Proteomics approach to investigate separation behavior of proteins in ion-exchange chromatography

by

Li Xu

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Chemical Engineering

Program of Study Committee:

Charles E. Glatz (Major Professor)

Surya K. Mallapragada

Daniel S. Nettleton

Iowa State University

Ames, Iowa

2008

TABLE OF CONTENTS

ACKNOWLEDGMENTS

Grateful thanks for the 3D characterization method provided by Dr. Zhengrong Gu and Dr. Charles. E. Glatz. In particular, I should like to thank for Dr. Charles. E. Glatz for his instruction on this project and his valuable comments on this thesis. I would also thanks to my research group members for all the help they gave me during past years.

GENERAL ABSTRACT

In the present study, protein isoelectric point (pI), molecular weight (MW) and partitioning coefficient (K) were characterized for a mixture of proteins using a three-dimensional (3D) method which combined hydrophobic partitioning with two-dimensional (2D) electrophoresis. Protein pI and MW were obtained from 2D gels directly using the known pI gradients and the positions of MW standards. Protein partition coefficients were obtained by quantifying protein spot mass for the top and bottom phase gels so that mass ratios could be calculated for matched spots. The characterized three protein molecular properties were used to represent protein charge, size and surface hydrophobicity, which showed different degrees of influence on protein separation behaviors in ion-exchange chromatography (IEC).

Statistical models correlated the three characterized protein properties to retention times in cation-exchange chromatography (CEC) using partial least squares (PLS) regression. The resulting models fit well (R^2=0.913 and 0.873 for SP and 15S, respectively) considering the limited property basis and the regression models were able to predict results for a small test set of proteins. The models showed that pI and MW correlated positively with CEC retention time, while the net influence of the partition coefficient depended on the base matrix type. This approach could be extended to host protein extracts to provide guidance for purification of recombinant proteins or choice of a suitable host for a particular recombinant protein.

CHAPTER 1.

GENERAL INTRODUCTION

1.1 Introduction

Recovery and purification of protein is the most costly part in recombinant protein production because multiple separation steps are usually needed to remove major contaminants such as host proteins, nucleic acids, and medium components. A great amount of experimental work is needed because of the lack of elemental knowledge as regards the molecular properties of the host proteins. The most frequently used protein separation methods usually operate based on the differences in surface hydrophobicity, charge and size among target and host proteins. Investigation of those protein properties for commonly used hosts and their influences on fundamental separation techniques could establish a database for rational selection of purification strategies for recombinant proteins.

In addition to being used as guidelines for protein purification, the properties and their relationship with separation behavior could also be used to select among alternative production hosts or expression target organs on the basis of simplified protein purification.

Ion-exchange chromatography (IEC) is one of the most frequently used protein separation technique because of its high resolution, modest cost and mild operating conditions. The main interaction for binding and elution in IEC is electrostatic attraction, so protein charge should be primarily responsible for binding in IEC. However, a number of studies showed other protein molecular properties such as protein size, hydrophobicity, shape, and charge distribution also influenced protein binding in IEC (Malmquist et al., 2006; Hallgren et al., 2000; Yao et al., 2004; Xu et al., 1998). Unfortunately, most of those studies were based on of a limited set of known proteins and their models could hardly estimate the behaviors of a large number of unknown proteins in IEC.

This project is trying to simplify recombinant protein purification strategies and select proper expression targeting by studying the measurable molecular properties of host proteins and their relationship to IEC. In order to achieve these goals, two things need to be done: first, find a way to

measure protein molecular properties of large numbers of host proteins; second, test the correlation between those measured protein molecular properties and IEC retention. The first task has been done by Gu and Glatz (2007) using a three-dimensional (3D) method to characterize the surface hydrophobicity (SH), isoelectric point (pI) and molecular weight (MW) of native corn proteins. The second task will focus on two aspects: 1) modeling the correlation of the measured protein properties and IEC retention times using a collection of model proteins for which IEC data are already available, which is addressed in this thesis; 2) extending that correlation to native corn proteins, which will be investigated in the future work.

1.2 Thesis Organization

This thesis includes three chapters. The first chapter is the general literature review, including the importance of downstream process for recovery of recombinant protein, how protein properties influence separation, relationships between protein properties and IEC behavior, modeling of IEC retention based on protein properties, and the methods available to measure the three fundamental protein properties.

In the second chapter, a statistical model is reported that correlates ion-exchange behavior with properties obtained via three-dimensional characterization of proteins. The pI, MW and aqueous two-phase partitioning coefficients of a set of model proteins were related to retention time in cation-exchange chromatography (CEC) using partial least squares regression. How well those same three properties could be determined for a mixture of proteins was also examined using methods reported previously (Gu and Glatz, 2007). The resulting model fit well (r^2=0.905) considering the limited property basis, and the regression model was able to predict results for a small test set of proteins. The model showed that pI and MW correlated positively with CEC retention time, while the partitioning coefficient was negatively correlated.

The third chapter gives the overall conclusions.

1.3 Literature Review

Importance of downstream recovery of recombinant proteins

With the rapid development of transgenic techniques, a variety of hosts, including prokaryotes, such as *Escherichia coli*, eukaryotes, such as yeast, mammalian and insect cells, animals, and plants, have been developed and utilized to produce recombinant protein (Kusnadi et al., 1997). By the end of 2007, 335 target proteins had been expressed in viruses, 8043 in *Achaea*, 58806 in bacteria, and 42239 proteins in *Eukarya* (Gräslund et al., 2008).

However, among those recombinant proteins, only 35% of the viruses expressed proteins, 36% of the *Achaea* expressed proteins, 30% of the bacteria expressed proteins and 19% of the *Eukarya* expressed proteins have been purified (Gräslund et al., 2008). Actually, recovery and purification of recombinant protein is more challenging than production, and it is usually made up of 50-80% of the total production cost for recombinant proteins. Mison and Evangelista (Menkhaus et al, 2004) calculated that 94% of the annual operating cost for producing ß-glucuronidase from transgenic maize was attributed to downstream steps, while the grain production was only 6%. Thus, downstream processing is very important in the whole recombinant protein production process.

Host cell proteins are one of the major contaminants to be removed to achieve highly purified recombinant proteins. However, extensive experimentation is needed for process development because of the lack of data on protein properties and limited models for how those proteins determine separation behavior. Remedying those deficiencies would reduce the costs of process development for large-scale recombinant protein production.

Processing and purification of recombinant proteins from transgenic corn

Tobacco, corn, soybean, canola and alfalfa have been used as plant hosts to produce recombinant protein (Whitelam et al., 1993; Khoudi et al., 1999; Menkhaus et al., 2004). Among those crops, corn was expected to be one of the most likely hosts for the commercial production of recombinant proteins. The first reason is its abundant protein production per planted acre, which is the same as

protein produced by soy (Menkhaus et al., 2004). Furthermore, as the plant is a natural storage organ, corn seeds can accumulate recombinant proteins and be stored for a long time without degradation (Menkhaus et al., 2004; Kusnadi, et al., 1997). Other advantages of corn hosts are established transformation methods, ability to target the recombinant proteins into the different corn grain fractions. Recombinant proteins, such as aprotinin (Azzoni et al., 2002; Zhong et al., 1999), avidin (Hood et al., 1997), ß-glucuronidase (Kusnadi et al., 1998a; Dharmadi et al., 2003) and dog gastric lipase (zhong et al., 2006; Gu and Glatz, 2007), have been reported to be separated or produced from transgenic corn. However, transgenic corn has their own disadvantages. The target proteins it produced usually have different glycosylation from cell culture products (Giddings, 2001). There are also public concerns about pollen and gene contamination, and food safety problem. In addition, the low accumulation levels of recombinant proteins and lack of data on downstream processing of plant systems (Goddijn and Pen, 1995) also make corn is not a so likely host in the future recombinant protein production.

Native corn proteins are the contaminants that are most difficult to be removed from recombinant proteins because of the high degree of the size and charge heterogeneity among the proteins that could be in a corn extract as shown in Table 1 (Menkhaus et al., 2004).

Table 1. Size and charge characteristics of the major proteins in corn (Menkhaus, 2004)

	Protein class			
Distinction	Albumins	Globulins	Glutelins	Prolamins
Molecular Weights	15-20 distinct fractions from 16-400 kDa	15-20 distinct fractions from 16-400 kDa	20 components from 11-127 kDa (in a reduced state)	Two major groups of 45 kDa and 22 kDa
Isoelectric Points	Most insoluble at pH 4.4; most soluble at pH 8.6	8 bands from pH 4-9	6 bands from pH 5-7	10 bands from pH 5-9

The general process for recovery of recombinant proteins from plants includes (1) fractionation of plant material; (2) protein extraction and solid/liquid clarification; and (3) protein purification.

There are well-established methods to isolate corn germ and endosperm from whole corn kernel (Johnson, 2000), and the fractionation of plant tissue can reduce the initial solids and enrich target proteins. Because of the solubility differences between native proteins of the specific kernel fractions, the purification process could be simplified if recombinant proteins are targeted to one of those fractions.

The purposes of extraction here are: (1) remove most solids such as insoluble carbohydrates (including fiber) and ash; (2) release the target protein into aqueous environment. Depending on where protein expression is targeted, two methods have been used to extract proteins from plant materials: One is the homogenization of green tissues, such as tobacco leaves or alfalfa; the other is dry-grinding followed by aqueous buffer extraction of seed or seed fractions (Kusnadi et al., 1997; Menkhaus et al., 2004). Solubility of both recombinant and host proteins should be compared to choose extraction conditions (i.e. pH, ionic strength, and using of detergent) for efficient releasing of recombinant proteins from plant materials while limiting the release of native proteins.

The costs and methods employed in the next stage of purification of the extract depend on the value and the intended usage of the products. For lower-value protein products such as industrial enzymes, lower-cost chromatography such as traditional ion-exchange chromatography, or non-chromatographic methods such as precipitation and aqueous two phase partitioning, could be employed to separate the final products. For the purification of high-value pharmaceutical proteins, costlier methods such as affinity chromatography can be employed to reach the required purity. Both sets of techniques are based on the differences among the molecular properties of native and recombinant proteins. Table 2 summarizes previous purification processes for recombinant proteins from corn seed.

Table 2. Downstream processing evaluations made on different recombinant proteins from corn (Adapted from Menkhaus et al., 2004)

Corn fraction	Protein	Purification	Reference
Corn (whole kernel or germ-rich)	Recombinant ß-glucuronidase	Adsorption on anion-exchange resin [a, b] Hydrophobic interaction chromatography [a, b] Anion exchange chromatography [a, b, c] Size exclusion chromatography [c]	[a] Kusnadi et al., 1998a [b] Kusnadi et al., 1998b [c] Witcher et al., 1998
Corn (whole kernel or germ-rich)	Recombinant avidin	Affinity (2-iminobiotin) [b, d]	[b] Kusnadi et al., 1998b [d] Hood et al., 1997
Corn (whole kernel or germ-rich)	Recombinant aprotinin	Immobilized metal affinity chromatography [e] Adsorption with cation exchange resin [f, g] Affinity chromatography (anti-aprotinin antibody) [f] Reverse phase chromatography [f] Heat precipitation [g] Hydrophobic interaction chromatography [g]	[e] Zhong et al., 1999 [f] Azzoni et al., 2002 [g] Zhong et al., 2007
Corn (endosperm-rich)	Recombinant dog gastric lipase	Aqueous two phase partitioning [h]	[h] Gu et al., 2007

Protein properties and their effects on purification methods

Protein purification usually includes several steps to reach the required product purity. Although additional steps generally improve purity, they also decrease the final product yield. So it is important

to minimize the purification steps by selecting of the proper protein separation techniques and arranging them in rational sequence. Table 3 shows protein properties and their influence on purification strategy. Knowing the properties of recombinant and host proteins will aid selection of proper separation methods and minimize purification steps.

