

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*

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的とする。具体的には、クエリー化合物に対し、遮蔽定数を量子化学計算プログラムである Gaussian で計算する。部分構造などの情報を活用して、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、多数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になる。量子化学計算や化学構造の計算結果から、重回帰分析等の手法で NMR 理論化学シフト計算の精密化を達成できると考えている。

2. 具体的な利用内容、計算方法

部分構造-理論遮蔽定数-NMR 実験値データベースは、NMR 理論データ予測ソフトウェア (図 1) のコマンド 1、コマンド 2、コマンド 3 の間でデータ仲介をするためのリレーショナルデータベースである。

まずコマンド 1 で、分子 ID、原子 ID、結合先原子 ID、結合タイプを新規 DB に登録する。コマンド 2 で、結合先原子への相対 xyz 座標、平均位置からの距離標準偏差、遮蔽定数平均値、遮蔽定数標準偏差を登録する。そして、コマンド 3 でそれらのデータを利用して化学シフトを予測する。

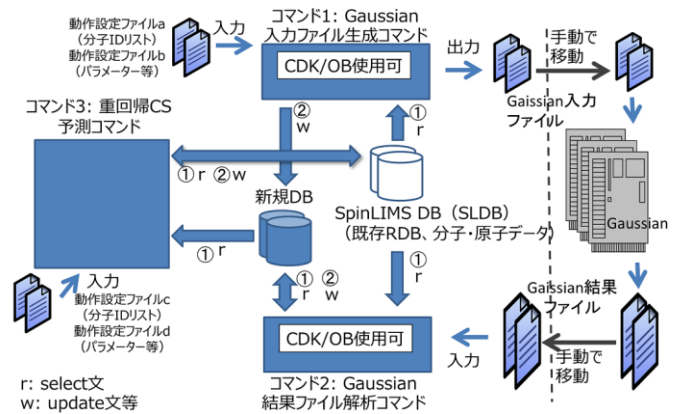


図 1 NMR 理論データ予測ソフトウェア

新規 DB の物理配置としては、parstrcta と parstrctb テーブルを、SLDB (図 1、SLDB) に構築する。つまり、SLDB 内に以下の parstrcta/b テーブルを create table する。図 1 では別の DBMS のように書いてあるが、実際の物理配置は同じ DBMS の同じデータベースになる。これは parstrcta/b と SLDB の他のテーブルのリレーションを利用するためである。

parstrctb テーブルについて記述する。このテーブルは 1 行が 1 個の 原子-結合先原子 を表す。末尾の b は bond の b である。bondtype は SDF からの取得値をそのまま設定する。

		主キー、外	
metid	INTEGER	部キー	代謝物 ID
atomid	INTEGER	主キー	分子内原子通番
atomidto	INTEGER	主キー、外 部キー	結合原子の分子内原子通番 このテーブルの同じ metid の 異なる atomid を指す。
bondtype	INTEGER		結合タイプ
rx	DOUBLE		結合先原子への相対 x 座標
ry	DOUBLE		結合先原子への相対 y 座標

rz	DOUBLE		結合先原子への相対 z 座標
rdv	DOUBLE		結合先原子の平均位置からの距離の不偏標準偏差

metid, atomid, atomidto, bondtype はコマンド 1 で登録し、rx, ry, rz, rdv はコマンド 2 で登録する。

Parstrcta テーブルについて説明する。1 行が 1 個の原子を表す。末尾の a は atom の a である。

metid	INTEGER	主キー、外部キー	代謝物 ID
atomid	INTEGER	主キー	分子内原子通番
sldavg	DOUBLE		遮蔽定数の平均値
slddv	DOUBLE		遮蔽定数の不偏標準偏差

metid, atomid はコマンド 1 で登録し、sldavg, slddv はコマンド 2 で登録する。

コマンド 1 は、Gaussian 入力ファイル生成と parstrcta/b への原子情報登録コマンドである。コマンド 1 は実行可能スクリプトが存在する場合

csps1

という名前にする。jar のみの場合

csps1.jar

という名前にする。

コマンド 1 の動作設定ファイル a について説明する。動作設定ファイル a は代謝物 ID のリストである。代謝物 ID=分子 ID で、以下 SLDB の対応する列名である metid と同じ記すことがある (代謝物 ID=metid=分子 ID)。単なる整数のテキストファイルになる。1 行に 1 つの metid をテキストで指定する。空白は不可。ファイルの終末は改行がなくても、1 個以上並んでもよいことにし、整数のないファイル終末に連なる改行は無視する。

