Application of Multi-wavelength Fluorometry to Monitoring Protein Ultrafiltration

By

Rand Elshereef

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Chemical Engineering

Waterloo, Ontario, Canada, 2009

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

ABSTRACT

Membrane filtration of protein solutions is influenced by a wide range of processing and physicochemical conditions. Monitoring and optimizing membrane filtration may have advantages for achieving, in a cost effective manner, improved bioproduct purification and membrane performance which is relevant to pharmaceutical and biochemical applications. The motivation of this work was to examine the feasibility of applying two-dimensional fluorescence spectroscopy in conjunction with chemometric techniques for monitoring and possibly optimizing the performance of membrane processes.

Preliminary work focused on assessing the use of multivariate calibration tools in conjunction with the sensitivity of intrinsic protein fluorescence towards changes in environmental conditions was to predict protein concentration and aggregation behavior. A model protein, β-lactoglobulin (β-LG), was used as a first simple case scenario. Results showed very good agreement between the fluorescence based predictions and measurements obtained by HPLC and gravimetric analysis regardless of the conditions. PLS analysis of excitation-emission matrices revealed unique spectral fingerprints that are most likely associated with the heat-induced denaturation and aggregation. Standard Normal Variate, a signal preprocessing and filtering tool, was shown to have a significant effect on enhancing the predictive accuracy and robustness of the PLS model as it reduced the effect of instrumental noise. The methodology was then extended to a two-component protein system consisting of α-lactlalbumin (α-LA) and β-lactoglobulin (β-LG). The process of thermal induced aggregation of β-LG and α-LA protein in mixtures, which involves the disappearance of native-like proteins, was studied under various treatment conditions including different temperatures, pH, total initial protein concentration and proportions of α-LA and β-LG. A Partial Least Squares (PLS) regression algorithm was used to correlate the concentrations of α-LA and β-LG to the fluorescence spectra obtained for mixtures.

The results illustrated that multivariate models could effectively deconvolute multi-wavelength fluorescence spectra collected for the protein mixtures and thereby provide a fairly accurate quantification of respective native-like α-LA and β-LG despite the significant overlap between their emission profiles. It was also demonstrated that a PLS model could be used as a black-box prediction tool for estimating protein aggregation when combined with simple mass balances.

Ultrafiltration experiments of the whey protein isolate solutions were carried out in dead-end filtration mode and fluorescence measurements of permeate and retentate solutions were acquired in synchronous scanning mode using a fiber optic probe. By implementing a dilution strategy for the retentate side, the fluorescence based PLS model encompassed a low protein concentration range where fluorescence was not expected to be significantly influenced by concentration-dependent interferences. It was also demonstrated that synchronous spectra can provide good predictions and consequently the use of the full spectrum may not be necessary for monitoring with corresponding savings in acquisition time. Membrane performance variables that are difficult to measure, such as individual protein transmission and membrane selectivity could be estimated directly from fluorescence-based predictions of protein concentrations in the retentate and permeate streams.

Multiwavelength light scattering spectra, acquired using the fiber optic probe, were shown to be a useful indicator for protein self-association behavior, which is known to influence the membrane filtration. High fouling potential were observed for protein solutions that exhibited significant Rayleigh scattering. A predictive PLS model for estimating protein aggregation from Rayleigh scattering measurements was developed and it was tested by using molecular weight experimental values obtained from the literature. Although this comparison was only partial due to the limited amount of molecular weight data available, the findings verified the possibility of estimating the aggregate size from multiwavelength Rayleigh scattering spectra acquired using a conventional spectrofluorometer. Thus, the results implied that both intrinsic

fluorescence and light scattering multiwavelength measurements could provide complementary information about the filtration process.

Acknowledgements

By the name of Allah, The Most Gracious, The Most Merciful ... All praise and thanks is due to Allah. He is the one who made this work possible. He is the one who gave me the strength and patience to accomplish this work.

I would like to extend my thanks and appreciation to my supervisors Profs Ray Legge and Hector Budman for helping me not to get lost during the development of this project. Both of them provided a motivating, enthusiastic, and critical atmosphere during the many discussions we had. It was a great pleasure to me to conduct this project under their supervision. I am also grateful to Prof Christine Moresoli for her valuable advice, guidance and encouragement throughout the project.

Financial support provided for this work by the following organizations is gratefully acknowledged: NSERC-Cellnet Research Network for, the School of Graduate Studies and the Department of Chemical Engineering in the University of Waterloo.

I would like to express my thanks and gratitude to Joe Clifford for his technical assistance in using the HPLC system, Julien Urfer and Chi Quang Huynh for their help with experimental work, Prof. L. Meiering and Prof. Pu Chen (University of Waterloo) for access to the DLS instrument, Prof. C. Moresoli for providing the whey protein powders, Mr. B. Habicher for modifying the stirred cell apparatus.

Finally, I wish to extend my heartfelt gratitude to my wife, Lara, my mother, Randa, my father, Shereef and my brothers, Ahmed and Noor. They have endured hardship, struggle, and sacrifices with me throughout the different stages of my studies. I personally would have not achieved this result without their valuable support, endurance, patience, sacrifices and encouragement. I also want to extend my thanks to my colleagues in the bio-group for creating a friendly work environment. My warmest thanks go to Ramila Peiris for being the way he is, a true friend and a great listener.

Table of Contents

List of Figures

List of Tables

Nomenclature

σ_i	Standard deviation of data in the ith row
σ_j	Standard Deviation of data in the jth column
τ_{ob}	Observed transmission coefficient
ψ_i	Membrane selectivity
y_{pred}	Predicted value (% of aggregation or protein concentration g/L)
y_{obs}	Observed or reference value(% of aggregation or protein concentration g/L)
x_{ij}^{VS}	Element in the X-matrix pretreated by variable scaling
x_{ij}^{MC}	Element in the X-matrix pretreated by mean centering
x_{ij}^{SNV}	Element in the X-matrix pretreated by SNV
x_{ij}^{NM}	Element in the X-matrix pretreated by normalization
$\bar{x}_i$	Average value of data in the ith row
$\bar{x}_j$	The average value of data in the jth column
(λ_{exci})	Excitation wavelength (nm)
(λ_{em})	Emission wavelength (nm)
$F_i(\lambda_{ex}, \lambda_{em})$	Fluorescence signal emitted by a fluorophore *i*.

β_j	Regression coefficient or weight associated with measurement at a given combination of excitation-emission pair.
$\phi(\lambda_{em})$	Quantum yield of the fluorophore at the emission wavelength
$\mu_{ai}(\lambda_{ex})$	Absorption coefficient at the excitation wavelength
ε	The heteroscedastic noise
BSA	Bovine Serum Albumin
C_b	Bulk concentration (g/l)
C_g	Gel concentration (g/l)
C_p	Permeate concentration (g/l)
E	The matrix containing the residuals of X
F	The matrix containing the residuals of Y
Fl	Fluorescence intensity
FOP	Fiber Optic Probe
HPLC	High Performance Liquid Chromatography
i	Sample index
j	Variable index
Lyz	Lysozyme
MLR	Multiple Linear Regression
n	The number of observations (i.e. samples in our case)
n	Number of samples
NM	Normalization
p	The number of samples in the test set.
p	Number of spectral variables

P	Loading matrix
pI	Isoelectric point (IEP)
PLS	Partial Least Square
Q^2	The squared predictive correlation coefficient
RMSEP	Root mean square error of prediction
SNV	Standard Normal Variate
T	The matrix containing latent variables scores of X
TMP	Transmembrane pressure
U	The matrix containing latent variables scores of Y
UV	Ultraviolet Light
w	PLS weight coefficient for x variables
W	Weight matrix
X	Matrix containing fluorescence data
x_{ij}	The intensity value of the j th spectral variable (excitation-emission wavelength pair) for the i th case (i.e. sample).
α-LA	α-lactoglobulin
β-LG	β-lactoglobulin
Δλ	Wavelength interval (nm)
x_{ij}°	Noise-free intensity value of the j th spectral variable (excitation-emission wavelength pair) for the i th case (i.e. sample). fluorescence signal for a given sample i

CHAPTER 1

Research Aim and Outline

The ability to purify proteins in a cost-effective manner on a commercial scale and to meet the required high purity for pharmaceutical and food products is an important technical goal that industry is striving to achieve. A large number of protein products are used as foods, food additives, therapeutic proteins that are recovered from various plant, microbial and animal sources; however, most protein-based products need to be purified before they can be used (Ghosh, 2003). As a result, protein separation technologies play a vital role as unit operations in the development and commercialization of high-value protein-based products. Membrane filtration is a technique that is commonly used in the biotech, food and beverage industries and it relies on the use of a synthetic membrane acting as a physical barrier to separate the target protein from other substances. Membrane pore size and membrane chemistry determine the type of molecules that can be processed (Zeman and Zydney, 1996).

During filtration, target proteins in the soluble phase are removed via the membrane to the permeate phase while macromolecules and particles larger than the membrane pores are retained by the membrane (i.e. in the retentate stream) as seen in Figure 1.1. High performance filtration is usually characterized by high protein product

transmission, high separation quality and high product quality (Zeman and Zydney, 1996).

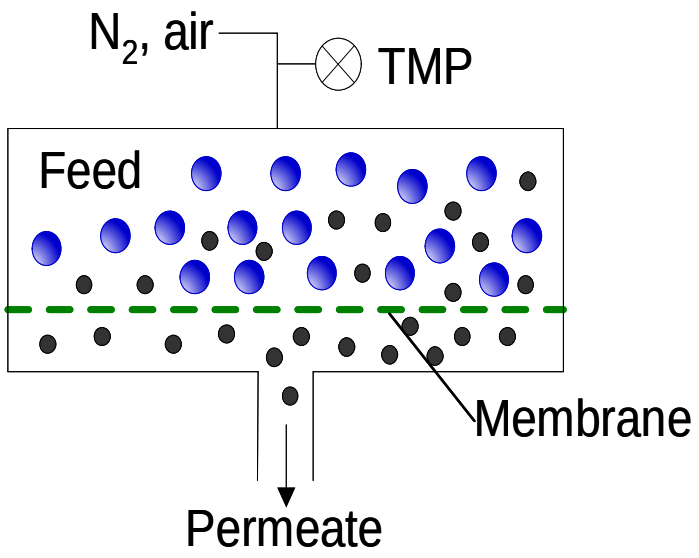


Figure 1.1 Schematic diagram for membrane-based protein separation module (as depicted in Skořepová, 2007).

1.1. Research Motivation

The most critical problem encountered in membrane filtration processes is membrane fouling which has an adverse effect on the performance of the separation. Membrane fouling is referred to as the accumulation of proteins and/or other particles in the feed stream inside the pores and on the external membrane surface, which will alter the sieving characteristics of the membrane, and add more resistance to the flow (Figure 1.2). Membrane fouling in protein separation processes is a very complicated phenomenon compared to fouling caused by dissolved solids in water treatment systems (Güell *et al.*, 1998). This is due to the complexity of protein mixtures. The

complexity of protein mixtures arises from the presence of surface interactions between species in the bulk solution as well as interactions between the membrane and these species (Palacio *et al*., 2003). The extent of membrane fouling by proteins involves the interplay of a large number of physicochemical conditions (i.e. solution pH and salt concentration) and hydrodynamic conditions (i.e. wall shear rate and permeate flux) (Tracey and Davis, 1994; Ghosh, 2003). Just a small variation in the above-mentioned conditions induces important changes in the way the process has to be operated (Arora and Davis, 1994). Membrane fouling involves three different patterns of matter-accumulation phenomena on its surface: (1) concentration polarization, (2) (followed by) cake/gel layer formation, (3) and aggregate cake formation (i.e., cake of retained aggregates) (Redkar and Davis, 1993).

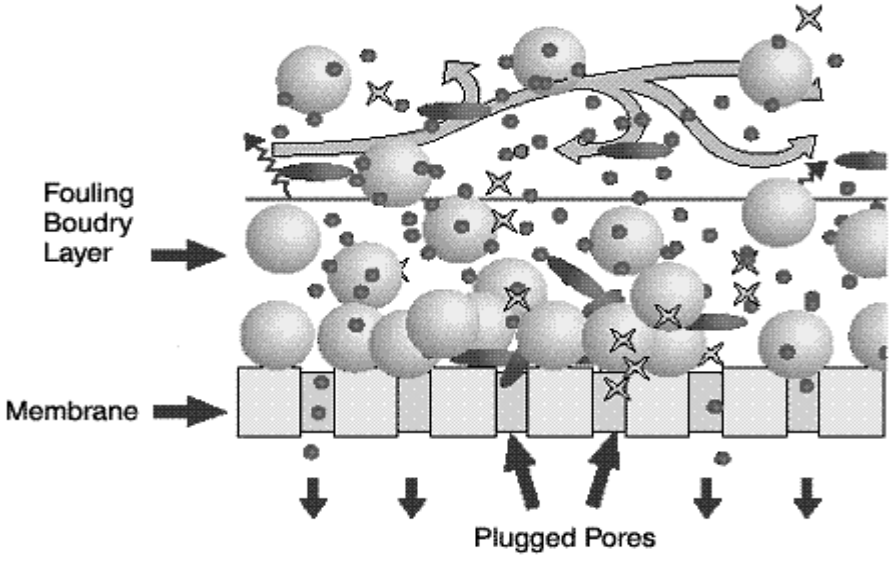


Figure 1.2 Fouling caused by the accumulation of proteins and aggregates on the membrane surface (http://www.dunwellgroup.com/ProductsServices/DNL/VSEP/Images/Figure1.gif).

Clearly, membrane fouling has a significant effect on product yield and productivity. The deposition of proteins inside pores and the tendency of these proteins to aggregate may cause yield losses of the target protein. In addition to product yield considerations associated with fouling, there are economical considerations. Protein fouling decreases the permeate flow and thereby drives up the cost of operation and maintenance (Davis, 1992).

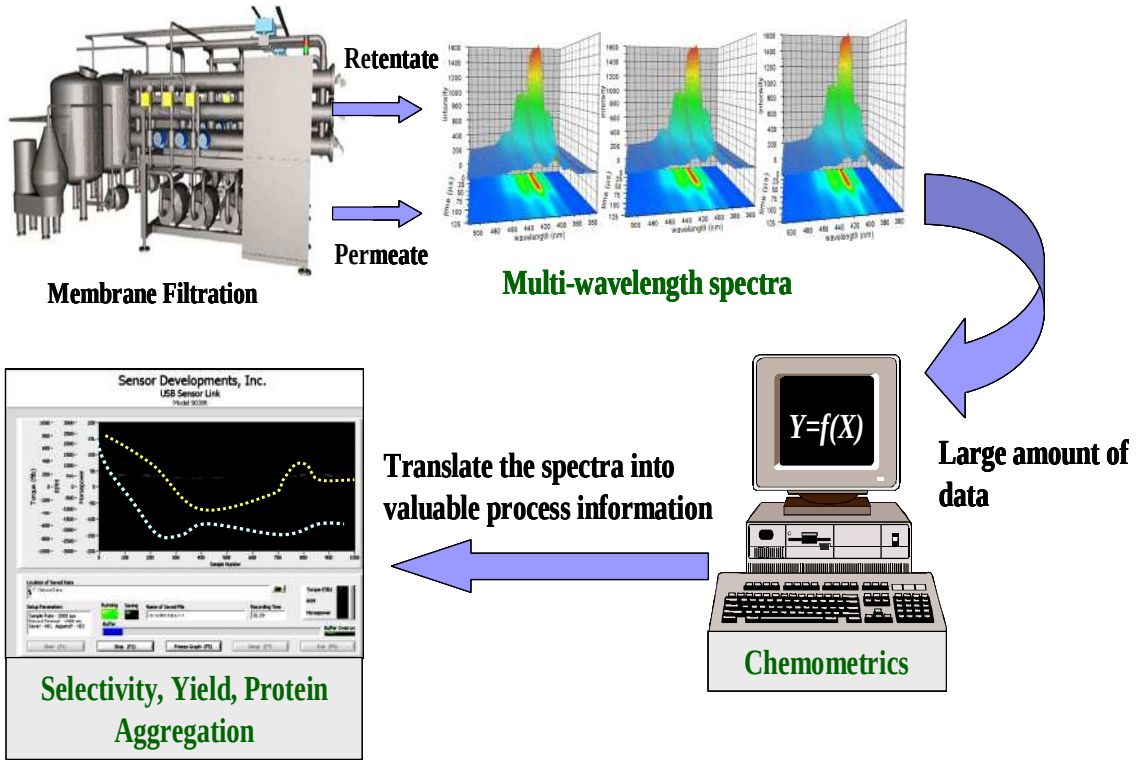


Figure 1.3 Feasibility of applying fluorescence spectroscopy in conjunction with chemometric techniques for monitoring the filtration process.

It would be highly beneficial to industry if an appropriate monitoring system could be implemented to determine the extent of protein fouling, feed composition, product purity, separation efficiency and yield in a minimum period of time. The traditional approach for determining product purity is based on off-line analysis techniques such

as analytical chromatography (e.g. HPLC and size exclusion measurements), which results in a time delay of several hours or days between sampling and obtaining the results. The traditional approach does not offer an option in terms of process control. There is a need for fast and reliable methods of monitoring protein membrane separation for process control and optimization.

1.2. Research Objectives

The first objective of this thesis was to examine the feasibility of applying multi-wavelength fluorescence spectroscopy in conjunction with chemometric techniques for monitoring and optimizing filtration processes as illustrated in Figure 1.3. A major advantage of fluorescence spectroscopy over other analytical techniques is that it is rapid, noninvasive and very sensitive. Chemometric techniques rely upon multivariate statistical and mathematical tools for decomposing a measurement into valuable process information.

The second objective was to use this information to identify the range of operating conditions (i.e. the transmembrane pressure and the pH value of the feed solution) that will lead to optimal membrane performance. Achieving this objective would most likely improve the quality of bioproducts, lower the cost of membrane replacement, control membrane fouling and thus prolong membrane stability and integrity. In addition, it could allow manufacturers to determine the necessary operating conditions for minimization of potential fouling phenomena and to address concurrent product and process engineering issues.

The third objective was to study the feasibility of utilizing multi-wavelength fluorescence data for predicting the level of protein aggregation that is considered to be a major factor that influences membrane fouling.

1.3. Thesis Structure

Chapter 2 explains many of the practical and theoretical fundamentals of membrane filtration, fluorescence spectroscopy and multivariate calibration methods. The first section of Chapter 2 begins by providing a theoretical and practical basis of membrane filtration and factors that influence the performance of membrane-based protein separation processes. Chapter 3 examines heat-induced aggregation behavior of a model protein β-LG under different conditions. A predictive model based on fluorescence measurements is developed for predicting protein solubility and its aggregation behavior and progress with time at different temperatures and pHs. Chapter 4 extends the studies conducted in Chapter 3 by considering a two-component system consisting of α-LA and β-LG as a model system. A predictive model based on fluorescence measurements is developed for predicting the solubility of individual components and their aggregation behavior after heat treatment. Chapter 5 presents a novel methodology for monitoring filtration process performance by using fluorescence measurements acquired using a fiber optic probe. Important parameters for evaluating membrane filtration performance (i.e. membrane selectivity, protein transmission behavior) were shown to be estimated using fluorescence-based measurements of the permeate, retentate and feed streams. Chapter 6 demonstrates how Rayleigh scattering signals of the feed stream can be used to gain more insight

into the potential of these mixtures for fouling and to provide quantitative information on average aggregate size of the protein in solution. Chapter 7 highlights conclusions from the thesis and suggestions for future work. Chapters 3, 4, 5 and 6 are arranged in publication format, each with its individual abstract, introduction, materials and methods, results and discussion as well as conclusion.

CHAPTER 2

Theoretical Background

2.1 Membrane Based Separation Processes

Membrane-based separation processes are those processes whose key component is a synthetic membrane, a thin barrier through which solutes and solvents are selectively transported (Ghosh, 2003). Such membranes are manufactured from organic polymers or inorganic materials resulting in membranes with different structural morphologies and chemical properties.

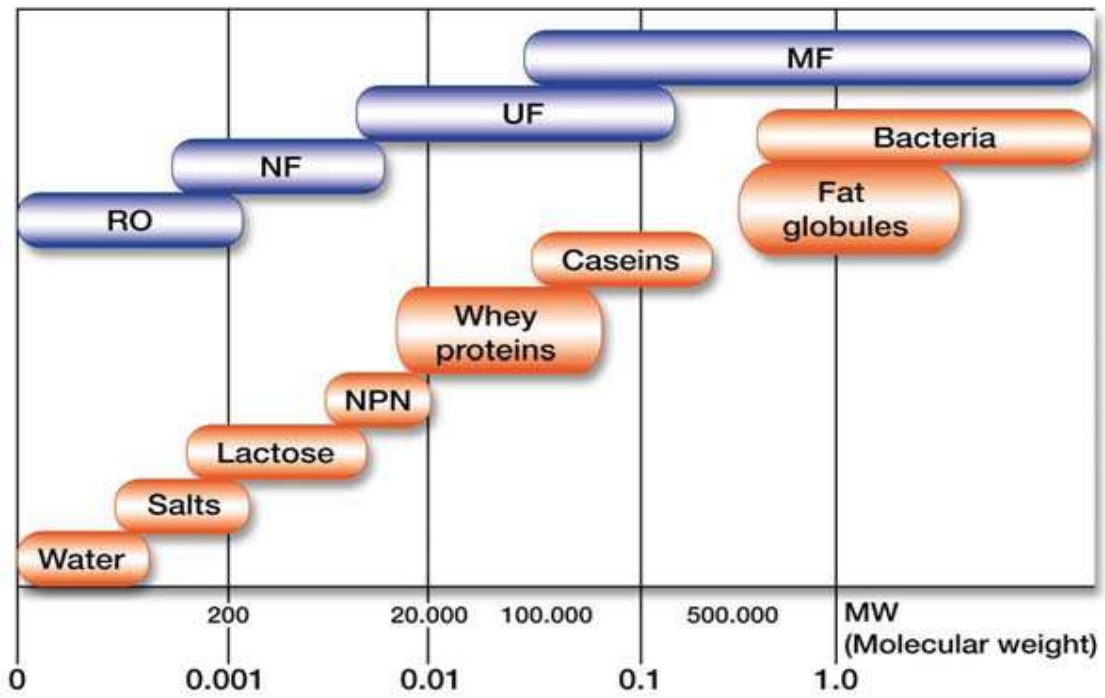


Figure 2.1 Classification of membrane processes for separating milk components based on membrane pore size and/or the type of material being processed (as depicted in http://www.idfa.org/meetings/presentations/milktechconf_cold.pdf)

Membrane characteristics such as membrane porosity, selectivity and hydraulic permeability are all dictated by its structural morphology and chemical characteristics (Ghosh, 2003). Membrane-based separation processes are generally classified based on the membrane pore size or on the type of material being processed. Figure 2.1 shows the classification of membrane processes for separating milk components based on membrane pore size and/or the type of material being processed.

