

進化的分類による*O*-GlcNAc修飾タンパク質の解析

田中純¹⁾、藤井正興¹⁾、植田竜太²⁾、小島寿夫³⁾、伊藤将弘^{1,3)}

Evolutionary analysis for *O*-GlcNAcylated proteins by clustering method

Jun Tanaka¹⁾, Masaoki Fujii¹⁾, Ryuta Ueda²⁾, Hisao Kojima³⁾, Masahiro Ito^{1, 3)}

O-GlcNAcylation, the attachment of a single *N*-acetylglucosamine to serine/threonine residues of proteins by *O*-GlcNAc transferase (OGT), is essential for protein function in eukaryotes. OGT was shown to be essential for the viability of embryonic stem cells. In spite of these important roles, no consensus sequence for *O*-GlcNAcylation had been found so far. There are many parallels between phosphorylation and *O*-GlcNAcylation; attachment occurs to serine/threonine residues, half-life is short, occurrence is infrequent in structured regions, whereas dominant in intrinsically disordered regions (IDRs).

In this study, we clustered *O*-GlcNAcylated proteins to investigate the relation between their evolutionary conservation and function for proteins. We used the experimental data of human *O*-GlcNAcylated proteins. Firstly, KEGG ortholog cluster was used for homology search in eukaryotes with genomes completely sequenced. Secondly, *O*-GlcNAcylated residues in orthologous proteins were checked for conservation using multiple sequence alignment. Finally, *O*-GlcNAcylated proteins were clustered by the degree of their conservation.

In the clustering result, *O*-GlcNAcylated proteins clustered into 3 clusters, Classes 1 to 3. In addition, each cluster further clustered into 3 subclusters, like subclasses 1-1 to 1-3. Both ratios of *O*-GlcNAcylation within IDRs and co-occurrence of *O*-GlcNAcylation/phosphorylation at identical region were high in evolutionarily young clusters. Comparison of the degree of conservation in OGT and *O*-GlcNAcase (OGA), the degrading enzyme of *O*-GlcNAc, showed that OGT was more widely conserved than OGA. In addition, approximately half of *O*-GlcNAcylated proteins were as conserved as OGA. These results suggest that a lot of *O*-GlcNAcylated proteins were acquired along with OGA.

Keywords: *O*-GlcNAc, Intrinsically disordered protein, phosphorylation, PTM

Email: maito@sk.ritsumei.ac.jp (M. Ito)

1) 立命館大学大学院生命科学研究科生命科学専攻生命情報科学コース

2) 立命館大学大学院テクノロジーマネジメント研究科

3) 立命館大学生命科学部生命情報学科

1) Dept. of Bioinfo., Grad. Sch. of Life Sci., Ritsumeikan Univ.

1-1-1 Noji-higashi, Kusatsu, Shiga 525-8577, Japan

2) Grad. Sch. of Technol. Mgmt., Ritsumeikan Univ.

2-150 Iwakura-cho, Ibaraki, Osaka 567-8570, Japan

3) Dept. of Bioinfo., Coll. of Life Sci., Ritsumeikan Univ.

1-1-1 Noji-higashi, Kusatsu, Shiga 525-8577, Japan

1. 緒言

タンパク質のセリンまたはスレオニンに単一の β -*N* アセチルグルコサミンが結合する *O*-GlcNAc 修飾は後生生物と植物の細胞内において必要不可欠な翻訳後修飾である[1,2]。 *O*-GlcNAc 修飾は主に核・細胞質で起こり、特に核に豊富である。 *O*-GlcNAc 修飾タンパク質の機能は代謝、シグナル伝達、転写および翻訳、さらにストレス応答など様々である[3]。このように *O*-GlcNAc 修飾はリン酸化のような可逆的な翻訳後修飾である[4]。 *O*-GlcNAc 修飾を受けるアミノ酸残基は、同一または近傍の残基がリン酸化されるなどリン酸化と広く相互作用して生命現象を制御している[5]。しかし、何百種類ものキナーゼやホスファターゼによって行われるリン酸化[6]とは対照的に、 *O*-GlcNAc 修飾は *O*-GlcNAc の付加を触媒する *O*-GlcNAc 転移酵素

