Better together – relative retention time plus spectral matching improves automated glycan characterization using PGC-nLC-IT-ESI-MS/MS

Kathirvel Alagesan^{1,2,3,*}, Falko Schirmeister^{2,3}, Uwe Möglinger^{2,3,#}, Arun Everest-Dass¹, Friedrich Altmann⁴, Peter H Seeberger^{2,3}, Mark von Itzstein¹, Nicolle H Packer^{1,5,6}, Daniel Kolarich^{1,2,6,*}

1 Institute for Glycomics, Griffith University Gold Coast Campus, QLD, 4222, Australia

2 Department of Biomolecular Systems, Max Planck Institute of Colloids and Interfaces, 14476 Potsdam, Germany

3 Institute of Chemistry and Biochemistry, Freie Universität Berlin, 14195 Berlin, Germany

4 Institute of Chemistry, University of Natural Resources and Applied Life Sciences, 1190 Vienna, Austria

5 Department of Molecular Sciences, Macquarie University, Sydney, NSW, Australia

6 ARC Centre of Excellence for Nanoscale BioPhotonics, Griffith University

*Corresponding Authors

Dr. Kathirvel Alagesan

Present address: Max Planck Unit for the Science of Pathogens, Charitéplatz 1, 10117 Berlin, Germany

T +49 30 28460 230

Email: alagesan@mpus.mpg.de

A/Prof. Daniel Kolarich

Institute for Glycomics, Griffith University, Gold Coast campus, QLD 4222

T +61 7 5552 7026

F +61 7 5552 9040

Email: d.kolarich@griffith.edu.au

Present address

Novo Nordisk, Novo Nordisk Park 2760 Måløv, Denmark

ORCID

KA	–	0000-0002-7596-5558
AD	–	0000-0002-3795-0291
FA	–	0000-0002-0112-7877
PHS	–	0000-0003-3394-8466
MVI	–	0000-0001-6302-7524
NHP	–	0000-0002-7532-4021
DK	–	0000-0002-8452-1350

CREDIT AUTHOR STATEMENT

Kathirvel Alagesan – Conceptualization, Methodology, Software, Validation, Investigation, Formal analysis, data curation, Writing – Original draft, Writing – Review & Editing, Visualization, Project administration

Falko Schirmeister – Investigation, Software, Validation, Writing – Review & Editing, Visualization

Uwe Möglinger – Investigation, Validation, Writing – Review & Editing, Visualization

Arun Everest-Dass - Validation, Writing – Review & Editing, Visualization

Friedrich Altmann – Resources, Writing – Review & Editing

Peter H Seeberger – Resources, Writing – Review & Editing

Mark von Itzstein – Resources, Writing – Review & Editing

Nicolas H Packer – Resources, Writing – Review & Editing

Daniel Kolarich – Conceptualization, Methodology, Resources, Visualization, Writing – Review & Editing, Funding acquisition

ACKNOWLEDGMENTS

We thank Max Planck Society for their generous financial support. KA was supported by PhD scholarship from Beilstein-Institut. This work was supported in part by Griffith University, European Union Seventh Framework Programme *IBD-BIOM* (project grant number 305479) and *Glycoproteomics* (project grant number 293847) and Australian Research Council Future Fellowship FT160100344 (to DK).

ABSTRACT

Porous Graphitized Carbon nano-liquid chromatography tandem mass spectrometry (PGC-nLC-MS/MS) is a glycomics technique with the unique capacity to differentiate isobaric glycans. The lack of suitable software tools integrating chromatography and MS-information delivered by PGC-nLC-MS/MS has been limiting fast and robust glycan identification and quantitation. We report a LC-system-independent strategy called GlycoRRT that combines relative retention time (RRT) and negative ion fragment spectra analyses for isobaric structure-specific glycomics of PGC-nLC-MS/MS data. The GlycoRRT toolset is fully customizable and easily adaptable enabling semi-automated high-throughput structural assignments. The current library contains over 200 entries and their individual meta-data (MS instrumentation, experimental conditions, retention times, fragmentation profiles and glycan structural diagnostic ion features) relevant for reliable data analyses. The GlycoRRT workflow was employed to map the N- and O-glycome in blood group matched human plasma and urine as well as decipher Immunoglobulin (IgG) glycosylation features from 13 different animal species. We have also developed visualization tools to enable a consistent, reliable, and reproducible analysis of large sets of multidimensional PGC-nLC-MS/MS glycomics data. This comprehensive glycan resource provides the glycan map of human and animal species, will serve as a reference in dissecting the role of glycans in host pathogen interaction and zoonotic disease transmission.

Key words: Glycomics - Porous Graphitized Carbon – Immunoglobulin G – Mass spectrometry – Plasma glycomics

Cell surface and body fluid proteins are extensively modified with specific sugar moieties, so-called glycans¹. These glycans build the basis for a universal language (glycome) that is used between cells but is also exploited by pathogens and cancer cells^{1, 2}. Glycomics seeks to understand how the entirety of glycans present on a protein, a cell or a body fluid relate to biological processes in health and disease by detailed structural and semi-quantitative analyses of glycans^{3, 4}.

