

ライフサイエンスデータベース統合推進事業  
統合データ解析トライアル  
研究開発課題  
「タンパク質－糖鎖間の糖鎖結合部位の解明  
のためのツール改良及び解析」

研究開発終了報告書

研究開発期間： 平成25年9月～平成26年1月

研究代表者 : 細田正恵

(創価大学大学院工学研究科生命情報工学専攻、  
大学院生)

## § 1 研究開発のねらい

糖鎖は、細胞表面のタンパク質や脂質に結合し、血液型の決定、細胞接着、抗原抗体反応、ウイルス感染などの生体反応で重要な役割を果たす生体分子である。糖鎖による多様な生物学的機能のうち、糖鎖結合タンパク質による糖鎖の特異的認識があり、糖鎖-タンパク質間の相互作用の解析に糖鎖アレイやレクチンアレイを用いたプロファイリング解析が行われている。

糖鎖に関する実験データや糖鎖構造の情報が KEGG GLYCAN<sup>\*1</sup> や CFG<sup>\*2</sup>、Glycosciences.de<sup>\*3</sup> などのデータベースや JCGGDB<sup>\*4</sup> のような糖鎖関連の情報を統合したデータベースで公開されている。分岐や修飾の多様さにより、形成される糖鎖構造は実験で扱う難しさがあるだけでなく、糖鎖構造の情報処理においても複雑で困難であることから、実験データを解析できるツールが多く存在していない状況である。そのため、インフォマティクスとしての糖鎖の解析はタンパク質の解析に比べて不十分な状態であると考えられることから、バイオインフォマティクスにおける糖鎖解析ツールを開発することを目的とする。

本研究機関では、アミノ酸配列の解析で良く用いられる BLAST や ClustalW のようなマルチプルアラインメントツールとして糖鎖構造のマルチプルアラインメントツール MCAW を開発した。このアラインメントツールを使用するためには更に改善が必要であることから、MCAW ツールの改良を目指す。現時点の MCAW ツールでは、複数の糖鎖構造の入力から糖鎖構造の単糖や結合情報の割合を表せるプロファイルを得ることが可能である。そのため、糖鎖-タンパク質間の相互作用の実験結果からタンパク質と相互作用する糖鎖部位を特定していく。本研究開発で対象とする糖鎖-タンパク質間相互作用のデータベースとして、レクチンフロンティアデータベースを用いる。レクチンフロンティアデータベースは、アヒニティークロマトグラフィーの実験より得られたレクチンに対する結合親和性を持つ糖鎖構造のデータが蓄積されている。これらの結合親和性データを用いて、レクチンに親和性を示す糖鎖構造部位を単糖のみならず、単糖間の結合情報を含めて予測・特定することを見出したいと考えている。また、ツールを用いた特定のレクチンに対する糖鎖の結合部位の発見ではなく、レクチンのファミリーごとのモチーフと糖鎖結合親和性のプロファイルの関連性を見出すことを目指す。

<sup>\*1</sup> KEGG GLYCAN, <http://www.genome.jp/kegg/glycan/>

<sup>\*2</sup> Consortium for Functional Glycomics, <http://www.functionalglycomics.org>

<sup>\*3</sup> Glycosciences.de, <http://www.glycosciences.de/>

<sup>\*4</sup> JCGGDB, <http://jcgddb.jp/>

## § 2 研究成果

本研究開発で行った、複数の糖鎖のアラインメントと糖鎖プロファイルを表示できる、Web ベースのツール MCAW (Multiple Carbohydrate Alignment with Weights) は [http://rings.t.soka.ac.jp/cgi-bin/tools/MCAW/mcaw\\_index.pl](http://rings.t.soka.ac.jp/cgi-bin/tools/MCAW/mcaw_index.pl) にて利用可能である。

タンパク質のモチーフと結合親和性を持つ糖鎖構造の特徴との関連性を見出していくため、レクチンフロンティアデータベースからクロマトグラフィーの結果と Pfam のリンクが表示されている結合親和性を持つ糖鎖構造データを取得した。そして、ツールで利用できるよう糖鎖構造データを KCF 形式に変換し、ツールを用いて糖鎖結合タンパク質と相互作用する糖鎖構造部位を観察した。

実験結果から抽出したデータを用いて、マメ科レクチンやキチン結合ファミリーの WGA など、既知の特異的結合親和性糖鎖部位をツールにより示すことができた。そして、レクチンフロンティアデータベースに示されている単糖のみの特異性だけではなく、ツールにより、単糖と結合の特異性を見ることができた。データベースに記載される実験で得られた糖鎖-タンパク質間で相互作用する糖鎖構造データは、ツールを用いることでプロファイルとして得ることができ、共通する構造パターンを可視化が可能となり構造の新しい特徴を発見できると考えられる。

