



Computer-aided gradient optimization of hydrophilic interaction liquid chromatographic separations of intact proteins and protein glycoforms[☆]



Guusje van Schaick^a, Bob W.J. Pirok^{b,c}, Rob Haselberg^{a,c}, Govert W. Somsen^{a,c},
Andrea F.G. Gargano^{a,b,c,*}

^a Division of Bioanalytical Chemistry, Amsterdam Institute for Molecules, Medicines and Systems, Vrije Universiteit Amsterdam, de Boelelaan 1085, 1081 HV Amsterdam, The Netherlands

^b University of Amsterdam, van 't Hoff Institute for Molecular Sciences, Analytical-Chemistry Group, Science Park 904, 1098 XH Amsterdam, The Netherlands

^c Centre for Analytical Sciences Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands

ARTICLE INFO

Article history:

Received 19 December 2018

Received in revised form 18 March 2019

Accepted 19 March 2019

Available online 3 April 2019

Keywords:

HILIC

Intact protein separation

Glycoform separations

Middle-up protein analysis

Computer-aided method development

ABSTRACT

Protein glycosylation is one of the most common and critical post-translational modification, which results from covalent attachment of carbohydrates to protein backbones. Glycosylation affects the physicochemical properties of proteins and potentially their function. Therefore it is important to establish analytical methods which can resolve glycoforms of glycoproteins. Recently, hydrophilic-interaction liquid chromatography (HILIC)-mass spectrometry has demonstrated to be a useful tool for the efficient separation and characterization of intact protein glycoforms. In particular, amide-based stationary phases in combination with acetonitrile-water gradients containing ion-pairing agents, have been used for the characterization of glycoproteins. However, finding the optimum gradient conditions for glycoform resolution can be quite tedious as shallow gradients (small decrease of acetonitrile percentage in the elution solvent over a long time) are required. In the present study, the retention mechanism and peak capacity of HILIC for non-glycosylated and glycosylated proteins were investigated and compared to reversed-phase liquid chromatography (RPLC). For both LC modes, $\ln k$ vs. φ plots of a series of test proteins were calculated using linear solvent strength (LSS) analysis. For RPLC, the plots were spread over a wider φ range than for HILIC, suggesting that HILIC methods require shallower gradients to resolve intact proteins. Next, the usefulness of computer-aided method development for the optimization of the separation of intact glycoform by HILIC was examined. Five retention models including LSS, adsorption, and mixed-mode, were tested to describe and predict glycoprotein retention under gradient conditions. The adsorption model appeared most suited and was applied to the gradient prediction for the separation of the glycoforms of six glycoproteins (Ides-digested trastuzumab, alpha-acid glycoprotein, ovalbumin, fetuin and thyroglobulin) employing the program PIOTR. Based on the results of three scouting gradients, conditions for high-efficiency separations of protein glycoforms varying in the degree and complexity of glycosylation was achieved, thereby significantly reducing the time needed for method optimization.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

Proteins are macromolecules with a complex and heterogeneous structure, which is partly due to post-translational

modifications (PTMs). One of the most critical PTMs is glycosylation. About half of the mammalian proteome is glycosylated, and approximately one-third of the approved biopharmaceuticals are glycoproteins [1]. Glycosylation is an enzyme-mediated process where carbohydrates (glycans) are covalently attached to proteins. The glycans can be attached to a serine or threonine residue (O-glycosylation) or an asparagine residue (N-glycosylation) of the backbone of proteins [2]. N-glycans share the same core structure and are classified into three different types: oligomannose, complex and hybrid [3]. In contrast, O-glycans do not have a distinct core structure. The attachment of glycans may affect the tertiary

[☆] Selected papers from the 32nd International Symposium on Chromatography (ISC 2018), September 23–27, 2018, Cannes-Mandelieu, France.

* Corresponding author at: Center for Analytical Sciences Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands.

E-mail address: a.gargano@uva.nl (A.F.G. Gargano).

structure and physicochemical properties of the protein, such as stability, solubility, and folding [4]. In biopharmaceuticals, these effects could result in a change of the quality, safety, and efficacy of the product. Therefore, it is of great importance to be able to monitor and check the glycosylation of these biopharmaceutical products [1].

There are three major approaches to study the glycosylation of proteins: analysis of released glycans [5], analysis of glycopeptides (either by bottom-up or middle-down approaches) [6,7], and analysis of intact glycoproteins [8]. Released glycans are obtained by enzymatic or chemical cleavage. This approach can help with the determination of the different glycan structures present, but results in a loss of information on the protein attachment sites of the glycans [9]. For the analysis of glycopeptides, the glycoproteins are digested by specific endoproteases (e.g., trypsin). With this approach, a complete overview of the glycosylation sites of proteins can be obtained. However, information on co-occurring glycosylation sites and the number and distribution of glycoforms is lost [10]. For the determination of the actual glycoforms, the analysis of intact glycoproteins is a more suitable approach [11]. So far, several analytical techniques have been described for the glycoform profiling of intact proteins, including capillary electrophoresis (CE) and liquid chromatography (LC) [8,12]. Coupling of CE or LC with high-resolution mass spectrometry (MS) enables the determination of the accurate mass of separated intact proteins. More structural information regarding protein sequence and PTMs can be obtained when employing tandem MS approaches [13].

When considering LC for the analysis of intact glycoproteins, different modes can be applied. For example, reversed-phase (RP) LC enables to resolve protein heterogeneity according to sequence (amino acid composition) but has limited selectivity toward glycoforms [9]. An attractive alternative for the separation and characterization of intact glycoproteins is hydrophilic-interaction liquid chromatography (HILIC) [14]. Currently, this technique is mostly applied to the chromatographic separation of small polar molecules [15] peptides and glycopeptides [16]. The precise retention mechanism of HILIC is debated, but the generally accepted idea is that retention derives from a combination of partitioning processes and electrostatic interactions (ion exchange and hydrogen bonding) between the analytes and the surface of the hydrophilic stationary phase. When analyzing proteins with HILIC, relatively high percentages of water are needed to assure protein solubilization and elution, leaving electrostatic interactions the predominant cause of retention.

Different types of stationary phases, such as hydroxylated stationary phases and weak ion-exchangers, have been adopted for the analysis of proteins by HILIC. These materials have proven to be useful for the separation of hydrophobic proteins (e.g., membrane proteins [17]) and charge variants of proteins [18]. Recently, amide-based stationary phases have demonstrated good performances for the separation of intact proteins [19] demonstrating interesting selectivity for the separation of glycoforms of glycoproteins [7,20,21]. The acetonitrile (ACN)-water mobile phases used for amide-based HILIC of proteins typically contain 0.05–0.1% trifluoroacetic acid (TFA). TFA lowers the pH of the mobile phase protonating the acidic residues of the protein and the free silanol groups of the stationary phase material. At the same time, TFA acts as an ion-pairing agent, interacting with the protonated basic residues of the protein. As a result of this, the HILIC separation of proteins is mainly driven by polar, but neutral, groups on the protein backbone (incl. glycans) and not by charged residues [11]. Such HILIC systems can separate a wide range of proteins, as demonstrated by the analysis of a cell lysate [19]. When applying shallow gradients (e.g., decreasing the ACN % of the mobile phase of 10% over 30 min), highly efficient glycoform separations

of glycoproteins such as monoclonal antibodies [7], therapeutical glycoproteins [11], and neo-glycoproteins [9,12] are obtained.

Unfortunately, the determination of the optimal gradient for a set of protein glycoforms can be a cumbersome process as it can be difficult to determine a suitable gradient program. To facilitate efficient method development in LC, computer-aided approaches, such as ChromSword, DryLab and PIOTR, have been developed [22–24]. These software use models (based on, e.g., linear solvent strength (LSS), ion exchange or mixed-mode) to describe retention and to make predictions of protein retention times based on a limited number of scouting gradients. The possibilities of automated method development in liquid chromatography for large biomolecules, such as therapeutic proteins, have been the topic of a recent review [25]. An example is the recent work of Bobaly et al. in which the DryLAB software was used to develop a generic HILIC method to study the glycoforms of Ides digested monoclonal antibodies and antibody drug conjugates [26]. In this study, the gradient steepness and temperature effects were tested, monitoring the increase of resolution of already resolved features and assuming a linear relationship between gradient time and gradient retention factor.

Retention models have been objects of recent studies to verify their applicability to different classes of compounds and separation modes. In particular, Tyteca et al. tested three retention models (LSS, quadratic, and Neue-Kuss) for the separation of small molecules, peptides and proteins in RPLC. In this study, the LSS model was described as the most suitable to describe the retention behavior [27]. Recently, five different retention models were applied to model the gradient elution in HILIC of small molecules and peptides using PIOTR [28], concluding that the adsorption model had the best fitting and prediction for the analyte set discussed.

