

LC-MS analysis combined with principal component analysis and soft independent modelling by class analogy for a better detection of changes in N-glycosylation profiles of therapeutic glycoproteins

Ana Planinc¹ · Bieke Dejaegher^{2,3} · Yvan Vander Heyden³ · Johan Viaene³ · Serge Van Praet⁴ · Florence Rappé⁴ · Pierre Van Antwerpen¹ · Cédric Delporte¹

Received: 15 April 2016 / Revised: 25 May 2016 / Accepted: 1 June 2016 / Published online: 10 June 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract Therapeutic proteins are among the top selling drugs in the pharmaceutical industry. More than 60 % of the approved therapeutic proteins are glycosylated. Nowadays, it is well accepted that changes in glycosylation may affect the safety and the efficacy of the therapeutic proteins. For this reason, it is important to char-

acterize both the protein and the glycan structures. In this study, analytical and data processing methods were developed ensuring an easier characterization of glycoprofiles. N-glycans were (i) enzymatically released using peptide-N-glycosidase F (PNGase F), (ii) reduced, and (iii) analyzed by hydrophilic interaction liquid chromatography coupled to a high-resolution mass spectrometer (HILIC-HRMS). Glycosylation changes were analyzed in human plasma immunoglobulin G samples which had previously been artificially modified by adding other glycoproteins (such as ribonuclease B and fetuin) or by digesting with enzyme (neuraminidase). Principal component analysis (PCA) and classification through soft independent modelling by class analogy (SIMCA) were used to detect minor glycosylation changes. Using HILIC-MS-PCA/SIMCA approach, it was possible to detect small changes in N-glycosylation, which had not been detected directly from the extracted-ion chromatograms, which is current technique to detect N-glycosylation changes in batch-to-batch analysis. The HILIC-MS-PCA/SIMCA approach is highly sensitive approach due to the sensitivity of MS and appropriate data processing approaches. It could help in assessing the changes in glycosylation, controlling batch-to-batch consistency, and establishing acceptance limits according to the glycosylation changes, ensuring safety and efficacy.

Published in the topical collection *Glycomics, Glycoproteomics and Allied Topics* with guest editors Yehia Mechref and David Muddiman.

Ana Planinc, Bieke Dejaegher, Yvan Vander Heyden, Johan Viaene, Pierre Van Antwerpen, and Cédric Delporte—Joint Research Group ULB-VUB

Electronic supplementary material The online version of this article (doi:10.1007/s00216-016-9683-9) contains supplementary material, which is available to authorized users.

✉ Ana Planinc
aplaninc@ulb.ac.be

¹ Analytical Platform of the Faculty of Pharmacy and Laboratory of Pharmaceutical Chemistry, Faculty of Pharmacy, Université Libre de Bruxelles (ULB), Boulevard du Triomphe, Campus Plaine, CP 205/05B, 1050 Brussels, Belgium

² Laboratory of Instrumental Analysis and Bioelectrochemistry, Faculty of Pharmacy, Université Libre de Bruxelles (ULB), Boulevard du Triomphe, 1050 Brussels, Belgium

³ Department of Analytical Chemistry and Pharmaceutical Technology (FABI), Center for Pharmaceutical Research (CePhAR), Faculty of Medicine and Pharmacy, Vrije Universiteit Brussel (VUB), Laarbeeklaan 103, 1090 Brussels, Belgium

⁴ Pharmacy, Centre Hospitalier Universitaire Saint-Pierre, Brussels, Belgium

Keywords Therapeutic glycoproteins · N-Glycans · Principal component analysis (PCA) · Soft independent modelling by class analogy (SIMCA) · High-resolution mass spectrometry (HRMS) · Hydrophilic interaction liquid chromatography (HILIC)

Introduction

Glycosylation characterization of therapeutic glycoproteins has rapidly gained importance in the pharmaceutical industry [1]. Therapeutic glycoproteins, such as monoclonal antibodies, antibody drug conjugates, Fc-fusion proteins, enzymes, hormones, and clotting factors, are undoubtedly one of the fastest growing fields in terms of pharmaceutical production because of their successful use in treatments of severe diseases such as cancers, immune disorders, and infections [2, 3]. More than 60 % of those therapeutic proteins are glycosylated. Today, it is widely accepted that changes in the structure of glycans can have a major impact on the safety, efficacy, and quality of therapeutic glycoproteins [4]. For example, the changes in the terminal monosaccharide residues (galactose, fucose, mannose, and sialic acid) may alter the functioning of the therapeutic glycoproteins, such as their half-life, immunogenicity, toxicity, stability, and solubility [5–7]. Therefore, special attention should be paid to the detection of changes in the glycosylation patterns. The arrival of biosimilars on the market due to the expiry of many patents of therapeutic proteins emphasizes the need for strong analytical tools for glycosylation characterization [8].

Oligosaccharides are attached to the therapeutic proteins either through the nitrogen atom binding, the asparagine amino acid residue (N-glycans), or the oxygen atom binding, the serine, or the threonine amino acid residue (O-glycans). The N-glycans have been more widely studied than the O-glycans, although the latter also play an important role [9]. The N-glycans are classified into groups, such as high mannose glycans and complex glycans (see Electronic Supplementary Material (ESM) Fig. S1) [10, 11].

Glycosylation characterization is a relatively recent research area. The guidelines, containing protocols and methods for the glycosylation characterization and defining reports of glyco-analysis, are still in progress. The European Medicines Agency (EMA) guidelines only suggest that “particular attention should be paid to the degree of mannosylation, galactosylation, fucosylation, and sialylation” and that “the distribution of the main glycan structures should be determined” [12]. For a biosimilar comparability exercise, EMA only suggests that the “proposed biosimilar and the reference product are highly similar in terms of primary, secondary, and (to the extent possible) tertiary and quaternary (if any) structure, taking into account glycosylation and other post-translational modifications” [13]. Considering the fact that the changes in glycosylation can crucially affect the safety and the efficacy of the therapeutic proteins, more extensive glycosylation studies should be performed for each of the therapeutic glycoprotein.

