

令和 6 年 6 月 4 日現在

機関番号：13901

研究種目：基盤研究(C) (一般)

研究期間：2021～2023

課題番号：21K12124

研究課題名(和文)大規模グライコミクスの自動データ解析への基盤構築

研究課題名(英文)Development for automatic data analysis of large-scale glycomics

研究代表者

三浦 信明(MIURA, Nobuaki)

名古屋大学・糖鎖生命コア研究所・特任講師

研究者番号：80372267

交付決定額(研究期間全体)：(直接経費) 3,200,000円

研究成果の概要(和文)：糖鎖は、生命現象における鍵となる分子で、様々な生体の状態における糖鎖の種類と量を分析することは、疾患や老化などの生命イベントを知るうえで不可欠な情報である。糖鎖の分析は質量分析計を用いる。そのデータ解析は従来手動で、ひどい場合には何週間もかかる場合があった。本課題では、インフォマティクスの支援によってデータ解析をおおむね自動化し、「分」の単位で解析可能なツールを整えた。また実施途中で、名古屋大学のヒューマングライコムアトラスプロジェクトに資する事になり、大規模な集団の糖鎖解析に対応可能なようにプログラムを変更し、これらを実現するためのプロトタイプを構築した。

研究成果の学術的意義や社会的意義

本研究は、これまで膨大な手間と時間を要した糖鎖解析のデータ処理を、きわめて短時間に処理することを可能とした。これによって、これまでのゲノム、トランスクリプトーム、プロテオームに加えて現実的に糖鎖、グリコムを加えた疾患機序解明のサイエンスを展開することが現実的となった。遺伝子、そこから生じるタンパク質の機能だけでなくタンパク質に付与される糖鎖についての知見を得ることが容易になる。つまり、生命分子全体として統合的に生命現象解明の本当のスタート地点に立てる事になる。

研究成果の概要(英文)：Glycans are key molecules in biological process, and the analysis of the types and amounts of glycans in various biological states is indispensable information for understanding biological events such as diseases and aging. Glycans are analyzed by mass spectrometry. The data analysis has conventionally been done manually, and in some cases, it takes weeks to complete the analysis. In this project, data analysis has been largely automated with the support of informatics, and tools that enable analysis in "minutes" have been developed. In the course of the project, we decided to contribute to the Human Glycome Atlas Project of Nagoya University, and modified the program to be able to analyze glycans of a large population, and constructed a prototype to realize these changes.

研究分野：質量分析インフォマティクス

キーワード：バイオインフォマティクス グライコミクス 質量分析 MALDI

1. 研究開始当初の背景

生命現象の多くは細胞の「顔」と言われる糖鎖の変化として現れる。糖鎖は分子認識を通して、細胞認識や疾患、細胞分化、シグナル伝達などに深く関与するので[1]、生命の変化に敏感に反応する。よって糖鎖全体の分布を解明することは疾患や分化などの生命現象の解明にゲノムやタンパク質の解析と同じくらい重要である。糖鎖全体の分布を解明するグライコミクスでは主に質量分析器を用いて分析される。ガラクトース、マンノース、*N*-アセチルグルコサミンなどの残基が鎖状に結合した糖鎖は、結合位置や分岐など結合様式が多様でDNAやタンパク質のように単鎖ではなく配列面でも構造面でも複雑である。糖鎖発現の仕方によって*N*型糖鎖や*O*型糖鎖、スフィンゴ糖脂質(GSL)糖鎖、遊離糖鎖など異なるクラスに分類され、まるで異なるオミックスであるかのように異なるバラエティを持つ。一方、クラス間でモチーフが保存され、ある糖鎖の欠損が異なるクラスの糖鎖によって補われることも頻繁である。試料からのエンリッチと質量分析(MS)測定において、還元末端のアルデヒドへの修飾によるラベル化と酸性糖鎖の中性化などの化学的分析手法の多様化で測定法が近年目覚しく発展した。しかし、そのバラエティによってデータ解析が困難である状況にある。

日本学会が策定した「第24期学術の大型研究計画によるマスタープラン」の「ヒューマングライコムプロジェクト」が学術大型研究計画に選定され、文科省ロードマップ2020に正式に採択された。今後コフォートなどより得られる大量のデータを高速に解析する状況が頻発する。Glycoblotting法[2]を用いると、MALDI-TOF-MSの測定までは96サンプル/2日で処理できる。しかし、そのデータ解析には相当の熟練が必要で、その熟練した研究者でも数週間を要する場合がある。申請者は、従来各研究者がエクセルなどを用いて手動で個別に行ってきた解析を自動化すべく研究に取り組んできた[3]。

