Application of sample displacement batch chromatography for fractionation of proteoforms

Siti Nurul Hidayah[1], Manasi Gaikwad[1], Ali Biabani[1], Paula Nissen[1], Bente Siebels[1], Hannah Voß[1], Maria Riedner[2], and Hartmut Schlüter[1]

[1]University Medical Center Hamburg-Eppendorf
[2]University of Hamburg

March 10, 2023

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.

1 **Application of sample displacement batch chromatography for fractionation of**
2 **proteoforms**

3

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

6 [1]Section Mass Spectrometric Proteomics, University Medical Center Hamburg-Eppendorf, Campus
7 Forschung N27-00.008, Martinistraße 52, 20246 Hamburg, Germany

8 [2]Department of Pharmaceutical Chemistry, Faculty of Pharmacy, Universitas Gadjah Mada, Jalan
9 Sekip Utara, Mlati, Sleman, 55281 Yogyakarta, Indonesia

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

12 **Abstract**

13 Fractionation of proteoforms is currently the most challenging topic in the field of protein purification.

14 The need for considering the existence of proteoforms into experimental approaches is not only

15 important in Life Science research in general but especially in the manufacturing of therapeutic

16 proteins (TPs) like recombinant therapeutic antibodies (mAbs). Some of the proteoforms of TPs have

17 significantly decreased actions or even cause side effects. The identification and removal of

18 proteoforms differing from the main species, having the desired action, is challenging because the

19 difference in the composition of atoms often is very small and their concentration in comparison to

20 the main proteoform can be small. In this study we demonstrate that sample displacement batch

21 chromatography (SDBC) is an easy to handle, economic and efficient method for fractionating

22 proteoforms. As a model sample a commercial ovalbumin fraction was used, containing many

23 ovalbumin proteoforms. The most promising parameters for the SDBC were determined by a screening

24 approach and applied for a 10-segment fractionation of the ovalbumin with cation exchange

25 chromatography resin. Mass spectrometry of intact proteoforms was used for characterizing the SDBC

26 fractionation process. By SDBC a significant separation of different proteoforms was obtained.

27 **Keywords**: sample displacement batch chromatography, proteoform fractionation, ovalbumin, post

28 translational modification, intact mass spectrometry

29

30 **Introduction**

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

32 single specific gene usually many different proteoforms arise [1, 2] . Different proteoforms are formed

33 by alternative RNA splicing as well as of all kinds of post translational modifications (PTMs) [3].

34 Aebersold et al. estimated the number of proteoforms coded by approximately 20.000 genes being in

35 the range of several billions [4]. Therapeutic proteins (TPs) are protein-based drugs that are

36 recombinantly engineered such as insulin [5], growth factor recombinant erythropoietin (rhEPO) [6],

37 or monoclonal antibodies (mAbs). During the production of TPs, the choice of organisms, cells, and

38 conditions affect the characteristic of the final products [7, 8]. Glycan analysis of same TPs produced

39 in HEK293 cells and CHO cells show differences in molecular weight, isoelectric point (pI), and

40 glycoprotein structure. CHO-derived proteins have a less complex glycoprotein profile, but more

41 sialylations compared to HEK293-derived proteins [7]. In addition, during the in-vitro steps of the

42 production of TPs, including harvesting and down-stream purification processes more modification

43 reactions such as oxidation of methionine residues or deamidation of asparagine or glutamine can

44 occur [9]. TPs can also be degraded by proteolysis thereby increasing the number of proteoforms[9].

45 In summary, a TP product, which is applied to patients, usually is containing many individual

46 proteoforms. A part of these proteoforms is very similar differing only in a few atoms whereas others

47 show large differences in size and/or the isoelectric points.

48 Although most proteoforms in TPs are not harmful [10], and their concentrations in relationship to the

49 main active proteoform usually is small, their effect should not be underestimated. Several studies

50 have shown that some proteoforms may lead to undesirable effects ranging from change in drug

51 activity, change in target-antigen recognition [11], alteration of product's half-life [12], reduction of

52 efficacy [13], and to potential effects on immunogenicity [14]. Therefore, identification of these

53 undesired proteoforms and their removal are paramount. However, analysis and purification of

54 proteoforms is one of the most challenging problems in analytical chemistry and in down-stream-

55 processing because proteoforms are coded by the same gene that makes most of them physically and

56 chemically very similar. Moreover, undesired proteoforms are often present in a low abundance.

57 As a starting point, the development of purification method for proteoforms can be approached in a

58 similar manner to protein purification. To date, liquid chromatography (LC) remains as the front runner

59 method. Different types of interactions of proteins to the stationary phase like electrostatic

60 interaction, coordinative bond formation (e.g., immobilized metal affinity chromatography, IMAC) or

61 hydrophobic interactions (e.g., reversed phase chromatography) offer a range of different

62 chromatographic mechanisms to be tested out for efficient protein purification [15, 16]. The stationary

63 phase and the mobile phase are key elements of liquid chromatography providing parameters for

64 optimizing protein separations. Besides gradient or isocratic elution, displacement chromatography is

65 a third basic elution mode that should be considered especially when facing challenging separation

66 problems like proteoform purification. In displacement chromatography analytes are loaded onto the

67 column using low flow rates compared to gradient or isocratic chromatography so that the system

68 comprising the distribution of the analyte between the stationary phase and the mobile phase will get

69 the opportunity to achieve an equilibrium [17]. This process is supporting the formation of bands of

70 different analytes on the stationary phase in a column, which are ordered according to their individual

71 binding affinity. The band with the analyte with the highest affinity is located on top of the column.

