SUPPLEMENTAL FIGURES

Discovery of O-GlcNAc proteins in published large-scale proteome data

Hannes Hahne[1], Amin Moghaddas Gholami[1], Bernhard Kuster[1, 2,*]

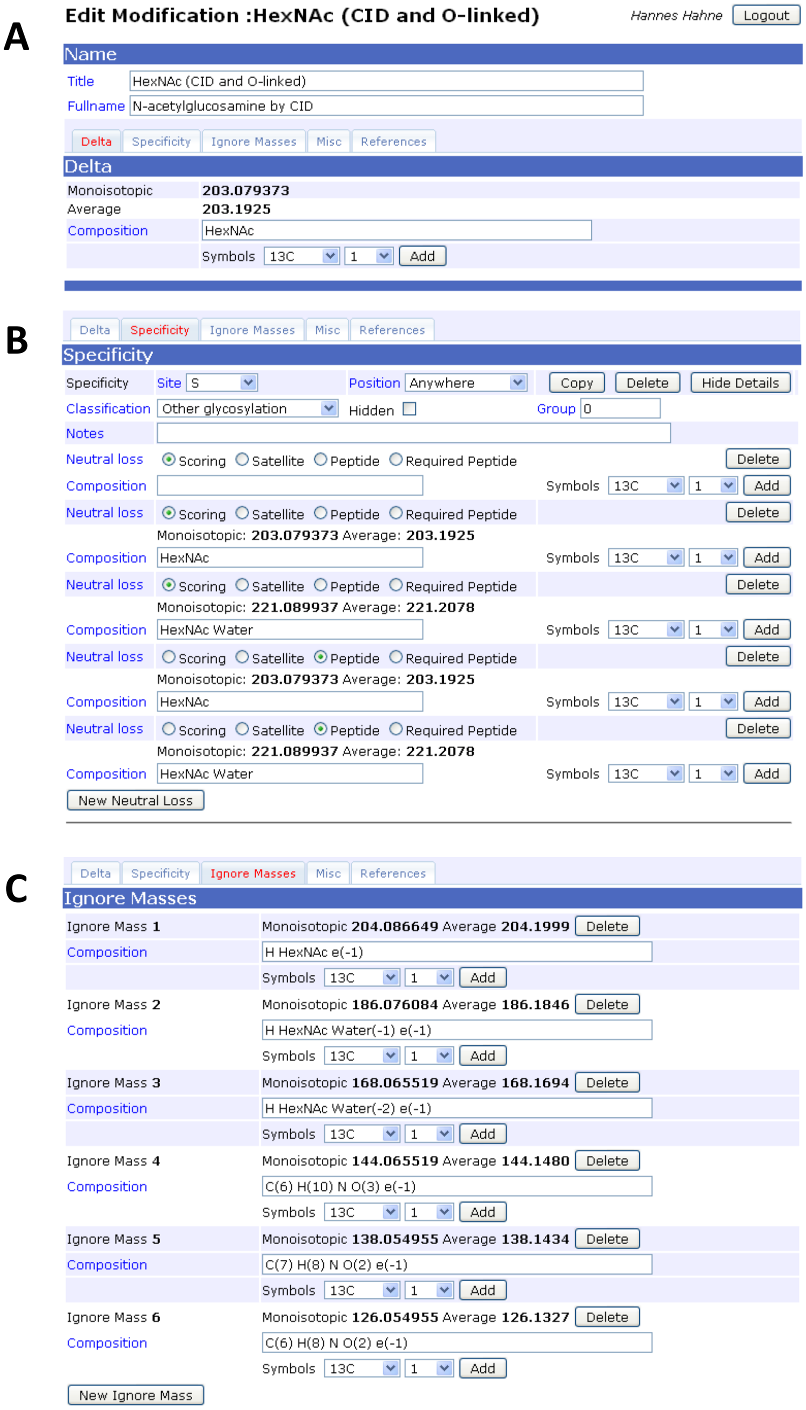
[1] Chair for Proteomics and Bioanalytics, Center of Life and Food Sciences Weihenstephan, Technische Universität München, Emil-Erlenmeyer-Forum 5, 85354 Freising, Germany

[2] Center for Integrated Protein Science Munich, Emil-Erlenmeyer-Forum 5, 85354 Freising, Germany

[*] Corresponding author
Bernhard Kuster
e-Mail: kuster@tum.de
Tel. +49 8161 715696
Fax: +49 8161 715931

Running Title: O-GlcNAc protein identification

Supplemental Figures



Edit Modification :HexNAc (CID and O-linked)

Hannes Hahne

Name

Title HexNAc (CID and O-linked)

