

科学研究費補助金研究成果報告書

平成24年5月30日現在

機関番号：32690

研究種目：若手研究A

研究期間：平成20年度～平成22年度

課題番号：20680017

研究課題名（和文） 糖鎖機能解明のためのウェブリソースの構築

研究課題名（英文） Development of a web resource for elucidating glycan function

研究代表者

木下 聖子 (KINOSHITA KIYOKO)

創価大学・工学部生命情報工学科・準教授

研究者番号：50440235

研究成果の概要（和文）：本プロジェクトは糖鎖の実験者が容易に利用できるウェブリソースを構築することを目的とした。特に、アルゴリズム開発及び機械学習の手法をウェブツールとして提供し、データもウェブ上で管理できるようにした。本プロジェクトの期間中、DrawRINGS と呼ぶ糖鎖構造の描画・検索ツール、Glycan Miner と呼ぶ糖鎖の特徴を抽出する機械学習を用いたツール、Glycan Pathway Predictor と呼ぶ N 型糖鎖の合成経路を予測するツール、Kernel Tool と呼ばれる、機械学習を用いて糖鎖構造の分類を行うツール、そして Profile PSTMM と呼ぶ糖鎖のプロファイルを抽出する確率モデルのツールを開発した。

研究成果の概要（英文）：This project focused on the development of a web resource to analyze glycans using algorithmic and machine learning tools which glycobiochemists could access and utilize easily. During the time of this project, we developed (a) DrawRINGS, a glycan structure drawing and search tool, (b) Glycan Miner, a tool to extract glycan features using machine learning, (c) Glycan Pathway Predictor, a tool to predict N-glycan biosynthesis pathways, (d) Kernel Tool, a tool for classifying different data sets of glycan structures, and (e) Profile PSTMM, a tool of a probabilistic model to extract glycan profiles.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
20年度	8,100,000	2,430,000	10,530,000
21年度	5,800,000	1,740,000	7,540,000
22年度	5,800,000	1,740,000	7,540,000
年度			
年度			
総計	19,700,000	5,910,000	25,610,000

研究分野：総合領域

科研費の分科・細目：情報学・生体生命情報学

キーワード：バイオインフォマティクス、糖鎖インフォマティクス

1. 研究開始当初の背景

タンパク質を修飾する糖鎖は免疫、神経経路、シグナル伝達等の生体内作用において不可欠な分子である。しかし、バイオインフォマティクスの分野において、DNA やタンパク質の解析手法やアルゴリズム等が活発に開

発されている一方、糖鎖の研究はわずかである。この理由として、いくつかの点を挙げられる。①糖鎖の分岐構造が複雑である。②糖鎖構造を決定するハイスループットな方法が存在しない。上記の理由で、③糖鎖構造の情報を電子化するための基準が存在しない。

最初の糖鎖構造データベースはジョージア大学の Complex Carbohydrate Structure Database (CCDB) であった。しかし、1996年以降、このデータベースは更新されず、中心的な糖鎖構造データベースは 2003 年まで開発されなかった。このとき、京都大学化学研究所バイオインフォマティクスセンターの KEGG データベースに糖鎖構造データベース(KEGG GLYCAN)が追加された。同時期に、The Consortium for Functional Glycomics (CFG)が糖鎖構造とタンパク質の相互作用データベースを開発し始めた。また、質量分析を主に研究しているドイツ癌研究センター(DKFZ)の Glycosciences.de と呼ばれる糖鎖データベースも(1997 年から開発されていたが)知られるようになった。欧州においては、EuroCarbDB と呼ばれる、糖鎖情報データベースの分散ネットワークを構築するプロジェクトも開始された。

これらのデータベースプロジェクトが開発されることにより、バイオインフォマティクスが糖鎖生物学(グリコバイオロジー)に始めて応用されるようになった。KEGG GLYCAN データベースの開発とともに、糖鎖構造の比較・一致・類似度を計算するアルゴリズム及びソフトが開発され、高効率かつ正確な糖鎖データベースの構造検索のために利用された。上述のような過程を経て糖鎖インフォマティクス分野が誕生し、様々な糖鎖構造及び機能の解析のためのアルゴリズムやモデルが、我々を含めた研究者により開発された。

