

科学研究費助成事業 研究成果報告書

平成 27 年 5 月 29 日現在

機関番号：12601

研究種目：基盤研究(B)

研究期間：2011～2014

課題番号：23300109

研究課題名(和文) タンパク質・糖鎖相互作用予測システムおよびデータベースの開発

研究課題名(英文) Development of protein-carbohydrate binding prediction systems and databases

研究代表者

清水 謙多郎 (Shimizu, Kentaro)

東京大学・農学生命科学研究科・教授

研究者番号：80178970

交付決定額(研究期間全体)：(直接経費) 14,600,000円

研究成果の概要(和文)：本研究では、タンパク質-糖鎖の相互作用に関して、糖鎖結合タンパク質予測(与えられたタンパク質が糖鎖と結合するかどうかを予測)、タンパク質-糖鎖結合部位予測(与えられたタンパク質が糖鎖と結合する残基、構造が得られている場合は空間上の位置を予測)、結合糖鎖予測(与えられたタンパク質に結合する糖鎖を予測)、タンパク質-糖鎖ドッキング予測、タンパク質-糖鎖相互作用解析、糖修飾部位予測(タンパク質が糖鎖修飾を受ける部位を予測)の各手法を実現し、統合的に利用できるようにするとともに、タンパク質-糖鎖の相互作用に関して、配列、構造、機能に関するさまざまな情報を登録した総合的なデータベースを構築した。

研究成果の概要(英文)：In this study, we developed the prediction methods of carbohydrate binding proteins (predicting whether a given protein binds to the carbohydrate), protein-carbohydrate binding sites (predicting which residues bind to the carbohydrate), binding carbohydrates (predicting what kinds of carbohydrate bind to a given protein), complex structures of protein and carbohydrate, glycosylation sites. We also developed the comprehensive databases of protein-carbohydrates which contain sequences, structures, binding ligands, and a variety of information about the function.

研究分野：生命情報科学

キーワード：タンパク質 糖鎖 相互作用予測 データベース バイオインフォマティクス

1. 研究開始当初の背景

糖鎖は、発生、分化、免疫、癌化、感染といった様々な生命現象の鍵を握る重要な役割を果たしているが、それらはタンパク質との相互作用によって制御されている。そのため、タンパク質-糖鎖の相互作用の研究は、非常に重要であり、さまざまな実験研究が推進され、近年では、グリカンアレイやレクチンアレイのような網羅的な解析も行われるようになってきた。このように実験データが蓄積される中、バイオインフォマティクスを利用したタンパク質-糖鎖相互作用の予測・解析に期待が高まっているが、現在、利用できるソフトウェアは非常に限定され、その精度も十分でないのが現状である。

2. 研究の目的

本研究では、タンパク質-糖鎖の相互作用に関して、糖鎖結合タンパク質予測（与えられたタンパク質が糖鎖と結合するかどうかを予測）、タンパク質-糖鎖結合部位予測（与えられたタンパク質が糖鎖と結合する残基、構造が得られている場合は空間上の位置を予測）、タンパク質-糖鎖ドッキング予測、タンパク質-糖鎖相互作用解析、糖修飾部位予測（タンパク質が糖鎖修飾を受ける部位を予測）の各手法を実現し、統合的に利用できるようにするとともに、タンパク質-糖鎖の相互作用に関して、配列、構造、機能に関するさまざまな情報を登録した総合的なデータベースを構築する。

本研究で開発する手法のうち、タンパク質-糖鎖のドッキング予測、結合糖鎖予測は、筆者の知る限り、現在利用できるソフトウェアがない。また、糖鎖結合タンパク質予測は、応募者らのグループが他に先駆けて開発し、AUC (Area Under the Curve) 値で 0.91 という高い予測精度を達成しているが、本研究では、さらなる精度の向上と構造情報も用いた新しい手法を開発する。タンパク質-糖鎖結合部位予測については、先行研究において、予測精度が AUC 値で 0.683 (Malik, 2010) という結果が得られているが、十分な精度が得られているとは言えない。タンパク質構造が既知の場合の結合部位（位置）予測については、グルコース、ガラクトースなどの単糖との結合部位予測 (Nassif, 2009; Scognamiglio, 2007) がいくつか提案されているが、糖鎖を対象としたものはさらに限られており、61 個の糖鎖結合部位のうち 69% の結合部位を予測した先駆的研究 (Mitsuyama, 2003) のほかは、近年になって AUC 値で 0.82 の予測精度を実現した研究 (Kulharia, 2009) が発表されている程度である。そのほか、一般のタンパク質-リガンド結合部位予測やドッキング予測の適用が考えられるが、糖鎖の特性を利用し対象を特化することで、予測精度を大幅に向上させることができるかと期待される。