In order to identify and quantify the cells' complete glycan repertoire, it is necessary to capture the enormous structural diversity and complexity. This daunting task requires orthogonal methods to accommodate both, in-depth structural data acquisition and medium-high throughput analyses⁵. Widely used Mass Spectrometry (MS) or Liquid Chromatography (LC)-based approaches provide good throughput but suffer from limited capacity to elucidate structure details and differentiate isobaric glycans. In consequence, the combination of LC separation with tandem mass spectrometry (MS/MS) has evolved to be one of the most powerful platforms to discriminate, distinguish, and quantify glycan structural isomers^{3, 4, 6}. Separation of structure isomers has been achieved by reversed-phase (RP) and hydrophilic liquid interaction chromatography (HILIC)^{7, 8}, but porous graphitized carbon (PGC) presents itself as the most versatile and effective approach to separate isobaric glycans^{6, 9, 10}. PGC can be applied universally to analyze not only N- and O-glycans, but also the glycan portion of glycolipids and glycosaminoglycan fragments¹¹⁻¹³. In negative ion mode, PGC-nLC-MS/MS allows for an unambiguous and detailed structural characterization of isobaric glycans and relative quantitation within a single LC-MS/MS experiment, drastically reducing the sample amount to even capture the glycome from as few as 1000 cells¹⁴⁻¹⁷.

Unlike HILIC, glycan retention in PGC cannot be correlated to the monosaccharide composition by multiple linear regression analyses¹⁸. The exact molecular nature of the PGC-retention mechanisms leading to the observed isobaric glycan separation is still not fully understood, further complicating computationally assisted retention time predictions. Nevertheless, specific glycan structural features are known to influence the PGC elution behavior, and the order of elution of individual structural isomers is also highly conserved¹⁹⁻²¹. Despite the availability of glycan retention time and elution order, efforts to utilize them for glycan identification have not yet been satisfactorily implemented as the absolute retention times can vary between individual analyses and drastically depend on the individual LC-setup (e.g. capillary lengths) and operational parameters (organic phase gradient) used.

Absolute retention time variations can be compensated by normalizing retention times based on calibration standards. This approach has been implemented successfully for the automated

structure assignment of HILIC-separated 2-AB-labeled glycans. The detected retention times are converted into "Glucose Units" (GU) determined by a separate analysis of a dextran ladder ranging from 1-15 glucose molecules²². A similar normalizing approach has been described recently for PGC-LC-MS/MS where the glycan structure assignment is based only on precursor mass and GU values with subsequent verification by fragment spectra²³. Beforehand, Pabst and co-workers were the first to demonstrate the use of relative retention times (RRT) for structural assignment of oligomannosidic N-glycans²⁴. However, this promising development has not been fully utilized for other classes of glycans that can be separated using PGC-LC. This is not surprising, as RRT and precursor MS is insufficient for reliable structure determination considering the tremendous structural diversity represented by glycans^{9, 25}.

To this end, a spectra library-based approach that allows for querying multiple tandem spectra against a well-defined library alongside precursor mass and RRT will improve glycan structure assignment and facilitate high-throughput (HTP) data analyses. Currently available software tools that predict glycan structures from MS data do not allow for a direct comparison of query spectra against library spectra²⁶. We addressed these challenges and developed a system-independent, semi-automated and easily adaptable toolset for automated glycan identification that combines MS/MS spectra matching, normalized glycan elution order and precursor mass. These distinct parameters along with peak intensity are acquired within a single PGC-nLC-MS/MS analysis for each glycan structure.

The herein described GlycoRRT tool incorporates and fully utilizes these distinct features to allow for glycan structure assignment and refinement as well as relative quantitation. The initial GlycoRRT library was established using data of 33 synthetic N-glycan standards and has been subsequently extended to now contain 200 annotated N-glycans from human plasma glycoproteins. Using the GlycoRRT tool, we established the human GlycoATLAS, a new resource that provides a baseline map of the N-glycans of human whole blood as well as blood group matched pooled plasma and urine. Finally, we also mapped species specific immunoglobulin (IgG) glycosylation features of 13 different animal species.

RESULTS

CREATION OF GLYCORRT SPECTRAL LIBRARY

A reliable, fully customizable GlycoRRT spectral library was developed based on PGC-nLC-ESI-MS/MS in negative ion mode to enable HTP automated glycan identification and characterization. The initial GlycoRRT spectral library was established utilizing 33 synthetic N-glycan standards (**Table S1**) incorporating two distinctive features: a) high-quality negative ion fragment spectra (**Figure S1&2**) and b) retention time (**Table S3**). In addition to the expected 33 N-glycans, we were also able to resolve other isomers present in these standards yielding a total of 41 distinctive tandem MS spectra (**Table S3**). Each entry in the GlycoRRT spectral library contained all crucial metadata including precursor mass, charge state, glycan composition, and structure represented in GlycoMod and Proglycan nomenclature ²⁷ (**Figure S3**). An overview of the GlycoRRT workflow is provided in **Figure 1** and described in the supplementary methods section.

