Glycobiology, 2019, vol. 00, no. 00, 1–3
doi: 10.1093/glycob/cwz102
Regular Manuscripts

Developmental Biology

Expanding the capillary electrophoresis-based glucose unit database of the GUcal app

Gabor Jarvas[1,2], Marton Szigeti[1], Matthew P Campbell[3], and Andras Guttman[1,2,*]

[1]Translational Glycomics Group, Research Institute of Biomolecular and Chemical Engineering, University of Pannonia, 10 Egyetem, Veszprem 8200, Hungary, [2]Horváth Csaba Memorial Laboratory of Bioseparation Sciences, Research Centre for Molecular Medicine, Faculty of Medicine, University of Debrecen, 98 Nagyerdei, Debrecen, 4032, Hungary, and [3]Institute for Glycomics, Griffith University at Gold Coast, Southport, QLD 4215, Australia

*To whom correspondence should be addressed: Tel: +36-88-624063; Fax: +36 52 255 539; e-mail: a.guttman@northeastern.edu

Received 3 September 2019; Revised 22 November 2019; Accepted 3 December 2019

Abstract

GUcal is a standalone application for automatically calculating the glucose unit (GU) values for separated *N*-glycan components of interest in an electropherogram and suggests their tentative structures by utilizing an internal database. We have expanded the original database of GUcal by integrating all publicly available capillary electrophoresis (CE) data in the GlycoStore collection (https://www.glycostore.org) and with in-house measured GU values. The GUcal app is freely available online (https://www.gucal.hu) and readily facilitates CE-based high throughput GU value determination for first line structural elucidation.

Key words: capillary electrophoresis, database, glucose unit, *N*-glycan, structure identification

Introduction

Glycan analysis is an important field of bioanalytics, especially in clinical and biopharmaceutical research as well as in production quality control of glycoprotein therapeutics. Structural elucidation of the glycan moieties of glycoproteins is a difficult task due to their inherited structural variety/complexity (linkage and positional isomers) and physical–chemical similarities (Sastre Torano et al. 2019). Currently, the necessary sample preparation methods and liquid chromatography and capillary electrophoresis (CE)-based separation workflows are well developed, but the related bioinformatics tools are not adequately addressed, especially not for CE (Lu et al. 2018; Shubhakar et al. 2015; Walsh et al. 2016).

Over the last few years, several large-scale initiatives have been launched to coordinate and support technology development in the glycomics field including the Human Glycome Project (https://human-glycome.org) that is exploring the role of glycans in the human body, their biological relevance and regulation mechanisms as well as analytical method development. Beside liquid chromatography, CE has major role in glycan analysis due to its high resolution capable of separating carbohydrates by charge-to-size ratio (Hajba et al. 2016). However, the major limiting factor is the lack of bioinformatics tools for CE-based analysis and subsequent structural elucidation. Manual calculation of GU values is a time consuming and complex task, thus not suitable for the automated workflow high-throughput glycan analysis requirement.

In 2015, we introduced GUcal (https://www.gucal.hu), an application that automatically calculates the glucose unit (GU) values for separated sample components of interest in an electropherogram and assigns putative structures by utilizing an internal database (Jarvas et al. 2015). The original GUcal database contained 32 *N*-glycan entries released from the conserved ASN_{297} site of human IgG. Features of the GUcal tool including the demonstration of the resolution have been published in a detailed tutorial paper (Jarvas et al. 2018). Recently, we have expanded this database by integrating all publicly available CE data from GlycoStore as well as in-house measured GU values. In this Application Note, we summarize the latest developments of GUcal with special attention on the expanded built-in database.

