CHAPTER

19

Database search assisted *N*-glycan structure identification

Gabor Jarvas[a,b], *Marton Szigeti*[a,b], *Matthew P. Campbell*[c], *and Andras Guttman*[a,b]

[a]Research Institute of Biomolecular and Chemical Engineering, University of Pannonia, Veszprem, Hungary, [b]Horváth Csaba Memorial Laboratory of Bioseparation Sciences, Research Centre for Molecular Medicine, Faculty of Medicine, University of Debrecen, Debrecen, Hungary, [c]Institute for Glycomics, Griffith University, Gold Coast, QLD, Australia

Abbreviations

η	viscosity
μ	electrophoretic mobility
v	velocity
A	preexponential factor
APTS	8-aminopyrene-1,3,6-trisulfonic acid
CE	capillary electrophoresis
CE-LIF	capillary electrophoresis with Laser-induced fluorescence
DP	degree of polymerization
E	electric field
E_a	activation energy
EOF	electro osmotic flow
F_e	electrical force
F_f	frictional force
GlcNAc	*N*-Acetylglucosamine
GU	glucose unit
MS	mass spectrometry
NMR	nuclear magnetic resonance
PNGaseF	peptide-*N*-glycosidase F
Q	charge

Carbohydrate Analysis by Modern Liquid Phase Separation Techniques
https://doi.org/10.1016/B978-0-12-821447-3.00010-X

1 Introduction

As an orthogonal separation technique to liquid chromatography, capillary electrophoresis (CE)-based glycan analysis with subsequent structural identification is among the most frequently used tools in the glycoanalytical field. In this chapter, we aim to introduce the relevant theoretical background, demonstrate the essentials of each methods and provide practical application examples. The biological relevance of glycans can be related to their *N*- and *O*-linked glycans, but this chapter is focused on structural elucidation of released asparagine linked oligosaccharides. Currently, several commercial kits and different methods reported by academic laboratories are available for high throughput *N*-glycan analysis, both exploiting the outstanding separation resolution obtained by CE. While the necessary instrumentation and the related sample preparation methods are well developed, related bioinformatics tools are only just starting to emerge. Glycan structure identification is a complex and challenging task, which combines physical-chemical, separation science, and bioinformatics-computational aspects. Therefore, special attention is given where the discussed topics are not closely related to the common subjects. The traditional glucose unit calculation approach, the virtual ladder method and glycan sequencing by serial or array based exoglycosidase digestion techniques are all considered for identifying glycans. Other methods, such us lectin arrays, nuclear magnetic resonance (NMR), and mass spectrometry (MS) are not discussed in this review.

2 Theory of glycan electromigration

Among many other advantages of CE, the capability of separating *N*-glycans with same monosaccharide sequence but different positional and linkage isomers (alpha or beta) makes it an ideal tool for analytical glycomics applications [1]. The unique separation means of CE is considered to be orthogonal to liquid chromatography separation mechanism. The deeper understanding of the nature of the electromigration behavior of glycans in free zone electrophoresis and in polymer networks allows the development of new glycoinformatics methods.

Generally, when a uniform electric field, E is applied to an analyte molecule with the net charge of Q, an electrical force (F_e) acts on the ion their relationship can be formalized as follows:

$$F_e = Q \cdot E \tag{1}$$

which is the fundamental equation of all electric field-mediated separation methods. Eq. (1) is opposed by a frictional force (F_f):

$$F_f = f \cdot \frac{dx}{dt} \tag{2}$$

where f is the translational friction coefficient and dx and dt are the respective distance and time increments.

When the resultant of the two forces (F_e and F_f) are not counterbalanced, the analyte ions migrate with a steady state velocity (v) as Eq. (3) describes:

$$v = \frac{dx}{dt} = E \cdot \frac{Q}{f} \tag{3}$$

The translational friction coefficient is considered to be proportional to the viscosity (η) of the background electrolyte [2]:

$$f = c \cdot \eta \tag{4}$$

The proportional constant c in Eq. (4) is influenced by molecular configuration, for example, for small spherical molecules, c is defined as $6\pi r$, where r is the radius of the ion [3].

Electrophoretic mobility (μ) is defined as the electric field normalized electromigration velocity (v/E) and can be expressed as:

$$\mu = \frac{Q}{c \cdot \eta} \tag{5}$$

As stated by the Eyring-Polanyi equation [4], the viscosity of the separation medium can be defined as:

$$\eta = A \cdot e^{E_a / RT} \tag{6}$$

where A is the preexponential factor, E_a is the activation energy of the viscous flow, and R is the universal gas constant. The electrophoretic mobility of a solute ion can be expressed by combining Eqs. (5) and (6),

$$\mu = \frac{Q}{const} \cdot e^{-E_a / RT} \tag{7}$$

where *const* represents a collection of constants including c and A from Eqs. (5) and (6).

