

タンパク質の糖鎖修飾に関するバイオインフォマテイクス解析と工学的応用

メタデータ	言語: Japanese 出版者: 公開日: 2018-11-16 キーワード (Ja): キーワード (En): 作成者: 越中谷, 賢治 メールアドレス: 所属:
URL	http://hdl.handle.net/10291/19694

明治大学大学院理工学研究科

2016年度

博士学位請求論文

タンパク質の糖鎖修飾に関する
バイオインフォマティクス解析と工学的応用

**Bioinformatics analysis and engineering applications
related to protein glycosylation**

学位請求者 電気工学専攻

越中谷 賢治

目次

第1章 序論	1
1-1. タンパク質の糖鎖修飾	1
1-2. O型糖鎖修飾	4
1-3. タンパク質糖鎖修飾研究のためのバイオツールおよびデータベース	6
1-4. 糖鎖研究の現状と問題点および本研究の目的	11
第2章 一次配列に基づく糖種判別とその限界	15
2-1. 背景	15
2-2. データセット作成および解析方法	16
2-3. 結果および考察	20
第3章 糖タンパク質の二次構造解析	24
3-1. 背景	24
3-2. データセット作成および解析方法	26
3-3. 結果および考察	28
第4章 糖鎖修飾残基周辺の空間的アミノ酸出現傾向解析	39
4-1. 背景	39
4-2. データセット作成および解析方法	41
4-3. 結果および考察	42
第5章 タンパク質細胞内局在経路に基づいた糖鎖修飾分布の解析	48
5-1. 背景	48
5-2. データセット作成および解析方法	50
5-3. 結果および考察	51
第6章 タンパク質O型糖鎖修飾の糖種判別	55
6-1. 背景	55
6-2. データセット作成および解析方法	56
6-3. 結果および考察	59
第7章 結論	61
謝辞	65
参考文献	66
研究業績	72
付録	76

略語

[アミノ酸]

Ala: アラニン

Asn: アスパラギン

Asp: アスパラギン酸

Cys: システイン

Glu: グルタミン酸

Phe: フェニルアラニン

Ser: セリン

Thr: スレオニン

Trp: トリプトファン

Val: バリン

[糖]

Fuc: L-フコース

Gal: D-ガラクトース

GalNAc: *N*-アセチル-D-ガラクトサミン

GlcNAc: *N*-アセチル-D-グルコサミン

Glc: D-グルコース

Hex: D-ヘキソース

HexNAc: *N*-アセチル-D-ヘキソサミン

Man: D-マンノース

NeuNAc: *N*-アセチルノイラミン酸

Xyl: D-キシロース

[その他]

PDB: Protein Data Bank

PSSM: 位置特異的アミノ酸スコアマトリクス

Swiss-Prot: UniProt
Knowledgebase/Swiss-Prot protein sequence
database

第1章 序論

1-1. タンパク質の糖鎖修飾

生体を構成している細胞の中では、多くのタンパク質がそれぞれの役割を果たしている。我々の健康的な生活は、50万種以上ものタンパク質による分子機械としての働きによって支えられているともいえよう。タンパク質が生体内で正確に機能するために、従来は、タンパク質自身のアミノ酸の並びや配列の長さ、それらが形成する構造のみが重要であると考えられてきた。しかし、近年ではタンパク質の機能を直接コントロールする他の生体分子の寄与に注目が集まるようになってきた。

糖鎖は、DNA・タンパク質に続く第三の生命鎖とも呼ばれ、タンパク質に結合する翻訳後修飾の一種であり、その機能をコントロールする生体分子である。タンパク質のアミノ酸配列データベースである UniProt Knowledgebase/Swiss-Prot protein sequence database (<http://www.uniprot.org/>) (The UniProt Consortium, *Nucl. Acids Res.*, 2017) に格納されている真核生物のタンパク質の半数以上が、糖鎖修飾を受けている (Apwailer, *et al.*, *Biochim. Biophys. Acta.*, 1999)。図 1-1 に、糖

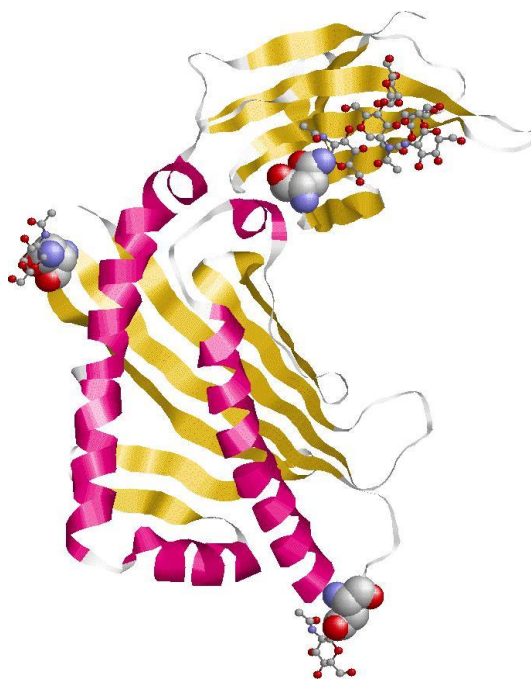


図 1-1. 3つの糖鎖修飾を受けたタンパク質の例
[PDBID: 1ZAG (human Zinc- α -2-glycoprotein)]
リボンモデルはタンパク質、ボール&スティックモデルは糖鎖、CPKモデルはタンパク質の糖鎖修飾残基を表す。

鎖修飾を受けているヒトの脂質移動因子タンパク質の立体構造を示す。糖鎖はタンパク質の折りたたみや機能発現、酵素活性を調節し、細胞の恒常性維持に寄与しており (Kornfeld, *et al.*, *Annu. Rev. Biochem.*, 1985; Varki, *Glycobiol.*, 1993; Hounsell, *et al.*, *Glycoconj. J.*, 1996; Lowe and Marth, *Annu. Rev. Biochem.*, 2003; Haltiwanger and Lowe, *Annu. Rev. Biochem.*, 2004)、それゆえ、糖鎖研究による新たな知見は、癌マーカーや糖鎖薬など疾患の治療へと応用されつつある。また、再生医療の分野においても、糖鎖は細胞分化の重要なマーカーとして知られている。今日では、我が国を中心に研究の進展がめざましい多能性細胞 (ES 細胞、iPS 細胞) の分野においても、発揮した多能性による分化先の細胞の種類を、糖鎖の解析によって確かめることができる。

以上で述べたように、糖鎖は我々の健康的な生活とも深く関わりのある生体分子であるが、真核細胞内の小胞体やゴルジ体に存在する糖転移酵素というタンパク質によって目的タンパク質の特定のアミノ酸に修飾されるという、大まかな修飾機構がわかっているにすぎない。糖鎖は結合する原子によって 3 パターンに大別されるが、主たる糖鎖修飾は *N* 型と *O* 型である。*N* 型糖鎖はタンパク質のアスパラギン (Asn) とセリン (Ser) またはスレオニン (Thr) 残基からなるモチーフ Asn-X-Ser/Thr における Asn 残基の N 原子に (Bause, *Biochem. J.*, 1983)、*O* 型糖鎖は Ser または Thr 残基の O 原子に結合する (Hansen, *et al.*, *Biochem. J.*, 1995)。わずかに見られる *C* 型糖鎖では、トリプトファン (Trp) 残基を中心とした Trp-X-X-Trp モチーフの最初の Trp 残基の C 原子に糖が結合する (Furmanek and Hofsteenge, *Acta Biochem. Pol.*, 2000)。

糖鎖は単糖が 2 つ以上結合してできるが、糖鎖修飾に見られる単糖には、L-フコース (Fuc)、D-ガラクトース (Gal)、D-グルコース (Glu)、D-マンノース (Man)、*N*-アセチル-D-ガラクトサミン (GalNAc)、*N*-アセチル-D-グルコサミン (GlcNAc)、*N*-アセチルノイラミン酸 (NeuNAc)、D-キシロース (Xyl) などが存在する。糖転移酵素ごとにタンパク質に修飾する糖の種類 (糖種) が決まっており、タンパク質のどの位置にどの糖種が修飾されるかということは、どの糖転移酵素によって標的タンパク質のどのような特徴が認識されているのかとい

うことに依存する。DNA やタンパク質に比較して糖鎖研究の歴史は浅く、今まさに機能解析が進められている最中であるが、生体内での役割は糖種特異的であることがわかってきた。したがって、それぞれのタンパク質に修飾される糖種を正確に把握することは、種々の生命現象を理解する上で本質的である。

本研究では、バイオインフォマティクス的手法により、3 パターンの糖鎖修飾のうち最も多く見られる O 型糖鎖が結合しているタンパク質の配列や構造の特徴を、糖の種類（糖種）ごとに抽出し、各種の O 型糖転移酵素が標的タンパク質のどのような性質を認識して選択特異的に糖鎖修飾を行っているかという点をふまえ、タンパク質 O 型糖鎖修飾の糖種判別法の開発を行った。糖鎖修飾における糖種選択性のメカニズムを把握することにより、適切な種類の糖鎖を標的タンパク質の任意の位置に導入してタンパク質の機能を人工的に制御でき、医療や創薬における糖鎖工学のさらなる応用展開につなげることも可能である。

1-2. O型糖鎖修飾

O型糖鎖修飾は主に細胞内の小胞体またはゴルジ体の2箇所で行われる(図1-2)。タンパク質が分泌タンパク質または膜タンパク質に成熟する過程で、翻訳後修飾として糖鎖修飾を受ける。それゆえ、分泌タンパク質や膜タンパク質のほとんどは、糖タンパク質である。O型糖鎖修飾は、最初に修飾された単糖を元に伸張されることで、タンパク質を

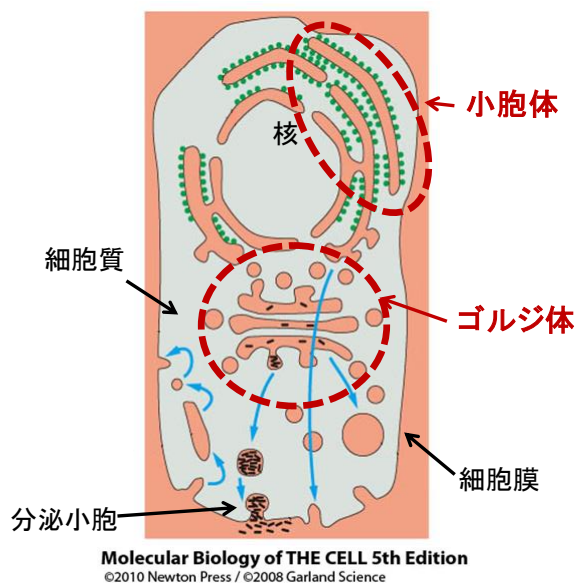


図1-2. タンパク質の細胞内局在と糖鎖修飾

コアとした巨大なビン洗浄ブラシ構造をとる。この構造は、プロテオグリカンと呼ばれ、負電荷を帯び親水性を持つことから耐衝撃性や粘着性を示し、さまざまな大きさの孔を持った巨大なゲルを形成する。生体成分として多様な機能性を持つプロテオグリカンは、最も重要な生体成分のひとつであり、主要な各種臓器、脳、皮膚を始めとした体組織中の細胞外マトリックスや細胞表面に存在するほか、関節などの軟骨の主

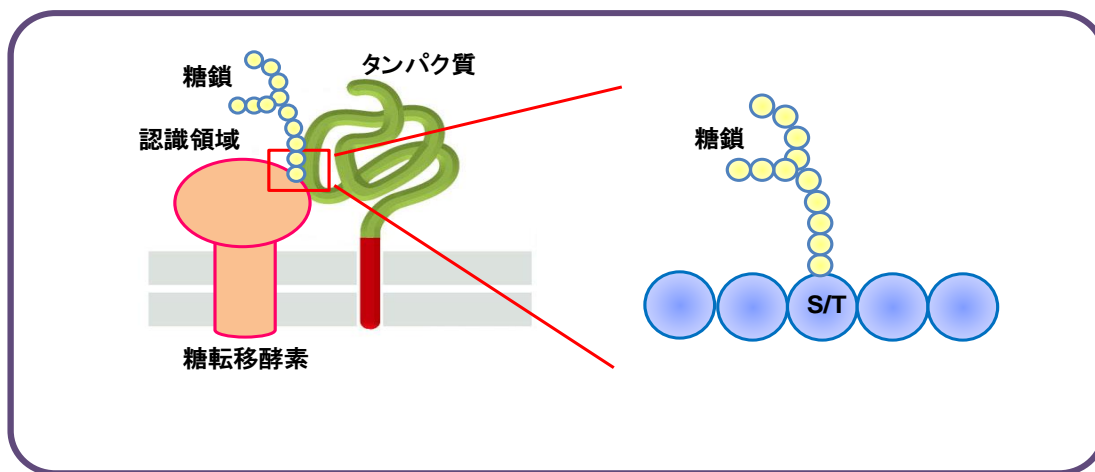


図1-3. O型糖鎖修飾およびそのモチーフ残基

成分としても存在している (Kolset, *et al.*, *Biochem. J.*, 2004; Bernfield, *et al.*, *Annu. Rev. Biochem.*, 1999; Esko, *et al.*, *Annu. Rev. Biochem.*, 2002)。前述の通り、O 型糖鎖修飾は Ser もしくは Thr 残基におけるヒドロキシ基の O 原子に糖を結合する反応であり、現在までに 7 糖種 (Gal, GalNAc, GlcNAc, Glc, Fuc, Xyl, Man) の結合が確認されている。

これらのセリン (Ser) またはスレオニン (Thr) 残基は O 型糖鎖修飾のモチーフ残基 (図 1-3) としてよく知られているが、配列中にモチーフ残基があれば必ずしも修飾が行われるとは限らないことから、O 型糖鎖修飾の明確なルールは明らかにされていない。

1-3. タンパク質糖鎖修飾研究のためのバイオツールおよびデータベース

[UniProt Knowledgebase/Swiss-Prot protein sequence database]

UniProt Knowledgebase/Swiss-Prot protein sequence database (Swiss-Prot) (The UniProt Consortium, *Nucl. Acids Res.*, 2017) とは、タンパク質の情報が Web 上で公開されているデータベースの一つである。タンパク質情報が公開されているデータベースは他にもあるが、Swiss-Prot はアミノ酸配列における注釈が豊富であり、信頼性が高いというメリットがある。

UniProtKB	Protein knowledgebase, consists of two sections: ★ Swiss-Prot, which is manually annotated and reviewed. ★ TrEMBL, which is automatically annotated and is not reviewed. Includes complete and reference proteome sets .
UniRef	Sequence clusters, used to speed up sequence similarity searches.
UniParc	Sequence archive, used to keep track of sequences and their identifiers.
Supporting data	Literature citations , taxonomy , keywords , subcellular locations , cross-referenced databases and more .

- [Text search](#)
- [Sequence similarity searches \(BLAST\)](#)

図 1-4. UniProt Knowledgebase トップページ (<http://www.uniprot.org/>)

Swiss-Prot のタンパク質情報を Text 形式で表示した (図 1-5)。Swiss-Prot のアノテーション情報はその信頼性によって、注釈なし、“By similarity”、“Potential”、“Probable” に分類される。注釈なしは何らかの実験的手法により確認が得られている情報、“By similarity” は実験的確認が得られた配列またはタンパク質と同

じファミリーに属する、または保存性の高い重要なドメインを持つなどの類似点が見られた場合のみ、“Probable” および “Potential” は配列解析や予測ツールを用いて、いくつかの論理的または決定的な確証から推測可能な情報を対象としており、“Probable” の方がより強く確証がある。

```

ID   FA7_HUMAN                               Reviewed:          466 AA.
AC   P08709; B0YJC8; Q14339; Q5JVF1; Q5JVF2; Q9UD52; Q9UD53; Q9UD54;
DT   01-JAN-1988, integrated into UniProtKB/Swiss-Prot.
DT   01-JAN-1988, sequence version 1.
DT   22-JAN-2014, entry version 200.
    (中略)
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
OC   Catarrhini; Hominidae; Homo.
    (中略)
CC   -!- FUNCTION: Initiates the extrinsic pathway of blood coagulation.
CC   Serine protease that circulates in the blood in a zymogen form.
CC   Factor VII is converted to factor VIIa by factor Xa, factor XIIa,
CC   factor IXa, or thrombin by minor proteolysis. In the presence of
CC   tissue factor and calcium ions, factor VIIa then converts factor X
CC   to factor Xa by limited proteolysis. Factor VIIa will also convert
CC   factor IX to factor IXa in the presence of tissue factor and
CC   calcium.
    (中略)
DR   PDB; 1BF9; NMR; -; A=105-145.
DR   PDB; 1CVW; X-ray; 2.28 A; H=213-466, L=150-204.
DR   PDB; 1DAN; X-ray; 2.00 A; H=213-466, L=61-212.
DR   PDB; 4JZE; X-ray; 1.52 A; H=213-466, L=150-204.
FT   SIGNAL          1      20      Potential.
FT   PROPEP          21     60
    (中略)
FT   CARBOHYD        112    112      0-linked (Glc...).
FT                                     /FTId=CAR_000007.
FT   CARBOHYD        120    120      0-linked (Fuc).
    (中略)
SQ   SEQUENCE 466 AA; 51594 MW; 9B5D501669D67B06 CRC64;
MVSQALRLLC LLLGLQGCLA AGGVAKASGG ETRDMPWKPG PHRVFVTQEE AHGVLHRRRR
ANAFLEELRP GSLERECKEE QCSFEEAREI FKDAERTKLF WISYSDGDQC ASSPCQNGGS
CKDQLQSYIC FCLPAFEGRN CETHKDDQLI CVNENGGCEQ YCS DHTGTKR SCRCHEGYSL
LADGVSTPT VEYPCGKIPI LEKRNASKPQ GRIVGGKVCP KGECPWQVLL LVNGAQLCGG
TLINTIWWVS AAHCFDKIKN WRNLIAYLGE HDLSEHDGDE QSRRVAQVII PSTYVPGTTN
HDIALLRHQ PVVLTDHVVP LCLPERTFSE RTLAFVRFSL VSGWGQLLDR GATALELMVL
NVPRLMTQDC LQSRKVGDS PNITEYMFCA GYSDGSKDSC KGDGGPHAT HYRGTWYLTG
IVSWGQGCAT VGHFGVYTRV SQYIEWLQKL MRSEPRPGVL LRAPFP
//

```

図 1-5. Swiss-Prot の表示 (Text 形式)

[CD-HIT]

CD-HIT (Huang, *et al.*, *Bioinformatics*, 2010) は、Web 上で公開されているクラスタリングソフトであり、40~100%までの配列類似性に基づいて冗長性の排除を行うことが可能である。

図 1-6. CD-HIT トップページ

(http://weizhong-lab.ucsd.edu/cdhit_suite/cgi-bin/index.cgi?cmd=cd-hit)

[Protein Data Bank (PDB)]

タンパク質と核酸の三次元座標（立体配座）を蓄積している国際的な公共のデータベースである。Protein Data Bank (PDB) (Berman, *et al.*, *Nucl. Acids Res.*, 2000; Berman, *et al.*, *Nucl. Acids Res.*, 2007) に蓄積されている構造データは X 線結晶解析法、NMR 法（核磁気共鳴法）などによって実験的に決定された三次元座標を扱っている。予測法などによって理論的に推定された三次元座標は対象外となっている。生物学の構造データの中心的なデータベース PDB は、構造生物学や、構造ゲノミクスといった研究の重要なデータソースとなっている。バ



イオインフォマティクスの分野においても構造・機能・進化などの視点から三次元座標を用いた構造解析が行われている。日本蛋白質構造データバンク (PDBj) では、2017-01-04 時点で 12 万 5526 件の構造データが蓄積されている。

図 1-7. Protein Data Bank (PDB) トップページ

(<http://www.rcsb.org/pdb/home/home.do>)

[PDBFINDER2]

上記の PDB にはタンパク質結晶構造から得られた三次元座標を中心に蓄積されている。一方で、PDBFINDER2 (Hooft, *et al.*, *CABIOS*, 1996) は、三次元座標の羅列を研究者が利用しやすく変換するために、種々の解析を用いて一元に情報をまとめているデータベースである。基本的には PDB、DSSP、HSSP が情報源となっている。代表的にはタンパク質のアミノ酸配列に沿って、タンパク質の二次構造情報や、タンパク質の表面に位置しているかの指標である露出度が記載されている。

LAY-OUT Concept Explanation FTP-PDBFINDER FTP-PDBFINDER2 Download	<div style="text-align: right;"> The PDBFINDER(s) </div> <h3>Concept</h3> <p>The PDB is a very information rich protein structure database. Unfortunately, the PDB people are not very good at making their data available for search engines. There are several reasons why search engines often fail on the PDB:</p> <ul style="list-style-type: none">• The PDB has zillions of small administrative errors• The PDB-format is search-engine unfriendly• Many PDB files are incomplete <p>The PDBFINDER project is a possible solution to these problems. The PDBFINDER holds for each PDB file a structured, search-engine-friendly-formatted entry that holds the data-items most likely needed for people search for certain types of PDB entries. The PDBFINDER is not useful to search in atomic coordinates; it is meant to aid searches in the administrative records of PDB files.</p>
Miscellaneous	

☒ 1-8. PDBFINDER2 トップページ

(<http://swift.cmbi.ru.nl/gv/pdbfinder/>)

1-4. 糖鎖研究の現状と問題点および本研究の目的

タンパク質の Ser や Thr に結合する O 型糖鎖修飾は、これまで細胞外マトリクスの主成分として知られてきたが、近年、豊富な糖種がさらに多様な働きを持ち、特異的な疾患にも影響を及ぼしていることが明らかにされつつある。そのため、修飾された糖種を知ることにより、タンパク質の機能や細胞内局在が予測できるばかりでなく、重要な疾患との関連を推測することが可能となり、疾患の治療や創薬への応用も期待できる。しかしながら、従来の糖鎖修飾予測法 (Caragea, *et al.*, *BMC Bioinform.*, 2007; Hamby and Hirst, *BMC Bioinform.*, 2008; Sasaki, *et al.*, *IP SJ Trans. Bioinform.*, 2009; Li, *et al.*, *Comput. Biol. Chem.*, 2006; Julenius, *et al.*, *Glycobiology*, 2004; Gupta and Brunak, *Pac. Symp. Biocomput.*, 2002; Blom, *et al.*, *Proteomics*, 2004; Malik and Ahmad, *BMC Struct. Biol.*, 2007) は、糖鎖修飾を受ける位置を予測するものであり、糖種の判別は不可能であった。また、タンパク質配列データベース Uniprot KB/Swiss-Prot には 8 種類の糖種アノテーションが記載されているにも関わらず、従来法ではわずか 3 種類の糖種の修飾位置に関してアミノ酸配列の学習結果で糖鎖修飾の有無が判定されるのみであった (Caragea, *et al.*, *BMC Bioinform.*, 2007; Hamby and Hirst, *BMC Bioinform.*, 2008; Zhou, *et al.*, *Glycoconjugate J.*, 2012)。一方、糖転移酵素の分子機構に関する実験研究が行われるようになり、糖転移酵素の糖種特異性が指摘されてきた。しかし、依然としてバイオインフォマティクスによる糖種の予測や判別は対応できておらず、糖種判別法の開発は急務である。

タンパク質糖鎖修飾における糖種判別には、糖タンパク質の配列データおよび立体構造データを網羅的に調査し、アミノ酸配列・二次構造・立体構造の特徴を糖種ごとにとらえ、糖種判別法に活用することが早道と考えられる。しかし、タンパク質の立体構造解析では一般的に、タンパク質の X 線結晶回折における結晶化実験の段階で、タンパク質よりもゆらぎが大きい糖の存在がタンパク質の結晶化に不具合とされるため、酵素による糖鎖のシェービングや、ポイントミューテーションによる糖鎖修飾モチーフ残基の変異が行われる。タンパ

ク質立体構造情報データベース PDB にも、糖が含まれているエント리는存在するものの、生体内では修飾されているはずの糖鎖が欠けているケースがほとんどであり、糖を完全に残したタンパク質の二次構造や立体構造の情報を得ることが困難であるという問題がある。

さらに、各糖種の修飾に関わる糖転移酵素と糖タンパク質との相互作用は、細胞内でどのようなシステムによって管理されているかという点も、大きな課題として残されている。タンパク質の機能発現には、厳密なプロセスに基づいて、多数の糖鎖修飾がミス無く行われなければならない。これらの糖鎖修飾の ON・OFF を管理する仕組みのひとつとして、細胞内局在化経路が考えられる。リボソームで合成されたタンパク質が、細胞内を輸送される過程で、どのような糖転移酵素から、どのような糖の修飾を受けるのか。糖種のパターンを調べることによって、その流れが視覚化される可能性がある。しかしやはり、タンパク質に修飾された複数の糖鎖が織りなすパターンや、細胞内局在経路との関係に焦点を当てたバイオインフォマティクス研究はほとんど行われていない。興味深いことに、実験的手法による解析では、GlcNAc が脳組織に多く、核や細胞質に存在する糖タンパク質に多く含まれること (Alfaro, *et al.*, *Proc. Natl. Acad. Sci.*, 2012; Okuyama and Marshall, *J. Neurochem.*, 2003) や、Man 修飾が脳や筋肉組織に特異的に見られること (Live, *et al.*, *ChemBioChem*, 2013) 等が報告されている。したがって、細胞内の糖鎖修飾システムを考える上で、細胞内局在性や組織特異性に応じて糖鎖修飾における糖種分布のパターンを分類し、解析することが必要不可欠であろう。

糖鎖研究の現状と問題点をふまえ、本研究では O 型糖鎖修飾を受けるタンパク質のアミノ酸配列・二次構造・立体構造・細胞内局在経路解析を行い、糖種特異的な特徴を抽出するとともに、O 型糖鎖修飾の糖種判別に利用可能なパラメータを探索し、高精度糖種判別法を開発することを課題とした。

