

PRESS RELEASE

 2022年2月7日
 理化学研究所
 金沢大学

計算機上で収集したデータの機械学習による不斉触媒設計

－有機合成 DX 化の基盤技術構築に向けて－

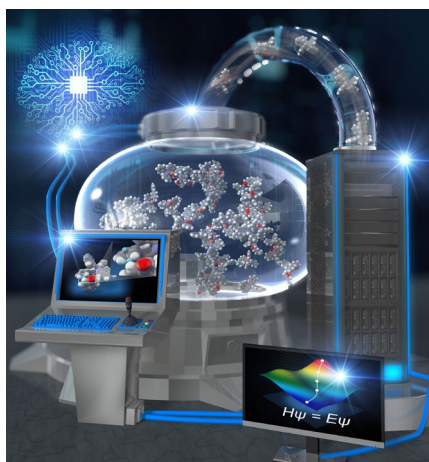
理化学研究所（理研）環境資源科学研究センター先進機能触媒研究グループの山口滋上級研究員（研究当時、現客員研究員）、金沢大学大学院医薬保健学総合研究科創薬科学専攻博士前期課程2年の向真潔、金沢大学医薬保健研究域薬学系の長尾一哲助教、大宮寛久教授らの共同研究グループ*は、遷移状態計算^[1]と機械学習^[2]を併用して、「エナンチオ選択性^[3]」が向上する不斉触媒^[3]を計算機上で設計することに成功しました。

本研究成果は、有機合成のデジタルトランスフォーメーション（DX）^[4]化基盤の構築に貢献すると期待できます。

医薬品などの開発においては、分子の立体構造が生物活性に大きく影響することから、エナンチオマー^[3]を選択的に生成できる不斉触媒が重要であり、近年、人工知能（AI）^[2]の活用が注目されています。現状では、AI構築には実験データが必要ですが、量子化学計算^[1]により計算機上（インシリコ）で収集したデータの解析を基に触媒が設計できれば、触媒が高価などの理由で解析用の実験データの収集が難しい反応系でも開発を効率化できるなど、さまざまな展開が考えられます。

今回、共同研究グループは、モデルとした不斉触媒反応^[3]において、遷移状態計算により計算機上で集めたわずか30個のサンプルから、エナンチオ選択性が向上する「データ駆動型不斉触媒設計^[5]」に成功しました。

本研究は、科学雑誌『*Bulletin of the Chemical Society of Japan*』（1月13日付）に掲載されました。



データ駆動型インシリコ不斉触媒設計の概念図

※共同研究グループ

理化学研究所 環境資源科学研究センター 先進機能触媒研究グループ
 上級研究員（研究当時） 山口 滋（やまぐち しげる）
 （現 客員研究員）

金沢大学

大学院医薬保健学総合研究科

創薬科学専攻博士前期課程 2 年 向 真潔（むかい まさきよ）

医薬保健研究域薬学系

助教 長尾 一哲（ながお かずのり）

教授 大宮 寛久（おおみや ひろひさ）

研究支援

本研究は、日本学術振興会（JSPS）科学研究費補助金新学術領域研究（研究領域提案型、研究領域名：分子合成オンデマンドを実現するハイブリッド触媒系の創製）「有機触媒と金属触媒のハイブリッドに基づく高次反応制御法の開発（研究代表者：大宮寛久）」（JP17H06449）、「ハイブリッド触媒開発を加速するデータ駆動型インシリコ分子設計法の構築（研究代表者：山口滋）」（JP20H04831）による支援を受けて実施されました。

1. 背景

触媒は医薬品やプラスチックなど、身の回りの化学製品やその原料となる有機分子を合成するのに不可欠です。特に、医農薬品などの開発においては、生成物の立体構造が生物活性に大きな影響を及ぼします。そのため、有機合成分野では生成物の立体構造の制御ができる不斉触媒反応の開発が重要な課題です。近年、不斉触媒反応開発に人工知能（AI）を活用することが注目されており、データ駆動型の触媒設計や反応最適化に関する研究が精力的に行われています。

