Application of sample displacement batch chromatography for fractionation of proteoforms

Siti Nurul Hidayah[1, 2)], Manasi Gailwad[1)], Ali Biabani [1)], Paula Nissen [1)], Bente Siebels [1)], Hannah Voß[1)], Maria Riedner [3)], Hartmut Schlüter*[1)]

[1)]Section Mass Spectrometric Proteomics, University Medical Center Hamburg-Eppendorf, Campus Forschung N27-00.008, Martinistraße 52, 20246 Hamburg, Germany

[2)]Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Universitas Gadjah Mada, Jalan Sekip Utara, Mlati, Sleman, 55281 Yogyakarta, Indonesia

3) Technology Platform Mass Spectrometry, University of Hamburg, Mittelweg 177, 20148 Hamburg, Germany

Abstract

Fractionation of proteoforms is currently the most challenging topic in the field of protein purification. The need for considering the existence of proteoforms into experimental approaches is not only important in Life Science research in general but especially in the manufacturing of therapeutic proteins (TPs) like recombinant therapeutic antibodies (mAbs). Some of the proteoforms of TPs have significantly decreased actions or even cause side effects. The identification and removal of proteoforms differing from the main species, having the desired action, is challenging because the difference in the composition of atoms often is very small and their concentration in comparison to the main proteoform can be small. In this study we demonstrate that sample displacement batch chromatography (SDBC) is an easy to handle, economic and efficient method for fractionating proteoforms. As a model sample a commercial ovalbumin fraction was used, containing many ovalbumin proteoforms. The most promising parameters for the SDBC were determined by a screening approach and applied for a 10-segment fractionation of the ovalbumin with cation exchange chromatography resin. Mass spectrometry of intact proteoforms was used for characterizing the SDBC fractionation process. By SDBC a significant separation of different proteoforms was obtained.

Keywords: sample displacement batch chromatography, proteoform fractionation, ovalbumin, post translational modification, intact mass spectrometry

Introduction

Proteoforms, formerly termed protein species, are defined as smallest units of the proteome. From a

single specific gene usually many different proteoforms arise [1, 2] . Different proteoforms are formed by alternative RNA splicing as well as of all kinds of post translational modifications (PTMs) [3]. Aebersold et al. estimated the number of proteoforms coded by approximately 20.000 genes being in the range of several billions [4]. Therapeutic proteins (TPs) are protein-based drugs that are recombinantly engineered such as insulin [5], growth factor recombinant erythropoietin (rhEPO) [6], or monoclonal antibodies (mAbs). During the production of TPs, the choice of organisms, cells, and conditions affect the characteristic of the final products [7, 8]. Glycan analysis of same TPs produced in HEK293 cells and CHO cells show differences in molecular weight, isoelectric point (pI), and glycoprotein structure. CHO-derived proteins have a less complex glycoprotein profile, but more sialylations compared to HEK293-derived proteins [7]. In addition, during the in-vitro steps of the production of TPs, including harvesting and down-stream purification processes more modification reactions such as oxidation of methionine residues or deamidation of asparagine or glutamine can occur [9]. TPs can also be degraded by proteolysis thereby increasing the number of proteoforms[9]. In summary, a TP product, which is applied to patients, usually is containing many individual proteoforms. A part of these proteoforms is very similar differing only in a few atoms whereas others show large differences in size and/or the isoelectric points.

Although most proteoforms in TPs are not harmful [10], and their concentrations in relationship to the main active proteoform usually is small, their effect should not be underestimated. Several studies have shown that some proteoforms may lead to undesirable effects ranging from change in drug activity, change in target-antigen recognition [11], alteration of product's half-life [12], reduction of efficacy [13], and to potential effects on immunogenicity [14]. Therefore, identification of these undesired proteoforms and their removal are paramount. However, analysis and purification of proteoforms is one of the most challenging problems in analytical chemistry and in down-stream-processing because proteoforms are coded by the same gene that makes most of them physically and chemically very similar. Moreover, undesired proteoforms are often present in a low abundance.

As a starting point, the development of purification method for proteoforms can be approached in a similar manner to protein purification. To date, liquid chromatography (LC) remains as the front runner method. Different types of interactions of proteins to the stationary phase like electrostatic interaction, coordinative bond formation (e.g., immobilized metal affinity chromatography, IMAC) or hydrophobic interactions (e.g., reversed phase chromatography) offer a range of different chromatographic mechanisms to be tested out for efficient protein purification [15, 16]. The stationary phase and the mobile phase are key elements of liquid chromatography providing parameters for optimizing protein separations. Besides gradient or isocratic elution, displacement chromatography is

a third basic elution mode that should be considered especially when facing challenging separation problems like proteoform purification. In displacement chromatography analytes are loaded onto the column using low flow rates compared to gradient or isocratic chromatography so that the system comprising the distribution of the analyte between the stationary phase and the mobile phase will get the opportunity to achieve an equilibrium [17]. This process is supporting the formation of bands of different analytes on the stationary phase in a column, which are ordered according to their individual binding affinity. The band with the analyte with the highest affinity is located on top of the column. The subsequent bands in the direction to the column outlet are containing analytes with decreasing affinity towards the stationary phase. Elution is performed using a displacer, a molecule that should ideally have a higher affinity towards the stationary phase than any of the analyte molecules. In competition about the binding sites on the stationary phase the displacer molecule is displacing analytes with lower affinity as soon as the displacer molecule enters the column. An analyte, which is displaced by the displacer, is then displacing a further different analyte, having a lower affinity then the analyte with the highest affinity. This process is continued throughout all bands of the analytes. If the displacer is pumped onto the column and thereby occupying more and more binding sites, the bands of the analytes move down the column, forming the “displacement train”. The separation is finished when all binding sites are saturated by the displacer [18]. Elution profiles of displacement chromatography are not Gaussian peaks as in gradient elution but sharp rectangular zones containing analytes of high purity and high concentration. Displacement chromatography is well suited for the preparative purification of synthesis products for removal of a minor amounts of purities from the target molecule [19]. The power of displacement in the separation of complex protein mixtures and to enrich trace amounts of proteins was shown in several studies [20–22]. A comparison of displacement versus gradient chromatography for the separation of a protein mixture clearly demonstrated the advantages of displacement chromatography [23]. If an appropriate displacer molecule is difficult to find, sample displacement chromatography (SDC) is an alternative, because in SDC no displacer is needed. SDC has been used for the separation and purification for peptides [22] , proteins [24], and human plasma protein fraction Cohn IV-4 [25]. In DC the quality of separation usually is better than in the gradient chromatography mode [23], as well as the enrichment effect of trace components and the concentrating effect yielding the highest concentration of analytes eluting from the stationary phase compared to any other elution modes [26–28]. Sample displacement chromatography using packed columns is requiring the use of segmented columns connected in series. During the loading of the columns the sample molecules will bind to the stationary phase of the columns according to their affinity: On the first column the molecules with the highest affinity will bind. If it is the aim to fractionate a total protein amount of less than 0.2 g it will be difficult to find appropriate columns