本研究を通して、O 型糖鎖修飾を受けるタンパク質のアミノ酸配列・二次構造・立体構造に基づいた連続的もしくは空間的なアミノ酸出現傾向解析、二次構造指向性解析により、糖転移酵素の認識と糖種特異性の具体的な要因を議論

することが可能となる。また、タンパク質の細胞内局在性という切り口で糖鎖修飾機構の解明を目指すバイオインフォマティクス研究は前例がなく、その新規性は高い。タンパク質の細胞内局在化と糖種の相関解析により、糖鎖修飾のエラーによる疾患の具体的な要因を議論することや、修飾を受けた糖種をパラメータとしたタンパク質の細胞内局在予測法の開発にも応用できる。さらに、本研究で得られる修飾機構の知見に基づいて、適切な種類の糖鎖をタンパク質へ効果的に導入することができれば、タンパク質の機能制御が可能となり、人工粘膜やがんワクチン、動物生体を用いない人工抗体など、医療・創薬分野への応用が見込まれる。

本研究は、糖鎖修飾の分子機構を明らかにする『基礎研究』から、その知見を応用して糖種判別法の開発を行うまでの『応用研究』に展開していることが特徴的である。今後、分子生物学実験と組み合わせることにより、糖転移酵素による標的タンパク質の認識機構や糖鎖修飾機構の詳細な解明や高機能性糖タンパク質の開発など、さらに幅広い応用展開が可能である。

本研究では、具体的に以下の項目を実施した。本研究の全体像を図 1-9 に示す。本論文では、実施内容の詳細を、第 2 章以降の各章にて述べている。

第 2 章 一次配列に基づく糖種判別

- 1) 糖種による修飾残基付近のアミノ酸出現傾向の違いを明らかにする。
- 2) 糖鎖修飾位置周辺のアミノ酸出現頻度から得られる糖種特異的な位置特異的アミノ酸スコアマトリクス (PSSM) (Staden, *Comput. Appl. Biosci.*, 1989) を用いて糖種判別を行う。
- 3) 一次配列では判別が難しい糖種を明らかにする。

第 3 章 糖タンパク質の二次構造解析

- 1) 糖鎖修飾を受けるアミノ酸周辺のタンパク質の二次構造を糖種ごとに調査し、糖転移酵素の糖種特異的な二次構造指向性を把握する。

- 2) 利用可能な糖タンパク質の立体構造数を糖種ごとに調査し、二次構造や立体構造を糖種判別に応用できる可能性を示す。

第4章 糖鎖修飾残基周辺の空間的アミノ酸出現傾向解析

- 1) 糖鎖修飾残基あるいは糖構成原子周辺の空間的アミノ酸出現傾向の解析を行う。
- 2) 糖鎖修飾位置や糖構成原子から一定距離内に存在するアミノ酸残基の出現頻度や、物理化学的な特性に基づいた共通ルールを抽出する。

第5章 タンパク質細胞内局在経路に基づいた糖鎖修飾分布の解析

- 1) 糖鎖修飾を受けているタンパク質について、細胞内局在・糖鎖修飾情報をもとに、細胞内での局在化経路および糖種に応じて分類する。
- 2) タンパク質の細胞内局在化パターンごとに、修飾される糖種や組合せを調査し、タンパク質の局在化経路と糖種の相関を明らかにする。

第6章 タンパク質O型糖鎖修飾の糖種判別

- 1) アミノ酸配列・二次構造・立体構造・細胞内局在経路解析から得られた糖種特異的な特徴を活用し、タンパク質O型糖鎖修飾における高精度糖種判別法の開発を行う。

ポストゲノム時代の課題: タンパク質の機能を制御する糖鎖修飾
糖鎖修飾機構の理解および糖種判別法の開発

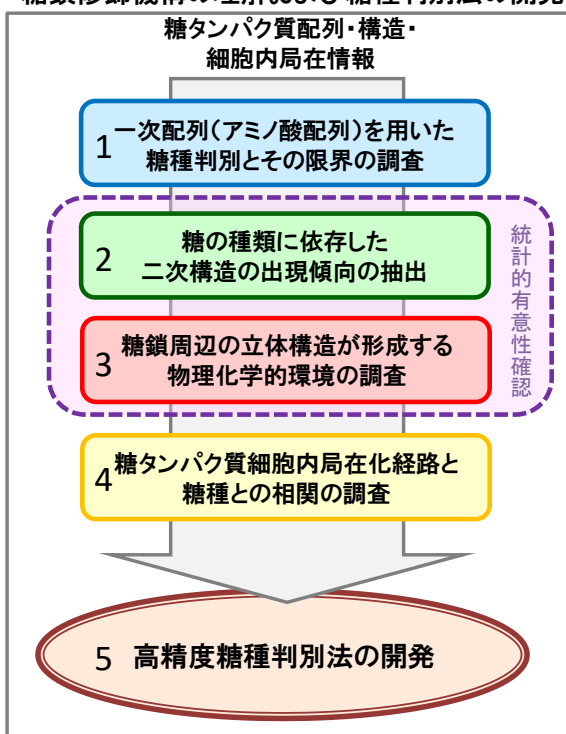


図 1-9. 本研究の全体像

第 2 章 一次配列に基づく糖種判別とその限界

(Etchuya, *et al.*, *Chem. Lett.*, 2013; Etchuya and Mukai, *J. Biomech. Sci. Eng.*, 2013)

2-1. 背景

O 型糖鎖修飾に関する情報は非常に少なく、Swiss-Prot にはモチーフ残基 (Ser または Thr) 上において O 型糖鎖を受けていないことがデータベースに明記されている O 型糖鎖 “非” 修飾のアノテーションは、わずか 4 サイトしかない。このように、O 型糖鎖修飾・非修飾位置を高精度で判別する予測法の開発は十分なデータ数が得られないことから困難を極める。GalNAc 修飾位置予測のための Oglyc (Li, *et al.*, *Comput. Biol. Chem.*, 2006) と NetOGlyc (Steentoft, *et al.*, *EMBO J.*, 2013)、GlcNAc 修飾予測のための Yin-O-Yang (Gupta and Brunak, *Pac. Symp. Biocomput.*, 2002) など、いくつかの O 型糖鎖修飾位置予測法が Web 上で利用可能であるが、ツールの予測精度はそれほど高くなく、また判別可能な糖の種類が 3 つ程度しかない。

そこで、本研究では O 型糖鎖修飾を受けるアミノ酸配列から位置特異的アミノ酸出現傾向解析を用いて糖種ごとの特徴を抽出し、抽出した特徴を用いて位置特異的アミノ酸スコアマトリクス (PSSM) (Staden, *Comput. Appl. Biosci.*, 1989) を作成し、O 型糖鎖修飾残基を対象とした新たに糖種を判別する方法を提案するとともに、一次配列のみを情報源とした判別の限界について議論する。

2-2. データセット作成および解析方法

Swiss-Prot 2012_06 から O 型糖鎖修飾を受けた哺乳類タンパク質 254 エントリを対象に、O 型糖鎖修飾のアノテーション情報に基づいて 757 サイトのデータを抽出した。抽出したサイトに “potential” または “probable” のアノテーションが記載されていた場合、データセットから削除した。それぞれのデータセットは O 型糖鎖修飾残基を中心に前後 10 残基ずつ、計 21 残基のアミノ酸配列として抽出した。O 型糖鎖修飾のアノテーション情報を基に、糖種によって 8 グループ (GlcNAc、Glc、GalNAc、Fuc、Hex、HexNAc、Man、Xyl) に分類した。Hex とは 6 単糖の総称であるが、Glc と Gal のどちらかという点が明らかにされていない。また HexNAc についても、GlcNAc と GalNAc のどちらかという点が明らかにされていない。CD-HIT (Huang, *et al.*, *Bioinformatics*, 2010) を用いて、各データセット内で 70%以上の配列類似性をもつ冗長配列を排除した。以下の組み合わせ (表 2-1) で、Positive data と Negative data とを判別するための PSSM を作成することとした。

表 2-1. PSSM 作成時のデータセット組み合わせ

		negative dataset
positive dataset	Fuc	Xyl + GalNAc + GlcNAc + Glc + Hex + HexNAc + Man
	Xyl	Fuc + GalNAc + GlcNAc + Glc + Hex + HexNAc + Man
	GalNAc	Fuc + Xyl + GlcNAc + Glc + Hex + HexNAc + Man
	GlcNAc	Fuc + Xyl + GalNAc + Glc + Hex + HexNAc + Man

図 2-1 は位置特異的アミノ酸出現傾向および PSSM を用いた糖の種類判別法の模式図である。糖種特異的な位置特異的アミノ酸出現傾向 (f_{jp})は、膜タンパク質の細胞内局在判別法 (Mukai, *et al.*, *Biosci. Biotech. Biochem.*, 2011) および GPI アンカー型タンパク質の判別法 (Mukai, *et al.*, *Biosci. Biotech. Biochem.*, 2013) 等の一次配列を用いた判別解析においてよく用いられている以下の式によって算出した。

$$f_{jp} = \frac{n_{jp}}{\sum_{j=1}^{20} n_{jp}} \quad (2.1)$$

p は O 型糖鎖修飾位置を基準とした距離を示し、 n_{jp} は p 位置の時のアミノ酸 j

の出現率を示している。しかし、PSSMを算出する際に0になることを避けるために、pseudo-count (Claverie and Audic, *Comput. Appl. Biosci.*, 1996)を導入した。

$$f_{jp} = \frac{n_{jp} + \frac{\varepsilon}{20}}{\sum_{j=1}^{20} n_{jp} + 20} \quad (2.2)$$

ε はpseudo-count (= 1)を示している。位置特異的スコア s_{jp} は全ての位置ごとに positive dataset のアミノ酸出現率 (f_{jp}^{positive})を negative dataset のアミノ酸出現率 (f_{jp}^{negative})で割ることによって求めた。

$$s_{jp} = \ln\left(\frac{f_{jp}^{\text{Positive}}}{f_{jp}^{\text{Negative}}}\right) \quad (2.3)$$

f_{jp}^{positive} は positive dataset のアミノ酸出現率を示しており、 f_{jp}^{negative} は negative dataset のアミノ酸出現率を示している。判別スコア (S)は M 残基目 ($-10 \leq M \leq 9$)から N 残基目 ($-9 \leq N \leq 10, M < N$)までの計算領域における位置特異的スコア s_{jp} の平均である。

$$S = \frac{1}{L} \sum_{p=M}^N s_{jp}, \quad (L = N - M + 1, M < N) \quad (2.4)$$

当該判別方法の精度として、感度 (Sensitivity)、選択性 (Specificity)、成功率 (Success rate)を以下の計算式に沿って算出した。

$$\text{Sensitivity} = \frac{\text{Correctly predicted true}}{\text{True}} \times 100 \quad (2.5)$$

$$\text{Specificity} = \frac{\text{Correctly predicted true}}{\text{Correctly predicted true} + \text{Incorrectly predicted false}} \times 100 \quad (2.6)$$

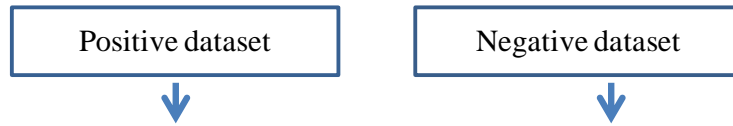
$$\text{Success rate} = \sqrt{\text{Sensitivity} \times \text{Specificity}} \times \frac{1}{100} \quad (2.7)$$

Positive dataset を negative dataset から判別する閾値を以下の条件で定義した。

(i) positive data と negative data を判別するための判別スコアの度数分布が重ならなかった場合、閾値は positive data のスコア最小値と negative data のスコア最大値の平均を示す。(ii) 判別スコアの度数分布に重なりが見られた場合、成功率が最大値を示す閾値を探し出し、判別の閾値として定義した。閾値の候補が多数見られた場合、最大値と最低値を示す閾値の平均を判別の閾値として定義した。

判別精度の算出には、膜タンパク質の細胞内局在判別法 (Mukai, *et al.*, *Biosci. Biotech. Biochem.*, 2011) および GPI アンカー型タンパク質の判別法 (Mukai, *et*

al., Biosci. Biotech. Biochem., 2013) において用いられている自己テストおよび5分割交差検定法を用いた。自己テストでは、学習データと試験データを同一にし、感度・選択性・成功率を求めた。5分割交差検定法では、5分割したデータセットのうち4つを学習データ、残りの1つを試験データとし、感度・選択性・成功率を求めた。この検定を20回繰り返し、感度・選択性・成功率の平均値を求めた。



Position-specific amino acid propensities

Residue	Alignment position							$f_{jp}^{Positive}$
	-10	-9	..	0	..	9	10	
A	5.62	5.62	9.24	8.84	8.03	6.83	2.81	$f_{jp}^{Negative}$
C	8.03	4.42	8.84	6.43	6.43	4.82	4.02	
D	0.40	0.80	1.20	1.61	1.20	0.40	0.80	
E	3.21	4.42	2.41	3.61	5.62	3.21	3.61	
.	5.62	8.43	3.61	5.62	6.43	2.81	3.21	
.	2.81	3.61	2.41	1.20	1.61	1.20	0.80	
.	4.82	6.02	6.02	2.41	2.41	4.82	3.61	
.	2.41	2.81	2.81	1.61	0.80	4.02	3.21	
Y	1.20	3.21	2.01	3.21	1.20	1.61	2.81	

Position-specific scoring matrix (PSSM)

$$S_{jp} = \ln(f_{jp}^{Positive} / f_{jp}^{Negative})$$

Residue	Alignment position						
	-10	-9	..	0	..	9	10
A	-0.17	0.14	0.8	-0.06	0.87	0.01	-0.5
C	-1.12	0.92	4.52	0.03	4.17	-0.26	-0.16
D	-0.32	0.06	-0.5	-0.6	-0.26	0.1	-0.27
E	0.35	-0.58	-0.05	-0.21	0.09	-0.26	-0.26
.	-1.12	0.26	0.26	-0.94	0.64	-1.51	-0.85
.	-0.74	-0.59	-0.5	-0.37	-0.23	-1.03	-0.14
.	-0.04	0.47	0.26	1.14	-0.44	0.22	0.64
Y	-0.03	0.11	-0.58	-0.03	0.07	-0.04	0.32

Calculating discrimination score (S)

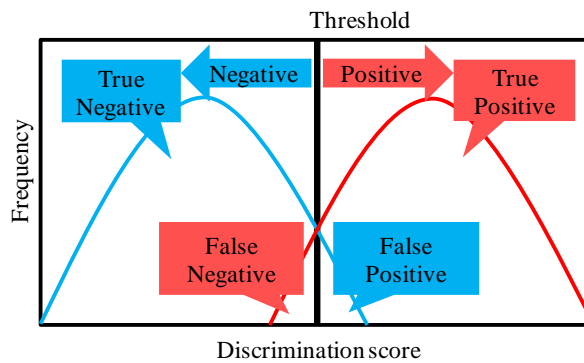


図 2-1. 位置特異的アミノ酸出現率の算出および PSSM を用いた判別のフローチャート

2-3. 結果および考察

Swiss-Prot 2012_06 から哺乳類タンパク質を対象に、O 型糖鎖修飾を受ける配列の冗長性を排除した数を示した (表 2-2)。哺乳類タンパク質中に 8 種類の糖が修飾されていることを確認した。データは Swiss-Prot のアノテーション情報を基にしているために、生体内の O 型糖鎖修飾の情報が直接反映されているとは言えない。なぜなら、データベースでは GalNAc および Xyl データが多数見つまっているが、それらの糖鎖が実験的プロセスの中に試料中で安定しているためと考えられる。それゆえ、データ数の少ない糖鎖は、実験中に不安定であるために剥離されてしまうことが推測される。CD-HIT を用いて、各糖種の代表配列を定義した。positive dataset と negative dataset の両方に跨いだ代表配列は positive dataset を優先とした。そのため、GalNAc および Xyl の negative dataset に差が生じている。

表 2-2. Swiss-Prot 2012_06 から抽出した
O 型糖鎖修飾糖種判別に用いたデータセット

Sugar Type	Positive data	Negative data								
		Total	GlcNAc	Glc	GalNAc	Fuc	Hex	HexNAc	Man	Xyl
GalNAc	249	158	57	4	0	12	5	8	1	71
GlcNAc	57	348	0	4	248	12	5	8	1	70
Fuc	12	391	57	4	245	0	5	8	1	71
Xyl	71	333	57	4	246	12	5	8	1	0

表 2-3 は自己テストおよび 5 分割交差検定法による最適な計算領域と糖の種類ごとの判別精度を示している。自己テストにおいて、GalNAc、Fuc、Xyl 修飾は高精度で negative dataset から判別することが可能であった。5 分割交差検定法を用いたテストにおいても高精度で判別することが可能なことを示し、特に Fuc および Xyl 修飾が極めて高い精度で判別できることが示された。

表 2-3. 自己テストおよび 5 分割交差検定法を用いた糖種の判別精度

糖種	計算領域	自己テスト			5 分割交差検定		
		感度 (%)	選択性 (%)	成功率	感度 (%)	選択性 (%)	成功率
GalNAc	-2 to 4	97.6	83.8	0.904	91.0	77.2	0.838
GlcNAc	-5 to 0	87.7	53.2	0.683	37.7	34.7	0.355
Fuc	-10 to 5	100	100	1.00	100	100	1.00
Xyl	-6 to 7	97.2	98.6	0.979	82.2	93.9	0.875

Xyl および Fuc データセットは他の O 型糖鎖修飾から高精度で判別できることを示した。Xyl 修飾を受けるアミノ酸配列は、Xyl だけがペントースであるという特徴が見られた。Fuc 修飾を受けるアミノ酸配列は、顕著な Cys 残基の出現が確認された。GlcNAc および GalNAc データは、5 分割交差検定における選択性が 80%を下回っており、他の糖種から見分けることが困難であると示唆された。GalNAc データの判別は false positive が高いために選択性が低いことが示された。GalNAc 修飾は様々な機能を発揮するために、様々な構造パターンが存在し、アミノ酸配列から特徴を抽出することが困難であると示唆された。GlcNAc 修飾を受けたアミノ酸配列は、Swiss-Prot から抽出した GlcNAc データは実験的プロセスの過程で剥離されてしまう可能性が高く、データ数が不十分であると考えられた。

アミノ酸残基の物理化学的特性に基づき、O 型糖鎖修飾周辺を対象に位置特異的アミノ酸出現傾向は糖の種類ごとに算出された (図 2-2)。オレンジに着色した領域は、5 分割交差検定法における成功率が最も高かった計算領域を示した。

疎水性領域は O 型糖鎖修飾を受ける残基周辺でいずれの糖種においても見られた。親水性、荷電アミノ酸を含む極性アミノ酸残基に異なる出現傾向が見られた。Fuc および Xyl 判別は計算領域内に顕著な特徴が見られた。Fuc 修飾を受けるアミノ酸配列は多少の荷電アミノ酸残基の出現が確認されたが、高い親水性アミノ酸残基の出現も見られた。Xyl 修飾を受けるアミノ酸配列は計算領域内で負電荷アミノ酸の高い出現傾向と、多少の正電荷アミノ酸残基の出現傾向が見られた。GlcNAc および GalNAc 修飾は、計算領域内に多数の類似点が見られる点から、判別精度が低いことが考えられた。

本研究では、タンパク質の一次配列に基づいた特徴を **Fuc** および **Xyl** のみ抽出することができ、高精度の判別法の開発が可能であったと示された。本研究の判別法と既存の *O* 型糖鎖修飾予測ツールを比較すると、既存の予測ツールでは予測することが不可能な **Fuc** および **Xyl** を新たに予測することが可能である。**GlcNAc** および **GalNAc** 修飾は、既存の予測ツールと大きな精度の向上などの違いが確認されなかった。*O* 型糖鎖修飾を受けるアミノ酸配列では、高い疎水性を示し、荷電アミノ酸の出現が確認され、それらの特徴が糖転移酵素の結合部位の形成に用いられている可能性を示した。**Fuc** および **Xyl** だけではなく多くの *O* 型糖鎖修飾に関わる糖種を判別するには、タンパク質の一次配列だけではなく *O* 型糖鎖修飾を受ける残基周辺の立体環境を考慮する必要があると考えられた。また今後、実験的に新たな *O* 型糖鎖修飾位置の同定が進み、学習データ数が増えることにより、当該判別法の精度の向上が期待できる。

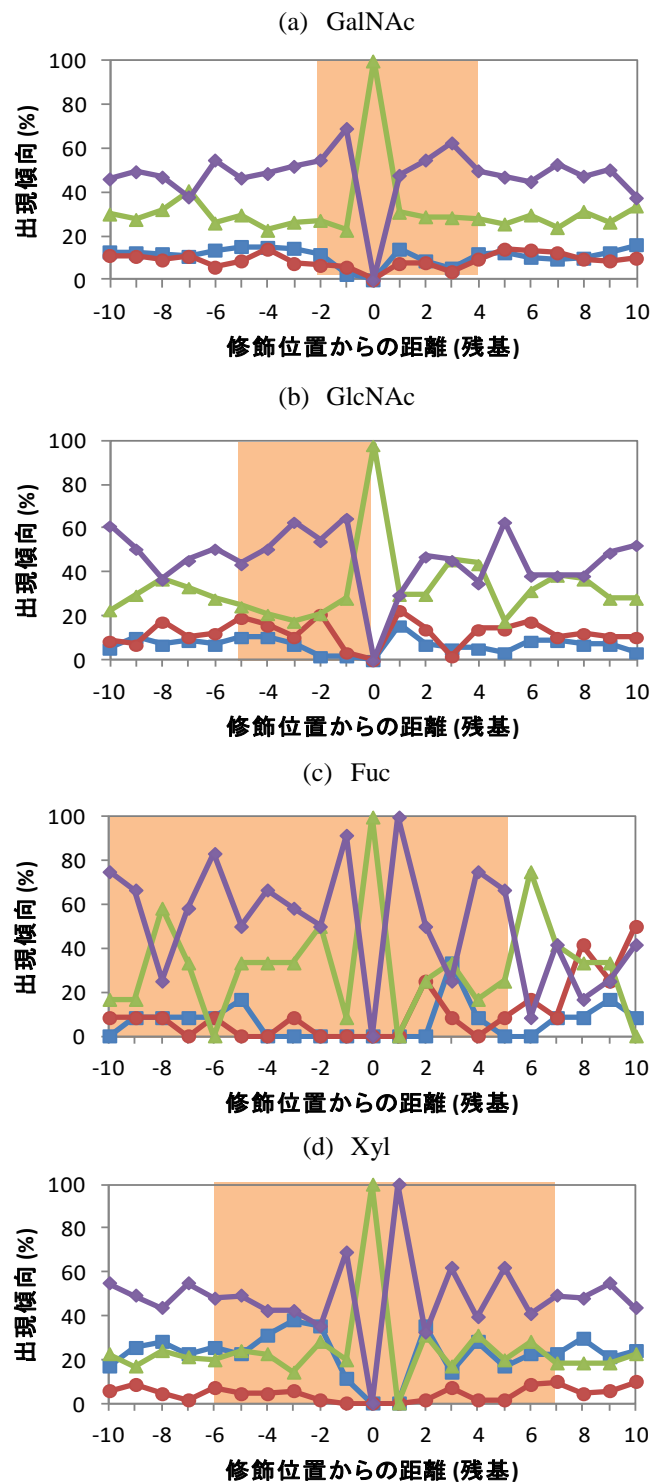


図 2-2. O 型糖鎖修飾を受ける残基周辺の位置特異的アミノ酸特性

赤 (●): 正電荷アミノ酸残基、青 (■): 負電荷アミノ酸残基、
 緑 (▲): 親水性アミノ酸残基、紫 (◆): 疎水性アミノ酸残基

第3章 糖タンパク質の二次構造解析

(Etchuya and Mukai, *J. Biomech. Sci. Eng.*, 2015)

3-1. 背景

本章では、タンパク質 *O* 型糖鎖修飾の糖種判別において、標的タンパク質の二次構造が判別の要因になり得る可能性を探索する。*O* 型糖鎖修飾周辺のアミノ酸配列にいくつかの構造上の特徴が見られ、GalNAc 修飾はコイルまたはターン構造中に修飾残基が含まれていることが報告された (Julenius, *et al.*, *Glycobiology*, 2004) ほか、GlcNAc 修飾においても主にストランド構造中に修飾残基が含まれていることが確認されている (Gupta and Brunak, *Pac. Symp. Biocomput.*, 2002)。本研究では新たに、Fuc および Man 修飾位置周辺に見られる二次構造上の特徴を見出し、*O* 型糖鎖修飾における糖類ごとの特徴について議論した。

また、その二次構造の糖種特異的特徴が、*O* 型の糖を修飾するための要因であるのか、あるいは、*O* 型糖鎖修飾が起こることにより標的タンパク質が特定の二次構造に変化する傾向にあるのか、という点については、慎重に考察する必要がある。そこで、複数の結晶構造が得られている同一タンパク質の構造情報ペア (図 3-1) を対象に、*O* 型糖鎖が修飾されている残基・同

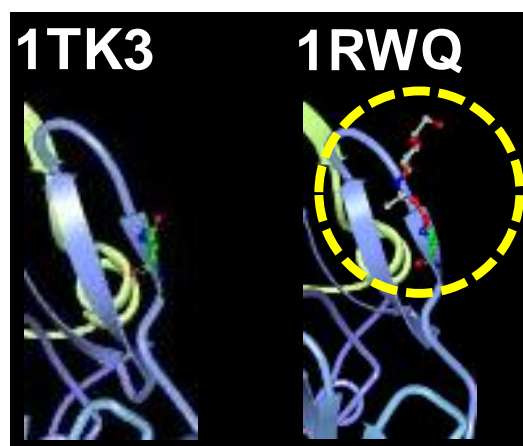


図 3-1. *O*-GlcNAc 修飾を含む立体構造・同じ残基位置にも関わらず *O*-GlcNAc が修飾されていない立体構造のペア (PDB ID: 1TK3/1RWQ)

