

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 24 年 6 月 6 日現在

機関番号：12608

研究種目：挑戦的萌芽研究

研究期間：2010～2011

課題番号：22650060

研究課題名（和文） 反応親和度の解析による糖鎖反応構造の推定

研究課題名（英文） Prediction of Glycan Interacting Structure using Glycan-Sample Interaction Database

研究代表者 瀬々 潤 (Jun Sese)

東京工業大学・大学院情報理工学研究科・准教授

研究者番号：40361539

研究成果の概要（和文）：

今まで少数の糖鎖とサンプル間に着目されてきた糖鎖研究は、近年の技術進歩により糖鎖とそれに反応するサンプルの大規模な観測が可能となり、新時代を迎えている。その一例が機能糖鎖オミックスコンソーシアム (Functional Glycomics Consortium; FGC) であり、現在同コンソーシアムのWebサイトには3000を超えるサンプルに関する反応情報が蓄積されている。この大きなFGCデータベースは我々に糖鎖反応に関する俯瞰的な様子を見せてくれ、このデータを用いて様々な糖鎖とサンプルの反応に関する仮説が導きだされることが期待できる。本研究ではこの仮説検証に向け、2つの手法を開発し評価した。第一に糖鎖アレイの反応データに着目したクラスタリング手法である。クラスタリングは類似の反応を示すサンプルもしくは糖鎖をグループ化するアルゴリズムであり、類似した反応を示すサンプルは同一のグループに、また、類似の反応を示す糖鎖は同一のグループに集めることが可能である事を示した。次に、サンプル（タンパク質）との反応に関与する糖鎖の部分構造を抽出する手法の開発を行った。サンプルが糖鎖に反応する際、糖鎖構造全体を認識する事は少なく、糖鎖の部分構造が重要である。本研究では、組み合わせ論的なアプローチによって、糖鎖の持つ部分構造の組み合わせと反応するサンプルの関係を調査するアルゴリズムを開発し、インフルエンザやガレクチンが認識する可能性の高い部位を特定した。

研究成果の概要（英文）：

Although most glycan research has focused on a few interactions between glycans and samples, recent technology has made it possible to observe many glycan-sample interactions at once. Consortium for Functional Glycomics (CFG) has collected data and compiled a database that contains glycan-sample interactions for more than three thousand samples. The large CFG database gives us “birds-eye view” of glycan interactions and the data from this database can be used to generate new hypothesis about interactions between glycans and samples or provide new rules about protein binding sites on glycans. In this work, we investigated the best combination of clustering technique and distance metrics for analyzing the large CFG datasets. We then developed an algorithm to extract glycan sub-structure which interacts target sample by combining the CFG data and glycan structure. The analysis results infer

substructures interacting with influenza or galectin.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010年度	1,700,000	0	1,700,000
2011年度	1,300,000	390,000	1,690,000
年度			
年度			
年度			
総計	3,000,000	390,000	3,390,000

研究分野：情報学

科研費の分科・細目：生体生命情報学

キーワード：機械学習，クラスタリング，クラス分類，糖鎖

1. 研究開始当初の背景

糖鎖は細胞表面に付着し、ウイルスや薬剤の認識に関わる重要な因子である。しかし、構造が木構造を成し複雑であること、またゲノムには糖鎖構造に関する情報は書かれておらず、機能解明は困難を極めていた。この困難を乗り越えるため、本研究では近年発達してきた300を超える糖鎖に対し化合物の結合度合いを調査できる糖鎖アレイのデータを利用し、糖鎖構造と化合物の反応予測を行うデータマイニング技術を確立することで、糖鎖の化合物・ウイルス反応部位を目指した。

2. 研究の目的

糖鎖は細胞表面のタンパク質などに付着し、薬剤に対する反応やウイルスの感染など細胞の対外的な認識に関連する。近年備蓄が問題になっているインフルエンザの薬タミフルが、インフルエンザの糖分解酵素を抑制するものであり糖鎖に関連しているだけでな

くノロウイルスも糖鎖に関連が深い。更に、がん化した細胞では、糖鎖の構造が変化する事も報告されている。このように生体に大きな影響を及ぼす糖鎖であるが、機能理解において情報的な視点から見て2点難しい事がある。1点目は、糖鎖は木構造をしており多様な形態を示している事である。このため全ての構造形態の挙動を、網羅的に調べることは現実的ではない。2点目はゲノム上には糖鎖に関する情報は書かれていないことである。このためゲノムが読破された種からも生体内に存在する糖鎖構造を絞る事ができない。%以上の性質から、糖鎖は実験対象を絞った上で解析が行われてきた。

しかし近年の技術進化により、糖鎖を大規模解析できるグリカンアレイが Consortium for Functional Glycomics (CFG)により開発されている。グリカンアレイでは、スライドガラス上に約300個の異なる糖鎖を作成し、

サンプル（化合物，ウイルス）が各糖鎖に対し結合した度合い（結合親和度）を一度に計測できる．本技術を用いて採取されたデータは Consortium for Functional Glycomics より蓄積・公開されている．本研究の目的は，このグリカンアレイで採取されたデータを統合し俯瞰することで，結合親和度を基にした糖鎖構造及びサンプルの分類を行い，更に各サンプルに対する糖鎖構造の反応部位を特定する事である．