In the present study, retention and chromatographic behavior of proteins in HILIC (using an amide stationary phase) were first compared to RPLC (C4 stationary phase). After that, accelerated optimization of HILIC methods for the separation of protein glycoforms was developed using a computer-aided approach. Different retention models (mixed-mode, Neue-Kuss, adsorption, LSS, and quadratic model) were compared for predicting the gradient conditions needed for optimal resolution. Finally, a PIOTR method employing the HILIC adsorption retention model and a Pareto-optimization approach was evaluated as a prediction tool to obtain gradient separation conditions of glycoforms from proteins varying in degree and complexity of glycosylation (Fc parts of trastuzumab, ovalbumin, fetuin, alpha-1-acid glycoprotein, and thyroglobulin).

Our results demonstrate that the adsorption model describes protein (and glycoprotein) elution in HILIC adequately. Gradient conditions resulting in efficient glycoform separations could be readily derived from scouting gradients, which by itself did not provide glycoform resolution.

2. Experimental

2.1. Chemicals and sample preparation

Deionized water (18.2 mΩ) was obtained from a Milli-Q purification system (Millipore, Bedford, USA). Acetonitrile (ACN; HPLC or MS grade) and trifluoroacetic acid (TFA; MS grade) were obtained from Biosolve B.V. (Valkenswaard, The Netherlands). Isopropanol (IPA; LC-MS grade), tris(hydroxymethyl)aminomethane ($\geq 99.8\%$), and hydrochloric acid (37%) were purchased from Sigma (Zwijndrecht, The Netherlands). All materials were used as received, and the mobile phases were not filtered before use. Alpha-acid glycoprotein (AGP) from human ($\geq 99\%$), alpha-chymotrypsin (c.tryp) from bovine, albumin from chicken egg white (ova) ($\geq 98\%$),

cytochrome C (cyt C) from equine heart ($\geq 95\%$), serum albumin (BSA) from bovine ($\geq 98\%$), carbonic anhydrase (CA) from bovine erythrocytes ($\geq 95\%$), fetuin (fet) from fetal calf serum, lysozyme (lys) from chicken egg white (95%), myoglobin (myo) from horse heart ($> 90\%$), ribonuclease A (RnA) from bovine pancreas Type I-A ($\geq 60\%$), ribonuclease B (RnB) from bovine pancreas ($\geq 80\%$), transferrin (trans) from human ($\geq 98\%$), trypsinogen (tryp) from bovine pancreas, and ubiquitin (ubi) from bovine erythrocytes ($\geq 98\%$) were acquired from Sigma. Thyroglobulin (thyro) from rabbit (polyclonal) was purchased from Biologend (San Diego, United States). Herceptin (trastuzumab) was acquired from Roche (Basel, Switzerland). The immunoglobulin-degrading enzyme of *Streptococcus pyogenes* (IdeS, FabRICATOR) was purchased from Genovis Inc. (Lund, Sweden).

Protein standard solutions (1 mg/mL) were prepared in deionized water. The IdeS-digestion of trastuzumab was performed following the protocol provided by the manufacturer (Genovis Inc.). Briefly, trastuzumab (100 μg) in 10 mM TRIS buffer (pH 7.5) was incubated with 100 units of IdeS enzyme at 37 °C overnight. No purification step was performed.

2.2. LC instrumentation, conditions and data analysis

Protein separations were performed on an Agilent HPLC 1290 Infinity II (Waldbronn, Germany), composed of an autosampler, column thermostat, variable wavelength detector, and Agilent HPLC 1100 binary pump. For the HILIC method optimization with PIOTR, an Agilent AdvanceBio glycan column (150 \times 2.1 mm i.d.; 2.7 μm particles with 125 Å pore size) was used (Etten-Leur, The Netherlands). For the comparison with RPLC, the HILIC column: Waters Acquity UPLC glycoprotein Amide column (50 \times 2.1 mm i.d.; 1.7 μm particles with 300 Å pore size) was used (Etten-Leur, The Netherlands). Both column materials are silica-based amide functionalized stationary phases. For RPLC, a Waters Acquity UPLC Protein BEH C4 column was used (50 \times 2.1 mm i.d.; 1.7 μm particles with 300 Å pore size). A Phenomenex SecurityGuard Ultra Cartridge (Widopore C4; 2 \times 4.6 mm i.d.) (Utrecht, The Netherlands) was installed before the analytical columns.

HILIC separations were performed using a mobile phase composed of solvent A (98% ACN, 2% water, 0.1% TFA) and solvent B (88% water, 10% IPA, 2% ACN, 0.1% TFA). For the analysis of the protein solutions the linear gradient was programmed as 20% B to 50% B in 10 or 30 min, followed by a cleaning step from 90% B to 10% B in 1 min, repeated three times, and final column equilibration at 20% B for 15 min. The flow rate and column temperature were 0.2 mL/min and 60 °C, respectively. The dwell volume of the applied system was 0.31 mL, and the hold-up volume was 0.26 mL. The HILIC linear scouting gradients for the glycoform separation were 10% B to 50% B in 15, 30 or 60 min. The hold-up volume, in this case, was 0.38 mL. For RPLC the flow rate, column temperature, and mobile phase solvents were the same as used for HILIC. The RPLC linear gradient went from 95% B to 40% B in 10 or 30 min, followed by a cleaning step from 10% B to 90% B in 1 min, repeated three times, and final column equilibration at 95% B for 10 min.

The retention times were determined using Openlab CDS Chem Station C.0107SR1. The software PIOTR (version 1.27) was installed on a standard PC to optimize the methods. A detailed description of the procedure to import data in PIOTR, the equation used for different models and the selection of optimized conditions is reported in Section S.3 and 4.

2.3. Mass spectrometry

For mass spectrometric (MS) detection of IdeS-digested trastuzumab, a Bruker Daltonics maXis HD high-resolution quadrupole time-of-flight (qTOF) mass spectrometer (Bremen,

Germany) was used, operating in positive-ion mode. The nebulizer was set at 0.8 bar, the dry gas at 8 L/min and the dry temperature of the nitrogen at 220 °C. The quadrupole ion and collision cell energies were 5 and 10 eV, respectively. The collision cell RF was 2000 Vpp. The in-source CID (isCID) was 120 eV. The funnel RF was set to 400 Vpp, and the multipole RF to 800 Vpp. The transfer and pre-pulse storage times were set at 190.0 and 20.0 μs , respectively. The monitored mass range was 600–5000 m/z. Data analysis was done using Compass data analysis (4.3) from Bruker and the charge state deconvolution using the Maximum Entropy algorithm.

2.4. Calculations

In Fig. 1a the retention times of both chromatographic modes were normalized using Equation 1, where RT_i is the retention time of the analyte (min), RT_{min} is the retention time (min) of the first eluting protein, and RT_{max} is the retention time (min) of the last eluting protein of the separation considered.

$$\text{normalized retention time} = \frac{RT_i - RT_{min}}{RT_{max} - RT_{min}} \quad (1)$$

The effective peak capacity (n_c) was calculated using Eq. (2), where $t_{G\text{eff}}$ is the effective window of the gradient (time of last eluting peak minus time of first eluting peak), and $\bar{w}_{1/2h}$ is the average peak width at half height.

$$n_c = \frac{t_{G\text{eff}}}{1.7 * \bar{w}_{1/2h}} + 1 \quad (2)$$

The gradient retention factor (k^*) was calculated using Eq. (3) where t_G is the gradient time programmed, F is the flow rate (mL/min), V_m is the hold-up volume (mL), S represents the change in $\ln k$ with increasing elution strength of the mobile phase (constant for a given solute), and $\Delta\varphi$ is the gradient range, i.e., the difference between fraction B at the start and the end of the gradient (e.g. if the gradient goes from 10 to 60% B, $\Delta\varphi$ is 0.5).

$$k^* = \frac{t_G F}{1.15 V_m * \Delta\varphi S} \quad (3)$$

3. Results and discussion

3.1. Protein selectivity and peak capacity of amide-based HILIC and C4-RPLC

The amide-HILIC and C4-RPLC retention of fifteen intact model proteins, covering a wide range of molecular weights and theoretical isoelectric points, was investigated. The test set included glycosylated proteins (AGP, fet, ova, RnB, trans, and thyro) and non-glycosylated proteins (BSA, CA, c.tryp, cyt C, lys, myo, RnA, tryp, and ubi). The HILIC and RPLC columns had the same dimensions and particle characteristics, and the same solvents A and B were used for both separation approaches. Protein samples were prepared in water and analyzed using linear gradients from 20% B to 50% B in 30 min and from 95% B to 40% B in 30 min when using HILIC and RPLC, respectively (see Section 2 for experimental details). These initial and final percentages of mobile phase B were chosen to allow elution of the model proteins within the gradient time, providing analysis methods with similar gradient volumes. The steepness and width of the gradient ($\Delta\%B$) needed for elution of the test proteins are different for HILIC and RPLC. HILIC separations were obtained using more shallow gradients ($\Delta\%B$, 30%) than RPLC ($\Delta\%B$, 55%). When comparing the HILIC and RPLC chromatograms of individual proteins, differences in retention order and selectivity are observed (see Fig. 1a–c). To further assess the orthogonality [29] of the two separation methods, the normalized retention times of the test proteins were calculated with Eq. (1). The obtained nor-