Different protocols and methods are in use in order to obtain glycosylation characteristics of therapeutic glycoproteins. Commonly, the glyco-analysis begins

with the release of the N-glycans from the proteins using a peptide-N-glycosidase F (PNGase F). Subsequently, the N-glycans are separated either by liquid chromatography (LC) or by capillary electrophoresis (CE), coupled to appropriate detection instruments, including fluorescence spectrophotometry or mass spectrometry (MS) [14–16]. Nowadays, LC using hydrophilic interaction chromatography (HILIC) enables one of the best separations of N-glycans and fluorescence or MS detectors are considered to be a gold standard for N-glycan analysis [17–20].

Data acquisition and data analysis software are usually used to determine the distribution of the main N-glycans. Absolute or relative abundances of the N-glycans are displayed in huge tables that are often difficult to interpret (see ESM Table S2–S4). However, it is not straightforward from those tables to conclude whether any changes in glycosylation occurred.

In our study, principal component analysis (PCA) and classification through soft independent modelling by class analogy (SIMCA) were applied in order to interpret the mass spectrometry results easier. The possibility to detect differences in N-glycosylation between different IgG samples that had artificially been modified by adding different glycoproteins (ribonuclease B or fetuin) or by digesting with an enzyme (neuraminidase) was evaluated. The developed approach was finally applied to the batch-to-batch studies, including three batches of human plasma IgG and three batches of trastuzumab.

Materials and methods

Reagents and chemicals

Sodium phosphate, sodium dodecyl sulfate (SDS), potassium hydroxide, monosodium phosphate dihydrate, and potassium chloride were obtained from Merck (Darmstadt, Germany). Ammonium acetate, 2-mercaptoethanol, sodium borohydride, PNGase F (from *Elizabethkingia meningoseptica*, P7367), ribonuclease B (R7884), fetuin (F6131), neuraminidase (from *Clostridium perfringens*, N2876), and trifluoroacetic acid were purchased from Sigma-Aldrich (Steinheim, Germany). Acetic acid and ammonium formate were obtained from VWR International (Leuven, Belgium), acetonitrile and methanol from Fluka (Steinheim, Germany), Nonidet P40 from Roche Diagnostics (Vilvoorde, Belgium), Nanosep 10K omega spin filters from Pall Corporation (New York, USA), and PGC-SPE columns from Supelco (Bellefonte, PA, USA).

Saint-Pierre University hospital (Brussels, Belgium) kindly donated samples of trastuzumab (Herceptin[®]) and human plasma immunoglobulin G (IgG) (Multigam[®]).

Sample preparation

Glycan preparation from proteins was adapted from previously established protocols [21, 22]. Briefly, each glycoprotein sample (equivalent of 1 mg of protein) was first incubated for 15 min at 100 °C in 220 µL of 50 mM sodium phosphate (pH 7.5) containing 5 % SDS and 1 % 2-mercaptoethanol. After cooling to room temperature, Nonidet P40 was added (22 µL) at a concentration of 10 %. Samples were diluted with 50 mM sodium phosphate to the final volume of 484 µL. Then, 5 µL PNGase F (0.5 U/µL) was added to release N-glycans. Afterwards, samples were incubated for 20 h at 37 °C. Proteins were removed by Nanosep 10K omega spin filters (Pall Corporation, New York, USA) using centrifugation at 10,400×*g* for 20 min. Samples were then dried using a centrifuge vacuum, and N-glycans were then reduced by adding 20 µL of 1 M sodium borohydride in 50 mM KOH for 3 h at 50 °C. Reduction was stopped by adding 2 µL acetic acid (100 %). After addition of 1 mL 0.1 % trifluoroacetic acid, samples were cleaned by PGC-SPE (100 mg, 1 mL tubes). Methanol (twice 1 mL) was used for column activation, trifluoroacetic acid (0.1 %) for column washing (3 times 1 mL), and a mixture of acetonitrile/0.1 % trifluoroacetic acid (40/60, V/V) for glycans elution (1 mL). Samples were then dried by means of vacuum centrifugation, and dried samples were dissolved in 50 µL solution of water/acetonitrile (50/50, V/V) prior to injection into the LC-MS system.

In this study, the IgG glycoprofile was artificially modified by adding other glycoproteins or enzymes. Then, it was tried to detect the created changes in glycosylation by LC-MS. We prepared 35 samples, i.e., 5 replicates for 7 sample types or IgG glycoprofile modifications (see ESM Table S2). Briefly, neuraminidase (50 µL, 0.35 U/mL) was added to create the first set of five modified IgG glycoprotein samples (1 mg). Neuraminidase was used to digest and remove terminal sialic acid of the IgG samples. Fetuin (1 mg) was added to form the second set of five modified IgG samples (1 mg). Fetuin is a glycoprotein which, compared to IgG, includes different sialylated complex N-glycans. Glycans A2G2S2 and A3G3S3 (see ESM Fig. S1) are largely present in fetuin but are not in IgG. Ribonuclease B was added in three different concentrations (1, 10, and 50 % (RiboB/IgG, m/m)) to create the third, fourth, and fifth set of five IgG-modified samples (1 mg). Ribonuclease B is a glycoprotein rich in high mannose N-glycans, such as M5, M6, M7, M8, and M9 (see ESM Fig. S1).

For the samples treated with neuraminidase (to artificially remove sialic acid), digestion was performed before the denaturation of the glycoproteins. Samples containing 1 mg glycoprotein (IgG) were diluted by adding 450 µL 50 mM ammonium acetate (pH 5.3). Then, 50 µL neuraminidase (0.35 U/mL) was added. Neuraminidase was prepared in 25 mM potassium chloride and 10 mM monosodium phosphate

dihydrate (pH 6.3). Samples were incubated with neuraminidase for 10 h at 37 °C.