(OGT) と付加された *O*-GlcNAc の切断を触媒する *O*-GlcNAc 分解酵素 (OGA) の 2 つの酵素によって調節されている[7,8]。 OGT の認識する配列には特徴が見られるが、絶対的なコンセンサス配列は見つかっておらず[9]、どのようにして OGT が修飾残基を選択しているのかについては不明な部分が多い。 OGT が多くの異なるタンパク質の修飾を行うことができる理由として、他の多様なタンパク質と結合することによるものであると示唆されており[10]、 *O*-GlcNAc 修飾を受ける配列にはいくつかのパターンがあると推察される。

天然変性タンパク質 (IDP) は、単独では立体構造を作らない天然変性領域 (ID 領域) を持つ[11,12]。 IDP は真核生物の核に豊富に存在し、転写、翻訳、および細胞周期制御を含む重要な細胞機能に関与している[13,14]。 また、ID 領域においてタンパク質のリン酸化は多く起こることが知られている[15]。

O-GlcNAc 修飾と IDP は細胞内局在および機能の面で共通点を有しており、ID 領域に多く見られるリン酸化との相互作用が知られている。 そこで本研究では、 *O*-GlcNAc 修飾の獲得の過程を明らかにすることを目的として *O*-GlcNAc 修飾を受ける配列を *O*-GlcNAc 修飾残基の保存度の観点からいくつかに分類し、 *O*-GlcNAc 修飾とリン酸化、ID 領域の関わりを解析した。

2. 方法

2-1. O-GlcNAc 修飾タンパク質の取得

O-GlcNAc 修飾タンパク質のデータは、OGTSite [16]より取得した。その中から、進化解析をヒトのタンパク質を基準として行うため、ヒトの O-GlcNAc タンパク質のみを対象とした。また、それぞれのタンパク質のアミノ酸配列データは UniProt データベース[17]から取得し、ID 領域に関するデータは DICHOT データベース[18]から、リン酸化残基に関するデータは PhosphoSitePlus [19]から取得した。

2-2. オーソログタンパク質の取得

O-GlcNAc 修飾タンパク質を進化的に解析するため、ヒトの O-GlcNAc 修飾タンパク質のオーソログを KEGG データベースのオーソログクラスター[20]より提供されているオーソログ情報をもとに取得した。生物種は KEGG Organisms の Complete Genomes 一覧 (http://www.kegg.jp/kegg/catalog/org_list.html) に登録されている生物種のうち、真核生物を対象とした。

2-3. OGT および OGA の保存確認

O-GlcNAc 修飾を触媒する酵素である OGT、O-GlcNAc を取り除く酵素である OGA に関しても O-GlcNAc 修飾タンパク質と同様に KEGG データベースのオーソログクラスターより提供されているオーソログ情報をもとにオーソログの有無を確認し、オーソログを持つ場合には1、持たない場合には0としてビット化した。

2-4. O-GlcNAc 修飾残基の保存

取得されたヒト O-GlcNAc 修飾タンパク質のオーソログタンパク質について O-GlcNAc 修飾残基が保存されているかを確認した。ヒトの O-GlcNAc 修飾タンパク質と他生物でのオーソログとのマルチプルアライメントを clustalw2 により行った。アライメント後に O-GlcNAc 修飾残基の保存度を解析する際、ヒトの O-GlcNAc 修飾タンパク質での修飾残基にあたるアミノ酸残基が他生物において他のアミノ酸に置換されていたもの、または欠損していたものは O-GlcNAc 修飾を受けないと仮定し、それ以外は O-GlcNAc 修飾を受けると仮定した。

