CRITICAL REVIEW

O-GlcNAc informatics: advances and trends

Chunyan Hou[1] · Weiyu Li[1,2] · Yaoxiang Li[1] · Junfeng Ma[1]

Abstract

As a post-translational modification, protein glycosylation is critical in health and disease. *O*-Linked β-*N*-acetylglucosamine (*O*-GlcNAc) modification (*O*-GlcNAcylation), as an intracellular monosaccharide modification on proteins, was discovered 40 years ago. Thanks to technological advances, the physiological and pathological significance of *O*-GlcNAcylation has been gradually revealed and widely appreciated, especially in recent years. *O*-GlcNAc informatics has been quickly evolving. Clearly, *O*-GlcNAc informatics tools have not only facilitated *O*-GlcNAc functional studies, but also provided us a unique perspective on protein *O*-GlcNAcylation. In this article, we review *O*-GlcNAc-focused software tools and servers that have been developed for *O*-GlcNAc research over the past four decades. Specifically, we will (1) survey bioinformatics tools that have facilitated *O*-GlcNAc proteomics data analysis, (2) introduce databases/servers for *O*-GlcNAc proteins/sites that have been experimentally identified by individual research labs, (3) describe software tools that have been developed to predict *O*-GlcNAc sites, and (4) introduce platforms cataloging proteins that interact with the *O*-GlcNAc cycling enzymes (i.e., O-GlcNAc transferase and O-GlcNAcase). We hope these resources will provide useful information to both experienced researchers and new incomers to the *O*-GlcNAc field. We anticipate that this review provides a framework to stimulate the future development of more sophisticated informatic tools for *O*-GlcNAc research.

Introduction

Glycosylation, the covalent addition of glycans onto proteins, is not only the most structurally diverse but also the most abundant post-translational modification (PTM) [1–3]. The classical glycosylation mainly occurs as oligosaccharides on proteins in the secretory pathway (commonly linked via Asn (N-linked) or Ser/Thr ("mucin-type" O-linked)). In comparison, O-linked β-N-acetylglucosamine modification (O-GlcNAcylation) on Ser/Thr residues of proteins was not reported until 1984 [4, 5]. Mounting evidence since then gradually revealed that

O-GlcNAc is distinct in multiple ways: (1) O-GlcNAc is a monosaccharide which is not elongated to complex sugar structures [6, 7]; (2) O-GlcNAc is a reversible and highly dynamic modification catalyzed by only a pair of enzymes: O-GlcNAc transferase (OGT) (which adds O-GlcNAc to Ser/Thr residues [8–10]) and β-D-N-acetylglucosaminidase (O-GlcNAcase) (which removes O-GlcNAc [11]); (3) O-GlcNAc exists almost exclusively on proteins localized in the nucleus, cytoplasm, and mitochondria [7, 12, 13]; (4) O-GlcNAc frequently interplays with other PTMs (particularly phosphorylation [14]); and (5) O-GlcNAc is most common in metazoans, despite the wide presence across all kingdoms of life (except Archaea) [15]; (6) due to the monosaccharide nature, functional roles of protein O-GlcNAcylation can be studied similar to phosphorylation and other PTMs, where site-specific mutagenesis as well as other targeted approaches [16–19] can be conducted. Indeed, O-GlcNAcylation as a single monosaccharide is of critical importance in numerous biological processes and intimately linked to physiological/pathological events [20–32]. Moreover, developing therapeutics and biomarker assays that target protein O-GlcNAcylation shows great promise [33].

✉ Junfeng Ma
 junfeng.ma@georgetown.edu

1 Department of Oncology, Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington, DC 20007, USA

2 Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD 21218, USA

In the past two decades, there have been tremendous technical advances, especially with high-throughput analytical techniques (tandem mass spectrometry in particular), allowing increasingly comprehensive analysis of glycans and glycoproteins [33–35]. Correspondingly, glycoproteomics data explosion promoted the rapid evolution of glycoinformatics, a branch of bioinformatics specifically tailored to information concerning the primary structure (composition, sequence, and linkages) and three-dimensional structure (including dynamics) of glycans and glycoproteins in tissues or cell types as well as their interaction with biological surroundings [36–40]. Until recently, the development of glycoinfomatics tools has been slow and difficult due to the complexity of conventional glycosylation, and O-GlcNAc, despite being a structurally well-defined monosaccharide that is found widespread throughout the nucelocytoplasmic proteome, has also been neglected. Of all the ~20,000 proteins encoded by the human genome, it is estimated that at least 25% (>5000 proteins) are localized in the secretory pathway (including ER, Golgi apparatus, and plasma membrane) [41], whereas the majority (~75%, 15,000 proteins) are localized in nucleus (nuclear and nucleolar structures), cytosol, mitochondria, cytoskeleton, and others. Among the ~15,000 proteins, >25% (>4000 proteins) have been found O-GlcNAcylated, with >11,000 Ser/Thr sites unambiguously identified to be O-GlcNAcylated thus far (O-GlcNAcAtlas, version 3.0) [42]. The numbers of O-GlcNAc proteins and modification sites clearly suggest its widespread presence (which may be far more than those modified by complex glycans). Arguably, O-GlcNAc represents a large population of glycosylated proteins in humans

and other species, which has led to substantially increased research attention over the past few decades, yielding >3400 O-GlcNAc-related publications (Fig. 1). Given the unique importance of O-GlcNAc in health and disease, increased efforts in developing O-GlcNAc informatic tools have been made. Undoubtedly, there is a need to bridge these endeavors to the glycoinformatics, glycobiology, and even a wider scientific community.

This article thus aims to summarize advances in informatics for O-GlcNAc research in the past 40 years. We will focus on several closely intertwined aspects: bioinformatics tools for O-GlcNAc proteomics; databases/servers cataloging O-GlcNAc proteins/peptides/sites experimentally identified; software tools for the prediction of O-GlcNAc sites; and databases for OGT/OGA interaction proteins (Fig. 2). Besides addressing the progress and challenges, we also illustrate how these O-GlcNAc informatics tools have shaped O-GlcNAc research, especially the elucidation of O-GlcNAc site-specific functional roles in physiology and pathology.

Bioinformatic tools for O-GlcNAc proteomics

As for other PTMs, chemical and biochemical tools/methods are indispensable for O-GlcNAc research. With the rapid maturation of high-resolution and sensitive mass spectrometry and accompanying techniques, tandem mass spectrometry-based proteomics has emerged as a cornerstone approach, which can be used for either low-throughput (e.g., immunoprecipitated proteins from cell lysates) or
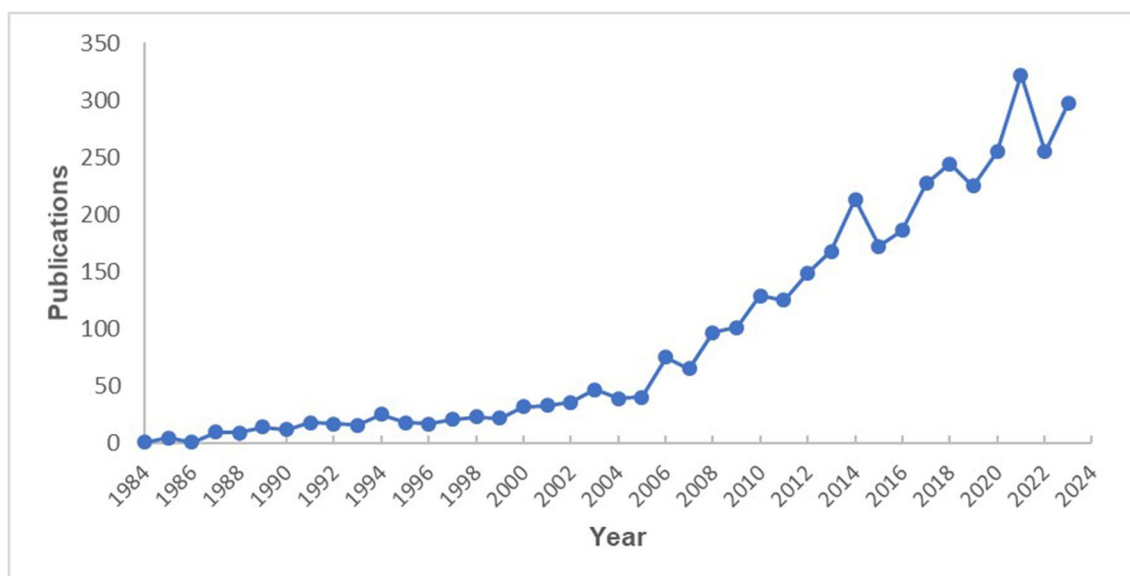
Fig. 1 Accumulated numbers of O-GlcNAc-focused publications indexed in PubMed from 1984 to 2023 (search queries: O-linked-β-N-acetylglucosamine, O-GlcNAc, O-GlcNAcylation)

