



Tokyo Tech



Press Release

2024年4月5日

東京工業大学

科学技術振興機構 (JST)

## 薬剤設計を支援する解釈性の高い AI 予測手法を開発

－ 持続可能な創薬を目指した合理的分子設計に向けて －

### 【要点】

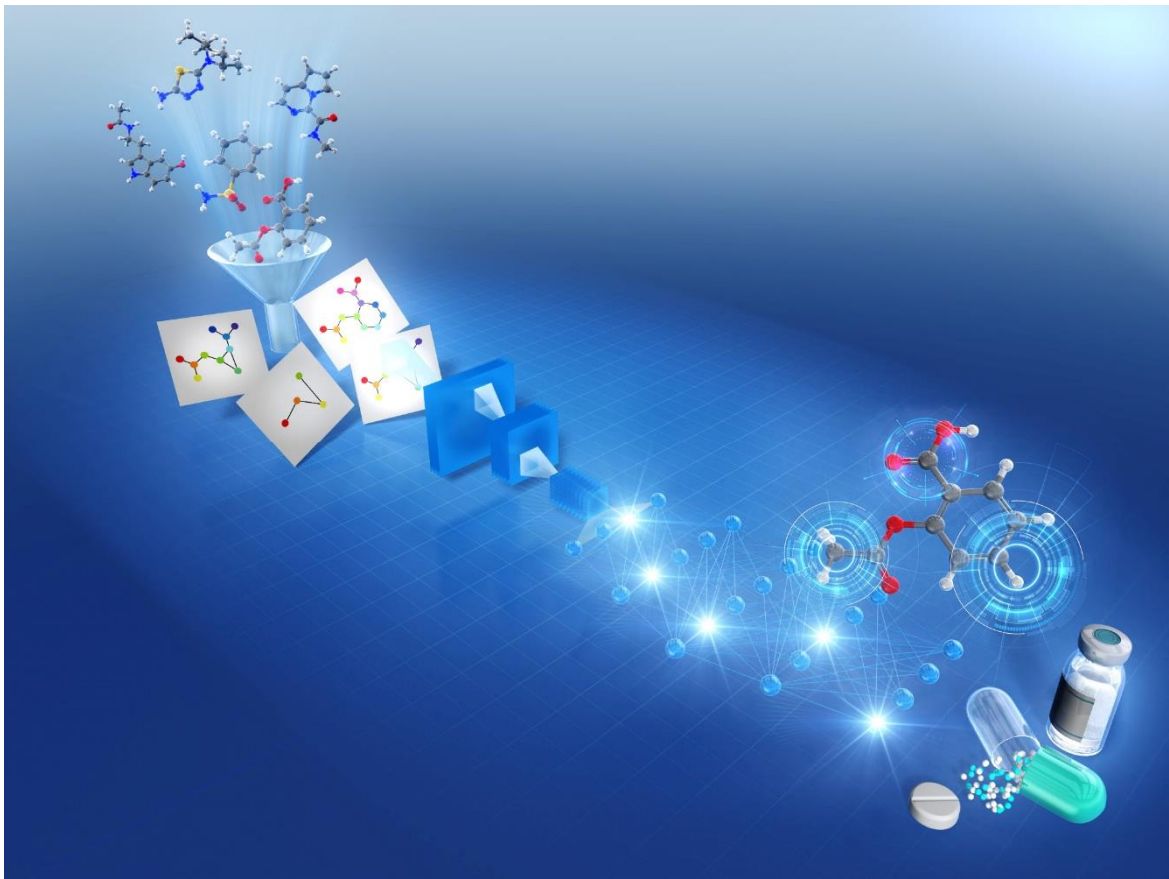
- 低分子医薬品開発のための AI 化合物設計技術 MMGX を開発
- 異なる分子グラフ表現の組み合わせにより、AI 予測結果に高い解釈性を付与
- 薬剤設計の効率化によって医薬品開発を加速させる AI 創薬の進展に期待

### 【概要】

東京工業大学 情報理工学院 情報工学系の大上雅史准教授と Kengkanna Apakorn (ケンカーンナー・アパーコーン) 大学院生は、創薬における低分子化合物の物性や活性を予測する、新たな AI 予測手法を開発した。

医薬品開発の加速を目指して、既知の実験データを活用して未知の物質の性質を予測する、人工知能 (AI) を利用した計算技術が数多く研究されてきた。特に、近年の**深層学習技術** (用語 1) の発展により、こうした AI による予測の精度は格段に上がっている。しかしその予測に至った理由を考えるための情報は乏しく、予測結果の妥当性の判断は熟練した専門家の知識と経験に委ねられる傾向にあった。

