

実験・理論・計算が融合したケム・インフォマティクス研究

研究代表者 清野 淳司

(先進理工学研究科 化学・生命化学専攻 准教授 (任期付))

1. 研究課題

理論・計算・実験化学と機械学習をはじめとするインフォマティクスの融合を図ることにより、計算化学と実験化学の間のギャップを埋め、計算化学の社会実装を加速させる。さらに、これまでの研究手法では困難であった化学分野の諸問題を解決へと導き、新たな研究領域ケム・インフォマティクスを定着させる。2024 年度は以下のテーマに関して研究を行った。(1) 量子化学計算に基づくスペクトル情報を用いた化合物の仮想的な自動同定、(2) 知識グラフと機械学習を利用した抗がん活性化合物提案システムの開発、(3) 係数最適化手法を導入したシンボリック回帰の開発と化学原理への適用。

2. 主な研究成果

2-1 量子化学計算に基づくスペクトル情報を用いた化合物の仮想的な自動同定

近年、ロボティクス技術やハイスループット実験、人工知能 (AI) 技術を活用した実験自動化が急速に進展している。このような自動実験において、化合物同定は構造情報の取得や物性の理解のための重要なステップである。化合物同定の自動化に向けて、これまでにデータベースや事前知識を利用し、スペクトルデータから化合物を同定する多様な手法が開発されてきた。しかし、既存の手法の多くは実験データを入力情報として用いるため、実験者や実験条件の違いなどに起因する揺らぎの影響を排除することが困難である。本研究では、膨大な化学空間の中から効率的に化合物の構造を同定するため、スペクトル情報に基づき機械学習と量子化学計算を活用した化合物自動同定システムを開発した。

本システムの流れを図 1 に示す。まず、①機械学習によりスペクトル情報から部分構造を予測する。次に、②分子量と一致するような部分構造の組み合わせを探索し、部分構造同士を繋ぎ合わせることで候補となる化合物を生成する。③生成した化合物すべてに対して安定構造を探索し、④量子化学計算により算出した仮想的な NMR スペクトルを参照スペクトルと比較する。最後に、⑤スペクトルの一致度に基づいて候補化合物をランク付けし、同定結果を出力する。

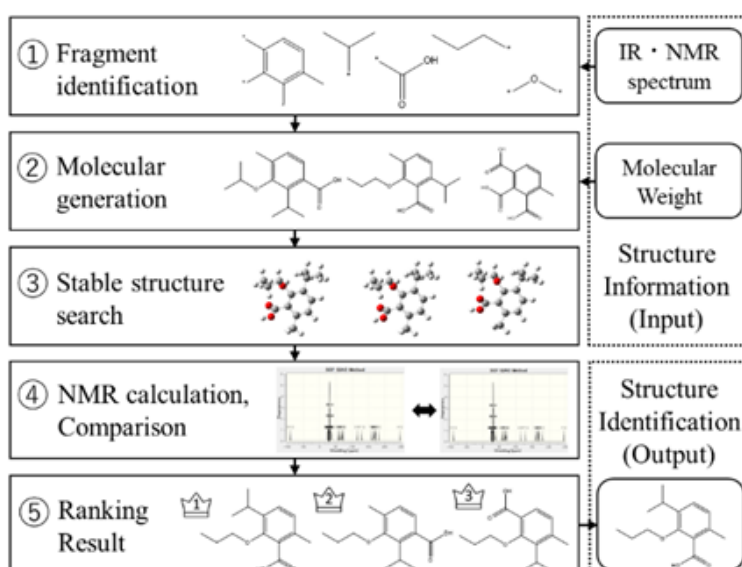


Fig. 1. Schematic flow for structure identification in this system.