不斉触媒反応の解析では多くの場合、実験により生成物の「エナンチオ選択性」の値（例えば「エナンチオマー過剰率（ee）^[6]」の値）を集める必要があります。一方で、計算機の性能向上や計算化学の発展により、量子化学計算手法を用いてエナンチオ選択性を推定することも可能になってきました。

計算機上で集めた訓練データ^[7]に基づくデータ駆動型の高性能触媒設計が実現できれば、触媒が高価だったり、合成が難しかったりするために実験による解析用のデータ収集が困難な反応系においても、最適化が容易になります。また、実験環境がない場合でも、大学や研究機関のスーパーコンピュータなどを用いることで、不斉触媒反応のデータ解析が可能になります。さらには計算機上で、全自動で高性能不斉触媒を設計するための人工知能の構築の足掛かりになるなど、さまざまな展開が期待できます。

しかし、現状では量子化学計算によって計算機上で生成物のエナンチオ選択性を推定するには、計算負荷の大きい高精度の計算化学手法を用いる必要があるため、時間がかかります。従って、上述のような展開を実現するには、少数の

サンプルを用いたデータ解析による不斉触媒の設計や最適化の方法論が必要です。

そこで共同研究グループは、計算機上で収集した少数データの機械学習によるデータ駆動型不斉触媒設計法の構築に取り組みました。

2. 研究手法と成果

共同研究グループは、データ駆動型不斉触媒設計法の構築のため、解析対象として不斉触媒「Nヘテロ環カルベン銅錯体」を用いた「シリルボロン酸エステルアルデヒドへの不斉付加反応」^{注1)}を選びました。遷移状態計算により計算機上で算出したエナンチオ選択性の値と、「分子場解析^[8]」と呼ばれるデータ解析手法を組み合わせることで、少数サンプルの訓練データを用いた、エナンチオ選択性が向上する触媒設計を目指しました。

分子場解析とは、生成物のエナンチオ選択性と分子の3次元構造から算出した分子場^[8]との間の回帰分析^[9]です。作成した回帰モデル^[9]の回帰係数^[9]から、エナンチオ選択性にとって重要な構造情報を可視化できます。山口滋上級研究員らは2019年に、エナンチオ選択性決定段階の中間体を用いて分子場解析を行うと、少数の実験データからどこに分子構造を導入すればエナンチオ選択性が向上するかに関する情報を抽出・可視化でき、選択性が向上する分子設計が可能になることを明らかにしました(図1A)^{注2)}。そこで、遷移状態計算により算出したエナンチオ選択性の値と、対応する遷移状態構造を用いて分子場解析を行えば、エナンチオ選択性が向上する分子設計が計算機上で実現できると考えました(図1B)。