The apparent electrophoretic mobility of the analyte molecules is influenced by differences in activation energy requirement. The activation energy is required by the solute molecule to overcome the obstacles created by the separation medium [5]. In practice, the activation energy values are usually derived from the slopes of the Arrhenius plots [6] of logarithmic electrophoretic mobility vs reciprocal absolute temperature:

$$\ln \mu = \ln \frac{Q}{const} - \frac{E_a}{RT} \tag{8}$$

As apparent from the theoretical treatment above, the measured electrophoretic mobility of glycans is the result of numerous physico-chemical phenomena. Composition of the separation matrices, which are almost

exclusively physical gels (noncross-linked entangled polymer solutions) in CE-based separation of glycans, as well as physico-chemical parameters such as the separation temperature, molecular characteristics of the solute molecules, etc., play essential roles in the electromigration of carbohydrates [5]. Pioneering research studies and textbooks stated that the CE separation of glycans is based on the charge-to-mass ratio of the solute molecules [7, 8], but recent results show that the separation is—with a more precise term—the function of charge-to-size/hydrodynamic volume of the analytes [9–11]. Considering the unique physico-chemical characteristics of, for example, maltooligosaccharides, which tend to be helically structured depending on their degree of polymerization, hydrophobic interactions can be anticipated between the separation matrix and the analyte molecule [12, 13]. Among others, Chiesa and Horváth [14], Nashabeh and El Rassi [15], Mittermayr and Guttman [16] investigated the influence of molecular configuration and conformation on the electromigration of oligosaccharides and concluded that special attention must be given when homooligosaccharide ladders are employed for the evaluation of CE separations of unknown species. The electromigration behavior of glycans can be better understood by using the so-called activation energy concept [8–10] originally developed by Eyring [11] and first applied to the field by Guttman and Cooke [12].

In a recent paper [17], Jarvas et al. studied the electromigration of charged fluorophore-labeled oligosaccharides in polyethylene oxide solutions utilizing the activation energy concept. The authors published evidence on the solute-matrix interaction, and interestingly found that it could be alleviated by the addition of an organic modifier to the background electrolyte. Other work from the Guttman group focused on the electromigration of oligosaccharides in viscosity modifier and polymeric additive containing background electrolytes [17, 18]. Activation energy curves were constructed based on the slopes of the Arrhenius plots from CE separation of APTS-labeled maltooligosaccharides in the temperature range of 20–50°C. As a consequence of the linear polyacrylamide addition, solute size-dependent activation energy variations were observed for the maltooligosaccharides with polymerization degrees below and above maltoheptaose. This is because of molecular conformation changes (first full helical turn) and possible matrix interaction effects.

The summarized theoretical considerations of glycan electromigration allow the understanding of some unexpected experimental findings that may support data interpretation to avoid any misinterpretation. One of the most important aspects of any liquid phase separation method is the change of the run-to-run repeatability caused by migration/retention time measurement inaccuracy [19].

CE is an outstandingly powerful separation technique, but without additional analytical dimensions, such as exoglycosidase digestion based

sequencing or MS, it does not provide all structural information unless accompanied by GU calculations relative to publicly available glycan structural databases. Obviously, the unprocessed migration times of peaks can be used for preliminary glycan structure assessment, however, because of unavoidable peak shifts, calibration of the migration time is necessary. The measured migration time of a peak of interest is the result of two superimposing transport mechanisms, namely electroosmotic flow (EOF) and electrophoresis. The EOF is the motion of the background electrolyte toward the cathode because of the formation of the electric double layer on the inner surface of the separation capillary. The effect of the EOF on the convective transport can be neglected at the pH under typical glycan separation conditions, especially when polymeric matrices are used [20]. Electrophoresis, on the other hand, is the motion of charged analytes under the influence of an applied electric field. Deposition of any sample or sample matrix molecules, microscopic, and macroscopic change in the composition of the background electrolyte, alteration of the pH or temperature may occur such surface charge heterogeneity. Furthermore, the apparent electrophoretic mobility may be further influenced by solute-matrix interactions [17]. Conversion of measured migration times to a kind of relative migration times can increase precision, which in the practice led to the utilization of the so-called glucose unit approach.

3 Internal/bracketing standard approach

The basic idea of the glucose unit (GU) glycan identification approach is to address the need for capillary/separation system calibration. The GU concept—analogously to the well-known Kovats retention index in gas chromatography [21]—is based on the comparison of the migration times of unknown analyte peaks with a mixture of homologous oligosaccharides (usually referred to as "ladder") to provide migration time normalization [19]. Direct comparison of the obtained GU values of commercially available standards (e.g., well characterized carbohydrate molecules) against GUs of unknown sample components of interest is one of the essential tasks of database mediated glycan structure assignment. GU based structural identification was originally suggested by Yamashita et al. [22] and later on widely utilized by others [16, 23–25].

For a certain unknown peak in an electropherogram the GU value can be calculated based on the migration time of its adjacent sugar ladder peaks:

$$GU_X = G_N + \frac{t_X - t_N}{t_{N+1} - t_N} \tag{9}$$

where GU_X is the glucose unit value of the peak of interest, G_N is the degree of polymerization of the preceding maltooligosaccharide

$(Glc(\alpha 1 \rightarrow 4)_n)$ ladder peak, t_X is the migration time of the peak of interest, t_N and t_{N+1} are the migration times of the maltooligomers immediately preceding and following the peak of interest, respectively [16]. The utilization of GU values instead of raw, unprocessed migration time data alleviates any discrepancies caused by separation parameter changes, column (capillary) history, or buffer composition alteration. This normalization of the migration times is a vertical alignment of the electropherograms, which is done by the exploitation of the co-injected internal standards.

