UPTEC X 05 032
JUN 2005

ISSN 1401-2138

THERESE GRANÉR

Understanding chromatographic behaviour of glycosylated proteins

Master’s degree project

UPPSALA UNIVERSITET

Molecular Biotechnology Programme

Uppsala University School of Engineering

UPTEC X 05 032	**Date of issue 2005-06**

Author

Therese Granér

Title (English)

Understanding chromatographic behaviour of glycosylated proteins

Title (Swedish)

Abstract

The chromatographic behaviour of glycosylated proteins was studied by separating protein/glycoprotein pairs on different chromatographic media. Two pairs were used: Ribonuclease A/ribonuclease B from bovine pancreas and deglycosylated avidin/native avidin from hen egg white. Avidin was successfully deglycosylated by α-mannosidase and endoglycosidase H treatment. Separation by cation chromatography at pH 4 resulted in an excellent separation of the two pairs. Cationic data were successfully predicted by cation exchange Quantitative Structure-Property Relationships (QSPR). The QSPR model predicted higher retention for non-glycosylated proteins than for glycoproteins. Separation by size exclusion chromatography showed that the apparent molecular weight difference by size exclusion chromatography is larger than the theoretical molecular weight difference.

Keywords

Glycoproteins, avidin, ribonuclease A, ribonuclease B, deglycosylation, chromatography, separation, Quantitative Structure-Property Relationships

Supervisors

Jinyu Zou and Ulrika Hjellström-Nilsson

R&D protein separations, GE Healthcare, Uppsala

Scientific reviewer

Jerry Ståhlberg

Department of Molecular Biology, Swedish University of Agricultural Sciences, Uppsala

Project name	Sponsors
Language **English**	Security
ISSN 1401-2138	Classification
Supplementary bibliographical information	Pages **41**

Biology Education Centre Biomedical Center Husargatan 3 Uppsala

Box 592, SE-75124 Uppsala Tel +46 (0)18 4710000 Fax +46 (0)18 555217

Understanding chromatographic behaviour of glycosylated proteins

Therese Granér

Sammanfattning

Informationen om hur proteiner skall bildas finns lagrad i DNA. Där finns noggranna ritningar om hur enskilda aminosyror skall sättas samman för att bilda proteiner. Att länka samman aminosyrorna är dock inte tillräckligt för att proteinerna skall nå sin fulla komplexitet. För att skapa större variation och funktionalitet kan aminosyrakedjorna genomgå ett flertal förändringar.

En mycket vanlig förändring hos proteiner är att sockerkedjor fästs på proteinet. Sockerkedjorna varierar till både storlek och form och består av många olika sockertyper. Ett visst protein kan oftast binda flera olika sockerkedjor, vilket innebär att ett antal olika glykoformer kan skapas från ett enda protein.

Olika glykoformer kan ha olika funktion och aktivitet. Detta gör att det finns ett behov av att kunna separera glykoformerna. Ett vanligt sätt att separera proteiner är att använda sig av kromatografiska packningsmaterial. Olika proteiner binder olika hårt till dessa packningsmaterial. När packningsmaterialet tvättas lossnar proteinerna och en separation uppstår beroende på hur hårt proteinerna bundit.

I denna studie har kromatografiska packningsmaterial med olika egenskaper använts för att separera proteiner med och utan socker. Packningsmaterialet som separerar med avseende på laddning gav den bästa separationen. Även packninsmaterial som separerar med avseende på storlek verkar lovande.

Examensarbete 20 p
Civilingenjörsprogrammet i molekylär bioteknik
Uppsala universitet, juni 2005

Table of Contents

1 Introduction

1.1 Glycosylation

Glycoproteins are formed by covalent attachment of carbohydrates to proteins, by a process called glycosylation. Glycosylation is the most extensively occurring natural modification in higher organisms [1] and it is both species and tissue specific. The glycosylation can be influenced by many physiological changes and diseases [2].

The majority of eukaryotic proteins are glycoproteins [3]. This is consistent with the presence of specific glycosylation sites on most proteins. The degree of glycosylation varies at a given glycosylation site and this can lead to heterogeneity in the glycoforms [2]. Glycoproteins exist as a diverse population of glycoforms, all of which share an identical protein backbone but are dissimilar in their carbohydrate structure or disposition. The number of different glycoforms for different glycoproteins varies from one to several thousand. One example of the enormous complexity of glycoproteins is recombinant tissue plasminogen activator (rtPA) which has been estimated to contain as many as 11,500 different glycoforms. Usually the number of glycoforms is much lower [4].

1.1.1 The Glycan Chain

The sugar chains of glycoproteins are built up from a number of different monosaccharides. The most common monosaccharides found in higher organisms are listed below.

- *Sialic acids*: Nine-carbon acidic sugars with a negative charge. *Example*: N-acetyl-neuraminic acid (NeuNAc).
- *Hexoses*: Six-carbon neutral sugars. Examples: Glucose (Glc), galactose (Gal) and mannose (Man).
- Hexosamines: Hexoses with an amino group at the 2-position. The amino group can be N-acetylated or free. *Examples*: N-acetylglucosamine (GlcNAc) and N-acetyl-galactosamine (GalNAc).
- *Deoxyhexoses*: Hexoses without the hydroxyl group at the 6-position. *Example*: Fucose (Fuc).
- Pentoses: Five-carbon natural sugars. *Examples*: Xylose (Xyl) and L-rhamnose.
- *Uronic acids*: Hexoses with a negatively charged carboxylate at the 6-position. *Examples*: Glucuronic acid (GlcA) and Iduronic acid (IdA).

The linkage between different sugar monomers can be formed between several positions on the molecules and be of either α or β configuration, making oligosaccharides very complex. Three hexoses can generate between 1,056 and 27,648 different trisaccharides, whereas three amino acids can produce only six different tripeptides; however, glycoproteins contain monosaccharides in a limited number of the possible combinations [2].

1.1.2 Glycan Protein Linkages

Oligosaccharides are covalently linked to proteins in two major ways representing two different classes of glycoconjugates: N-linked and O-linked [5].

1.1.2.1 N-linked Glycans

N-linked glycans are attached to asparagine residues of a polypeptide chain within the consensus sequence Asn-X-Ser/Thr motif, where X can be any amino acid except proline [3]. The glycosylation of N-linked proteins is a co-translational process and it is estimated that the degree of glycosylation of this motif is 70-90%. Whether or not the site is glycosylated is probably due to the three-dimensional structure of the protein [6]. N-glycans share the common core pentasaccharide $Man_3(GlcNAc)_2$, and can be divided into three main classes: high-mannose type, hybrid type, and complex type (*Figure 1*) [3].

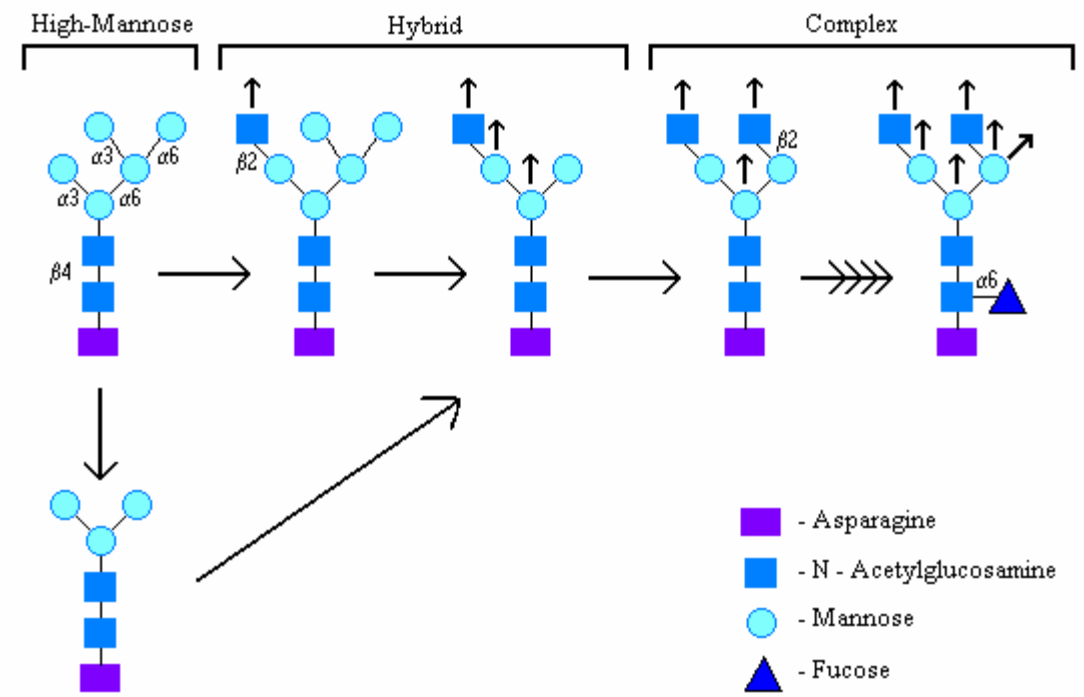

Figure 1. Diversification of the three classes of N-linked glycans found in vertebrates: high-mannose, hybrid, and complex type. The small arrows (↑) indicate where branch formation can occur. The illustration was adapted from Varki *et al.* [2].

In bacterial glycoproteins, N-linkages have also been characterized between asparagine and other sugars such as glucose, N-acetylgalactosamine and L-rhamnose [7].

1.1.2.2 O-linked Glycans

O-linked glycans are attached to proteins through an oxygen atom on the protein. The glycans are extremely diverse both in structure and function. They are attached to proteins in a number of different protein glycan linkages, in which GalNAc, fucose, GlcNAc, mannose, xylose, or galactose can be linked to serine, threonine, or hydroxylysine residues [5]. Usually, the glycan is linked to the protein via GalNAc to the hydroxyl group of serine or threonine [2]. The glycosylation of O-linked proteins is a post-translational process and does not, unlike N-linked glycans, require a consensus sequence. O-linked glycans are very heterogeneous [3] and are more common in higher organisms [2]. Due to their extensive heterogeneousity, O-linked glycans are generally classified by their core structure. A large number of different core structures exist [3].

1.1.3 Functions of Glycoproteins

There is no single predominant function of glycans. The diverse functions can be divided into at least five different classes: providing structural components, modifying protein properties, directing trafficking of glycoconjugates, mediating and modulating cell adhesion, and finally, mediating and modulating signalling [5].

The carbohydrate moiety of a glycoprotein can vary greatly, from being very minor in amount to being the dominant component of the molecule [2]. A large carbohydrate portion may result in shielding of functionally important areas of the protein to which they are attached. This might result in significant changes in the properties of the protein [6]. Studies of glycosylated and non-glycosylated proteins have revealed differences in solubility [8], thermal stability [9], and susceptibility towards proteolytic attack [6].

A correct working glycan biosynthesis is essential to humans. A number of different inherited diseases exist including carbohydrate-deficient glycoprotein syndrome (CDGS), leukocyte adhesion deficiency syndrome II (LAD II), congenital dyserythropoietic anemia type II (HEMPAS II), galactosemia, and abnormalities in proteoglycan synthesis. These diseases are in general very severe and result in malfunction of multiple organ systems [2, 10].

As time goes by, the function and structure of more and more useful glycoproteins are revealed. New glycoprotein drugs are being developed and an increasing number of proteins with therapeutic value are produced. Different glycoforms often have different activities and sometimes glycosylation of a protein results in a protein with an entirely different function than the native one. As a result of this, one variant of a glycoprotein might be excellent pharmaceutically for human treatment while others may appear immunogenic and toxic [7]. Taking these facts into consideration, the importance of developing fast and uncomplicated methods for separation of glycoproteins becomes evident. In order to develop such methods, the possibility to predict retention times of chromatographic separations would be of great value.

1.2 Aim

The aim of this study is to optimise chromatographic conditions for separating different glycoforms. Two different protein/glycoprotein pairs will be studied on different chromatographic systems. The protein/glycoprotein pairs have an identical peptide backbone but only one of the proteins has a carbohydrate moiety. Once the chromatographic data has been acquired, attempts will be made to understand the chromatographic behaviour of various descriptors based on the molecular properties of the proteins and their attached carbohydrates.

1.3 Choice of Proteins

A number of criteria have been considered in order to find suitable glycoproteins for the study. Suitable glycoproteins should be stable, inexpensive and have a well-characterized glycosylation pattern. In addition, the glycoproteins must either be available as proteins without carbohydrate moieties or, alternatively, be possible to deglycosylate. Two glycoproteins have been found that fulfil these criteria. The first one is ribonuclease (RNase) B from bovine pancreas and the second one is avidin from egg white.

1.3.1 Bovine Pancreatic Ribonuclease A and B

Bovine pancreatic ribonuclease A and B are two endoribonucleases of a family of at least four related molecules that seem to differ only in their glycan chains. They are very stable and have a isoelectric point (pI) of approximately 9.6. RNase A is the most predominant type, 87-95% of the total molecule (RNase A-D) in bovine pancrease, whereas RNase B only comprises 5-8% [11]. Both enzymes have the same specificity and hydrolyze RNA by cleaving phosphodiester bonds [5].

RNase A and B have an identical amino acid composition of 124 amino acids. The only structural difference between them is that RNase B is N-glycosylated with high-mannose type, whereas RNase A is not glycosylated at all [12]. Glycosylation of RNase has shown to decrease the activity while it increases the stability towards proteinases [1].

