

平成 26 年 6 月 10 日現在

機関番号：32682

研究種目：若手研究(B)

研究期間：2012～2013

課題番号：24700298

研究課題名(和文)位置特異的スコアおよび機械学習を用いたGPI修飾タンパク質予測とデータベース構築

研究課題名(英文) Prediction of GPI-anchored protein using position-specific scoring matrix and development of GPI-anchored protein database

研究代表者

池田 有理 (Yuri, Ikeda)

明治大学・理工学部・准教授

研究者番号：30371082

交付決定額(研究期間全体)：(直接経費) 1,800,000円、(間接経費) 540,000円

研究成果の概要(和文)：本研究では、位置特異的アミノ酸スコアマトリクス(PSSM)と誤差逆伝搬型ニューラルネットワーク(BP-ANN)を用いて、高精度のGPIアンカー型タンパク質判別方法を開発することができた。また、GPIが修飾される位置(オメガサイト)とその周辺のアミノ酸配列の物理化学的特徴を抽出し、オメガサイト位置を高精度に予測することもできた。また、ヒト・マウス・ウシのORFから新規GPI-AP候補遺伝子を予測し、データベース化した。本研究を通して得られた知見から、タンパク質の糖鎖修飾予測法の開発、人工GPI-APの作製など、新しい複数の研究に発展させることができた。

研究成果の概要(英文)：In this study, new methods for the detection of GPI-anchored proteins (GPI-APs) and for the prediction of the location of omega-sites by physicochemical properties, PSSM and back-propagation on artificial neural networks (BP-ANNs) were developed. PSSs were calculated based on amino acid propensities around the omega-sites. PSSs were applied to BP-ANNs which consist of a three-layered structure. This method could distinguish GPI-APs from non-GPI-APs (92.9% sensitivity and 94.8% specificity) and also could discriminate omega-sites from non-omega-sites with higher accuracy (95.6% sensitivity and 97.0% specificity) than other GPI-APs detection tools reported previously.

研究分野：生命情報科学

科研費の分科・細目：生体生命情報学

キーワード：バイオインフォマティクス タンパク質 翻訳後修飾 GPIアンカー PSSM 機械学習

## 1. 研究開始当初の背景

糖脂質 GPI は翻訳後修飾のひとつであり、GPI アンカー型タンパク質 (GPI-AP) は難治疾患や植物の生体防御反応などの主要因であることが知られている。我々はこれまでに、アミノ酸配列特性を利用することにより、GPI-AP を他のタンパク質から判別可能であることを見出した。

本研究では、GPI-AP 予測法を改良し、ゲノム ORF から新規 GPI-AP を予測する。また、それらのオミクス情報のアノテーションを行い、データベース化することにより、GPI 修飾のメカニズム解明をはじめ、難治疾患の治療や植物の生体防御反応の分子機構研究に先立つ知識基盤整備を行う。

## 2. 研究の目的

本研究では GPI-AP 予測法の改良および新規 GPI-AP の予測を行い、データベース化することを目的とした。詳細は以下である。

### 1) GPI-AP 予測法の確立

我々が開発してきた位置特異的アミノ酸スコアマトリクス (PSSM) による GPI-AP 予測法に、機械学習を適用する。予測精度を求め、申請者の従来法と比較検討することにより、ゲノム情報からより多くの GPI-AP 候補を予測することが可能となる方法を明らかにする。公共データベースや文献調査から、生化学実験により新規に GPI-AP と同定されたタンパク質のデータを得て試験データとし、申請者の GPI-AP 予測法の客観的な精度を明らかにする。

### 2) 新規 GPI-AP 候補予測

既に配列が明らかにされている生物種のゲノム ORF から、上記 1) で開発された予測法を用いて新規の GPI-AP 候補を予測し、データベース化する。

## 3. 研究の方法

本研究では、以下の方法で GPI-AP 予測法の改良および新規 GPI-AP の予測、データベース化を目指した。

### 1) GPI-AP 予測法の確立

本研究では以下の方法で、GPI-AP 予測法の改良および新規 GPI-AP の予測とデータベース化を行った。詳細は以下の通りである。

我々は位置特異的アミノ酸スコアマトリクス (PSSM) を用いた GPI-AP 予測法を開発してきた。本研究計画では、GPI 修飾残基付近のアミノ酸出現頻度より PSSM を作成し、PSSM を用いてスコア算出を行うが、スコアの分布から GPI-AP か否かを判別する際に誤差逆伝搬型ニューラルネットワーク (BP-ANN) やサポートベクターマシン