Figure 1: PGC-LC-ESI-MSMS and GlycoRRT spectral-library matching glycomics workflow. (A) The four main steps of glycomics sample preparation: The proteins were extracted from the desired biological material using suitable methods and then immobilized onto a PVDF membrane. N-linked glycans were released enzymatically using PNGase F, followed by the O-glycan release from the same sample by reductive β -elimination. Finally, N- and O-glycan alditols were desalted and analyzed by PGC-nLC-ESI MSMS. **(B)** Glycan spectral library design: The Glycan RRT LC-MS/MS library contains annotated, negative ion ESI-MS product ion spectra, precursor ion information including charge state, retention time information, glycan composition and isomer information, as well experimental metadata. **(C)** Spectral matching of query spectra against the well-annotated reference spectra database enables automated glycan identification, providing as %fit score for each glycan that informs on positive or negative identification. The results can be easily exported and are readily visualized using our custom R packages.

Spectral matching evaluation

We demonstrated previously that it is possible to use the spectral matching approach for automated glycan structure characterization based on the dot-product function available in the R package OrgMassSpecR²⁸. Nevertheless, the spectral matching function was not available within the data analysis workflow for HTP analyses. We first evaluated the feasibility of the spectral matching tool incorporated within the data analysis software by re-searching the dataset used to create the initial spectral library. The query compounds were identified in high confidence as evident by the high RFit' score (max 1000). Next, the applicability and feasibility of the spectral library to identify N-glycans derived from a randomly selected human Immunoglobulin (IgG3) was evaluated. Details on the workflow for the sequential release of N- and O-glycans adapted from a previous protocol¹⁰ are described in the online methods. The initial analyses resulted in the identification of 15 different N-glycan structures from 500 ng protein in less than minute via single click data analysis (**Table S4, Supplementary results - Spectral matching automation in data analysis**). The RFit' scores of the identified N-glycans ranged from 918 for GnA to 988 for Na(2-6)Na(2-6) indicating high confidence in the structure assignment (**Table S4**). Nevertheless, the entire data file contained several spectra that did not result in any confident library match at that time point. The following manual analyses focused on these unidentified spectra, adding another six N-glycans to the spectra library. We confidently identified 21 different N-glycan structures from 500 ng human IgG3 in less than 15 minutes using this automated spectral matching and manual identification approach (**Table S6 – for additional identifications**).

Relative Retention time inclusion

Spectral matching provides an enormous step forward in automated structure identification; however, closely related, isobaric glycans often share multiple product ion spectra features that can result in incorrect structure assignments. Such wrong assignments can be reduced significantly when glycan elution profile information is included^{11, 19} (**Figure S4**). Thus, the integration of glycan structure isomer information derived from the relative order of PGC LC-elution together with spectral matching can facilitate and improve HTP automated glycan structure assignment.

Run-to-run variation in the absolute retention time and among various laboratories, however, impedes the integration of absolute retention time information that will assist in the structural assignment. We first evaluated the influence of column temperature and grounding on the separation capacity offered by PGC to better understand the impact of these conditions on PGC separation in our LC-MS setup (**Supplementary Results: Glycan elution Rules of PGC**

& Factors influencing glycan separation by PGC-LC, Figure S5 & S6)²⁹. Next, we sought to normalize the variation in absolute retention times by converting these values into relative retention times (RRT) and incorporated this feature into our spectral matching tool. The RRT for each glycan species was calculated using the AAF (***N-glycan ID20, Table S5***) isomer elution time as a reference value, as it elutes in the middle of the gradient. The RRT values of all other synthetic N-glycan standards were then established in relation to AAF (***Table S5***). The robustness of RRT normalization was then evaluated using 12 independent samples obtained from three different users containing 21 N-glycan species that were all identified by using the spectral matching tool as described above with high accuracy (***Table S6, Figure 2A***). The RRT approach corrected absolute retention time shifts with less than 5% coefficient of variation (***Figure 2-B***).

Figure 2: (A) Absolute retention times of 21 N-glycans collected from 12 independent samples that were obtained from three different users and their relative retention times (RRT). **(B)** The RRT for each identified N-glycan was calculated using the AAF isomer elution time as reference using the expression $RRT = (\text{Sample RT} / \text{Reference RT})$. The use of RRT normalizes the variations in absolute retention time for accurate incorporation of retention time features in glycan structure assignment. Glycans are labeled with their simple “Proglycan” nomenclature to allows for differentiation of glycan isomers.