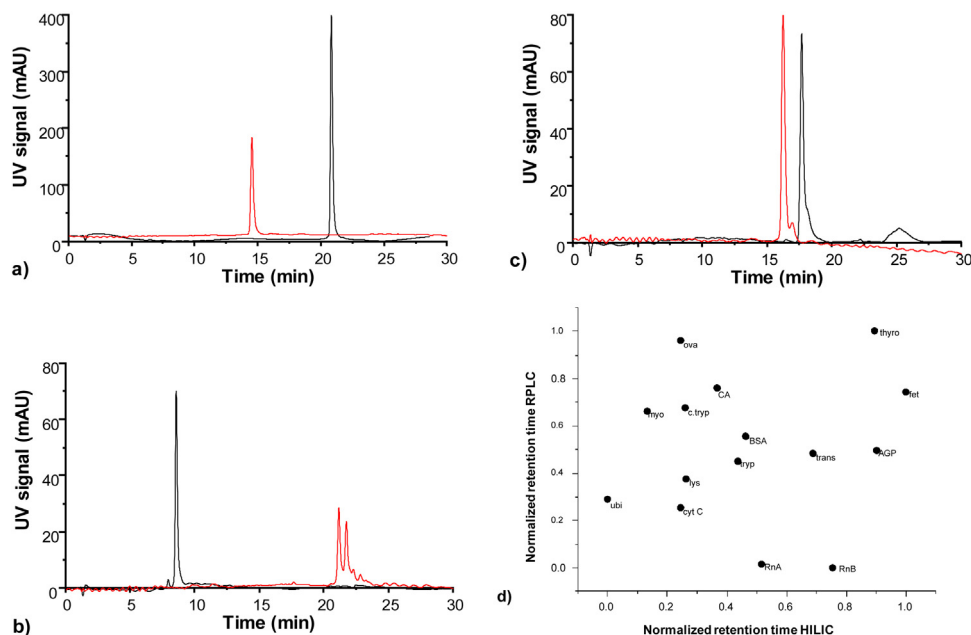


Fig. 1. (a–c) HILIC chromatograms (red) and RPLC chromatograms (black) of (a) CA (b) RnB, and (c) BSA. (d) normalized retention times of test proteins obtained during C4-RPLC and amide-based HILIC. The normalized data points are labeled with the abbreviation of the corresponding protein. For C4-RPLC analysis the linear gradient was from 5% to 60% A in 30 min, and for HILIC analysis the linear gradient was from 10% to 50% B in 30 min. The flow rate, column temperature, and absorbance detection wavelength were 0.2 mL/min, 60 °C, and at 214 nm, respectively. An overview of the chromatographic data obtained for the proteins analyzed is reported in S.1, Table S1. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

malized retention times with HILIC were plotted against the ones from RPLC (Fig. 1d). The individual proteins are scattered over the plot, indicating uncorrelated elution properties. This suggests that coupling HILIC to RPLC (i.e., two-dimensional LC) may represent an attractive option for increasing the peak capacity of LC(-MS)-based methods for the separation of complex protein mixtures (e.g., for top-down analysis of cell lysates [30,31]). This two-dimensional column coupling has been successfully applied to various types of sample including the study of the (micro)heterogeneity of single proteins or protein groups [31,32] as well as scorpion venom [33] Ginseng extract [34] lipids [35,36] and surfactants [37,38].

To further investigate the chromatographic behavior of the 15 proteins in HILIC and RPLC, the influence of the amino acid composition on the retention was examined. For both HILIC and RPLC, the amino acid composition of the protein influences retention as shown by the different elution times of the non-glycosylated proteins investigated (S.1, Table S1).

The retention of proteins in HILIC separations using amide stationary phases and TFA is thought to be based mainly on the overall polarity of the proteins. However, we did not observe a clear trend when looking at the protein elution order in HILIC and the number of polar amino acids. Moreover, no correlation was observed for RPLC between observed protein retention and the relative number of non-polar amino acids (Section S.1, Tables S2, and S3).

The HILIC and RPLC analyses were performed with the same column and stationary phase dimensions and mobile phase solvents, allowing the direct comparison of the average peak width at half height ($\bar{w}_{1/2h}$) and the peak capacity (Eq. (2)) of the two LC modes. The $\bar{w}_{1/2h}$ for the peaks of the measured proteins in RPLC was 0.37 min, resulting in a peak capacity of 31 for an effective gradient window (i.e., the gradient time in which the separation occurs) of 18 min. For HILIC, the $\bar{w}_{1/2h}$ for the peaks of the test proteins was 0.61 min, (effective gradient window, 18 min) resulting in a peak capacity of 18. Overall, the peaks in HILIC were broader than in RPLC. In particular, glycoproteins, such as AGP, fetuin, ovalbumin, RNase B, and thyroglobulin, showed broader peaks in HILIC, possibly due to partial separation of their glycoforms under non-

optimized gradient conditions that could not be distinguished using UV detection. For instance, the glycoforms of RNase B were partially separated in HILIC, while with RPLC no separation was obtained (Fig. 1b). Non-glycosylated proteins showed similar average width at half height, i.e., 0.24 min for RPLC and 0.28 min for HILIC (Section S.1, Table S1). MS data analysis using extracted-ion chromatograms (EICs) to reveal single glycoforms, confirmed the selectivity of HILIC toward glycosylation (results shown in section S6 of the supporting information). Yet, when UV absorbance detection is performed, the different proteoforms are not distinguished and therefore give overall broader peak profiles. Further evidence for this is provided in the discussion of the HILIC-MS results (Section 3.5).

3.2. LSS modeling of gradient elution of intact proteins in amide-based HILIC and C4-RPLC

The retention of individual test proteins was studied applying two linear gradients of 10 and 30 min in both HILIC (10% to 50% B) and RPLC (5% to 60% A). To be able to model the separation conditions, the software requires to carefully determine the system parameters: flow rate and initial/final mobile phase composition (percentage solvent B), the dwell volume and the hold-up volume. The hold-up volume was determined by HILIC analysis of ubiquitin under non-retaining conditions using an isocratic mobile phase of ACN-water (50:50, v/v). The dwell volume was calculated from the gradient delay for gradients of different times without a column installed.

The retention times of the main peak of each protein were used to construct LSS plots. The plots were compared for both LC modes. The LSS model describes analyte retention in RPLC as a function of mobile phase composition, but it has also shown useful for other LC modes [39]. LSS presumes a linear relationship between the natural logarithm of the retention factor ($\ln k$) and the volume fraction (φ) of the strong solvent in a binary eluent (Eq. (4)). In this equation, k_0 is the extrapolated (not necessarily real) k of the analyte in pure weak solvent (i.e., φ equals 0) and S the slope of the plot rep-

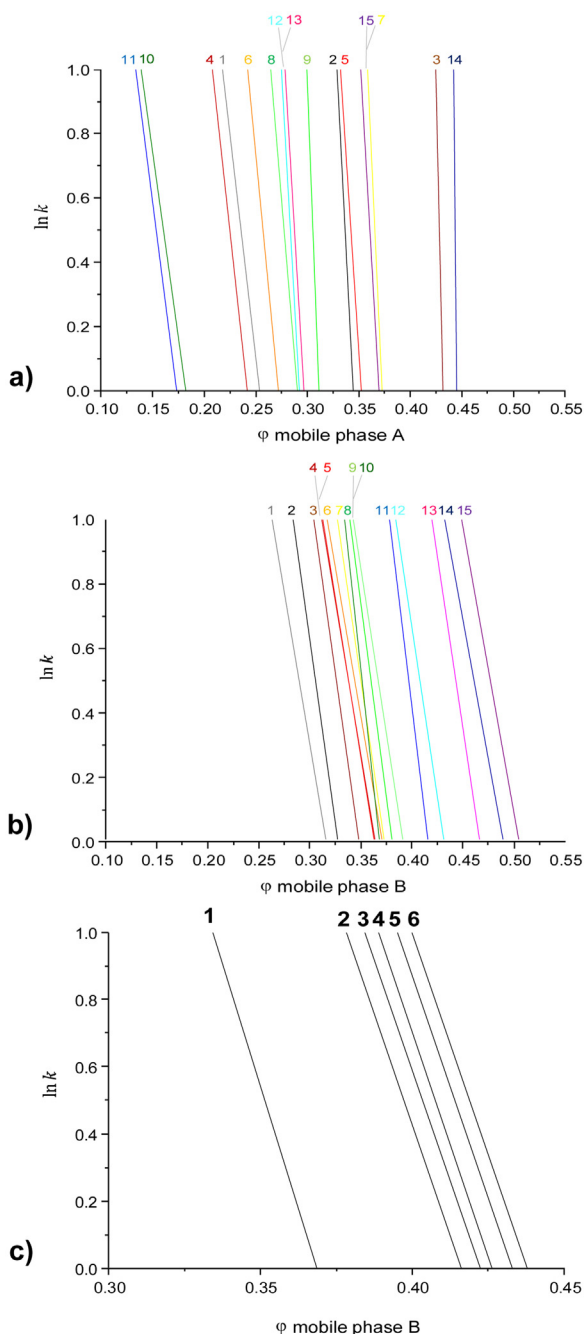


Fig. 2. $\ln k$ vs. ϕ plots for all test proteins using (a) C4-RPLC, and (b) HILIC; (c) $\ln k$ vs. ϕ plot for RnA (1) and RnB (2–6) using HILIC. The numbers in plot 2a and 2b correspond to the test proteins: 1 = ubi, 2 = myo, 3 = ova, 4 = cyt C, 5 = c.tryp, 6 = lys, 7 = CA, 8 = tryp, 9 = BSA, 10 = RnA, 11 = Trans, 12 = RnB, 13 = thyro, 14 = AGP, 15 = fet.