大上准教授らは、化合物の構造式を**グラフ** (用語 2) で表現して処理する**グラフニューラルネットワーク** (用語 3) に着目した。そのうえで、原子と結合の関係を表現する一般的なグラフと、化合物の複数の原子や結合 (部分構造) を 1 つのノードに縮約するグラフ表現を組み合わせ、グラフニューラルネットワークの一種である**グラフアテンションネットワーク構造** (用語 4) によって学習する MMGX (Multiple Molecular Graph eXplainable discovery) という予測手法を提案した。この手法により、化合物の物性や活性を高精度に予測すると同時に、**アテンション機構** (用語 5) を用いて部分構造表現から算出される値によって、「どの部分に着目してその予測結果としたのか」という情報を得ることができるようになった。MMGX による化合物の予測と解釈は、AI によって医薬品開発を加速させる AI 創薬の進展に大きく貢献する。この研究成果は 2024 年 4 月 5 日 (現地時間) に英科学誌「*Communications Chemistry*」でオンライン公開される。



MMGX 法によって化合物の重要な部分構造を予測・提示する AI のイメージ図

## ●背景

医薬品開発においては、開発コストの削減や時間の短縮は喫緊の課題とされている。計算機を用いて所望の性質を持つ分子を設計する方法の確立は、医薬品開発プロセスを迅速に進めるための一助となる。

これまでに、既知の実験データを活用して未知の物質の性質を予測する**教師あり学習**（用語 6）による機械学習・AI 手法が数多く研究されてきた。特に近年の深層学習技術の発展により、こうした AI による予測の精度は格段に上がってきており、創薬支援を目的として活用されている。しかし、AI の予測は「なぜその予測結果となったのか？」という理由を考えるための情報に乏しく、予測結果の解釈は熟練した専門家の知識と経験に委ねられる傾向にあり、解釈性の向上が望まれていた。

## ●研究成果

本研究では、低分子化合物の性質の予測と解釈を行う新たな AI 手法を提案するにあたり、化合物の構造式をグラフで表現して処理するグラフニューラルネットワークに着目した。原子と結合の関係を表現したグラフ（原子グラフ）のほかに、化合物の複数の原子や結合（部分構造）を 1 つのノードに縮約するグラフ表現を用いた（図 1）。化合物の性質は、化合物に含まれる部分構造によって決まるものが多く、部分構造の情報をうまく活用することで高度な予測が可能になると考えた。

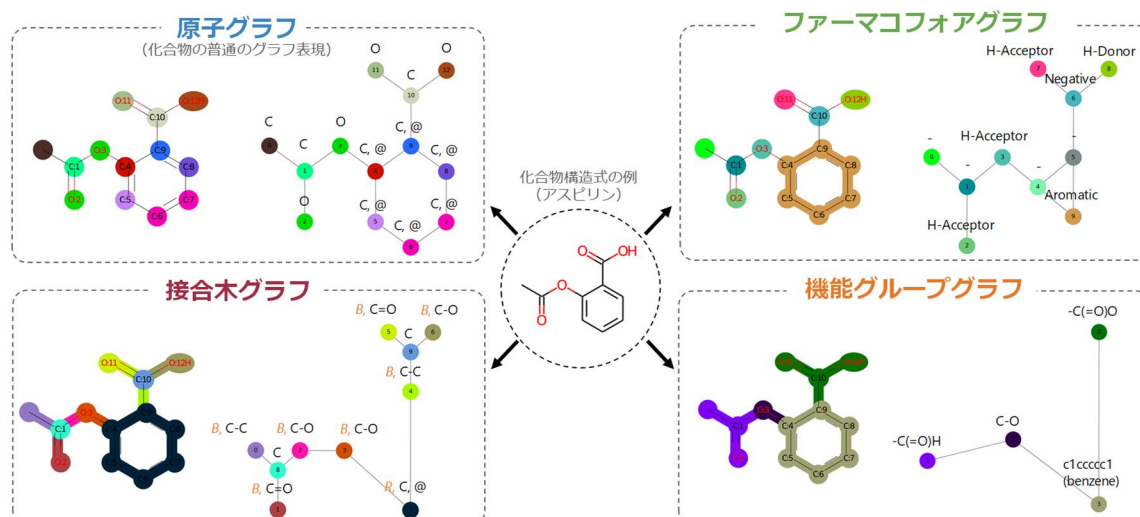


図1 化合物の構造式のグラフ表現。この図では化合物の例として、アスピリン（アセチルサリチル酸、 $C_6H_4(COOH)OCOCH_3$ ）の各グラフ表現を記載した。原子と結合の一般的なグラフ表現である原子グラフのほかに、一部の原子集団を縮約する複数のグラフ表現を用いている。

