Monitoring antigenic protein integrity during glycoconjugate vaccine synthesis using capillary electrophoresis-mass spectrometry

Sara Tengattini[1,2], Elena Domínguez-Vega[2], Caterina Temporini[1], Marco Terreni[1], Govert W. Somsen[2]

[1]*Department of Drug Sciences, University of Pavia, via Taramelli 12, I-27100 Pavia, Italy*
[2]*Division of BioAnalytical Chemistry, Department of Chemistry and Pharmaceutical Sciences, Vrije Universiteit Amsterdam, de Boelelaan 1085, 1081 HV Amsterdam, The Netherlands*

Corresponding author:
Elena Domínguez-Vega
E-mail: e.dominguezvega@vu.nl
Telephone: +31 (0)20 59 89182

Abstract
A capillary electrophoresis-mass spectrometry (CE-MS) method was developed for the characterization and integrity assessment of the *Mycobacterium tuberculosis* (MTB) antigens TB10.4 and Ag85B and their chemically produced glycoconjugates, which are glycovaccine candidates against tuberculosis (TB). In order to prevent protein adsorption to the inner capillary wall and to achieve efficient separation of the antigen proteoforms, a polyionic multilayer coating of polybrene-dextran sulfate-polybrene (PB-DS-PB) was used in combination with 1.5 M acetic acid as background electrolyte (BGE). Coupling of CE to high-resolution time-of-flight MS was achieved by a coaxial interface employing a sheath liquid of isopropanol-water (50:50, v/v) containing 0.1% formic acid. The MTB antigens were exposed to experimental conditions used for chemical glycosylation (but no activated saccharide was added) in order to investigate their stability during glycovaccine production. CE-MS analysis revealed the presence of several closely-related degradation products, including truncated, oxidised and conformational variants, which were assigned by accurate mass. Analysis of synthesized mannose conjugates of TB10.4 and Ag85B allowed the determination of the glycoform composition of the *neo*-glycoproteins next to the characterization of degradation products which were shown to be partly glycoconjugated. Moreover, the selectivity of CE-MS allowed specific detection of deamidated species (protein mass change of 1.0 Da only), indicating that chemical glycosylation increased susceptibility to deamidation. Overall, the results show that CE-MS represents a useful analytical tool for the detailed characterization and optimization of *neo*-glyconjugate products.

Keywords
CE-MS; intact protein analysis; protein characterization; neo-glycoproteins; antigenic proteins; glycoconjugate vaccines

1. Introduction

Prevention of infectious diseases by vaccination is considered one of the most successful health treatments. Glycoconjugate vaccines, in which oligosaccharides are conjugated to carrier proteins (*neo*-glycoproteins), are among the safest and most efficacious vaccines available [1-3]. Since the introduction of the first glycoconjugate vaccine against *Haemophilus influenzae type b* (Hib) in 1987, this strategy has been successfully applied against numerous bacterial-infectious diseases [4].

Tuberculosis (TB) still remains one of the world's deadliest human diseases, ranked second only after immune deficiency syndrome (AIDS) in the number of people killed every year. In 2013, about 9.0 million people developed TB and 1.5 million died [5, 6]. Even though TB is curable, the efficacy of available chemotherapies is limited due to emerging multidrug-resistant (MDR) and extensively drug-resistant (XDR) strains of *Mycobacterium tuberculosis* (MTB) [5]. Therapeutic vaccination against MTB has potential for the treatment of MDR- and XDR-TB [5]. At present, *Bacillus Calmette-Guérin* (BCG), the only TB vaccine in use, has major limitations: BCG protects only infants and [7] its efficacy wanes significantly over a period of 10-15 years [8]. Thus, the design of an effective vaccine against TB remains an international research priority [6]. In this context, conjugation of antigenic proteins from MTB with arabino-mannan polysaccharides was considered for the development of highly immunogenic glycoconjugate vaccines [9].

