## 科学研究**費**補助金研究成果報告書

平成 22 年 6月 28 日現在

研究種目:基盤研究(B) 研究期間:2007~2009 課題番号:19300080

研究課題名(和文)ニューラルネットワーク及びサポートベクタマシンによる

蛋白質 O型糖鎖修飾部位の予測

研究課題名(英文) Prediction of the Protein O-glycosylation site by Neural Networks and Support Vector Machine

研究代表者

西川 郁子 (NISHIKAWA IKUKO) 立命館大学・情報理工学部・教授

研究者番号:90212117

研究成果の概要(和文): 哺乳類蛋白質に対して, O 型糖鎖修飾を受ける部位を機械学習により予測し, その結果を通して修飾機構を解析した. 特に,多くの修飾部位が群発して発生することに着目し,その場合には周辺アミノ酸の組成比が,そうでない場合にはアミノ酸配列が修飾に影響を与えることを見出し,その特徴を具体的に解析した.蛋白質の天然変性領域に顕著に多く,進化的保存性や構造安定化,真核生物に多いこと等との関連を議論した.予測結果の実験的検証も行った.

研究成果の概要 ( 英文 ): O-Glycosylation sites in the mammalian protein is predicted by the machine learning to elucidate the glycosylation mechanisms. Especially, the crowded and the isolated glycosylation sites are separately considered, and it is found that the former is affected by the amino acid composition of the neighboring sites while the latter is affected by the amino acid sequence, and more detailed characteristics are investigated. O-Glycosylation sites are often found in the intrinsically disordered proteins, which lead to the biological understandings of the conservation property, the role in the structural stability and so on. Biological experiments are also executed for a verification of the prediction.

#### 交付決定額

(金額単位:円)

	直接経費	間接経費	合 計
2007 年度	9,100,000	2,730,000	11,830,000
2008 年度	4,100,000	1,230,000	5,330,000
2009 年度	1,600,000	480,000	2,080,000
年度			
年度			
総計	14,800,000	4,440,000	19,240,000

研究分野:知能情報学

科研費の分科・細目:情報学・感性情報学・ソフトコンピューティング

キーワード:サポート・ベクタ・マシン, O型糖鎖修飾,機械学習,哺乳類タンパク質,

天然変性領域,膜貫通タンパク質,ニューラルネットワーク,

#### 1.研究開始当初の背景

ポストゲノム時代における生命システム 理解の主要課題の一つに,蛋白質への糖鎖修 飾とその機能の解明が挙げられる.哺乳類の 主要な糖鎖修飾には二種あるが,その一つに も関わらず,機構やコンセンサス配列が未だ 知られていない *O*型糖鎖修飾に着目した.

の型糖鎖修飾は,関与する酵素が多様で複雑なこともあり機構の実験的解明は困難だったが,修飾部位に関する実験的報告は急速に蓄積され,最新データが公開されている.その結果,機械学習による修飾予測に十分な数の信頼性の高いデータが得られつつあり,機械学習による有意な予測が可能となった.そこで機械学習を通じた,機構の解析を目指した.

## 2. 研究の目的

- (1)哺乳類蛋白質に対する最新の実験データを用いて,機械学習により,O型糖鎖修飾を受ける蛋白質部位を予測する.アミノ酸の配列や組成比,生化学的性質,蛋白質の構造情報,進化的情報など各種データを学習器へ入力とし,それらから最も高精度な予測を実現する.
- (2)高精度な予測器が得られれば,未知の修飾部位を予見しうる.それにより,創薬や糖鎖修飾の不全に起因する疾患の治療にも寄与しうる.
- (3)修飾に寄与する要因を明らかにすることで、修飾機構の解析を行う.
- (4)生物学実験を実施することで,予測結果の検証する.実験データ,情報学的予測,実験による検証のループを廻すことで,情報学と生物学が緊密に連携した研究を推進する.

#### 3.研究の方法

- (1)UniProt や PDB などの公開されている 蛋白質データベースから,最新の実験データ を取得する.O型糖鎖修飾が実験的に確認さ れている全哺乳類蛋白質,全修飾部位に対し て,そのアミノ酸配列や立体構造,進化的情 報,生化学的特性などを取得する.
- (2)それらを入力データおよび教師信号として,予測器の学習を行う.具体的には,サポート・ベクタ・マシン(SVM)と階層型ニューラルネットワークを用いる.修飾の有無を判定するニクラス分類課題とし,ネガティブデータとしては,修飾報告がないセリンおまびスレオニンを用いる.ただし,ネガティブであることの実験的検証が可能なものは多くないため,一部を教師信号なしのデータとして扱う半教師あり学習も行う.