次に、グラフ表現の入力から学習を行う、グラフニューラルネットワークによる教師あり機械学習手法「MMGX」の構築を行った。このとき、原子グラフと縮約されたグラフを組み合わせる情報抽出を行う仕組みを採用した（図2）。さらに、アテンション機構を取り入れたグラフアテンションネットワーク構造を用いることで、アテンションウェイト（用語7）の値から「どの部分構造に着目してその予測結果としたのか」という情報を得ることができるようになった（図3）。

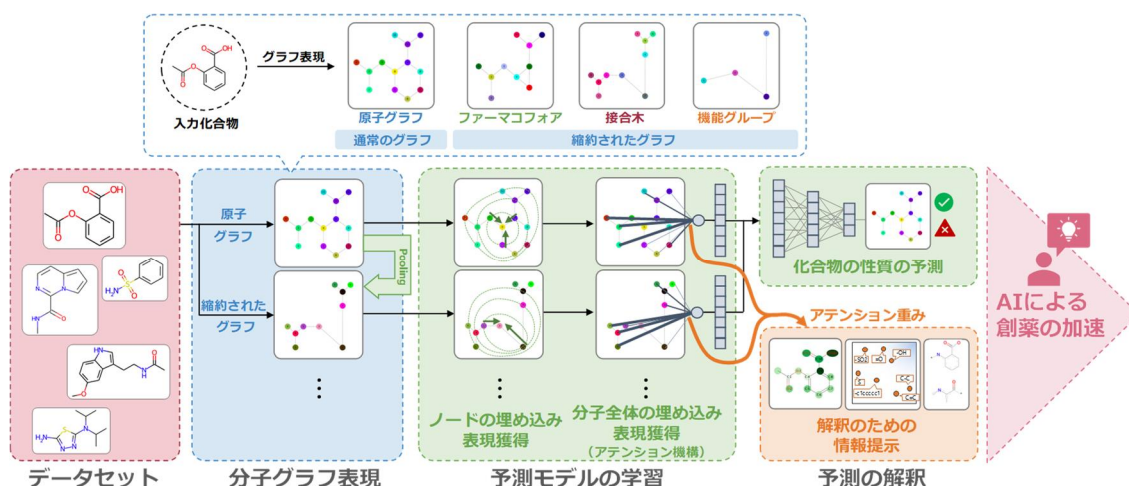


図2 提案手法であるMMGXの全体像。複数の分子グラフ表現を入力とし、グラフのノード情報を変換しながら情報抽出を行うグラフニューラルネットワークにより、化合物の部分構造に基づく性質の学習を行っていく。アテンション機構により抽出される重み（アテンションウェイト）は予測結果の解釈に用いられる。

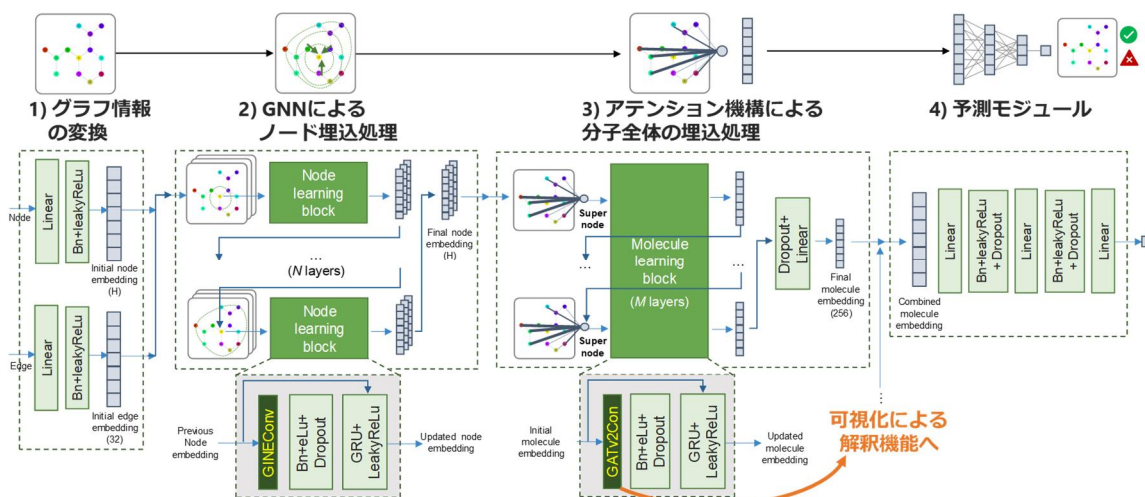


図 3 提案手法である MMGX のニューラルネットワーク構造。グラフニューラルネットワーク (GNN) とグラフアテンションネットワークによるアテンション機構を活用して、分子の部分構造の情報抽出が可能な予測モデルを構築した。

