BEILSTEIN INSTITUT

Discovering the Subtleties of Sugars
June 10th – 14th, 2013, Potsdam, Germany

Utilising the Carbohydrate Fragmentation Database UniCarb-DB for Glyco Research

Catherine A. Hayes, Sarah A. Flowers, Liaqat Ali, Samah M. A. Issa, Chunsheng Jin and Niclas G. Karlsson*

Medical Biochemistry, Sahlgrenska Academy, University of Gothenburg, Box 440 405 30 Gothenburg, Sweden

E-Mail: *niclas.karlsson@medkem.gu.se

Received: 5th December 2013 / Published: 22nd December 2014

Abstract

UniCarb-DB is an experimental database consisting of structural information of *O*-linked and *N*-linked oligosaccharides and associated LC-MS/MS fragmentation data. This report illustrates how the database can be useful for future software development for the interpretation of glycomic mass spectrometric data. The information about rich fragmentation spectra generated for *O*-linked oligosaccharides in negative ion mode allows matching with candidate spectra in the database for accurate assignment of oligosaccharide sequence and in some cases, linkage information. Furthermore, peak matching can be used for selected *m/z* regions to identify sites of sulfation and identification of the sequence of the neutral oligosaccharide backbone of sialylated structures. The reproducibility of the fragmentation patterns of oligosaccharides present in the database suggests that targeted mass spectrometric approaches can be developed for glycomic discovery and validation, using such methods as multiple/selected reaction monitoring (MRM/SRM).

This article is part of the Proceedings of the Beilstein Glyco-Bioinformatics Symposium 2013.
www.proceedings.beilstein-symposia.org

Introduction

The lack of structural databases for glycomic research is a key obstacle for successful high throughput glycomic analysis to identify the role of glycosylation in life science. Progress has been made in the last couple of years. While the pioneering work to establish CarbBank [1] as a universal carbohydrate databank was discontinued in the 90s, other groups have now taken up the challenge that will allow glycomic structural databases to grow again. For instance, the UniCarbKB project [2] is building on the framework established for the GlycoSuite database [3], capturing structural information of glycoproteins from the scientific literature. Another example is GlycomeDB [4] which aims to amalgamate the information about carbohydrate structures available in existing databases into one depository. Publically available experimental glycomic databases, where structural data are associated with characterisation information, have been established for NMR [5] HPLC [6] and LC-MS^2 data [7]. The latter, named UniCarb-DB, is based on *de novo* sequencing of both *N*-linked and *O*-linked oligosaccharides primarily based on fragmentation analysis in negative ion mode. To date, the database contains over 500 MS/MS spectra representing 416 uniquely defined structures. The association of structures with MS/MS spectra allows the database to be utilised in downstream glycomic MS based analysis to match sample spectra with consensus spectra from the database, and provides a pathway to design targeted MS experiments of predetermined components in SRM. An SRM approach has been shown to be an effective pathway to monitor low abundant sulfated oligosaccharides present in inflammatory diseases [8].

Methods

MS analysis of O-linked oligosaccharides

Ion trap LC-MS/MS spectra of neutral and sialylated structures were generated from oligosaccharides released from human gastric mucins as described [9]. LTQ orbitrap (Thermo Electron, San Jose, CA, US) Higher energy Collision induced Dissociation (HCD) was performed on *O*-linked oligosaccharides released from porcine gastric mucin (PGM) (Sigma, St Lois, MO, US) or human salivary MUC 5B [10]. The Obritrap was calibrated and tuned in negative ion mode with the manufacturer's standard mixture. Released *O*-linked oligosaccharides were fragmented in the HCD cell with a normalised energy of 90. The eluate from a graphitized carbon capillary HPLC column (produced and run as described [9]) system was introduced using the standard ion max source. Ions were generated and focused using an ESI voltage of – 3,750 V, a sheath gas flow of 20 L/min, and a capillary temperature of 250 °C. MS data acquisition was carried out with the LTQ-Orbitrap mass spectrometer scanning in negative ion mode over *m/z* 700 – 1650, with a resolution of 30,000 at *m/z* 400. This was followed by data dependent MS^2 scans of the three most abundant ions in each scan

(2 microscans, maximum 500 ms, target value of 100,000). The signal threshold for MS^2 was set to 5,000 counts with an isolation window of 3 Da, and an activation time of 10 ms was used.