Glycoconjugate vaccines are produced by chemical binding of oligosaccharides to carrier proteins forming *neo*-glycoproteins [10]. For this purpose, glycans are activated targeting nucleophilic groups of amino acid residues (lysines, aspartic/glutamic acids or cysteines) of antigen proteins [11]. One of the most diffused and consolidated method implies the use of 2-iminomethoxyethyl thioglycosides (IME) that selectively react with ε-amino groups of lysine residues [12]. These glycosylation strategies, however, commonly lead to a mixture of *neo*-glycoproteins with different saccharide loading number and/or saccharide positioning, which may have different properties such as clearance kinetics and/or antigenicity [11]. Moreover, isolated or recombinantly produced carrier proteins, such as antigenic proteins, may not be entirely pure and contain degradation products. In addition, these proteins may undergo unwanted chemical modifications during the glycovaccine production process. Characterization of heterogeneous *neo*-glycoprotein products evidently is a fundamental, but challenging, task requiring not only determination of the generated glycoforms, but also characterization of potential degradation products formed during the conjugation process in order to ensure the production of pure and stable products during the glycovaccine synthesis.

Mass spectrometry (MS)-based approaches are widely used for the structural characterization of novel glycoconjugates [13, 14]. Bottom-up approaches in which proteins are first digested into fragments, i.e. (glyco)peptides, are mainly used for the localization of glycosylation sites of novel (*neo*-)glycoproteins [15, 16]. We have shown the usefulness of bottom-up liquid chromatography (LC)-MS to characterize the glycosylation reactivity of surface amino acids of antigens using different linkers (e.g. IME or homobifunctional (4-nitrophenyl ester)). This allowed selection of the optimal activation and conjugation conditions preserving the protein antigenic epitope after glycosylation [17]. Still, quality control of the intact *neo*-glycoprotein is an important issue in biopharmaceutical analysis as it provides essential information on (glycoform) composition and protein integrity which cannot be revealed by bottom-up approaches only. Moreover intact analysis is relatively fast, requires minimum sample treatment, and avoids unwanted modifications induced by enzymatic treatments [18, 19]. We have applied direct infusion MS methods for the characterization of the glycoform composition of intact novel glycovaccines [17]. However, direct MS characterization of intact *neo*-glycoproteins can be quite challenging due to their complex (micro)heterogeneity. In addition to the inevitable heterogeneity related to the carbohydrate component (number, structure and position), one also has to consider the potential antigen heterogeneity caused by the presence of *e.g.* proteoforms and degradation products. In particular, distinguishing closely-related variants, such deamidated forms or conformers, might be no possible with standalone MS. Therefore, separation of intact *neo*-glycoprotein variants prior to MS detection is needed in order to reduce the complexity of mass spectra and will facilitate the characterization of all present species, including minor components. Conventional reversed-phase liquid chromatography (RPLC) of intact glycoproteins may be troublesome due to adverse interactions with the hydrophobic stationary phase, leading to poor performance. Moreover, RPLC often lacks the selectivity and efficiency to assess subtle protein modifications and proteoforms. Recently, HILIC has shown promising possibilities for resolving intact glycoforms including neo-glycoproteins [20-22], but coupling of protein HILIC with MS still has to be demonstrated. Capillary electrophoresis (CE)-MS has proved to be a useful and powerful analytical tool for the characterization of intact proteins, glycoproteins and drug-protein conjugates, combining high

protein separation efficiency with mass-selective detection [23-26]. Unwanted interactions between proteins and negatively-charged silanol groups on the inner wall of fused-silica capillaries can be prevented by applying MS-compatible coatings, allowing efficient separation of proteins in a wide range of conditions [26]. CE has shown particularly useful in revealing charge heterogeneity among protein degradation products and impurities, but also has shown the ability to separate positional isomers and conformers [25, 27, 28].

In this work, we studied the potential of CE-MS for the characterization of the antigenic MTB proteins TB10.4 (11 kDa) and Ag85B (31 kDa) which were conjugated with 2-iminomethoxyethyl-mannose (Man-IME) and 2-minomethoxyethyl-mannose(1-6)mannose (Man(1-6)Man-IME). These *neo*-glycoconjugates are candidate vaccines against TB [17]. In order to investigate the stability of the MTB antigens during glycovaccine production, the proteins were analysed by the developed CE-MS method before and after exposure to experimental conditions used for chemical glycosylation, but with no activated saccharide added. Next, the *neo*-glycoproteins were analysed by CE-MS in order to determine glycoform composition as well as to monitor the presence of glycosylation-related degradation products.

