Biochimica et Biophysica Acta 1844 (2014) 416–421

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbapap

Validation of the reliability of computational O-GlcNAc prediction

Ramona Jochmann [a], Patrick Holz [a], Heinrich Sticht [b], Michael Stürzl [a,*]

[a] Division of Molecular and Experimental Surgery, University Medical Center Erlangen, Friedrich-Alexander University of Erlangen-Nuremberg, Schwabachanlage 10, 91054 Erlangen, Germany
[b] Institute of Biochemistry, University of Erlangen-Nuremberg, Fahrstr. 17, 91054 Erlangen, Germany

ARTICLE INFO

Article history:
Received 28 June 2013
Received in revised form 29 November 2013
Accepted 2 December 2013
Available online 9 December 2013

Keywords:
O-GlcNAc
Prediction program
YinOYang server
dbOGAP database
OGlcNAcScan

ABSTRACT

O-GlcNAcylation is an inducible, highly dynamic and reversible posttranslational modification, which regulates numerous cellular processes such as gene expression, translation, immune reactions, protein degradation, protein–protein interaction, apoptosis, and signal transduction. In contrast to N-linked glycosylation, O-GlcNAcylation does not display a strict amino acid consensus sequence, although serine or threonine residues flanked by proline and valine are preferred sites of O-GlcNAcylation. Based on this information, computational prediction tools of O-GlcNAc sites have been developed. Here, we retrospectively assessed the performance of two available O-GlcNAc prediction programs YinOYang 1.2 server and OGlcNAcScan by comparing their predictions for recently discovered experimentally validated O-GlcNAc sites. Both prediction programs efficiently identified O-GlcNAc sites situated in an environment resembling the consensus sequence P-P-V-[*ST*]-T-A. However, both prediction programs revealed numerous false negative O-GlcNAc predictions when the site of modification was located in an amino acid sequence differing from the known consensus sequence. By searching for a common sequence motif, we found that O-GlcNAcylation of nucleocytoplasmic proteins preferably occurs at serine and threonine residues flanked downstream by proline and valine and upstream by one to two alanines followed by a stretch of serine and threonine residues. However, O-GlcNAcylation of proteins located in the mitochondria or in the secretory lumen occurs at different sites and does not follow a distinct consensus sequence. Thus, our study indicates the limitations of the presently available computational prediction methods for O-GlcNAc sites and suggests that experimental validation is mandatory. Continuously update and further development of available databases will be the key to improve the performance of O-GlcNAc site prediction.

1. Introduction

O-GlcNAcylation describes the reversible posttranslational addition of the single monosaccharide β-1,4-N-acetylglucosamine (GlcNAc) in an O-glycosidic linkage to the hydroxyl groups of serine and threonine residues. O-GlcNAc has been found on a myriad of cytoplasmic and nuclear proteins and has the ability to modulate molecular processes such as transcription, translation, protein stability, and signal transduction, as well as cellular processes including proliferation, apoptosis and development [1]. O-GlcNAc serves as a metabolic sensor, as O-GlcNAcylation is linked to the glucose metabolism of a cell: up to 3% of the incoming glucose is converted to UDP-GlcNAc, the donor sugar nucleotide for the addition of O-GlcNAc to proteins [2].

The addition and removal of O-GlcNAc are catalyzed by two unique enzymes called O-GlcNAc transferase (OGT) and O-GlcNAc hexosaminidase (O-GlcNAcase) [3]. To date, it is still not completely understood, how OGT recognizes and O-GlcNAcylates hundreds of individual protein substrates. The acceptor site for O-GlcNAc does not display a strict amino acid consensus sequence. The sequence previously suggested as consensus sequence is P-P-V-[*ST*]-T-A [4], although this sequence is not sufficient to define an O-GlcNAc site. Beside the consensus sequence, adjacent amino acids also affect O-GlcNAcylation. These include β-branched amino acids like threonine, valine, and isoleucine, as well as the structural disruptor proline, which enforce an extended conformation [4]. In addition, OGT seems to be recruited via “bridging proteins” to target substrates. As such, interaction of OGT with Tet2, an enzyme which regulates gene transcription by converting 5-methylcytosine to 5-hydroxymethylcytosine, facilitates O-GlcNAcylation of histones [5], possibly by exposing the target amino acid and making it accessible for O-GlcNAcylation or by enhancing binding affinity.

Working with O-GlcNAc-modified proteins is challenging due to the dynamic and posttranslational nature of O-GlcNAc modification, as well as laborious methods to enrich and detect O-GlcNAcylated proteins. Unbiased identification of posttranslational modification sites by experimental methods is expensive and time consuming. In contrast, advanced computational algorithms trained on available experimental data may identify functional candidates leading to reduced experimental efforts. To date, two algorithms are available, which recognize and predict O-GlcNAc sites: The YinOYang prediction program developed 2001 [6,7] and the OGlcNAcScan prediction program developed in 2011 [4].