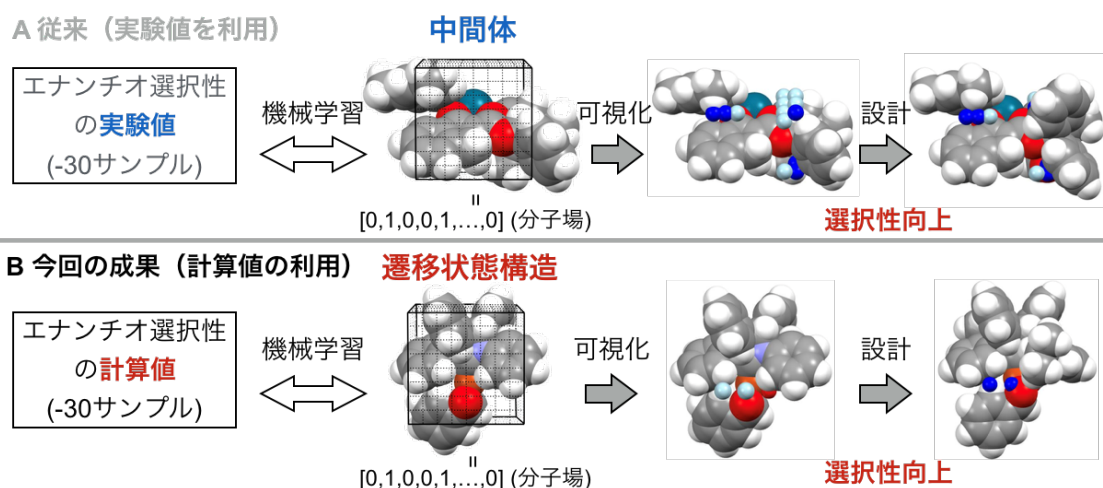


図1 従来法と今回の成果の比較

A: 従来法。エナンチオ選択性の実験値と、反応中間体との間で機械学習・分子場解析を行い、抽出・可視化した重要構造情報(水色の球)に重なるように分子設計を行うことで、エナンチオ選択性が向上する。

B：今回開発した方法論。Aにおけるエナンチオ選択性の実験値および中間体の代わりに、エナンチオ選択性の計算値および計算に伴って得られる「遷移状態構造」を用いた分子場解析に基づき、エナンチオ選択性が向上する分子設計を行う。原理的に、設計に実験を必要としないことが特徴である。

まず、対象とする不斉触媒（Nヘテロ環カルベン銅錯体）を用いたシリルボロン酸エステルのアルデヒドへの不斉付加反応において、これまでに低～中程度のエナンチオ選択性を示すことが分かっている不斉配位子（不斉源となる配位子で、今回は銅イオンに配位している）L1～L3を用いて、基質（アルデヒド）6種と組み合わせた計18サンプルに対して遷移状態計算を行いました（図2上段）。

次に、算出したエナンチオ選択性の値と、対応する遷移状態の構造を用いて分子場解析を行い、重要構造情報（分子構造上エナンチオ選択性が向上する領域：図では青または水色の球で示す）を抽出・可視化しました。解析によりエナンチオ選択性発現のメカニズムに関する知見が得られ、この知見を基に、遷移状態構造が重要構造情報に重なるように、新たに二つの不斉配位子L4とL5を設計しました。設計した不斉配位子を用いて実験を行ったところ、エナンチオ選択性の値は、訓練データの最大値（73%ee）を大きく超える87%ee（L4の場合）と82%ee（L5の場合）を示しました（図2下段左）。

さらなるエナンチオ選択性の向上を目指し、設計した不斉配位子の遷移状態計算を行い、計30サンプル（L1～L5と6種のアルデヒドの組み合わせ）の訓練データを用いて分子場解析を行いました。その結果、再解析により抽出・可視化した重要構造情報を基に設計した不斉配位子L6とL7は、それぞれ訓練データの最大値（89%ee）を超える90%eeと96%eeを示すことを実験で確認しました（図2下段右）。L7はデータ解析から得られた知見がなければ、検討しなかった不斉配位子です。