72 The subsequent bands in the direction to the column outlet are containing analytes with decreasing

73 affinity towards the stationary phase. Elution is performed using a displacer, a molecule that should

74 ideally have a higher affinity towards the stationary phase than any of the analyte molecules. In

75 competition about the binding sites on the stationary phase the displacer molecule is displacing

76 analytes with lower affinity as soon as the displacer molecule enters the column. An analyte, which is

77 displaced by the displacer, is then displacing a further different analyte, having a lower affinity then

78 the analyte with the highest affinity. This process is continued throughout all bands of the analytes. If

79 the displacer is pumped onto the column and thereby occupying more and more binding sites, the

80 bands of the analytes move down the column, forming the "displacement train". The separation is

81 finished when all binding sites are saturated by the displacer [18]. Elution profiles of displacement

82 chromatography are not Gaussian peaks as in gradient elution but sharp rectangular zones containing

83 analytes of high purity and high concentration. Displacement chromatography is well suited for the

84 preparative purification of synthesis products for removal of a minor amounts of purities from the

85 target molecule [19]. The power of displacement in the separation of complex protein mixtures and to

86 enrich trace amounts of proteins was shown in several studies [20–22]. A comparison of displacement

87 versus gradient chromatography for the separation of a protein mixture clearly demonstrated the

88 advantages of displacement chromatography [23]. If an appropriate displacer molecule is difficult to

89 find, sample displacement chromatography (SDC) is an alternative, because in SDC no displacer is

90 needed. SDC has been used for the separation and purification for peptides [22] , proteins [24], and

91 human plasma protein fraction Cohn IV-4 [25]. In DC the quality of separation usually is better than in

92 the gradient chromatography mode [23], as well as the enrichment effect of trace components and

93 the concentrating effect yielding the highest concentration of analytes eluting from the stationary

94 phase compared to any other elution modes [26–28]. Sample displacement chromatography using

95 packed columns is requiring the use of segmented columns connected in series. During the loading of

96 the columns the sample molecules will bind to the stationary phase of the columns according to their

97 affinity: On the first column the molecules with the highest affinity will bind. If it is the aim to

98 fractionate a total protein amount of less than 0.2 g it will be difficult to find appropriate columns

99   because the smallest commercially available columns are containing 1 ml of resin. In case of a binding

100   capacity of 200 mg/ml all proteins may bind to the first column and no separation will be obtained.

101   With batch chromatography, resin in suspension in reaction vials, this problem can be circumvented

102   because segments with resins much smaller than 1 ml are realizable. Sample displacement batch

103   chromatography (SDBC) was already shown to be a powerful method for protein fractionation [24]. A

104   detailed review of sample displacement chromatography for protein separation can be found in [22].

105   The Structural Genomics Consortium proposed in 2008 the following protein purification routine for

106   recombinant proteins: proteins should be loaded onto the column until the column is fully saturated,

107   implying that the amount of protein must slightly exceed the binding capacity of the column resin. All

108   components with a lower affinity to the stationary phase would be displaced by the target protein [20].

109   In theory, this approach seems straight forward and promising on the first view. However, if a part of

110   the impurities has a higher affinity than the target protein, its purification will not work, because the

111   impurities are not displaced by the target protein. Furthermore, it is difficult to estimate the binding

112   capacity of the target protein to the chosen column. If the applied amount of the target protein is too

113   small, there is the risk that also impurities with a lower affinity than that of the target protein will bind.

114   Here, we demonstrate that fractionation of proteoforms with SDBC is possible and advantageous. The

115   success of the fractionation was monitored with quantitative mass spectrometry detecting intact

116   proteins. The presented method offers the utility and possibility of SDBC for simple and effective

117   separation of proteoforms. In this study, only 1 µL of chromatography resin and 1 mg of sample were

118   used in SDBC. As a model sample protein ovalbumin from chicken egg white was used as its

119   composition is representing a large diversity of proteoforms. Using optimised conditions of SDBC, a

120   two-fold increase in the number of detectable proteoforms in comparison to the original sample and

121   significant enrichment of low-abundant basic ovalbumin proteoforms were obtained.

122   **Materials and methods**

123   1.1   Chemical

124   Strong cation exchange chromatography resins Eshmuno CPX was provided by Merck KGaA

125   (Darmstadt, Germany). Ovalbumin from chicken egg white as model sample was purchased from

126   Sigma-Aldrich (St. Louis, USA). Sodium formate, sodium acetate, and LC-MS grade water were

127   purchased from Merck KGaA (Darmstadt, Germany). Formic acid, acetic acid, dithiothreitol (DTT),

128   iodoacetamide (IAA), dimethyl sulfoxide (DMSO), NaCl, and sulfolane were purchased from Sigma-

129   Aldrich (St. Louis, USA). Trypsin and PNGase F were purchased from Promega (Mannheim, Germany).

130   1.2   Sample Displacement Batch Chromatography

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

153
154 1.3   Mass spectrometric quantitation of intact proteoforms

155 All samples collected from SDBC experiments were desalted with 100% HPLC-grade water using 3K
156 centrifugal filter 0.5 mL (Amicon Ultra, Merck KGaA, Darmstadt, Germany) for 6 times. 5% sulfolane
157 was added to each sample as a supercharging agent. All samples were then analysed by mass
158 spectrometry using flow injection analysis (FIA-MS). Therefore, LC autosampler from Waters ACQUITY
159 UPLC system (Waters, Manchester, UK) was used to directly inject 1.5 µL of each sample into the MS.
160 A hybrid-quadrupole-orbitrap mass spectrometry (Q-Exactive, Thermo Scientific,Bremen, Germany)
161 was used. Mobile phase of 100% water was used for 4 minutes run in a flow rate of 0.075 mL/min for
162 1.5 minutes, then 0.1 mL/min for 1.5 minutes and back to 0.075 mL/min for 1 minute. MS parameters
163 were set at scan range 1200 to 3600, resolution was set to 17500, AGC target was 5×10^{6}, micro scan
164 count was 4, capillary temperature was set 275°C and in-source CID was 30eV.