Matching of fragmentation data with spectral library

MS/MS peak lists from fragmentation of oligosaccharides were centroided using Xcalibur and compared with the spectra in UniCarb-DB. The R package OrgMassSpecR[3] was used, specifically the SpectrumSimilarity function. This function takes two spectra and compares them. The output is a similarity score based on the normal dot product.

A subset library of 231 MS/MS spectra from UniCarb-DB was created for matching with the neutral and sialylated structures. A test-set of seven unknown spectra was used. This set included neutral and sialylated structures. Matching of the human gastric neutral and sialylated MS/MS spectra with the database against the 231-structure database was performed using a precursor mass filter, $t = 0.25$ and $b = 5$, where t is the mass tolerance (m/z) used to align the spectra, and b is the baseline threshold for peak identification (expressed as % of max. intensity). Sulfated oligosaccharides from PGM and MUC 5B was matched with a smaller spectral library consisting of sulfated *N*-acetyllactosamine with known sulfate position (generous gift from James Paulson, Scripps Research Institute, La Jolla, CA, USA).

Results

MS fragmentation of neutral, sialylated and sulfated oligosaccharides provides different information that needs to be considered when utilising UniCarb-DB fragmentation data

The chemical nature of *O*-linked oligosaccharide subclasses i. e., neutral, sialylated and sulfated oligosaccharides, influences their fragmentation in negative ion mode (Figure 1). The negative charge introduced on neutral oligosaccharides can be spread throughout the molecules by removing a proton from either of the omnipresent alcohol groups throughout the oligosaccharide chain. The close proximity of the charge to the neighbouring carbons allows the fragmentation to progress via both charge remote and charge induced fragmentation [11]. The acidic sulfate and sialic acid residues, on the other hand, allow the charge to be specifically localised to these residues. The close proximity of the carboxyl group of sialic acid residues to a glycosidic bond promotes loss of the sialic acid as the main fragmentation pathway as seen in Figure 1 middle panel, where the fragment ion of m/z 1040 is the dominating ion. Further fragmentation after removal of the sialic acid then progresses similarly to fragmentation of neutral oligosaccharides, that is, according to a decentralised charge carried by a deprotonated poly-ol. For sulfated oligosaccharides, the sulfate group appears to be more stable during the fragmentation process compared to sialic

acid. This stability promotes charge–remote fragmentation and progresses mainly *via* glycosidic Y and B cleavages. The knowledge of how different oligosaccharide subclasses fragment will impact on how the database can be used for glycol research.

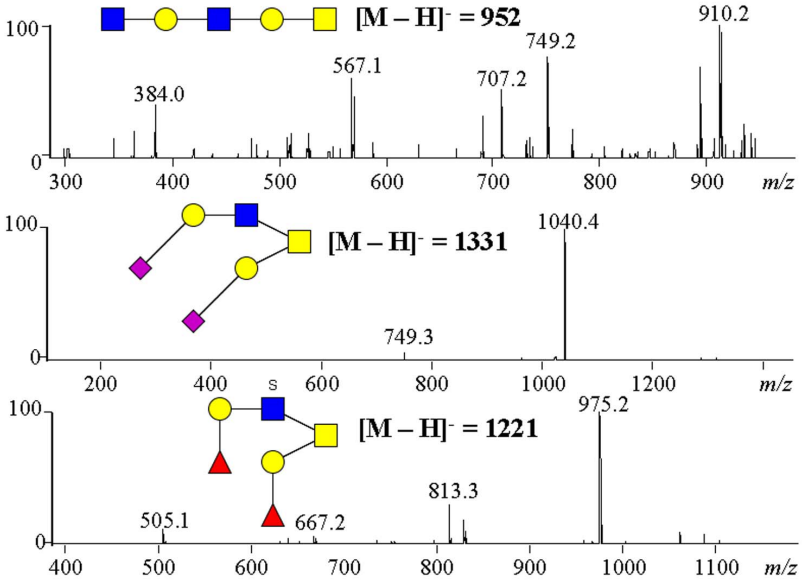


Figure 1. Illustration of the differences between the fragmentation in negative ion CID of neutral (upper panel), sialylated (middle panel) and sulfated (lower panel) *O*-linked oligosaccharides.