3. 研究の方法

タンパク質-糖鎖結合部位の特性に関して、既存のデータベースおよび文献等により網羅的に解析を行い、初年度中に、糖鎖結合タンパク質予測、タンパク質-糖鎖結合部位予測、タンパク質-糖鎖ドッキング予測、タンパク質-糖鎖相互作用解析の各手法の開発に着手する。単に機械学習等を使って予測精度を上げるだけでなく、タンパク質-糖鎖の結合のメカニズム、機能の解明を目指し、そのための新規の手法・アルゴリズムの開発を行う。また、解析によって得られた結果を実験事実とともに（両者は区別して）データベース化する。データベースは、糖鎖結合部位の配列特徴や、ドメイン構成、進化情報、不規則領域、代謝経路など、機能に関わる広範な情報を検索できるようにし、そこから構造や機能の比較、分類、その他高次の検索ができるようにする。開発した手法（ツール）とデータベースは Web を介して一般に公開する。

4. 研究成果

本研究では、タンパク質-糖鎖の相互作用に関して、糖鎖結合タンパク質予測（与えられたタンパク質が糖鎖と結合するかどうかを予測）、タンパク質-糖鎖結合部位予測（与えられたタンパク質が糖鎖と結合する残基、構造が得られている場合は空間上の位置を予測）、結合糖鎖予測（与えられたタンパク質に結合する糖鎖を予測）、タンパク質-糖鎖ドッキング予測、タンパク質-糖鎖相互作用解析、糖修飾部位予測（タンパク質が糖鎖修飾を受ける部位を予測）の各手法を実現し、統合的に利用できるようにするとともに、タンパク質-糖鎖の相互作用に関して、配列、構造、機能に関するさまざまな情報を登録した総合的なデータベースを構築した。

(1) タンパク質-糖結合部位予測システムの開発

SVM を用いて配列情報のみからタンパク質の糖鎖結合部位を予測するシステムを開発した。タンパク質に結合する糖を、そのタンパク質結合部位に現れる糖結合残基の組成をもとに、酸性の官能基を持つ酸性糖と、酸性の官能基を持たない非酸性糖の 2 つに分類し、それぞれの結合部位を予測するシステムを開発した。データセットとして使用した各アミノ酸配列について、NCBI の非冗長データベースを対象に PSI-BLAST を実行し、作成されたマルチプルアライメントを基に PSSM (位置特異的スコアマトリックス) を計算し特徴量として用いた。酸性糖と非酸性糖の結合部位の予測器を、それらの decision value の線形結合により結合することで新たな合併予測器を開発した。その結果、酸性糖の場合は酸性糖で学習させた予測器、非酸性糖の場合、および酸性糖か非酸性糖か最初からわからない場合は、合併予測器が最も高い性能を示した。表 1 に糖結合部位予測の結果

を示す。

表 1 糖結合部位予測のまとめ

予測対象データセット	最も良い性能を示した予測器	AUC	MCC
糖結合タンパク質	合併予測器	0.760	0.185
酸性糖結合タンパク質	酸性糖結合部位予測器	0.787	0.221
酸性糖結合タンパク質	合併予測器	0.752	0.178

表 1 より、糖結合部位予測では、合併予測器が、糖結合タンパク質と酸性糖結合タンパク質で最も良い予測結果を実現し、酸性糖タンパク質では酸性糖結合予測器が最も良い予測結果を実現した。このことは多数派の非酸性糖（データセット全体の 7 割以上）に有利なパラメータが選ばれたことにより、少数派の酸性糖結合残基予測の性能を落とす結果となったとことが原因であると考えられる。実際、糖結合残基予測器、合併予測器ともに、非酸性糖結合残基予測器に近い予測傾向を示していた。実際の利用では、対象となる糖が未知である場合、合併予測器が最も実用的であり、一方、結合する糖に酸性基がついているとわかっている場合、酸性糖結合残基予測器を用いると良いことがわかる。