165   1.4 N-glycan analysis

166   N-glycans of all ovalbumin proteoforms of the different SDBC fractions were released and analysed

167   separately. To obtain the free N-glycans, proteoforms were reduced by 20mM dithiothreitol (DTT) at

168   56°C for 30 min and alkylated by 40mM iodoacetamide (IAA) at room temperature for 30 min in the

169   dark. Buffer exchange were then performed using 3K centrifugal filter 0.5 mL (Amicon Ultra) to a 100

170   mM ammonium bicarbonate solution. 1:30 (v/w) PNGase F was added and incubated at 37°C for 24 h

171   to release N-glycans, followed by addition of 1:100 (w/w) trypsin and incubated for another 20 h at

172   37°. N-glycans and tryptic peptides were separated using a RP-SPE C18 cartridge as described by

173   Morelle at al.,[30]. The purified N-glycans were then permethylated using optimized solid phase

174   permethylation method. For all methods of N-glycans permethylation, N-glycans LC-MS analysis, and

175   N-glycans data analysis were performed using method developed and optimized by by Guan, Y., et. al.,

176   [31].

177   1.5 Identification and relative quantitation of the proteoforms using Intact mass spectrometric data

178   The MS1 spectra was deconvoluted using UniDec GUI deconvoluted software ver. 4.0.2 from Marty et.

179   al., [32]. Parameters were set at data processing m/z range from 1500 to 3000, charge range from 10

180   to 35, sample mass every 0.5 Da, peak detection range was 15 Da, and peak detection threshold was

181   set to 0.15. The relative quantities of the proteoforms were determined with UniDec using the

182   signal intensities of the deconvoluted spectra. Results were exported and processed further in Excel

183   spreadsheet. From the observed experimental masses, identification of ovalbumin proteoforms was

184   performed by mass-matching

185   **2.   Results**

186   2.1 Parameter screening for SDBC

187   For successful proteoform purification, the most optimal conditions must be determined prior to the

188   SDBC by performing protein purification parameter screening (PPS). Parameters, including different

189   pH values of sample application buffer, different resins, and different additional sodium chloride in

190   sample application buffer were screened. Details about PPS results are given in the supplement. PPS

191   conditions yielding the best results with respect to the number of detected proteoforms was chosen

192   for SDBC, comprising a combination of strong cation exchange resin Eshmuno CPX and sample

193   application buffer of 25 mM acetate buffer with a pH 5. Eshmuno CPX resin performed better than

194   other resins, which may be due to the combination of high cross-linked rigid matrix that causes the

195   proteoforms to bind stronger and the tentacle structure in the resin which allows multi-point

196   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

263 a lower affinity to the resin. These proteoforms are significantly purified from the other proteoforms

264 eluting in the fractions 1 - 3. The identification of the proteoforms in both categories revealed that

265 highly enriched proteoforms in early fraction are most probably ovalbumin with one N-glycan, one

266 acetyl, and one phosphate attached on it, while the proteoforms in the later fractions are most

267 probably ovalbumin with one N-glycan, one acetyl, and two phosphates attached on it.

268 Phosphorylation of these proteoforms reduce their isoelectric point making them less basic. This gives

269 a proof that SDBC is effective even for the separation of closely related proteoforms from a single

270 protein.

271 In addition, with SDBC it was possible to remove proteoforms not belonging to the ovalbumin

272 proteoform family. By bottom-up proteomics of the proteins in the flow-through-fraction ovomucoid

273 proteoforms beside some proteoforms of ovalbumin as main abundant proteins were identified

274 (Figure S3; details are given in the supplement). The ovalbumin proteoforms detected in the flow-

275 through fraction may have low isoelectric points and therefore did not bind to the CEX resin.

276 **3. Discussion**

277 In this study it was demonstrated that SDBC is suitable for fractionation of proteoforms. SDBC

278 simplifies the set-up of typical SDC which is requiring the connection of several columns in series. SDBC

279 can be easily performed in different devices, such as microtiter 384-wells plate, 96 deep well-plates or

280 reaction vials. Using cost effective, simple, and easy to handle batch chromatography performed in

281 reaction vials, low-abundant ovalbumin proteoforms were enriched. The separation of different

282 proteoforms of ovalbumin into different fractions, thereby reducing the complexity of the composition

283 of proteoforms, resulted in a two-fold increase of the number of detected proteoforms compared to

284 the analysis of original ovalbumin sample. Compared to previous attempts on proteoforms

285 purification, the need of high sample amounts in a typical DC or SDC was eliminated. In a study by

286 Zhang et al. [36] using DC to separate antibody charge variants using a cation exchange column 200

287 mg of monoclonal antibody was needed, while another study by Khanal et al. [37] has used around 10

288 mg of monoclonal antibody for one SDC experiment using in-house-packed columns. This need of high

289 amount of sample may hinder the experiment when dealing with limited amount of sample. In our

290 method, only 1 µL of resin per segment and 1 mg of ovalbumin sample were used thus giving benefit

291 when dealing with precious samples with limited availability. If a preparative scale is desired, the scale-

292 up of a SDBC method also is straightforward and easy [25, 29]. SDBC offers an economical benefit as it

293 is much less expensive than preparative column gradient LC, because with SDBC no expensive gradient

294 pumps are required and the binding capacity of the stationary phase is completely used whereas in

295 gradient chromatography it is recommendable not to exceed 20 % of the binding capacity. In this study

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

297

298   **Conclusion**

299   In summary, we demonstrated that fractionation of proteoforms by liquid chromatography, based on

300   sample displacement batch chromatography (SDBC) is possible. As a result of the successful

301   fractionation the total number of proteoforms detectable with mass spectrometry significantly

302   increased to 42 compared to 16 proteoforms in the original ovalbumin sample prior to its fractionation.