because the smallest commercially available columns are containing 1 ml of resin. In case of a binding capacity of 200 mg/ml all proteins may bind to the first column and no separation will be obtained. With batch chromatography, resin in suspension in reaction vials, this problem can be circumvented because segments with resins much smaller than 1 ml are realizable. Sample displacement batch chromatography (SDBC) was already shown to be a powerful method for protein fractionation [24]. A detailed review of sample displacement chromatography for protein separation can be found in [22]. The Structural Genomics Consortium proposed in 2008 the following protein purification routine for recombinant proteins: proteins should be loaded onto the column until the column is fully saturated, implying that the amount of protein must slightly exceed the binding capacity of the column resin. All components with a lower affinity to the stationary phase would be displaced by the target protein [20]. In theory, this approach seems straight forward and promising on the first view. However, if a part of the impurities has a higher affinity than the target protein, its purification will not work, because the impurities are not displaced by the target protein. Furthermore, it is difficult to estimate the binding capacity of the target protein to the chosen column. If the applied amount of the target protein is too small, there is the risk that also impurities with a lower affinity than that of the target protein will bind.

Here, we demonstrate that fractionation of proteoforms with SDBC is possible and advantageous. The success of the fractionation was monitored with quantitative mass spectrometry detecting intact proteins. The presented method offers the utility and possibility of SDBC for simple and effective separation of proteoforms. In this study, only 1 µL of chromatography resin and 1 mg of sample were used in SDBC. As a model sample protein ovalbumin from chicken egg white was used as its composition is representing a large diversity of proteoforms. Using optimised conditions of SDBC, a two-fold increase in the number of detectable proteoforms in comparison to the original sample and significant enrichment of low-abundant basic ovalbumin proteoforms were obtained.

Materials and methods

1.1 Chemical

Strong cation exchange chromatography resins Eshmuno CPX was provided by Merck KGaA (Darmstadt, Germany). Ovalbumin from chicken egg white as model sample was purchased from Sigma-Aldrich (St. Louis, USA). Sodium formate, sodium acetate, and LC-MS grade water were purchased from Merck KGaA (Darmstadt, Germany). Formic acid, acetic acid, dithiothreitol (DTT), iodoacetamide (IAA), dimethyl sulfoxide (DMSO), NaCl, and sulfolane were purchased from Sigma-Aldrich (St. Louis, USA). Trypsin and PNGase F were purchased from Promega (Mannheim, Germany).

1.2 Sample Displacement Batch Chromatography

SDBC was performed following the procedure described by Kotasinska et al. [24, 29]. Prior to the SDBC several parameters like resin, type of ion-exchange chromatography and pH were screened for obtaining a set of parameters giving the largest number of identifiable ovalbumin proteoforms. Details are described in the supplement. For the SDBC experiment of ovalbumin cation exchange Eshmuno CPX resin (Merck KGaA, Darmstadt, Germany) was used. The binding capacity of the resin was calculated based on the information provided by the manufacturer which is 120 µg/µL resin suspension. The resin was then diluted to get a binding capacity of 2µg/µL resin suspension and equilibrated in sample application buffer 25 mM acetate buffer pH 5 (buffer A). Experiments were performed in three replicates. In each replicate, ten reaction vials (1.5 mL Eppendorf tubes) were used as ten segments. Diluted Eshmuno CPX resin suspension was transferred into each reaction vial in equal amounts (50 µL of resin suspension in each reaction vial). An overview about the steps of SDBC is given in **Figure 1.** One mg commercial ovalbumin (Sigma-Aldrich, St. Louis, USA) was dissolved in 1 mL buffer A (25mM acetate buffer pH 5), loaded into the first segment, incubated, and continuously shaken for 30 min in rotational shaker. After sedimentation of the resin, the supernatant from the first segment was transferred to the second segment, incubated, and continuously shaken for 30 min with a rotational shaker. The process was repeated until the sample passed the last segment. A washing step for each segment was performed using buffer A and the resulting supernatant was discarded. Each washing step was repeated 3 times. Proteoforms from each of the individual segments were eluted by adding 200 µL of buffer B (1M NaCl in buffer A) per segment, each incubating 30 minutes and thereafter sedimenting. The supernatant was collected from each segment. The total protein amount in the different fractions was then measured using Pierce™ BCA protein assay (Thermo Scientific, Bremen, Germany).

1.3 Mass spectrometric quantitation of intact proteoforms

All samples collected from SDBC experiments were desalted with 100% HPLC-grade water using 3K centrifugal filter 0.5 mL (Amicon Ultra, Merck KGaA, Darmstadt, Germany) for 6 times. 5% sulfolane was added to each sample as a supercharging agent. All samples were then analysed by mass spectrometry using flow injection analysis (FIA-MS). Therefore, LC autosampler from Waters ACQUITY UPLC system (Waters, Manchester, UK) was used to directly inject 1.5 µL of each sample into the MS. A hybrid-quadrupole-orbitrap mass spectrometry (Q-Exactive, Thermo Scientific,Bremen, Germany) was used. Mobile phase of 100% water was used for 4 minutes run in a flow rate of 0.075 mL/min for 1.5 minutes, then 0.1 mL/min for 1.5 minutes and back to 0.075 mL/min for 1 minute. MS parameters were set at scan range 1200 to 3600, resolution was set to 17500, AGC target was 5×10^6, micro scan count was 4, capillary temperature was set 275°C and in-source CID was 30eV.

1.4 N-glycan analysis

N-glycans of all ovalbumin proteoforms of the different SDBC fractions were released and analysed separately. To obtain the free N-glycans, proteoforms were reduced by 20mM dithiothreitol (DTT) at 56°C for 30 min and alkylated by 40mM iodoacetamide (IAA) at room temperature for 30 min in the dark. Buffer exchange were then performed using 3K centrifugal filter 0.5 mL (Amicon Ultra) to a 100 mM ammonium bicarbonate solution. 1:30 (v/w) PNGase F was added and incubated at 37°C for 24 h to release N-glycans, followed by addition of 1:100 (w/w) trypsin and incubated for another 20 h at 37°. N-glycans and tryptic peptides were separated using a RP-SPE C18 cartridge as described by Morelle at al.,[30]. The purified N-glycans were then permethylated using optimized solid phase permethylation method. For all methods of N-glycans permethylation, N-glycans LC-MS analysis, and N-glycans data analysis were performed using method developed and optimized by by Guan, Y., et. al., [31].