RNase B exists as five different glycoforms, Asn^{34}-$(GlcNAc)_2Man_{5-9}$ [6], with relatively flexible carbohydrate chains [13]. The glycosylation results in an increase of the molecular mass from 13,683 to 14,899-15,547 Da depending on the number of mannoses [1]. This is equivalent with a carbohydrate fraction of 8-12%. Earlier studies have shown that the glycan chains on RNase B, due to their dynamic nature and flexibility, shields charged residues on the surface of the protein. Shielding of charged residues makes the glycosylated protein appear less charged than the protein without carbohydrate moiety; thus causing it to elute earlier than RNase A on a cation exchange column [14]. The same study also showed that RNase A undergoes selective deamidation at Asn^{67}-Gly^{68} and that RNase A can form different oligomers, usually dimers. RNase B has shown similar characteristics.

1.3.2 Avidin from Hen Egg White

Avidin is a glycoprotein mostly known for its extremely high affinity for biotin [15]. Due to this characteristic, avidin is useful in many biotechnological applications, including purification, labelling and targeting of various materials [15]. Natural avidin is extremely stable, has a pI of approximately 10 and a molecular weight of 63,872 Da, as calculated from the amino acid composition and analysis of the carbohydrate moiety [16]. The structure is tetrameric consisting of four identical subunits each with 128 amino acid residues. One single glycosylation site exists on each subunit at Asn^{17}. The glycan chains exhibit extensive glycan microheterogeneity [17] and are mainly N-linked of high-mannose type [16]. The carbohydrate portion accounts for about 10% of the molecular mass and has shown to usually consist of four to five mannoses and three N-acetylglucosamine residues. It has also been reported that some glycopeptides contain small amounts of galactose [17].

1.3.2.1. Deglycosylation of Avidin

Unlike ribonuclease, no natural variant of hen egg white avidin without carbohydrate moiety exists. Hence, deglycosylation of the glycoprotein has to be performed. The removal of carbohydrates requires either enzymatic or chemical methods. Since chemical methods often can be harmful to proteins, enzymatic deglycosylation strategies is preferable [3]. In this study two different enzymes are used. Jack bean alpha-mannosidase readily cleaves off α-1.2, α-1.3, α-1.6 and more slowly α-1.4-linked terminal mannose residues. The purpose of using this enzyme is to increase the accessibility of the second enzyme, endoglycosidase H (Endo H). Endo H is specific for high-mannose and most hybrid types of glycans and cleaves

between the two N-acetylglucosamine residues of N-linked glycans, leaving one N-acetylglucosamine residue attached to the asparagines (*Figure 2*) [3].

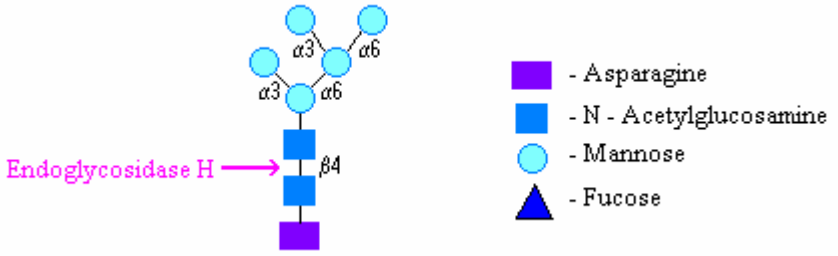

Figure 2. Endoglycosidase H is specific for high-mannose and most hybrid types of glycans and cleaves between the two N-acetylglucosamine residues of N-linked glycans, leaving one N-acetylglucosamine residue attached to the asparagines.

1.3.3 Ovalbumin from Hen Egg White

Ovalbumin is the main hen egg white protein. It is a 43 kDa glycoprotein of 385 amino acids with a pI of 4.5 [18]. Ovalbumin has two N-glycosylation sites at asparagines 292 and 311. Most glycans have a high-mannose and hybrid type. Glycans of complex type have also been reported [19].

1.4 Verification of Deglycosylation

A number of different methods can be used to control whether a deglycosylation experiment has been successful. Two commonly used methods are SDS-PAGE and MALDI-ToF MS. Both methods determine protein molecular weight before and after deglycosylation. The first method, SDS-PAGE, is easy to handle but gives a low resolution while the other method, MALDI-ToF MS, is more complex but has a better resolution and a much higher mass accuracy determination.

1.4.1. SDS-PAGE

SDS-PAGE stands for sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Proteins to be analysed by SDS-PAGE are mixed with sodium dodecyl sulfate (SDS), which is an anionic detergent that binds quantitatively to proteins causing them to denaturate and acquire a negative charge proportional to the number of amino acids in the protein. After denaturation, proteins are applied to a polyacrylamide gel in an electric field. The negatively charged proteins move through the electric field but encounter meanwhile resistance from the gel. Due to the resistance, smaller molecules are able to move faster than larger molecules, resulting in a separation according to molecular size. The molecular size is closely related to molecular weight. Molecular weight is determined by simultaneously running marker proteins of known molecular weight [20].

1.4.2. MALDI-ToF MS

Before analysis with Matrix-Assisted Laser Desorption and Ionization–Time-of-Flight Mass Spectrometry (MALDI-ToF MS), proteins are mixed with a specific, UV absorbing matrix and dried on a sample slide. As the matrix crystallises, proteins are incorporated into the crystals. The matrix is necessary to ensure that proteins become ionized when subjected to laser pulses [21].

The MALDI-ToF mass spectrometer consists of an ion source, a flight tube that separates the molecular ions according to their mass-to-charge ratio (m/z), and an ion detector (*Figure 3*). In the ion source, i.e. the sample slide and the laser, the crystals are subjected to nanosecond pulses of laser light causing part of the crystal's surface to be volatilized and ionized. The ionized molecules enter the flight tube, where they are accelerated by a strong electric field to a fixed kinetic energy.

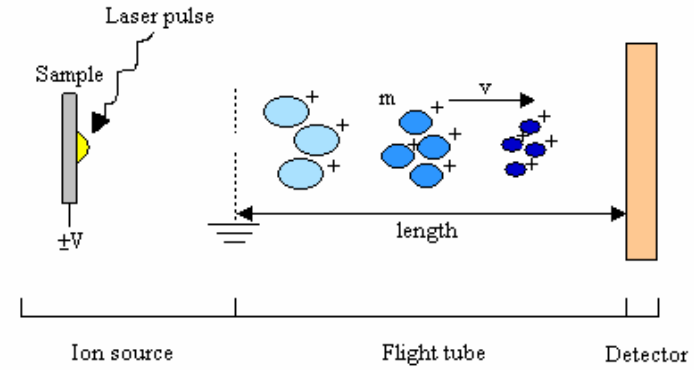

Figure 3. Schematic view of a Matrix-Assisted Laser Desorption/Ionization–Time-of-Flight Mass Spectrometer (MALDI-ToF MS). The illustration was adapted from GE Healthcare [21].

Fixed kinetic energy combined with varying molecular weights of the ionized proteins cause the proteins to travel with different velocities. Consequently, the time needed to reach the detector at the end of the flight tube differs for different proteins. In other words, the molecular weight of a protein can be determined, according to *Equation 1*, by simply measuring the time of flight [21].

$$m = \frac{2\,q\,V\,\Delta t_m^2}{l^2} \qquad \text{(Equation 1)}$$

where Δt_m = time between the application of the electrostatic field and the arrival of the ion at the detector, m = ion mass, q = ion charge, V = acceleration potential, and l = length of flight tube.

1.5 Separation by Different Chromatographic Techniques

Chromatography is the most common method used to separate proteins and other substances according to different characteristics. During separation, the components to be separated are distributed between a stationary phase bed and a mobile phase that flows through the stationary bed (*Figure 4*). Separation is dependent on the difference in affinity of the substances for the two phases. The higher the affinity for the stationary phase, the slower the substance can move through the separation unit.

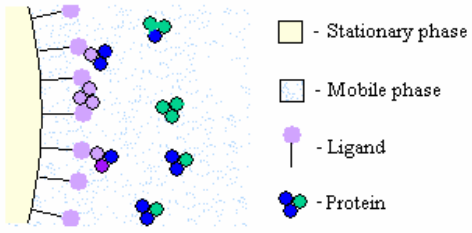

Figure 4. Chromatographic separation is based on the principle that components to be separated are distributed between a stationary phase and a mobile phase.

In general proteins can vary in size, isoelectric point (pI), hydrophobicity and specific binding. Ion exchange chromatography separates molecules based on differences in surface charge, hydrophobic interaction chromatography separates molecules according to differences in hydrophobicity, and size exclusion chromatography separates molecules according to size difference. Other media separate molecules according to other principles [22, 24, 25].

Little is known about how glycosylation affects the chromatographic behaviour of proteins. A number of different media will be tested for their ability to separate glycosylated and non-glycosylated proteins under different chromatographic conditions.

1.5.1 Ion Exchange Chromatography

In ion exchange chromatography (IEC), charged molecules bind reversible to a chromatographic media. The chromatographic media is either positively or negatively charged and the interaction between the molecule and the media increases with the size of the protein charge.

The isoelectric point, pI, of a protein is the pH where the net charge of the protein equals zero, i.e. the number of positive and negative charges are equal. Different proteins have different pI, mainly due to different amino acid compositions. Generally, if proteins are separated in a buffer system at a pH below their pI, the proteins will be positively charged and thus bind to a negatively charged media or cation exchanger. If the proteins instead are separated at a pH above their pI, the proteins will be negatively charged and thus bind to an anion exchanger [22]. Some proteins have shown to bind to ion exchangers at pH values where they theoretically not are supposed to. This is probably due to that proteins are not uniformly charged, i.e. some areas of the protein might interact with the ion exchanger even though the net charge of the protein indicates that no interaction should appear.

Bound proteins are usually released by increasing the ionic strength. The added ions will compete with the adsorbed proteins for the charged particles on the matrix causing the proteins to elute. By gradually increasing the ionic strength proteins will elute in a specific order starting with the protein with the lowest surface net charge. Proteins can also be eluted by changing the pH [22].

1.5.1.1 Mono S™

Mono S is a strong cation exchange media. The media is composed of rigid, monodisperse, polystyrene/divinylbenzene particles with an average particle size of 10 μm. In general resolution increases with decreasing particle size. Methyl sulfonate groups, $-CH_2SO_3^-$, are substituted on the base matrix [22]. Mono S is characterized by high resolution, dynamic capacity, reproducibility, and durability [23]. Mono S has an ionic capacity of 0.12–0.15 mmol H^+/ml and it is preferable used for capture or intermediate purification when milligram quantities are required [22].

1.5.1.2 Mono Q™

Mono Q is a strong anion exchange media very similar to Mono S. The base matrix is identical for both exchangers making them useful under the same conditions. Unlike Mono S, Mono Q is substituted with quaternary ammonium groups, $-CH_2N^+(CH_3)_3$ [22] and has a lower ionic capacity of 0.27–0.37 mmol Cl^-/ml.

1.5.2 Size Exclusion Chromatography

In size exclusion chromatography (SEC), or gel filtration, molecules are separated according to differences in size and to some extent shape. Unlike most other chromatographic systems, molecules separated by SEC do not bind to the medium. SEC medium is composed of porous gel beds. Depending mainly on molecular size, molecules passing through the media are allowed to penetrate the beads to different degrees. Small molecules will have access to a larger volume inside the beads causing them to elute late. Consequently, molecules are eluted in order of size starting with the largest molecule. SEC can only separate molecules within a certain size range. Molecules that are too large are excluded from the pores and molecules that are too small have full access to the pores [24].

1.5.2.1 Superdex™ 200

Superdex 200 consists of a matrix composed of cross-linked porous agarose particles and covalently bound dextran. It has a fractionation range of 10-600 kDa and a loading capacity of <50 µl [24]. Superdex is characterized by high resolution and stability [23] and is preferably used for the polishing step in a purification procedure [24].

1.5.3 Hydrophobic Interaction Chromatography

Hydrophobic interaction chromatography (HIC) separates molecules based on differences in their surface hydrophobicity. The principle is the same as for IEC, i.e. molecules are reversibly bound to a matrix and eluted under conditions that gradually lower the interaction between the molecules and the matrix, ideally, causing the most hydrophobic molecule to elute last. In the case of proteins, the relationship between protein hydrophobicity and HIC elution is not fully understood. Highly charged and soluble proteins, which possess hydrophobic surface regions, may elute late in HIC [25].

Most hydrophobic amino acids on a protein are hidden in the interior of the protein. Fortunately, the number of exposed amino acids required for interaction with the hydrophobic ligands on the matrix is low. The hydrophobic interactions are enhanced by high ionic strength buffers [23, 25].

1.5.3.1 SOURCE™ PHE

SOURCE PHE consists of a matrix composed of rigid, monodisperse 15 µm beads made of polystyrene/divinylbenzene substituted with phenyl groups. The phenyl groups have been shown to have a potential for π-π interactions. There have also been reports on stacking interactions between carbohydrates and phenyls making this particular chromatographic media interesting for separation of glycoproteins [23, 26].

SOURCE PHE is suitable for fast and high-resolution separations and has a binding capacity of >25 mg/ml. SOURCE PHE is preferably used for final polishing [23].

1.6 Quantitative Structure-Property Relationships

A Quantitative Structure-Property Relationships (QSPR) model is a multivariate mathematical relationship between a set of independent numerical structure-derived properties, called descriptors, and a dependent property of the system being studied (*Figure 5*). Examples of dependent properties are biological activity, solubility, or, as in this study, retention time. The technique is based on the hypothesis that the effect a molecule has is somehow linked to its molecular properties.

