



2024年3月6日

報道機関 各位

国立大学法人東北大学

## 機械学習で迅速に有効な新規抗菌薬候補分子を見つける 新技術を開発

—これまで報告された膨大な候補分子情報を記述子に変換利用—

### 【発表のポイント】

- 過去のデータ利用を簡易にし、迅速に有効な新規抗菌薬候補の発見を可能とする置換基<sup>(注1)</sup>に着目した記述子（Combined substitute number: CSN）<sup>(注2)</sup>を提案しました。
- CSNを用いて過去に報告された分子約1万の構造と抗菌活性データを学習し、未知の薬剤候補分子約3200万を生成・活性予測を行いました。その結果、予測値の傾向は実験結果によく一致し、本手法の有効性が示されました。
- 本手法は、有効な新規抗菌薬候補分子を迅速に見つけ出すことができ、また過去の蓄積データの利用が容易なため、抗菌薬のみならず様々な分野への貢献も期待できます。

### 【概要】

新しい抗菌薬の開発は、薬剤耐性菌<sup>(注3)</sup>（AMR）に対抗するため常に必要とされています。しかし開発コストの上昇などといった昨今の開発環境の悪化による開発ペースの低下が問題になっています。

東北大学多元物質科学研究所の笠井均教授と中辻博貴助教（研究当時 株式会社コンポン研究所）らによる共同研究グループは、このたび分子の置換基情報のみを用いた記述子（Combined substitute number: CSN）による情報処理技術の利用を提案しました。過去の抗菌薬候補分子の情報を本記述子に変換して機械学習を行うことにより、情報収集コストと計算コストを大きく低減した上で有意な抗菌活性の予測モデルを構築することに成功しました。また本記述子を用いることで、可能性の高い約3200万個の候補分子も容易に提案することができました。実際に予測の傾向と実験結果はよく一致しており、本手法の有効性が明らかになっています。

本研究成果は、2024年2月19日（現地時間）付けで、学術誌 Scientific Reports（電子版）に掲載されました。

## 【詳細な説明】

### 研究の背景

新しい抗菌薬は、抗菌薬の継続使用により薬剤耐性を獲得した薬剤耐性菌（AMR）が発生するリスクがあることなどから、継続的に開発が求められています。しかし、研究の成熟とともに開発コストの上昇やターゲット分子の枯渇から開発環境は年々悪化しており、開発速度の低下につながっています。一方で、AMRの出現は増加傾向にあり、最悪の場合2050年までにAMR感染による死亡者数は1000万人、GDPに対する影響は100兆ドルに達すると予測されています。そのため、抗菌薬の開発環境の改善は喫緊の課題となっています。

新規薬剤の開発において、機械学習などの情報処理技術を利用したマテリアルインフォマティクス（MI）の有用性が叫ばれ、昨今盛んに研究が行われています。MIを抗菌薬候補分子などに用いるためには、分子の構造・物性情報を数学的に解析可能な情報である分子記述子として変換する必要があります。この時、多くの情報を記述子に含めることで、予測精度の高いモデルの構築が可能になります。一方で、情報量が増えると計算に必要な時間や機械の性能が増大し、学習データを準備する際にも不足しているデータを集めるために膨大な労力が必要となってしまいます。また、新規化合物を提案させる際にも現実には作るのが難しいような不合理な分子を提案してしまうことも多くなります（図1上）。

