平成 23 年度 RICC 利用報告書

課題名（タイトル）：

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

利用者氏名：近山 英輔
理研での所属研究室名：
横浜研究所 植物科学研究センター メタボローム研究推進部門 先端 NMR メタボミクスチーム

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

植物科学研究センター先端 NMR メタボミクスチームでは NMR では世界最高度のメタボロミクス解析プラットフォームを構築してきた。そこでは代謝物の標準化合物による化学シフトデータベースが重要な要素となっており、現在 270 代謝物以上の実験的計測データが蓄積されている。しかし、代謝物総体は例えば植物で 20 万種以上あると言われており、その全ての化合物を入手することはできない。また、例え入手できたとしても、その全てを NMR で計測することは不可能である。従って、計算機を用いて大量の代謝化合物についてその化学シフト予測値を計算し、データベース化することは、予測値の精度を犠牲にしても、未知試料の候補代謝物同定時に非常に有用となる。RICC には Gaussian が利用可能となっており、Gaussian では NMR 化学シフトの理論値が複数の量子化学計算アルゴリズムで計算可能となっていることから、本課題の最終目的は、量子化学計算による大量の化合物の化学シフトデータベースを構築することである。しかし、その過程においても問題が存在する。継続 2 年目の今回は、前年度行った計算を、29 個の化学分子の系に拡大し、より計算量の多い MP2 計算を追加して評価することを目的とした。

近年、様々な post-Hartree-Fock(post-HF)法／密度汎関数法(DFT)による遮蔽定数の高負荷計算が可能になってきているものの、実験値に比する高精度の遮蔽定数計算法を確立するためには、メソッド／基底セットを含む理論レベルの選定、溶媒効果、分子構造のアンサンブル、実験値との補正法など数々の要因を最適化する必要がある。特に多数の分子構造アンサンブルと第一原理計算による化学シフト予測値の正確性との関係についての研究例はまだ少なく、その検討を行った。

２． 具体的な利用内容、計算方法

前年度の報告書の今後の計画・展望の箇所に記載したように、29 化合物で、理論レベルとして MP2 を加えた系の計算を行った。具体的には、各々の化合物で 10 ns の古典定温定積分子動力学(MD)計算により生成された 100 構造に対し、HF/B3LYP(DFT)/MP2 のメソッドと 6-31G/aug-cc-pVDZ の基底セットによる理論レベルを用いた第一原理量子化学計算により遮蔽定数を求めた。主成分分析による遮蔽定数・分子構造相関解析を行った。各化合物は、化合物ファイルで与えられている初期座標から、エネルギー最小化を真空中で行い、1 ns の平衡化の後、10 ns の MD シミュレーションにかけられた。MD に用いた力場は CACTUS (http://cactus.nci.nih.gov/)でエタノール分子の Mol2 フォーマット(Tripos)ファイルを取得するか、又は PubChem(http://pubchem.ncbi.nlm.nih.gov/)にて取得した sdf ファイルを Open Babel(http://openbabel.org/wiki/Main_Page)を用いて Mol2 に変換し，acpype (http://code.google.com/p/acpype/) を用いて GROMACS 用 GAFF 力場を生成した。MD は倍精度の GROMACS 4.0.5 で修正 Wolf 法を静電力計算に用い、真空中、298.15 K の定温シミュレーションを Nose-Hoover 法で行った。100 ps 毎に各化合物の分子構造をサンプルし、各化合物につき計 100 構造を得た。MD で得られた各構造の座標についての遮蔽定数の第一原理計算は、RICC 上の Gaussian09 で HF、B3LYP、MP2 の 3 種の理論レベルと 6-31G、aug-cc-pVDZ の 2 種の基底セットについて計算した。故に 1 理論レベル、1 基底セットにつき 100 構造の各化合物の分子内原子の遮蔽定数を得た。計 2900 構造になった。29 化合物は、in-house で作成している化学シフトデータベースの中から、分子量 200 以下の化合物を全て選択した。

３． 結果

29 化合物×100 構造に対し得られた全ての遮蔽定数を各原子毎に標準化し、得られた標準得点を正方向を赤、負方向を青で可視化すると、大まかには理論レベル／基底セットごとに赤、青が色分けされる結果となった。これは、遮蔽定数の計算では理論レベル／基底