Manual calculation of GU values for a complex sample is time consuming and labor intensive. To ease this work, a freely available application, GUcal was developed (www.gucal.hu) to automatically calculate the GU values for any peaks of interest in an electropherogram [26]. Depending on the actual needs of data interpretation, GUcal is capable to process traces having one internal standard (dedicated peak, e.g., maltose) or for increased precision the bracketing standard approach (two dedicated peaks, e.g., maltose and maltopentadecaose). Practical aspects of the use of GUcal has been demonstrated in detail in a recent tutorial [19]. After successful GU value calculation, the results are readily applicable for database search assisted glycan structure elucidation, see details in Section 5 of this chapter.

4 The virtual ladder approach

As described earlier, in order to achieve maximum precision and alleviate the effect of drifting migration times on single spectral imaging detection systems (i.e., using a single dye type), GU value calculation requires an accompanying maltooligosaccharide ladder for each sample, which doubles sample processing time. Another alternative approach is, when a single ladder is used for a batch of samples, however, in this scenario the subsequent runs may have slightly different separation conditions resulting migration time shifts. Assuming the most frequently used single wavelength detection systems, co-injection of the maltooligosaccharide ladder with the sample has no practical relevance since multiple ladder and sample peaks could co-migrate making structural identification difficult. On the other hand, in multispectral imaging (i.e., using multiple wavelengths to detect differently labeled species at the same time), co-injection of the ladder could cause displacement effects, thus making data evaluation ambiguous [27]. Therefore, the development of a method with co-injected multiple internal standards that do not interfere with the sample components of interest and do not increase analysis time fulfilled the need for a new, high throughput glycan structure identification technology.

The theoretical treatment on the electromigration of glycans in Section 2 inspired the development of an accurate prediction method to establish a virtual ladder, using information from early and late migrating co-injected components [28]. The virtual ladder is a set of calculated migration times, equivalent to the use of an actual maltooligosaccharide ladder and thus can be used to replace it. Glucose α1–4 linked oligomers of DP2, DP3, and DP15 were used for generating the virtual ladder since they adequately represent the actual (measured) ladder elements and their separation behavior. Furthermore, they migrate prior and after the range of most peaks of interest in biological samples. Maltooligosaccharides with DP < 7 migrate in nonstructured configuration, while DP > 7 form a helical structure [16]. Distinct behavior of maltooligosaccharides suggests that the electromigration of the prehelical and helical forms should be modeled separately. The butterfly diagram in Fig. 1 compares the typical electropherograms of APTS labeled human IgG *N*-glycan pool with the three internal standards DP2, DP3, and DP15 and a maltooligosaccharide ladder. This demonstrates the alignment of peaks on the relative migration time basis using the triple-internal standard approach.

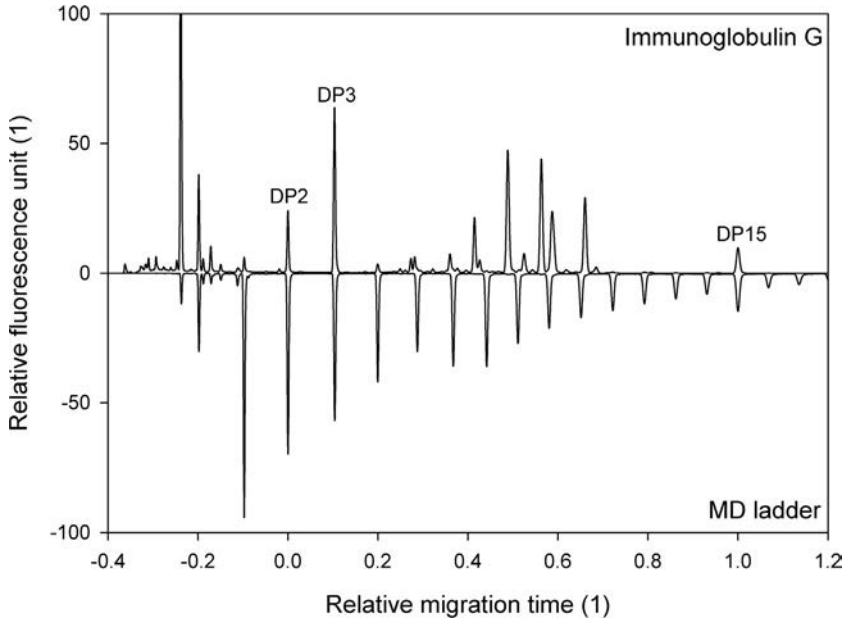

FIG. 1 Comparative mirror diagram of the capillary electrophoresis separations of APTS-labeled human immunoglobulin G *N*-glycans (upper trace) and the linear maltooligosaccharide ladder (lower trace) including the three suggested internal standards of DP2, DP3, and DP15 in both. *Reprinted from G. Jarvas, et al., Triple-internal standard based glycan structural assignment method for capillary electrophoresis analysis of carbohydrates, Anal. Chem. 88(23) (2016) 11364–11367, with permission from American Chemical Society.*

By comprehensively analyzing numerous CE runs, it was observed that the ratio distribution of migration time increments of consecutive peaks was constant. This observation was based on 500 + measured data points including dedicated experiments with altered separation voltage, temperature, capillary type (uncoated and coated), capillary length, injection methods, as well as day-to-day repetition and instrument dependency investigation. Thus, the information about the migration time difference between the DP2 and DP3 peaks enabled accurate calculation of the migration time of shorter structures utilizing this constant distribution of migration time differences.