Matching fragment spectra of neutral O-linked oligosaccharides with UniCarb-DB spectral library

Peak matching of *m/z* and intensity data of unidentified components with a spectral library has become one of the most successful approaches for mass spectrometry, originally implemented for small molecules and GC-MS with electron impact (EI) fragmentation. This type of fragmentation is very reproducible between different instruments. Hence, a database of small molecule standards and associated fragmentation has been developed and currently the National Institute of Standards and Technology contains over 200,000 EI spectra (http://www.nist.gov/srd/nist1a.cfm). Collision induced dissociation (CID) used in combination with ES and MALDI is less standardized and factors such as charge state, collision gas, collision energy and type of mass spectrometer (triple quadrupole instruments, quadrupole-time of flight or ion trap) will influence the fragment spectra. Hence, the current initiative for Minimum Information Required for A Glycomics Experiment (MIRAGE) [12] is important for any glycomic spectral MS library containing all this meta data, to be able to compare data generated under similar conditions. In Figure 2, MS/MS spectra from ion trap CID of

two isomeric *O*-linked oligosaccharides from human gastric mucins with the $[M–H]^-$-ions of *m/z* 749 is compared with the UniCarb-DB spectral library (ion trap data) using the dot-product algorithm. For the structure identified as Galβ1-3(Galβ1-4GlcNAcβ1-6)GalNAcol (left panel) the dot product scoring indicates almost identical spectra in regards of intensities of all *m/z* values. The isomer matched in the right panel of Figure 2 showed to be less consistent spectra with its match. Closer examination of this spectrum and the meta data showed that it was generated under similar conditions and is probably not the reason for this difference. Hence, the data suggests that the sequence of the matched unknown has a similar linear sequence Hex-HexNAc-Hex-HexNAcol, but differs from its match either on linkage configuration, linkage type and/or difference between isomeric monosaccharide units. The two examples illustrate the caution in utilising the database for anything else but to identify candidate structures for the oligosaccharide of interest. This also suggests that a statistically significant match does not have to be found in order to get insight into the structure of unidentified oligosaccharides.

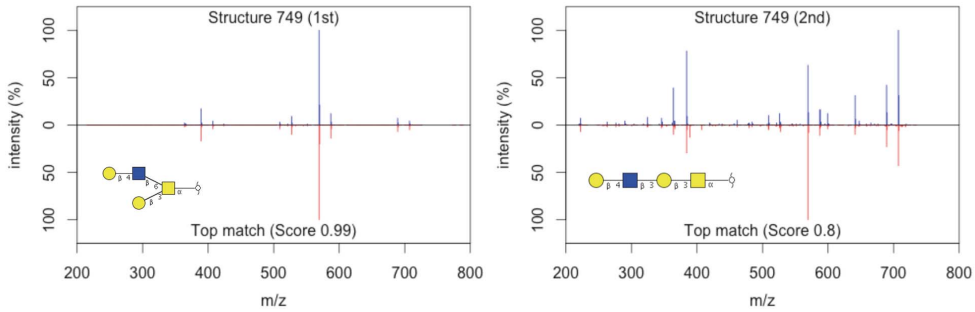


Figure 2. Results of the two neutral *O*-linked oligosaccharides queried against the database. Shown are the top-to-tail plots of the query (top) versus the best match (bottom) MS/MS along with the score (based on normal dot product).