Fig. 2 Application of O-Glc-NAc informatics tools for O-GlcNAc site-specific functions on protein(s) of interest
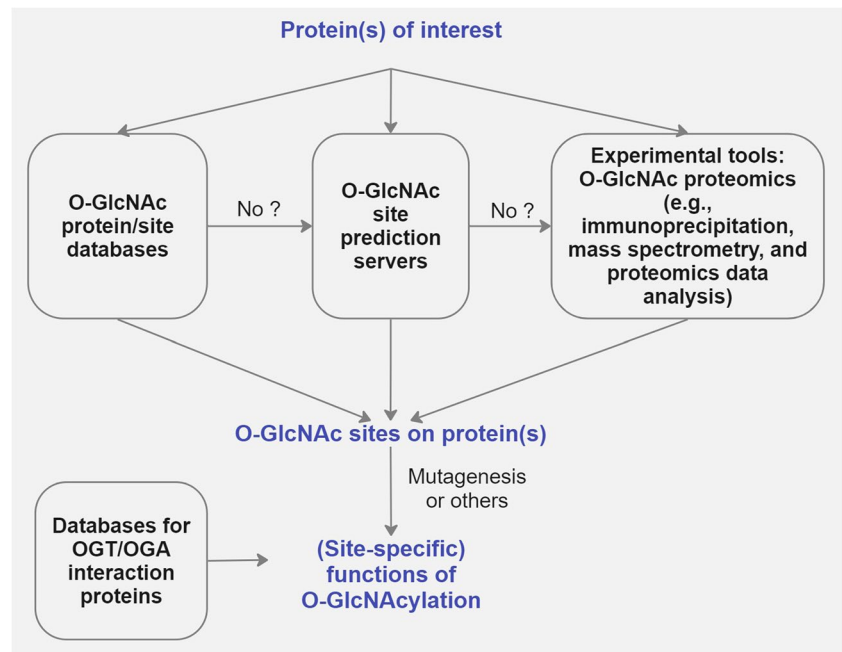
high-throughput (e.g., whole cell lysates) analysis. Due to the low stoichiometric nature, enrichment of O-GlcNAc proteins/peptides is generally needed for O-GlcNAc proteomics. Moreover, the O-GlcNAc moiety is extremely labile in gas phase, hindering accurate identification of O-GlcNAc sites. Thus, quite a few O-GlcNAc proteomics studies (especially in earlier days) reported numerous O-GlcNAcylated proteins as potential or putative (even after certain types of enrichment) and without information on the specific O-GlcNAc peptides or sites. It is problematic to consider any of such proteins as modified by O-GlcNAc. Therefore, identification of O-GlcNAc peptides/sites provides direct evidence for confident assignment of proteins as O-GlcNAcylated. And mapping O-GlcNAc sites generally serves as the first step towards elucidation of their functional importance [33, 43–49].

Despite the structural simplicity, site-specific O-GlcNAc proteomics has never been an easy task. Due to the inherent challenges of O-GlcNAc aforementioned (e.g., low stoichiometry and extreme lability in gas phase), it is not surprising that substantial efforts have been focused on sample preparation (e.g., development of enrichment methods/materials) and mass spectrometry data acquisition. For example, electron transfer dissociation (ETD) and hybrid fragmentation modes (such as electron transfer/higher-energy collision dissociation (EThcD), and higher-energy collisional dissociation (HCD) product dependent EThcD (HCD-pd-EThcD)) can retain O-GlcNAc moiety on peptides during fragmentation, largely facilitating O-GlcNAc site mapping [33, 45]. As an indispensable part, O-GlcNAc proteomics data analysis is not trivial. So far, a number of software packages have

been applied for O-GlcNAc site mapping. Instead of introducing all proteomics software tools available, we aim to focus on those that have been used more often for O-GlcNAc proteomics (below). It is notable that other tools (e.g., the OScore software [50], the Open Mass Spectrometry Search Algorithm (OMSSA) [51–53], and MS-GF+ [54]) have also been utilized for site-specific O-GlcNAc analysis.

Proteome Discoverer (PD) software, a commercial platform of Thermo Fisher Scientific, provides streamlined peptide/protein identification and quantification. Since its first release in 2007, PD has been evolving to integrate multiple search algorithms such as Sequest HT, Mascot, MSAmanda, Byonic, and ProSightPD [55]. PD is also commonly used for PTM analysis, since it includes ptmRS (the successor of PhosphoRS) as a node for measuring the confidence of modification localization in peptide sequences in terms of probability. In order to calculate individual probability values for each putatively modified site, the algorithm optimizes individual peak depths for different regions of the tandem mass spectrum with a limit of eight peaks per 100 m/z window [56]. Since PhosphoRS has been validated using synthetic phosphopeptides, it is not fully clear about the performance of ptmRS for other types of modifications [56]. Nevertheless, as the accompanying software for Orbitrap instruments, PD is a widely used software by many labs. As a seamlessly embedded module, ptmRS has been adopted for O-GlcNAc site localization analysis in many studies (e.g., [57–61]).

MaxQuant, a freely available proteomics software package integrated with search engine Andromeda, has gained wide adoption for analyzing large mass-spectrometric datasets [62]. The PTM scoring algorithm in Andromeda was initially

developed for phosphorylation and has been expanded to other types of modification. Andromeda divides the entire MS/MS spectrum into 100 m/z bins and evaluates the experimental peaks in each bin using a binominal distribution probability formula, giving priority to peaks with higher intensities [62]. Specifically, the PTM localization score is determined by selecting the four highest-intensity fragment ions from each 100 m/z segment, and this selection is performed for every possible combination of PTM sites within the peptide sequence [63]. MaxQuant has been applied for site-specific O-GlcNAc proteomics data analysis, typically with 0.75 as the PTM score threshold for unambiguously localization assignment [64–66].

Protein Prospector is another freely accessible suite of tools developed for analysis of MS-based proteomics. For ambiguous modified positions, Site Localization In Peptide (SLIP) scores are reported, which are determined by comparing probability and expectation values for the same peptide with different site assignments in terms of $-\log P$ [67]. So far, Protein Prospector has been used in a number of O-GlcNAc proteomics studies [68–72].

The relatively effective identification of O-GlcNAc peptides/sites by the aforementioned search engines shows that it is possible to use the traditional proteomics software packages for O-GlcNAc proteomics by including HexNAc as a variable modification. To bear in mind, the HexNAc moiety might be lost from peptides during collision-based fragmentation (especially HCD), producing peptide fragments without the glycan mass and resulting in low confident or erroneous site assignment. Thus, data analysis programs must anticipate the potential HexNAc loss to correctly calculate the peptide fragment masses. Besides, in recent years, there has been an emergence of a series of software tools targeted for glycoproteomics data analysis, due to the failure of complex glycan analysis by using the commonly used proteomics software. For example, to facilitate the structural education and site mapping of complex O-glycans on peptides, a series of packages have been developed, including Byonic [73], MSFragger-Glyco [74], pGlyco [75], O-pair [76], and GlycReSoft [77]. Although promising, only a few of them (e.g., Byonic [78, 79]) have been tentatively applied for O-GlcNAc proteomics analysis thus far.