HILIC coupled to ESI-MS

Analyses were performed using a 1200 series rapid resolution liquid chromatograph (RRLC) coupled to a 6520 series electrospray ionization (ESI)-quadrupole time-of-flight (QTOF) mass spectrometer from Agilent Technologies (Waldbronn, Germany). Separation of the glycans was performed using an XBridge BEH Amide column (130 Å, 2.5 µm, 150 mm × 2.1 mm I.D.), obtained from Waters (Zellik, Belgium). The column temperature was set at 35 °C. The mobile phases used in all experiments were composed of 10 mM ammonium formate (solvent A) and acetonitrile (solvent B). The applied gradient was as follows: 0–25 min, 68 % B, 0.5 mL/min; 25–26 min, 57 % B, 0.5 mL/min; 26–31 min, 20 % B, 0.25 mL/min; 31–35 min, 68 % B, 0.25 mL/min; and 35–40 min, 68 % B, 0.5 mL/min. ESI-QTOF parameters were as follows: negative mode, high-resolution (4 GHz) mode, mass range 100–3200 *m/z*, gas temperature 325 °C, drying gas flow 5 L/min, nebulizer pressure 6 psi, and capillary voltage 3500 V. Nitrogen was used as the nebulizer gas. Data acquisition and data analysis were carried out by MassHunter Acquisition[®] software for QTOF (Version B.04 SP3) and MassHunter Qualitative Analysis[®] (Version B.06) software (both from Agilent Technologies). N-glycans from IgG, ribonuclease B, and fetuin were analyzed by extraction of ions (*m/z* values recorded in ESM Table S1). Major ions observed for glycans were [M-H]⁻, [M-H+COOH]⁻, [M-2H]²⁻, [M-H+COOH]²⁻. Each N-glycan was quantified using integration (area under the curve) of the merged *m/z* peak corresponding to the glycans in extracted-ion chromatogram (EIC *m/z* ± 0.01). Areas of the N-glycans were recorded. For data analysis, relative abundances of the N-glycans were taken into account.

Chemometric approaches

According to the chemometric approaches, PCA and SIMCA were selected. In the present study, PCA was applied to test whether it is possible to visualize N-glycosylation differences between the modified IgG sample types and between the batches of human plasma IgG and trastuzumab. PCA, which is a simple data analysis technique, allowed us to rapidly evaluate whether N-glycosylation changes between the sample types appeared. Moreover, PCA was applied to help identifying within which N-glycans changes occurred. SIMCA, which is a more complex data analysis technique, was applied to build and validate a classification model that is able to distinguish the modified IgG sample types according to the sample type. SIMCA was also applied to verify whether it is possible to use this model to classify real samples, in our case

samples of batches of human plasma IgG and trastuzumab. PCA and SIMCA chemometric approaches were so combined to (1) quickly visualize N-glycosylation changes (PCA) and (2) determine, e.g., if a released batch is assigned to the same model class as the reference product (SIMCA).

Data pre-processing

Prior to data analysis, the data were organized in an $n \times p$ data matrix X (n = number of samples and p = number of N-glycans). In each cell of the matrix, the relative abundance of a given N-glycan is presented (see ESM Table S2–S4). Depending on the examined data, the relative abundances of some or all N-glycans (using integrated areas under the curve for related peaks) from IgG, Ribo B, and fetuin (FA2, FA3, FA2G1, FA3G1, FA2G2, FA3G2, FA2G1S1, FA2G2S1, FA3G2S1, FA2G2S2, FA3G2S2, A2, A2G1, A2G2, M5, M6, M7, M8, M9, A2G2S1, A2G2S2, A3G3S3, and A3G3S4) were considered as variables. Column centering was applied as data pre-processing for PCA and for SIMCA.

PCA

PCA is an unsupervised multivariate data analysis technique used for exploratory data analysis purposes [23]. PCA creates so-called latent variables or principal components (PCs). These PCs are new axes defined in the initial data space. PC1 is constructed in such a way that it describes the largest possible data variation. PC2 describes the largest possible remaining data variation, etc. All PCs are orthogonal to each other. A score is a linear combination of the original variables and represents the projection of an object (sample) on the considered PCs. The weights used to make the linear combination are called the loadings on the considered PCs. The loadings represent the importance (weights) of the individual variables in the score. A 2D score plot represents the scores on two PCs, respectively, and provides information regarding the data structure, for instance, on the (dis)similarity of the samples [23].

Selection of a training set and test set

The MS data were divided according to the Duplex algorithm in a training set, used to build the model, and a test set, used to estimate the prediction accuracy of the model [24]. For the gathered IgG controls (IgGa and IgGb), 22 samples were selected for the training set and 8 for the test set, while for the artificially modified sample types (IgG+N, IgG+F, IgG+R1, IgG+R10, and IgG+R100), 11 were chosen for the training set and 4 for the test set. This method starts by selecting the two profiles furthest from each other (Euclidean distance) and puts them both in a first set, i.e., the training set. Then, the next two profiles furthest from each other in the remaining data set are

put in a second set, i.e., the test set. The procedure then continues by alternatively placing objects in the training and test set. Finally, when the test set contains the specified number of objects, the remaining samples are added to the training set [24].

SIMCA

SIMCA is a supervised multivariate classification technique that is especially used for data with high within-class variability. SIMCA requires a training set of data in order to build the training classes. For each of the training classes, the optimal number of PCs to consider is determined using cross-validation [25]. The Euclidean distance and Mahalanobis distance define a restricted space around the samples of the training classes [26]. Objects from the test set are assigned to a given training class, if they are situated within the restricted space around the samples of the respective training classes. Confidence limits were 95 %.