(SVM) を適用し、予測精度の向上を試みた。5 分割交差検定法等のクロスバリデーションテストによって予測精度の客観的評価を行い、従来法と比較検討した。

また、公共データベースや文献調査から、生化学実験により新規に GPI-AP と同定されたタンパク質のデータを得て、評価試験のための差分データとした。差分データを用いて申請者の改良法、従来法、および“PredGPI”や“Big-Pi”等の既存法の予測精度を客観的に評価し、比較検討を行った。

### 2) 新規 GPI-AP 候補予測

従来法で成功している大規模ゲノムスクランの自動化および高速化を、改良法に適用した。次に、UniProtKB/Swiss-Prot、UniProtKB/TrEMBL に格納されているゲノム ORF を生物種ごとに分類し、新規の GPI-AP 候補を予測した。ヒト難治疾患の分子機構解明の観点からヒト・マウス・ウシの遺伝子解析を優先した。UniProtKB/Swiss-Prot、UniProtKB/TrEMBL の記載内容およびリンク先からのアミノ酸配列、遺伝子・cDNA 配列、立体構造、機能、細胞内局在等のオミクス情報の検索と付加を行いデータベース化し、Web 公開を行った。

## 3. 研究成果

本研究では GPI-AP 予測法の改良および新規 GPI-AP の予測を行った。また、GPI-AP 候補遺伝子をデータベース化した。詳細は以下の通りである。

### 1) GPI-AP 予測法の確立

我々が開発してきた位置特異的アミノ酸スコアマトリクス (PSSM) による GPI-AP 予測法に、BP-ANN を適用した。5-fold crossvalidation test を用いて予測精度の客観的評価を行った結果、92.9%の感度と 94.8%の選択性で、GPI-AP の配列を非 GPI-AP の配列から判別できることがわかった。

また、本研究の遂行過程において、以下の知見が得られた。

[1] タンパク質の翻訳後修飾予測に PSSM や BP-ANN が効果的であることが明らかにされた。

[2] GPI-AP の配列解析により、GPI-AP が N 末端側に持つシグナルペプチドの配列にバリエーションがあることが明らかにされた。

[3] GPI-AP が N 末端側に持つシグナルペプチドと、C 末端側に持つ GPI アタッチメント配列とを構成するアミノ酸の物理化学的特徴が明らかにされた。

上記の知見は、下記 3) で述べる研究に発展させることができた。

### 2) 新規 GPI-AP 候補予測

ヒト・マウス・ウシのゲノム ORF から、上記 1) で開発された GPI-AP 予測法を用いて新

規の GPI-AP 遺伝子候補を予測し、データベース化した。新規に GPI-AP と予測された候補遺伝子はヒトで 57 エントリ、マウスで 44 エントリ、ウシで 15 エントリとなった。予測された遺伝子については予測時のパラメータや信頼性を示すスコアとともにデータベース化し、株式会社情報数理バイオの HP へ公開した (<http://imsbio.jp/GPIAPsPred/>)。

### 3) 本研究成果から発展した研究

本研究で得られた知見やデータを基に、以下の複数の新しい研究に発展させることができた。

#### [1. GPI アンカー型タンパク質統合データベースの開発]

本研究で得られた GPI-AP 予測法および GPI-AP 候補遺伝子検索法について、株式会社情報数理バイオとの共同研究に発展した。情報数理バイオ社と協力し、ゲノム配列が明らかにされている生物種で網羅的な GPI-AP 候補遺伝子の発見とデータベース化を目指す。

#### [2. タンパク質の糖鎖修飾の分子機構解明と予測法の開発]

本研究で、タンパク質の翻訳後修飾予測に PSSM や BP-ANN が効果的であることが示された。タンパク質の糖鎖修飾予測にも PSSM と BP-ANN を適用し、高精度予測を目指す。