Of special note, despite the significant progress to identify O-GlcNAc peptides and localize O-GlcNAc sites, it is a challenge to distinguish it from O-GalNAc (especially GalNAcα1-S/T, i.e., Tn antigen), since the two isomers of HexNAc result in the same delta mass on the modified peptides when fragmentated in fragmentation modes such as HCD. Although subcellular localization of proteins is helpful to distinguish GlcNAc and GalNAc, its performance might be compromised since O-GlcNAc also appears to be on certain extracellular proteins (e.g., mediated by the atypical enzyme EOGT) [80] and a small portion of O-GalNAc proteins appears to be localized in nucleus [81]. To that end, by leveraging the differential intensities of oxonium ions (i.e., m/z 126.055, 138.055, 144.066, 168.0655, and 186.0761) for O-GlcNAc and O-GalNAc, a binary logistic regression model HexNAcQuest was developed recently [82]. Results from independent validation datasets demonstrate that HexNAcQuest can accurately discern O-GlcNAc from O-GalNAc modification mainly based on the intensities of oxonium ions that have resulted from HCD, EThcD, or HCD-pd-EThcD mass spectrometry. Moreover, HexNAcQuest is a more accurate and general model in comparison to criteria based on empirical observations [45, 79, 83, 84]. In addition, a detailed protocol has recently been described to integrate HexNAcQuest with commonly used proteomics data analysis workflows [85], which will further facilitate distinguishing HexNAc isomers on peptides from complex samples.

Taken together, a series of software packages have been adopted for O-GlcNAc proteomics data analysis. However, their performance (especially those of the recently developed software tailored for O-glycoproteomics) has not been rigorously evaluated for O-GlcNAc proteomics. Each algorithm used may identify overlapping as well as different populations of O-GlcNAC peptides from the same mass spectrometry data files. Furthermore, although some software (e.g., PD and MaxQuant) allows custom settings (e.g., adding a certain chemical tag onto Ser/Thr residues) for data analysis, many others do not. This serves as an important feature when selecting appropriate software tools, as chemical tagging (e.g., the metabolic labeling- or chemozymatic labeling-based enrichment) appears to be a powerful strategy for O-GlcNAc enrichment [51, 53, 64, 65, 78]. Besides identification of O-GlcNAc peptides, accurate O-GlcNAc site localization is a critical aspect. Although many software tools provide site localization scores, unambiguous site assignment is still challenging (especially for peptides containing multiple O-GlcNAc sites). Clearly, this field is still in need of a widely accepted software tool and metric that can accurately describe the certainty of site localization in a user-friendly manner for O-GlcNAc proteomics data analysis of large datasets.

Databases/servers for O-GlcNAc proteins/sites

Technical advances in O-GlcNAc proteomics have produced a large amount of data. Several databases (e.g., PhosphoSite Plus [86], dbPTM [87], and MS-viewer [88]) attempted to collect O-GlcNAc sites and proteins. Unfortunately, these databases only covered a limited amount of useful information. Thus,

several databases have been created to specifically accommodate the rapid accumulation of O-GlcNAc information on proteins.

The database dbOGAP, established in 2011, is the first public bioinformatics resource dedicated to O-GlcNAcylated proteins and sites [89]. The initial version contained ~800 O-GlcNAcylated proteins and ~400 sites experimentally determined in about 500 articles published since 1984 to 2010. Unfortunately, the website became inaccessible due to the lack of maintenance. Consequently, we took the initiative to establish a new one, O-GlcNAcAtlas, to integrate O-GlcNAc sites and proteins from literature [42]. Stringent selection criteria were applied to select O-GlcNAc sites and proteins, and proteins identified in large-scale proteomics studies without O-GlcNAc site localization were not included. Besides unambiguously identified sites (with a localization score >0.75), ambiguously identified sites (mainly due to the low localization scores for peptides with clustered Ser/Thr residues especially when processed by the software tool Protein Prospector) were also included. Besides O-GlcNAc sites, related information (including species, sample type, peptide sequence, protein name, and site mapping methods used) was also extracted and presented. Meanwhile, the O-GlcNAc database was developed as another online resource for O-GlcNAc proteins (first focusing on human then extended to other species) [90]. Since many proteins without information on O-GlcNAc peptides/sites have been included, the O-GlcNAc database appears to have substantially more proteins. However, as aforementioned, it is hard to consider such proteins (mostly from O-GlcNAc proteomics studies without peptides/sites) as O-GlcNAcylated (lacking basic information). Interestingly, an O-GlcNAc score (between 0 and 100) is used as a quantifier to estimate the level of O-GlcNAc confidence for each protein in the database, with advertised caution for any score below 10.

Of note, the interplay between Ser/Thr O-GlcNAcylation and phosphorylation has been extensively studied in the past few decades [14]. With the public availability of specifically tailored databases of O-GlcNAc sites/proteins (as aforementioned) and phosphorylation sites/proteins (such as PhosphoSitePlus [86] and EPSD [91]), investigators can integrate such information easily, in whichever way they like (e.g., explore the potential cross-talk on modification sites on proteins of interest). The integration would also allow exploration of the intricate interaction between these two PTMs (e.g., co-localization evaluation of the two PTMs by doing meta-analysis [15]).

Bioinformatic tools for O-GlcNAc prediction

So far, >16,000 unambiguous sites and >10,000 ambiguous sites have been identified on a total number of >7000 proteins in multiple species (O-GlcNAcAtlas, version 3.0).

Despite the technological progress, sensitive and robust methods for global and site-specific O-GlcNAc analysis are still lacking in many cases. Since protein O-GlcNAcylation functions in a site-specific manner and accurate determination of O-GlcNAc sites is also technically demanding, there is a need to obtain O-GlcNAc site information to afford site-specific O-GlcNAc functional studies of proteins (without known modification sites reported). To that end, a number of computational tools (based on machine learning and deep learning methods) have been developed to predict O-GlcNAc sites (Table 1).

YinOYang, an artificial neuronal network-based predictor developed in 2002, is a pioneering effort in this area [92]. It was trained on 40 experimentally determined O-GlcNAc acceptor sites on human proteins to recognize the sequence context and surface accessibility. Although the positive dataset is limited, it serves as a benchmark for comparison in evaluating the effectiveness and reliability of newly emerging predictors. With the growing knowledge of O-GlcNAcylation and the development of its public resources, several predictors have been developed by using larger positive datasets. When the first O-GlcNAc database, i.e., dbOGAP, was launched, a site prediction system named OGlcNAcScan was also introduced to the scientific community, which was based on a support vector machine (SVM) and trained on nearly 400 O-GlcNAcylated sites in dbOGAP [89]. However, this predictor and related database have not been accessible. PGlcS is another O-GlcNAcylated sites predictor trained on the dbOGAP dataset [93]. It used k-means cluster and SVM classifier combined with multiple features to improve the performance. Although it claimed that PGlcS presented better sensitivity compared to O-GlcNAcScan when tested on an independent dataset, its public access was not provided. Potential OGT substrate motifs have also been used in developing a two-layered machine learning-based predictive model, i.e., OGTSite [94]. The positive training data composed of 410 experimentally verified O-GlcNAcylation sites was investigated using the maximal dependence decomposition (MDD) method to discover substrate motif signatures. Despite the promising accuracy provided by OGTSite, its web portal is no longer available. Since 2013, Jia and colleagues developed a series of predictors to capture O-GlcNAcylated sites on proteins. The first one, i.e., O-GlcNAcPRED, is an SVM-based model trained on 41 mer peptide sequences built from 167 proteins in dbOGAP with application of the adapted normal distribution bi-profile Bayes (ANBPB) feature encoding scheme [95]. In 2018, Jia et al. [96] introduced an improved predictor, O-GlcNAcPRED-II, based on larger training datasets (e.g., 945 O-GlcNAc sites) and the rotation forest algorithm.