Software

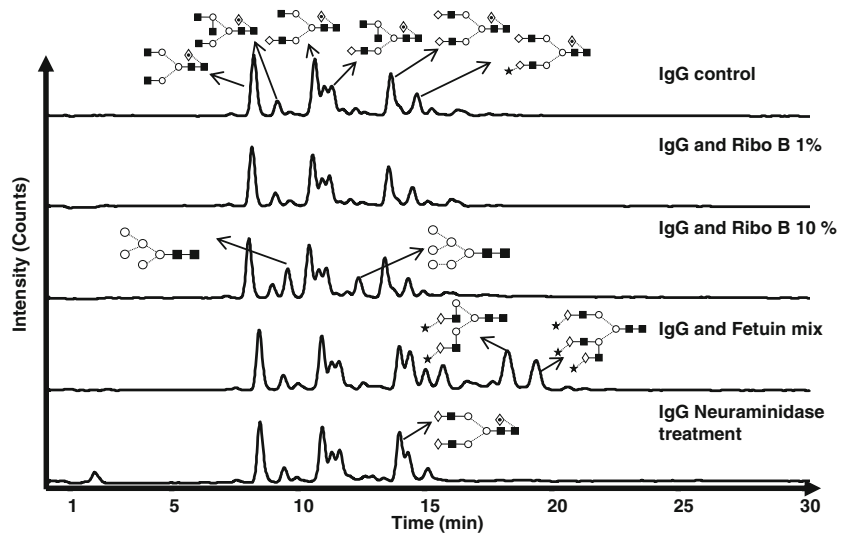
PCA was performed using m-files written for Matlab 2013b (The Mathworks, Natick, MA). SIMCA was performed using a trial version of PLS Toolbox 8.1 (Eigenvector Research, Manson, WA).

Results

Detection of glycosylation changes in IgG samples

N-glycans were detected using HILIC-HRMS, as previously established protocols [21, 22, 27]. At first, it was tried to find possible changes in glycosylation directly from the chromatograms of IgG samples, which had previously been artificially modified by adding other glycoproteins or enzymes (such as ribonuclease B, fetuin, and neuraminidase). EICs, where all N-glycans were merged into one chromatogram, were recorded. Then, the EICs of modified IgG sample types were compared to EIC of the control IgG (see Fig. 1). The EICs of modified IgG sample type, where ribonuclease B in the amount of 1 % of the IgG amount had been added, was very similar to the control EIC, which means that we are unable to detect small changes in N-glycosylation (a change of only 1 %) by comparing EICs. However, compared to the control EIC, two new clearly visible peaks in the EIC were noticed, where ribonuclease B in the amount of 10 % of the IgG amount had been added. These two peaks appeared due to the increased amounts of M5 and M6 N-glycans. In the EIC of the sample, where fetuin had been added to IgG, two major peaks appeared, representing A2G2S2 and A3G3S3 N-glycans. In the last EIC sample, which had been digested with

Fig. 1 Extracted-ion chromatograms (EICs) of IgG control and artificially changed sample types: Differences between the chromatograms of IgG (control), IgG with ribonuclease B (1 and 10 %), IgG with fetuin, and IgG with neuraminidase are shown. The five chromatograms represent the extracted-ion chromatograms (EICs), where all *m/z* values of known N-glycans were merged into one chromatogram



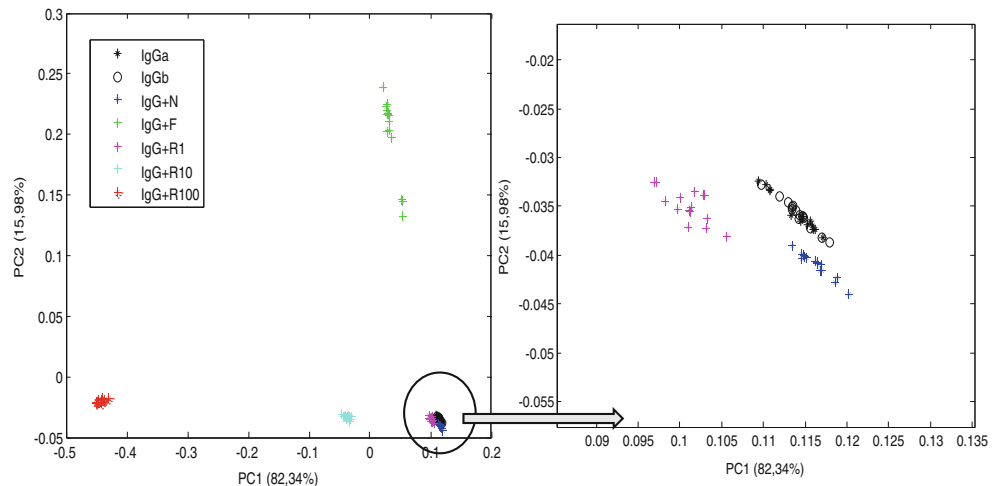
neuraminidase, differences in the second part of the EIC were observed. Accordingly, the peak representing FA2G2 increased, because of the prior removal of the sialic acid residue from FA2G2S1 and FA2G2S2.

Then, we extracted the N-glycans of interest in EICs (using *m/z* values presented in ESM Table S1). Peaks corresponding to the N-glycans were selected and integrated (taking into account the area under the curve for each N-glycan). Areas under the curve were collected for N-glycans of interest, and the relative abundances of those N-glycans were determined (see ESM Table S2). Then, the relative abundances of the modified IgG sample types N-glycans were compared to the relative abundances of the control IgG N-glycans. In ESM Table S2, differences in relative abundances of high mannose N-glycans (M5, M6, M7, M8, and M9) were noticed for samples, where ribonuclease B was added. In samples artificially modified with fetuin, four new complex N-glycans were

detected, which do not contain core-fucose. These four N-glycans are A2G2S2, A3G3S3, A2G2S1, and A3G3S4. Finally, the treatment with neuraminidase decreased the relative abundances of the sialylated glycans, such as FA2G1S1, FA2G2S1, FA3G2S1, FA2G2S2, and FA3G2S2. These decreases are reflected by the increase of the relative abundances of N-glycans, which previously lost sialic acid, such as FA2G2 and FA3G2.