The overlap between different types of membrane-based separation processes is so significant that, in some cases, classification becomes difficult. For instance, in spite of the fact that ultrafiltration is generally applied to filtration of macromolecules such as proteins, smaller molecules or even particulate material can also be processed by ultrafiltration (Ghosh, 2003).

2.2 Major Concerns Associated with Ultrafiltration-Based Protein Fractionation

2.2.1. Membrane Fouling

Fouling is defined in general as adsorption and deposition of material present in the feed which results in reduction in the permeate flux due to an increase in the membrane resistance and decrease in hydraulic permeability. There are three types of protein fouling classified according to where it occurs: internal, pore blockage and external (Figure 2.2). Internal fouling (pore narrowing/constriction) is defined as the deposition and adsorption of feed particles inside the membrane surface (Güell *et al.,* 1998). This

leads to pore narrowing and constriction of the flow and eventually reduction of the permeate flux through the membrane.

The extent of internal fouling is dependent on the membrane characteristics such as morphology, pore size distribution and average pore size. For example, very porous membrane surfaces are highly susceptible to internal fouling because proteins can easily penetrate and accumulate on the interior surface of the membrane pores (Davis, 1992). Pore plugging refers to the accumulation of the larger protein aggregates or particles on the pore entrance thereby increasing the resistance of the membrane. Pore plugging, however, does not appear to contribute significantly to the decline in flux that occurs during microfiltration of protein solutions, as the average protein particle size is much smaller than the average membrane pore size (Belfort *et al*., 1994; Zeman and Zydney, 1996). Protein fouling can also be classified as reversible or irreversible (Palacio *et al.*, 2003).

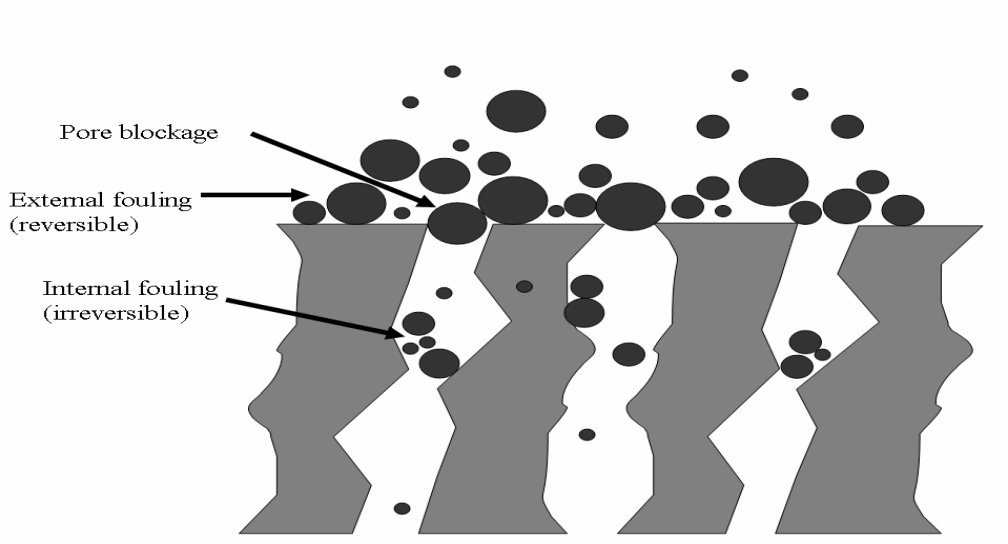


Figure 2.2: Mechanisms of membrane fouling (as depicted in Guettler, 2006)

2.2.2 Concentration Polarization

Concentration polarization is referred to as the accumulation of retained solute at the upstream surface of the membrane (Zeman and Zydney, 1996). Concentration polarization can affect the process through the following mechanisms (Zeman and Zydney, 1996):

- A localized high protein concentration can result in protein/protein interactions leading to concentration polarization and a high osmotic pressure.
- When the concentration polarization is severe, the wall protein concentration can reach a value called gelation-concentration at which a protein gel layer can form on the membrane surface as shown in Figure 2.3. That gel layer provides an additional hydraulic resistance to the solvent flow in series with that provided by the membrane (Ghosh, 2003).
- The accumulation of solutes in the immediate vicinity of the membrane surface increases the amount of protein adsorption, which may result in irreversible fouling of the membrane, altering its sieving characteristics (Babu and Gaikar, 2001; Zeman and Zydney, 1996).

All of these mechanisms can change the apparent sieving coefficient of proteins and therefore the selectivity of the membrane towards permeable protein species could decrease with filtration time. Permeate flux decreases with time also as the extent of concentration polarization increases resulting in decreasing process productivity. Techniques useful in minimizing concentration polarization would also be useful in minimizing fouling (Ghosh, 2003).

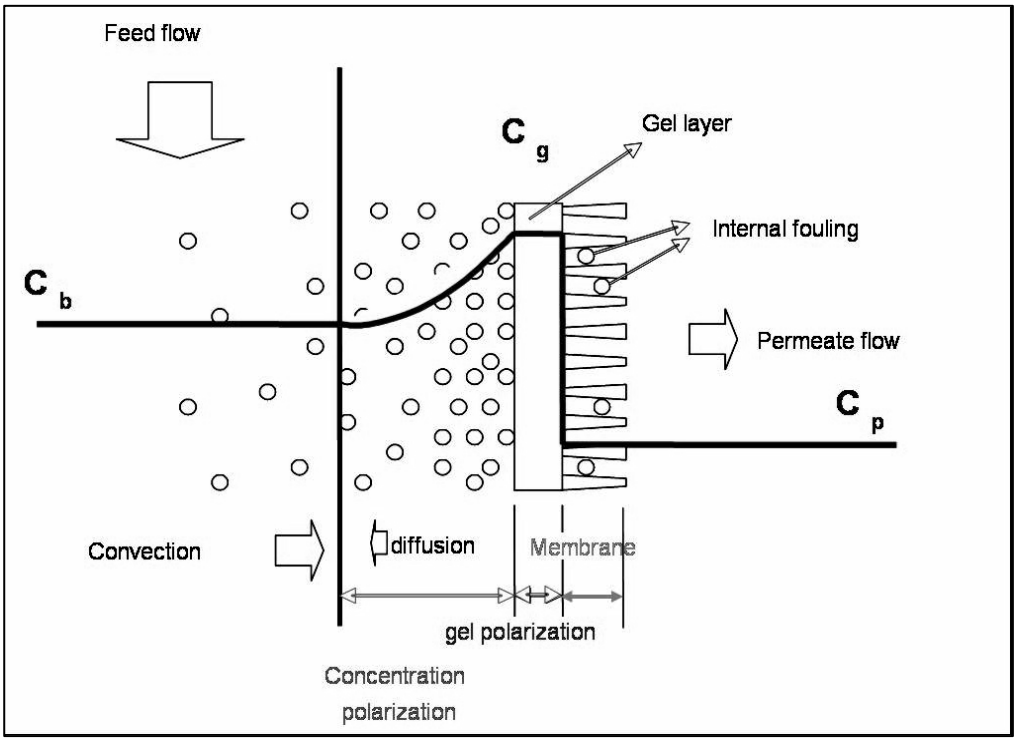


Figure 2.3: A schematic diagram showing the accumulation of retained solute at the upstream surface of the membrane (i.e. the concentration polarization) where C_b, C_p and C_g denote the bulk, permeate and gel concentrations respectively (as depicted in http//:te-webserver.cce.iastate.edu/courses/ce525/Membrane)

2.2.3. *Protein denaturation and Aggregation Behavior*

Achieving optimal performance of protein separation by membrane filtration is a challenging task due to the complexity of protein behavior in solution. This behavior can be explained by the occurrence of several events such as protein-protein interactions, denaturation and aggregation over the course of the filtration process. The likelihood of occurrence of any of these events is determined by the surrounding conditions, as previously described. Monitoring protein stability and aggregation behavior in the feed solution and in the feed line is essential to ensure cost efficient and reliable operation over the course of filtration. Protein denaturation generally involves conformational and structural changes related to partial unfolding of the native protein

that leads to the exposure of some hydrophobic amino acid residues. The degree of denaturation is very complicated as it depends on a number of physical and chemical parameters such as temperature, protein concentration, protein-protein interactions, ionic strength, pH and shear (Vetri and Militello, 2005). In their study of microfiltration fouling behaviour of a β-lactoglobulin solution, Marshall *et al.* (1997) observed an increase in fouling resistance with increasing flux, which was accompanied by a decrease in protein transmission across the membrane due to protein aggregates blocking a majority of membrane pores. According to Marshall *et al.* (1997), higher mechanical shear will favor protein denaturation and aggregation as more protein molecules might undergo conformational changes in their structure. Protein denaturation is usually followed by aggregation of the partially unfolded protein molecules via the formation of new intermolecular bonds between the exposed amino acid residues (Mulvihill and Donovan, 1987). Aggregation can occur due to non-covalent bonding (electrostatic and dipolar Van der Waals attractive forces).

2.3 Performance of the Protein-protein Fractionation

The performance of membrane filtration process can be assessed in terms of the selectivity of the membrane, product recovery (yield) and the protein transmissions. Filtration performance is affected over time as a result of protein fouling and concentration polarization across the membrane. In Chapter 5 the use of fluorescence spectroscopy for monitoring these parameters during the filtration process will be illustrated. A detailed description of all of these filtration performance parameters is provided here for information.

2.3.1 Product Yield

Product yield (also called recovery) of the targeted protein (i.e. the component that needs to be purified) is defined as the ratio of the total mass of the recovered protein in the product stream to its initial total mass in the feed solution (Cheang and Zydney, 2004). The definition of the product stream can be either the permeate or retentate depending on the objective of the separation (Mulder, 1996). If the objective is to concentrate the targeted component in the feed by removing smaller solutes through the membrane, then the retentate is considered to be the product stream. However, when membrane filtration is used for purifying the target component by removing larger solutes, then the permeate stream is considered to be the product stream since it contains the target components that selectively pass through the membrane (Mulder, 1996). The overall protein yield for a given protein *i* is usually estimated using Equation 2-1 (Cheang and Zydney., 2004):

$$Y_i = \left(\frac{V_p C_p}{V_f C_f} \right)_i \times 100 \ \% \qquad \textbf{(2-1)}$$

where V_p and V_f are the volume of the product (either permeate or retentate) and the initial feed solutions, respectively.

Product losses can be attributed to numerous sources such as membrane fouling, protein solubility reduction, protein denaturation and aggregation. Such sources of product loss are influenced by the operating conditions and the feed solution properties. Therefore, product yield can be sensitive to changes in these conditions. A thorough evaluation of each of these areas should result in a process with stable and continuous

high yields. It is also possible that over time mechanical and or chemical degradation of the membrane can occur and retention could change over time (http://www.millipore.com/techpublications/tech1/an1026en00). Trends in process yields which change over time are indicative of this type of behavior. Achieving optimal product yield for membrane separation processes is one of the challenges encountered in the food and pharmaceutical industries (Tutunjian., 2006). It is important to take samples and measure increased product losses in order to confirm what is happening (http://www.millipore.com/techpublications/tech1/an1026en00). For industrial processes where the primary purpose of filtration is the concentration of product present in feed solution from a previous purification or separation step, real time measurement can allow accurate determination of concentration end point and hence maximize yield.

2.3.2 Protein Sieving Coefficient

The transmission coefficient (also called the protein sieving coefficient), which is equal to the ratio of the concentration of a component in the permeate (C_{pi}) to the bulk (C_{bi}), is an important quantity for monitoring membrane performance (Ghosh, 2003). It is given by the following:

$$\tau_{ob} = \frac{C_{pi}}{C_{bi}} \qquad \textbf{(2-2)}$$

where C_{pi} and C_{bi} are the permeate and the bulk concentrations respectively of a given protein *i*.

2.3.3 *Membrane Selectivity*

The efficiency of the binary protein fractionation is commonly expressed in terms of the selectivity which is defined as (Ghosh, 2002):

$$\psi_i = \frac{\tau_i}{\tau_j} \qquad \textbf{(2-3)}$$

Where τ_i stands for the sieving coefficient of the preferentially transmitted protein while τ_j stands for the sieving coefficient of the preferentially retained protein. The selectivity in ultrafiltration is mainly determined by the molecular size of the proteins to be separated if there is a significant difference in their size. In the case of proteins having comparable sizes, other factors such as physicochemical conditions can be manipulated to play significant roles in determining selectivity. Eq (2-3) is valid only for binary protein mixture and it can not be used for assessing the efficiency of protein fractionation of complex protein mixtures that consist of more than two proteins (Ghosh, 2002). Instead, a new parameter termed effective selectivity can be introduced for accurate description of selectivity in ultra filtration of complex protein mixtures (Ghosh, 2000). Such parameter is given as (Ghosh, 2003):

$$\psi_i = \frac{\tau_i}{\sum \tau_{j \neq i}} \qquad \textbf{(2-4)}$$

It is obvious that the selectivity value is dependent upon the sieving coefficients of individual proteins, which in turn, would depend on parameters such as pH, ionic strength, permeate flux and the degree of concentration polarization as will be described in section 2.4. The selectivity of separations generally changes with process

time as a consequence of membrane fouling, concentration polarization and protein-protein interactions. The selectivity of separation can be enhanced by the optimization of pH, feed concentration, salt concentration or permeate flux (Ghosh, 2002). The effect of permeate flux on the selectivity coefficient for ultrafiltration of BSA and IgG is illustrated by Figure 2.4(Ghosh, 2002). Rapid monitoring of the selectivity value during operation as a result of changing these variables could help the operator to find the optimal conditions either in pilot-lab scale or at industrial scale to achieve highest selectivity.

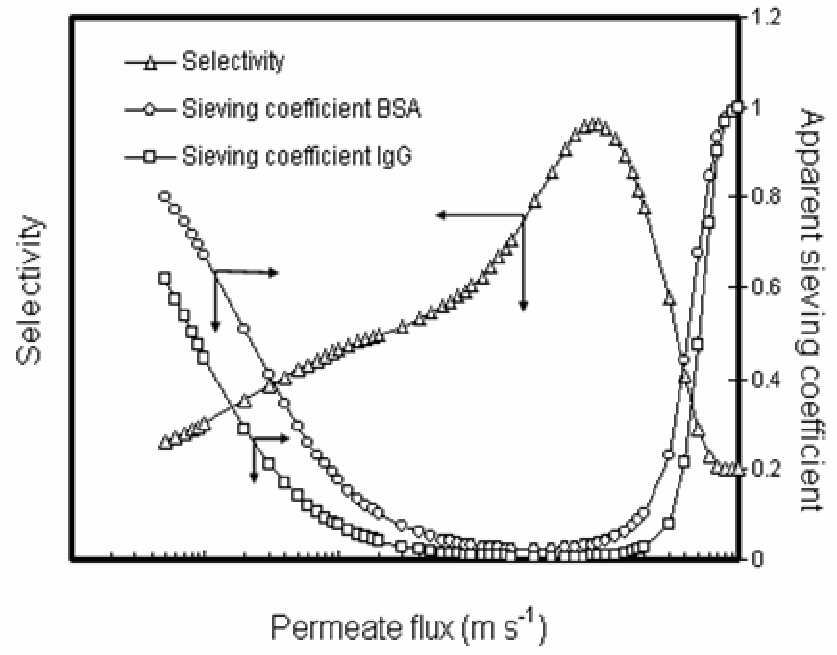


Figure 2.4: The effect of permeate flux on the selectivity coefficient for ultrafiltration of BSA and IgG is illustrated by Figure (Ghosh, 2002).

2.4 Factors Affecting the Separation Performance

The efficiency and cost of membrane processing is dependent on flux and transmission, which is a function of different factors. The membrane type, processing parameters and

the feed solution properties will determine flux and concentration of components in the permeate and retentate. Protein fractionation can be affected by numerous factors that are relevant to the process. These factors can be divided into two categories: operating factors and properties of the feed solution (Balakrishnan and Agarwal., 1996; Chan, 2002). Operating parameters refer to the pressure drop across the membrane and the stirring speed, while the properties of the feed solution involve physiochemical properties of the feed (i.e. pH and ionic strength), total protein concentration and feed composition. Those variables have to be manipulated in order to achieve optimal performance of the protein separation process (Chan, 2002).

2.4.1. *Processing Parameters (Operating Conditions)*

2.4.1.1 Transmembrane Pressure *TMP*.

Transmembrane pressure is defined as the difference in pressure between the permeate side and the feed side of the membrane. The transmembrane pressure provides the driving force for the separation. In general, the initial permeate flux increases linearly with an increase in the transmembrane pressure. Increasing the transmembrane pressure can increase the driving force acting on the permeable molecules, and eventually permeate flux also increases (Sulaiman and Aroua, 2002). However, the occurrence of such increase in the permeate flux as a result of increasing the *TMP* is only temporary and it is usually followed by a rapid flux decline soon after (Chan, 2002). At the same time, the increase in the permeate flux may be accompanied by an associated momentary increase in the solute permeability through the membrane before a sharp decline (Chan, 2002). Such a time dependent decrease in flux and solute

permeability are induced by the accumulation of solutes on the membrane leading to the deposition and adsorption of solutes on the membrane surface as well as within the pores (Güell *et al.*, 1998, Chan, 2002).

2.4.1.2. Stirring Speed

It is known that stirring reduces the concentration polarization of solutes at the membrane surface since it increases the rate of back-diffusion of solutes from the membrane surface towards the bulk. The consequence of such an effect is greater protein rejection and at the same time an enhancement in the permeate flux (Chan, 2002). Some researchers have observations shown that higher stirring speeds may not necessarily improve the separation process. For instance, Franken *et al.* (1989) showed that higher stirring speeds caused the transmission of BSA to decrease with respect to time. According to the authors, such a decrease in BSA transmission can be attributable to the shear induced aggregation of BSA caused by the stirring action, resulting in plugging of the pores and hence increased rejection (Chan, 2002). Mukai *et al.*, (1998) performed a stirred-cell ultrafiltration study to investigate the effect of stirring on the separation of BSA and egg white lysozyme. Although the filtration rate was improved by increasing the stirring speed, lysozyme transmission across the membrane was reduced (Chan, 2002, Mukai *et al.*, 1998). A possible explanation of this is that stirring reduced the concentration polarization of lysozyme and therefore, the concentration of lysozyme at the membrane surface decreased. Since the protein transport through membrane pores is influenced by the concentration gradient between the retentate and

the permeate sides, decreasing the concentration at the membrane wall would result in reducing the transmission of the protein (Chan, 2002, Mukai *et al.*, 1998).

2.4.2. *Feed-Stock Conditions:*

2.4.2.1. Physicochemical Conditions: pH and ionic strength

Physiochemical conditions (pH and ionic strength) have a profound influence on membrane fouling as they affect both intermolecular protein-protein interactions as well as electrostatic protein-membrane interactions (Redkar and Davis, 1993). These interactions can be manipulated for the purpose of reducing membrane fouling by adjusting the feed solution pH and ionic strength (Chan, 2002). Each protein has a certain isoelectric point. The isoelectric point (IEP) is defined as the neutral pH value at which the protein molecule acquires zero net charge and as a result, electrostatic attractive forces prevail in the protein solution. At pH values below the isoelectric point, protein molecules acquire net positive charges while above the isoelectric point they acquire net negative charges. These charges increase in magnitude with increasing distance away from the isoelectric point (Chan, 2002; Zeman and Zydney, 1996; Ghosh, 2003). Physiochemical conditions (pH and ionic strength) have a profound influence on membrane fouling as they affect both protein-protein interactions as well as protein-membrane interactions (Redkar and Davis, 1993). These interactions can be manipulated for the purpose of reducing membrane fouling by adjusting the feed solution pH and ionic strength (Chan, 2002; Heath and Belfort, 2006). Each protein has a certain isoelectric point. The isoelectric point (IEP) is defined as the neutral pH value at which the protein molecule acquires zero net charge and as a result,

electrostatic attractive forces prevail in the protein solution. Away from the IEP, intermolecular electrostatic repulsive forces dominate protein-protein interactions as protein molecules in solution acquire similar charges. Below the isoelectric point, a net positive charge prevails while a net negative charge is present above the isoelectric point (Chan, 2002; Zeman and Zydney, 1996; Ghosh, 2003). Generally speaking, membrane fouling becomes significant at the protein's isoelectric point due to the electrostatic attractive forces developing between protein molecules. Palecek and Zydney (1994) studied the fouling behavior of five protein solutions: hemoglobin, lysozyme, ribonuclease A, immunoglobulin G and BSA in a stirred cell filtration device at pH 7. Flux decline data were obtained as a function of time. The immunoglobulin G, hemoglobin, and ribonulease A solutions appeared to have approximately identical membrane fouling behavior while less significant flux reduction was observed during the filtration of the other two protein solutions. The reason for this difference in membrane fouling behavior is due to their different isoelectric points. A protein in solution that has an isoelectric point close to the feed solution pH would have a greater tendency to aggregate, leading to a significant decline in flux as in the case of immunoglobulin G, hemoglobin, and ribonuclease A. The magnitude of the electrostatic interactions between charged protein molecules and the membrane is also influenced by the total salt concentration (also called ionic strength) of the solution (Zeman and Zydney, 1996). Salt ions bind to ionized groups on charged protein molecules and produce a charge screening effect on the electrical interactions between these molecules, significantly reducing any electrostatic attractive or repulsive forces (Chan, 2002; Zeman and Zydney, 1996; Ghosh, 2003). Electrostatic repulsion

between the membrane surface and the protein molecules in solution occurs as they acquire similar charge sign and thus the likelihood of protein fouling and protein adsorption on the membrane surface is reduced. On the other hand, the tendency of protein deposition and adsorption on the membrane surface is increased due to attractive electrostatic interactions if both the proteins and the membrane are oppositely charged (Chan, 2002; Zeman and Zydney, 1996; Ghosh, 2003).

2.4.2.2. Feed Concentration

In all studies undertaken so far, it has been found that, the flux decline is higher when increasing the feed concentration (Chan 2002). By increasing the feed concentration, a greater amount of solute will be transported towards the membrane surface and hence greater accumulation of solutes in the membrane boundary layer will occur. This results in greater concentration polarization and increased likelihood of membrane fouling (Chan, 2002).