### § 3 研究開発計画および計画に対する達成状況

#### (1) 達成状況

本研究では、複数の糖鎖のアラインメントと糖鎖プロファイルを表示できるツールを実装し、レクチンフロンティアデータベースの糖鎖構造データを用いて実行できた。データベースで記載されている単糖のみの特異性だけでなく糖の結合も合わせて特徴を見つけることができた。当初計画にあった、糖鎖結合タンパク質のモチーフと結合親和性を持つ糖鎖構造の特徴との関連性を見出していくために、レクチンフロンティアデータベースに関連付けられている Pfam の HMMlogo のデータをモチーフとのツールにより得られる糖鎖プロファイルとの関連を見るまでには至らなかった。本研究開発期間中にタンパク質のモチーフとの関連を見ることはできなかったが、個々のタンパク質との糖鎖構造の特徴の関連を見ることができた。

#### (2) ツールの将来性への展望

今まで糖鎖の多重アラインメントはなかったため、MCAW ツールによって、糖鎖が持つ新たな共通構造の発見、機能特定のための糖鎖モチーフ生成や単糖同士の類似度を表すスコア行列の生成が可能になり、さらに精度の高いアラインメントを計算できるなど、解析の幅が広がると考えられます。また、糖鎖アラインメントより機械学習を用いたプロファイルの生成を行うこともできる。糖鎖プロファイル学習の結果をモチーフとしてタンパク質配列のプロファイルデータベースである Pfam のような糖鎖プロファイルデータベースの構築を行うことも可能となり、より糖鎖の特異的認識部位解明の発展に役立つと考えられる。

### § 4 研究参加者

| 氏名    | 所属             | 役職   | 研究開発項目          | 参加時期         |
|-------|----------------|------|-----------------|--------------|
| ○細田正恵 | 創価大学院<br>工学研究科 | 大学院生 | ツール開発・ツールを用いた解析 | H25.10-H26.1 |

### § 5 成果発表等

(1) 原著論文発表 (国内(和文)誌 0 件、国際 (欧文) 誌 0 件)

(2) その他の著作物(総説、書籍など)

なし

(3) 国際学会発表及び主要な国内学会発表

① 招待講演 (国内会議 0 件、国際会議 0 件)

② 口頭発表 (国内会議 1 件、国際会議 0 件)

1. 細田正恵(創価大学大学院工学研究科生命情報工学専攻)、糖鎖構造アラインメントツールの開発、2014 年糖鎖インフォマティクス若手の会 合同セミナー、理化学研究所(和光市)、2 月 12 日

③ ポスター発表 (国内会議 0 件、国際会議 0 件、国外会議 1 件)

1. Masae Hosoda, Kiyoko F. Aoki-Kinoshita (Division of Bioinformatics Graduate School of Engineering Soka University)、A Web tool for visualizing common patterns among glycans、2013 Annual Meeting of the Society for Glycobiology、アメリカ フロリダ州 セント・ピーターズバーグ、11 月 17 日～20 日

(4)知財出願

- ①国内出願 (0 件)
- ②海外出願 (0 件)
- ③その他の知的財産権  
なし

(5)受賞・報道等

なし

## § 6 自己評価

ツールの開発により、糖鎖結合タンパク質に親和性を示す糖鎖構造部位を単糖のみならず、単糖間の結合情報を含めて予測・特定することを見出すという目標は達成に近いと考えられる。ツールの改良とデータベースからの糖鎖構造データの取得に時間を要してしまい、タンパク質のモチーフとの解析までには至らなかったため、特定のタンパク質に対する糖鎖の結合部位の発見ではなく、データベースのレクチンのファミリーごとのモチーフと糖鎖結合親和性のプロファイルの関連性を見出すことは達成できなかった。自身の評価として、当初の目的の達成度は 60%くらいだと思う。

以上

# ツールの詳細資料

# MCAW (Multiple Carbohydrate Alignment with Weights) ツール

MCAW は複数の糖鎖構造から共通する単糖・結合をアラインメントするツールである。そして、アラインメント結果を糖鎖構造のプロファイルとして共通する部分構造の割合を表して構造の特徴を表示する。

複数のアミノ酸配列から配列の保存された領域や類似パターンを比較する際によく使用されている ClustalW\* でアミノ酸配列アラインメントが行われているように、複数の糖鎖構造をアラインメントする。