Furthermore, PCA was applied in order to more easily detect the changes in glycosylation between the modified IgG sample types and the control IgG. Figure 2 represents the clustering of the IgG sample types (IgG, IgG with addition of neuraminidase, IgG with addition of fetuin, and IgG with addition of three different concentrations of ribonuclease B) in the PC1–PC2 score plot. The IgG sample types with modified glycoprofiles are clearly separated from the IgG control samples (black spots and dark blue spots, two controls). The

Fig. 2 PCA score plots of IgG controls and artificially changed sample types: PCA score plots of the 105 samples (7 sample types × 5 replicates × 3 injections) × 24 N-glycans. Differences between sample types IgGa (control 1), IgGb (control 2), IgG+N, IgG+F, IgG+R1, IgG+R10, and IgG+R100 are visualized. PC1–PC2 score plot and zoom on the IgGa, IgGb, IgG+N, and IgG+R1 sample types in the PC1–PC2 score plot are shown



samples containing the smallest amount of the ribonuclease B (light blue spots) are the most similar to the control samples as already observed in Fig. 1.

The contribution of the original variables to the scores in the different PCs can be observed on 2D loading plots (see ESM Fig. S2). Separation along PC1 is mainly caused by variables M5 and M6 and also by variables FA2G1, FA2, and FA2G2. Separation along PC2 is mainly caused by A2G2S2, A3G3S3, and also by variables FA2G1, FA2, FA2G2, M5, and M6.

Figure 3 represents the SIMCA-predicted class plot of the test set samples (IgG, IgG with addition of neuraminidase, IgG with addition of fetuin, and IgG with addition of three different concentrations of ribonuclease B). All test samples are assigned to their real class, except one sample of the IgG+R1 sample type (addition of ribonuclease B to the IgG in the amount of 1 % of IgG amount), which is predicted as a control IgG sample (class model 1). This sample is not assigned to the IgG+R1 sample type (class 4), because its probability of being a class 4 member is below the 95 % confidence limit. These results show that we were able to differentiate minor modifications of the glycoprofiles of the modified IgG sample types. Therefore, this approach could be applied to the real samples, such as different batches of therapeutic glycoproteins, in order to assess batch-to-batch consistency, described below.

Glycosylation characterization of different batches of human plasma IgG and trastuzumab

One possible application of the LC-MS-PCA/SIMCA approach is the detection of changes in glycosylation between

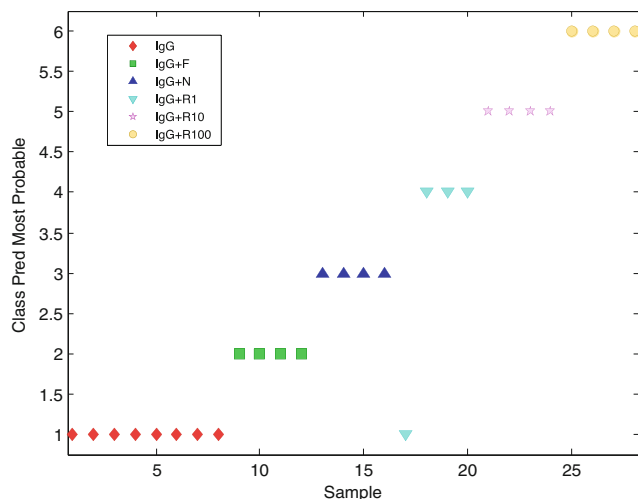


Fig. 3 SIMCA-predicted class plot of IgG control and artificially changed sample types: SIMCA-predicted class plot of 28 test samples \times 24 N-glycans as variables. Classification of gathered IgG controls (eight samples were selected for the test set) and the artificially modified sample types IgG+N, IgG+F, IgG+R1, IgG+R10, and IgG+R100 (four samples were chosen for the test set) is shown

batches. Three batches of human plasma IgG and three batches of trastuzumab were tested.

Data of the human plasma IgG batches and data of the trastuzumab batches were added to the PCA score plot introduced in Fig. 2. From the PC1–PC2 score plot in Fig. 4, it can be observed that samples of human plasma IgG (lot 1, lot 2, and lot 3) correspond to the control IgG, while samples of trastuzumab (lot 1, lot 2, and lot 3) differ largely from the control IgG. Moreover, there is no difference between the human plasma IgG batches, while the zoom on PC1–PC2 score plot reveals that the trastuzumab batch-to-batch differences are slightly larger than the within-batch differences. ESM Table S4 reveals that trastuzumab batches differ in the quantity of certain N-glycans, such as FA2, FA2G1, and M5 (e.g., relative abundance of M5 in lot 1 is about 1.2 %, in lot 2 it is about 2.5 %, and in lot 3 it is about 2.9 %).