Table 3. Protein properties and their effect on development of purification strategies (Adapted from Protein Purification Handbook, Amersham Pharmacia Biotech, 2001)

protein properties	**Influence on purification strategy**
Molecular weight (protein size)	Selection of pore size of gel filtration resin or ultrafiltation membrane
Charge	Selection of conditions for ion exchange chromatography or isoelectric focusing
Hydrophobicity	Selection of medium for hydrophobic interaction chromatography or reverse phase chromatography
Biospecific affinity	Selection of ligand for affinity medium
Co-factors for stability or activity	Selection of additives, pH, salts, buffers
Protease sensitivity	Need to add protease inhibitors
Sensitivity to metal ions	Need to add EDTA or EGTA to buffers
Redox sensitivity	Need to add reducing agents
Temperature stability	Need to work rapidly at lowered temperature
pH stability	Selection of buffers for extraction or selection of conditions for chromatography
Detergent requirement	Selection types of detergent
Post translational modifications	Selection of group-specific affinity medium
Organic solvent stability	Selection of conditions for reversed phase chromatography

Among different separation techniques, chromatography is the most commonly used because of its high resolution and relatively gentle operating conditions. Most chromatographic methods separate proteins based on the differences in four specific protein properties: **charge, size, hydrophobicity and ligand specificity**. Unfortunately, insufficient information on those properties of host proteins

makes it difficult to select the proper purification technique (Kusnadi et al., 1997). Hence, investigation of those properties of host proteins and how they influence chromatographic behavior is necessary before selection of purification process.

Among available chromatographic techniques, ion-exchange chromatography, hydrophobic interaction chromatography, reverse phase chromatography, and gel filtration chromatography are the most frequently used ones. All of those chromatographic methods separate proteins according to one or several particular properties mentioned above except for ligand affinity. Relatively costly affinity chromatography separates proteins according to their ligand specificity. Thus, protein charge, size and hydrophobicity can be considered as the important characteristics that influence protein behavior in commonly used separation techniques. Therefore, in order to simplify the downstream processing and improve the yields of recombinant proteins, it is desirable to characterize those three protein properties and investigate how they affect separation behavior prior to selecting proper chromatography techniques and experimental conditions.

Effect of protein properties on ion-exchange chromatography behavior

IEC is one of the most frequently used techniques in the field of protein purification because of its high resolution and mild running conditions. IEC depends on the reversible adsorption of charged solute molecules to immobilized ion exchange groups of opposite charge. The mobile phase is usually a buffer solution containing an eluting salt, and the protein is strongly adsorbed when the stationary phase and protein are of opposite charge at low salt concentrations. When the salt concentration in the buffer increases, the adsorption strength will decrease rapidly and the proteins will elute.

When considering protein separation by IEC, it is assumed the net charge is the major factor which influences protein elution in IEC. However, a number of studies found that protein surface charge (Regnier, 1987; Malmquist et al., 2006), charge asymmetry (Kopaciewicz, et. al., 1983), and average potential over the molecular surface (Haggerty and Lenhoff, 1991) are all responsible for the protein's elution behavior in IEC. In addition to electrostatic interaction, Stahlberg (1999) also

mentioned van der Waals interaction as the other important interaction in IEC. Yao et al. (2004) found that not only the charge, but protein structural properties, such as size, shape, and charge distribution, may play significant roles in protein's retention time. Hallgren et al. (2000) and Xu et al. (1998) also found protein charge distribution and molecular shape were responsible for the protein performance in IEC. In addition, Xu and Regnier (1998) found that protein-protein interaction influenced IEC retention when the column was overloaded. Mazza et al. (2001) used a statistical model to reduce several hundred molecular descriptors to 8-19 that were significant to IEC behavior. Ladiwala et al (2005) used a quantitative structure–property relationship models to identify molecular properties of protein influencing IEC behavior. Later, Malmquist et al. (2006) correlated 58 protein molecular descriptors with protein behavior in IEC. The majority of those descriptors are related to the charge, the rest were three size and shape descriptors, and eight surface hydrophobicity descriptors. In general, protein behavior in IEC is not solely determined by its charge, but also by its shape, charge distribution, hydrophobic patches and even the loading conditions.

Recent research which compared IEC and pI from two-dimensional (2D) gel electrophoresis found that one could not select IEC running conditions only based on the difference between denatured pI and solution pH (Butt et al., 2001). They suggested that the retention is modified by surface charge features, by formation of protein-protein complexes (in which the combined pI is different to the individual pI of the constituents) or by non-ionic interactions with the chromatographic matrix.

Most of the above studies investigated how IEC performance relates to protein properties by using a set of known proteins, for which extensive information of protein molecular properties was available. For large numbers of unknown proteins, it is not practical to characterize so many properties and relate them to IEC behavior. However, the literature reviewed above cumulatively notes the influence of pI, size and hydrophobicity. Since it is practical to characterize those three

protein properties for large numbers using the method of Gu and Glatz (2007), it is worth seeing to what degree separation behavior prediction can be based on those three.

Modeling of protein retention in ion-exchange chromatography

The stoichiometric displacement model (SDM) is the earliest model to depict protein binding to ion-exchangers (Boardman and Partridge, 1955; Roth et al., 1996). The fundamental idea of SDM is that when adsorption happens, a protein molecular displaces ions stoichiometrically. At equilibrium this exchange can be expressed as equation 1 (Regnier, 1987; Stahlberg et al., 1991)

$$P_0 + Z \cdot C_b \Leftrightarrow P_b + Z \cdot C_0 \quad (1)$$

where P_0 and P_b refer to free and bound protein concentrations in the mobile and stationary phase, respectively; C_b and C_0 are the bound and free counterion concentrations, and Z is the number of charges displaced during the ion exchange process. The chromatography retention factor k can be related to Z and C_0 by

$$\log k = \log I + \log(C_0)^{-Z} \quad (2)$$

where $k = (t_r - t_0)/t_0$, and t_r is the retention time for the retained sample while t_0 is the time for an unretained solute to flow through the column. I is a constant which relates to both specific affinity and non-specific interaction between the solute and the stationary phase (Chicz and Regnier, 1989). Equation 2 can be used to find the Z number of a solute and thus associate protein charge to the IEC elution behavior. Because of the steric effects, this Z number only represents those protein charges which interact with the stationary phase (Regnier, 1987).

The SDM does not directly address the electrostatic interaction force, which has major influence on protein retention in ion-exchange chromatography (Stahlberg and Jonsson, 1996). To compensate for this limitation of SDM, Stahlberg et al (1991) proposed a "slab model" by solving the linearized Poisson-Boltzmann equation of two charged planar surfaces in a solution of electrolytes. The retention factor k was expressed as

$$\ln k = \frac{A_p \sigma_p^{\,2}}{F\left(2RT\varepsilon_0\varepsilon_r\right)^{1/2}} \frac{1}{\sqrt{I}} + \ln \Phi \qquad (3)$$

where A_p is the interacting area between the protein surface and the charged stationary phase, σ_p is the protein charge density, I is the total buffer ionic strength, ε_r is the dielectric constant of the medium between the surfaces, ε_0 is the permittivity of vacuum, R is the universal gas constant, and Φ is the column phase ratio. As this model assumes the protein is a sphere with uniform distributed surface charge, and only half of the total protein area interacts with the stationary phase, the protein net charge can be expressed as $2 A_p \sigma_p$, which can be determined from IEC retention data by the slope (k') of $\ln k$ vs $1/\sqrt{I}$ giving, by rearrangement of equation 3, $2 A_p \sigma_p =$ $2\sqrt{k' A_p F\left(2RT\varepsilon_0\varepsilon_r\right)^{1/2}}$). Thus, this model relates both protein size (as A_p can be used to calculate protein volume) and charge (as protein net charge can be determined by $2 A_p \sigma_p$) to IEC retention. Later, Jonsson and Stahlberg (1999) extended this model by treating the binding protein as a charged sphere rather than a planar surface. This gave better prediction of protein net charge based on IEC elution data than the previous model. However, the assumption of a homogeneous surface charge distribution will result in error if the charge distribution of a protein is highly asymmetrical. For example, Hallgren et al. (2000) found that two mutants of staphylococcal nuclease A showed different retention behavior although they had same net charge with the difference in retention time attributed to the different distribution of charge.

Recently, Quantitative Structure-Property Relationships (QSPR) have been successfully applied to IEC for correlating protein molecular structure properties with retention behavior (Mazza, et al., 2001; Ladiwala et al., 2005; Malmquist et al., 2006). The basic theory of QSPR is the possibility of quantitatively correlating the performance of a process (such as chemical or biological reaction process) to the molecular structures of the substances which are in that process. Those QSPR model-

related studies have correlated protein IEC elution behaviors with descriptors derivable from known protein three-dimensional structure. The QSPR model gave better prediction of protein IEC retention than earlier models but does require more protein molecular properties. For example, Malmquist et al. (2006) used 58 descriptors which including 17 surface charge distribution descriptors, 28 electrostatic potential distribution descriptors, 8 surface hydrophobicity descriptors and 3 size and shape descriptors.

Availability and measurement of protein properties

For proteins that have been previously purified and crystallized for structure determination, a number of molecular properties can be found in protein databases such as EXPASY.org. However, for the larger numbers of host proteins, especially for hosts which are not commonly used, most of their properties have never been determined. Fortunately, there are different techniques for measuring those protein properties of interest to us.

The size and charge dimensions:

Two-dimensional electrophoresis is the traditional and most frequently used method to measure the pI and MW of proteins. A property indicative of the net charge of a protein would be the difference between pI from 2D electrophoresis and the solution pH (pH-pI); the protein size can be represented by measured MW. 2D electrophoresis separates proteins in two steps: the first dimension, isoelectric focusing, separates proteins based on their pI, which is the pI of denatured, unfolded protein; and the second dimension, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), separates proteins according to their MW. Thus, each resolved spot on the 2D electrophoresis gel represents a single protein species (within the limitations of protein resolution in the gel) and the associated protein MW and pI.

In addition to 2D electrophoresis, protein size can be measured by mass spectrometry (MS). MS is often combined with 2D electrophoresis to obtain sequence information permitting identification of the protein. MS is an expensive method and used more often to identify limited number of proteins

rather than characterize large numbers of host proteins. MS is usually used as the last step of protein purification, while 2D electrophoresis could be used either to obtain host protein properties before protein purification or identify target proteins at the end of purification. A chromatography-based method, chromatofocusing, also can provide a measure of pI of proteins (Pirondini et al., 2006; Qin et al., 2005). This method is a variant of ion-exchange chromatography that elutes from ion exchangers exclusively by the means of pH. The advantage of chromatofocusing is that it can measure protein pI without denaturing the protein, thus it can provide a direct measure of the pI of the folded protein. So, the protein charges indicated by the difference between those surface pIs and solution pH are surface charges which could give better guidelines for IEC running conditions than pI of denatured proteins. However, as chromatofocusing collects groups of proteins in fractions, it can not provide specific pI for individual proteins. Also, charged residues are not often found internally so the denatured protein and native protein pI may not differ in many cases.

The hydrophobic dimension:

As for the characterization of protein hydrophobicity, there are several methods available. Hydrophobic interaction chromatography (HIC) (Huddleston et al., 1996) and reverse phase chromatography (RPC) (Stroink et al., 2005) are two of them. Non-chromatographic techniques such as potassium phosphate precipitation (Huddleston et al., 1996), $(NH_4)_2SO_4$ precipitation (Hachem et al., 1996), ANS (1-(anilino) naphthalene-8-sulfonate) fluorescent staining (Haskard et al., 1998) and aqueous two-phase partitioning system (Andrews et al., 2005; Hachem et al., 1996) also have been used to evaluate hydrophobicity of proteins.