2. Materials and methods

2.1. Reagents and chemicals

All chemicals were of analytical grade. Ammonium hydroxide (30% solution), acetic acid, formic acid, isopropanol (IPA), acetonitrile (ACN), polybrene (hexadimethrine bromide, PB; average molecular weight (MW) 15,000), dextran sulfate (DS, average MW>500,000), benzamidine hydrochloride, sodium tetraborate and phosphate-buffered saline (PBS) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Deionized water was obtained from a Milli-Q purification system (Millipore, Bedford, MA, USA). IME thioglycosides (Man-IME and Man(1-6)Man-IME) were prepared according to a previously reported procedure [29]. The TB10.4 and Ag85B immunogenic proteins were obtained as recombinant forms in *Escherichia coli* as reported by Piubelli *et al.* [8] and finally collected in 20 mM 3-(N-morpholino)propanesulfonic acid (MOPS), 0.4 M NaCl, pH 7.0 at different concentrations.

2.2. Protein glycosylation

According to the protocol previously reported [17], the glycosylation reaction was carried out in sodium tetraborate buffer, 100 mM, pH 9.5. TB10.4 and Ag85B were dissolved in the buffer in a concentration of 5.5 mg/mL and the solution was subsequently mixed with IME-glycoside in a glycoside/protein molar ratio of 200/1. Benzamidine chloride was added to the reaction mixture in order to reduce protein digestion by residual enterokinase derived from the antigen protein production process. The reaction mixture was vortexed for 1 min and incubated for 24 h at 25 or 37 °C under continuous stirring.

2.3 Sample preparation

Neo-glycoconjugate samples and antigenic proteins exposed to glycosylation conditions for stability studies were purified in order to remove reagents and/or salts. For this, the solutions were submitted to four 20 min steps of ultrafiltration at 13000 g and 4 °C using Amicon® Ultra filters (Millipore, Billerica, MA, USA) with nominal molecular weight limit (NMWL) of 3 or 10 kDa, and a load capacity of 500 µL. Proteins and *neo*-glycoproteins were finally collected and stored in PBS. Prior to CE-MS analysis, the buffer of proteins and their glycoconjugates were exchanged for water by two 20-min steps of ultrafiltration at 13000 g and 4 °C. All the samples were diluted with water to a final concentration of 1 mg/mL.

2.4. Capillary electrophoresis

CE analyses were performed using a Beckman PA 800 plus instrument (Beckman Coulter, Brea, CA, USA.). Bare fused-silica capillary with an internal diameter of 50 µm were obtained from Polymicro Technologies (Phoenix, AZ, U.S.A.) and cut to a total length of 90 cm. Hydrodynamic sample injections were performed at 1 psi for 12 s corresponding to an injection volume of 12.5 nL and an injected protein amount of 12.5 ng. UV detection was performed at 214 nm.

For the optimized CE-MS method the separation voltage was -25 kV, the capillary temperature 15 °C, and the background electrolyte (BGE) 1.5 M acetic acid. For the PB-DS-PB-coating, new bare fused-silica capillaries were firstly rinsed with successively 1 M NaOH and water (both 30 min at 20 psi). Subsequently the capillaries were flushed with 10% (w/v) PB solution (60 min at 5 psi), deionized water (30 min at 10 psi), 0.5% (w/v) DS solution (60 min at 5 psi), deionized water (30 min at 10 psi), 10% (w/v) PB solution (50 min

at 5 psi), deionized water (30 min at 10 psi), and BGE (10 min at 20 psi). Overnight, coated capillaries were filled with water and tips were immersed in vials with water.