However, this predictable domain was restricted to shorter maltooligosaccharides (DP < 7) since the migration behavior (and consequently the migration time differences) changed when the sugar oligomer formed a full helical turn at DP > 7. This change resulted in a shift in the calculated migration times at DP > 7, which was addressed by utilizing the migration information of the DP15 peak for accurate ladder step prediction in the range of DP8–DP15. A combination of these two approaches led to a robust calculation algorithm, which provided distinct and very reliable migration times for the virtual ladder. Fig. 2 shows the accuracy of the virtual ladder compared to an actual ladder, as well as its use in the structural analysis of human IgG *N*-glycans.

Similar to the traditional approach, calculated GU values can be searched against database entries to assist glycan structure elucidation. Here, matching or similar GU values (within a defined GU range) suggest the possible glycan structures. Therefore, this novel triple-internal standard approach has been implemented in the GUcal app for single trace evaluation [28] and also applicable for batch processing.

5 Analytical databases

Analytical glycomics has been a rapidly developing and growing glycobiology field [29, 30] that has been boosted by the introduction of glycoprotein-based biotherapeutic drugs. Over the last decade, several bioinformatics initiatives have emerged to support analytical data collections including the Consortium for Functional Glycomics [31], Glycosciences.de [32], formerly BCSDB, now CSDB [33, 34], GlycoSuiteDB [35], EUROCarbDB [36], UniCarb-DB [37], and GlycoBase [38]. A majority of these resources were developed to support MS, liquid chromatography, and NMR data. However, availability of similar tools for CE data processing has been restricted. Typically, each company or academic group developed their own CE separation based GU database and utilized them in their studies.

Currently, the largest data collection is the GlycoStore database [39], built on publicly available experimental data sets from GlycoBase, devel-

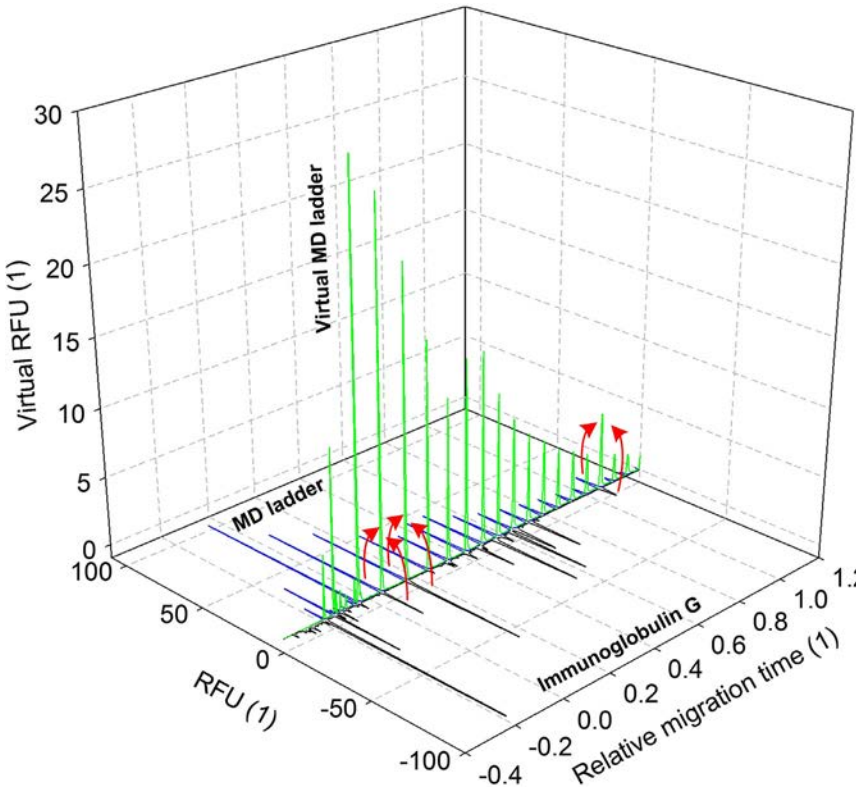

FIG. 2 Comparison of the virtual ladder (calculated) against an actual maltooligosaccharide ladder (measured). *Reprinted from G. Jarvas, et al., Triple-internal standard based glycan structural assignment method for capillary electrophoresis analysis of carbohydrates, Anal. Chem. 88(23) (2016) 11364–11367, with permission from American Chemical Society.*

oped in the Oxford Glycobiology Institute and then the National Institute for Bioprocessing Research and Training (Dublin). It has now been extended to include published and in-house collected data sets from the Bioprocessing Technology Institute, Macquarie University, Institute for Glycomics (Griffith University) and from an industrial partner (Ludger Ltd). GlycoStore provides access to approximately 850 unique glycan structure entries both from liquid chromatography and CE separations, supported by over 10,000 data records.

The GUcal app has been linked with the CE relevant part of the GlycoStore collection as the coverage, reliability, and relevance of the utilized database is essential for the GU value based structural elucidation [40]. Furthermore, the database has been merged with in-house analyzed capillary electrophoresis based GU values for 65 *N*-glycan structures. The resulted data set provides access to over 90 unique *N*-glycan structure entries with more than 500 CE migration and GU values [39, 41], where

multiple entries for certain structures represent repeated measurements or independent values from different sources. In addition, it provides details on experimental conditions and acquisition modes for the glycan entries. It is important to note that high fidelity structural elucidation always require cross-validation of the suggested structures either by exoglycosidase array based sequencing or hyphenated techniques with MS [42].