図 1 は、酸性糖結合タンパク質に対する、糖結合部位予測器、酸性糖結合部位予測器、合併予測器、糖結合部位予測器の性能を示したものである。

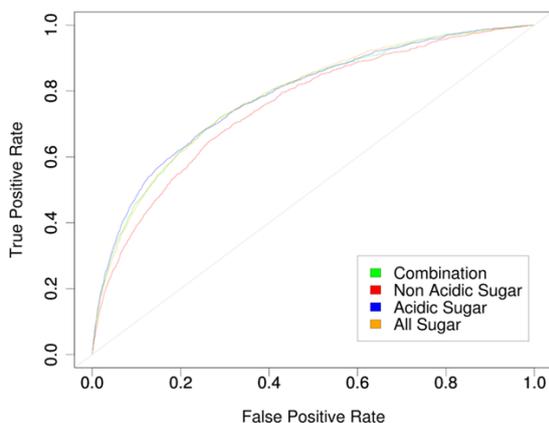


図 1 酸性糖結合タンパク質に対する性能評価

(2) タンパク質-糖結合予測システムの開発

我々は、また、与えられたタンパク質が糖に結合するかどうかを予測するシステムを開発した。(1)と同様、酸性の官能基を持つ酸性糖と、酸性の官能基を持たない非酸性糖の 2 つに分類し、それぞれの結合部位を予測するシステムを開発した。表 2 に糖結合部位予測の結果を示す。

表 2 糖結合予測のまとめ

予測対象データセット	最も良い性能を示した予測器	AUC	MCC
糖結合タンパク質	糖結合予測器	0.869	0.654
酸性糖結合タンパク質	酸性糖結合予測器	0.846	0.565
酸性糖結合タンパク質	非酸性糖結合予測器	0.901	0.708

表 2 より、糖が酸性糖または非酸性糖であるとわかっているときは、それぞれの糖クラス向けの予測器を適用し、わからない場合は、糖全体の予測器を適用するのがよいことがわかる。

(3) 糖修飾されたタンパク質に対する予測

タンパク質-糖結合予測糖鎖修飾予測手法については、Protein Data Bank を探索し、糖と 1.5 Å 以内に近接した残基および共有結合している糖を収集して学習データセットを作成し、SVM による予測を行い、高い AUC 値を得ることができた。しかし、(1)の糖結合部位予測における酸性糖予測器では、MCC が 0.221 と、高い予測性能は得られなかった。これはタンパク質に付加された糖はタンパク質の表面に現れることが多く、近接した残基の多くはクリスタルパッキングである可能性がある。これら残基が結合残基特有の保存パターンを持っていなかったため予測できなかったものと考えられる。

(4) タンパク質-リガンドドッキング予測

タンパク質の構造を用いて、リガンドが結合する空間を予測するシステムを開発した。本システムの予測手順は以下の通りである。タンパク質の周辺にプローブを配置し、タンパク質とプローブとのエネルギー値を計算、エネルギー値の低いプローブをクラスタリングする。また、あわせて、類縁タンパク質の各残基の配列保存度を求め、上記エネルギーと合わせて評価を行う。

タンパク質の冗長性を排除し、リガンド結合構造、リガンド非結合構造の両方が得られているもの 342 個を用いて評価を行った。その結果、リガンド結合空間予測について、各正解条件において本手法の success rate が Q-SiteFinder の success rate を上回ったことが確認できた。具体的な数値は、予測順位 1 位のクラスターで precision $\geq 25\%$ のものが 52.0% から 74.0%、予測順位 3 位以内で precision $\geq 25\%$ のものが 71.9% から 88.0% と改善されていた。average precision は、正解条件により Q-SiteFinder の方が上回ることがあったが、その条件においても両者は近い値をとっていることが分かった。アミノ酸保存度を用いた場合と用いなかった場合の比較では、用いた場合の方が、予測順位 1 位の

クラスで precision $\geq 25\%$ のものが 57.0% から 74.0%、予測順位 3 位以内で precision $\geq 25\%$ のものが 73.7% から 88.0% と改善されており、アミノ酸保存率を用いた効果が非常に大きいことがわかる。