グライコミクスのデータ解析ソフトとしては、糖鎖アレイの解析用のものや[4]、質量分析に関してはGlyco-Peakfinder[5] MultiGlycan[6]などがあったが、前者は開発中止で後者は現在ソフトウェアのリンクに行き当たれない。GlycoWorkbench[Ceroniら, Proteome, 2008]はMSでのイオンをさらに分解するMS/MS解析によって詳細な解析を行えるが、本課題が志向しているMSのみの解析で大量のデータを連続して処理するという用途にはあまり向いていない。その他、各研究室レベルのインハウスのものはあるであろうが、公共に資するグライコミクスの解析ソフトは皆無に等しい。TAGは*N*型については一部公開を開始している。

ゲノムやプロテオームなどは良い解析ソフトウェアに恵まれて目覚ましい発展した一方で、糖鎖に関しては、様々な難しい問題から世界標準的と言えるようなソフトウェアが存在しなかったことがグライコミクスの進展が遅い原因の一つである。申請者のTAGをBiomolecules誌に発表した際もその号の表紙に採用されるなど注目の高さが窺える。糖鎖は古くから日本が世界をリードしてきた分野である。本課題でグライコミクス用の情報ツールを発展させることでグライコミクス研究を加速すべき段階にきている。

2. 研究の目的

申請者はこれまでの研究でTAGの開発を進めてきた。TAGは*N*型糖鎖と*N*型遊離糖鎖に限定ではあるが、既知の生合成経路を反映させた(a)準包括的な糖鎖リストを自動的に生成し(TAG List)、(b)質量分析器の出力データを入力として自動で糖鎖同定と定量、統計を行い(TAG Expression)、(c)決まったフォーマットでエクセル可読(CSV)ファイルに出力(TAG Expression)する。さらに発現変動を生合成経路上にマッピングする(TAG Pathway)。現在のTAGでは質量分析の*m/z*とエリア強度からの糖鎖同定・定量のアルゴリズムは糖鎖リストをデータベースとした検索エンジンとして確立している。一般に質量分析の*m/z*は理論値からはずれるが、糖鎖の場合はそのずれが比較的一定であるという知見によりクラスターリングによって糖鎖シグナルを同定する。内部標準として*m/z*がわかっている糖鎖(A2GN1糖鎖)を含めて定量を行う。*N*型限定であるが、糖鎖リストと発現変動のパスウェイへのマッピングが可能である。本研究では(1)糖鎖リストの拡充、(2)機種依存の解消、(3)発現変動パスウェイの拡張を行う。

3. 研究の方法

(1) 糖鎖リストの拡充(TAG List)

糖鎖リストの構築はawk言語で書かれており、糖残基と質量情報、修飾やラベルの情報をリストとして保持し、生合成経路に基づいて、糖鎖構造の特徴からアルゴリズムを作り、質量とともにカンマ区切りファイル(CSV)に出力する。複雑すぎて計画通りにアルゴリズム化できない場合は原始的だが手で書き下す事で代表的なリストを作成する。*N*型糖鎖から開始し、*O*型糖鎖、GSL糖鎖の糖鎖リスト構築の構築へと進める。

(2) 機種依存の解消(TAG Expression)

TAG Expressionでは、質量分析データと(1)の糖鎖リストを入力として、*m/z*とエリア、から

帰属・定量を行い、各種統計を行う。機種依存の低減と Python による新規インターフェイスの構築。当初 TAG Expression は Windows 上の FORTRAN77 で書かれた 3000 行程度のコードである。結果は html と CSV で出力しウェブブラウザやエクセルで閲覧できる。出力だけでなく入力も含め全ての操作をブラウザ上で行う事で機種依存を解消すべく考えている。システム構築やウェブ実装に広く用いられている Python 言語にウェブ用のフレームワーク Django を用いて TAG 全体を移植する。うまく進まない場合は Flask や Bottle, web2Py などの他のフレームワークや、申請者がプログラム経験のある JavaScript などを用いてシステムの構築を試みる。MS の一般的な mgf や mzML 形式にも対応し公共データの活用も可能にする。FDR など多重検定に関する機能の追加(R4 年度以降)

(3) 発現変動マップの拡張(TAG Pathway)

TAG Pathway は予め CSV ファイル納めた経路情報をもとに、糖鎖組成を鍵にして TAG Expression の発現情報を貼り付けて html の table タグで出力する awk スクリプトである。シアルさんはバラエティが大きいためマップ上での表示に工夫が必要である。この点について検討を行う。また O 型糖鎖、GSL 糖鎖についてもマップ構築を試みる。