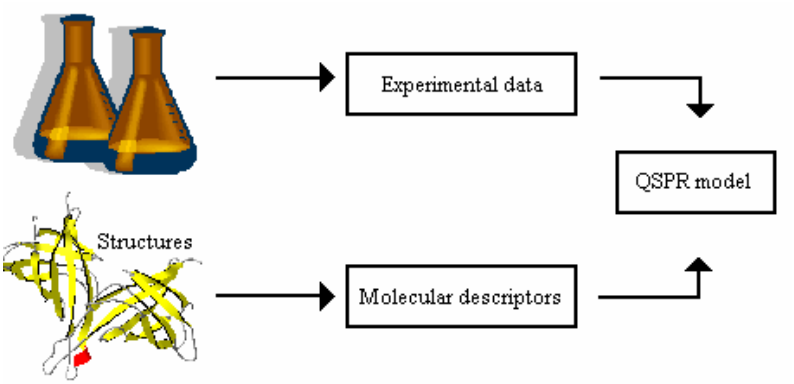

Figure 5. Simplified schematic view of Quantitative Structure-Property Relationships (QSPR).

Different regression methods can be used to correlate numerical molecular descriptors with the dependent property to be studied. Commonly used regression types are Multiple Regression, Principal Component Regression (PCR) and Partial Least Squares (PLS) [28]. The relationship between these numerical properties and the dependent property is generally described by an equation of the general form:

$$y = f(x) \qquad \text{(Equation 2)}$$

where y is the dependent property, x are the descriptors and f is some function. The derived equation can be used to predict properties of new molecules [27].

1.6.1 Molecular Descriptors

Molecular descriptors are mathematical values derived from the structure of a molecule. Commonly used descriptors are different constitutional descriptors such as molecular weight and number of hydrogen atoms, different electrostatic descriptors such as net charge, average surface potential and dipole moment, and different geometrical descriptors such as molecular volume and surface area.

1.6.2 Partial Least Squares

Partial Least Squares (PLS) is a multivariate projection method. Multivariate projection methods are characterized by their ability to project multivariate data into a space of lower dimensions, with a minimal loss of information, thereby making the large amount of data more manageable and comprehensive. PLS is particularly good when analysing strongly correlated, noisy, or numerous sets of data [28].

1.6.2 Cross Validation

Cross validation is a widely used technique for validation of QSPR models. The technique involves removing some of the structural data and deriving a new model using the reduced data set. The new model is then used to predict values for the excluded data. This procedure is repeated until all compounds have been excluded and predicted once. One commonly used cross validation method is leave-one-out, in which one element of the data is excluded each time. An average error is computed from the predicted values and used to evaluate the model [27, 28].

2 Materials and Methods

2.1 Materials Used for Deglycosylation and Chromatographic Separation

Four different chromatographic media were used in the chromatographic study (*Table 1*).

Table 1. Different chromatographic media used and their respective suppliers.

Chromatographic media	Supplier
Mono S HR 5/5	GE Healthcare, Amersham Biosciences, Uppsala, Sweden
Mono Q HR 5/5	GE Healthcare, Amersham Biosciences, Uppsala, Sweden
Superdex 200 HR 10/30	GE Healthcare, Amersham Biosciences, Uppsala, Sweden
RESOURCE PHE, 1 ml	GE Healthcare, Amersham Biosciences, Uppsala, Sweden

The chromatographic experiments were performed using ÄKTA™ Explorer 10 with Autosampler A-900 and Fraction Collector Frac-900 (all three from GE Healthcare), all three controlled by the software UNICORN 5.01.

Six different proteins were obtained from their respective suppliers (*Table 2*).

Table 2. Used proteins, their biological source, and their respective suppliers and product numbers.

Protein	Source	Supplier and Product Number
α-mannosidase	Jack Beans	SIGMA, M-7257
Albumin	Hen Egg White	SIGMA, A-5503
Avidin	Hen Egg White	SIGMA, A-9275
Endoglycosidase H	Streptomyces plicatus[a]	Kind gift of Genencor International, Inc. [b]
Ribonuclease A	Bovine Pancreas	SIGMA, R-5125
Ribonuclease B	Bovine Pancreas	SIGMA, R-7884

[a] Endoglycosidase H from Streptomyces plicatus—recombinant material expressed in Bacillus subtilis.
[b] Endoglycosidase H is not a commercial product of Genencor International, Inc.

2.2 Deglycosylation of Proteins

2.2.1 Enzymatic Treatment

Avidin and ovalbumin were treated with α-mannosidase and endoglycosidase H in order to remove glycan chains.

2.2.1.1 Removal of Mannose by α-Mannosidase

Three avidin samples of approximately 10 mg each were separately dissolved in 7.5 ml of 0.1 M sodium acetate, pH 5.5. 80 µl of 0.5 mg α-mannosidase in 3.0 M $(NH_4)_2SO_4$ and 0.1 mM zinc acetate, pH 7.5 were added to the samples. The zinc concentration of the samples was increased by adding zinc chloride to a final concentration of 20 mM. Zinc chloride has been shown to effectively increase the activity of α-mannosidase [29]. The samples were sterile filtered through a 0.22 µm pore size membrane and incubated overnight at 37˚C.

Four samples of ovalbumin were prepared for deglycosylation as described above for avidin with a few differences: no zinc chloride was added to sample number two, three and four; 0.05 M sodium citrate, pH 4.6 was used as sample buffer for sample three, and no α-mannosidase was added to sample number four.

2.2.1.2 Removal of N-linked Glycans of High-Mannose Type by Endoglycosidase H

The α-mannosidase treated samples, where no precipitation was visible, were treated with Endoglycosidase H (Endo H). 220 µl of 1.1 mg/ml Endo H in 0.1 M sodium acetate was added to the samples through a 0.22 µm pore size sterile filter. The sterile filter was washed with additional 600 µl 0.1 M sodium acetate. The samples were incubated at 37˚C for two days.

2.2.2 Concentration and Purification of Deglycosylated Protein

The enzyme treated proteins were concentrated using Amicon Ultra centrifugal filter devices (Millipore) in a Hettich ROTIXA/A with a speed of 2300 rpm to a final volume of 4 ml. This step is performed to reduce the volume before separation by gel filtration. A sample volume of more than 4% of the total size exclusion column volume is not recommended for high-resolution fractionation.

The concentrated samples were separated by gel filtration using a pre-packed Superdex 200 HR 10/30 column in order to remove α-mannosidase and Endo H. Samples of 1 ml each were injected manually on the column and run according to method 1 (*Appendix A*). Buffer containing 10 mM sodium phosphate, pH 8.0 was used in the separation. The method was repeated four times each in order to purify all deglycosylated avidin and deglycosylated ovalbumin. The separated proteins were collected using Fraction Collector Frac-900 (GE Healthcare).

2.2.3 Verification

2.2.3.1. Molecular Weight Determination of Proteins by SDS-PAGE using PhastSystem

To analyse if the glycan chains have been successfully removed, SDS-PAGE was performed. Protein samples to be analysed were taken before the enzymatic treatment and after each of the two deglycosylation steps. Samples were also taken after the gel filtration purification in order to determine in which fractions the purified proteins could be found. Samples containing 1 mg/ml untreated ribonuclease A and B in Milli-Q water were also analysed by SDS-PAGE. 15 µl of each sample was heated at 98°C for five minutes in the presence of 5 µl sample buffer consisting of 5 ml Tris buffer pH 6.8, 15 ml 10% (w/v) SDS solution, 5 ml glycerol (87%), 1.12 ml Milli-Q water and 0.5 ml 1-mercaptoethanol. The denatured samples were applied to the gel together with low molecular weight marker proteins (LMW; GE Healthcare). The electrophoresis was performed on an 8-25% gradient gel using PhastSystem™ Separation-Control Unit (GE Healthcare) according to *Appendix B*. Staining was performed with Coomassie Blue using PhastSystem™ Development unit.

2.2.3.2. Molecular Weight Determination of Proteins by MALDI-ToF MS

Ribonuclease A, ribonuclease B, avidin, deglycosylated avidin, ovalbumin and deglycosylated ovalbumin were analysed by MALDI-ToF MS. The proteins had a concentration of approximately 0.5 mg/ml. Matrix solution was prepared by making a saturated solution of sinapinic acid in 50% acetonitrile and 0.9% trifluoroacetic acid. Equal volumes of matrix solution and protein sample were mixed and 0.3 µl of the mixture was applied to Ettan™ MALDI-ToF (GE Healthcare) sample slide where they were allowed to air-dry. MALDI-ToF MS was performed using Ettan MALDI-ToF mass spectrometer in linear mode.

2.3 Separation on Chromatographic Media

Avidin, deglycosylated avidin, ribonuclease A and ribonuclease B were chromatographed on four different chromatographic media (see *Table 1*) at varying pH. The protein concentrations were 2 mg/ml for the native proteins and approximately 1 mg/ml for the deglycosylated avidin. Avidin was dissolved in 10 mM sodium phosphate, pH 8.0 and ribonuclease A and B were dissolved in 90% loading buffer (*Table 3, 4 and 6*). Ovalbumin, 2 mg/ml in 10 mM sodium phosphate, pH 8.0, and ovalbumin from the deglycosylation study were separated by Mono Q HR 5/5.

2.3.1 Mono S HR 5/5

Proteins were separately separated on a 1 ml Mono S HR 5/5 column at pH 4.0, 5.0, 6.0, 7.0, 8.0 and 8.7. A sample volume of 50 µl was injected and loaded on the equilibrated column by Autosampler A-900 (GE Healthcare). Flow rate was set to 1 ml/min and elution was achieved by increasing the conductivity. The linear salt gradient started five column volumes (CV) after the protein injection and proceeded for 40 CV reaching a final value of 1 M sodium chloride. Ordinary regeneration and re-equilibration steps followed the elution. Additional details of the method are given in *Appendix C*. The method was run as a scouting scheme. Different loading buffers used are listed in *Table 3*.

Table 3. Loading buffers used during separation of avidin, deglycosylated avidin, ovalbumin, deglycosylated ovalbumin, ribonuclease A and ribonuclease B, on a Mono S HR 5/5 column at pH 4.0, 5.0, 6.0, 7.0, 8.0 and 8.7.

Name	pH	Buffer	Concentration (mM)
A1	4.0	Formic acid	50
A2	5.0	Acetic acid	50
A3	6.0	Sodium phosphate	25
A4	7.0	Sodium phosphate	15
A5	8.0	Sodium phosphate	12
A6	8.7	BICINE	50

To prepare the elution buffer (Buffer Bx), 1000 mM sodium chloride was added to the loading buffers in *Table 3*, i.e.:

$$Buffer\ Bx = Buffer\ Ax + 1.0\ M\ NaCl \qquad \text{(Equation 3)}$$

where x is any integer from 1 to 6.

2.3.2 Mono Q HR 5/5

Proteins were separated on a 1 ml Mono Q HR 5/5 column at pH 5.0, 6.0, 7.0, 8.0 and 9.0 using the same method as for Mono S HR 5/5 (see *Appendix C*) with the following modification: an additional equilibration step was added to the method before the loading buffer equilibration step. In the new equilibration step, elution buffer is pumped at a constant flow of 1 ml/min for 7 minutes to prevent unspecific binding. The loading buffers used at different pH are listed in *Table 4*.

Table 4. Loading buffers used during separation of avidin, deglycosylated avidin, ribonuclease A and ribonuclease B on a Mono Q HR 5/5 column at pH 5.0, 6.0, 7.0, 8.0 and 9.0.

Name	pH	Buffer	Concentration (mM)
A1	5.0	Piperazine	20
A2	6.0	Bis-Tris	20
A3	7.0	Bis-Tris Propane	20
A4	8.0	Tris	20
A5	9.0	Ethanolamine	20

The corresponding elution buffers (Bx buffers) were prepared according to *Equation 3* using the loading buffers in *Table 4*.

2.3.3 Superdex 200 HR 10/30

Proteins were separated on a Superdex 200 HR 10/30 column at pH 7.4 using Phosphate-Buffered Saline (PBS) as buffer. A sample volume of 120 µl was injected using Autosampler A-900 (GE Healthcare). The autosampler was washed with three ml of PBS buffer after the injection. The sample loading flow and the elution flow were set to 0.4 ml/min. Additional details of the method are given in *Appendix D*.

2.3.3.1 Molecular Weight Determination by Size Exclusion Chromatography

The apparent molecular weight difference between proteins with and without sugar was determined by constructing a calibration curve. The calibration curve was prepared by separating cytochrome c, chymotrypsinogen, ovalbumin and albumin (*Table 5*) by Superdex 200 HR 10/30 as described above, measuring their elution volumes, calculating their corresponding distribution coefficients (K_{AV} values), and, finally, plotting these values versus the logarithm of their molecular weights seen in *Table 5*. K_{AV} values are calculated according to *Equation 4:*

$$K_{AV} = \frac{V_e - V_0}{V_t - V_0} \qquad \text{(Equation 4)}$$

where V_e is the elution volume, V_0 is the column void volume determined by measuring V_e for blue dextran, and V_t is the total bed volume of the column determined by measuring V_e for ethanol.

The selectivity curve can be used to determine the molecular weight of unknown proteins by measuring their elution volumes while separated by Superdex 200 HR 10/30 as described above and calculating their corresponding K_{AV} values.

The apparent molecular weight difference between avidin and deglycosylated avidin proteins was calculated and compared to the theoretical molecular weight difference. The same calculations were performed for ribonuclease A and ribonuclease B.