2. 研究の目的

糖鎖インフォマティクスが発展してきたが、糖鎖の研究者がこれらのモデルやアルゴリズムを利用して解析できるような環境は今までなかった。糖鎖インフォマティクスを更に発展させるために、本研究は、これらのアルゴリズムをツール化してウェブ上で容易に利用できるようにする糖鎖インフォマティクスリソースの構築を目的とする。更に、本ウェブリソースを利用することで、糖鎖機能を推定できるようにしたいと計画している。このためには、特に、人工知能的アプローチが役立つと考えられる。なぜなら、糖鎖構造のデータにはノイズにすぎない不明な情報が多く含まれているからである。現時点において、糖鎖のデータに最も期待できるモデルは確率モデルであり、糖鎖結合タンパク質分子であるレクチンの糖鎖認識部位を予測できる結果を我々は既に得ている。また、カーネル法を用いることで、糖鎖の複雑な構造や関連しているタンパク質、パスウェイ情報等をコンパクトに扱うことができる。そし

て、入力情報の特徴を導き出すことができる。最近の研究で、糖鎖の部分構造の分布を学習し、白血球に関連する糖鎖マーカーの候補を抽出するモデルが発表された。このように、糖鎖マーカーを効率的に抽出することによって、新たなドラッグターゲットの研究にも結びつく。従って、本研究の目的は、糖鎖インフォマティクスのツールをウェブ上で利用できるようリソースを構築することである。

3. 研究の方法

本研究では、(1)糖鎖構造の確率モデルを用いた結合親和力データからのタンパク質の糖鎖認識部位の予測ツール及びカーネルを利用した糖鎖の機能予測と構造分類ツールを最優先に開発する。同時に、(2)ユーザーにより快適なサービスを提供できるように、フィードバックフォーム、バグレポートなどのユーザーの要望への対応できるシステムも開発する。また、ユーザーのプライバシーを考慮し、ユーザーアカウントの管理システムも開発する必要がある。アカウント登録・管理や個人データの保存等、ユーザーが安心して研究を行える環境を構築する。具体的な研究方法として、次の作業を行う。

(1) 糖鎖情報処理

KEGG GLYCAN、CFG、Glycosciences.de の三つの主要な糖鎖構造データベースの情報はもと CCDB のデータであるにもかかわらず、互いにリンクは張っておらず、独立に開発されている。従って、本研究の代表者も含めて、委員会が設立され、GLYDE-II と呼ぶ、糖鎖情報の XML 標準を決定することになった。これにより糖鎖構造の情報が自由に交換できるようになると予想される。GLYDE-II を踏まえて、本研究にもこれらのデータベースから適切な情報を取り入れることにする。糖鎖を解析するために、糖鎖に関連するタンパク質や脂質の情報も取り入れて、糖鎖及び他の生体内物質との関連情報を収集する。これらの情報の取得には、まず文献を利用する。文献からこれらの情報を抽出するためには、能力も時間も非常に掛かると予想される。従って、技術員の協力で文献調査に加えて、LipidBank, LipidMaps, LectinDB, Animal LectinsDB そして Lectines の糖鎖関連データを関連付けていく。

(2) ツール開発・検証

糖鎖生物学の情報に対しては人工知能的アプローチが最も役立つと考えられるため、本研究において、次のようにツールの開発を行う。

①ユーザー対応システムの開発

・フィードバックフォーム及び内容の送信、

ツールの開発者が利用できるバグレポートシステム

- ・ユーザーからのフィードバックを効率的に受け入れられる工夫として、ウェブ上のフィードバックシステム、ヘルプ、およびマニュアルなどの文章化を行う。

②糖鎖構造の確率モデルを用いた結合親和力データからのタンパク質の糖鎖認識部位の予測ツールの開発

- ・プロフィール確率兄弟依存木マルコフモデル(Profile PSTMM)のウェブインターフェース

- ・糖鎖構造情報の入力からアラインメントを行い、最大共通部分より状態モデルを作成し、ProfilePSTMM の学習を行う。そして学習した結果を表示する。

③カーネルを利用した糖鎖の機能予測と構造分類ツールの開発

- ・糖鎖マーカーを抽出するためのカーネルのインターフェース

- ・糖鎖構造情報の入力から、カーネルを学習させ、特徴的構造を表示し、サーバーに結果を保存する。また学習されたカーネルから分類を行うツールも同様に開発する。

(3)ユーザー対応システムの開発

ユーザーアカウントの管理システムは、個人情報を守るためのシステムである。学習されたカーネルやプロフィール情報を個人別のユーザーアカウントで管理し、個人情報を守るシステムを開発する。