* Corresponding author. Tel.: +49 9131 8533109; fax: +49 9131 8532077.
E-mail addresses: r.jochmann@gmx.de (R. Jochmann), patrick.holz@uk-erlangen.de (P. Holz), heinrich.sticht@fau.de (H. Sticht), michael.stuerzl@uk-erlangen.de (M. Stürzl).

1570-9639/$ – see front matter © 2013 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.bbapap.2013.12.002

The YinOYang prediction program is based on a neural network trained on 40 experimentally determined O-GlcNAc acceptor sites. It recognizes and predicts potential O-GlcNAc sites depending on the sequence context and the surface accessibility of the respective sites, which also defines the threshold. The YinOYang prediction program differentiates between four different prediction strengths, from low [marked with "+"] to very high [marked with "++++"], depending on the O-GlcNAc prediction strength and the threshold. Moreover, the YinOYang prediction program is linked to the NetPhos prediction program, which recognizes phosphorylation sites. Therefore, the YinOYang algorithm can predict as well potential Yin-Yang sites, which can be alternately modified by O-GlcNAc and O-phosphate, often resulting in a reciprocal relationship in respect to protein function [8,9].

OGlcNAcScan prediction program is based on the recently developed dbOGAP database [4], which includes a very large number of O-GlcNAc-modified proteins collected from the literature. The published version of dbOGAP contains 1163 entries, composed of 798 experimentally determined O-GlcNAcylated proteins and 365 proteins with inferred O-GlcNAc sites from validated orthologs [4]. The OGlcNAcScan prediction program is based on Support Vector Machine, a machine learning program able to categorize objects into two classes: O-GlcNAcylated and not O-GlcNAcylated. The data set used to train the OGlcNAcScan prediction program consisted of 373 experimentally identified O-GlcNAc sites from 167 proteins. The negative group consisted of the rest of the non-identified serine and threonine sites in the same protein sequences (29,897 negative sites).

As the deposited information on O-GlcNAc-modified proteins has strongly increased within the last decade and computational methods are used more frequently, the aim of this study was to assess the performance of the YinOYang and OGlcNAcScan prediction program in predicting O-GlcNAc sites. For this purpose, we performed a retrospective analysis by comparing experimentally validated O-GlcNAc sites with the predictions using the YinOYang and OGlcNAcScan prediction programs.

2. Materials and methods

2.1. Data set

Data shown in this paper were derived from recent publications of newly identified O-GlcNAc sites on O-GlcNAcylated proteins [10–21]. An O-GlcNAc site had to fulfill several criteria to be included into this study. First, it had to be unambiguously confirmed by mass spectrometry (MS) to contain an O-GlcNAc modification. Second, to enable an unbiased evaluation, the O-GlcNAcylated protein should not be included into the dbOGAP database, which was used to design the OGlcNAcScan prediction program. To rule this out, only O-GlcNAc sites published since 2011 were evaluated, as they were not part of the training data set for O-GlcNAcScan. Third, the O-GlcNAc site had to be surrounded by at least 5 amino acids in both directions, to be included in the evaluation. A total of 1181 O-GlcNAc sites on more than 520 proteins were collected. The proteins are listed in the supplementary Table S1 in an alphabetical order together with their accession number, protein function, subcellular localization and corresponding O-GlcNAc site evaluated by MS.

2.2. Prediction of O-GlcNAc-modified sites

Previously published O-GlcNAc sites were retrospectively compared to O-GlcNAc predictions using the YinOYang [6,7] and OGlcNAcScan prediction programs [4]. To this goal, the complete amino acid sequence (obtained from the UniProt database) was uploaded into the program template and submitted to O-GlcNAc prediction. For predictions using the OGlcNAcScan prediction program, the default threshold of 0.1 was used.

2.3. Estimation of the predictive performance of both algorithms

The predictive performance of a prediction program can be measured by calculating the prediction sensitivity (Sn) and specificity (Sp). The sensitivity of a prediction program indicates its ability to correctly identify positive results (here referred to as O-GlcNAcylated sites), while the specificity indicates the ability of a prediction program to identify negative results (here referred to as non-O-GlcNAcylated sites). However, conclusions about non-O-GlcNAcylated sites cannot be made with certainty, as the negative data set is usually composed of the rest of the serine and threonine residues from the known O-GlcNAcylated proteins, which may include true, but not yet identified, O-GlcNAcylation sites. Therefore, only the MS data available for O-GlcNAcylated sites were used for the calculation of the prediction sensitivity according to the following equation:

$$\mathrm{Sn} = \frac{\text{true positive predictions}}{(\text{true positive predictions} + \text{false negative predictions})}.$$

2.4. Sequence logo-based representations

The sequence logo-based representations of the sequence patterns surrounding true positively and false negatively predicted sites were generated by the *Motif* & *Logo Analysis Tool* from PhosphoSitePlus on http://www.phosphosite.org/homeAction.do [22]. The respective sequences were inserted into the mask of the *Sequence Logo Generator* and the PhosphoSitePlus Logo was generated.