この metid をコマンド 1 は読み込み、各 metid に関する原子情報を parstrcta/b に登録し、各 metid について Gaussian 入力ファイルを生成する。

コマンド 1 の動作設定ファイル b について説明する。分子動力学 (MD) →遮蔽定数計算の 1 個の組を 1 ターンと呼ぶことにする。

動作設定ファイル b には、Gaussian 入力ファイル生成に必要な、入力ファイルの命令行の文字列などを記述する。他に、並列数、使用メモリ量の指

定などがある。最低限、以下の項目を設定可能にする

コマンド 2 は、Gaussian 結果から抽出した、結合原子座標、遮蔽定数情報の parstrcta/b への登録コマンドである。

コマンド 2 は実行可能スクリプトが存在する場合 csps2

という名前にする。jar のみの場合

csps2.jar

という名前にする。コマンド 2 には動作設定ファイルはない。コマンド 2 はコマンド 1 の \$outdir に cd してから実行する。そのカレントディレクトリ上のファイルを解析する。コマンド 1 で出力した使用した csps1.molecules (動作設定ファイル a)、csps1.params (動作設定ファイル b) を読む。metid の各々について対応する Gaussian の結果ファイルを読み込んで解析する。

コマンド 3 は、常時稼働している SpinLIMS DB の一部に存在する理論化学シフト予測値に関わるレコードを、SLDB を停止させずに、コマンド 3 が動的に計算する新たな理論化学シフト予測値で上書きするシステムである。コマンド 3 は新規 DB から回収した遮蔽定数データと SLDB から回収した実験化学シフトデータから動的に新しい理論化学シフト予測値を計算する。コマンド 3 は新規 DB と SLDB からデータを読み、SLDB に結果を追加・上書きする。コマンド 3 のユーザーはそのデータベースの運用管理者である。SLDB と新規 DB は同じコンピュータ上で稼働しているが、コマンド 3 も同じコンピュータ上で稼働させる。SLDB は、化合物・NMR スペクトル・実験室情報・化学シフトデータベースである。

また、昨年度に引き続き、藻類の混合成分の新規帰属法の構築を行うため、NMR の実測データと Gaussian で計算された化学シフト値およびスピン結合定数の比較解析を行った。まず、混合物の多様な多次元 NMR 計測を行い、既往研究および化学シフトデータベースから化合物のアノテーションを行った。さらに、3D J-resolved HSQC および IPAP-HSQC-TOCSY 等の計測を行い、nJHH および nJCH、1JCC の実測値を得た。ここでアノテーションが完了した藻類の 90 の代謝物 (糖、アミノ酸、有機酸等)

について、B3LYP/6-311++G(d, p) 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およびスピン結合定数の理論値を算出した。また、この時に SCRF 理論に基づいた溶媒効果(メタノール)を取り入れて計算を行った。これらの実測値および理論値の比較解析では、二乗平均平方根偏差 (RMSD) および両側検定により、帰属の妥当性および立体構造の評価を行った。