(4)講習会の開催

本研究で構築されたツールをより効率的に利用できるために、ユーザーの教育も必要となる。そのために、マニュアルやフィードバックから集められた情報を基にユーザーのための講習会や e-learning システムを開発する。これは本研究のリソースの教育を重視するが、関連のデータベースやツールも紹介し、糖鎖インフォマティクス全体の教育も行う。

4. 研究成果

本プロジェクトの一番の成果として、RINGS と呼ぶリソースが構築でき、次のツール及び機能を開発できた。

(1) DrawRINGS と呼ぶ糖鎖構造の描画・検索ツールを Java applet として開発した。糖鎖構造を描画でき、CFG の標準表記法で分

かりやすく表示される。また、描画した糖鎖構造を RINGS に登録した糖鎖構造と比較し、検索結果として類似した順番にデータベースの情報を表示することができる。

(2) Glycan Miner と呼ぶ糖鎖の特徴を抽出する機械学習を用いたツールを Ruby 言語で開発した。α-closed frequent mining と呼ぶアルゴリズムを実装し、ウェブで利用できるように CGI-Perl を用いてウェブインターフェースを開発した。従って、ユーザーは糖鎖構造及び必要なパラメータを入力するだけで、ウェブ上で特徴的な糖鎖の部分構造を得ることができる。

(3) Glycan Pathway Predictor と呼ぶ N 型糖鎖の生合成経路を予測するツールを Perl で開発した。N 型糖鎖の生合成経路を数学的なモデルとして表すことができる。そこで、一つの糖鎖構造及びいくつかの糖転移酵素を選択することで、ウェブ上で合成される糖鎖構造のパスウェイを Java applet で表示できるようにした。この Java applet にはズームや検索機能が備えているため、LinearCode 形式で表した糖鎖構造を調べることができる。また、一つの糖鎖構造をダブルクリックすることで、その構造を合成する酵素の詳細も表示される。従って、N 型糖鎖の合成シミュレーションなどを可能にした。

(4) Kernel Tool と呼ばれる、機械学習を用いて糖鎖構造の分類を行うツールを開発した。Matlab の SVM や Kernel 学習機能を利用し、プログラムを開発し、独立した実行可能ファイルにまとめた。本ツールでは、二つの糖鎖構造のデータセットを分類するための最も特徴的な糖鎖構造が抽出される。従って、糖鎖の質量分析データの解析の用いることができ、大量の糖鎖構造情報の区別が短時間でできるようになる。このツールを実行するためのウェブインターフェースを作成し、ウェブから実行されると、実行 ID が振られ、計算が行われる。そして実行 ID より結果をウェブから閲覧できるように開発した。実行 ID は学習に時間がかかることも予想したため、付けることにした。また、実行番号が振られるため、ユーザー登録していなくても実行番号より結果を取得できるようにした。

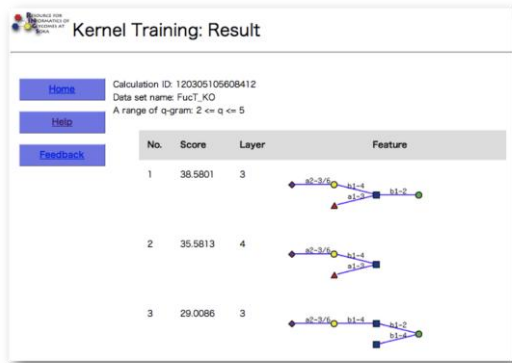


図 1 : Kernel Tool の結果画面の一例。Calculation ID が自動的に振られ、二つの糖鎖構造のデータセットを分類するための最も特徴的な糖鎖構造が抽出される。

(5) Profile PSTMM と呼ぶ糖鎖のプロファイルを抽出する確率モデルのツールを開発した。このツールは、糖鎖アレイなどの認識・結合される糖鎖構造群を入力することで、認識されている可能性のある糖鎖の部分構造をプロファイルとして表示する。図 2 に、Siglec-7 と呼ぶレクチンが認識される糖鎖のプロファイル結果を示す。ここで、Siglec-7 の特徴として知られているジシアル酸構造を認識することを確認できた。