4. 研究成果

- (1) 糖鎖リストについては、近年シアル酸の結合様式を質量差に変換する SALSa 法の導入にとともに、残基の記述等を根本的に書き換え、従来法であるシアル酸の MTT キャップに加えて SALSa に対応した。非還元末端のラベルとしてこれまでの aoWR に加えて BOA にも対応させた。これらの変更によって、現在の HGA プロジェクトに資するシステムとして大きなアドバンテージとなった。

N 型糖鎖について、近年ヒト血清中で発見された LacDiNAc や M5 より多くのマンノースを有するハイブリッド構造などを生成できるようにリストを拡充した。

(2)の成果である linked-glycans という考え方を発展させ、質量分析スペクトルから、残基の接続(linked-glycans)をたどる事により、可能な限りの糖鎖残基の接続情報を抽出、糖鎖リストを自動生成する事に成功した。“Glyco Spectral Harvest”と命名し、改良を続けながらこのシステムの論文投稿の準備を進めている。

- (2) HGP プロジェクトに資する目的で大量の入力データの処理を簡便に行うために、MALDI プレートへの点着情報を入力として自動で TAG Expression の入力ファイル作成する機能を追加した。これによってこれまで masslist の 1 つ 1 つのタブに手入力していたデータを一括で入力できるようになり、大規模解析を可能とした。定量性を検証するための検量線を作成する機能を追加した。同定した糖鎖組成から可能な糖鎖構造を調べるために Glyconnect (<https://glyconnect.expasy.org/>)への検索リンクを付与した。これまでインテル製のチップを搭載した PC のみで動作していた TAG を M1 チップを搭載した Mac で動作するように対応し使用できるプラットフォームを拡張した。

糖鎖帰属の安定性の向上、帰属した糖鎖から 1 および 2 残基増減した糖鎖(linked-glycan)の帰属情報をサマリーファイルに追加した。この情報は生合成経路と発現量変動などを結び付けるための基本情報にもなる。生命科学的な糖鎖の機能解明に役立つだけでなく、帰属上も生合成の流れを追う事でより信頼できる帰属を可能とする。プラットフォームの共通化については、当初から一般的なプログラムの記述を心がけていた事、gnu-fortran や gawk, python など広く用いられている言語を使っていた点で、実際にはどのプラットフォームでも問題なく動作する事が確認できた。

- (3) 糖鎖残基のデータ構造を変更した事に対応して内部データの取り扱いを変更した。また N 型糖鎖の 2 本差については SALSa や NeuGc に対応したパスウェイマップを試作として構築した。その後の検討の末、シアル酸全体がどのように変動したかを把握することが重要であるという見解に到達した。シアル酸の情報をひとまとめにして生合成経路上の糖鎖の発現変動をマップする指針を立てた。(1)で見出した Glyco Spectral Harvest の類型で、糖鎖の接続情報の一部を Cytoscape などの一般のネットワーク可視化ツールに読み込ませることによって一部表現できる仕組みを実装した

全体として、解析の自動化という観点からは、Glyco Spectral Harvest の開発に行きついた点が大きな成果と考えている。糖鎖は、ゲノムの産物ではない代謝物なので、包括的な糖鎖リストが存在しない。これを、その時に測定した質量分析スペクトルから抽出できる技術であり、これによって、どんなサンプルに対しても必要十分な糖鎖リストをその場で構築できる可能性がある。Glyco Spectral Harvest の発展は今後の課題に移っていくので、これらの提案の際にはご支援頂ければありがたい。

また、当初の目的に加えて、AI を用いた糖鎖パターンからの疾患や細胞の状態を分類する新機能 TAG classifier を考案した。今後、HGA プロジェクトにおいて膨大なデータが蓄積された場合に、様々なデータドリブンな解析を行うプラットフォームとなるので、今後も発展させていきたい。