2-5. 進化情報によるクラスタリング

まず、進化情報に基づいて図1のようにスコア付けを実施した。図1Aのように縦にヒト O-GlcNAc 修飾タンパク質（例ではタンパク質 A,B,C,D）、横に対象とした生物種を並べ（例では生物 1,2,3）、対応する生物におけるタンパク質の保存度によりスコアを与えた。スコアの定義として、ヒト O-GlcNAc 修飾タンパク質のオーソログが該当の生物で確認できなかったものは-1、タンパク質のオーソログが該当の生物で確認できたものについては式（1）により算出した。各ヒト O-GlcNAc 修飾タンパク質のスコアのパターンに対してワード法を用いることでヒト O-GlcNAc 修飾タンパク質を分類した。

$$\text{スコア} = \frac{\text{オーソログ上で保存されたO-GlcNAc 修飾残基数}}{\text{クエリタンパク質上のO-GlcNAc 修飾残基数}} \quad (1)$$

(A)

	ヒト	生物1	生物2	生物3
タンパク質A 				×
タンパク質B 				×
タンパク質C 				×
タンパク質D 			×	

(B)

	ヒト	生物1	生物2	生物3
タンパク質A	1	1	0.5	-1
タンパク質B	1	1	1	-1
タンパク質C	1	0	0	-1
タンパク質D	1	1	-1	0

図 1. 進化情報によるスコア付け方法

(A) オーソログタンパク質において *O*-GlcNAc 修飾残基の保存度が異なる 4 種類のタンパク質の例。各マス内の横長の図形はタンパク質の 1 次構造を表し、図形上のローマ字はアミノ酸残基を表し、ヒトの列における S および T は *O*-GlcNAc 修飾される残基、他の生物種の列ではヒトの *O*-GlcNAc 修飾残基にあたる残基を表す。×はオーソログが存在しないことを表す。(B) (A)で示された四種類のタンパク質における算出スコアおよびスコアを基にしたビットパターンをクラスタリングすることで得られる系統樹。オーソログタンパク質が存在し、かつ全ての *O*-GlcNAc 修飾残基が保存されている場合（例：生物 1 のタンパク質 A）のスコアは 1 で、オーソログタンパク質が存在するものの半分の *O*-GlcNAc 修飾残基が置換されている場合（例：生物 2 のタンパク質 A）のスコアは 0.5、オーソログタンパク質が存在するもののどの *O*-GlcNAc 修飾残基も保存されていない場合（例：生物 1 のタンパク質 C）はスコア 0、オーソログタンパク質が存在しない場合はスコア -1 とした。また、スコアパターンを見やすくするためにスコア 1 の場合には黒色、スコア 0.5 なら薄い黒、スコア 0 は灰色、スコア -1 は白色のようにヒートマップで表示した。

3. 結果と考察

3-1. ヒトの O-GlcNAc 修飾タンパク質の取得

OGTSite より提供されているタンパク質から、ヒトのタンパク質が 162 個取得された。取得された 162 の O-GlcNAc 修飾タンパク質は 408 個の O-GlcNAc 修飾残基を含み、また 162 個すべてのタンパク質がリン酸化されることが知られているタンパク質であった。ID 領域の割合は平均で約 49.6%であったが、O-GlcNAc 修飾を受ける 408 個のアミノ酸残基のうち 319 個(78.2%)が ID 領域内に位置していた。このことから、O-GlcNAc 修飾は ID 領域に多く起こることが示唆された。これは O-GlcNAc 修飾と相互作用することが知られているリン酸化と同様の傾向である。O-GlcNAc を付加する酵素である OGT の立体構造解析では、OGT と O-GlcNAc 修飾を受けるペプチドの立体構造が解明されているが、O-GlcNAc 修飾を受けるペプチドは伸びた構造を取っている[21]。これらより、ID 領域の方が OGT との結合が起りやすいために ID 領域に O-GlcNAc 修飾が起りやすいと推察される。

