HOKKAIDO UNIVERSITY

Title	Development of a Method to Reduce the Number of Synthetic Experiments by Combining Bayesian Optimization and Reaction Barrier Calculations [an abstract of dissertation and a summary of dissertation review]
Author(s)	岡田, 拓明
Degree Grantor	北海道大学
Degree Name	博士(理学)
Dissertation Number	甲第15860号
Issue Date	2024-03-25
Doc URL	https://hdl.handle.net/2115/92253
Rights(URL)	https://creativecommons.org/licenses/by/4.0/
Type	doctoral thesis
File Information	OKADA_Hiroaki_abstract.pdf, 論文内容の要旨

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

学 位 論 文 内 容 の 要 旨

博士の専攻分野の名称　博士（理学）　氏名　岡田　拓明

学 位 論 文 題 名

Development of a Method to Reduce the Number of Synthetic Experiments by Combining Bayesian Optimization and Reaction Barrier Calculations
（ベイズ最適化と反応障壁計算を組み合わせた合成実験数削減法の開発）

有機合成において、基質や反応条件のスクリーニングは重要である。従来、化学者は自身の勘や経験に基づいて化学空間を探索し、反応開発を進めてきた。しかし、そのようなヒューリスティクスなアプローチは、新規性の高い反応に対して事前知識が適用しにくく、化学者のバイアスがかかった探索をしてしまうという欠点がある。この課題に対し、インフォマティクスや量子化学計算を活用した反応スクリーニングが行われている。

インフォマティクスでは、既知のデータから隠れたルールや関係性を見つけ出し、目的変数の予測や最適化を行う。特に有機合成では、得られた実験結果を用いてベイズ最適化を行うことで、少ない実験回数で効率的に最適な基質や反応条件を発見できることが知られている。また、量子化学計算では、化学反応の反応障壁であるギブス自由エネルギー変化を計算することで、反応性や選択性を評価できる。即ち、障壁計算を行うことで、実験をせずに計算だけで化学反応の結果を予測することが可能である。

このような背景を踏まえて本研究では、ベイズ最適化と障壁計算を組み合わせることで、より効率的に反応スクリーニングできると考えた。より具体的には、実験データの代わりに計算で予測された反応障壁データを用いてベイズ最適化を行えば、最適化に要する実験データをさらに削減できると考えた。しかし、障壁計算を利用したベイズ最適化がどの程度実験回数を削減できるかは明らかでない。

そこで本研究では、ベイズ最適化と反応障壁計算を組み合わせた合成実験数削減法を開発した。本手法では、実験データだけでなく、障壁計算で得られた反応性・選択性のデータも用いてベイズ最適化を行い、基質や反応条件の最適化に要する実験回数を削減する。開発した手法の有用性を検証するため、開発手法の性能を系統的にベンチマークすると共に、本手法を実際の有機合成に適用して反応スクリーニングを行った。

本学位論文は全 6 章で構成されている。第 1 章は general introduction であり、有機反応の概要や反応スクリーニングの難しさについて述べた。また、量子化学計算を用いた反応経路探索や機械学習、最適化アルゴリズムの概念および有機合成への活用例についてまとめた。また、それらの背景を元に、本研究で取り組む課題について述べた。

第 2 章では、本研究で利用した手法についてまとめた。初めに、人工力誘起反応（AFIR）法の概念や方法論についてまとめた。続いて、分子構造に対して原子の置換・置換基の追加・分子の配置を行うアルゴリズムの詳細を記載した。さらに、遺伝的アルゴリズムとベイズ最適化の概要およびアルゴリズムについて述べた。最後に、自動合成装置である Chemspeed

ISYNTH robotic platform の概要について紹介した。

第 3 章では、最適化アルゴリズムの性能を評価できる反応障壁のデータセットを作成した。具体的には Claisen 転位に着目し、異なる置換基組み合わせを持つアリルビニルエーテルを 100,000 分子作成した。各分子の Claisen 転位における反応障壁を AFIR 法および半経験的計算手法である PM7 を用いて計算し、反応障壁のデータセットを作成した。また、データセットにおいて反応障壁を小さくする置換基のトレンドを系統的に解析し、化学的・数値的な観点から妥当なデータが得られていることを確認した。続いて、作成したデータセットを用いて遺伝的アルゴリズムの性能評価を行い、ランダムサーチの結果と比較した。その結果として、作成した反応障壁のデータセットが最適化アルゴリズムの性能評価に利用できることを示した。

第 4 章では、ベイズ最適化と反応障壁計算を組み合わせた合成実験数削減法を開発した。量子化学計算で得られた反応障壁データは、それそのものが化学反応の反応性や選択性を表す指標である。そこで、実験データと共に障壁計算で得られた反応性・選択性のデータも用いてベイズ最適化を行うことで、最適化に要する実験回数を削減できると考えた。しかし、計算誤差の大きさや計算のバッチサイズと削減できる実験回数の関係性は明らかでない。そこで本研究では、Claisen 転位の反応障壁データセットを用いた仮想的な実験・計算により、本手法の性能をベンチマークした。具体的には、仮想的な計算誤差の大きさや計算バッチサイズを系統的に変化させ、本手法と従来のベイズ最適化の性能を比較した。その結果、20 kJ/mol 程度の計算誤差がある場合にも、本手法は従来のベイズ最適化よりも少ない実験回数で最適化を完了できることが示唆された。

第 5 章では、開発した手法を熱的アジド-アルキン環化付加の位置選択性最大化へ適用した。熱的アジド-アルキン環化付加は、アジドとアルキンから 1,2,3-トリアゾールを生成する反応であり、一般に 2 種類の異性体（1,5-位置換異性体と 1,4-位置換異性体）の混合物が得られる。そこで、異なるデータを用いる 3 つの最適化スキームを用いてアジドとアルキンの置換基組み合わせを最適化し、本反応の位置選択性を最大化する組み合わせを探索した。具体的には、アジドとアルキンの置換基の候補を 16 種類ずつ用意し、256 通りの組み合わせの中から最適な組み合わせを探索した。実験においては、Chemspeed ISYNTH robotic platform を用いて効率的に合成を行った。最適化の結果、74.7% の位置選択性で 1,5-位置換生成物を与える置換基組み合わせ、および 86.4% の位置選択性で 1,4-位置換生成物を与える置換基組み合わせが発見された。

第 6 章は general conclusion であり、本研究の総括と今後の展望について述べた。

以上、本研究ではベイズ最適化と障壁計算を組み合わせた合成実験数削減法を開発し、その性能を反応障壁のデータセットを用いてベンチマークした。また、開発した手法をアジド-アルキン環化付加に適用し、位置選択性を最大化する置換基組み合わせを探索した。本手法は、異なる有機反応にも適用可能な汎用的な手法であり、反応スクリーニングに伴う人的・実験的コストを削減することができる。今後、本手法を自動合成ロボットと連動させた closed-loop な反応開発への応用が期待できる。