

# 反応機構を語る収率予測 AI

## —自動反応経路探索・自動実験・機械学習の融合—

### 概要

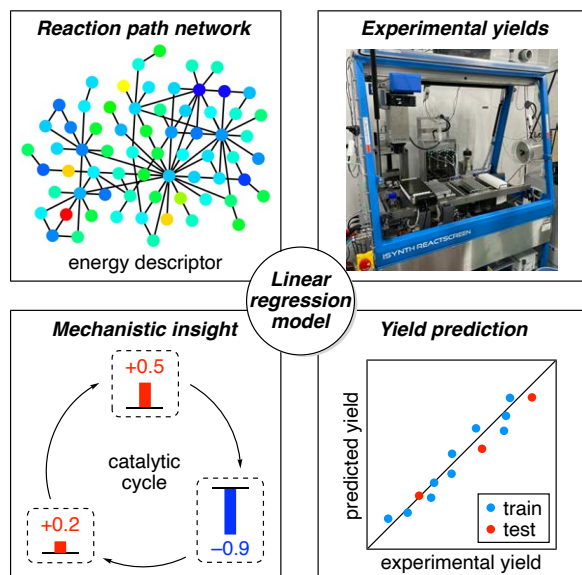
化学反応の収率を予測する人工知能 (AI) は近年発展していますが、多くは「なぜそう予測できるのか」という反応の仕組み (反応機構) までは説明できませんでした。

京都大学化学研究所 道場貴大 助教、北海道大学総合イノベーション創発機構化学反応創成研究拠点 (WPI-ICReDD) 前田理 教授らの研究グループは、反応の途中に現れる多数の中間体のエネルギーを入力 (記述子) として用いることで、収率を高い精度で予測しつつ、反応機構まで読み解ける「解釈可能な AI モデル」を構築しました。

反応経路を自動で網羅的に探索する独自の計算手法 (SC-AFIR 法)、ロボットによる自動実験、線形回帰を組み合わせた手法です。パラジウム触媒による炭素-水素結合の変換反応に適用し、未知の配位子の収率を平均誤差約 6% で予測し、収率を左右する重要な中間体を特定できることを示しました。

本成果は、反応の最適化や機構解明を予測と理解の両面から支える新しい枠組みになると期待されます。

本研究成果は、2026 年 6 月 19 日に米国の国際学術誌「*Journal of the American Chemical Society*」にオンライン掲載されました。



本研究の概要図：反応経路の自動探索で得た中間体のエネルギーを記述子とし、ロボット実験で得た収率と線形回帰でつなぐことで、収率予測と反応機構の解明を同時に実現する。(作成：道場 貴大)

## 1. 背景

反応の収率や選択性をあらかじめ予測できれば、化学反応の開発や最適化を大きく加速できます。そのため近年、機械学習 (AI) を用いて反応の結果を予測する研究が盛んに行われています。一般的な機械学習では、分子を数値の並び (特徴量・記述子) に変換し、その記述子と実験で得られた収率との関係をモデルに学習させます。

しかし、従来用いられてきた記述子は、分子の構造的・電子的な特徴を表すものが多く、研究分野ごとに経験的・恣意的に選ばれてきました。そのため、たとえ予測がよく当たっても、「なぜその収率になるのか」という反応機構と直接結び付けて解釈することが難しいという根本的な課題がありました。モデルから特徴量の重要度を取り出すことはできても、それが化学的に意味のある情報に対応しているとは限らなかったのです。

予測精度だけでなく、反応の仕組みそのものへの理解 (機構的洞察) を与えてくれるモデルが求められていました。

## 2. 研究手法・成果

研究グループは、反応系に現れうる「中間体のエネルギー」そのものを記述子に用いるという発想に着目しました。これは、反応の起こりやすさ (活性化エネルギー) が中間体の安定性 (エネルギー) と直線的に関係するという経験則 (BEP 則) に基づくもので、計算が難しい遷移状態を直接扱わずに反応性を捉えられる利点があります。

一方で、反応混合物中に存在する多数の中間体を漏れなく見つけ出すことは容易ではありません。そこで、ポテンシャルエネルギー曲面上の反応経路を自動かつ網羅的に探索する独自手法「SC-AFIR 法」(GRRM プログラムに実装) を用いました。初期の原料を入力するだけで、人工的な力を加えながら反応経路のネットワークを自動構築できます。今回のモデル反応では、約 4000 個の安定構造と約 7000 個の遷移状態を含むネットワークを構築し、その中から反応に重要な 52 個の中間体を抽出しました。各中間体に置換基を導入してエネルギーを計算し、「エネルギー記述子」を組み立てました。