3-2. ヒトの O-GlcNAc 修飾タンパク質に相同なタンパク質の進化情報に基づくクラスタリング

取得された 162 個のヒトの O-GlcNAc 修飾タンパク質に含まれる 408 個の O-GlcNAc 修飾残基についてのスコアに対してワード法を用いることでヒト O-GlcNAc 修飾タンパク質をクラスタリングした(図2)。ワード法によるクラスタリングによって得られた樹形図をもとに 408 個の修飾残基をクラス分けした。樹形図の根より分岐する 3 本の枝を基部とする系統群をクラス 1 から 3 とし、各クラスを表す系統群を構成する 3 つのサブ系統群をサブクラスに分けた。クラス 1 (47 タンパク質) は脊椎動物、クラス 2 (94 タンパク質) は後生動物で相同タンパク質が見られるもので、クラス 3 は広く相同タンパク質がみられるものが分類された。サブクラス 1-1 は主に哺乳類、サブクラス 1-2 とサブクラス 1-3 は主に脊椎動物で保存されている。クラス 1 の中では O-GlcNAc 修飾残基はサブクラス 1-3 では脊椎動物全体に広く保存されていた。サブクラス 1-2 では相同なタンパク質は脊椎動物まで確認されたが O-GlcNAc 修飾残基まで保存されていたのは主に哺乳類までであった。クラス 2 の中では相同なタンパク質は動物において確認されたが、O-GlcNAc 修飾残基までの保存はサブクラス 2-1 では主に哺乳類、サブクラス 2-2 では脊椎動物、サブクラス 2-3 では後生動物で保存が見られた。クラス 3 ではサブクラス 3-1 は後生動物と植物で保存されている。サブクラス 3-2 と 3-3 は真核生物全体で広く保存され、サブクラス 3-3 では O-GlcNAc 修飾残基も広く保存されているがサブクラス 3-2 では多くの生物種で保存されていなかった。

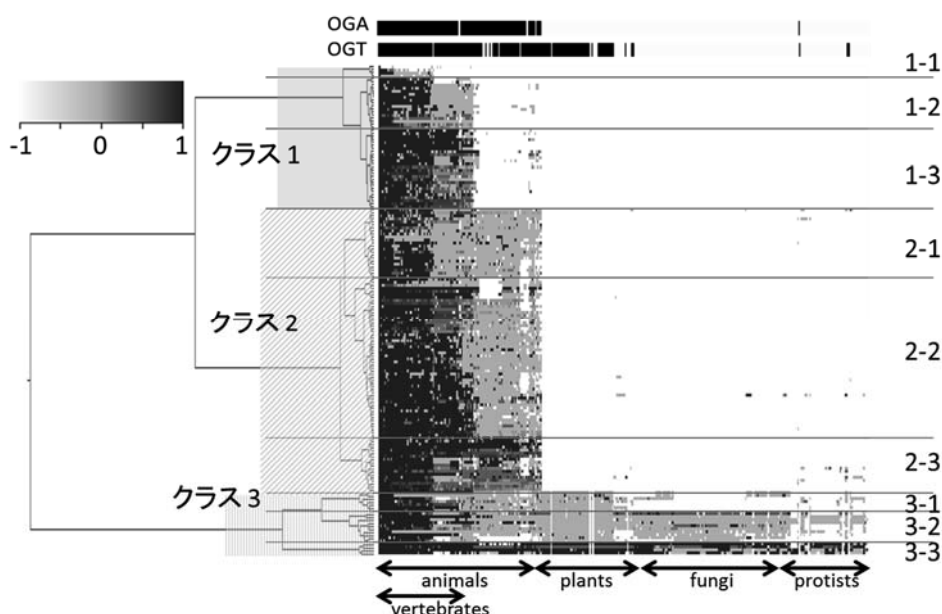


図 2. ヒトの *O*-GlcNAc 修飾タンパク質に相同なタンパク質のスコア付け後のクラスタリング

図の縦方向には、OGTSite より取得されたヒトの *O*-GlcNAc 修飾タンパク質に含まれる *O*-GlcNAc 修飾残基についてのスコア付けに伴うビットパターンがクラスタリングされて配置されている。横軸は生物種を表す。クラスタリング後に得られる樹形図をもとに修飾残基をクラス分けした。ヒートマップで表示された保存パターンは左上のスケールに従う。ヒートマップ上部に示された横長の図形は、OGT および OGA について保存されている生物種では黒、保存されていない生物種では白で表したものである。

