OTKA K-81839: FINAL REPORT

TITLE: Carbohydrate Structure Data Base for Capillary Electrophoresis Profile Allocation

THE PROJECT:

With the support of the OTKA K-81839 grant, we established a glucose unit (GU) database for capillary electrophoresis profile allocation. The work started with the development of a reliable sample preparation protocol including purification, enzymatic N-glycan release, fluorescent labeling, post-derivatization cleanup and consecutive capillary electrophoresis (CE) analysis of glycans. The method was then optimized by using both capillary zone electrophoresis (CZE) and capillary gel electrophoresis (CGE) modes, towards high-resolution carbohydrate analysis. A large number of glycans was analyzed and hence a relational database containing CE migration positions, expressed as GU values, was established. The database allows full structural assignments by allocating structures to CE profiles based on their relative migration position to a maltooligosaccharide ladder of Glc(α1-4) homooligomers that was apparently appropriate for structural elucidation, un-necessitating further comparison with other homooligomers (e.g., GβU and GNU). The database also contains further important information on exoglycosidase digestion profiles and mass spectrometry data that increases the precision of structural assignment in ambiguous cases. Each glycan database entry includes (i) a graphical and symbolical representation of the structure depicting monosaccharide sequence and linkages, (ii) a display of CE migration time expressed as GU values with the corresponding standard deviations (calculated from 6 parallel runs), (iii) optionally further information containing exoglycosidase digest product and mass values, and (iv) an extensive literature reference list was also added to all database entries. The project was divided into two phases such as data generation and database development. The data generation phase utilized capillary electrophoresis with laser induced fluorescence detection (CE-LIF) to analyze a large variety of glycans in respect to their migration in reference to a standard maltooligosaccharide ladder, while the database development (in silico part) covered the computational work based on the experimental results and user interface design for a web based database.

SHORT DESCRIPTION OF THE RESULTS

Glycans are among the most complex groups of biopolymers with their chemical natures ranging from neutral, linear polymer structures to branched, highly polar or even charged molecules. This great structural diversity of glycans makes carbohydrate analysis extremely challenging, especially as glycans lack chromophore / fluorophore moieties and, in many instances, easily ionizable groups. Therefore, high performance analytical techniques, such as hydrophilic interaction chromatography (HILIC), graphitized carbon chromatography, microfluidics, ultra-performance liquid chromatography and mass spectrometry usually need some kind of pre-analysis derivatization. Among the electric field mediated separation techniques capillary electrophoresis (CE) is one of the most powerful methods to analyze complex glycans. Fluorescence-tagging procedures in conjunction with laser-induced fluorescence detection satisfy sensitivity requirements, and thus represent an established detection technique in high-sensitivity profiling of glycan mixtures by capillary electrophoresis [Guttman 2012, 2013]. A novel bioanalytical device was also developed as part of this project [Kerekgyarto et al 2012].

First, we developed a robust separation method for fast, accurate and reproducible carbohydrate analysis by capillary electrophoresis. N-glycans were enzymatically released by

in solution PNGaseF digestion from glycoproteins and fluorescently derivatized with aminopyrene trisulfonate (APTS) for subsequent CE analysis with laser induced fluorescent detection (LIF). Excess fluorophore removal by HILIC bead containing pipette tips was optimized. To increase injection-to-injection reproducibility and eliminate possible experimental variance, a lower (APTS labeled maltose) and upper (2-aminoacridone labeled glucuronic acid) bracketing standard was introduced. Relative migration times were calculated by using the bracketing standard boundaries. These normalized migration times were subsequently converted to glucose unit (GU) values using a fifth order time based standardization against a maltooligosaccharide ladder. In order to automate this process, a script was written in Matlab version 7.6 (The Math Works, Natick, MA). The precision of the applied method was evaluated using n=6 repetitions and the GU value deviations were < 0.1% RSD. The high precision of the developed separation method and parameters allowed input the resulting values to a comprehensive glycan GU database for capillary electrophoresis [Mittermayr et al 2012; Szekrenyes et al 2013].