2.5. Process Analysis and Monitoring

A commonly used phrase “if you do not measure it, you cannot control it” applies as much to membrane filtration processes as to any other part of pharmaceutical and food products manufacturing (Harrington, 1987). The previous section demonstrated that the performance of membrane separation processes is influenced by many factors. Accordingly, in order to develop and optimize membrane processes and in order to operate them efficiently, it is critical that the overall state of the process be monitored and controlled in an appropriate manner and that the response of the process towards

changes in the processing conditions be determined. Achieving these goals requires three different functions: measurement, analysis of measurements and control. This section reviews currently available instrumentations and monitoring methodologies implemented in membrane separation processes.

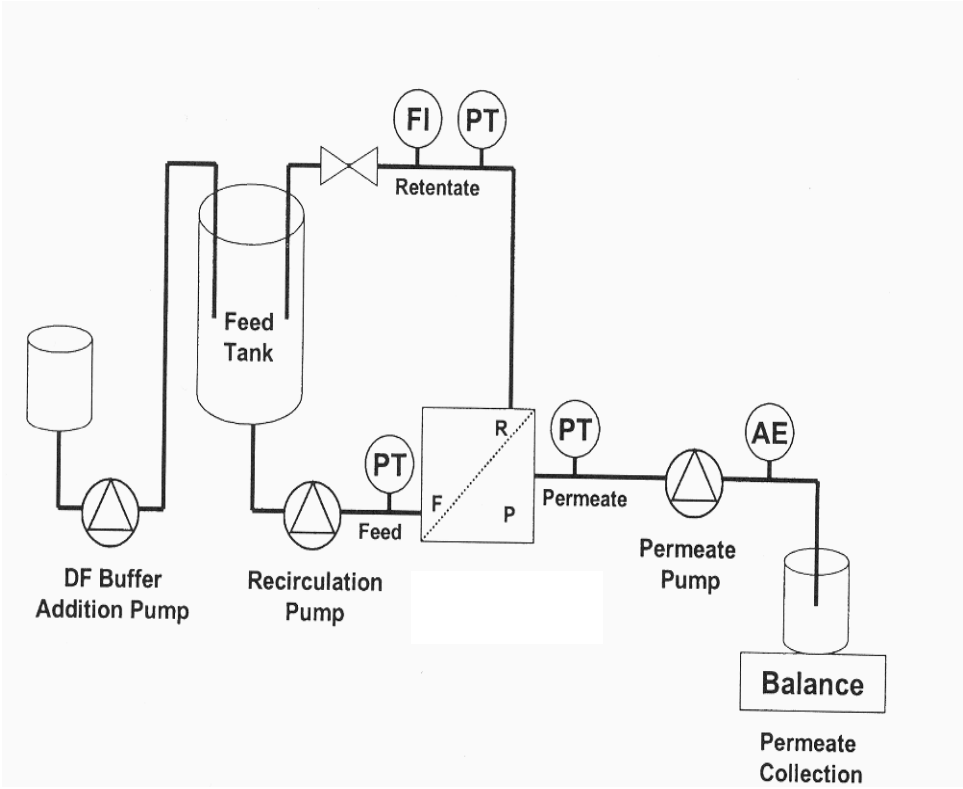


Figure 2.5: A schematic summary of membrane filtration instrumentation (Kelly and Peterson, 2001). The system consists of a feed tank, membrane module, recirculation pump, pressure sensors (PT) and flowmeters (FI) on the feed, retentate and permeate lines.

Figure 2.5 shows a schematic summary of membrane filtration instrumentation. Pressure sensors denoted by (PT) are placed in the feed, retentate and permeate lines to monitor the transmembrane pressure and retentate pressure drop of the membrane operation. Flowmeters (Fl) are installed in both retentate and permeate streams to monitor the stream flow rates (Kelly and Peterson, 2001).

One of the major goals, if not requirements, of obtaining data from membrane filtration is estimation of product recovery and membrane selectivity based on the available measurements of permeate, retentate and feed streams. Process analysis and monitoring systems are classified into two categories based on the location and the speed of analysis as described in the following subsections.

2.5.1 *Off-line Analytical Techniques*

Analysis that is done on samples removed from the process and sent to analytical device for testing is called *off-line*. The objective of sampling in bioprocesses is to acquire representative samples that correspond to the overall state of the process at a certain time (Vojinovic *et al.,* 2006). Traditional laboratory techniques are implemented for the analysis of samples withdrawn from the process such as size-exclusion chromatography. HPLC has been extensively used for analyzing individual proteins in permeate and retentate and for ascertaining the consistency and quality of the desired product stream (Folta-Stogniew & Williams, 1999; Chirino & Mire-Sluis, 2004). Another advantage of size-exclusion HPLC is that it gives a reliable, quantitative determination of the level of aggregation in retentate, feed and permeate, and may also distinguish multimers form the product of interest (Kelly and Peterson, 2001; Wang *et al*., 2003). Due to these advantages, HPLC is an established analytical tool for decades and available in any laboratory dealing with process development and quality control (Folta-Stogniew & Williams, 1999). However, HPLC has proven to be excessively time consuming and labor-intensive and hence it cannot be used for continuous monitoring of the filtration process progress.

2.5.2 *On-line Optical Sensors*

The second type of process analysis is known as *on-line* analysis since the analytical instrument is positioned directly in close proximity to the process line (Vojinovic *et al*., 2006). Optical methods based upon light absorbance (spectrophotometry), scattering (turbiditimeters) or fluorescence measurements (spectrofluorometry) have started to find more application for continuous monitoring of bioprocesses. They provide increased efficiency and productivity through effective process automation without violating integrity of the system and without wasting valuable products through sampling. This section provides a discussion of application, advantages and limitations of these methods in relation to membrane separation processes.

2.5.2.1. UV photometers

Ingham *et al.* (1979) was one of the earliest works that investigated the feasibility of UV photometers for continuous monitoring of a filtration process *in-line*. In a closed loop filtration unit where the retentate and permeate lines were combined together and returned to the stirred cell containing the feed solution, a small part of the returning fluid was diverted to a UV detector, which allowed the UV absorbance of the returning solution to be continuously monitored. Ingham *et al*, found that a stepwise increase in trans-membrane pressure was accompanied by a progressive decrease in the UV absorbance reading, reflecting the lower concentration of the protein in the fluid returning to the stirred cell. They found that such a decrease in protein concentration is due to protein adsorption and deposition on the membrane. However, they confined their emphasis to the steady state portions of the UV time-curves while ignoring the

real-time transient changes of UV readings which accompanied the step changes in the transmembrane pressure.

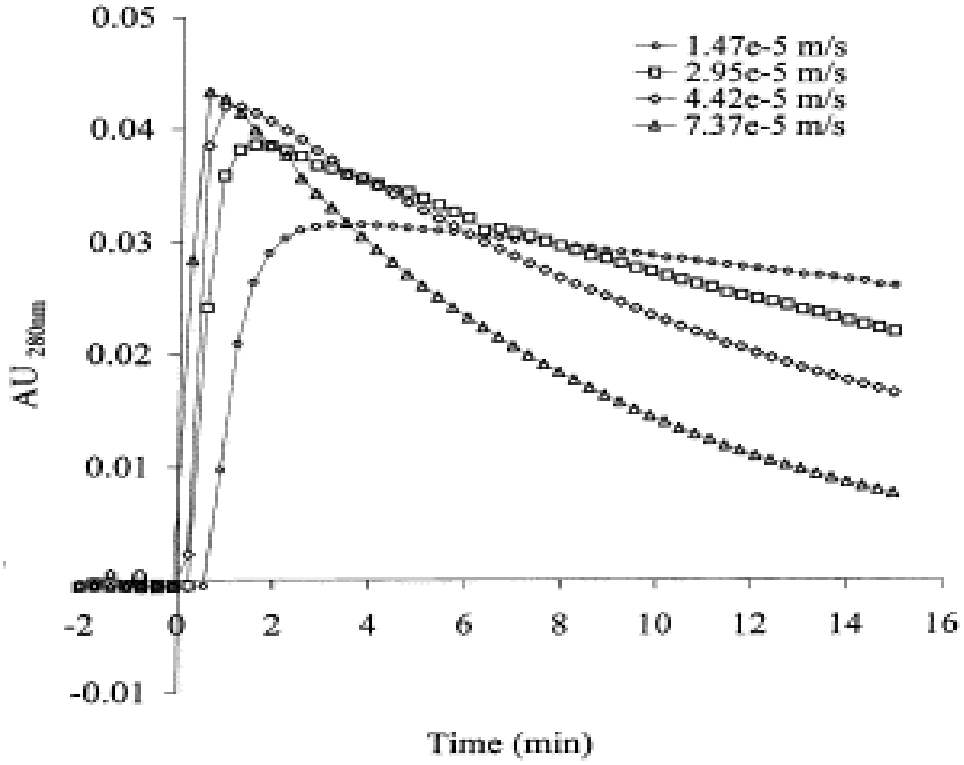


Figure 2.6: Continuous monitoring of the permeate using UV detectors at different permeate fluxes for an ultrafiltration system (Ghosh et al., 2000).

In another study (Ghosh *et al.*, 2000b), an on-line flow-through UV detector was implemented for continuous monitoring of the permeate line under constant-flux ultrafiltration conditions. It was found that the UV time-curve is sensitive to step changes in the permeate flux as seen in Figure 2.6. They observed a faster decrease in the UV value (reflecting protein concentration) at higher permeate flux, which can be attributed to the rapid accumulation of proteins on the membrane surface and thereby hindered protein transport across the membrane. Ingham *et al.* (1974) pointed out that a

more elaborate system containing monitors installed in both lines would allow necessary parameters related to membrane separation performance to be monitored in real time.

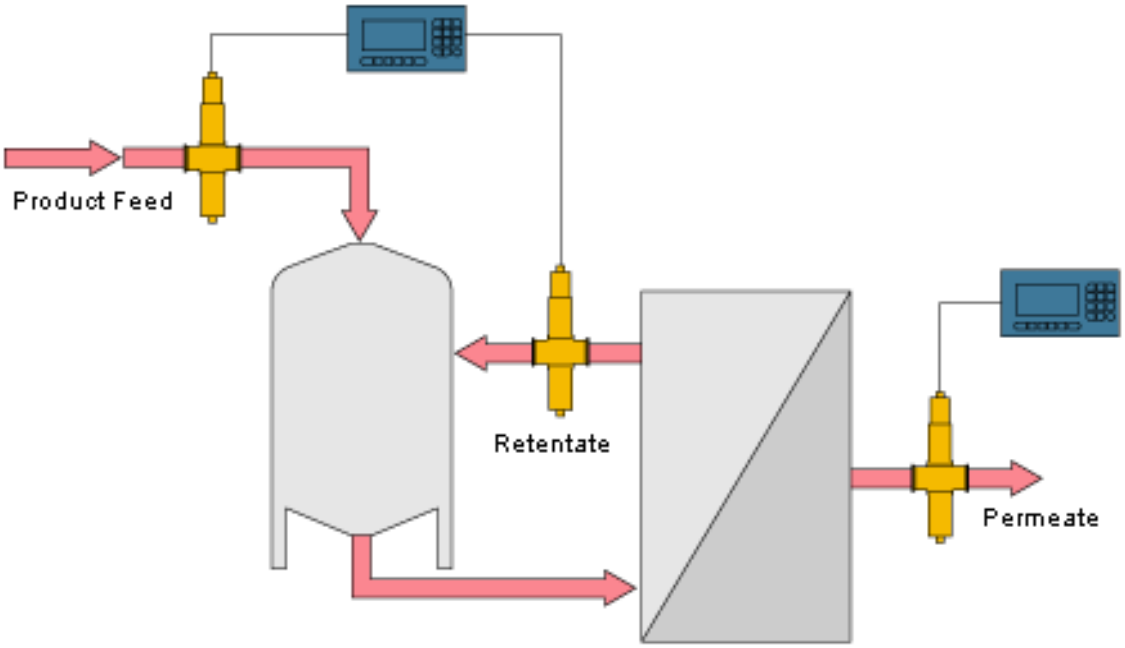


Figure 2.7: Schematic diagram of a typical membrane filtration system showing the positions of in-line photometers installed for monitoring the process streams (as depicted in http://www.optek.com/images/app_Tangential_Filtration.gif)

Figure 2.7 is a schematic diagram of the membrane unit identifying the three process flows that should be monitored: product feed to the recirculation tank, retentate return to the recirculation tank and permeate flow. In systems with an overall control scheme, UV absorption sensors installed at the feed line to the second stage can control the feed concentration by adjusting the operating conditions such as flow rates and pressures to achieve optimal performance of the process.

UV sensors installed at the retentate line can be used for signaling or indicating the desired end point at which acceptable product yield is achieved during the production run (Meltzer and Jornitz, 2003). Although UV monitoring is rapid and non-invasive, it cannot give reliable and quantitative determination of individual protein concentrations in permeate and retentate during ultrafiltration of complex protein mixtures since all proteins absorb UV light in the same wavelength range. In addition, there may be cases where the UV monitor cannot provide useful information, for example, at extremely low or high concentrations of protein, or in the presence of interfering compounds (Kelly and Peterson, 2001).

2.5.2.2 Turbidimeters

The presence of protein aggregates with relatively high concentrations in the influent stream to the filtration unit can contribute to high rates of membrane fouling and thus reduced membrane life. Achieving consistent, efficient and reliable performance for membrane filtration systems is dependent upon maintaining the membrane at the best possible conditions, which can be done by controlling membrane fouling. In order to control membrane fouling, the level of aggregates and colloidal particles incoming to the membrane unit has to be monitored continuously (Orchard, 2006). Turbidity measurements provide a reading for the amount of light scattered by colloidal particles and protein aggregates dispersed in the solution. Turbiditimeters are basically consist of a light source, producing a light beam of known intensity at a single wavelength that is directed into a vial or flow cell containing the turbid medium to be analyzed. Part of this light is either scattered or absorbed by the suspended particles and aggregates. The

amount of scattered light is then recorded and is proportional to the quantity of aggregates suspended in the medium. A turbidity meter that measures scattered light at an angle of 90° is known as a nephelometer (Orchard, 2006). Other types of turbidity monitors employ a "forward scatter" technique. It is also known as a “small-angle scattering” technique since the scattered light detector is placed at a position where the angle between the path of the incident light and the point of detection is smaller than 90-degree angle. The forward scatter measurement is intended to be more sensitive to larger sized colloidal particles, while it has reduced sensitivity for smaller macromolecules (Cowie, 1991; Hiemenz, 1984).

Not only does the size of the particle affect the scattering pattern but also the particle’s shape; non-spherical particles will scatter light differently than spherical particles (Cowie, 1991; Hiemenz, 1984; Nakagaki, 1980). Turbidity and light scattering monitoring are employed in the influent (feed) line as well as in the effluent line in order to evaluate the performance of the membrane modules. Also for large-scale purification and separation processes, where multiple membrane filtration units are used for product recovery, it is recommended that each membrane unit has a separate individual turbidity monitoring system.

By installing turbidity photometric sensors at the influent stream of the filtration process and between each filtering step, flow can be stopped, recirculated or switched to an alternate membrane filtration unit if turbidity reaches an unacceptable level (http://www.optek.com). In-line turbidity measurements permit optimal control of the recirculation loop during the cake layer buildup.

2.5.2.3. In-situ Fluorometry

Fluorescence is a type of optical phenomena that involves absorption of electromagnetic (ultraviolet, visible, or near infrared) radiation by a certain type of molecules (called fluorophores) which results in exciting the molecule to a higher electronic energy level. This is followed by returning of the excited molecule to the ground state, or to lower electronic energy level by losing energy through emission of light (Shea**,** 1997). One of the advantages of fluorescence spectrometry is its high sensitivity that offers detection limits lower than those achievable using current techniques such as UV spectroscopy (Shea**,** 1997; Deshpande, 2001). Because of the low detection limits, fluorescence spectroscopy is widely employed for the analysis of biological and environmental samples where trace constituents in these samples can be quantified (Shea**,** 1997). The importance of this technique is also reflected in its higher selectivity compared to other techniques. Fluorescence spectroscopy can provide more information than UV/Vis absorption spectrometry since the fluorescence signal is expressed as a function of two wavelengths (excitation and emission), while only one wavelength is available in UV-absorbance measurements (Shea**,** 1997; Deshpande, 2001). If two compounds in a sample with similar absorption spectra emit light at different wavelengths, they may be distinguished from one another by appropriate choice of emission wavelength. Similarly, two compounds that have similar fluorescence spectra but absorb strongly at different wavelengths may be distinguished by proper choice of excitation wavelength (selective excitation) (Shea**,** 1997). As such, fluorescence spectroscopy has potentially higher information content for resolving mixtures than UV-VIS absorbance spectroscopy (Baker, 1991). Fluorometry is being

increasingly used in bioprocess and wastewater monitoring applications such as fermentation processes and bioreactors (Hilmer and Scheper., 1996; Tartakovsky *et al*., 1996; Hagedorn *et al*., 2004). However, the use of fluorometry in monitoring membrane separation processes is rarely found in literature. The earliest work that utilized fluorescence spectroscopy for continuous monitoring of protein fractionation is that of Crespo *et al.* (1999). Crespo *et al.* (1999) developed an on-line fluorescence detection technique for monitoring the transmission of β-LG and γ-globulin by using protein labeling with fluorescent markers. The two proteins were labeled with two different fluorescent markers which strongly absorb and emit light at different wavelength ranges. Thus on-line detection of protein-fluorescent labelled conjugates was performed with a fluorescence detector that was programmed at appropriate excitation and emission wavelengths. The transient transmission behaviors of β-LG and γ-globulin through the membrane were identified by the transmission of the corresponding protein-fluorescent label conjugate. The drawback of protein labeling is that it may introduce changes to the protein surface chemistry and to the overall protein charge which may alter protein folding properties and ultimately its aggregation behavior during separation (Crespo *et al.*, 1999). In addition, the technique has some practical limitations since it requires removal of the fluorescent label downstream. Hence, there is a strong motivation to avoid the use of fluorescent labeling in order to preserve the native state of the protein product. What is proposed here in this thesis is to use intrinsic protein fluorescence. The challenge in using intrinsic fluorescence is that the spectra can significantly overlap. To handle the complex fluorescence signals obtained when analyzing multicomponent protein solutions and to resolve the issue of

overlapping information, multi-wavelength fluorometry in conjunction with chemometrics is applied and developed in this thesis.

2.6. Multiwavelength Fluorometry for Process Monitoring

Simultaneous measurements of various fluorophores are possible thanks to the development of fluorometers capable of rapidly recording two-dimensional excitation-emission spectra. Previous studies have shown that multiwavelength fluorometry can be used for on-line monitoring of recombinant protein production from mammalian cells in upstream processes where the presence of intrinsic biogenic fluorophores such as vitamins, NADH, NADPH, FAD and fluorescent amino-acids give important information about the process (physiological state of cells) (Boehl *et al*., 2003; Hisiger & Jolicoeur, 2005). In this work the feasibility of utilizing multiwavelength fluorescence spectroscopy as a tool for monitoring the protein filtration utilizing their natural (intrinsic) fluorescence was investigated. The following sections provide information on the basic principles of fluorescence spectroscopy including the nature of the fluorescence, how it is influenced by environmental conditions and how it is measured.

2.6.1. *The Nature of Fluorescence*

The optical phenomenon, known as *fluorescence* is a three-stage process that occurs for certain types of molecules that are called *fluorophores* (Christensen *et al.*, 2006). This process is represented by a *Jablonski* Diagram as given in Figure 2.8. *First*, the *fluorophore* is excited from a ground energy state (singlet state, S_0) to either first

excited state (singlet state S_1) or second excited state (S_2) by the absorption of an external photon (i.e. light) with a frequency of ν and an energy of hv_A, which takes about 10^{-15} seconds. *Second*, some internal conversion of that absorbed energy occurs by a number of different ways including vibrational relaxation, quenching and energy transfer (Mycek and Pogue, 2003; Christensen *et al.*, 2006) which results in relaxation of the fluorophore into the lowest excited energy state (singlet state S_1). *Third*, the fluorophore then relaxes from the singlet-excited state S_1 to the ground state S0 by emission of a photon with energy hv_F at a longer wavelength relative to the absorbed photon. The energy of the emitted photon varies, depending on the S_0 ground state level it returns to.

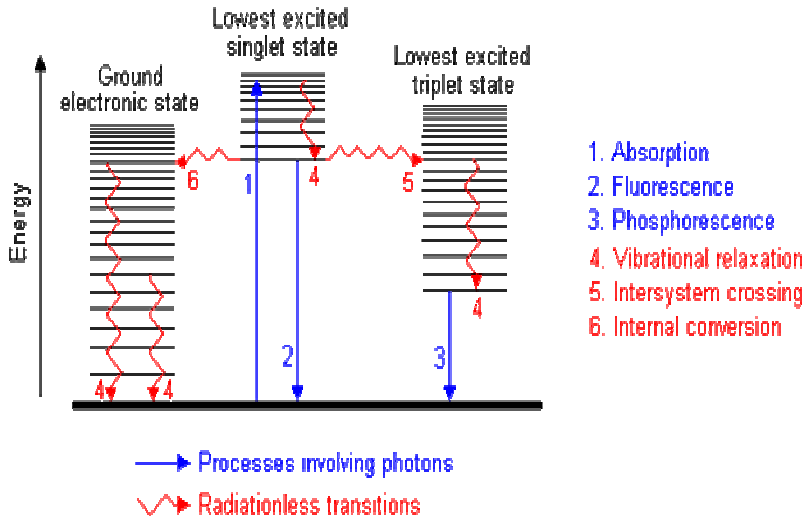


Figure 2.8: Jablonski diagram showing the energy levels and various processes in an electronically excited molecule (http://teaching.shu.ac.uk/hwb/chemistry/tutorials/)

If a fluorophore does not fluoresce upon absorption of the UV radiation it means that it must have lost its energy some other way. These processes are called *radiationless transfer of energy*. The difference in energy or wavelength between the absorbed and emitted photon is called the Stoke's shift (Mycek and Pogue, 2003; Christensen *et al.*, 2006).