#### [3. シグナルペプチドの物理化学的性質と成熟タンパク質の細胞内局在性の相関]

本研究で GPI-AP の配列解析を行うことによって、GPI-AP が N 末端側に持っているシグナルペプチドの配列にバリエーションがあることが明らかにされた。成熟タンパク質の細胞内局在性とシグナルペプチド配列の相関を明らかにする。

#### [4. 人工 GPI アンカー型タンパク質の設計と真核細胞内での発現解析]

本研究で、GPI-AP が N 末端側に持っているシグナルペプチドと、C 末端側に持っている GPI アタッチメント配列とを構成するアミノ酸の物理化学的特徴が明らかにされた。シグナルペプチドとアタッチメント配列に特異的な物理化学的特徴を持たせるために効果的なアミノ酸を用いて、人工配列を設計して GFP と融合させ、HeLa 細胞内での発現解析を行っている。

## 5 . 主な発表論文等

### [雑誌論文](計 6 件)

- 1) K. Etchuya, R. Nambu, Y. Mukai (2013) Discrimination of xylose O-glycosylation sites in mammalian proteins, *Chemistry Letters*, 42, 9, 1043-1045.
- 2) K. Etchuya, Y. Mukai (2013) Sugar type

discrimination in O-glycosylation based on protein primary sequences, *Journal of Biomechanical Science and Engineering*, 8, 3, 225-232.

- 3) Y. Mukai, M. Ikeda, H. Tanaka, T. Konishi, O. Oura, T. Sasaki (2013) Discrimination of mammalian GPI-anchored proteins by their hydropathy and amino acid propensities, *Bioscience, Biotechnology, and Biochemistry*, 77, 3, 526-533.

- 4) Y. Mukai, K. Machino, K. Etchuya, H. Tanaka, T. Sasaki (2013) Correlation between Proline-Proline/hydroxy-Proline in collagen proteins and the moisture content of the stratum corneum after oral ingestion, *International Journal on Applied Bioengineering*, 7, 1, 71-75.

- 5) Y. Mukai, T. Sasaki (2012) Structural change and stability of solubilized bacteriorhodopsin, *International Journal on Applied Bioengineering*, 6, 2, 59-65.

- 6) Y. Mukai, H. Tanaka, M. Yoshizawa, O. Oura, T. Sasaki, M. Ikeda (2012) A computational identification method for GPI-anchored proteins by artificial neural network, *Current Bioinformatics*, 7, 125-131.

### [学会発表](計 33 件)

- 1) K. Etchuya, Y. Mukai: Structural aspects of proteins modified by oligosaccharides, 10th International Conference on Flow Dynamics, Sendai, Japan, Nov. 2013.

- 2) K. Etchuya, Y. Mukai: Correlations between protein secondary structure and glycosylation, 21st Annual International Conference on Intelligent Systems for Molecular Biology & 12th European Conference on Computational Biology, Berlin, Germany, Jul. 2013.

- 3) R. Nambu, Y. Mukai: Sequence analysis and discrimination of subcellular localization of type II membrane proteins, 21st Annual International Conference on Intelligent Systems for Molecular Biology & 12th European Conference on Computational Biology, Berlin, Germany, Jul. 2013.

- 4) K. Etchuya, Y. Mukai: Characteristics for sugar modification from protein structure, 38th FEBS Congress, St. Petersburg, Russia, Jul. 2013.

- 5) R. Nambu, Y. Mukai: Sequence analysis around transmembrane regions and discrimination of subcellular localization of type II membrane proteins, 38th FEBS Congress, St. Petersburg, Russia, Jul. 2013.

- 6) T. Ogawa, K. Etchuya, R. Nambu, Y. Mukai: Physicochemical properties around cleavage sites by Caspase family proteases, 38th FEBS Congress, St. Petersburg, Russia, Jul. 2013.