For full elucidation of the glycan structures, carbohydrate sequencing was accomplished by consecutive and/or parallel digestion of the glycan pool of interest, mediated by an array of carefully designed mixtures of enzymes consisting of a set of exoglycosidases with accurately defined linkage and monosaccharide unit specificity. After enzymatic digestion, the resulting fragments were reanalyzed by CE-LIF and the migration time shifts between the traces were carefully evaluated. Electrophoretic migration time shifts between the undigested and the digested structures provided the means for monosaccharide type, linkage, anomericity and sequence determination. Both, commercially available glycan standards with known structures and glycans derived from standard glycoproteins (fetuin, transferrin, alpha1-acid glycoprotein, IgG) were analyzed by CE-LIF and their glucose unit values calculated accordingly. Specific enzymatic digestion of the glycan standards resulted in predictable loss of corresponding substructures and the difference in migration time was observed after reanalyzing the product by CE-LIF. These migration time shifts, originated from the loss of specific monosaccharides (possibly even linkage specific), as well as the migration time of the product were converted to GU and recorded. Consequently, this information was employed during the structural elucidation of sequential enzymatic digest of unknown glycans for database entries [Mittermayr et al 2011].

Particular attention was paid to the presence of sialic acid residues that caused increased charge and subsequently faster migration time in electrophoresis. Digestion with various sialidases and successive monitoring of the peak shifts to the asialo counterpart positions can be difficult if not impossible for lower abundant sialoforms due to only minor increases in their respective relative peak areas. Exhaustive structural annotation of the peaks corresponding to sialylated oligosaccharides in the CE-LIF traces was achieved by the combination of a first dimensional anion exchange fractionation and targeted exoglycosidase digests. In the first dimension, the oligosaccharide pool was separated based upon the degree of sialylation present, resulting in neutral, mono- (S1) di- (S2), tri- (S3) and tetra-sialylated (S4) fractions. Each fraction was then profiled by CE-LIF. This two dimensional approach using weak anion exchange fractionation as first dimension proved to be a powerful technique for structural elucidation of sialylated components [Mittermayr et al 2013].

During the experimental data generation we also started the evaluation and analysis of the generated data including examination of completeness and correctness by statistical approaches. Parallel with this endeavor, we have annotated additional structures by including

IgG, haptoglobin and other high abundant serum glycoproteins in the work. At this stage of the work we have identified all neutral and sialylated glycans of IgG. Also, the neutral glycans of haptoglobin including their possible glycosylation changes in different diseases (pneumonia, COPD and lung cancer) were structurally elucidated in order to identify disease specific features for the database. Following our standard protocol, the haptoglobin glycans were enzymatically released, fluorescently labeled and profiled by capillary electrophoresis. Disease associated changes such as core and antennary fucosylation were investigated by targeted exoglycosidase digestions and their levels were compared in the different patient groups. The important phenomena of branching and fucosylation degrees were established and measured. Statistical analysis was used to examine the structures responsible for the observed differences in fucosylation and branching [Varadi et al 2013]. Comparative core fucosylation analysis of some major therapeutic antibody N-glycans by direct infusion electrospray ionization mass spectrometry and capillary electrophoresis – laser induced fluorescence detection was also accomplished [Wang et al 2013]. Neoglycoproteins were also generated and evaluated within this project [Kerekgyarto et al 2012] along with genetic profiling of the relevant genes [Mesko et al 2012].

On the computational part, locally implemented database and algorithms were transferred to a remote server environment. After completing the setup and implementation, database, algorithms and user interface was thoroughly tested and adapted in case of deficiencies. Regarding this *in silico* part of the project, database scaffold design was successfully applied. The database scaffold was set up locally and populated with the experimentally obtained data. Furthermore, the database was prepared for public dissemination on the internet providing the experimentally obtained GU values from our study about the glycosylation of several important glycoproteins of haptoglobin (18 structures), human IgG (32 structures), ribonuclease B (9 structures), transferrin (18 structures) and also with commercially available glycan standards (20 structures). Pursuing increased throughput and associated structural annotation confidence, we compared 1.7 μm hydrophilic interaction phase in ultrahigh performance liquid chromatography (UPLC) with CE-LIF for rapid and comprehensive characterization of N-glycans released from polyclonal IgG of healthy human serum. Combination of the data individually generated using both techniques demonstrated that comprehensive structural annotation was possible within the analysis time of only 20 minutes due to the advantageous orthogonality of the separation mechanisms [Mittermayr et al 2011].