Table 5. Standard molecules used to construct calibration curve for molecular weight determination by size exclusion chromatography (SEC).

Molecule	Source	Molecular Weight	Supplier
EtOH	—	46	Kemetyl
Cytochrome c	Bovine Heart	12000	SIGMA
Chymotrypsinogen	Bovine Pancreas	25000	GE Healthcare
Ovalbumin	Hen Egg White	43000	GE Healthcare
Albumin	Bovine Serum	67000	GE Healthcare
Blue Dextran 2000	—	~2,000,000	GE Healthcare

2.3.4 RESOURCE PHE

Proteins were separated on a 1 ml RESOURCE PHE column at pH 6.5, 7.5, 8.5 and 9.5 using the same method as for Mono S HR 5/5 (see *Appendix C*) with the following modifications: the high pressure alarm was set to 1.5 MPa and no cleaning of the column was performed between the separations. The elution buffers used at different pH are listed in *Table 6*.

Table 6. Elution buffers used during separation of avidin, deglycosylated avidin, ribonuclease A and ribonuclease B on a 1 ml RESOURCE PHE column at pH 6.5, 7.5, 8.5 and 9.5.

Name	pH	Buffer	Concentration (mM)
B1	6.5	Sodium phosphate	50
B2	7.5	Sodium phosphate	50
B3	8.5	Boric acid	50
B4	9.5	Boric acid	50

The corresponding loading buffers (A buffers) were prepared by adding ammonium sulphate to the elution buffers. Separate loading buffers were prepared for avidin and ribonuclease since different amounts of ammonium sulphate are required for avidin and ribonuclease to bind to RESOURCE PHE. Loading buffers (A buffers) used for ribonuclease at pH 6.5, 7.5 and 8.5 were prepared according to *Equation 5* and the loading buffer used at pH 9.5 was prepared according to *Equation 6*. Corresponding loading buffers (A buffers) used for avidin at pH 6.5, 7.5 and 8.5 were prepared according to *Equation 6* and the loading buffer used at pH 9.5 was prepared according to *Equation 5*.

$$Buffer\ Ax = Buffer\ Bx + 2\ M\ (NH_4)_2SO_4 \qquad \text{(Equation 5)}$$

$$Buffer\ Ax = Buffer\ Bx + 2{,}5\ M\ (NH_4)_2SO_4 \qquad \text{(Equation 6)}$$

where x is any integer from 1 to 6 and Bx are the elution buffers in *Table 6*.

2.4 Calculation of Elution Ionic Strength and Elution Conductivity

Elution ionic strengths (I_M) were calculated according to *Equation 7*.

$$I_M = \frac{(t_R - t_0 - t_s)}{t_G}(I_1 - I_0) + I_0 \qquad \text{(Equation 7)}$$

where t_R is the retention volume calculated by Unicorn, t_0 is the void volume calculated according to *Equation 8*, t_S is the dwell volume of mixer and tubing, t_G is the gradient volume, I_1 is the ionic strength or conductivity at the end of the gradient, and I_o is the ionic strength or conductivity at the beginning of the gradient.

$$t_0 = \frac{m_{H_2O} - m_{EtOH}}{\delta_{H_2O} - \delta_{EtOH}} \qquad \text{(Equation 8)}$$

where m_{H2O} is the weight of the chromatographic column equilibrated in water, m_{EtOH} is the weight of the chromatographic column equilibrated in ethanol, δ_{H2O} is the density of water, and δ_{EtOH} is the density of ethanol.

Ionic strength values at the beginning and the end of the gradient during separation on Mono S HR 5/5 were determined using the internet website buffer design [30].

Conductivity values during the separations on other chromatographic media were determined by measuring the conductivity in the different chromatograms.

2.5 Quantitative Structure-Property Relationships

2.5.1 Protein Model Preparation

All protein-model preparations listed below were performed by Enrique Carredano and Jinyu Zou, GE Healthcare, Uppsala, Sweden.

2.5.1.1 Ribonuclease A from Bovine Pancreas

The protein model of ribonuclease A (ID XAFK) from bovine pancrease was prepared using the crystal structure produced by Leonidas *et al.* (PDB code: 1AFK) [31] as an initial model. The original model was modified by removing part B, water and a cofactor.

2.5.1.2 Ribonuclease B from Bovine Pancreas

Four protein models were prepared for ribonuclease B. GBKA01 was prepared by linking $(GlcNAc)_2Man_5$ to Asn^{34} in the model prepared for ribonuclease A described above. GBKA02 was prepared as GBKA01 by adding one additional mannose residue to the glycan chain. GBKA03 and GBKA04 were prepared as GBKA01 and GBKA02, respectively, changing Asn^{67} to Asp^{67}. The glycan chains were modelled using the program BIOPOLYMER.

2.5.1.3 Deglycosylated Avidin from Hen Egg White

The crystal structure of avidin from egg white produced by Nardone *et al.* 1998 (PDB code: 1RAV) [32] was used as an initial model. The model has been modified by modelling the missing C-terminal segment using BIOPOLYMER and generating the two missing subunits using the program O. The model has also been modified by removing water. The new model was assigned the ID XRAV.

2.5.1.4 Avidin from Hen Egg White

The protein model of avidin from egg white, ID GRAV, was prepared by linking $(GlcNAc)_3Man_5$ to Asn^{17} in each subunit of the model prepared for deglycosylated avidin described above. The glycan chain was modelled using BIOPOLYMER.

2.5.2 Calculation of Descriptors

58 different descriptors were calculated using the proprietary program SCARP version 2.6 [33]. The descriptors describe surface charge distribution, hydrogen bonding features, surface electrostatic potential, surface hydrophobicity, size and shape. Most of the descriptors are pH dependent.

2.5.3 QSPR Modelling

The elution ionic strengths obtained using 13 different proteins separated by Mono S at pH 4.0, 5.0, 6.0, 7.0, 8.0 and 8.7 as described above [34] were used as response values for QSPR modelling. The QSPR was performed using the program 'The Unscrambler' with the partial least square regression PLS1 [35]. Martens' uncertainty test was performed in order to exclude
non-significant descriptors [36]. Non-significant descriptors were removed and the regression was repeated until all descriptors were significant. Leave-one-out cross validation was performed to check the quality of the model.

2.5.4 QSPR Prediction

The elution ionic strengths of ribonuclease A, ribonuclease B, avidin and deglycosylated avidin were predicted using the QSPR model and compared to experimental data.

3 Results

3.1 Deglycosylation of Proteins

3.1.1 Precipitation of Ovalbumin During Deglycosylation

Ovalbumin precipitated in precence of 0.35 mM zinc chloride. The glycoprotein also precipitated when 0.05 M sodium citrate was used as sample buffer. The two remaining samples did not precipitate. Sample number 4 dissolved in 0.1 M sodium acetate, pH 4.6 treated with α-mannosidase and Endo H were chosen for further studies.

3.1.2 Molecular Weight Determination of Proteins by SDS-PAGE Using PhastSystem

An SDS-PAGE analysis was made to investigate changes in molecular weight during deglycosylation of avidin (*Figure 6*) and ovalbumin (figure not shown). The size difference between the bands in lane 4 and 6, representing avidin before and after treatment with Endo H, is consistent with a decrease in molecular weight after Endo H treatment. No difference in molecular weight could be discerned between the bands in lane 2 and 4, representing avidin before and after treatment with α-mannosidase.

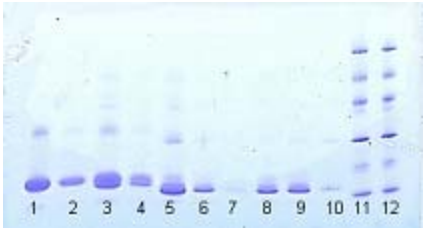

Figure 6. Deglycosylation analysis of avidin. Native avidin was treated with α-mannosidase overnight followed by endoglycosidase H (Endo H) treatment for two days. The deglycosylated avidin was purified by size exclusion chromatography (SEC). Change in molecular weight was analysed with an 8-25% gradient gel using PhastSystem. Staining was performed with Coomassie Blue. *Lane 1*: Native avidin, ~1.3 mg/ml. *Lane 2*: Native avidin, ~0.3 mg/ml. *Lane 3*: α-mannosidase treated avidin, ~1.3 mg/ml. *Lane 4*: α-mannosidase treated avidin, ~0.3 mg/ml. *Lane 5*: Endo H treated avidin, ~1,3mg/ml. *Lane 6*: Endo H treated avidin, ~0.3 mg/ml. *Lane 7-10*: Fractions collected during SEC purification of avidin after α-mannosidase and Endo H treatment. *Lane 11-12*: Low molecular weight marker (GE Healthcare, Uppsala, Sweden).

The purity and amount of avidin in fractions collected after two size exclusion chromatography purifications performed after α-mannosidase and Endo H treatment were analysed on SDS-PAGE (*Figure 7*, lane 1-8). Lane 1, 2, 5 and 6 contained the highest amounts of deglycosylated avidin with a satisfactory purity. Corresponding fractions were pooled and kept for chromatographic studies. The purity and amount of avidin after two additional SEC purifications were also analysed on SDS-PAGE (figure not shown) and the fractions with highest amount of satisfactory pure avidin were pooled together with the fractions described above.

Molecular weights of ribonuclease (RNase) A and B were studied by SDS-PAGE (*Figure 7*, lane 9-10). The gel showed that RNase B in lane 10 has a higher molecular weight than RNase A in lane 9. The broad band seen in lane 10 indicated that RNase B exists as several variants.

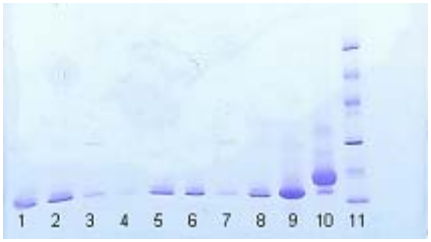

Figure 7. Analysis of purity and amount of deglycosylated avidin after two SEC purifications and molecular weight determination of ribonuclease A and B. Deglycosylated avidin was purified by four separate SEC runs. Fractions were collected during purification and analysed with an 8-25% gradient gel using PhastSystem. Molecular weight determination of RNase A and B by SDS-PAGE was performed on the same gradient gel. *Lane 1-8*: Fractions collected during two different SEC purifications of avidin after α-mannosidase and Endo H treatment. *Lane 9*: RNase A, ~2 mg/ml. *Lane 10*: RNase B, ~2 mg/ml. *Lane 11*: Low molecular weight marker (GE Healthcare, Uppsala, Sweden).

3.1.3 Molecular Weight Determination of Proteins by MALDI-ToF MS

MALDI-ToF MS was performed to determine molecular weight and heterogeneity of RNase A, RNase B, avidin, deglycosylated avidin, ovalbumin and deglycosylated ovalbumin. The spectrum of ribonuclease A (*Figure 8*, left side) shows that the protein has a molecular weight of 13700 Da and that almost no heterogeneity exists. The presence of several different peaks in the spectrum of RNase B (*Figure 8*, right side) indicates that RNase B exists in at least five different glycoforms with molecular weights between 14915 Da to 15577 Da.

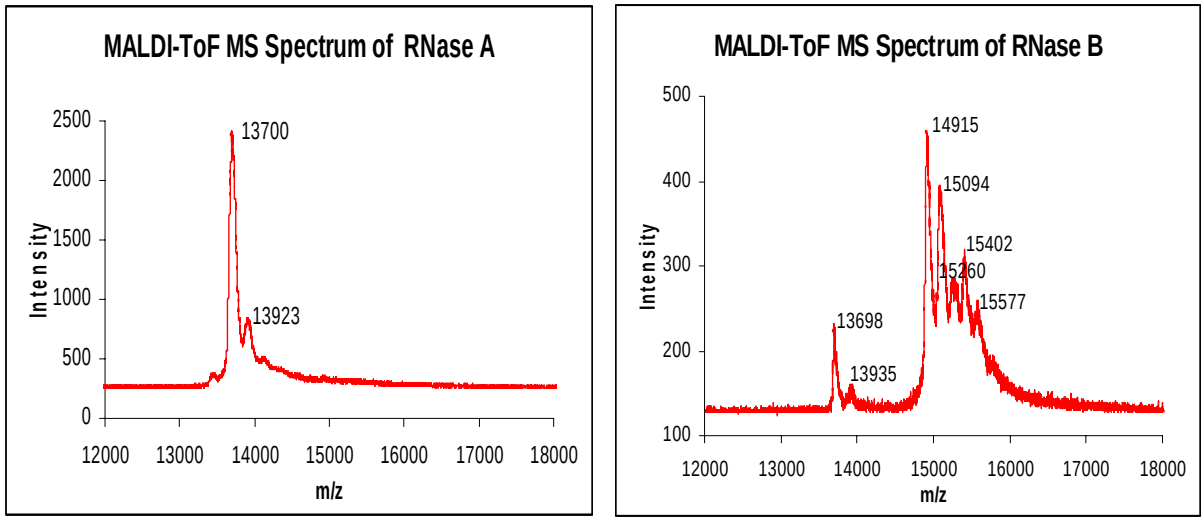

Figure 8. MALDI-ToF MS spectra of ribonuclease A (left) and ribonuclease B (right). MALDI-ToF MS of ribonuclease A and ribonuclease B was performed using Ettan MALDI-ToF mass spectrometer in linear mode.