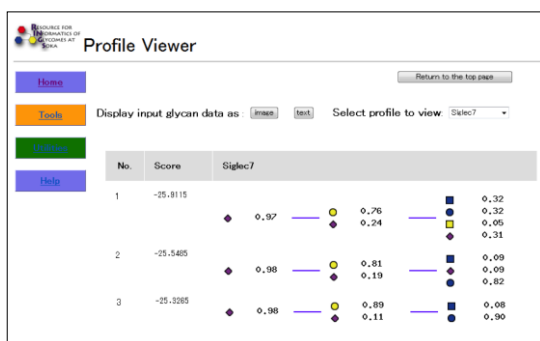


図 2: Profile PSTMM ツールの Siglec-7 と呼ぶレクチンが認識される糖鎖プロファイルの結果画面である。

(6) ユーザー登録・データ管理システムを開発した。ユーザーが無料でメールアドレス及びパスワードを登録し、ツールを実行した際の入力及び結果データを自動的にユーザーのフォルダに保存できるようにした。従って、ユーザーのプライバシーを守りながら、過去の実行結果を何時でも閲覧することができ、コメントなどの記録も付けることが可能になった。さらに、一つの実行結果を他のツールで入力として利用したい場合は、ウェブ上のプルダウンメニューから入力可能の

ツール一覧を設置した。そのため、選択されたツールに入力データが記入された状態で、入力画面が表示されるようにした。

(7) 糖鎖構造の変換ユーティリティも開発した。糖鎖構造を表記する形式が様々な存在し、主に GLYDE-II, GlycoCT, IUPAC, KCF, LinearCode® と LINUCS が良く知られている。RINGS のツールを利用する際、どの形式からでも入力できるようにするため、これら形式間で糖鎖構造情報を変換できるユーティリティを用意した。

(8) eLearning of glycomics と呼ぶ、高校生向けの糖鎖科学を学べる教育資料も用意した。図 3 にそのメイン画面を示す。5つの章で構成され、糖鎖の概要、生物学的な基礎、認識機構、糖鎖プロファイリング、そして RINGS のマニュアルを作成した。付録も用意され、単語集や参考文献も含まれている。



図 3 : eLearning of glycomics のメインページで、高校生が糖鎖科学について簡単に勉強できるように開発した。

(9) また、糖鎖に関連するレクチン、脂質及び新しい糖鎖構造も文献や他の公開データベースから取得し、RINGS のデータベースに導入した。これらの情報を全て RINGS のメインページの上部に設置した「Top Search」から検索できるようにした。

(10) さらに、RINGS を含め、国内の糖鎖関連のウェブリソースの講習会・シンポジウムも平成 23 年 2 月 25 日にライフサイエンス統合データベースセンターにて開催した。約 30 名の参加者で活発な議論ができた。

最終的に、本プロジェクトは予想以上の数のツールをウェブに公開することができ、またユーザー登録数も徐々に増えつつある。今後もツールの改良、情報の増加、そして最新技術のセマンティックウェブの導入を予定しているため、糖鎖インフォマティクス、そして糖鎖科学の発展に貢献していきたい。

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 3 件)

- ① Jiang, H., Aoki-Kinoshita, K.F. and Ching, W. Extracting Glycan Motifs using a Biochemically-Weighted Kernel, *Bioinformatics*, 査読有、Vol. 7 No. 8, 2011, pp. 405-412.
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3280441/>
- ② Li, L., Ching, W.K., Yamaguchi, T., Aoki-Kinoshita, K.F. A weighted q-gram method for glycan structure classification. *BMC Bioinformatics*. 査読有、Vol. 11, Suppl 1, No. S33, 2010.
DOI:10.1186/1471-2105-11-S1-S33
- ③ Akune, Y., Hosoda, M., Kaiya, S., Shinmachi, D., Aoki-Kinoshita, K.F. The RINGS resource for glycome informatics analysis and data mining on the Web. *OMICS*. 査読有、Vol. 14, No. 4, 2010, pp. 475-86.
<http://www.ncbi.nlm.nih.gov/pubmed/20726803>

[学会発表] (計 18 件)