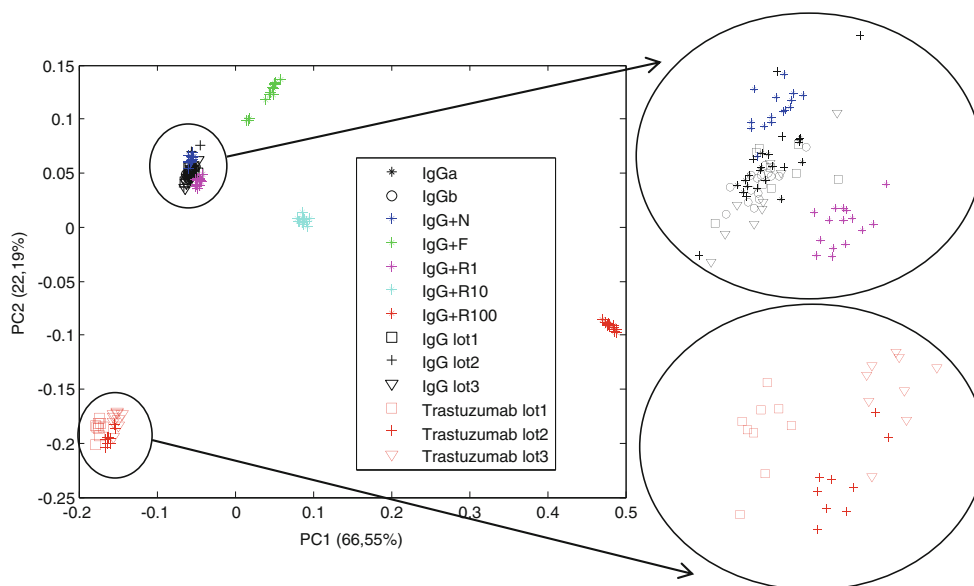
Finally, the SIMCA class model, which was created in the previous experiment (see Fig. 3) for the six classes, was used to test data of the human plasma IgG batches and data of the trastuzumab batches. All test samples of the human plasma IgG were predicted as a control IgG (class 1, control IgG), while trastuzumab test samples were not predicted in any of the six classes.

Discussion

Nowadays, huge amounts of data are obtained by glycoanalysis of therapeutic glycoproteins. Therefore, it is necessary to find an approach that simplifies the way of describing the entire glycoprofile of the therapeutic glycoproteins. This problem was pointed out before. Gervais et al. introduced two parameters describing the glycosylation of human recombinant gonadotrophins [28]. The first parameter, named the hypothetical charge Z number, characterizes the terminal sialylation level and is calculated as the sum of the areas under the curve for related glycans each multiplied by the corresponding charge (due to the sialylation). The Z number is currently used by the European Pharmacopoeia (eighth edition) to characterize the glycosylation of follitropin. The second parameter, called the hypothetical antennarity index A , describes the antennarity level and is calculated as the sum of the areas under the curve for related glycans each multiplied by the corresponding number of antennas. Calculated values of both parameters were used in order to assess consistency between the different batches of the human recombinant gonadotrophins describing the terminal sialylation and the antennarity level [28].

Describing the terminal sialylation and the antennarity levels is not enough to assess the safety and the efficacy of the therapeutic glycoproteins. As mentioned before, also terminal galactosylation, core-fucosylation, and mannosylation may change the functioning of the therapeutic glycoproteins.

Fig. 4 PCA score plots of three batches of human plasma IgG and of trastuzumab: Data of the 27 samples of human plasma IgG (3 batches \times 3 replicates \times 3 injections) and of the 27 samples of trastuzumab (3 batches \times 3 replicates \times 3 injections) \times 24 N-glycans were added to the PCA score plots introduced in Fig. 2. Differences between three batches of human plasma IgG (lot 1, lot 2, and lot 3) and trastuzumab (lot 1, lot 2, and lot 3) are presented. PC1–PC2 score plot and zoom on the control IgG region and trastuzumab region are shown



For example, an increase in terminal galactosylation largely increases clearance (decreases half-life) from the blood circulation due to the asialoglycoprotein receptor in the liver. This receptor binds and removes glycoproteins containing terminal galactose or *N*-acetylgalactosamine [29–31]. Moreover, an increase in terminal galactosylation increases the anti-inflammatory effect due to the stronger binding of the therapeutic glycoprotein to the Fc receptors expressed on natural killer cells [32]. Therapeutic glycoproteins lacking in core-fucosylation are also associated with a much higher anti-inflammatory effect (for the same reason as terminal galactosylated therapeutic glycoproteins). Therefore, afucosylated therapeutic glycoproteins can potentially become the next-generation therapeutic glycoproteins with improved efficacy in certain applications [33, 34]. It was also described that an increased presence of high mannose glycans is associated with an increased clearance of the therapeutic glycoproteins (due to the mannose binding receptor in the liver) and with off-target hepatic toxicities. At the same time, an increased presence of high mannose glycans can indicate the unfinished synthesis of the complex glycans, because high mannose glycans are highly presented in the first stages of the synthesis of complex glycans [35]. In this context, new approaches are required for a better discrimination of glycosylation which take into account all characteristics (terminal sialylation, terminal galactosylation, core-fucosylation, and mannosylation) in a simple and unique analysis process.

In this study, we illustrated that a simple comparison of chromatographic profiles (see Fig. 1) without a good data handling is insufficient to detect changes in glycosylation despite the fact that this approach is widely used by pharmaceutical industries. For instance, a small increase in high mannose glycans could not be observed when 1 % of ribonuclease B was added to IgG. Indeed, it is crucial to find and use an

approach which is capable to detect small changes in relative glycan amount, which may occur during the industrial recombinant production. Therefore, we continued with the chemometric analysis of our data hoping to be able to discriminate glycosylation changes according in only 1 %. Application of PCA and SIMCA analyses allowed to distinguish between all modified IgG sample types. It means that the PCA and SIMCA developed tools are capable to detect even very small changes in glycosylation (1 %).

Using PCA, it is possible to observe differences in glycoprofiles (changes in mannosylation, galactosylation, sialylation, and fucosylation) (see Fig. 2). PCA allows to determine within which N-glycans the changes in glycosylation occurred (see ESM Fig. S2). In this study, the largest differences (described by PC1) were detected between the M5 and M6 N-glycans and also between the FA2G1, FA2, and FA2G2 N-glycans, which means that PCA detected the major changes within mannosylation and galactosylation. Furthermore, PC2 detected the changes between the A2G2S2 and A3G3S3 which suggests that PCA detected the changes within sialylation as well. The largest advantage of the developed approach compared to previously described approaches is that the LC-MS-PCA approach allows the detection of changes in glycosylation taking into account all characteristics (terminal sialylation, terminal galactosylation, core-fucosylation, and mannosylation) simultaneously.