The spectra of deglycosylated avidin and native avidin are shown in *Figure 9*. The molecular weight of one subunit of deglycosylated avidin was determined to be 14603 Da by studying the main peak in the spectrum (*Figure 9*, left side) and the molecular weight of one subunit of native avidin was determined to be approximately 15940 Da (*Figure 9*, right side). Presence of additional peaks in the spectrum of deglycosylated avidin indicates that not all of the proteins has been completely deglycosylated. No MALDI-ToF MS spectra could be obtained for ovalbumin and deglycosylated ovalbumin.

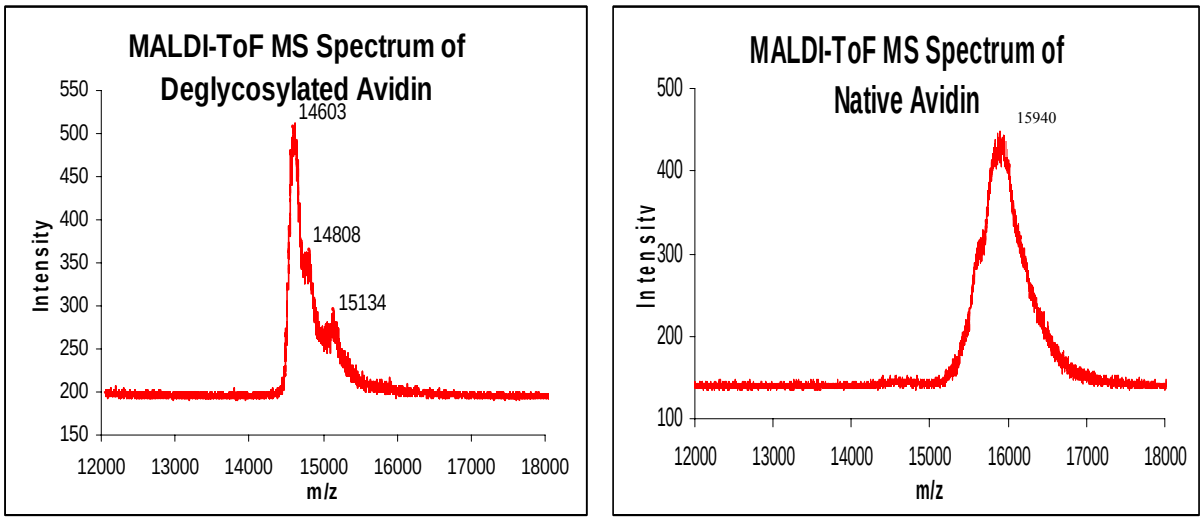

Figure 9. MALDI-ToF MS spectra of deglycosylated avidin (left) and native avidin (right). MALDI-ToF MS of native avidin and deglycosylated avidin was performed using Ettan MALDI-ToF mass spectrometer in linear mode.

3.2 Separation on Chromatographic Media

3.2.1 Mono S

RNase A, RNase B, avidin and deglycosylated avidin were separated at six different pH values on a Mono S HR 5/5 column (see *Table 3*). The proteins were eluted with a 40 column volumes linear gradient of 0 to 1 M NaCl. Chromatograms showing the best and worst separation of native avidin and deglycosylated avidin at pH 4.0 and pH 8.7, respectively, are shown in *Figure 10*, top. Corresponding chromatograms of RNase A and RNase B at pH 4.0 and pH 8.0, respectively, are shown in *Figure 10*, bottom. RNase B eluted as a double peak.

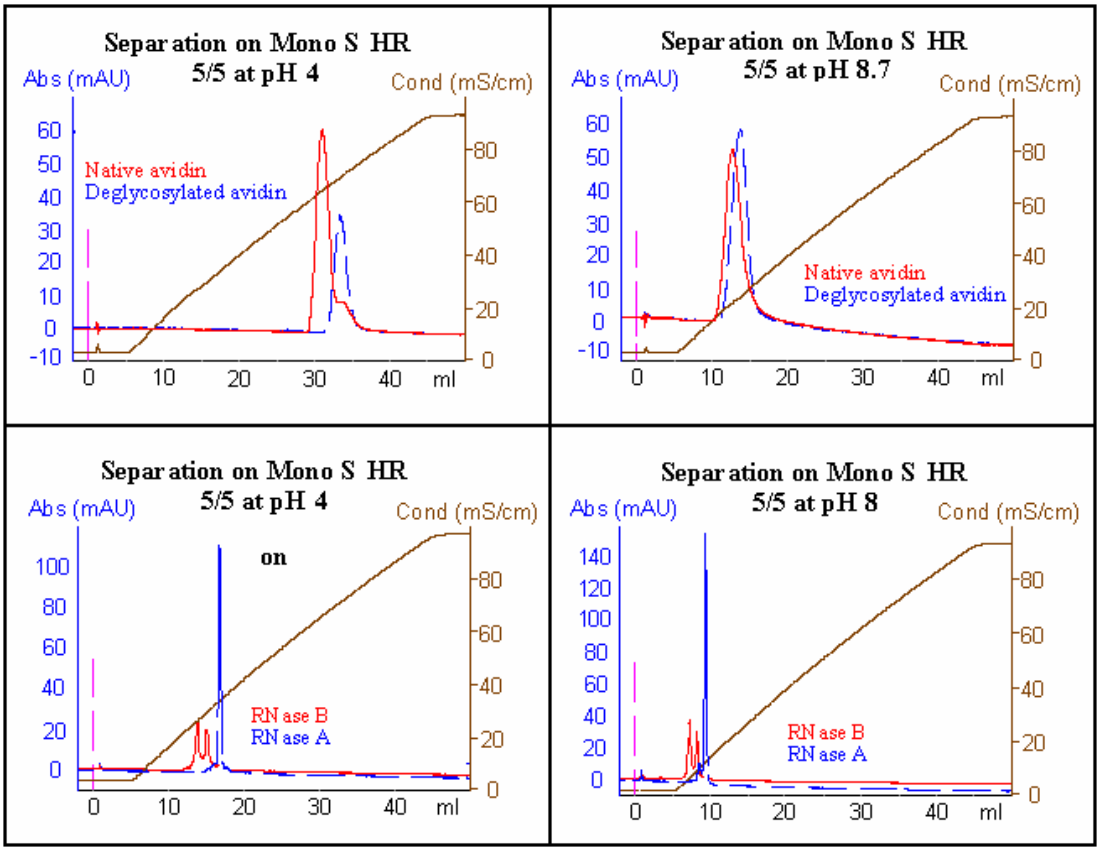

Figure 10. Separation on a Mono S HR 5/5 column using a linear elution gradient of 0 to 1 M NaCl. Top: Chromatograms of native avidin and deglycosylated avidin separated at pH 4.0 and 8.7, respectively. **Bottom**: Chromatograms of RNase A and RNase B separated at pH 4.0 and 8.0, respectively.

Elution ionic strengths, I_M, for the separations by Mono S were calculated according to *Equation 7*. If a protein eluted as a double peak, I_M were calculated and tabulated for both peaks. The calculated I_M values can be seen in *Table 7*. All separations of native avidin and deglycosylated avidin on the Mono S column implied that separation of glycosylated protein and protein without sugar is improved with decreasing pH. The separations of RNase A and RNase B also support this theory but a more extensive analysis of the chromatograms is required.

Table 7. Experimental elution ionic strength of RNase A, RNase B, native avidin and deglycosylated avidin on a Mono S column run at pH 4.0, 5.0, 6.0, 7.0, 8.0 and 8.7. Elution ionic strengths are calculated according to *Equation 7*.

Protein	I_M, pH 4.0 (mM)	I_M, pH 5.0 (mM)	I_M, pH 6.0 (mM)	I_M, pH 7.0 (mM)	I_M, pH 8.0 (mM)	I_M, pH 8.7 (mM)
RNase A	442	317	194	164	132	102
RNase B, peak 1	391	276	163	134	105	85
RNase B, peak 2	378	246	139	109	79	—
Avidin, native	675	461	389	331	285	218
Avidin, deglycosylated	736	498	420	365	312	244

3.2.2 Mono Q

RNase A, RNase B, avidin and deglycosylated avidin were separated at five different pH values on a Mono Q HR 5/5 column (see *Table 4*). The proteins were eluted with a linear gradient of 0 to 1 M NaCl. Chromatograms showing the chromatographic runs of the four proteins at pH 7.0 are shown in *Figure 11*. The proteins did not bind to Mono Q at any pH.

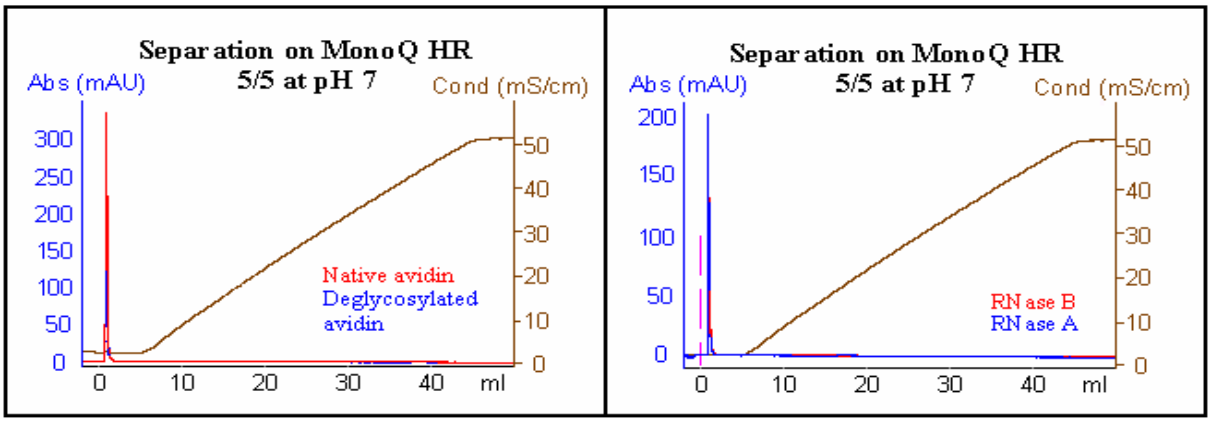

Figure 11. Separation on a Mono Q HR 5/5 column using a linear elution gradient of 0 to 1 M NaCl. Left: Chromatogram of native avidin and deglycosylated avidin separated at pH 7.0. **Bottom**: Chromatogram of RNase A and RNase B separated at pH 7.0.

Ovalbumin and deglycosylated ovalbumin were separated on five different pH values by Mono Q HR 5/5. The proteins were eluted with a linear gradient of 0 to 1 M NaCl. Chromatogram showing the separation at pH 7.0 is shown in *Figure 12*. Both ovalbumin and deglycosylated ovalbumin eluted as multiple peaks.

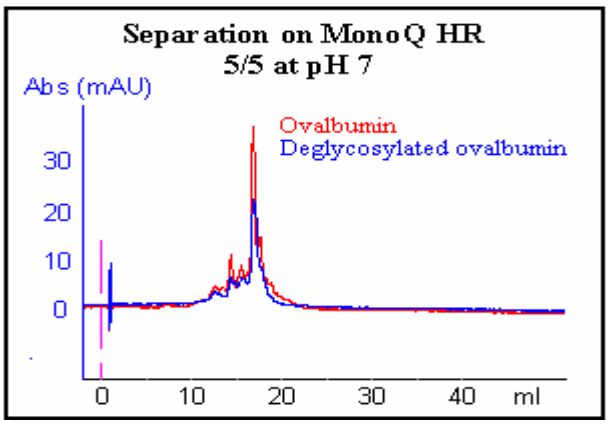

Figure 12. Chromatogram of native ovalbumin and deglycosylated ovalbumin separated on a Mono Q HR 5/5 column at pH 7.0 using a linear elution gradient of 0 to 1 M NaCl.

3.2.3 Superdex 200 HR 10/30

RNase A, RNase B, avidin and deglycosylated avidin were separated at pH 7.4 on a Superdex 200 column. Chromatogram showing the separation of native avidin and deglycosylated avidin is shown in *Figure 13*, left. Corresponding chromatogram of ribonuclease A and ribonuclease B is shown in *Figure 13*, right. The glycoproteins had a shorter retenrion than the proteins without sugar as can be expected from their larger size.

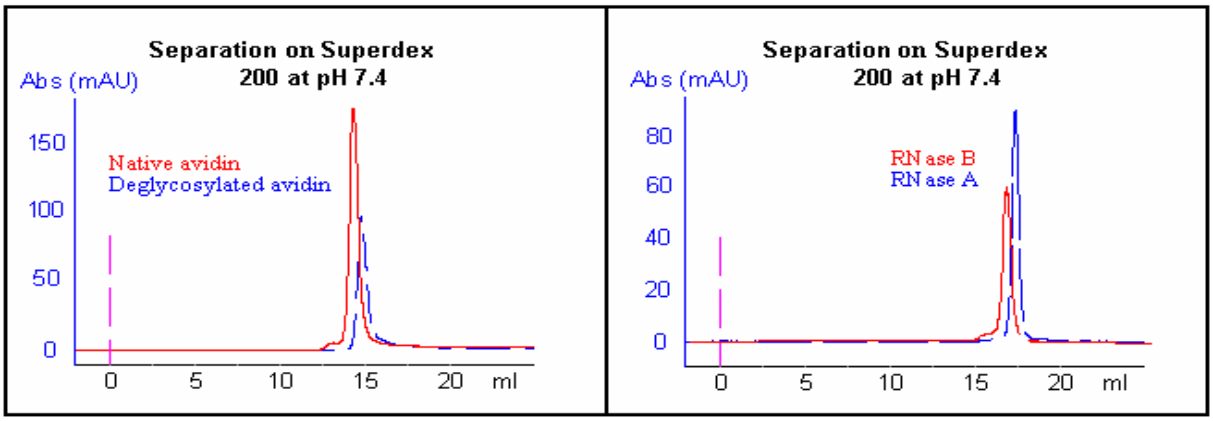

Figure 13. Separation on a Superdex 200 column at pH 7.4. Left: Chromatogram of native avidin and deglycosylated avidin. **Bottom**: Chromatogram of RNase A and RNase B.