Matching fragment spectra of sialylated O-linked oligosaccharides with spectral library

The fragmentation observed for sialylated structures, with the predominant loss of sialic acid and additional low intense fragments from fragmentation of the neutral backbone, makes the localisation of the sialic acid within the molecule difficult. We have previously devised a method using sialidase treatment in combination with LC-MS^2 to generate high quality fragmentation of its neutral counterpart. An alternative approach would be to utilise the inherent ability of the mass spectrometer to perform desialylation during the fragmentation, and acknowledge that there is little information in the spectra that tells where the sialic acid is located. This involves the partial matching of the sialylated fragment data against a spectral library containing desialylated (neutral) versions of the structures. In Figure 3, the fragment spectra of the neutral branched core 2 structure provides a better match to the monosialylated [M–H] ion of *m/z* 1040 than the neutral linear core 1 structure. This is possible by using the neutral backbone mass-range for the matching and excluding the dominating Y fragment of *m/z* 749 (loss of sialic acid) (Figure 1B).

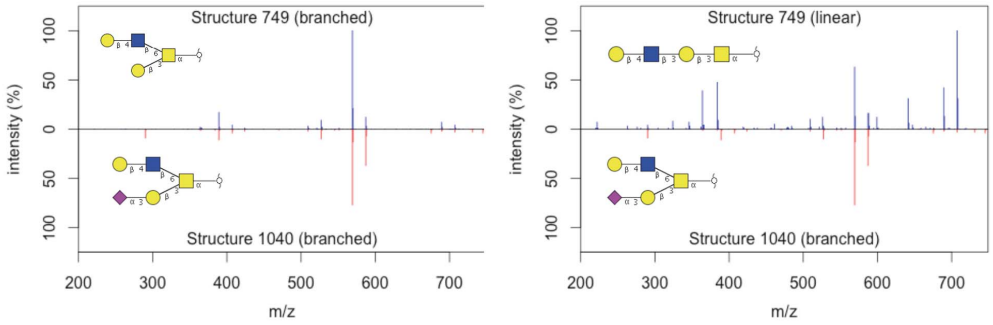


Figure 3. Results of comparing the neutral back-bone region of a sialylated structure with [M–H] ion of *m/z* 1040 structure to two neutral MS/MS spectra. This enabled the identification of the neutral nature of this particular structure; score of branched core 2 structure of [M–H] of *m/z* 749 to 1040 was 0.94 (left panel), whereas it was only 0.4 for the linear core 1 structure (right panel).

The partial matching of sialylated spectra with neutral spectra from the UniCarb-DB library illustrates how the fragmentation can be used intelligently for querying the library. However, care must be taken not to over interpret results. In this particular case the isomers NeuAcα2-3Galβ1-3(Galβ1-4GlcNAcβ1-6)GalNAcol and Galβ1-3(NeuAcα2-3Galβ1-4GlcNAcβ1-6)GalNAcol, [M–H] ion of *m/z* 1040, display almost identical MS/MS spectra, but the structure of the two pairs of can be distinguished by the difference in LC-MS retention time [13].

Matching fragment spectra of sulfated O-linked oligosaccharides with spectral library

Knowledge of the structure of sulfated oligosaccharides is still scarce, and only a limited number of ms/ms are recorded in UniCarb-DB. We set out to investigate if we could build a small library of known sulfation positions available on mucin type oligosaccharides, and explored if we can predict the type of sulfation present (6-linked GlcNAc or 3-linked Gal). The sulfate group of the oligosaccharide will carry the negative charge in the electrospray and promote charge remote fragmentation. This provides cross ring fragments in the low molecular mass region (*m/z* < 350) from single and multiple (internal) fragmentations using HCD fragmentation. Figure 4 (left panel) shows the matching of the low mass region of a sulfated oligosaccharide with [M–H] ion of *m/z* 1324 (HSO_3-$Fuc_2Hex_2HexNAc_2HexNAcol$) from human salivary MUC 5B with the MS/MS of a sulfated standard HSO_3-3Galβ1-4GlcNAc. The data generated by HCD fragmentation [14] is clearly distinguished from the HCD fragmentation of the 6 sulfated GlcNAc from porcine gastric mucin ([M–H] ion of *m/z* 1178, HSO_3-$Fuc_1Gal_2HexNAc_2HexNAcol$) and the standard Galβ1-4(HSO_3-6) GlcNAc (Figure 4 right panel).