1.5 Identification and relative quantitation of the proteoforms using Intact mass spectrometric data

The MS1 spectra was deconvoluted using UniDec GUI deconvoluted software ver. 4.0.2 from Marty et. al., [32]. Parameters were set at data processing m/z range from 1500 to 3000, charge range from 10 to 35, sample mass every 0.5 Da, peak detection range was 15 Da, and peak detection threshold was set to 0.15. The relative quantities of the proteoforms were determined with UniDec using the signal intensities of the deconvoluted spectra. Results were exported and processed further in Excel spreadsheet. From the observed experimental masses, identification of ovalbumin proteoforms was performed by mass-matching

2. Results

2.1 Parameter screening for SDBC

For successful proteoform purification, the most optimal conditions must be determined prior to the SDBC by performing protein purification parameter screening (PPS). Parameters, including different pH values of sample application buffer, different resins, and different additional sodium chloride in sample application buffer were screened. Details about PPS results are given in the supplement. PPS conditions yielding the best results with respect to the number of detected proteoforms was chosen for SDBC, comprising a combination of strong cation exchange resin Eshmuno CPX and sample application buffer of 25 mM acetate buffer with a pH 5. Eshmuno CPX resin performed better than other resins, which may be due to the combination of high cross-linked rigid matrix that causes the proteoforms to bind stronger and the tentacle structure in the resin which allows multi-point interaction and reduces sterical hindrance between the functional group and the proteoforms.

3.2 Detection of ovalbumin proteoforms

MS1 spectra were deconvoluted with UniDec deconvolution software. **Figure 2** shows deconvoluted spectra of ovalbumin original sample and ovalbumin eluted from SDBC experiment fraction 1. Two clusters of proteoforms were detected in the molecular weight areas of 40 kDa (lower molecular weight/LMW) and 44 kDa (higher molecular weight/HMW). The mass difference of a proteoform with a molecular weight of 44,167 Da (HMW proteoforms) and a proteoform with a mass of 39,832 Da (LMW proteoforms) is 4,335 Da, which corresponds to the loss of the C-terminal peptide 348 to 386 of ovalbumin (EAGVDAASVSEEFRADHPFLFCIKHIATNAVLFFGRCVSP), presumably cleaved of by a protease, which is in agreement with the observation of Füssl et al., [33] who reported ovalbumin proteoforms with an identical truncation. Further assignments of the composition of proteoforms to the measured molecular weights were obtained by adding the mass of known ovalbumin post-translational modifications (PTMs) and the mass of experimental N-glycan composition to the ovalbumin average mass. The initial methionine of ovalbumin was cleaved from mature protein thus the average mass of ovalbumin was derived from the amino acid sequences position number 2 to386 that is 42,750 Da [34]. Then, the mass of known PTMs in ovalbumin named N-acetylation of amino acid (delta mass 42 Da) and one or two phosphorylation (delta mass 79.99 Da) were added. The mass of all N-glycan compositions obtained from experimental released N-glycan using PNGase was also added. The possibility of ovalbumin proteoforms lacking N-acetylation and lacking phosphorylation were also considered for mass-matching identification. In addition, potential succinimide formation of an aspartic acid (mass decrease 18 Da) was also considered. **Table 1** lists all assignments of ovalbumin proteoforms towards measured molecular weights. In this study, a total of at least 43 ovalbumin proteoforms were detected. In this identification approach, 4 types of the most often reported PTMs in ovalbumin including N-glycosylation, phosphorylation, N-terminal acetylation, and succinamate formation were used. One N-glycan site in amino acid asparagine position 293 contributes to largest number of ovalbumin proteoforms as its biosynthesis is template free-driven process which results in various forms of branching, glycan composition and other types of isomerism. In this study, 67 different N-glycan structures were identified from one N-glycosylation site of ovalbumin. It can be assumed that the commercial ovalbumin sample used in this study was purified from of eggs derived from thousands of individual hens. Thus, genetic variations resulting in amino acid substitutions combined with significant differences in PTMs are responsible for a large number of individual proteoforms of ovalbumin. Therefore, the total number of ovalbumin proteoforms is probably far larger than what has been detected in this study.

3.3 SDBC of ovalbumin proteoforms

In this study ovalbumin proteoforms were separated by cation-exchange SDBC into 10 fractions applying the separation workflow in **Figure 1**. The total protein amount of the individual fractions eluted from the different SDBC segments is shown in **Figure 3a**. The decrease of protein amount from the first segment to the last segment is in agreement with the theory in displacement chromatography, since in the first segment molecules with the highest affinity are binding resulting in the highest concentration per resin segment. In fraction number 10, no proteins were detectable, which is showing that all proteins with an affinity to the resin were already adsorbed to the segments 1 to 9. The total number of detected proteoforms in each fraction is given in **figure 3b,** while all possible proteoforms calculated by mass matching in each SDBC fraction are listed in **supporting information table 2**. Twice the number of proteoforms were detected in SDBC fraction 1-4 compared to the original sample, which was applied to to SDBC, demonstrating a significant enrichment of low-abundant proteoforms as well as a significant reduction of the complexity in the segments compared to the original sample. The number was decreasing with the increasing fraction number because the affinity of the proteoforms is decreasing, and consequently lower numbers of proteins have been binding to the resin.

The relative abundance of each proteoform in all SDBC fractions was quantified from the deconvoluted spectra by extracting the intensities of their corresponding signals. It was hypothesized that the proteoforms have similar desorption and ionization probabilities and thus that their signal intensities are representing their relative abundance. In total 43 masses were quantified, and the relative quantities were plotted as chromatogram. **Figure 4** shows the chromatograms of individual proteoforms in the original sample and SDBC fractions. In **Figure 4** those 24 proteoforms are highlighted which were not detectable in the original ovalbumin sample. Thus, it can be assumed that these proteoforms were enriched in the early SDBC fractions. As the SDBC experiment was performed using CEX chromatography resin, those low-abundant proteoforms are more basic ovalbumin proteoforms than those eluting in the later fractions. Identification of these proteoforms by mass-matching was conducted and it was observed that 12 out of 24 proteoforms show a mass decrease of 18 Da which potentially corresponds to succinimide formation of an aspartic acid [35].

Furthermore, in this study, SDBC was also able to separate closely related ovalbumin proteoforms into two categories: enriched proteoforms in early fractions and purified main ovalbumin proteoforms in later fractions. **Figure 5** focuses on enriched ovalbumin proteoforms. These proteoforms were detected in relatively low abundance in the original sample. Due to their high affinity toward the cation exchange resin, they occupied the binding sites of the resins in the early segments. **Figure 6** shows the chromatogram highlighting the most abundant ovalbumin proteoforms. These proteoforms show the highest abundance in the original sample. Obviously these proteoforms eluting in fraction 4 to 9 have

a lower affinity to the resin. These proteoforms are significantly purified from the other proteoforms eluting in the fractions 1 - 3. The identification of the proteoforms in both categories revealed that highly enriched proteoforms in early fraction are most probably ovalbumin with one N-glycan, one acetyl, and one phosphate attached on it, while the proteoforms in the later fractions are most probably ovalbumin with one N-glycan, one acetyl, and two phosphates attached on it. Phosphorylation of these proteoforms reduce their isoelectric point making them less basic. This gives a proof that SDBC is effective even for the separation of closely related proteoforms from a single protein.