We also evaluated the influence of the ion trap instrument's Smart Parameter Setting (SPS) target parameter on glycan identification and, specifically, quantitation. Typically, the SPS is set in the middle of the intended acquisition mass range to support the auto-adjustment of the specific acquisition mass window to a particular target m/z . Our results indicate that the selection of an SPS target m/z oriented slightly towards the higher end of the intended m/z scan range improved the identification and quantitation of larger glycan structures compared to the traditional mid-range SPS selection (**Figure S7, see supplementary results - Influence of SPS parameter on quantification**).

THE HUMAN GLYCOATLAS

The results of the initial GlycoRRT approach for HTP glycan analyses motivated us to further expand the library using N-glycans released from human serum glycoproteins.

Plasma/serum vs whole blood N-glycan atlas

The serum GlycoRRT spectral library was used to profile plasma and whole blood N-glycans. In total, 70 different compounds corresponding to N-glycans were identified in less than 5 min, including various charge states (**Table S7, Figure S8-A**). In addition to the RFit' score, relative retention times as well as any deviations (dRRT, **Formula: $dRRT = RRT_{library} - RRT_{Run}$**) were used as a means of quality control for unambiguous glycan structure identification and characterization. Of 70 N-glycan compounds identified (60 by automated identification, two via manual identification, the remaining eight are different charge state variants of glycans already present in the library), the relative abundances of 68 N-glycan structures were quantified by integrating the area under the curve obtained for the respective EICs. For two tetra-sialylated N-glycans the EIC-peak was unsuitable for reliable quantitation. For each measurement, the values were normalized to the total quantity of all detected N-glycans.

Next, we sought to visualize the difference in relative abundance at both individual glycan and group level. The relative abundances of individual N-glycan structures were assessed and visualized in a heatmap by unsupervised hierarchical cluster analysis **Figure S8-A (See Table S7 for identification results)**. The most abundant N-glycan in serum, plasma, and whole blood was the Na(2-6)Na(2-6) glycan contributing to approximately 50% in all three types of samples (**Figure S8-A**). The N-glycans were also grouped based on their structural features: Oligomannose, neutral complex and sialylated complex. Sialylated N-glycans accounted for approximately 90% of the total N-glycome in the analyzed body fluids. The majority of these were di-sialylated (75% in the case of plasma and serum, 70% in whole blood), followed by mono- and tri-sialylated. Tetra-sialylated N-glycans were just present in trace amounts (**Figure S8-B**). Overall, the N-glycans derived from plasma and serum exhibited similar relative abundances for all glycan classes, whereas serum/plasma vs whole blood exhibited marked differences especially in the case of mono- and di-sialylated N-glycans (**Figure S8-B**).

Blood-group specific features of the plasma and urinary N-glycome

The GlycoRRT tool was further employed to establish an average profile for the global plasma and urine N-glycome (600 ng - injected) after pooling these samples according to blood groups (A, B and O). After automated and manual identification, 46 N-glycan isomers corresponding to 29 glycan compositions were detected in plasma, whereas 33 N-glycan isomers corresponding to 20 glycan compositions were identified in urine by a single click analysis that

took less than 15 min analysis time for 18 samples on a standard PC (**Table S8 and S9**). The N-glycan structures initially not present in the library such as tri-antennary N-glycans from plasma (Glycan ID: 32 and 33 – **see Table S9**) and N-glycans carrying the LacdiNAc epitope (Glycan ID: U23 and U24, see Table S11) from urine were manually identified, validated, added to the library, and quantified and visualized in a heatmap (**Figure 3 and 4**). Na(2-6)Na(2-6) was the most abundant glycan in both plasma and urine irrespective of the blood group.

Figure 3: (A) Heatmap of log-transformed relative abundances of N-glycans from blood group matched plasma. Rows indicate the glycan structures whereas the column indicate the three technical replicates from different blood group (A, B and O). The samples were ordered using unsupervised hierarchical clustering based the Euclidean distance matrix with complete linkage. Remarkably, the overall glycome signature allowed the algorithm to independently differentiate the pooled plasma samples based on their blood groups indicating a blood group dependent impact on the global N-glycome profile. (B-F) Differences in relative abundance of different N-glycan classes are shown as box plots. Multiple comparisons among various blood groups were performed using paired T-test. The segregation of blood group B and O from A is largely driven by differences in the abundance of mono and di-sialylated N-glycans. See source data table S1 for pair-wise comparison and results table.

Figure 4: (A) Heatmap of log-transformed relative abundances of N-glycans from blood group matched urine. The samples were ordered by hierarchical clustering using the Euclidean distance matrix with complete linkage. Rows indicate the glycan structures (Supplementary Table S9). Similar to plasma glycome, also the urinary N-glycome enabled the algorithm to cluster the samples based on the ABO blood group status. We also observed two clusters based on unsupervised hierarchical clustering yet in an opposite trend compared to plasma indicating source dependent differential action of glycan modifying enzymes. (B-F) Differences in relative abundances of different N-glycan classes are shown as box plots. Multiple comparisons among various blood groups were performed using paired T-test. As evident, the segregation of N-glycans derived from blood group B and O from A is largely driven by mono and di-sialylated N-glycans. See source data table S2 for pair-wise comparison and results table.