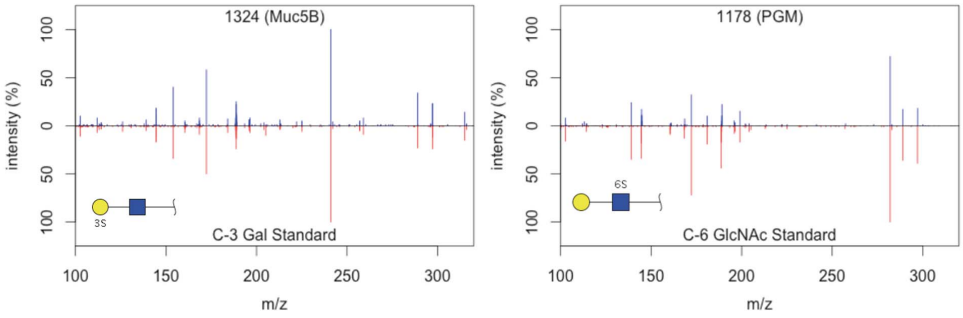


Figure 4. Top two matching structures from sulfated human salivary MUC 5B structure with the [M–H] ion of *m/z* 1324 (left top) with in the C-3 linked sulfated Gal standard in the *m/z* 100 – 320 region (0.89) (left bottom) and the PGM structure with the [M–H] ion of *m/z* 1178 (right top) with the C-6 linked sulfated GlcNAc standard (right bottom) with the C-6 linked GlcNAc standard (0.85).

The charge remote fragmentation seen for sulfated oligosaccharides makes it possible to identify the site of sulfation without actually having the identical oligosaccharide available in the spectral library. The nature of the fragmentation distinguishing the neutral, sialylated and sulfated structures allows a spectral library, such as UniCarb-DB, to be utilised intelligently to identify oligosaccharide sequence, including at times, linkage information within a sample analysed by LC-MS^2.

Target driven glycomic-Selective Reaction Monitoring (SRM)

ELISA and western blotting methods for individual proteins and protein families are widely used in life science and routinely for clinical diagnostics. In a similar fashion, the monitoring of individual oligosaccharides and/or oligosaccharide families to address biological questions would be useful in glycobiology. SRM for quantification of peptides have been shown to be a very efficient way for quantification in targeted proteomics [15]. This methodology requires insight into the fragmentation of the molecule to accurately determine transitions that would be used for monitoring protein signature peptides. In proteomics these transitions can be verified by standard synthetic peptides. However, commercial standards are not available for most *O*-linked oligosaccharides.

SRM using triple quadrupole mass spectrometry in negative ion mode and the excellent separation using graphitised carbon would provide a pathway for setting up targeted glycomics. This approach has been utilised to show how sulfation influences the extension of core 1 *O*-linked glycosylation in inflammation [8]. In Table 1 it can be seen that the data in UniCarb-DB could be used to design this type of targeted approach for investigating relationship between families of *O*-linked oligosaccharides. The table shows the 11 *O*-linked oligosaccharides that would be found in plasma (core 1 and core 2 oligosaccharides). Without access to standards the table allows access to transitions for SRM from the UniCarb-DB

spectra library, since all these structures have been recorded in the database with their associated negative ion mode MS/MS spectra. Most of the glycans provide specific fragmentation/precursor ion transition pairs that would allow them to be specifically identified. In regards to some of the isomeric structures like the pairs NeuAcα2-3Galβ1-3GalNAcol/Galβ1-3(NeuAcα2-6)GalNAcol with [M–H] of *m/z* 675 and the NeuAcα2-3Galβ1-3(Galβ1-4GlcNAcβ1-6)GalNAcol/Galβ1-3(NeuAcα2-3Galβ1-4GlcNAcβ1-6)GalNAcol with [M–H] of *m/z* 1040, within each isomeric pair the most intense transitions would also have the same mass. Since these structures are clearly separated by chromatography, their identity may not only be based on the transitions and their individual intensity relationship but also on the retention time.