複数のアミノ酸配列のアラインメント結果に共通にみられる高度な保存領域を同様にアミノ酸配列プロファイルを参考にして。

## アミノ酸配列

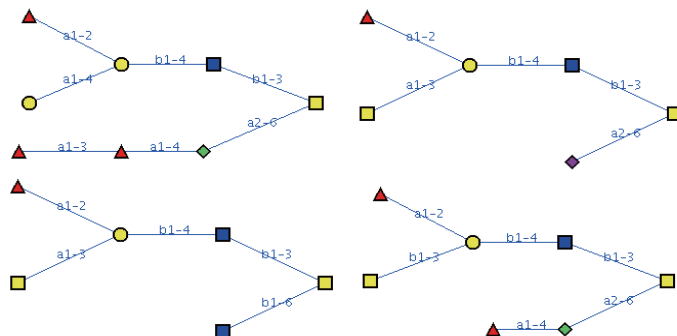
```
1. proteinA_RAT      L R V L E A Q A A T G G L I D
2. proteinA_MOUSE   Q K L L E A Q L A T G G I I D
3. proteinA_HUMAN   L R L L Q A Q L A T G G I V L
4. proteinA_HORSE   L R L L E V Q A A T G G L V D
```

## アミノ酸配列のプロファイル

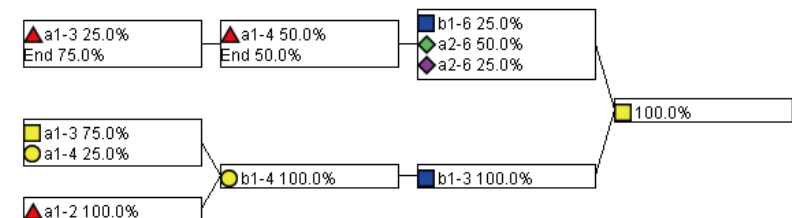


<http://pfam.sanger.ac.uk/>

## 糖鎖構造



## 糖鎖構造のプロファイル



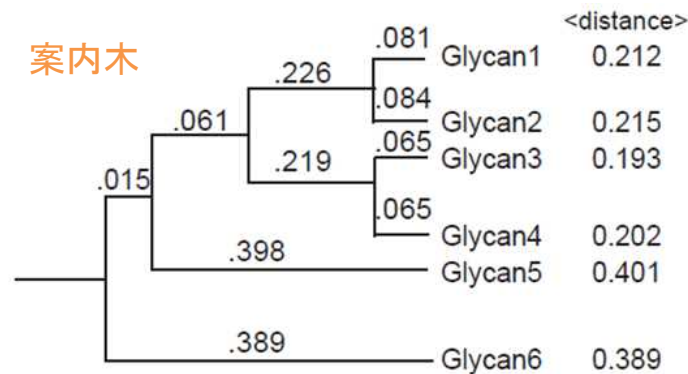
# MCAW(Multiple Carbohydrate Alignment with Weights)ツール

糖鎖のアラインメントはKCaM\*(KEGG Carbohydrate Matcher)を参考にし、動的計画法を用いてアラインメントを行っている。MCAWは、ClustalWを参考に入力糖鎖構造の類似順に整列し糖鎖多重木アラインメントを行う。

1. 全糖鎖のペアワイズアラインメント (KCaM)を行い、距離行列を得る

|          |     |      |     |     |     |   |
|----------|-----|------|-----|-----|-----|---|
| Glycan 1 | -   | 距離行列 |     |     |     |   |
| Glycan 2 | .17 | -    |     |     |     |   |
| Glycan 3 | .59 | .60  | -   |     |     |   |
| Glycan 4 | .59 | .59  | .13 | -   |     |   |
| Glycan 5 | .77 | .77  | .75 | .75 | -   |   |
| Glycan 6 | .81 | .82  | .73 | .74 | .80 | - |
|          | 1   | 2    | 3   | 4   | 5   |   |

2. 距離行列からFM法を用いて系統樹(案内木)を作成し、根からの距離を計算する



3. 案内木にしたがって糖鎖の多重アラインメントを行う

## アラインメントのオーダー

- (1) Glycan1+Glycan2
- (2) Glycan3+Glycan4
- (3) (1)+(2)
- (4) (3)+ Glycan5
- (5) (4) + Glycan6

## MCAWの利用

複数の糖鎖構造が共通する構造パターンを見つけ出すためのツールであり、共通するパターンの割合をみることができる。糖鎖アレイやアフィニティークロマトグラフィーの実験などにより得られた、タンパク質と結合親和性を示す糖鎖構造を入力として、タンパク質の認識・結合に関与する糖鎖構造の特徴パターンを検索するためにツールを利用した。

# MCAWツールの利用方法

ツール入力画面

[http://rings.t.soka.ac.jp/cgi-bin/tools/MCAW/mcaw\\_index.pl](http://rings.t.soka.ac.jp/cgi-bin/tools/MCAW/mcaw_index.pl)

ツールの入力画面で、複数の糖鎖構造をKCF(KEGG Chemical Function)形式で入力またはファイルからロードする。

RESOURCE FOR  
INFORMATICS OF  
GLYCOMES AT  
SOKA

## MCAW (Multiple Carbohydrate Alignment with Weights)

Home

Help

Feedback

Data set name:

Enter a glycan structure in KCF format:

|             |                                           |        |     |
|-------------|-------------------------------------------|--------|-----|
| ENTRY       | G04845                                    | Glycan |     |
| COMPOSITION | (Gal)3 (Glc)1 (GlcNAc)2 (LFuc)2 (Neu5Ac)1 |        |     |
| MASS        | 1656.5                                    |        |     |
| DBLINKS     | CCSD: 23949                               |        |     |
|             | GlycomeDB: 20420                          |        |     |
|             | JCGGDB: JCGG-STR011245                    |        |     |
| NODE        | 9                                         |        |     |
|             | 1 Glc                                     | 0      | 0   |
|             | 2 Gal                                     | -10    | 0   |
|             | 3 GlcNAc                                  | -20    | 10  |
|             | 4 GlcNAc                                  | -20    | -10 |
|             | 5 Gal                                     | -30    | 15  |
|             | 6 LFuc                                    | -30    | 5   |
|             | 7 LFuc                                    | -30    | -5  |
|             | 8 Gal                                     | -30    | -15 |
|             | 9 Neu5Ac                                  | -40    | 15  |
| EDGE        | 8                                         |        |     |
|             | 1 2:b1 1:4                                |        |     |

Or load KCF from a file:  選択されていません

Advanced weighting options

Gap penalty:

Monosaccharide:

- Linkage information -

Anomer:

Non reducing side carbon number:

Reducing side carbon number:

複数の糖鎖構造を  
KCFで入力する

入力はファイルを選択す  
ることも可能である

糖鎖構造  
の入力

KCF 形式

|       |            |        |    |
|-------|------------|--------|----|
| ENTRY | G10652     | Glycan |    |
| NODE  | 6          |        |    |
|       | 1 Asn      | 20     | 0  |
|       | 2 GlcNAc   | 10     | 0  |
|       | 3 GlcNAc   | 0      | 0  |
|       | 4 Man      | -10    | 0  |
|       | 5 Man      | -20    | 5  |
|       | 6 Man      | -20    | -5 |
| EDGE  | 5          |        |    |
|       | 1 2 1      |        |    |
|       | 2 3:b1 2:4 |        |    |
|       | 3 4:b1 3:4 |        |    |
|       | 4 5:a1 4:6 |        |    |
|       | 5 6:a1 4:3 |        |    |

///



# MCAWツールの利用方法

ツール入力画面

[http://rings.t.soka.ac.jp/cgi-bin/tools/MCAW/mcaw\\_index.pl](http://rings.t.soka.ac.jp/cgi-bin/tools/MCAW/mcaw_index.pl)

単糖や結合情報アラインメントの計算スコアの調節をする場合は“Advanced weighting options”のアラインメント時のオプションスコアを設定してからSubmit ボタンをクリックする。

RESOURCE FOR  
INFORMATICS OF  
GLYCOMES AT  
SOKA

## MCAW (Multiple Carbohydrate Alignment with Weights)

[Home](#)

[Help](#)

[Feedback](#)

Data set name

Enter a glycan structure in KCF format:

```
ENTRY      G04845                      Glycan
COMPOSITION (Gal)3 (Glc)1 (GlcNAc)2 (LFuc)2 (Neu5Ac)1
MASS       1656.5
DBLINKS    CCSD: 23949
            GlycomeDB: 20420
            JCGGDB: JCGG-STR011245
NODE
  1  Glc      0      0
  2  Gal     -10     0
  3  GlcNAc  -20    10
  4  GlcNAc  -20   -10
  5  Gal     -30    15
  6  LFuc    -30     5
  7  LFuc    -30    -5
  8  Gal     -30   -15
  9  Neu5Ac  -40    15
EDGE
  1      2:b1  1:4
```

Or load KCF from a file:  選択されていません

Advanced weighting options

Gap penalty:

Monosaccharide:

- Linkage information -

Anomer:

Non reducing side carbon number:

Reducing side carbon number:

実行

複数の糖鎖構造を  
KCFで入力する

入力はファイルを選択す  
ることも可能である

糖鎖構造  
の入力

ギャップが入る時のペナルティスコア

単糖が一致したときのスコア

アノマーが一致したときのスコア

非還元末端側の炭素番号が一致  
したときのスコア

還元末端側の炭素番号が一致  
したときのスコア

アラインメント時  
のオプションス  
コアの設定

# MCAWツールの利用方法

ツール結果画面

結果画面には、入力した複数の糖鎖構造のアラインメント結果を表示する。整列された単糖と結合の情報が、入力構造中でどのくらい一致したのかを表している。割合が高ければ入力構造中に共通して見られる構造パターンだということがわかる。単糖はCFG\* で提案されている単糖の標準シンボルを用いて表している。"Gap"はアラインメントで入ったギャップを示し、"End"はアラインメント時に構造の端にノードがないことを表す。