図2の比較的分子量の小さい有機化合物 20 分子について化合物同定を実施した。結果、ID1 を除く 19 個の化合物に対して、正しい構造を同定し、95%という高い正解率を達成した。

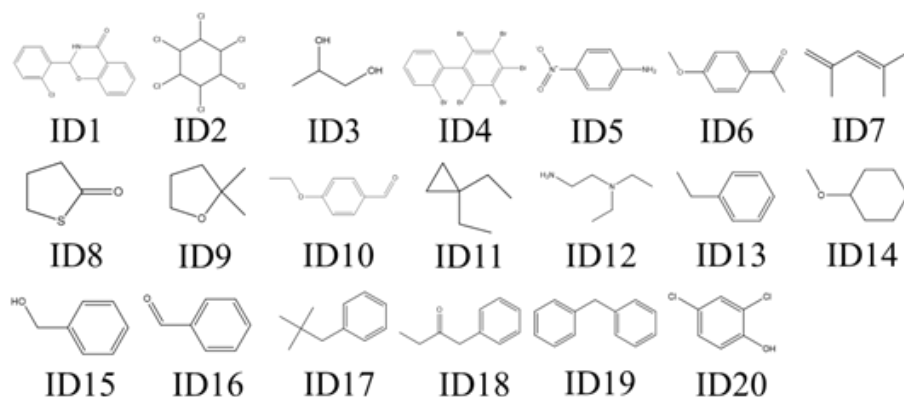


Fig. 2. Structures of 20 compounds for automated identification.

2-2 知識グラフと機械学習を利用した抗がん活性化合物提案システムの開発

急速に発展する AI 技術と創薬ビッグデータの活用により、効能・効果の高い抗がん剤の開発を加速させる技術のさらなる進展が期待されている。これまでに我々は、広範な化合物や生物活性種・細胞種に対して、抗がん活性を高精度に予測可能な AI システムを開発してきた。本研究では、現実世界の情報や単語間の関係性を視覚的かつ構造的に表現する知識グラフと当該 AI システムを組み合わせることで、作用機序を間接的に考慮しつつ、高い安全性と目的の活性を持つ可能性のある化合物を提案可能なシステムを開発した。

本システムは、目的とする活性を示す既知化合物の活性パターンをもとに、有望な化合物を選別する。まず、①抗がん活性データベースから、目的の活性パターンに類似した化合物を抽出する。次に、②それらの阻害情報や関連するタンパク質の情報をもとに、知識グラフを構築する。続いて、③構築した知識グラフに基づいて、目的の活性に関連するタンパク質を選定し、それらのタンパク質に対する阻害剤を候補化合物群として取得する。その後、④AI システムを用いて候補化合物群の潜在的な活性パターンを予測し、目的の活性パターンに類似する化合物を選出する。最後に、⑤毒性や安全性に関わる指標を予測・評価し、最終的な候補化合物を選別する。

開発したシステムにおける化合物提案の性能を評価した。図 3, 4 に、抗がん剤であるシタラビンおよびドキソルビシンの活性パターンを基準として提案された候補化合物の一部をそれぞれ示す。候補化合物のうち、ヒト経口バイオアベイラビリティ (HOB) 値が 50 を超えると予測された化合物を赤枠で囲った。これらは 60 種のがん細胞の少なくとも 1 種において、正常細胞と比較して pIC_{50} の予測値が 2 以上高かった。いずれのケースにおいても、提案された化合物のうち慣用名のあるものはすべて抗がん作用が報告されており、実際の活性パターンも基準の活性パターンと類似する。

図 3 に示したシタラビンの活性パターンを基準とした提案化合物のうち、(a)は DNA 塩基対間に挿入し、DNA ポリメラーゼを阻害することで抗がん活性を示す化合物であり、シタラビンの作用機序と一致する。また、(b)~(e)についてもトポイソメラーゼを阻害し、DNA 合成を阻害することで抗がん活性を示す化合物であることから、シタラビンと類似した作用機序を有している。

図 4 に示したドキソルビシンの活性パターンを基準とした提案化合物のうち、(c)は抗がん作用が確認されている既存化合物であり、いずれも臨床段階のフェーズ 1 において確認中または完了している段階であった。一方、(a), (b), (d), (e)は、抗がん剤としてすでに実用化されている化合物である。これらはいずれもドキソルビシンと同様に II 型トポイソメラーゼの阻害剤であり、ドキソルビシンと作用機序が一致する。中でも例えば(a)はドキソルビシンの立体異性体であり、ドキソルビシンよりも毒性が小さいことが知られている。