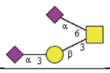
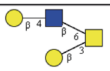
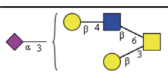
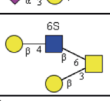
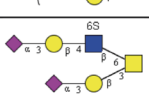
[M – H]$^-$	Structure	*Glycan ID*
384		*367*
675		*1141*
675		*572*
966		*1142*
587		*568*
749		*581*
1040		*3696*
1331		*2250*
829		*28901*
1120		*3239*
1411		*9673*

Table 1. *O*-linked oligosaccharides present in human serum/plasma and their ID as recorded in UniCarb-DB

Discussion

Spectral libraries of oligosaccharides (such as UniCarb-DB) can be useful for glycomic analysis. The matching of query fragment spectra with the spectra recorded in the experimentally generated spectral library facilitates a high-throughput glycomic identification workflow. With the increasing number of spectra generated, this process will be easier, and the focus of *O*-linked glycomic discovery will be able to shift away from structural characterisation and focus on structural verification and quantitative aspects related to regulation of various glyco-epitopes. Spectral matching, used intelligently, can also aid in the structural assignment of novel structures, as was illustrated by the identification of the sulfate position of unknown sulfated oligosaccharides in PGM and MUC 5B. The UniCarb-DB spectral library is currently primarily based on CID in negative ion mode using ion trap fragmentation. As this database becomes an acknowledged site for storage of oligosaccharide fragmentation spectra in the research community, we hope to include other fragmentation methods, ion modes, mass spectrometer types and various modification (various adducts, reducing end, permethylation, peracetylation and other ways of modification such as methyl ester formation of sialylation).

Going beyond discovery into biomarker screening, we expect that the spectral library would be useful in the design of SRM experiments for *O*-linked oligosaccharides (Figure 4). This would allow hypothesis driven glycomic research, where certain structures or family of structures can be targeted. Beyond this, a structural spectral library would be a prerequisite for querying very large amounts of LC-MS based oligosaccharide fragment data, as for instance can be generated with novel type of broad range simultaneous precursor fragmentations [16]. As the synthetic generation of all of the complex *O*-linked structures present in mammalia is still in the future, there is no choice but to use the natural sources to generate a valuable resource such as UniCarb-DB. To be successful, we are convinced that this resource needs the contribution from the whole glycomic MS society. In addition to the information about SRM transitions, it also provides information about biological sources from where individual structures can prepared. Access to sources where oligosaccharides of interest are present is important for the MS optimization of transitions.

Acknowledgements

This work was supported by the Swedish Research Council (621 – 2010 – 5322) and the Swedish Foundation for International Cooperation in Research and Higher Education. The mass spectrometer was obtained by a grant from the Swedish Research Council (342 – 2004 – 4434) and from Knut and Alice Wallenberg's Foundation (KAW2007.0118).

References

[1] Doubet, S., Bock, K., Smith, D., Darvill, A., Albersheim, P. (1989) The Complex Carbohydrate Structure Database. *Trends in Biochemical Sciences* **14**:475 – 477. doi: http://dx.doi.org/10.1016/0968-0004(89)90175-8

[2] Campbell, M.P., Hayes, C.A., Struwe, W.B., Wilkins, M.R., Aoki-Kinoshita, K. F., Harvey, D.J., Rudd, P.M., Kolarich, D., Lisacek, F., Karlsson, N.G., Packer, N.H. (2011) UniCarbKB: Putting the pieces together for glycomics research. *Proteomics* **11**:4117 – 4121. doi: http://dx.doi.org/10.1002/pmic.201100302

[3] Cooper, C.A., Harrison, M.J., Wilkins, M.R., Packer, N.H. (2001) GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources. *Nucleic Acids Research* **29**:332 – 335. doi: http://dx.doi.org/10.1093/nar/29.1.332

[4] Ranzinger, R., Frank, M., von der Lieth, C.-W., Herget, S. (2009) Glycome-DB.org: a portal for querying across the digital world of carbohydrate sequences. *Glycobiology* **19**:1563 – 1567. doi: http://dx.doi.org/10.1093/glycob/cwp137

[5] Lundborg, M., Widmalm, G. (2011) Structural Analysis of Glycans by NMR Chemical Shift Prediction. *Analytical Chemistry* **83**:1514 – 1517. doi: http://dx.doi.org/10.1021/ac1032534