The quickly evolving deep learning methods, together with the availability of much larger datasets, have prompted the development of several new prediction tools for

Table 1 Computational tools developed for the prediction of O-GlcNAc sites on proteins

Tool name	Feature information	Model training	Positive instances for training	Species	Web server	Ref.
YinOYang	Local sequence and surface accessibility	Artificial neuronal networks	40	Human	services.healthtech.dtu.dk/services/YinOY-ang-1.2	[92]
OGlcNAcScan	Local sequence	Support vector machine	373	Not specified	cbsb.lombardi.george-town.edu/OGAP.html	[89]
PGlcS	Local sequence, secondary structure, disorder, position specifice scoring matrix, AAindex	Support vector machine	339	Not specified	N/A	[93]
OGTSite	Conserved motifs	Support vector machine	410	Not specified	csb.cse.yzu.edu.tw/OGTSite	[94]
O-GlcNAcPRED	Local sequence	Support vector machine	339	Not specified	N/A	[95]
O-GlcNAcPRED-II	Local sequence	Rotation forest	945	Not specified	N/A	[96]
O-GlcNAcPRED-DL	Local sequence, evolutionary information, AAindex	Deep neural networks	4424	Human and mouse	https://oglcnac.org/pred_dl/	[97]
LM-OGlcNAc-Site	Local sequence	Feedforward neural network	12,644	Not specified	kcdukkalab.org/LMOGlcNAcSite	[98]

O-GlcNAc sites in recent years. Among the >9000 unambiguously identified O-GlcNAc sites on proteins from multiple species (O-GlcNAcAtlas, version 2.0), humans and mice contain the most O-GlcNAcylated sites. Thus, Jia and collaborators built ensemble models based on deep learning networks (named O-GlcNAcPRED-DL) for these two species separately [97]. In short, O-GlcNAcPRED-DL employed a one-hot encoding approach, BLOSUM62 (a matrix reflecting sequence similarity), AAindex (reflecting amino acid physical and chemical properties), and Word2Vec (reflecting global characteristics and contextual information). Moreover, four network frameworks were built based on the connection of the convolutional neural network (CNN) and the bidirectional long short-term memory method (BiLSTM). In comparison to the traditional machine learning-based tools reported previously, the ensembled O-GlcNAcPRED-DL models showed substantially enhanced performance for the O-GlcNAc site prediction on human and mouse proteins. Meanwhile, also based upon the datasets from O-GlcNAcAtlas (version 2.0), Pokharel et al. developed another predictor LM-OGlcNAc-Site [98]. In brief, it integrates embeddings from multiple protein language models (Ank, ESM-2, and ProtT5) by adopting a decision-level fusion approach. LM-OGlcNAc-Site appears to outperform the models trained on these individual models as well as the integrated models using score-level fusion and O-GlcNAcPRED-II.

Collectively, a number of computational models have been developed to predict O-GlcNAc sites in the past few decades. It is fortunate that O-GlcNAc prediction, coupled with downstream experimental approaches (e.g., site-directed mutagenesis), has been applied for many site-specific functional studies. Despite the success obtained, some challenges remain. For example, there is no consensus motif for protein O-GlcNAcylation, although there appear to be OGT-preferred amino acid sequences within several species [15]. In addition, secondary and tertiary structures also appear to affect O-GlcNAcylation [15]. Therefore, there may still be challenges to accurately predict O-GlcNAc sites on certain proteins. Moreover, the developed models are mostly trained on human and mouse datasets, due to the limited experimental datasets for other species. Thus, the performance of current prediction models for proteins in other species might be compromised. With the further development of advanced artificial intelligence (e.g., advanced deep learning and protein language models) and further increased growth of experimental datasets for training and testing, we anticipate that current models will be refined and novel models will be developed for enhanced prediction, sensitivity, and accuracy of O-GlcNAc sites on proteins from human, mouse, and other species. Prediction of O-GlcNAc sites will continue to serve as a valuable strategy by many researchers to expedite site-specific functional elucidation in diverse scenarios.

Databases for OGT/OGA interaction proteins

As the main executors of biological function, proteins are not working alone. But rather, proteins work closely with their interaction partners. Defining the interaction proteins

of OGT/OGA, the O-GlcNAc cycling enzymes, is of critical importance.

Technologies to map protein–protein interactions (PPIs) have enabled the global characterization of PPIs of many proteins in recent years. Some high-throughput methods (including affinity purification coupled tandem mass spectrometry (AP-MS), immunoprecipitation-MS (IP-MS), cross-linking MS (XL-MS), and proximity labeling MS (PL-MS)) have also been tailored for the characterization of OGT/OGA-interacting proteins [99–107].

To accommodate the exponentially increased datasets of PPIs, several comprehensive databases (e.g., BioGRID [108], APID [109], IntAct [110], and STRING [111]) were constructed in the past years. Although these public repertoires categorize hundreds and thousands of PPIs from many species, they covered only a limited number of OGT-interacting proteins experimentally described. Recently, we compiled OGT-Protein Interaction Network (OGT-PIN) [112]. As a specifically designed portal, OGT-PIN is a rigorously curated and comprehensive database for interaction proteins of OGT and its orthologues identified in several species of intense research (e.g., SXC in *Drosophila melanogaster*, SEC in plants, and OGT-1 in Caenorhabditis elegans). Although over 2500 were experimentally identified as OGT interaction proteins, >1000 can be regarded as high-stringent interacting proteins of OGT and orthologues (OGT-PIN, version 2.0). Among them, it appears that human OGT has >800 high-stringent interacting proteins, suggesting that it is truly one of the hub proteins in cellular interaction network. Interestingly, only a small portion (~39%) of OGT-interacting proteins have been found to be OGT substrates, supporting the notion that they are not necessarily OGT substrates. The cataloging of OGT interaction proteins will facilitate functional studies of OGT-catalyzed O-GlcNAcylation on protein(s) of interest. In addition, it will provide clues to further understand the non-canonical roles of OGT which are yet to be characterized [113]. In contrast, OGA interactomes have been much less explored [106, 107, 114], and a specifically tailored database of OGA interaction proteins is not available yet. Investigators who are interested in that aspect might have to retrieve information from comprehensive databases aforementioned (such as BioGRID) and/or related publications.

With the further technical advances (e.g., in vivo proximity labeling), it is anticipated that more proteins will be identified as OGT/OGA-interacting proteins. The identification of weak and transient OGT/OGA interaction proteins will be anticipated, which will further expand the repositories. However, how to accurately define true hits and avoid false positives continues to be a question. In addition, as other proteins, interactomes of OGT/OGA are a dynamic system; thus, cataloging temporal interaction partners of OGT/OGA may provide us a dynamic view of the OGT/OGA interaction networks upon perturbation in different biological settings.

Conclusions

Four decades of research on O-GlcNAcylation has been very fruitful! In this article, we summarized great progress in informatics for O-GlcNAc research, by focusing on several aspects (including bioinformatics tools for O-GlcNAc proteomics; databases/servers cataloging O-GlcNAc proteins/peptides/sites experimentally identified; software tools for the prediction of O-GlcNAc sites; and databases for OGT/OGA interaction proteins). Not only have O-GlcNAc studies created large amounts of datasets (e.g., O-GlcNAc proteomics and interactomics), urging the development of an array of O-GlcNAc-focused databases/servers and software tools, but O-GlcNAc informatics resources have been instrumental in facilitating O-GlcNAc studies along the years (Fig. 2). Besides elucidating functional roles of O-GlcNAcylation on many proteins, they have provided us valuable insights into multiple aspects such as a global and detailed view of protein O-GlcNAcylation and OGT biology. Of note, informatics analysis can also be applied to other O-GlcNAc-relevant research topics (e.g., prediction of putative conveyers of OGT intellectual disability [115] and cancer biomarkers [116]), which are beyond the scope of this article.

Clearly, there are aspects of O-GlcNAc informatics that need to be improved. For example, performance of software packages for O-GlcNAc proteomics needs to be rigorously evaluated and/or refined. A key point is to further improve O-GlcNAc site localization so that unambiguous and accurate site assignment can be achieved (especially for peptides containing multiple O-GlcNAc sites). Although certain success regarding O-GlcNAc prediction has been obtained, there is room to further enhance prediction sensitivity and accuracy of O-GlcNAc sites. The continuously expanding repositories of O-GlcNAc sites from experiments can be used to further improve O-GlcNAc prediction performance. The rapid evolution of artificial intelligence capabilities (e.g., via machine learning and deep learning-based algorithms) has begun to unleash unprecedented potential to modern biology research (including glycobiology) and precision medicine [117, 118]. With the implementation of these technological advances, we believe that O-GlcNAc informatics resources will become more sophisticated. The maturation of these resources will undoubtedly encourage more researchers to join the glycosciences and particularly O-GlcNAc field. We anticipate that the advances in O-GlcNAc informatics will make it a handy and indispensable tool for biomedical scientists in O-GlcNAc-targeted basic and translation research in the coming years.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Spiro RG. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. Glycobiology. 2002;12:43R–56R. https://doi.org/10.1093/glycob/12.4.43r.

2. Hart GW, Copeland RJ. Glycomics Hits the Big Time. Cell. 2010;143:672–6. https://doi.org/10.1016/j.cell.2010.11.008.

3. Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, Mohnen D, Kinoshita T, Packer NH, Prestegard JH, Schnaar RL, Seeberger PH. Essentials of Glycobiology. 4th ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2022.

4. Torres CR, Hart GW. Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc. J Biol Chem. 1984;259:3308–17.

5. Holt GD, Hart GW. The subcellular distribution of terminal N-acetylglucosamine moieties. Localization of a novel protein-saccharide linkage. O-linked GlcNAc J Biol Chem. 1986;261:8049–57.

