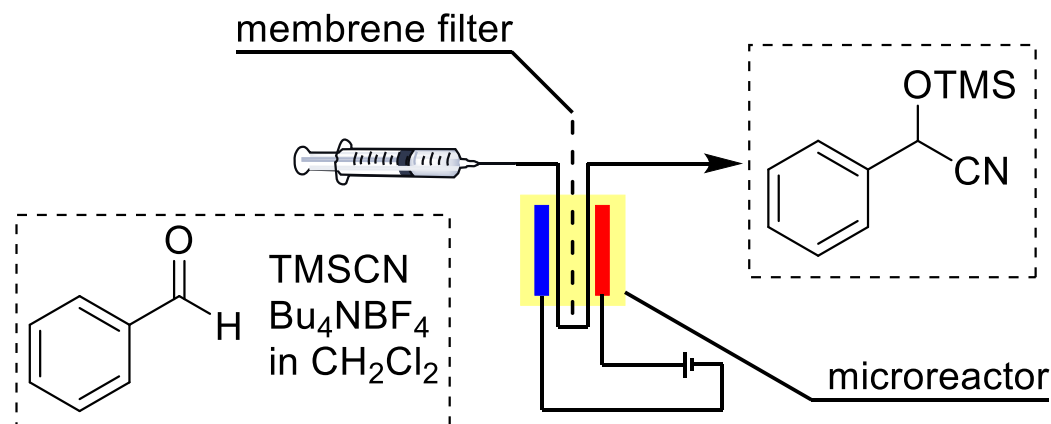
4.0 mL/min, 0.5 mAにて最大生産性を予測

シアノシリル化反応の最大収率条件探索



	流速	電流値	収率	生産性
収率	0.25 mL/min	0.2 mA	99%	0.012 g/min
生産性	4.0 mL/min	0.5 mA	86%	0.17 g/min

最適化条件を利用した大量合成(グラムスケール合成への適用)



ベンズアルデヒド615 mg (5.8 mmol)の原料に対して、最適化したそれぞれの条件下でシアノシリル化反応を試みた。

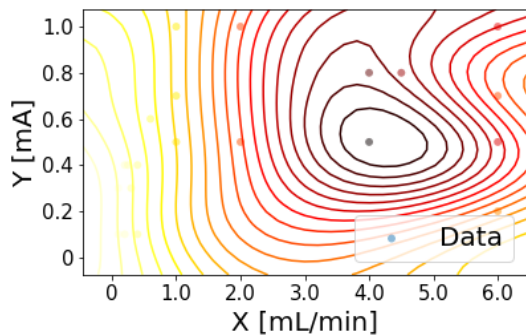
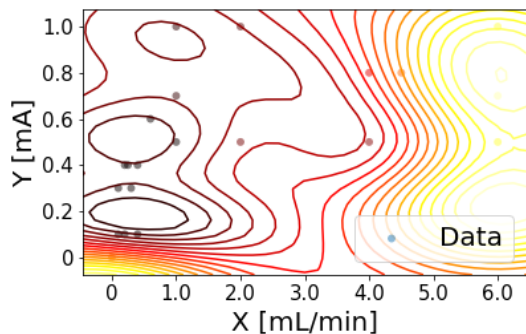
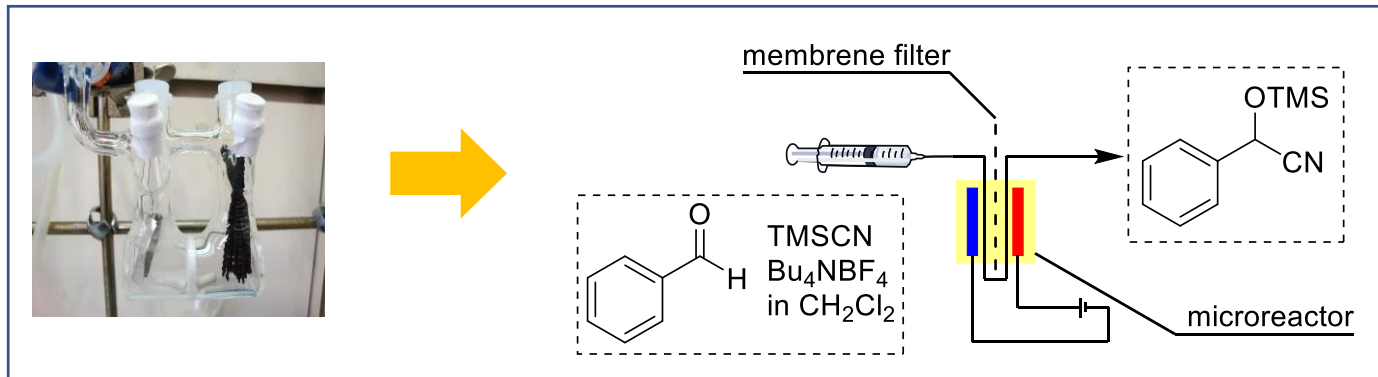
> 最大収率条件