2.5. Gene ontology analysis

Gene ontology analyses regarding molecular function and biological processes were performed by submitting the protein accession numbers to the Panther database (http://www.pantherdb.org/) [23,24]. The subcellular localization of the analyzed proteins was attributed with the help of the UniProt/SwissProt server (http://www.uniprot.org/uniprot/) [25].

3. Results and discussion

A total of 1181 experimentally validated O-GlcNAc sites on more than 520 proteins were collected [10–21]. Selection criteria are summarized in the materials and methods section. The proteins are listed in the supplementary Table S1 in an alphabetical order together with their accession number, protein function, subcellular localization and corresponding O-GlcNAc site evaluated by MS. Using the YinOYang prediction program, 518 out of the 1181 O-GlcNAcylated sites were predicted with at least low prediction strength to be O-GlcNAcylated, while 663 sites were false negative predictions (Table 1 and supplementary Table S2). Similarly, using the OGlcNAcScan algorithm, 354 out of the 1181 evaluated O-GlcNAcylated sites were predicted to be O-GlcNAc-modified using the

Table 1
Evaluation of the predictive performance of the two O-GlcNAc prediction programs YinOYang and OGlcNAcScan.

	Prediction of O-GlcNAc–modified site		Sensitivity
	Predicted	Not predicted	
YinOYang	518 [44%]	663 [56%]	44%
Taking into account O-GlcNAc prediction at adjacent sites	588 [50%]	592 [50%]	50%
O-GlcNAcScan	354 [30%]	827 [70%]	30%
Taking into account O-GlcNAc prediction at adjacent sites	437 [37%]	744 [63%]	37%

default threshold of 0.1, while the remaining 827 sites were false negative predictions (Table 1 and supplementary Table S3).

In order to evaluate the prediction performance of the two prediction programs, the ability to correctly predict true O-GlcNAcylated sites (sensitivity, Sn) was calculated (Table 1). The YinYang prediction program exhibits a sensitivity of 44%, whereas the OGlcNAcScan prediction program performs with a sensitivity of 30%, which is rather low compared to the values of more than 70% reported for other sequence-based predictions of protein features [26–29]. These relatively low values might be explained by the fact that the thresholds used by prediction programs generally represent a compromise between a reasonable sensitivity and specificity. Hence, a high sensitivity (correct prediction of all true positive O-GlcNAc sites) goes to the expense of a high specificity (correct prediction of all true negative O-GlcNAc sites), and vice versa. In this context it is important to mention that the OGlcNAcScan algorithm provides an estimated precision threshold as a measure of prediction accuracy. The default threshold is set at 0.1, but can be adjusted manually. Therefore, it is conceivable that a decrease in the threshold leads to an improved sensitivity of the algorithm, but unfortunately also to lower specificity. Thus, the problem of low sensitivity cannot readily be solved by reducing the threshold for the prediction of glycosylation sites because this would also significantly increase the number of false positive hits.

However, OGT exhibits some inaccuracy when it comes to adjacent O-GlcNAc sites. Therefore, we investigated the predictive performance of both programs under the condition that the positive prediction of a site adjacent (+/−1) to a negatively predicted O-GlcNAc site is regarded as a positive prediction. The predictions by YinOYang server included in total 204 false negative predicted O-GlcNAc sites containing adjacent serine or threonine residues. Of these, 70 adjacent sites were predicted to be O-GlcNAc-modified (supplementary Table S4), which leads in total to 588 true positive predictions and to an increased sensitivity of 50% (Table 1). The predictions by the O-GlcNAcScan prediction program included 296 false negative predicted O-GlcNAc sites containing adjacent serine or threonine residues. Of these, 83 adjacent sites were predicted to be O-GlcNAcylated (supplementary Table S4), leading in total to 437 true positive predictions and to a still very low sensitivity of 37% (Table 1). Thus, even when looking at the predictive performance on adjacent sites, both prediction programs exhibit a very low sensitivity.