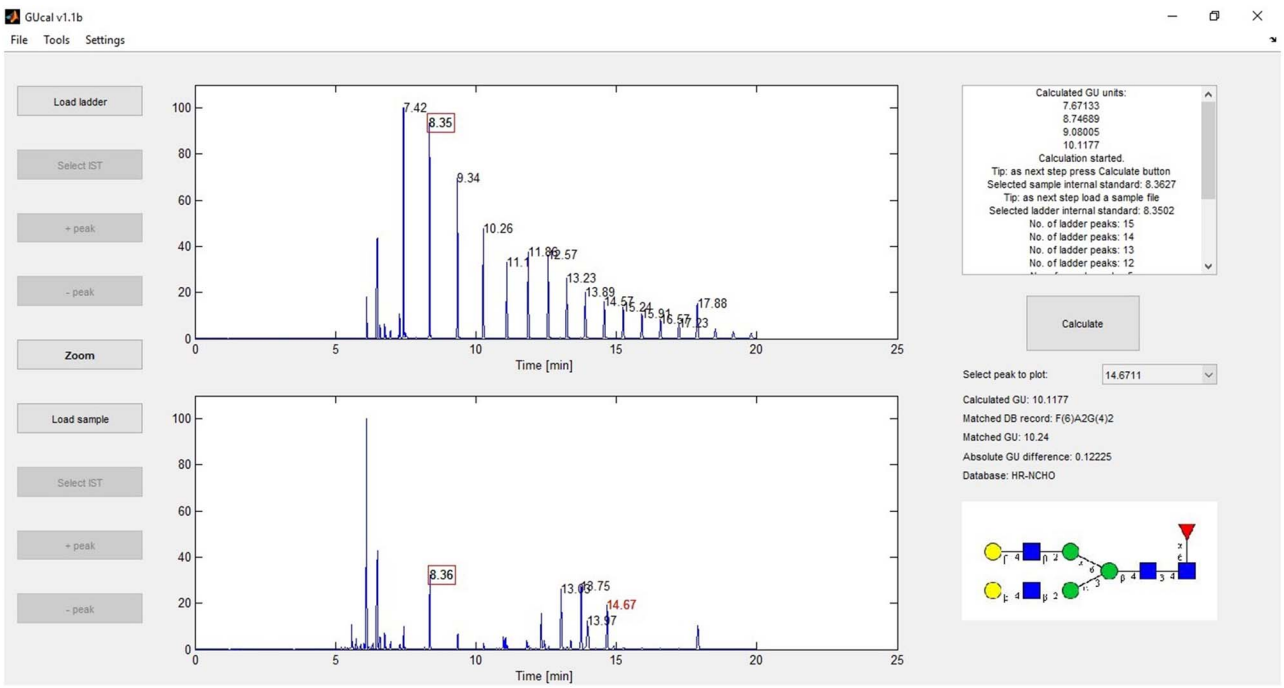

Fig. 1. Screenshot of the GUcal app showing a database match after a complete GU calculation session. The upper panel shows the maltooligosaccharide ladder trace while the lower electropherogram is the human immunoglobulin G sample. Separation conditions: as noted in paragraph "completing the database." The peak of interest is marked in red at 14.67 min. Peaks with migration times of 8.35 (ladder) and 8.36 (sample) min are the internal standards and boxed red. This figure is available in black and white in print and in colour at Glycobiology online.

Materials and methods

Coverage, reliability and relevance of the utilized database is essential for any GU value-based structural elucidation exercise. Thus, the original internal database of the GUcal app has been expanded by integrating all CE data from GlycoStore together with in-house measured and validated GU values from frequently used glycan libraries.

Link the GlycoStore database

GlycoStore is a glycan separation-based database that integrates data collections from academic and commercial partners (Zhao et al. 2018). It provides access to over 90 unique *N*-/*O*-glycan structure entries with more than 500 CE migration and GU values (Abrahams et al. 2018; Campbell et al. 2008), where multiple entries for certain structures represent repeated measurements or independent values from different sources. Furthermore, it reports details on the experimental conditions and acquisition modes for the glycan entries.

Completing the database

To expand the coverage, diversity and application of GUcal and GlycoStore, we have integrated in-house analyzed CE-based GU values for 65 *N*-glycan structures (abbreviated names) from standard partitioned libraries and carbohydrates labeled with aminopyrenetrisulfonate (APTS). CE separations were performed on a PA 800 Plus Pharmaceutical Analysis system (SCIEX, Brea, CA), equipped with a solid state laser-induced fluorescent detector (excitation: 488 nm, emission: 520 nm). CE measurements were carried out using two different, commercially available CE kits, namely the Carbohydrate Labeling and Analysis Kit and the Fast Glycan Sample Preparation and Analysis kit (both from SCIEX). Human immunoglobulin G sample was from Sigma-Aldrich (St. Louis, MO). The maltooligosaccharide ladder was from Grain Processing Corporation (Muscatine, IA). The effective length of the separation capillary in the Carbohydrate Labeling kit was 40 cm (50-cm total length, 50 µm i.d.). The applied electric field strength was 600 V/cm in reversed polarity mode (cathode at the injection side) at 20°C. Samples were pressure injected by applying 1 psi (6.89 kPa) for 5 s. The bracketing standard mixture, required for the GU value calculation, was electrokinetically injected by applying 3 kV for 3 s.