Fullname N-acetylglucosamine by CID

Delta

Monoisotopic **203.079373**

Average **203.1925**

Composition HexNAc

Symbols 13C 1

Specificity

Specificity Site S Position Anywhere

Classification Other glycosylation Hidden ☐ Group 0

Notes

Neutral loss ● Scoring ○ Satellite ○ Peptide ○ Required Peptide

Composition Symbols 13C 1

Neutral loss ● Scoring ○ Satellite ○ Peptide ○ Required Peptide

Monoisotopic: **203.079373** Average: **203.1925**

Composition HexNAc Symbols 13C 1

Neutral loss ● Scoring ○ Satellite ○ Peptide ○ Required Peptide

Monoisotopic: **221.089937** Average: **221.2078**

Composition HexNAc Water Symbols 13C 1

Neutral loss ○ Scoring ○ Satellite ● Peptide ○ Required Peptide

Monoisotopic: **203.079373** Average: **203.1925**

Composition HexNAc Symbols 13C 1

Neutral loss ○ Scoring ○ Satellite ● Peptide ○ Required Peptide

Monoisotopic: **221.089937** Average: **221.2078**

Composition HexNAc Water Symbols 13C 1

Ignore Masses

Ignore Mass **1** Monoisotopic **204.086649** Average **204.1999**

Composition H HexNAc e(-1)

Symbols 13C 1

Ignore Mass **2** Monoisotopic **186.076084** Average **186.1846**

Composition H HexNAc Water(-1) e(-1)

Symbols 13C 1

Ignore Mass **3** Monoisotopic **168.065519** Average **168.1694**

Composition H HexNAc Water(-2) e(-1)

Symbols 13C 1

Ignore Mass **4** Monoisotopic **144.065519** Average **144.1480**

Composition C(6) H(10) N O(3) e(-1)

Symbols 13C 1

Ignore Mass **5** Monoisotopic **138.054955** Average **138.1434**

Composition C(7) H(8) N O(2) e(-1)

Symbols 13C 1

Ignore Mass **6** Monoisotopic **126.054955** Average **126.1327**

Composition C(6) H(8) N O(2) e(-1)

Symbols 13C 1

Figure S1. HexNAc modification definition for Mascot. In the standard setting, Mascot provides the proper composition and exact mass for HexNAc. However, the standard modification definition should be extended as depicted in the “Specificity” (B) and in the “Ignore Masses” tab (C). To achieve proper matching of sequence-specific peptide fragments, it is important to set the

HexNAc fragment neutral loss as “scoring” as depicted in panel B. These fragment neutral loss peaks are then considered for scoring and significantly improve the matching of sequence-specific ions and the Mascot score. Otherwise, these fragment neutral loss peaks would be treated as noise and reduce the Mascot ion score. The “peptide” neutral loss setting matches the neutral loss from the intact peptide, which is then not treated as noise peak for scoring. As depicted in panel C, the HexNAc oxonium ion as well as its fragments should be included in the “Ignore Masses” list. A detailed explanation of the modification definitions can also be found at the Mascot website (www.matrixscience.com/help/pt_mods_help.html).

