

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, 伊藤 研悟

所属 : 環境資源科学研究センター 統合メタボロミクス研究グループ 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

環境資源科学研究センター環境代謝分析研究チームでは、NMR で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 800 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。

2. 具体的な利用内容、計算方法

昨年度までに用いた手法で解析が終了していない部分の精密化を行った。また、応用として、海藻類の未知成分の帰属法の研究も進めている。具体的には、本継続課題では、29 化合物に対し、RICC の計算リソースを用いて、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ている。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 構造を得ている。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ている。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ている。計 2900 構造である。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ている。

また、海藻類の未知成分の帰属法の研究については、方法論の確からしさを検証するにあたり、混合試料の既知の 12 成分および標準物質を対象に Gaussian を使用して構造最適化および遮蔽定数・スピン結合定数を算出し、実測値との比較を行った。この時の計算レベルは B3LYP/6-311++G**であり、真空中および SCRF 理論に基づいた溶媒効果を取り入れて計算を行っ

た。しかし、連続誘電体近似では、溶媒を分子として取り扱っていないため、溶媒和構造が得られない。実際に溶媒分子を考慮して計算することにより精度が向上すると考え、QM/MM 法を用いて計算を行った。また、水溶媒中では、分子がイオン化している可能性があり、イオン化を考慮した計算も行った。QM/MM 法およびイオン化を考慮した計算には混合試料中のギ酸を対象に行った。QM/MM 法では、溶質を B3LYP/6-311++G** レベルで、溶媒(水分子 20 個)を UFF 力場で計算を行った。また、安定構造の存在確率による化学シフトの補正を行うためエネルギー計算を行った。イオン化していないギ酸は安定構造が 2 つしか存在しないと考えられたため MD による計算は行わず、2 つの安定構造についてエネルギー計算を行い、ボルツマン分布により存在確率を算出後、加重平均により化学シフトを補正した。

3. 結果

29 個の化合物に対し、6 つの理論レベルによる計算が終了した。昨年度の計算結果において、Ascorbate の構造ファイルを取り違えていたことが今年度判明し、正しい構造ファイルを取得して、計算をやり直した。さらに、昨年度終了していなかった、NMR 実験値の原子帰属を、29 の化合物に対し、全て終了させた。帰属には、内部で取得したスペクトル、および公開されている NMR スペクトルデータベースを参考にした。そのデータベースは HMDB、BMRB、SDBS である。その結果を分析し、¹H/¹³C 化学シフトについての、理論遮蔽定数と実験値との回帰直線を求めた(図 1)。回帰直線を理論レベル毎に評価した結果、最も相関の良い理論レベルは ¹³C では DFT1、¹H では MP2_2 であった。

藻類未知成分の帰属研究では、算出されたギ酸の ¹H および ¹³C の化学シフトを実測値と比較した結果、SCRF 理論に基づいた溶媒効果を取り入れた計算が良い近似解を得られた。真空中および QM/MM 法では計算結果があまり変わらなかった。イオン化を考慮したものとしていないものでは、化学シフトが大きく変化する結果となった。

4. まとめ

今年度は図 1 の回帰直線を、全ての 29 化合物の ¹H 原子、¹³C 原子で決定することができた。理論レベル毎の評価は途中である。

今回はほとんどがギ酸を用いての計算方法の選定であったため、他の分子での誤差が分からなかった。多くの分子の結果を得ることで回帰分析、重回帰分析、パターン認識によるスケーリングファクターを算出することで、より良い近似解が得られることが予想される。また、今回の計算に用いた標準物質は TMS であったが、実際には DSS を用いているため、これによる誤差要因も考えられた。

5. 今後の計画・展望

費用対効果の高い理論レベルを選定し、化合物数を 10 倍程度にスケールアップして NMR メタボロミクス用理論化学シフトデータベース理論データベースの試作データの構築を行いたい。

未知成分帰属において、DSS の遮蔽定数を計算する。混合試料中における他分子においても計算を行う。誤差の範囲が大きい場合には、複数の分子について同様の計算を行い、統計解析によりスケーリングファクターを算出する。最終目標としては、海藻成分の未知シグナルの理論的帰属を試みる。

6. 利用がなかった場合の理由

利用あり