Although, we cannot completely rule out false positive identification of O-GlcNAcylated sites by MS, the high percentage of false negative predictions of both prediction programs (up to 70% of all predictions) led to the hypothesis that many of the experimentally validated O-GlcNAc sites might be situated in a sequence surrounding differing from the known consensus motif. In order to identify differences in the sequence patterns of true positively and false negatively predicted O-GlcNAcylation sites, graphical sequence logo-based representations were performed. To this goal, the experimentally validated O-GlcNAcylated sites were divided into two groups: true positively predicted O-GlcNAc sites and false negatively predicted O-GlcNAc sites (supplementary Table S2 and S3). The sequence logo-based representation of the sequence pattern surrounding the true positively predicted O-GlcNAcylation sites identified for both prediction programs a sequence motif x-x-P-P-V-[*ST*]-[AT]-A-S-[ST]-[STV], which strongly resembles the previously described sequence motif (P-P-V-[*ST*]-T-A), where letters in italics indicate the position of O-GlcNAcylation (Fig. 1A and C). Especially the accumulation of serine/threonine residues has also been observed by others [10]. Unexpectedly, the sequence pattern surrounding the false negatively predicted O-GlcNAcylation sites does not exhibit a clearly distinct recognition motif (Fig. 1B and D). Common features are the two alanine residues at the +1/+2 positions and the stretch of serine and threonine residues at +3–+5 positions. By searching for a common sequence motif, we found that O-GlcNAcylation preferably occurs at serine and threonine residues flanked downstream by proline and valine and upstream by two alanines followed by a stretch of serine and threonine residues (Fig. 1E). However, from this evaluation, it also becomes evident that O-GlcNAcylation does not follow a stringent recognition motif. The lack of such a defined recognition motif might explain the poor predictive outcome of the two prediction programs.

In order to investigate whether there are any correlations between the motif properties and the protein function, gene ontology classifications using the panther gene ontology software were performed. Gene ontology analysis showed that proteins containing either true positively or false negatively predicted O-GlcNAc sites are involved in a variety of molecular functions and biological processes (Supplementary Fig. S1). However, although the distribution is slightly different, no significant correlation between the predictive performance and molecular function (Supplementary Fig. 1A) or biological process (Supplementary Fig. 1B) was detected.

Second, the subcellular localization of the studied proteins was addressed (see supplemental Table S1, 4th column) using the UniProt server. To date, several OGT isoforms are known, which localize to the mitochondria (mOGT) or the nucleocytoplasmic compartment (ncOGT). Moreover, an additional gene has recently been identified, encoding an OGT variant which localizes into the lumen of the ER (eOGT) and is responsible for O-GlcNAcylation of extracellular proteins [30]. As such proteins are co-translationally integrated into the ER and Golgi, they are unlikely to be modified by ncOGT. We thus investigated whether different OGT variants prefer different sites. Therefore, sequence logo-based representations of the respective peptides located either on nucleocytoplasmic, mitochondrial or secreted proteins were performed (supplemental Table S6-S8 and Fig. 2). Interestingly, O-GlcNAc peptides from proteins localizing to the nucleocytoplasmic (nc) compartment exhibited a preference for the motif x-x-P-P-V-[*ST*]-[AT]-A-S-T-T (Fig. 2A), which is strongly similar to the described motif P-P-V-[*ST*]-T-A. In contrast, O-GlcNAc peptides from proteins localizing to the mitochondria or the extracellular site (secretory lumen) exhibited a rather different sequence motif (Fig. 2B and C). The only common characteristics are the valine at position −1 and the alanine at position +2. This indicates that all three OGT variants have different target sequences. Thus, it might be helpful to include this information into the prediction programs.

Based on the present analysis, however, we conclude that both prediction programs have a rather low sensitivity, which is due to (1) the low number of O-GlcNAc sites used to train both programs (40 O-GlcNAc sites for YinOYang, 373 O-GlcNAc sites for dbOGAP), and (2) the existence of different protein substrate recognition patterns of OGT. The variety of substrate recognition patterns is not adequately covered by the present prediction tools. Even a very recently developed prediction program named O-GlcNAcPRED, based on a support vector machine (SVM) model using the novel adapted normal distribution bi-profile Bayes (ANBPB) feature extraction method, performed only slightly better than the two other prediction programs [31]. This implies that prediction of O-GlcNAc sites based on the primary amino acid sequence requires the continuous improvement of prediction tools based on the larger data sets that became recently available. Hence, the benefit of computational prediction methods to identify O-GlcNAc sites is yet limited and experimental validation is mandatory.

4. Conclusion

O-GlcNAcylation occurs on serine/threonine residues embedded into a stretch of neutral or non-polar amino acids. Specifically O-GlcNAcylation of nucleocytoplasmic proteins follows a more or less stringent sequence motif characterized by downstream proline and valine residues and upstream by one to two alanines followed by a stretch of serine and threonine residues. However, O-GlcNAcylation of mitochondrial and secreted proteins does not follow the same pattern. Our data enforces the assumption that O-GlcNAcylation does not occur at a strict amino acid consensus sequence and suggests that prediction of O-GlcNAc sites solely based on the peptide sequence requires