2.5 Mass spectrometry

MS detection was performed using a maXis HD ultra-high resolution quadrupole-time-of-flight (QTOF) mass spectrometer (Rs, *80,000*) (Bruker Daltonics, Bremen, Germany) equipped with an electrospray ionisation (ESI) source. CE-MS coupling was performed using a sheath-liquid electrospray interface from Agilent Technologies. The sheath liquid was a mixture of IPA-water (50:50, v/v) containing 0.1% of formic acid and was delivered at a flow rate of 3 μL/min by a syringe pump from Cole-Parmer (Vernon Hill, IL, USA). The mass spectrometer was operated in positive-ion mode with an electrospray voltage of 4.5 kV. The nebulizer gas (nitrogen) pressure was 5 psi and the dry gas flow rate and temperature was 4 L/min nitrogen and 200 °C, respectively. Quadrupole ion and collision cell energies were 5 and 8 eV, respectively. The monitored mass range was 250-4000 *m/z*. Data were analyzed using Bruker Daltonics Data Analysis software (Compass DataAnalysis, version 3.2). Extracted-ion electropherograms (EIEs) were obtained with an extraction window of ±0.1 *m/z* and using the smooth option of the software (Gaussian at 1 point). Mass determinations of proteins were performed using the "*Maximum entropy deconvolution*" utility. Mass accuracy employed for assignments was <20 ppm. Theoretical masses were calculated from the amino acid sequence using the "*Peptide mass calculator*" on IonSourceMS (www.ionsource.com).

3 Results and discussion

3.1 CE-MS method development

A BGE of acetic acid was selected for the CE-MS analysis of TB10.4 and Ag85B, as this low-pH BGEs has previously shown good separation and ESI performance for intact proteins [23]. TB10.4 and Ag85B have an isoelectric point (pI) of 4.44 and 4.77, and therefore will be positively charged when using an acidic BGE. In order to minimize adsorption of the proteins to the inner wall of the capillary, a MS-compatible non-covalent multilayer capillary coating, consisting of PB-DS-PB, was employed. Using 200 mM acetic acid as BGE a single symmetrical peak for TB10.4 and two partially resolved peaks for Ag85B were observed in CE-UV. Increase of the acetic acid concentration to 2 M (pH 2.2) resulted in an improved resolution of the Ag85B. When using these conditions for the analysis of the Ag85B-Man sample, two additional small peaks migrating before the main peak were observed (Fig. S1 in Electronic Supplementary Material (ESM)). The influence of voltage and temperature on separation was investigated. Voltage variation between -25 and -15 kV did not provide a significant change in the separation, whereas decreasing the temperature from 25 to 15 °C allowed to improve considerably the resolution of peaks observed for the Ag85B and Ag85B-Man samples. The addition of ACN (up to 20% v/v) to the BGE was considered, but no improvement in separation was obtained. Based on the CE-UV experiments, the finally selected CE conditions comprised a PB-DS-PB coated capillary in combination with 2 M acetic acid as BGE, a capillary temperature of 15 °C, and a separation voltage of -25 kV. For coupling CE to MS, a coaxial sheath-liquid interface was employed. A sheath liquid consisting of IPA-water (50:50, v/v) containing 0.1% of formic acid was selected as it provided good sensitivity and current stability. However, the BGE of 2 M acetic acid resulted in considerable ionization suppression of the proteins. As a compromise between CE resolution and MS sensitivity, a BGE of 1.5 M acetic acid (pH 2.3) was finally selected.

3.2 Characterization and stability evaluation of TB10.4 and Ag85B

In order to investigate the stability of TB10.4 and Ag85B during glycoconjugate synthesis, the protein antigens were analysed by CE-MS before and after exposure for 24 h to glycosylation reaction conditions (see Section 2.2), but with no activated saccharides added. TB10.4 exposure was performed at 37 °C, being the optimum glycosylation temperature [17]. For Ag85B, glycosylation yields were 100% at both 25 and 37 °C and, therefore, both temperatures were employed in the exposure study.

3.2.1 TB10.4

CE-MS analysis of TB10.4 in water showed a single symmetric peak at a migration time of approximately 20 min. Deconvolution of the mass spectrum obtained in the apex of the peak revealed a single protein molecular mass of 11076.1 Da, which agrees well with the expected molecular mass for TB10.4 (11076.3 Da; see Fig. S2 for protein sequence and Fig. S3 for deconvoluted spectrum in ESM). CE-MS analysis of the