Albeit, the virtual ladder approach attempts to overcome some of the bottlenecks of traditional GU calculation approaches, which were sensitive to experimental conditions, special attention is still required on the choice of the appropriate database. Unfortunately, the glycomics community lacks such large CE based structural databases. The more similar the separation conditions between the database entries and the actual measurement, the more reliable the suggested structure is.

A frequent criticism against the GU value based structural elucidation of glycans is that the freely available apps have no ability to process multiple runs at the same time in an automated fashion. While the glyXera suite supports multilane CE migration time alignment for capillary gel electrophoresis-based *N*-glycan analysis [43] unfortunately, it does not address GU value calculation or direct structural elucidation. For a description of glycan structure and glycoprotein centric databases readers are referred to GlyGen [44] and GlyCosmos [45] activities.

6 *N*-Glycan sequencing

Due to limited access to commercially available glycan standards [46], partial ambiguity of database assisted structural determination, and co-migration of structures exoglycosidase digestion is often required for high fidelity peak identification and structural elucidation. The background of an entire CE based sequencing method using exoglycosidases was published earlier [47, 48]. Briefly, *N*-linked carbohydrate sequencing starts with the release of the sugar moieties from the polypeptide backbone of the glycoprotein, in most cases using PNGase F [10]. CE based *N*-glycan profiling requires the labeling of the oligosaccharides with a charged fluorophore so support proper electromigration and fluorescence detection option. The most often used dye is the 8-aminopyrene-1,3,6-trisulfonic acid (APTS). Current carbohydrate sequencing strategies involve the use of specific exoglycosidase enzymes to remove the sugar residues from the nonreductive end of the glycan structures. The enzymes can be used sequentially, or arrays (combinations) of enzymes can be applied in a parallel fashion [47]. In both instances, following the reaction steps the resulting glycan pools are analyzed by CE and the GU shifts of all the peaks are recorded. Glycan sequencing starts with cleaving the outermost monosaccharide residues from the nonreducing end (in most instances sialic acids),

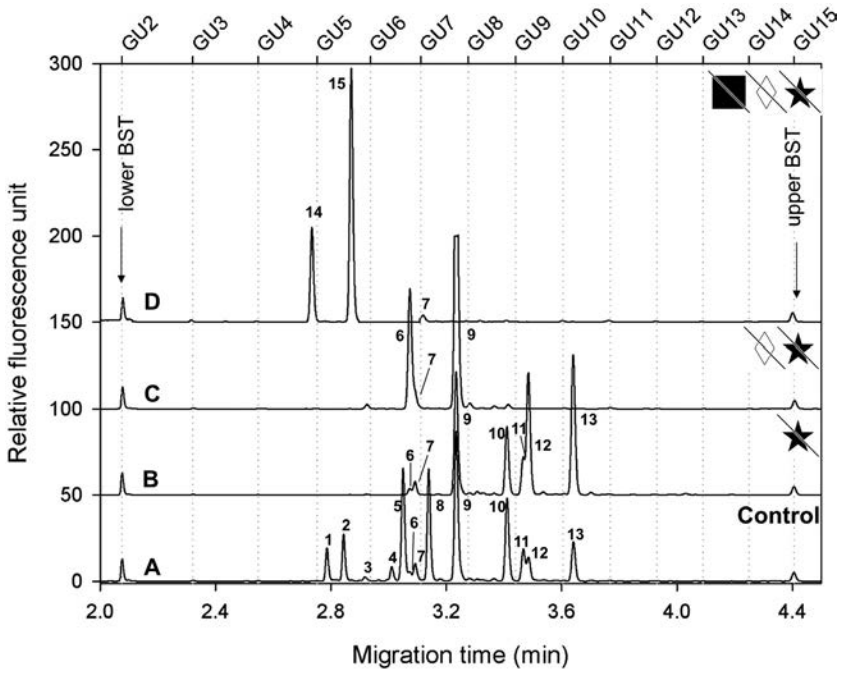

FIG. 3 Representative electropherograms of exoglycosidase sequencing based structural elucidation of APTS labeled glycan sample. Symbols: ★—sialidase, ◇—galactosidase, ■—hexosaminidase. *Arrows* depict the lower and upper bracketing standards. *Reprinted from M. Szigeti, A. Guttman, Automated N-glycosylation sequencing of biopharmaceuticals by capillary electrophoresis, Sci. Rep. 7(1) (2017) 11663, with permission from Springer Nature.*

followed by removal of all galactose, fucose and *N*-acetylhexosamine residues (including antennary and bisecting) until the trimannosyl-chitobiose core structure of all *N*-linked sugars is reached (fucosylated and/or non-fucosylated). Independent of the number of the applied enzymes used, glycan sequencing is a complex process.