じ位置にも関わらず *O* 型糖鎖修飾が行われていない残基のペアを見出し、当該残基周辺の二次構造を比較することによって、*O* 型糖鎖修飾の立体構造への影響を調査した。

3-2. データセット作成および解析方法

Swiss-Prot 2013_04 から哺乳類由来タンパク質を対象に、PDB の立体構造情報を抽出した。PDB 2012_08 (Berman, *et al.*, *Nucl. Acids Res.*, 2000; Berman, *et al.*, *Nucl. Acids Res.*, 2007) から立体構造情報中の”LINK”行に”MAN”、”NAG”、”NGA”、”FUC”が、”Ser” または ”Thr” 残基に結合している結晶構造を対象に、三次元座標データを抽出した。DSSP アルゴリズムにより二次構造を 8 種 (alpha-helix: H, 3¹⁰-helix: G, π -helix: I, turn: T, extended strand: E, beta-bridge: B, bend: S and coil: C) に定義して網羅的に格納したデータベース PDBFINDER2 (Hooft, *et al.*, *CABIOS*, 1996) を用いて、上記で抽出した PDB エントリの二次構造を参照し、糖鎖修飾残基を中心に前後 5 残基ずつ、計 11 残基分の二次構造情報を抽出した。

位置特異的二次構造データは、糖鎖修飾を受けた位置を基準にアラインメントした。位置特異的二次構造出現傾向 (f_{jp}) は、以下の式を用いて算出した。

$$f_{jp} = \frac{n_{jp}}{\sum_{j=1}^4 n_{jp}} \quad (3.1)$$

p は O 型糖鎖修飾を受けた位置を基準とした距離を示し、 n_{jp} は p の時の二次構造 j の出現傾向を示している。二次構造は 4 種類 (Helix (H, G, I), Strand (B, E), Turn (T) and Coil (C, S)) に分類した (Hooft, *et al.*, *CABIOS*, 1996; Kabsch and Sander, *Biopolymers*, 1983)。

次に、上記で抽出した PDB データのうち、糖鎖修飾を含む立体構造と、同じ残基位置にも関わらず糖鎖を受けていない立体構造のペアを検索し、糖鎖修飾を含む構造と含まな

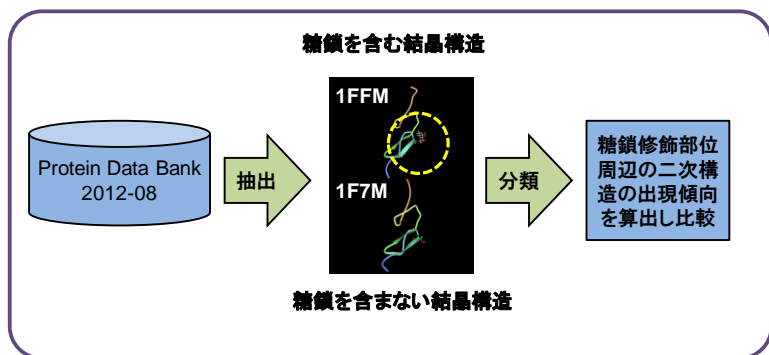


図 3-2. O 型糖鎖修飾の立体構造への影響解析

い構造の双方から、修飾残基を中心とした前後 11 残基分の二次構造情報を抽出し、糖種および糖鎖の有無別に、式 3.1 を用いて位置特異的二次構造出現傾向を算出した。

3-3. 結果および考察

[二次構造と糖鎖修飾の相関解析]

PDB 2012_08 より、タンパク質の結晶構造中に糖鎖修飾を受けている残基位置を抽出した。表 3-1 に、PDB 中の O 型糖鎖修飾残基の数を糖種ごとに示した。GalNAc や GlcNAc では、Swiss-Prot にそれぞれ 57 および 249 もの修飾サイトが含まれているにも関わらず (表 2-2)、PDB にはこれらがほとんど含まれていないことが示された。一方、Fuc や Man は、Swiss-Prot に報告されている件数に対して、立体構造データはより多く得られていることが示された。

表 3-1. Protein Data Bank 中の O 型糖鎖修飾残基の数

Sugar type	Number
GalNAc (NGA)	3
GlcNAc (NAG)	8
Fuc (FUC)	34
Man (MAN)	28

糖種ごとに、位置特異的二次構造出現傾向を算出した。GalNAc 修飾は全てコイル構造中に見つかったが (図 3-3c)、わずか 3 エントリのデータしか存在しないため、統計学的有意性を示すことは困難であった。GalNAc 修飾がコイルおよびターン構造中に確認された実験研究も報告されており、今後、立体構造解明による GalNAc 修飾情報の増加が望まれる。一方、GlcNAc 修飾は、ストランド・ターン・コイル構造中に見いだされたが、特に、ストランド構造中の修飾が多く確認された (図 3-3b)。この結果は、GlcNAc 修飾がストランド構造を形成する領域に起こりやすいことを示している。Fuc 修飾についても、ターンおよびコイル構造に比べ、ストランド構造で顕著に見いだされた (図 3-3c)。Swiss-Prot のアノテーションによると、Epidermal Growth Factor (EGF) ドメイン中に多くの Fuc が修飾されていることがわかった (Rana, *et al.*, *J. Biol. Chem.*, 2011; Rao, *et al.*, *Cell*, 1995; Schurph, *et al.*, *FASEB J.*, 2012; Tan, *et al.*, *J. Cell. Biol.*, 2002)。複数の

EGF ドメインにおけるストランド構造末端領域で、Fuc 修飾が確認された。ヘリックス構造中には、Fuc 修飾が全く見られなかった。Man は、ストランドおよびコイル構造中に修飾されていたが (図 3-3d)、ストランド構造への修飾の割合が最も高かった。

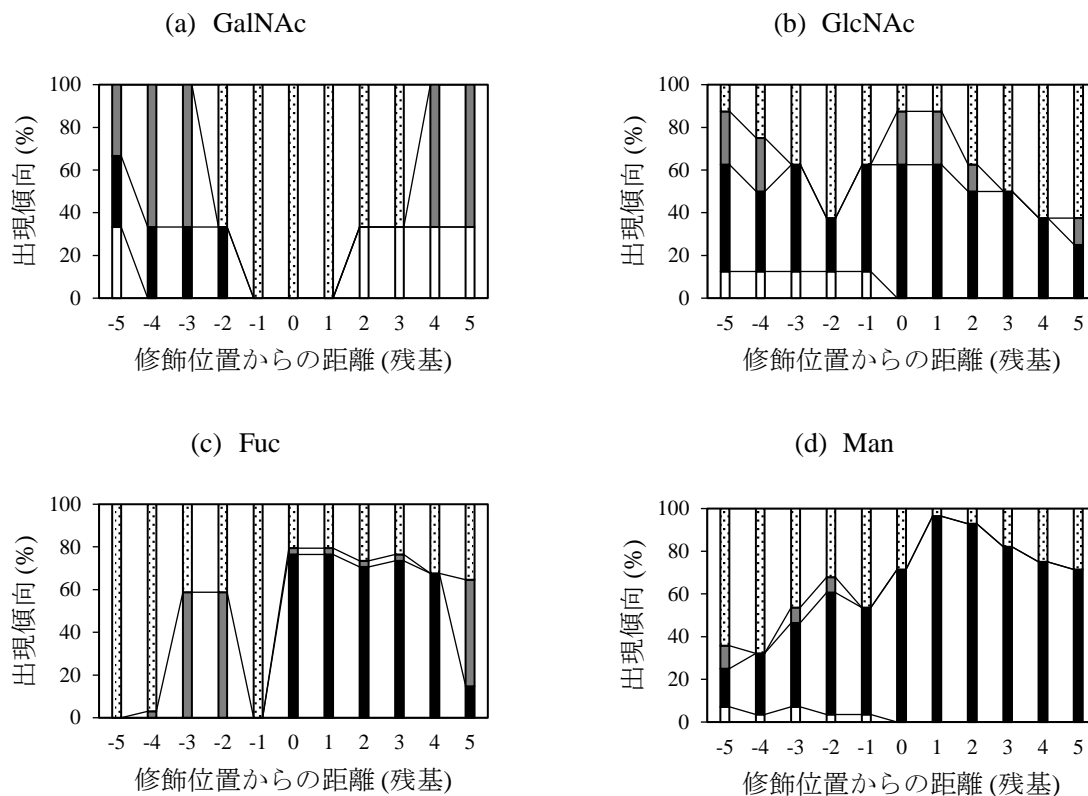


図 3-3. 糖鎖修飾を受けた残基周辺の二次構造出現傾向
 白: ヘリックス構造 (H, G, I), 黒: ストランド構造 (E, B),
 灰: ターン構造 (T), ドット: コイル構造 (S, C).

O 型糖鎖修飾のほとんどは、ストランドまたはコイル構造中に見られた。ストランドおよびコイル構造のみの出現傾向を図 3-4 に示した。GlcNAc および Man では、修飾残基周辺のストランド構造の出現傾向が高いことから、ストランド構造中に修飾される傾向にあることが示唆された。また、EGF ドメイン内のストランド構造の最も N 末端側に、Fuc が修飾されやすいことが明らかにさ

れ、Fuc 転移酵素がストランド構造の境界を認識している可能性が示唆された (図 3-5)。また、GlcNAc および Man 修飾位置周辺で観察されたコイル構造の近傍には、ストランド構造が見られた。この結果から、GlcNAc および Man の転移酵素は、ストランド構造とそれに連なるコイル構造中のモチーフ残基を認識している可能性が示唆された (図 3-6)。

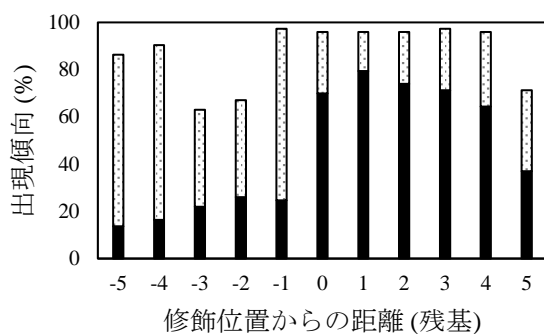


図 3-4. O 型糖鎖修飾を受ける残基周辺のストランドおよびコイル構造の出現傾向
 黒: ストランド構造 (E, B), ドット: コイル構造 (C, S)

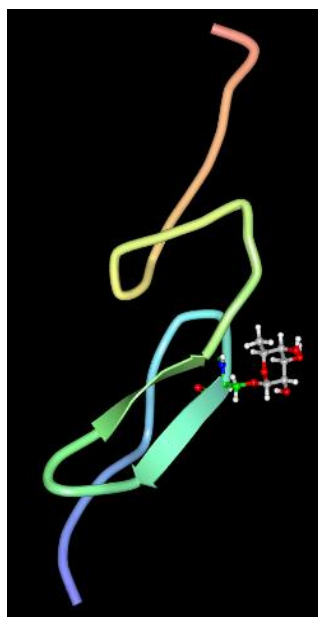


図 3-5. EGF ドメインに見られる代表的な Fuc 修飾 (PDBID: 1FFM)
 ストランド構造中の最も N 末端側である Ser-60 に Fuc 修飾

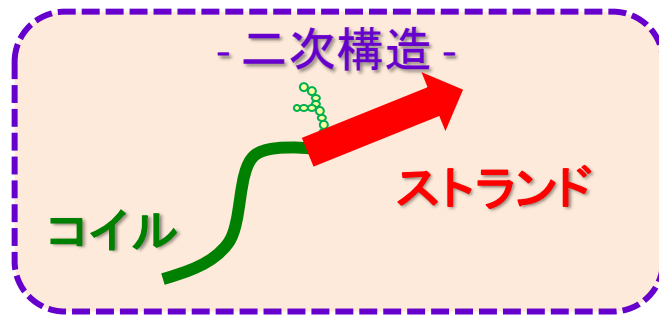


図 3-6. O 型糖鎖修飾を受けやすい二次構造の模式図

[O 型糖鎖修飾の立体構造への影響]

結晶構造が複数得られているタンパク質を対象に、PDB 2012_08 から同一残基に糖鎖修飾を受けている結晶構造・修飾を受けていない結晶構造のペアを抽出した（表 3-2）。糖鎖修飾を受けている結晶構造 1 つに対し、を受けていない結晶構造が複数対応していた。しかし、GalNAc および Man に関しては、該当の構造ペアが見出せなかった。

表 3-2. O 型糖鎖修飾の有無によるデータ数

Sugar type	Number
GlcNAc (NAG) 修飾あり	8
GlcNAc (NAG) 修飾なし	81
Fuc (FUC) 修飾あり	34
Fuc (FUC) 修飾なし	56

PDBFINDER2 (Hoof, *et al.*, CABIOS, 1996) から該当する領域の二次構造情報を抽出し、糖種ごとに位置特異的二次構造の出現傾向を算出した。

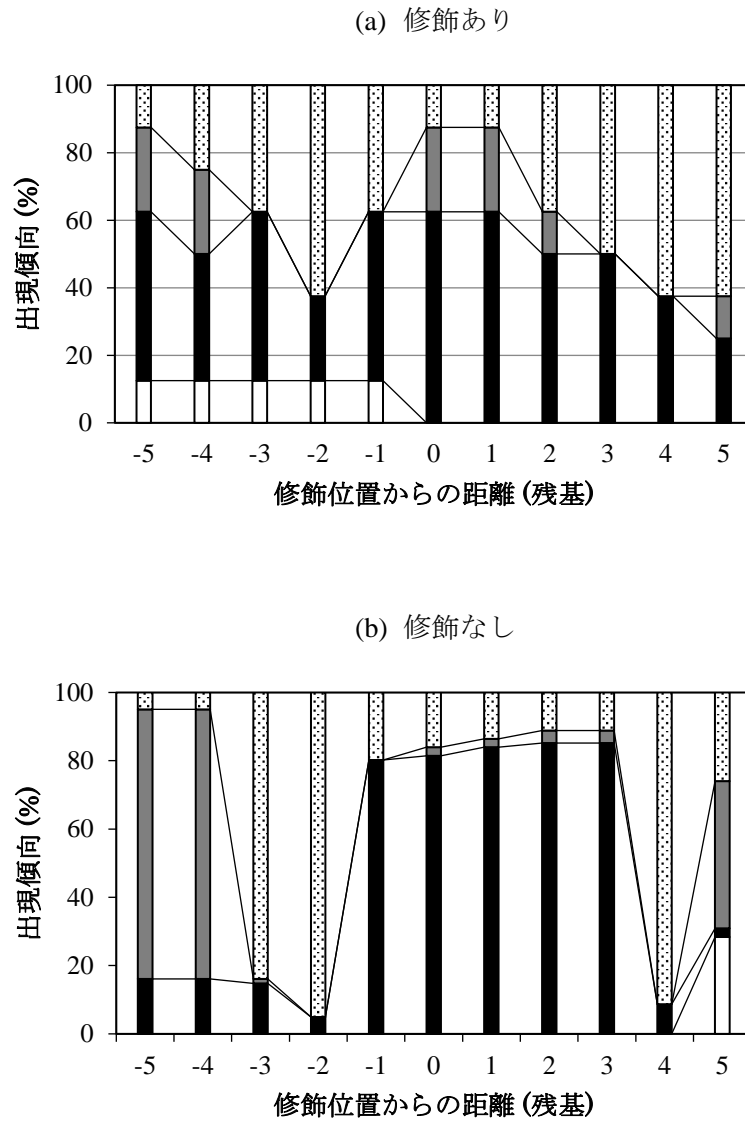


図 3-8. GlcNAc 修飾の有無による位置特異的二次構造出現傾向の変化

白: ヘリックス構造 (H, G, I), 黒: ストランド構造 (E, B),

灰: ターン構造 (T), ドット: コイル構造 (S, C).

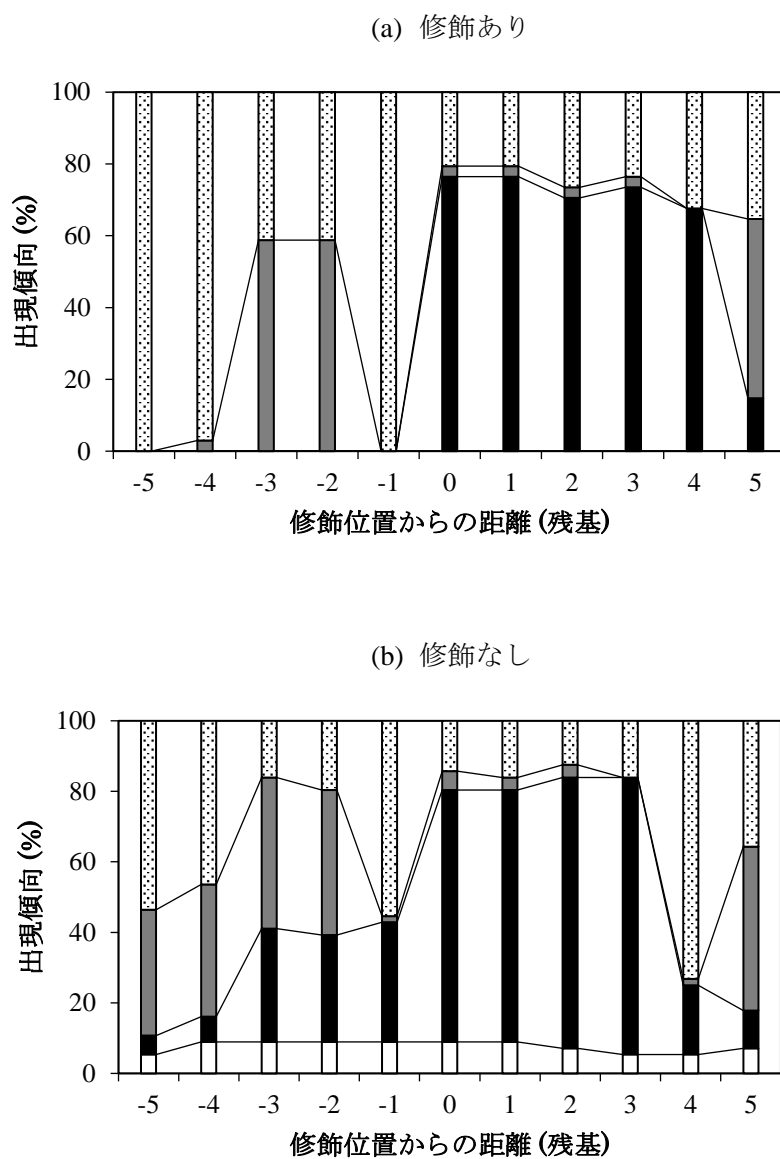


図 3-9. Fuc 修飾の有無による位置特異的二次構造出現傾向の変化

白: ヘリックス構造 (H, G, I), 黒: スtrand構造 (E, B),

灰: ターン構造 (T), ドット: コイル構造 (S, C).

GlcNAc 修飾 (図 3-8) において、非修飾では修飾残基より N 末端側領域内にコイル構造の出現が見られ、修飾残基以降ではストランド構造の出現が確認さ

れた。全体的にヘリックス構造の出現は極めて低い傾向にあった。修飾を受けた構造では、N 末端領域内にストランドおよびコイル構造に加え、ヘリックス構造の出現が見られた。修飾位置以降では、ストランドおよびコイル構造の出現が確認された。ストランド構造を中心にそれに連なるコイルまたはターン構造内に GlcNAc 修飾が行われていることが推測された。

Fuc 修飾 (図 3-9) において、非修飾では N 末端領域内にターンおよびストランド構造が出現しており、修飾残基以降ではストランドの出現が確認された。修飾を受けた構造では、N 末端領域にコイルおよびターン構造の出現が見られ、修飾残基以降ではストランド構造が出現していた。非修飾と修飾を受けた構造とを比較すると、ストランド構造を中心とした出現が見られるが、修飾を受けた構造では-1 の位置でストランド構造の出現がないことから、ストランド構造中の最も N 末端側の残基に Fuc 修飾が行われていることが示唆された。一方で非修飾では、-4 の位置まで緩やかにストランド構造の出現が見られることから、Fuc 修飾はストランド構造を中心に認識していることが示唆された。

これらの結果から、O 型糖鎖修飾を受けるタンパク質の立体構造は、修飾位置以降では立体構造に糖鎖修飾の有無による影響はほとんどないことが示唆された。しかしながら、修飾残基より N 末端側では、二次構造の出現傾向に差異が見られることから、修飾残基の N 末端側の立体構造には O 型糖鎖修飾が影響を与える可能性が考えられた。今後、O 型糖鎖修飾を含む結晶構造が増えることにより統計的評価が可能となり、糖種特異性の有無についても議論できるようになることが期待される。

[二次構造上の特徴の糖種判別への応用]

表 3-3 より、GlcNAc 修飾は Swiss-Prot 中の数に比べ、PDB 中に含まれる数が大幅に少ない。一方、Fuc 修飾では Swiss-Prot、PDB とともにほぼ同数のデータが存在した。GlcNAc 修飾は他の O 型糖鎖修飾より剥離されやすいという特徴が確

認められており、結晶構造を作製する際に GlcNAc 修飾が剥離されている可能性が考えられる。

表 3-3. データセット

Sugar type	Origin	Number
GlcNAc	Swiss-Prot 2013_04	59
	PDB 2012_08	8
Fuc	Swiss-Prot 2012_08	29
	PDB 2013_04	25

Swiss-Prot 由来の糖鎖修飾情報を基にアミノ酸配列を抽出し、PDB 結晶構造から二次構造情報を抽出した。GlcNAc 修飾（図 3-10）においては、Swiss-Prot の“DRPDB” 行に記載されている PDB ID の構造情報に基づくと、ヘリックス構造の出現傾向が高く、約半数の GlcNAc 修飾サイトはヘリックス構造を形成していることが示された。一方、PDB より抽出した結晶構造から GlcNAc 修飾残基周辺の二次構造の出現傾向を算出すると、ヘリックス構造の出現が極めて低いことが示された。Swiss-Prot 由来の二次構造出現傾向のうち、ヘリックス構造を除くストランド、ターン、コイル構造の出現傾向と PDB 由来の二次構造出現傾向が類似していた。それゆえ、GlcNAc 修飾のうち、ヘリックス構造に結合する GlcNAc と、ストランドやコイル構造に結合する GlcNAc とに分類できる可能性が考えられ、異なる標的をもつ複数の GlcNAc 転移酵素の存在が示唆された。

Fuc 修飾（図 3-11）においては、Swiss-Prot と PDB の双方で Fuc 修飾残基が一致している場合のみ、Swiss-Prot と PDB のそれぞれから二次構造情報を抽出した。Swiss-Prot と PDB とともに、Fuc 修飾残基周辺の二次構造の出現傾向に高い類似性が見られた。いずれにおいても、ストランド構造中の最も N 末端側の残基に Fuc 修飾が起りやすい傾向が見られた。一方、GlcNAc 修飾（図 3-10）においては、Swiss-Prot の情報では約半数がヘリックス構造に特異的に修飾されているが、PDB の立体構造情報に基づくと、ストランドまたはコイル構造に修飾されていた。ヘリックス構造に行われる GlcNAc 修飾は修飾・切断のサイクルが極

めて早いか、結合が不安定で結晶構造を作成する過程で切断・剥離してしまう可能性が推測された。ヘリックス構造中の特殊な GlcNAc 修飾を、学習データセットから除外することにより、GlcNAc 修飾の判別精度を向上することも可能であると考えられた。

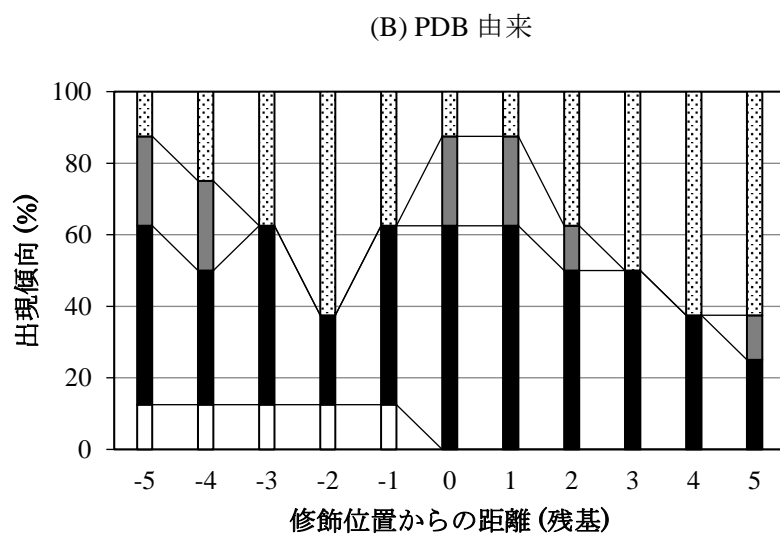
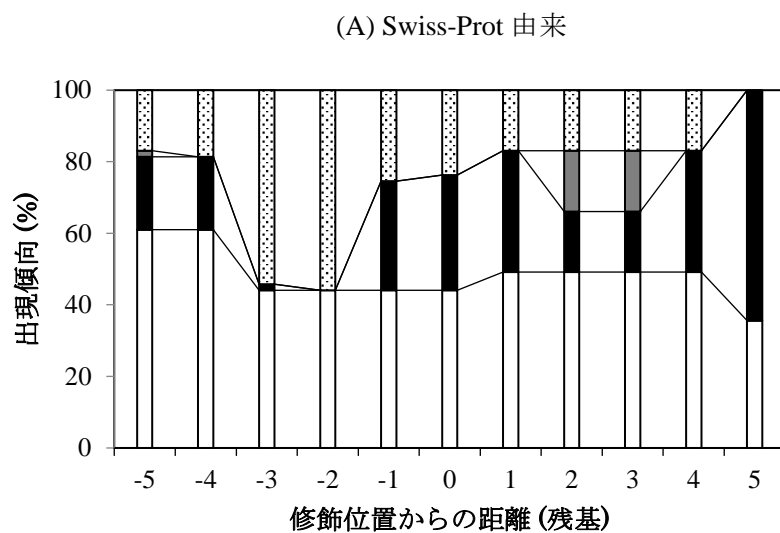


図 3-10. (A) Swiss-Prot および (B) PDB 由来の
 GlcNAc 修飾残基周辺の位置特異的二次構造出現傾向
 白: ヘリックス構造 (H, G, I), 黒: ストランド構造 (E, B),
 灰: ターン構造 (T), ドット: コイル構造 (S, C).