In addition, with SDBC it was possible to remove proteoforms not belonging to the ovalbumin proteoform family. By bottom-up proteomics of the proteins in the flow-through-fraction ovomucoid proteoforms beside some proteoforms of ovalbumin as main abundant proteins were identified (Figure S3; details are given in the supplement). The ovalbumin proteoforms detected in the flow-through fraction may have low isoelectric points and therefore did not bind to the CEX resin.

3. Discussion

In this study it was demonstrated that SDBC is suitable for fractionation of proteoforms. SDBC simplifies the set-up of typical SDC which is requiring the connection of several columns in series. SDBC can be easily performed in different devices, such as microtiter 384-wells plate, 96 deep well-plates or reaction vials. Using cost effective, simple, and easy to handle batch chromatography performed in reaction vials, low-abundant ovalbumin proteoforms were enriched. The separation of different proteoforms of ovalbumin into different fractions, thereby reducing the complexity of the composition of proteoforms, resulted in a two-fold increase of the number of detected proteoforms compared to the analysis of original ovalbumin sample. Compared to previous attempts on proteoforms purification, the need of high sample amounts in a typical DC or SDC was eliminated. In a study by Zhang et al. [36] using DC to separate antibody charge variants using a cation exchange column 200 mg of monoclonal antibody was needed, while another study by Khanal et al. [37] has used around 10 mg of monoclonal antibody for one SDC experiment using in-house-packed columns. This need of high amount of sample may hinder the experiment when dealing with limited amount of sample. In our method, only 1 µL of resin per segment and 1 mg of ovalbumin sample were used thus giving benefit when dealing with precious samples with limited availability. If a preparative scale is desired, the scale-up of a SDBC method also is straightforward and easy [25, 29]. SDBC offers an economical benefit as it is much less expensive than preparative column gradient LC, because with SDBC no expensive gradient pumps are required and the binding capacity of the stationary phase is completely used whereas in gradient chromatography it is recommendable not to exceed 20 % of the binding capacity. In this study

it is shown that SDBC can then be applied for therapeutic proteins such as monoclonal antibodies.

Conclusion

In summary, we demonstrated that fractionation of proteoforms by liquid chromatography, based on sample displacement batch chromatography (SDBC) is possible. As a result of the successful fractionation the total number of proteoforms detectable with mass spectrometry significantly increased to 42 compared to 16 proteoforms in the original ovalbumin sample prior to its fractionation. With SDBC enrichment of low-abundant ovalbumin proteoforms was obtained. We also successfully separated some of ovalbumin proteoforms which proves that SDBC is effective even for the separation of closely related proteoforms. Despite several challenges regarding the complexity of proteoforms and limitation regarding the availability of well-established analytical system, we demonstrated the utility and possibility of SDBC for simple and effective separation of proteoforms. This in the future can be used for removal of harmful proteoforms and as a tool for enrichment of low abundant proteoforms for further analysis or to support and enhance molecular characterization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [38] partner repository with the dataset identifier PXD040682.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No [765502], A4B project. This study was supported by grants from the Deutsche Forschungsgemeinschaft (DFG) (INST 337/15-1, INST 337/16-1 & INST 152/837-1).

References

1. Schlüter, H., Apweiler, R., Holzhütter, H. G., & Jungblut, P. R. (2009). Finding one's way in proteomics: A protein species nomenclature. *Chemistry Central Journal*, *3*(1), 1–10.
2. Jungblut, P. R., Holzhütter, H. G., Apweiler, R., & Schlüter, H. (2008). The speciation of the proteome. *Chemistry Central Journal*, *2*(1), 1–10.
3. Smith, L. M., & Kelleher, N. L. (2013). Proteoform: A single term describing protein complexity. *Nature Methods*, *10*(3), 186–187.
4. Aebersold, R., Agar, J. N., Amster, I. J., Baker, M. S., Bertozzi, C. R., Boja, E. S., … Zhang, B. (2018). How many human proteoforms are there? *Nature Chemical Biology*, *14*(3), 206–214.
5. Rege, N. K., Wickramasinghe, N. P., Tustan, A. N., Phillips, N. F. B., Yee, V. C., Ismail-Beigi, F., & Weiss,

M. A. (2018). Structure-based stabilization of insulin as a therapeutic protein assembly via enhanced aromatic–aromatic interactions. *Journal of Biological Chemistry*, *293*(28), 10895–10910.