Among those different protein hydrophobicity measuring techniques, ATP has advantages over others when characterizing large numbers of proteins. First of all, the partition coefficients, K, (the ratio of concentrations of top to bottom phase) of individual proteins can be determined in a one-step partitioning of the mixture when combined with 2D electrophoresis quantification of the proteins in each phase (Gu and Glatz, 2007). Second, ATP generates fewer fractions (two fractions) when

compared to RPC, HIC and $(NH_4)_2SO_4$ precipitation. The fewer fractions will decrease the possibility of error because it will reduce experiments in the other two dimensions. The third advantage is that the two phases of ATP system are both aqueous solutions with high water content (70% - 90%), so both phases can maintain protein stability.

There are ATP systems which partition proteins based on the difference of protein hydrophobicity, charge and MW, while polyethylene glycol (PEG)-salt system with added NaCl shows partitioning dominated by protein hydrophobicity (Gu and Glatz, 2007). In a hydrophobic force dominant ATP system, the partition coefficients K could be expressed by equation 4 (Gu and Glatz, 2007; Hachem et al., 1996),

$$Log\ K = R \times Log\ \frac{P_i}{P_0} \qquad (4)$$

Here R is the hydrophobic resolution and P_0 is the intrinsic hydrophobicity of the ATP system.

As P_0 is identical for all the proteins in a particular ATP system, equation (4) can be arranged to equation (5)

$$\Delta Log\ P = \frac{\Delta Log\ K}{R} \qquad (5)$$

R in a particular ATP system can be characterized by a set of proteins with known P. Then, P of an unknown protein can be calculated from its own K plus K and P of one of model proteins in the same system using equation (5). Hence, log K, via equation (5) becomes a measure of surface hydrophobicity.

K values for individual proteins in a complex mixture can be obtained by matching of the corresponding spots in 2D electrophoresis gels from the two phases followed by using the matched spot protein contents to calculate the K values with equation $K_i = C_{i,\,top}/C_{i,\,bottom}$ (Gu and Glatz, 2007). At the same time, the 2D electrophoresis gels can provide the other two protein characteristics.

Thus, by combining ATP together with 2D electrophoresis, the three protein properties-size, charge and hydrophobicity-of a large number of host proteins can be determined.

2D liquid chromatography (chromatofocusing and RPC) followed by MALDI-TOF/MS can also measure protein charge, hydrophobicity and size (Pirondini et al., 2006; Stroink et al., 2005), but the equipment is costly and not available in most protein facilities.

1.4 References

Andrews BA, Schmidt AS, Asenjo JA. 2005. Correlation for the partition behavior of proteins in aqueous two-phase systems: effect of surface hydrophobicity and charge. Biotechnol Bioengr 90:380-390.

Azzoni AR, Kusnadi AR, Nikolov ZL. 2002. Recombinant aprotinin produced in transgenic corn seed: extraction and purification studies. Biotechnol Bioengr 80:269-276.

Boardman NK, Partridge SM. 1955. Separation of neutral proteins on ion-exchange resins. Bioch 59: 543-552.

Butt A, Davison MD, Smith GJ, Young JA, Gaskell SJ, Oliver SG, Beynon RJ. 2001. Chromatographic separations as a prelude to two-dimensional electrophoresis in proteomics analysis. Proteomics 1:42-53.

Chicz RM, Regnier FE. 1989. Single amino acid contributions to protein retention in cation-exchange chromatography: resolution of genetically engineered subtilisin variants. Anal Chem 61:2059-2066.

Dharmadi Y, Chang QL, Glatz CE. 2003. Recovery of enzyme byproducts from potential plant hosts for recombinant protein production. Enzyme Microb Technol 33:596-605.

Giddings G. 2001. Transgenic plants as protein factories. Curr Opin Biotechnol 12: 450 - 454.

Goddijn OJM, Pen J. 1995. Plants as bioreactors. Trends Biotechnol 13:379 -387.

Gräslund S, Nordlund P, Weigelt J. 2008. Protein production and purification. Nature Methods 5:135-146.

Gu Z, Glatz CE. 2007. Recovery of recombinant dog gastric lipase from corn endosperm extract. Sep Sci Technol. 42:1195-1213.

Gu Z, Glatz CE. 2007. A method for three-dimensional protein characterization and its application to a complex plant (corn) extract. Biotechnol Bioengr 97:1158-1169.

Hachem F, Andrews BA, Asenjo JA. 1996. Hydrophobic partitioning of proteins in aqueous two-phase systems. Enzyme Microb Tech 19:507-517.

Haggerty L, Lenhoff AM. 1991. Relation of protein electrostatics computations to ion-exchange and electrophoretic behavior. J Phys Chem 95:1472-1477.

Hallgren E, Kalman F, Farnan D, Horvath C, Stahlberg J. 2000. Protein retention in ion-exchange chromatography: effect of net charge and charge distribution. J Chromatogr A 877:13 - 24.

Haskard CA, Eunice CY, Li C. 1998. Hydrophobicity of bovine serum albumin and ovalbumin determined using uncharged (PRODAN) and anionic (ANS-) fluorescent probes. J Agric Food Chem 46:2671–2677.

Hood EE, Witcher DR, Maddock S, Meyer T, Baszezynski C, Bailey M, Flynn P. 1997. Commercial production of avidin from transgenic maize: characterization of transformant, production, processing, extraction, and purification. Mol Breed 3:291-306

Huddleston J, Abelaira JC, Wang R, Lyddiatt A. 1996. Protein partition between the different phases comprising poly (ethylene glycol)-salt aqueous two-phase systems, hydrophobic interaction chromatography and precipitation: a generic description in terms of salting-out effects. J Chromatogr B 680:31-41.

Johnson LA. 2000. Corn: The major cereal of the Americas. In handbook of cereal science and technology, 2nd Ed., New York:Marcel Dekker Inc. 31-80p.

Khoudi H, Laberge S, Ferullo JM, Bazin R, Darveau A, Castonguay Y, Allard G, Lemieux R, Vezina LP. 1999. Production of monoclonal antibody in perennial alfalfa plants. Biotechnol Bioengr 64:135-143.

Kopaciewicz W, Rounds MA, Fausnaugh J, Regnier FE. 1983. Retention model for high-performance ion-exchange chromatography. J Chromatogr 266:3-21.

Kusnadi AR, Nikolov ZL, Howard JA. 1997. Production of recombinant proteins in transgenic plants, practical considerations. Biotechnol Bioengr 56:473 - 484.

Kusnadi AR, Evangelista RL, Nikolov ZL, Howard JA.1998a. Processing of transgenic corn seed and its effect on the recovery of recombinant ß-glucuronidase. Biotechnol Bioengr 60:44-52.

Kusnadi AR, Hood EE, Witcher DR, Howard JA, Nikolov ZN.1998b. Production and purification of two recombinant proteins from transgenic corn. Biotechnol Prog 14:149-155.

Ladiwala A, Rege K, Breneman CM, Cramer SM. 2005. A priori prediction of adsorption isotherm parameters and chromatographic behavior in ion-exchange systems. PNAS 102:11710 -11715.

Malmquist G, Nilsson UH, Norrman M, Skarp U, Stromgren M, Carredano E. 2006. Electrostatic calculations and quantitative protein retention models for ion exchange chromatography. J Chromatogr A 1115:164 - 186.

Mazza CB, Sukumar N, Breneman CM, Cramer SM. 2001. Prediction of protein retention in ion-exchange systems using molecular descriptors obtained from crystal structure. Anal Chem 73:5457-5461.

Menkhaus TJ, Bai Y, Zhang C, Nikolov ZL, Glatz CE. 2004. Considerations for the recovery of recombinant proteins from plants. Biotechnol Prog 20:1001-1014.

Pirondini A, Visioli G, Malcevschi A, Marmiroli N. 2006. A 2-D liquid-phase chromatography for proteomic analysis in plant tissues. J Chromatogr B 833:91-100.

Qin S, Ferdinand AS, Richie JP, O'Leary MP, Mok SC, Liu BCS. 2005. Chromatofocusing fractionation and two-dimensional difference gel electrophoresis for low abundance serum proteins. Proteomics 5:3183-3192.

Regnier FE. 1987. The role of protein structure in chromatographic behavior. Science 238:319-323.

Roth CM, Unger KK, Lenhoff AM. 1996. Mechanistic model of retention in protein ion-exchange chromatography. J Chromatogr A 726:45-56.

Stahlberg J, Jonsson B, Horvath C. 1991. Theory for electrostatic interaction chromatography of proteins. Anal Chem 63:1867-1874.

Stahlberg J, Jonsson B. 1996. Influence of charge regulation in electrostatic interaction chromatography of proteins. Anal Chem 68:1536-1544.

Stahlberg J. 1999. Review. Retention models for ions in chromatography. J Chromatogr A 855:3-55.

Stroink T, Ortiz MC, Bult A, Lingeman H, Jong GJ, Underberg WJM. 2005. On-line multidimensional liquid chromatography and capillary electrophoresis systems for peptides and proteins. J Chromatogr B 817:49-66.

Whitelam GC, Cockburn B, Gandecha AR, Owen MRL. 1993. Heterologous protein production in transgenic plants. Biotechnol Genet Eng Rev 11:1-29.

Witcher DR, Hood EE, Peterson D, Bailey M, Bond D, Kusnadi A, Evangelista R, Nikolov Z. 1998. Commercial production of α-glucuronidase (GUS): a model system for the production of proteins in plants. Mol Breed 4:301-312.

Xu W, Regnier FE. 1998. Protein–protein interactions on weak-cation-exchange sorbent surfaces during chromatographic separations. J Chromatogr A 828:357-64.

Yao Y, Lenhoff AM. 2004. Electrostatic contributions to protein retention in ion-exchange chromatography.1.cytochrome c variants. Anal Chem 76:6743-6752.

Zhong G, Peterson D, Delaney DE. 1999. Commercial production of aprotinin in transgenic maize seeds. Mol Breed 5:345-356.

Zhong Q, Gu Z, Glatz CE. 2006. Extraction of recombinant dog gastric lipase from transgenic corn. J Agric Food Chem 54:8086-8092.

Zhong Q, Xu L, Zhang C, Glatz CE. 2007. Purification of recombinant aprotinin from transgenic corn germ fraction using ion exchange and hydrophobic interaction chromatography. Appl Microbiol Biotechnol 76:607-613.

Protein purification handbook. 2001. Uppsala, Sweden:Amersham Biosciences.

CHAPTER 2.

PREDICTING PROTEIN RETENTION TIME IN ION-EXCHANGE CHROMATOGRAPHY BASED ON THREE-DIMENSIONAL PROTEIN CHARACTERIZATIONS

A manuscript for submission to Journal of Chromatography A

Li Xu and Charles E. Glatz

2.1 Abstract

Recovery and purification is the most costly part of recombinant protein production. A great amount of experimental work is generally needed for process development because of the lack of knowledge of the molecular properties of the host proteins as well as the means to use such knowledge to predict behavior. The most frequently used protein separation methods are based on the differences in charge, molecular weight and surface hydrophobicity among target and host proteins. Investigation of those protein properties for commonly used hosts and their influences on separation performance could establish a database for rational selection of purification strategies for recombinant proteins or could guide selection among alternative production hosts or expression target organs on the basis of simplifying protein purification.

The isoelectric point (pI), molecular weight (MW) and aqueous two-phase partitioning coefficients of a set of model proteins were related to retention time in cation-exchange chromatography (CEC) using partial least squares regression. A three-dimensional (3D) method which combined hydrophobic partitioning and two-dimensional (2D) electrophoresis was used to determine those three properties for a mixture of proteins. The resulting models fit well (R^2=0.913 and 0.873 for SP and 15S, respectively) considering the limited property basis and the regression models were able to predict results for a small test set of proteins. The models showed that pI and

MW correlated positively with CEC retention time, while the net influence of the partition coefficient depended on the base matrix type

2.2 Introduction

Ion-exchange chromatography (IEC) is one of the most frequently used techniques in the field of protein purification because of its high resolution, mild running conditions and relatively low cost. When considering protein separation by IEC, it is assumed the net charge is the major factor influencing protein binding and elution. However, protein surface charge [1, 2], charge asymmetry [3], and average potential over the molecular surface [4] are all distinct charge-related influences. In addition to charge distribution, Yao et al. [5] found that protein structural properties, such as size and shape may play significant roles in protein retention. Hallgren et al. [6] and Xu et al. [7] also found a combination of charge distribution and molecular shape to be responsible for elution behavior. Malmquist et al. [2] included eight protein surface hydrophobicity descriptors among 58 molecular descriptors found to influence IEC behavior. In general, protein behavior in IEC is not solely determined by its charge, but also by its shape, charge distribution, size and hydrophobic patches.