モデル反応には、パラジウム触媒による単純な芳香族化合物の酸化的炭素-水素結合アルケニル化反応を選びました。配位子としてアミノ酸誘導体とピリドンの二種を組み合わせ、計 104 通りの条件をロボット合成プラットフォーム (Chemspeed SWING) で自動的に実験し、収率 (17~70%) を取得しました。

得られたエネルギー記述子と収率を線形回帰 (Ridge 回帰など) で結び付け、再帰的特徴量削減 (RFE) と交差検証 (LOGOCV) により評価しました。その結果、未知の配位子に対する収率を二乗平均平方根誤差 (RMSE) 約 6% という高い精度で予測できることを示しました。

さらに重要なのは、モデルの係数から反応機構を読み解ける点です。各中間体に割り当てられた係数の

符号と大きさは、その中間体を安定化・不安定化させることが収率向上に効くかどうかを表します。これらをつなぎ合わせると妥当な触媒サイクルが描け、パラジウム触媒による炭素-水素結合活性化や溝呂木・ヘック反応に関するこれまでの知見ともよく一致しました。加えて、立体・電子的パラメータに基づく従来型の記述子と比べても、予測精度は同等でありながら、反応機構に対する解釈性で優れていることを明らかにしました。

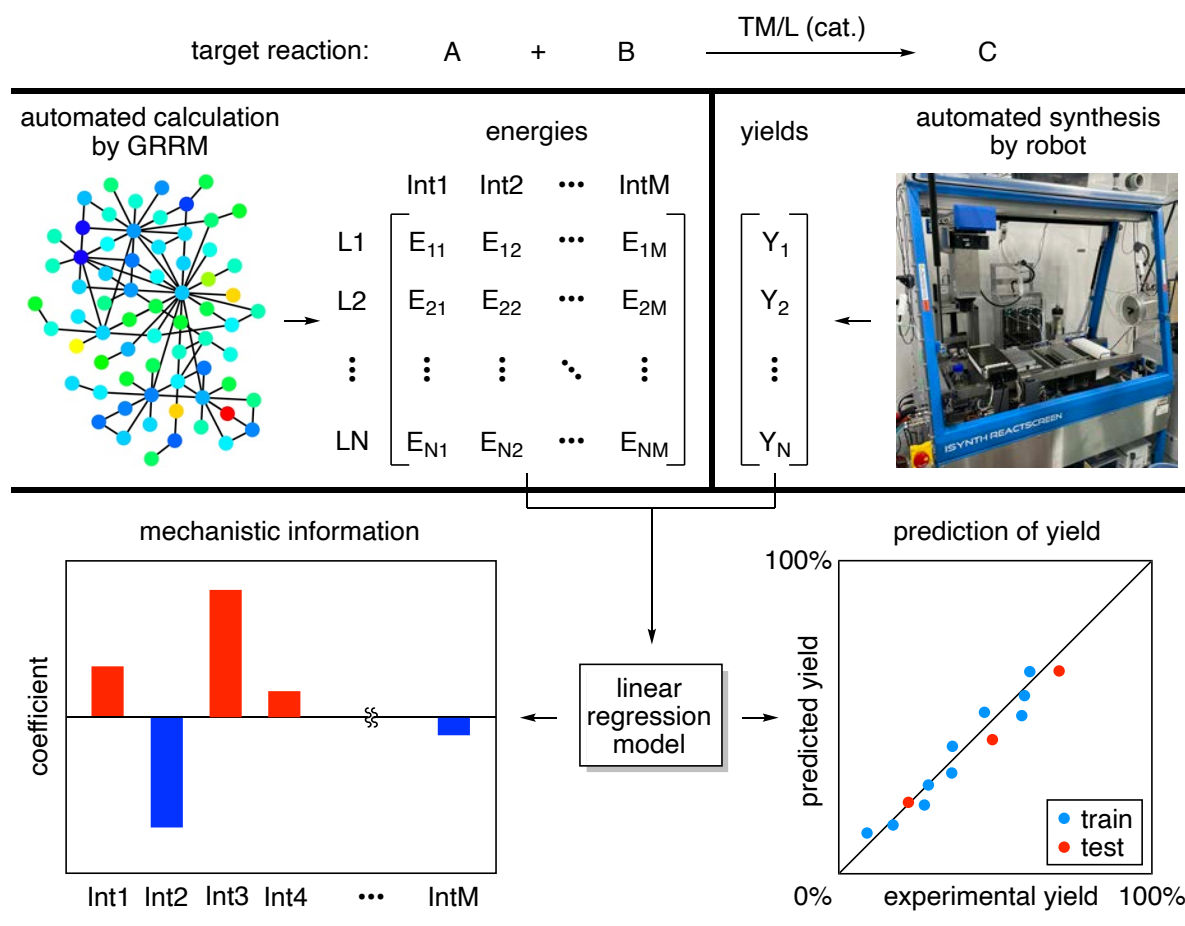


図 1. 本研究のワークフロー

### 3. 波及効果、今後の予定

本研究は、自動反応経路探索（理論計算）、自動合成（ロボット実験）、回帰モデル（機械学習）を統合し、収率の予測と反応機構の理解を同時に与える枠組みを示したものです。とくに、単一の律速段階が明確でない多段階反応では、どの中間体が結果を左右するかを実験や分光法だけで突き止めることが難しく、本手法はこうした系の機構解明に威力を発揮すると期待されます。

今後は、本手法をより優れた触媒や反応条件の探索へ展開する手法や、捉えた反応機構を類似反応へ転移する手法の開発を進める予定です。反応開発の効率化を通じて、持続可能で省力的な物質生産プロセスへの貢献も期待されます。

## 4. 研究プロジェクトについて

本研究は、科学技術振興機構（JST）戦略的創造研究推進事業 ERATO「前田化学反応創成知能プロジェクト」（JPMJER1903）、日本学術振興会（JSPS）科研費（JP22KJ0043、JP23B203）、JST 創発的研究支援事業（JPMJFR2221）の支援を受けて実施されました。理論計算の一部は北海道大学情報基盤センターのスーパーコンピュータシステムを利用して行われました。

### <用語解説>

- ・ **記述子（特徴量）**：機械学習において、分子や反応を数値の並びとして表現したもの。モデルはこの数値と実験結果との関係を学習する。どのような記述子を選ぶかが予測精度と解釈性を大きく左右する。