6. Hassiki, R., Behi, J., Said, N. ben, Boujbel, L., & Bouhaouala-Zahar, B. (2017). Biochemical and Biological Quality Control of Two Recombinant Human Erythropoietin Biosimilar Products. *Current Pharmaceutical Analysis*, *13*(6), 566–573.
7. Goh, J. B., & Ng, S. K. (2018). Impact of host cell line choice on glycan profile. *Critical Reviews in Biotechnology*, *38*(6), 851–867.
8. Tyshchuk, O., Gstöttner, C., Funk, D., Nicolardi, S., Frost, S., Klostermann, S., … Mølhøj, M. (2019). Characterization and prediction of positional 4-hydroxyproline and sulfotyrosine, two post-translational modifications that can occur at substantial levels in CHO cells-expressed biotherapeutics. *mAbs*, *11*(7), 1219–1232.
9. Gervais, D. (2016). Protein deamidation in biopharmaceutical manufacture: Understanding, control and impact. *Journal of Chemical Technology and Biotechnology*, *91*(3), 569–575.
10. Beck, A., & Liu, H. (2019). Macro- and Micro-Heterogeneity of Natural and Recombinant IgG Antibodies. *Antibodies*, *8*(1), 18.
11. Moss, C. X., Matthews, S. P., Lamont, D. J., & Watts, C. (2005). Asparagine deamidation perturbs antigen presentation on class II major histocompatibility complex molecules. *Journal of Biological Chemistry*, *280*(18), 18498–18503.
12. Liu, Y. D., van Enk, J. Z., & Flynn, G. C. (2009). Human antibody Fc deamidation in vivo. *Biologicals*, *37*(5), 313–322.
13. Lee, H. H. L., Heo, C. E., Seo, N., Yun, S. G., An, H. J., & Kim, H. I. (2018). Accurate Quantification of N-Glycolylneuraminic Acid in Therapeutic Proteins Using Supramolecular Mass Spectrometry. *Journal of the American Chemical Society*, *140*(48), 16528–16534.
14. Hermeling, S., Crommelin, D. J. A., Schellekens, H., & Jiskoot, W. (2004). Structure-immunogenicity relationships of therapeutic proteins. *Pharmaceutical Research*, *21*(6), 897–903.
15. Riguero, V., Clifford, R., Dawley, M., Dickson, M., Gastfriend, B., Thompson, C., … O'Connor, E. (2020). Immobilized metal affinity chromatography optimization for poly-histidine tagged proteins. *Journal of Chromatography A*, *1629*.
16. Kopp, J., Zauner, F. B., Pell, A., Hausjell, J., Humer, D., Ebner, J., … Pell, R. (2020). Development of a generic reversed-phase liquid chromatography method for protein quantification using analytical quality-by-design principles. *Journal of Pharmaceutical and Biomedical Analysis*, *188*.
17. McAtee, C. P. (2010). Displacement Chromatography of Proteins. *Current Protocols in Protein Science*, (SUPPL. 59), 1–14.
18. Evans, S. T., Freed, A., & Cramer, S. M. (2009). Displacer concentration effects in displacement chromatography: Implications for trace solute detection. *Journal of Chromatography A*, *1216*(1), 79–85.
19. Jankowski, ¨ J, Potthoff, W., Zidek, W., & Schluter¨medizinische, H. (1998). *Purification of chemically synthesised dinucleoside(59,59) polyphosphates by displacement chromatography*. *Journal of Chromatography B* (Vol. 719).
20. Gräslund, S., Nordlund, P., Weigelt, J., Hallberg, B. M., Bray, J., Gileadi, O., … Gunsalus, K. C. (2008). Protein production and purification. *Nature Methods 2008 5:2*, *5*(2), 135–146.
21. Josic, D., Breen, L., Clifton, J., Gajdosik, M. S., Gaso-Sokac, D., Rucevic, M., & Müller, E. (2012). Separation of proteins from human plasma by sample displacement chromatography in hydrophobic interaction mode. *ELECTROPHORESIS*, *33*(12), 1842–1849.
22. Srajer Gajdosik, M., Clifton, J., & Josic, D. (2012). Sample displacement chromatography as a method for purification of proteins and peptides from complex mixtures. *Journal of Chromatography A*, *1239*, 1–9.
23. Ahrends, R., Lichtner, B., Buck, F., Hildebrand, D., Kotasinska, M., Kohlbacher, O., … Schlüter, H. (2012). Comparison of displacement versus gradient mode for separation of a complex protein mixture by anion-exchange chromatography. *Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences*, *901*, 34–40.
24. Kotasińska, M., Richter, V., Thiemann, J., & Schlüter, H. (2012). Cation exchange displacement batch chromatography of proteins guided by screening of protein purification parameters. *Journal of Separation Science*, *35*(22), 3170–3176.

25. Heikaus, L., Hidayah, S., Gaikwad, M., Kotasinska, M., Richter, V., Kwiatkowski, M., & Schlüter, H. (2021). Sample displacement batch chromatography of proteins. In *Methods in Molecular Biology* (Vol. 2178, pp. 285–299). Humana Press Inc.
26. Horváth, C., Nahum, A., & Frenz, J. H. (1981). High-performance displacement chromatography. *Journal of Chromatography A*, *218*(C), 365–393.
27. Trusch, M., Tillack, K., Kwiatkowski, M., Bertsch, A., Ahrends, R., Kohlbacher, O., … Schlüter, H. (2012). Displacement chromatography as first separating step in online two-dimensional liquid chromatography coupled to mass spectrometry analysis of a complex protein sample-The proteome of neutrophils. *Journal of Chromatography A*, *1232*, 288–294.
28. Kwiatkowski, M., Krösser, D., Wurlitzer, M., Steffen, P., Barcaru, A., Krisp, C., … Schlüter, H. (2018). Application of Displacement Chromatography to Online Two-Dimensional Liquid Chromatography Coupled to Tandem Mass Spectrometry Improves Peptide Separation Efficiency and Detectability for the Analysis of Complex Proteomes. *Analytical Chemistry*, *90*(16), 9951–9958.
29. Kotasinska, M., Richter, V., Kwiatkowski, M., & Schlüter, H. (2014). Sample displacement batch chromatography of proteins. *Methods in Molecular Biology*, *1129*, 325–338.
30. Morelle, W., & Michalski, J. C. (2007). Analysis of protein glycosylation by mass spectrometry. *Nature Protocols*, *2*(7), 1585–1602.
31. Guan, Y., Zhang, M., Wang, J., & Schlüter, H. (2021). Comparative Analysis of Different N-glycan Preparation Approaches and Development of Optimized Solid-Phase Permethylation Using Mass Spectrometry. *Journal of Proteome Research*, *20*(5), 2914–2922.
32. Marty, M. T., Baldwin, A. J., Marklund, E. G., Hochberg, G. K. A., Benesch, J. L. P., & Robinson, C. V. (2015). Bayesian deconvolution of mass and ion mobility spectra: From binary interactions to polydisperse ensembles. *Analytical Chemistry*, *87*(8), 4370–4376.
33. Füssl, F., Criscuolo, A., Cook, K., Scheffler, K., & Bones, J. (2019). Cracking Proteoform Complexity of Ovalbumin with Anion-Exchange Chromatography-High-Resolution Mass Spectrometry under Native Conditions. *Journal of Proteome Research*, *18*(10), 3689–3702.
34. Thompson, E. O. P., & Fisher, W. K. (1978). *A Correction and Extension of the Acetylated Amino Terminal Sequence of Ovalbumin*. *J. Bioi. Sci* (Vol. 31).
35. Nakayoshi, T., Kato, K., Fukuyoshi, S., Takahashi, O., Kurimoto, E., & Oda, A. (2021). Molecular mechanisms of succinimide formation from aspartic acid residues catalyzed by two water molecules in the aqueous phase. *International Journal of Molecular Sciences*, *22*(2), 1–11.
36. Zhang, T., Bourret, J., & Cano, T. (2011). Isolation and characterization of therapeutic antibody charge variants using cation exchange displacement chromatography. *Journal of Chromatography A*, *1218*(31), 5079–5086.
37. Khanal, O., Kumar, V., Westerberg, K., Schlegel, F., & Lenhoff, A. M. (2019). Multi-column displacement chromatography for separation of charge variants of monoclonal antibodies. *Journal of Chromatography A*, *1586*, 40–51.
38. Perez-Riverol, Y., Bai, J., Bandla, C., Hewapathirana, S., García-Seisdedos, D., Kamatchinathan, S., Kundu, D., Prakash, A., Frericks-Zipper, A., Eisenacher, M., Walzer, M., Wang, S., Brazma, A., Vizcaíno, J.A. (2022). The PRIDE database resources in 2022: A Hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res 50*(D1), D543-D552
.