303   With SDBC enrichment of low-abundant ovalbumin proteoforms was obtained. We also successfully

304   separated some of ovalbumin proteoforms which proves that SDBC is effective even for the separation

305   of closely related proteoforms. Despite several challenges regarding the complexity of proteoforms

306   and limitation regarding the availability of well-established analytical system, we demonstrated the

307   utility and possibility of SDBC for simple and effective separation of proteoforms. This in the future can

308   be used for removal of harmful proteoforms and as a tool for enrichment of low abundant proteoforms

309   for further analysis or to support and enhance molecular characterization.

310   **Declaration of competing interest**

311   The authors declare that they have no known competing financial interests or personal relationships

312   that could have appeared to influence the work reported in this paper.

321

322   **References**

323

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

333   M. A. (2018). Structure-based stabilization of insulin as a therapeutic protein assembly via enhanced
334   aromatic–aromatic interactions. *Journal of Biological Chemistry*, *293*(28), 10895–10910.
335 6. Hassiki, R., Behi, J., Said, N. ben, Boujbel, L., & Bouhaouala-Zahar, B. (2017). Biochemical and
336   Biological Quality Control of Two Recombinant Human Erythropoietin Biosimilar Products. *Current
337   Pharmaceutical Analysis*, *13*(6), 566–573.
338 7. Goh, J. B., & Ng, S. K. (2018). Impact of host cell line choice on glycan profile. *Critical Reviews in
339   Biotechnology*, *38*(6), 851–867.
340 8. Tyshchuk, O., Gstöttner, C., Funk, D., Nicolardi, S., Frost, S., Klostermann, S., … Mølhøj, M. (2019).
341   Characterization and prediction of positional 4-hydroxyproline and sulfotyrosine, two post-
342   translational modifications that can occur at substantial levels in CHO cells-expressed
343   biotherapeutics. *mAbs*, *11*(7), 1219–1232.
344 9. Gervais, D. (2016). Protein deamidation in biopharmaceutical manufacture: Understanding, control
345   and impact. *Journal of Chemical Technology and Biotechnology*, *91*(3), 569–575.
346 10. Beck, A., & Liu, H. (2019). Macro- and Micro-Heterogeneity of Natural and Recombinant IgG
347   Antibodies. *Antibodies*, *8*(1), 18.
348 11. Moss, C. X., Matthews, S. P., Lamont, D. J., & Watts, C. (2005). Asparagine deamidation perturbs
349   antigen presentation on class II major histocompatibility complex molecules. *Journal of Biological
350   Chemistry*, *280*(18), 18498–18503.
351 12. Liu, Y. D., van Enk, J. Z., & Flynn, G. C. (2009). Human antibody Fc deamidation in vivo. *Biologicals*,
352   *37*(5), 313–322.
353 13. Lee, H. H. L., Heo, C. E., Seo, N., Yun, S. G., An, H. J., & Kim, H. I. (2018). Accurate Quantification of
354   N-Glycolylneuraminic Acid in Therapeutic Proteins Using Supramolecular Mass Spectrometry.
355   *Journal of the American Chemical Society*, *140*(48), 16528–16534.
356 14. Hermeling, S., Crommelin, D. J. A., Schellekens, H., & Jiskoot, W. (2004). Structure-immunogenicity
357   relationships of therapeutic proteins. *Pharmaceutical Research*, *21*(6), 897–903.
358 15. Riguero, V., Clifford, R., Dawley, M., Dickson, M., Gastfriend, B., Thompson, C., … O'Connor, E. (2020).
359   Immobilized metal affinity chromatography optimization for poly-histidine tagged proteins. *Journal
360   of Chromatography A*, *1629*.
361 16. Kopp, J., Zauner, F. B., Pell, A., Hausjell, J., Humer, D., Ebner, J., … Pell, R. (2020). Development of a
362   generic reversed-phase liquid chromatography method for protein quantification using analytical
363   quality-by-design principles. *Journal of Pharmaceutical and Biomedical Analysis*, *188*.
364 17. McAtee, C. P. (2010). Displacement Chromatography of Proteins. *Current Protocols in Protein
365   Science*, (SUPPL. 59), 1–14.
366 18. Evans, S. T., Freed, A., & Cramer, S. M. (2009). Displacer concentration effects in displacement
367   chromatography: Implications for trace solute detection. *Journal of Chromatography A*, *1216*(1),
368   79–85.
369 19. Jankowski, ¨J, Potthoff, W., Zidek, W., & Schluter¨medizinische, H. (1998). *Purification of chemically
370   synthesised dinucleoside(59,59) polyphosphates by displacement chromatography*. *Journal of
371   Chromatography B* (Vol. 719).
372 20. Gräslund, S., Nordlund, P., Weigelt, J., Hallberg, B. M., Bray, J., Gileadi, O., … Gunsalus, K. C. (2008).
373   Protein production and purification. *Nature Methods 2008 5:2*, *5*(2), 135–146.
374 21. Josic, D., Breen, L., Clifton, J., Gajdosik, M. S., Gaso-Sokac, D., Rucevic, M., & Müller, E. (2012).
375   Separation of proteins from human plasma by sample displacement chromatography in
376   hydrophobic interaction mode. *ELECTROPHORESIS*, *33*(12), 1842–1849.
377 22. Srajer Gajdosik, M., Clifton, J., & Josic, D. (2012). Sample displacement chromatography as a method
378   for purification of proteins and peptides from complex mixtures. *Journal of Chromatography A*,
379   *1239*, 1–9.
380 23. Ahrends, R., Lichtner, B., Buck, F., Hildebrand, D., Kotasinska, M., Kohlbacher, O., … Schlüter, H.
381   (2012). Comparison of displacement versus gradient mode for separation of a complex protein
382   mixture by anion-exchange chromatography. *Journal of Chromatography B: Analytical Technologies
383   in the Biomedical and Life Sciences*, *901*, 34–40.
384 24. Kotasińska, M., Richter, V., Thiemann, J., & Schlüter, H. (2012). Cation exchange displacement batch
385   chromatography of proteins guided by screening of protein purification parameters. *Journal of
386   Separation Science*, *35*(22), 3170–3176.