- (3)入力データとして,対象部位周辺の近傍 範囲,データの種類,コーディング法などを 変えた予測実験を網羅的に行い,予測精度の 向上,ひいては修飾に寄与する要因を明らか にする.予測精度は,マシューズ相関係数や 正答率などにより定量化する.
- (4)寄与する要因として,殆どの従来研究では,アミノ酸残基配列の一次元情報,あるいは蛋白質全体にわたる統計量のいずれであるかを論じている.ここでは,その単純な二分法のみでなく,異なる修飾機構が複数存在する可能性も含めて,解明を目指す.
- (5)寄与する要因が得ることで,修飾の条件を具体的に求める.例えば,修飾に有意に作用するアミノ酸配列や組成比,生化学的条件や蛋白質の構造などの諸条件であり,それを得ることで,O型糖鎖修飾がいかにして蛋白質の構造安定性や多様性に寄与し,その不全がいかにして疾患に結びつくかの手がかりが得られることになる.
- (6)機械学習では修飾が予測されるにも関わらず、実験的な報告事例がない事例に対して、実験的検証を行う. すなわち、該当するポリペプチドを実験的に合成し、GalNAcとOglycosyltransferaseを加えて反応させ、生成された反応物に対して質量分析を行い、修飾の有無を確認する.

## 4. 研究成果

(1)各修飾部位をアミノ酸配列上で群発的に発生するものと散発的なものに二分類し, 予測に有効な情報や,修飾部位周辺の統計性に違いを見い出した.それぞれに予測器を構築し,組合せることで,最も高精度な予測を実現した.

の型糖鎖修飾の報告がある全哺乳類蛋白質の全修飾部位に対して,一定の基準で群発部位と散発部位に分け,各々で SVM による予測を行った.散発部位では比較的近傍の位置特異的なアミノ酸の存在が,群発部位では 15近傍程度の比較的広い周辺部位のアミノ酸組成比が,有効であることを示した.

散発的修飾に関与する位置特異的なアミノ酸を具体的に求め,既知のP+1,-3以外に複数の存在を見い出した.いずれも,散発的修飾により顕著であった.さらに,アミノ酸配列に対する独立成分分析により,それらが独立に存在していることを明らかにした.

得られた 2 種の SVM を組み合わせることで,単一の SVM より高い予測性能を実現した.

#### (2)タンパク質の構造的特徴との関連

修飾部位の分布を調べた結果,タンパク質のドメインには少なく,天然変性(intrinsically disordered)領域に顕著に多いこと,それらは特に群発的修飾に特徴的なことを見出した.膜貫通タンパク質では,いずれも膜外で修飾されていた.

定まった構造を持たない天然変性領域は,機能的にも意味がないと考えられていたが,近年,その機能が注目されている. *O*型糖鎖修飾は,タンパク質の構造安定化や機能の多様性の発現に関与し,真核生物に多く見られ,進化的保存性が悪い領域に多く見られるなど,天然変性領域に共通する特徴が多いことを指摘し,その生物学的意味を議論した.

## (3)分子生物学的実験の実施

SVMによる予測結果より,修飾の可能性が高いと思われる部位を,群発と散発から各1部位を選び,合成した20ポリペプチドを用いて,分子生物学的実験を行った.すなわち,GalNAcと酵素を加えて反応物を得た.ただしカラム洗浄の不備のため質量分析による確認には至らなかった.反応物は保管してあり,実験の再開で再度の確認は可能である.

## 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7件)

著者名: Yasushi Tani, Touri Funatsu, Hisashi Ashida, Masahiro Ito, Saki Itonori, Mutsumi Sugita, Kenji Yamamoto、論文標題:Novel neogala-series glycosphingolipids with terminal mannose and glucose residues from Hirsutella rhossiliensis, an aureobasidin A-resistant ascomycete fungus.、雑誌 名:Glycobiology、査読:有、巻:20(4)、 発行年:2010、ページ:433-441 著者名:Hisao Kojima, Takashi Inoue, Mutsumi Sugita, Saki Itonori, Masahiro Ito.、論文標題:Biochemical Studies on Sphingolipid of Artemia franciscana (I) Isolation and Characterization of Sphingomyelin、雑誌名:Lipid、査読: 有、巻:掲載決定(in press) 発行年: 2010、ページ:掲載決定 (in press) 著者名: Hu-Chuan Lu, Guo-Liang Fang, Chao Wang and Yen-Wei Chen、論文標題: A novel method for gaze tracking by local pattern model and support vector regressor