Figures and table

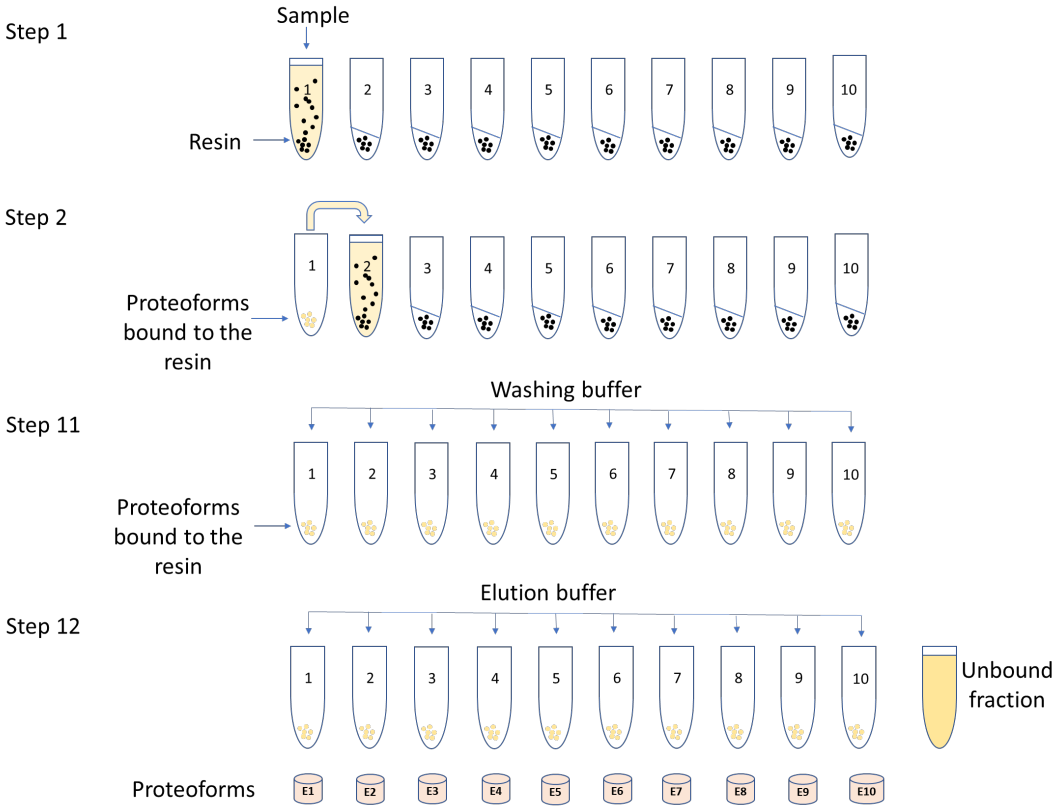

Figure 1. Steps of sample displacement batch chromatography (SDBC) using 10 segments. Step 3 comprises the transfer of the sample from segment 2 to segment 3, and so on.

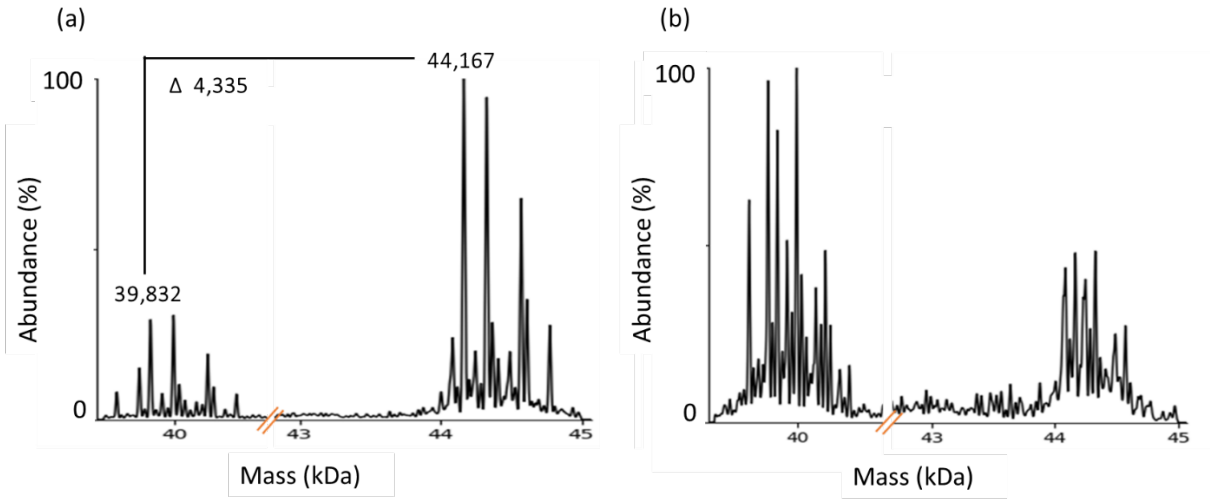

Figure 2. Deconvoluted spectra of ovalbumin from the original sample (**a**) and from the SDBC fraction 1 (**b**).

454 **Table 1.** List of ovalbumin proteoforms identified by mass-matching.

No.	Mass (Da)	*PTMs	**N-glycan composition
1	39593	GPPS	HexNAc2Hex4
2	39674	GPS	HexNAc2Hex5
3	39753	GPA	HexNAc2Hex5
4a	39832	GPPA	HexNAc2Hex5
4b		GA	HexNAc2Hex6
5	39915	GS	Neu5Ac1HexNAc3Hex3Fuc1
6	39954	GAS	Neu5Ac1HexNAc3Hex3Fuc1
7	39995	GPPA	HexNAc2Hex6
8	40036	GPAS	Neu5Ac1HexNAc3Hex3Fuc1
9a	40076	GA	Neu5Ac1HexNAc3Hex3Fuc1
9b		GPPS	Neu5Ac1HexNAc3Hex3Fuc1
10	40158	GP	Neu5Ac1HexNAc3Hex3Fuc1
11	40200	GPA	Neu5Ac1HexNAc3Hex3Fuc1
12	40238	GPP	Neu5Ac1HexNAc3Hex3Fuc1
13	40279	GPPA	Neu5Ac1HexNAc3Hex3Fuc1
14	40441	GS	Neu5Ac1HexNAc4Hex5Fuc1
15a	44070	GP	Neu5Ac1HexNAc3Hex3Fuc1
15b		GPAS	HexNAc2Hex5
16a	44086	GPA	HexNAc2Hex5
16b		GP	HexNAc3Hex4
17a	44123	GPP	HexNAc2Hex5
17b		GAS	Neu5Ac1HexNAc2Hex4
18a	44167	G	HexNAc3Hex5
18b		GPPA	HexNAc2Hex5
18c		GPP	HexNAc3Hex4
19a	44208	G	HexNAc4Hex4
19b		GPPA	HexNAc3Hex4
19c		GPP	HexNAc4Hex3
19d		GP	HexNAc2Hex6
20a	44231	GPS	HexNAc3Hex5
20b		GPAS	HexNAc2Hex6
20c		GPPAS	HexNAc4Hex3
21a	44248	GPA	HexNAc2Hex6
21b		GPPA	HexNAc4Hex3
21c		GP	HexNAc3Hex5
22a	44289	G	Neu5Ac1HexNAc3Hex4
22b		GPA	HexNAc3Hex5
22c		GPP	HexNAc2Hex6
22d		GP	HexNAc4Hex4
23a	44328	GA	Neu5Ac1HexNAc3Hex3Fuc1
23b		GPPA	HexNAc2Hex6
23c		GPP	HexNAc3Hex5
24a	44370	G	HexNAc4Hex5
24b		GPPA	HexNAc3Hex5