このようにして構築した MMGX について、20 種類の異なる予測タスクからなるデータセットによって予測性能の検証を行ったところ、いずれのタスクでも現時点での最高精度に比肩する予測結果を示した。さらに、アテンション機構の重み値による可視化を行い、分子の性質を解釈する方法を提案した (図 4)。実際に、**変異原性** (用語 8) の情報を集めたデータセットから学習した MMGX の予測モデルでは、化合物グラフの入力情報を用いて変異原性毒性が示唆される部分構造を複数抽出することができ、その一部は実際に文献等で変異原性が報告されている部分構造であった (図 5)。

### 3つの可視化方法の提案

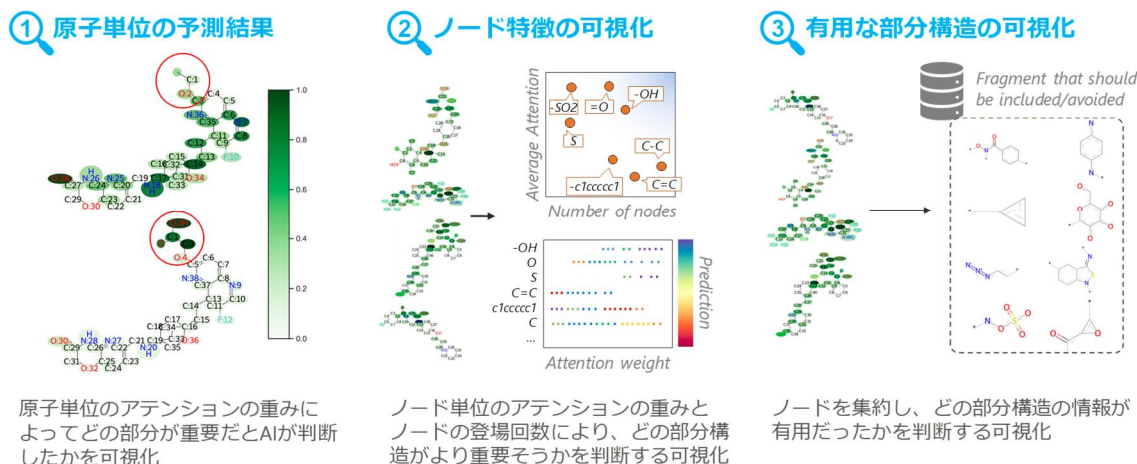


図 4 提案手法 MMGX による分子の解釈機能。



## 変異原性毒性を予測するモデルの出力 (AmesMutagデータセットによる学習)

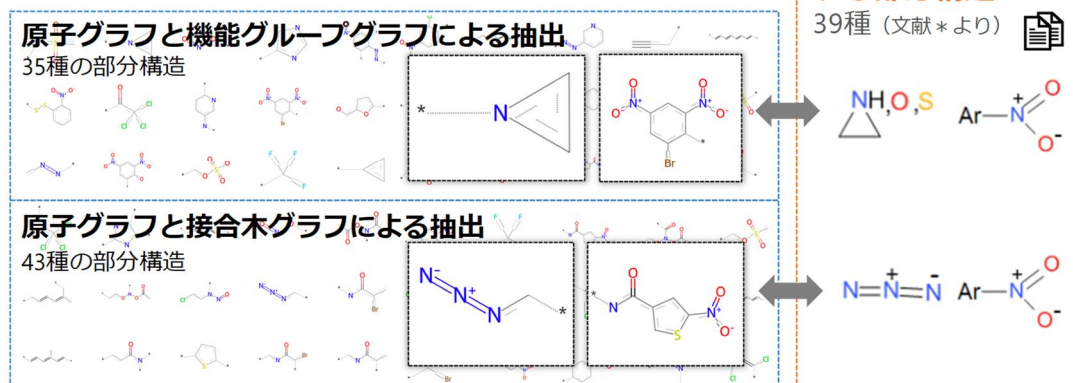


図5 MMGXによる予測例。変異原性毒性を予測するための学習を行い、予測結果を解釈するために可視化を行ったところ、既に毒性が知られている部分構造が複数抽出された。また、未報告だが毒性のある可能性が示唆される部分構造も提示された。