As an extension of the project, we applied combined chromatographic and fractionation techniques to generate a large set of individual fractions representing the human plasma proteome (referred to as Analyte Library), with the goal to use the relevant library fractions in conjunction with mAb proteomics to characterize glycosylation differences. As a proof of concept, the separation route of vitamin D-binding protein (an antibody proteomics lead) was followed in all major fractionation levels by dot blot assay in order to identify the library fractions it accumulated in and the identity of the antigen was verified by Western blot. A mini dot blot array system based assay was applied to identify and compare the fractionation routes of various disease specific proteins by using monoclonal antibodies raised against them. In the majority of the cases the distribution of the antigens in the fractions matched the expectations predicted based on physical, chemical and biochemical properties, like hydrophilicity, hydrophobicity, molecular mass, complex forming ability, etc. [Kovacs et al 2011]. The Library was interrogated with 10 mAbs raised against complex protein mixtures where the exact antigen was not known. Antigens of biomarker potential were identified following the workflow of high-throughput dot blot array screening, immunoprecipitation,

Western blotting and mass spectrometry. As a continuation of the project, all Library fractions will be evaluated for their glycosylation profile with the consideration that the deglycosylated forms in cancer patients might eliminate the molecular precursors [Kovacs et al 2013].

SUMMARY

The development of a robust separation method for fast, accurate and reproducible glycoprotein derived carbohydrate analysis by capillary electrophoresis with laser induced fluorescent detection (CE-LIF) for a glycan migration library generation was successfully accomplished. The excellent resolving power and high detection sensitivity of CE-LIF offered the ability to separate and identify both positional and/or linkage glycan isomers, based on their hydrodynamic volume to charge ratio differences. Structural elucidation of glycans by consecutive enzymatic digestion of carbohydrates using exoglycosidase combinations, followed by capillary electrophoresis separation of the digests was also implemented. In addition, offline weak anion exchange chromatography (WAX) fractionation was introduced and proved to be a powerful technique for deeper structural studies of highly sialylated glycan structures. Comparison of the migration times of the exoglycosidase digestion products to a standard oligosaccharide ladder enabled to build the migration shifts library (GU values), due to cleavage based on the actual exoglycosidases used; thus the exact sequence of each oligosaccharide in a glycan pool to be deciphered with high confidence. The carbohydrate structure data base for capillary electrophoresis profile allocation was posted online as planned and publicly accessible through a user friendly web-based interface bringing CE separation based glycan analysis within the reach of the research community. (http://glycobase.nibrt.ie/glycobase/show_nibrt.action).

With the support of this grant we:

- Developed a robust, fast and reproducible glycan analysis method and corresponding parameters for capillary electrophoresis
- Determined carbohydrate GU values for over 80 distinct glycan structures (further 85 is in process)
- Evaluated the completeness and correctness of generated data by statistical analysis
- Database layout and scaffold design was finalized and a beta version posted online (GlycoBase 3.2.3).
- Some of the annotated glycans were examined in various diseases to prove that the developed method can be efficiently applied
- Identified differently linked sialic acids and their antennary structures
- Calculated branching and fucosylation degrees to better characterize disease associated changes
- An extension of the project was implemented via applying combined chromatographic and precipitation techniques to generate a large set of individual fractions representing the human plasma proteome (referred to as Analyte Library), with the goal to use the relevant library fractions in conjunction with mAb proteomics to characterize significant glycosylation differences
- Public dissemination: Publications: 26 (ΣIF=90.831), lectures: 26 and posters: 31.