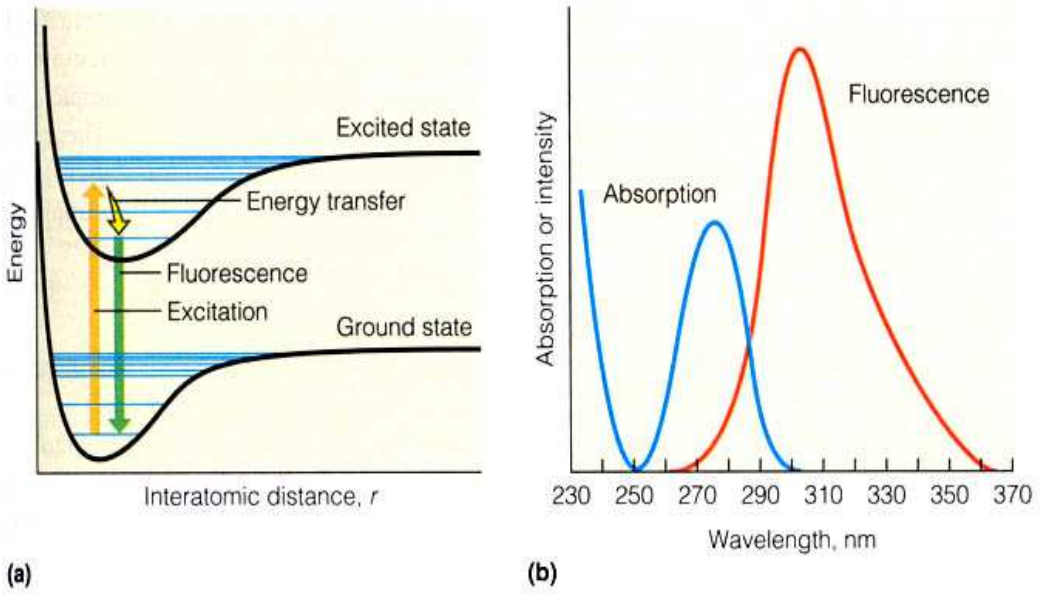


Figure 2.9: Excitation and emission spectra showing the energy levels and various processes in an electronically excited molecule
(http://teaching.shu.ac.uk/hwb/chemistry/tutorials/)

Each electronic state has several associated vibrational levels (Figure 2.9), which implies that fluorophores does not absorb excitation radiation at one wavelength but over a distribution of wavelengths corresponding to different vibrational transitions (Christensen *et al.*, 2006). Emission also occurs at several wavelengths as it may reach different vibrational levels in the electronic ground state. The result is that all fluorophores have specific and independent spectral excitation and emission profiles

characterizing their unique fluorescent properties. The distribution of wavelength-dependent intensity that causes fluorescence is known as the fluorescence excitation spectrum, and the distribution of wavelength-dependent intensity of emitted energy is known as the fluorescence emission spectrum (Christensen *et al.*, 2006).

2.6.2. *Fluorescence Scanning Modes*

In practice, fluorescence spectra can be recorded using a spectrofluorometer by irradiating a sample in three different scan modes (excitation, emission and synchronous scan). The three resulting types of spectra are normally presented on a wavelength scale calibrated in nanometers (nm). In excitation scan, the excitation wavelength is changed through the desired range of wavelengths while fixing the emission wavelength, which results in an excitation spectrum. An emission scan is made by fixing the excitation wavelength and changing the emission wavelength through the desired range of wavelengths, thereby recording an emission spectrum. A synchronous scan can be recorded by changing both the excitation and emission wavelengths in a stepwise manner with a constant offset between them. Such an offset or interval between the excitation and emission wavelengths is designated by the symbol $\Delta\lambda$. In a single synchronous spectrum, the intensity profile of the fluorescence is dependent on both the excitation and the emission wavelengths (Sharma and Schulman, 1999). Synchronous scanning mode yields fluorescence spectra with narrower and sharper peaks than emission or excitation spectra (Sharma and Schulman, 1999).

2.6.3. *Extrinsic and Intrinsic Fluorophores*

Not all molecules in nature display fluorescence properties upon excitation because the molecular structure and the environment dictate whether the compound is fluorescent or not. In other words, fluorescence is often exhibited by organic polyaromatic and heterocyclic compounds with rigid molecular skeletons that contains large conjugated π-electron systems (Shea**,** 1997; Christensen *et al.*, 2006). Fluorophores are divided into two classes: intrinsic and extrinsic fluorophores. Extrinsic fluorophores, also known as exogenous fluorophores, are dyes or fluorescent labels that are attached to the biological analyte of interest. Most commercially available fluorophores come with chemical groups that make labeling of biological species quite easy. Application of extrinsic fluorophores in monitoring ultrafiltration of protein mixtures was first demonstrated by Crespo *et al.* (1999), where β-LG and γ-globulin proteins were labeled with extrinsic fluorophores to make them easily spectrally resolvable and thus distinguishable. Intrinsic fluorophores are those substances that occur naturally in a variety of biological systems such as vitamins, NADH, NADPH, FAD and fluorescent amino acids (Vojinovic *et al.*, 2006). The use of such intrinsic fluorophores for monitoring recombinant protein production from mammalian cells upstream has been addressed in many studies (Hisiger & Jolicoeur, 2005). In contrast, using intrinsic fluorophores for monitoring membrane-based protein separation has not been reported before, which is the focus of this thesis.

The most common example of intrinsic fluorophores are the three aromatic amino acids, tryptophan, tyrosine and phenylalanine, that are primarily responsible for the

inherent fluorescence of proteins (Lakowicz, 1999). The aromatic amino acids are present in almost all proteins, enzymes and antibodies. These amino acid residues have distinct absorption and emission wavelengths and differ in the quantum yields as given in Table 2.1 and Figure 2.10.

Table 2.1. Fluorescent Characteristics of the Aromatic Amino Acids (Lakowicz, 1999).

Amino Acid	Excitation Wavelength (nm)	Absorptivity	Emission Wavelength (nm)	Fluorescence Quantum Yield
Tryptophan	280	5,600	348	0.20
Tyrosine	274	1,400	303	0.14
Phenylalanine	257	200	282	0.04

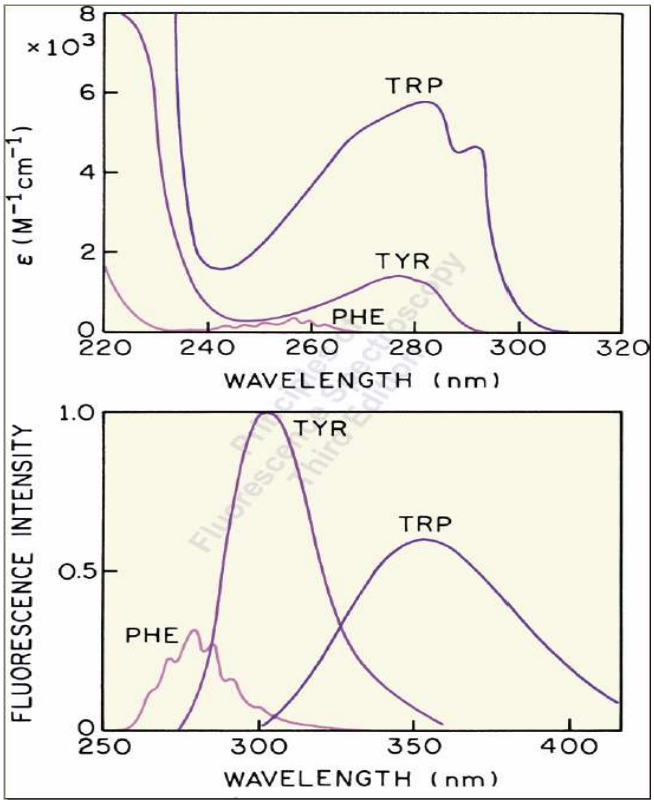


Figure 2.10: Absorption (top) and emission spectral profiles (bottom) for Tryptophane (Trp), Tyrosine (Tyr) and Phenylalanine (Phe) (taken from http://www.embl-hamburg.de/~tucker/JGS/fluorescence1.pdf)

As seen in Figure 2.10, tryptophan is much more fluorescent than either tyrosine or phenylalanine, accounting for more than 90% of the total fluorescence from proteins (Lakowicz, 1999). While tyrosine is less fluorescent than tryptophan, its contribution cannot be neglected as it is often present in large amounts in many proteins.

2.6.4. *Multidimensional Fluorescence Spectra*

Fluorescence is multi-dimensional containing a wealth of independent information related to the fluorophore type, its amount and its molecular environment. Multidimensional fluorescence signals recorded from a given multi-fluorophoric solution can be presented as *Excitation Emission Matrix (EEM)* or *Total Synchronous Fluorescence Spectra (TSFS).*

2.6.4.1 Excitation Emission Matrix

One approach to presenting three-dimensional fluorescence data is in the form of an Excitation Emission Matrix (EEM), which is a matrix of fluorescence intensities expressed as a function of excitation and emission wavelengths. Such a matrix can be collected by recoding a series of fluorescence emission scans at different excitation wavelengths. Spectrofluorometer software normally allows the user to select the range of excitation and emission wavelengths and the excitation wavelength increment between scans: the emission wavelength range measured will be the same for all scans, but the excitation wavelength is increased at a constant increment along the range of excitation wavelengths selected. Once defined, all scans can be recorded automatically

and without user intervention**.** Figure 2.11 shows a two-dimensional fluorescence scan visualized using a contour plot, where lines represent equal emission intensities.

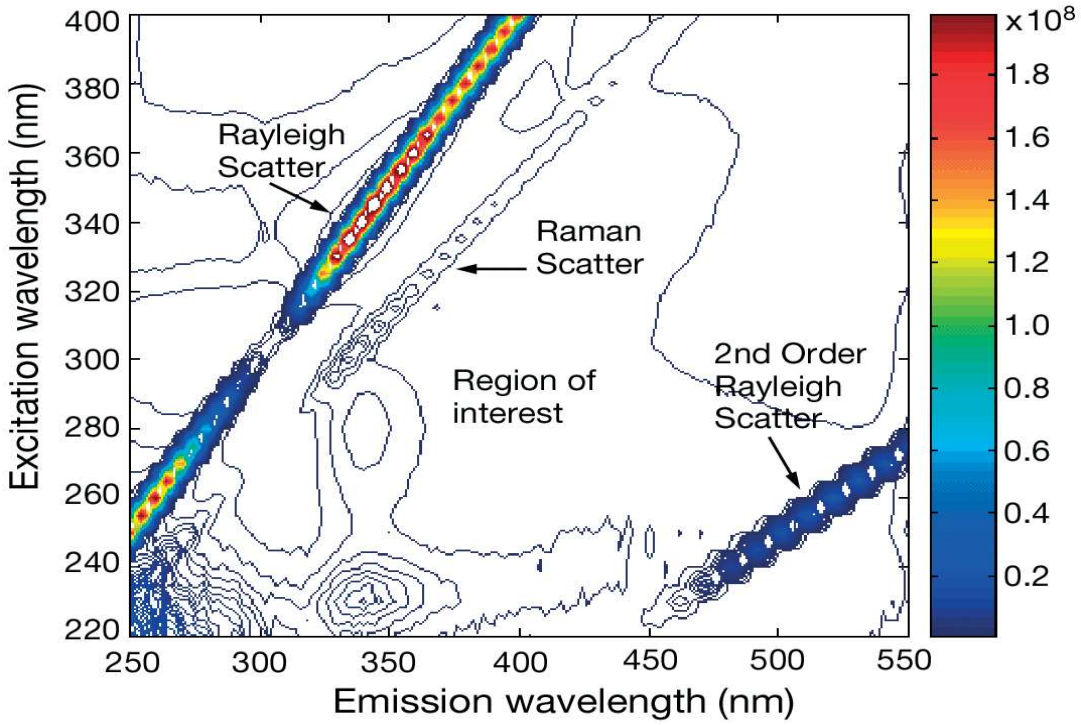


Figure 2.11 An excitation-emission matrix contour plot for a typical water sample containing low protein concentration generated using multiple emission scans, at excitation increments of 5 nm from 220 to 400 nm (as depicted in Hunt *et al*.,2007).

The complete excitation-emission matrix EEM is also known as a fluorescence landscape. A complete excitation-emission matrix EEM is a very sensitive and fast tool for the quantitative analysis of a biological solution consisting of multi-fluorophores such as a protein solution, where each fluorophore can be distinguished by its own spectral features. An EEM can be divided into three different zones (Figure 2.11) in terms of the hypothesized relevance to the protein-fractionation processes as described below:

i) **Fluorescence Spectral Zone** ($\lambda_{em} > \lambda_{ex}$)**:** The region of interest containing the intrinsic fluorescence spectral fingerprints of the proteins is located in the window between excitation range (220-320 nm) and the emission wavelength (250-450 nm) (Figure 2.11). The protein intrinsic fluorescence is identified by two distinct peaks located at approximately 220/330 nm and 275/330nm (excitation/emission) that are attributed mostly to both tyrosine and tryptophan fluorescence. These two fluorophores each have two excitation wavelengths with tryptophan at 215-220 nm and 275-280 nm and tyrosine at 220-225 nm and 275-280 nm as seen for pure proteins, alone or in combination, and for real protein mixtures (Mayer *et al.*, 1999). The relevance of this to protein-fractionation is addressed in Chapter 5.

ii) **Light Scattering Spectral Zones** ($\lambda_{em} = \lambda_{ex}$),($\lambda_{em} = 2\lambda_{ex}$)**:** While scanning and recording an EEM, signal components may arise from optical phenomena other than fluorescence such as light scattering. Scattered light appears both in clear and in turbid solutions and it has a substantial effect on fluorescence measurements. In clear solutions, there is Rayleigh scatter, second-order Rayleigh scatter and the Raman scatter. In opaque solutions there is, in addition, the Tydnal scatter or scattering by large particles. In the case of 3D-fluorescence (Figure 2.11), an EEM will typically have areas that are dominated by each type of scatter. These areas are represented by diagonal lines in the landscapes as seen in Figure 2.11. Raman scattering (also called inelastic) arises from the interaction of the exciting incident light with the solvent molecules producing an inelastic scattering of photons with lower energy (i.e. longer wavelength) relative to the exciting ones

(Deshpande.,2001). Such scattering is seen at low concentrations of fluorophores when the instrument is set at its greatest sensitivity. Rayleigh scattering usually occur at the emission wavelength equal to the excitation wavelength ($\lambda_{em} = \lambda_{ex}$) while second order Rayleigh scattering appears at the emission wavelength twice as long as the excitation wavelength ($\lambda_{em} = 2\lambda_{ex}$) as seen in Figure 2.11. Chapter 6 will describe how the scattering component in an EEM can be utilized to provide quantitative information about protein aggregation and its affect on membrane separation performance (Deshpande., 2001).

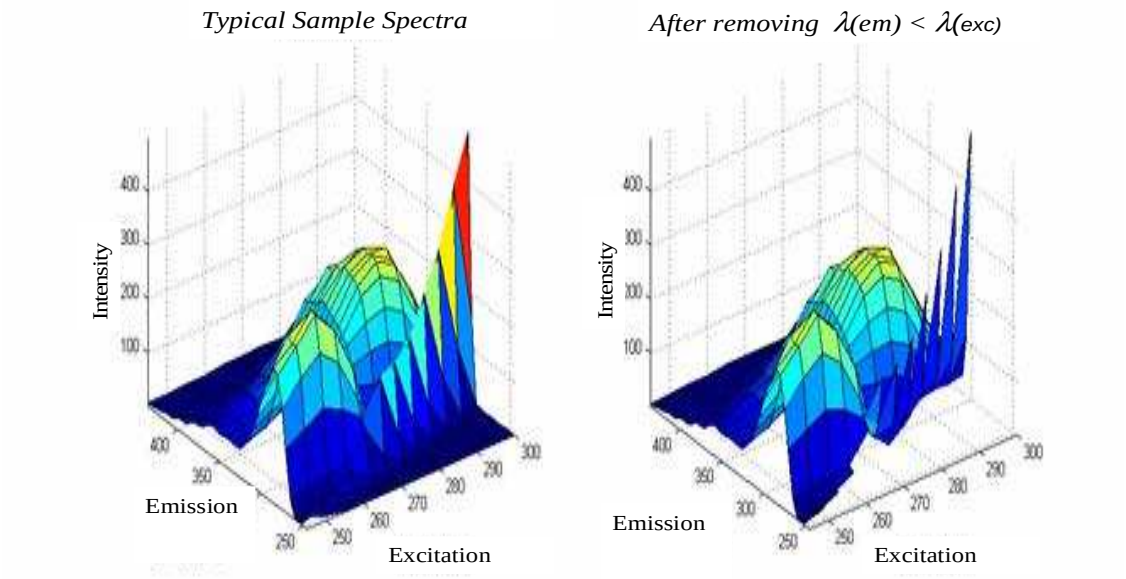


Figure 2.12: Subtraction of the triangular-shaped region where the emission wavelength is less than the excitation wavelength (upper left-hand side)(as depicted in Bro *et al.*, 2002)

iii) **Triangular-shaped region** ($\lambda_{em} < \lambda_{ex}$): The triangular-shaped region is where the emission wavelength is less than the excitation wavelength (upper left-hand side) and is considered to be non-informative since it contains physically impossible data points. It is usually removed by subtracting the background or by setting the

intensity values to zero as represented by Fig 2.12. The non-informative part of the EEM can be selectively avoided during data acquisition by using a synchronous scanning mode with ($\Delta\lambda > 0$) instead of the conventional excitation-emission scanning mode which will be demonstrated in Chapter 5.

2.6.4.2. Total Synchronous Fluorescence Spectra (TSFS)

Total Synchronous Fluorescence Spectra (TSFS) is another form of multidimensional fluorescence data where fluorescence intensities are expressed as a function of excitation wavelength and the interval between the excitation and emission wavelengths ($\Delta\lambda$). Such a matrix can be collected by recording a series of synchronous scans at different wavelength intervals. The contour profiles generated for EEM and TSFS are different (Patra and Mishra, 2002) due to the difference in the way of arranging the data structure; however, both EEM and TSFS should contain the same information content as long as the scanning mode covers the full spectral range of interest (Patra and Mishra, 2002).

2.6.5. *Quantitative Analysis of the Fluorescence Spectra*

The fluorescence intensity (*Fl*) emitted by a given fluorophore *i* at a particular excitation (λ_{exci}) and emission wavelength (λ_{em}) is expressed as a function of the absorption coefficient and the quantum yield of the fluorophore as given by Equation (2-5):

$$Fl_i(\lambda_{ex},\lambda_{em}) = k.I_o(\lambda_{ex}).\mu_{ai}(\lambda_{ex})\phi_i(\lambda_{em}) \qquad \textbf{(2-5)}$$

where *k* is a proportionality constant that is related to instrumental parameters (the detector collection efficiency, the path length, the sample geometry) and $\phi(\lambda_{em})$ is the quantum yield of the fluorophore at the emission wavelength (λ_{em}). I_o is the intensity of the incident light (Ramanujam, 2000).

The absorption coefficient $\mu_{ai}(\lambda_{ex})$ is a linear function of the extinction coefficient, denoted by $\varepsilon(\lambda_{ex})$ and the concentration of the fluorophore, denoted by C*i* as given by Equation 2-6:

$$\mu_a(\lambda_{ex}) = 2.303\varepsilon_i(\lambda_{ex}) \cdot C_i \qquad \textbf{(2-6)}$$

By combining Equations (2-5) and (2-6), the fluorescence intensity (*Fl*) at a particular excitation (λ_{exc}) and emission wavelength (λ_{em}) for a dilute solution containing a fluorophore can be described by the following equation (Ramanujam, 2000; Christensen *et al.*, 2006):

$$Fl_i(\lambda_{ex}, \lambda_{em}) = 2.303k.I_o(\lambda_{ex}).\varepsilon_i(\lambda_{ex})\phi_i(\lambda_{em})C_i \qquad \textbf{(2-7)}$$

Equation (2-7) implies that the relationship between the fluorescence signal and the concentration of the fluorophore is approximately linear for dilute solutions. According to Equation (2-7), the intensity of the fluorescence depends on the concentration, the molar absorptivity, and the quantum yield of the fluorophore (Christensen *et al.*, 2006). Deviation from linearity occurs in concentrated solutions due to self-quenching and inner-filter effects that are explained in section (2.6.6). The fluorescence spectra are additive in mixtures for extremely dilute solutions, i.e. the overall fluorescence spectra acquired for a given mixture can be expressed as a linear

contribution of all individual spectra from inherent fluorophores in the appropriate proportions as given in Equation (2-8) (Ramanujam,2000;. Christensen *et al.*, 2006).

$$Fl(\lambda_{ex},\lambda_{em}) = \sum_{i=1}^{n} Fl_i(\lambda_{ex},\lambda_{em}) \quad \text{(2-8)}$$

From Eq (2-8) and Eq (2-7),

$$Fl\ (\lambda_{ex},\lambda_{em}) = 2.303k.I_o(\lambda_{ex})\sum_{i=1}^{n}\varepsilon_i(\lambda_{ex}).\phi_i(\lambda_{ex}).C_i \quad \text{(2-9)}$$

where Fl_i denotes the fluorescence signal emitted by a fluorophore *i*. In complex mixtures, such as biological samples, the fluorescence may not be additive due to quenching phenomena and interactions with the molecular environment of the fluorophore (Christensen *et al.*, 2006).

2.6.6. *Factors Affecting Fluorescence Measurements*

This section is an overview of several factors that can influence fluorescence measurements for biological samples. These factors are related to the composition of the biological sample as well as the concentration and the molecular environment that contribute to the complexity and variability of fluorescence measurements (Christensen *et al.*, 2006). According to Equation (2-9), the intensity of the fluorescence depends on the concentration, the molar absorptivity, and the quantum yield of the fluorophore (Christensen *et al.*, 2006). The effect of quenching, inner-filter effects, the molecular environment of fluorophores and the light scattering phenomena will be addressed in the following subsections.

2.6.6.1. Quenching

Spectrofluorometer readings do not necessarily change in step with the known concentration change of the fluorophore. The factor that may be responsible for this is called "concentration quenching", or sometimes just "quenching". Fluorescence quenching can be referred to any process that reduces the fluorescence intensity of a sample. Such a phenomenon occurs when the excited fluorophore returns to the ground state without emitting a photon (Christensen *et al.*, 2006). A variety of processes can result in quenching, such as energy transfer, complex-formation and collisional quenching where the excited state of the fluorophore can be deactivated by contact with other molecules or by either intra- or intermolecular interactions.