TB10.4 sample after exposure to glycosylation conditions (37 °C for 24 h) showed the presence of several degradation products (Fig. 1) which could be assigned (except peak 1) based on the masses derived from the deconvoluted spectra (Table 1). Peaks 2 and 3 were identified as composed of truncated protein forms that lost 4 to 17 amino acids from the C-terminus side: A1-N86, A1-T87, A1-M88, A1-A89, A1-M90, A1-M91, A1-A92, A1-E97 and A1-A99. The shorter migration times of these compounds with respect to TB10.4 most probably can be explained by the loss of two (R93, K100) or one (K100) basic amino acids, causing a decrease of net positive charge. The peaks 5 and 6 were assigned to truncated forms that lost amino acids from the N-terminal side: M18-G103, N14-G103, Y13-G103 and S48-G103. In this case, the longer migration times compared to native TB10.4 might be the consequence of the loss of one (D5) or four (D5, D24, E38, E42) acidic amino acids. As indicated previously [8], these truncated species might result from aspecific protein digestion due to residual enterokinase activity.

The base peak electropherogram (BPE) also shows a component at 20.2 min (Fig. 1, peak 4), which is not well resolved from TB10.4. The mass observed for this component is 16 Da higher than the mass of TB10.4, suggesting an oxidation product. Oxidation induces a relatively small change in protein mass (+16 Da), but no change in the protein charge, resulting in a small decrease in migration time only.

3.2.2 Ag85B

CE-MS analysis of Ag85B in water showed the presence of four different species (Fig. 2a). The mass spectrum obtained for the main peak at a approximately 22.5 min (peak 0), revealed, after deconvolution, a protein molecular mass of 31345.7 Da, corresponding to the expected molecular weight of Ag85B (31345.6 Da; see Fig. S4 for protein sequence and Fig. S5 for deconvoluted spectrum in ESM). The deconvoluted mass spectrum of the first component (peak 1) migrating after the main peak, indicated the same mass as Ag85B, but the recorded mass spectra showed a clearly different charge state distribution (*cf*. Fig. $2a_1$ and $2a_2$). The mass spectrum of the main peak (Fig. $2a_1$) seems to comprise two charge state distributions, which might be explained by partial unfolding of the protein during ESI, leading to increased charging and a shift of the distribution to lower *m/z* values. The mass spectrum corresponding to peak 1 (Fig. $2a_2$) shows a significantly lower relative abundance of the charge states at lower *m/z* values, indicating less susceptibility to unfolding. Peak 1 might be caused by a conformer of Ag85B exhibiting a different electrophoretic mobility and ESI.

Deconvolution of the mass spectrum obtained for peak 2 revealed a molecular weight of 30424.3 Da, which was assigned to the truncated form R10-G292. The mass spectrum of low abundant peak 3 indicates the same molecular mass as peak 2. However, the mass spectra of peaks 2 and 3 differ in their charge state distributions in a similar fashion as with peaks 0 and 1, again suggesting the presence of two protein conformers.

The Ag85B sample was exposed for 24 h to glycosylation conditions at both 25 and 37 °C. CE-MS analysis (Fig. 2b and 2c) showed the appearance of two additional degradation products (peaks 4 and 5), while the abundance of peaks 2 and 3 increased with respect to non-exposed Ag85B. Table 2 summarizes the peak assignment. The considerable increase of peak 2 is mainly caused by the formation of a new degradation product identified as the truncated form S9-G292, which co-migrates with R10-G292. Moreover, the abundance of R10-G292 relative to Ag85B slightly increased from 8.6% in the non-exposed sample to 11.1% in the sample incubated at 37 °C. The compound migrating as peak 5 at 21.9 min differs from Ag85B by +0.9 Da in mass only. Most probably, this modification relates to a single deamidation of the protein. Deamidation is a common protein degradation that results in a small change of protein molecular mass (+0.984 Da) and a shift of the pI due to the transformation of an amide residue into an acidic residue. The relative abundance of the degradation products was significantly lower at 25 °C. Therefore, this temperature was selected for Ag85B glycoconjugate synthesis.

3.3 Characterization of TB10.4 and Ag85B glycoconjugates

TB10.4 and Ag85B were conjugated with Man and Man(1-6)Man activated as IME-glycosides according to the protocol previously reported (see materials and methods section and Fig. S6 in ESM for the reaction scheme). The resulting glycoconjugates were analysed with the optimized CE-MS method.