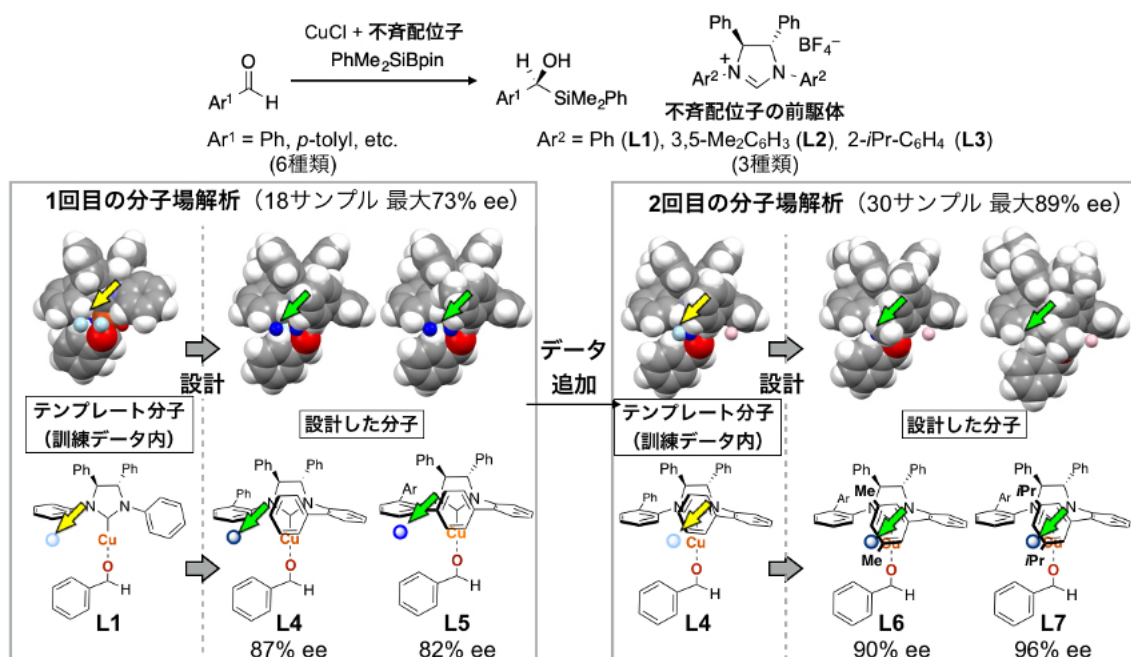


図 2 解析対象の反応と実際の分子設計

上段：左側の反応式は、解析対象とした不斉触媒 N ヘテロ環カルベン銅錯体 (CuCl+不斉配位子) を用いたシリルポロン酸エステル (PhMe₂SiBpin) のアルデヒド (Ar¹-CHO) への不斉付加反応。右側は不斉配位子 L1~L3 の前駆体の構造式。

下段：L1 などのラベルで示す不斉配位子およびベンズアルデヒドを含む遷移状態構造を、分子場解析により抽出・可視化した重要構造情報 (水色および青色の球) とともに示す。遷移状態構造に本来含まれているケイ素 (Si) 置換基は省略している。水色の球に重なるように (青の球に変わるように) 訓練データ内のテンプレート分子に置換基を導入することで、エナンチオ選択性の向上に成功した。L6 および L7 を含む遷移状態構造は、緑色の矢印の先にある青色の球と完全に重なっている。ピンク色の球は重ると選択性が下がる重要構造情報を示す。

以上の結果から、実験値で最大 73%ee を示す 18 サンプルの訓練データの解析から始めて、分子場解析に基づくデータ駆動型不斉触媒設計を 2 サイクル繰り返すことで、最大 96%ee を示す不斉触媒中の不斉配位子の設計に成功しました (図 3)。