The importance of GU value shift calculation in glycan sequencing is demonstrated in the following example. The *N*-linked glycan pool of etanercept was enzymatically released by PNGase F digestion. The liberated carbohydrates were then labeled by APTS and purified with magnetic beads based technology [29] followed by capillary electrophoresis separation with laser-induced fluorescence (CE-LIF) detection. The resulted trace is shown in Fig. 3 and used as a control for subsequent sequencing runs. First, the terminal sialic acids were removed using sialidase α(2→3,6,8,9) enzyme (trace B, Fig. 3). This treatment resulted in the shift of the affected peaks from the higher mobility regime (GU = 4–6, sialylated glycans) to the neutral regime (GU = 7–10). Sialic acid removal has a double effect: it decreases the hydrodynamic volume of the structures and in parallel

also decreases the resulted overall charge, due to the inherent negative charge of sialic acid. Next, the desialylated glycans were further digested by galactosidase β(1 → 3,4,6) (trace C, Fig. 3). Removal of the galactose residues caused further mobility shift, however, in this case toward the higher mobility region (i.e., shorter migration times). Please note, the total charge of the glycans was not affected in this instance, only their hydrodynamic volume decreased. Finally, all antennary and bisecting GlcNAc residues were removed by hexosaminidase β(1 → 2,3,4,6) digestion resulting in two large remaining peaks as observed in trace D Fig. 3. One is the *N*-linked trimannosyl chitobiose core (peak #14) and the other a fucosylated trimannosyl chitobiose (peak #15). Consequently, fucosidase treatment was skipped as the GU values are well defined. In addition, incomplete fucosidase treatment may lead to the appearance of misleading peaks. This partial removal of the core fucose from *N*-linked glycans is most likely due to steric hindrance.

Exoglycosidase sequencing starts from bottom up, that is, by identifying peaks in the simplest electropherogram (trace D in Fig. 3). From its GU value, peak 14 was identified as the trimannosyl chitobiose structure (M3) with GU = 4.94, while peak 15 was its core fucosylated version (F(6)M3) with GU = 5.67. Distinction between peaks 14 and 15 was based on the fact that CE separated these molecules according to their charge to hydrodynamic volume ratio, thus the fucosylated larger structure migrated slower. Peak 7 was probably a high mannose structure since all other residues were cut-off by reaching this step. Considering its GU = 6.89 very close to 7 it was considered as M5. Comparing trace D against C, GU values of the peaks of 6 and 9 in trace C were 6.78 and 7.75, corresponding to the shifts caused by the removal of two antennary GlcNAc residues (0.92 each) and (1.04), depending on their linkage states of 1–6 and 1–3, respectively. Thus, peak 6 was identified as A2, while peak 9 was assigned as FA2. Trace B represents the glycan structure profile after sialidase digestion having several additional major peaks compared to trace C with the GU values of 8.82, 9.16, 9.27, and 10.22. Taking into consideration the GU shifts caused by the removal of the galactose residues, these structures were identified as monogalactosylated biantennary structures with 1–6 (GU = 8.82) and 1–3 (GU = 9.16) linkages. Furthermore, peaks 12 and 13 were assigned to as digalacto biantennary structures without (GU = 9.27) and with core fucosylation (GU = 10.22). Interesting to note that similarly to the above GlcNAc shifts, the galactoses on the 1–6 and 1–3 antennas contributed different GU shifts values of 1.07 and 1.41, respectively, which is probably because of hydrodynamic volume differences caused by the linkage orientation. The shifts both for the digalactosylated A2G2 and FA2G2 structures was practically the sum of these two GU = 2.49 and GU = 2.47. The last step of the peak identification process was the comparison of the electropherogram of the control glycan pool with the sialidase treated trace

(B). These shifts were due to the presence of several disialo and monosialo structures. Peak 1 was the fastest migrating glycan with GU = 5.18 with probably two sialic acid residues on each arms of A2G2, that is, A2G2S2 structure. This was further verified by the apparent high GU shift (4.09), which is not surprising if one considers the additional charge of sialic acid. Similarly, peak 2 (GU = 5.5) was identified as FA2G2S2, the doubly sialylated form of FA2G2 with a GU shift 4.72. Peaks 3 and 4 with GU values of 5.90 and 6.42 represented the GU shifts of 2.92 and 2.74 relative to FA2(6)G1 and FA2(3)G1 structures and identified FA2(6)G1S1 and FA2(3)G1S1. Please note, distinction between 2 and 3 or 2 and 6 linkages was based on assumption that similarly to as explained earlier, structures with 2–6 linkages electromigrated faster than their 2–3 counterparts. Finally, peak 5 was identified as A2G2S1 (GU = 6.65) with the GU shift of 2.62 relative to A2G2, which corresponded to the contribution of one sialic acid.

Recently, a CE based, fully automated *N*-glycosylation sequencing technique was published [49], which radically reduced the wet laboratory tasks and decreased analysis time. The evaluation of the obtained electropherograms is also very time consuming as all steps require GU value calculation for all peaks of interest, thus it is an essential need of applications capable to easy GU and shift calculations. Please note, both profile based GU value calculation and exoglycosidase sequencing mediated structural identification of an unknown glycan pool is rather complex and often requires additional separations of glycan standard libraries. Finally, it should be emphasized that high fidelity structural elucidation based on CE separations may require cross-checking of the suspected structures with orthogonal techniques such as LC-MS and CE-MS [42].

7 Conclusion

In this chapter, the recent advances of database search assisted glycan structure identification techniques are discussed. We have demonstrated the underlying theories together with the principles of the different approaches namely the conventional glucose unit calculation, the virtual ladder approach and exoglycosidase mediated glycan sequencing. Thanks to the unique advantages of the capillary electrophoresis technique (high efficiency, excellent resolution, short separation times, very low sample consumption, and femtomolar range detection limit) the related instrumentation and sample preparation methods have been well developed. This development was tightly followed by the improvement of the necessary bioinformatics tools, which allows high throughput data evaluation with high confidence in the obtained results. It is envisioned that CE together with the automated database search assisted glycan structure identification will be readily accepted

by the glycomics community and will meet the growing challenges in glycoscience research.