3. 結果

以下にコマンド 1 実行後の parstrcta の更新結果の一部を示す。代謝物毎に原子 ID が登録されていることが分かる。

```
...
| 48 | 10 | NULL | NULL |
| 48 | 11 | NULL | NULL |
| 48 | 12 | NULL | NULL |
| 48 | 13 | NULL | NULL |
| 48 | 14 | NULL | NULL |
| 48 | 15 | NULL | NULL |
| 48 | 16 | NULL | NULL |
| 48 | 17 | NULL | NULL |
| 49 | 1 | NULL | NULL |
| 49 | 2 | NULL | NULL |
| 49 | 3 | NULL | NULL |
| 49 | 4 | NULL | NULL |
| 49 | 5 | NULL | NULL |
| 49 | 6 | NULL | NULL |
| 49 | 7 | NULL | NULL |
| 49 | 8 | NULL | NULL |
| 49 | 9 | NULL | NULL |
| 49 | 10 | NULL | NULL |
| 49 | 11 | NULL | NULL |
| 49 | 12 | NULL | NULL |
| 49 | 13 | NULL | NULL |
| 49 | 14 | NULL | NULL |
| 49 | 15 | NULL | NULL |
| 49 | 16 | NULL | NULL |
| 50 | 1 | NULL | NULL |
| 50 | 2 | NULL | NULL |
| 50 | 3 | NULL | NULL |
| 50 | 4 | NULL | NULL |
```

```
| 50 | 5 | NULL | NULL |
| 50 | 6 | NULL | NULL |
| 50 | 7 | NULL | NULL |
| 50 | 8 | NULL | NULL |
| 50 | 9 | NULL | NULL |
| 50 | 10 | NULL | NULL |
| 50 | 11 | NULL | NULL |
| 50 | 12 | NULL | NULL |
| 50 | 13 | NULL | NULL |
| 50 | 14 | NULL | NULL |
| 50 | 15 | NULL | NULL |
| 50 | 16 | NULL | NULL |
| 50 | 17 | NULL | NULL |
| 50 | 18 | NULL | NULL |
| 50 | 19 | NULL | NULL |
| 50 | 20 | NULL | NULL |
| 50 | 21 | NULL | NULL |
| 50 | 22 | NULL | NULL |
| 50 | 23 | NULL | NULL |
```

```
+-----+-----+-----+-----+
1152 rows in set (0.03 sec)
mysql>
```

以下にコマンド 1 実行後の parstrctb の更新結果の一部を示す。原子 ID と結合先原子 ID、結合タイプが登録されていることが分かる。

```
...
| 50 | 2 | 9 | 1 | NULL
| NULL | NULL | NULL |
| 50 | 9 | 2 | 1 | NULL
| NULL | NULL | NULL |
| 50 | 2 | 20 | 1 | NULL
| NULL | NULL | NULL |
| 50 | 20 | 2 | 1 | NULL
| NULL | NULL | NULL |
| 50 | 3 | 10 | 1 | NULL
| NULL | NULL | NULL |
| 50 | 10 | 3 | 1 | NULL
| NULL | NULL | NULL |
| 50 | 3 | 21 | 1 | NULL
| NULL | NULL | NULL |
| 50 | 21 | 3 | 1 | NULL
```

平成 28 年度 利用報告書

NULL	NULL	NULL		
50	4	11	1	NULL
NULL	NULL	NULL		
50	11	4	1	NULL
NULL	NULL	NULL		
50	4	22	1	NULL
NULL	NULL	NULL		
50	22	4	1	NULL
NULL	NULL	NULL		
50	5	12	2	NULL
NULL	NULL	NULL		
50	12	5	2	NULL
NULL	NULL	NULL		
50	6	13	1	NULL
NULL	NULL	NULL		
50	13	6	1	NULL
NULL	NULL	NULL		
50	6	23	1	NULL
NULL	NULL	NULL		
50	23	6	1	NULL
NULL	NULL	NULL		
50	7	13	2	NULL
NULL	NULL	NULL		
50	13	7	2	NULL
NULL	NULL	NULL		
50	8	9	1	NULL
NULL	NULL	NULL		
50	9	8	1	NULL
NULL	NULL	NULL		
50	8	10	1	NULL
NULL	NULL	NULL		
50	10	8	1	NULL
NULL	NULL	NULL		
50	8	14	1	NULL
NULL	NULL	NULL		
50	14	8	1	NULL
NULL	NULL	NULL		
50	9	11	1	NULL
NULL	NULL	NULL		
50	11	9	1	NULL
NULL	NULL	NULL		
50	9	15	1	NULL