これらの結果は、開発したシステムによる化合物提案の妥当性を支持しており、目的とする活性を有し、安全性の高い抗がん活性化合物を提案できる可能性が示された。本システムは、新規抗がん剤の創出や既存薬の再配置への応用が期待される。

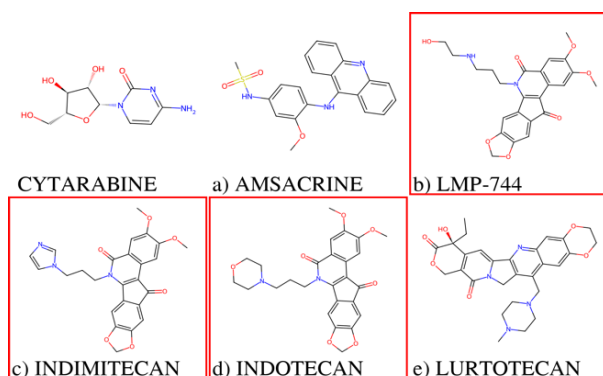


Fig. 3. Candidate compounds with generic names proposed based on activity pattern of cytarabine.

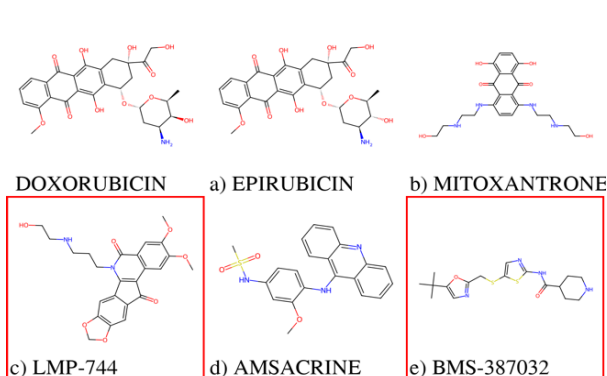


Fig. 4. Candidate compounds with generic names proposed based on activity pattern of doxorubicin.

2-3 係数最適化手法を導入したシンボリック回帰の開発と化学原理への適用

近年の AI 技術の急速な発展により、データ駆動型の数式の自動導出法による新たな法則や原理の発見が期待されている。我々はこれまで、データから解釈可能な数式を得るシンボリック回帰を化学法則の導出に適用し、その性能を検証してきた。しかし、シンボリック回帰では一般的に、数式の構造探索と係数最適化を同時に行う必要があり（図 5）、探索空間が広大で計算負荷が高く、最適解への収束困難性や局所解への捕捉といった課題がある。さらに化学法則に含まれる係数は値の幅が非常に広く、係数のスケールの多様性に対応する必要がある。そこで本研究では化学法則の自動導出精度の向上を目的に、遺伝的プログラミング（GP）に高度な係数最適化手法を導入し、その有効性を検証した。

まず係数最適化として、勾配法、メタヒューリスティクス、ベイズ最適化、その他に分類される 42 種の手法を用い、数式の形状を固定した上で、複数の化学法則に対する精度を検証した。その結果、最も精度が高かった人工蜂コロニー（ABC）アルゴリズムを GP に導入した（GP-ABC）。

GP-ABC の有効性確認のため、アトキンス物理化学に記載されている 52 種類の化学法則に対してシンボリック回帰を適用した。表 1 に化学法則を正しく導出できた割合（%）を示す。従来のシンボリック回帰手法である GP、再帰的 LASSO 型シンボリック回帰（RLS）、AI-Feynman（AIF）と比較して、GP-ABC は正解率が 69.2%と高精度であった。この結果より、シンボリック回帰においては、高度な局所・大域探索能を備えた係数最適化手法の導入が有効であることが確認された。

Table 1. Percentage of chemical laws correctly derived in symbolic regression.