3.2.3.1 Molecular Weight Determination by Size Exclusion Chromatography

The elution volumes (V_e) after separation on a Superdex 200 column of cytochrome c, chymotrypsinogen, ovalbumin and albumin were determined and corresponding K_{AV} values were calculated according to *Equation 4* (*Table 8*).

Table 8. Experimental elution volumes (V_e) and corresponding K_{AV} values of cytochrome c, chymotrypsinogen, ovalbumin and albumin on a Superdex 200 column run at pH 7.4.

Protein	V_e (ml)	K_{AV}
Cytochrome c	18.95	0.86
Chymotrypsinogen	16.68	0.68
Ovalbumin	14.66	0.52
Albumin	13.62	0.44

A selectivity curve was prepared by plotting the K_{AV} values for each protein against the logarithm of their theoretical molecular weight in *Table 5* (*Figure 14*).

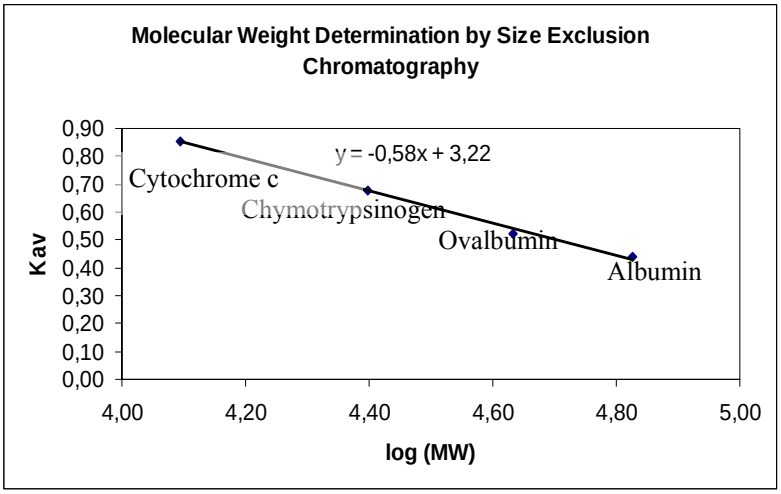

Figure 14. A selectivity curve was prepared by measuring the elution volumes of cytochrome c, chymotrypsinogen, ovalbumin and albumin separated on a Superdex 200 column, calculating the corresponding K_{AV} values and plotting the K_{AV} values against the logarithm of the corresponding molecular weight. Finally, a straight line was fitted to the data.

The selectivity curve was used to calculate the apparent molecular weight of RNase A, RNase B, native avidin and deglycosylated avidin. Apparent molecular weights as well as theoretical molecular weights can be seen in *Table 9*.

Table 9. Theoretical and experimental apparent molecular weights of RNase A, RNase B, native avidin and deglycosylated avidin. The experimental apparent molecular weights have been determined using the selectivity curve in *Figure 14* and the experimental K_{AV} values have been calculated using *Equation 4* and the experimental elution times in *Table 8*. Theoretical molecular weights (MW) are determined by MALDI-ToF MS.

Protein	Ve (ml)	Theoretical MW	Apparent MW
RNase A	17.39	13700	17513
RNase B	16.29	14930	20793
Avidin, native	14.35	63760	50803
Avidin, deglycosylated	14.48	58410	42047

The apparent and theoretical molecular weight difference between glycoprotein and protein without sugar were calculated according to *Table 10*.

Table 10. Apparent and theoretical molecular weight differences between glycoprotein and protein without sugar. The molecular weights can be seen in *Table 9*.

Proteins	Theoretical MW Difference	Apparent MW Difference
RNase B – RNase A	1230	3280
Native avidin – deglycosylated avidin	5320	8756

3.2.4 RESOURCE PHE

RNase A, RNase B, avidin and deglycosylated avidin were separated at four different pH values on a RESOURCE PHE column(see *Table 6*) using a linear gradient of $(NH_4)_2SO_4$. The best separation of avidin and deglycosylated avidin was achieved at pH 6.5 (*Figure 15*, left) and the best separation of RNase A and B was achieved at pH 8.5 (*Figure 15*, right).

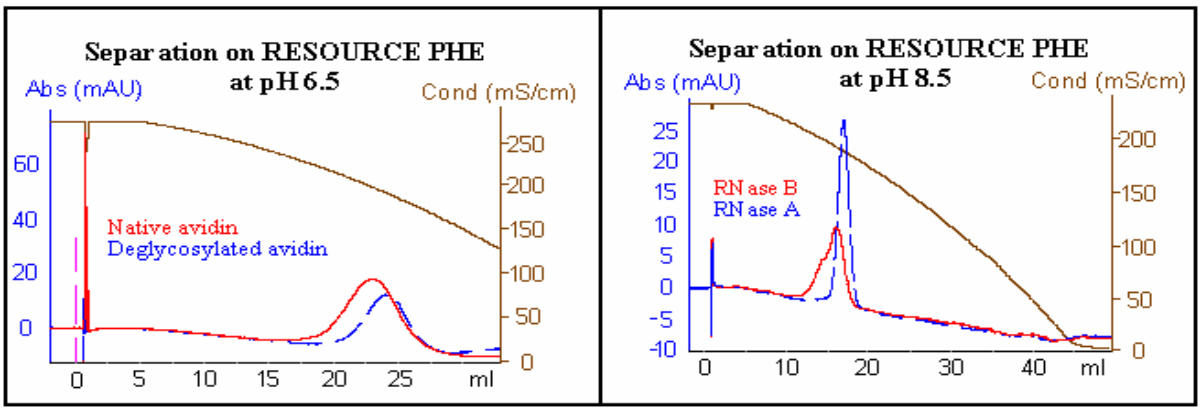

Figure 15. Separation on a RESOURCE PHE column using a linear elution gradient of $(NH_4)_2SO_4$. Left: Chromatogram of native avidin and deglycosylated avidin separated at pH 6.5. **Right**: Chromatogram of RNase A and RNase B separated at pH 8.5.

The glycosylated proteins eluted as broader peaks than the proteins without sugar during all separations on RESOURCE PHE. The non-symmetric peak present in the chromatogram of RNase B is probably due to heterogeneity. All separations performed on RESOURCE PHE indicate that proteins without sugar are retained more on the RESOURCE PHE column than glycoproteins but that the separation does not improve with decreasing or increasing pH.

Measured elution conductivity of RNase A, RNase B, native avidin and deglycosylated avidin on RESOURCE PHE can be seen in *Table 11*.

Table 11. Experimental elution conductivity (Cond) of RNase A, RNase B, native avidin and deglycosylated avidin during separation on RESOURCE PHE at pH 6.5, 7.5, 8.5 and 9.5.

Protein	Cond, pH 6.5 (mS/cm)	Cond, pH 7.5 (mS/cm)	Cond, pH 8.5 (mS/cm)	Cond, pH 9.5 (mS/cm)
RNase A	151.16	147.05	145.12	112.28
RNase B	154.92	149.84	149.80	114.49
Avidin, native	126.17	126.12	127.24	118.15
Avidin, deglycosylated	118.10	118.56	120.07	111.67

3.3 Quantitative Structure-Property Relationships

The previously obtained QSPR model [34] was used to predict the elution ionic strengths (I_M) of the protein/glycoprotein pairs during cation exchange chromatography. Experimental I_M values (see *Table 7*) plotted versus predicted I_M values can be seen in *Figure 16*. The cation exchange chromatography data used for the calibration of the model [34] was also included in the plot.

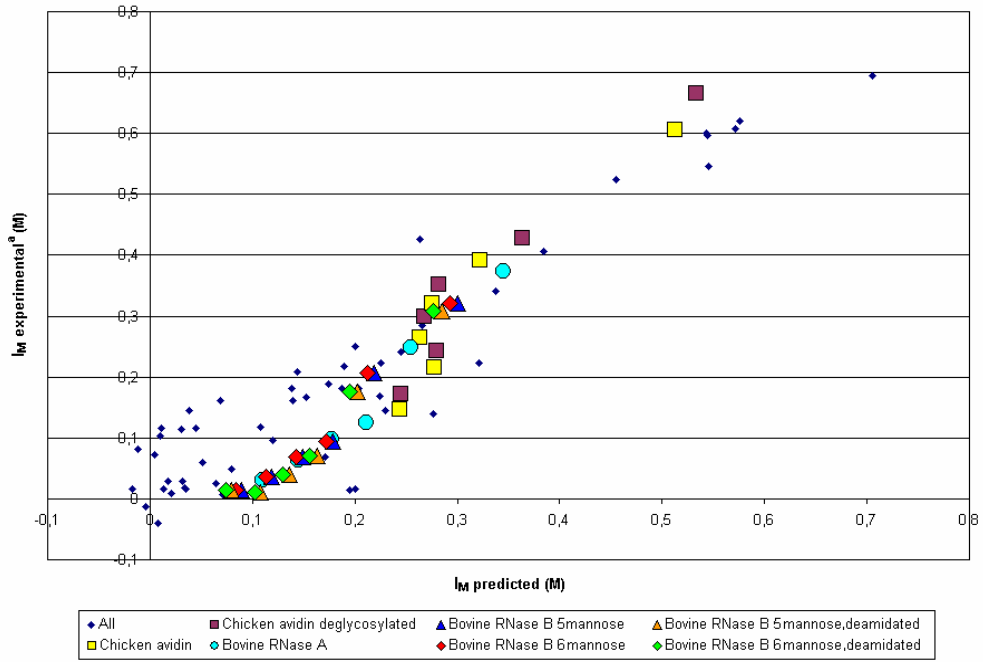

Figure 16. Experimental elution ionic strength (I_M) during purification on Mono S HR 5/5 plotted against predicted I_M. The I_M values have been predicted by a QSPR model, derived from the data produced by Malmquist *et al.* 2005 [34], using leave-one-out cross validation.

[a] The highlighted points represent ribonuclease A, ribonuclease B, avidin and deglcosylated avidin (see *Table 7*); the other data have been produced by Malmquist *et al.* 2005 [34].

The experimental I_M values (see *Table 7*) were also plotted against the calculated average surface potential for ribonuclease A, ribonuclease B, avidin and deglycosylated avidin in *Figure 17*. The cation exchange chromatography data set produced by Malmquist *et al.* 2005 [34] is also included in the plot.

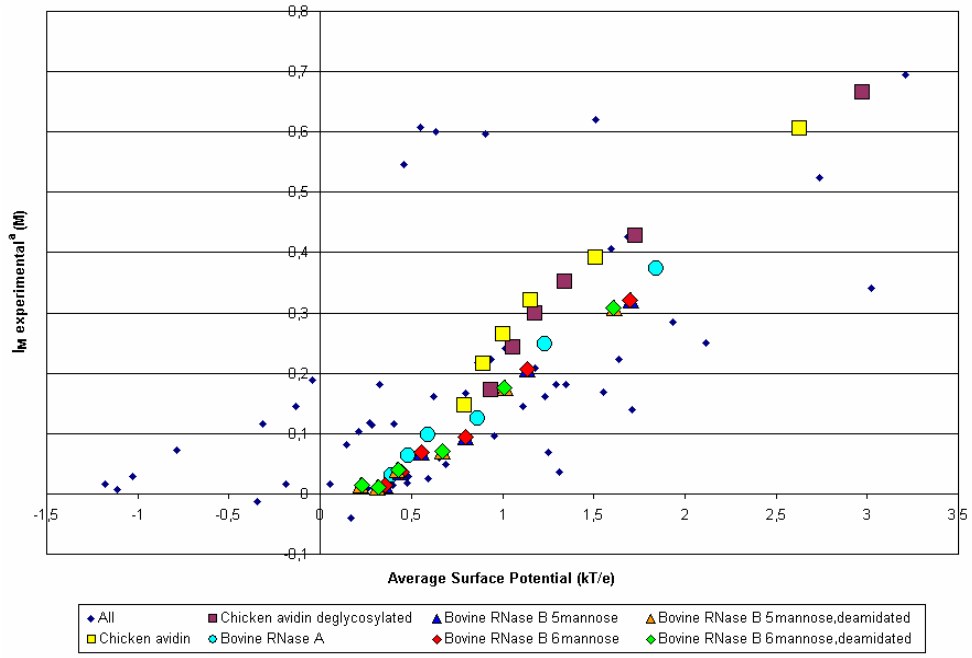

Figure 17. Experimental elution ionic strength (I_M) during purification on Mono S HR 5/5 plotted against average surface potential calculated by the proprietary program SCARP.

[a] The highlighted points represent ribonuclease A, ribonuclease B, avidin and deglcosylated avidin (see *Table 7*); the other data have been produced by Malmquist *et al.* 2005 [34].

4 Discussion

4.1 Molecular Weight Analysis of Protein/Glycoprotein Pairs

4.1.1 Avidin from Hen Egg White

The deglycosylation of avidin worked reasonably well. The SDS-PAGE analysis (see *Figure 6*) performed to investigate changes in molecular weight during deglycosylation suggests that the molecular weight decreases after treatment with endoglycosidase H (Endo H), whereas no decrease in molecular weight can be distinguished after treatment with α-mannosidase. According to Bruch *et al.* [16], the glycan chain of avidin consists of approximately three N-acetylglucosamine residues and five mannoses. The non-distinguishable size difference after α-mannosidase treatment suggests that most of the mannoses are still attached to the protein.