5. 主な発表論文等

〔雑誌論文〕(計 8 件)

① M. Komatsu, Y. Sato, T. Negami, T. Terada, O. Sasaki, J. Yasuda, A. Arakawa, C. Yoshida, H. Takahashi, Malau-Aduli, A.E.O., K. Suzuki, K. Shimizu, Overdominance effect of the bovine ghrelin receptor (GHSR1a)-DelR242 locus on growth in Japanese Shorthorn weaner bulls: Heterozygote advantage in bull selection and molecular mechanisms, *G3: Genes, Genomes, Genetics*, 査読有, Vol.5, 2015, 271-279
DOI:10.1534/g3.114.016105

② T. Negami, K. Shimizu, T. Terada, Coarse-Grained Molecular Dynamics Simulations of Protein-Ligand Binding, *Journal of Computational Chemistry*, 査読有, Vol.35, No.25, 2014, 1835-1845
DOI:10.1002/jcc.23693

③ 込山 悠介, 番野 雅城, 鎌水 優行, 加藤 文彦, 大向 一輝, 武田 英明, 清水 謙多郎, 生命科学の複数 LOD の統合による目的別タンパク質分子間相互作用 Linked Open Data の構築, *人工知能学会論文誌*, 査読有, Vol. 29, 2014, 356-363
DOI:10.1527/tjsai.29.356

④ N. T. Vu, Y. Moriwaki, J. M. Caaveiro, T. Terada, H. Tsutsumi, I. Hamachi, K. Shimizu, and K. Tsumoto, Selective binding of antimicrobial porphyrins to the heme-receptor IsdH-NEAT3 of *Staphylococcus aureus*. *Protein Science*, 査読有, Vol.22, 2013, 942-953
DOI:10.1002/pro.2276

〔学会発表〕(計 16 件)

① 清水 謙多郎, 寺田 透, タンパク質の複合体構造・相互作用に関する総合的な予測・解析/Prediction and analysis of protein complex structures and protein interactions, 第 14 回日本蛋白質科学会年会(招待講演), 2014 年 06 月 25 日, ワークピア横浜(神奈川県、横浜市)

② Kazuya Sumikoshi, Teppei Ikeya, Yutaka Ito, Kentaro Shimizu, Effect of basis selection in reconstructing NMR spectra using compressed sensing, 第 52 回日本生物

物理学会年会, 2014 年 09 月 25 日, 札幌コンベンションセンター(北海道、札幌市)

③ Tatsuki Negami, Tohru Terada, Kentaro Shimizu, PROTEIN-LIGAND BINDING SIMULATION WITH THE MARTINI COARSE-GRAINED FORCE FIELD. Biophysical Society, 58th Annual Meeting, 2014 年 02 月 18 日, サンフランシスコ(アメリカ)

④ 清水 謙多郎, 蛋白質立体構造および相互作用の予測, 第 4 回神経科学と構造生物学の融合研究会(招待講演), 2013 年 11 月 19 日, 岡崎コンファレンスセンター(愛知県岡崎市)

⑤ Z. Ge, W. Cao, K. Sumikoshi, S. Nakamura, K. Shimizu, Prediction of the types of carbohydrate-binding proteins, 日本生物物理学会 第 50 回年会, 2012 年 09 月 22 日, 名古屋大学(愛知県名古屋市)

〔その他〕

ホームページ等

lipid-binding protein predictor
<http://ec2-54-64-162-67.ap-northeast-1.compute.amazonaws.com/LBPredictor/default/index>

東京大学大学院農学生命科学研究科応用生命工学専攻生物情報工学研究室ホームページ

<http://www.bi.a.u-tokyo.ac.jp/papers/>
レクチン予測システム

<http://www.bi.a.u-tokyo.ac.jp/software.html>

6. 研究組織

(1) 研究代表者

清水 謙多郎 (SHIMIZU, Kentaro)
東京大学・大学院農学生命科学研究科・教授
研究者番号: 80178970

(2) 研究分担者

寺田 透 (TERADA, Tohru)
東京大学・大学院農学生命科学研究科・准教授
研究者番号: 40359641

角越 和也 (SUMIKOSHI, Kazuya)
東京大学・大学院農学生命科学研究科・助教
研究者番号: 90431832