雑誌名:Signal Processing、查読:有、

巻:90、発行年:2010、ページ:1290-1299 著者名: Masahiro Ito, Kazuhito Yokoi. Takashi Inoue, Shogo Asano, Rei Hatano, Ryota Shinohara, Saki Itonori, Mutsumi Sugita、論文標題:Sphingomyelins in Four Ascidians, Ciona intestinalis, Halocynthia roretzi, Halocynthia aurantium, and Styela clava、雑誌名: J. Oleo Sci.、 查読:有、巻:58(9) 発行年:2009、ページ:473-480 著者名:Rui Xu and Yen-Wei Chen、論文 標題:Generalized N-dimensional Principal Component Analysis (GND-PCA) and Its Application on Construction of Statistical Appearance Models for Medical Volumes with Fewer Samples. 雑誌名:Neurocomputing、査読:有、 巻:72、発行年:2009、ページ:2276-2287 著者名: Xu Qiao, Rui Xu, Yen-Wei Chen, Takanori Igarashi, Keisuke Nakao and Akio Kashimoto、論文標題:Generalized N-dimensional Principal Component Analysis (GND-PCA) Based Statistical Appearance Modeling of Facial Images with Multiple Modes、雑誌名: IPSJ Trans. On Computer Vision and Applications 查読:有、巻:1、発行年:2009、 ページ: 231-241 著者名:谷口忠大,高木圭太,榊原一紀,

著者名:谷口忠大、高木圭太、<u>榊原一紀</u>、 西川郁子、論文標題:地産地消型電力ネットワークの為の Natural Actor-Critic を用いた自動取引 エージェントの構築、 雑誌名:日本知能情報ファジィ学会論文 誌、査読:有、巻:21、発行年:2009、 ページ:567-577

#### [学会発表](計38件)

発表者名: Nakajima, Y., <u>Sakakibara, K., Ito, M.</u>, <u>Nishikawa, I.</u>、発表標題: Prediction of the O-glycosylation by Support Vector Machine and Independent Component Analysis for Amino Acid Sequence around O-glysocylation、学会名等: 20th International Conference on Genome Informatics、発表年月日: 2009年12月14日、発表場所:パシフィコ横浜(神奈川県) 発表者名: Mazaki, Y., Mori, H., <u>Ito, M.</u>,

発表者名: Mazaki, Y., Mori, H., Ito, M., Tohsato, Y.、発表標題: Kinetic Modeling of Central Metabolic Pathway for Phenotype MicroArray Analysis、学会名等: 20th International Conference on Genome Informatics、発表年月日: 2009年12月14日、発表場所: パシフィコ横浜(神奈川県)

発表者名: Shizuka Hino, Dai Sasahara,

Ayako Terasawa, Yusuke Ishido, Toshya Hayano, <u>Masahiro Ito</u>、発表標題: Quantititative proteome analysisof successive protein expression profiles during early embryogenesis in Caenorhabditis elegans、学会名等:第32回日本分子生物学年会、発表年月日:2009年12月9日、発表場所:パシフィコ横浜(神奈川県)発表者名:Ayako Terasawa, D. Sasahara,

発表者名: Ayako Terasawa, D. Sasahara, Shizuka Hino, Toshiya Hayano, <u>Masahiro Ito</u>、発表標題: Identification and quantitative analysis of proteins expressed during Caenorhabditis briggsae development. 学会名等:第32回日本分子生物学年会、発表年月日:2009年12月9日、発表場所:パシフィコ横浜(神奈川県)

発表者名: Dai Sasahara, S. Hino, A. Terasawa, Toshiya Hayano, Masahiro Ito, 発表標題:Protein expression profile in the larva 1 stage differ considerably between Caenorhabditis elegans and Caenorhabditis briggsae. 学会名等:第32回日本分子生物学年会、 発表年月日:2009年12月9日、発表場 所:パシフィコ横浜(神奈川県) 発表者名:Xianhua Han, Yen-Wei Chen and Xiang Ruan、発表標題:Object Class Recognition with Supervised Nonlinear Neighborhood Embedding of Visual Words, 学会名等: the First International Conference on Internet Multimedia Computing and Service (ICIMCS2009), 発表年月日: 2009 年 11 月 24 日、発表場 所:Yunnan(中国)

発表者名:<u>榊原一紀</u>,前田忠之,谷口忠大,<u>西川郁子</u>、発表標題:太陽光発電による分散型電力ネットワーク構築のための静的最適化モデルに基づく挙動解析、学会名等:第52回自動制御連合講演会、発表年月日:2009年11月21-22日、発表場所:大阪大学(大阪府)