6. Wells L, Vosseller K, Hart GW. Glycosylation of nucleocytoplasmic proteins: signal transduction and O-GlcNAc. Science. 2001;291:2376–8. https://doi.org/10.1126/science.1058714.

7. Hart GW, Housley MP, Slawson C. Cycling of O-linked beta-N-acetylglucosamine on nucleocytoplasmic proteins. Nature. 2007;446:1017–22. https://doi.org/10.1038/nature05815.

8. 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;265:2563–8.

9. Haltiwanger RS, Blomberg MA, Hart GW. Glycosylation of nuclear and cytoplasmic proteins. Purification and characterization of a uridine diphospho-N-acetylglucosamine:polypeptide beta-N-acetylglucosaminyltransferase. J Biol Chem. 1992;267:9005–13.

10. Hanover JA, Yu S, Lubas WB, Shin SH, Ragano-Caracciola M, Kochran J, Love DC. Mitochondrial and nucleocytoplasmic isoforms of O-linked GlcNAc transferase encoded by a single mammalian gene. Arch Biochem Biophys. 2003;409:287–97. https://doi.org/10.1016/s0003-9861(02)00578-7.

11. Dong DL, Hart GW. Purification and characterization of an O-GlcNAc selective N-acetyl-beta-D-glucosaminidase from rat spleen cytosol. J Biol Chem. 1994;269:19321–30.

12. Ma J, Liu T, Wei A-C, Banerjee P, O'Rourke B, Hart GW. O-GlcNAcomic profiling identifies widespread O-Linked β-N-Acetylglucosamine modification (O-GlcNAcylation) in oxidative phosphorylation system regulating cardiac mitochondrial function. J Biol Chem. 2015;290:29141–53. https://doi.org/10.1074/jbc.M115.691741.

13. Banerjee PS, Ma J, Hart GW. Diabetes-associated dysregulation of O-GlcNAcylation in rat cardiac mitochondria. PNAS. 2015;112:6050–5. https://doi.org/10.1073/pnas.1424017112.

14. Hart GW, Slawson C, Ramirez-Correa G, Lagerlof O. Cross talk between O-GlcNAcylation and Phosphorylation: roles in signaling, transcription, and chronic disease. Annu Rev Biochem. 2011;80:825–58. https://doi.org/10.1146/annurev-biochem-060608-102511.

15. Ma J, Hou C, Wu C. Demystifying the O-GlcNAc Code: A Systems View. Chem Rev. 2022. https://doi.org/10.1021/acs.chemrev.1c01006.

16. Gorelik A, van Aalten DMF. Tools for functional dissection of site-specific O-GlcNAcylation. RSC Chem Biol. 2020;1:98–109. https://doi.org/10.1039/d0cb00052c.

17. Ramirez DH, Aonbangkhen C, Wu H-Y, Naftaly JA, Tang S, O'Meara TR, Woo CM. Engineering a proximity-directed O-GlcNAc transferase for selective protein O-GlcNAcylation in cells. ACS Chem Biol. 2020;15:1059–66. https://doi.org/10.1021/acschembio.0c00074.

18. Moon SP, Javed A, Hard ER, Pratt MR. Methods for Studying site-specific O-GlcNAc modifications: Successes, Limitations, and important future goals. JACS Au. 2022;2:74–83. https://doi.org/10.1021/jacsau.1c00455.

19. Ma B, Khan KS, Xu T, Xeque Amada J, Guo Z, Huang Y, Yan Y, Lam H, Cheng AS-L, Ng BW-L. Targeted Protein O-GlcNAcylation using bifunctional small molecules. J Am Chem Soc. 2024;146:9779–89. https://doi.org/10.1021/jacs.3c14380.

20. Bond MR, Hanover JA. O-GlcNAc cycling: a link between metabolism and chronic disease. Annu Rev Nutr. 2013;33:205–29. https://doi.org/10.1146/annurev-nutr-071812-161240.

21. Yang X, Qian K. Protein O -GlcNAcylation: emerging mechanisms and functions. Nat Rev Mol Cell Biol. 2017;18:452–65. https://doi.org/10.1038/nrm.2017.22.

22. Hart GW. Nutrient regulation of signaling and transcription. J Biol Chem. 2019;294:2211–31. https://doi.org/10.1074/jbc.AW119.003226.

23. Chatham JC, Zhang J, Wende AR. Role of O-Linked N-Acetylglucosamine protein modification in cellular (Patho)Physiology. Physiol Rev. 2020;101:427–93. https://doi.org/10.1152/physrev.00043.2019.

24. Saunders H, Dias WB, Slawson C. Growing and dividing: how O-GlcNAcylation leads the way. J Biol Chem. 2023;299:105330. https://doi.org/10.1016/j.jbc.2023.105330.

25. Minh GL, Esquea EM, Young RG, Huang J, Reginato MJ. On a sugar high: Role of O-GlcNAcylation in cancer. J Biol Chem. 2023;299:105344. https://doi.org/10.1016/j.jbc.2023.105344.

26. Pratt MR, Vocadlo DJ. Understanding and exploiting the roles of O-GlcNAc in neurodegenerative diseases. J Biol Chem. 2023;299:105411. https://doi.org/10.1016/j.jbc.2023.105411.

27. Nelson ZM, Leonard GD, Fehl C. Tools for investigating O-GlcNAc in signaling and other fundamental biological pathways. J Biol Chem. 2024;300:105615. https://doi.org/10.1016/j.jbc.2023.105615.

28. Liu X, Cai YD, Chiu JC. Regulation of protein O-GlcNAcylation by circadian, metabolic, and cellular signals. J Biol Chem. 2024;300:105616. https://doi.org/10.1016/j.jbc.2023.105616.

29. Zhang J, Wang Y. Emerging roles of O-GlcNAcylation in protein trafficking and secretion. J Biol Chem. 2024;300:105617. https://doi.org/10.1016/j.jbc.2024.105677.

30. Wu C, Li J, Lu L, Li M, Yuan Y, Li J. OGT and OGA: Sweet guardians of the genome. J Biol Chem. 2024;300:107141. https://doi.org/10.1016/j.jbc.2024.107141.

31. Umapathi P, Aggarwal A, Zahra F, Narayanan B, Zachara NE. The multifaceted role of intracellular glycosylation in cytoprotection and heart disease. J Biol Chem. 2024. https://doi.org/10.1016/j.jbc.2024.107296.

32. Wells L, Hart GW. O-GlcNAcylation: A major nutrient/stress sensor that regulates cellular physiology. J Biol Chem. 2024. https://doi.org/10.1016/j.jbc.2024.107635.

33. Ma J, Wu C, Hart GW. Analytical and biochemical perspectives of protein O-GlcNAcylation. Chem Rev. 2021;121:1513–81. https://doi.org/10.1021/acs.chemrev.0c00884.

34. Schjoldager KT, Narimatsu Y, Joshi HJ, Clausen H. Global view of human protein glycosylation pathways and functions. Nat Rev Mol Cell Biol. 2020;21:729–49. https://doi.org/10.1038/s41580-020-00294-x.

35. Bagdonaite I, Malaker SA, Polasky DA, Riley NM, Schjoldager K, Vakhrushev SY, Halim A, Aoki-Kinoshita KF, Nesvizhskii AI, Bertozzi CR, Wandall HH, Parker BL, Thaysen-Andersen M, Scott NE. Glycoproteomics. Nat Rev Methods Primers. 2022;2:1–29. https://doi.org/10.1038/s43586-022-00128-4.

36. Pérez S, Mulloy B. Prospects for glycoinformatics. Curr Opinion Struct Biol. 2005;15:517–24. https://doi.org/10.1016/j.sbi.2005.08.005.

37. Egorova KS, Toukach PV. Glycoinformatics: Bridging Isolated Islands in the Sea of Data. Angewandte Chemie International Edition. 2018;57:14986–90. https://doi.org/10.1002/anie.201803576.

38. Klein J, Zaia J. glypy: An Open source glycoinformatics library. J Proteome Res. 2019;18:3532–7. https://doi.org/10.1021/acs.jproteome.9b00367.

