

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、また、それらの物性や空間的な存在分布を評価することも可能である。検出された代謝物群の同定には、一般的にデータベースを用いることが多いが、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。そのため、代謝物群の解析においては、多くの未同定シグナルの存在が問題となっている。この問題を解決する手段として、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同定支援法が着目されているものの、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考えられていた。そこで、我々はこれまでに、人工知能(AI)の基幹技術である機械学習と量子化学計算を組み合わせた化学シフト予測法を開発し、世界最高精度のシグナル同定支援法を報告した。

この背景を踏まえ、今年度は機械学習と量子化学計算を組み合わせた代謝物群のスピン結合定数の予測法の構築を目指した。スピン結合定数は化合物の立体配座を決めるうえで重要な物理的パラメータであり、高精度に予測が可能となれば、物質同定のみならず、立体構造の決定にも大きく役立つ技術となることが予想される。

2. 具体的な利用内容、計算方法

まず、当研究室で開発された InterSpin 用のアーカイブに登録されている低分子化合物群の立体構造を PubChem ウェブサイトから収集した。

Gaussian16 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*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。実測値においては、水およびメタノール溶媒中の各化合物の 2D J -res. NMR スペクトルから取得した。

得られた実測値と理論値間の誤差を目的変数、量子化学的パラメータ、分子記述子・指紋を説明変数とし、RF および XGB による機械学習を行い、理論スピン結合定数補正モデルを作成し、理論スピン結合定数補正因子(SF)を算出した。補正された理論スピン結合定数は、実測スピン結合定数と比較し、決定係数および平均平方二乗誤差(RMSD)によって評価した。また、未知の混合試料を用いて、作成した予測モデルの性能を評価した。

3. 結果

量子化学計算により算出された化合物群のスピン結合定数の理論値と実測値を比較したところ、決定係数は 0.31 であり、RMSD は 3.64 Hz となった。環状構造をもつ化合物のスピン結合定数の理論値は実測値と良く一致していたが、直鎖分子の理論値は実測値との一致度が非常に低かった。この原因は、直鎖分子の運動性に起因し、実験上はスピン結合定数がボルツマン平均化されるのに対し、計算上は静止状態の評価になるからである。次に、機械学習と量子化学計算を組み合わせて算出したスピン結合定数の予測値と実測値を比較した。比較対象の化合物は前述のものと同様で、機械学習における 3

分割交差検証時の検証データで評価した。その結果、決定係数は 0.69 であり、RMSD は 1.59 Hz となった。このことから、機械学習を組み合わせることで、スピン結合定数の予測性能が大幅に向上することが分かった。また、機械学習により算出した重要度から、この補正要因として、直鎖分子の運動性に関連する項目が浮かび上がった。最終的に、未知の混合試料を用いて予測モデルの性能を評価した結果、実測スペクトルと良く近似した予測スペクトルを得ることが可能であった。

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、化学シフトのみならず、スピン結合定数においても良好な結果を示した。検討中の課題ではあるが、現時点でも、従来法と比べ、大幅な予測精度の向上を達成している。

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。

2019 年度 利用研究成果リスト

【口頭発表】

- ・伊藤研悟, “機械学習と理論化学を組み合わせた NMR スペクトルの予測” 第 13 回メタボロームシンポジウム, 筑波, 2019 年 10 月 16-18 日
- ・伊藤研悟, “AI と演繹的 NMR による分子複雑系のマルチスケール構造・物性解析” 第5回理研－北大－産総研－物材機構 「キャタリストインフォマティクスシンポジウム」, 東京, 2019 年 11 月 12 日

【ポスター発表】

- ・X. Xu, K. Ito, J. Kikuchi, “Prediction of 2D-J parameters by combination of machine-learning and quantum chemistry computations for identification of substances in molecular complexity” Materials Research Meeting 2019, Yokohama, 10-14 December 2019