- ・ **SC-AFIR 法**：単成分人工力誘起反応（Single Component Artificial Force Induced Reaction）法。分子に仮想的な力を加えることで、化学反応の経路を自動的に探索する手法。GRRM プログラムに実装されており、初期の原料を与えるだけで反応経路のネットワークを網羅的に構築できる。
- ・ **反応経路ネットワーク**：原料から生成物に至る多数の中間体と、それらをつなぐ反応経路を網状に表したものの。反応の全体像を俯瞰できる。
- ・ **BEP 則（ベル・エバンス・ポラニー則）**：反応の起こりやすさを決める活性化エネルギーが、反応に伴うエネルギー変化と直線的な関係をもつという経験則。これにより、計算の難しい遷移状態を介さずに、中間体のエネルギーから反応性の傾向を見積もることができる。
- ・ **線形回帰・Ridge 回帰**：入力（記述子）と出力（収率）の関係を直線的な式で表す統計モデル。Ridge 回帰は、過度に複雑なモデルになることを抑える「正則化」を加えた手法で、少ないデータでも安定した予測と解釈が可能になる。
- ・ **交差検証**：モデルの予測性能を公平に評価する方法の一つ。本研究では、ある配位子に関わるデータをすべて学習から除外して予測させる「leave-one-group-out 交差検証」を用い、未知の配位子に対する真の予測力を厳しく検証した。
- ・ **酸化的炭素-水素結合アルケニル化**：芳香族化合物の炭素-水素結合を直接利用して炭素-炭素結合をつくり、アルケン（二重結合をもつ部位）を導入する反応。あらかじめ反応点を作り込む必要がなく、効率的な合成手法として注目されている。

### <研究者のコメント>

本研究は、機械学習による収率予測を「当たればよい」道具にとどめず、反応の仕組みを理解するための道具にできないか、という思いから始まりました。膨大な中間体を自動で洗い出し、その中から収率を左右する鍵を統計的に選び出せたときは、計算化学・実験化学・情報科学がひとつにつながる手応えを感じました。反応機構を理解した「AI 化学者」を開発する足掛かりにしたいと考えています。（道場 貴大）

<論文タイトルと著者>

タイトル：Construction of an Interpretable Regression Model for Yield Prediction and Mechanistic Insight Enabled by Automated Reaction Path Exploration (自動反応経路探索による収率予測と反応機構の理解のための解釈可能な回帰モデルの構築)

著者：Takahiro Doba (道場 貴大), Yu Harabuchi (原渕 祐), Yuuya Nagata (長田 裕也), Satoshi Maeda (前田 理)

掲載誌： *Journal of the American Chemical Society* DOI : 10.1021/jacs.6c05203