3.3.1 TB10.4 glycoconjugates

CE-MS analysis of TB10.4 conjugated with Man-IME and Man(1-6)Man-IME showed the presence of several components. The BPE obtained for TB10.4 conjugated with Man(1-6)Man-IME is shown in Fig. 3a; the mass spectra of the two main peaks are given in Fig. S7 and Fig. S8 in ESM. A number of additional

peaks was observed in comparison to the unconjugated TB10.4 sample which was exposed to glycosylation conditions, but with no activated saccharide added (*cf.* Fig. 1). These peaks resulted mostly from the glycoconjugation of TB10.4. The main glycoconjugate species detected for the two samples had masses corresponding to TB10.4-Man and TB10.4-(Man)$_2$, and TB10.4-Man(1-6)Man and TB10.4-[Man(1-6)Man]$_2$ respectively. TB10.4 comprises one lysine residue as potential glycosylation site, but as demonstrated in our previous work [17], under the applied glycosylation conditions, the amino functionality of the N-terminus of the protein also can be conjugated, leading to the formation of diglycosylated antigen. IME conjugation does not induce a change in the protein charge (see Fig. S2 in ESM), resulting in a partial CE separation of the unmodified protein and the somewhat larger glycoforms. For TB10.4 conjugated with Man(1-6)Man-IME, a better glycoconjugate resolution was reached due to the larger increase in antigen size (+397 Da respect to +235 Da of Man-IME) upon conjugation (Fig. 3b).

The truncated species identified in the antigen stability studies were also detected in the conjugated samples. The truncated forms that lost a lysine residue (peaks 2 and 3 in Table 1) with respect to native TB10.4, were detected both unmodified and glycosylated (at the N-terminal residue). The degradation products that still contained the lysine residue were mainly observed as glycosylated forms (Table 1).

Interestingly, samples of TB10.4 conjugated with Man-IME as well as with Man(1-6)Man-IME showed two additional degradation products which were not observed during the stability studies (Section 3.2). These two compounds appeared to have molecular weights differing only 1 Da from the mono- and diglycosylated TB10.4, respectively, and were detected at shorter migration times (Fig. 3c). Based on the molecular mass difference and migration time, we suppose the degradation products to be deamidated forms of the glycoconjugates. In order to test this hypothesis, the deconvoluted mass spectra of TB10.4-[Man(1-6)Man]$_2$ and deamidated TB10.4-[Man(1-6)Man]$_2$ were *in-silico* simulated (Fig. 4a) and compared with the experimentally obtained deconvoluted mass spectra (Fig. 4b). The striking correspondence between the simulated and the experimental data, both showing 1 Da shift of the isotopic distribution for the deamidated form in comparison with the non-deamidated one, strongly supports our assumption that deamidation occurred. As can be seen in Fig. 3, no deamidated form was detected for the unconjugated form of TB10.4. This result is in agreement with the stability study (Section 3.2.1) where no deamidated species were observed for TB10.4. Assuming equimolar ESI efficiencies for the deamidated and nondeamidated glycoconjugates, deamidation ratios were estimated (Table 3). Higher rates of deamidation were obtained for the diglycosylated species (7.1-7.3%) in comparison with the monoglicosylated ones (0.5-0.6%). All together, these semi-quantitative results indicate a correlation between deamidation incidence and degree of glycosylation.

3.3.2 Characterization of Ag85B glycoconjugates

Ag85B has 8 lysine residues in the amino acid sequence as potential glycosylation sites. CE-MS analysis of Ag85B conjugated with Man-IME or with Man(1-6)Man-IME revealed the presence of six glycoforms in both samples. Based on the molecular weights indicated by the deconvoluted mass spectra (see Fig. S9 and S10 in ESM), these Ag85B glycoconjugates appeared to containing 3 to 8 saccharide units for the Man sample and 2 to 7 for the Man(1-6)Man sample. Summed extraction of the three most abundant ions for each glycoform generated the EIEs reported in Fig. 5. The most glycosylated species migrated faster. As with TB10.4, higher resolution was obtained for the Man(1-6)Man glycoforms. The broad and non-Gaussian shape of the observed peaks, as appeared more evident for the Man(1-6)Man sample, is, most probably related to the overlap of different *neo*-glycoprotein isomers characterized by the same saccharide loading number but different location in the amino acid sequence. Higher conjugation efficiency was observed for the Man with an average of 5.9 saccharide units per protein whereas only an average of 4.6 was obtained in the case of Man(1-6)Man.