Using SIMCA, it is also possible to observe a difference in glycosylation between all modified IgG sample types (see Fig. 3). Accordingly, all test samples match the SIMCA class models, except one sample of the IgG+R1 sample type, which was found most similar to the control IgG. This classification approach can be used to verify whether a new released batch has the same glycoprofile than a referenced batch and could be used for quality control statements.

Like Gervais' approach [28], PCA and SIMCA help to assess consistency between the different batches of therapeutic glycoproteins monitoring all glycosylation changes. In present study, three batches of human plasma IgG and three batches of trastuzumab were selected in order to test PCA and the SIMCA model, which were created in the previous experiment of the artificially modified IgG sample types (IgG control, IgG+N, IgG+F, IgG+R1, IgG+R10, and IgG+R100). PCA and SIMCA suggest that all samples of the human plasma IgG batches match the control IgG sample type. However, the samples of trastuzumab batches are badly predicted using previously developed SIMCA model, because the glycoprofile of the trastuzumab greatly differs from the glycoprofiles of all modified sample types (see PCA score plot Fig. 4). The results also suggest that there is no difference between the three human plasma IgG batches, while the differences between the trastuzumab batches are slightly larger than the within-batch differences. The differences in glycosylation between the trastuzumab batches can be explained by the fact that trastuzumab is produced in Chinese hamster ovary recombinant cells. While IgG is extracted from human plasma, recombinant production of therapeutic glycoproteins is definitely more demanding than obtaining therapeutic glycoproteins from human plasma (as human plasma IgG). In recombinant production, glycosylation patterns are highly sensitive to the culture conditions (e.g., nutrient levels, dissolved oxygen level, pH, and temperature) [36]. Further studies are needed in order to evaluate the acceptable variation limits in relative abundances of certain N-glycans between the batches in order to ensure the safety and the efficacy of the therapeutic proteins. Accordingly, mathematical and statistical approaches should be used as illustrated in the present study.

Conclusions

In conclusion, LC-MS-PCA/SIMCA approach was introduced for the characterization of glycosylation of the therapeutic glycoproteins. This approach enables to apply MS results of glyco-analyses of therapeutic glycoproteins to PCA score and loading plots. Furthermore, it also enables to build SIMCA models using MS data. By evaluating the PCA plots, it is possible to detect small changes in glycosylation of the therapeutic glycoproteins (a change of only 1 % in relative glycan amount), which cannot be detected using the classical chromatographic profile comparison. PCA can also help to identify within which N-glycans changes occurred. Furthermore, the SIMCA model allows classification of the differently glycosylated profiles when the representative training set is available. SIMCA can be used to determine if a released batch is assigned to the same class as the reference product. Certainly, the proposed LC-MS-PCA/SIMCA

approach can help manufacturers in assessing the changes in glycosylation between different batches of therapeutic glycoproteins as well as establishing acceptance limits for the glycosylation changes.

According to the N-glycosylation changes of the therapeutic proteins, pharmaceutical industry needs robust and fast answers, because even small N-glycosylation changes can affect the safety and the efficacy of the therapeutic proteins. The present approach (HILIC-MS-PCA/SIMCA) contributes to the N-glycosylation characterization in a way that it facilitates N-glycosylation data interpretation. In this manner, N-glycosylation data interpretation could become more and more automated. Thanks to the better data handling approaches, LC-MS analysis could gain even more important place in the N-glycosylation analysis.

Determination of glycosylation changes is indispensable for providing safety, efficacy, and quality of the therapeutic glycoproteins. In the future, optimization of the glycan sample preparation and data processing should also be investigated, focusing on getting the results in a shorter timeframe, because in our study, sample preparation and data extraction were still time consuming.

Acknowledgments This study was supported by grants from the Belgian Fund for Scientific Research (FRS-FNRS), no. 34553.08, grant from the FER 2007 (ULB), and grant from Slovenia (Ad futura). CD is a postdoctoral researcher of the FRS-FNRS.

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interests.

References

1. Walsh G. Biopharmaceutical benchmarks 2014. *Nat Biotechnol.* 2014;32(10):992–1000.
2. Carter PJ. Introduction to current and future protein therapeutics: a protein engineering perspective. *Exp Cell Res.* 2011;317(9):1261–9.
3. Walsh G. Biopharmaceutical benchmarks 2010. *Nat Biotechnol.* 2010;28(9):917–24.
4. Hossler P, Khattak SF, Li ZJ. Optimal and consistent protein glycosylation in mammalian cell culture. *Glycobiology.* 2009;19(9):936–49.
5. Delobel A, Cantais F, Catrain A, Dereux E, Van Vyncht G. Therapeutic antibody glycosylation analysis: a contract research organization perspective in the frame of batch release or comparability support. *Methods Mol Biol.* 2013;988:115–43.
6. Kobata A. The N-linked sugar chains of human immunoglobulin G: their unique pattern, and their functional roles. *Biochim Biophys Acta.* 2008;1780(3):472–8.
7. Raju TS, Scallion B. Fc glycans terminated with N-acetylglucosamine residues increase antibody resistance to papain. *Biotechnol Prog.* 2007;23(4):964–71.