発表者名: <u>Yen-Wei Chen</u> and Guifang Duan、 発表標題: Independent Component Analysis Based Ring Artifact Reduction in Cone-Beam CT Images、学会名等: 2009 IEEE 16th International Conference on Image Processing (ICIP2009)、発表年月 日: 2009 年 11 月 9 日、発表場所: Cairo (エジプト)

発表者名: Huchuan Lu, Yunyun Liu, Zhipeng Sun and <u>Yen-Wei Chen</u>、発表標題: An Active Contours Method Based on Intensity and Reduced Garbor Features for Texture Segmentation、学会名等: 2009 IEEE 16th International

Conference on Image Processing (ICIP2009)、発表年月日: 2009年11月9 日、発表場所: Cairo (エジプト) 発表者名: Guifang Duan and Yen-Wei Chen、 発表標題: Improved Active Shape Model for Automatic Optical Phase Identification of Microdrill Bits in Printed Circuit Board Production、学 会名等: 2009 IEEE 16th International Conference on Image Processing (ICIP2009)、発表年月日: 2009年11月9 日、発表場所:Cairo(エジプト) 発表者名: Yen-Wei Chen, Aya Mimori and Chen-Lun Lin、発表標題: Hybrid Particle Swarm Optimization for 3-D Image Registration、学会名等: 2009 IEEE 16th International Conference on Image Processing (ICIP2009)、発表年月日: 2009年11月8日、発表場所: Cairo (工 ジプト)

発表者名: Tomoko Tateyama, Hossein Foruzan and <u>Yen-Wei Chen</u>、発表標題: 2D-PCA based Statistical Shape Model from few Medical Samples、学会名等: The Fifth Int. Conf. on Intelligent Information Hiding and Multimedia Signal Processing、発表年月日: 2009年10月13日、発表場所:メルパルク京都(京都府)

発表者名: Farshid Babapour Mofrad, Reza Aghaeizadeh Zoroofi, Yen-Wei Chen, Ali Abbaspour、発表標題: Evaluation of Liver Shape Approximation and Characterization、学会名等: The Fifth Int. Conf. on Intelligent Information Hiding and Multimedia Signal Processing、発表年月日: 2009年10月12日、発表場所:メルパルク京都(京都府)

発表者名: Nakajima, Y., <u>Sakakibara, K.</u>, <u>Ito, M., Nishikawa I.</u>、発表標題: Prediction of the 0-glycosylation by Support Vector Machines and Characteristics of the Crowded and Isolated 0-glycosylation Sites、学会名等: The Fifth Int. Conf. on Intelligent Information Hiding and Multimedia Signal Processing、発表年月日: 2009年10月12日、発表場所:メルパルク京都(京都府)

発表者名:Xianhua Han and <u>Yen-Wei Chen</u> 発表標題:Image Categorization by Learned PCA Subspace of Combined Visual-words and Low-level Features 学会名等:The Fifth Int. Conf. on Intelligent Information Hiding and Multimedia Signal Processing、発表年 月日:2009 年 10 月 11 日、発表場所:メルパルク京都(京都府)

発表者名: Masataka Seo and <u>Yen-Wei Chen</u> 発表標題: Face Image Metamorphosis with an Improved Multilevel B-Spline Approximation、学会名等: The Fifth Int. Conf. on Intelligent Information Hiding and Multimedia Signal Processing、発表年月日: 2009 年 10 月 11 日、発表場所:メルパルク京都(京都府)

発表者名:A. Takashima, S. Namiki, T. Kazawa, Stephan Haupt, <u>I. Nishikawa</u>, H. Ikeno, R. Kanzaki、発表標題:Simulation of a small neural network involved in pheromone orientation in male silkmoths、学会名等:Asia Simulation Conference 2009 (JSST2009) 発表年月日:2009年10月9日、発表場所:立命館大学(滋賀県)

発表者名: I. Nishikawa, Y. Igarashi, A. Takashima, S. Namiki, T. Kazawa, Stephan Haupt, H. Ikeno, R. Kanzak、発表標題: A simulation of the dynamics of the premotor center in an insect brain to generate the programmed behavior for the pheromone orientation、学会名等: Asia Simulation Conference 2009 (JSST2009) 発表年月日: 2009年10月9日、発表場所:立命館大学(滋賀県)