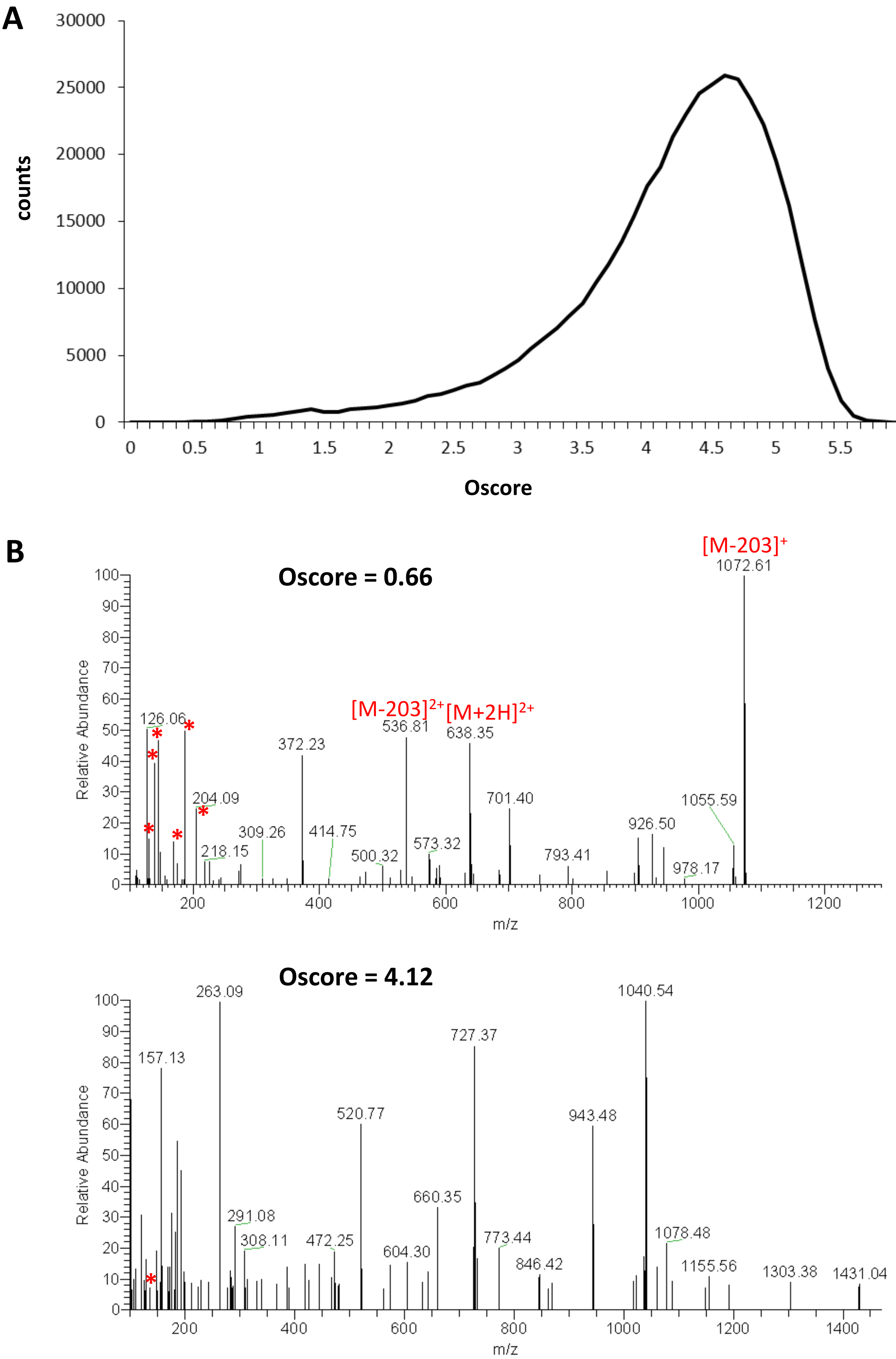


Figure S2. Typical Oscore distribution and representative spectra for low and high Oscores. (**A**) The Oscore distribution is based on 414,599 spectra (complete Geiger *et al.* data set), for which an Oscore could be calculated. (**B**) The upper spectrum represents a typical O-GlcNAc case with a low Oscore, where most of the abundant fragment ions point to the O-GlcNAc modification (indicated in red). The spectrum below represents a typical “high Oscore” case and only contains one weak fragment ion (m/z 126.05) diagnostic for O-GlcNAc.

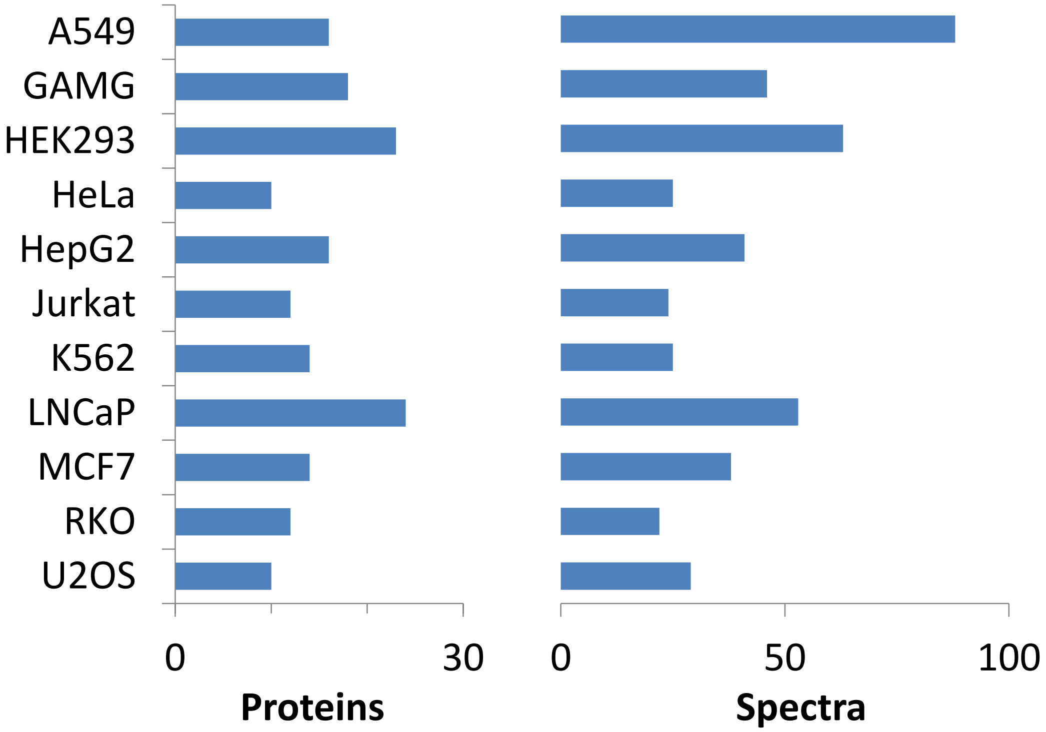


Figure S3. O-GlcNAc protein identifications and O-GlcNAc spectra across all eleven cell lines analyzed by Geiger *et al.* (Geiger, T., Wehner, A., Schaab, C., Cox, J., and Mann, M. Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. Mol. Cell. Proteomics 11, M111 014050).