24c		GPP	HexNAc4Hex4
24d		GP	Neu5Ac1HexNAc3Hex4
25a	44394	GPPA	HexNAc4Hex3Fuc1
25b		GPPS	HexNAc5Hex3
25c		GPPAS	HexNAc4Hex4
26a	44411	G	HexNAc4Hex4Red-HexNAc1
26b		GPA	Neu5Ac1HexNAc3Hex4
26c		GPPA	HexNAc4Hex4
26d		GPP	HexNAc5Hex3
26e		GP	HexNAc3Hex6
27a	44476	GPA	HexNAc4Hex4Fuc1
27b		GPS	HexNAc5Hex4
27c		GPAS	HexNAc4Hex5
27d		GPPAS	Neu5Ac1HexNAc3Hex4
28a	44493	GPA	HexNAc4Hex5
28b		GPPA	Neu5Ac1HexNAc3Hex4
28c		GPP	HexNAc3Hex6
28d		GA	HexNAc5Hex4
29a	44531	GPPA	HexNAc3Hex6
29b		GPP	HexNAc4Hex5
30a	44573	G	HexNAc5Hex5
30b		GPPA	HexNAc4Hex5
30c		GPP	HexNAc5Hex4
31a	44613	GPPA	HexNAc5Hex4
31b		GPP	HexNAc6Hex3
31c		GP	HexNAc4Hex6
32a	44697	GA	HexNAc7Hex3
32b		GPA	HexNAc5Hex5
32c		GPP	HexNAc4Hex6
32d		GP	HexNAc6Hex4
33a	44776	G	HexNAc6Hex5
33b		GPPA	HexNAc5Hex5
33c		GPP	HexNAc6Hex4
34	33639		
35	39630		
36	39711		
37	39791		
38	39872		
39	40361		
40	43638		
41	43713		
42	43878		

*PTMs, N-glycosylation (N), phosphorylation (P), acetylation (A), and succinimide (S). **Proposed N-glycan compositions, Hex (hexose, galactose/mannose), HexNAc (N-acetylhexosamine), Neu5Ac (N-acetylneuraminic acid), and Fuc (fucose). Preferred N-glycan structure for each composition can be found in the supplement 3.

455

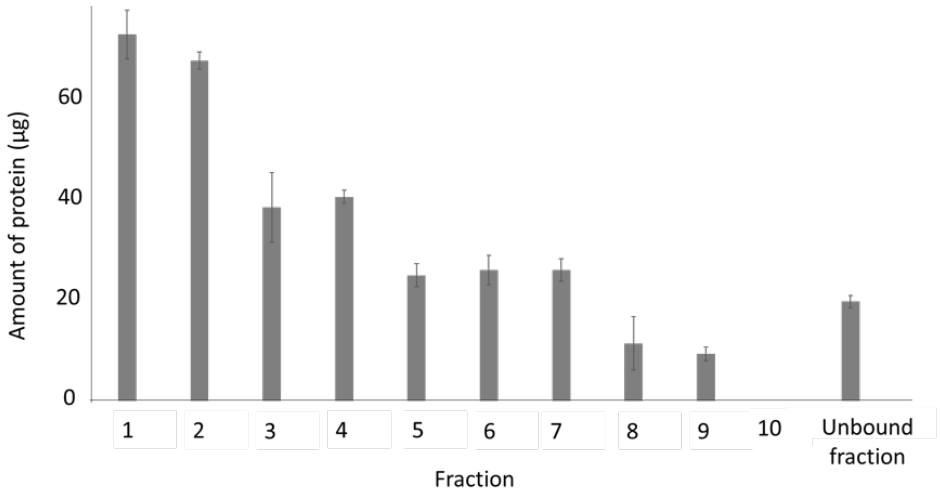

Figure 3. Total protein amount (measured with a BCA assay) eluted from each SDBC fraction.