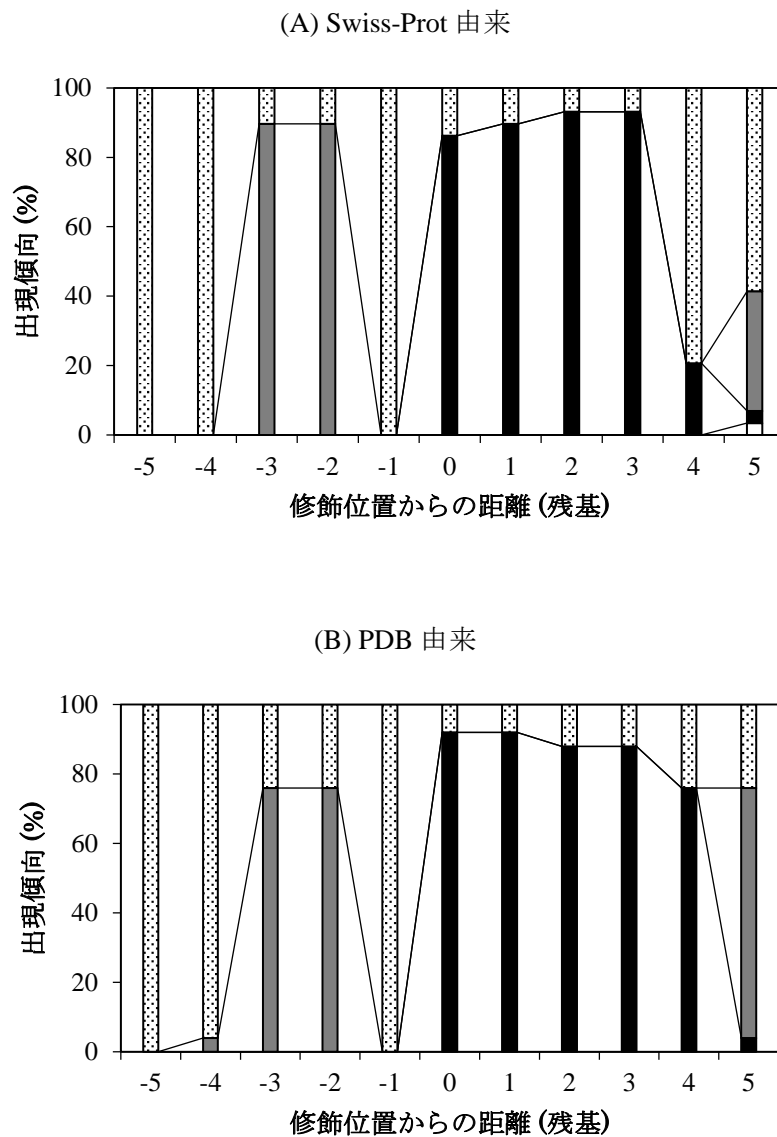


図 3-11. (A) Swiss-Prot および (B) PDB 由来の Fuc 修飾残基周辺の位置特異的二次構造出現傾向
 白: ヘリックス構造 (H, G, I), 黒: ストランド構造 (E, B),
 灰: ターン構造 (T), ドット: コイル構造 (S, C).

第 4 章 糖鎖修飾残基周辺の空間的アミノ酸出現傾向解析

(Etchuya and Mukai, unpublished paper 1)

4-1. 背景

Fuc と Xyl に関しては、タンパク質一次配列を用いて高精度で他の糖種から判別が可能であった。しかしながら、他の糖種に関しては、タンパク質の一次配列を用いて判別することは困難であることが示された (Etchuya, *et al.*, *Chem. Lett.*, 2013; Etchuya and Mukai, *J. Biomech. Sci. Eng.*, 2013)。タンパク質二次構造の出現傾向解析では、Fuc 修飾タンパク質において顕著な特徴が見られた (Etchuya and Mukai, *J. Biomech. Sci. Eng.*, 2015)。一方で、他の糖種に関しては糖種特異的な特徴は得られなかったものの、O 型糖鎖修飾全体での特徴が得られた。上記の特徴はタンパク質の機能部位における、糖鎖修飾の意味を明らかにするにあたり非常に有用であり、生物学的な意味付けの大きな助けになる。しかしながら、タンパク質結晶構造のデータ数の偏りから、統計的有意性を議論することは難しく、最適なパラメータの探索が必要である。それゆえ、糖鎖修飾の詳細なメカニズムを明らかにすると同時に、さらに多くの種類の糖を正確に判別するために、糖鎖修飾位置周辺の立体的な環境を考慮に加える必要があることが示唆された。糖鎖修飾の詳細なメカニズムを明らかにするにあたり、糖転移酵素と直接相互作用する糖鎖修飾部位周辺の空間的な環境を考慮に入れることは必要不可欠であるといえる。しかしながら、タンパク質結晶構造を対象とした統計的な糖鎖研究はほとんど行われておらず、タンパク質の結晶回折の実験が困難であることを理由とした結晶構造データ数の偏りが十分に想定できる。それゆえ本研究では、立体構造情報量から糖種特異性の統計的有意性を十分に考

慮しつつ、アミノ酸配列や二次構造では判別パラメータとしての特徴に乏しい **GlcNAc** に着目した。本章では **GlcNAc** 修飾を受けたタンパク質における修飾残基周辺の立体的なアミノ酸出現傾向を、**Fuc** 修飾残基周辺の傾向と比較することにより、糖種特異的な特徴の抽出を目指した。

4-2. データセット作成および解析方法

Swiss-Prot 2016_07 より、結晶構造が明らかにされている哺乳類由来のタンパク質を抽出した。さらに、当該タンパク質中のアミノ酸の三次元座標および糖鎖修飾情報を、PDB 2016_03 より抽出した。

表 4-2. データセット

	アミノ酸配列	二次構造	立体構造	
			糖鎖修飾残基周辺 (図 4-3A 参照)	糖構成原子周辺 (図 4-3B 参照)
Fuc	12	36	9	9
GlcNAc	57	31	14	14

糖鎖修飾残基周辺の立体的な環境を明らかにするために、糖鎖修飾残基および糖鎖分子と相互作用しうる距離に存在するアミノ酸残基の特徴抽出を目指した。本研究では、一定の半径を持つ単位球を想定し、単位球内にアミノ酸残基を構成する原子のうち、いずれかが存在した場合、対象となる糖鎖またはアミノ酸残基と相互作用しているとし、それらのアミノ酸出現傾向を算出した。修飾されたモチーフ残基は除外した。単位球の中心には、糖鎖修飾残基 (図 4-1A) と糖鎖分子 (図 4-1B) を採用した。

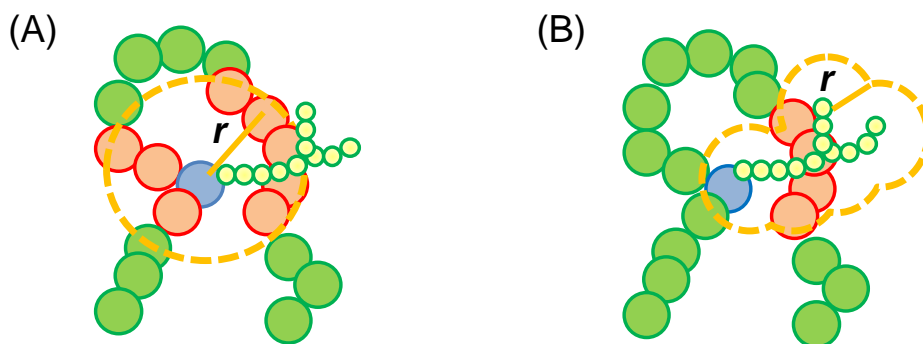


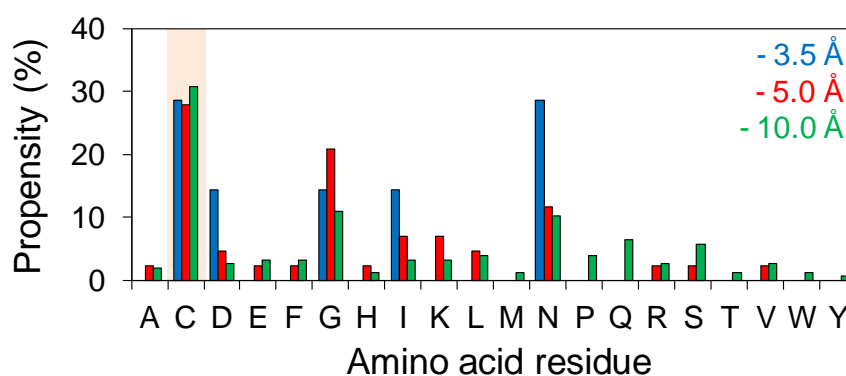
図 4-1. 立体空間内のアミノ酸出現傾向解析のための計算領域
 (A) 糖鎖修飾を受けたアミノ酸を中心とした場合
 (B) 糖を構成する原子を中心とした場合

4-3. 結果および考察

[糖鎖修飾位置周辺および糖を構成する原子周辺のアミノ酸出現傾向解析]

O 型糖鎖修飾を取り巻く空間において特徴的に出現するアミノ酸残基を観察するために、糖鎖修飾残基周辺の空間的位置特異的アミノ酸出現傾向を解析した。

(A) Fuc



(B) GlcNAc

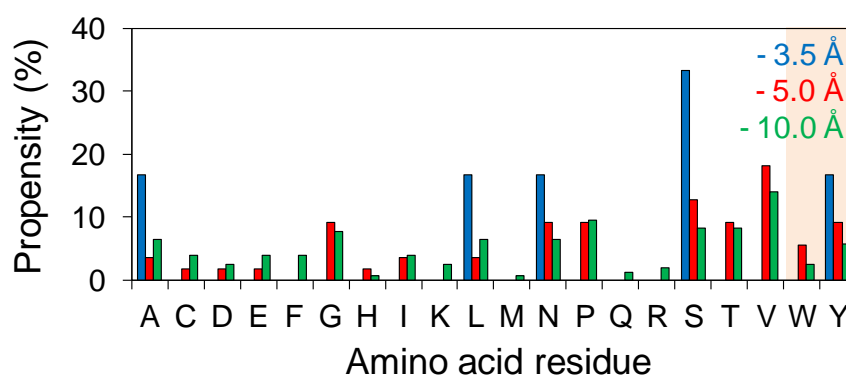


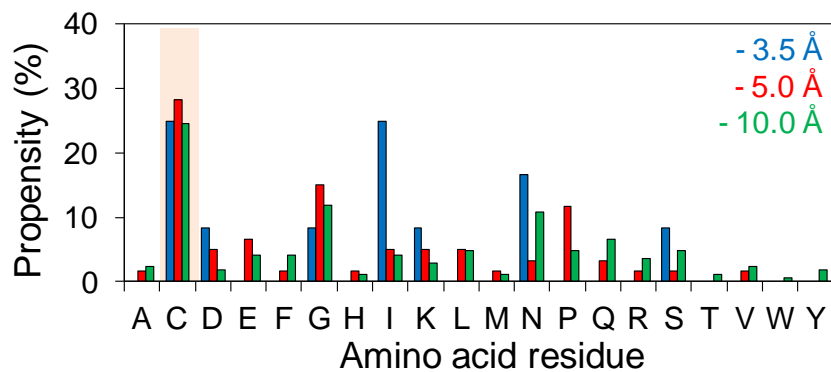
図 4-2. 糖鎖修飾残基周辺のアミノ酸出現傾向
(A) Fuc 修飾 (B) GlcNAc 修飾

図 4-2 より、糖種に応じて、異なるアミノ酸残基が糖鎖修飾残基周辺に近接していることがわかった。Fuc 修飾 (図 4-2A) では、Cys 残基が高い出現傾向を示していた。Cys 残基はタンパク質の構造形成でも初期に S-S 結合を形成するため

に、Fuc 転移酵素が修飾残基を認識する際に必要である可能性が考えられる。また、Asn 残基も極めて高い出現傾向を示していた。親水性を示す Asn 残基や負電荷を持つ Asp 残基が Fuc 修飾残基周辺で顕著に確認された。一方で、GlcNAc 修飾 (図 4-2B) では、Tyr 残基の出現傾向が高かった。また、Ser 残基や Asn 残基などの親水性アミノ酸残基の高い出現傾向が確認された。Ala 残基や Leu 残基などの疎水性残基が高い出現傾向を示していた。芳香族アミノ酸残基はアミノ酸残基の中でも極めて大きな体積を有するが、糖鎖認識タンパク質であるレクチンの糖鎖認識部位に見られ、糖の選択性に重要であることが報告されている (Sato, *et al.*, *Mol. Cell*, 2010)。GlcNAc 周辺の Tyr 残基も、目的の糖種や糖転移酵素の選択性に寄与している可能性が考えられる。また、Fuc 修飾でも GlcNAc 修飾でも親水性アミノ酸残基が見られた。

また、O 型の糖を構成する原子それぞれを中心に単位球を想定し、糖分子と相互作用しうる距離に存在するアミノ酸残基を算出し、空間的位置特異的アミノ酸出現傾向を以下の図 4-3 に示した。

(A) Fuc



(B) GlcNAc

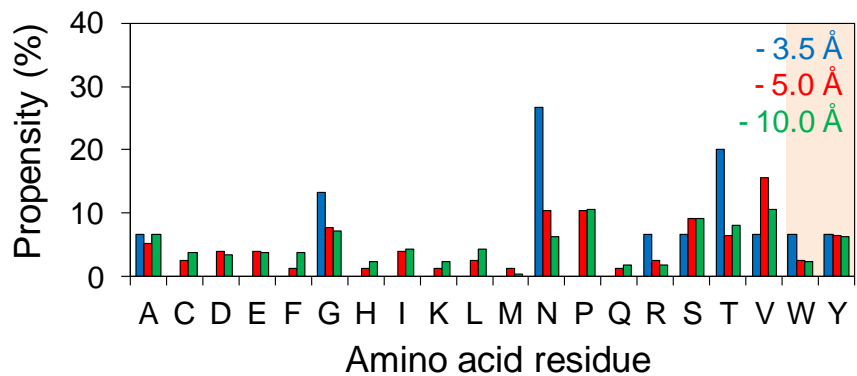


図 4-3. 糖構成原子周辺のアミノ酸出現傾向
(A) Fuc (B) GlcNAc

図 4-3 は糖分子と相互作用しうる距離内のアミノ酸残基の出現傾向を示している。Cys 残基は糖分子に近接した距離内に高い頻度で出現していた。Cys 残基は他の Cys 残基と S-S 結合を構成し、これらの構造が Fuc 修飾において、重要である可能性が示された。加えて、Ile 残基も高い出現傾向を示しており、構造異性体である Leu 残基はほとんど見られなかった。そのため、Ile 残基の構造が重要であることが示唆された。また、Asn 残基の出現傾向が高いことが確認された。Cys 残基や Ile 残基はストランド構造によく含まれることや、Asn 残基はターン構造中の出現傾向が高いことが報告されており、二次構造解析を支持する結果が得られた (Etchuya and Mukai, *J. Biomech. Sci. Eng.*, 2015)。Fuc 分子と比べ

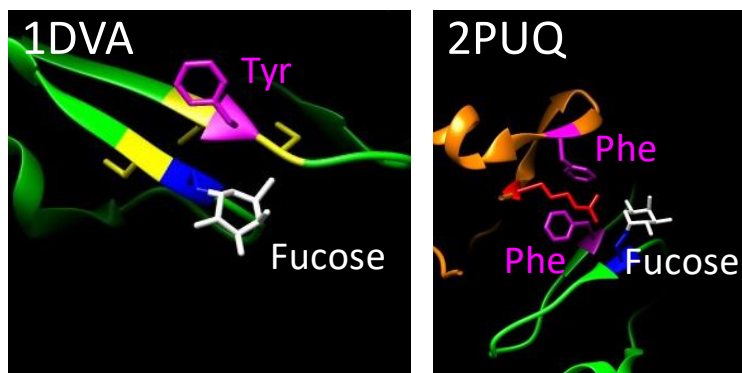
て、GlcNAc 分子の周辺では、Asn 残基や Thr 残基などの親水性アミノ酸残基や Gly 残基の出現傾向が高かった。Asn 残基や Gly 残基はターン構造中によく見られることから、二次構造解析と近い結果が得られた。また、Trp や Tyr 残基など芳香族アミノ酸残基の出現傾向が高かった。Trp や Tyr などの芳香族アミノ酸残基が GlcNAc 分子の周辺に存在していることは、GlcNAc 分子の保持や GlcNAc 分子以外の修飾を阻害するなど GlcNAc の選択性に本質的な要素であることが考えられた。

糖種に応じて、特徴的なアミノ酸残基がそれぞれの糖分子周辺に空間的に存在しており、一次配列では近接していないが、構造的に近接しているアミノ酸残基の糖種特異的な特徴を明らかにすることができた。

[糖鎖認識タンパク質の認識メカニズム解析]

上記の解析から、Fuc 修飾では周辺に Cys 残基や親水性アミノ酸残基の高い出現傾向が確認され、GlcNAc 修飾では親水性アミノ酸残基や Tyr や Trp などの芳香族アミノ酸残基の高い出現傾向が見られた。図 4-4 からも、GlcNAc 転移酵素や GlcNAc 分解酵素、Fuc 分子を介したタンパク質複合体の糖鎖を認識する領域には、たくさんの芳香族アミノ酸残基や電荷を持ったアミノ酸残基が空間的に近接していることがわかる。とりわけ、芳香族アミノ酸残基は O 型糖鎖認識・修飾メカニズムにおいて、共通して、本質的に重要な要素であることが考えられた。

(A) Fuc



(B) GlcNAc

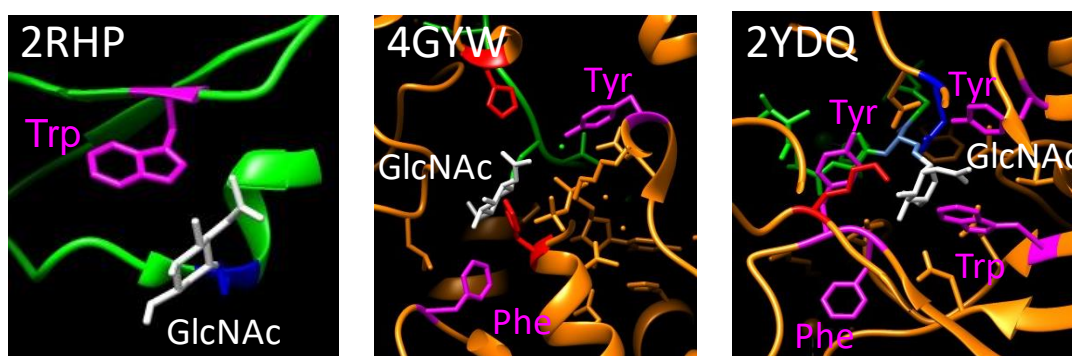


図 4-4. 糖鎖修飾残基 3.5 Å 以内の芳香族性残基と荷電残基
(A) Fuc 修飾タンパク質 (1DVA), Fuc 修飾タンパク質複合体(2PUQ)
(B) GlcNAc 修飾タンパク質 (2RHP), GlcNAc 転移酵素 (4GYW),
GlcNAc 分解酵素 (2YDQ),
紫: 芳香族性残基, 赤: 正電荷残基, 青: 糖修飾残基, 白: 糖

Fuc と GlcNAc 修飾を対象に、相互作用が見られる 3.5 Å 以内の領域では、Cys 残基や Tyr 残基の出現が確認された (図 4-4)。糖鎖を認識し結合するタンパク質であるレクチンのうち、Man 分子を特異的に認識するレクチンでは、Trp-Trp 残基からなる認識モチーフ (図 4-5) が Man 分子の選択性に重要であることが報告されている (Sato, *et al.*, *Mol. Cell*, 2010)。この結果に基づくと、糖鎖修飾残基に空間的に近接した芳香族アミノ酸残基は、目的の糖または糖転移酵素の選択性に必要であることが考えられる。とりわけ、GlcNAc 修飾は、一次配列を使っ

た判別では困難であったが、今後 PDB データ数の拡大に伴い GlcNAc 修飾情報が蓄積され、統計的有意性が担保されれば、上記で見られた芳香族アミノ酸残基の特徴を応用することによって、高精度に GlcNAc 修飾が判別できる可能性が示された。

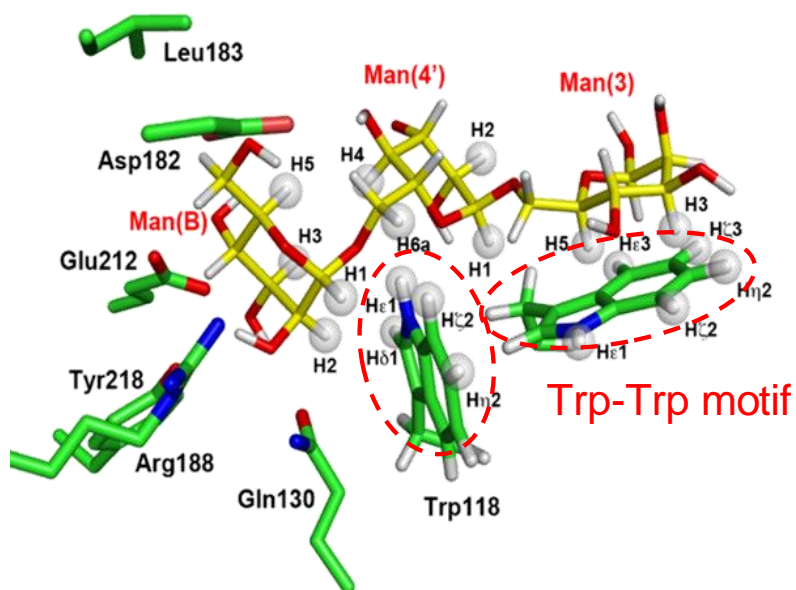


図 4-5. Man を認識するレクチンにおける糖認識 Trp-Trp モチーフ (Sato, *et al.*, *Mol. Cell*, 2010)

第 5 章 タンパク質細胞内局在経路に基づいた糖鎖修飾分布の解析

(Etchuya, *et al.*, unpublished paper 2)

5-1. 背景

真核細胞に含まれる多くのタンパク質は、翻訳後に小胞体やゴルジ体などでさまざまな修飾を受け、成熟タンパク質としての成長を遂げる。糖鎖修飾は主要な翻訳後修飾のひとつとして知られ (Hansen, *et al.*, *Biochem. J.*, 1995)、タンパク質の構造形成、機能発現、酵素活性などに関与している。O 型糖鎖修飾では、Ser と Thr がモチーフ残基として知られ、ゴルジ体でそれぞれの糖転移酵素によってさまざまな種類の糖が修飾される。ゴルジ体において O 型糖鎖修飾によって成熟したタンパク質の多くは、細胞外に分泌される。また、一部のタンパク質は糖鎖により細胞外マトリクスの基礎が構築され、細胞外マトリクスを形成する (Hollingsworth and Swanson, *Nat. Rev.*, 2004; Fukuda, *Biochim. Biophys. Acta*, 2002)。しかしながら、近年ではさらに複数種の糖鎖の発見

や、これまで糖鎖修飾が存在しないと考えられていた核や細胞質、ミトコンドリアなどの細胞内小器官でも、O 型糖鎖修飾が見いだされた (Cao, *et al.*, *PLOS ONE*, 2013; Kreppel, *et al.*, *J. Biol. Chem.*, 1997; Hart, *Annu. Rev. Biochem.*, 1997; Whelan and Hart, *Circ. Res.*,

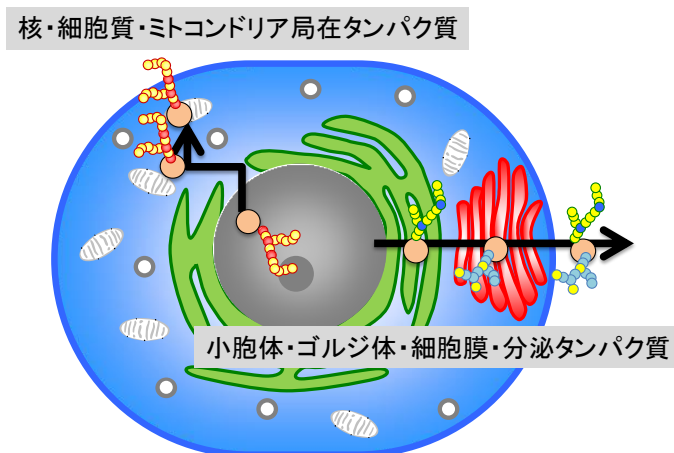


図 5-1. タンパク質の細胞内局在化経路と糖鎖修飾

2003; Comer and Hart, *J. Biol. Chem.*, 2000)。また、それぞれの糖種が、特異的な生化学反応や異なる役割を担っていることが明らかにされつつある(Lowe and Marth, *Annu. Rev. Biochem.*, 2003; Haltiwanger and Lowe, *Annu. Rev. Biochem.*, 2004)。このような背景から、糖種と生化学的な機能との相関を明らかにするための基盤が、除々に構築されてきたといえよう。

