

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*、小松 功典*

所属 : *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。そのため、「NMR メタボロミクス用理論化学シフトデータベースの構築実験」の課題を進めていたが、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指す。具体的には、新規予測アルゴリズムを開発する。それは、クエリー化合物に対し、遮蔽定数を Gaussian で計算、部分構造に分割し、あらかじめ Gaussian により作成した部分構造-理論遮蔽定数-NMR 実験値データベースに対し照合する。このデータベースは、数百の化合物の計算結果の構築によるため、スーパーコンピュータを用いた第一原理量子化学計算が必要である。ヒットした部分構造から重回帰分析等の手法で検量線を作成することで精密化を達成する。実用としては、¹H 化学シフトの実験値と理論計算値との誤差が 0.05 ppm 以内であることが必要と見積もっているが、この値は文献上の様々な化合物の予測精度の上限より 2 倍程度高精度のため、現段階ではまだ困難であると考えられる。そこで、¹H 化学シフト予測精度の目標値を 0.10 ppm (実験値との誤差の標準偏差) とする。

2. 具体的な利用内容、計算方法

新規アルゴリズムの予備調査には、HOKUSAI GW において、`rscunit=gwacsl`、`rscgrp=gaussian` を用いた。エネルギー最小化は `Opt` 命令をデフォルトパラメーターで `%Chk` 指定で用いた。`Opt` における理論レベルは `B3LYP/aug-cc-pVDZ` を用いた。分子力学は Born-Oppenheimer 分子力学モデルよりも計算負荷の低い、Atom Centered Density Matrix Propagation 分子力学モデルを実行するために `ADMP` 命令を用いた。

時間ステップは 0.1 fs を用いた。`ADMP` の計算負荷を抑えるため、予備分子力学としては、`RHF/STO-3G` を用い、引き続き構造緩和のための分子力学として、`B3LYP/aug-cc-pVDZ` を用いた。遮蔽定数計算は `NMR` 命令をデフォルトパラメーターで用いた。理論レベルは `B3LYP/aug-cc-pVDZ` を用いた。

混合物の多様な多次元 NMR スペクトルを用いたネットワーク理論に基づく半自動連鎖帰属では、グループ化されたピーク群を用いてグラフ理論に基づき構造を推定し、化合物を予測した。推定した化合物の確からしさを検証するため、量子化学計算を用いて化学シフトの理論値を算出した。標準物質の混合試料中の 12 成分および基準物質の計算を行った。海藻類の 18 の代謝物 (糖、アミノ酸、有機酸など) の計算を行った。`Grid search` 法を用いて各化合物の配座を複数発生させ、半経験的分子軌道法を用いて構造最適化を行うと共にエネルギー値を算出し、エネルギー的に安定した上位 10 配座を抽出し、`B3LYP/6-311++G**` レベルで構造最適化を行い、同じ計算レベルで `GIAO` 法を用いて化学シフト値を算出した。また、`SCRF` 理論に基づいた溶媒効果 (水) を適用した。イオン化を考慮しているものといないものの両方を計算し、また、ボルツマン分布および回帰分析を用いて補正を行った。

3. 結果

新規予測アルゴリズムでは、構造アンサンブルから理論化学シフトを計算する手法を用いる。構造アンサンブルの生成のため分子力学法を用いる。そのため、HOKUSAI GW 上で Gaussian を連続動作させるための予備調査を行った。Gaussian 入力ファイルからエネルギー最小化、予備分子力学、分子力学、理論遮蔽定数計算を連続して行い、生成された構造アンサンブルに対する各原子の遮蔽定数を抽出するフローを調査した結果、エネルギー最小化、予備分子力学、分子力学、理論遮蔽定数計算の部分は、HOKUSAI GW の Gaussian 上でジョブとして一括投入できることが分かった。ま