[6] Royle, L., Campbell, M.P., Radcliffe, C.M., White, D.M., Harvey, D.J., Abrahams, J.L., Kim, Y.-G., Henry, G.W., Shadick, N.A., Weinblatt, M.E., Lee, D.M., Rudd, P.M., Dwek, R.A. (2008) HPLC-based analysis of serum *N*-glycans on a 96-well plate platform with dedicated database software. *Analytical Biochemistry* **376**:1 – 12. doi: http://dx.doi.org/10.1016/j.ab.2007.12.012

[7] Hayes, C.A., Karlsson, N.G., Struwe, W.B., Lisacek, F., Rudd, P.M., Packer, N.H., Campbell, M.P. (2011) UniCarb-DB: a database resource for glycomic discovery. *Bioinformatics* **27**:1343 – 1344. doi: http://dx.doi.org/10.1093/bioinformatics/btr137

[8] Flowers, S.A., Ali, L., Lane, C.S., Olin, M., Karlsson, N.G. (2013) Selected Reaction Monitoring to Differentiate and Relatively Quantitate Isomers of Sulfated and Unsulfated Core 1 *O*-Glycans from Salivary MUC 7 Protein in Rheumatoid Arthritis. *Molecular & Cellular Proteomics* **12**:921 – 931. doi: http://dx.doi.org/10.1074/mcp.M113.028878

[9] Kenny, D.T., Skoog, E.C., Lindén, S.K., Struwe, W.B. Rudd, P.M., Karlsson, N.G. (2012) Presence of Terminal *N*-acetylgalactosamineβ1-4*N*-acetylglucosamine Residues on *O*-linked Oligosaccharides from Gastric MUC 5AC: Involvement in *Helicobacter pylori* Colonization? *Glycobiology* **22**:1077 – 1785.
doi: http://dx.doi.org/10.1093/glycob/cws076

[10] Thomsson, K.A., Schulz, B.L., Packer, N.H., Karlsson, N.G. (2005) MUC 5B glycosylation in human saliva reflects blood group and secretor status. *Glycobiology* **15**:791 – 804.
doi: http://dx.doi.org/10.1093/glycob/cwi059

[11] Doohan, R.A., Hayes, C.A., Harhen, B., Karlsson, N.G. (2011) Negative ion CID fragmentation of *O*-linked oligosaccharide aldoses-charge induced and charge remote fragmentation. *Journal of The American Society for Mass Spectrometry* **22**(6):1052 – 1062.
doi: http://dx.doi.org/10.1007/s13361-011-0102-3

[12] Kolarich, D., Rapp, E., Struwe, W.B., Haslam, S.M., Zaia, J., McBride, R., Agravat, M.P., Campbell, M.P., Kato, M., Ranzinger, R., Kettner, C., York, W.S. (2013) The Minimum Information Required for A Glycomics Experiment (MIRAGE) Project: Improving the standards for reporting Mass spectrometry-based Glycoanalytic Data. *Molecular & Cellular Proteomics* **12**:991 – 995.
doi: http://dx.doi.org/10.1074/mcp.O112.026492

[13] Karlsson, N.G., McGuckin, M.A. (2012) *O*-Linked glycome and proteome of high-molecular-mass proteins in human ovarian cancer ascites: Identification of sulfation, disialic acid and *O*-linked fucose. *Glycobiology* **22**:918 – 929.
doi: http://dx.doi.org/10.1093/glycob/cws060

[14] Olsen, J.V., Macek, B., Lange, O., Makarov, A., Horning, S., Mann, M. (2007) Higher-energy C-trap dissociation for peptide modification analysis. *Nature Methods* **4**:709 – 12.
doi: http://dx.doi.org/10.1038/nmeth1060

[15] Gallien, S., Duriez, E., Domon, B. (2011) Selected reaction monitoring applied to proteomics. *Journal of Mass Spectrometry* **46**:298 – 312.
doi: http://dx.doi.org/10.1002/jms.1895

[16] Law, K.P., Lim, Y.P. (2013) Recent advances in mass spectrometry: data independent analysis and hyper reaction monitoring. *Expert Review of Proteomics* **10**:551 – 566.
doi: http://dx.doi.org/10.1586/14789450.2013.858022