Method	GP	RLS	AIF	GP-ABC
Rate [%]	36.5	56.0	50.0	69.2

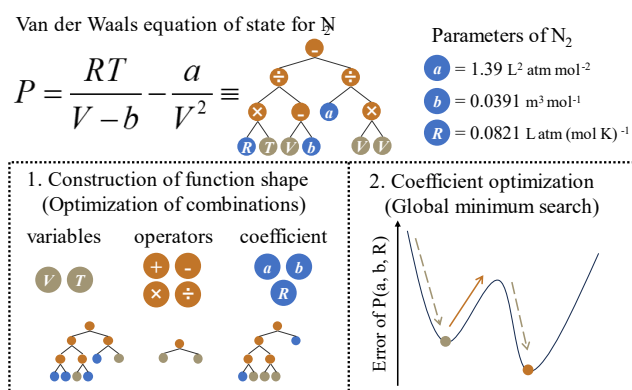


Fig. 5. Schematic image of mathematical model construction in symbolic regression.

3. 共同研究者

中尾 洋一（先進理工学部・化学・生命化学科・教授）
町田 光史（先進理工学部・化学・生命化学科・講師（任期付））
神平 梨絵（先進理工学部・化学・生命化学科・助教）
中野 匡彦（理工学術院総合研究所・招聘研究員）
中嶋 裕也（理工学術院総合研究所・客員次席研究員）
速水 雅生（理工学術院総合研究所・嘱託）

4. 研究業績

4.1 特許

- (1) “学習装置、学習方法、予測装置、予測方法、およびプログラム”、清野淳司、中尾洋一、中嶋裕也、金子武史、町田光史、神平梨絵、特願 2024-091295 (2024).
- (2) “情報処理装置、情報処理方法、およびプログラム”、清野淳司、中尾洋一、中嶋裕也、金子武史、増田幹太、町田光史、神平梨絵、特願 2025-048766 (2025).

4.2 総説・フォーラム

- (1) 清野淳司、“シンボリック回帰による化学データの解釈と外挿的な材料探索の可能性”、フロンティア、6(2), 28 (2024).

4.3 招待講演

- (1) “計算・実験・AI 技術の融合による化学研究”、清野淳司、講演会（明治大学理工学部応用化学 学科 データ化学工学研究室主催）、2024 年 7 月、明治大学生田キャンパス
- (2) “Systematic Accuracy Assessment of Quantum Chemical Calculations Using Machine Learning”, J. Seino, Asia Pacific Association of Theoretical and Computational Chemistry (APATCC11), April 21-25, 2025, Kobe (Japan)

4.4 受賞・表彰

1. 第 14 回 CSJ 化学フェスタ 2024、優秀ポスター発表賞、関屋鯨達
「機械学習・量子化学計算・多変量スペクトル分解法を用いた混合物スペクトルの分離」

4.5 学会および社会的活動

- (1) JST 研究成果展開事業 START 大学・エコシステム推進型「抗がん活性予測 AI システムの開発と検証」、清野淳司（研究代表、2024 年度）
- (2) ENEOS 株式会社との共同研究「超高速 AI 分子シミュレータの活用に向けた評価および利用法の検討(2)」、清野淳司（研究代表、2024 年度）
- (3) 文部科学省科学研究費助成金 若手研究「機械学習を利用した量子化学計算誤差データベースの開発」、中嶋裕也（研究代表、2024 年度－2026 年度）
- (4) 理論化学会 幹事、2021 年－現在
- (5) 日本化学会ケモインフォマティクス部会 幹事、2024 年－現在

5. 研究活動の課題と展望

本研究の2-1で実施してきた、化合物自動同定システムを完成させ、近年の発展が目覚ましいロボティクスと融合させることで、人の手を介さない全自動の化合物合成技術が確立される。このためには量子化学計算と実験値のギャップを埋める手法の開発が不可欠である。2-2の抗がん活性化合物提案システムの開発においては、さらなるシステムの拡張や実験実証を進めると同時に、効率的に社会実装を進めるために、事業化も考慮した活動を継続する。2-3のシンボリック回帰の開発においては、膨大に生成される化学データから化学原理を創出、つまりは化学に関する知識をAIが獲得可能とするために重要な技術となる。