3-3. OGT および OGA の保存度との比較

O-GlcNAc 修飾を触媒する酵素である OGT、*O*-GlcNAc を取り除く酵素である OGA の保存をデータセットのタンパク質と比較した (図 2 上部)。OGT の相同タンパク質は動物および植物で存在することが確認されるが、これと類似したパターンで保存されているクラスはサブクラス 3-1 のみであり、数は他のクラスと比較して少なかった。しかし、OGA の相同タンパク質の保存のパターンに類似して保存されているクラスはクラス 2 全体であり、データセットの半数以上を占めていた。このことから、OGA の *O*-GlcNAc を遊離させる能力の獲得と時期を同じくして数多くの *O*-GlcNAc 修飾タンパク質が現れたことが推察される。一方、OGA を持たずに OGT を持っている生物種が見られることからこれらの生物種については次の 2 つの仮説が考えられる。一つ目は、ヒトの OGT のアミノ酸配列と似ているために OGT のオーソログとされているだけで、機能が異なるために *O*-GlcNAc 修飾がほとんど起こっていないという仮説である。この仮説では OGA の獲得により、*O*-GlcNAc を取り除くことが出来るようになったことによって *O*-GlcNAc 修飾を調節する準備が整い、OGT のオーソログが *O*-GlcNAc を付加する酵素に進化したと推察される。二つ目は、OGT のオーソログが *O*-GlcNAc 修飾を付加する働きを持っており OGA とは別に *O*-GlcNAc 脱離酵素が存在するという仮説である。この仮説では、別の *O*-GlcNAc 脱離酵素よりも OGA の方が *O*-GlcNAc 修飾の調整に適していたため、OGA の獲得により別の *O*-GlcNAc 脱離酵素が不必要になり、多くのタンパク質を *O*-GlcNAc 修飾することができるようになったと推察される。

3-4. ID 領域への O-GlcNAc 修飾

O-GlcNAc 修飾を受ける 408 個のアミノ酸残基のうち大半(78.2%)を占める 319 個が ID 領域内に位置していたことから、クラスごとに ID 領域の割合を求めた (図 3)。各クラスのタンパク質が有する ID 領域の割合はクラス 1 とクラス 2 では同等であったが、クラス 3 に分類されたタンパク質では他のクラスに比べて ID 領域の割合が低い傾向が見られた (図 3A)。さらに、O-GlcNAc 修飾のうち ID 領域内で起こるものの割合は進化的に新しいクラスほど多くなる傾向が見られた (図 3B)。このことから ID 領域への O-GlcNAc 修飾を多数行うことで複雑な制御を行えるようになったと推察される。

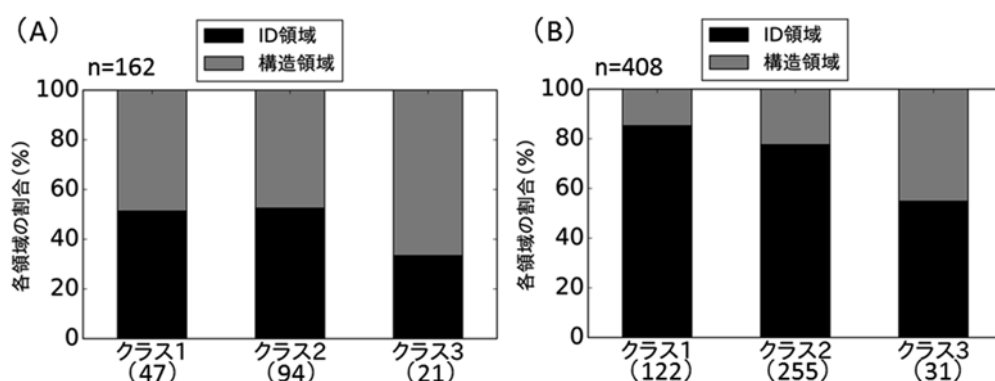
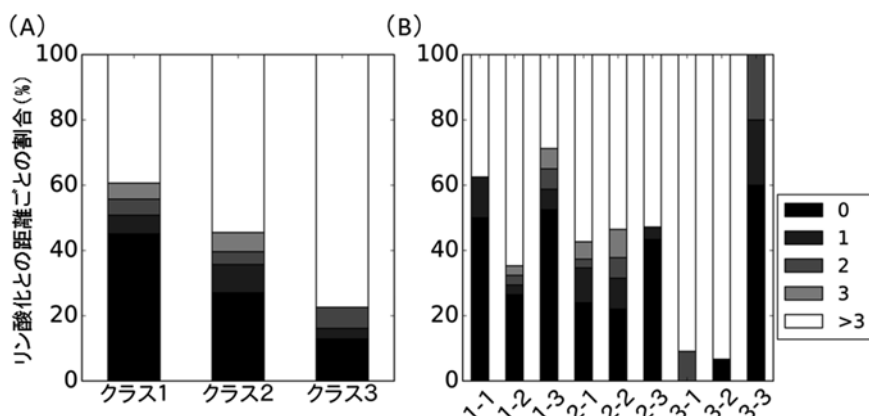