### 今回の取り組み

そこで、笠井均教授らによる共同研究グループは化合物の置換基に着目した分子記述子（Combined substitute number: CSN）を提案しました。

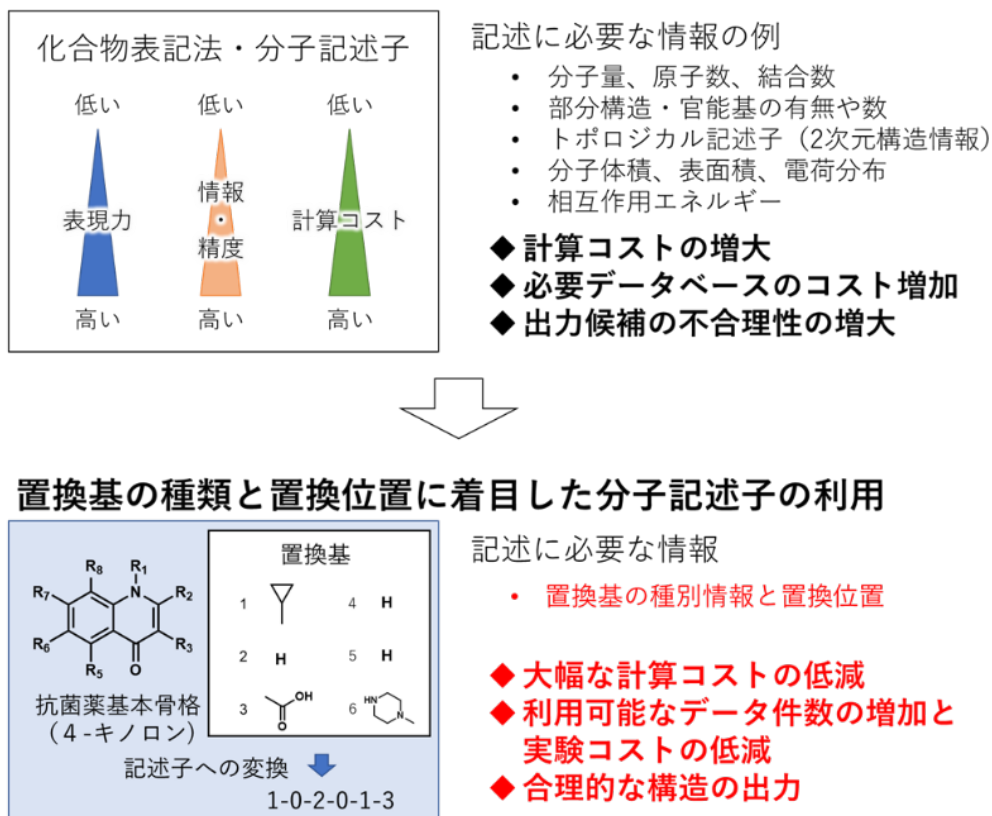


図 1. 化合物表記法・分子記述子の課題と本研究の取り組み

一般的に現在の新規抗菌薬の開発は、すでに効果の確認できている抗菌薬分子の中心骨格をベースに、側鎖の置換基の一部を変更することで抗菌活性の向上や毒性の回避が試みられています。それに着目し、基本骨格の側鎖に存在する置換基の種類とその置換位置の情報を分子の構造を表す記述子として用いる方法を提案しました (図 2a)。本記述子と抗菌活性のデータのみを用いて機械学習を行うことで、大幅な計算コストの低減を行うことが可能になります。また必要な情報を絞っているため、過去にデータベースに記録された分子の情報を、大きな労力を払わずに利用することが可能になります。また、新規な抗菌薬候補の分子を提案する際にも、置換基の組み合わせを変えることによって容易に高い実現可能性を持つ分子を提案できます。

このコンセプトを検証するため、キノロン系抗菌薬を実験対象に選びました。キノロン系抗菌薬は 4-キノロン骨格を有する抗菌薬で、その有効性から盛んに研究が行われてきた抗菌薬です。まず、4-キノロン骨格を持つ化合物 11689 個の構造情報と大腸菌に対する抗菌活性値を収集し、キノロン骨格の置換位置ごとに出現頻度の順に置換基に ID (数字) を割り当てました。次に、それぞれの化合物の持つ置換基の ID を置換位置の順に記録することで、それぞれの分子の構造の情報を記述子に変換しました。得られた分子の構造情報と抗菌活性の情報を、機械学習モデルに学習させて抗菌活性の予測モデルを構築しました。このモデルによって予測された抗菌活性は、本記述子が分子の構造の情報のみしか含まない軽量なデータであるにもかかわらず、有効な予測精度を達成しました。