目的物1.17 g (98%), 93 min

> 最大生産性条件

目的物1.07 g (90%), 6 min

小まとめ：ガウス過程回帰を用いたシアノシリル化反応の最適条件探索



「ベンズアルデヒドのシアノシリル化反応」

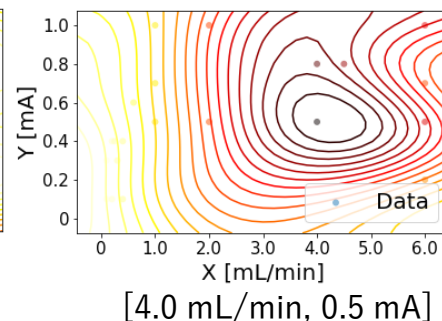
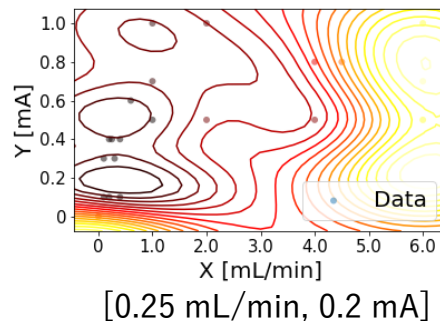
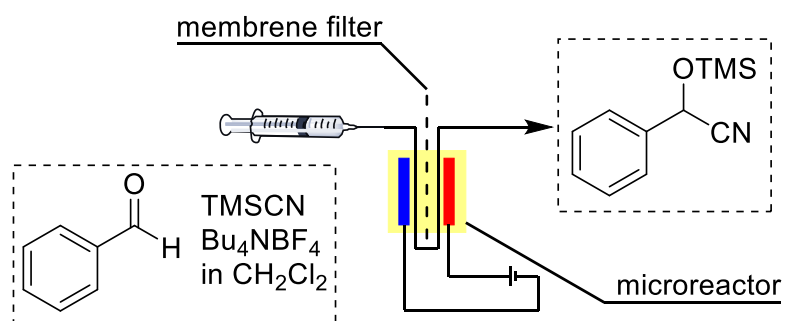
- ✓ マイクロフロー電解合成法の確立
- ✓ 機械学習を用いた反応条件最適化