A snapshot of Vertebrata Immunoglobulin G (IgG) glycosylation

The developed PGC-nLC-ESI-MS/MS glycomics workflow was also utilized to investigate IgG N-glycans derived from thirteen different Vertebrata species (**Table S10**). Samples from twelve mammalian and one avian species were compared with regard to the similarity of their qualitative and relative quantitative IgG N-glycan signatures (**Figure S9 & Figure 5**). The hierarchical cluster analysis sorted the animals in an unbiased manner, solely based on the determined N-glycan profiles. Most species showed a preference for incorporation of either NeuAc or NeuGc. The human IgG N-glycans separated clearly from the other mammals due to the lack of N-glycans carrying NeuGc. Duck IgG N-glycans showed the highest degree of differentiation due to the presence of unique N-glycan structures exhibiting sialylated LacdiNAc motifs not found in any other species analyzed (**Figure S9**). Vertebrate IgG contained predominantly contained biantennary neutral N-glycan followed by sialylated N-glycan except in the case of Donkey (**Figure 5 A-C**).

Figure 5 A-K: Rigdeline plot of Glyco-epitope features of IgG N-glycan derived from twelve different mammals and one bird. Glyco-epitope features detected exhibited a strong species dependency. Irrespective of the fucosylation status, the bulk of the IgG N-glycans were neutral ones lacking NeuAc or NeuGc, with the exception of donkey who predominately contained sialylated N-glycans (A-C). The majority of the neutral N-glycans were core-fucosylated except for duck who was lacking fucosylation on its IgG (D,G). Incorporation of either NeuAc or NeuGc was dependent on the species. Human and Duck contained only NeuAc. Nevertheless, the NeuAc present in duck was just found attached to a LacdiNAc motif which was not found in any other species. Human and dog had similar NeuAc levels but exhibited different levels of core fucosylation (E-K).

DISCUSSION

Despite the unique analytical opportunities provided by various emerging LC-MS technologies to characterize glycans, reliable interpretation of repetitive glycan MS/MS data is a major limiting factor in HTP data analysis. **Figure 1-B&C** summarizes our GlycoRRT pipeline for HTP qualitative and quantitative assessment of multi-dimensional glycomics data obtained using PGC-nLC-ESI-MSMS in negative ion mode. The conserved diagnostic fragment ions observed in product ion spectra of deprotonated glycan ions³⁰⁻³² in combination with the high separation capacity provided by the PGC was the basis to develop this reliable, fully customizable GlycoRRT tool. To best of our knowledge, this study represents the first examination towards optimizing and improving the existing tool for HTP glycan structural assignment. Our GlycoRRT tool utilizes (i) retention time and glycan elution order, (ii) precursor mass, and (iii) fragment spectra to achieve reliable structure identification.

Current bioinformatics tools for MS/MS-based glycan sequencing rely on inferring glycan composition from the precursor mass using tools such as GlycoMod³³ and use information derived from diagnostic fragment ions in the product ion spectra for structure determination. Despite the availability of the Glycoworkbench³⁴ software that assists in manual interpretation of tandem MS data, the process is tedious and repetitive. Spectral library matching for N-glycans has been reported previously to show good spectral reproducibility and glycan fragmentation in negative ion mode and is fully suitable for facilitating cross-platform analyses²⁸. To this end, a spectral matching approach including standardized retention time data for structure assignment enables reliable HTP data analysis while minimizing repetitive tasks. Retention time information is regularly used in normal phase separation for glycan identification, but it is less frequently an integrative parameter in LC-MS/MS based glycomics to support structure identification and/or additional validation. However, the absolute retention time of glycan analytes can vary substantially when samples are analyzed across a long time span as it is required when investigating large cohorts. Here we relied on relative retention times to normalize the unavoidable variation in the absolute retention time. We show that the RRTs are system independent and remain stable over an extended time period for years (**Figure 2A**). RRT values are calculated by dividing each individual retention time value by the retention time of one selected reference glycan structure (we selected AAF for our analyses). In principle, however, any glycan structure and its corresponding retention time can be used as reference as long as they are present in all the samples and as such, the proposed system is highly flexible and can be optimized for any sample cohort.