NULL	NULL	NULL		
50	15	9	1	NULL
NULL	NULL	NULL		
50	10	12	1	NULL
NULL	NULL	NULL		
50	12	10	1	NULL
NULL	NULL	NULL		
50	10	16	1	NULL
NULL	NULL	NULL		
50	16	10	1	NULL
NULL	NULL	NULL		
50	11	17	1	NULL
NULL	NULL	NULL		
50	17	11	1	NULL
NULL	NULL	NULL		
50	11	18	1	NULL
NULL	NULL	NULL		
50	18	11	1	NULL
NULL	NULL	NULL		
50	12	13	1	NULL
NULL	NULL	NULL		
50	13	12	1	NULL
NULL	NULL	NULL		

```

+-----+-----+-----+-----+-----+
--+-----+-----+-----+
2264 rows in set (0.04 sec)
mysql>

```

コマンド 1 に metid が 1~50 までを指定して実行した結果、正しく 50 個の Gaussian 用ファイルが出力されていることを確認した。以下はその一部である。

```

out1/10:
合計 24
-rw-rw-r--. 1 chika chika 153 4月 1 03:00
2016 admph.com
-rw-rw-r--. 1 chika chika 147 4月 1 03:00
2016 admpl.com
-rw-rw-r--. 1 chika chika 281 4月 1 03:00
2016 gau.sh
-rw-rw-r--. 1 chika chika 126 4月 1 03:00

```

```

2016 nmr0.com
-rw-rw-r--. 1 chika chika 126 4 月 1 03:00
2016 nmr1.com
-rw-rw-r--. 1 chika chika 1924 4 月 1 03:00
2016 opt.com

out1/11:
合計 24
-rw-rw-r--. 1 chika chika 153 4 月 1 03:00
2016 admph.com
-rw-rw-r--. 1 chika chika 147 4 月 1 03:00
2016 admpl.com
-rw-rw-r--. 1 chika chika 281 4 月 1 03:00
2016 gau.sh
-rw-rw-r--. 1 chika chika 126 4 月 1 03:00
2016 nmr0.com
-rw-rw-r--. 1 chika chika 126 4 月 1 03:00
2016 nmr1.com
-rw-rw-r--. 1 chika chika 349 4 月 1 03:00
2016 opt.com

out1/12:
合計 24
-rw-rw-r--. 1 chika chika 153 4 月 1 03:00
2016 admph.com
-rw-rw-r--. 1 chika chika 147 4 月 1 03:00
2016 admpl.com
-rw-rw-r--. 1 chika chika 281 4 月 1 03:00
2016 gau.sh
-rw-rw-r--. 1 chika chika 126 4 月 1 03:00
2016 nmr0.com
-rw-rw-r--. 1 chika chika 126 4 月 1 03:00
2016 nmr1.com
-rw-rw-r--. 1 chika chika 597 4 月 1 03:00
2016 opt.com

```

これらの各ディレクトリには Gaussian の HOKUSAI 実行用ファイルがセットアップされている。

次に、藻類の 90 の代謝物中の 12 成分の化学シフト値およびスピン結合定数を正則化した実測値および理論値の比較解析では、ほぼ全ての値が両側検定の採択域に入った。それぞれの RMSD は、

0.282ppm (δ 1H)、8.047 ppm (δ 13C)、8.497Hz (nJCH)、1.802Hz (nJHH)、4.553Hz (1JCC)であった。また、藻類中に豊富に含まれていた Eicosapentaenoic acid (EPA) について、今回の実測値と、既往研究で知られているタンパク質中に結合している EPA の 4 つの構造の理論値を比較し、機能性のある構造との類似点、異なる点を調べた。スピン結合定数の誤差が少ないほど、同様の部分的な相対配座を持っていることが分かる。このことから、藻類中の EPA は、4 つの構造と部分的に類似の相対配座を持つことが示された。