ガウス過程回帰によるモデル構築

- > 最大収率条件と最大生産性条件
- > グラムスケール合成への展開

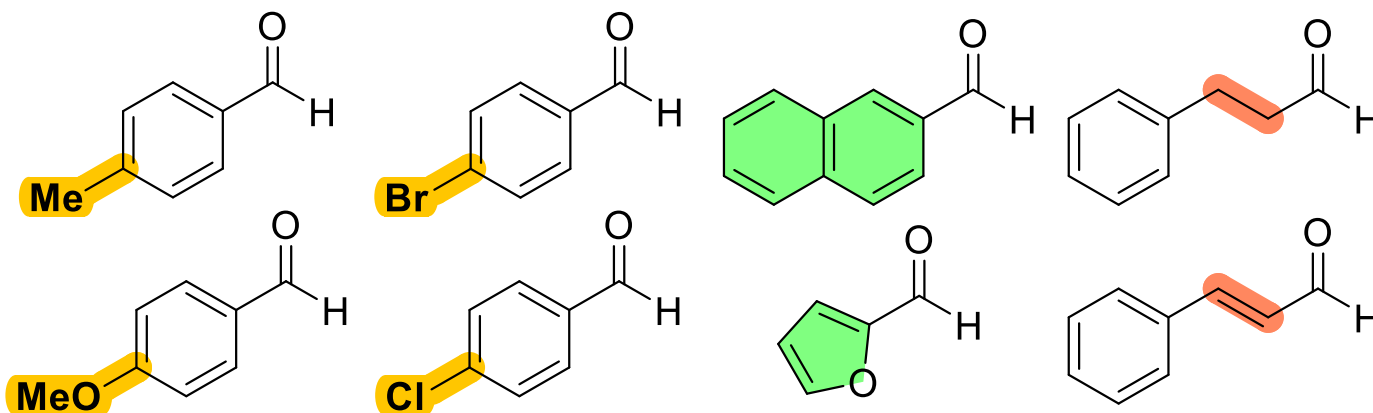
現在までの問題点と今後のやりたいこと

「ベンズアルデヒド」のシアノメチル化における最適な「流速」と「電流値」



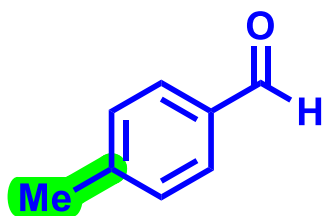
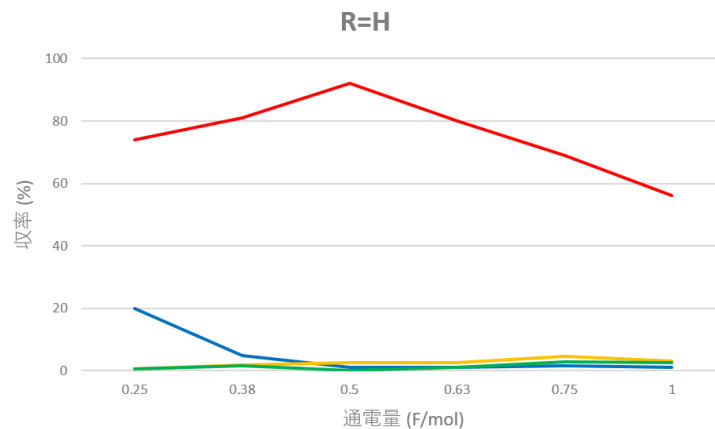
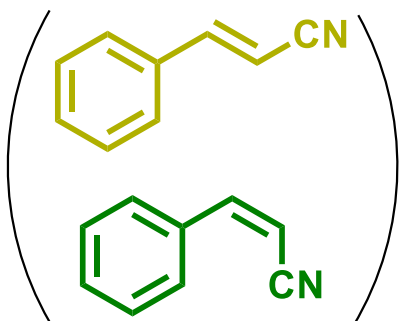
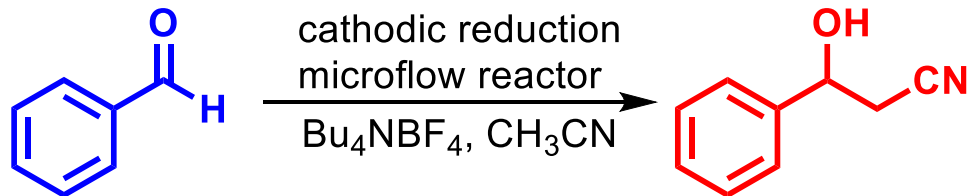
有機合成化学者はモデル基質に対する最適条件を用いて
出発原料を変えただけの「基質一般性調査」で結論を出してしまうことが多い

様々なベンズアルデヒド誘導体

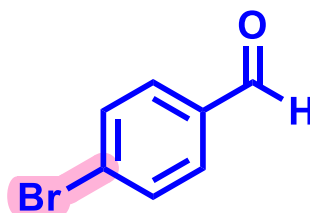
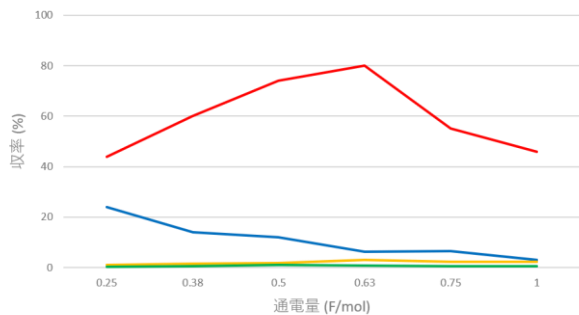