We created the GlycoRRT tool using the LibraryEditor available within the DataAnalysis software (Bruker) that allows creation of glycan spectra data bases as well as fast spectra matching to facilitate a fast and reliable, semi-automated data analysis. Nonetheless, any available spectral matching tool such as MSdial³⁵ can be refined for automated glycan identification by incorporating retention time, associated experimental, MS and MS/MS acquisition metadata. Caveats of any software-assisted approach relying on spectral library matching are false assignments or low scores, especially if fragment ions are of low intensity, absent or very limited (as it is frequently found for highly sialylated glycan precursors). However, the approach presented here enables easy identification of such wrong or no assignments based on RFit score and dRRT values. Upon manual validation, the new structures can be easily added as a new entry to the library. Including GlycoRRT also allows identifying highly sialylated N-glycans that deliver poor product ion spectra with improved accuracy. Overall, our work represents a first step towards automated qualitative glycan analysis by incorporating spectral matching and RRT.

Relative quantitation allows for the sensitive monitoring of changes in glycan abundances across various samples and biological conditions³⁶. Here, we present several ways to analyze and visualize large-scale multi-dimensional LC-MS glycomics data as a part of the GlycoRRT glycomics workflow. The developed R package *BCQA* covers the data handling from import of raw quantitation values from Bruker's QuantAnalysis, their normalization, statistical analysis and comprehensive data visualization (**Figure S11**). Furthermore, the developed functions allow seamless conversion of glycan composition to other formats to allow for visualization using other tools available in Glycomics ExPASy server (<https://www.expasy.org/glycomics>, **See supplementary Figure S12**).

In line with recent global initiatives (<https://human-glycome.org/>), the developed PGC glycomics workflow was used to capture an overview of the human plasma and urinary N-glycome from blood-group matched plasma and urine pooled from 20 individuals (**Figure 3&4**). Our analyses have demonstrated that the plasma and urinary N-glycome exhibited an ABO blood-group dependent profile (**Figure 3&4**). It has been reported previously that sialylated glycan clusters on the red blood cell membrane are unique for each blood type³⁷. We observed a similar inclination on the secreted glycoproteins present in plasma with a significant increase in di-sialylated N-glycans in plasma from blood group A individuals, while monosialylated were increased in blood group B individuals and blood group O individuals led in neutral N-glycan abundances (**Figure 3 D&E**). In urine, remarkably, blood group B exhibited the highest levels of di-sialylated N-glycans (**Figure 4-E**). To the best of our knowledge plasma

glycan biomarker studies have not, included any data stratification based on blood-group status. Our data, however, hint towards the necessity to include blood group stratification and appropriate blood group stratified controls in any type of plasma glyco-marker studies.

Revealing deep N-glycan structure details of species-specific IgGs using our GlycoRRT approach also enabled us to reconstruct the enzymatic steps in N-glycan biosynthesis pathways among the 13 different species analyzed (**Figure 6**). Our results clearly indicate that the mechanisms regulating IgG glycosylation are conserved across the different species to deliver predominately biantennary N-glycans (**Figure S10**). The terminal glyco-epitope features, however, are highly species specific (Na in humans vs Na/Ng in other mammals vs LacDiNAc in duck). This is likely the consequence of differences in monosaccharide donor availability, transferase substrate specificities as well as transferase expression.

Glyco-epitopes such as fucose (as in blood group antigens such as A, B, H, Lewis a, etc.) and/or sialic acids (Na, Ng) are known to be involved in host-pathogen interactions³⁸. Our results clearly demonstrate that our GlycoRRT-PGC-LC-MS/MS strategy can be successfully applied to study the glyco-space in less well characterized species to understand the role of glycans in interspecies transmission of pathogens. As a first step, the IgG glyco-epitope map across 13 species presented here will augment our understanding of the molecular mechanisms of cross-species transmittable diseases, where species-specific glycosylation features play undoubtedly an important role ³⁹.

The glycan spectral matching and scoring tool (GlycoRRT) including both, negative mode fragmentation and relative retention time, represents an important step forward towards reliable automated glycan identification. Overall, the library now contains more than 200 annotated N-glycans structures derived from synthetic N-glycan standards and human Immunoglobulins, plasma, blood, and urine. All N-glycan structures present in the library were characterized using (i) molecular monoisotopic mass and (ii) CID-MS/MS *de novo* sequencing. The results demonstrated the general applicability of the PGC-nLC-ESI MS/MS approach for detailed qualitative and quantitative glycomics analyses across species.

Figure 6: Consolidated representation of various sequential steps involved in the IgG N-glycosylation pathway. For simplicity, N-glycans carrying bisecting GlcNAc and α 2-3 sialic acid linkage were omitted. The initial steps of N-glycosylation occurring within ER are highly conserved, but further downstream processing occurring in the Golgi exhibits a high species-specificity resulting in the synthesis of different (terminal) glyco-epitopes. In mammals, over 250 glycosyltransferases are known to reside within the Golgi⁴⁰. Their mission is to catalyze the transfer of a specific activated monosaccharide on a particular non-reducing end monosaccharide that is part of a larger glycan molecule. The specificity of individual glycosyltransferases is not only defined by the respective monosaccharide that is being transferred but also by the specific glycosidic linkage this monosaccharide is attached to. PGC-nLC-ESI-MS/MS analysis of N-glycans released from in vivo produced glycoproteins allows a highly representative interrogation of glycosidic pathway reactions as they occur within an organism that can be much more informative compared to the use of defined, synthetic glycan standards that might not ideally reflect the in vivo substrate. The ability to precisely differentiate these glyco-epitope features as such linkage, position, type of monosaccharide within a single analytical experiment provides a unique opportunity to interrogate the N-glycosylation pathways across different species.