resenting the elution strength of the strong solvent for the analyte [40].

$$\ln k = \ln k_0 - S\phi \quad (4)$$

The protein gradient elution times were used to calculate the $\ln k_0$ and S for each protein applying the LSS retention model using the software PIOTR. These values were used to generate the $\ln k$ vs. ϕ plots for both RPLC (Fig. 2a) and HILIC (Fig. 2b). The model reliably described the elution of the proteins for both chromatographic modes as indicated by a low Akaike Information Criterion (AIC; results reported in Table S1, the significance of this parameter is described in the next chapter). As can be seen from Fig. 2, the slope S is relatively large, implying that for proteins, a relatively

small change in ϕ (%B) may strongly affect their retention [39]. For HILIC, the obtained slopes (S) are generally smaller (i.e., less steep curves) than for RPLC and appear to be only weakly correlated to the protein molecular weight. For example, the S value of thyroglobulin (having a molecular weight of about 660 kDa) is 160 in RPLC and only 16 in HILIC. Another considerable difference between the two LC modes is the spread of the x -intercepts (range of ϕ values corresponding to $\ln k=0$) in the protein plots, which is considerably smaller for HILIC as compared to RPLC. This indicates that the separation of protein mixtures with HILIC may need more detailed optimization and requires more shallow gradients than in RPLC.

Fig. 2c shows the LSS plots for RnA and its corresponding glycoprotein RnB (single N-glycosylation site), which comprises five glycoforms differing in number (5–9) of mannose residues. The glycosylated RnB elutes at a higher water percentage than the non-glycosylated RnA. Moreover, the percentage water at which the RnB glycoforms elute increases with the size of the glycan (curves 2–6 in Fig. 2c), clearly showing that protein retention in amide-based HILIC depends on (the degree of) glycosylation. The slope (S) of the curves of the RnA and the RnB glycoforms are very similar, indicating that these proteins have a comparable gradient response behavior, probably because the proteins have the same backbone. Notably, with RPLC, the glycoforms of RnB were not separated, and protein glycosylation does not seem to influence protein retention significantly.

3.3. Computer-aided optimization of HILIC-UV methods for glycoproteins: evaluating retention models

Next, we investigated the possibility of performing computer-aided method development for the separation of glycoforms of proteins that so far had not been characterized by HILIC (ova, fet, AGP, and thyro) using the software program PIOTR [23]. These glycoproteins have different sites of glycosylation (only N, or N and O glycosylation) and different glycosylation complexity. To be able to model the separation conditions, the software requires experimental retention data, system parameters, and a proper retention model. Each glycoprotein was analyzed by HILIC using several scouting gradients with different slopes and the obtained analyte retention times were imported in the software.

Pirok et al. [23] suggested using large gradient ranges for accurate modeling with PIOTR. In the present study, three HILIC scouting gradients were performed in triplicate for each test protein from 10 to 50% solvent B – in 15, 30, and 60 min. In our experience, these gradient times allow for the elution of a wide range of proteins (see Section 3.1). Under these gradient conditions, a number of the tested glycoproteins did not give a narrow symmetric peak, but rather broad bands comprised of partially separated glycoforms, which could not be differentiated reliably using UV absorbance detection, as exemplified for thyro in Fig. 3.

A detailed explanation of the peak picking and analysis process is provided in S.4. In general, if glycoform features could be distinguished during the longest gradient time ($t_G = 60$ min), these were chosen as retention times and also assigned in the shorter gradient times. These values were then used to model protein retention and optimize the separation. When this was not possible (i.e., the protein eluted as a single featureless band) the retention times of the peak at its maximum and at half height (front and tail) were measured, as shown for thyro (Fig. 3). Scouting results of the other proteins of interest (i.e., *IdeS*-digested trastuzumab and the intact glycoproteins AGP, fet, and ova) can be found in S.2 (Figures S1–4).

Five different retention models were compared: mixed-mode [41], Neue-Kuss [42], adsorption [43], LSS [40], and quadratic model [44]. The equation and parameters of the LSS model can be found in Section 3.2 (Eq. (4)). For the other models, the equations and parameters are stated in S.3 (Equation S2 to S5). The retention times

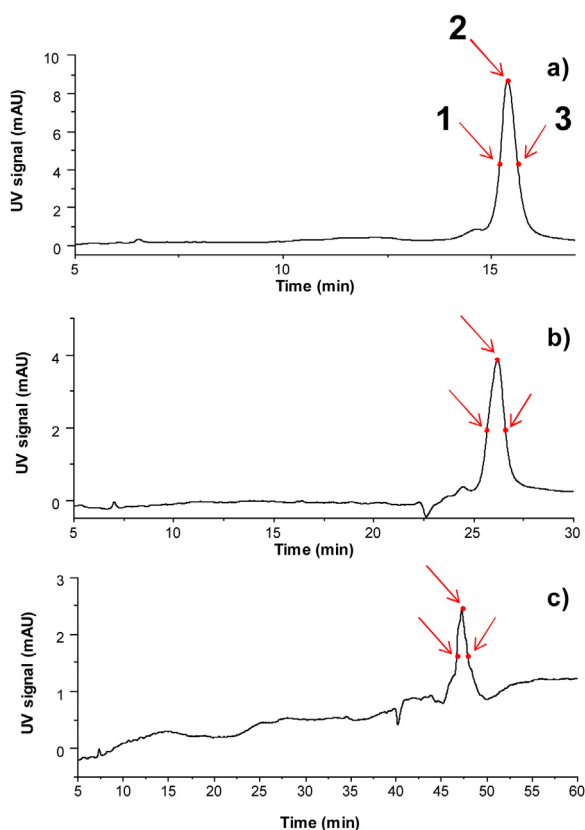


Fig. 3. HILIC-UV of thyro using a linear gradient from 10 to 50% B in (a) 15, (b) 30 and (c) 60 min. The red dots and arrows indicate the retention times corresponding to the peak maximum (2) and to the peak width at half height (front (1) and tail (3) of the peak) which were used for modeling and optimization by PIOTR. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

of the main features (minimum 3) of each protein band obtained under the different gradient conditions were used. The goodness-of-fit of the different models was determined by calculating the AIC values, which describe the quality of fit of the selected model and the given experimental dataset, relative to the other models used [43,45]. The AIC allows comparison of models that use a different number of parameters. This is convenient for the present study as the LSS and the adsorption model employ two parameters, whereas the quadratic, Neue-Kuss and the mixed-mode model comprise three parameters. For AIC calculation, Eq. (5) was used, where p is the number of parameters of the model, n is the number of analyses, and SSE is the sum-of-squares error from the retention times. Three different gradient analyses (15, 30 and 60 min) were performed, each in triplicate ($n=9$). The lower the AIC, the better the model describes the retention of analytes [28].

$$AIC = 2p + n \left[\ln \left(\frac{2\pi * SSE}{n} \right) + 1 \right] \quad (5)$$

For each model, the AICs obtained for the test proteins were binned in five ranges of which the frequency (N) was plotted (Fig. 4). Notably, our results regarding the modeling of protein HILIC retention align rather well with results of Pirok et al. obtained for small molecules and peptides [28]. Considering the specific chemico-physical properties of proteins as well as the characteristic mobile phase conditions for protein HILIC, this is not evident. Still, also for proteins, the adsorption model showed optimal to model and predict the retention of HILIC on amide stationary phases. Moreover, our results show that the AICs for glycoproteins are lower than the AICs found for small molecules [28], indicating a better fitting of the model in this study. This result can be at least partially ascribed to the fact that the relatively high percentage of water (20–50%) used for protein elution diminishes the stagnant water layer on the surface of the stationary phase, and thus minimizes the contribution of analyte partitioning to the retention.