我々は、PSSMを用いてO型糖鎖修飾の糖種を判別する予測法の開発を行い、アミノ酸配列情報のみを用いて、FucとXylを他の糖種から高精度で判別することができた(Etchuya, *et al.*, *Chem. Lett.*, 2013; Etchuya and Mukai, *J. Biomech. Sci. Eng.*, 2013)。しかしながら、タンパク質の一次配列を用いた糖種判別では、とりわけGlcNAc修飾の判別が困難であることが示された。また、標的タンパク質の二次構造(第3章)(Etchuya and Mukai, *J. Biomech. Sci. Eng.*, 2015)および立体構造(第4章)(Etchuya and Mukai, unpublished paper 1)に関して、糖種判別の要素としての可能性を検討したが、糖種特異性を見出すことは可能であったものの、データセット数が問題となり統計的有意性が担保できず、糖種判別法への応用は難しいことが明らかになった。

そこで本章では、糖鎖修飾がタンパク質の成熟化の一部である点に着目し、タンパク質の細胞内輸送経路と糖種の相関を解析した。特に、タンパク質は合成された後、大まかには、小胞体を経由しない細胞内在性タンパク質と、小胞体を経由する分泌型タンパク質・シグナルアンカー型タンパク質・膜貫通型タンパク質の、2つの輸送経路に分かれる(図5-1)。そこで、この2つの細胞内輸送経路と糖種特異性、および、その特徴には統計的有意性が担保されるか否かという点について調査し、O型糖鎖修飾の糖種判別の要素として応用できる可能性を検討した。

5-2. データセット作成および解析方法

Swiss-Prot 2015_06 より哺乳類タンパク質を対象に“FT CARBOHYD”行に O 型糖鎖修飾の情報を持つタンパク質を抽出した。抽出したタンパク質群から“FT SIGNAL”と“FT TRANSMEM”行のシグナルの有無に関する情報に基づいて、細胞内在型タンパク質・分泌型タンパク質・シグナルアンカー型タンパク質・膜貫通型タンパク質の 4 種類に分類した (表 5-1)。それぞれのタンパク質に修飾されている糖鎖の出現傾向を解析し、タンパク質の細胞内輸送経路と糖鎖の修飾過程との相関を調査した。

表 5-1. O 型糖鎖修飾を受けたタンパク質の数

糖タンパク質のトポロジー	シグナルペプチドの有無	膜貫通領域の有無	エントリ数
細胞内在型	—	—	59
分泌型	+	—	105
シグナルアンカー型	—	+	18
膜貫通型	+	+	48

5-3. 結果および考察

糖鎖修飾されたタンパク質エントリを、タンパク質が有するシグナル領域を基に、細胞内在型タンパク質・分泌型タンパク質・シグナルアンカー型タンパク質・膜貫通型タンパク質の4種類に分類した。各グループのタンパク質群に修飾された糖種の出現傾向を解析した。細胞内在型タンパク質の多くは核や細胞質、ミトコンドリアに局在していた。これらのタンパク質はGlcNAc修飾のみを受けていることを明らかにした(図5-2)。シグナルペプチドやシグナルアンカーなどの小胞体移行シグナル配列を持たず、小胞体を経由しない局在化経路を持つタンパク質は、GlcNAc修飾を選択的に受けていた。核や細胞質で特異的に発見されたGlcNAc修飾は、リン酸化と共役しタンパク質の機能発現のコントロールをしていることが報告されている(Wang, *et al.*, *Sci. Signal.*, 2010) ほか、近年では、ミトコンドリアにもGlcNAc修飾が見られることが確認されている(Cao, *et al.*, *PLOS ONE*, 2010)。これらの実験報告も、小胞体を経由しない局在化経路を持つタンパク質が、GlcNAc修飾を受けていることを裏付けている。

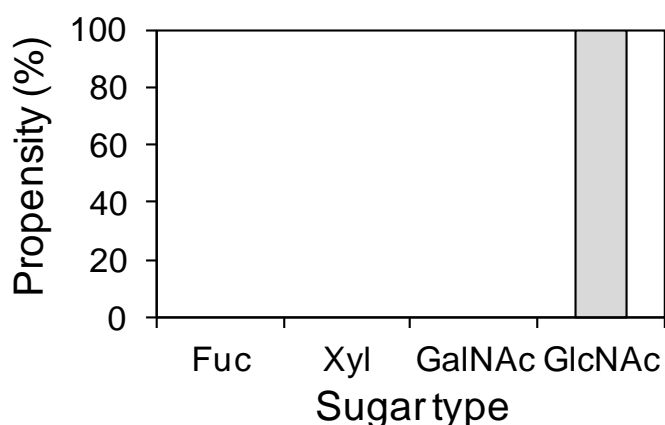


図 5-2. 細胞内在型糖タンパク質の糖種出現傾向

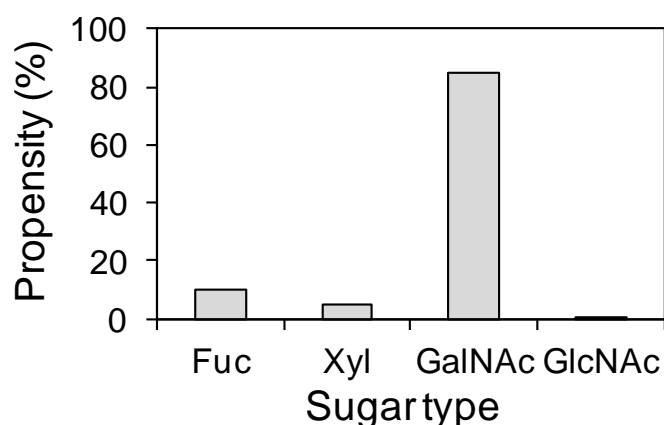


図 5-3. 分泌型糖タンパク質の糖種出現傾向

シグナルペプチドを持ち、膜貫通領域を持たない糖タンパク質は一般的には細胞外に分泌され、一部の糖タンパク質は細胞外マトリクスを形成する。これらのタンパク質は外部からのシグナルを受信する受容器としての働き(Kao, *et al.*, *Biochem.*, 1999; Stanley, *Curr. Opin. Struct. Biol.*, 2007) や、自身がシグナルタンパク質として他の細胞にシグナルを伝達する。とりわけ、Fuc 修飾は、シグナル伝達を開始する因子であることも知られている。図 5-3 によると、分泌型糖タンパク質では GalNAc 修飾の出現傾向が高く、Fuc 修飾および Xyl 修飾も確認された。分泌タンパク質のうち 2 種のタンパク質においては GlcNAc 修飾も見られたが、1 つはシグナルペプチドを持ち、エキソソームに局在化するタンパク質であった。エキソソームは細胞全体に見られるが、特に核内に集中して存在している。そのために、エキソソームに分泌された後に、核内で GlcNAc 修飾を受けることが予測される。もう 1 つの GlcNAc 修飾タンパク質については、Swiss-Prot 中の引用論文 (Schjoldager, *et al.*, *J. Biol. Chem.*, 2010) を確認したところ、GlcNAc が修飾されているとのアノテーションがある残基には実際には GalNAc が修飾されていることが報告されており、データベースへの記載ミスだと考えられた。

シグナルアンカー型タンパク質は、Xyl や GalNAc、GlcNAc 修飾を受けていた (図 5-4)。Fuc 修飾はシグナルアンカー型タンパク質には修飾されていなかった。GlcNAc 修飾タンパク質が 2 種見つかったものの、これらは核やミトコンドリア膜に局在化していることがわかり、細胞内局在化経路を考慮すると小胞体

を經由しないグループに所属する可能性が高い。

シグナルペプチドと膜貫通領域をもつ膜貫通型タンパク質では、Fuc や Xyl、GalNAc 修飾が確認された (図 5-5)。膜タンパク質のほとんどは GalNAc 修飾を受けており、細胞膜に局在化していた。また、図 5-3、5-4 との比較から、シグナルペプチドがなければ、Fuc 修飾を受けない要因であることが示唆された。

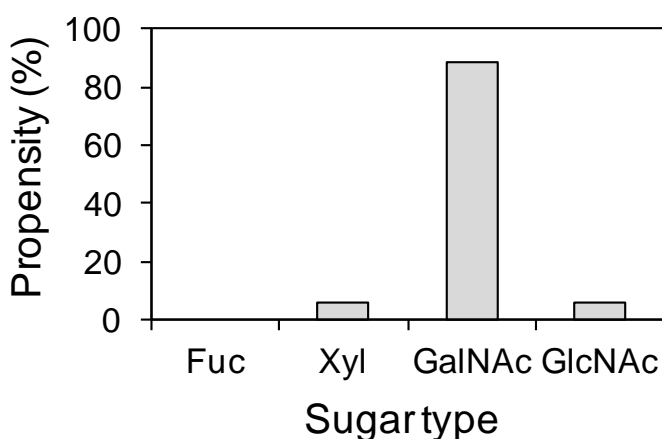


図 5-4. シグナルアンカー型糖タンパク質の糖種出現傾向

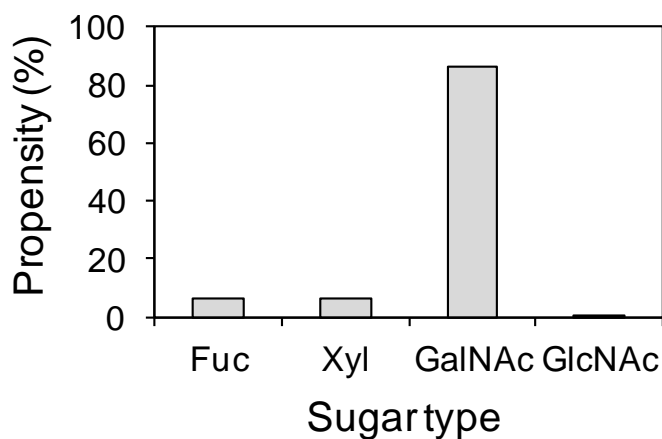


図 5-5. 膜貫通型糖タンパク質の糖種出現傾向

O 型糖鎖修飾はたいてい細胞外基質を構成し、細胞膜の保護やシグナル伝達のコントロールなどその生化学的な役割から GalNAc 修飾のみが着目され、研究されていた (Hatstrup and Gendler, *Annu. Rev. Physiol.*, 2008; Carraway and Hull, *Glycobiology*, 1991; Strous and Dekker, *Crit. Rev. Biochem. Mol. Biol.*, 1992)。本章で、

これらの分泌タンパク質には GalNAc ばかりでなく、Fuc や Xyl などの糖種多様性があることを示した。特に、シグナルアンカー型タンパク質では、GalNAc の他に Xyl のみが修飾されることを示した。また、シグナルペプチドを持つ分泌タンパク質・膜貫通タンパク質のみが Fuc 修飾を受けており、シグナルペプチドを認識する過程で Fuc 修飾が起きる可能性を示した。シグナルペプチドや膜貫通領域を持ち、小胞体やゴルジ体を経由する膜貫通型タンパク質においては、Fuc や Xyl、GalNAc が修飾されることを示した。小胞体に輸送されない細胞内存在タンパク質は GlcNAc 修飾のみを選択特異的に受けていることが明らかにされた。これらは、糖転移酵素の細胞内局在性（図 5-6）ともよく一致した結果となった。本章では、糖鎖修飾をタンパク質の成熟過程と捉える切り口で、細胞内での多様な糖種分布の存在と、細胞内局在経路の糖種特異性を示すことができた。細胞内局在経路は、O 型糖鎖修飾における GlcNAc 判別の有力なパラメータとなり得る。

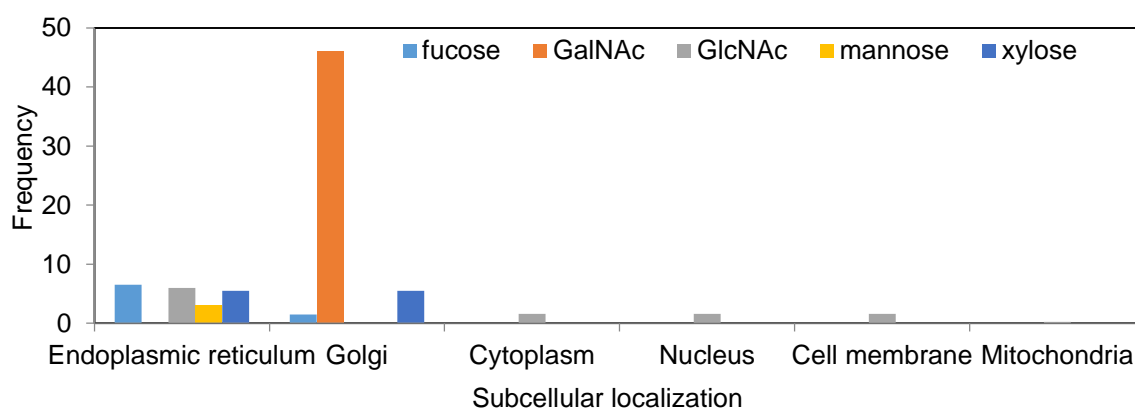


図 5-6. 各糖種の糖転移酵素の細胞内局在性

第6章 タンパク質 O 型糖鎖修飾の糖種判別

(Etchuya, *et al.*, unpublished paper 2)

6-1. 背景

第2章で、一次配列を用いた O 型糖鎖修飾の糖種判別法では Fuc や Xyl が高精度で判別することが可能であることがわかった。一方、一次配列を用いた判別法ではとりわけ GlcNAc の判別が困難であることが示された (Etchuya, *et al.*, *Chem. Lett.*, 2013; Etchuya and Mukai, *J. Biomech. Sci. Eng.*, 2013)。第3章から第4章では、二次構造 (Etchuya and Mukai, *J. Biomech. Sci. Eng.*, 2015)・立体空間 (Etchuya and Mukai, unpublished paper 1) におけるアミノ酸出現傾向の糖種特異的な特徴抽出には成功したものの、修飾される糖の情報が未だ不明瞭なタンパク質に糖種判別を適用するにあたり、立体構造が不明なケースがほとんどであることから、二次構造や立体構造に基づく情報を糖種判別法に応用することは困難であった。また、二次構造や立体構造から得られた糖種特異的特徴は、データ数の不足により統計的有意性が担保できていないという問題も含んでいた。第5章では、タンパク質の細胞内局在化経路に基づいた糖種分布を解析することで、核や細胞質、ミトコンドリアに局在化する小胞体に輸送されないタンパク質群は GlcNAc のみが修飾されることを明らかにした (Etchuya, *et al.*, unpublished paper 2)。データベースでの蓄積量やアノテーションの信頼性も高い、小胞体輸送シグナル (シグナルペプチド・シグナルアンカー) の有無により、GlcNAc の判別が可能であり、O 型糖鎖修飾の糖種判別法への応用が可能であることが示唆された。本章では、第2章で開発した PSSM を用いる糖種判別法に、輸送シグナルの有無による GlcNAc 判別を加えるために、GlcNAc 以外の糖種がそれぞれ判別可能であることを示しつつ、GlcNAc 判別が困難であることとした。

6-2. データセット作成および解析方法

Swiss-Prot 2016_07 から *O* 型糖鎖修飾を受けた哺乳類タンパク質 254 エントリを対象に、“FT CARBOHYD” 行に *O*-Fuc, *O*-Xyl, *O*-GalNAc, *O*-GlcNAc のいずれかのアノテーションが付与されており、かつ、“potential” または “probable” のアノテーションの記載のない *O* 型糖鎖修飾位置周辺のデータを抽出した。それぞれのデータについて、*O* 型糖鎖修飾残基を中心に前後 10 残基ずつ、計 21 残基のアミノ酸配列を抽出した。また、“FT SIGNAL” と “FT TRANSMEM” の有無の情報を、配列データに加えて抽出した。データセットは、第 2 章で用いたデータセットから最新のバージョンを使用し、糖種に基づいて 4 グループ (Fuc, Xyl, GalNAc, GlcNAc) に分類した (表 6-1)。CD-HIT (Huang, *et al.*, *Bioinformatics*, 2010) を用いて、各データセット内で 70% 以上の配列類似性をもつ冗長配列を排除した。以下、各グループ総当たりの組み合わせ (表 6-2) で、PSSM を作成することとした。

表 6-1. 糖種別データ数

	冗長	非冗長
Fuc	37	36
Xyl	21	18
GalNAc	525	304
GlcNAc	110	92

表 6-2. PSSM 作成時のデータセット組み合わせ

		positive dataset			
		Fuc	Xyl	GalNAc	GlcNAc
negative dataset	Fuc	—	○	○	○
	Xyl	—	—	○	○
	GalNAc	—	—	—	○
	GlcNAc	—	—	—	—

糖種特異的な位置特異的なアミノ酸出現傾向 (f_{jp}) の算出には以下の式を用いた。

$$f_{jp} = \frac{n_{jp}}{\sum_{j=1}^{20} n_{jp}} \quad (6.1)$$

p は O 型糖鎖修飾位置を基準とした距離を示し、 n_{jp} は p 位置の時のアミノ酸 j の出現率を示している。しかし、PSSM を算出する際に 0 になることを避けるために、pseudo-count (Claverie and Audic, *Comput. Appl. Biosci.*, 1996) を導入した。

$$f_{jp} = \frac{n_{jp} + \frac{\varepsilon}{20}}{\sum_{j=1}^{20} n_{jp} + 20} \quad (6.2)$$

ε は pseudo-count (= 1) を示している。位置特異的スコア s_{jp} は全ての位置ごとに positive dataset のアミノ酸出現率 (f_{jp}^{positive}) を negative dataset のアミノ酸出現率 (f_{jp}^{negative}) で割ることによって求めた。

$$s_{jp} = \ln\left(\frac{f_{jp}^{\text{Positive}}}{f_{jp}^{\text{Negative}}}\right) \quad (6.3)$$

f_{jp}^{positive} は positive dataset のアミノ酸出現率を示しており、 f_{jp}^{negative} は negative dataset のアミノ酸出現率を示している。判別スコア (S) は計算領域における位置特異的スコア s_{jp} の平均である。

$$S = \frac{1}{L} \sum_{p=M}^N s_{jp}, \quad (L = N - M + 1, M < N) \quad (6.4)$$

当該判別方法の精度として、感度 (Sensitivity)、選択性 (Specificity)、成功率 (Success rate) を以下の計算式に沿って算出した。

$$\text{Sensitivity} = \frac{\text{Correctly predicted true}}{\text{True}} \times 100 \quad (6.5)$$

$$\text{Specificity} = \frac{\text{Correctly predicted true}}{\text{Correctly predicted true} + \text{Incorrectly predicted false}} \times 100 \quad (6.6)$$

$$\text{Success rate} = \sqrt{\text{Sensitivity} \times \text{Specificity}} \times \frac{1}{100} \quad (6.7)$$

Positive dataset を negative dataset から判別する閾値を以下の条件で定義した。
 (i) positive data と negative data を判別するための判別スコアの度数分布が重ならなかった場合、閾値は positive data のスコア最小値と negative data のスコア最大値の平均を示す。
 (ii) 判別スコアの度数分布に重なりが見られた場合、成功率が最大値を示す閾値を探し出し、判別の閾値として定義した。閾値の候補が多数見られた場合、最大値と最低値を示す閾値の平均を判別の閾値として定義した。

判別精度の算出には、自己テストおよび 5 分割交差検定法を用いた。自己テストでは、学習データと試験データを同一にし、感度・選択性・成功率を求め

た。5分割交差検定法では、5分割したデータセットのうち4つを学習データ、残りの1つを試験データとし、感度・選択性・成功率を求めた。この検定を20回繰り返し、感度・選択性・成功率の平均値を求めた。

6-3. 結果および考察

本章では、前章までの検討において、O型糖鎖修飾の糖種判別のパラメータとして利用可能であると考えられた PSSM および細胞内局在経路を用いて、糖種判別を試みた。

まず、O型糖鎖修飾の糖種判別のパラメータとして、糖鎖修飾残基周辺のアミノ酸出現傾向から求められる PSSM のみを用いて、糖種判別を行った。第2章で行った糖種判別と異なる点は、Swiss-Prot のバージョンが更新されている点と、各糖種総当たりの組み合わせで判別精度を算出している点である。総当たりの利点は、アミノ酸配列情報で高精度に判別することが困難な糖種の組み合わせ、すなわち、糖鎖修飾残基周辺のアミノ酸出現傾向が類似している糖種の組み合わせを探し当てることができることである。また、細胞内局在性やシグナルペプチド・膜貫通領域の有無が不明のタンパク質配列の糖種判別を行う可能性も鑑み、糖種特異的 PSSM のみをパラメータとして用いた際の信頼性を算出した (表 6-3)。

表 6-3. 自己テストおよび5分割交差検定法を用いた糖種の判別精度

糖種 ペア	計算領域	自己テスト			5分割交差検定		
		感度 (%)	選択性 (%)	成功率	感度 (%)	選択性 (%)	成功率
Xyl, Fuc	-1 to 2	100	100	1.00	100	100	1.00
GalNAc, Fuc	-1 to 0	97.2	94.6	0.959	97.1	95.0	0.960
GalNAc, Xyl	-2 to 0	99.7	100	0.998	83.3	96.0	0.883
GlcNAc, Fuc	-10 to 1	100	100	1.00	100	100	1.00
GlcNAc, Xyl	-3 to 1	83.3	88.2	0.857	88.3	83.0	0.847
GlcNAc, GalNAc	-3 to 3	68.5	78.8	0.734	55.4	54.674	0.549

PSSM を用いた判別においては、Fuc・Xyl・GalNAc 内、および Fuc・Xyl・GlcNAc 内は、どの組合せにおいても成功率 85%~100%の判別が可能であった。しかし、GalNAc と GlcNAc の組み合わせでは成功率 55%と、判別精度は極めて低かったことから、アミノ酸配列情報のみの判別が困難である組み合わせは GalNAc と GlcNAc であることがわかった。GlcNAc 修飾は細胞内局在経路情報が存在すれば完全に判別できるため、GalNAc と GlcNAc の判別には輸送シグナル配列の有

無を利用することが望ましいと考えられた。

そこで、糖種特異的 PSSM および細胞内局在経路（シグナルペプチド・膜貫通領域の有無）をパラメータとし、タンパク質 O 型糖鎖修飾における糖種判別を行った（表 6-4）。

表 6-4. PSSM およびシグナル配列の有無を用いた糖種判別精度
(5 分割交差検定)

糖種 ペア	計算領域	5 分割交差検定		
		感度 (%)	選択性 (%)	成功率
Xyl, Fuc	-1 to 2	100	100	1.00
GalNAc, Fuc	-1 to 0	97.1	95.0	0.960
GalNAc, Xyl	-2 to 0	83.3	96.0	0.883
GlcNAc, Fuc	—	100	100	1.00
GlcNAc, Xyl	—	100	100	1.00
GlcNAc, GalNAc	—	100	100	1.00

その結果、細胞内局在経路（シグナルペプチド・膜貫通領域の有無）をパラメータに加えることにより、GlcNAc 修飾は完全に分離することが可能となることがわかった。したがって、まず細胞内局在経路（シグナルペプチド・膜貫通領域の有無）によって GlcNAc と Fuc・Xyl・GalNAc の 2 グループに分類し、Fuc・Xyl・GalNAc 内において PSSM を用いて判別スコアを算出することにより、高精度に O 型糖鎖修飾糖種を判別することが可能となった。本研究における提案手法による O 型糖鎖修飾の糖種判別戦略を図 6-1 に示した。

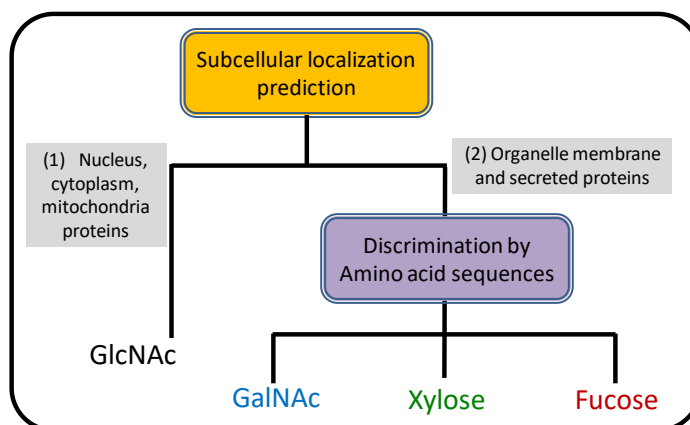


図 6-1. O 型糖鎖修飾糖種判別の戦略

第7章 結論

本研究では *O* 型糖鎖修飾を受けるタンパク質のアミノ酸配列・二次構造・立体構造・細胞内局在経路における糖種特異的な特徴を抽出し、*O* 型糖鎖修飾の糖種判別に用いることができるパラメータの探索、および、高精度糖種判別法の開発を行った。各章の概要を以下に示す。

第2章 一次配列に基づく糖種判別

O 型糖鎖修飾を受けている糖種によって、修飾残基付近のアミノ酸出現傾向が異なることを明らかにした。また、糖鎖修飾位置周辺のアミノ酸出現頻度から得られた PSSM を用いることにより、一部の糖種については高精度な糖種判別の開発に成功した。糖鎖修飾の有無を判別する精度が 85%程度であった従来法と比較し、特に Xyl は 5 分割交差検定において成功率 88%、Fuc は 100%と極めて高精度に判別することができた。従来の *O* 型糖鎖修飾予測法では対応していない糖種が多かったが、本研究では糖種特異的な PSSM を用いることにより、より多くの糖種に対応した判別を可能にした。一方で、GalNAc と GlcNAc を他の糖種の修飾サイトから判別する精度は低く、一次配列に基づく糖種判別には限界があることがわかった。