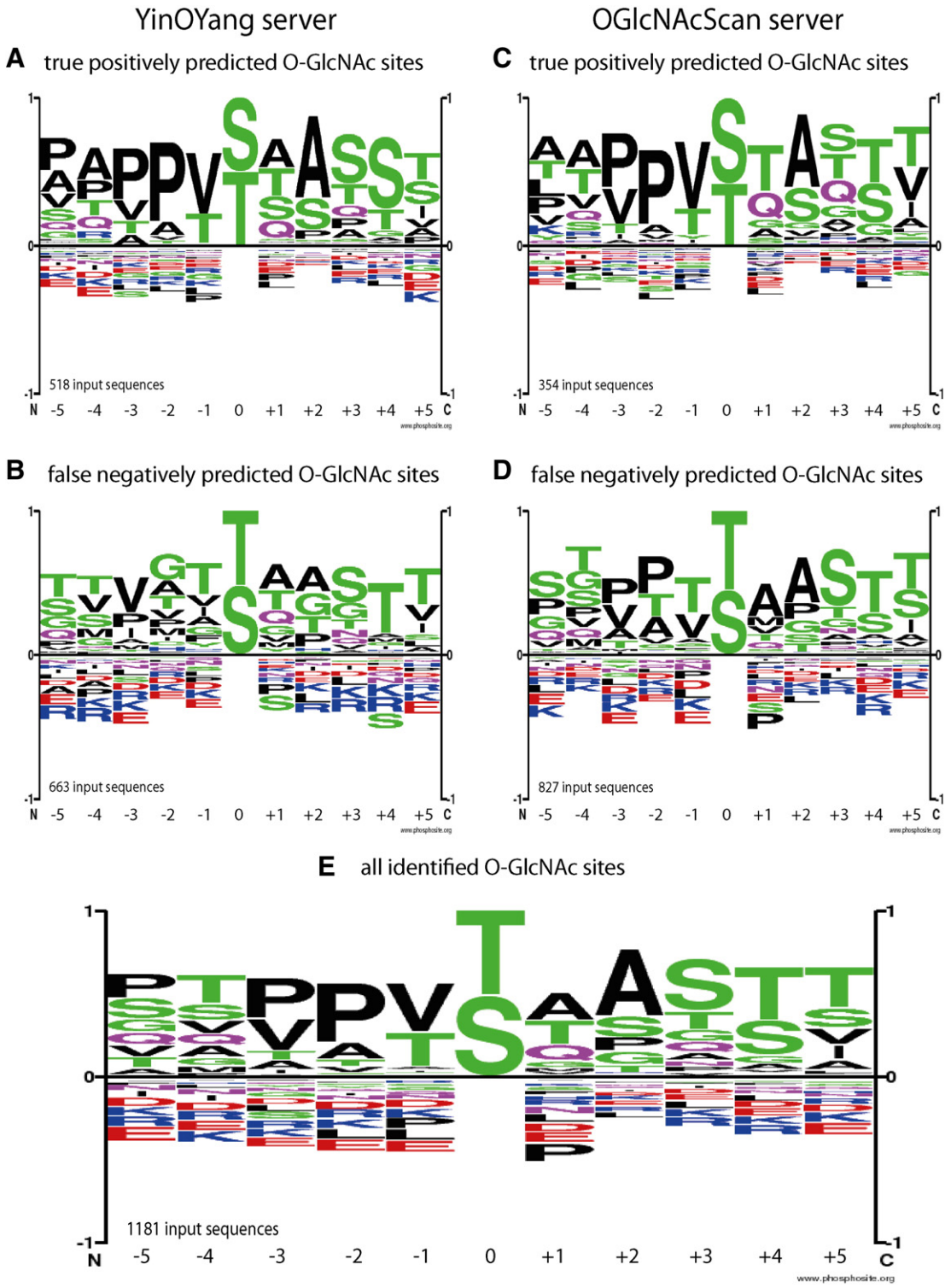

Fig. 1. Sequence-logo-based graphical representation of the sequence patterns surrounding experimentally validated O-GlcNAc sites. The upper panels represent the sequence motifs true positively predicted to be modified with O-GlcNAc using the YinOYang (A) or the OGlcNAcScan prediction program (C). The middle panels represent the sequence motifs of the sites which failed to be recognized by the YinOYang (B) and OGlcNAcScan prediction program (D) as O-GlcNAc-modified (false negatively predicted). The lower panels represent the common sequence motif of all experimentally validated O-GlcNAc sites. The modified serine or threonine residue is at position 0, indicated on the x-axis (from −5 to +5). The height of the amino acid represents the relative frequency in the surrounding sequence. The graphical representation was obtained from www.phosphosite.org by uploading the sequences of the O-GlcNAc-modified sites presented in the supplementary Tables S2 and S3.

the continuous improvement of prediction tools based on the larger data sets that became recently available. In addition, several other processes might influence O-GlcNAcylation of a protein, such as tyrosine phosphorylation [32] and binding to bridging factors [5]. Especially the latter might bring OGT in the spatial proximity to potential O-GlcNAc sites, which would not be targeted by OGT alone. Altogether, these factors yet hamper the reliable prediction of O-GlcNAc sites.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.bbapap.2013.12.002.