REFERENCE

1. Varki, A., Kannagi, R., Toole, B. & Stanley, P. in *Essentials of Glycobiology*, Edn. 3rd. (eds. rd et al.) 597-609 (Cold Spring Harbor (NY), Cold Spring Harbor (NY); 2015).
2. Rodríguez, E., Schetters, S.T.T. & van Kooyk, Y. The tumour glyco-code as a novel immune checkpoint for immunotherapy. *Nat Rev Immunol* (2018).
3. Everest-Dass, A.V., Moh, E.S.X., Ashwood, C., Shathili, A.M.M. & Packer, N.H. Human disease glycomics: technology advances enabling protein glycosylation analysis - part 2. *Expert Rev Proteomics* **15**, 341-352 (2018).
4. Everest-Dass, A.V., Moh, E.S.X., Ashwood, C., Shathili, A.M.M. & Packer, N.H. Human disease glycomics: technology advances enabling protein glycosylation analysis - part 1. *Expert Rev Proteomics* **15**, 165-182 (2018).
5. Kolarich, D. & Packer, N.H. in *Structural Glycobiology*. (eds. E. Yuriev & P.A. Ramsland) 141-161 (CRC Press, Boca Raton, FL 33487-2742; 2012).
6. Alagesan, K., Everest-Dass, A. & Kolarich, D. Isomeric Separation and Characterisation of Glycoconjugates. *Adv Exp Med Biol* **1104**, 77-99 (2018).
7. Natsuka, S. & Hase, S. Analysis of N- and O-glycans by pyridylamination. *Methods Mol Biol* **76**, 101-113 (1998).
8. Rudd, P.M. et al. Oligosaccharide sequencing technology. *Nature* **388**, 205-207 (1997).
9. Everest-Dass, A.V. et al. Comparative structural analysis of the glycosylation of salivary and buccal cell proteins: innate protection against infection by *Candida albicans*. *Glycobiology* **22**, 1465-1479 (2012).
10. Jensen, P.H., Karlsson, N.G., Kolarich, D. & Packer, N.H. Structural analysis of N- and O-glycans released from glycoproteins. *Nature protocols* **7**, 1299-1310 (2012).
11. Pabst, M., Bondili, J.S., Stadlmann, J., Mach, L. & Altmann, F. Mass + retention time = structure: a strategy for the analysis of N-glycans by carbon LC-ESI-MS and its application to fibrin N-glycans. *Analytical chemistry* **79**, 5051-5057 (2007).
12. Stadlmann, J., Pabst, M., Kolarich, D., Kunert, R. & Altmann, F. Analysis of immunoglobulin glycosylation by LC-ESI-MS of glycopeptides and oligosaccharides. *Proteomics* **8**, 2858-2871 (2008).
13. Fan, J.Q., Kondo, A., Kato, I. & Lee, Y.C. High-performance liquid chromatography of glycopeptides and oligosaccharides on graphitized carbon columns. *Analytical biochemistry* **219**, 224-229 (1994).
14. Hinneburg, H. et al. Unlocking Cancer Glycomes from Histopathological Formalin-fixed and Paraffin-embedded (FFPE) Tissue Microdissections. *Mol Cell Proteomics* **16**, 524-536 (2017).
15. Schulz, B.L., Packer, N.H. & Karlsson, N.G. Small-scale analysis of O-linked oligosaccharides from glycoproteins and mucins separated by gel electrophoresis. *Analytical chemistry* **74**, 6088-6097 (2002).
16. Thomsson, K.A., Karlsson, N.G. & Hansson, G.C. Liquid chromatography-electrospray mass spectrometry as a tool for the analysis of sulfated oligosaccharides from mucin glycoproteins. *Journal of chromatography. A* **854**, 131-139 (1999).
17. Karlsson, N.G. et al. Negative ion graphitised carbon nano-liquid chromatography/mass spectrometry increases sensitivity for glycoprotein oligosaccharide analysis. *Rapid communications in mass spectrometry : RCM* **18**, 2282-2292 (2004).
18. Melmer, M., Stangler, T., Premstaller, A. & Lindner, W. Comparison of hydrophilic-interaction, reversed-phase and porous graphitic carbon chromatography for glycan analysis. *Journal of chromatography. A* **1218**, 118-123 (2011).
19. Abrahams, J.L., Campbell, M.P. & Packer, N.H. Building a PGC-LC-MS N-glycan retention library and elution mapping resource. *Glycoconj J* **35**, 15-29 (2018).
20. Pabst, M. & Altmann, F. Influence of electrosorption, solvent, temperature, and ion polarity on the performance of LC-ESI-MS using graphitic carbon for acidic oligosaccharides. *Analytical Chemistry* **80**, 7534-7542 (2008).