The quadratic and the mixed-mode model also showed quite favorable AIC values. These models performed somewhat worse than the adsorption model but slightly better than the LSS model. The Neue-Kuss model performed poorly for most proteins, pro-

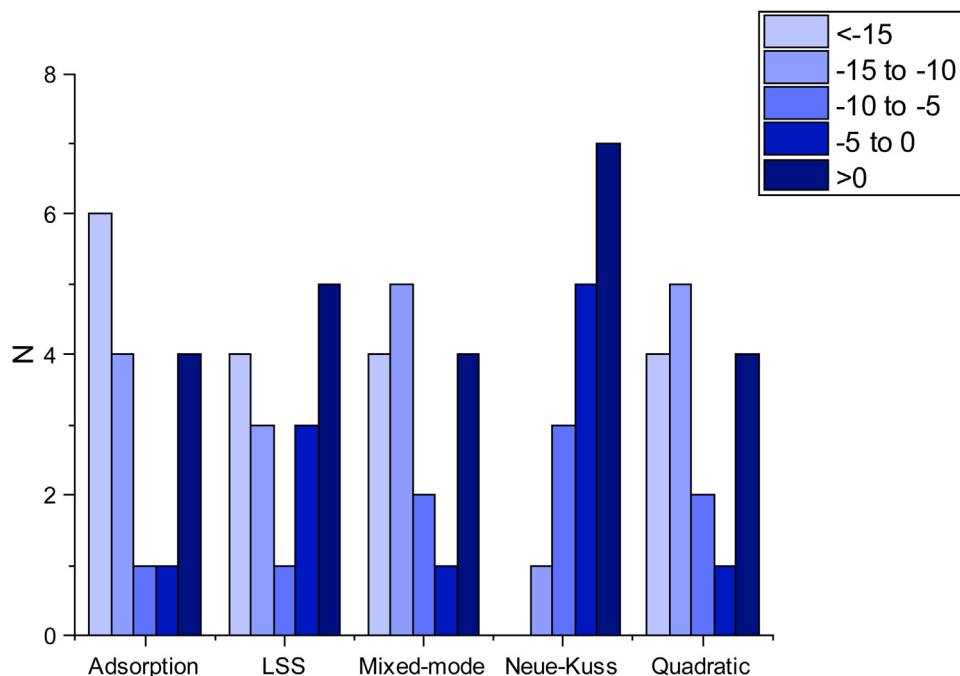


Fig. 4. Assessment of the retention models' goodness-of-fit based on the AIC values calculated for all peaks detected for the tested glycoproteins. AIC ranges were from light to dark blue: <-15, -15 to -10, -10 to -5, -5 to 0 and >0. See S.3 (Table S4) for the AIC values of each analyte.

viding no proper fits. A possible explanation is that this model is empirical and, therefore, needs many data points to make a reliable prediction. In the present study, only three points (gradient times) were used to make the model. Based on the results described above, the adsorption model was selected as the retention model for the computer-aided method optimization for the separation of the glycoproteins. The results of the parameters calculated for each retention model are reported in Table S4.

3.4. Computer-aided optimization of HILIC gradient conditions for glycoproteins

Using the adsorption model to describe protein retention, we studied the effect of the following parameters: the starting percentage solvent B (φ_{init}), gradient time (up to 60 min), and final percentage solvent B (φ_{final}). For each factor, we selected a range (i.e., the starting and final value) and the number of increments (steps) to take into account during the calculations. First, broad ranges for solvent B with a low number of steps were chosen to get an indication of the optimal conditions. In this case, the φ_{init} was from 0.20 to 0.45 in 10 steps of 2.5% B, the φ_{final} was from 0.25 to 0.50 in 10 steps of 2.5% B, and the gradient time was from 15 to 60 min in 45 steps of 1 min. Thereafter, the ranges were further specified per protein to achieve shallower gradients. The specified ranges of each protein can be found in S.4 (Table S5). Then PIOTR calculated the results for all possible methods within those ranges using a Pareto-optimization approach. For the Pareto optimization, all possible combinations of factors were plotted considering two chosen objectives (i.e., gradient time, last eluted peak, φ_{init} , φ_{final} or resolution of the predicted separation). As an example, in S.4 (Figure S5), the Pareto-plots of thyro are depicted.

Finally, the optimized gradient selecting points within the Pareto-optimal conditions was selected. A condition is Pareto-optimal when it is not possible to improve one of the objectives without making the other one worse, which results in a Pareto front that represents the performance limit within the specified constraints [46]. In the present study, the resolution score of the predicted separation was selected as an important objective. To calculate the resolution score, a procedure as described in [23] was used. The resolution of each predicted peak with all other peaks was calculated. The obtained resolutions were normalized between 0 and 1, where a score of 1 means a minimum predicted resolution of 1.5 between two peaks and 0 means complete overlap. Lastly, the resolution scores of all the peak pairs were multiplied, resulting in a measure of the overall predicted separation power. Of all the solutions reported, the one having the highest value of resolution (i.e., resolution score is 1) was selected meeting the following criteria: the peaks eluted within the gradient time, have the lowest gradient time in the interval between 40 and 55 min, and a maximum total analysis time of 70 min. A detailed description of the selection procedure can be found in Section S.4.

The described approach was first evaluated for the separation of the Fc glycoforms of *IdeS*-digested trastuzumab, which contains a conserved N-glycosylation site on the Fc part of each heavy chain [47]. The digestion with *IdeS* results in three fragments: $F(ab)_2$ of about 100 kDa and two *Fc/2* fragments of approximately 25 kDa [7]. Fig. 5a shows the UV chromatogram of the optimized method (26.5 to 34.0% B in 53 min). Starting from general elution conditions and using only three scouting gradients, our approach established an optimal gradient slope of 0.14%/min. The obtained method is in agreement with the one used by D'Atri et al. [7] for HILIC of *IdeS*-digested monoclonal antibodies. Furthermore, the retention times for the *Fc/2* glycoform peaks (corresponding to the peaks between 24 and 31 min in Fig. 5a) were accurately predicted by the adsorption model (error below 1 min in a gradient time of 53 min (S.4, Tables S6 and S7).

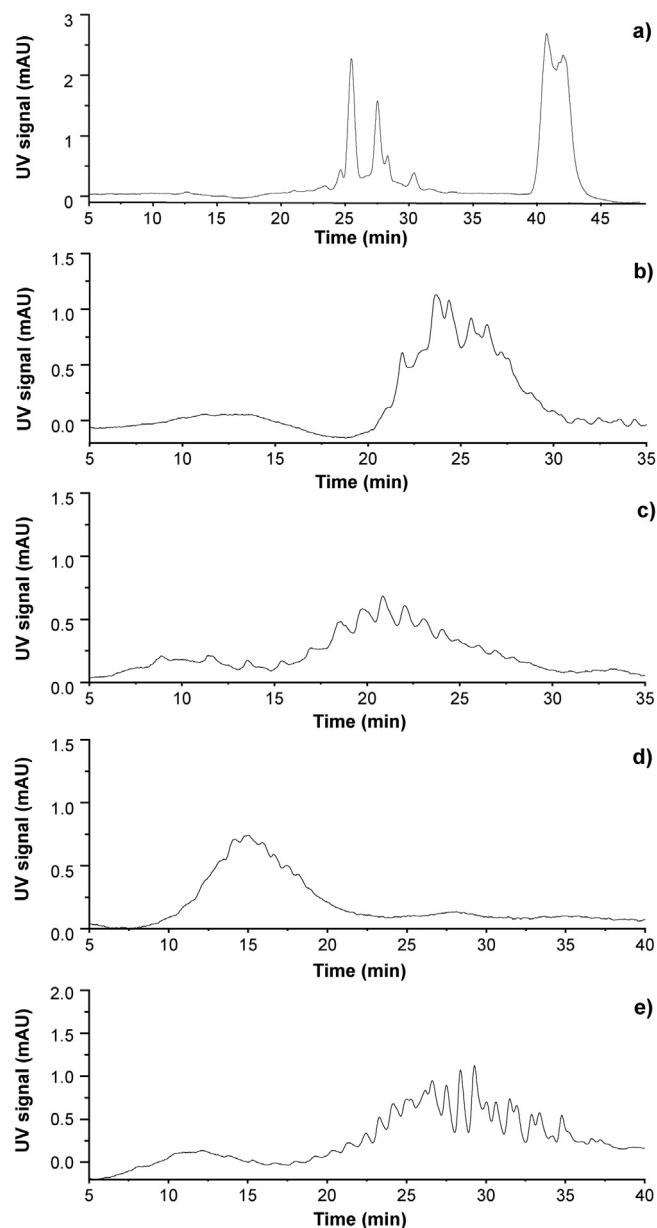


Fig. 5. HILIC-UV chromatograms obtained with the PIOTR-optimized methods for (a) *IdeS*-digested trastuzumab (gradient, 26.5 to 34.0% B in 53 min), (b) ova (gradient, 22.3 to 29.5% B in 45 min), (c) fet (gradient, 33.0 to 42.0% B in 53 min), (d) AGP (gradient, 40.0 to 47.7% B in 41 min), and (e) thyro (gradient 33.5 to 40.5% B in 44 min). Flow rate, column temperature, and UV detection wavelength were 0.2 mL/min, 60 °C, and 280 nm, respectively. Injected protein concentration, 2 mg/mL each. The scouting gradients of these proteins can be found in Fig. 3 and Figs. S1–S5.