39. Aoki-Kinoshita KF, Campbell MP, Lisacek F, Neelamegham S, York WS, Packer NH. Glycoinformatics. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, Mohnen D, Kinoshita T, Packer NH, Prestegard JH, Schnaar RL, Seeberger PH (eds) Essentials of Glycobiology, 4th ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY); 2022.

40. Martinez K, Agirre J, Akune Y, Aoki-Kinoshita KF, Arighi C, Axelsen KB, Bolton E, Bordeleau E, Edwards NJ, Fadda E, Feizi T, Hayes C, Ives CM, Joshi HJ, Krishna Prasad K, Kossida S, Lisacek F, Liu Y, Lütteke T, Ma J, Malik A, Martin M, Mehta AY, Neelamegham S, Panneerselvam K, Ranzinger R, Ricard-Blum S, Sanou G, Shanker V, Thomas PD, Tiemeyer M, Urban J, Vita R, Vora J, Yamamoto Y, Mazumder R. Functional implications of glycans and their curation: insights from the workshop held at the 16th Annual International Biocuration Conference in Padua. Italy Database. 2024;2024:baae073. https://doi.org/10.1093/database/baae073.

41. Thul PJ, Åkesson L, Wiking M, Mahdessian D, Geladaki A, Ait Blal H, Alm T, Asplund A, Björk L, Breckels LM, Bäckström A, Danielsson F, Fagerberg L, Fall J, Gatto L, Gnann C, Hober S, Hjelmare M, Johansson F, Lee S, Lindskog C, Mulder J, Mulvey CM, Nilsson P, Oksvold P, Rockberg J, Schutten R, Schwenk JM, Sivertsson Å, Sjöstedt E, Skogs M, Stadler C, Sullivan DP, Tegel H, Winsnes C, Zhang C, Zwahlen M, Mardinoglu A, Pontén F, von Feilitzen K, Lilley KS, Uhlén M, Lundberg E. A subcellular map of the human proteome. Science. 2017;356:eaal1321. https://doi.org/10.1126/science.aal3321.

42. Ma J, Li Y, Hou C, Wu C. O-GlcNAcAtlas: A database of experimentally identified O-GlcNAc sites and proteins. Glycobiology. 2021;31:719–23. https://doi.org/10.1093/glycob/cwab003.

43. Ma J, Hart GW. Protein O-GlcNAcylation in diabetes and diabetic complications. Expert Rev Proteomics. 2013;10:365–80. https://doi.org/10.1586/14789450.2013.820536.

44. Ma J, Hart GW. O-GlcNAc profiling: from proteins to proteomes. Clin Proteomics. 2014;11:8. https://doi.org/10.1186/1559-0275-11-8.

45. Maynard JC, Chalkley RJ. Methods for enrichment and assignment of N-Acetylglucosamine modification sites. Mol Cell Proteomics. 2021;20: 100031. https://doi.org/10.1074/mcp.R120.002206.

46. Riley NM, Bertozzi CR, Pitteri SJ. A pragmatic guide to enrichment strategies for mass spectrometry-based glycoproteomics. Mol Cell Proteomics. 2021;20: 100029. https://doi.org/10.1074/mcp.R120.002277.

47. Saha A, Bello D, Fernández-Tejada A. Advances in chemical probing of protein O-GlcNAc glycosylation: structural role and molecular mechanisms. Chem Soc Rev. 2021;50:10451–85. https://doi.org/10.1039/d0cs01275k.

48. Xu S, Sun F, Tong M, Wu R. MS-based proteomics for comprehensive investigation of protein O-GlcNAcylation. Mol Omics. 2021;17:186–96. https://doi.org/10.1039/d1mo00025j.

49. Hu W, Zhang G, Zhou Y, Xia J, Zhang P, Xiao W, Xue M, Lu Z, Yang S. Recent development of analytical methods for disease-specific protein O-GlcNAcylation. RSC Adv. 2022;13:264–80. https://doi.org/10.1039/d2ra07184c.

50. Hahne H, Kuster B. A novel two-stage tandem mass spectrometry approach and scoring scheme for the identification of O-GlcNAc modified peptides. J Am Soc Mass Spectrom. 2011;22:931–42. https://doi.org/10.1007/s13361-011-0107-y.

51. Wang Z, Udeshi ND, O'Malley M, Shabanowitz J, Hunt DF, Hart GW. Enrichment and site mapping of O-linked N-acetylglucosamine by a combination of chemical/enzymatic tagging, photochemical cleavage, and electron transfer dissociation mass spectrometry. Mol Cell Proteomics. 2010;9:153–60. https://doi.org/10.1074/mcp.M900268-MCP200.

52. Wang Z, Udeshi ND, Slawson C, Compton PD, Sakabe K, Cheung WD, Shabanowitz J, Hunt DF, Hart GW. Extensive crosstalk between O-GlcNAcylation and phosphorylation regulates cytokinesis. Sci Signal. 2010;3:ra2. https://doi.org/10.1126/scisignal.2000526.

53. Ma J, Wang W-H, Li Z, Shabanowitz J, Hunt DF, Hart GW. O-GlcNAc site mapping by using a combination of chemoenzymatic labeling, Copper-free click chemistry, Reductive Cleavage, and electron-transfer dissociation mass spectrometry. Anal Chem. 2019;91:2620–5. https://doi.org/10.1021/acs.analchem.8b05688.

54. Wang S, Yang F, Petyuk VA, Shukla AK, Monroe ME, Gritsenko MA, Rodland KD, Smith RD, Qian W-J, Gong C-X, Liu T. Quantitative proteomics identifies altered O-GlcNAcylation of structural, synaptic and memory-associated proteins in Alzheimer's disease. J Pathol. 2017;243:78–88. https://doi.org/10.1002/path.4929.

55. Orsburn BC. Proteome discoverer—a community enhanced data processing suite for protein informatics. Proteomes. 2021;9:15. https://doi.org/10.3390/proteomes9010015.

56. Taus T, Köcher T, Pichler P, Paschke C, Schmidt A, Henrich C, Mechtler K. Universal and confident phosphorylation site localization using phosphoRS. J Proteome Res. 2011;10:5354–62. https://doi.org/10.1021/pr200611n.

57. Zhao P, Viner R, Teo CF, Boons G-J, Horn D, Wells L. Combining high-energy C-trap dissociation and electron transfer dissociation for protein O-GlcNAc modification site assignment. J Proteome Res. 2011;10:4088–104. https://doi.org/10.1021/pr2002726.

58. Griffin ME, Jensen EH, Mason DE, Jenkins CL, Stone SE, Peters EC, Hsieh-Wilson LC. Comprehensive mapping of O-GlcNAc modification sites using a chemically cleavable tag. Mol Biosyst. 2016;12:1756–9. https://doi.org/10.1039/c6mb00138f.

59. Wu C, Shi S, Hou C, Luo Y, Byers S, Ma J. Design and preparation of novel nitro-oxide-grafted nanospheres with enhanced hydrogen bonding interaction for O-GlcNAc analysis. ACS Appl Mater Interfaces. 2022;14:47482–90. https://doi.org/10.1021/acsami.2c15039.

60. Mitchell CW, Galan Bartual S, Ferenbach AT, Scavenius C, van Aalten DMF. Exploiting O-GlcNAc transferase promiscuity to dissect site-specific O-GlcNAcylation. Glycobiology. 2023;33:1172–81. https://doi.org/10.1093/glycob/cwad086.

61. Zhang Y, Zhou S, Kai Y, Zhang Y, Peng C, Li Z, Mughal MJ, Julie B, Zheng X, Ma J, Ma CX, Shen M, Hall MD, Li S, Zhu W. O-GlcNAcylation of MITF regulates its activity and CDK4/6 inhibitor resistance in breast cancer. Nat Commun. 2024;15:5597. https://doi.org/10.1038/s41467-024-49875-w.

62. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: A peptide search engine integrated into the maxquant environment. J Proteome Res. 2011;10:1794–805. https://doi.org/10.1021/pr101065j.

63. Olsen JV, Blagoev B, Gnad F, Macek B, Kumar C, Mortensen P, Mann M. Global, In Vivo, and site-specific phosphorylation dynamics in signaling networks. Cell. 2006;127:635–48. https://doi.org/10.1016/j.cell.2006.09.026.

64. Li J, Li Z, Duan X, Qin K, Dang L, Sun S, Cai L, Hsieh-Wilson LC, Wu L, Yi W. An isotope-coded photocleavable probe for quantitative profiling of protein O-GlcNAcylation. ACS Chem Biol. 2019;14:4–10. https://doi.org/10.1021/acschembio.8b01052.