- 7) N. Takachio, T. Konishi, H. Takata, T. Terasaki, N. Kato, T. Sasaki, Y. Mukai: Subcellular localization of GFP-signal sequence fusion

proteins, 38th FEBS Congress, St. Petersburg, Russia, Jul. 2013.

8) S. Tsunemoto, R. Nambu, K. Etchuya, Y. Mukai: Prediction of secondary structure for loop domains in transmembrane proteins, 38th FEBS Congress, St. Petersburg, Russia, Jul. 2013.

9) K. Nagashima, T. Kimura, Y. Miyazaki, Y. Mukai, T. Terasaki: Purification of Apoptosis Protein using II-D Cell Blot Method, 38th FEBS Congress, St. Petersburg, Russia, Jul. 2013.

10) K. Hamada, K. Etchuya, R. Nambu, Y. Mukai: Correlation between signal peptide sequences and physicochemical properties of mature formed proteins, 38th FEBS Congress, St. Petersburg, Russia, Jul. 2013.

11) K. Etchuya, H. Tanaka, T. Terasaki, T. Sasaki, Y. Mukai: Solvent accessibility and amino acid propensities around glycosylation sites depending on the kinds of oligosaccharides, Pacific Symposium on Biocomputing 2013, Hawaii, USA, Jan. 2013.

12) R. Nambu, T. Sasaki, Y. Mukai: Analysis and discrimination of subcellular localization based on amino acid sequences of membrane proteins, Pacific Symposium on Biocomputing 2013, Hawaii, USA, Jan. 2013.

13) Y. Mukai, H. Tanaka, T. Konishi, H. Takata, N. Takachio, M. Ikeda, T. Sasaki: Computational detection of organelle sorting signals in membrane localized proteins, 9th International Conference on Flow Dynamics (Invited), Sendai, Japan, Sep. 2012.

14) H. Tanaka, T. Konishi, T. Sasaki, M. Ikeda, Y. Mukai: A computational method for detecting GPI-anchored proteins and their GPI modification sites by neural networks, 11th European Conference on Computational Biology, Basel, Switzerland, Sep. 2012.

15) T. Konishi, H. Takata, N. Takachio, T. Sasaki, N. Kato, T. Terasaki, Y. Mukai: Subcellular localization and transport efficiency of protein sorting signal sequence, 22nd IUBMB & 37th FEBS Congress, Sevilla, Spain, Sep. 2012.

16) H. Tanaka, T. Sasaki, Y. Mukai: Computational prediction of GPI modification sites in GPI-anchored proteins, 22nd IUBMB & 37th FEBS Congress, Sevilla, Spain, Sep. 2012.

17) 永島啓矢, シティ・マイサラ, 向井有理, 寺崎武夫: 二次元セルプロット法の開発, 第3回日本生物物理学会関東支部会, 東京, 2014年3月.

18) 濱田康太, 越中谷賢治, 南部龍平, 向井有理: シグナルペプチド配列と成熟タンパク質の細胞内局在との相関性, 第3回日本生物物理学会関東支部会, 東京, 2013年12月.

19) 越中谷賢治, 南部龍平, 向井有理: 糖転移酵素のタンパク質立体構造認識および糖鎖修飾の立体構造への影響, 第36回日本分子生物学会年会, 神戸, 2013年12月.

20) 南部龍平, 越中谷賢治, 向井有理: 膜タ

ンパク質の膜貫通領域周辺のアミノ酸配列に着目した細胞内局在性解析, 第36回日本分子生物学会年会, 神戸, 2013年12月.

21) 小川翼, 越中谷賢治, 向井有理: タンパク質のパルミトイル化要因の検討, 第36回日本分子生物学会年会, 神戸, 2013年12月.

22) 高知尾尚志, 小西達也, 高田浩武, 寺崎武夫, 加藤徳剛, 佐々木貴規, 向井有理: 人工シグナル配列を用いたGPIアンカー型タンパク質のGPI修飾能の調査, 第36回日本分子生物学会年会, 神戸, 2013年12月.

23) 常本沙和子, 南部龍平, 越中谷賢治, 向井有理: 膜貫通タンパク質のループ領域二次構造予測のための条件検討, 第36回日本分子生物学会年会, 神戸, 2013年12月.