21. Pabst, M., Bondili, J.S., Stadlmann, J., Mach, L. & Altmann, F. Mass plus retention time = structure: A strategy for the analysis of N-glycans by carbon LC-ESI-MS and its application to fibrin N-glycans. *Analytical chemistry* **79**, 5051-5057 (2007).
22. Royle, L., Radcliffe, C.M., Dwek, R.A. & Rudd, P.M. Detailed structural analysis of N-glycans released from glycoproteins in SDS-PAGE gel bands using HPLC combined with exoglycosidase array digestions. *Methods Mol Biol* **347**, 125-143 (2006).
23. Ashwood, C., Pratt, B., MacLean, B.X., Gundry, R.L. & Packer, N.H. Standardization of PGC-LC-MS-based glycomics for sample specific glycotyping. *The Analyst* **144**, 3601-3612 (2019).
24. Pabst, M. et al. Isomeric analysis of oligomannosidic N-glycans and their dolichol-linked precursors. *Glycobiology* **22**, 389-399 (2012).
25. Everest-Dass, A.V., Kolarich, D., Pascovici, D. & Packer, N.H. Blood group antigen expression is involved in *C. albicans* interaction with buccal epithelial cells. *Glycoconj J* **34**, 31-50 (2017).
26. Abrahams, J.L. et al. Recent advances in glycoinformatic platforms for glycomics and glycoproteomics. *Curr Opin Struct Biol* **62**, 56-69 (2020).
27. Wilson, I.B. et al. Analysis of Asn-linked glycans from vegetable foodstuffs: widespread occurrence of Lewis a, core alpha1,3-linked fucose and xylose substitutions. *Glycobiology* **11**, 261-274 (2001).
28. Campbell, M.P. et al. Validation of the curation pipeline of UniCarb-DB: building a global glycan reference MS/MS repository. *Biochim Biophys Acta* **1844**, 108-116 (2014).
29. Pabst, M. & Altmann, F. Influence of electrosorption, solvent, temperature, and ion polarity on the performance of LC-ESI-MS using graphitic carbon for acidic oligosaccharides. *Analytical chemistry* **80**, 7534-7542 (2008).
30. Everest-Dass, A.V., Abrahams, J.L., Kolarich, D., Packer, N.H. & Campbell, M.P. Structural feature ions for distinguishing N- and O-linked glycan isomers by LC-ESI-IT MS/MS. *Journal of the American Society for Mass Spectrometry* **24**, 895-906 (2013).
31. Everest-Dass, A.V., Kolarich, D., Campbell, M.P. & Packer, N.H. Tandem mass spectra of glycan substructures enable the multistage mass spectrometric identification of determinants on oligosaccharides. *Rapid communications in mass spectrometry : RCM* **27**, 931-939 (2013).
32. Harvey, D.J., Royle, L., Radcliffe, C.M., Rudd, P.M. & Dwek, R.A. Structural and quantitative analysis of N-linked glycans by matrix-assisted laser desorption ionization and negative ion nanospray mass spectrometry. *Anal Biochem* **376**, 44-60 (2008).
33. Cooper, C.A., Gasteiger, E. & Packer, N.H. GlycoMod--a software tool for determining glycosylation compositions from mass spectrometric data. *Proteomics* **1**, 340-349 (2001).
34. Ceroni, A. et al. GlycoWorkbench: a tool for the computer-assisted annotation of mass spectra of glycans. *Journal of proteome research* **7**, 1650-1659 (2008).
35. Tsugawa, H. et al. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nature methods* **12**, 523-526 (2015).
36. Moh, E.S., Thaysen-Andersen, M. & Packer, N.H. Relative versus absolute quantitation in disease glycomics. *Proteomics. Clinical applications* **9**, 368-382 (2015).
37. Cohen, M., Hurtado-Ziola, N. & Varki, A. ABO blood group glycans modulate sialic acid recognition on erythrocytes. *Blood* **114**, 3668-3676 (2009).
38. Poole, J., Day, C.J., von Itzstein, M., Paton, J.C. & Jennings, M.P. Glycointeractions in bacterial pathogenesis. *Nat Rev Microbiol* **16**, 440-452 (2018).
39. Moonens, K. & Remaut, H. Evolution and structural dynamics of bacterial glycan binding adhesins. *Curr Opin Struc Biol* **44**, 48-58 (2017).
40. Stanley, P. Golgi glycosylation. *Cold Spring Harb Perspect Biol* **3** (2011).