図 3. 各クラスにおける ID 領域の割合と O-GlcNAc 修飾残基の領域別割合

(A) 各クラスのタンパク質が有する ID 領域と構造領域の割合を示す。(B) 各クラスのタンパク質における O-GlcNAc 修飾残基の領域別割合を示す。どちらも ID 領域を黒、構造領域をグレーで表示している。各クラスごとのカッコ内の数字は (A) はタンパク質、(B) は O-GlcNAc 修飾残基の数を示す。

3-5. O-GlcNAc 修飾とリン酸化

O-GlcNAc 修飾は近傍のリン酸化と相互作用することが知られている。リン酸化との競合は転写因子 c-Myc、エストロゲンレセプタ β などのように同一残基での相互作用が知られる[22]ほか、カゼインキナーゼ CK2 のように近傍のリン酸化と O-GlcNAc 修飾によって調節される例も存在する[23]。このため、データセット中の O-GlcNAc 修飾残基のうちリン酸化と同一残基または近傍で起こるものの割合を求めた (図 4)。O-GlcNAc 修飾残基のうち、リン酸化残基の近傍で起こるものの割合は進化的に新しいタンパク質のほうが高いことが確認された (図 4A)。このことから進化的に新しいタンパク質ほど O-GlcNAc 修飾とリン酸化による調節が必要になったと推察される。また、クラス 3 においてのみサブクラス間での大きな差が見られた (図 4B)。クラス 3-1 と 3-2 はほとんどがリン酸化と競合しないと考えられるが、クラス 3-3 では多くの O-GlcNAc 修飾残基がリン酸化される残基の近傍で起こっていた。サブクラス 3-3 では図 2 より多くの生物種で保存されているがリン酸化との競合が多いことから、アミノ酸配列の変化ではなく翻訳後修飾による少ない変化によって適応してきたタンパク質中の重要部位であることが示唆される。この仮説を検証するために、サブクラス 3-3 に属するタンパク質の O-GlcNAc 修飾残基周辺の翻訳後修飾を調べた (図 5)。その結果、サブクラス 3-3 のタンパク質の O-GlcNAc 修飾残基周辺はリン酸化を始めとした多くの翻訳後修飾が起こる重要な部位であることが分かった。

図 4. リン酸化の近傍で起こる *O*-GlcNAc 修飾残基の割合

各クラスのタンパク質における *O*-GlcNAc 修飾のうち、リン酸化残基の近傍にあるものの割合を (A) クラスごとおよび (B) サブクラスごとに示した。凡例の数字は *O*-GlcNAc 修飾残基の最も近くにあるリン酸化残基までの距離 (単位はアミノ酸残基) である。0 は同一残基、1 はリン酸化残基に隣接するアミノ酸残基に *O*-GlcNAc 修飾が起こることを示す。