The effective length of the separation capillary in the Fast Glycan kit was 20 cm (30-cm total length, 50 µm i.d.). The applied electric field strength was 1 kV/cm in reversed polarity mode (cathode at the injection side) with 0.17-min ramp-up time at 30°C. Samples were injected as follows: water by 3.0 psi (20.67 kPa) for 5.0 s water, then the sample by applying 1 kV for 1 s and finally the bracketing standard mix by applying 1 kV for 1 s. Please note that the internal standards were parts of the CE kits. The 32 Karat version 9.1 software package (SCIEX) was used for data acquisition and analysis. Sixty-five glycan structures together with their GU values, neutral mass and doubly charged APTS labeled mass are listed in the Supplementary data.

The new built-in database consists of 131 unique glycan structure entries (there are some overlapping records between the GlycoStore and our in-house measured GU value sets) with the corresponding GU values calculated based on more than 630 CE migration time data together with their Consortium for Functional Glycomics (CFG) notations. The database covers a wide GU value range from 2.34 to 14.81, which covers all major classes of glycans originating from common biological samples, biologics and biosimilars.

Discussion

GUcal is an efficient tool for interpreting CE glycan data and we demonstrate its improved utility with the inclusion of an extended

database, which is continuously curated and further development is planned by adding glycans from human serum, aberrant glycosylation forms due to malignant transformation as well as O-linked glycans. In addition, this updated version provides significant improvements such as: 1) users can choose the database subset to search; 2) the output file lists the three closest matching glycan structures; 3) a minimum peak height can be defined for automatic peak detection in the sample trace; and 4) CFG notation of the matched database entries, as shown in Figure 1. The GUcal app suggests possible structures for each marked peak of the sample trace; however, these suggestions are relevant only when the analyzed CE traces were measured with identical separation conditions as noted for the corresponding database subset.

Funding

National Research, Development and Innovation Office (NN127062, 2018-2.1.17-TÉT-KR-2018-00010); Institute for Glycomics; BIONANO_GINOP-2.3.2-15-2016-00017 project of the Hungarian Government; János Bolyai Research Scholarship of the Hungarian Academy of Sciences and ÚNKP-19-4 New National Excellence Program of The Ministry for Innovation and Technology. This is contribution # 151 from the Horváth Csaba Memorial Laboratory of Bioseparation Sciences.

Supplementary data

Supplementary data for this article is available online at http://glycob.oxfordjournals.org/.

Conflict of interest statement

The authors have declared no conflict of interest.

Abbreviations

APTS, aminopyrenetrisulfonate; CE, capillary electrophoresis; CFG, Consortium for Functional Glycomics; GU, glucose unit.

References

Abrahams JL, Campbell MP, Packer NH. 2018. Building a PGC-LC-MS N-glycan retention library and elution mapping resource. *Glycoconj J*. 35:15–29.

Campbell MP, Royle L, Radcliffe CM, Dwek RA, Rudd PM. 2008. GlycoBase and autoGU: Tools for HPLC-based glycan analysis. *Bioinformatics*. England: Oxford, 24:1214–1216.

Hajba L, Csanky E, Guttman A. 2016. Liquid phase separation methods for N-glycosylation analysis of glycoproteins of biomedical and biopharmaceutical interest. A critical review. *Anal Chim Acta*. 943:8–16.

Jarvas G, Szigeti M, Guttman A. 2015. GUcal: An integrated application for capillary electrophoresis based glycan analysis. *Electrophoresis*. 36:3094–3096.

Jarvas G, Szigeti M, Guttman A. 2018. Structural identification of N-linked carbohydrates using the GUcal application: A tutorial. *J Proteome*. 171:107–115.

Lu G, Crihfield CL, Gattu S, Veltri LM, Holland LA. 2018. Capillary electrophoresis separations of glycans. *Chem Rev*. 118:7867–7885.

Sastre Torano J, Ramautar R, de Jong G. 2019. Advances in capillary electrophoresis for the life sciences. *J Chromatogr B, Anal techn biomed life sci*. 1118–1119:116–136.

Shubhakar A, Reiding KR, Gardner RA, Spencer DI, Fernandes DL, Wuhrer M. 2015. High-throughput analysis and automation for glycomics studies. *Chromatographia*. 78:321–333.

Walsh I, Zhao S, Campbell M, Taron CH, Rudd PM. 2016. Quantitative profiling of glycans and glycopeptides: An informatics' perspective. *Curr Opin Struct Biol*. 40:70–80.

Zhao S, Walsh I, Abrahams JL, Royle L, Nguyen-Khuong T, Spencer D, Fernandes DL, Packer NH, Rudd PM, Campbell MP. 2018. GlycoStore: A database of retention properties for glycan analysis. *Bioinformatics*. 34:3231–3232.