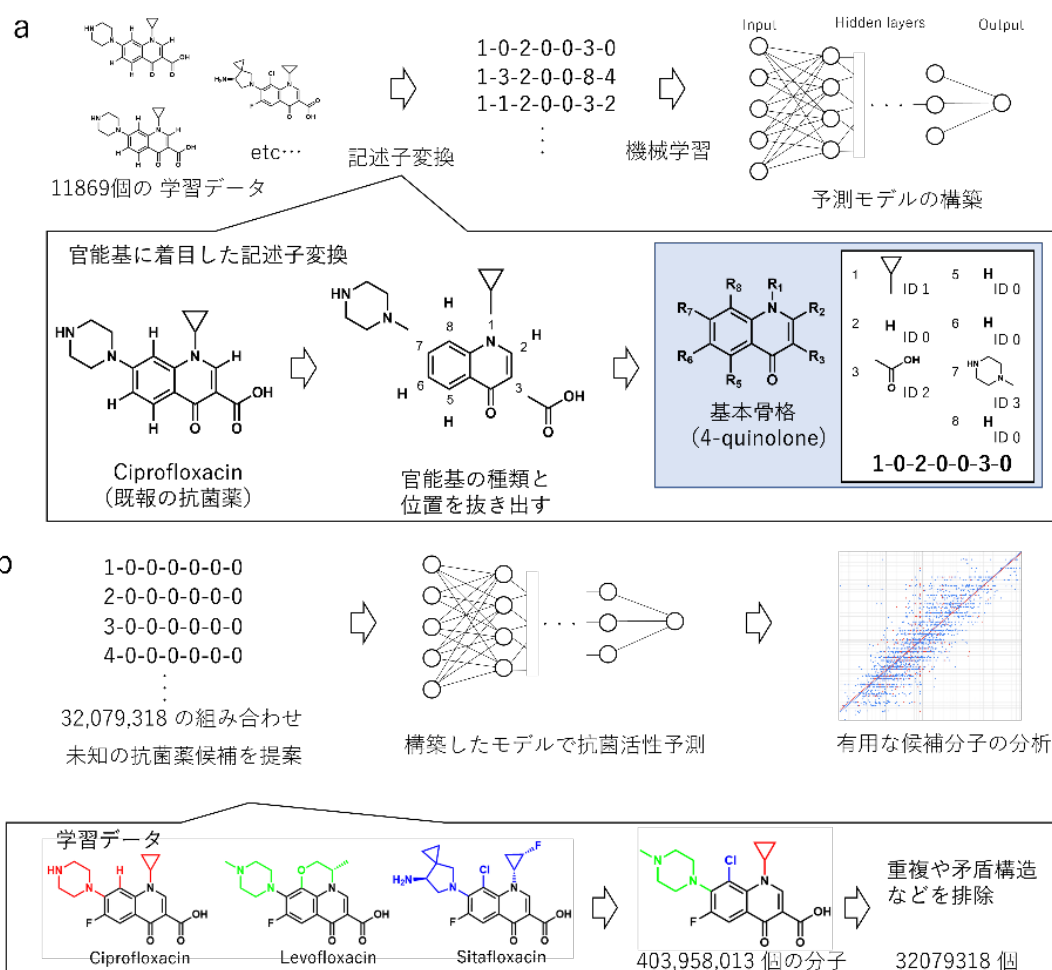


図 2. (a) 分子記述子 (Combined substitute number: CSN) を用いた機械学習によるモデル構築 (b) 未知の抗菌薬候補分子の提案と抗菌活性予測

さらに、得られた既知の置換基情報から未知化合物の組み合わせデータを生成し、約3000万種の合理性の高い化合物を提案できました。本モデルで提案された分子の予測された抗菌活性の傾向は実験で得られた結果とよく一致しており、本記述子による新規抗菌薬候補の提案の有効性が確認されました。

本研究では、キノロン系抗菌薬の置換基の種類と位置情報のみに着目した記述子を用いることで、迅速に有意な抗菌活性の予測モデルが構築できることが示されました。また、この記述子を用いることで容易に十分な数の有効な新規抗菌薬候補分子を提案できることも明らかになりました。またこの記述子は、変換に必要な情報を置換基に絞っており、十分な物性データが記録されていない過去の蓄積データも学習データとして利用することが容易であり、抗菌薬開発のみならず開発の成熟した様々な材料への機械学習モデルの導入にも貢献すると考えています。

#### 【用語説明】

##### 注1. 置換基

有機化合物において、基本骨格の持つ水素から置換された原子または原子の集団のことを指す。

##### 注2. 記述子 (Combined substitute number: CSN)

分子記述子とも言う。分子の特徴を化学構造や物理化学的性質に基づく数値ベクトルとして表現したもの。

##### 注3. 薬剤耐性菌

抗菌薬への耐性を獲得した細菌。

#### 【論文情報】

タイトル : Combined Substituent Number Utilized Machine Learning for the Development of Antimicrobial Agent

著者 : Keitaro Yamauchi, Hirotaka Nakatsuji\*, Takaaki Kamishima, Yoshitaka Koseki, Masaki Kubo, and Hitoshi Kasai\*

掲載誌 : Scientific Reports

DOI : 10.1038/s41598-024-53888-2

\*責任著者 : 東北大学 多元物質科学研究所 教授 笠井均、助教 中辻博貴

#### 【問い合わせ先】

(研究に関すること)

東北大学大学院多元物質科学研究所

教授 笠井 均 (かさい ひとし)

電話番号 : 022-217-5612

Email : kasai@tohoku.ac.jp

(報道に関すること)

東北大学多元物質科学研究所 広報情報室

電話 : 022-217-5198

Email : press.tagen@grp.tohoku.ac.jp