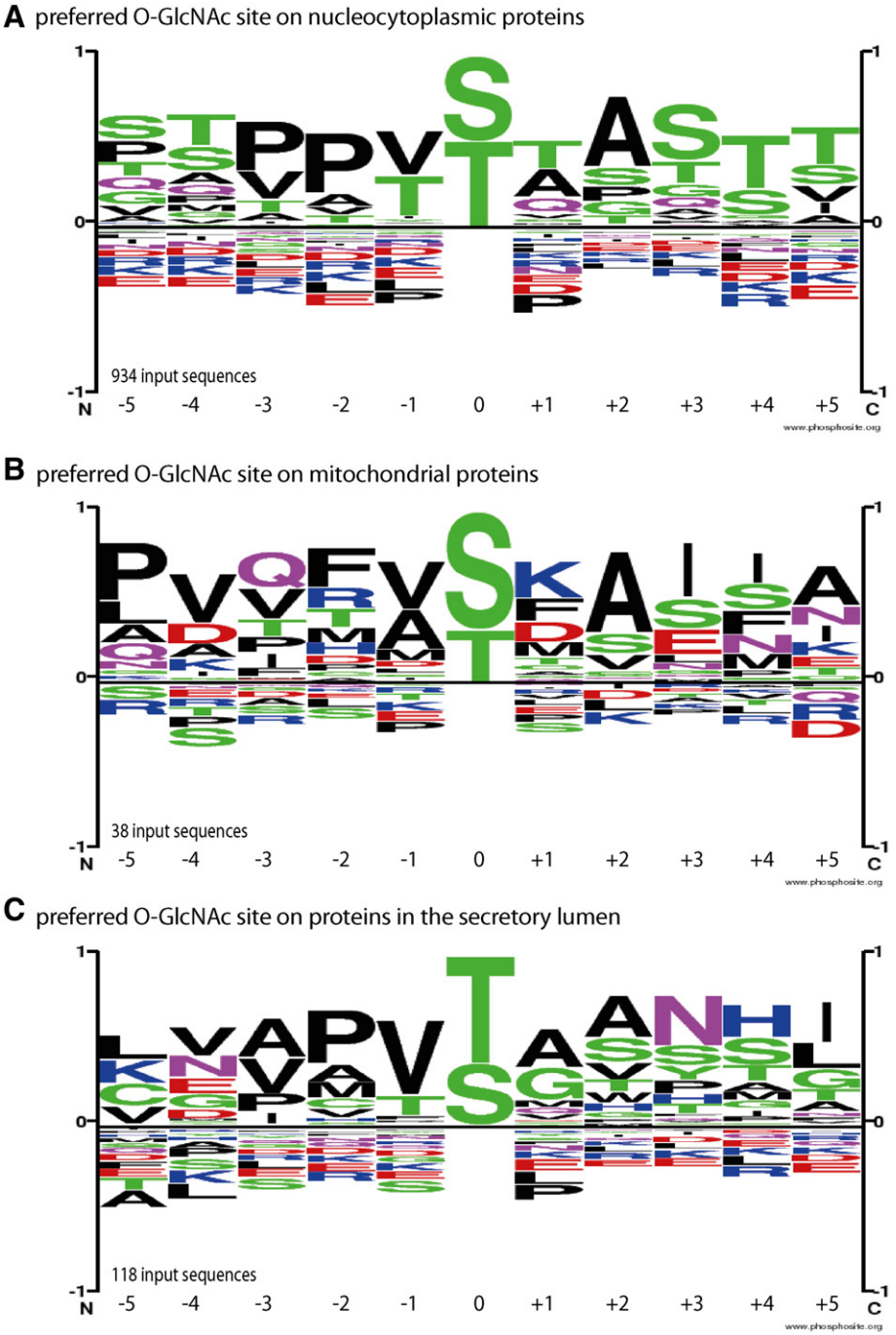

Fig. 2. Sequence-logo-based graphical representation of the O-GlcNAc-peptides clustered by subcellular localization. The sequence motif was obtained from all peptides from proteins localized either in the nucleocytoplasmic compartment (A), mitochondria (B), or secretory lumen (C). The modified serine or threonine residue is at position 0, indicated on the x-axis (from −5 to +5). The height of the amino acid represents the relative frequency in the surrounding sequence. The graphical representation was obtained from www.phosphosite.org by uploading the sequences of the O-GlcNAc-modified sites presented in the supplementary Tables S6–S8.