Regarding antigen degradation (see section 3.2.2), only the most abundant degradation product (S9-G292; peak 2 in Fig. 2 and Table 2), was detected with low intensity as a mixture of glycoconjugates (Table 2 for Ag85B conjugated with Man-IME). EIE traces of Ag85B conjugated with Man-IME and Man(1-6)Man-IME (Fig. 5a and 5b) show the presence of a cluster of broadened peaks at lower migration times comprising glycoform ions with similar *m/z* values (± 0.5). These peaks most probably correspond to a number of glycoforms differing 1 Da, indicating the occurrence of deamidation. Different peaks were observed in each trace suggesting the occurrence of multiple deamidations in the same protein. Due to the broadened peak shapes and low intensity it was not possible to estimate the percentage of deamidation of each glycoform. Still, the EIEs in Fig. 5 indicate that glycoforms with a higher number of conjugated saccharides show a greater susceptibility for deamidation, as was also observed for TB10.4.

4. Conclusion

The potential of CE-MS for the characterization of intact *neo*-glycoproteins prepared from two MTB antigens was demonstrated. The use of a non-covalently coated (PB-DS-PB) capillary in combination with an acidic BGE allowed efficient separation of different antigen proteoforms and degradation products. The coupling with high resolution TOF-MS permitted mass assignment for most of the variants observed. The method was suitable for the characterization of protein antigens allowing to establish the occurrence of modifications at different stages of the glycovaccine production. For example, CE-MS revealed the presence of conformers in Ag85B. The method also permits evaluation of antigen integrity and stability under glycosylation conditions and showed multiple degradation products resulting from e.g. deamidation and truncation. For *neo*-glycoprotein samples, the efficiency of glycosylation, number of glycoforms and saccharide units can be assessed, also showing conjugation of degradation products. In addition, glycoform specific deamidation was observed suggesting a relation between the number of saccharide units and the rate of protein deamidation.

The developed CE-TOF-MS method was found suitable for the assessment of the detailed composition of *neo*-glycoprotein samples. As such this method represents a valuable tool for obtaining well-characterized products that will further drive the optimization of the *neo*-glycoprotein production process.

Conflict of interest

The authors have declared no conflict of interest.

References

[1] Astronomo RD, Burton DR. Carbohydrate vaccines: Developing sweet solutions to sticky situations? Nat Rev Drug Discov. 2010;9:308–324.

[2] Boltje TJ, Buskas T, Boons GJ. Opportunities and challenges in synthetic oligosaccaride and glycoconjugate research. Nat Chem. 2009;1:611–622.

[3] Finco O, Rappuoli R. Designing vaccines for the twenty-first century society. Front Immunol. 2014;5:1–6.

[4] Hütter J, Lepenies B. Carbohydrate-Based Vaccines: An Overview. Methods Mol Biol. 2015;1331:1–10.

[5] Bourzac K. Infectious disease: Beating the big three. Nature. 2014;507:S4–7.

[6] Global tuberculosis report. World Health Organization (WHO). 2015. http://www.who.int/tb/publications/global_report/en/

[7] Andersen P, Kaufmann SH. Novel vaccination strategies against tuberculosis. Cold Spring Harb Perspect Med. 2014;26:533–542.

[8] Piubelli L, Campa M, Temporini C, Binda E, Mangione F, Amicosante M, Terreni M, Marinelli F, Pollegioni L. Optimizing Escherichia coli as a protein expression platform to produce *Mycobacterium tuberculosis* immunogenic proteins. Microb Cell Fact. 2013;12:115.

[9] Kallenius G, Pawlowski A, Hamasur B, Svenson SB. Mycobacterial glycoconjugates as vaccine candidates against tuberculosis. Trends Microbiol. 2008;16:456–462.

[10] Grayson EJ, Bernardes GJL, Chalker JM, Boutureira O, Koeppe JR, Davis BG. A coordinated synthesis and conjugation strategy for the preparation of homogeneous glycoconjugate vaccine candidates. Angew Chem Int Ed. 2011;50:4127–4132.