Acknowledgment

The authors gratefully acknowledge the support from the National Research, Development and Innovation Office (BIONANO_GINOP-2.3.2-15-2016-00017; 2018-2.1.17-TÉT-KR-2018-00010) grants of the Hungarian Government. This work was supported by the TKP2020-IKA-07 project financed under the 2020-4.1.1-TKP2020 Thematic Excellence Programme by the National Research, Development and Innovation Fund of Hungary. This work was also supported by the UNKP-20-5 New National Excellence Program Hungarian Ministry of Human Capacities and the Janos Bolyai Research Scholarship of the Hungarian Academy of Sciences. This is contribution #179 of the Horváth Csaba Memorial Laboratory of Bioseparation Sciences.

References

[1] Z. El Rassi, Recent developments in capillary electrophoresis and capillary electrochromatography of carbohydrate species, Electrophoresis 20 (15–16) (1999) 3134–3144.
[2] A.T. Andrews, Electrophoresis, Theory, Techniques and Biochemical and Clinical Applications, second ed., Claredon Press Oxford, Oxford, England, 1986.
[3] G.G. Stokes, On the theories of the internal friction in motion, and of the equilibrium and motion of elastic solids, Trans. Camb. Phil. Soc. 8 (1845) 287–305.
[4] H. Eyring, M. Polanyi, Uber einfache gasreaktionen, Z. Phys. Chem. Abt. B 12 (1931) 279–311.
[5] H. Cottet, P. Gareil, On the use of the activation energy concept to investigate analyte and network deformations in entangled polymer solution capillary electrophoresis of synthetic polyelectrolytes, Electrophoresis 22 (4) (2001) 684–691.
[6] S.A. Arrhenius, Über die Dissociationswärme und den Einflusß der Temperatur auf den Dissociationsgrad der Elektrolyte, Z. Physik. Chem. 4 (1889) 96–116.
[7] A. Tiselius, A new apparatus for electrophoretic analysis of colloidal mixtures, Trans. Faraday Soc. 33 (1937) 524–531.
[8] K. Kakehi, S. Suzuki, 2.12 – Capillary electrophoresis of carbohydrates, in: H. Kamerling (Ed.), Comprehensive Glycoscience, Elsevier, Oxford, 2007, pp. 389–426.
[9] F.T.A. Chen, R.A. Evangelista, Analysis of mono- and oligosaccharide isomers derivatized with 9-aminopyrene-1,4,6-trisulfonate by capillary electrophoresis with laser-induced fluorescence, Anal. Biochem. 230 (2) (1995) 273–280.
[10] G. Lu, et al., Capillary electrophoresis separations of glycans, Chem. Rev. 118 (17) (2018) 7867–7885.
[11] A. Guttman, High-resolution carbohydrate profiling by capillary gel electrophoresis, Nature 380 (6573) (1996) 461–462.
[12] J. Szejtli, Helical and cyclic structures in starch chemistry, in: R. Friedman (Ed.), Biotechnology of Amylodextrin Oligosaccharides, American Chemical Society, Washington, DC, 1991, pp. 2–10.
[13] J. Szejtli, M. Richter, S. Augustat, Molecular configuration of amylose and its complexes in aqueous solutions. Part 11. Relation between the DP of helical segments of the amylose-iodine complex and the equilibrium concentration of free iodine, Biopolymers 5 (1967) 5–16.
[14] C. Chiesa, C. Horváth, Capillary zone electrophoresis of malto-oligosaccharides derivatized with 8-aminonaphthalene-1,3,6-trisulfonic acid, J. Chromatogr. 645 (2) (1993) 337–352.