3. 研究の方法

研究成果内に記載

4. 研究成果

本研究では2つの手法を開発し評価した．第一に糖鎖アレイのデータにのみ着目したクラスタリング手法である．クラスタリングは類似の反応を示すサンプルもしくは糖鎖をグループ化するアルゴリズムであり，類似した反応を示すサンプルは同一のグループに，また，類似の反応を示す糖鎖は同一のグループに集めることが可能である事を示した．鳥や豚のインフルエンザはそれぞれ特徴的な糖鎖に結合する事が知られているが，本アレイの情報を用いてもインフルエンザによって異なる糖鎖に反応することを示すことができた．これにより，本結果を利用することで異なる糖鎖構造を有していても類似のサンプルに反応する場合などを発見する事が可能となった．

解析結果の一例として，糖鎖アレイバージョン 4.0 を用いて行った結果を示す．図は大変大きさが大きいため，<http://www.se-se.jp/maerials/glycancluster/> に掲載する．図 3(A) に階層的クラスタリングの結果を示す．このデータは 392 のサ

ンプル 450 の糖鎖からなる．この図で赤のマス目は着目するサンプルと糖鎖の間に強い相互作用が認められたことを，青のマス目はそのペアの間に相互作用が認められなかった事を示している．図の上と左にある木構造はサンプル間，あるいは糖鎖間の近さで構築したクラスタの階層構造を表している．短い枝はそれを結ぶクラスタが反応する糖鎖もしくはサンプルが互いに近いことを示しており，この図で赤い部分が固まっている事から反応の近いグループが抽出できている事がわかる．

図 3B は，図 3A の左中央にある赤の多い部分を拡大し詳細に表したものである．この部分では多くのペアが高い相互作用を示しており，ほとんどのサンプルはインフルエンザ A，糖鎖はシアリルガラクトース (Neu5Ac-Gal) である．このクラスタの詳細を調べると，上半分と下半分の糖鎖が異なる反応を示していることが分かる．上半分のクラスタに属する糖鎖の多くは， α 2-3 結合のシアリルガラクトースであり，それに対し，した半分は α 2-6 結合のシアリルガラクトースである．この糖鎖の末端 Neu5Ac α 2-3Gal は野鳥から採取されたものであり， α 2-6 結合はヒト由来のものである．クラスタ解析により，糖鎖構造が何由来のものであるかを伏せていても，反応するサンプルの傾向から由来が分かる可能性があることを示唆している．

更にクラスタ間の違いに着目することで，相互作用の相手の違いを詳細に見よう．図 4 右側三分の一のクラスタは， α 2-6 結合のシアリルガラクトースには非常によく反応するものの， α 2-3 結合のものにはあまり反応しないことが分かる．このクラスタに属するサンプルは全て H1N 型インフルエンザ A であり，このインフルエンザは α 2-6 結合のシアリルガラクトースを認識して結合する

事が知られている。この解析により、既知の結果と同様の計算結果を得ることが出来た。同様に、左側三分の一は α 2-3結合に相互作用し、 α 2-6結合とはあまり反応しないサンプルである。このクラスタは空胞化サイトトキシン、HN1型以外のH2N8型、H4N8型などの様々なインフルエンザ、Alexa488でラベルしたタンパク質の3種類が混ざっている。興味深いことに、空胞化サイトトキシンとインフルエンザは糖鎖アレイ上で非常に類似した糖鎖と反応をしている。この観察から、これらのサンプルが認識する糖鎖構造は類似している事が示唆される。

次に、サンプル（タンパク質）との反応に関する糖鎖の部分構造を抽出する手法の開発を行った。サンプルが糖鎖に反応する際、糖鎖構造全体を認識する事は少なく、糖鎖の部分的な構造が重要である。主として末端構造が重要であるが、分岐などを認識しているケースも存在する。本研究では、組み合わせ論的なアプローチによって、糖鎖の持つ部分構造の組み合わせと反応するサンプルの関係を調査するアルゴリズムを開発し、インフルエンザやガレクチンが認識する可能性の高い部位を特定した。

1

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計2件)

(1) Data Analysis of Large-Scale Glycan-Sample Interaction. Jun Sese. Trends in Glycoscience and Glycotechnology. *in press*.

(2) クラスタリングによる糖鎖認識パター

ン解析. 伊藤真和吏, 寺井はるな, 瀬々潤. DBSJ Journal. Vol. 2 0109 . p p. 59-63.

[学会発表] (計0件)

[図書] (計0件)

[産業財産権]

○出願状況 (計0件)

名称:

発明者:

権利者:

種類:

番号:

出願年月日:

国内外の別:

○取得状況 (計0件)

名称:

発明者:

権利者:

種類:

番号:

取得年月日:

国内外の別:

[その他]

ホームページ等

6. 研究組織

(1)研究代表者

瀬々潤 (Jun Sese)

東京工業大学・大学院情報理工学研究科・准教授

研究者番号: 40361536