置換基が変われば、反応のしやすさも変わってくるはず
→それぞれ基質に応じた「最適条件」をできるだけ少ない検討で見出したい

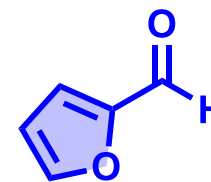
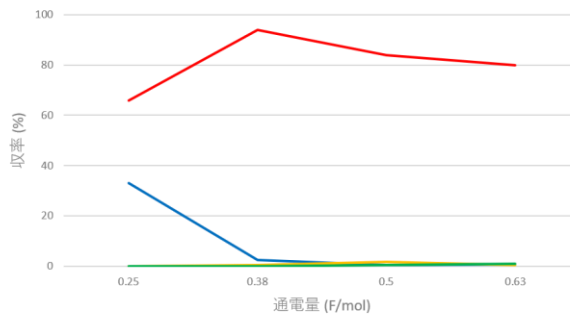
同じ化学変換でも出発原料を変えると最適な条件が異なる (シアノメチル化反応)



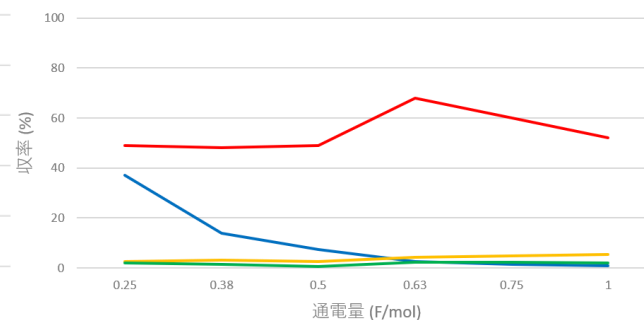
R=Me



R=Br



フルフラール



現在までの問題点と今後のやりたいこと：化合物の数値化とそれを用いた最適化

「流速」と「電流値」のみを入力変数としたモデル構築

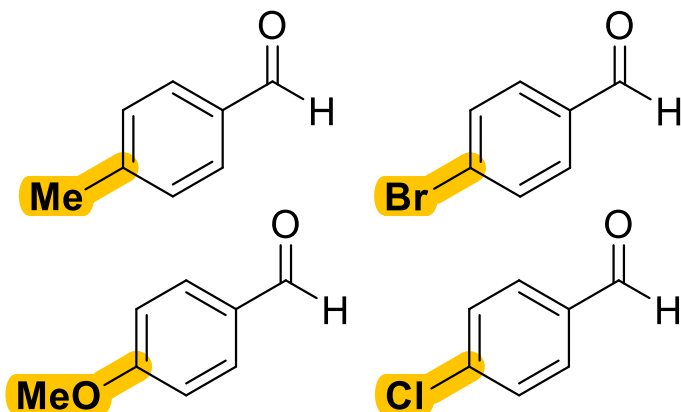


「化合物に関する情報」を入力変数とすることで、
モデルの構築に用いていない化合物を用いた時の
最適な反応条件を予測できるようなモデルを構築したい

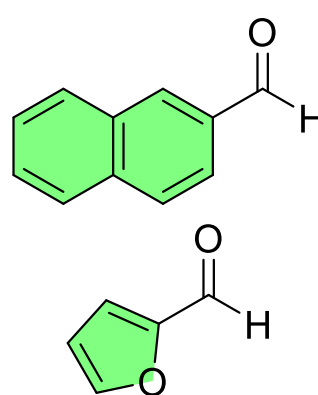


収率・生産性予測モデルの一般化と条件探索の効率化

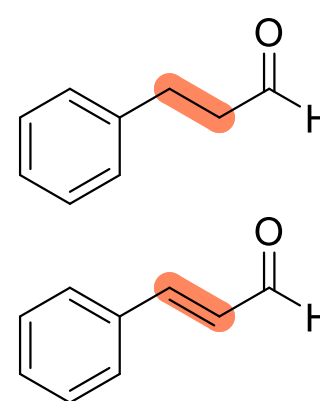
芳香環上の置換基



芳香環の種類



非芳香族アルデヒド



➡ どういうにして、化合物の特徴を数値化していくか？

有機化合物をどのようにしてパラメータ化していくのか

実験的なデータを用いる方法

核磁気共鳴(NMR)などの測定データ
普段取り扱っている実験データをそのまま使用することができる

量子化学計算を用いる方法

Gaussianを用いた量子化学計算
一つの分子に対する計算コストが高い