第3章 糖タンパク質の二次構造解析

立体構造データベース (PDB) の原子座標データに *O* 型糖鎖を含むタンパク質を対象とし、糖種ごとに、糖鎖修飾位置周辺の二次構造のバリエーションを調査した。その結果、糖種ごとに二次構造の出現傾向に特徴が見られ、各糖を修飾する糖転移酵素の認識における二次構造の指向性を示すことができた。*O* 型糖鎖修飾はストランド構造とコイル構造の境目に受けやすいことを明らかに

したほか、特に Fuc はストランド中の最も N 末端側の残基に修飾されやすい傾向にあることがわかった。

一方、PDB に含まれる O 型の糖の情報を調査したところ、GalNAc, GlcNAc 修飾位置は Swiss-Prot に数多く報告されているにも関わらず、いずれも結晶構造数は 1 桁しかなく、二次構造・立体構造の特徴を糖種判別法に活用するには、統計的有意性が乏しいことも示された。

第 4 章 糖鎖修飾残基周辺の空間的アミノ酸出現傾向解析

Fuc 修飾と GlcNAc 修飾を受けた糖鎖周辺の物理化学的環境を調査した。構造既知の糖タンパク質を対象に、糖鎖修飾残基および糖構成原子周辺の出現傾向を調べた。Fuc 修飾の周辺では、とりわけ Phe および Cys 残基の出現傾向が高かった。これら Cys 残基は他の Cys とジスルフィド結合しており、Fuc 転移酵素がジスルフィド結合からなる初期構造を認識している可能性を示すことができた。一方、GlcNAc 修飾では Tyr 残基および疎水性残基 (Ala, Val) の高頻度出現を見だし、糖鎖修飾位置周辺に高疎水性環境を形成していることを明らかにした。糖鎖認識残基の周辺に芳香族アミノ酸が存在することは、糖転移酵素や糖分解酵素にも共通しており、Fuc や GlcNAc 修飾位置周辺で見られた芳香族アミノ酸が、それぞれの糖鎖修飾の選択性に寄与している可能性を示すことができた。今後、PDB に含まれる GlcNAc 修飾タンパク質の構造情報が増加すれば、糖種判別法にも活用できるであろう。

第 5 章 タンパク質細胞内局在経路に基づいた糖鎖修飾分布の解析

O 型糖鎖修飾を受けている糖タンパク質について、細胞内局在・糖鎖修飾情報をもとに、細胞内での局在化経路および糖種に応じて分類した。糖タンパク質をシグナルペプチドや膜貫通領域の有無で、細胞内在型・分泌型・シグナルアンカー型・膜貫通型の 4 グループに分類し、O 型糖鎖修飾の糖種を調査した。シグナルペプチドもしくは膜貫通領域を有する分泌型・シグナルアンカー型、あるいはその両方を有する膜貫通型の糖タンパク質は、トランスロコンから小

胞体に挿入され、小胞体を経由する局在化経路（小胞体・ゴルジ体・細胞膜・細胞外分泌）を通る。小胞体経由の3グループにおいては、主として Fuc・Xyl・GalNAc の修飾が見られた。一方、小胞体を経由しない局在化経路（核・細胞質・ミトコンドリア）を通る細胞内在型タンパク質においては、O型糖鎖修飾の100%が GlcNAc によることがわかった。上記の結果は、O-Fuc・O-Xyl・O-GalNAc・O-GlcNAc 転移酵素の細胞内局在性とよく一致した。本項目では特に、一次配列では判別が難しかった GlcNAc 修飾タンパク質の細胞内局在化経路の特異性が明らかになり、糖種判別のための強力なパラメータになることが示された。

第6章 タンパク質 O 型糖鎖修飾の糖種判別

O 型糖鎖修飾を受けるタンパク質の糖鎖修飾位置周辺のアミノ酸配列から得られた糖種特異的 PSSM および細胞内局在経路（シグナルペプチド・膜貫通領域の有無）をパラメータとし、タンパク質 O 型糖鎖修飾における糖種判別を行った。本項目では、判別が難しい糖種の組み合わせを明らかにするために、各糖種総当たりの組み合わせ判別を行った。また、細胞内局在性やシグナルペプチド・膜貫通領域の有無が不明のタンパク質配列の糖種判別を行う可能性も鑑み、糖種特異的 PSSM のみをパラメータとして用いた際の糖種判別の精度も求めた。

PSSM を用いた判別においては、Fuc・Xyl・GalNAc 内、および Fuc・Xyl・GlcNAc 内は、どの組合せにおいても成功率 85%~100% の判別が可能であった。しかし、GalNAc と GlcNAc の組み合わせでは成功率 55% と、判別精度は極めて低かった。つまり、アミノ酸配列情報のみの判別が困難である組み合わせは GalNAc と GlcNAc であることがわかった。しかし、細胞内局在経路（シグナルペプチド・膜貫通領域の有無）をパラメータに加えることにより、GlcNAc 修飾位置は GalNAc 修飾位置から完全に分離することが可能である。したがって、まず細胞内局在経路（シグナルペプチド・膜貫通領域の有無）によって GlcNAc と Fuc・Xyl・GalNAc の2グループに分類し、Fuc・Xyl・GalNAc 内において PSSM から

スコアを算出することにより、高精度に O 型糖鎖修飾糖種を判別することが可能となった。

今後、新たな糖種についても、利用可能な公共データベースに情報が蓄積されることにより、本研究で行ったアミノ酸配列・立体構造・細胞内局在化経路の糖種特異的特徴を抽出して利用することにより、糖種を判別できると考えられる。修飾を受けた糖種をパラメータとしたタンパク質の細胞内局在予測法の開発や、糖鎖修飾のエラーによる疾患の具体的な要因を議論することにも応用できる。

第 5 章で行った細胞内局在化経路と糖種との相関解析を、人工タンパク質を用いて検証することは急務である。ヒト培養細胞に発現させた人工タンパク質の細胞内局在化経路の観察を行うとともに、修飾を受けた糖種を分析し、局在化経路と糖種の相関を評価することを計画している。まず、細胞内局在化経路を変更した人工タンパク質を設計する。次に、**HeLa** 細胞で人工タンパク質を発現させ、抗体染色法により局在化経路を調査し、目的の経路をたどっているかを確認する。続いて、人工タンパク質を精製し、ウェスタンブロッティングと質量分析法を用いて糖種の分析を行う予定である。

さらに、タンパク質の細胞内局在化経路や組織特異性、生物種特異性に基づいた糖種をバイオインフォマティクス解析と実験によって明らかにすることにより、網羅的な糖鎖修飾の知識ベースを構築でき、包括的な糖鎖修飾のメカニズムを議論できる。適切な種類の糖鎖をタンパク質へ効果的に導入し、タンパク質の機能制御を可能にすることにより、人工粘膜やがんワクチン、動物生体を用いない人工抗体など、医療・創薬分野への工学的応用が実現できる。

謝辞

本論文は筆者が明治大学大学院理工学研究科電気工学専攻の博士後期課程在籍中に行った研究成果を博士論文としてまとめたものです。

本研究を博士論文という形でまとめるにあたり、指導教員である池田有理准教授には、本研究の全般に渡りご指導をいただいたこと、海外・国内で多くの研究発表の機会をいただいたこと、論文発表や学会発表では原稿をきめ細かく確認していただいたことに、甚大なる感謝を申し上げます。

石田義久教授には、博士後期課程進学の際に多くのご協力を賜りましたことに、厚く御礼申し上げます。

本論文の審査に際して、鎌田弘之教授、加藤徳剛教授には大変多くのご指導を賜りました。厚く御礼申し上げます。

博士後期課程に進学するにあたり、日本学生支援機構による奨学金や明治大学の助手、日本学術振興会の特別研究員 (DC2) に採用していただいたことで、研究に集中できる環境に身をおくことができました。深く感謝申し上げます。また、本研究を遂行するにあたり、研究費として特別研究員奨励費を賜ったことに感謝いたします。

博士後期課程在籍中、本論文を作成する際に、田中諒先生に多くのご助言、ご協力をいただきましたことに深く感謝申し上げます。

最後に、研究活動を共にした生命情報科学研究室の仲間たちに、深く感謝申し上げます。

参考文献

- Alfaro, J. F., Gong, C.X., Monroe, M. E., Aldrich, J. T., Clauss, T. R., Purvine, S. O., Wang, Z., Camp, D. G. 2nd, Shabanowitz, J., Stanley, P., Hart, G. W., Hunt, D. F., Yang, F., Smith, R. D. (2012) Tandem mass spectrometry identifies many mouse brain *O*-GlcNAcylated proteins including EGF domain-specific *O*-GlcNAc transferase targets., *Proc. Natl. Acad. Sci. USA*, 109, 7280-7285.
- Apweiler, R., Hermjakob, H., Sharon, N. (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database, *Biochim. Biophys. Acta.*, 1473, 4-8.
- Bause, A. (1983) Structural requirements of *N*-glycosylation of proteins, *Biochem. J.*, 209, 331-336.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank, *Nucleic Acids Research*, 28, 1, 235-242.
- Berman, H., Henrick, K., Nakamura, H., Markley, J. L. (2007) The worldwide Protein Data Bank wwPDB: ensuring a single uniform archive of PDB data, *Nucleic Acids Research*, 35, D301-D303.
- Bernfield, M., Götte, M., Park, P. W., Reizes, O., Fitzgerald, M. L., Lincecum, J., and Zako, M. (1999) Functions of cell surface heparin sulfate proteoglycans, *Annu. Rev. Biochem.*, 68, 729-777.
- Blom, N., Sicheritz-Pontén, T., Gupta, R., Gammeltoft, S., and Brunak, S. (2004) Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence, *Proteomics*, 4, 1633-1649.
- Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S., Schneider, M. (2003) SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003, *Nucleic Acids Res.*, 31, 365-370.

- Cao, W., Cao, J., Huang, J., Yao, J., Yan, G., Xu, H., Yang, P. (2013) Discovery and confirmation of O-GlcNAcylated proteins in rat liver mitochondria by combination of mass spectrometry and immunological methods, *PLOS ONE*, 8, e76399.
- Caragea, C., Sinapov, J., Silvescu, A., Dobbs, D., Honavar, V. (2007) Glycosylation site prediction using ensembles of Support Vector Machine classifiers, *BMC Bioinform.*, 8, 438.
- Carraway, K. L. and Hull, S. R. (1991) Cell surface mucin-type glycoproteins and mucin-like domains, *Glycobiology*, 1, 131-138.
- Claverie, J. M. and Audic, S. (1996) The statistical significance of nucleotide position-weight matrix matches, *Comput. Appl. Biosci.*, 12, 431-439.
- Comer, F. I. and Hart, G. W. (2000) O-glycosylation of nuclear and cytosolic proteins, *J. Biol. Chem.*, 275, 29179-29182.
- Esko, J. D. and Selleck, S. B. (2002) Order out of chaos: Assemble of ligand binding sites in heparin sulfate, *Annu. Rev. Biochem.*, 71, 435-471.
- Etchuya, K. and Mukai, Y. (2013) Sugar type discrimination in O-glycosylation based on protein primary sequences, *J. Biomech. Sci. Eng.*, 8, 3, 225-232.
- Etchuya, K., Nambu, R., Mukai, Y. (2013) Discrimination of xylose O-glycosylation sites in mammalian proteins, *Chemistry Letters*, 42, 9, 1043-1045.
- Etchuya, K. and Mukai, Y. (2015) Structural characteristics around O-glycosylation sites in mammalian proteins, *J. Biomech. Sci. Eng.*, 10, 14-00249, 1-6.
- Etchuya, K. and Mukai, Y. (2016) Structural environment factor around O-glycosylation sites in mammalian proteins, submitted.
- Etchuya, K., Miyasaka, G., Mukai, Y. (2016) Sugar type discrimination of O-glycosylation in mammalian proteins based on protein subcellular localization and amino acid propensity, submitted.
- Fukuda, M. (2002) Roles of mucin-type O-glycans in cell adhesion, *Biochim. Biophys. Acta*, 1573, 394-405.
- Furmanek, A. and Hofsteenge, J. (2000) Protein C-mannosylation: Facts and questions, *Acta Biochem. Pol.*, 47, 781-789.
- Gupta, R. and Brunak, S. (2002) Prediction of glycosylation across the human proteome and the correlation to protein function, *Pac. Symp. Biocomput.*, 7, 310-322.
- Haltiwanger, R. S. and Lowe, J. B. (2004) Role of glycosylation in development, *Annu.*

- Rev. Biochem.*, 73, 491-537.
- Hamby, S. E. and Hirst, J. D. (2008) Prediction of glycosylation sites using random forests, *BMC Bioinform.*, 9, 500.
- Hansen, J. E., Lund, O., Engelbrecht, J., Bohr, H., Nielsen, J., Hansen, J. -E. S., Brunak, S. (1995) Prediction of *O*-glycosylation of mammalian proteins: specificity patterns of UDP-GalNAc: polypeptide *N*-acetylgalactosaminyltransferase, *Biochem. J.*, 308, 801-813.
- Hart, G. W. (1997) Dynamic *O*-linked glycosylation of nuclear and cytoskeletal proteins, *Annu. Rev. Biochem.*, 66, 315-335.
- Hattrup, C. L. and Gendler, S. J. (2008) Structure and function of the cell surface tethered mucins, *Annu. Rev. Physiol.*, 70, 431-457.
- Hollingsworth, M. A. and Swanson, B. J. (2004) Mucins in cancer: Protection and control of the cell surface, *Nat. Rev.*, 4, 45-60.
- Hooft, R. W. W., Sander, C., Scharf, M., Vriend, G. (1996) The PDBFINDER database: a summary of PDB, DSSP and HSSP information with added value, *CABIOS*, 12, 525-529.
- Hounsell, E. F., Davies, M. J., Renouf, D. V. (1996) *O*-linked protein glycosylation structure and function, *Glycoconj. J.*, 13, 19-26.
- Huang, Y., Niu, G., Gao, Y., Fu, L., Li, W. (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences, *Bioinformatics*, 26, 680-682.
- Julenius, K., Molgaard, A., Gupta, R., Brunak, S. (2004) Prediction, conservation analysis, and structural characterization of mammalian mucin-type *O*-glycosylation sites, *Glycobiology*, 15, 153-164.
- Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers*, 22, 2577-2637.
- Kao, Y. H., Lee, G. F., Wang, Y., Starovasnik, M. A., Kelley, R. F. (1999) The Effect of *O*-Fucosylation on the First-like Domain from Human Blood Coagulation Factor VII, *Biochem.*, 38, 7097-7110.
- Kolset, S. O., Prydz, K., Pekler, G. (2004) Intracellular proteoglycans, *Biochem. J.*, 379, 217-227.
- Kornfeld, R. and Kornfeld, S. (1985) Assembly of asparagine-linked oligosaccharides,

- Annu. Rev. Biochem.*, 54, 631-664.
- Kreppel, L., Blomberg, M., Hart G. (1997) Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and characterization of a unique O-GlcNAc transferase with multiple tetratricopeptide repeats, *J. Biol. Chem.*, 272, 14, 9308-15.
- Live, D., Wells, L., Boons, G. (2013) Dissecting the molecular basis of the role of the O-mannosylation pathway in disease: α -dystroglycan and forms of muscular dystrophy, *ChemBioChem*, 14, 18, 2392-402.
- Li, S., Liu, B., Zeng, R., Cai, Y., Li, Y. (2006) Predicting O-glycosylation sites in mammalian proteins by using SVMs, *Comput. Biol. Chem.*, 30, 203-208.
- Lowe, J. B. and Marth, J. (2003) A genetic approach to mammalian glycan function, *Annu. Rev. Biochem.*, 72, 643-691.
- Malik, A. and Ahmad, S. (2007) Sequence and structural features of carbohydrate binding in proteins and assessment of predictability using a neural network, *BMC Structural Biology*, 7, 1.
- Mukai, Y., Yoshizawa, M., Sasaki, T., Ikeda, M., Tomii, K., Hirokawa, T., Suwa, M. (2011) Discrimination of golgi type II membrane proteins based on their hydropathy profiles and the amino acid propensities of their transmembrane regions, *Biosci. Biotech. Biochem.*, 75, 1, 82-88.
- Mukai, Y., Ikeda, M., Tanaka, H., Konishi, T., Oura, O., Sasaki, T. (2013) Discrimination of mammalian GPI-anchored proteins by their hydropathy and amino acid propensities, *Biosci. Biotech. Biochem.*, 77, 3, 526-533.
- Okuyama, R. and Marshall, S. (2003) UDP-N-acetylglucosaminyl transferase (OGT) in brain tissue: temperature sensitivity and subcellular distribution of cytosolic and nuclear enzyme, *J. Neurochem.*, 86, 5, 1271-80.
- Rana, N.A., Nita-Lazar, A., Takeuchi, H., Kakuda, S., Luther, K.B. (2011) O-glucose trisaccharide is present at high but variable stoichiometry at multiple sites on mouse Notch1, *J. Biol. Chem.*, 286, 36, 31623-31637.
- Rao, Z., Handford, P., Mayhew, M., Knott, V., Brownlee, G.G., Stuart, D. (1995) The structure of a Ca²⁺-binding epidermal growth factor-like domain: its role in protein-protein interactions, *Cell*, 82, 131-141.
- Sasaki, K., Nagamine, N., Sakakibara, Y. (2009) Support vector machine prediction of N- and O-glycosylation sites using whole sequence information and subcellular

- location, *IPSI Trans. Bioinform.*, 2, 25-35.
- Satoh, T., Chen, Y., Hu, D., Hanashima, S., Yamamoto, K., Yamaguchi, Y. (2010) Structural basis for oligosaccharide recognition of misfolded glycoproteins by OS-9 in ER-associated degradation., *Mol. Cell*, 40, 6, 905-16.
- Schjoldager, K., Vester-Christensen, M., Bennett, E., Levery, S., Schwientek, T., Yin, W., Blixt, O., Clausen, H. (2010) O-glycosylation modulates proprotein convertase activation of angiopoietin-like protein 3: possible role of polypeptide GalNAc-transferase-2 in regulation of concentrations of plasma lipids., *J. Biol. Chem.*, 285, 47, 36293-303.
- Schurph, T., Chen, Q., Liu, J.H., Wang, R., Springer, T. A., Wang, J. H. (2012) The RGD finger of Del-a is a unique structural feature critical for integrin binding, *FASEB J.*, 26, 3412-3420.
- Staden, R. (1989) Methods for calculating the probabilities of finding patterns in sequences, *Comput. Appl. Biosci.*, 5, 89-96.
- Stanley, P. (2007) Regulation of notch signaling by glycosylation, *Curr. Opin. Struct. Biol.*, 17, 530-535.
- Stentoft, C., Vakhrushev, S., Joshi, H., Kong, Y., Vester-Christensen, M., Schjoldager, K., Lavrsen, K., Dabelsteen, S., Pedersen, N., Marcos-Silva, L., Gupta, R., Bennett, E., Mandel, U., Brunak, S., Wandall, H., Levery, S., Clausen, H. (2013) Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology, *EMBO J.*, 32, 10, 1478-88.
- Strous, G. J. and Dekker, J. (1992) Mucin-type glycoproteins, *Crit. Rev. Biochem. Mol. Biol.*, 27, 43-52.
- Tan, K., Duquette, M., Liu, J.H., Dong, Y., Zhang, R., Joachimiak, A., Lawler, J., Wang, J.H. (2002) Crystal structure of the TSP-1 type 1 repeats: a novel layered fold and its biological implication, *J. Cell. Biol.*, 159, 2, 373-382.
- The UniProt Consortium (2017) UniProt: the universal protein knowledgebase, *Nucl. Acids Res.*, 45, D158-D169.
- Varki, A. (1993) Biological roles of oligosaccharides: all of the theories are correct, *Glycobiology*, 3, 97-130.
- Wang, Z., Udeshi, N. D., Slawson, C., Compton, P. D., Sakabe, K., Cheung, W. D., Shabanowitz, J., Hunt, D. F., Hart, G. W. (2010) Extensive Crosstalk Between

- O-GlcNAcylation and Phosphorylation Regulates Cytokinesis. *Sci. Signal.* 3, ra2.
- Whelan, S. A. and Hart, G. W. (2003) Proteomic approaches to analyze the dynamic relationships between nucleocytoplasmic protein glycosylation and phosphorylation, *Circ. Res.*, 93, 1047-1058.
- Zhou, K., Ai, C., Dong, P., Fan, X., Yang, L. (2012) A novel model to predict O-glycosylation sites using a highly unbalanced dataset, *Glycoconjugate J.*, 29, 551-564.

研究業績

[著書]

1. 越中谷賢治, 濱田康太, 向井有理「バイオインフォマティクス入門 (第 1 章 生命科学)」, 日本バイオインフォマティクス学会編, 慶應義塾大学出版会, pp. 10-49, 2015.

[査読付き論文]

0. Kenji Etchuya, Yuri Mukai: Structural environment factor around *O*-glycosylation sites in mammalian proteins, submitted.
0. Kenji Etchuya, Go Miyasaka, Yuri Mukai: Sugar type discrimination of *O*-glycosylation in mammalian proteins based on protein subcellular localization and amino acid propensity, submitted.
1. Kenji Etchuya, Yuri Mukai (2015) Structural characteristics around *O*-glycosylation sites in mammalian proteins, *Journal of Biomechanical Science and Engineering*, 10, 14-00249, 1-6.
2. Kenji Etchuya, Ryohei Nambu, Yuri Mukai (2013) Discrimination of xylose *O*-glycosylation sites in mammalian proteins, *Chemistry Letters*, 42, 9, 1043-1045.
3. Kenji Etchuya, Yuri Mukai (2013) Sugar type discrimination in *O*-glycosylation based on protein primary sequences, *Journal of Biomechanical Science and Engineering*, 8, 3, 225-232.
4. Yuri Mukai, Daiki Takahashi, Keiya Inoue, Hiromu Sugita, Tatsuki Kikegawa, Kenji Etchuya (2016) Secondary structure of GPI attachment signal region monitored by circular dichroism, *Chemistry Letters*, 45, 10, 1153-1155.

5. Yuri Mukai, Daiki Takahashi, Tsubasa Ogawa, Kota Hamada, Kenji Etchuya (2016) Study of molecular recognition mechanism in protein GPI modification: a bioinformatics analysis of interaction between GPI-anchored proteins and modification enzyme, *Journal of Biomechanical Science and Engineering*, 11, 15-00361, 1-7.
6. Yuri Mukai, Keitaro Machino, Kenji Etchuya, Hirotaka Tanaka, Takanori 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.

[査読付き国際学会プロシーディングス]

1. Kenji Etchuya, Tatsuki Kikegawa, Yuri Mukai (2016) Correlation between protein subcellular localization and sugar variation, Proceedings of the 13th International Conference on Flow Dynamics, Sendai, Japan, pp. 306-307.
2. Kenji Etchuya, Hiromu Sugita, Tatsuki Kikegawa, Kota Hamada, Naoyuki Takachio, Noritaka Kato, Makoto Ohta and Yuri Mukai (2015) Correlation between physicochemical properties of protein signal sequence variation and subcellular transportation, Proceedings of the 15th International Symposium on Advanced Fluid Information, Sendai, Japan, pp. 86-87.
3. Kenji Etchuya, Makoto Ohta, Yuri Mukai (2014) Study of 3D recognition by glycosyltransferase in protein sugar modification, Proceedings of the 14th International Symposium on Advanced Fluid Information, Sendai, Japan, pp. 92-93.
4. Kenji Etchuya, Yuri Mukai (2013) Structural aspects of proteins modified by oligosaccharides, Proceedings of tenth International Conference on Flow Dynamics, Sendai, Japan, pp. 616-617.
5. Tatsuki Kikegawa, Tomonao Iibuchi, Kenji Etchuya, Makoto Ohta, Noritaka Kato, Yuri Mukai (2016) Correlation between physicochemical properties of protein signal sequence variation and subcellular transportation, Proceedings of the 16th International Symposium on Advanced Fluid Information, Sendai, Japan, pp. 88-89.

6. Daiki Takahashi, Tsubasa Ogawa, Kenji Etchuya, Kota Hamada, Yuri Mukai (2015) Protein recognition mechanisms for GPI modification, Proceedings of the 12th International Conference on Flow Dynamics, Sendai, Japan, pp. 334-335.
7. Kota Hamada, Naoyuki Takachio, Hiromu Sugita, Noritaka Kato, Makoto Ohta, Kenji Etchuya, Yuri Mukai, Correlation between physicochemical properties of protein signal sequence variation and subcellular transportation, Proceedings of the 14th International Symposium on Advanced Fluid Information, Sendai, Japan, pp. 90-91.

[査読付き国際学会発表 (筆頭のみ)]

1. Kenji Etchuya, Yuri Mukai: Environment factor depending on each sugar type around *O*-glycosylation sites in mammalian proteins, The 15th European Conference on Computational Biology, The Hague, Netherlands, Sep. 2016.
2. Kenji Etchuya, Yuri Mukai: Comparative study of the sequences and structures around *O*-glycosylation sites between each sugar type in mammalian proteins, The Pacific Symposium on Biocomputing 2016, Hawaii, USA, Jun. 2016.
3. Kenji Etchuya, Yuri Mukai: Interaction analysis between sugar chain and aromatic residue in mammalian, The 40th FEBS Congress, Berlin, Germany, Jul. 2015.
4. Kenji Etchuya, Yuri Mukai: Structural characteristics of Fuc modification sites in mammalian proteins, The 13th European Conference on Computational Biology, Strasbourg, France, Sep. 2014.
5. Kenji Etchuya, Yuri 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.
6. Kenji Etchuya, Yuri Mukai: Characteristics for sugar modification from protein structure, 38th FEBS Congress, St. Petersburg, Russia, Jul. 2013.
7. Kenji Etchuya, Hirotaka Tanaka, Takeo Terasaki, Takanori Sasaki, Yuri 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.