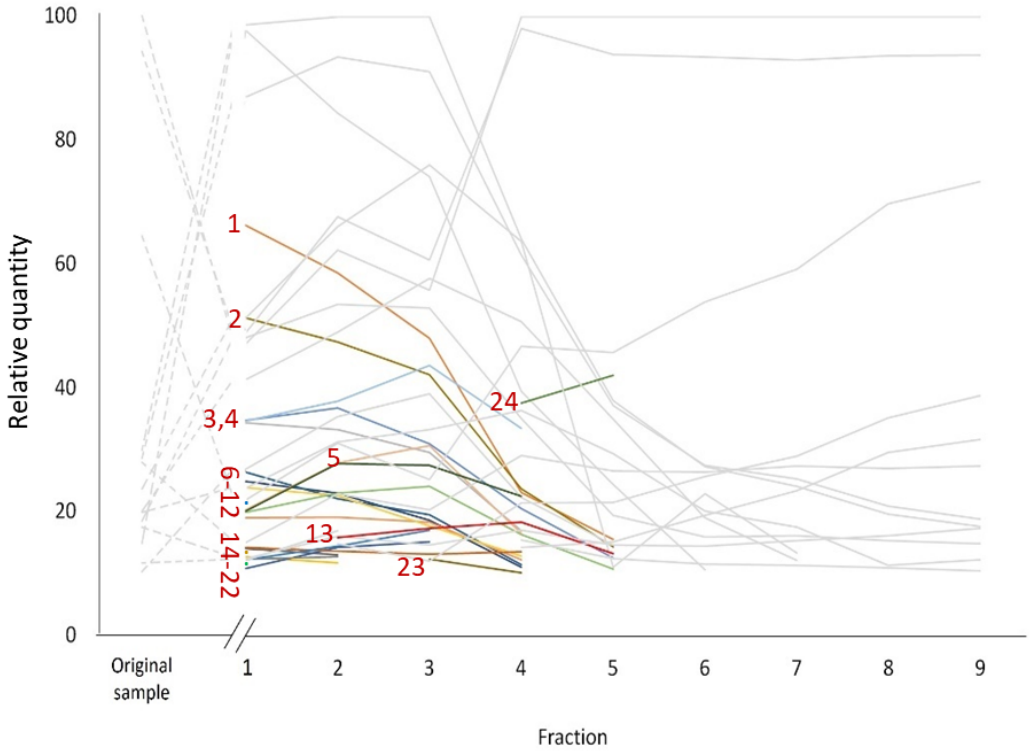

Figure 4. Chromatograms of proteoforms detected in the ovalbumin original sample and in the eluates of SDBC fractions. Highlighted in color-lines are proteoforms enriched and detected in SDBC early fractions only and not in the original sample. All other proteoforms are presented in grey lines. Mass of proteoforms in 1-24: 39593; 39630; 39674; 39711; 39791; 39872; 39915; 39954; 40036; 40076; 40158; 40200; 40279; 40361; 40441; 43638; 43713; 43878; 44070; 44123; 44231; 44394; 44476; 44531 Da. Relative abundance of the proteoforms was obtained from the signals of the deconvoluted spectra. in which the highest signal intensity in each fraction was given a value of 100 and the abundance of other proteoforms was calculated relatively to the most intense signal. Note that in the eluate of segment 10 no signals of any proteoform were detected.

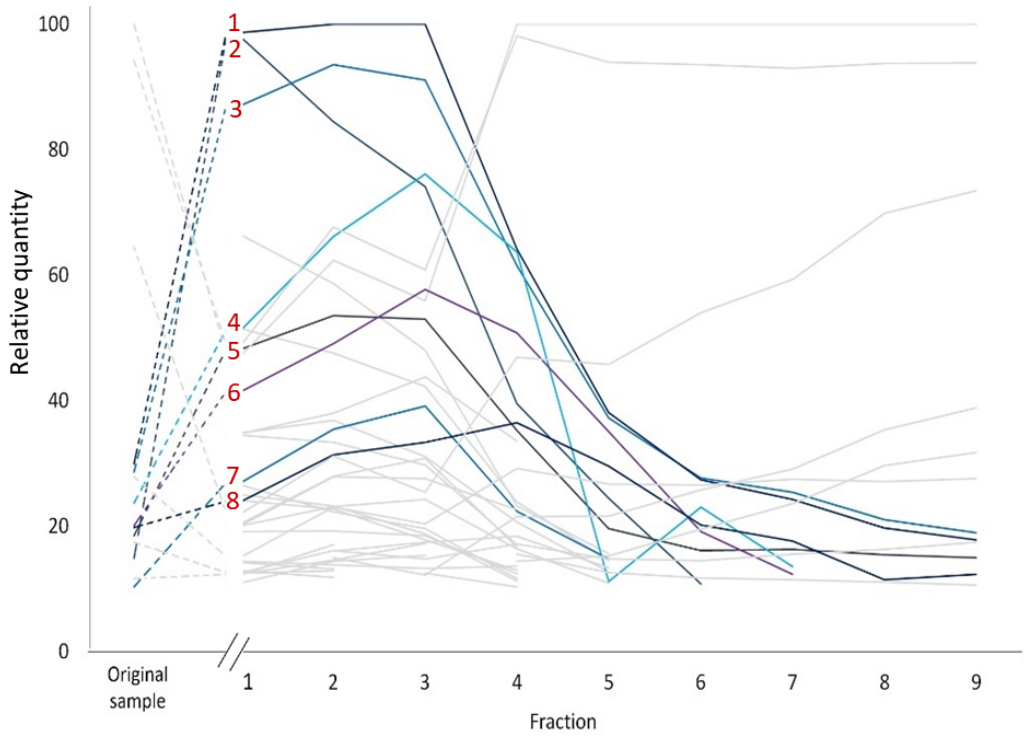

Figure 5. Chromatograms of proteoforms detected in the ovalbumin original sample and in the eluates of SDBC fractions, highlighting significantly enriched proteoforms represented in color-lines. All other proteoforms are presented in grey-lines. Mass of proteoform 1-8: 39995; 39753; 39832; 44086; 44248; 40237; 44288; 44493 Da. Note that in the eluate of segment 10 no signals of any proteoform were detected.

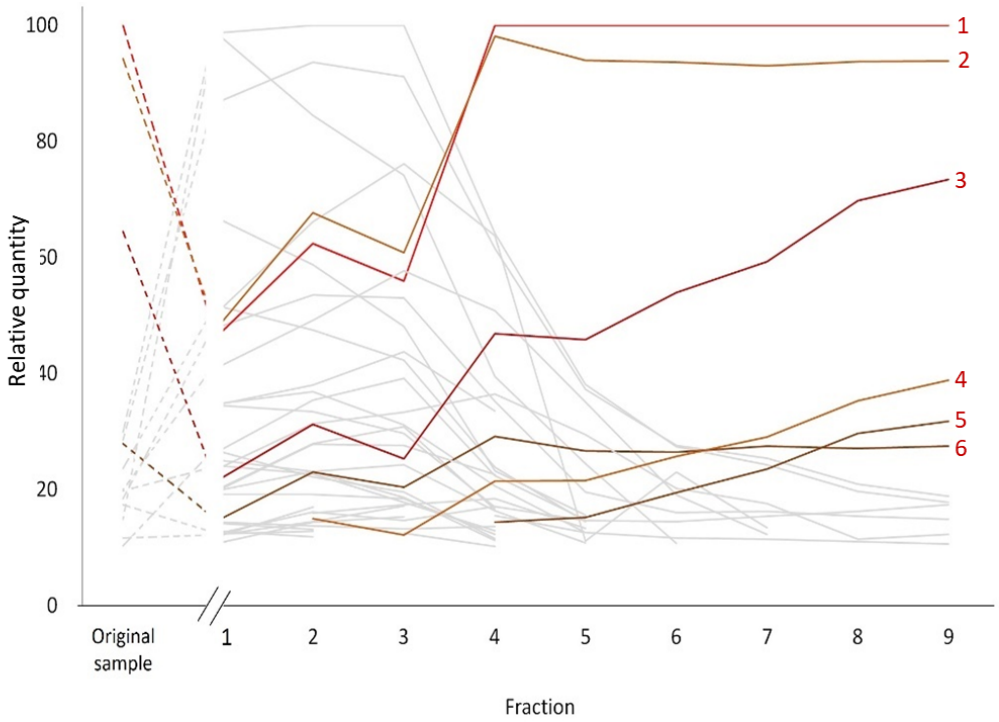

Figure 6. Chromatograms of all proteoforms detected in ovalbumin original sample and in the eluates of SDBC fractions, focus on proteoforms detected in higher abundance in the later SDBC fractions (represented in color-lines) while other proteoforms were presented in grey-lines. Mass of proteoform in 1-6: 44167; 44328; 44573; 44613; 44775; 44369 Da. Note that in the eluate of segment 10 no signals of any proteoform were detected.