- ① Aoki-Kinoshita KF. "RINGS: a web resource of data mining and analytical tools for glycomics research", The 15th Annual San Diego Glycobiology Symposium, San Diego, CA, USA, March 16-17, 2012.
- ② Hosoda, M, Aoki-Kinoshita KF. "Development of a Multiple Glycan Alignment Tool", Annual Meeting for the Japanese Society for Bioinformatics. Kobe, Japan, P-53, November 8-10, 2011.
- ③ Aoki-Kinoshita KF. "Methods and applications of glycome informatics" The 71st Okazaki Conference on New Perspectives on Molecular Science of Glycoconjugates, Okazaki, Japan, October 12-14, 2011.
- ④ Fujita A, Mori T, Aoki-Kinoshita KF. "RINGS: a web resource for glycome informatics analysis", The 71st Okazaki Conference on "New perspectives on molecular science of glycoconjugates", Okazaki, Japan, October 12-14, 2011.
- ⑤ Hosoda, M, Aoki-Kinoshita KF. "Development of a tool for profile analysis of glycan recognition

- patterns", The Japanese Biochemical Society Annual Meeting, Kyoto, Japan, 4P-0456, September 24, 2011.
- ⑥ Fujita, A, Mori, T, Aoki-Kinoshita KF. "RINGS:A web resource for glycome informatics", The Japanese Biochemical Society Annual Meeting, Kyoto, Japan, 4P-0458, September 24, 2011.
- ⑦ Akune, Y, Aoki-Kinoshita KF. "Glycan biomarker prediction using Glycan Kernel Tool", The Japanese Biochemical Society Annual Meeting, Kyoto, Japan, 4P-0454, September 24, 2011.
- ⑧ Aoki-Kinoshita KF. "Mining meaningful glycan structures from CFG resources" CFG Participating Investigators Meeting, NIH campus, Bethesda, MD, USA, July 27-29, 2011.
- ⑨ Aoki-Kinoshita KF, Ichimiya T, Takase-Yoden S, Nishihara S. "Identifying key glycan components in influenza virus infection using glycan structure mining techniques" 2nd Beilstein Symposium on Glyco-Bioinformatics, Potsdam, Germany, June 27-July 1, 2011.
- ⑩ Mori T, Fujita A, Aoki-Kinoshita KF. "Development of a web resource for glycan structure analysis" P-085. 30th Japanese Carbohydrate Symposium, Nagaoka, Japan, July 11-13, 2011.
- ⑪ Hosoda M, Aoki-Kinoshita KF. "Development of a tool for prediction of glycan recognition patterns" P-087. 30th Japanese Carbohydrate Symposium, Nagaoka, Japan, July 11-13, 2011.
- ⑫ Akune Y, Aoki-Kinoshita KF. "A tool for glycan biomarker prediction using kernel" P-088. 30th Japanese Carbohydrate Symposium, Nagaoka, Japan, July 11-13, 2011.
- ⑬ Mori T, Uchida T, Yasuoka H, Fujita A, Aoki-Kinoshita KF. "Improved usability of RINGS for glycome informatics research" S01. 2010 Annual Conference of the Japanese Society for Bioinformatics, Fukuoka, Japan, December 13-15, 2010.
- ⑭ Akune Y, Jiang H, Aoki-Kinoshita KF. "Development of a new kernel tool for glycan biomarker prediction" P039. 2010 Annual Conference of the Japanese Society for Bioinformatics, Fukuoka, Japan, December 13-15, 2010.
- ⑮ Hosoda M, Aoki-Kinoshita KF. "Multiple Tree Alignment for Glycan Profile Analysis of Glycan Binding Proteins"

- P023. 2010 Annual Conference of the Japanese Society for Bioinformatics, Fukuoka, Japan, December 13-15, 2010.
- ⑯ Akune Y, Aoki-Kinoshita KF. "Development of a New Glycan Score Matrix" P031. 20th International Conference on Genome Informatics, Yokohama, Japan, December 14-16, 2009.
- ⑰ Kaiya S, Hosoda M, Aoki-Kinoshita KF. "Improvements to Profile PSTMM for Glycan Recognition Profile Prediction" P101. 20th International Conference on Genome Informatics, Yokohama, Japan, December 14-16, 2009.
- ⑱ Hosoda M, Katoda T, Shinmachi D, Kaiya S, Akune Y, Aoki-Kinoshita KF. "A Web Resource for Glycan Function Analysis: RINGS" P-075. 29th Japanese Carbohydrate Symposium, Takayama, Japan, September 9-11, 2009.

[図書] (計 0 件)

[産業財産権]

○出願状況 (計 0 件)

[その他]

ホームページ等

<http://www.rings.t.soka.ac.jp>

6. 研究組織

(1) 研究代表者

木下 聖子 (KINOSHITA KIYOKO)

創価大学・工学部生命情報工学科・準教授

研究者番号：5 0 4 4 0 2 3 5