The MALDI-ToF MS spectra of avidin before and after Endo H treatment (see *Figure 9*) also support that deglycosylation has been satisfactory. The peak present in the spectrum of avidin correlates well with the theoretical molecular weight of avidin and the size difference between this peak and the major peak seen in the spectrum of deglycosylated avidin was determined to be 1337 Da. This is equivalent with the weight of two N-acetylglucosamine residues and five mannoses, i.e. the weight of the glycan chain to be removed. Two additional peaks of higher weight are present in the spectrum of deglycosylated avidin indicating that not all avidin has been completely deglycosylated. A probable explanation for incomplete deglycosylation is that the glycosylation site is inaccessible to the enzyme—denaturing the protein is often necessary for deglycosylation to be successful. This might be due to the presence of glycoforms of hybrid or complex types. Another explanation suggests that Endo H is specific

for N-linked oligosaccharides of high-mannose type possessing at least three mannose residues [37], i.e. it might not be possible to deglycosylate avidin molecules if N-linked glycan chains of hybrid or complex types exist or if more than two mannoses have been removed by α-mannosidase. Consequently, deglycosylation by Endo H alone might improve the outcome. The deglycosylation may also be improved by increasing the incubation time with Endo H.

4.1.2 Ovalbumin from Hen Egg White

No size difference could be distinguished in the SDS-PAGE analysis performed to investigate changes in molecular weight during deglycosylation of ovalbumin. Nor was it possible to obtain a MALDI-ToF MS spectrum of ovalbumin. Whether or not the degycosylation was successful had to be investigated in some other way (see below).

4.1.3 RNase from Bovine Pancreas

The SDS-PAGE performed to analyse the molecular weights of RNase A and RNase B (see *Figure 7*) showed that the apparent molecular weight difference between them is larger than the corresponding difference between avidin before and after deglycosylation. The theoretical molecular weight difference concludes that the size difference between the glycosylated and non-glycosylated protein should be about the same. This phenomenon could possibly be explained by differences in the flexibility and the conformation of the glycan chains.

The more reliable and exact MALDI-ToF MS spectrum of RNase A (see *Figure 8*) confirm that RNase A has a molecular weight of 13700 Da. The spectrum of RNase B is consistent with the published result that RNase B exist as five different glycoforms, $(GlcNAc)_2Man_{5-9}$ [6]. The small peak determined to be 13698 Da in the spectum of RNase B shows that the sample of RNase B purchased from Sigma-Aldrich Corporation actually contains a mixture of RNase A and RNase B.

4.2 Separation on Chromatographic Media

4.2.1 Ion Exchange Chromatography

The best chromatographic separation with a satisfactory high resolution was achieved by Mono S at pH 4.0 (see *Figure 10*). This is not a surprising result considering that Mono S is a high-resolution column and both avidin and RNase have high pI values. A high pI implies that a protein is positively charged at neutral pH and that the positive charge increases with decreasing pH. This causes the interaction between the protein and a cation exchanger to increase as pH decreases.

There are at least two possible explanations to why proteins without carbohydrate moiety are retained longer on a cation exchanger. Plummer *et al.* [13] have reported that glycan chains can shield part of the charged surface causing the protein to appear less charged. Another contributing factor could be the larger size of the glycoproteins. A larger size makes the average surface potential to decrease. Earlier studies have shown that average surface potential is closely related to retention on cation exchangers [34, 38].

RNase B elutes as a double peak during separation on Mono S. The two peaks do not represent different glycoforms of RNase B but are explained by another type of post-translational modification. Gotte *et al.* [12] have shown that the double peak is due to deamidation of Asn^{67} to Asp^{67}. The resolution on Mono S is not sufficient to separate the various glycoforms of avidin and RNase B.

To increase the cationic separation several approaches can be used. If the proteins still are stable, the pH during separation can be further lowered resulting in a larger charge and probably a better separation. Another approach is to optimise the elution gradient that can be modified in a number of ways. Knowing the elution ionic strengths of the proteins makes it possible to design a much shallower gradient resulting in a better separation. It might also be possible to improve the separation of the protein/glycoprotein pair by use of different type of salt ions, a step-wise gradient, or by elution with an increasing pH instead of increasing salt concentration. Changing the type of ligand is another way of optimizing the separation conditions in IEC. Mini S (GE Healthcare) is another available cation exchanger that yields exceptional resolution thanks to 3 μm nonporous beads [21]. The drawback of Mini S is that the binding capacity is somewhat lower than the capacity of Mono S. During analytical separation a low binding capacity is of less importance.

As expected, avidin and RNase did not bind to the anion exchanger column Mono Q (see *Figure 11*) due to their high pI. The net charges of the proteins were positive at all chromatographic runs, i.e. the only possibility for the proteins to bind to the column would be if their charge asymmetries were large enough to allow a sufficient part of the protein surfaces to be negatively charged and thus enable interaction with the anionic column. Obviously, this was not the case.

Since ovalbumin has a pI of approximately 4.5 it interacted with the anion exchanger column at neutral pH. Both native and deglycosylated ovalbumin eluted as similar multiple peaks (see *Figure 12*). It is clear that the deglycosylation has not been successful. The unsuccessful deglycosylation procedure can be explained by the fact that the glycosylation site is inaccessible to the enzyme, or by existence of glycans of hybrid and complex type [39], which are not possible to remove with Endo H. An alternative deglycosylation procedure has to be applied in order to deglycosylate ovalbumin. Treatment with Endoglycosidase F1 might be a more effective way to deglycosylate ovalbumin. Chemical deglycosylation using trifluoromethanesulfonic acid (TFMS) hydrolysis might also be effective [2]. TMFS removes all O-linked and N-linked glycans but might also cause cleavage of the polypeptide chain and denaturation of the protein [5]. Moreover, TMFS is harmful to human and must be handled with great care.

4.2.2 Size Exclusion Chromatography

The resolution of the separation on Superdex 200 (see *Figure 13*) has to be improved in order for gel filtration to be useful. The easiest way to improve separation is to lower the separation flow rate. A flow rate of 0.05 ml/min is not unusual during size exclusion chromatography. Increasing the length of the gel filtration column increases the resolution by the square root of the length. The disadvantage of decreasing the flow rate or increasing the column length is that the time required for a separation increases. Changing the size exclusion chromatographic media is an alternative approach for increasing separation. The only available column from GE Healthcare with higher resolution is Superdex™ 75. Superdex 75 has a separation range

of 3-70 kDa, i.e. separation of avidin and RNase should be possible. The sample volume is also of great importance, but since a relative sample volume of less than 0.5% of the column volume not normally increase resolution, decreasing the sample volume used in this study would not increase separation further.

4.2.2.1 Molecular Weight Determination by Size Exclusion Chromatography

It is difficult to make any direct conclusions regarding the apparent molecular weights determined by size exclusion chromatography (see *Table 9*). None of the proteins, neither with sugar nor without, fit the selectivity curve (see *Figure 14*). The apparent molecular weight of avidin is much lower than the theoretical molecular weight and the apparent molecular weight of RNase is much higher than the theoretical molecular weight. This difference in theoretical and apparent molecular weight might be due to that the proteins are not ideal globular proteins. It should be mentioned that apparent molecular weights determined by size exclusion chromatography only are estimations of molecular weights. The retention times are dependent not only on molecular weight but also, to a certain extent, on the shape of the protein. Relating protein retention times to Stokes radii might result in a better correlation than relating protein retention times to molecular weights [24].

Comparing the apparent molecular weight difference and the theoretical molecular weight difference of the protein/glycoprotein pairs (see *Table 10*) indicates that the apparent molecular weight difference is significantly greater than the theoretical molecular weight, i.e. the carbohydrate moiety makes the glycoproteins appear larger than they actually are, resulting in a better separation of protein/glycoprotein pairs than expected.

4.2.3 Hydrophobic Interaction Chromatography

Both selectivity and resolution is unsatisfactory in the separations by hydrophobic interaction chromatography (HIC) (see *Figure 15*). Phenyl groups have been shown to interact with carbohydrates through stacking interactions [40]. The expectation was therefore that separation on the RESOURCE PHE column would result in a separation of protein/glycoprotein pairs with the glycoprotein interacting strongest with the column. Instead, separation on RESOURCE PHE showed that the interaction between hydrophobic ligands and proteins decreases when glycan chains are attached to the proteins. The decrease in interaction is most likely due to that carbohydrates are hydrophilic.

It is complicated to predict how proteins will behave during HIC. Finding the optimal separation parameters is generally an experimental procedure. Changing the type of ligand, type of base matrix, ligand density, type and concentration of salt, temperature, pH, or adding additives can all affect separation.

4.3 Quantitative Structure-Property Relationships

The previously obtained QSPR model [34] used to predict the elution ionic strengths for RNase A, RNase B, avidin and deglycosylated avidin seems to be working well (see *Figure 16*). A prediction close to the experimental value was expected since both avidin and RNase A are included in the training set used to build the QSPR model. However, the ability of the model to predict in which order the proteins would elute was unknown. The model managed to predict lower retention time of RNase B compared to RNase A and higher retention of deglycosylated avidin compared to native avidin. In addition, the model was capable of predicting lower retention of the deamidated variants of RNase B as compared to

the non-deamidated variant of RNase B. This is in agreement with the experimental results, which validated the QSPR model.

The differences in elution ionic strengths (I_M) are explained by normalised charge or electrostatic descriptors. From *Figure 17* it can be seen that the experimental elution ionic strengths of all proteins could not be explained solely by their average surface potential. Other descriptors may be important when predicting experimental elution ionic strength for more unusual types of proteins. However, for the protein/glycoprotein pairs studied the average surface potential is related to the experimental elution ionic strength in an almost linear manner.

5 Acknowledgements

My deepest thanks to my supervisor Jinyu Zou for all his enthusiastic ideas and guiding throughout the project. Special thanks to my second supervisor Ulrika Hjellström-Nilsson for all her support and practical advice in the laboratory. I would also like to thank Enrique Carredano for giving me the opportunity to do my Master Thesis at R&D Protein Separations, GE Healthcare and for guiding me through the QSPR analysis. Thanks to Jerry Ståhlberg for his generosity and for all the good advice concerning deglycosylation. Thanks also to Tomas Björkman for introducing me to MALDI-ToF MS and to Anna Grönberg for lending me her ÄKTA™ Explorer 10. I would like to thank all the members of RUCB, RUCC and RULA for always being helpful and making me feel welcome. Finally, special thanks to my opponents Joanna Källgren and Ronnie Jansson for all your help.

6 Abbreviations

CDGS	Carbohydrate-deficient glycoprotein syndrome
Cond	Elution conductivity
CV	Column volumes
Endo	Endoglycosidase
Fuc	Fucose
Gal	Galactose
GalNAc	N-acetylgalactosamine
Glc	Glucose
GlcA	Glucuronic acid
GlcNAc	N-acetylglucosamine
HEMPAS II	Congenital dyserythropoietic anemia type II
HIC	Hydrophobic interaction chromatography
HR	High resolution
IdA	Iduronic acid
IEC	Ion exchange chromatography
I_M	Elution ionic strength
LADII	Leukocyte adhesion deficiency syndrome II
LMW	Low molecular weight marker
MALDI-ToF	Matrix assisted laser desorption/ionization–time-of-flight
Man	Mannose
MS	Mass spectrometry
MW	Molecular weight
m/z	Mass-to-charge ratio
NeuNAc	N-acetylneuraminic acid
PBS	Phosphate-buffered saline
PCR	Principal Component Regression
PDB	Protein data bank
pH-HIC	Hydrophobic interaction chromatography with pH-sensitive polymers
pI	Isoelectric point
PLS	Partial least squares
QSPR	Quantitative structure-property relationships
RNase	Ribonuclease
rtPA	Recombinant tissue plasminogen activator
SDS	Sodium dodecylsulphate
SDS-PAGE	Sodium dodecylsulphate–polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
TFMS	Trifluoromethanesulfonic acid
w/v	Weight per volume
Xyl	Xylose