実験では得られない化合物の特徴を数値化できる

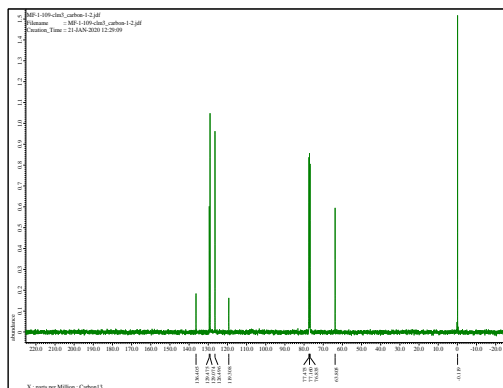
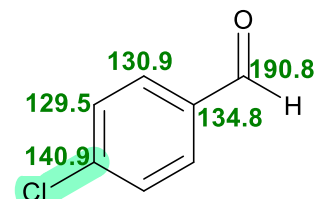
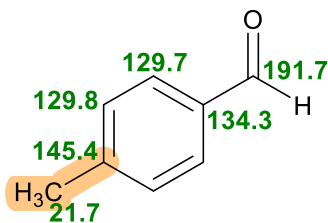
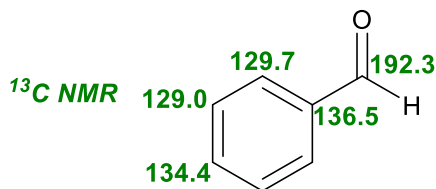
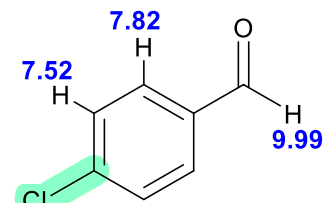
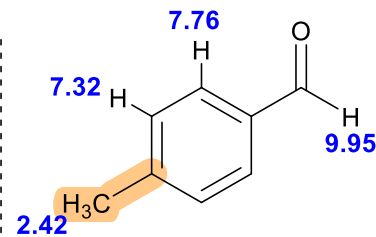
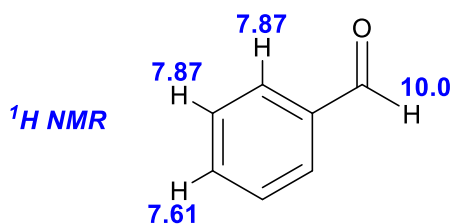
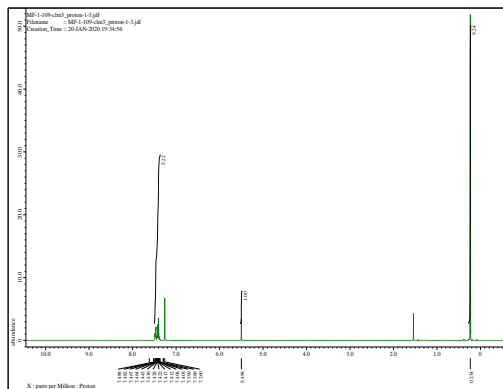
記述子を用いる方法

化合物の構造式を数値化する手法
構造式情報などをルールに基づいて数値化する手法
機械学習の利用が盛んになるにつれて、多くの表現方法が開発
フィンガープリントなど

有機化合物をどのようにしてパラメータ化していくのか：実験的なデータを用いる方法

核磁気共鳴(NMR)測定

- > 最も頻繁に使用される有機化合物の分析手法
- > ^1H や ^{13}C などの原子核がおかれている状態が分かる
- > ^1H 測定ならば数分で完結するため測定コストが少ない
- > 汎用な手法なので、データベースも豊富
- > 同じ位置に ^1H や ^{13}C といった核がなければ比較することができない



産総研のデータベースSDBSより引用
(https://sdb.sdb.aist.go.jp/sdb/cgi-bin/cre_index.cgi)