発表者名: K. Tanabe, D. Yamada, <u>K. Sakakibara</u>, and <u>I. Nishikawa</u>、発表標題: Acquisition of Negotiation Rules for AGV Transportation System by Genetic Based Machine Learning Techniques、学会名等: Asia Simulation Conference 2009 (JSST2009) 発表年月日: 2009 年 10 月 8 日、発表場所: 立命館大学(滋賀県)

発表者名: Danni Ai, Xianhua Han, Xiang Ruan and <u>Yen-Wei Chen</u>、発表標題: A Novel Color Independent Components based SIFT Descriptor、学会名等: Asia Simulation Conference 2009 (JSST2009) 発表年月日: 2009 年 10 月 8 日、発表場所:立命館大学(滋賀県)

- 21 発表者名: Xu Qiao, Yen-Wei Chen, Xian-Hua Han, T. Igarashi and K. Nakao、 発表標題: Synthesis of Multiple Pose Facial Images Using Tensor-Based Subspace Learning Method、学会名等: 2009 IEEE 12th International Conference on Computer Vision Workshops (ICCV2009 workshops)、発表年月日: 2009 年 9 月 30 日、発表場所: 京都国際会議場・京都大学(京都府)
- 22 発表者名: Takashima A., Namiki S.,

- Kazawa T., Haupt Stephan Shuichi, Nishikawa I., Ikeno H. and Kanzaki R. 発表標題: Simulation of a small neural network related to programmed behavior in pheromone orientation in male silkmoths、学会名等: 2nd INCF Congress of Neuroinformatics、発表年月日: 2009年9月7日、発表場所: Plzen (Czech)

Reconstruction、学会名等:Fifth International Conference on Natural Computation、発表年月日:2009年8月 15日、発表場所:Tianjin(中国)

- 32 発表者名: 西尾匡、小島寿夫、糸乗前、 杉田陸海、<u>伊藤將弘</u>、発表標題: エボヤ Styela clava 中性スフィンゴ糖脂質の構 造解析、学会名等: 第51回日本脂質生化 学会年会、発表年月日: 2009年7月30 日、発表場所: ウィルあいち(愛知県)
- 33 発表者名:H. Ikeno, T. Kazawa, S. Namiki, Stephan Haupt, A. Takashima, R. Fukushima, <u>I. Nishikawa</u>, R Kanzaki、発表標題: Development of standard brain for silkworm moth, bombyx mori, linked with a neuron database、学会名等: 18th Annual Computational Neuroscience Meeting CNS2009、発表年月日: 2009 年7月20日、発表場所: Berlin(ドイツ)
- 34 発表者名:谷口忠大,高木圭太,<u>榊原</u> 一紀,西川郁子、発表標題:地産地消型 電力ネットワークの為のNatural Actor-Criticを用いた自動取引 エージェントの構築、学会名等:第25回ファジィシステムシンポジウム、発表年月日:2009年7月14 16日、発表場所:筑波大学(茨城県)

Electrophoresis、学会名等:17th International C. elegans Meeting、発 表年月日:2009年6月25日、発表場所: Los Angeles(米国)

- 37 発表者名: <u>榊原一紀</u>, 伊山宗吉, <u>西川</u> <u>郁子</u>、発表標題: 配送計画問題に対する 問題分解と分枝限定法およびメタヒューリスティクス のハイブリッド法、学会名等: 第53 回システム制御情報学会研究発表講演会、発表年月日: 2009 年5月 20-22 日、発表場所: 神戸市産業振興センター(兵庫県)
- 38 発表者名:中島有紀子,平昌憲,飯野修平,坂本浩隆,<u>榊原一紀</u>,<u>伊藤將弘</u>,西川郁子、発表標題: 0-型糖鎖修飾のサポートベクターマシンを用いた予測と独立成分分析による解析、学会名等:第53回システム制御情報学会研究発表講演会、発表年月日:2009年5月20-22日、発表場所:神戸市産業振興センター(兵庫県)

#### 〔図書〕(計0件)

# 〔産業財産権〕

出願状況(計0件) 取得状況(計0件)

### [その他]

http://sys.ci.ritsumei.ac.jp/

## 6.研究組織

## (1)研究代表者

西川 郁子(NISHIKAWA IKUKO) 立命館大学・情報理工学部・教授 研究者番号:90212117

#### (2)研究分担者

伊藤 將弘 (ITO MASAHIRO)

立命館大学・生命科学部・准教授 研究者番号:50388112

陳 延偉 (CHIN ENI)

立命館大学・情報理工学部・教授

研究者番号:60236841

榊原 一紀(SAKAKIBARA KAZUTOSHI)

立命館大学・情報理工学部・講師

研究者番号:60236841