

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
飯倉智弘所属 : 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。今回は、簡便に予測精度を上げるためには化学シフト予測のどの箇所が重要かを探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

エタノール分子/セロビオース分子の 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) のメソッドと 6-31G(d)/cc-pVTZ 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、主成分分析による遮蔽定数・分子構造相関解析を行った。エタノール分子のエネルギー最小化を真空中で行い、1 ns の平衡化の後、1 ns、2 ns、10 ns、の MD シミュレーションを行った。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 法を静電力計算に使い、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。10 ps (1 ns/2 ns シミュレーション時)、100 ps (10 ns) 毎にエタノールの分子構造をサンプルし、計 100 構造を得た。エタノール分子の配位を一意に決定する 2 個の 2 面角 (水素原子による等価な二面角を除いたもの) は H(Mol2 ファイル中で自動割当された原子インデックス 8)-C(2)-C(1)-O(4) と C(2)-C(1)-O(4)-H(3) で定義し計算した。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、Gaussian09 で HF、B3LYP、MP2 の 3 種の理論レベルと 6-31G(d)、cc-pVTZ の 2 種の基

底セットについて計算した(ただし MP2 については 6-31G(d)のみ)。故に 1 理論レベル、1 基底セットにつき 100 構造のエタノール分子内原子の遮蔽定数を得た。10 ns の MD シミュレーションの結果から計算された 4 種の理論レベルと基底セットの全 400 構造の遮蔽定数を R 2.11.1 の prcomp 関数を用いて相関行列の主成分分析を計算し、主成分スコアとローディングを求めた。その中から、H-C 結合を持つものを抽出し、我々の研究室で計測・構築した HSQC データベースからの実験値化学シフトを用いて遮蔽定数を化学シフトに変換する検量線を作成し、解析に用いた。

3. 結果

真空中の 1 ns の平衡化後 1 ns と 2 ns の MD シミュレーションでは十分にエタノール分子の 2 面角空間をサンプリングしないが、10 ns では十分に全ての 2 面角の安定配置(ゴーシュ型、 -60° , 60° , 180°)に行きわたっていることがわかった。従って、10 ns の MD シミュレーションで得られた 100 構造を第一原理量子化学計算の構造セットとして用いることにした。4 種の組み合わせによる理論レベルを用いて遮蔽定数を計算し、検量線を用いて化学シフトを計算した。それによるエタノールとセロビオースの H-C 原子の化学シフトを 4 種の理論レベルの全 400×2 構造を用いて行列データを作成し、相関行列による主成分分析を行った。その結果、第 1 主成分(PC1)方向に 10.5%、第 2 主成分(PC2)方向に 9.6%の寄与率を持つスコアプロットが得られたが、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ことがわかった。

4. まとめ

今回の予備結果は、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を提起したことを意味している。しかし、今回はわずか 2 化合物しか用いていないため、さらなる一般性を証明

するために、現在我々の化学シフトデータベースに登録されている化合物の中から 29 化合物を選択し、メソッドとして MP2 を加えた 6 種の理論レベル用いた大量の計算を行っているところである。

5. 今後の計画・展望

NMR の量子化学計算による化学シフトの理論計算には構造アンサンブルを熟考することが重要である、29 化合物を用いて示す予定であり、その成果を物理化学の査読付き専門誌に投稿する予定である。さらに、効率的な化学シフトの計算法を検討し、最終目的である、数万の化合物の理論化学シフトデータベースの構築を行う予定である。

6. RICC の継続利用を希望の場合の具体的な内容

前述したように、現在 29 化合物での証明を行う計算を行っており、その成果の論文投稿と受理を行うことを次年度の目標としている。

その後の数万化合物計算の時期についての見通しは、次年度後半以降になると思われる。

平成 22 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】

近山 英輔、尾形 善之、森岡 祐介、菊地淳

"第一原理量子化学計算と古典分子動力学計算による化学シフト・構造相関の評価"

第 49 回 NMR 討論会, 東京, 2010 年 11 月 15 日～17 日

飯倉 智弘、伊達 康博、山澤 哲、菊地 淳

"廃水・廃棄物処理系バイオプロセスにおける ^{13}C 結晶セルロース分解過程の固・液時系列変動データマイニング"

第 49 回 NMR 討論会, 東京, 2010 年 11 月 15 日～17 日

【その他】

飯倉智弘

"廃水・廃棄物処理系バイオプロセスにおけるセルロース分解過程の解明"

横浜市立大学大学院生命ナノシステム科学研究科生体超分子システム科学専攻平成 22 年度修士論文