4. まとめ

部分構造—理論遮蔽定数—NMR 実験値データベースは、NMR 理論データ予測ソフトウェアのコマンド 1、コマンド 2、コマンド 3 の間でデータ仲介をするためのリレーショナルデータベースであり、まずコマンド 1 で、分子 ID、原子 ID、結合先原子 ID、結合タイプを新規 DB に登録する。コマンド 2 で、Gaussian の結果をまとめる。そして、コマンド 3 でそれらのデータを利用して化学シフトを予測する。

parstrcta と parstrctb テーブルを、SLDB に構築した。parstrctb テーブルについては 1 行が 1 個の原子—結合先原子 を表す。末尾の b は bond の b である。bondtype は SDF からの取得値をそのまま設定する。metid, atomid, atomidto, bondtype はコマンド 1 で登録する。parstrcta テーブルについては、1 行が 1 個の原子を表す。末尾の a は atom の a である。metid, atomid はコマンド 1 で登録する。コマンド 1 は、Gaussian 入力ファイル生成と parstrcta/b への原子情報登録コマンドである。コマンド 1 の動作設定ファイルは代謝物 ID のリストである。代謝物 ID=分子 ID で、以下 SLDB の対応する列名である metid である。各 metid に関する原子情報を parstrcta/b に登録し、各 metid について Gaussian 入力ファイルを生成した。コマンド 1 の動作設定ファイル b については、Gaussian 入力ファイル生成に必要な、入力ファイルの命令行の文字列などを記述する。コマンド 3 は新規 DB から回収した遮蔽定数データと SLDB から回収した実験化学シフトデータから動的に新しい理論化学シ

フト予測値を計算する。コマンド 1 実行後の parstrcta の更新結果を確認し、代謝物毎に原子 ID が登録されていることが分かった。また、コマンド 1 実行後の parstrctb の更新結果を確認し、原子 ID と結合先原子 ID、結合タイプが登録されていることが分かった。コマンド 1 に metid が 1~50 までを指定して実行した結果、正しく 50 個の Gaussian 用ファイルが出力されていることを確認した。これらの各ディレクトリには Gaussian の HOKUSAI 実行用ファイルがセットアップされている。

化学シフト予測法の精密化では、理論化学シフトに加え、nJHH、nJCH、1JCC といった理論スピン結合定数も含めて総合的に解析することで、代謝混合物に対しての帰属精度の向上と立体構造の評価が可能になると考えられた

5. 今後の計画・展望

部分構造－理論遮蔽定数－NMR 実験値データベースと NMR 理論データ予測ソフトウェアのコマンド 1、コマンド 2、コマンド 3 の実装で部分的に終了していない部分があるため、改良が必要である。改良後は、実際に運用している化学シフトデータベースの数百の化合物に適用して、理論化学シフト計算方法の評価を行う。

藻類の 90 の代謝物の化学シフト値およびスピン結合定数の実測値および理論値をまとめ、総合的な帰属および立体構造の評価を行う。

平成 28 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. Eisuke Chikayama, Yudai Shimbo, Keiko Komatsu, Jun Kikuchi, "The Effect of Molecular Conformation on the Accuracy of Theoretical ^1H and ^{13}C Chemical Shifts Calculated by Ab Initio Methods for Metabolic Mixture Analysis", *Journal of Physical Chemistry B* 120, 3479-3487(Mar 2016)

【国際会議、学会などでの口頭発表】

1. 伊藤研悟, 小松功典, 坪井裕理, 伊達康博, 菊地淳, 代謝混合物のフラグメント集積法による構造生成と計算化学による推定構造の総合評価, 日本農芸化学会 2016, 札幌, 2016 年 3 月 27-30 日
2. K. Ito, T. Komatsu, Y. Tsuboi, Y. Date, J. Kikuchi, Structure determination of metabolites in biological extract, ICMRBS 2016, Kyoto, 21-26 August 2016
3. 尾瀨由佳, 伊藤研悟, 菊地淳, 理論化学シフト・スピン結合定数と構造相関モデル式を用いた代謝混合物同定法の構築, 第 55 回 NMR 討論会, 広島, 2016 年 11 月 16-18 日