Models have been developed to correlate protein molecular properties and IEC retention. Mazza et al. [8] used quantitative structure retention relationship (QSRR) models of IEC behavior based on partial least square (PLS) approach to reduce several hundred molecular descriptors to 8 and 19 descriptors for two different resins. Ladiwala et al. [9] used quantitative structure-property relationship models to determine which protein molecular surface properties influenced IEC behavior; later, Malmquist et al. [2] correlated 58 protein molecular descriptors with protein behavior in IEC. The majority of those descriptors were related to the charge, the rest were three non-pH dependent size and shape descriptors, and eight surface hydrophobicity descriptors. The above studies were based on a set of known proteins, for which extensive information of protein molecular properties was available. For large numbers of unknown host proteins, such detailed information is not available.

Gu and Glatz [10] proposed a means of characterizing three properties related to IEC behavior for large numbers of unknown proteins that includes measures of pI, MW and hydrophobicity. In this study, (pH-pI) is used to represent protein net charge, MW for protein size and partitioning coefficient (K) in an aqueous two phase partitioning system is used to represent protein hydrophobicity. We then examine to what degree separation behavior prediction in IEC can be based on these three measures for a set of model proteins. PLS regression was applied to correlate those three properties to CEC retention.

2.3 Experimental

Material

Bovine hemoglobin, α-chymotrypsin, α-chymotrypsinogen A, bovine heart cytochrome c,chicken egg white lysozyme, hen egg white avidin, horse heart cytochrome c, rabbit muscle pyruvate kinase, bee phospholipase, bovine pancreatic ribonuclease A, and bovine pancreatic ribonuclease B were purchased from Sigma (St. Louis, MO). Sodium acetate and sodium chloride were purchased from Aldrich (Milwaukee, WI). PEG 3350, sodium sulfate, trichloroacetic acid (TCA), sodium dodecyl sulfate (SDS), hydroxymethyl aminomethane (Tris), urea, acetone, glacial acetic acid, and methanol were ACS-certified grade from Fischer Scientific (Pittsburgh, PA). Eleven cm Immobiline^{TM} Drystrip (pH 6-11), IPG buffer (pH 6-11) and Destreak Rehydration buffer were purchased from GE Healthcare (Piscataway, NJ). DL-dithiothreitol (DTT), iodoacetamide, and 4-20% (w/v) acrylamide gels were from Bio-Rad (Hercules, CA). Coomassie brilliant blue R-250 stain was purchased from Bio-Rad (Hercules, CA). The Coomassie Plus® protein assay reagent kit was from Pierce Biotechnology (Rockford, IL). Water was deionized.

Cation-exchange chromatography

Retention times were taken from reported values [11] for the set of proteins used here. Some of the proteases were eliminated from their set to avoid proteolysis when combined in the mixture used here. Their data were for gradient elution of single-protein samples at pH 5 from Fast Flow

Sepharose SP and Source 15S columns (1 mL) with a linear gradient slope of 20 mM sodium per column volume.

Aqueous two-phase partitioning

The aqueous two-phase partitioning (ATPP) used the procedure established by Gu and Glatz [10]. The final ATPP system overall composition (after sample addition) was 15.7% (wt %) polyethylene glycol 3350 (PEG 3350) (Sigma-Aldrich Corp., St. Louis, MO), 8.9% (wt %) sodium sulfate (Na_2SO_4) (Sigma-Aldrich Corp., St. Louis, MO), 3% (wt %) sodium chloride (NaCl) (Sigma-Aldrich Corp., St. Louis, MO). For single model protein, the protein loading ratio was 0.2 mg total protein/ g ATP system; for protein mixture, the protein loading ratio was 0.9 mg total protein/ g ATP system with equal amounts of each individual protein.

Partitionings were performed at pH 7 in 15 ml polypropylene centrifuge tubes with 3 g ATP system with gentle mixing s in an end to end shaker (Lab Industries Inc., Berkeley, CA) at room temperature (22 °C) for 1 hr, followed by phase separation by centrifugation at 1000 × g for 20 min (Sorvall RC5B Plus centrifuge, DuPont, Wilmington, DE).

After measuring the volumes of the top phase and bottom phase, the top phase was withdrawn with a 1 ml pipette and the bottom phase was collected by using a syringe through the bottom of the centrifuge tube. To avoid contamination of the phase samples, the region near the interface was left in the tube. The protein concentrations in samples from each phase were analyzed using the Coomassie Plus® protein assay reagent kit (Pierce, Rockford, IL).

2D electrophoresis

Proteins from the initial protein mixture and phase samples were isolated using the TCA precipitation, washing and redissolution procedures of Gu and Glatz [10]. Total protein concentrations were determined by Coomassie Plus® protein assay reagent kit.

The isoelectric focusing (IEF) procedure was modified from that of Gu and Glatz [10] to accommodate the more basic set of proteins used here. Eleven cm Immobiline™ Drystrip (pH 6-11)

was rehydrated in Destreak Rehydration buffer with added 0.5% IPG buffer (pH 6-11) for 12-16 hr which consisted of 8 M urea, 2% CHAPS, 50 mM DTT, 0.5% IPG buffer (pH 6-11), and trace bromophenol blue. Each sample with 100 µl of protein was applied in a localized region through an open-bottom loading cup (GE Healthcare, Piscataway, NJ) during IEF. IEF was performed in an Ettan™ IPGphor™ Isoelectric Focusing System (Amersham, Piscataway, NJ) for a total of 20,000 V-h.

The second dimension SDS-PAGE, staining, destaining, spot matching and quantitation procedures were also those used previously [10]. The normalized protein spot quantity of matched spots from protein mixture, top phase and bottom phase gels were used to calculate the concentration, mass balance and partition coefficient (K) of proteins. Combined with pI and MW obtained from 2D gels, the three molecular properties of the model proteins were characterized. MW was determined by interpolation with migration of Precision Plus Protein Unstained Standards (Bio-Rad, Hercules, CA) in each gel; pI was determined using the linearity of the pH gradient in the IEF strip. MW and pI of individual protein were assigned by PDQuest software (Bio-Rad, Hercules, CA) after the determination of MW and pI coordinates of the 2D gels.

Protein concentration assay

Concentration of single-protein solutions was measured at 280 nm, except for horse and bovine cytochrome c, which were assayed at 550 nm. The total concentration of the model protein mixture, the phase samples and the redissolved TCA precipitates were determined using the Coomassie Plus® protein assay reagent kit (Pierce, Rockford, IL) with BSA as standard. The ATP phases and extract buffer were employed in standard curves for corresponding samples to correct for interference from salt and PEG. All ATP partitioning experiments were replicated three times and replicate samples were assayed twice.

Data analysis

Partial least square (PLS) regression was used to model the relationship between the IEC retention time and the three measurable proteins molecular properties (pI, MW and partitioning coefficient). PLS is a statistical method which generalizes and combines features of principal component analysis and multiple linear regression. PLS regression aims to predict dependent variables based on some other measurable variables. In general, the PLS model can be expressed as:

$$y = a_0 + a_1 C_1 + a_2 C_2 + \ldots + a_i C_i \quad (1)$$

Where y is the dependent variable (protein retention time), C_i is the i^{th} PLS component, and a_i is the regression coefficient corresponding to the i^{th} PLS component. The first PLS component usually accounts for the greatest portion of the variance in the data. C_i can be expressed as:

$$C_i = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_j x_j \quad (2)$$

Where x_j is the j^{th} independent variable (each protein molecular property), and α_j is the j^{th} regression coefficient corresponding to the independent variable. When the maximum number of components is chosen to explain the variation, the PLS regression is equivalent to the multiple linear regression (MLR). In this study, the maximum number of components that can be estimated in the model is three. The numbers of the components in this study were chosen by leave-one-out cross validation using training data.

The PLS regression was used here for three reasons: (1) to account for multicollinearity in the dataset; (2) because there was not a well-understood relationship between the measured protein properties and the IEC retention time; and (3) to avoid over-fitting the data.

In order to validate the usefulness of the regression model, a number of proteins must be used as "blind test set". Those test set proteins were excluded from the training set, such that their data were not used in the PLS regression. It has been recommended that a minimum of 10% of the data be used for the blind test set [12], thus, two model proteins served to test the validity of the PLS regression model, which was based on 9 other model proteins. The two test proteins are proteases taken from model proteins in Ladiwala et al. [11]. These were chosen as test proteins because they would

degrade other proteins if used in training mixture; however, they did not autolyse during when measuring their single protein partitioning coefficient.

2.4 Results and Discussion

Three-dimensional characterization of model proteins

Table 1 shows the properties of the eleven model proteins. The pI and MW values are from the work of others, while the partition coefficients (K) are those obtained from single protein partitioning experiments in this work.

Table 1. Properties of the model proteins [a]

Protein	pI [b]	MW [b]	Log K [c]
Bovine hemoglobin	7.0 [13]	64.4 [13]	-1.04 ± 0.06
Pyruvate kinase	7.4 [14]	237.0 [15]	-0.46 ± 0.04
Bee phospholipase	9.5 [16]	14.5 [17]	-0.94 ± 0.02
Ribonuclease B	9.6 [18]	15.0 [18, 19]	-1.38 ± 0.08
Ribonuclease A	9.6 [20]	13.7 [19]	-1.16 ± 0.11
Horse cytochrome c	10.0 [21]	12.4 [22]	-1.47± 0.14
Bovine cytochrome c	10.0 [21]	12.6 [23]	-1.39± 0.19
Avidin	10.0 [24]	66.0 [25]	-1.72 ± 0.12
Chicken lysozyme	11.3 [26]	14.3 [27]	1.32 ± 0.13
α-chymotrypsinogen A	9.0 [28]	25.6 [29]	0.82 ± 0.02
α-chymotrypsin	8.8 [30]	25.0 [31]	-0.36 ± 0.02

[a] Proteins in boldface type were used as the test set.

[b] pI and MW were obtained from the sources cited with each entry. MW had been calculated from amino acid sequences.

[c] Log K obtained via single protein partitioning experiments. The ± represents the 95% confidence intervals (3 replicates).

Figs. 1a-c show the 2D gels from the initial model protein mixture and the two equilibrium phases after partitioning. The numbers identifying spots in those figures correspond to the entries in

Table 2, which displays identity, properties and mass balance of protein spots as determined by the 3D method. Protein identification was established by matching these spots with the 2D gels of the individual proteins (not shown here).

Several proteins appeared as multiple spots on the 2D gels (Figs 1a-c) as a result of four different causes. Variable phosphorylation results in the same apparent MW but different pI [32] and apparently little difference in K. All such forms were all included in the calculation of protein properties in Table 2. Examples are spots 1a-1c for the β subunit and spots 1d-1f for the α subunit of bovine hemoglobin, spots 2a-2c for pyruvate kinase, spots 3f and 3h for bee phopholipase, spots 5a-5b for ribonuclease A, spots 4a-4b for ribonuclease B and spots 8a-8d for avidin. Heterogeneous multimers appear as the individual subunits with all three properties reflecting the nature of the subunits. Here this occurs with α and β subunits of bovine hemoglobin and again an averaging of the separate properties are used to estimate the MW and pI of the intact multimer. Because the multimeric form was present at the partitioning step (confirmed by size-exclusion chromatography (SEC) of the solutions partitioned), both subunit K's should be the same. Impurities in the source material result in spots with different pI, MW and Log K. In this case the minor spots have been ignored since the reported retention time [11] was determined by the major peak (S. Cramer, personal communication). Examples here are spots 3a-3e and 3g for bee phospholipase and with only spots 3f and 3h being used for property estimations. The last situation is lysozyme (9a) which dimerizes (9b) giving two spots showing similar pI and Log K but two-fold difference in MW. In this case only the monomer properties were used.