セットの選択による遮蔽定数のオフセットの寄与が大きく、100 構造の分子動力学的なゆらぎによる遮蔽定数のずれが少ないことを意味しており、このことは前年度報告した 2 分子の結果と整合している。この可視化した遮蔽定数のデータは正確には、29 化合物 536 原子分の遮蔽定数を含んでおり、結果的に、536×100×5(理論レベル)＝268000 の遮蔽定数を含んだ行列である。MP2/aug-cc-pDZ の結果は膨大な計算量が必要であったため、この行列を計算した時点で 29 化合物分の結果を得られていなかったため、計算に入れていない。この行列を主成分分析(PCA)した結果、PC1/PC2 平面を用いて、5 つの理論レベル/基底セットのクラスタが明確に分離していた。このことは先の可視化の結果と整合している。また、今後、遮蔽定数を適切な検量線を用いて化学シフトに変換し、このオフセットを調整すれば、前年度に得た結論である、理論レベルによる変動よりも、分子動力学計算由来の構造のゆらぎによる変動の方が化学シフトを大きく変化させる、という結論を得られると期待できるものになった。

全 29 化合物の MP2/aug-cc-pVDZ の計算には、膨大な計算時間がかかったが、RICC の性能により、全て計算を終了することができた。一時ファイルやメモリの制限などの関係で、初期のスクリプトでは、効率良く分子量が 200 に近い相対的に原子数の多い化合物の MP2/aug-cc-pVDZ 計算ができなかったことで、入力ファイルを分割してサブミットするスクリプトの開発などが必要となり、予定よりも大きな作業時間を要してしまい、まだ論文用のデータ解析まで進んでいないのが現状である。論文のためには実験値との比較が必要であり、一部不十分な実験値しか現在持っていないため、NMR の実験を今後合わせて行う必要性が生じており、その箇所は今後の論文投稿の律速段階になってしまうかもしれない。

４．　まとめ

今回の結果は、前年度得た 2 化合物での結論を 29 化合物へ拡張した結果、理論レベル／基底セットのオフセットを調整すると、遮蔽定数から検量線で変換される化学シフトのゆらぎは、分子のダイナミクスによる構造のゆらぎが主になる、という同じ結論を得られると期待できるものになった。その結論は、現在計算が終了しているデータのさらなる解析によらなければならない。前年度と同様に、今後理論化学シフトデータベースを構築する際に、構造のアンサンブルを以下に効率的に生成するかという問題が一般化されたことを意味している。

５．　今後の計画・展望

前・今年度を通じて、理論レベル／基底セットの選択と、分子のダイナミクスによる構造のゆらぎが、化学シフトの理論値計算にどのように影響するか、という問題を追及してきたが、この結論は今年度得た計算の今後のさらなる解析によって決定する予定である。従って、今後の RICC を用いた計算は次のフェーズに入る予定である。即ち、次のフェーズとは、課題のタイトルにもなっている、大量の代謝化合物について、化学シフトを計算し、NMR メタボロミクス用理論化学シフトデータベースを構築することである。この大量計算を遂行する前に、上記の結論から、最適な理論レベルと基底セットを選択しなければならない。

６．　RICC の継続利用を希望の場合は、これまで利用した状況（どの程度研究が進んだか、研究においてどこまで計算出来て、何が出来ていないか）や、継続して利用する際に行う具体的な内容

理論レベル／基底セット／構造ゆらぎが理論化学シフトに与える影響の調査のための 29 化合物の計算は全て終了した。

継続利用する際は、数万の代謝化合物の理論化学シフト計算のための、今期と同様の手法による遮蔽定数の計算を行う予定である。ただし、最適な理論レベル／基底セットの選択が、データ解析が遅れていることから、まだ未決定であるため、それが終了した後になる。

平成 23 年度　RICC 利用研究成果リスト

【国際会議、学会などでの口頭発表】

E. Chikayama, T. Mori, T. Iikura, Y. Date, J. Kikuchi, “THEORETICAL CHEMICAL SHIFT DATABASE AND STRUCTURAL INVESTIGATION OF CELLULOSE BY NMR AND SUPERCOMPUTER”, Metabolomics 2011, June 27-30, 2011, Cairns, Australia