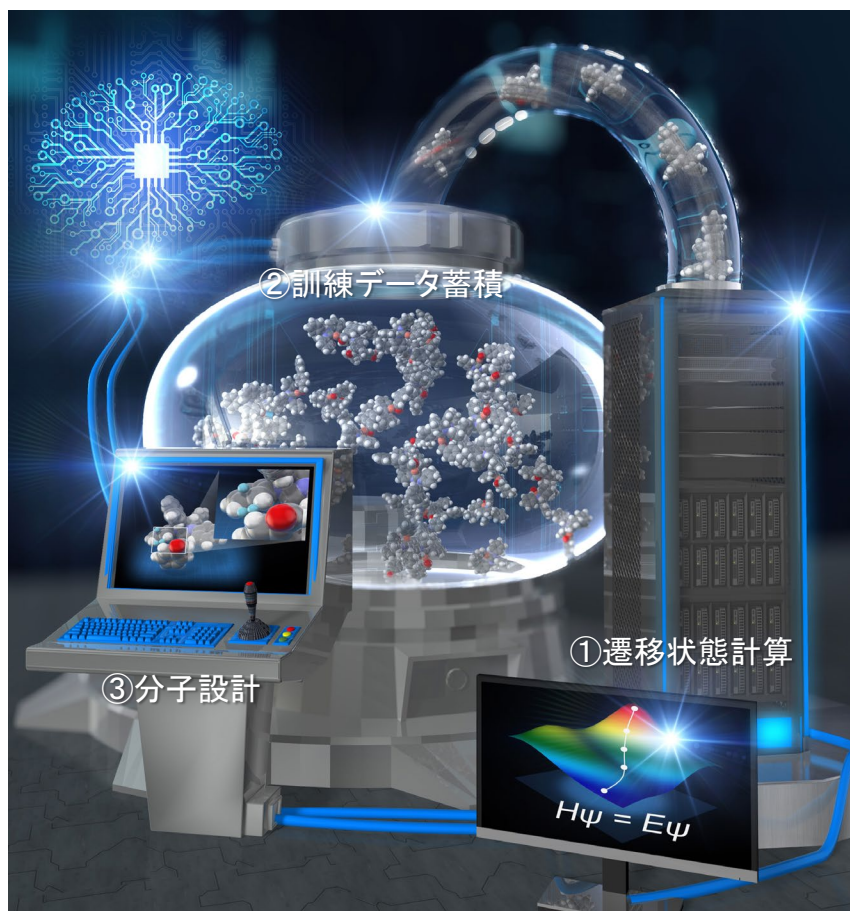


図 3 本研究で構築したデータ駆動型インシリコ不斉触媒設計の概念図

右下の計算機でエナンチオ選択性の値と対応する遷移状態構造を算出し (①)、訓練データを中央のドームに蓄積 (②)、集めた訓練データを用いて機械学習を行い、選択性支配因子を抽出・可視化する (③)。可視化した重要構造情報を基に研究者が設計した分子を用いて、①で再び遷移状態計算を行い、訓練データを追加する。最大 73%ee を示す触媒を含む訓練データ 18 サンプルから始めて、①～③のサイクルを繰り返すことで、96%ee を示す不斉触媒および不斉配位子の設計に成功した。

注 1) 2018 年 12 月 21 日金沢大学プレスリリース「キラルな医薬品の合成に向けた新しい戦略」

<https://www.kanazawa-u.ac.jp/rd/63183>

注 2) 2019 年 9 月 13 日理化学研究所プレスリリース「触媒反応におけるデータ駆動型分子設計に成功」

https://www.riken.jp/press/2019/20190913_4/index.html

3. 今後の期待

今回、遷移状態計算で得られた少数のエナンチオ選択性のデータと、対応する遷移状態の構造を基に分子場解析を行い、エナンチオ選択性が向上する触媒のデータ駆動型設計に成功しました。本手法を用いることで、実験による解析用データの収集が難しい場合や、合成が難しい不斉触媒を用いる反応系でも、ある程度の計算資源があれば、データ駆動型触媒設計を展開できます。

また、本手法は Gaussian^[10]や Python^[10]、R 言語^[10]といった研究者に一般的に普及しているツールを用いた解析で、実験化学者をはじめさまざまな研究者が汎用的に手法を活用し改善できるため、有機合成デジタルトランスフォーメーション (DX) 化のための基盤技術の構築につながります。

今回の研究成果は、触媒反応の開発における開発期間の短縮や環境負荷の低減を可能にし、2016年に国際連合が発表した「持続可能な開発目標 (SDGs)^[11]」にある「7 エネルギーをみんなに、そしてクリーンに」や「9 産業と技術革新の基盤を作ろう」に貢献するものです。

4. 論文情報

<タイトル>

Molecular Field Analysis Using Computational-Screening Data in Asymmetric N-Heterocyclic Carbene-Copper Catalysis toward Data-driven in silico Catalyst Optimization

<著者名>

Masakiyo Mukai, Kazunori Nagao, Shigeru Yamaguchi, and Hirohisa Ohmiya

<雑誌>

Bulletin of the Chemical Society of Japan

<DOI>

[10.1246/bcsj.20210349](https://doi.org/10.1246/bcsj.20210349)

5. 補足説明

[1] 遷移状態計算、量子化学計算

化学反応の進行につれて反応系が始状態から終状態に向かって原子配置を変えていく過程で、自由エネルギーの最も高い状態を遷移状態と呼ぶ。ここでの「遷移状態計算」とは遷移状態の構造やエネルギーを求める「量子化学計算」を指す。量子化学計算は、原子や分子の構造や性質を電子状態から解析する手法のこと。

[2] 機械学習、人工知能 (AI)

ここでは、データを基にコンピュータにその特徴やパターンを学習させることを機械学習とし、機械学習によりデータの分析や予測が可能になったコンピュータを人工知能 (AI) とする。