[11] Adamo R, Nilo A, Castagner B, Boutureira O, Berti F, Bernardes GJL. Synthetically defined glycoprotein vaccines: Current status and future directions. Chem Sci. 2013;4:2995–3008.

[12] Gamblin DP, Scanlan EM, Davis BG. Glycoprotein synthesis: an update. Chem Rev. 2009;109:131–163.

[13] Staub A, Guillarme D, Schappler J, Veuthey JL, Rudaz S. Intact protein analysis in the biopharmaceutical field. J Pharm Biomed Anal. 2011;55:810–822.

[14] Novotny MV, Mechref Y. New hyphenated methodologies in high-sensitivity glycoprotein analysis. J Sep Sci. 2005;28:1956–1968.

[15] Mariño K, Bones J, Kattla JJ, Rudd PM. A systematic approach to protein glycosylation analysis: a path through the maze. Nat Chem Biol. 2010;6:713–723.

[16] Thaysen-Andersen M, Packer NH. Advances in LC–MS/MS-based glycoproteomics: Getting closer to system-wide site-specific mapping of the *N*- and *O*-glycoproteome. Biochim Biophys Acta. 2014;1844: 437–1452.

[17] Temporini C, Bavaro T, Tengattini S, Serra I, Marrubini G, Calleri E, Fasanella F, Piubelli L, Marinelli F, Pollegioni L, Speranza G, Massolini G, Terreni M. Liquid chromatography–mass spectrometry structural characterization of *neo* glycoproteins aiding the rational design and synthesis of a novel glycovaccine for protection against tuberculosis. J. Chromatogr. A. 2014;1365:57–67.

[18] Balaguer E, Neususs C. Glycoprotein characterization combining intact protein and glycan analysis by capillary electrophoresis-electrospray ionization-mass spectrometry. Anal Chem. 2006;78:5384–5393.

[19] Sandra K, Vandenheede I, Sandra P. Modern chromatographic and mass spectrometric techniques for protein biopharmaceutical characterization. J Chromatogr A. 2014;1335:81–103.

[20] Tetaz T, Detzner S, Friedlein A, Molitor B, Mary JL, J. Chromatogr. A. 2014; 1218: 5892–5896.

[21] Pedrali A, Tengattini S, Marrubini G, Bavaro T, Hemstrom P, Massolini G, Terreni M, Temporini C, Molecules 2014; 19: 9070–9088.

[22] Zhang Z, Wu Z, Wirth M J, J. Chromatogr. A 2013; 1301:156–161.

[23] Haselberg R, de Jong GJ, Somsen GW. CE-MS for the analysis of intact proteins 2010–2012. Electrophoresis. 2013;34:99–112.

[24] Biacchi M, Gahoual R, Said N, Beck A, Leize-Wagner E, Francois YN. Glycoform separation and characterization of cetuximab variants by middle-up off-line capillary zone electrophoresis-UV/electrospray ionization-MS. Anal Chem. 2015;87:6240–6250.

[25] Haselberg R, de Jong GJ, Somsen GW. Low-Flow Sheathless Capillary Electrophoresis–Mass Spectrometry for Sensitive Glycoform Profiling of Intact Pharmaceutical Proteins. Anal Chem. 2013;85:2289−2296.

[26] Zhao SS, Chen DDY. Applications of capillary electrophoresis in characterizing recombinant protein therapeutics. Electrophoresis. 2014;35:96–108.

[27] Bush DR, Zang L, Belov AM, Ivanov AR, Karger BL. High Resolution CZE-MS Quantitative Characterization of Intact Biopharmaceutical Proteins: Proteoforms of Interferon-β1. Anal Chem. 2016;88:1138–1146.

[28] Bertoletti L, Bisceglia F, Colombo R, Giorgetti S, Raimondi S, Mangione PP, De Lorenzi E. Capillary electrophoresis analysis of different variants of the amyloidogenic protein β2 -microglobulin as a simple tool for misfolding and stability studies. Electrophoresis. 2015;36:2465–2472.

[29] Bavaro T, Filice M, Temporini C, Tengattini S, Serra I, Morelli CF, Massolini G, Terreni M. Chemoenzymatic synthesis of neoglycoproteins driven by the assessment of protein surface reactivity. RSC Adv. 2014;4:56455–56465.