387 25. Heikaus, L., Hidayah, S., Gaikwad, M., Kotasinska, M., Richter, V., Kwiatkowski, M., & Schlüter, H.
388   (2021). Sample displacement batch chromatography of proteins. In *Methods in Molecular Biology*
389   (Vol. 2178, pp. 285–299). Humana Press Inc.
390 26. Horváth, C., Nahum, A., & Frenz, J. H. (1981). High-performance displacement chromatography.
391   *Journal of Chromatography A*, *218*(C), 365–393.
392 27. Trusch, M., Tillack, K., Kwiatkowski, M., Bertsch, A., Ahrends, R., Kohlbacher, O., … Schlüter, H.
393   (2012). Displacement chromatography as first separating step in online two-dimensional liquid
394   chromatography coupled to mass spectrometry analysis of a complex protein sample-The proteome
395   of neutrophils. *Journal of Chromatography A*, *1232*, 288–294.
396 28. Kwiatkowski, M., Krösser, D., Wurlitzer, M., Steffen, P., Barcaru, A., Krisp, C., … Schlüter, H. (2018).
397   Application of Displacement Chromatography to Online Two-Dimensional Liquid Chromatography
398   Coupled to Tandem Mass Spectrometry Improves Peptide Separation Efficiency and Detectability
399   for the Analysis of Complex Proteomes. *Analytical Chemistry*, *90*(16), 9951–9958.
400 29. Kotasinska, M., Richter, V., Kwiatkowski, M., & Schlüter, H. (2014). Sample displacement batch
401   chromatography of proteins. *Methods in Molecular Biology*, *1129*, 325–338.
402 30. Morelle, W., & Michalski, J. C. (2007). Analysis of protein glycosylation by mass spectrometry. *Nature
403   Protocols*, *2*(7), 1585–1602.
404 31. Guan, Y., Zhang, M., Wang, J., & Schlüter, H. (2021). Comparative Analysis of Different N-glycan
405   Preparation Approaches and Development of Optimized Solid-Phase Permethylation Using Mass
406   Spectrometry. *Journal of Proteome Research*, *20*(5), 2914–2922.
407 32. Marty, M. T., Baldwin, A. J., Marklund, E. G., Hochberg, G. K. A., Benesch, J. L. P., & Robinson, C. V.
408   (2015). Bayesian deconvolution of mass and ion mobility spectra: From binary interactions to
409   polydisperse ensembles. *Analytical Chemistry*, *87*(8), 4370–4376.
410 33. Füssl, F., Criscuolo, A., Cook, K., Scheffler, K., & Bones, J. (2019). Cracking Proteoform Complexity of
411   Ovalbumin with Anion-Exchange Chromatography-High-Resolution Mass Spectrometry under
412   Native Conditions. *Journal of Proteome Research*, *18*(10), 3689–3702.
413 34. Thompson, E. O. P., & Fisher, W. K. (1978). *A Correction and Extension of the Acetylated Amino
414   Terminal Sequence of Ovalbumin*. *J. Bioi. Sci* (Vol. 31).
415 35. Nakayoshi, T., Kato, K., Fukuyoshi, S., Takahashi, O., Kurimoto, E., & Oda, A. (2021). Molecular
416   mechanisms of succinimide formation from aspartic acid residues catalyzed by two water molecules
417   in the aqueous phase. *International Journal of Molecular Sciences*, *22*(2), 1–11.
418 36. Zhang, T., Bourret, J., & Cano, T. (2011). Isolation and characterization of therapeutic antibody
419   charge variants using cation exchange displacement chromatography. *Journal of Chromatography
420   A*, *1218*(31), 5079–5086.
421 37. Khanal, O., Kumar, V., Westerberg, K., Schlegel, F., & Lenhoff, A. M. (2019). Multi-column
422   displacement chromatography for separation of charge variants of monoclonal antibodies. *Journal
423   of Chromatography A*, *1586*, 40–51.
424 38. Perez-Riverol, Y., Bai, J., Bandla, C., Hewapathirana, S., García-Seisdedos, D., Kamatchinathan, S.,
425   Kundu, D., Prakash, A., Frericks-Zipper, A., Eisenacher, M., Walzer, M., Wang, S., Brazma, A., Vizcaíno,
426   J.A. (2022). The PRIDE database resources in 2022: A Hub for mass spectrometry-based proteomics
427   evidences. *Nucleic Acids Res 50*(D1), D543-D552
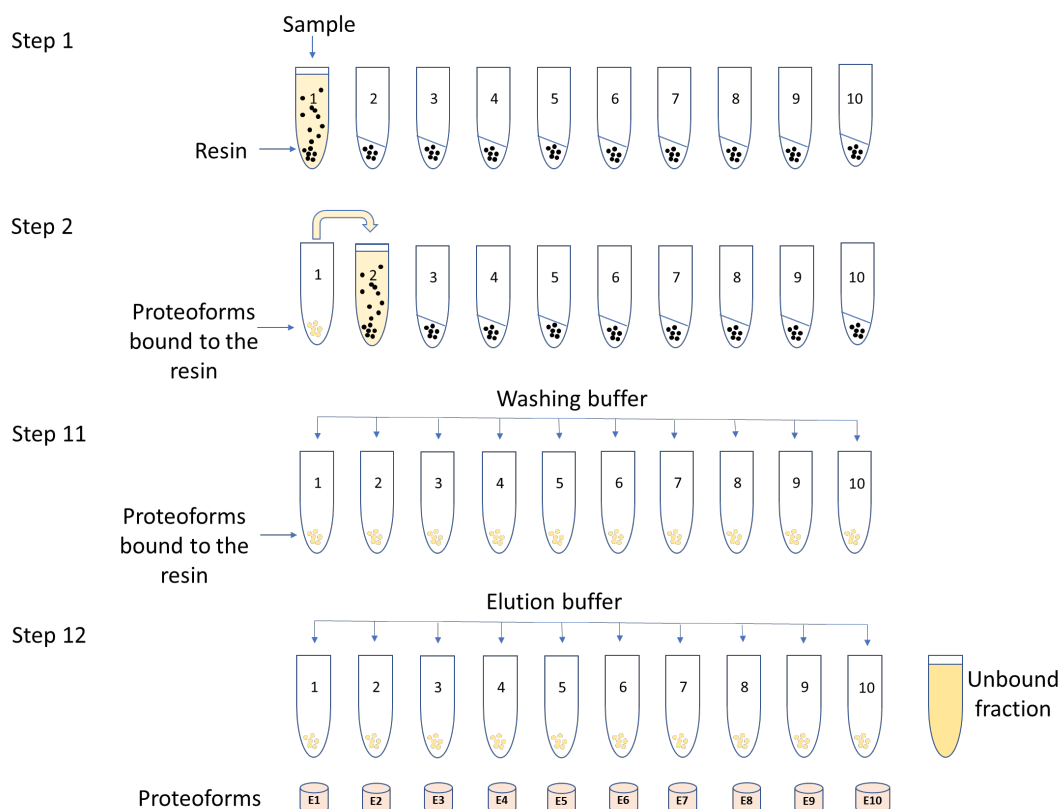428   .