たフロー全体を自動化するには、HOKUSAI 外の部分で初期時の PubChem の SDF ファイルからの Gaussian 入力ファイル、および HOKUSAI ジョブ投入ファイルの自動生成と、終了時の Gaussian 結果ファイルからの理論遮蔽定数計算部分を自動抽出するプログラムを用意すればよいことが分かった。Gaussian では、エネルギー最小化として Opt、第一原理分子動力学として ADMP、遮蔽定数計算として NMR を用いることが可能であることが分かった。これらをチェックポイントファイルを経由して繰り返し動作させることにした。1 分子につき投入するジョブファイルは以下のように記述することにした。

```
#!/bin/sh
#----- pjsub option -----#
#PJM -L rscunit=gwacs1
#PJM -L rscgrp=gaussian
#PJM -L vnode=1
#PJM -L vnode-core=8
#PJM -L elapse=23:55:00
#PJM -g Q15279
#---- Program execution ----#
module load gaussian
g09 opt.com
g09 nmr0.com
g09 admpl.com
g09 admph.com
g09 nmr1.com
g09 admpl.com
g09 admph.com
g09 nmr2.com
. . .
g09 admpl.com
g09 admph.com
g09 nmr100.com
```

ここで#PJM 部分は実行時に変更する。後半の opt.com がエネルギー最小化で最初の 1 回のみ行う。その後、admpl.com が予備分子動力学、admph.com が分子動力学、nmr*.com が遮蔽定数計算であり、例えば上記では 101 回の遮蔽定数計算を行う。admpl は計算負荷の低い理論レベルを用い、一方、admph は目的の理論レベルを用いる。今回のプロジェクトでは、admpl は RHF/STO-3G、admph は B3LYP/aug-cc-pVDZ を想定している。チェック

ポイントファイルは全て同じファイル名で上書きしながらシリアルに実行されることも確認できた。admpl と admph をこのように分割して行う理由は、#T で出力を抑えても大量の計算結果が出力されるため、不要な分子動力学の計算結果ファイルによるディスク容量の使用を 1/100 程度に抑えるためである。このように、Gaussian 入力ファイルからエネルギー最小化、予備分子動力学、分子動力学、理論遮蔽定数計算を連続して行い、生成された構造アンサンブルに対する各原子の遮蔽定数を抽出するフローの実現に近づいたが、一方で、第一原理分子動力学の計算負荷が高く、前年度までに調査していた 10 ns レベルの分子動力学シミュレーションが小さな ethanol 分子でさえ、困難であることも分かった。これは、並列コア計算でも計算時間の短縮が判定できなかったこともあり、今後の調査を要する。

理論化学シフトデータベースを今後整備してゆくためには、NMR 実験スペクトルに対し、原子帰属を行わなければならない。そのために、海藻類の混合成分に対する NMR の実測データと Gaussian で計算された化学シフト値の比較解析を行った。海藻成分の 18 の予測した代謝物の化学シフトの計算値と実測値を比較した結果、誤差の指標である RMSD 値の平均は、0.317ppm (δ 1H) と 4.773ppm (δ 13C) であった。これは、イオン化を考慮した計算値と実測値の RMSD 値の平均よりも低かった。ボルツマン分布による補正結果では、RMSD 値の平均は、0.287ppm (δ 1H) と 4.716ppm (δ 13C) となり、回帰分析による補正結果では、0.302ppm (δ 1H) と 3.483ppm (δ 13C) であった。全て、RMSD 値 (δ 1H) は 0.9ppm 以下に収まる結果であった。イオン化を考慮した場合の RMSD 値 (δ 13C) の最大値は 12.438ppm であり、考慮しない場合は 6.713ppm であった。また、95%予測区間を用いた評価では、ほぼ全ての結果が区間内に入った。そのため、差が有意の場合は、化合物予測の失敗に対応づけてきた。これらの結果から、量子化学計算を用いた代謝混合物の理論的な帰属の可能性が示され、さらに有効な補正方法や高い計算レベルの導入により、精密な評価が可能であると考えられた。

その他では、2次元 J 分解法と呼ばれる多次元 NMR 計測を利用したメタボローム用データベースを公開した (Kikuchi et al. *Anal. Chem.*, 2016) ため、スピン結合定数の計算についても本課題に関連してくる可能性

があった。2次元J分解法は、化学シフトおよびスピン結合定数の2次元相関を検出する。そのため、理論化学シフトに加え、理論スピン結合定数の蓄積が有用になる。そこでB3LYP/6-31G* レベルにおける芳香族化合物の理論スピン結合定数の蓄積を行った。24の芳香族化合物に対して、理論スピン結合定数の蓄積を行った。共通部分構造に着目し、その理論スピン結合定数から、カーネル密度推定を使って、その分布を推定した。得られた推定分布を用いて実測の代謝産物スペクトルから特定の部分構造を推定することができるかどうかの検証の必要性が見いだされた。