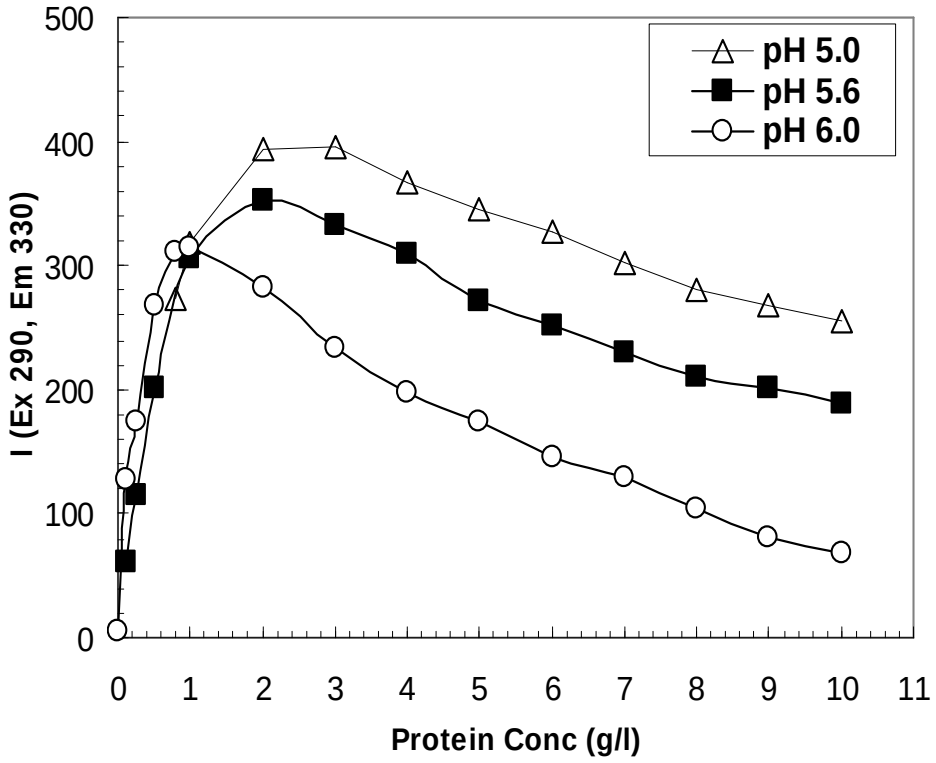


Figure 2.13 Concentration quenching effect on emission intensity at 330 nm observed for the whey protein isolate solution.

An example of quenching due to intermolecular interactions is called "concentration quenching" or “self-quenching”. At low sample concentrations, the fluorescence intensity is directly proportional to the sample concentration; however, as the concentration increases beyond the linear range for fluorescence, the fluorescence intensity decreases with concentration (Figure 2.13).

With increasing protein concentration, the proximity of the protein macromolecules allows the light emitted by the tryptophan residues of the protein to be re-absorbed by the same fluorophore (i.e. tryptophan) in the adjacent protein macromolecule and each time the light is re-absorbed, there is a chance for the energy to be dissipated non-radiatively.

Concentration quenching does not only affect the magnitude of fluorescence intensity but also the overall shape of the fluorescence spectra as seen in Figure 2.14. In Chapter 5, spectral differences between the retentate with high protein concentrations and the permeate with low protein concentrations can most likely attributed to self-quenching that is related to the intermolecular interactions (Christensen *et al.*, 2006).

In addition to quenching by intermolecular interactions, quenching can occur through intramolecular interactions. Interaction of adjacent fluorophores occur within the protein macromolecule itself where the fluorescence from tyrosine can be easily quenched by the presence of nearby tryptophan moieties via resonance energy transfer, as well as by ionization of its aromatic hydroxyl group (Christensen *et al.*, 2006).

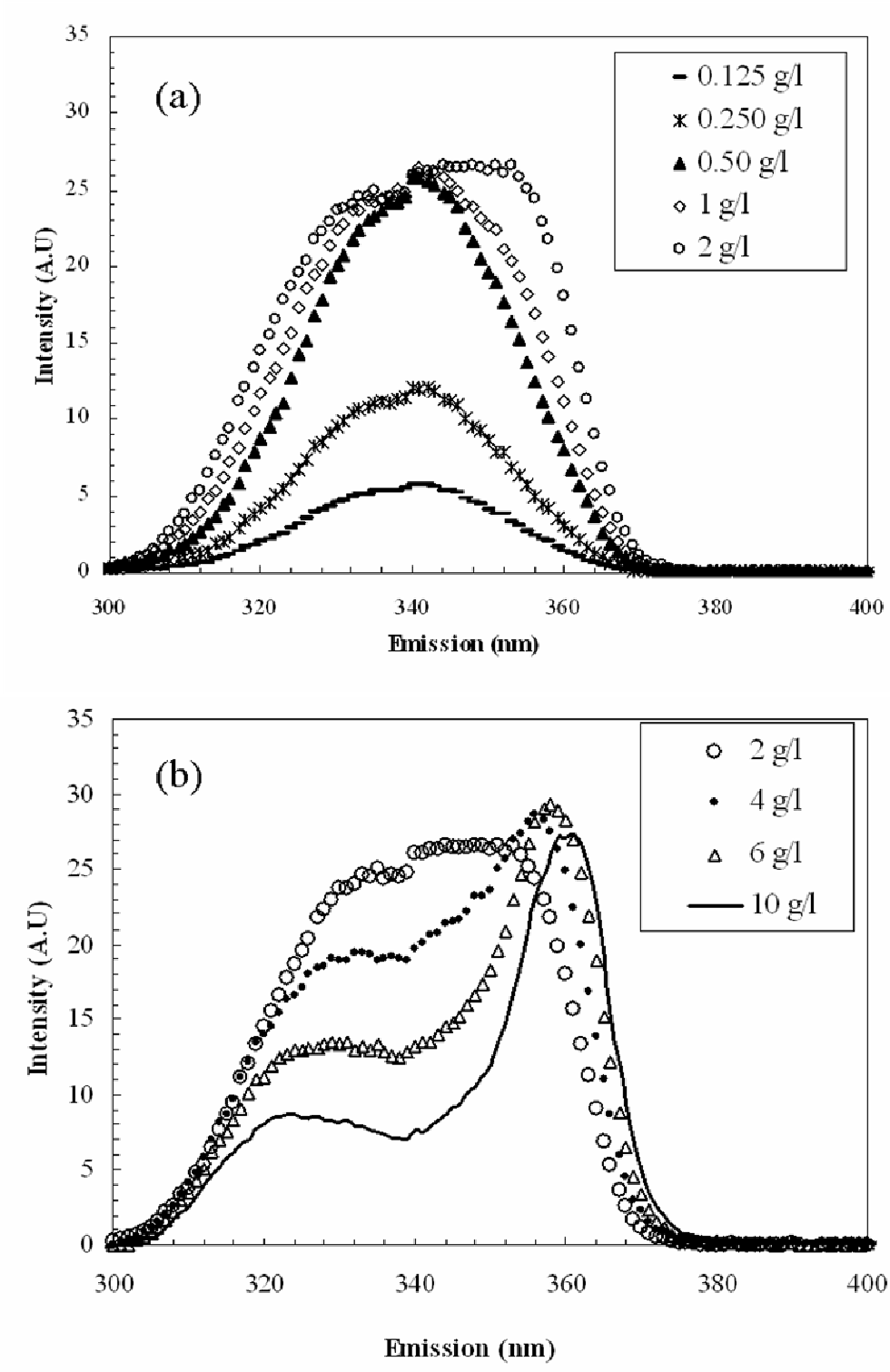


Figure 2.14 Synchronous spectra at Δλ =60 nm collected for whey protein isolate at different concentrations from 0.125 g/l to 2 g/l (a) and from 2 g/l to 10 g/l (b).

2.6.6.2. Molecular Environment.

Native protein fluorescence is the result of intrinsic fluorophores in proteins that consist of hydrophobic amino acid side chains as illustrated in Figure 2.15. Most of the native protein fluorescence is generally due to the amino acids tryptophan and tyrosine.

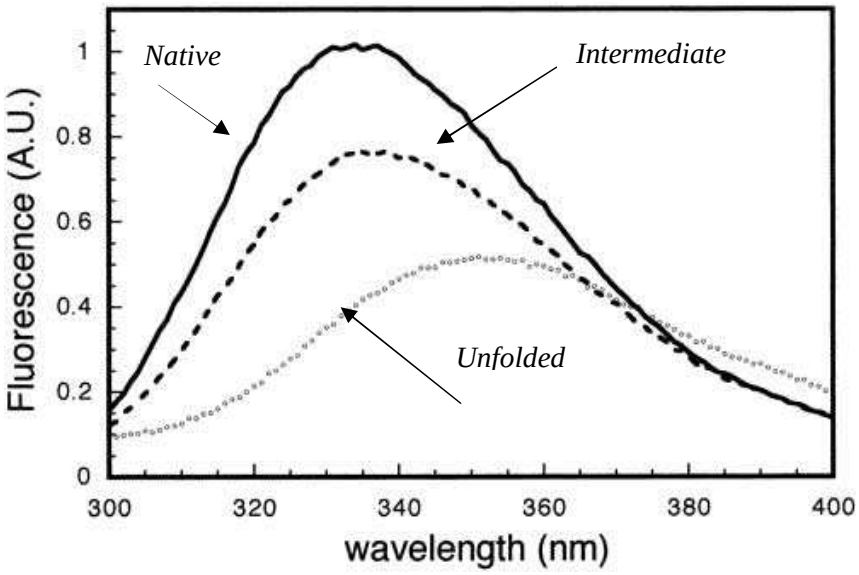


Figure 2.15 Fluorescence spectroscopy as a tool for tracking protein conformational changes (www.soc.nii.ac.jp/jbiochem).

These hydrophobic amino acid residues are so sensitive to the local environment that it is possible to see changes in emission spectra when they are exposed to the solvent or bulk phase and therefore information about protein conformation, subunit association and denaturation can be obtained (Lakowicz, 1999). In this respect, intrinsic fluorescence measurements of proteins have been used to study the effects of the protein microenvironment, pH, ionic strength, and temperature on their association properties, degree of unfolding and aggregation behavior (Lakowicz, 1999). Multi-wavelength fluorescence spectroscopy was employed in this study in Chapter 3 and

Chapter 4 for simultaneous determination of whey proteins solubility as well as their aggregation behavior induced by heat treatment at different conditions including pH, ionic strength and temperature.

2.6.6.3 Light Scattering

Rayleigh scattering can constitute a significant interference to fluorescence emission from fluorophores with a small Stoke's shift (Christensen *et al*., 2006). Since interference from Rayleigh scattering cannot be avoided or eliminated, mathematical corrections of the fluorescence signal can be performed instead by addressing the scatter in the modeling and analysis of the 2-D fluorescence data (Christensen *et al*., 2006).

2.6.7. Why Chemometrics?

The intrinsic fluorescence of proteins can be easily distinguished using fingerprints of the fluorescent amino acid residues. Although EEM allows detection of the presence or the absence of protein in a biological mixture, discriminating between different proteins or identification of the type of protein in a biological mixture is challenging because of the subtle differences in their fluorescence spectra. Fluorescence signals are complex as the spectra are the result of interferences, scatter and overlapping signals. Furthermore, the fluorescence intensity of a given compound is influenced by environmental variables like pH, ionic strength, total concentration and physical factors like the inner filter effects and the energy transfer processes. Therefore, before the measured intrinsic fluorescence can be related to a filtration process an awareness of the factors that could affect the measured fluorescence signal is necessary. To tackle

the complex nature of the contributions to the of fluorescence signal, chemometrics is proposed in this thesis.

2.7. Chemometrics

Monitoring, controlling and optimizing membrane-based filtration processes of complex biological solutions is difficult to achieve practically because all of the components of a complex biological solution interfere with the performance of the membrane filtration process (Darnon *et al*., 2002). During filtration of a complex biological solution, tracking transient changes in product yield and the selectivity of separation cannot be performed without information about the transport of the various feed components through the membrane. It will be demonstrated in Chapter 5 that such information can be extracted from multi-wavelength fluorescence spectra collected for the feed, permeate and retentate. Extraction of information from spectra can be achieved using mathematical analysis tools known as chemometric tools. Since not all fluorescence data collected is meaningful, consisting mainly of both unwanted variations (i.e. noise) and information relevant to the process (Eriksson *et al.,* 2002), chemometric tools can be used to reveal the information in these large data sets. The field of chemometrics was found by Bruce Kowlaski and Svante Wold in the early 1970s (Eriksson *et al.,* 2001). Chemometrics has been defined as "the chemical discipline that uses mathematics, statistics and formal logic (a) to design or select optimal experimental procedures; (b) to provide maximum chemical information by analyzing chemical data; and (c) to obtain knowledge about chemical systems". The name chemometrics is a combination of the two suffixes: chemo (i.e. chemical) and

metric (i.e. measurements) because it deals with extracting information from chemical data by means of multivariate data analysis (Eriksson *et al.,* 2002; Wiberg, 2004). In chemometrics, turning these large data sets into knowledge about the process is performed by the use of a mathematical model. Various examples of the application of chemometrics for extracting relevant information from fluorescence spectra can be found in literature. Hagedorn *et al.* (2003) evaluated spectrofluorometry as a tool for monitoring a fermentation process, including substrate, biomass and product predictions by making use of multivariate calibration models for extracting the underlying variations in the multi-wavelength fluorescence spectra that were most correlated with the important process variables in the process. The next section gives a brief description of the multivariate techniques that are most widely used to regress spectral data.

2.7.1. *Multivariate Nature of Fluorescence Data*

Multivariate data analysis methods are part of the chemometric techniques that are used to analyze data sets consisting of multiple variables measured from many samples. Chemical data can often be arranged as a table, a data matrix as given by Equation 2-10 that contains measurements of *m* variables on *n* objects. Typical chemical objects are analytical samples. $\boldsymbol{X}$ is usually used to denote the data, *i* denotes the index for objects or samples and *j* denotes the index for variables.

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \quad \textbf{(2-10)}$$

In the context of fluorescence, x_{ij} represents the intensity value of the *j*th spectral variable (excitation-emission wavelength pair) for the *i* th case (i.e. sample).

For example,$X_1=[x_{11}\ x_{12}\ x_{13}\ldots x_{1m}]$ is the row vector containing the fluorescence intensities measured at m excitation/emission wavelength pairs for the first sample. $X_2=[x_{21}\ x_{22}\ x_{23}\ldots x_{2m}]$ is the row vector containing the fluorescence intensities measured at *m* excitation/emission wavelength pairs for the second sample and so on.

Multivariate methods can be classified into two categories: (i) Multivariate methods that find the relationship between x and y variables are generally called regression methods such as Partial Least Squares (PLS). (ii) Multivariate methods that are used for explanatory analysis and survey of the **X** data, finding trends, groups and outliers. The next section gives a brief review of the PLS regression method. Multivariate data analysis tools, such as principal component analysis and Partial least squares (PLS), are considered to be powerful for extracting high quality information from the less resolved high-dimensional spectroscopic data. It will be shown in Chapter 5 that even extremely subtle spectral differences between three different whey proteins can be distinguished using chemometrics. Multivariate methods are intrinsically more robust and accurate with respect to peak shifts and instrumental noise than univariate

methods, because such multivariate methods use area under the whole curves (called scores in chemometrics) rather than just single wavelength intensity for an excitation/emission pair (Christensen *et al*., 2006). Such areas are much less influenced by moderate peak shifts and instrumental noise than single intensity values. Additionally, robustness is also obtained from the general noise reduction obtained from using the above mentioned areas. The method for multivariate data calibration employed in Chapters 3, 4, 5, and 6 was PLS.

2.7.2. *Calibration and Regression Analysis in Fluorescence Spectroscopy.*

Calibration is one of the most important tasks in quantitative spectrochemical analysis The term calibration model can be generally defined as the process of deriving a mathematical relationship between available process measurements **X** and quantitative information **Y** that allows predicting to the best possible degree unknown quantitative information about the process from future **X** measurements. The underlying model for the relationship between the measured variable x and a dependent variable y is generally presented by the mathematical function in Eq (2-11):

$$Y = f(X) \qquad \textbf{(2-11)}$$

The application of different classes of regression methods in spectroscopy can be found in the literature. Regression models implemented in spectroscopy range from simple linear univariate regression (with one x predictor variable related to one y response variable) to complex multivariate regression (with so many x predictor variables related simultaneously to one or more y response variables) that require the use of a software package. The purpose of this section is to compare different classes of

regression models and to explore their scope of applicability to quantitative spectral analysis of multi-component systems. The emphasis here is confined to cases in which the relationship between the response y variable and a set of x-variables (i.e. spectroscopic measurements) is linear or close to linear. Herein, the existing regression methods have been arranged according to their level of complexity (i.e. the number of variables involved) as well as their scope of applicability to meet the stated objectives of this thesis. Mathematical formulations, advantages and limitations of these models have been discussed in order to understand the connection between them

2.7.2.1 Univariate Calibtation

Assume that we want to build a calibration model to determine the concentration of a specific protein, β-LG in the permeate, retentate and feed lines during ultrafiltration of a whey protein isolate solution. Figure 2.16 shows two-dimensional fluorescence spectra acquired in synchronous scanning mode for whey protein isolate solution. This fluorescence landscape consists of 1660 distinct fluorescence measurements at different combinations of excitation/emission wavelengths. More generally, each fluorescence landscape could be thought of as a set of 1660 different univariate measurements (Bro, 2003). The simplest form of calibration is a univariate calibration model, which can be built by choosing one out of these 1660 distinct measurements. A typical choice could be to select a wavelength corresponding to a peak maximum related to the specific protein of interest. In this example, the signal at an excitation of 295 nm and an emission at 330 nm is chosen which corresponds to the maximum emission peak of tryptophan and so a linear univariate regression model can be

established between one x-variable (i.e. intensity measurement at 295nm/330nm) and one- y-variable (i.e., the concentration of β-LG in the solution) as given by Equation 2-12.

$$y = \beta_1 x + \beta_0 \quad \textbf{(2-12)}$$

Where β_1 is the regression coefficient and β_o is the intercept of the linear correlation. The univariate approach is valid provided that the following conditions are satisfied (Torgrip., 2003):

- *The linearity condition:* The instrument's response must be linearly correlated with the measured feature. Deviations in the absorption coefficient can occur at high concentrations due to electrostatic interactions between molecules in close proximity, as referred to as analyte association.
- *The interferent condition*. The instrument response must not exhibit any wavelength shift for the measured constituent. Fluorescence red-shift can occur as a result of increasing concentration.
- *The selectivity condition:* The spectral peak of interest must be fully separated from spectral peaks belonging to other components in the mixture..
- *The noise:* The measurement process will always yield noise in the measured data. The structure of the noise varies, depending on the analytical system involved.
- *The scatter condition:* There must be no scattering of light due to particulates in the sample.

Univariate calibration suffers from some major disadvantages that make it unsuitable for application to real processes (Bro, 2003). Accuracy of the univariate calibration

model is only possible if the analyte of interest (e.g. β-LG) contributes to the measured signal

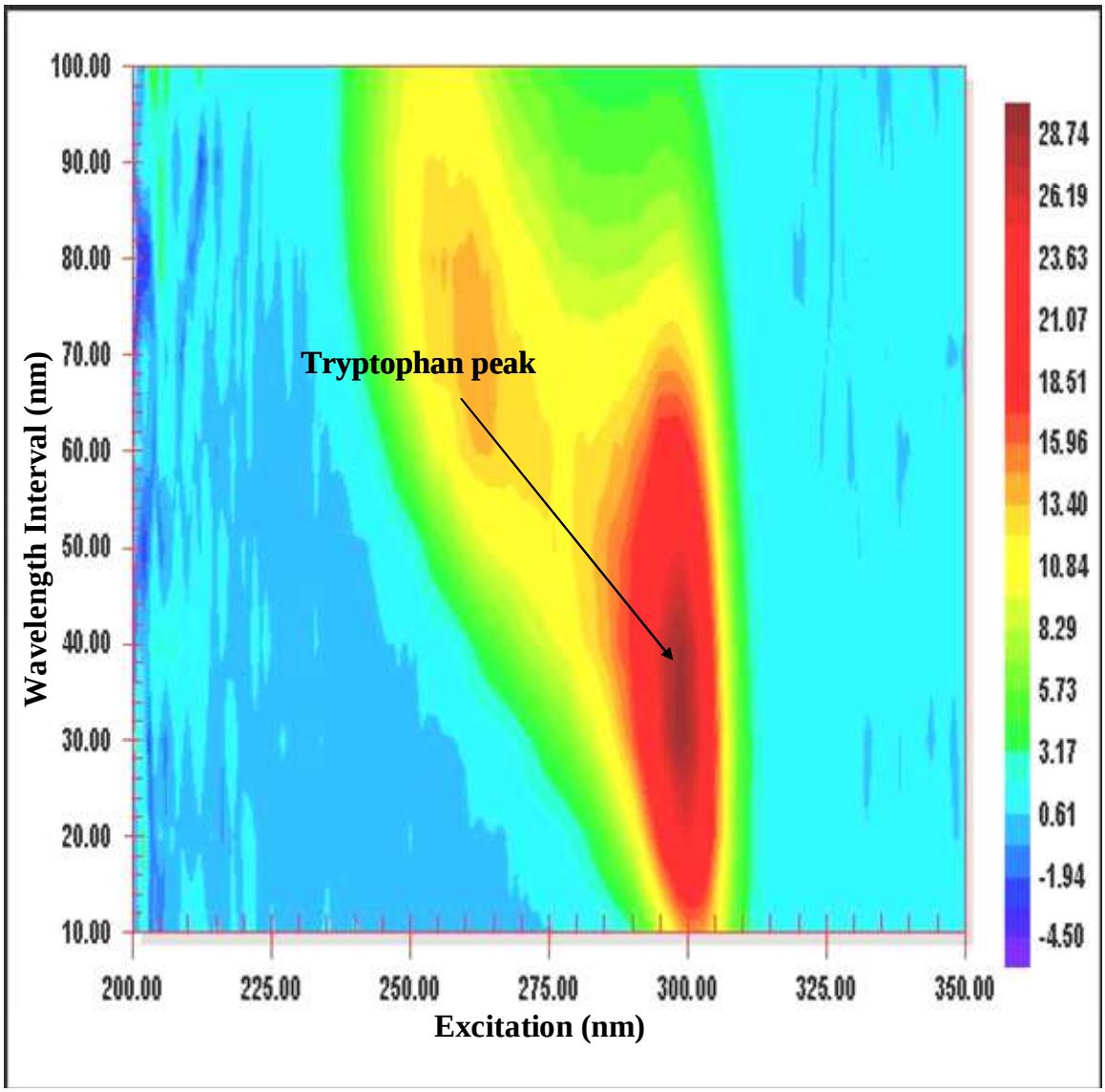


Figure 2.16: Fluorescence spectra acquired in synchronous scanning mode for whey protein isolate solution at room temperature and pH of 4.5.