[15] W. Nashabeh, Z. El Rassi, Capillary zone electrophoresis of linear and branched oligosaccharides, J. Chromatogr. 600 (2) (1992) 279–287.
[16] S. Mittermayr, A. Guttman, Influence of molecular configuration and conformation on the electromigration of oligosaccharides in narrow bore capillaries, Electrophoresis 33 (6) (2012) 1000–1007.
[17] G. Jarvas, M. Kerekgyarto, A. Guttman, On the electromigration of charged fluorophore-labeled oligosaccharides in polyethylene oxide solutions, Electrophoresis 37 (17–18) (2016) 2347–2351.
[18] M. Kerékgyártó, et al., Activation energy associated with the electromigration of oligosaccharides through viscosity modifier and polymeric additive containing background electrolytes, Electrophoresis 37 (4) (2016) 573–578.
[19] G. Jarvas, M. Szigeti, A. Guttman, Structural identification of N-linked carbohydrates using the GUcal application: a tutorial, J. Proteomics 171 (2018) 107–115.
[20] G. Jarvas, M. Szigeti, A. Guttman, Effect of the flow profile on separation efficiency in pressure-assisted reversed-polarity capillary zone electrophoresis of anions: simulation and experimental evaluation, J. Sep. Sci. 41 (11) (2018) 2473–2478.
[21] E. Kováts, Gas-chromatographische Charakterisierung organischer Verbindungen. Teil 1: Retentionsindices aliphatischer Halogenide, Alkohole, Aldehyde und Ketone, Helv. Chim. Acta 41 (7) (1958) 1915–1932.
[22] K. Yamashita, T. Mizuochi, A. Kobata, Analysis of oligosaccharides by gel filtration, Methods Enzymol. 83 (1982) 105–126.
[23] A. Guttman, F.-T.A. Chen, R.A. Evangelista, Separation of 1-aminopyrene-3,6,8-trisulfonate-labeled asparagine-linked fetuin glycans by capillary gel electrophoresis, Electrophoresis 17 (2) (1996) 412–417.
[24] A. Guttman, S. Herrick, Effect of the quantity and linkage position of mannose (alpha 1,2) residues in capillary gel electrophoresis of high-mannose-type oligosaccharides, Anal. Biochem. 235 (2) (1996) 236–239.
[25] K. Mariño, et al., A systematic approach to protein glycosylation analysis: a path through the maze, Nat. Chem. Biol. 6 (10) (2010) 713–723.
[26] G. Jarvas, M. Szigeti, A. Guttman, GUcal: an integrated application for capillary electrophoresis based glycan analysis, Electrophoresis 36 (24) (2015) 3094–3096.
[27] P. Bocek, et al., Analytical isotachophoresis, in: B.J. Radola (Ed.), Electrophoresis Library, Basel, Switzerland, Weinheim, 1988.
[28] G. Jarvas, et al., Triple-internal standard based glycan structural assignment method for capillary electrophoresis analysis of carbohydrates, Anal. Chem. 88 (23) (2016) 11364–11367.
[29] C. Varadi, C. Lew, A. Guttman, Rapid magnetic bead based sample preparation for automated and high throughput N-glycan analysis of therapeutic antibodies, Anal. Chem. 86 (12) (2014) 5682–5687.
[30] J.F. Rakus, L.K. Mahal, New technologies for glycomic analysis: toward a systematic understanding of the glycome, Annu. Rev. Anal. Chem. 4 (2011) 367–392.
[31] R. Raman, et al., Advancing glycomics: implementation strategies at the consortium for functional glycomics, Glycobiology 16 (5) (2006) 82r–90r.
[32] T. Lutteke, et al., GLYCOSCIENCES.de: an Internet portal to support glycomics and glycobiology research, Glycobiology 16 (5) (2006) 71r–81r.
[33] P. Toukach, et al., Sharing of worldwide distributed carbohydrate-related digital resources: online connection of the bacterial carbohydrate structure DataBase and GLYCOSCIENCES.de, Nucleic Acids Res. 35 (Database issue) (2007) D280–D286.
[34] P.V. Toukach, K.S. Egorova, Carbohydrate structure database merged from bacterial, archaeal, plant and fungal parts, Nucleic Acids Res. 44 (D1) (2016) D1229–D1236.
[35] C.A. Cooper, et al., Data standardisation in GlycoSuiteDB, in: Pacific Symposium on Biocomputing, 2002, pp. 297–309.

[36] C.W. von der Lieth, et al., EUROCarbDB: an open-access platform for glycoinformatics, Glycobiology 21 (4) (2011) 493–502.
[37] C.A. Hayes, et al., UniCarb-DB: a database resource for glycomic discovery, Bioinformatics 27 (9) (2011) 1343–1344.
[38] M.P. Campbell, L. Royle, P.M. Rudd, GlycoBase and autoGU: resources for interpreting HPLC-glycan data, Methods Mol. Biol. 1273 (2015) 17–28.
[39] J.L. Abrahams, M.P. Campbell, N.H. Packer, Building a PGC-LC-MS N-glycan retention library and elution mapping resource, Glycoconj. J. 35 (1) (2018) 15–29.
[40] G. Jarvas, et al., Expanding the capillary electrophoresis-based glucose unit database of the GUcal app, Glycobiology 30 (6) (2020) 362–364.
[41] M.P. Campbell, et al., GlycoBase and autoGU: tools for HPLC-based glycan analysis, Bioinformatics 24 (9) (2008) 1214–1216.
[42] J.L. Abrahams, et al., Recent advances in glycoinformatic platforms for glycomics and glycoproteomics, Curr. Opin. Struct. Biol. 62 (2020) 56–69.
[43] A. Behne, et al., glyXalign: high-throughput migration time alignment preprocessing of electrophoretic data retrieved via multiplexed capillary gel electrophoresis with laser-induced fluorescence detection-based glycoprofiling, Electrophoresis 34 (16) (2013) 2311–2315.
[44] W.S. York, et al., GlyGen: computational and informatics resources for glycoscience, Glycobiology 30 (2) (2020) 72–73.
[45] I. Yamada, et al., The GlyCosmos Portal: a unified and comprehensive web resource for the glycosciences, Nat. Methods 17 (7) (2020) 649–650.
[46] L. Zhang, S. Luo, B. Zhang, Glycan analysis of therapeutic glycoproteins, MAbs 8 (2) (2016) 205–215.
[47] A. Guttman, Multistructure sequencing of N-linked fetuin glycans by capillary gel electrophoresis and enzyme matrix digestion, Electrophoresis 18 (7) (1997) 1136–1141.
[48] A. Guttman, K.W. Ulfelder, Exoglycosidase matrix-mediated sequencing of a complex glycan pool by capillary electrophoresis, J. Chromatogr. A 781 (1) (1997) 547–554.
[49] M. Szigeti, A. Guttman, Automated N-glycosylation sequencing of biopharmaceuticals by capillary electrophoresis, Sci. Rep. 7 (1) (2017) 11663.