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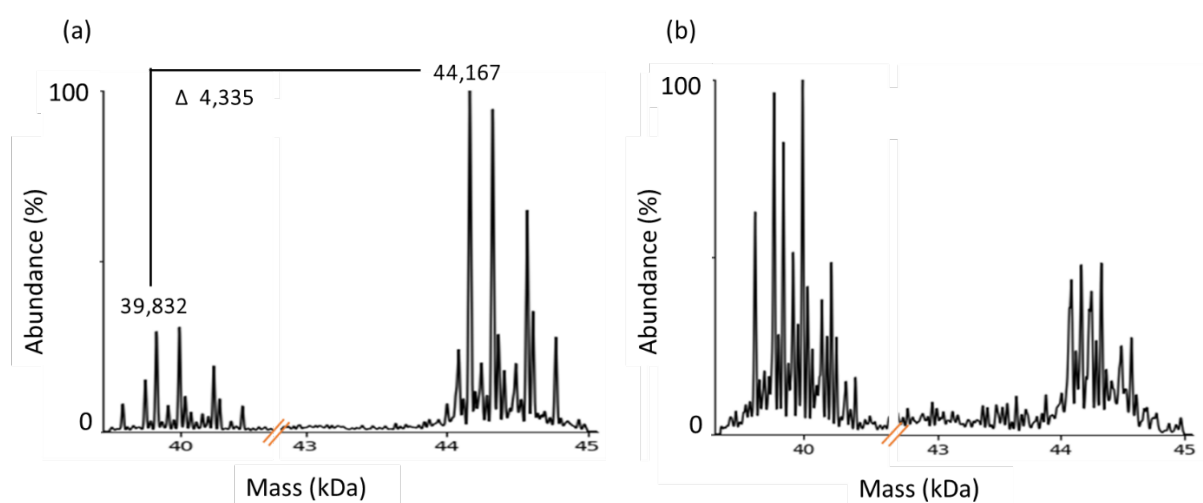
Figure 2. Deconvoluted spectra of ovalbumin from the original sample (**a**) and from the SDBC fraction 1 (**b**).

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		G	HexNAc3Hex5
18b	44167	GPPA	HexNAc2Hex5
18c		GPP	HexNAc3Hex4
19a		G	HexNAc4Hex4
19b	44208	GPPA	HexNAc3Hex4
19c		GPP	HexNAc4Hex3
19d		GP	HexNAc2Hex6
20a	44231	GPS	HexNAc3Hex5
20b		GPAS	HexNAc2Hex6
20c		GPPAS	HexNAc4Hex3
21a		GPA	HexNAc2Hex6
21b	44248	GPPA	HexNAc4Hex3
21c		GP	HexNAc3Hex5
22a		G	Neu5Ac1HexNAc3Hex4
22b	44289	GPA	HexNAc3Hex5
22c		GPP	HexNAc2Hex6
22d		GP	HexNAc4Hex4
23a		GA	Neu5Ac1HexNAc3Hex3Fuc1
23b	44328	GPPA	HexNAc2Hex6
23c		GPP	HexNAc3Hex5
24a	44370	G	HexNAc4Hex5
24b		GPPA	HexNAc3Hex5