4. 今後の計画・展望

新規予測アルゴリズムは、予測対象原子を中心原子とした部分構造からの重回帰による予測を想定している。そのアルゴリズムの概要は以下である。まず部分構造に注目する。部分構造は結合1階層目のみを考慮する。この部分構造と同じ部分構造を持つ分子の部分構造をデータベースから検索する。その結果、4個の部分構造がヒットしたとする。データベースは、Gaussianによる様々な分子に対する遮蔽定数の計算結果のデータベースであるため、ヒットした4個の部分構造の全ての原子1個につき、1個の遮蔽定数が登録されている。従来の簡単な予測値の計算法では、ヒットの中で、単に遮蔽定数の平均値を使用するものがあつたであろう。つまり、遮蔽定数の平均値を計算し、その平均値を単回帰の直線で化学シフトに変換して化学シフト予測値とする。しかし、今回の新規アルゴリズムでは、部分構造の5つの構成原子の遮蔽定数を用いて計算する。つまり、1番目の部分構造の5個の遮蔽定数、2番目のそれ、3番目のそれ、4番目のそれというように用いる。部分構造1個に対し定まるこの遮蔽定数の組を、遮蔽定数ベクトルと、以降呼ぶ。データベースには、実験化学シフトが登録されている。予測対象原子の実験化学シフトはデータベースに登録されていることが前提になっている。この実験値と対応する遮蔽定数ベクトルを、例えば

$$x_1=(x_{11}, x_{12}, x_{13}, x_{14}, x_{15})$$

$$=(193, 197, 31.2, 31.7, 31.1)$$

$$x_2=(x_{21}, x_{22}, x_{23}, x_{24}, x_{25})$$

$$=(191, 207, 32.1, 32.1, 31.9)$$

$$x_3=(x_{31}, x_{32}, x_{33}, x_{34}, x_{35})$$

$$=(187, 188, 31.5, 31.5, 31.6)$$

$$x_4=(x_{41}, x_{42}, x_{43}, x_{44}, x_{45})$$

$$=(184, 194, 31.1, 31.6, 32.6)$$

のようにし、対応する実験値である目的変数を y_i 、説明変数を x_{i1}, \dots, x_{i5} とした重回帰分析を行う。その結果、 $y=Ax+B$ の係数ベクトル A と切片 B を決定でき、これが部分構造に対応する遮蔽定数ベクトル x から、理論化学シフト予測値 y を計算する式になる。部分構造の遮蔽定数ベクトル x は、データベースに登録されているものとする。つまり、この分子に対し、Gaussianで前述のエネルギー最小化、(予備+)分子動力学、構造アンサンブルに対する遮蔽定数が計算され、データベースに各原子の遮蔽定数の平均値がすでに登録されているとする。これを図で表すと、図1のようになる。独立変数の5個の x 軸は、部分構造の5個の原子の遮蔽定数を表し、その任意の値の組に対し、1個の理論化学シフト予測値が縦軸の y として定まる。この関数を描けば、図1の平面のようになる。

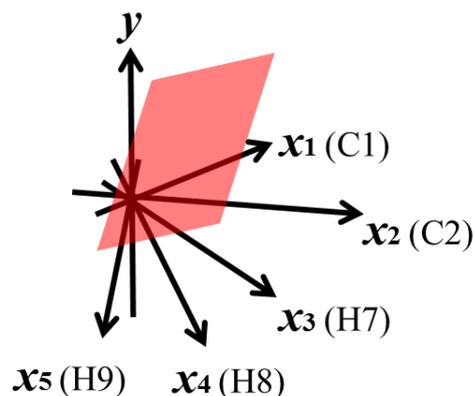


図1 重回帰による回帰式
括弧内は部分構造の原子名の例

平成 27 年度 利用研究成果リスト

【論文、学会報告・雑誌などの論文発表】

1. Ito, K., Tsutsumi, Y., Date, Y. and Kikuchi, J. Fragment Assembly Approach Based on Graph/Network Theory with Quantum Chemistry Verifications for Assigning Multidimensional NMR Signals in Metabolite Mixtures. ACS Chem. Biol. 2016. (in press)

【国際会議、学会などでの口頭発表】

1. 伊藤研悟、代謝混合物のフラグメント集積法による構造生成とその理論化学シフト・スピン結合定数の評価、第 54 回 NMR 討論会、2014.11.6-8、千葉