

課題名(タイトル):

代謝混合物の NMR シグナルの *in silico* 同定法の高度化

利用者氏名:

○伊藤 研悟*

近山 英輔*

坪井 裕理*

理研における所属研究室名:

*環境資源科学研究センター 環境代謝分析研究チーム

1. 本課題の研究の背景、目的、関係するプロジェクトとの関係

核磁気共鳴法(NMR)は、生体試料に含まれる代謝物群を網羅的に検出することが可能であり、実測化学シフトデータベース等を用いた解析により、代謝物の同定や構造決定が可能である。しかし、現状では存在する全ての代謝物をデータベースに収録することは困難であるため、多くの未帰属シグナルの存在が多く報告されている。そのため、量子化学計算を用いた実測値に依存しない代謝物のシグナル帰属支援法が開発されているが、理論値と実測値の間の誤差は大きく、現状では援用するのに精度が不十分であるといえ、有用な理論値補正技術が必要とされている。そこで我々は、人工知能(AI)の基幹技術である機械学習を用いてこの誤差を学習させ、理論値を補正させることで化学シフト予測精度の向上を目指した。

また、NMR による代謝物群の包括的な検出と同定、構造評価に加え、物性評価も生物資源を有効活用する上では極めて重要であると考えられる。そこで我々は、非侵襲的なシグナル分離や同定、物性評価に有用な拡散及び緩和 NMR 実験を代謝物群の解析に適用し、古典力学及び量子化学といった演繹的手法により、代謝混合物の *in silico* 同定及び物性評価を行った。

2. 具体的な利用内容、計算方法

当研究室で開発された SpinAssign 用のアーカイブに登録されている 150 の低分子化合物を機械学習の訓練用化合物、34 の低分子化合物及び 16 の海藻成分を機械学習のテスト化合物とし、各化合物の

水およびメタノール溶媒中の NMR スペクトルから化学シフトの実測値を取得した。これらの低分子化合物について、PubChem ウェブサイトから立体構造情報を収集し、Gaussian09 プログラムを用いて、構造最適化、化学シフトとスピン結合定数の理論値を算出した。計算レベルは、B3LYP/6-31G* および B3LYP/6-311++G**を選択し、NMR パラメータの算出には GIAO 法を用いた。また、分極性連続体モデル(PCM)法を用いて、溶質に対する溶媒効果を考慮した。得られた実測値と理論値間の誤差を目的変数、物理化学的パラメータ群を説明変数とし、91 の機械学習アルゴリズムを探索し、最適な理論化学シフト補正モデルを作成し、理論化学シフト補正因子(SF)を算出した。補正された理論化学シフトは、実測化学シフトと比較したプロットおよび平均平方二乗誤差(RMSD)によって評価した。他の既存の予測手法(Gaussian09、Spartan、NMRShiftDB、Mnova)も同様に比較することで、本手法の予測性能を評価した。

代謝混合物の *in silico* 同定及び物性評価を行うにあたり、重水及び重メタノールを用いて生体試料中の極性代謝物群の抽出を行い、拡散・緩和 NMR 実験に供試し、拡散係数及び磁気緩和時間の実測値を得た。溶液中の運動性を理論的に評価するため、SEGWE 法及び BRW 理論から拡散係数及び緩和時間の理論値を算出した。

3. 結果

訓練データを用いた機械学習アルゴリズムの探索の結果、最良の機械学習アルゴリズムで補正した理論値と実測値の RMSD 値は、 δ 13C で約 0.3ppm、

δ 1H で 0.025ppm となった。補正を行っていない理論値と実測値の RMSD 値は、 δ 13C で約 7ppm、 δ 1H で 0.35ppm となったことから、量子化学計算に機械学習を援用することで化学シフト予測精度の大幅な向上が可能となることが示唆された。機械学習を用いて作成した予測モデルには、過学習が起きている可能性があるため、学習データセットに使われていない 34 化合物のテストデータを用いて化学シフト予測モデルの汎用性の評価も行った。量子化学計算による理論値、機械学習による予測値、本研究の量子化学計算と機械学習の組み合わせによる補正值と実測値の相関を比較したところ、本研究の手法の平均誤差が一番小さくなった。さらに、本研究の手法を用いて海藻成分のシグナル帰属を試みた。理論化学シフトは実験シグナルに対して大きな誤差を有するが、補正化学シフトは高精度に実測シグナルの位置を特定した。

また、生体試料の代謝物群の溶液中の拡散係数及び磁気緩和時間の理論値の分布は実測値に近い傾向であった。

4. まとめ

本研究で開発した量子化学計算と機械学習を組み合わせた予測法は、従来の量子化学計算のみの手法及び機械学習のみの手法よりも、精度の高い、世界最高精度の化学シフトの予測が可能であることが明らかになった。また、拡散係数及び磁気緩和時間の理論値と実測値の比較を行うことで、物性の推定及び評価が可能となることが示唆された。

5. 今後の計画・展望

化学構造と化学的・物理的特性をも補正できるようになれば、企業では分子構造解析や材料開発の時間とコストを最小限にするマテリアルズインフォマティクスへと展開することができると予想された。

平成 30 年度 利用研究成果リスト

【雑誌に受理された論文】

Kengo Ito, Yuka Obuchi, Eisuke Chikayama, Yasuhiro Date, and Jun Kikuchi. “Exploratory machine-learned theoretical chemical shifts can closely predict metabolic mixture signals” *Chemical Science* **9**, 8213–8220 (2018).

【口頭発表】

- Kengo Ito, “AI and deductive NMR approach for prediction of structure and mobility of metabolites”, 理研 CSRS 中間報告会, 2018 年 11 月 27 日
- 伊藤研悟, 坪井裕理, 伊達康博, 菊地淳, “空間・構造・運動性相関スペクトルを用いた複雑系代謝物群の俯瞰的プロファイリング技術の開発” 日本農芸化学会 2019, 東京, 2019 年 3 月 24-27 日

【ポスター発表】

- 伊藤研悟, 坪井裕理, 伊達康博, 菊地淳, “代謝混合物の化学シフト・運動性2次元相関スペクトルの解析技術高度化” 18-1 NMR 研究会, 理研・横浜, 2018 年 5 月 9 日
- 伊藤研悟, 坪井裕理, 伊達康博, 菊地淳, “複雑系代謝物群の構造・運動性相関スペクトルの俯瞰的解析技術の開発” 第 57 回 NMR 討論会, 札幌, 2018 年 9 月 18-20 日 (震災のため中止、誌面発表)
- Kengo Ito, Yuuri Tsuboi, Yasuhiro Date, Jun Kikuchi, “Development of a Comprehensive Analytical Method for Structure and Mobility Correlation Spectra of Chemical Complexity”, 理研 CSRS 中間報告会, 2018 年 11 月 27 日
- 伊藤研悟, 坪井裕理, 伊達康博, 菊地淳, “空間・構造・運動性相関スペクトルを用いた複雑系代謝物群の俯瞰的プロファイリング技術の開発” 日本農芸化学会 2019, 東京, 2019 年 3 月 24-27 日 (優秀発表選出)

【その他(著書、プレスリリースなど)】

- ”AI で世界最高精度の NMR 化学シフト予測を達成”(理研プレスリリース, 2018 年 9 月 12 日)