In this work, although β-LG constitutes 75% of the whey protein isolate, other proteins that exist in smaller amounts (e.g. α-LA, BSA and IgG) may have significant

contributions to the measured fluorescence signal due to their tryptophan content. If other proteins in the solution contribute to the measured signal, the results will be biased (Bro, 2003). Therefore, instead of using one out of the 1660 measurements, it will be more reliable to use all the available data in the fluorescence landscape for calibration (Bro, 2003), which is known as multivariate calibration. In multivariate calibration, use of many x-variables automatically corrects for each other's selectivity and thus the x-variables used do not need to be totally selective. High precision of multivariate calibration can be generally achieved as long as the relationship between x and y variables is linear. Multivariate calibration is also generally more robust and less sensitive to small changes in the experimental or instrumental parameters such as pH, temperature or lamp intensity (Wiberg, 2004).

2.7.2.2 Multivariate Calibration

As discussed in the previous section, it would be necessary to extend the regression to include cases in which several variables contribute to the measured response y. In the simplest example, the dependent response is expressed as a function of two such independent variables x_1 and x_2.

$$y = \beta_o + \beta_1 x_1 + \beta_2 x_2 \qquad \textbf{(2-13)}$$

Again β_o is the intercept on the y-axis, β_1 and β_2 are the partial regression coefficients.

The following example illustrates the usefulness of multivariate calibration for quantitative analysis of fluorescence spectra. Figure 2.17 presents synchronous

fluorescence spectra acquired for β-LG, α-LA and BSA protein solutions by irradiating the samples at Δλ=100 nm. For quantitative analysis of β-LG, measurements at a single wavelength would be adequate if no interfering species are present. In the presence of other absorbing species such as BSA and α-LA., however, more measurements would be required in order to account for such interferences in the regression model. For example, by looking at Figure 2.17, it can be seen that the fluorescence signal attributed by BSA is more than two folds greater than that produced from β-LG at (λ_{ex}= 270 nm, λ_{em} =350 nm).

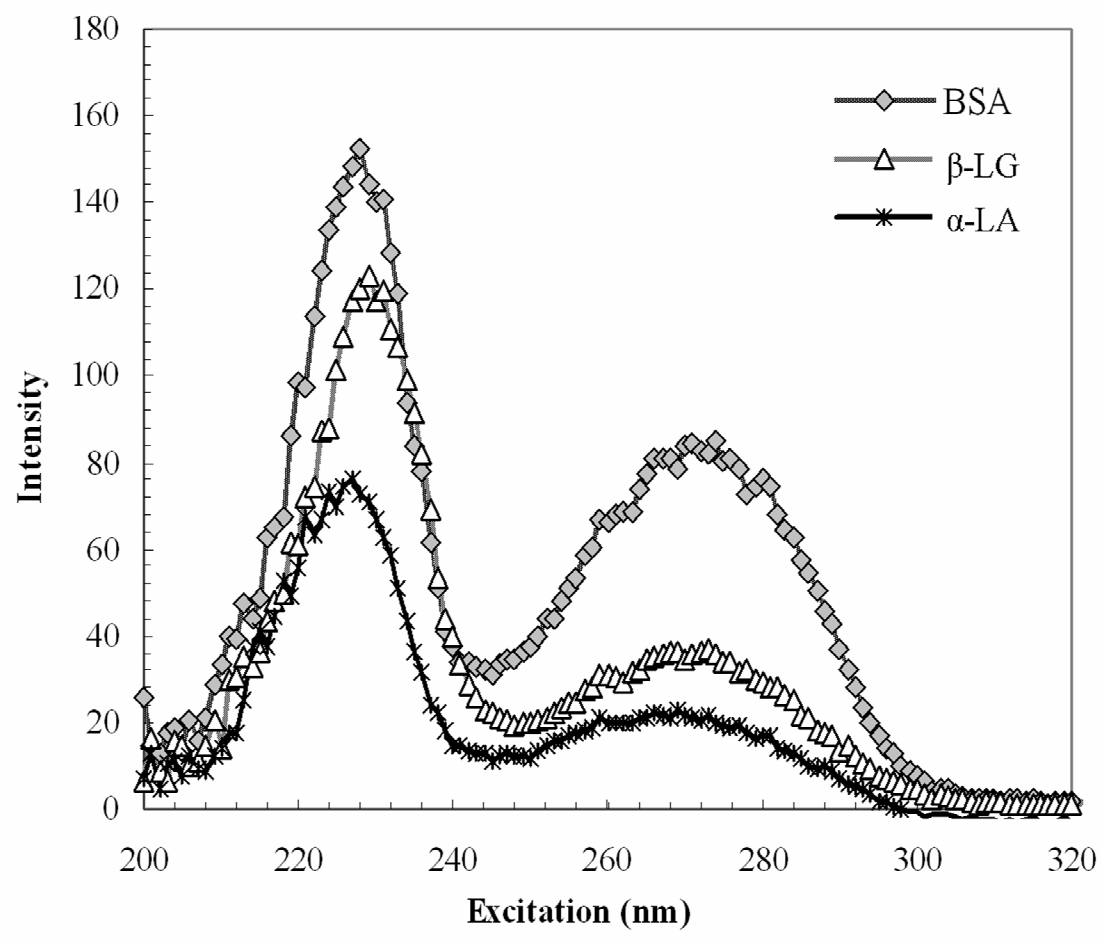


Figure 2.17 Synchronous fluorescence excitation spectra of β-LG, α-LA and BSA protein solutions acquired at Δλ=100 nm

Meanwhile, fluorescence signals produced from BSA and β-LG are relatively comparable at (λ_{ex}=220 nm, λ_{em} =350 nm). This implies that including measurements at (λ_{ex}= 270 nm, λ_{em} =350 nm) in the calibration model for predicting β-LG concentration could serve to compensate for the fluorescence due to BSA since this excitation-emission pair is in the spectral region of BSA with little interference from β-LG. In order to develop a reliable and robust calibration model, other information from the whole spectral data could be included. Formally, the model for multivariate regression , given *n* observations, is:

$$y_i = \beta_o + \sum_{j=1}^{m} \beta_j x_{ij} + \varepsilon_i \qquad \text{for i = 1,2, ... } n. \qquad \textbf{(2-14)}$$

Where y is the concentration of some analyte (in our case α-LA, β-LG or BSA), x_1 is the measured fluorescence intensity at *j* specific combination of excitation and emission wavelengths, β_j is the regression coefficient or weight associated with measurement at a given combination of excitation-emission pair. For a complete spectrum, *m* may take on values of several hundreds depending on the resolution of the fluorescence scan. Multivariate calibration can be expressed in matrix notation as given in Eq 2-15:

$$Y = X.B + E \qquad \textbf{(2-15)}$$

In which ***E*** is the matrix containing the residuals (variations not described by the model). Mathematical methods for achieving multivariate calibration between an ***X*** and

a ***Y*** matrix are generally classified into two categories: Least squares modeling methods and factor based methods.

1. Least Squares Modeling:

Least squares methods attempt to model the relationship between ***X*** and ***Y*** matrices by finding the regression coefficients that minimize the sum of the squares error SSE as given by Equation (2-16)

$$SSE = \min\left(\left[Y - X\beta\right]^2\right) \quad \textbf{(2-16)}$$

The least squares estimates of the regression coefficients are given by

$$\beta = (X'X)^{-1}X'Y \quad \textbf{(2-17)}$$

The desired property y for a given sample can be predicted when multiplying the regression coefficients by the spectrum acquired for that sample. The major limitation of ordinary least squares regression is related to the nature of finding the inverse for $(X'X)$ in the solution of Equation (2-17). When the number of x-variables exceeds the number of samples or/and when there is a high degree of collinearity among the variables, the estimated regression coefficients may be unreliable. This mathematically implies that high collinearity between the data at different wavelengths in the ***X*** matrix comprising the spectroscopic measurements could result in an $(X'X)$ matrix that is singular or close to singular. Consequently, the regression coefficients become large and this makes the model more sensitive to instrumental noise in ***X***. This, in turn, causes degradation of the model performance (Torgrip, 2003).

2. Factor-based regression methods-Chemometrics

Factor based regression methods such as Principal components regression PCR and Partial Least Squares Regression PLSR handle the problem of collinearity by compressing the ***X*** matrix containing *p* spectral variables for *n* samples into a small representative set of new variables which summarizes most of the spectral information in the original ***X*** matrix. Such set of variables are linear combination of the original variables in the data set and orthogonal (completely uncorrelated) to one another. These new variables are called Principal Components, Latent variables or eigenvectors. The use and applications of principal components in regression analysis has been extensively reported in the literature (Torgrip, 2003).. Mathematically, the matrix ***X*** is approximated in terms of the product of two smaller matrices; ***T*** and ***P'***. These two matrices capture the essential data patterns in ***X*** as given in Equation (2-18).

$$X = TP' + E \qquad \textbf{(2-18)}$$

, where ***T*** and ***P*** are referred to as the score and the loading matrices respectively and ***E*** is the modeling error matrix or the matrix of residuals. The principal component scores of the first, second, third components ($\mathbf{t_1}$, $\mathbf{t_2}$, $\mathbf{t_3}$, …) are columns of the score matrix ***T***. As mentioned before, these scores may be considered as new variables, which summarize the original ones. In their derivation, the scores are sorted in descending importance ($\mathbf{t_1}$ explains more variation than $\mathbf{t_2}$, $\mathbf{t_2}$ explains more variation than $\mathbf{t_3}$, and so on). In general, 2 to 5 principal components are usually sufficient to approximate a large data set of spectrochemical measurements.

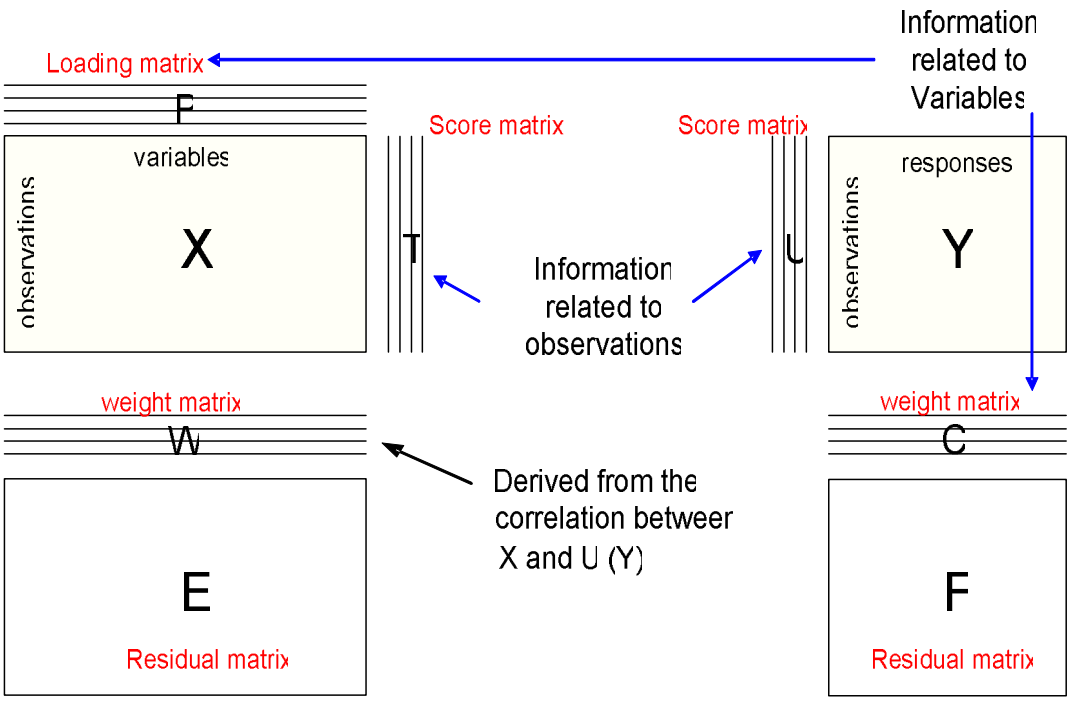


Figure 2-18 The matrix relationship in PLS regression. The score, weight and loading matrices are derived during the development of the PLS regression model. Source: Eriksson *et al.* (2001).

The meaning of the scores is given by the loadings. The loading vectors of the first, second, third, components ($\mathbf{p_1}$, $\mathbf{p_2}$, $\mathbf{p_3}$,..) build up the loading matrix ***P***. ***P*** is the matrix of loadings that express the relationship between ***T*** and ***X***. The original spectra are reconstructed when the scores are multiplied by the loading vectors, and the results summed, as described by Equation (2-19).

$$x_{ij} = \sum_{a=1}^{A} t_{ia} p'_{aj} \qquad \textbf{(2-19)}$$

where $\mathbf{p_{1j}}$ is the loading of the variable x_j in the loading vector $\mathbf{p}_1$ of the first latent variable. Using the matrix ***T*** of this smaller number of principal components, rather than the entire ***X*** matrix in Eq (2-17), is known as principal components regression

(PCR). Partial Least Squares Regression PLSR is more superior than PCR. The method uses two outer relations and one inner relation. The outer relation describes the decomposition of ***X*** and ***Y*** matrices.

$$X = TP' + E = \sum_{a=1}^{A} t_a p_a' + E \qquad \textbf{(2-20)}$$

$$Y = UC' + F = \sum_{a=1}^{A} u_a c_a' + F \qquad \textbf{(2-21)}$$

The inner relation is written as

$$U = TW \qquad \textbf{(2-22)}$$

In essence the inner relation is a least squares fit between the ***X*** block scores and the ***Y*** block scores. This implies that:

(i) PLS regression consists of simultaneous orthogonal decomposition of both ***X*** and ***Y*** matrices so that the ***X*** and ***Y*** matrices can be well-approximated using Equation (2-20) and Equation (2-21) respectively.

(ii) The PLS algorithm also derives a ***W*** weight matrix that maximizes the correlation between ***X*** and ***Y*** .as given by Equation (2-22). The inner relationship is improved by exchanging the scores ***T*** and ***U*** in an iterative calculation (Figure 2-18). This allows information from one block to be used to adjust the orientation of the latent vectors in the other block, and vice versa.

A detailed explanation of the iterative method is available in the literature (Torgrip, 2003). When all scores and loadings are calculated, the ultimate PLSR model becomes:

$$\hat{Y} = X\hat{B} \quad \textbf{(2-23)}$$

$$\hat{B} = P(P^T P)^{-1} WC^T \quad \textbf{(2-24)}$$

Where ***B*** is the matrix of regression vectors.

The number of latent variables in the model A is of crucial importance and its optimum value is derived by cross-validation. The ***W*** weight matrix represents how the ***X***-variables are linearly combined to form any score vector $\mathbf{t_i}$. Hence, by examining the ***W*** matrix, one could understand which original variable in ***X*** space would dominate the latent variables $\mathbf{t_i}$ (Eriksson *et al.,* 2001). The variation in the data that was left unexplained by the PLS modeling is given by the ***E*** and ***F*** residual matrices.

CHAPTER 3

Fluorescence Spectroscopy as a Tool for Monitoring Solubility and Aggregation Behavior of β-Lactoglobulin after Heat Treatment*

Denaturation and aggregation of whey proteins is of interest to the food and pharmaceutical industry due to the importance of final structure in functionality, impact on food texture, and the chemical stability of the final product. In this study, we demonstrate the potential of fluorescence spectrometry combined with multivariate chemometric methods for quantifying solubility and aggregation behavior of β-lactoglobulin; a major whey protein and a frequent food ingredient. Heat induced aggregation of β-lactoglobulin was studied under different conditions including pH, temperature and heating durations. Results showed very good agreement between the fluorescence-based predictions and measurements obtained by HPLC and gravimetric analysis regardless of the conditions. Standard Normal Variate (SNV), a signal preprocessing and filtering tool, was found to enhance the predictive accuracy and robustness of the fluorescence-based model.

---

* Elshereef R, Budman H, Moresoli C, Legge R. (2006). Fluorescence spectroscopy as a tool for monitoring solubility and aggregation behavior of β-lactoglobulin after heat treatment. *Biotech Bioeng* **95**:863-874.

3.1 INTRODUCTION

Denaturation and aggregation behavior of β-lactoglobulin, one of the major whey proteins and a frequent food ingredient, is of interest to the food industry. This is related in part to its effect on the final structure and texture of food as well as the chemical stability of the final product (Euston *et al.,* 2001). The protein aggregation process involves generally two steps: first, conformational and structural changes related to partial unfolding of the native protein that leads to the exposure of some hydrophobic amino acid residues and second, the subsequent aggregation of the unfolded molecules via the formation of new intermolecular bonds between the exposed amino acid residues in different peptides (Mulvihill and Donovan, 1987). The degree of aggregation is very complicated as it depends on a number of physicochemical parameters such as temperature, protein concentration, protein-protein interactions, ionic strength and pH (Vetri and Militello, 2005).

It is apparent therefore that monitoring of whey protein aggregate formation during processing is critical to the development of highly functional products. In recent years, fluorescence spectroscopy has been a useful tool for chemical analysis of diverse pharmaceutical, food and biotechnological products. A major advantage of fluorescence spectroscopy over other analytical techniques is that it is rapid, noninvasive and very sensitive to biological components and is amenable to development as an on-line sensor.

Protein fluorescence is related to intrinsic fluorophores in the protein largely due to the tryptophan and tyrosine amino acid residues. These hydrophobic amino acid residues and their fluorescence are sensitive to the local environment so changes in the

fluorescence emission spectra can provide information about protein conformation, subunit association and denaturation (Lakowicz, 1999). As a result, intrinsic fluorescence measurements of proteins have been used to study the effects of the protein microenvironment, pH, ionic strength, and temperature on protein association properties, degree of unfolding and aggregation behavior (Lakowicz, 1999).

Fluorescence spectroscopy, like all types of spectrometric methods (UV/VIS, IR and NMR), have become a common tool for exploratory analysis in most science and engineering fields such as medicine, biotechnology, food, toxicology and applied pharmacology. However, not all the data collected from scanning spectrofluorometry is relevant for every measurement. Hence, there is a significant body of literature on the use of chemometric methods to extract meaningful and relevant information for the purpose of quantifying and predicting a set of desired quality variables.

For example, Herbert *et al.* (2000) were able to discriminate eight different soft cheeses using their fluorescence spectra by applying the multivariate chemometric methods such as principal component analysis and factorial discriminant analysis. They found that the spectral patterns associated with principal components provide characteristic wavelengths, which are suitable for classifying the eight different soft cheeses Becker *et al.* (2003) demonstrated the use multi-wavelength fluorescence spectroscopy and chemometrics for predicting riboflavin content in plain yogurt during storage. Hagedorn *et al.,* (2004) evaluated spectrofluorometry as a tool for monitoring bioreactor fermentations, including substrate, biomass and product predictions by making use of multivariate calibration models in extracting the underlying variations in

the multi-wavelength fluorescence spectra that are mostly correlated with the important process variables in bioreactor fermentations.

In this study multivariate chemometric tools were applied to the analysis of intrinsic protein fluorescence measurements to investigate and monitor solubility of β-lactoglobulin (β-LG) and its aggregation behavior caused by changes in pH, temperature and heating duration. β-LG was used as a model protein because it is the most abundant protein component in bovine whey (consisting up to 50% of the total whey protein) and it is largely responsible for whey protein functionality (Schokker *et al.,* 1999). This approach included the development of a fluorescence-based chemometric model for monitoring the solubility of β–LG and its aggregation behavior, validated by two independent methods for the estimation of protein concentration: HPLC and dry weight (gravimetric) analysis.

3.2 MATERIALS AND METHODS

3.2.1 Materials and Sample Preparation

β-Lactoglobulin (β-LG), in powder form (lot JE 007-3-921 and JE 003-3-922) of 95% purity was donated by Davisco Foods International (LeSueur, MN) and was used without further purification. All other chemicals were of analytical grade. Solutions of 9.3 g/l β-LG were prepared in two different media: 0.1 M acetate buffer (pH 3.5-7.0) and 0.1 M citrate buffer adjusted with HCl (pH 3.4-4.5).

3.2.2 Heat Treatment and Gravimetric Analysis

A 25 ml volume of the desired protein solution was placed in a temperature controlled water bath at the desired temperature and treatment duration. A summary of all the experimental conditions is presented in Table 3-1. After heat treatment at the desired temperature, 25 ml samples were cooled to room temperature then centrifuged at 22000×g in a Beckman L7 Ultracentrifuge for thirty minutes. The supernatant was decanted and analyzed for the final protein *(Cf)* content by HPLC and the protein aggregate (pellet) used for dry weight determinations. Pre-tared centrifuge tubes containing the pellet were oven dried at 90°C for approximately 20 hours until constant weight and the dry weight of the pellet determined.

3.2.3. HPLC Analysis of Soluble β-LG

The initial *(C_i)* and final protein *(C_f)* content for all samples was done using High Pressure Liquid Chromatography (HPLC). The chromatography system consisted of a Waters 600 E systems controller, Waters 700 Satellite WISP, and a Waters 486 Tunable Absorbance Detector set a 280 nm. 10 μL of supernatant was injected onto a Zorbax GF-250 (9.4×250 mm) analytical column and eluted with 200 mM phosphate buffer at pH of 7. A calibration curve was prepared using different concentrations of pure untreated β-LG. The protein concentrations were estimated from the peak height. Percentage of protein aggregation was calculated as:

$$Aggregate\ \ \% = 100 * \left(\frac{C_i - C_f}{C_i} \right) \qquad \textbf{(3-1)}$$

Table 3-1: Summary of the heat treatment experiments of β-LG protein-solutions performed at different conditions (temperatures, durations and mediums).