65. Liu J, Hao Y, Wang C, Jin Y, Yang Y, Gu J, Chen X. An optimized isotopic photocleavable tagging strategy for site-specific and quantitative profiling of protein O-GlcNAcylation in colorectal cancer metastasis. ACS Chem Biol. 2022;17:513–20. https://doi.org/10.1021/acschembio.1c00981.

66. Li X, Lei C, Song Q, Bai L, Cheng B, Qin K, Li X, Ma B, Wang B, Zhou W, Chen X, Li J. Chemoproteomic profiling of O-GlcNAcylated proteins and identification of O-GlcNAc transferases in rice. Plant Biotechnol J. 2023;21:742–53. https://doi.org/10.1111/pbi.13991.

67. Baker PR, Trinidad JC, Chalkley RJ. Modification site localization scoring integrated into a search engine. Mol Cell Proteomics. 2011;10(M111): 008078. https://doi.org/10.1074/mcp.M111.008078.

68. Chalkley RJ, Thalhammer A, Schoepfer R, Burlingame AL. Identification of protein O-GlcNAcylation sites using electron transfer dissociation mass spectrometry on native peptides. Proc Natl Acad Sci U S A. 2009;106:8894–9. https://doi.org/10.1073/pnas.0900288106.

69. Myers SA, Panning B, Burlingame AL. Polycomb repressive complex 2 is necessary for the normal site-specific O-GlcNAc distribution in mouse embryonic stem cells. Proc Natl Acad Sci U S A. 2011;108:9490–5. https://doi.org/10.1073/pnas.1019289108.

70. Trinidad JC, Barkan DT, Gulledge BF, Thalhammer A, Sali A, Schoepfer R, Burlingame AL. Global identification and characterization of both O-GlcNAcylation and phosphorylation at the murine synapse. Mol Cell Proteomics. 2012;11:215–29. https://doi.org/10.1074/mcp.O112.018366.

71. Nagel AK, Schilling M, Comte-Walters S, Berkaw MN, Ball LE. Identification of O-linked N-acetylglucosamine (O-GlcNAc)-modified osteoblast proteins by electron transfer dissociation tandem mass spectrometry reveals proteins critical for bone formation. Mol Cell Proteomics. 2013;12:945–55. https://doi.org/10.1074/mcp.M112.026633.

72. Xu S-L, Chalkley RJ, Maynard JC, Wang W, Ni W, Jiang X, Shin K, Cheng L, Savage D, Hühmer AFR, Burlingame AL, Wang Z-Y. Proteomic analysis reveals O-GlcNAc modification on proteins with key regulatory functions in Arabidopsis. Proc Nat Acad Sci. 2017;114:E1536–43. https://doi.org/10.1073/pnas.1610452114.

73. Bern M, Kil YJ, Becker C (2012) Byonic: advanced peptide and protein identification software. Curr Protoc Bioinformatics Chapter 13:13.20.1-13.20.14. https://doi.org/10.1002/0471250953.bi1320s40.

74. Polasky DA, Yu F, Teo GC, Nesvizhskii AI. Fast and comprehensive N- and O-glycoproteomics analysis with MSFragger-Glyco. Nat Methods. 2020;17:1125–32. https://doi.org/10.1038/s41592-020-0967-9.

75. Zeng W-F, Cao W-Q, Liu M-Q, He S-M, Yang P-Y. Precise, fast and comprehensive analysis of intact glycopeptides and modified glycans with pGlyco3. Nat Methods. 2021;18:1515–23. https://doi.org/10.1038/s41592-021-01306-0.

76. Lu L, Riley NM, Shortreed MR, Bertozzi CR, Smith LM. O-pair search with metamorpheus for o-glycopeptide characterization. Nat Methods. 2020;17:1133–8. https://doi.org/10.1038/s41592-020-00985-5.

77. Klein J, Carvalho L, Zaia J. Expanding N-glycopeptide identifications by modeling fragmentation, elution, and glycome connectivity. Nat Commun. 2024;15:6168. https://doi.org/10.1038/s41467-024-50338-5.

78. Woo CM, Lund PJ, Huang AC, Davis MM, Bertozzi CR, Pitteri SJ. Mapping and quantification of over 2000 O-linked glycopeptides in activated human T cells with isotope-targeted glycoproteomics (Isotag). Mol Cell Proteomics. 2018;17:764–75. https://doi.org/10.1074/mcp.RA117.000261.

79. Burt RA, Dejanovic B, Peckham HJ, Lee KA, Li X, Ounadjela JR, Rao A, Malaker SA, Carr SA, Myers SA. Novel antibodies for the simple and efficient enrichment of native O-GlcNAc modified peptides. Mol Cell Proteomics. 2021;20: 100167. https://doi.org/10.1016/j.mcpro.2021.100167.

80. Sakaidani Y, Nomura T, Matsuura A, Ito M, Suzuki E, Murakami K, Nadano D, Matsuda T, Furukawa K, Okajima T. O-Linked-N-acetylglucosamine on extracellular protein domains mediates epithelial cell–matrix interactions. Nat Commun. 2011;2:583. https://doi.org/10.1038/ncomms1591.

81. Cejas RB, Lorenz V, Garay YC, Irazoqui FJ. Biosynthesis of O-N-acetylgalactosamine glycans in the human cell nucleus. J Biol Chem. 2019;294:2997–3011. https://doi.org/10.1074/jbc.RA118.005524.

82. Li W, Hou C, Li Y, Wu C, Ma J. HexNAcQuest: A Tool to Distinguish O-GlcNAc and O-GalNAc. J Am Soc Mass Spectrom. 2022;33:2008–12. https://doi.org/10.1021/jasms.2c00172.

83. Halim A, Westerlind U, Pett C, Schorlemer M, Rüetschi U, Brinkmalm G, Sihlbom C, Lengqvist J, Larson G, Nilsson J. Assignment of saccharide identities through analysis of oxonium ion fragmentation profiles in LC–MS/MS of glycopeptides. J Proteome Res. 2014;13:6024–32. https://doi.org/10.1021/pr500898r.

84. Malaker SA, Penny SA, Steadman LG, Myers PT, Loke JC, Raghavan M, Bai DL, Shabanowitz J, Hunt DF, Cobbold M. Identification of glycopeptides as post-translationally modified neoantigens in leukemia. Cancer Immunol Res. 2017;5:376–84. https://doi.org/10.1158/2326-6066.CIR-16-0280.

85. Hou C, Li W, Li Y, Ma J. Integrating HexNAcQuest with glycoproteomics data analysis software to distinguish HexNAc isomers on peptides. In: Lisacek F, editor. Protein Bioinformatics. US, New York, NY: Springer; 2024. p. 67–76.

86. Hornbeck PV, Kornhauser JM, Latham V, Murray B, Nandhikonda V, Nord A, Skrzypek E, Wheeler T, Zhang B, Gnad F. 15 years of PhosphoSitePlus®: integrating post-translationally modified sites, disease variants and isoforms. Nucl Acids Res. 2019;47:D433–41. https://doi.org/10.1093/nar/gky1159.

87. Huang K-Y, Lee T-Y, Kao H-J, Ma C-T, Lee C-C, Lin T-H, Chang W-C, Huang H-D. dbPTM in 2019: exploring disease association and cross-talk of post-translational modifications. Nucleic Acids Res. 2019;47:D298–308. https://doi.org/10.1093/nar/gky1074.

88. Baker PR, Chalkley RJ. MS-viewer: a web-based spectral viewer for proteomics results. Mol Cell Proteomics. 2014;13:1392–6. https://doi.org/10.1074/mcp.O113.037200.

89. Wang J, Torii M, Liu H, Hart GW, Hu Z-Z. dbOGAP - An integrated bioinformatics resource for protein O-GlcNAcylation. BMC Bioinformatics. 2011;12:91. https://doi.org/10.1186/1471-2105-12-91.

90. Wulff-Fuentes E, Berendt RR, Massman L, Danner L, Malard F, Vora J, Kahsay R, Olivier-Van Stichelen S. The human O-GlcNAcome database and meta-analysis. Sci Data. 2021;8:25. https://doi.org/10.1038/s41597-021-00810-4.