To express the steepness of a method we calculated the gradient retention factors (k^*) and compared its value for scouting gradients and the optimal method. The parameter k^* is the median value of k during gradient elution (i.e., the k when the analyte band has reached the middle of the column) and can be calculated with Eq. (3). For optimal resolution, k^* should be between 1 and 10 [39]. The calculated k^* values are listed in S.5 (Table S8). On average, the k^* of the general gradients were around 0.5, 1 or 2 for gradient times of 15, 30 or 60 min, respectively. The gradient conditions of the optimized method for the separation of the *IdeS*-digested Trastuzumab correspond to a k^* of 8.2, showing the importance of shallow gradients to enable efficient separation of protein glycoforms.

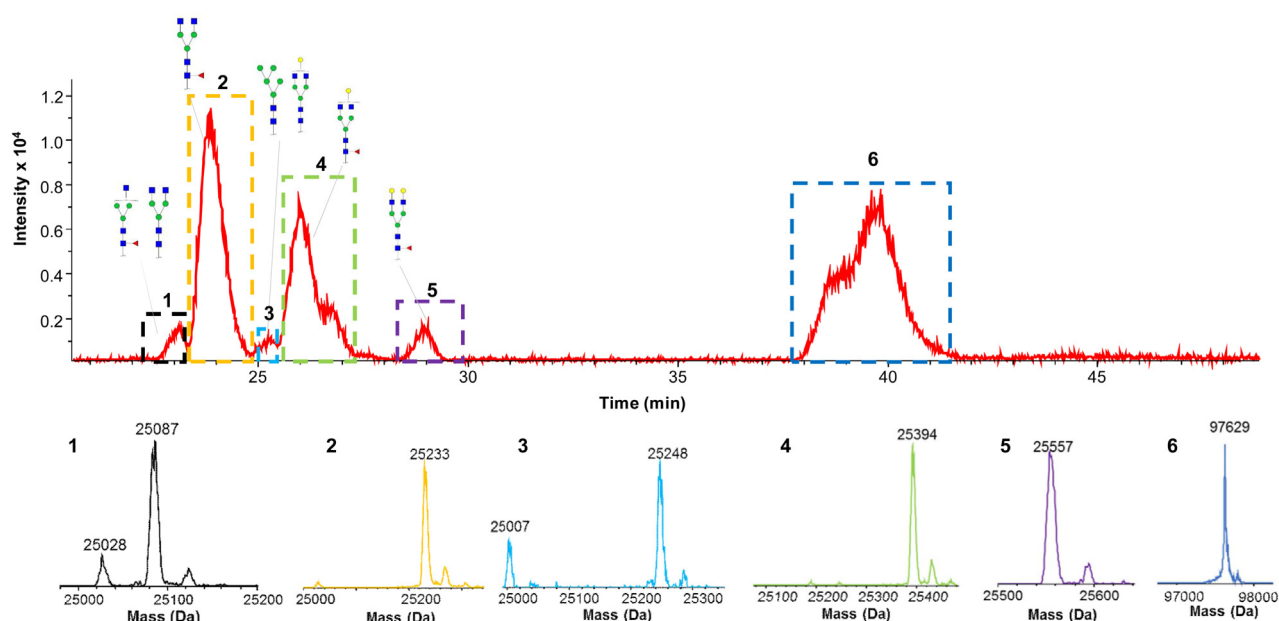


Fig. 6. HILIC-MS of IdeS-digested trastuzumab (1 $\mu\text{g}/\mu\text{L}$). Base peak chromatogram including the proposed glycan structures and deconvoluted mass spectra of the fragments (indicated by number 1–6) (1–5) $Fc/2$ fragments and (6) $F(ab)_2$ fragment. The linear gradient was from 26.5 to 34.0% B in 53 min. The flow rate and column temperature were 0.2 mL/min and 60 °C, respectively. Injection volume was 2 μL .

Next, PIOTR was used to predict optimal HILIC gradient conditions for separating glycoforms of intact proteins of increasing complexity: ova, fet, AGP, and thyro. Ovalbumin is a 45 kDa-glycoprotein from chicken egg white and has one N-glycosylation site [48]. Yang et al. identified 45 glycoforms using native MS. Bovine fetuin (42 kDa) has three N-glycosylation and two O-glycosylation sites [49]. AGP is a 41 kDa glycoprotein of which the glycan content represents 45% of the molecular weight, including highly sialylated complex-type N-glycans [50]. Imre et al. identified 80 different AGP-derived glycopeptides using MS(/MS) [51]. Bovine thyroglobulin is a dimeric glycoprotein of approximately 660 kDa and one of the largest glycoproteins known. Rawitch et al. showed that bovine thyroglobulin has thirteen N-glycosylation sites. Nine of these sites are complex or hybrid type glycans, and the other four are oligomannose-type. Besides N-glycosylation, also phosphorylation and sulfation sites occur [52].

Fig. 5b–e shows the HILIC-UV chromatograms using the optimal methods for the analyzed glycoproteins as proposed by PIOTR based on three general scouting gradients. The chromatograms clearly show a multitude of features. For all proteins, shallow gradients with an overall change of only 7–9% in solvent B over a time of 40–53 min were predicted (0.13 to 0.22%B/min) with k^* between 7.8 and 15.6. Under these conditions, the peak of glycoproteins that appeared only as broad peaks in the scouting gradients were resolved into profiles with distinct features applying the predicted optimal gradients.

3.5. Assignment of glycoforms of IdeS-digested trastuzumab using HILIC-MS

In the previous section, we assumed that based on the UV data, the different glycoforms were resolved by the optimized methods. To confirm that the observed peaks of IdeS-digested trastuzumab (Fig. 5a) indeed are different glycoforms, we also analyzed the sample with HILIC-MS using the same HILIC conditions (Fig. 6). The glycosylation of IgG class therapeutic monoclonal antibodies is well characterized with glycans comprised of galactose or mannose (H), N-acetyl glucosamine (N), and fucose (F). The base peak chromatogram (BPC) obtained with HILIC-MS (including pro-

posed glycan structures) is depicted in Fig. 6. The deconvoluted mass spectra of the fragments are indicated with 1–6. The first five peaks corresponded to the different glycoforms of the $Fc/2$ part and the last peak to the $F(ab)_2$ part. MS-based assignment of the glycoforms indicated that neutral glycan units significantly contribute to glycoform separation. The two most abundant glycoforms (Fig. 6, deconvoluted spectra 2 and 4) correspond to H3N4F1 and H4N4F1. The peaks with lower intensity could be assigned to H3N3F1, H5N4F1, H3N4, H4N4, and H5N2 glycoforms. Extracted ion chromatograms of the glycoforms described in Fig. 6 as well as for ova are reported in Section S6 of the supporting information. The observed glycoforms have approximately the same peak widths, but distinct elution times. This explains the broadened peaks observed for the glycoproteins during HILIC-UV. The HILIC-MS analysis of ova revealed several protein masses. However, because of the simultaneous presence of sequence variants and numerous proteoforms, the assignment of the masses observed was not trivial and not further attempted in this study. In order to aid assignment of the glycoforms observed, released glycan studies and/or bottom-up characterization of the protein could be performed.

For fet, AGP, and thyro, no satisfactory MS results were obtained. The presence of TFA in the mobile phase and the relatively high molecular weight (and thus distribution into multiple charge states) of the proteins probably hindered an appreciable MS response of these proteins. Optimization of LC and MS conditions allowing the characterization of the proteoforms of ova, fet, AGP, and thyro is currently under investigation.

4. Conclusions

We have investigated the quality-of-fit for five retention models applied to the modeling of the retention behavior of glycoproteins in HILIC using amide stationary phases and TFA based mobile phases. The adsorption model demonstrated robust performance in terms of its ability to describe HILIC retention of glycoproteins using three gradient times having a wide solvent composition.

We used the gradient elution modeling as the strategy to rapidly obtain shallow gradient conditions that allow for the resolution of

glycoform of glycoproteins. This is demonstrated by the separation of proteins with high degree and complexity of glycosylation. Features of ovalbumin, fetuin, AGP, and thyroglobulin (proteins that were not previously studied using HILIC) were resolved using shallow gradients (overall change of 7–9% B over gradient times up to 1 h) calculated using the computer-aided method development described here.

Acknowledgments

This work was financially supported by The Netherlands Organization for Scientific Research by the NWO-VENI grant IPA (722.015.009).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.chroma.2019.03.038>.