[査読付き国内学会発表 (筆頭のみ)]

1. 越中谷賢治, 向井有理: O 型糖鎖修飾を受ける哺乳類タンパク質の構造的解析, 第 37 回日本分子生物学会年会, 横浜, 2014 年 11 月.
2. 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.
3. 越中谷賢治, 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.
4. 越中谷賢治, 向井有理: タンパク質の二次構造と糖鎖修飾の相関(口頭発表), 第 2 回日本生物物理学会関東支部会, 東京, 2013 年 3 月.
5. 越中谷賢治, 田中大貴, 寺崎武夫, 佐々木貴規, 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.

付録

[第 2 章]

表 A-1. Fuc 修飾位置を他の糖の修飾位置から判別するための PSSM

表 A-2. Xyl 修飾位置を他の糖の修飾位置から判別するための PSSM

表 A-3. GalNAc 修飾位置を他の糖の修飾位置から判別するための PSSM

表 A-4. GlcNAc 修飾位置を他の糖の修飾位置から判別するための PSSM

[第 6 章]

表 A-5. Xyl 修飾位置と Fuc 修飾位置を判別するための PSSM

表 A-6. GalNAc 修飾位置と Fuc 修飾位置を判別するための PSSM

表 A-7. Xyl 修飾位置と GalNAc 修飾位置を判別するための PSSM

表 A-8. GlcNAc 修飾位置と Fuc 修飾位置を判別するための PSSM

表 A-9. GlcNAc 修飾位置と Xyl 修飾位置を判別するための PSSM

表 A-10. GlcNAc 修飾位置と GalNAc 修飾位置を判別するための PSSM

表 A-1. Fuc 修飾位置を他の糖の修飾位置から判別するための PSSM (第 2 章)

Amino Acid	Alignment Position										
	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0
A	-2.70	-0.12	-3.04	-2.90	-3.06	0.05	-3.21	-3.06	-3.09	-0.44	3.41
C	2.70	-1.22	-1.56	0.95	-1.22	3.22	-1.55	4.64	-1.54	3.41	3.41
D	-2.44	-2.97	-2.90	-2.78	-0.02	0.12	-3.15	-2.93	-2.30	-0.99	3.41
E	-3.07	0.22	0.18	0.18	-2.89	0.01	-3.06	-3.23	-3.28	-2.16	3.41
F	-2.44	-2.09	-1.81	-1.69	-1.68	-0.99	1.25	-1.99	-1.21	-2.16	3.41
G	-0.03	-2.78	1.41	-3.04	-2.93	-2.89	-2.85	-2.69	1.73	1.31	3.41
H	2.04	-1.56	1.49	-1.40	-1.68	-0.71	-2.08	-1.68	-0.70	-1.54	3.41
I	-1.92	-2.55	-2.17	-1.91	-2.37	-2.31	0.61	-2.08	-2.30	-2.48	3.41
K	-2.55	0.87	-1.81	-1.91	-1.55	-2.24	-2.49	1.14	-2.30	-1.21	3.41
L	-0.30	-2.78	-2.55	-2.70	-3.00	-2.64	1.03	-2.89	-2.92	-2.37	3.41
M	-0.72	-1.22	-0.72	-1.56	-2.00	-1.21	-0.99	-0.99	-1.99	-0.99	3.41
N	-1.41	-1.91	0.80	0.80	-1.55	-2.00	-1.68	3.00	-2.43	-1.21	3.41
P	1.15	-3.16	-3.29	1.86	0.68	-3.53	0.91	-3.54	-0.50	-3.92	3.41
Q	-2.25	-2.38	-2.25	-2.60	-2.24	-2.54	1.48	-3.03	-2.73	-2.16	3.41
R	-2.38	-2.44	-2.70	-2.55	0.88	-2.59	-2.49	-1.90	-2.16	-1.39	3.41
S	-0.44	0.32	1.28	0.26	-3.24	0.40	0.68	-2.99	1.68	-3.12	-0.40
T	0.00	-2.83	-3.13	-0.49	-2.86	0.62	-2.93	-2.64	-2.64	-0.04	0.16
V	-2.44	-2.90	-2.78	-2.44	0.11	-2.54	-2.43	-3.23	-2.64	1.22	3.41
W	0.34	3.54	-0.72	-0.32	4.08	-1.21	0.36	-0.99	-0.31	-0.70	3.41
Y	-1.81	-1.56	-2.01	-1.22	-1.80	-1.39	-0.99	-0.99	-1.79	-1.39	0.36

Amino Acid	Alignment Position									
	1	2	3	4	5	6	7	8	9	10
A	-2.96	-3.58	-2.92	0.19	-3.18	-2.60	-2.65	-2.91	-3.22	-2.79
C	4.78	0.36	-1.68	0.36	0.35	-1.00	-1.69	-0.33	-2.01	2.88
D	-2.08	-3.06	2.12	0.36	-2.86	-2.65	-2.70	-2.91	0.66	-2.98
E	-2.99	-2.64	0.74	-3.03	-2.60	-2.90	0.26	0.39	-0.15	0.00
F	-0.99	-1.90	-2.37	-1.39	-1.22	-2.00	2.03	-2.10	-2.25	1.81
G	-4.01	1.76	0.51	2.17	-3.03	-2.94	0.01	-2.94	0.25	-3.02
H	-0.70	1.15	-1.21	-2.00	-1.55	0.87	-2.00	-0.72	1.48	-1.92
I	-1.39	-2.08	-1.79	-1.39	1.28	1.13	-1.81	-1.56	1.61	0.79
K	-2.08	2.32	2.74	-2.16	0.50	-2.50	0.87	-2.75	0.87	-2.18
L	-2.08	-2.54	0.68	-3.00	0.71	-2.74	0.18	-2.87	-2.79	-2.39
M	-0.99	-1.54	-1.21	1.65	-1.55	-1.56	1.82	-2.01	-1.70	-1.01
N	-1.54	-2.16	1.92	1.25	-2.24	1.82	-1.40	-1.81	0.79	-1.92
P	-3.44	-3.20	-4.25	-3.71	0.13	-3.36	-3.66	0.42	-3.46	-3.25
Q	-2.43	1.15	0.68	0.67	0.80	1.65	0.55	-2.55	0.95	-2.39
R	-2.48	-2.43	-1.68	-2.08	-2.49	0.40	-2.60	2.61	0.66	2.35
S	-3.20	0.08	-0.35	-3.33	0.01	-0.32	0.53	0.42	-0.03	-3.25
T	-3.20	-2.99	-3.06	-3.38	-0.02	-0.03	0.58	-3.32	-3.05	-3.30
V	-2.77	0.79	-2.30	-2.73	1.07	-2.86	-2.90	-2.94	-2.45	-2.95
W	-0.70	-0.31	3.41	3.40	0.35	-0.72	-0.32	-0.33	-1.23	-0.33
Y	-0.70	-0.99	-1.68	-1.39	-1.00	2.03	-1.56	2.15	1.48	-2.10

表 A-2. Xyl 修飾位置を他の糖の修飾位置から判別するための PSSM (第 2 章)

Amino Acid	Alignment Position										
	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0
A	0.02	-0.30	-1.13	-0.22	-1.89	-0.69	0.08	0.07	-0.44	0.25	1.53
C	-0.52	0.17	-3.43	0.33	-3.09	-3.87	1.24	-0.51	-3.41	1.53	1.53
D	1.40	0.88	1.29	1.18	0.97	0.39	0.75	1.24	1.67	2.60	1.53
E	-0.42	0.71	0.65	0.27	0.28	0.55	0.88	1.06	1.23	1.07	1.53
F	0.64	0.83	0.83	1.52	-0.38	2.59	1.13	0.98	0.18	1.07	1.53
G	1.08	0.19	0.52	0.78	0.23	0.48	-0.49	1.35	0.07	1.18	1.53
H	-3.10	-0.23	-3.57	-3.28	-0.38	-2.58	-0.83	-3.55	-2.58	-3.41	1.53
I	-0.64	-0.60	0.32	-0.63	0.44	0.84	0.22	-0.82	-1.13	0.29	1.53
K	-4.43	-0.25	-0.52	-0.63	-0.23	-0.99	-1.26	0.05	-1.06	-3.08	1.53
L	-1.06	0.42	0.20	1.34	0.15	0.62	-0.40	-0.97	0.24	0.08	1.53
M	-2.60	-3.28	0.85	0.62	0.55	1.13	0.46	-2.86	0.56	-2.86	1.53
N	-0.06	0.15	0.51	-0.33	0.62	0.56	-0.38	-3.77	-1.19	0.18	1.53
P	0.04	-0.05	-0.42	-1.93	-0.44	-0.92	-0.97	-2.35	-2.37	-1.64	1.53
Q	-1.01	-0.41	0.23	0.43	-4.11	-1.32	-0.40	-0.11	-0.01	-0.91	1.53
R	0.42	-0.01	-0.76	-4.47	0.24	-0.65	-1.26	0.16	-4.03	-3.26	1.53
S	-0.45	-0.20	-0.27	-0.89	-0.14	-0.49	0.39	-1.79	-0.01	0.00	1.00
T	-0.74	-4.70	-1.23	-1.26	-0.50	-0.04	-0.25	-0.25	0.10	-0.47	-3.11
V	0.34	-0.22	0.42	-0.01	0.39	0.22	-1.20	-0.60	-0.69	-1.20	1.53
W	-1.53	-0.17	-2.59	1.52	-3.67	0.18	-1.52	0.47	1.53	0.87	1.53
Y	1.73	-0.23	-3.88	0.17	0.84	0.85	0.46	0.47	1.31	0.85	-1.51

Amino Acid	Alignment Position									
	1	2	3	4	5	6	7	8	9	10
A	-4.83	-0.36	-1.01	-0.21	0.11	-0.66	-0.71	-0.23	-0.59	0.18
C	-4.17	-1.51	0.45	-1.52	-1.52	-2.88	0.44	5.23	0.03	0.42
D	-3.95	1.78	0.63	1.09	0.72	1.04	0.76	1.59	0.91	0.37
E	-4.86	0.84	1.03	0.93	0.15	0.66	0.95	0.69	0.38	0.72
F	-2.86	0.17	1.03	-0.04	1.13	-3.88	1.52	1.18	-4.13	0.83
G	3.13	-0.89	2.34	-0.11	1.53	-0.26	0.87	0.60	0.36	0.66
H	-2.58	-0.72	0.18	-3.87	-3.42	-1.00	-0.74	-2.60	-0.39	0.14
I	-3.26	-0.06	-3.66	-0.04	-0.14	-3.88	-3.68	0.61	-1.01	-0.34
K	-3.95	-3.55	-2.58	-4.04	-4.47	0.28	0.23	-0.37	-1.01	-0.18
L	-3.95	0.22	0.01	0.15	-0.15	1.08	-0.55	0.32	0.76	0.34
M	-2.86	-0.37	0.18	0.63	0.62	1.24	-3.28	-3.88	-0.39	0.44
N	-3.41	-0.15	0.05	0.69	-0.99	0.62	-3.28	-0.52	-0.34	-3.80
P	-5.31	-0.88	-2.00	-0.68	-1.03	-0.75	-0.63	-0.48	0.33	-0.62
Q	-4.30	0.05	-1.19	0.36	-1.07	-0.55	-0.14	-1.33	0.71	-0.42
R	-4.35	-4.30	1.53	-0.83	-1.26	-1.48	-0.26	-4.26	-0.48	-0.38
S	-5.07	0.63	-0.81	0.06	0.07	0.02	-0.92	-0.09	-0.46	-0.37
T	-5.07	-1.79	-1.15	-0.13	-0.44	-0.20	-0.08	-1.01	-0.71	-0.43
V	-1.60	-1.79	-4.17	-0.35	-0.02	0.09	0.05	-1.03	0.34	-1.04
W	0.87	1.53	1.53	1.53	4.57	2.19	1.52	-2.20	1.11	-2.21
Y	-2.58	-2.86	1.03	-3.27	1.53	0.15	-0.23	-0.52	-3.57	0.82

表 A-3. GalNAc 修飾位置を他の糖の修飾位置から判別するための PSSM (第 2 章)

Amino Acid	Alignment Position										
	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0
A	-0.17	0.14	0.80	-0.06	0.87	0.01	-0.50	0.49	-0.15	-0.02	-0.45
C	-1.12	0.92	4.52	0.03	4.17	-0.26	-0.16	-2.49	-0.74	-0.45	-0.45
D	-0.32	0.06	-0.50	-0.60	-0.26	0.10	-0.27	-0.59	-0.86	-1.52	-0.45
E	0.35	-0.58	-0.05	-0.21	0.09	-0.26	-0.26	-0.25	-0.25	-0.92	-0.45
F	-1.12	0.26	0.26	-0.94	0.64	-1.51	-0.85	-1.00	0.90	-0.61	-0.45
G	-0.74	-0.59	-0.50	-0.37	-0.23	-1.03	-0.14	-1.42	-0.33	-1.08	-0.45
H	-0.04	0.47	0.26	1.14	-0.44	0.22	0.64	0.63	0.22	1.30	-0.45
I	-0.03	0.11	-0.58	-0.03	0.07	-0.04	0.32	-0.45	1.47	0.24	-0.45
K	1.23	0.48	0.26	0.94	-0.72	-0.16	0.24	-0.27	-0.86	-0.06	-0.45
L	0.79	-0.26	0.66	-0.80	-0.44	-0.54	0.65	0.80	0.13	0.15	-0.45
M	3.68	0.47	-1.10	-0.72	0.11	-0.05	-1.51	0.62	-0.27	0.61	-0.45
N	0.25	0.41	0.26	-0.18	-0.16	-0.26	0.64	-0.45	1.07	-0.06	-0.45
P	-0.20	0.14	0.08	0.17	0.02	0.84	0.06	0.36	0.59	1.68	-0.45
Q	-0.43	-0.43	0.16	0.18	0.56	0.58	-0.09	0.06	-0.19	-0.74	-0.45
R	-0.43	0.17	-0.61	0.18	-0.44	-0.24	0.50	-0.05	0.01	1.12	-0.45
S	0.29	-0.04	-0.50	0.32	0.03	-0.02	-0.56	0.78	-0.26	0.15	-0.97
T	0.83	0.95	0.92	0.90	0.76	0.12	0.85	0.46	0.24	0.28	1.04
V	-0.55	0.10	-0.09	-0.32	-0.23	0.58	0.43	0.32	0.70	-0.57	-0.45
W	2.61	-1.23	3.68	-4.15	-1.13	0.91	2.60	0.62	-4.17	0.22	-0.45
Y	-0.66	-0.14	-0.61	-0.83	-0.66	0.24	0.62	-0.45	-0.67	-1.13	-3.50

Amino Acid	Alignment Position									
	1	2	3	4	5	6	7	8	9	10
A	0.51	0.24	-0.31	0.08	-0.33	0.40	0.71	-0.07	0.29	0.25
C	-1.83	2.59	-1.53	2.60	2.60	0.62	0.06	-4.16	-0.77	-0.94
D	0.64	-0.96	-0.74	-0.36	-0.13	-0.54	0.11	-0.67	-0.56	0.57
E	0.82	-0.55	-0.70	-0.25	0.94	-0.22	-0.68	-0.36	0.29	-0.31
F	3.94	0.39	-0.20	1.13	-0.84	1.04	-1.28	-0.78	2.08	-0.44
G	-2.55	-0.37	-2.15	-0.90	-1.08	-0.16	-1.09	-0.45	-0.29	-0.98
H	0.22	-0.63	0.90	1.81	1.31	0.14	1.04	3.67	0.06	0.12
I	1.12	-0.45	0.23	-0.16	-0.76	-0.26	0.93	-0.73	-0.73	-0.03
K	0.35	0.05	0.22	-0.60	0.72	-0.44	-0.44	-0.18	-0.44	-0.28
L	0.64	0.64	-0.10	-0.04	0.07	0.00	0.60	0.02	0.31	-0.79
M	3.94	0.63	0.90	-0.16	-0.16	-0.16	0.24	0.24	-0.95	0.63
N	-0.17	0.46	-1.01	0.11	-0.15	-1.34	1.13	0.79	0.65	0.93
P	1.91	0.68	1.38	0.86	0.41	0.62	0.70	0.47	-0.05	0.03
Q	0.15	-0.63	-0.57	-0.56	0.25	-0.22	-0.44	-0.13	-1.63	0.35
R	-0.67	-0.10	-0.96	-0.29	0.25	0.18	0.47	-0.45	-0.32	-0.70
S	1.00	-0.59	0.37	-0.40	0.25	-0.13	-0.23	-0.40	0.12	0.88
T	1.00	0.93	0.48	0.11	0.39	0.86	0.01	0.85	0.61	0.41
V	0.69	0.93	0.24	0.59	-0.35	-0.13	-0.01	1.34	-0.09	0.15
W	0.22	-0.45	-0.45	-0.44	-3.49	-1.11	-4.16	3.27	-0.04	3.28
Y	0.22	0.61	-0.67	0.24	-1.51	-1.00	-0.73	-0.22	0.64	-0.44

表 A-4. GlcNAc 修飾位置を他の糖の修飾位置から判別するための PSSM (第 2 章)

Amino Acid	Alignment Position										
	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0
A	0.82	0.36	0.14	0.53	0.29	0.78	0.93	-0.89	0.82	-0.21	1.79
C	0.56	-2.84	-3.17	-0.66	-2.83	-3.61	-3.16	-3.41	0.04	1.79	1.79
D	-4.04	-1.51	-0.73	-0.60	-4.71	-0.79	-0.56	-0.75	-0.80	-2.60	1.79
E	-0.46	0.40	-0.73	-0.24	0.05	-0.18	-0.46	-1.13	-4.89	-0.65	1.79
F	0.93	-3.71	-3.42	-3.30	-0.12	-2.61	-0.37	0.82	-2.82	0.11	1.79
G	0.08	0.68	-0.69	0.14	0.00	0.27	0.09	0.30	-0.46	-0.30	1.79
H	0.44	0.03	0.54	-3.02	1.28	1.12	0.20	0.71	1.13	0.04	1.79
I	-0.36	0.12	0.19	0.94	-0.06	-0.80	-1.06	0.70	-3.98	-0.26	1.79
K	-1.05	-0.74	-3.42	-0.37	1.50	0.50	0.84	0.31	1.66	-2.82	1.79
L	-0.04	-0.15	-1.12	-0.05	0.18	0.63	-1.30	-0.27	-0.31	-0.13	1.79
M	-2.32	0.88	1.11	0.03	-0.47	-2.83	0.72	0.73	0.43	-2.60	1.79
N	-3.00	-0.38	-3.93	-0.81	0.03	-0.47	-0.12	-3.51	-4.04	-2.82	1.79
P	-0.02	-0.01	0.05	-0.54	-0.18	-0.04	-0.09	0.80	0.34	-1.78	1.79
Q	0.51	0.69	0.01	-0.40	0.78	-1.06	0.26	-1.56	-0.09	0.89	1.79
R	0.34	-0.22	1.22	0.69	-0.73	1.17	0.19	0.42	1.33	0.22	1.79
S	0.17	0.04	0.68	0.27	0.27	0.32	0.25	0.37	0.01	0.26	0.30
T	-0.89	-0.20	-0.23	-0.47	-0.24	-0.28	-0.71	0.01	-0.43	0.04	-0.34
V	0.62	0.04	0.18	0.61	-0.04	-4.15	0.26	0.16	0.02	1.27	1.79
W	-1.25	-3.79	-2.33	1.78	-3.41	-2.83	-1.26	-2.60	-1.92	-2.32	1.79
Y	-3.40	0.88	0.29	1.38	0.55	-3.01	-2.61	0.73	-0.24	0.22	-1.25

Amino Acid	Alignment Position									
	1	2	3	4	5	6	7	8	9	10
A	0.46	0.43	1.05	0.05	0.54	0.42	0.02	0.34	0.53	-0.12
C	-3.91	-1.25	-0.11	-1.26	-1.26	0.73	-0.10	-1.90	0.22	-3.27
D	0.71	-0.89	-0.73	-1.26	-0.68	0.02	-1.20	-1.40	-0.18	-4.59
E	0.37	-0.43	-0.13	-0.83	-4.21	-0.70	0.15	-0.07	-1.03	-0.88
F	-2.60	-0.36	-0.87	-3.01	0.44	-0.46	-3.50	0.23	-3.83	-2.99
G	-0.93	0.42	-0.68	0.15	-0.15	0.28	-0.13	0.03	-0.17	0.80
H	1.13	0.82	-2.82	-0.47	0.03	0.51	-0.46	-2.30	-0.09	-0.34
I	0.22	1.11	0.56	0.89	0.76	0.32	0.43	0.06	0.53	-0.04
K	0.99	-0.11	-2.32	1.64	0.06	0.56	0.03	0.54	0.53	0.62
L	-0.56	-1.10	-0.20	-0.39	-0.11	-4.29	-0.70	0.12	-4.40	0.94
M	-2.60	-0.11	-2.82	-3.16	0.03	-3.15	0.23	0.33	1.31	-2.58
N	0.89	-3.69	0.82	-3.61	-0.73	0.05	-3.00	-0.22	-0.85	0.44
P	-1.99	-0.62	-1.22	-0.86	0.12	-0.22	-0.79	-0.71	0.10	0.14
Q	-0.20	0.31	1.44	-0.20	-0.06	0.33	0.71	0.79	1.35	-0.12
R	1.57	0.62	0.04	-0.56	0.54	0.25	0.02	-0.11	-0.18	-0.52
S	0.16	0.41	0.08	0.95	-0.50	0.42	0.77	0.49	0.31	-0.63
T	0.16	-0.39	0.53	0.29	-0.50	-1.62	0.02	-0.18	-0.09	0.26
V	0.15	-0.82	0.79	-0.09	0.51	0.11	0.33	-0.73	-0.18	0.72
W	-2.32	-1.92	1.79	1.79	-1.26	-2.31	-1.91	-1.90	-2.80	-1.90
Y	1.13	-2.60	0.56	1.11	0.72	0.43	1.52	-0.22	-0.09	0.22

表 A-5. Xyl 修飾位置と Fuc 修飾位置を判別するための PSSM (第 6 章)

Amino Acid	Alignment Position										
	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0
A	0.00	-0.42	-3.44	1.34	-2.38	-3.05	0.67	0.67	1.34	0.45	0.67
C	-4.90	0.67	0.67	-2.38	0.67	-5.10	0.67	-5.33	0.67	0.67	0.67
D	0.67	0.00	4.38	0.00	1.73	-3.05	3.71	5.28	0.67	0.67	0.67
E	0.67	0.27	1.73	0.67	4.38	1.73	1.90	5.62	5.46	3.71	0.67
F	0.67	4.38	3.71	4.38	0.00	0.67	-3.05	3.71	0.67	0.67	0.67
G	2.02	1.34	-1.37	1.73	0.67	-0.90	3.71	4.78	-4.90	0.16	0.67
H	-3.05	0.67	-2.38	0.67	0.67	0.67	0.67	-3.44	0.67	0.67	0.67
I	0.67	-2.38	3.71	0.67	0.67	0.67	-2.38	0.67	0.67	0.67	0.67
K	-2.38	-2.38	3.71	-2.38	0.67	0.67	-2.38	-2.38	0.67	0.67	0.67
L	0.00	0.27	3.71	1.73	-0.01	4.38	-1.08	0.67	4.38	0.67	0.67
M	0.67	0.67	-2.38	3.71	4.38	0.67	0.67	0.67	4.38	0.67	0.67
N	3.71	-2.38	0.67	-1.08	0.67	3.71	0.67	-4.82	0.67	0.67	0.67
P	-0.68	3.71	0.27	-4.53	-0.62	3.71	-1.59	0.67	0.67	0.67	0.67
Q	-2.38	-2.38	0.67	1.34	0.67	3.71	-0.90	0.67	-2.38	0.67	0.67
R	1.34	3.71	3.71	-2.38	-3.73	3.71	-3.05	3.71	0.67	-3.44	0.67
S	0.67	0.67	-1.93	-4.13	0.67	-0.90	-0.40	0.67	-2.12	0.67	0.92
T	1.34	-3.05	-3.05	-3.44	4.38	0.67	4.38	3.71	-2.38	0.67	-5.42
V	3.71	0.27	4.38	0.67	0.67	0.67	-2.38	0.67	3.71	-4.64	0.67
W	-3.44	-1.59	0.67	0.67	-4.73	3.71	0.67	0.67	3.71	0.67	0.67
Y	0.67	0.67	-2.38	0.67	3.71	3.71	0.67	0.67	0.67	3.71	0.67

Amino Acid	Alignment Position									
	1	2	3	4	5	6	7	8	9	10
A	0.67	3.71	0.67	0.67	1.34	-3.44	3.71	1.34	-2.38	-2.38
C	-5.89	-2.38	0.67	0.67	0.67	0.67	0.67	-2.38	3.71	-1.59
D	0.67	2.24	-1.49	1.57	0.67	-2.38	3.71	4.38	1.34	0.00
E	0.67	1.73	1.34	0.67	0.00	4.38	-0.01	1.75	1.34	2.02
F	0.67	0.67	0.67	3.71	4.38	0.67	-0.40	0.67	0.67	-3.05
G	6.56	-2.06	0.20	-2.28	1.34	-0.68	-0.01	-2.38	0.00	5.06
H	0.67	-3.44	0.67	-2.38	0.67	-3.05	3.71	0.67	-2.38	0.67
I	0.67	-0.40	-2.38	0.67	-0.42	-2.38	0.67	0.67	-3.44	0.67
K	0.67	-3.44	-2.38	0.67	-2.38	3.71	1.34	3.71	-3.73	-2.38
L	-2.38	1.06	-3.95	-3.05	-1.08	1.34	0.00	0.67	1.06	0.00
M	0.67	0.67	3.71	-3.05	0.67	0.67	-2.38	0.67	0.67	0.67
N	0.67	0.67	0.27	0.67	0.67	-3.05	0.67	0.67	-3.05	0.67
P	0.67	0.67	4.38	4.38	0.27	3.71	4.38	-0.40	1.06	0.67
Q	0.67	0.00	0.00	0.67	-2.38	-4.41	0.67	0.67	-3.05	3.71
R	0.67	0.67	0.00	0.67	0.67	-0.90	-0.68	-4.90	-0.90	-4.53
S	0.67	0.67	-3.05	-3.05	1.34	0.67	-4.13	0.27	-3.44	4.38
T	0.67	0.67	3.71	4.38	-0.24	1.34	-0.42	0.00	1.06	3.71
V	0.67	-3.05	-2.38	0.67	-4.13	1.06	3.71	0.67	0.67	-0.90
W	0.67	3.71	0.67	0.67	3.71	0.67	-2.38	0.67	0.67	0.67
Y	0.67	0.67	3.71	0.67	-3.05	0.00	-3.44	-4.28	-3.44	-2.38