Funding

This work was supported by grants of the Erlangen's fund for performance-based start-up funding and promotion of young researchers (ELAN) of the University Medical Center Erlangen to R.J, by grants of the Interdisciplinary Centre for Clinical Research (IZKF) of the University Medical Center Erlangen to M.S. and to R.J., and by grants of the German Federal Ministry of Education and Research (BMBF, Polyprobe-Study), the German Research Foundation [DFG-GRK1071 (subproject A2), STU238/6-1, SFB796 (subproject B9), KFO257 (subproject 4)] and the German Cancer Aid (109510) to M.S.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

[1] F.I. Comer, G.W. Hart, O-GlcNAc and the control of gene expression, Biochim. Biophys. Acta 1473 (1999) 161–171.
[2] S. Marshall, V. Bacote, R.R. Traxinger, Discovery of a metabolic pathway mediating glucose-induced desensitization of the glucose transport system. Role of hexosamine biosynthesis in the induction of insulin resistance, J. Biol. Chem. 266 (1991) 4706–4712.
[3] S.P. Iyer, G.W. Hart, Dynamic nuclear and cytoplasmic glycosylation: enzymes of O-GlcNAc cycling, Biochemistry 42 (2003) 2493–2499.

[4] J. Wang, M. Torii, H. Liu, G.W. Hart, Z.Z. Hu, dbOGAP — an integrated bioinformatics resource for protein O-GlcNAcylation, BMC Bioinforma. 12 (2011) 91.
[5] Q. Chen, Y. Chen, C. Bian, R. Fujiki, X. Yu, TET2 promotes histone O-GlcNAcylation during gene transcription, Nature 493 (2013) 561–564.
[6] R. Gupta, Prediction of glycosylation sites in proteomes: from post-translational modifications to protein function, (Ph.D. thesis at CBS) 2001.
[7] R. Gupta, S. Brunak, Prediction of glycosylation across the human proteome and the correlation to protein function, pacific symposium on biocomputing, Pac. Symp. Biocomput. (2002) 310–322.
[8] G.W. Hart, K.D. Greis, L.Y. Dong, M.A. Blomberg, T.Y. Chou, M.S. Jiang, E.P. Roquemore, D.M. Snow, L.K. Kreppel, R.N. Cole, et al., O-linked N-acetylglucosamine: the "yin-yang" of Ser/Thr phosphorylation? Nuclear and cytoplasmic glycosylation, Adv. Exp. Med. Biol. 376 (1995) 115–123.
[9] G.W. Hart, L.K. Kreppel, F.I. Comer, C.S. Arnold, D.M. Snow, Z. Ye, X. Cheng, D. DellaManna, D.S. Caine, B.J. Earles, Y. Akimoto, R.N. Cole, B.K. Hayes, O-GlcNAcylation of key nuclear and cytoskeletal proteins: reciprocity with O-phosphorylation and putative roles in protein multimerization, Glycobiology 6 (1996) 711–716.
[10] J.C. Trinidad, D.T. Barkan, B.F. Gulledge, A. Thalhammer, A. Sali, R. Schoepfer, A.L. Burlingame, Global identification and characterization of both O-GlcNAcylation and phosphorylation at the murine synapse, Mol. Cell. Proteomics 11 (2012) 215–229.
[11] R. Jochmann, J. Pfannstiel, P. Chudasama, E. Kuhn, A. Konrad, M. Stürzl, O-GlcNAc transferase inhibits KSHV propagation and modifies replication relevant viral proteins as detected by systematic O-GlcNAcylation analysis, Glycobiology 23 (2013) 1114–1130.
[12] H. Hahne, A. Moghaddas Gholami, B. Kuster, Discovery of O-GlcNAc-modified proteins in published large-scale proteome data, Mol. Cell. Proteomics 11 (2012) 843–850.
[13] H. Hahne, N. Sobotzki, T. Nyberg, D. Helm, V.S. Borodkin, D.M. van Aalten, B. Agnew, B. Kuster, Proteome wide purification and identification of O-GlcNAc-modified proteins using click chemistry and mass spectrometry, J. Proteome Res. 12 (2013) 927–936.
[14] D.F. Allison, J.J. Wamsley, M. Kumar, D. Li, L.G. Gray, G.W. Hart, D.R. Jones, M.W. Mayo, Modification of RelA by O-linked N-acetylglucosamine links glucose metabolism to NF-κB acetylation and transcription, Proc. Natl. Acad. Sci. 109 (2012) 16888–16893.