References

- [1] E. Higgins, Carbohydrate analysis throughout the development of a protein therapeutic, *Glycoconj. J.* 27 (2010) 211–225, <http://dx.doi.org/10.1007/s10719-009-9261-x>.
- [2] A. Planinc, J. Bones, B. Dejaeger, P. Van Antwerpen, C. Delporte, Glycan characterization of biopharmaceuticals: updates and perspectives, *Anal. Chim. Acta* 921 (2016) 13–27, <http://dx.doi.org/10.1016/j.aca.2016.03.049>.
- [3] A. Varki, R.D. Cummings, J. Esko, H. Freeze, P. Stanley, C. Bertozzi, G. Hart, M. Etzler, *Essentials of Glycobiology*, 3rd edition, 1999, Chapter 9, Available at: <https://www.ncbi.nlm.nih.gov/books/NBK310274/>.
- [4] H. Geyer, R. Geyer, Strategies for analysis of glycoprotein glycosylation, *Biochim. Biophys. Acta – Proteins Proteomics* 1764 (2006) 1853–1869, <http://dx.doi.org/10.1016/j.bbapap.2006.10.007>.
- [5] H.J. An, T.R. Peavy, J.L. Hedrick, C.B. Lebrilla, Determination of N-glycosylation sites and site heterogeneity in glycoproteins, *Anal. Chem.* 75 (2003) 5628–5637, <http://dx.doi.org/10.1021/ac034414x>.
- [6] M. Wührer, C.A.M. Koeleman, C.H. Hokke, A.M. Deelder, Protein glycosylation analyzed by normal-phase nano-liquid chromatography–mass spectrometry of glycopeptides, *Anal. Chem.* 77 (2005) 886–894, <http://dx.doi.org/10.1021/ac048619x>.
- [7] V. D'Atri, S. Fekete, A. Beck, M. Lauber, D. Guillaume, Hydrophilic interaction chromatography hyphenated with mass spectrometry: a powerful analytical tool for the comparison of originator and biosimilar therapeutic monoclonal antibodies at the middle-up level of analysis, *Anal. Chem.* 89 (2017) 2086–2092, <http://dx.doi.org/10.1021/acs.analchem.6b04726>.
- [8] E. Balaguer, C. Neusüss, Glycoprotein characterization combining intact protein and glycan analysis by capillary electrophoresis–electrospray ionization–mass spectrometry, *Anal. Chem.* 78 (2006) 5384–5393, <http://dx.doi.org/10.1021/ac060376g>.
- [9] A. Pedrali, S. Tengattini, G. Marrubini, T. Bavaro, P. Hemström, G. Massolini, M. Terreni, C. Temporini, Characterization of intact neo-glycoproteins by hydrophilic interaction liquid chromatography, *Molecules* 19 (2014) 9070–9088, <http://dx.doi.org/10.3390/molecules19079070>.
- [10] O.S. Skinner, N.A. Haverland, L. Fornelli, R.D. Melani, L.H.F. Do Vale, H.S. Seckler, P.F. Doubleday, L.F. Schachner, K. Srzentić, N.L. Kelleher, P.D. Compton, Top-down characterization of endogenous protein complexes with native proteomics, *Nat. Chem. Biol.* 14 (2018) 36–41, <http://dx.doi.org/10.1038/nchembio.2515>.
- [11] E. Domínguez-Vega, S. Tengattini, C. Peintner, J. van Angeren, C. Temporini, R. Haselberg, G. Massolini, G.W. Somsen, High-resolution glycoform profiling of intact therapeutic proteins by hydrophilic interaction chromatography–mass spectrometry, *Talanta* 184 (2018) 375–381, <http://dx.doi.org/10.1016/j.talanta.2018.03.015>.
- [12] G.W. Somsen, C. Temporini, G. Massolini, L. Pollegioni, E. Domínguez-Vega, F. Rinaldi, S. Tengattini, L. Piubelli, T. Bavaro, Hydrophilic interaction liquid chromatography–mass spectrometry as a new tool for the characterization of intact semi-synthetic glycoproteins, *Anal. Chim. Acta* 981 (2017) 94–105, <http://dx.doi.org/10.1016/j.aca.2017.05.020>.
- [13] T.K. Toby, L. Fornelli, N.L. Kelleher, Progress in top-down proteomics and the analysis of proteoforms, *Annu. Rev. Anal. Chem.* 9 (2016) 499–519, <http://dx.doi.org/10.1146/annurev-anchem-071015-041550>.
- [14] M. Wührer, A.M. Deelder, C.H. Hokke, Protein glycosylation analysis by liquid chromatography–mass spectrometry, *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 825 (2005) 124–133, <http://dx.doi.org/10.1016/j.jchromb.2005.01.030>.
- [15] B. Buszewski, S. Noga, Hydrophilic interaction liquid chromatography (HILIC) – a powerful separation technique, *Anal. Bioanal. Chem.* 402 (2012) 231–247, <http://dx.doi.org/10.1007/s00216-011-5308-5>.
- [16] P.J. Boersema, S. Mohammed, A.J.R. Heck, Hydrophilic interaction liquid chromatography (HILIC) in proteomics, *Anal. Bioanal. Chem.* 391 (2008) 151–159, <http://dx.doi.org/10.1007/s00216-008-1865-7>.
- [17] J. Carroll, I.M. Fearnley, J.E. Walker, Definition of the mitochondrial proteome by measurement of molecular masses of membrane proteins, *Proc. Natl. Acad. Sci.* 103 (2006) 16170–16175, <http://dx.doi.org/10.1073/pnas.0607719103>.
- [18] Z. Tian, N. Tolić, R. Zhao, R.J. Moore, S.M. Hengel, E.W. Robinson, D.L. Stenoien, S. Wu, R.D. Smith, L. Paša-Tolić, N. Toli, R. Zhao, R.J. Moore, S.M. Hengel, E.W. Robinson, D.L. Stenoien, S. Wu, R.D. Smith, L. Pa, Enhanced top-down characterization of histone post-translational modifications, *Genome Biol.* 13 (2012) R86, <http://dx.doi.org/10.1186/gb-2012-13-10-r86>.
- [19] A.F.G. Gargano, L.S. Roca, R.T. Fellers, M. Bocxe, E. Domínguez-Vega, G.W. Somsen, Capillary HILIC–MS: a new tool for sensitive top-down proteomics, *Anal. Chem.* 90 (2018) 6601–6609, <http://dx.doi.org/10.1021/acs.analchem.8b00382>.
- [20] Z. Zhang, Z. Wu, M.J. Wirth, Polyacrylamide brush layer for hydrophilic interaction liquid chromatography of intact glycoproteins, *J. Chromatogr. A* 1301 (2013) 156–161, <http://dx.doi.org/10.1016/j.chroma.2013.05.076>.
- [21] A. Periat, S. Fekete, A. Cusumano, J.-L. Veuthey, A. Beck, M. Lauber, D. Guillaume, Potential of hydrophilic interaction chromatography for the analytical characterization of protein biopharmaceuticals, *J. Chromatogr. A* 1448 (2016) 81–92, <http://dx.doi.org/10.1016/j.chroma.2016.04.056>.
- [22] L.R. Snyder, J.W. Dolan, D.C. Lommen, Drylab[®] computer simulation for high-performance liquid chromatographic method development. I. Isocratic elution, *J. Chromatogr. A* 485 (1989) 65–89, [http://dx.doi.org/10.1016/S0021-9673\(01\)89133-0](http://dx.doi.org/10.1016/S0021-9673(01)89133-0).
- [23] B.W.J. Pirok, S. Pous-Torres, C. Ortiz-Bolsico, G. Viv??-Truyols, P.J. Schoenmakers, Program for the interpretive optimization of two-dimensional resolution, *J. Chromatogr. A* 1450 (2016) 29–37, <http://dx.doi.org/10.1016/j.chroma.2016.04.061>.
- [24] T. Baczek, R. Kaliszczan, H.A. Claessens, M.A. van Straten, Computer-assisted optimization of reversed-phase HPLC isocratic separations of neutral compounds Tomasz, *LC GC Eur.* (2001) 2–6.
- [25] E. Tyteca, J.L. Veuthey, G. Desmet, D. Guillaume, S. Fekete, Computer assisted liquid chromatographic method development for the separation of therapeutic proteins, *Analyst* 141 (2016) 5488–5501, <http://dx.doi.org/10.1039/c6an01520d>.
- [26] B. Bobály, V. D'Atri, A. Beck, D. Guillaume, S. Fekete, Analysis of recombinant monoclonal antibodies in hydrophilic interaction chromatography: a generic method development approach, *J. Pharm. Biomed. Anal.* 145 (2017) 24–32, <http://dx.doi.org/10.1016/j.jpba.2017.06.016>.
- [27] E. Tyteca, J. De Vos, N. Vankova, P. Cesla, G. Desmet, S. Eeltink, Applicability of linear and nonlinear retention-time models for reversed-phase liquid chromatography separations of small molecules, peptides, and intact proteins, *J. Sep. Sci.* 39 (2016) 1249–1257, <http://dx.doi.org/10.1002/jssc.201501395>.
- [28] B.W.J. Pirok, S.R.A. Molenaar, R.E. van Outersterp, P.J. Schoenmakers, Applicability of retention modelling in hydrophilic-interaction liquid chromatography for algorithmic optimization programs with gradient-scanning techniques, *J. Chromatogr. A* (2017), <http://dx.doi.org/10.1016/j.chroma.2017.11.017>.
- [29] M. Gilar, P. Olivova, A.E. Daly, J.C. Gebler, Orthogonality of separation in two-dimensional liquid chromatography, *Anal. Chem.* 77 (2005) 6426–6434, <http://dx.doi.org/10.1021/ac050923i>.
- [30] J.C. Tran, L. Zamborg, D.R. Ahlf, J.E. Lee, A.D. Catherman, K.R. Durbin, J.D. Tipton, A. Vellaichamy, J.F. Kellie, M. Li, C. Wu, S.M.M. Sweet, B.P. Early, N. Siuti, R.D. LeDuc, P.D. Compton, P.M. Thomas, N.L. Kelleher, Mapping intact protein isoforms in discovery mode using top-down proteomics, *Nature* 480 (2011) 254–258, <http://dx.doi.org/10.1038/nature10575>.
- [31] A.F.G. Gargano, J.B. Shaw, M. Zhou, C.S. Wilkins, T.L. Fillmore, R.J. Moore, G.W. Somsen, L. Paša-Tolić, Increasing the separation capacity of intact histone proteoforms chromatography coupling online weak cation exchange–HILIC to reversed phase LC UVPD–HRMS, *J. Proteome Res.* (2018), <http://dx.doi.org/10.1021/acs.jproteome.8b00458>.
- [32] D.R. Stoll, D.C. Harnes, G.O. Staples, O.G. Potter, C.T. Dammann, D. Guillaume, A. Beck, Development of comprehensive online two-dimensional liquid chromatography/mass spectrometry using hydrophilic interaction and reversed-phase separations for rapid and deep profiling of therapeutic antibodies, *Anal. Chem.* 90 (2018) 5923–5929, <http://dx.doi.org/10.1021/acs.analchem.8b00776>.
- [33] J. Xu, X. Zhang, Z. Guo, J. Yan, L. Yu, X. Li, X. Xue, X. Liang, Short-chain peptides identification of scorpion *Buthus tianshanensis* Karsch venom by employing high orthogonal 2D-HPLC system and tandem mass spectrometry, *Proteomics* (2012) 3076–3084.
- [34] Q. Xing, T. Liang, G. Shen, X. Wang, Y. Jin, X. Liang, Comprehensive HILIC × RPLC with mass spectrometry detection for the analysis of saponins in *Panax notoginseng*, *Analyst* 137 (2012), 2339–2249.
- [35] M. Holčápek, M. Ovčáčíková, M. Lisa, E. Cífková, T. Hájek, Continuous comprehensive two-dimensional liquid chromatography–electrospray ionization mass spectrometry of complex lipidomic samples, *Anal. Bioanal. Chem.* 407 (2015) 5033–5043, <http://dx.doi.org/10.1007/s00216-015-8528-2>.
- [36] A. Baglai, A.F.G. Gargano, J. Jordens, Y. Mengerink, M. Honing, S. van der Wal, P.J. Schoenmakers, Comprehensive lipidomic analysis of human plasma using multidimensional liquid- and gas-phase separations: two-dimensional liquid chromatography–mass spectrometry vs. liquid chromatography–trapped-ion-mobility–mass spectrometry, *J. Chromatogr. A* 1530 (2017) 90–103, <http://dx.doi.org/10.1016/j.chroma.2017.11.014>.