<引用文献>

1. Varki, A. Essentials of Glycobiology, 4th ed.; Varki, A., Ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2022
2. Fujitani N, Furukawa J, Araki K, Fujioka T, Takegawa Y, Piao J, Nishioka T, Tamura T, Nikaido T, Ito M, Nakamura Y, Shinohara Y. Total cellular glycomics allows characterizing cells and streamlining the discovery process for cellular biomarkers. Proc. Natl. Acad. Sci. U S A., 110, 2105-10, 2013
3. Miura, N.; Hanamatsu, H.; Yokota, I.; Okada, K.; Furukawa, J.-I.; Shinohara, Y. Toolbox Accelerating Glycomics (TAG): Glycan Annotation from MALDI-TOF MS Spectra and Mapping Expression Variation to Biosynthetic Pathways. Biomolecules 2020, 10,1383.
4. Agravat SB, Saltz JH, Cummings RD, Smith DF. GlycoPattern: a web platform for glycan array mining. Bioinformatics. 2014, 30, 3417-8
5. Maass K, Ranzinger R, Geyer H, von der Lieth CW, Geyer R. "Glyco-peakfinder"--de novo composition analysis of glycoconjugates. Proteomics 2007, 7, 4435-44.
6. Hu Y, Zhou S, Yu CY, Tang H, Mechref Y. Automated annotation and quantitation of glycans by liquid chromatography/electrospray ionization mass spectrometric analysis using the MultiGlycan-ESI computational tool. Rapid Commun Mass Spectrom. 2015, 29, 135-42

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 3件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 Miura Nobuaki, Hanamatsu Hisatoshi, Yokota Ikuko, Akasaka-Manyu Keiko, Manyu Hiroshi, Endo Tamao, Shinohara Yasuro, Furukawa Jun-ichi	4. 巻 23
2. 論文標題 Toolbox Accelerating Glycomics (TAG): Improving Large-Scale Serum Glycomics and Refinement to Identify SALSA-Modified and Rare Glycans	5. 発行年 2022年
3. 雑誌名 International Journal of Molecular Sciences	6. 最初と最後の頁 13097 ~ 13097
掲載論文のDOI (デジタルオブジェクト識別子) 10.3390/ijms232113097	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Sugiura Kanako, Kawai Yuho, Yamamoto Arisa, Yoshioka Hiroki, Kiyohara Yuika, Iida Ayaka, Ozawa Yurika, Nishikawa Mai, Miura Nobuaki, Hanamatsu Hisatoshi, Furukawa Jun-ichi, Shinohara Yasuro	4. 巻 1867
2. 論文標題 Exposure to brefeldin A induces unusual expression of hybrid- and complex-type free N-glycans in HepG2 cells	5. 発行年 2023年
3. 雑誌名 Biochimica et Biophysica Acta (BBA) - General Subjects	6. 最初と最後の頁 130331 ~ 130331
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.bbagen.2023.130331	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Morikawa Chie, Sugiura Kanako, Kondo Keina, Yamamoto Yurie, Kojima Yuma, Ozawa Yurika, Yoshioka Hiroki, Miura Nobuaki, Piao Jinhua, Okada Kazue, Hanamatsu Hisatoshi, Tsuda Masumi, Tanaka Shinya, Furukawa Jun-ichi, Shinohara Yasuro	4. 巻 1866
2. 論文標題 Evaluation of the context of downstream N- and free N-glycomic alterations induced by swainsonine in HepG2 cells	5. 発行年 2022年
3. 雑誌名 Biochimica et Biophysica Acta (BBA) - General Subjects	6. 最初と最後の頁 130168 ~ 130168
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.bbagen.2022.130168	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計4件（うち招待講演 0件/うち国際学会 0件）

1. 発表者名 三浦信明, 横田育子, 花松久寿, 篠原康郎, 古川潤一
2. 発表標題 グライコミクス解析ソフトウェアTAG Expressionに対する血清糖鎖の大規模定量
3. 学会等名 第41回糖質学会年会
4. 発表年 2022年

1. 発表者名 1.三浦信明, 花松久寿, 古川潤一, 篠原康郎
2. 発表標題 TAG Classifier:機械学習によるカテゴリ分類とバイオマーカー探索
3. 学会等名 第40回日本糖質学会年会(鹿児島)
4. 発表年 2021年

1. 発表者名 三浦信明, 横田育子, 花松久寿, 篠原康郎, 木下聖子, 古川潤一
2. 発表標題 Toolbox Accelerating Glycomics:リストフリーデータ解析とパスウェイ構築の取り組み
3. 学会等名 第42回日本糖質学会年会
4. 発表年 2023年

1. 発表者名 三浦信明, 花松久寿, 横田育子, 篠原康郎, 木下聖子, 古川潤一
2. 発表標題 グライコミクスの解析ソフトToolbox Accelerating Glycomics (TAG): 血清の大規模解析とスペクトルからの糖鎖リスト抽出とパスウェイ構築の試み
3. 学会等名 第96回日本生化学会大会
4. 発表年 2023年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

githubリポジトリ (TAGの最新版をダウンロード可能・マニュアルを掲載)
<https://github.com/nmiura3/tag>

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------