[15] T. Gawlowski, J. Suarez, B. Scott, M. Torres-Gonzalez, H. Wang, R. Schwappacher, X. Han, J.R. Yates III, M. Hoshijima, W. Dillmann, Modulation of dynamin-related protein 1 (DRP1) function by increased O-linked-beta-N-acetylglucosamine modification (O-GlcNAc) in cardiac myocytes, J. Biol. Chem. 287 (2012) 30024–30034.
[16] S. Wang, X. Huang, D. Sun, X. Xin, Q. Pan, S. Peng, Z. Liang, C. Luo, Y. Yang, H. Jiang, M. Huang, W. Chai, J. Ding, M. Geng, Extensive crosstalk between O-GlcNAcylation and phosphorylation regulates Akt signaling, PLoS One 7 (2012) e37427.
[17] Z.E. Floyd, J.M. Stephens, Controlling a master switch of adipocyte development and insulin sensitivity: covalent modifications of PPARγ, Biochim. Biophys. Acta (BBA) - Mol. Basis Dis. 1822 (2012) 1090–1095.
[18] S. Ji, S.Y. Park, J. Roth, H.S. Kim, J.W. Cho, O-GlcNAc modification of PPARgamma reduces its transcriptional activity, Biochem. Biophys. Res. Commun. 417 (2012) 1158–1163.
[19] J.F. Alfaro, C.X. Gong, M.E. Monroe, J.T. Aldrich, T.R. Clauss, S.O. Purvine, Z. Wang, D.G. Camp II, J. Shabanowitz, P. Stanley, G.W. Hart, D.F. Hunt, F. Yang, R.D. Smith, Tandem mass spectrometry identifies many mouse brain O-GlcNAcylated proteins including EGF domain-specific O-GlcNAc transferase targets, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 7280–7285.
[20] S. Pathak, V.S. Borodkin, O. Albarbarawi, D.G. Campbell, A. Ibrahim, D.M.F. van Aalten, O-GlcNAcylation of TAB1 modulates TAK1-mediated cytokine release, EMBO J. 31 (2012) 1394–1404.
[21] T. Overath, U. Kuckelkorn, P. Henklein, B. Strehl, D. Bonar, A. Kloss, D. Siele, P.M. Kloetzel, K. Janek, Mapping of O-GlcNAc sites of 20S proteasome subunits and Hsp90 by a novel biotin-cystamine tag, Mol. Cell. Proteomics 11 (2012) 467–477.
[22] P.V. Hornbeck, J.M. Kornhauser, S. Tkachev, B. Zhang, E. Skrzypek, B. Murray, V. Latham, M. Sullivan, PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse, Nucleic Acids Res. 40 (2012) D261–D270.
[23] P.D. Thomas, M.J. Campbell, A. Kejariwal, H. Mi, B. Karlak, R. Daverman, K. Diemer, A. Muruganujan, A. Narechania, PANTHER: a library of protein families and subfamilies indexed by function, Genome Res. 13 (2003) 2129–2141.
[24] H. Mi, B. Lazareva-Ulitsky, R. Loo, A. Kejariwal, J. Vandergriff, S. Rabkin, N. Guo, A. Muruganujan, O. Doremieux, M.J. Campbell, H. Kitano, P.D. Thomas, The PANTHER database of protein families, subfamilies, functions and pathways, Nucleic Acids Res. 33 (2005) D284–D288.
[25] E. Jain, A. Bairoch, S. Duvaud, I. Phan, N. Redaschi, B.E. Suzek, M.J. Martin, P. McGarvey, E. Gasteiger, Infrastructure for the life sciences: design and implementation of the UniProt website, BMC Bioinforma. 10 (2009) 136.
[26] X. Zhao, W. Zhang, X. Xu, Z. Ma, M. Yin, Prediction of protein phosphorylation sites by using the composition of k-spaced amino acid pairs, PLoS One 7 (2012) e46302.
[27] P. Duckert, S. Brunak, N. Blom, Prediction of proprotein convertase cleavage sites, Protein Eng. Des. Sel. 17 (2004) 107–112.
[28] N. Blom, S. Gammeltoft, S. Brunak, Sequence and structure-based prediction of eukaryotic protein phosphorylation sites, J. Mol. Biol. 294 (1999) 1351–1362.
[29] L. Kiemer, J.D. Bendtsen, N. Blom, NetAcet: prediction of N-terminal acetylation sites, Bioinformatics 21 (2005) 1269–1270.
[30] Y. Sakaidani, N. Ichiyanagi, C. Saito, T. Nomura, M. Ito, Y. Nishio, D. Nadano, T. Matsuda, K. Furukawa, T. Okajima, O-linked-N-acetylglucosamine modification of mammalian Notch receptors by an atypical O-GlcNAc transferase Eogt1, Biochem. Biophys. Res. Commun. 419 (2012) 14–19.
[31] C.-Z. Jia, T. Liu, Z.-P. Wang, O-GlcNAcPRED: a sensitive predictor to capture protein O-GlcNAcylation sites, Mol. BioSyst. 9 (2013) 2909–2913.
[32] S.R. Ande, S. Moulik, S. Mishra, Interaction between O-GlcNAc modification and tyrosine phosphorylation of prohibitin: implication for a novel binary switch, PLoS One 4 (2009) e4586.