7 References

1. Arnold U., Schierhorn A., Ulbrich-Hofmann R., 1998. Influence of the Carbohydrate Moiety on the Proteolytic Cleavage Site in Ribonuclease G. *Journal of Protein Chemistry.* **17**, 397-405.
2. Varki A., Cummings R., Esko J., Freeze H., Hart G., Marth J., 1999. Essentials of Glycobiology. *Spring Harbour Laboratory Press*. La Jolla.
3. Gates R., Rathbone E., Masterson L., Wright I., Electricwala., 2004. Glycoprotein Analysis: MANUAL, 1st edition. *SIGMA-Aldrich Corporation.*
4. Chakel JA., Pungoe E Jr., Hancock WS., Swedberg SA., 1997. Analysis of recombinant DNA-derived glycoproteins via high-performance capillary electrophoresis coupled with off-line matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *Journal of Chromatografi B: Biomedical Sciences and Applications.* **689**, 215-220.
5. Taylor M., Drickamer K., 2003. Introduction to Glycobiology. *Oxford University Press.* Oxford.
6. Rudd PM., Woods RJ., Wormald., Opdenakker G., Downing AK., Campbell ID., Dwek RA., 1995. Review: The effects of variable glycosylation on the functional activities of ribonuclease, plasminogen and tissue plasminogen activator. *Biochemica et Biophysica Acta* **1248,** 1-10.
7. Lis H., Sharon N., 1993. Protein glycosylation. Structural and functional aspects. *European Journal of Biochemistry*. **218**, 1-27.
8. Jaenicke R., 1991. Protein folding: local structures, domains, subunits, and assemblies. *Biochemistry* **30**, 3147-3161.
9. Mer G., Hietter H., Lefevre JF., 1996. Stabilization of proteins by glycosylation examined by NMR analysis of a fucosylated proteinase inhibitor. *Nature Structural Biology.* **3**, 45-53.
10. Wren-Gray C., The CDG Family Network. http://*www.cdgs.com* (Feb 2005).
11. Williams DB., Lennarz WJ., 1984. Control of Asparagine-linked Oligosaccharide Chain Processing: Studies of Bovine Pancreatic Ribonuclease B. *The Journal of Biological Chemistry.* **259,** 5105-5114.
12. Plummer TH., Tarentino A., Maley F., 1968. The glycopeptide linkage of ribonuclease B. *The Journal of Biological Chemistry.* **243**, 1459-1465.
13. Gotte G., Libonati M., Laurents., 2003. Glycosylation and Specific Deamidation of Ribonuclease B Affect the Formation of Three-dimensional Domain-swapped Oligomers. *The Journal of Biological Chemistry.* **278**, 46241-46251.
14. Plummer TH., Hirs CHW., 1963. The isolation of Ribonuclease B, a Glycoprotein, from Bovine Pancreatic Juice. *Journal of Biological Chemistry.* **238,** 1396–1401.
15. Laitinen OH., Hytönen VP., Ahlroth MK., Pentikäinen OT., Gallagher C., Nordlund HR., Ovod V., Marttila AT., Porkka E., Heino S., Johnson MS., Airenne KJ., Kulomaa MS., 2002. Chicken avidin-related proteins show altered biotin-binding and physica-chemical properties as compared with avidin. *Biochemical Journal.* **363,** 609-617.
16. SIGMA, ProductInformation, Avidin from Egg White, 2005.
17. Bruch RC., White HB., 1982. Compositional and Structural Heterogeneity of Avidin Glycopeptides. *Biochemistry.* **21,** 5334-5341.
18. An HJ., Peavy TR., Hedrick JL., Lebrilla CB., 2003. Determination of N-glycosylation sites and site heterogeneity in glycoproteins. *Analytical Biochemistry.* **75(20),** 5628-5637.
19. Lattova E., Perreault H., Krokhin O., 2004. Matrix-assisted laser desorption/ionization tandem mass spectrometry and post-source decay fragmentation study of phenylhydrazones of N-linked oligosaccharides from ovalbumin. *Journal of the American Society for Mass Spectrometry*. **15(5),** 725-735.
20. Wilson K., Walker J., 2000. Principles and Techniques of Practical Biochemistry, 5th edition. *Cambridge University Press.*Cambridge.
21. Amersham Biosciences, 2002. Ettan MALDI-ToF, Method Handbook, edition AB.
22. Amersham Biosciences, 2004. Ion Exchange Chromatography & Chromatofocusing, Principles and Methods, edition AA.
23. Amersham Biosciences, 2004. Biodirectory. Product Catalog.
24. Amersham Biosciences, 2002. Gel Filtration, Principles and Methods, edition AI.
25. Amersham Biosciences, 2000. Hydrophobic Interaction Chromatography, Principles and Methods, edition AB.
26. Haga K., Kanai R., Sakamoto O., Aoyagi M., Harata K., Yamane K., 2003. Effects of essential carbohydrate/aromatic stacking interaction with Tyr100 and Phe259 on substrate binding of cyclodextrin glycosyltransferase from alkalophilic Bacillus sp. 1011. *Journal of Biochemistry, Tokyo*. **134,** 881-891.

27. Leach A.R., 2001. Molecular Modelling: Principles and Applications, 2nd edition. *Prentice Hall, Pearson Education EMA.* New York.
28. Gallegos Saliner A., 2004. Molecular Quantum Similarity in QSPR: Applications in Computer-Aided Molecular Design. Doctoral thesis. University of Girona, Spain.
29. Kishimoto T., Hori H., Takano D., Nakano Y., Watanabe M., Mitsui T., 2001. Rice alpha-mannosidase digesting the high mannose glycopeptide of glutelin. *Physiologia Plantarum.* **112(1),** 15-24.
30. Beynon R., Buffers for pH control. http://www.bi.umist.ac.uk/users/mjfrbn/buffers/makebuf.asp (Jan-Jun 2005).
31. Leonidas DD., Shapiro R., Irons LI., Russo N., Acharya KR., 1997. Crystal structures of ribonuclease A complexes with 5'-diphosphoadenosine 3'-phosphate and 5'-diphosphoadenosine 2'-phosphate at 1.7 A resolution. *Biochemistry.* **36(18),** 5578-5588.
32. Nardone E., Rosano C., Santambrogio P., Curnis F., Corti A., Magni F., Siccardi AG., Paganelli G., Losso R., Apreda B., Bolognesi M., Sidoli A., Arosio P. 1998. Biochemical characterization and crystal structure of a recombinant hen avidin and its acidic mutant expressed in Escherichia coli. *European Journal of Biochemistry.* **256(2),** 453-460.
33. Skarp U., 2003. Relating surface potentials to cation and anion exchange retention of protein. Master's degree project. ISSN: 1401-2138, UPTEC: 03 015.
34. Malmquist G., Hjellström-Nilsson U., Zou J., Granér T., Norrman M., Skarp U., Strömgren M., Carredano E., 2005. Electrostatic calculations and QSPR protein retention models for ion exchange chromatography. Manuscript submitted to *Protein Science*.
35. CAMO 2005. CAMO PROCESS AS. Nedre Vollgate 8, N-0158 OSLO, Norway.
36. Wested F., Martens H. 2000. Variable selection in near infrared spectroscopy based on significance testing in partial least square regression. *Journal of Near Infrared Spectroscopy,* 8, 117-124.
37. Hoja-Lukowicz D., Ciolczyk D., Bergquist J., Litynska A., Laidler P., 2000. High-mannose-type oligosaccharides from human placental arylsulfatase A are core fucosylated as confirmed by MALDI MS. *Glycobiology.* **10(6),** 551-557.
38. Haggerty L., Lenhoff AM., 1991. Relation of protein electrostatics computations to ionexchange and electrophoretic behaviour. *Journal of Physical Chemistry.* **95,** 1472-1477.
39. SIGMA, ProductInformation, Albumin from Chicken Egg White, 2005.
40. Quiocho FA., 1990. Atomic structures of periplasmic binding proteins and the high-affinity active transport systems in bacteria. *Philos Trans R Soc Lond B Biol Sci.* **326(1236),** 341-51.
41. Heukeshoven J., Dernick R., 1988. Increased sensitivity for Coomassie staining of sodium dodecyl sulfate-polyacrylamide gels using PhastSystem Development Unit. *Electrophoreses*. **9**, 60-61.

8 Appendices

8.1 Appendix A

Unicorn method for Size Exclusion Chromatography
Method: v:\UNICORN\Local\Fil\Therese\Method\Gelfiltration.m20

Log Format

Base: Time
Unit: min

Main method:
(Main)
0.00 Base CV 23.5 {ml} Any
0.00 ColumnPosition (Position5)#Col_pos
0.00 Wavelength 280 {nm} 254 {nm} 215 {nm}
0.00 AveragingTimeUV 2.56 {sec}
0.00 Block Equilibration
(Equilibration)
0.00 Base SameAsMain
0.00 AutoZeroUV
0.00 Flow 0.50 {ml/min}
0.00 BufferValveA1 A11
14.10 End_Block
14.10 Block Sample_load
(Sample_load)
14.10 Base Volume
14.10 Flow (0.5)#sample_flow {ml/min}
14.10 InjectionMark
14.10 InjectionValve Inject
16.50 InjectionValve Load
28.10 Fractionation_900 0.4 {ml}
74.10 Fractionation_Stop_900
74.10 End_Block
End_Method

8.2 Appendix B

The electrophoresis experiments were performed on 8-25% gradient gels using PhastSystem Separation-Control Unit (GE Healthcare) and Development Unit.

8.2.1 Separation Method

SAMPLE APPL. DOWN AT	1.1	1	Vh
SAMPLE APPL. UP AT	1.1	10	Vh
EXTRA ALARM TO SOUND AT	1.1	100	Vh

Table 12. Separation method used for SDS-PAGE on 8-25% gradient gels. The method was programmed into the control unit.

SEP	Volt (V)	mA	Watt (W)	Temp. (°C)	Volt hours (Vh)
1.1	250	10	3	15	120
1.2	50	0.1	0.5	15	80

8.2.2 Development Method

The staining method is according to [41].

DEV 8 Ct (5, 30, 40, 50) °C = (1.0, 1.0, 1.0, 1.0)

EXTRA ALARM TO SOUND AT	1.0	0	min

Table 13. Method used for development of SDS-PAGE PhastGel Gradient 8-25%. The method was programmed into the control unit.

DEV	Stains/Solutions	IN-port	OUT-port	Time (min)	Temp. (°C)
1.01	30% ethanol, 10% Acetic acid solution	5	0	4.0	50
1.02	30% ethanol, 10% Acetic acid solution	5	0	4.5	50
1.03	Coomassie™ Stain solution	6	0	20.0	50
1.04	30% ethanol, 10% Acetic acid solution	5	0	0.1	50
1.05	10% Acetic acid solution	4	0	5.0	50
1.06	10% Acetic acid solution	4	0	10.0	50
1.07	20% glycerol solution	9	0	5.0	50
1.08	Milli-Q water	1	0	0.4	50

8.3 Appendix C

Unicorn method for Ion Exchange Chromatography

Method:
v:\UNICORN\Local\Fil\Therese\Method\Retentionautosamplerscouting\Retentionautosamplerscouting.m20

Log Format
Base: Time, Unit: min

Main method:

(Main)
0.00 Base CV (1.0)#cv {ml} Any
0.00 Block Start_conditions
(Start_conditions)
0.00 Base SameAsMain
0.00 Flow 1.0 {ml/min}
0.00 Wavelength 280 {nm} 254 {nm} 215 {nm}
0.00 ColumnPosition Position2
0.00 AveragingTimeUV 2.56 {sec}
0.00 Alarm_Pressure Enabled 8.00 {MPa} 0.00 {MPa}
0.00 PumpWashExplorer OFF OFF OFF (OFF)#B_pump_wash
0.00 End_Block
0.00 Block Equilibration
(Equilibration)
0.00 Base SameAsMain
0.00 PumpWashExplorer (A11)#equilibration_buffer_wash OFF OFF OFF
0.00 BufferValveA1 (A11)#equilibration_buffer
4.5 AutoZeroUV
4.6 Set_Mark ()#column_name
7.00 End_Block
0.00 Block Sample_loading
(Sample_loading)
0.00 Base SameAsMain
0.00 InjectionPartial (1)#Injection_vial 50 {µl}Yes NoAir
0.50 Set_Mark ()#Sample
3.00 End_Block
0.00 Block Elute
(Elute)
0.00 Base SameAsMain
0.00 PumpBInlet (B2)#B_inlet
0.00 Gradient 100.00 {%B} 40.00 {base}
50.00 Gradient 0.00 {%B} 0.00 {base}
50.00 End_Block
0.00 Block CIP
(CIP)
0.00 Base Volume
0.00 PumpWashExplorer A11 OFF OFF OFF
0.00 BufferValveA1 (A11)#cippningsbuffert
0.00 Flow 0.5 {ml/min}
10.00 End_Block
0.00 Block Re_equilibration
(Re_equilibration)
0.00 Base SameAsMain
0.00 Flow 1.0 {ml/min}
0.00 PumpWashExplorer (A11)#reequilibration_buffer_wash OFF OFF OFF
0.00 BufferValveA1 (A11)#reequilibration_buffer

(7.00) # re_equilibration_volume End_Block
0.00 End_Method

8.4 Appendix D

Unicorn method for Size Exclusion Chromatography

Method: v:\UNICORN\Local\Fil\Therese\Method\Gelfiltrering Scouting.m20

Log Format

Base: CV
Unit: 23.5 ml

Main method:

(Main)
0.00 Base CV 23.5 {ml} Any
0.00 ColumnPosition (Position5) #Col_pos
0.00 Wavelength 280 {nm} 254 {nm} 215 {nm}
0.00 AveragingTimeUV 2.56 {sec}
0.00 Flow 0.40 {ml/min}
0.00 Alarm_Pressure Enabled 1.50 {MPa} 0.00 {MPa}
0.00 PumpWashExplorer (OFF) #PumpWash OFF OFF OFF
0.00 Block Equilibration
(Equilibration)
0.00 Base SameAsMain
0.00 Flow 0.65 {ml/min}
(1.00) #Equilibration_Volume Hold_Until UV1 Stable_Baseline 3.00 {minutes} 1 {base}
1.00 Set_Mark ( ) #Column_Name
1.00 AutoZeroUV
1.00 End_Block
0.00 Block Sample_load
(Sample_load)
0.00 Base Volume
0.00 Flow (0.40)#sample_flow {ml/min}
0.50 Set_Mark ( ) #Sample
3.00 InjectionPartial (1) #Injection_vial 120 {µl} Yes NoAir
6.00 End_Block
0.00 Block Elution
(Elution)
0.00 Base SameAsMain
0.00 Flow (0.40) #Elution_flow {ml/min}
1.00 End_Block
End_Method