24c		GPP	HexNAc4Hex4
24d		GP	Neu5Ac1HexNAc3Hex4
25a		GPPA	HexNAc4Hex3Fuc1
25b	44394	GPPS	HexNAc5Hex3
25c		GPPAS	HexNAc4Hex4
26a		G	HexNAc4Hex4Red-HexNAc1
26b		GPA	Neu5Ac1HexNAc3Hex4
26c	44411	GPPA	HexNAc4Hex4
26d		GPP	HexNAc5Hex3
26e		GP	HexNAc3Hex6
27a		GPA	HexNAc4Hex4Fuc1
27b	44476	GPS	HexNAc5Hex4
27c		GPAS	HexNAc4Hex5
27d		GPPAS	Neu5Ac1HexNAc3Hex4
28a		GPA	HexNAc4Hex5
28b	44493	GPPA	Neu5Ac1HexNAc3Hex4
28c		GPP	HexNAc3Hex6
28d		GA	HexNAc5Hex4
29a	44531	GPPA	HexNAc3Hex6
29b		GPP	HexNAc4Hex5
30a		G	HexNAc5Hex5
30b	44573	GPPA	HexNAc4Hex5
30c		GPP	HexNAc5Hex4
31a		GPPA	HexNAc5Hex4
31b	44613	GPP	HexNAc6Hex3
31c		GP	HexNAc4Hex6
32a		GA	HexNAc7Hex3
32b	44697	GPA	HexNAc5Hex5
32c		GPP	HexNAc4Hex6
32d		GP	HexNAc6Hex4
33a		G	HexNAc6Hex5
33b	44776	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.
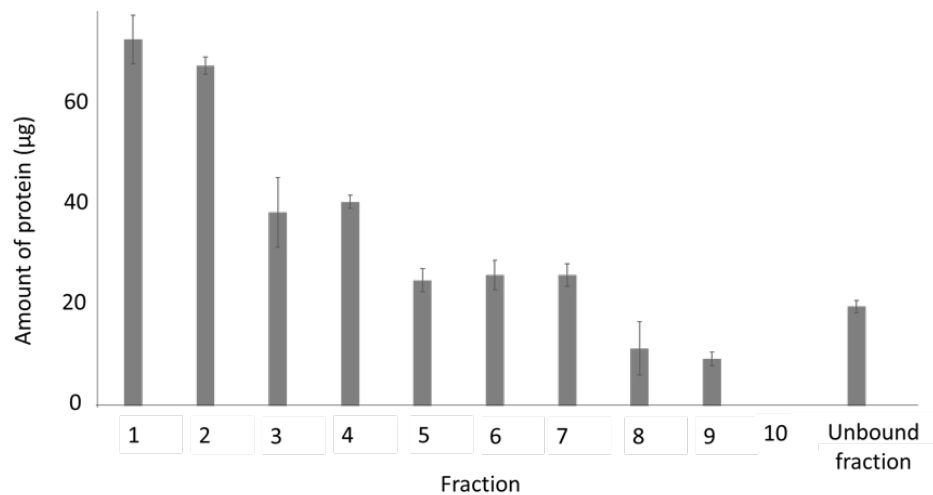
455