[3] エナンチオ選択性、不斉触媒、エナンチオマー、不斉触媒反応

鏡に映した右手と左手のように、重ね合わせることができない立体異性体を「エナンチオマー (鏡像異性体)」という。通常の有機合成ではエナンチオマーの関係にある物質が 1:1 で生成される (ラセミ体) が、片方のエナンチオマーを選択的に作り出す反応を不斉反応という。触媒としてわずかな量の不斉源 (不斉触媒) を用いた不斉反応のことを「不斉触媒反応」という。不斉反応の結果、片方のエナンチオマーがどれだけ得られたかを「エナンチオ選択性」という。

[4] デジタルトランスフォーメーション (DX)

ここでは機械学習・人工知能などのデジタル技術による研究のあり方の変革を指す。

[5] データ駆動型不斉触媒設計

機械学習による予測や、抽出したパターン・知見をもとに不斉触媒を設計すること。

[6] エナンチオマー過剰率 (ee)

エナンチオ選択性の指標で、次の式で求められる。

$$ee (\%ee) = ([R] - [S]) / ([R] + [S])$$

[R]および[S]は、各エナンチオマーを R および S としたときのそれぞれのモル分率。100%ee に近いほど純粋なエナンチマーであることを示す。ee は enantiomeric excess の略。

[7] 訓練データ

回帰モデルを作成するために使う反応データのこと。

[8] 分子場解析、分子場

比較分子場解析 (CoMFA: Comparative Molecular Field Analysis) を代表とする 3D-QSAR のことを指す。特に、ここでは格子空間に基づき計算した分子の 3 次元構造情報である分子相互作用場を記述子として用い、また目的変数として生物活性ではなく不斉収率を用いた 3D-QSAR のことを指している。格子空間に配置した分子の 3 次元構造から計算した分子の性質を表す数値を、分子相互作用場あるいは分子場と呼ぶ。

[9] 回帰分析、回帰モデル、回帰係数

回帰分析は、ある一つの量 (目的変数) を別の量 (説明変数) を用いて予測するための関係式を導く統計的手法。推定された関係式を回帰モデルと呼び、説明変数にかかるパラメータ・数値のことを回帰係数と呼ぶ。目的変数をうまく予測できるように回帰係数を決定することで、回帰モデルを作成する。

[10] Gaussian、Python、R 言語

Gaussian は量子化学計算のためのソフトウェア。Python や R 言語は機械学習に使われることが多いプログラミング言語。

[11] 持続可能な開発目標 (SDGs)

2015 年 9 月の国連サミットで採択された「持続可能な開発のための 2030 アジェンダ」にて記載された 2016 年から 2030 年までの国際目標。持続可能な世界を実現するための 17 のゴール、169 のターゲットから構成され、発展途上国のみならず、先進国自身が取り組むユニバーサル (普遍的) なものであり、日本としても積極的に取り組んでいる (外務省ホームページから一部改変して転載)。

6. 発表者・機関窓口

<発表者> ※研究内容については発表者にお問い合わせください。
理化学研究所 環境資源科学研究センター 先進機能触媒研究グループ
上級研究員（研究当時） 山口 滋 （やまぐち しげる）
（現 客員研究員）
E-mail : shigeru.yamaguchi.hw[at]a.riken.jp

金沢大学
大学院 医薬保健学総合研究科 創薬科学専攻
博士前期課程2年 向 真潔 （むかい まさきよ）
医薬保健研究域薬学系
助教 長尾 一哲（ながお かずのり）
教授 大宮 寛久（おおみや ひろひさ）
TEL : 076-234-4411（大宮）
E-mail : ohmiya[at]p.kanazawa-u.ac.jp（大宮）

<機関窓口>

*今般の新型コロナウイルス感染症対策として、理化学研究所では在宅勤務を実施しておりますので、メールにてお問い合わせ願います。

理化学研究所 広報室 報道担当
E-mail : ex-press[at]riken.jp

金沢大学薬学・がん研支援課企画総務係
E-mail : y-somu[at]adm.kanazawa-u.ac.jp

※上記の[at]は@に置き換えてください。