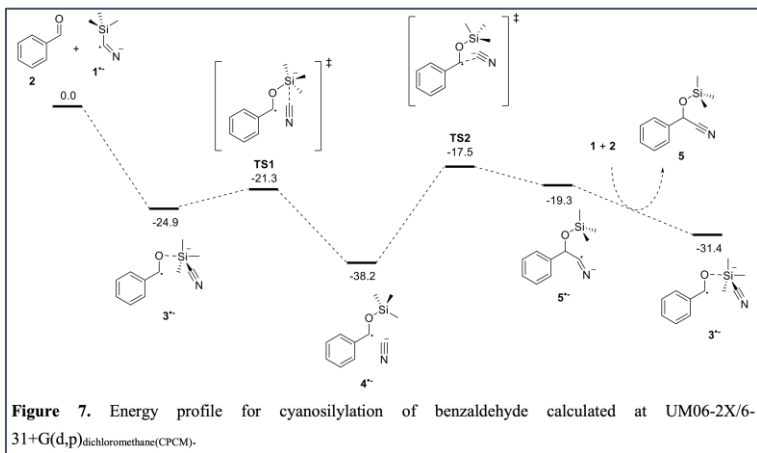
有機化合物をどのようにしてパラメータ化していくのか：量子化学計算を用いる方法

量子化学計算 (DFT計算など)

- > 最近では実験化学者も実験結果を支持するために使用頻度が上がってきた
- > 化合物の電子状態や軌道係数、内部エネルギーなどを数値化できる
- > 汎用コンピューターでも可能であるが、用いる関数によっては計算コストが高め
- > 一般的にはGaussianやSpartanといったソフトウェアを用いる(自動化?)











Psi4

- > Python上で動かせるオープンソースの量子化学計算アプリケーション
- > 速度が必要な部分はC++で記述されている。
- > 小さな分子であれば実用的に用いることは可能
- > 機械学習に利用する際に、計算の自動化も



反応機構探索(左図)などは得意だが、
大量の化合物に対しての
網羅的な計算はコストがかかりすぎる

有機化合物をどのようにしてパラメータ化していくのか：記述子を用いる方法

	Fsp3	Tanimoto	smiles	MW	PEOE_VSA1
	0	1	<chem>O=Cc1ccccc1</chem>	106.124	0
	0.125	0.5	<chem>Cc1ccc(C=O)cc1</chem>	120.151	0
	0.125	0.434782609	<chem>COc1ccc(C=O)cc1</chem>	136.15	4.736862954
	0	0.5	<chem>O=Cc1ccc(F)cc1</chem>	124.114	0
	0	0.5	<chem>O=Cc1ccc(Cl)cc1</chem>	140.569	0
	0	0.5	<chem>O=Cc1ccc(Br)cc1</chem>	185.02	0
	0.111111111	0.37037037	<chem>COC(=O)c1ccc(C=O)cc1</chem>	164.16	4.736862954
	0.2	0.37037037	<chem>CN(C)C(=O)c1ccc(C=O)cc1</chem>	177.203	4.899909731
	0	0.416666667	<chem>O=Cc1ccc([N+](=O)[O-])cc1</chem>	151.121	0
	0	0.434782609	<chem>N#Cc1ccc(C=O)cc1</chem>	131.134	0

ベンズアルデヒド誘導体についての記述子出力

Fsp3: sp³炭素の数

Tanimoto: 谷本係数によるベンズアルデヒドとの類似性

PEOE_VSA1: 部分電荷と表面積の尺度

RDKitを用いた記述子計算

ChemDraw(化合物の描画ソフト)で構造式を描く
SDF(Structure-Data file)形式で保存



RDKitで読み込み、記述子の出力
0次元から4次元まで数多くの種類

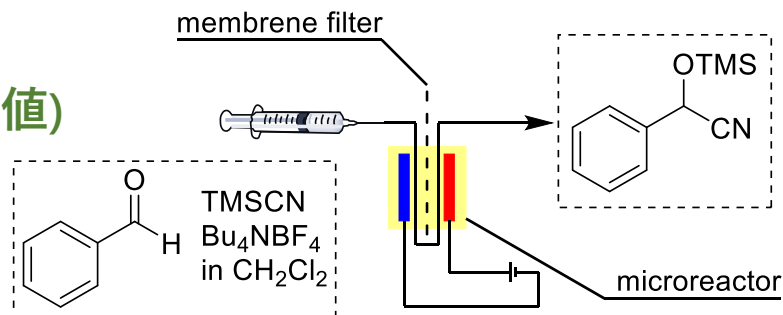


そのまま機械学習などの情報処理も可能

短い計算時間、計算コストで無数の記述子
フィンガープリントなどの出力も可能

現在までの問題点と今後のやりたいこと：化合物の数値化とそれを用いた最適化

これまでの実験化学的な条件探索(流速や電流値)