8. Gaugh D. The biosimilar price is right. *Nat Biotechnol.* 2013;31(11):971.
9. Warren CE. Glycosylation. *Curr Opin Biotechnol.* 1993;4(5):596–602.
10. Raju TS, Briggs JB, Borge SM, Jones AJS. Species-specific variation in glycosylation of IgG: evidence for the species-specific sialylation and branch-specific galactosylation and importance for engineering recombinant glycoprotein therapeutics. *Glycobiology.* 2000;10(5):477–86.
11. Harvey DJ, Merry AH, Royle L, Campbell MP, Dwek RA, Rudd PM. Proposal for a standard system for drawing structural diagrams of N- and O-linked carbohydrates and related compounds. *Proteomics.* 2009;9(15):3796–801.
12. Guideline on monoclonal antibodies adopted by CHMP—WC500003074. 2008. from:http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003074. Accessed 05 April 2016.
13. Overview of comments overarching guideline—WC500177967. 2014. http://www.ema.europa.eu/docs/en_GB/document_library/Overview_of_comments/2014/12/WC500177967.pdf. Accessed 05 April 2016.
14. Tharmalingam T, Adameczyk B, Doherty MA, Royle L, Rudd PM. Strategies for the profiling, characterisation and detailed structural analysis of N-linked oligosaccharides. *Glycoconj J.* 2013;30(2):137–46.
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(10):713–23.
16. Jensen PH, Karlsson NG, Kolarich D, Packer NH. Structural analysis of N- and O-glycans released from glycoproteins. *Nat Protoc.* 2012;7(7):1299–310.
17. Guo Y, Yuan Q, Li R, Huang Y. Hydrophilic interaction chromatography and its recent applications in environmental analysis. *Chin J Chromatogr.* 2012;30(3):232–8.
18. Yuen C-T, Gee CK, Jones C. High-performance liquid chromatographic profiling of fluorescent labelled N-glycans on glycoproteins. *Biomed Chromatogr.* 2002;16(4):247–54.
19. Hu Y, Mechref Y. Comparing MALDI-MS, RP-LC-MALDI-MS and RP-LC-ESI-MS glycomic profiles of permethylated N-glycans derived from model glycoproteins and human blood serum. *Electrophoresis.* 2012;33(12):1768–77.
20. Nishikaze T, Fukuyama Y, Kawabata S, Tanaka K. Sensitive analyses of neutral N-glycans using anion-doped liquid matrix G3CA by negative-ion matrix-assisted laser desorption/ionization mass spectrometry. *Anal Chem.* 2012;84(14):6097–103.
21. Morelle W, Michalski J-C. Analysis of protein glycosylation by mass spectrometry. *Nat Protoc.* 2007;2(7):1585–602.
22. Millán Martín S, Delporte C, Farrell A, Navas Iglesias N, McLoughlin N, Bones J. Comparative analysis of monoclonal antibody N-glycosylation using stable isotope labelling and UPLC-fluorescence-MS. *Analyst.* 2015;140(5):1442–7.
23. Esbensen KH, Geladi P. Principal component analysis: concept, geometrical interpretation, mathematical background, algorithms, history, practice. *Comprehensive chemometrics.* 2nd ed. Amsterdam: Elsevier; 2002. p. 211–26.
24. Snee RD. Validation of regression models: methods and examples. *Technometrics.* 1977;19(4):415–28.
25. Bylesjö M, Rantalainen M, Cloarec O, Nicholson JK, Holmes E, Trygg J. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *J Chemometrics.* 2006;20(8–10):341–51.
26. Custers D, Krakowska B, De Beer JO, Courselle P, Daszykowski M, Apers S, et al. Testing of complementarity of PDA and MS detectors using chromatographic fingerprinting of genuine and counterfeit samples containing sildenafil citrate. *Anal Bioanal Chem.* 2016;408(6):1643–56.
27. Huhn C, Selman MHJ, Ruhaak LR, Deelder AM, Wührer M. IgG glycosylation analysis. *Proteomics.* 2009. <http://www.readcube.com/articles/10.1002%2Fpmic.200800715>. Accessed 29 Jan 2016.
28. Gervais A, Hammel Y-A, Pelloux S, Lepage P, Baer G, Carte N, et al. Glycosylation of human recombinant gonadotrophins: characterization and batch-to-batch consistency. *Glycobiology.* 2003;13(3):179–89.
29. Park EI, Manzella SM, Baenziger JU. Rapid clearance of sialylated glycoproteins by the asialoglycoprotein receptor. *J Biol Chem.* 2003;278(7):4597–602.
30. Park EI, Mi Y, Unverzagt C, Gabius H-J, Baenziger JU. The asialoglycoprotein receptor clears glycoconjugates terminating with sialic acid alpha 2,6GalNAc. *Proc Natl Acad Sci U S A.* 2005;102(47):17125–9.
31. Mi Y, Shapiro SD, Baenziger JU. Regulation of lutropin circulatory half-life by the mannose/N-acetylgalactosamine-4-SO₄ receptor is critical for implantation in vivo. *J Clin Invest.* 2002;109(2):269–76.
32. Goulabchand R, Vincent T, Batteux F, Eliaou J-F, Guilpain P. Impact of autoantibody glycosylation in autoimmune diseases. *Autoimmun Rev.* 2014;13(7):742–50.
33. Iida S, Kuni-Kamochi R, Mori K, Misaka H, Inoue M, Okazaki A, et al. Two mechanisms of the enhanced antibody-dependent cellular cytotoxicity (ADCC) efficacy of non-fucosylated therapeutic antibodies in human blood. *BMC Cancer.* 2009;9:58.
34. Okazaki A, Shoji-Hosaka E, Nakamura K, Wakitani M, Uchida K, Kakita S, et al. Fucose depletion from human IgG1 oligosaccharide enhances binding enthalpy and association rate between IgG1 and FcγRIIIa. *J Mol Biol.* 2004;336(5):1239–49.
35. Gorovits B, Krinos-Fiorotti C. Proposed mechanism of off-target toxicity for antibody-drug conjugates driven by mannose receptor uptake. *Cancer Immunol Immunother.* 2013;62(2):217–23.
36. Nyberg GB, Balcarcel RR, Follstad BD, Stephanopoulos G, Wang DI. Metabolic effects on recombinant interferon-gamma glycosylation in continuous culture of Chinese hamster ovary cells. *Biotechnol Bioeng.* 1999;62(3):336–47.