Experiment	# Samples	T[°C]	Medium	pH	Heat Treatment Period (minutes)
1	16	85	0.1 M Sodium Acetate	4.5	5 -110
2	16	85	0.1 M Sodium Acetate	4.5	5 -110
3	24	82	0.1 M Sodium Acetate	4.5	0-60
4	12	80	0.1 M Sodium Acetate	4.5	0-120
5	13	75	0.1 M Sodium Acetate	4.5	0-90
6	15	65	0.1 M Sodium Citrate adjusted with HCl	4.5	0-100
7	13	45	0.1 M Sodium Citrate adjusted with HCl	3.5	0-180
8	9	85	Sodium Acetate	Different pH (3.5-7.0)	90

3.2.4 Fluorescence Analysis

Fluorescence measurements were conducted in a 1.0-cm cuvette using a Varian Cary Eclipse Fluorescence Spectrophotometer (Palo Alto, CA). Excitation and emission slit widths were set to 5 nm and 10 nm, respectively. Excitation was conducted over a wavelength range from 280 to 320 nm at a scan interval of 2 nm; the emission spectra were recorded in the region 300-420 nm with a resolution of 1 nm producing a 20×120 excitation-emission wavelength pairs matrix consisting of 2400 intensity data points

for a given sample measurement. A typical plot of fluorescence intensity data versus emission and excitation wavelengths as a landscape layout is presented in Figure 3-1. The maximum peak was located at an excitation wavelength of 300 nm and emission of 330 nm, which corresponds to tryptophan (Lakowicz, 1999; Renard *et al.* 1998).

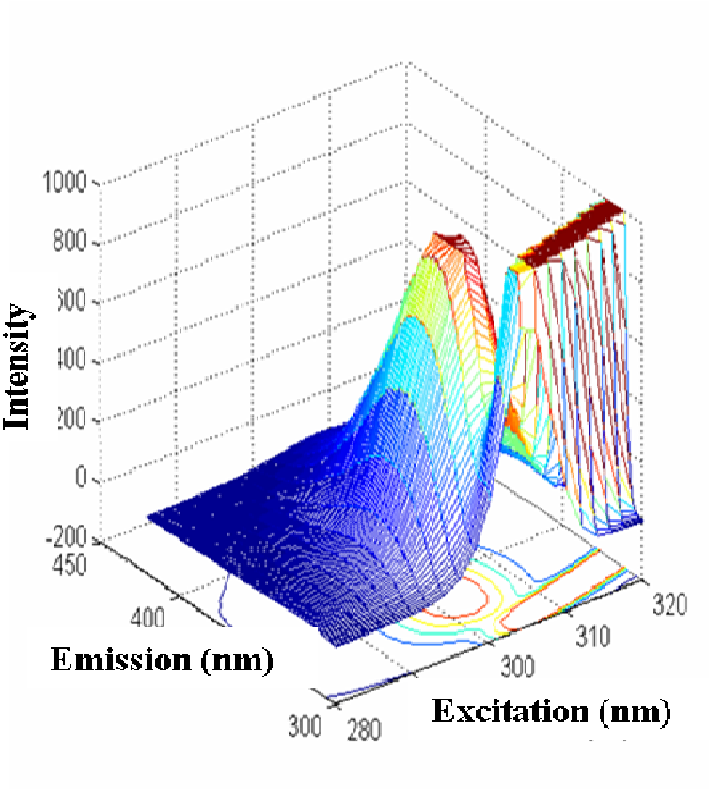


Figure 3.1: Fluorescence intensity data shown in landscape layout for a solution of 9.3 g/L β-LG at room temperature and pH of 4.5.

3.3 MATHEMATICAL METHODS

The experimental data in this study were divided into two categories: input data, the fluorescence spectrometric measurements and output data, protein concentration obtained by dry weight and HPLC analysis. This section provides a brief description of the mathematical tools and data analysis techniques that were implemented in this

study. All computations were carried out using MATLAB 5.3 (MathWorks, Natick, MA) along with the PLS Toolbox 3.5 (Eigenvector Research Inc., Manson, WA).

3.3.1 Preparation of Fluorescence Data for Analysis

The input data obtained in this study can be described by a three-way data structure as illustrated in Figure 3-2. 118 samples were measured using fluorescence spectroscopy with a 20 by 120 excitation/emission wavelength pairs matrix producing a three-way data array (118×120×20). Raw spectral data were collected and transformed into a form suitable for the PLS analysis where each 120×20 excitation-emission wavelength pairs matrix was unfolded to a 1×2400 matrix.

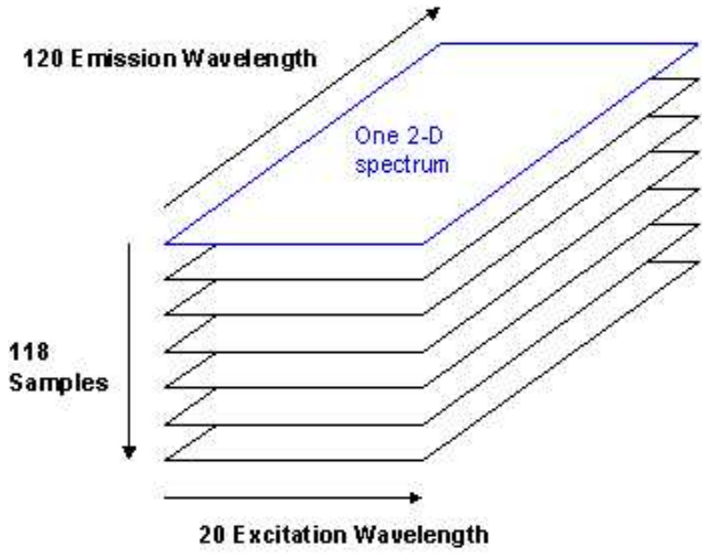


Figure 3.2: A three-way data structure consisting of 118 excitation-emission matrices with 20 excitation wavelengths and 120 emission wavelengths (120x20).

These unfolded excitation-emission matrices of dimension 1×2400 can subsequently be arranged into one single two-way matrix of dimension 118×2400 as given below:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \quad \textbf{(3-2)}$$

where x_{ij} represents the intensity value of the j th spectral variable (excitation-emission wavelength pair) for the i th case (i.e. sample). $X_1=[x_{11}\ x_{12}\ x_{13}\ldots.x_{1m}]$ is the row vector containing the fluorescence intensities measured at m excitation/emission wavelength pairs for the first sample. $X_2=[x_{21}\ x_{22}\ x_{23}\ldots.x_{2m}]$ is the row vector containing the fluorescence intensities measured at m excitation/emission wavelength pairs for the second sample and so on where m and n are equal to 2400 and 118, respectively.

3.3.2 PLS Regression

Like other spectrometric methods (NMR, UV/VIS and IR), fluorescence is characterized with data sets having a high degree of interaction, redundancy and collinearity (i.e. correlation) between the columns (variables). Collinearity, a situation where measurements at different wavelengths are strongly correlated, is considered to be a problem because it diminishes the information content of the data. Collinearity results in the spectral data being poorly handled by the traditional calibration methods such as Multiple Linear Regression (MLR), which assumes that the X-variables are independent and not correlated. Furthermore, MLR tends to deteriorate drastically if there are only a limited number of observations compared to the dimension of the

variable space n (in our case the number of samples is 118, the number of variables are 2400). One possible approach to overcome such problems is the use of Partial Least Squares (PLS). PLS regression is a well-known multivariate data analysis method that is capable of handling collinearity among the input variables in the X-matrix and projecting the spectral data into a reduced dimensional space. Hence, the original variables *x* are replaced by a smaller set of underlying new variables that are uncorrelated, mutually independent (orthogonal) and linear combinations of original descriptors.

These new variables, known as latent variables or principal components, are calculated to both provide good representation of the X-matrix and maximize the relationship between the input and the output (Qin and McAvoy, 1992). This can be expressed mathematically as $T=XW$, where T is the matrix containing scores that are the linear combinations of the original X values. PLS calculates the weight matrix W that maximizes the covariance between Y and T (Qin and McAvoy, 1992). The weight matrix shows the important excitation-emission pairs for each PLS component.

3.3.3 Cross-Validation

To avoid over-fitting in PLS calibration, cross-validation is performed with the aim to determine the optimal model complexity and the minimum number of components that describes the underlying relationship between the input and the output. The optimal model complexity has been determined from a leave-one-out cross validation approach (Geladi and Kowalski, 1986).

3.3.4 External Validation

The evaluation of the predictive capability of a multivariate calibration model can be made by comparing the concentrations and protein aggregation calculated by the calibration models with those obtained experimentally by dry weight analysis and HPLC measurements. The actual validation is done by comparing the model predictions to dry weight and HPLC data points that have not been used for calibration of the model. The two measures of model predictive capability that were used in this study are the root mean square error of prediction (RMSEP) and the squared predictive correlation coefficient (Q^2).

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{P}(y_{pred} - y_{obs})^2}{p}} \quad \textbf{(3-3)}$$

$$Q^2 = 1 - \left[\frac{\sum_{i=1}^{p}(y_{pred} - y_{obs})^2}{\sum_{i=1}^{p}(y_{obs} - \bar{y}_{obs})^2}\right] \quad \textbf{(3-4)}$$

y_{pred} can be either the predicted β-LG concentration (i.e. solubility) in the sample or the predicted protein aggregation (%), y_{obs} is the observed or reference value of the concentration (i.e. solubility) in that sample or the observed protein aggregation (%) and p is the number of samples in the test set. RMSEP gives an estimate of the prediction error in the same unit as the initial data. On the other hand, the squared predictive correlation coefficient (Q^2) measures the strength of the correlation between the values obtained by the model and the reference values obtained experimentally.

3.3.5 Pretreatment Methods of Spectroscopic Data

The X-matrix (Equation 3-2) that contains spectral data was pretreated by four different techniques (mean centering, scaling, standard normal variate and normalization). These pretreatment methods were performed prior to PLS regression in an effort to improve the correlation between the input and response. Then, the regression models obtained by using these different pretreatment techniques were compared on the basis of prediction accuracy.

Columns Mean Centering (MC)

Mean centering is useful in that it can be applied to remove a common background variation or an offset in the data (Bro and Smilde, 2003) that is irrelevant to the predicted response. The X data matrix given by Equation 2 is mean centered by calculating the average value for data in a column and subsequently subtracting that average value from every element in that column. This results in a mean-centered data matrix that has new columns with zero means. Such transformation can be expressed mathematically as follows:

$$x_{ij}^{MC} = x_{ij} - \bar{x}_j \qquad \textbf{(3-5)}$$

where $\bar{x}_j$ is the average value in a column; j is the variable index and i is the row index.

Variable (Column-wise) Scaling (VS)

Variable or column-wise scaling is usually performed by dividing every measurement in a column (i.e. excitation-emission pair) by the standard deviation of

that column (Bro and Smilde, 2003) as given by Equation 3-6a. Thereby, all columns will have the same variance and every variable will have the same chance of influencing the regression model.

$$x_{ij}^{VS} = \frac{x_{ij}}{\sigma_j} \qquad \textbf{(3-6a)}$$

where σ_j is the standard deviation of data in the jth column as given below:

$$\sigma_j = \sqrt{\sum_{i=1}^{n} \frac{(x_{ij} - \bar{x}_j)^2}{n}} \qquad \textbf{(3-6b)}$$

$\bar{x}_j$ is the average value of data in the *j*th column. *j* is the variable index and *i* is the row index. *n* is the number of observations.

Standard Normal Variate (SNV)

The standard normal variate approach has been used for near-infrared spectra to reduce the multiplicative interferences of scatter and particle size (Geladi *et al*., 1985). No literature was found regarding its application to correcting fluorescence data. SNV corrects the spectra by centering each row and then scaling it by its own standard deviation as given by Equation 3-7a. In our case, each row corresponds to the fluorescence spectrum of a given sample.

$$x_{ij}^{SNV} = \frac{x_{ij} - \bar{x}_i}{\sigma_i} \qquad \textbf{(3-7a)}$$

, where σ_i is the standard deviation of data in the *i*th row as given below:

$$\sigma_i = \sqrt{\sum_{j=1}^{m} \frac{(x_{ij} - \bar{x}_i)^2}{m}} \quad \textbf{(3-7b)}$$

$\bar{x}_i$ is the average value of data in the *i*th row, *j* is the variable index. *m* is the number of columns (i.e. excitation emission pairs).

Normalization (NM)

Normalization of the spectra prior to multivariate calibration has been used for path length correction (Geladi *et al.,* 1985). Each element in the row (Equation 3-2) is divided by the square root of the sums of squares for that row and consequently, each normalized spectrum will be represented by a unit vector as given by Equation (3-8).

$$x_{ij}^{NM} = \frac{x_{ij}}{\sqrt{\sum_{j=1}^{m} x_{ij} x_{ij}}} \quad \textbf{(3-8)}$$

where *m* is the number of columns, *j* is the variable index and *i* is the row index.

3.4. RESULTS AND DISCUSSION

3.4.1 Effect of Heat Treatment Duration and Temperature on β-LG Aggregation

The time course of β-LG aggregation as measured by precipitation at 85°C based on dry weight and HPLC analysis is presented in Figure 3-3. The two methods were assumed reliable as for two replicates the experimental error was estimated to be 5.93%.

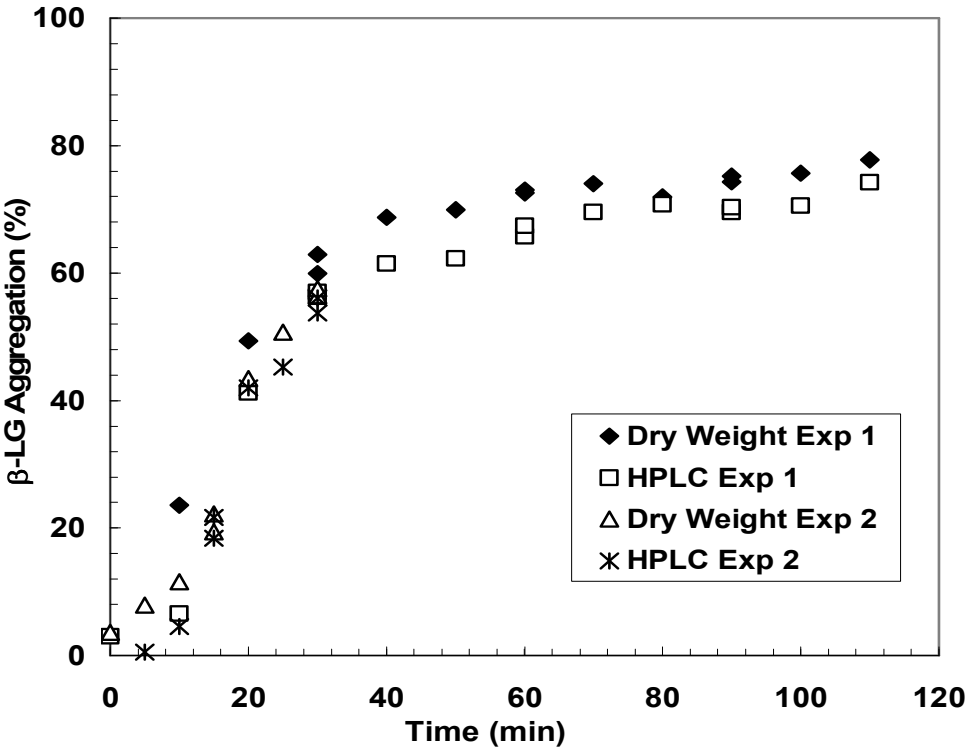


Figure 3-3. Percentage of β-LG protein aggregation based on dry weight and HPLC analysis for two replicates (experiment 1 and 2) with heat treatment at 85°C, pH 4.5.

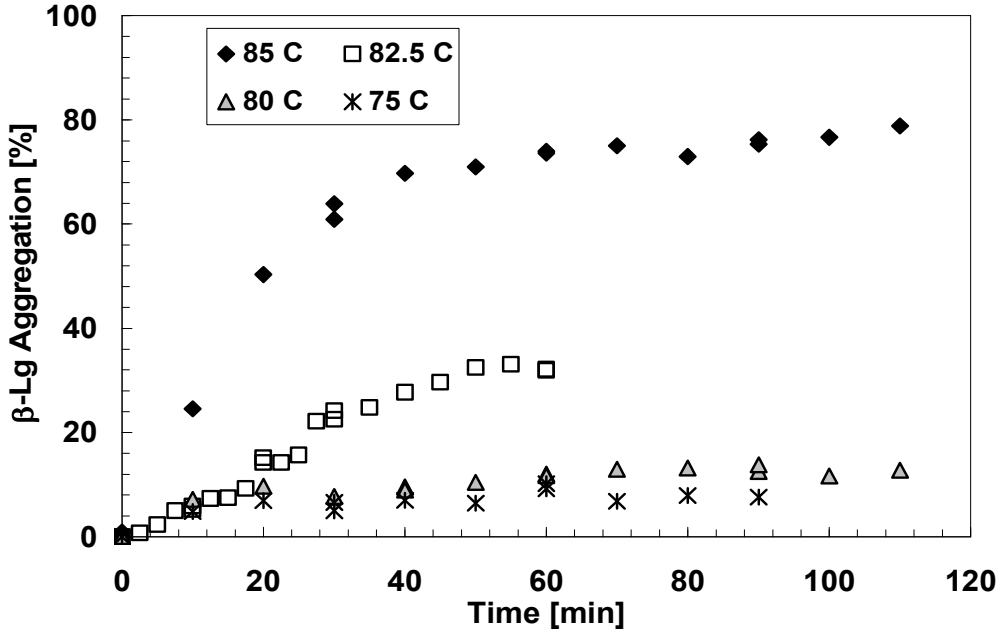


Figure 3.4: Percentage of β-LG aggregation based on dry weight analysis plotted versus heating time at four different temperatures 75, 80, 82.5 and 85°C, sodium acetate buffer, pH 4.5.

At 85°C the maximum amount of protein precipitation that could be achieved was in the order of 80%. The time course of precipitation following thermal treatment of β-LG in solution at various temperatures is given in Figure 3-4. The extent of protein precipitation increased over time and with increasing temperature. At 85°C, a plateau was observed at around 80% after one hour, whereas at 82.5°C a plateau appeared at a similar time but at 35% β-LG precipitation. At 80°C and lower temperatures, the amount of β-LG precipitation was less than 15%. It was observed that the rate of protein aggregation was strongly temperature dependent over a temperature range of 80-85 °C where two-degree rise in temperature resulted in a two-fold increase in the rate of β-LG precipitation.

3.4.2. Effect of pH on β-LG Aggregation

Along with temperature and heating time, pH is considered to be one of the key factors that influence the heat-induced aggregation behaviour of whey proteins and their functional properties (Fang and Dalgleish, 1998; Hoffmann and van Mil, 1999; Hunt and Dalgleish, 1994). The vast majority of formulated and processed dairy products, including whey protein end products, are manufactured under acidic conditions (Xiong *et al.,* 1993). Eight β-LG solutions of identical protein concentration (9.3 g/L) and the same buffer (sodium acetate buffer), but different pH values in the acidic range, were prepared and then subjected to heat treatment at 85°C for 90 minutes. Based on both HPLC analysis and dry weight, pH was found to have a significant effect on the amount of β-LG aggregation (Fig. 3-5).

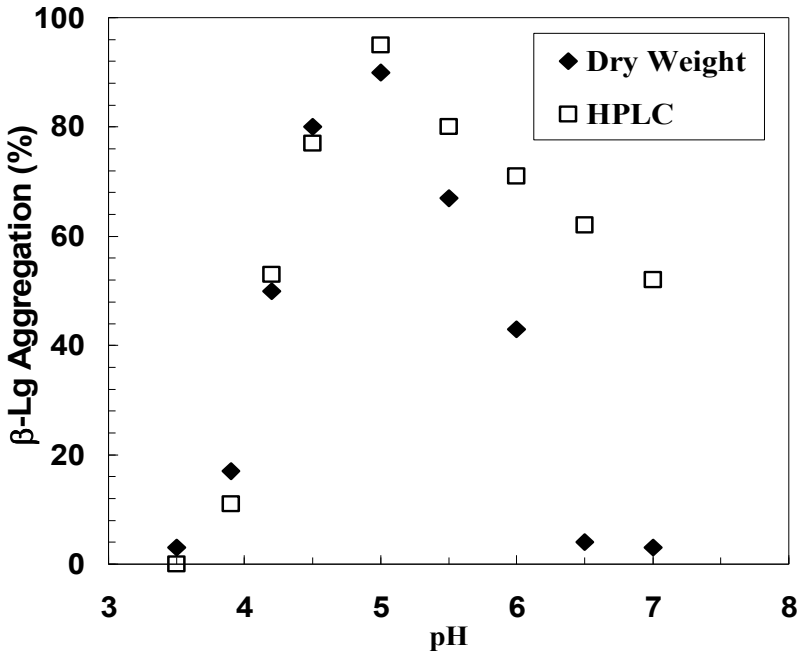


Figure 3.5 Effect of pH on the β-LG aggregation after a 90-minute heat treatment at 85°C in acetate buffer (Experiment 8, Table I): Percentage of β-LG aggregation is based on dry weight and HPLC analysis.

At pH values below 4.0, precipitation is very low, in agreement with previous observations of Renard *et al.* (1998). Renard *et al.* (1998) attributed the very low β-LG protein aggregation observed at pH 2.0 to the inhibitory effect of that pH on the formation of disulphide bonds. The electrostatic repulsion between positively charged protein molecules at low pH values are strong enough that thiol/disulphide interchange reactions between monomers are inhibited resulting in a small degree of aggregation (De la Fuente *et al.,* 2002). These results show that protein aggregation reaches a maximum value at a pH of approximately 5.0, which is very close to the theoretical β-LG isoelectric point (pI 5.3) (Kelly and Zydney, 1997). Similar results have been reported by others (De Rham and Chanton, 1984; Renard *et al.,* 1998; Xiong *et al.,* 1993). There is very good agreement between HPLC and dry weight measurements at

pH values lower than 5.0; however, the two methods give different levels of aggregation at pH values greater than 5.0. The HPLC chromatograms for β-LG solutions at pH values greater than 5.5 show a peak that is probably due to the presence of larger molecular weight aggregates that were in suspension. It was not possible to take these peaks into account in the calculation of the amount of aggregation since it is not clear at this point how these peaks are related to the aggregates in solution. However, these additional peaks could explain why the predicted amount of aggregation is higher based on dry weight determinations versus the HPLC method.

3.4.3 Fluorescence Analysis

Visual inspection of the fluorescence excitation-emission matrices collected during experiment 1 (85 °C), where a new sample was used for each time point, reveals that the fluorescence landscape of β-LG solution changes over the time course of heat treatment. Figure 3-6 shows fluorescence measurements collected during experiment 1 for three samples (with different heat treatment times of 5, 40 and 110 minutes) with an initial β-LG concentration of 9.3 g/L. The change in the fluorescence landscape of β-LG solution over the time course of heat treatment involves, first, a progressive decline in the intrinsic fluorescence intensity of tryptophan (Figure 3-6) and second, a shift of the emission peak of tryptophan from 328 nm towards longer wavelengths (340-350 nm).