The "estimated from 2D gels" column of Table 2 is based on the weighted average of the retained spot values using Eqns. 3-5.

$$\text{Estimated } MW = \frac{\sum_i mass_i \times MW_i}{\sum_i mass_i} \qquad (3)$$

$$\text{Estimated } pI = \frac{\sum_i mass_i \times pI_i}{\sum_i mass_i} \quad (4)$$

$$\text{Estimated } Log\ K = \frac{\sum_i mass_i^t}{\sum_i mass_i^b} \quad (5)$$

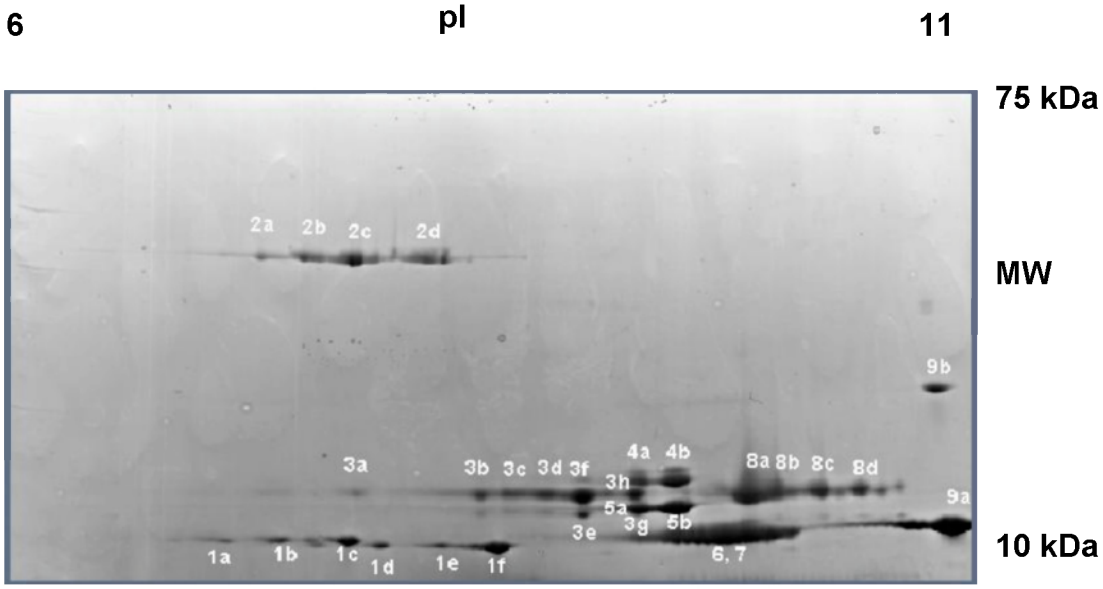
where *i* is the number of the retained spots for the specific protein as shown in Table 2, *t* and *b* refer to top phase and bottom phases, respectively. The $mass_i$ is that obtained from the normalized spot volumes from image analysis. For $mass_i^{t,b}$ the mass is that in the partitioning experiment obtained by correcting the normalized spot volumes from image analysis for the appropriate dilutions/concentrations occurring during transfer to the gels.

For the multimers, bovine hemoglobin ($\alpha_2\beta_2$), pyruvate kinase and avidin (α_4) are tetramers [15, 33, 34] with estimated MW calculated by Eqns. 6 (α_4) and 7 ($\alpha_2\beta_2$) as appropriate.

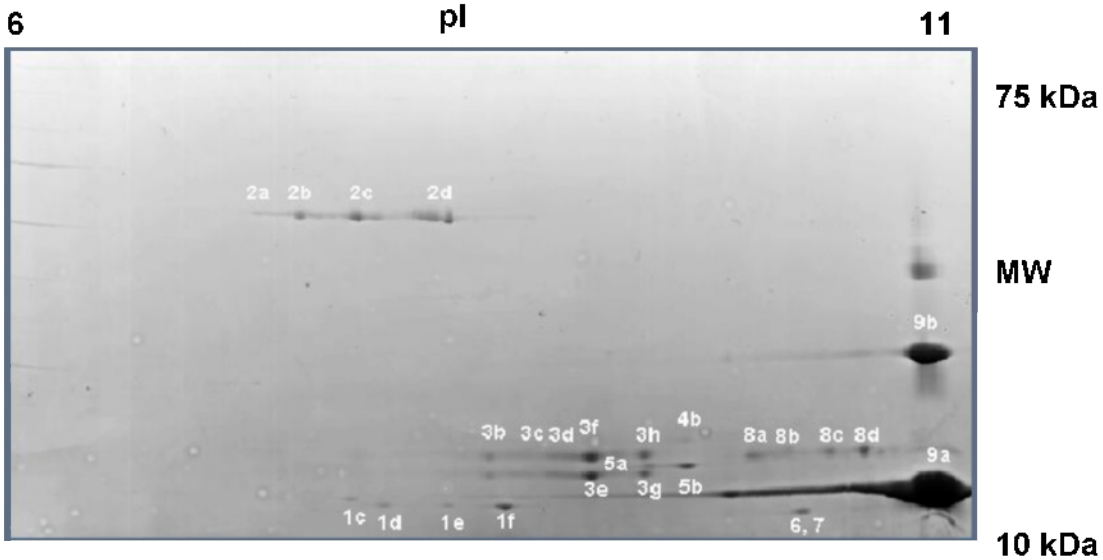
$$\text{Estimated } MW = \frac{\sum_i mass_i \times MW_i}{\sum_i mass_i} \times 4 \quad (6)$$

$$\text{Estimated } MW = \frac{\sum_{\alpha i} mass_{\alpha i} \times MW_{\alpha i}}{\sum_{\alpha i} mass_{\alpha i}} \times 2 + \frac{\sum_{\beta i} mass_{\beta i} \times MW_{\beta i}}{\sum_{\beta i} mass_{\beta i}} \times 2 \quad (7)$$

(a)


6
pI
11
75 kDa
MW
10 kDa
2a 2b 2c 2d
9b
3a
3b 3c 3d 3f
4a 4b
3h
5a
3e
3g 5b
8a 8b 8c 8d
9a
1a 1b 1c 1d 1e 1f
6, 7

(b)


6
pI
11
75 kDa
MW
10 kDa
2a 2b 2c 2d
9b
3b 3c 3d 3f
3h
4b
5a
3e
3g 5b
8a 8b 8c 8d
9a
1c 1d 1e 1f
6, 7

(c)

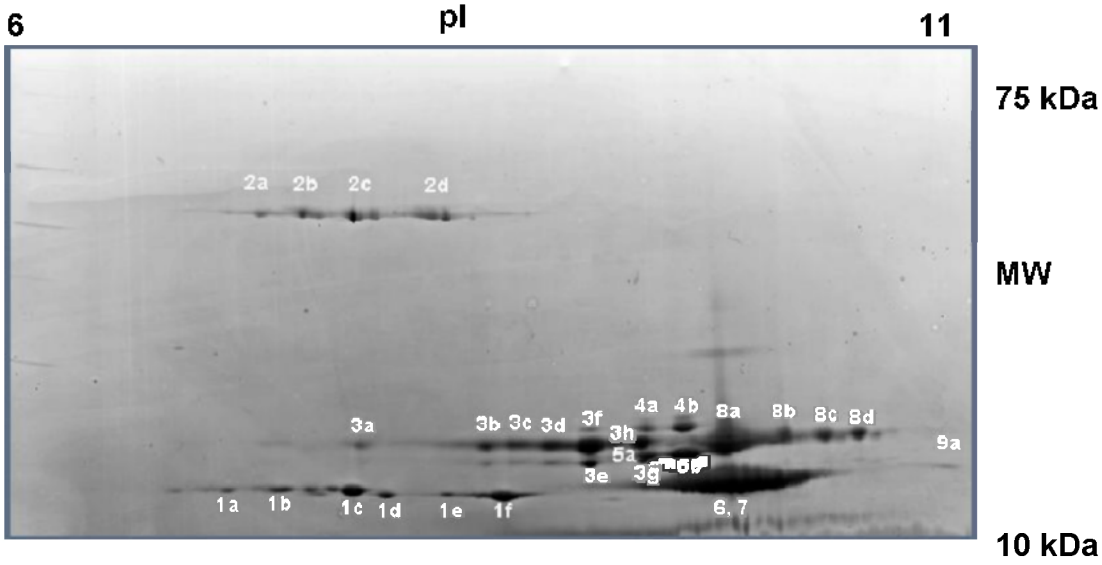


Figure 1. 2D gels of model proteins. (a) Initial protein mixture; (b) Equilibrium top phase; (c) Equilibrium bottom phase.

To assess the validation of the 3D characterization method, the pI, MW and Log K data from Table 2 were compared with those in Table 1. The results show no appreciable difference between the estimated pI and MW values obtained by 2D electrophoresis experiments from those in Table 1. The difference between the pI and MW values obtained from the 3D method and those from other sources in Table 2 are in the range of ±15% except the MW of bovine hemoglobin (30% less than MW calculated from amino acid sequence). The greater difference for hemoglobin MW might be because the tightly bound heme group constrained the extent of unfolding, which would lead to it running as if having smaller MW [35]. The estimated Log K values from the 3D method also agreed with those obtained from single partitioning experiments. The small difference might indicate the discrepancy between total protein assays and image analysis of the 2D gels, but also could result from the former including protein impurities that were excluded in the latter determination.

Table 2. Characterization of the three protein properties by multiple protein partitioning experiments and 2D gel electrophoresis with spot matching, and calculation of mass balance for individual protein spots from the 2D gels using image analysis.

Protein [a]		from 2D gels		Estimated from 2D gels [b]			MW from SEC-HPLC (kDa) [c]	Mass balance from 2D gels [d] (%)
		pI	MW (kDa)	pI	MW (kDa)	Log K		
Bovine hemoglobin [c]		-	-	7.9	45.1	-1.03	61.4	-
Valid Spots for subunit β	1a	7.0	11.6	-	-	-	-	98
	1b	7.2	11.6	-	-	-	-	96
	1c	7.6	11.4	-	-	-	-	110
Valid Spots for subunit α	1d	7.8	11.2	-	-	-	-	105
	1e	8.1	11.2	-	-	-	-	96
	1f	8.5	11.0	-	-	-	-	91
Pyruvate kinase [c]		-	-	7.7	204.0	-0.60	200.9	-
Valid Spots	2a	7.1	51.8	-	-	-	-	88
	2b	7.4	51.3	-	-	-	-	94
	2c	7.6	50.6	-	-	-	-	95
	2d	8.1	51.1	-	-	-	-	101
Bee phospholipase		-	-	9.1	15.2	-0.80	-	-
Impurity	3a	7.6	15.5	-	-	-	-	110
	3b	8.4	15.3	-	-	-	-	95
	3c	8.6	15.4	-	-	-	-	96
	3d	8.8	15.4	-	-	-	-	88
	3e	9.0	13.8	-	-	-	-	112
Valid Spots	3f	9.0	15.2	-	-	-	-	98
Impurity	3g	9.2	13.8	-	-	-	-	109
Valid Spots	3h	9.2	15.3	-	-	-	-	115
Ribonuclease B		-	-	9.4	16.3	-1.90	-	-
Valid Spots	4a	9.3	16.3	-	-	-	-	85
	4b	9.5	16.3	-	-	-	-	86

Table 2. (continued)

Protein [a]		from 2D gels		Estimated from 2D gels [b]			MW from SEC-HPLC (kDa) [c]	Mass balance from 2D gels [d] (%)
		pI	MW (kDa)	pI	MW (kDa)	Log K		
Ribonuclease A		-	-	9.4	14.4	-1.23	-	86
Valid Spots	5a	9.3	14.3	-	-	-	-	87
	5b	9.5	14.4	-	-	-	-	85
Horse cytochrome c		-	-	9.7	12.1	-1.98	-	115
Bovine cytochrome c		-	-	9.7	12.1	-1.98	-	115
Avidin [c]		-	-	10.1	62.4	-1.44	60.4	-
Valid Spots	8a	9.9	15.6	-	-	-	-	113
	8b	10.2	15.6	-	-	-	-	98
	8c	10.4	15.6	-	-	-	-	90
	8d	10.6	15.6	-	-	-	-	90
Chicken lysozyme		-	-	11.0	12.7	1.74	-	-
Valid Spots	9a	11.0	12.7	-	-	-	-	115
Dimer	9b	11.0	25.8	-	-	-	-	113

[a] Protein identification was established by comparison with 2D gels of the individual proteins (those gels not shown). Multiple spots were due to variable phosphorylation, dimerization or impurities.