91. Lin S, Wang C, Zhou J, Shi Y, Ruan C, Tu Y, Yao L, Peng D, Xue Y. EPSD: a well-annotated data resource of protein phosphorylation sites in eukaryotes. Brief Bioinfor. 2021;22:298–307. https://doi.org/10.1093/bib/bbz169.

92. Gupta R, Brunak S. Prediction of glycosylation across the human proteome and the correlation to protein function. Pac Symp Biocomput. 2002;2002:310–22.

93. Zhao X, Ning Q, Chai H, Ai M, Ma Z. PGlcS: Prediction of protein O-GlcNAcylation sites with multiple features and analysis. J Theoretic Biol. 2015;380:524–9. https://doi.org/10.1016/j.jtbi.2015.06.026.

94. Kao H-J, Huang C-H, Bretaña NA, Lu C-T, Huang K-Y, Weng S-L, Lee T-Y. A two-layered machine learning method to identify protein O-GlcNAcylation sites with O-GlcNAc transferase substrate motifs. BMC Bioinformatics. 2015;16:S10. https://doi.org/10.1186/1471-2105-16-S18-S10.

95. Jia C-Z, Liu T, Wang Z-P. O-GlcNAcPRED: a sensitive predictor to capture protein O-GlcNAcylation sites. Mol BioSyst. 2013;9:2909–13. https://doi.org/10.1039/C3MB70326F.

96. Jia C, Zuo Y, Zou Q. O-GlcNAcPRED-II: an integrated classification algorithm for identifying O-GlcNAcylation sites based on fuzzy undersampling and a K-means PCA oversampling technique. Bioinformatics. 2018;34:2029–36. https://doi.org/10.1093/bioinformatics/bty039.

97. Hu F, Li W, Li Y, Hou C, Ma J, Jia C. O-GlcNAcPRED-DL: Prediction of protein O-GlcNAcylation sites based on an ensemble model of deep learning. J Proteome Res. 2024;23:95–106. https://doi.org/10.1021/acs.jproteome.3c00458.

98. Pokharel S, Pratyush P, Ismail HD, Ma J, Kc DB. Integrating embeddings from multiple protein language models to improve protein O-GlcNAc site prediction. Int J Mol Sci. 2023;24:16000. https://doi.org/10.3390/ijms242116000.

99. Ruan H-B, Han X, Li M-D, Singh JP, Qian K, Azarhoush S, Zhao L, Bennett AM, Samuel VT, Wu J, Yates JR, Yang X. O-GlcNAc transferase/host cell factor C1 complex regulates gluconeogenesis by modulating PGC-1α stability. Cell Metab. 2012;16:226–37. https://doi.org/10.1016/j.cmet.2012.07.006.

100. Deng R-P, He X, Guo S-J, Liu W-F, Tao Y, Tao S-C. Global identification of O-GlcNAc transferase (OGT) interactors by a human proteome microarray and the construction of an OGT interactome. PROTEOMICS. 2014;14:1020–30. https://doi.org/10.1002/pmic.201300144.

101. Yu S-H, Boyce M, Wands AM, Bond MR, Bertozzi CR, Kohler JJ. Metabolic labeling enables selective photocrosslinking of O-GlcNAc-modified proteins to their binding partners. Proc Natl Acad Sci U S A. 2012;109:4834–9. https://doi.org/10.1073/pnas.1114356109.

102. Hu C-W, Worth M, Fan D, Li B, Li H, Lu L, Zhong X, Lin Z, Wei L, Ge Y, Li L, Jiang J. Electrophilic probes for deciphering substrate recognition by O-GlcNAc transferase. Nat Chem Biol. 2017;13:1267–73. https://doi.org/10.1038/nchembio.2494.

103. Martinez M, Renuse S, Kreimer S, O'Meally R, Natov P, Madugundu AK, Nirujogi RS, Tahir R, Cole R, Pandey A, Zachara NE. Quantitative proteomics reveals that the OGT interactome is remodeled in response to oxidative stress. Mol Cell Proteomics. 2021;20:100069. https://doi.org/10.1016/j.mcpro.2021.100069.

104. Stephen HM, Praissman JL, Wells L. Generation of an interactome for the tetratricopeptide repeat domain of O-GlcNAc transferase indicates a role for the enzyme in intellectual disability. J Proteome Res. 2021;20:1229–42. https://doi.org/10.1021/acs.jproteome.0c00604.

105. Liu Y, Nelson ZM, Reda A, Fehl C. Spatiotemporal Proximity Labeling Tools to Track GlcNAc Sugar-Modified Functional Protein Hubs during Cellular Signaling. ACS Chem Biol. 2022;17:2153–64. https://doi.org/10.1021/acschembio.2c00282.

106. Groves JA, Maduka AO, O'Meally RN, Cole RN, Zachara NE. Fatty acid synthase inhibits the O-GlcNAcase during oxidative stress. J Biol Chem. 2017;292:6493–511. https://doi.org/10.1074/jbc.M116.760785.

107. Singh JP, Qian K, Lee J-S, Zhou J, Han X, Zhang B, Ong Q, Ni W, Jiang M, Ruan H-B, Li M-D, Zhang K, Ding Z, Lee P, Singh K, Wu J, Herzog RI, Kaech S, Wendel H-G, Yates JR, Han W, Sherwin RS, Nie Y, Yang X. O-GlcNAcase targets pyruvate kinase M2 to regulate tumor growth. Oncogene. 2020;39:560–73. https://doi.org/10.1038/s41388-019-0975-3.

108. Oughtred R, Rust J, Chang C, Breitkreutz B-J, Stark C, Willems A, Boucher L, Leung G, Kolas N, Zhang F, Dolma S, Coulombe-Huntington J, Chatr-aryamontri A, Dolinski K, Tyers M. The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. Protein Science. 2021;30:187–200. https://doi.org/10.1002/pro.3978.

109. Alonso-López D, Campos-Laborie FJ, Gutiérrez MA, Lambourne L, Calderwood MA, Vidal M, De Las RJ. APID database: redefining protein–protein interaction experimental evidences and binary interactomes. Database. 2019;2019:0baz005. https://doi.org/10.1093/database/baz005.

110. Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, Broackes-Carter F, Campbell NH, Chavali G, Chen C, del-Toro N, Duesbury M, Dumousseau M, Galeota E, Hinz U, Iannuccelli M, Jagannathan S, Jimenez R, Khadake J, Lagreid A, Licata L, Lovering RC, Meldal B, Melidoni AN, Milagros M, Peluso D, Perfetto L, Porras P, Raghunath A, Ricard-Blum S, Roechert B, Stutz A, Tognolli M, van Roey K, Cesareni G, Hermjakob H (2014) The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. Nucl Acids Res 42:D358–D363. https://doi.org/10.1093/nar/gkt1115.

111. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P, Jensen LJ, von Mering C. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res. 2021;49:D605–12. https://doi.org/10.1093/nar/gkaa1074.

112. Ma J, Hou C, Li Y, Chen S, Wu C. OGT Protein Interaction Network (OGT-PIN): A curated database of experimentally identified interaction proteins of OGT. Int J Mol Sci. 2021;22:9620. https://doi.org/10.3390/ijms22179620.

113. Levine ZG, Walker S. The Biochemistry of O-GlcNAc Transferase: Which functions make it essential in mammalian cells? Annu Rev Biochem. 2016;85:631–57. https://doi.org/10.1146/annurev-biochem-060713-035344.

114. Hu C-W, Xie J, Jiang J. The emerging roles of protein interactions with O-GlcNAc cycling enzymes in cancer. Cancers (Basel). 2022;14:5135. https://doi.org/10.3390/cancers14205135.

115. Mitchell CW, Czajewski I, van Aalten DMF. Bioinformatic prediction of putative conveyers of O-GlcNAc transferase intellectual disability. J Biol Chem. 2022;298: 102276. https://doi.org/10.1016/j.jbc.2022.102276.

116. Hou C, Wu C, Zhu W, Pei H, Ma J. Systematic pan-cancer analysis reveals OGT and OGA as potential biomarkers for tumor microenvironment and therapeutic responses. Genes Dis. 2023;11: 101089. https://doi.org/10.1016/j.gendis.2023.101089.

117. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596:583–9. https://doi.org/10.1038/s41586-021-03819-2.

118. Bojar D, Lisacek F. Glycoinformatics in the Artificial Intelligence Era. Chem Rev. 2022;122:15971–88. https://doi.org/10.1021/acs.chemrev.2c00110.