図 5. サブクラス 3-3 に属するタンパク質上の *O*-GlcNAc 修飾残基近傍の特徴

サブクラス 3-3 に属するタンパク質上の *O*-GlcNAc 修飾残基近傍のアミノ酸配列と翻訳後修飾が起こる残基を示した。左端はタンパク質名、その右側に修飾残基番号を示す。アミノ酸配列中央のグレーの背景で示されたアミノ酸が *O*-GlcNAc 修飾を受ける残基、白抜き文字で示されたアミノ酸がリン酸化を受ける残基、四角で囲われたアミノ酸がユビキチン化を受ける残基、網掛けで示されたアミノ酸はメチル化・アセチル化を受ける残基である。*O*-GlcNAc 修飾残基の周辺に多くの翻訳後修飾が起こるアミノ酸があることが分かる。PHB は DNA の合成を阻害する Prohibitin タンパク質、H2B1B はヒストン、ACTS はアクチン、RPB1 は RNA ポリメラーゼ II のサブユニットである。RPB1 の T1618 は ID 領域、それ以外は構造領域への *O*-GlcNAc 修飾である。

4. まとめ

O-GlcNAc 修飾タンパク質の進化解析から、はじめに OGT の獲得による *O*-GlcNAc を付加する機能の獲得が考えられる。しかし、OGT と同時期に出現したタンパク質自体が少なく、その中で残基が保存され *O*-GlcNAc 修飾されると推定される残基は少量である。よって、この時点では数少ないタンパク質を *O*-GlcNAc 修飾していた、または OGT に相同なタンパク質に機能の同一性ではなく *O*-GlcNAc 修飾が行われていなかったと考えられる。次に OGA の獲得によって *O*-GlcNAc を適切に取り除くことが可能になってダイナミックな調節が行えるようになり、多数のタンパク質が *O*-GlcNAc 修飾を受けるようになったと推察される。また、ダイナミックな調節が行えるようになったことおよび進化軸における ID 領域比率の増大により構造をとらない ID 領域による機能発現の調節が重要度を増し、リン酸化との相互作用も加わることでタンパク質の高度な機能調節が可能になったと推察される。

参考文献

- [1] Lubas WA, Frank DW, Krause M, Hanover JA. *O*-Linked GlcNAc transferase is a conserved nucleocytoplasmic protein containing tetratricopeptide repeats. *J Biol Chem*. 1997 Apr 4;272(14):9316-24.
- [2] Olszewski NE, West CM, Sassi SO, Hartweck LM. *O*-GlcNAc protein modification in plants: Evolution and function. *Biochim Biophys Acta*. 2010 Feb;1800(2):49-56. doi: 10.1016/j.bbagen.2009.11.016.
- [3] Love DC, Hanover JA. The hexosamine signaling pathway: deciphering the "O-GlcNAc code". *Sci STKE*. 2005 Nov 29;2005(312):re13.
- [4] Hart GW, Kreppel LK, Comer FI, Arnold CS, Snow DM, Ye Z, Cheng X, DellaManna D, Caine DS, Earles BJ, Akimoto Y, Cole RN, Hayes BK. *O*-GlcNAcylation of key nuclear and cytoskeletal proteins: reciprocity with *O*-phosphorylation and putative roles in protein multimerization. *Glycobiology*. 1996 Oct;6(7):711-6.
- [5] Hu P, Shimoji S, Hart GW. Site-specific interplay between *O*-GlcNAcylation and phosphorylation in cellular regulation. *FEBS Lett*. 2010 Jun 18;584(12):2526-38. doi: 10.1016/j.febslet.2010.04.044. Epub 2010 Apr 22.
- [6] Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. *Science*. 2002 Dec 6;298(5600):1912-34.
- [7] Haltiwanger RS, Holt GD, Hart GW. Enzymatic addition of *O*-GlcNAc to nuclear and cytoplasmic proteins. Identification of a uridine diphospho-N-acetylglucosamine:peptide beta-N-acetylglucosaminyltransferase. *J Biol Chem*. 1990 Feb 15;265(5):2563-8.
- [8] Gao Y, Wells L, Comer FI, Parker GJ, Hart GW. Dynamic *O*-glycosylation of nuclear and cytosolic proteins: cloning and characterization of a neutral, cytosolic beta-N-acetylglucosaminidase from human brain. *J Biol Chem*. 2001 Mar 30;276(13):9838-45. Epub 2001 Jan 8.
- [9] Liu X, Li L, Wang Y, Yan H, Ma X, Wang PG, Zhang L. A peptide panel investigation reveals the acceptor specificity of *O*-GlcNAc transferase. *FASEB J*. 2014 Aug;28(8):3362-72. doi: 10.1096/fj.13-246850. Epub 2014 Apr 23.
- [10] Cheung WD, Sakabe K, Housley MP, Dias WB, Hart GW. *O*-linked beta-N-acetylglucosaminyltransferase substrate specificity is regulated by myosin phosphatase targeting and other interacting proteins. *J Biol Chem*. 2008 Dec 5;283(49):33935-41. doi: 10.1074/jbc.M806199200. Epub 2008 Oct 7.
- [11] Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol*. 1999 Oct 22;293(2):321-31.
- [12] Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang C, Kissinger CR, Bailey RW, Griswold MD, Chiu W, Garner EC, Obradovic Z. Intrinsically disordered protein. *J Mol Graph Model*. 2001;19(1):26-59.
- [13] Uversky VN. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci*. 2002 Apr;11(4):739-56.
- [14] Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol*. 2005 Mar;6(3):197-208.
- [15] Gsponer J, Futschik ME, Teichmann SA, Babu MM. Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. *Science*. 2008 Nov 28;322(5906):1365-8. doi: 10.1126/science.1163581.
- [16] Kao HJ, Huang CH, Bretaña NA, Lu CT, Huang KY, Weng SL, Lee TY. A two-layered machine learning method to