[b] pI and MW of the intact multimers were estimated using Eqns. 3-7.

[c] Multimeric proteins shown by SEC-HPLC to be intact for the partitioning experiments.

[d] Mass balance accounted for phase volumes and dilutions during analysis.

All the data are the average of three replicate experiments. The average 95% confidence intervals of the estimated pI, MW and Log K are 0.06, 2.38 and 0.10, respectively (three replicates).

In addition to the three characterized protein properties, mass balance of individual spots on the 2D gels from initial mixture and the two equilibrium phases were also checked to further evaluate quantitative validity. Summation of spot masses from the gels of the two phases agreed within ±15% of those calculated from the initial solution (Table 2). Indirect justification of the partitioning coefficients was also accomplished by comparing Log Ks for total protein determined from the total

spot masses on the equilibrium top phase and bottom phase gels with the value calculated from total protein assay of the top and bottom phases (Table 3). Mass balances were better here than reported for the initial use of the method on a more complex extract [10]. The lower complexity likely helped, but the change to cup loading of samples to the IEF strip also decreased sample loss from the original rehydration method.

However, the successful characterization of the three properties for this model protein mixture did depend on knowing where subunits appeared. Thus, unknown multimers in any mixture being characterized will lead to mischaracterization. This limitation constrains the goal of developing purification strategies using those characterized properties to one of improving the likelihood of making good choices.

Table 3. Comparison of determination of total mixture protein partition coefficient via 2D gel spot matching with that from total protein concentration assay of top and bottom phase

Method		
2D Gel [a]	Log K	-0.67 ± 0.03 [b]
	Mass balance (%)	104 ± 9
Total protein assay	Log K	-0.66 ± 0.02 [b]
	Mass balance (%)	98 ± 5

[a] Calculated from the summation of the individual spot masses.

[b] 95% confidence intervals (three replicates).

PLS regression modeling

Table 4 summarizes the estimated pI, MW and Log K from Table 2 that were used to investigate the correlation between the characterized properties and CEC retention times on FF Sepharose SP and Source 15S resins.

Table 4. Protein retention times and their corresponding properties used for PLS regression

Protein	Protein retention time (min) on FF Sepharose SP [a]	Protein retention time (min) on Source 15S [a]	pI	MW (kDa)	Log K
Bovine hemoglobin	31.50	26.11	7.9	45.1	-1.03
Pyruvate kinase	32.61	27.41	7.7	204.0	-0.60
Bee phospholipase	34.50	33.48	9.1	15.2	-0.80
Ribonuclease B	36.14	29.33	9.4	16.3	-1.90
Ribonuclease A	39.40	32.08	9.4	14.4	-1.23
Horse cytochrome c	44.15	33.65	9.7	12.1	-1.98
Bovine cytochrome c	45.22	33.14	9.7	12.1	-1.98
Chicken lysozyme	47.34	39.70	11.0	12.7	1.74
Avidin	53.12	43.09	10.1	62.4	-1.44

[a] Values were provided by Ladiwala et al. [11].

Figs. 2a-c and Figs. 3a-c show PLS models for protein retention on FF Sepharose SP resin and Source 15S resin, respectively. The plots demonstrate PLS-correlated protein retention data for the training set and predicted retention for the test set versus experimental retention data based on the three protein properties measured by 3D method in Table 4. For the regression, pI-pH was used to represent protein net charge, and MW was normalized to Log MW in order to provide comparable scaling to the other two factors. The best cross-validated correlation coefficients, R^2, were generated when including all three protein molecular properties during the regression. For both resins, only two PLS components were needed to explain the variation in the data.

These improvements in the correlation are seen as each variable is added to the correlation. Single property regression for Log MW or Log K generated correlation coefficients less than 0.1, while pI-pH alone exhibited correlation coefficients R^2 of 0.699 and 0.705 for FF Sepharose SP resin and Source 15S resin, respectively. This demonstrates that protein charge (pI-pH) was more important

than the other two properties for the determination of protein IEC retention. However, the relatively low R^2 value indicates that this measure is not sufficient to correlate IEC behavior. When Log MW was added to pI-pH, an additional 17% and 23% of variability was accounted for, and when Log K was also included, the R^2 value was increased to 0.913 and 0.873 for FF Sepharose SP and Source 15S, respectively (Figs. 2a-c and 3a-c). Those results are comparable to the regression result reported by Ladiwala et al. (R^2 = 0.937 and 0.959 for FF Sepharose SP and Source 15S, respectively.) using a larger number of factors [11]. The regression model was also checked by comparing the predicted retention times of the test proteins with their experimental values (Figs. 2a-c and 3a-c). The MLR model based on the three properties (not shown) gave very similar results to the cross-validated PLS model using two components.

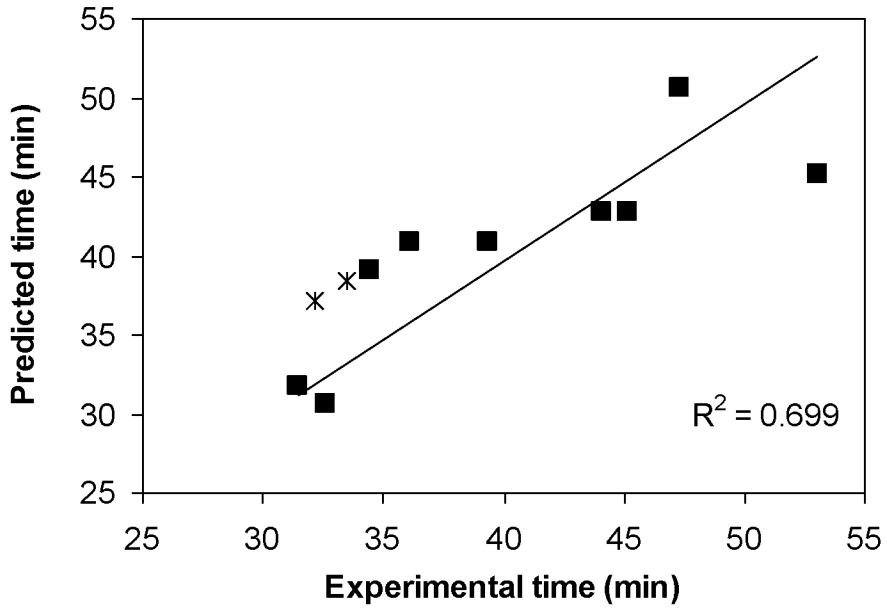
The PLS models recast in terms of the protein properties are shown in Eqns. 8 and 9.

$$t_{R,SP} = -16.74 + 9.45 \times (pI - pH) + 10.00 \times Log\ MW - 2.20 \times Log\ K \quad (8)$$

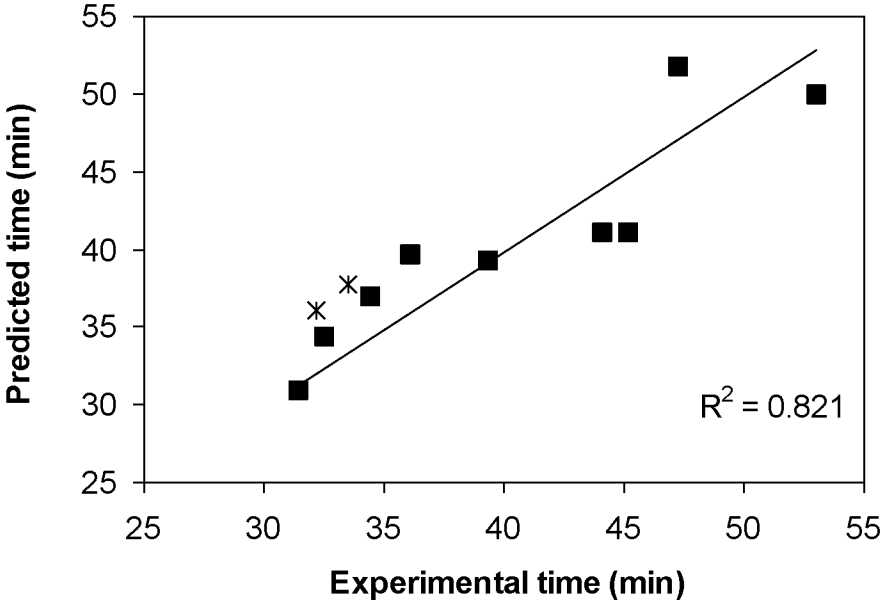
$$t_{R,15S} = -1.43 + 5.91 \times (pI - pH) + 6.74 \times Log\ MW + 0.49 \times Log\ K \quad (9)$$

The importance of the charge measure (pI-pH) is expected for IEC. The impact of protein MW may also indirectly reflect a charge influence as larger proteins are more likely to have higher net charge for the same (pI – pH) and could more easily accommodate favorable charge distributions. Since FF Sepharose SP resin is hydrophilic, some negative influence of hydrophobicity is reasonable; for the relatively hydrophobic Source 15S resin shows positive hydrophobicity slightly increases IEC retention time.

(a)


Predicted time (min)
Experimental time (min)
R^2 = 0.699
25
30
35
40
45
50
55

(b)


Predicted time (min)
Experimental time (min)
R^2 = 0.821
25
30
35
40
45
50
55

(c)