- [37] A.F.G. Gargano, M. Duffin, P. Navarro, P.J. Schoenmakers, Reducing dilution and analysis time in online comprehensive two-dimensional liquid chromatography by active modulation, *Anal. Chem.* 1 (2016) 1–16, <http://dx.doi.org/10.1021/acs.analchem.5b04051>.
- [38] G. Groeneveld, M.N. Dunkle, M. Rinken, A.F.G. Gargano, A. de Niet, M. Pursch, E.P.C. Mes, P.J. Schoenmakers, Characterization of complex polyether polyols using comprehensive two-dimensional liquid chromatography hyphenated to high-resolution mass spectrometry, *J. Chromatogr. A* 1569 (2018) 128–138, <http://dx.doi.org/10.1016/j.chroma.2018.07.054>.
- [39] L.R. Snyder, J.W. Dolan, High-Performance Gradient Elution, 2006, <http://dx.doi.org/10.1002/0470055529>.
- [40] L.R. Snyder, J.W. Dolan, J.R. Gant, Gradient elution in high-performance liquid chromatography. I. Theoretical basis for reversed-phase systems, *J. Chromatogr. A* 165 (1979) 3–30, [http://dx.doi.org/10.1016/S0021-9673\(00\)85726-X](http://dx.doi.org/10.1016/S0021-9673(00)85726-X).
- [41] G. Jin, Z. Guo, F. Zhang, X. Xue, Y. Jin, X. Liang, Study on the retention equation in hydrophilic interaction liquid chromatography, *Talanta* 76 (2008) 522–527, <http://dx.doi.org/10.1016/j.talanta.2008.03.042>.
- [42] U.D. Neue, H.J. Kuss, Improved reversed-phase gradient retention modeling, *J. Chromatogr. A* 1217 (2010) 3794–3803, <http://dx.doi.org/10.1016/j.chroma.2010.04.023>.
- [43] P. Česla, N. Vaňková, J. Křenková, J. Fischer, Comparison of isocratic retention models for hydrophilic interaction liquid chromatographic separation of native and fluorescently labeled oligosaccharides, *J. Chromatogr. A* 1438 (2016) 179–188, <http://dx.doi.org/10.1016/j.chroma.2016.02.032>.
- [44] P.J. Schoenmakers, Á. Bartha, H.A.H. Billiet, Gradient elution methods for predicting isocratic conditions, *J. Chromatogr. A* 550 (1991) 425–447, [http://dx.doi.org/10.1016/S0021-9673\(01\)88554-X](http://dx.doi.org/10.1016/S0021-9673(01)88554-X).
- [45] H. Akaike, A new look at the statistical model identification, in: K. K.G. Parzen Emanueland Tanabe (Eds.), *Sel. Pap. Hirotugu Akaike*, Springer New York, New York, NY, 1998, pp. 215–222, http://dx.doi.org/10.1007/978-1-4612-1694-0_16.
- [46] G. Vivó-Truyols, S. van der Wal, P.J. Schoenmakers, Comprehensive study on the optimization of online two-dimensional liquid chromatographic systems considering losses in theoretical peak capacity in first- and second-dimensions: a pareto-optimality approach, *Anal. Chem.* 82 (2010) 8525–8536, <http://dx.doi.org/10.1021/ac101420f>.
- [47] R.M. O’Flaherty, I. Trbojević-Akmačić, G. Greville, P.M. Rudd, G. Lauc, The sweet spot for biologics: recent advances in characterization of biotherapeutic glycoproteins, *Expert Rev. Proteomics* 15 (2018) 13–29, <http://dx.doi.org/10.1080/14789450.2018.1404907>.
- [48] Y. Yang, A. Barendregt, J.P. Kamerling, A.J.R. Heck, Analyzing protein micro-heterogeneity in chicken ovalbumin by high-resolution native mass spectrometry exposes qualitatively and semi-quantitatively 59 proteoforms, *Anal. Chem.* 85 (2013) 12037–12045, <http://dx.doi.org/10.1021/ac403057y>.
- [49] S.A. Carr, M.J. Huddleston, M.F. Bean, Selective identification and differentiation of N- and O-linked oligosaccharides in glycoproteins by liquid chromatography–mass spectrometry, *Protein Sci.* 2 (1993) 183–196, <http://dx.doi.org/10.1002/pro.5560020207>.
- [50] T. Fournier, N. Medjoubi, D. Porquet, Alpha-1-acid glycoprotein 1, *Biochim. Biophys. Acta.* 1482 (2000) 157–171, [http://dx.doi.org/10.1016/S0167-4838\(00\)00153-9](http://dx.doi.org/10.1016/S0167-4838(00)00153-9).
- [51] T. Imre, G. Schlosser, G. Pocsfalvi, R. Siciliano, E. Molnár-Szöllosi, T. Kremmer, A. Malorni, K. Vékey, Glycosylation site analysis of human alpha-1-acid glycoprotein (AGP) by capillary liquid chromatography-electrospray mass spectrometry, *J. Mass Spectrom.* 40 (2005) 1472–1483, <http://dx.doi.org/10.1002/jms.938>.
- [52] A.B. Rawitch, H.G. Pollock, S.X. Yang, Thyroglobulin glycosylation: location and nature of the N-linked oligosaccharide units in bovine thyroglobulin, *Arch. Biochem. Biophys.* 300 (1993) 271–279, <http://dx.doi.org/10.1006/abbi.1993.1038>.