### ●社会的インパクト

今回発表した MMGX は、オープンソースソフトウェアとしてプログラム共有サイト GitHub からダウンロード可能である (<https://github.com/ohuelab/MMGX>)。昨今、医薬品開発プロセスの効率化を目指して、各所で AI を活用した創薬手法 (AI 創薬) の実用化が叫ばれているが、本研究成果は AI と人間が協力して高度な成果を目指す **ヒューマン・イン・ザ・ループ** (用語9) の考え方を加速するものと言える。

### ●今後の展開

研究グループは既に、今回発表した手法を改善し、より大きな分子量 (分子量 500 超) の化合物にも適用できる AI 予測手法の構築を進めている。また、分子シミュレーションによる結合自由エネルギー計算手法と組み合わせて、訓練データによる学習にとらわれずに重要な部分構造を推定できる枠組みを検討している。医薬品開発コストが増大しつつある昨今において、本研究成果や今後の研究が持続可能な創薬への一助となることを期待している。

### ●付記

本研究は以下の事業の支援を受けて実施された。

- 科学技術振興機構 (JST) 創発的研究支援事業「マルチモダリティ創薬を拓くインフォマティクス基盤」(JPMJFR216J)
- 日本学術振興会 (JSPS) 科学研究費助成事業 学術変革領域研究 (A) 計画研究「天然物が織り成す化合物潜在空間が拓く生物活性分子デザイン」(JP23H04880)
- 日本学術振興会 (JSPS) 科学研究費助成事業 学術変革領域研究 (A) 計画研究「bRO5 化合物の潜在空間構築と応用のための情報科学」(JP23H04887)

○日本医療研究開発機構（AMED） 生命科学・創薬研究支援基盤事業（BINDS）  
（JP23ama121026）

#### 【用語説明】

- (1) **深層学習技術**：ニューラルネットワークに基づいた機械学習アルゴリズムの一種。ネットワークを多層化させることで高い予測性能を得ることが可能である。
- (2) **グラフ**：鉄道の路線図のような、点や丸で表された「ノード」と、それらの間に張られる「エッジ」によって表現される数学的な構造。化合物のグラフ表現においては、原子を「ノード」、原子間の結合を「エッジ」とした原子グラフが通常よく用いられる。
- (3) **グラフニューラルネットワーク**：グラフ情報を扱うニューラルネットワークを指す。
- (4) **グラフアテンションネットワーク構造**：グラフニューラルネットワークの一種。アテンション機構と呼ばれる仕組みを採用している。
- (5) **アテンション機構**：ニューラルネットワークに入力されるデータの重要な部分を認識させるためのネットワーク構造。
- (6) **教師あり学習**：訓練データの正解を教えて学習させる、機械学習手法の枠組みの一種。
- (7) **アテンションウェイト**：アテンション機構における、入力されたデータから学習された各注目部分についての重み（重要度）。
- (8) **変異原性**：生物の遺伝情報を変化させる性質。細胞のがん化の誘発要因となることがある。
- (9) **ヒューマン・イン・ザ・ループ**：AIなどで自動化が行われているシステムなどで、一部の判断や制御に人間を介在させる仕組み。

#### 【論文情報】

掲載誌：*Communications Chemistry*

論文タイトル：Enhancing property and activity prediction and interpretation using multiple molecular graph representations with MMGX

著者：Kengkanna Apakorn, Masahito Ohue

DOI：10.1038/s42004-024-01155-w

#### 【問い合わせ先】

東京工業大学 情報理工学院 情報工学系 准教授

大上 雅史

Email: info[at]li.c.titech.ac.jp

TEL: 045-924-5530 FAX: 045-924-5523

**【JST 事業に関すること】**

科学技術振興機構 創発的研究推進部

東出 学信

Email: souhatsu-inquiry[at]jst.go.jp

TEL: 03-5214-7276 FAX: 03-6268-9413

**【取材申し込み先】**

東京工業大学 総務部 広報課

Email: media[at]jim.titech.ac.jp

TEL: 03-5734-2975 FAX: 03-5734-3661

科学技術振興機構 広報課

Email : jstkoho[at]jst.go.jp

TEL: 03-5214-8404 FAX: 03-5214-8432