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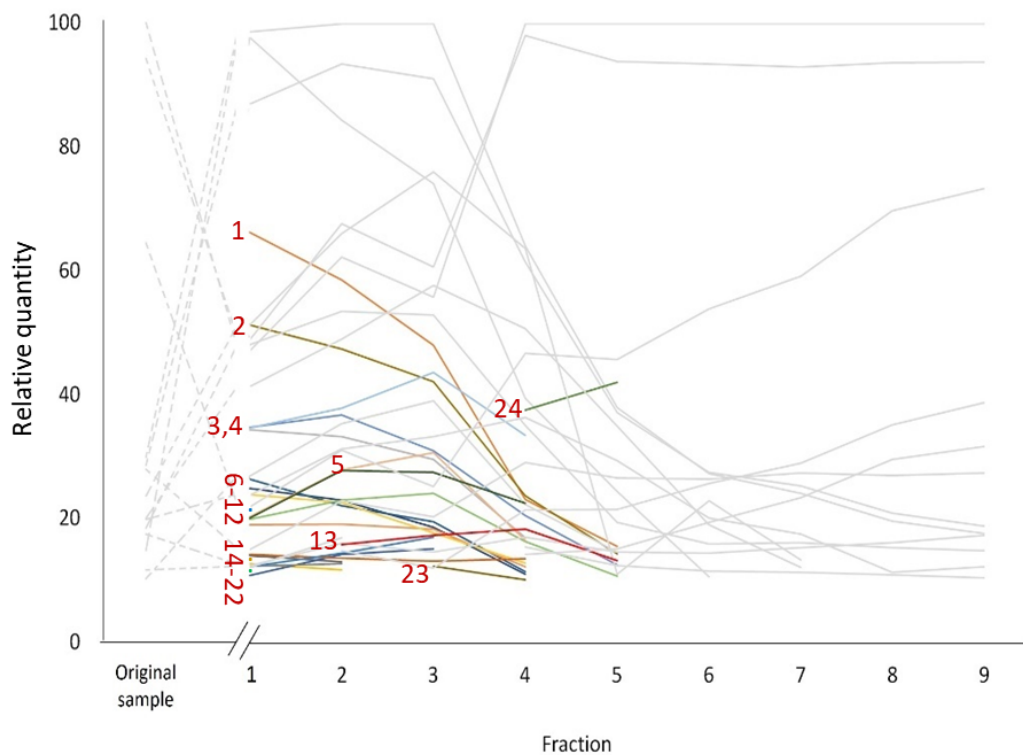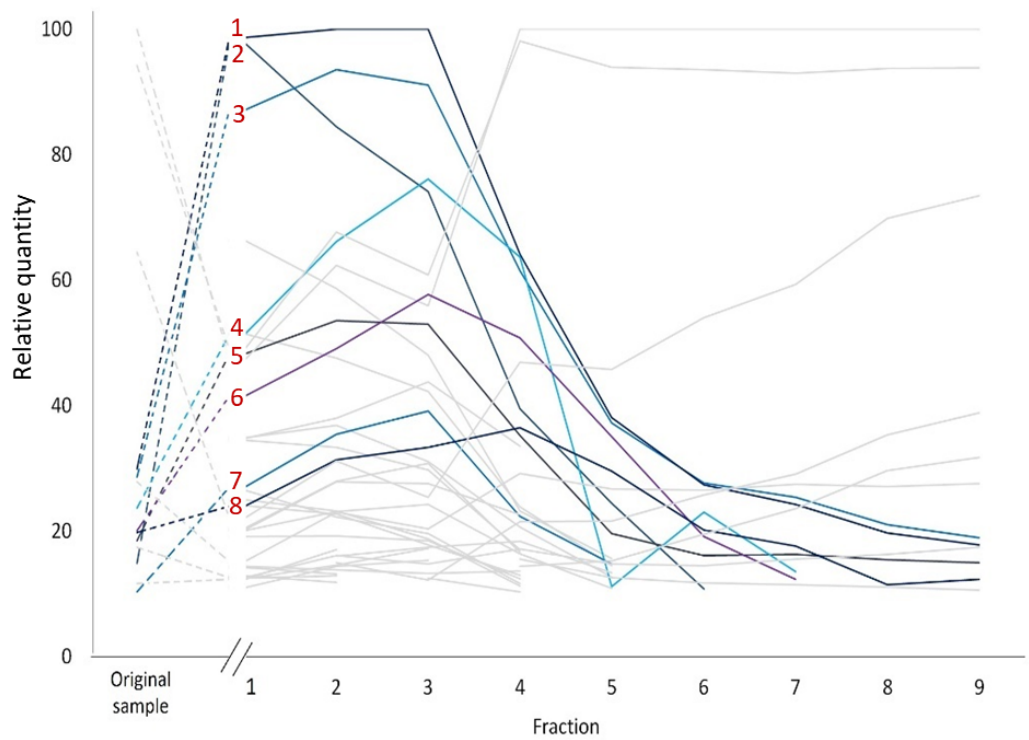

Figure 4. Chromatograms of proteoforms detected in the ovalbumin original sample and in the eluates of SDBC fractions. Highlighted in color-lines are proteoforms enriched and detected in SDBC early fractions only and not in the original sample. All other proteoforms are presented in grey lines. Mass 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.

16

474

475
476
477 **Figure 5.** Chromatograms of proteoforms detected in the ovalbumin original sample and in the eluates of SDBC
478 fractions, highlighting significantly enriched proteoforms represented in color-lines. All other proteoforms are
479 presented in grey-lines. Mass of proteoform 1-8: 39995; 39753; 39832; 44086; 44248; 40237; 44288; 44493 Da.
480 Note that in the eluate of segment 10 no signals of any proteoform were detected.
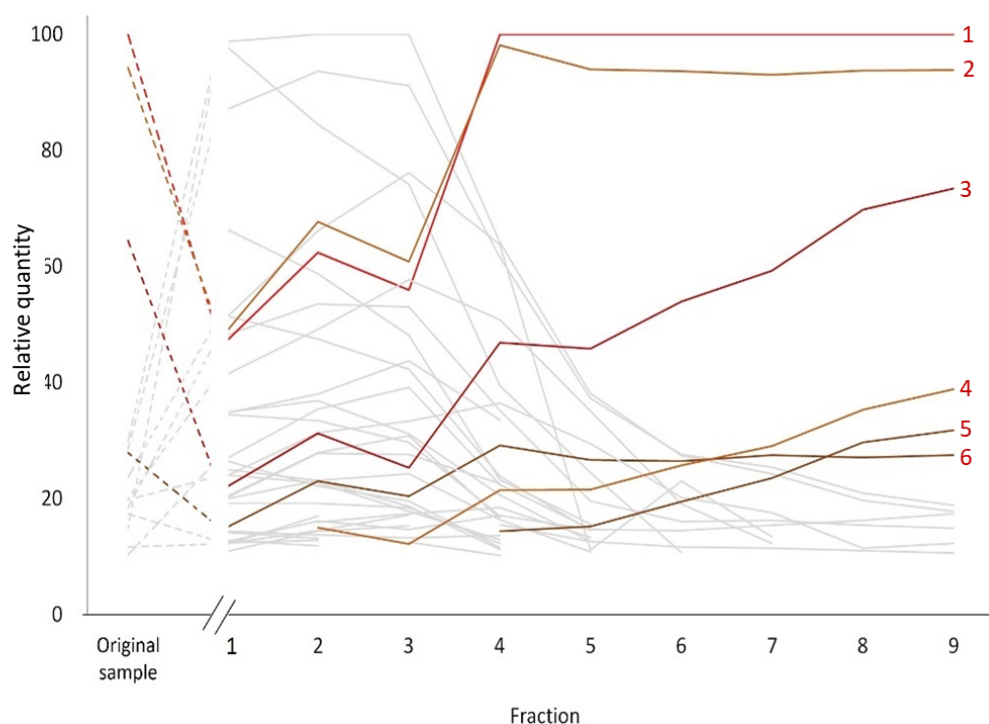481
482
483
484
485
486
487
488

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.