数値化された化合物情報

実験的な測定データ

量子化学計算

構造式の特徴量



化合物情報を内包する反応性モデル構築

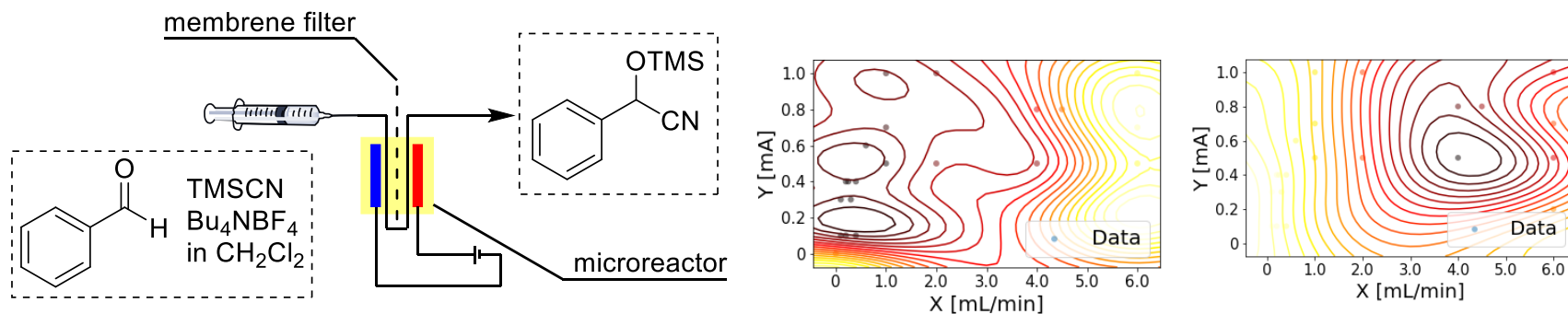
これまでは2次元でのガウス過程回帰を行ってきたが、これを多次元化することはできないか？

最近では、ベイズ最適化を用いた実験計画法(DoE)のアプローチも

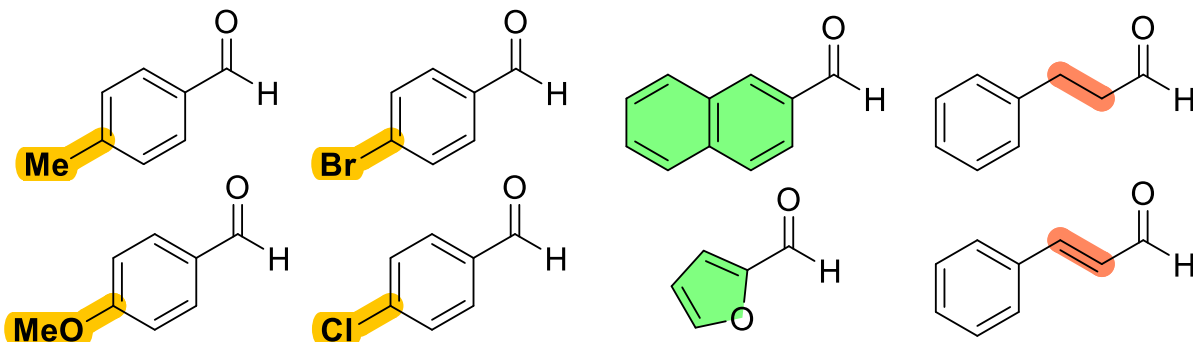
→2021年度 JST ACT-Xに申請中
「AI活用で挑む学問の革新と創成」

まとめ

マイクロフロー電解合成によるシアノシリル化と機械学習を用いた条件探索(論文化)



化合物の情報を組み込んだ最適化モデルの構築によって一般性を獲得する(計画中)



→2021年度 JST ACT-Xに申請中
「AI活用で挑む学問の革新と創成」

謝辞

工学部合成プロセス化学研究室 菅教授、光藤准教授、藤井さん(修士2年)
大阪大学産業科学技術研究所 笹井教授、滝澤教授、近藤助教(茨城大)
サイバーフィジカル情報応用研究コア(Cypher) Society5.0 研究支援プログラム