

課題名 (タイトル) :

## NMR 理論化学シフト予測法の精密化

利用者氏名 :

○近山 英輔\*

伊藤 研悟\*

坪井 裕理\*

理研での所属研究室名 :

\*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

NMR メタボロミクス用化学シフトデータベースには実験データの登録数が十分でない問題があり、第一原理量子化学計算による理論データを登録してデータベース化する期待が高まっている。その中で、計算精度の点で、現在良い手法がないことが問題になっていた。本研究は、この解決のため、高品質 NMR 実験データを有効活用した NMR 理論データ予測法の精密化を目的としている。

NMR は生体試料中の代謝物を包括的に検出することが出来るが、その同定作業は実測化学シフトデータベースに依存している。データベースは全ての代謝物を網羅することは困難であることから、未帰属シグナルの存在が多く報告されている。また、実測値に依存しない代謝物のシグナル帰属支援法として、量子化学計算や機械学習による予測法が開発されているが、実際には様々な要因から予測および実測化学シフトの差が著しいため、代謝混合物中の未帰属シグナルにおいては、さらに正確なシグナル予測技術が必要であるといえる。そこで、我々は代謝物シグナルを高精度に帰属するための新しい化学シフト予測法を開発を試みた。本手法は、量子化学計算と機械学習を組み合わせることによって達成されたと考えられた。

## 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、機械学習アルゴリズムを使用して、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 (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。予測モデルの有効性は、生体試料の実スペクトルの帰属に適用することで検証された。

## 3. 結果

機械学習による補正なしの量子化学計算の結果では、 $\delta^1\text{H}$  の RMSD が約 0.3ppm、 $\delta^{13}\text{C}$  の RMSD が約 7ppm となり、補正した場合は、 $\delta^1\text{H}$  で 10 倍程度、 $\delta^{13}\text{C}$  で 10 倍以上の改善が見られた。既存および本研究の化学シフト予測法を比較し、その予測性能を評価した結果、既存の化学シフト予測法で最も RMSD が低かった Mnova の結果と比較したところ、RMSD は  $\delta^1\text{H}$  および  $\delta^{13}\text{C}$  で 1/10 程であった。量子化学計算後にボルツマン分布に基づいた補正を行っている Spartan の結果は、補正をしない結果よりも誤差が小さくなるが、機械学習程の補正効果を得ることが難しいと考えられた。NMRShiftDB を用いた  $\delta^1\text{H}$  の予

測結果の RMSD は、他の手法と比較して最も大きくなった。これらの結果から、量子化学計算と機械学習を組み合わせた化学シフト予測法は、従来の手法に比べて、大幅な予測精度の向上を期待できることが示唆された。最後に、生体試料中の成分の帰属に化学シフト予測モデルを適用することで有効性を検証した。我々が以前に報告した海藻の代謝物群の実測化学シフト、帰属支援に使用した理論化学シフトおよび本研究の手法から得た理論値補正化学シフトを比較した。本研究で開発した手法は、補正なしの手法に比べて実測スペクトルに近い疑似シグナルが得られ、代謝混合物の帰属支援の性能が向上したといえた。

#### 4. まとめ

本研究における、量子化学計算と機械学習を組み合わせた化学シフト予測法は、従来の経験的および非経験的な化学シフト予測法に比べ、高精度に予測が可能であった。また、作成した予測モデル式を代謝混合物のシグナル帰属に適用することで、その有用性が検証された。

#### 5. 今後の計画・展望

学習データセットや説明変数を充実化することで、実測値に匹敵するほどの高精度な化学シフト予測法に改良することが可能であると考えられる。標品の入手や実測値の取得が困難なために帰属を行うことが出来なかったシグナルも、本手法を用いることで、帰属が可能となることが予想された。

平成 29 年度 利用研究成果リスト

【国際会議、学会などでの口頭発表】

- ・ K. Ito, Y. Obuchi, E. Chikayama, Y. Date, J. Kikuchi, Predictive modeling of chemical shift using quantum chemical calculation and machine learning, EUROMAR 2017, Poland, 2-6 July 2017
- ・ 伊藤研悟, 尾瀨由佳, 近山英輔, 伊達康博, 菊地淳, 量子化学計算と機械学習を用いた化学シフト予測法の高度化, 第 56 回 NMR 討論会, 東京, 2017 年 11 月 14-16 日
- ・ 伊藤研悟, 坪井裕理, 伊達康博, 菊地淳, 拡散 NMR と演繹的手法による代謝物群の *in silico* 同定法の高度化, 日本農芸化学会 2018 年度大会, 名古屋, 2018 年 3 月 15-18 日