24) 永島啓矢, シティ・マイサラ, 向井有理, 寺崎武夫: 二次元セルプロット法を用いたアポトーシスタンパク質の解析, 第36回日本分子生物学会年会, 神戸, 2013年12月.

25) 濱田康太, 越中谷賢治, 南部龍平, 向井有理: シグナルペプチド配列と成熟タンパク質の膜透過との相関, 第36回日本分子生物学会年会, 神戸, 2013年12月.

26) 越中谷賢治, 池田(向井)有理: タンパク質の二次構造と糖鎖修飾の相関性, 第13回蛋白質科学会年会, 鳥取, 2013年6月.

27) 永島啓矢, 木村尊斗, 宮崎吉弘, 池田(向井)有理, 寺崎武夫: 二次元セルプロット法を用いたアポトーシスタンパク質の分離, 第13回蛋白質科学会年会, 鳥取, 2013年6月.

28) 田中大貴, 小西達也, 佐々木貴規, 池田(向井)有理: GPI修飾要因の検討および修飾位置予測法の確立, 第2回日本生物物理学会関東支部会, 東京, 2013年3月.

29) 越中谷賢治, 池田(向井)有理: タンパク質の二次構造と糖鎖修飾の相関, 第2回日本生物物理学会関東支部会, 東京, 2013年3月.

30) 越中谷賢治, 田中大貴, 寺崎武夫, 佐々木貴規, 池田(向井)有理: 一次配列から見た糖転移酵素の選択性, 第35回日本分子生物学会年会, 福岡, 2012年12月.

31) 田中大貴, 小西達也, 佐々木貴規, 池田(向井)有理: アミノ酸の物理化学的・構造的特徴を用いたタンパク質のGPI修飾位置予測, 第35回日本分子生物学会年会, 福岡, 2012年12月.

32) 小西達也, 高田浩武, 高知尾尚志, 佐々木貴規, 加藤徳剛, 寺崎武夫, 池田(向井)有理: 人工シグナル配列を導入したGFPの細胞内局在性観察, 第35回日本分子生物学会年会, 福岡, 2012年12月.

33) 南部龍平, 佐々木貴規, 池田(向井)有理: 膜タンパク質の疎水性領域に着目した細胞内局在性解析, 第35回日本分子生物学会年会, 福岡, 2012年12月.

〔図書〕(計0件)

〔産業財産権〕

○出願状況(計2件)

名称：位置特異的スコアの算出装置、算出方法及びプログラム、GPI アンカー修飾部位の特定装置、特定方法及びプログラム、並びに GPI アンカー修飾部位の判定装置、判定方法及びプログラム

発明者：田中大貴, 池田有理, 佐々木貴規

権利者：明治大学

種類：特願

番号：2012-189907

出願年月日：2012 年 9 月

国内外の別：国内

名称：ポリペプチド、遺伝子、発現ベクター、及びタンパク質を小胞体に輸送する方法、並びに、シグナルペプチドの設計方法

発明者：小西達也, 高田浩武, 池田有理, 高知尾尚志, 佐々木貴規

権利者：明治大学

種類：特願

番号：2012-192318

出願年月日：2012 年 9 月

国内外の別：国内

○取得状況（計 1 件）

名称：立体構造データ帰属方法、立体構造データ帰属プログラム及び立体構造データ帰属装置

発明者：池田修己, 池田有理, 矢葺幸光, 諏訪牧子

権利者：池田修己, 池田有理, 矢葺幸光, 諏訪牧子（個人）

種類：特許

番号：第 5433894 号

取得年月日：2013 年 12 月

国内外の別：国内

〔その他〕

本研究成果は情報数理バイオ株式会社との共同研究に発展した。情報数理バイオ株式会社の実績紹介ページより、本研究の成果の一部を以下の URL で公開している。

<http://imsbio.jp/GPIAPsPred/>

## 6 . 研究組織

### (1) 研究代表者

池田 有理 (IKEDA, Yuri)

明治大学工学部電気電子生命学科・准教授

研究者番号：30371082

### (2) 研究分担者：なし

### (3) 連携研究者：なし