

# 二部グラフを利用した糖鎖反応可視化ツールの開発

吉武雅侑子 (指導教員：瀬々 潤)

## 1 はじめに

近年、核酸 (DNA), アミノ酸に続く第三の鎖状生体分子として糖鎖が着目されている。この糖鎖は生体内のほとんどのタンパク質や脂質に結合し、免疫、細胞内輸送など生命機能に重要な役割を果たしている。例えば、インフルエンザウィルスは、細胞表面にある特定の糖鎖を認識して感染する。

DNA やタンパク質の鎖は分岐構造が無く直線的に結合しているのに対し、糖鎖は図 1 に示すように、糖が木構造になって結合している。図の四角や丸で表した頂点が糖を示しており、色や形は糖の種類を表している。また、辺に付けられたラベルは糖と糖が結合している位置と立体構造を示している。また、一般に糖鎖を書いた場合の右側がタンパク質に結合し、左端が細胞外の反応部位である。生体内で現在確認されている糖鎖は約 1,000 種類程度である。

このように生体内で重要な役割を示す糖鎖であるがタンパク質、ウィルスの種類などの実験サンプル (以下、サンプル) によって認識する糖鎖の構造が異なっている事が知られている。この機構解明に向け、近年開発されたグリカンアレイ [1] が利用されている。グリカンアレイは、スライドガラス上に多数の糖鎖をスポットし、その上からサンプルを流すことで糖鎖とサンプルを反応させ、一種類のサンプルが結合する糖鎖とその結合度合い (結合親和度) を一度に大量に計測することが可能な実験である。このグリカンアレイ実験結果は Consortium for Functional Glycomics (CFG) に蓄積されている。実験者が得たい情報は、着目する糖鎖やサンプルがどのような反応をするかであるが、CFG では各サンプル毎の数値データしか記載されていないため、探索が難しい。本研究では、CFG のデータを利用し、糖鎖とサンプル間の反応を探索するツールを開発した。

## 2 提案手法

提案手法の概観を図 2 に示す。図 2 (a) に CFG から得られるデータの一部で、糖鎖数 2, サンプル数 4, の例を示した。これは糖鎖 (行) とサンプル (列) の反応度の値を示している表である。図 2 (a) を見ても、糖鎖とサンプルの反応の強さはわかりにくい。

提案手法は、ブラウザを利用することでユーザーが視覚的にわかりやすく、動的に、必要な情報を手に入れることができるようにする。このため、糖鎖とサンプルを頂点とした二部グラフを作成する (図 2 (b))。図 2 (b) では左の点が糖鎖、右の点がサンプルの要素、線がそれぞれの反応度を表している。二部グラフを生成する際、同じ太さの線を引くのではなく、基準値を決め、反応度の値と基準値との大小によって、線の太さを変化させる。

特定の糖鎖とサンプルの関係性は二部グラフで示すが、どの種類のサンプルかは特定出来ない。そのため、反応するサンプルの種類で円グラフを作成した (図 2 (c))。サンプルは、ウィルス、血清、毒素、タンパ

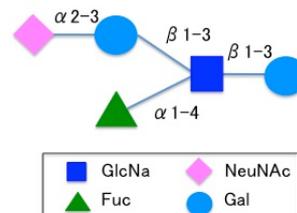


図 1: 糖鎖の木構造。

サンプルの名前	IH6C4-Ab	Influenza-A-virus-VN1194_Hemagglutinin-wild-type-WT-with-NA-Anti-Pentahis	E-Selectin-Mouse-50ug-ml	MAb-141-B-Cell-Antigen-Receptor
(GlcNAcb1-4)5b-Sp8	424.865925	65.7909	5030.521725	482.251125
(GlcNAcb1-4)6b-Sp8	151.31245	120.447325	414.7993	769.841975

(a) 糖鎖数 2, サンプル数 4 のデータ例

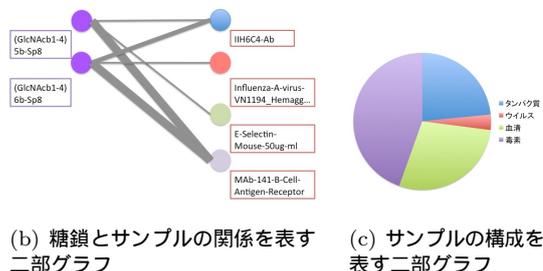


図 2: 提案手法の概観

ク質, などの種類に分けることができる。特定の糖鎖と反応しているサンプルの反応度を合計し、各サンプルの種類が、合計に対してどれくらいの割合を占めているか、を明らかにしたものが円グラフである。図 2 (c) は糖鎖の「(GlcNAcb1-4)5b-Sp8」とサンプルの反応度に関する円グラフを示している。円グラフの色は、サンプルの種類によって特定される。作成した二部グラフで特定の糖鎖と反応しているサンプルが、どの種類なのかわかるように円グラフの色と二部グラフのサンプルの点の色を合わせた。ブラウザ上では、糖鎖の点をクリックすることが出来、クリックした糖鎖に対応する円グラフが現れるようになっている。つまりこの例では、図 2 (b) の「(GlcNAcb1-4)5b-Sp8」と反応するサンプルの割合を示す図 2 (c) の円グラフが現れる。