表 A-6. GalNAc 修飾位置と Fuc 修飾位置を判別するための PSSM (第 6 章)

Amino Acid	Alignment Position										
	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0
A	0.41	-0.53	0.19	0.99	1.29	0.32	0.44	4.19	1.43	-0.36	-2.11
C	-2.86	2.52	2.52	-0.19	2.85	-3.26	2.01	-4.39	2.00	-2.11	-2.11
D	3.47	0.13	3.54	-0.25	0.69	0.07	3.89	3.73	-0.34	2.00	-2.11
E	1.23	-0.16	0.85	0.18	4.08	1.07	0.44	4.19	3.98	3.09	-2.11
F	-0.52	2.85	2.70	2.85	-1.42	-0.75	-0.73	1.61	2.29	2.69	-2.11
G	0.63	0.57	-1.29	0.69	0.56	-1.15	3.84	3.46	-1.68	-1.73	-2.11
H	-0.61	2.70	-0.19	2.70	2.85	2.29	0.05	-1.60	0.94	2.51	-2.11
I	2.70	0.34	2.99	2.85	3.30	3.20	0.42	2.98	3.53	3.60	-2.11
K	0.91	0.63	3.21	0.26	3.47	3.79	0.62	-0.07	3.60	2.84	-2.11
L	0.75	-0.38	3.85	0.80	-0.50	3.94	-0.61	4.03	4.07	3.38	-2.11
M	2.02	2.85	-1.43	2.52	2.29	-1.04	0.94	2.00	2.51	1.60	-2.11
N	2.99	0.26	-0.06	-1.69	2.85	2.98	2.51	-2.03	3.53	2.69	-2.11
P	-0.12	4.30	0.35	-1.12	-0.74	4.58	-0.65	4.72	1.65	5.36	-2.11
Q	-0.52	0.57	0.34	0.75	3.47	3.46	-1.08	3.89	0.94	2.97	-2.11
R	0.26	2.99	3.47	0.50	-1.55	3.46	-0.25	3.46	3.09	-1.61	-2.11
S	0.40	0.71	-1.21	-0.34	4.19	-0.58	-0.38	4.03	-1.72	0.31	-0.23
T	1.12	0.37	0.68	0.70	4.30	0.22	4.03	3.73	0.48	1.24	0.15
V	3.21	-0.38	3.68	0.50	0.49	3.99	0.49	4.39	3.83	-1.05	-2.11
W	-3.16	-3.69	2.01	-2.10	-3.39	2.01	0.94	-2.11	-2.11	1.60	-2.11
Y	2.30	2.52	-0.19	0.95	2.29	2.51	2.29	2.00	-2.11	2.00	-2.11

Amino Acid	Alignment Position									
	1	2	3	4	5	6	7	8	9	10
A	4.15	4.58	0.98	0.80	1.04	-0.31	4.13	0.58	1.13	0.71
C	-4.95	-5.15	1.61	0.95	0.95	2.02	2.31	-2.08	3.00	-2.76
D	3.29	0.33	-2.36	-0.51	0.50	0.43	3.62	3.81	0.27	0.30
E	4.39	0.49	-0.07	1.11	0.33	3.74	-1.00	0.10	1.05	0.77
F	1.61	2.69	3.09	2.52	2.02	2.86	-1.12	-0.51	3.56	-1.68
G	3.38	-2.31	-1.70	-2.05	0.69	-0.54	-0.54	0.76	0.38	3.32
H	2.00	-1.82	2.29	0.05	2.86	-0.32	2.86	2.32	-0.33	3.23
I	2.51	-1.14	-0.35	2.70	-1.59	-0.34	2.86	2.32	-1.24	0.18
K	3.09	-0.39	0.05	2.98	0.75	3.68	0.58	3.81	-0.84	0.59
L	0.25	0.01	-1.32	0.44	-0.71	0.64	0.49	0.92	0.25	-0.23
M	2.00	2.69	2.00	-1.42	2.30	-0.74	-0.74	3.00	0.97	2.54
N	1.61	3.38	-1.42	-0.41	3.31	-1.19	2.86	2.54	-0.39	3.23
P	5.05	4.58	5.52	5.01	0.51	4.71	4.69	0.57	0.64	1.40
Q	3.53	-0.87	-0.74	-0.41	0.57	-1.18	0.35	3.41	-0.71	3.41
R	3.20	3.67	-1.71	2.98	3.68	-0.76	-0.54	-2.08	-1.50	-1.88
S	4.29	0.58	0.73	0.32	1.19	0.06	-0.55	-0.20	-0.15	4.29
T	4.60	4.22	4.47	4.08	-0.31	1.23	-0.67	0.94	0.87	4.36
V	3.89	0.36	0.25	0.99	-1.40	0.03	3.80	0.97	-0.49	-0.40
W	2.00	2.29	2.00	-2.10	-2.09	0.95	-5.13	2.03	2.32	2.03
Y	1.61	2.51	2.51	2.52	-2.76	-1.19	-3.15	-2.41	-1.39	-0.04

表 A-7. Xyl 修飾位置と GalNAc 修飾位置を判別するための PSSM (第 6 章)

Amino Acid	Alignment Position										
	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0
A	-0.41	0.11	-3.64	0.34	-3.67	-3.36	0.23	-3.52	-0.09	0.81	2.78
C	-2.04	-1.85	-1.85	-2.19	-2.18	-1.85	-1.34	-0.94	-1.34	2.78	2.78
D	-2.81	-0.14	0.84	0.24	1.04	-3.12	-0.18	1.56	1.00	-1.34	2.78
E	-0.56	0.43	0.88	0.48	0.30	0.66	1.46	1.43	1.48	0.62	2.78
F	1.19	1.53	1.01	1.53	1.42	1.42	-2.31	2.10	-1.62	-2.02	2.78
G	1.38	0.77	-0.08	1.04	0.10	0.25	-0.13	1.32	-3.22	1.89	2.78
H	-2.44	-2.03	-2.19	-2.03	-2.18	-1.63	0.62	-1.84	-0.27	-1.84	2.78
I	-2.04	-2.72	0.73	-2.19	-2.63	-2.53	-2.80	-2.31	-2.86	-2.93	2.78
K	-3.28	-3.01	0.50	-2.64	-2.80	-3.12	-3.00	-2.31	-2.93	-2.17	2.78
L	-0.75	0.64	-0.14	0.93	0.49	0.44	-0.48	-3.36	0.31	-2.71	2.78
M	-1.35	-2.19	-0.95	1.19	2.09	1.70	-0.28	-1.34	1.87	-0.94	2.78
N	0.72	-2.64	0.73	0.61	-2.18	0.73	-1.85	-2.79	-2.86	-2.02	2.78
P	-0.56	-0.59	-0.08	-3.41	0.12	-0.87	-0.94	-4.05	-0.98	-4.70	2.78
Q	-1.86	-2.94	0.32	0.59	-2.80	0.25	0.18	-3.22	-3.31	-2.31	2.78
R	1.07	0.73	0.24	-2.88	-2.18	0.25	-2.80	0.25	-2.42	-1.84	2.78
S	0.26	-0.05	-0.72	-3.79	-3.53	-0.32	-0.02	-3.36	-0.40	0.36	1.15
T	0.22	-3.41	-3.73	-4.15	0.08	0.45	0.35	-0.01	-2.86	-0.58	-5.58
V	0.50	0.64	0.70	0.17	0.17	-3.32	-2.87	-3.72	-0.12	-3.59	2.78
W	-0.29	2.09	-1.35	2.76	-1.34	1.70	-0.28	2.77	5.82	-0.94	2.78
Y	-1.64	-1.85	-2.19	-0.28	1.42	1.20	-1.63	-1.34	2.78	1.71	2.78

Amino Acid	Alignment Position									
	1	2	3	4	5	6	7	8	9	10
A	-3.48	-0.87	-0.32	-0.13	0.30	-3.13	-0.42	0.75	-3.51	-3.09
C	-0.94	2.77	-0.94	-0.28	-0.29	-1.35	-1.64	-0.30	0.71	1.17
D	-2.63	1.90	0.87	2.08	0.17	-2.81	0.09	0.57	1.06	-0.30
E	-3.72	1.25	1.40	-0.45	-0.33	0.64	0.98	1.65	0.28	1.25
F	-0.94	-2.02	-2.43	1.20	2.36	-2.19	0.72	1.18	-2.89	-1.37
G	3.18	0.25	1.90	-0.23	0.64	-0.14	0.52	-3.14	-0.39	1.74
H	-1.34	-1.62	-1.62	-2.43	-2.19	-2.73	0.85	-1.65	-2.05	-2.56
I	-1.84	0.74	-2.02	-2.03	1.17	-2.04	-2.20	-1.65	-2.20	0.48
K	-2.43	-3.06	-2.43	-2.32	-3.13	0.03	0.76	-0.10	-2.89	-2.96
L	-2.63	1.05	-2.63	-3.49	-0.37	0.70	-0.50	-0.25	0.81	0.22
M	-1.34	-2.02	1.71	-1.63	-1.64	1.41	-1.64	-2.34	-0.30	-1.87
N	-0.94	-2.71	1.69	1.08	-2.64	-1.86	-2.20	-1.87	-2.65	-2.56
P	-4.38	-3.91	-1.14	-0.63	-0.24	-1.00	-0.31	-0.97	0.43	-0.74
Q	-2.87	0.87	0.74	1.08	-2.95	-3.24	0.31	-2.74	-2.34	0.30
R	-2.53	-3.00	1.71	-2.32	-3.01	-0.14	-0.15	-2.82	0.59	-2.66
S	-3.63	0.09	-3.78	-3.37	0.14	0.61	-3.58	0.47	-3.30	0.09
T	-3.94	-3.56	-0.76	0.30	0.07	0.10	0.25	-0.94	0.20	-0.64
V	-3.22	-3.40	-2.63	-0.32	-2.73	1.03	-0.09	-0.30	1.16	-0.51
W	-1.34	1.42	-1.34	2.77	5.80	-0.29	2.75	-1.36	-1.65	-1.37
Y	-0.94	-1.84	1.20	-1.85	-0.29	1.19	-0.29	-1.87	-2.05	-2.34

表 A-8. GlcNAc 修飾位置と Fuc 修飾位置を判別するための PSSM (第 6 章)

Amino Acid	Alignment Position										
	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0
A	0.34	-0.40	0.18	1.13	0.99	0.57	0.85	3.47	1.60	-0.14	-0.92
C	-2.75	2.13	-0.91	-0.24	-0.91	-6.68	2.12	-3.20	2.12	-0.92	-0.92
D	2.14	-0.91	2.80	-1.58	-0.24	-0.51	3.87	3.19	-0.52	-0.92	-0.92
E	0.45	-0.63	0.66	-0.23	3.70	0.99	0.31	3.69	2.79	3.19	-0.92
F	0.85	2.80	2.13	2.13	-0.51	0.16	-0.92	4.03	2.12	3.47	-0.92
G	1.14	1.13	-2.28	0.99	0.66	-0.44	3.69	3.19	-1.40	-1.68	-0.92
H	-1.57	2.13	-0.91	2.80	3.48	-0.91	0.14	-1.32	2.12	2.79	-0.92
I	3.90	0.44	3.20	3.88	2.80	3.70	-0.92	4.03	3.19	3.87	-0.92
K	-0.23	-0.91	3.48	1.13	4.04	3.70	0.98	0.43	4.48	2.79	-0.92
L	0.58	-0.63	3.20	1.13	-0.11	3.88	-1.32	4.03	3.47	3.19	-0.92
M	2.81	3.20	-0.24	2.80	3.48	-3.96	3.19	2.12	3.19	2.12	-0.92
N	3.21	0.66	-0.24	-2.66	3.48	2.13	2.12	-3.36	2.12	-0.92	-0.92
P	-0.10	4.29	0.38	-1.49	-1.11	4.17	-0.74	4.97	1.67	3.19	-0.92
Q	0.45	0.44	0.16	-0.24	3.70	3.70	-1.43	2.79	0.14	3.47	-0.92
R	0.17	3.70	3.88	0.44	-1.19	4.65	-0.92	2.79	4.03	-1.99	-0.92
S	0.55	0.86	-0.56	-0.14	4.39	-0.13	0.17	3.87	-1.67	0.76	0.46
T	0.17	0.17	0.57	-0.07	3.48	-0.23	3.69	4.16	0.14	1.52	-0.49
V	4.05	-0.07	3.70	0.84	0.99	2.13	1.12	4.16	3.69	-0.18	-0.92
W	-5.01	-6.21	-0.91	2.13	-6.31	-0.91	-0.92	-0.92	-0.92	-0.92	-0.92
Y	-0.90	3.20	0.16	3.20	2.13	-0.91	2.12	2.79	2.12	2.12	2.79

Amino Acid	Alignment Position									
	1	2	3	4	5	6	7	8	9	10
A	4.48	4.64	1.12	0.43	1.35	-0.40	3.21	0.67	1.36	0.85
C	-4.43	-0.92	-0.92	-0.92	2.13	2.81	2.14	-0.90	2.14	-3.16
D	3.47	-0.25	-2.01	-1.59	0.66	0.45	3.72	3.21	0.67	-0.50
E	4.16	0.65	0.43	0.43	-4.62	3.90	-0.35	0.00	1.00	0.67
F	2.12	2.79	3.19	2.79	3.48	2.81	-1.30	0.67	-0.90	-4.61
G	4.03	-1.29	-1.61	-1.72	1.13	0.01	-1.18	0.17	0.00	4.05
H	-0.92	-0.92	-0.92	-0.92	2.13	-0.22	2.14	2.81	-0.90	2.14
I	2.79	-0.24	1.12	2.12	-0.76	-0.23	3.21	3.72	-0.40	0.85
K	3.69	-1.32	-3.97	4.03	0.84	3.72	0.45	4.05	-0.68	0.67
L	-0.25	-0.24	-0.74	-0.02	-0.76	-0.90	0.00	0.67	-0.50	0.47
M	2.12	2.12	-0.92	-4.64	3.20	-3.94	0.45	2.81	3.21	-0.90
N	3.47	2.79	-0.64	-1.59	2.80	-1.57	3.49	2.14	-4.61	3.21
P	3.19	4.16	3.87	4.38	0.06	4.40	3.49	-0.06	0.69	1.00
Q	3.47	-0.52	0.31	-0.02	0.44	-0.90	0.45	4.18	0.18	3.21
R	4.28	3.47	-0.24	3.19	3.70	-1.12	-0.90	-2.35	-1.40	-1.99
S	4.38	0.93	1.14	1.14	0.16	0.47	-0.13	0.47	0.29	4.18
T	4.56	4.03	4.38	4.56	-0.58	0.45	-0.30	0.47	0.47	4.40
V	4.03	-0.52	0.83	1.12	-0.76	0.47	4.18	0.45	0.00	-0.21
W	2.12	2.12	-0.92	2.12	-0.91	-0.90	-3.94	-0.90	-0.90	-0.90
Y	-0.92	2.12	3.19	2.79	-0.51	-0.90	-0.62	-1.45	-0.90	0.45

表 A-9. GlcNAc 修飾位置と Xyl 修飾位置を判別するための PSSM (第 6 章)

Amino Acid	Alignment Position										
	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0
A	0.34	0.01	3.62	-0.21	3.37	3.62	0.18	2.81	0.26	-0.58	-1.59
C	2.15	1.47	-1.58	2.14	-1.58	-1.58	1.46	2.13	1.46	-1.59	-1.59
D	1.48	-0.91	-1.58	-1.58	-1.97	2.53	0.16	-2.09	-1.19	-1.59	-1.59
E	-0.22	-0.90	-1.07	-0.90	-0.68	-0.74	-1.59	-1.92	-2.67	-0.52	-1.59
F	0.18	-1.58	-1.58	-2.25	-0.51	-0.51	2.13	0.32	1.46	2.81	-1.59
G	-0.88	-0.21	-0.91	-0.74	-0.01	0.46	-0.02	-1.59	3.49	-1.84	-1.59
H	1.48	1.47	1.47	2.14	2.82	-1.58	-0.52	2.13	1.46	2.13	-1.59
I	3.23	2.82	-0.51	3.22	2.14	3.04	1.46	3.36	2.52	3.21	-1.59
K	2.15	1.47	-0.23	3.50	3.37	3.04	3.36	2.81	3.81	2.13	-1.59
L	0.59	-0.90	-0.51	-0.61	-0.09	-0.50	-0.24	3.36	-0.91	2.52	-1.59
M	2.15	2.53	2.14	-0.91	-0.90	-4.62	2.52	1.46	-1.19	1.46	-1.59
N	-0.50	3.04	-0.91	-1.58	2.82	-1.58	1.46	1.46	1.46	-1.59	-1.59
P	0.59	0.58	0.11	3.04	-0.49	0.46	0.85	4.30	1.01	2.52	-1.59
Q	2.83	2.82	-0.51	-1.58	3.04	-0.01	-0.52	2.13	2.52	2.81	-1.59
R	-1.17	-0.01	0.17	2.82	2.53	0.94	2.13	-0.92	3.36	1.46	-1.59
S	-0.11	0.19	1.37	3.99	3.73	0.78	0.57	3.21	0.45	0.10	-0.46
T	-1.17	3.22	3.62	3.37	-0.90	-0.89	-0.69	0.45	2.52	0.85	4.94
V	0.34	-0.34	-0.68	0.17	0.33	1.47	3.49	3.49	-0.02	4.45	-1.59
W	-1.57	-4.62	-1.58	1.47	-1.58	-4.62	-1.59	-1.59	-4.63	-1.59	-1.59
Y	-1.57	2.53	2.53	2.53	-1.58	-4.62	1.46	2.13	1.46	-1.59	2.13

Amino Acid	Alignment Position									
	1	2	3	4	5	6	7	8	9	10
A	3.81	0.93	0.45	-0.24	0.01	3.05	-0.50	-0.66	3.74	3.23
C	1.46	1.46	-1.59	-1.59	1.47	2.15	1.48	1.48	-1.57	-1.57
D	2.81	-2.49	-0.52	-3.16	-0.01	2.83	0.00	-1.17	-0.66	-0.50
E	3.49	-1.08	-0.91	-0.24	-4.62	-0.48	-0.33	-1.75	-0.33	-1.35
F	1.46	2.13	2.52	-0.92	-0.90	2.15	-0.90	0.00	-1.57	-1.57
G	-2.53	0.77	-1.81	0.57	-0.21	0.69	-1.17	2.54	0.00	-1.01
H	-1.59	2.52	-1.59	1.46	1.47	2.83	-1.57	2.15	1.48	1.48
I	2.13	0.16	3.49	1.46	-0.34	2.15	2.54	3.05	3.05	0.18
K	3.03	2.13	-1.59	3.36	3.22	0.00	-0.89	0.34	3.05	3.05
L	2.13	-1.30	3.21	3.03	0.33	-2.24	0.00	0.00	-1.57	0.47
M	1.46	1.46	-4.63	-1.59	2.53	-4.61	2.83	2.15	2.54	-1.57
N	2.81	2.13	-0.91	-2.26	2.14	1.48	2.83	1.48	-1.57	2.54
P	2.52	3.49	-0.51	0.00	-0.21	0.69	-0.89	0.34	-0.37	0.34
Q	2.81	-0.52	0.32	-0.69	2.82	3.51	-0.22	3.51	3.23	-0.50
R	3.61	2.81	-0.24	2.52	3.04	-0.22	-0.22	2.54	-0.50	2.54
S	3.72	0.26	4.18	4.18	-1.18	-0.19	4.00	0.20	3.74	-0.20
T	3.90	3.36	0.67	0.18	-0.34	-0.89	0.12	0.47	-0.60	0.69
V	3.36	2.52	3.21	0.45	3.37	-0.60	0.47	-0.22	-0.66	0.69
W	1.46	-1.59	-1.59	1.46	-4.62	-1.57	-1.57	-1.57	-1.57	-1.57
Y	-1.59	1.46	-0.52	2.13	2.53	-0.90	2.83	2.83	2.54	2.83

表 A-10. GlcNAc 修飾位置と GalNAc 修飾位置を判別するための PSSM (第 6 章)

Amino Acid	Alignment Position										
	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0
A	-0.07	0.12	-0.01	0.13	-0.29	0.26	0.41	-0.71	0.17	0.22	1.19
C	0.11	-0.39	-3.43	-0.05	-3.76	-3.42	0.11	1.18	0.12	1.19	1.19
D	-1.33	-1.04	-0.74	-1.33	-0.93	-0.59	-0.02	-0.54	-0.19	-2.92	1.19
E	-0.78	-0.46	-0.19	-0.41	-0.37	-0.08	-0.13	-0.49	-1.19	0.10	1.19
F	1.37	-0.05	-0.57	-0.72	0.91	0.91	-0.19	2.42	-0.16	0.79	1.19
G	0.50	0.56	-0.99	0.30	0.10	0.71	-0.15	-0.27	0.28	0.05	1.19
H	-0.96	-0.57	-0.72	0.10	0.63	-3.20	0.09	0.28	1.19	0.29	1.19
I	1.19	0.09	0.21	1.03	-0.50	0.50	-1.34	1.05	-0.34	0.28	1.19
K	-1.14	-1.54	0.28	0.87	0.57	-0.08	0.36	0.50	0.88	-0.05	1.19
L	-0.16	-0.25	-0.65	0.32	0.39	-0.05	-0.72	0.00	-0.59	-0.19	1.19
M	0.80	0.35	1.19	0.28	1.19	-2.92	2.25	0.12	0.68	0.52	1.19
N	0.22	0.40	-0.18	-0.97	0.63	-0.84	-0.39	-1.34	-1.41	-3.61	1.19
P	0.03	-0.01	0.03	-0.38	-0.37	-0.41	-0.09	0.25	0.02	-2.18	1.19
Q	0.97	-0.13	-0.19	-0.99	0.24	0.24	-0.35	-1.10	-0.79	0.50	1.19
R	-0.09	0.72	0.42	-0.06	0.35	1.19	-0.67	-0.67	0.94	-0.38	1.19
S	0.15	0.15	0.66	0.20	0.20	0.46	0.55	-0.15	0.05	0.45	0.69
T	-0.95	-0.19	-0.11	-0.77	-0.82	-0.44	-0.34	0.43	-0.34	0.27	-0.64
V	0.84	0.30	0.03	0.34	0.50	-1.85	0.62	-0.23	-0.14	0.87	1.19
W	-1.85	-2.53	-2.93	4.23	-2.92	-2.92	-1.86	1.18	1.19	-2.53	1.19
Y	-3.20	0.68	0.35	2.25	-0.16	-3.42	-0.17	0.79	4.23	0.12	4.90

Amino Acid	Alignment Position									
	1	2	3	4	5	6	7	8	9	10
A	0.33	0.06	0.13	-0.37	0.31	-0.08	-0.92	0.09	0.23	0.14
C	0.52	4.23	-2.53	-1.87	1.18	0.79	-0.16	1.18	-0.86	-0.39
D	0.18	-0.59	0.35	-1.08	0.16	0.02	0.09	-0.60	0.40	-0.80
E	-0.23	0.16	0.50	-0.68	-4.95	0.15	0.65	-0.09	-0.05	-0.10
F	0.52	0.10	0.10	0.28	1.47	-0.05	-0.18	1.18	-4.46	-2.93
G	0.65	1.02	0.09	0.33	0.43	0.55	-0.65	-0.60	-0.38	0.73
H	-2.93	0.90	-3.21	-0.98	-0.72	0.10	-0.72	0.50	-0.57	-1.08
I	0.28	0.90	1.47	-0.57	0.83	0.11	0.35	1.40	0.85	0.67
K	0.60	-0.93	-4.01	1.05	0.09	0.03	-0.13	0.24	0.16	0.08
L	-0.50	-0.25	0.58	-0.46	-0.05	-1.54	-0.49	-0.25	-0.75	0.69
M	0.12	-0.57	-2.93	-3.22	0.90	-3.20	1.19	-0.19	2.25	-3.44
N	1.87	-0.59	0.78	-1.18	-0.50	-0.38	0.63	-0.39	-4.22	-0.02
P	-1.86	-0.42	-1.64	-0.63	-0.45	-0.31	-1.20	-0.63	0.05	-0.40
Q	-0.06	0.35	1.05	0.39	-0.13	0.28	0.10	0.78	0.89	-0.20
R	1.08	-0.19	1.47	0.21	0.03	-0.36	-0.36	-0.27	0.09	-0.11
S	0.09	0.35	0.40	0.82	-1.04	0.42	0.42	0.67	0.44	-0.11
T	-0.04	-0.20	-0.09	0.49	-0.27	-0.78	0.37	-0.47	-0.40	0.05
V	0.14	-0.88	0.58	0.13	0.65	0.44	0.38	-0.52	0.49	0.19
W	0.12	-0.17	-2.93	4.22	1.18	-1.85	1.19	-2.93	-3.21	-2.93
Y	-2.53	-0.39	0.68	0.28	2.25	0.29	2.54	0.96	0.49	0.49