- identify protein *O*-GlcNAcylation sites with *O*-GlcNAc transferase substrate motifs. BMC Bioinformatics. 2015;16 Suppl 18:S10. doi: 10.1186/1471-2105-16-S18-S10. Epub 2015 Dec 9.
- [17] The UniProt Consortium. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 2017 Jan 4;45(D1):D158-D169. doi: 10.1093/nar/gkw1099. Epub 2016 Nov 29.
- [18] Fukuchi S, Homma K, Minezaki Y, Gojobori T, Nishikawa K. Development of an accurate classification system of proteins into structured and unstructured regions that uncovers novel structural domains: its application to human transcription factors. BMC Struct Biol. 2009 Apr 30;9:26. doi: 10.1186/1472-6807-9-26.
- [19] Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res. 2015 Jan;43(Database issue):D512-20. doi: 10.1093/nar/gku1267. Epub 2014 Dec 16.
- [20] Nakaya A, Katayama T, Itoh M, Hiranuka K, Kawashima S, Moriya Y, Okuda S, Tanaka M, Tokimatsu T, Yamanishi Y, Yoshizawa AC, Kanehisa M, Goto S. KEGG OC: a large-scale automatic construction of taxonomy-based ortholog clusters. Nucleic Acids Res. 2013 Jan;41(Database issue):D353-7. doi: 10.1093/nar/gks1239. Epub 2012 Nov 27.
- [21] Pathak S, Alonso J, Schimpl M, Rafie K, Blair DE, Borodkin VS, Schüttelkopf AW, Albarbarawi O, van Aalten DM. The active site of *O*-GlcNAc transferase imposes constraints on substrate sequence. Nat Struct Mol Biol. 2015 Sep;22(9):744-50. doi: 10.1038/nsmb.3063. Epub 2015 Aug 3.
- [22] Hu P, Shimoji S, Hart GW. Site-specific interplay between *O*-GlcNAcylation and phosphorylation in cellular regulation. FEBS Lett. 2010 Jun 18;584(12):2526-38. doi: 10.1016/j.febslet.2010.04.044. Epub 2010 Apr 22.
- [23] Tarrant MK, Rho HS, Xie Z, Jiang YL, Gross C, Culhane JC, Yan G, Qian J, Ichikawa Y, Matsuoka T, Zachara N, Etzkorn FA, Hart GW, Jeong JS, Blackshaw S, Zhu H, Cole PA. Regulation of CK2 by phosphorylation and *O*-GlcNAcylation revealed by semisynthesis. Nat Chem Biol. 2012 Jan 22;8(3):262-9. doi: 10.1038/nchembio.771.