## 3 実行結果

### 3.1 実装環境

本研究では実装に PHP, JavaScript, HTML5 を利用した。PHP でデータベースとブラウザを連携させることで、Web サーバとして Apache HTTP Server 2.2.15 を利用し、PHP 5.3.3 で実装した。また、ユーザーインターフェースの作成には、HTML5 を利用している。ユーザとの動的なインタラクティブには HTML5

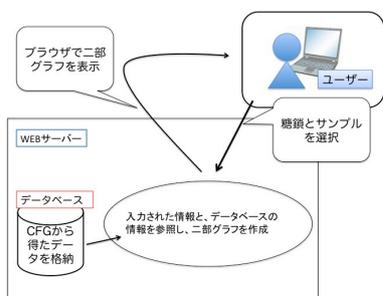
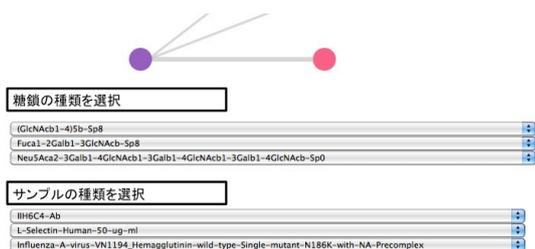


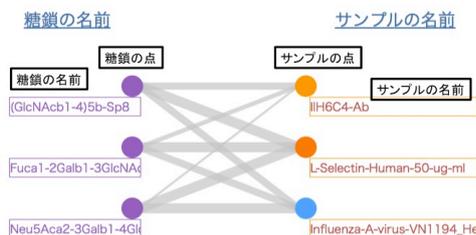
図 3: 操作の流れ



(a) 糖鎖とサンプルを選択

#### 糖鎖反応インターフェース

このサイトは、糖鎖とサンプルの反応を示すページです



(b) 二部グラフ

図 4: 実行結果

の Canvas を利用することで、ブラウザに別途プラグインを入れること無く利用可能にした。また、糖鎖の反応親和度は、MySQL 5.1.50 に格納し、PHP を介してデータベースにアクセスしている。

### 3.2 実行結果

CFG に蓄積されているデータに対し本可視化手法を用いて有用性を検証する。用いたデータは、糖鎖の種類数が 263、サンプルの種類数が 794 のデータである。操作の流れを図 3 に示した。利用者がプルダウンで糖鎖とサンプルを選択すると (図 4 (a)) データベースから情報が送られ、二部グラフを形成される (図 4 (b))

#### 糖鎖とサンプルの種類を表示

二部グラフの左側が選択した糖鎖の個数分の点と名前、右側が選択したサンプルの個数分の点とサンプルの名前である (図 4 (b))。選択したものを確認できるが、サンプルは名前が長いものも多く、全てを表示するのではなく、はじめの方のみを表示することで、視認性を良くした。

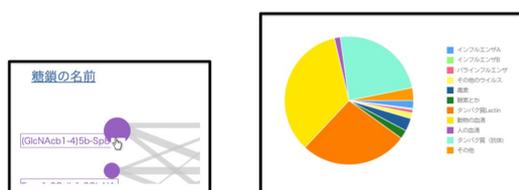


図 5: ユーザーインターフェースの改善

#### ユーザーインターフェースの改善

本研究では開発したツールの利用者は生物、化学、薬学など情報系以外の研究者を想定している。このため、利用者に次の動作をわかりやすく示す必要がある。本ツールでは糖鎖の点、サンプルの点、それぞれにマウスを近づけると点が大きくなり、利用者に対してクリックを促している。糖鎖の点をマウスでクリックすることで、二部グラフの右に円グラフを表示させる (図 5)。円グラフでサンプルの種類を確認することができる。

#### 動的な糖鎖、サンプルの探索

生成した二部グラフに新しい糖鎖やサンプルを付け加えることもでき、利用者自身で求める二部グラフを作成していくことが出来ることで、利用者が動的にデータを見る事ができるようになった。

### 4 考察と今後の課題

本提案手法を用いれば、特定の糖鎖やサンプルを選択するだけで反応度の強さを知ることが可能である。それだけではなく、選択した糖鎖と反応しやすいサンプルの種類も円グラフで確認することが出来る。このように二部グラフの作成だけでなく、それに付随する情報を別途表示することで、膨大なデータから目的の情報を的確に見せることが可能となった。

本研究の実行結果より、インターフェース利用者の要望に応じた糖鎖とサンプルの反応を二部グラフで表示することが出来た。ユーザーに糖鎖を選択してもらったり、点を付け加えてもらうことでユーザの希望通りの二部グラフを生成することが可能となった。今回は、少数の特定した糖鎖とサンプルについて二部グラフを生成することが目的であったため、選択する要素数が多くなると明確さが失われてしまうことがある。また、プルダウンメニューの選択肢が多いため選択する時間がかかることがある。今後は、データの拡張に伴い糖鎖とサンプルの関係性を保持することが課題である。それに加え、よりよいインターフェースの作成のために各機能の改善をしてしていく必要がある。例えば、プルダウンメニューをもっと細かく区切ったり、円グラフをサンプルだけでなく糖鎖に関しても作成したり、工夫を加えていくことが出来る。

#### 参考文献

[1] Ola Blixt *et al.* Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proc Natl Acad Sci*, Vol. 101, No. 49, pp. 17033-17038, 2004.