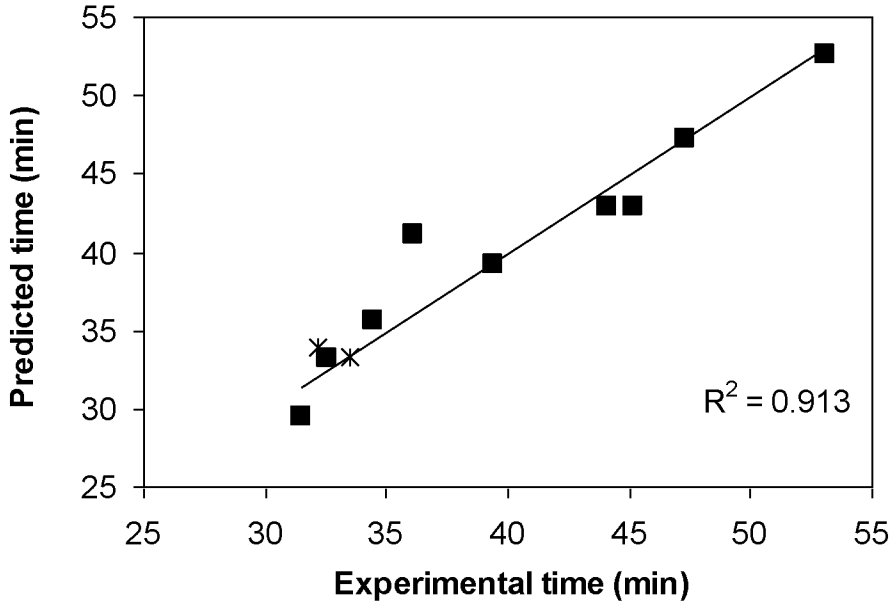
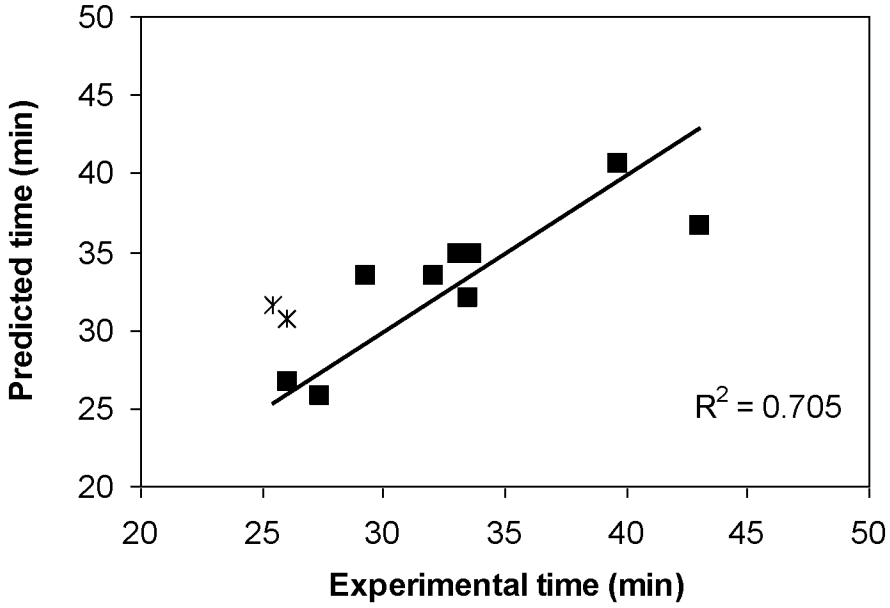
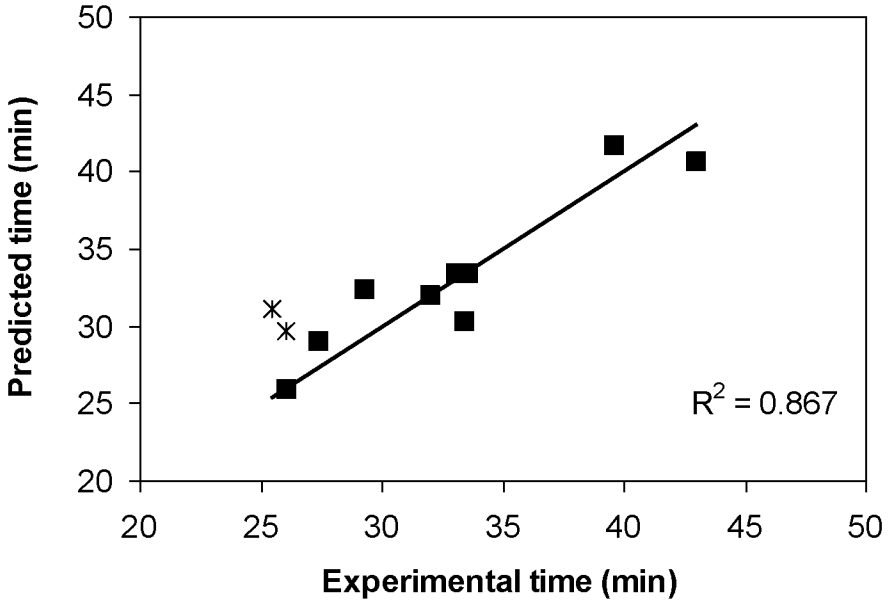


Figure 2. Plots of PLS-predicted protein retention time versus experimental retention time on FF Sepharose SP resin. (a) PLS-prediction based on protein pI-pH only. (b) PLS-prediction based on protein pI-pH and Log MW. (c) PLS-prediction based on protein pI, Log MW and Log K. All the PLS models were generated based on the training set proteins; test set proteins not included in the reported R^2.

(a)


Predicted time (min)
Experimental time (min)
R^2 = 0.705
50
45
40
35
30
25
20
20
25
30
35
40
45
50

(b)


Predicted time (min)
Experimental time (min)
R^2 = 0.867
50
45
40
35
30
25
20
20
25
30
35
40
45
50

(c)

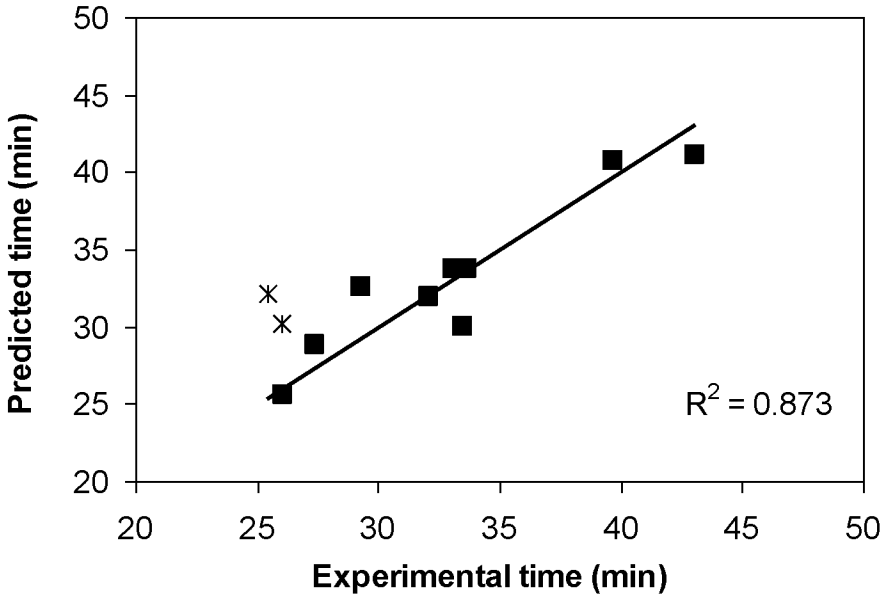


Figure 3. Plots of PLS-predicted protein retention time versus experimental retention time on Source 15S resin. (a) PLS-prediction based on protein pI-pH only. (b) PLS-prediction based on protein pI-pH and Log MW. (c) PLS-prediction based on protein pI, Log MW and Log K. All the PLS models were generated based on the training set proteins; test set proteins not included in the reported R^2.

2.5 Conclusion

That protein retention in IEC likely depends on protein charge, size and surface hydrophobicity has been established previously. This work has shown that pI, MW, and ATP partition coefficients can be simultaneously measured for individual proteins in a mixture of nine proteins. These three properties were able to correlate IE chromatographic retention time, accounting for variability to a degree comparable to previous models based on larger numbers of properties derived from protein crystal structures. The PLS regression models obtained showed that greater MW and (pI – pH) increased retention on both resins while the net influence of Log K depended on the base matrix type. Application to two test proteins provided tentative validation for the predictive power of the model.

2.6 Acknowledgment

We thank Prof. Steven Cramer and his students at RPI for providing the retention values and source information for the model proteins used in this work and Prof. Dan Nettleton of the ISU Statistics Department for guidance using regression models. This work was sponsored by the USDA CREES Grant # 2005-34496-16003 and Grant # 2006-34496-17122.

2.7 References

[1] F.E. Regnier, The role of protein structure in chromatographic behavior. Science 238 (1987) 319-323.

[2] G. Malmquist, U.H. Nilsson, M. Norrman, U. Skarp, M. Stromgren, E. Carredano, Electrostatic calculations and quantitative protein retention models for ion exchange chromatography. J. Chromatogr. A 1115 (2006) 164-186.

[3] W. Kopaciewicz, M.A. Rounds, J. Fausnaugh, F.E. Regnier, Retention model for high-performance ion-exchange chromatography. J. Chromatogr. 266 (1983) 3-21.

[4] L. Haggerty, A.M. Lenhoff , Relation of protein electrostatics computations to ion-exchange and electrophoretic behavior. Phys. Chem. 95 (1991) 1472-1477.

[5] Y. Yao, A.M. Lenhoff, Electrostatic contributions to protein retention in ion-exchange chromatography.1.Cytochrome c variants. Anal. Chem. 76 (2004) 6743-6752.

[6] E. Hallgren, F. Kalman , D. Farnan, C. Horvath, J. Stahlberg, Protein retention in ion-exchange chromatography: effect of net charge and charge distribution. J. Chromatogr. A 877 (2000) 13-24.

[7] W. Xu, F.E. Regnier, Protein–protein interactions on weak-cation-exchange sorbent surfaces during chromatographic separations. J. Chromatogr. A 828 (1998) 357-364.

[8] C.B. Mazza, N. Sukumar, C.M. Breneman, S.M. Cramer, Prediction of protein retention in ion-exchange systems using molecular descriptors obtained from crystal structure. Anal. Chem. 73 (2001) 5457-5461.

[9] A. Ladiwala, K. Rege, C.M. Breneman, S.M. Cramer, A priori prediction of adsorption isotherm parameters and chromatographic behavior in ion-exchange systems. PNAS 102 (2005) 11710-11715.

[10] Z. Gu, C.E. Glatz, A method for three-dimensional protein characterization and its application to a complex plant (corn) extract. Biotechnol. Bioengr. 97 (2007) 1158-1169.

[11] A. Ladiwala, K. Rege, C.M. Breneman, S.M. Cramer, Investigation of mobile phase salt type effects on protein retention and selectivity in cation-exchange systems using quantitative structure retention relationship models. Langmuir 19 (2003) 8443-8454.

[12] C.B. Mazza, C.E. Whitehead, C.M. Brenernan, S.M. Cramer, Predictive quantitative structure retention relationship models for ion-exchange chromatography. Chromatographia 56 (2002) 147-152.

[13] H. Sakai, Y. Masada, S. Takeoka, E. Tsuchida, Characteristics of bovine hemoglobin as a potential source of hemoglobin-vesicles for an artificial oxygen carrier. J. Biochem. 131 (2002) 611-617.

[14] N. Muroya, Y. Nagao, K. Miyazaki, K. Nishikawa, T. Horio, Pyruvate kinase isozymes in various tissues of rat, and increase of spleen-type pyruvate kinase in liver by injecting chromatins from spleen and tumor. J. Biochem. 79 (1976) 203-215.

[15] F.J. Kayne, in The Enzymes, 3rd ed., vol. 8, Group transfer, pt. A: Nucleotidyl transfer, nucleosidyl transfer, acyl transfer, phosphoryl transfer. P.D. Boyer (Ed.), Academic Press, New York, 1973.

[16] G.H. De Haas, Purification and properties of phospholipase A from porcine pancreas. Biochim. Biophys. Acta. 159(1968) 103-117.

[17] G.H. De Haas, Studies on phospholipase A and its zymogen from porcine pancreas. II. The assignment of the position of the six disulfide bridges. Biochim. Biophys. Acta, 221(1970) 54-61.

[18] T. Yang, G. Malmquist, B.L. Johansson, J.L. Maloisel, S. Cramer, Evaluation of multi-modal high salt binding ion exchange materials. J. Chromatogr. A 1157 (2007) 171-177.

[19] D.G. Smyth, W.H. Stein, S. Moore, The sequence of amino acid residues in bovine pancreatic ribonuclease: revisions and confirmations. J. Biol.Chem. 238 (1963) 227-234.

[20] C. Tanford, J.D.Hauenstein, Hydrogen ion equilibria of ribonuclease. J. Am. Chem. Soc. 78 (1956) 5287-5291.

[21] L. Malmgren, Y. Olsson, T. Olsson, K. Kristensson, Uptake and retrograde axonal transport of various exogenous macromolecules in normal and crushed hypoglossal nerves. Brain Research 153 (1978) 477-493.

[22] E. Margoliash, E.L. Smith, G. Kreil, H. Tuppy, Amino-acid sequence of horse heart cytochrome *c*: The complete amino-acid sequence. Nature 192 (1961) 1125-1127.

[23] T. Nakashima, H. Higa, H. Matsubara, A.M. Benson, K.T. Yasunobu, The amino acid sequence of bovine heart cytochrome c. J. Biol.Chem. 241 (1966) 1166-1177.

[24] D.W. Woolley, H.B. White, Isolation of an antibiotin factor from egg white. J. Biol. Chem. 142 (1942) 285-290.

[25] M.J. O'Neil, A. Smith, P.E. Heckelman, J.R. Obenchain, J.A.R. Gallipeau,M.A. D'Arecca, S. Budavari (Ed.), Merck Index—An Encyclopedia of Chemicals,Drugs and Biologicals, 13th ed., Merck & Co., Rahway, NJ, 2001.

[26] L.R. Wetter, H.F. Deutsch, Immunological studies on egg white proteins. IV. Immunochemical and physical studies of lysozyme. J. Biol. Chem. 192 (1951) 237-242.

[27] R.E. Canfield, The amino acid sequence of egg white lysozyme. J. Biol. Chem. 238 (1963) 2698-2707.

[28] J.A. Beeley, S.M. Stevenson, J.G. Beeley, Polyacrylamide gel isoelectric focusing of proteins: determination of isoelectric points using an antimony electrode. Biochim. Biophys. Acta. 285(1972) 293-300.

[29] P.E. Wilcox, Chymotrypsinogens-chymotrypsins. Meth. Enzymol. 19 (1970) 64-108.

[30] N. Ui, Isoelectric points and conformation of proteins. II. Isoelectric focusing of α-chymotrypsin and its inactive derivative. Biochim. Biophys. Acta. 229 (1971) 582-589.

[31] M.M. Burrell (Ed.), Enzymes of Molecular Biology, vol. 16, Humana Press, Totowa, NJ, 1993.

[32] S. Siwkoa, D. Mochly-Rosenb, Use of a novel method to find substrates of protein kinase C delta identifies M2 pyruvate kinase. Int. J. Biochem. Cell Biol. 39 (2007) 978-987.

[33] T. Yamaguchi, K. Adachi1, Hemoglobin equilibrium analysis by the multiangle laser light-scattering method. Biochemical and Biophysical Research Communications 290 (2002) 1382-1387.

[34] R.J. DeLange, T.S. Huang, Egg white avidin. III. Sequence of the 78-residue middle cyanogens bromide peptide. Complete amino acid sequence of the protein subunit. J. Biol. Chem. 246 (1971) 698-709.

[35] M. Sallantin, J.C. Huet, C. Demarteau, J.C. Pernollet, Reassessment of commercially available molecular weight standards for peptide sodium dodecyl sulfatepolyacrylamide gel electrophoresis using electroblotting and microsequencing. Electrophoresis 11 (1990) 34-36.

CHAPTER 3.

GENERAL CONCLUSIONS

3.1 General Conclusion

The relationship between protein properties and protein selectivity in ion-exchange chromatography (IEC) was investigated. IEC is one of the most frequently used techniques in protein purification; however, a great amount of experimental work is generally needed for IEC process development because of the lack of knowledge of the molecular properties of the host proteins as well as the means to use such knowledge to predict behavior. Studies found that IEC usually operated based on the differences in surface hydrophobicity, charge and size among target and host proteins [1-5], so exploit the relationship between those protein molecular properties and IE chromatographic retention time can be useful in designing more efficient separations. Previous models to correlate protein molecular properties with IE chromatographic retention time were all based on sets of known proteins, for which extensive information of protein molecular properties was available. For large numbers of host proteins, it is impractical to investigate detailed protein molecular properties of individual host protein for solely separation purpose.

In the present study, protein pI, MW, and ATP partition coefficients were simultaneously measured for a mixture of nine proteins using a three-dimensional (3D) method which combined hydrophobic partitioning with two-dimensional (2D) electrophoresis. Statistical models were established that correlates protein IE chromatographic retention time with these three protein properties using PLS regression. The models had correlation coefficients of 0.913 and 0.873 for FF Sepharose SP and Source 15S, respectively, which were comparable to those models based on much larger numbers of properties obtained from protein crystal structures. The predictive power of the models were evaluated using two test proteins not included in the model training set, and the predictive retention data showed good agreement with experimental data. The PLS regression models

obtained showed that greater MW and (pI – pH) increased cation-exchange chromatographic retention time for both resins, while the net influence of Log K depended on the base matrix type. The established PLS model will provide a tool for (1) investigating of those protein properties for commonly used hosts and their influences on IEC; (2) developing rational selection of purification strategies for recombinant proteins expressed in a particular host; (3) selecting among alternative production hosts or expression target organs on the basis of simplified protein purification.

However, the successful characterization of the three properties for this model protein mixture did depend on knowing where subunits appeared. Thus, unknown multimers in any mixture being characterized will lead to mischaracterization. As a recommendation for the future application, this limitation could be partly solved using size exclusion chromatography (SEC) combined with 2D electrophoresis. SEC fractions with high MW values could be analyzed using 2D electrophoresis. By matching and comparing those SEC fraction gels with the 2D gel from original mixture, multimers could be identified and their MW values could be estimated.

The statistical model building needs a reliable characterization of all the three protein properties. In order to maintain the accuracy of the 3D method when apply it to more complicated host protein systems in the future, larger gels with 18cm or 24 cm strips and 26 × 20 cm gels or more highly sensitive fluorescence staining could be used to improve individual spot matching and quantification.

3.2 Reference

[1] F.E. Regnier, The role of protein structure in chromatographic behavior. Science 238 (1987) 319-323.

[2] G. Malmquist, U.H. Nilsson, M. Norrman, U. Skarp, M. Stromgren, E. Carredano, Electrostatic calculations and quantitative protein retention models for ion exchange chromatography. J. Chromatogr. A 1115 (2006) 164-186.

[3] W. Kopaciewicz, M.A. Rounds, J. Fausnaugh, F.E. Regnier, Retention model for high-performance ion-exchange chromatography. J. Chromatogr. 266 (1983) 3-21.

[4] L. Haggerty, A.M. Lenhoff , Relation of protein electrostatics computations to ion-exchange and electrophoretic behavior. Phys. Chem. 95 (1991) 1472-1477.

[5] Y. Yao, A.M. Lenhoff, Electrostatic contributions to protein retention in ion-exchange chromatography.1.Cytochrome c variants. Anal. Chem. 76 (2004) 6743-6752.

APPENDIX:

ADDITIONAL TABLE AND FIGURE

Table 1. Protein properties from the 3D characterization method including calculated Log K values of individual spots, shown as averages in Chap. 2, Table 2.

Protein		pI	MW (kDa)	Log K
Bovine hemoglobin		-	-	-
Valid Spots for subunit β	1a	7.0	11.6	-1.52
	1b	7.2	11.6	-1.77
	1c	7.6	11.4	-1.51
Valid Spots for subunit α	1d	7.8	11.2	-0.83
	1e	8.1	11.2	-0.91
	1f	8.5	11.0	-0.76
Pyruvate kinase		-	-	-
Valid Spots	2a	7.1	51.8	-0.83
	2b	7.4	51.3	-0.69
	2c	7.6	50.6	-0.56
	2d	8.1	51.1	-0.56
Bee phospholipase		-	-	-
Impurity	3a	7.6	15.5	-1.99
	3b	8.4	15.3	-0.71
	3c	8.6	15.4	-1.21
	3d	8.8	15.4	-1.56
	3e	9.0	13.8	-0.52
Valid Spots	3f	9.0	15.2	-0.76
Impurity	3g	9.2	13.8	-0.39
Valid Spots	3h	9.2	15.3	-0.90
Ribonuclease B		-	-	-
Valid Spots	4a	9.3	16.3	-1.89
	4b	9.5	16.3	-1.90

Table 1. (continued)

Protein		pI	MW (kDa)	Log K
Ribonuclease A		-	-	-
Valid Spots	5a	9.3	14.3	-1.28
	5b	9.5	14.4	-1.22
Horse cytochrome c		9.7	12.1	-1.98
Bovine cytochrome c		9.7	12.1	-1.98
Avidin		-	-	-
Valid Spots	8a	9.9	15.6	-1.50
	8b	10.2	15.6	-1.69
	8c	10.4	15.6	-1.24
	8d	10.6	15.6	-1.18
Chicken lysozyme		-	-	-
Valid Spots	9a	11.0	12.7	1.74
Dimer	9b	11.0	25.8	1.78

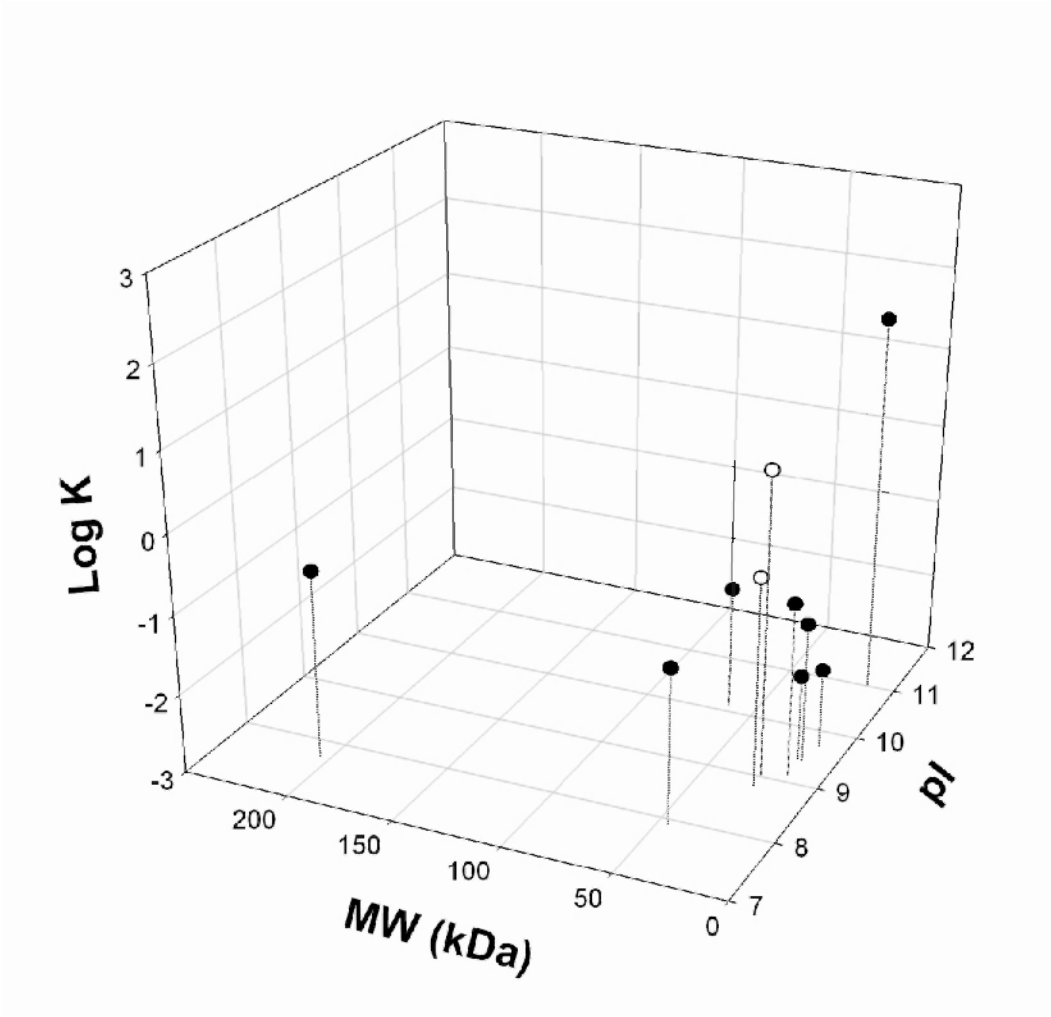


Figure 1. 3D Plot of the protein properties of the model proteins used in this study. Black balls represent proteins in training set; white balls represent proteins in test set.