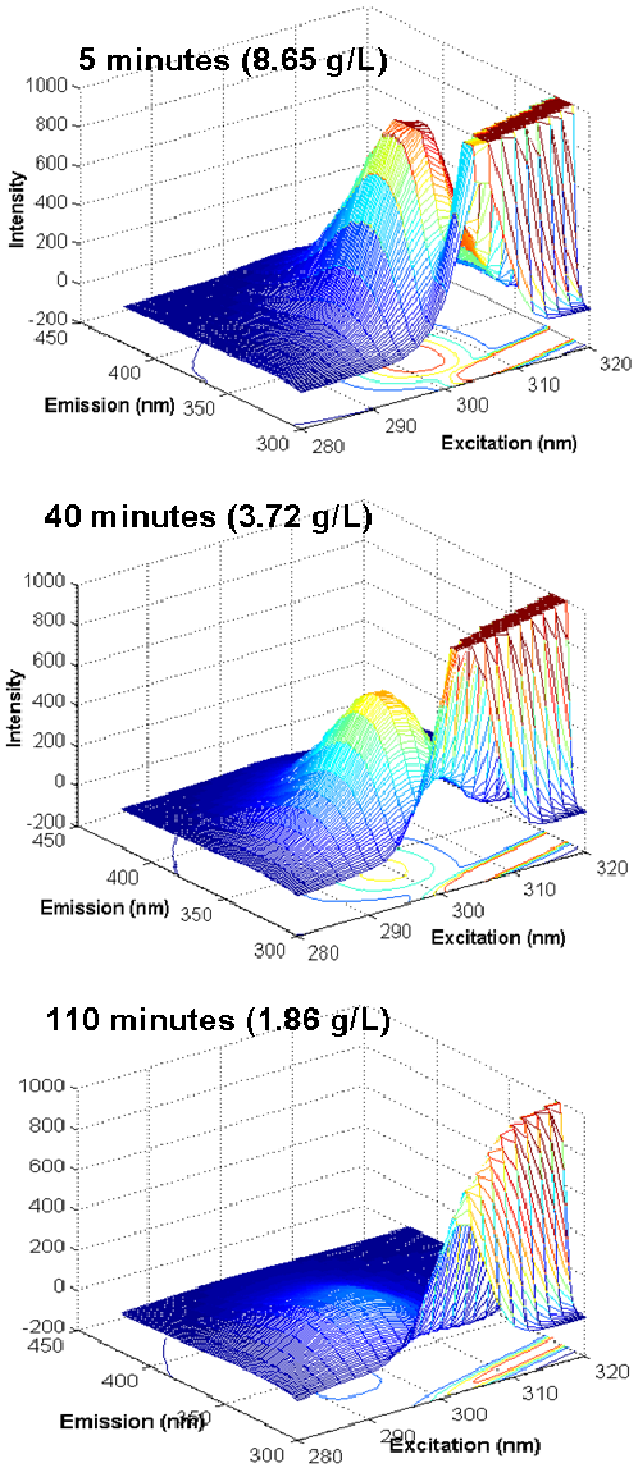


Figure 3.6: Fluorescence profile for a 9.5 g/L β-LG solution (Experiment 1) after thermal treatment at 85 °C at different time intervals (5, 40 and 110 minutes): β-LG concentrations were determined by HPLC and dry weight.

3.4.4 PLS Regression

Partial Least Squares regression was employed to determine the underlying components (also called latent variables) in the fluorescence spectra that are relevant to the measured solubility of β-LG and its aggregation percentage calculated from Equation (3-1). The experimental data for 118 different β-LG samples with different thermal treatments (one fluorescence spectrum per case) has been divided into a calibration set and a validation set. The first experiment (Table 3-1) consisting of 16 samples was used for calibration. The X matrix for the calibration has dimensions of 16×2400 and contains in its rows the individual spectral samples. The X and Y matrices were both mean-centered prior to PLS regression. PLS regression applied to the calibration set (i.e. the data set with known concentrations) provided four latent variables or PLS components that are statistically significant, with a goodness of prediction by cross validation (Q^2) of about 92%. These PLS components capture 91% of the variance in the X matrix (fluorescence intensity). The first PLS component is the most significant since it accounts for 58% of the variance in the X-matrix and it has strong linear correlation with the observed extent of aggregation (Figure 3-3). The second PLS component accounts for 23.8% of the variance in the X-matrix. The 3rd and 4th components are less important but they were retained for PLS modeling since they were determined necessary based on cross-validation. The four PLS components contain the projections (scores) of the fluorescence landscapes that belong to 16 thermally treated protein solutions onto the low dimensional space determined by PLS. PLS scores for the first two significant components are plotted versus heating time

(Figure 3-7). Important spectral regions were identified by plotting the PLS weights of component 1 and 2 versus excitation-emission wavelength (Figure 3-8a and 3-8b).

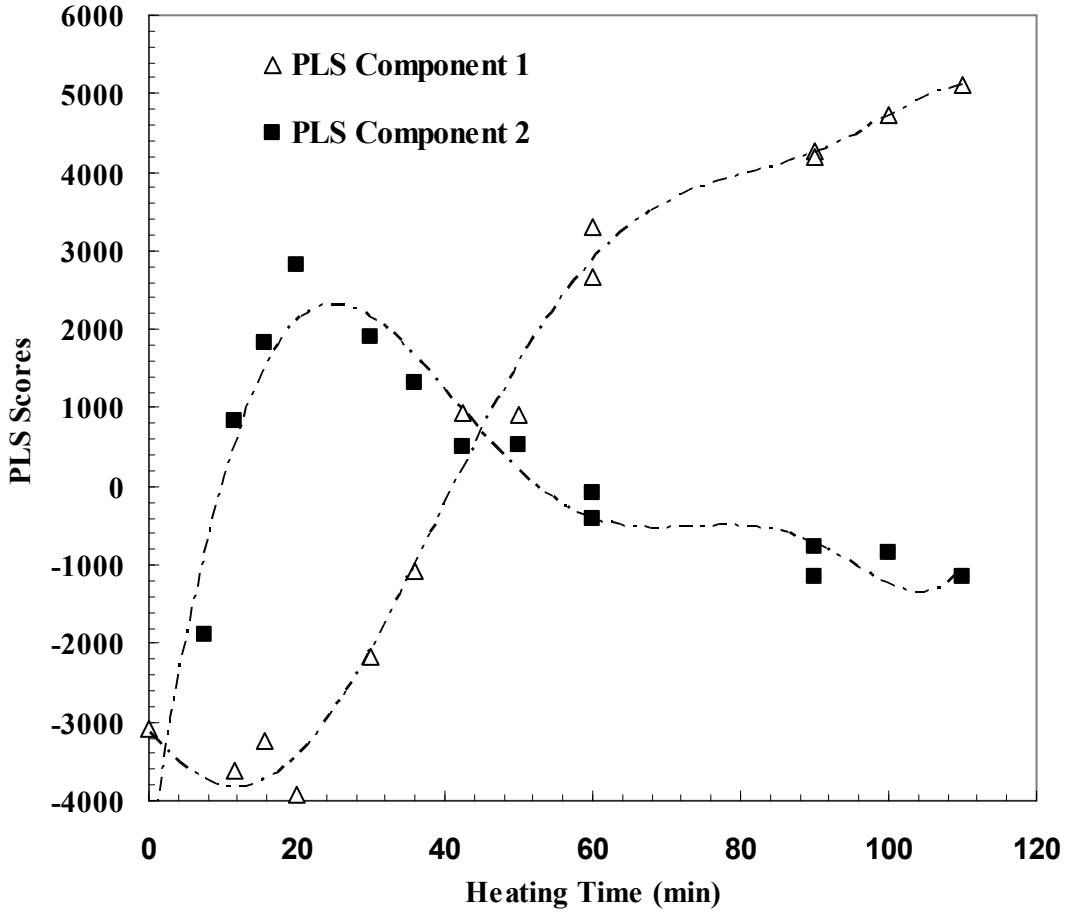


Figure 3-7. Plot of PLS scores for the first two significant latent variables versus heating time (Experiment 1: 85 °C, 0.1 M acetate buffer pH 4.5).

A plot of PLS scores for the first component versus heating time (Figure 3-7) shows an increasing trend that is strongly correlated with the observed extent of aggregation. According to Figure 3-8a, the first PLS component captures the fluorescence change in the spectral region at around 328-330 nm emission and 300 nm excitation.

Figure 3-8.a

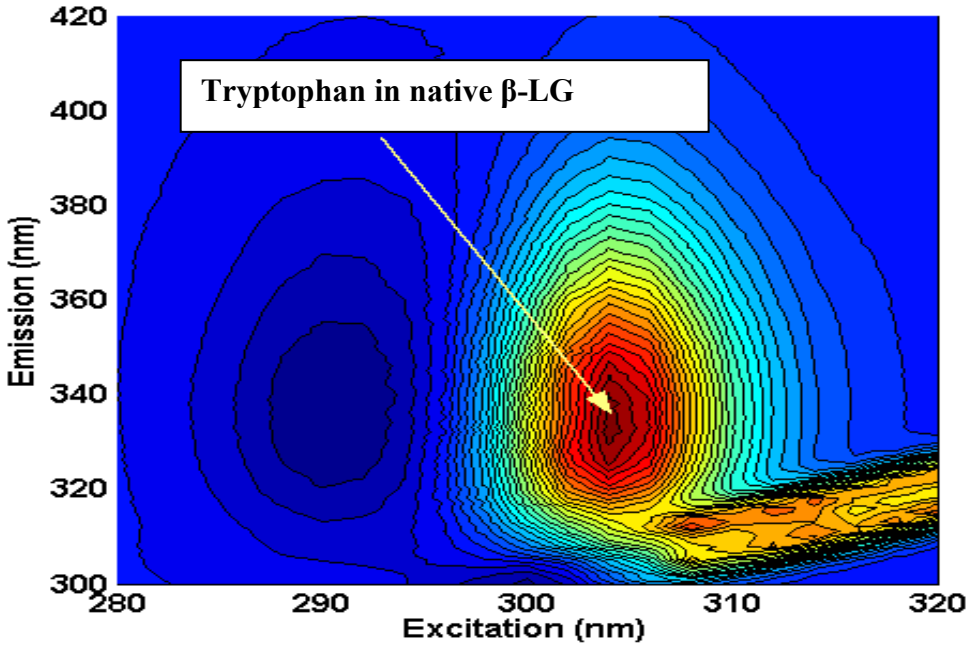


Figure 3-8.b

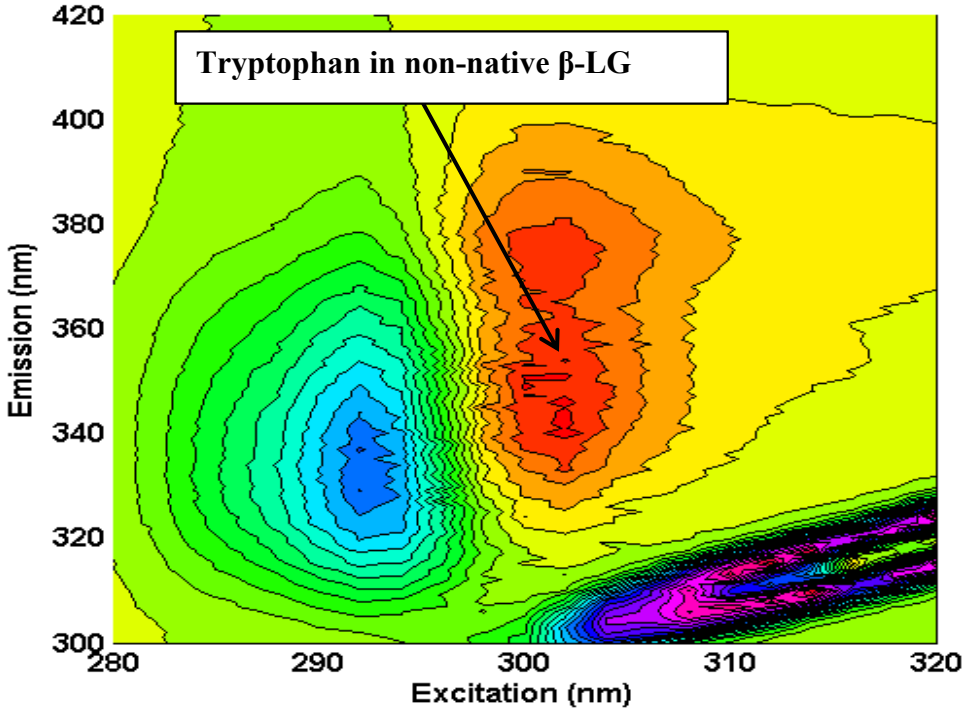


Figure 3-8. (a) PLS weights of component 1 versus excitation-emission wavelengths. (b) PLS weights of component 2 versus excitation-emission wavelengths for Experiment 1 (85 °C, 0.1 M sodium acetate buffer pH 4.5).

Comparing the spectral region seen in Figure 3-8a with the emission wavelength corresponding to the maximum intensity of 328 nm for β-LG reported at room temperature by Renard *et al.* 1998, this spectral region can be most likely attributed to tryptophan buried in hydrophobic cores of the native protein (Reshetnyak and Burstein, 2001).

Intuitively, the observed formation of insoluble aggregates over the course of heat treatment is accompanied by the decrease in the soluble amount of native β-LG. On the other hand, a plot of PLS scores for the second component (Figure 3-7) versus heating time shows an increase up to 20 min heating time that is followed by a slow decline with further increase in heating time. PLS weight plot for the second PLS component provides a spectral signature (Figure 3-8.b) that is different from that of the native β-LG (Figure 3-8a). As seen in Figure 3-8.b, that spectral signature at around emission 340-350 nm is possibly of tryptophan residue in the non-native state of the protein (Reshetnyak and Burstein, 2001).

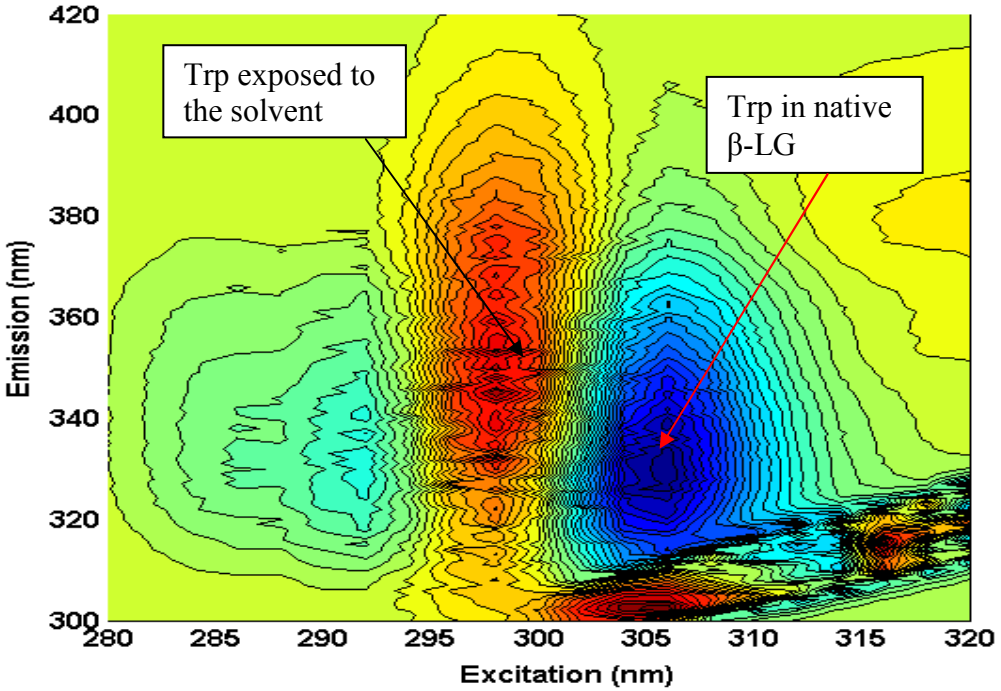
Recent studies using CD spectroscopy (Prabakaran and Damodaran, 1997) support the hypothesis of (Qi et al., 1995) that the primary cause of the initiation reaction in heat induced aggregation of β-LG involves critical conformational changes in β-LG to form reactive monomers which then react with each other via sulphhydryl-disulphide exchange reactions to form dimers and other aggregates (i.e. trimers and tetramers). Mandenson et al. (1998) noted that these dimers and aggregates could be important intermediates in the further aggregation of β-LG. The native protein and its aggregates are believed to involve a very large, heterogeneous population of partially unfolded molecules that interact differently with the solvent and with the neighboring

molecules (Vetri and Militello, 2005). The plot of PLS scores for the second component may reflect the formation and depletion of some non-native intermediates over the course of heating.

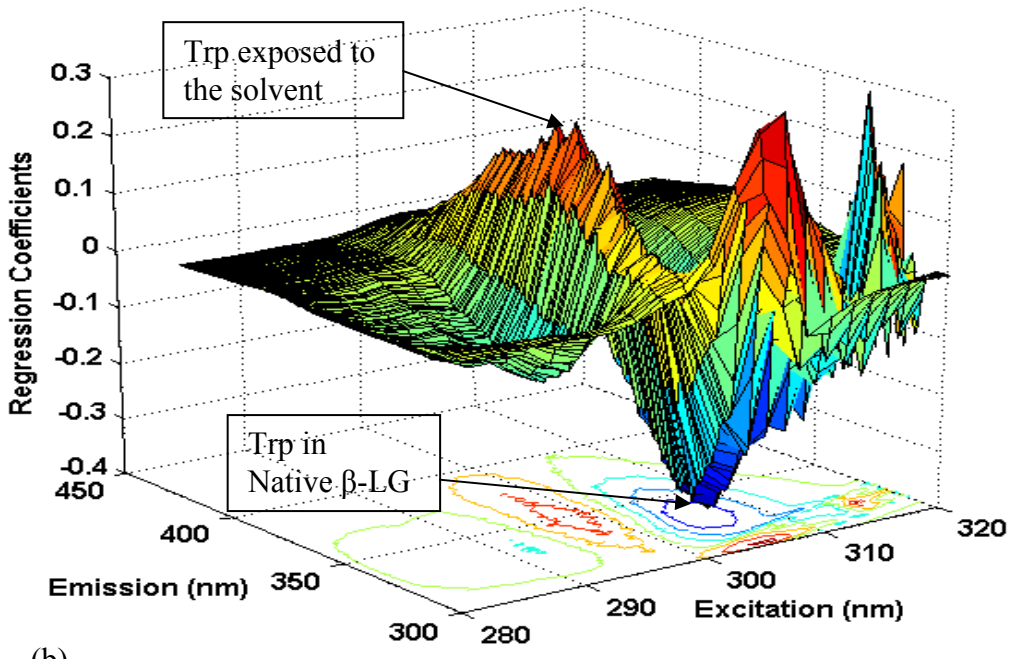
Formation of non-native β-LG intermediates after short heating times has been observed by Iametti *et al.* (1996) and Schokker *et al.* (1999) and it is considered to be the first step in the heat-induced aggregation of β-LG. Schokker *et al.* (1999) monitored the formation of irreversibly altered monomers and non-native dimers throughout the aggregation of β-LG. Their study showed that the increase in the amount of aggregates is accompanied by a progressive loss of native-like β-LG throughout the heating process. At the same time, Schokker *et al.* (1999) observed that the amount of non-native monomers, dimers and trimers increased during the early stages of heat treatment after which a slow decrease with time was observed. Similar results were obtained by other researchers (McSwiney *et al.,* 1994; Prabakaran and Damodaran, 1997). This analysis suggests that several simultaneous kinetic events associated with the protein structural change, such as partial unfolding and aggregation, can be captured using the PLS components that were extracted from the multiwavelength spectra.

In addition to the PLS weight plots, the sensitivity spectrum (Boehl *et al.*, 2003) is another method which allows extraction of qualitative information in the model by analysis of the wavelength pairs that are relevant to the predicted variable (i.e. β-LG aggregation). The sensitivity spectrum consisting of the scaled regression coefficients for every wavelength combination in the model are shown in Figure 3-9. From Figure 3-9, two major areas can be identified that are relevant to predicting the extent of β-LG

aggregation for a given heat treatment condition. The significant peak at around 303 nm excitation and 330 nm emission, which is due to tryptophan buried in hydrophobic cores of the native β-LG, has a negative correlation with the predicted variable (i.e. β-LG aggregation). Hence, the higher the amount of aggregates, the lower the amount of native β-LG residing in the solution. The area of large peaks at around excitation 297 nm and emission 340-350 nm corresponding to tryptophan residues at different exposure levels to the surrounding solvent (Reshetnyak and Burstein, 2001) contribute positively to the predicted β-LG aggregation. In other words, this positive correlation implies that the exposure of the protein hydrophobic residues to the surrounding solvent upon unfolding gives rise to β-LG aggregation, which is in agreement with the literature (De la Fuente *et al.,* 2002; McSwiney *et al.* 1994; Prabakaran and Damodaran, 1997; Schokker *et al.* 1999).



(a)



(b)

Figure 3-9: Sensitivity spectra (regression coefficients plot) used for modeling β-LG aggregation shown as a contour map (a) or landscape layout (b).

3.4.5 PLS Model Testing

The PLS calibration model was tested against three independent data sets that were not used in the calibration. The fluorescence spectra for experiments 2, 3 and 4 (Table 3-1) were fed into the PLS model to calculate the β-LG concentration that corresponds to each sample. The model gives a RMSEP of 0.912 g/L, which corresponds to 10.1 % prediction error in terms of aggregation percentage. While this prediction error is larger than the overall experimental error of HPLC measurements (5.93 %), the current model may not be suitable for reliable quantitative determination of the protein content in real-life application such as quality control and on-line monitoring of food processes. As this prediction error is relatively large, attempts to improve the calibration model have been made and will be discussed in the next section.

3.4.6 Spectra Preprocessing Prior to PLS Regression

Understanding the origins and characteristics of measurement error may suggest approaches of improving the quality of input data that is fed into the PLS model. Therefore, the fluorescence measurements for four replicates of identical protein content were collected at an excitation wavelength of 300 nm as seen in Figure 3-10a. A close examination of the fluorescence spectra of the four replicates reveals that the variance of measurements is proportional to the measurement itself (Figure 3-10b). Such an error pattern is often referred to in literature as heteroscedastic noise or error that possibly arises from counting statistics (i.e. shot noise) or fluctuations in source intensity with wavelength channel (Schreyer *et al.,* 2002).

Figure 3-10.a

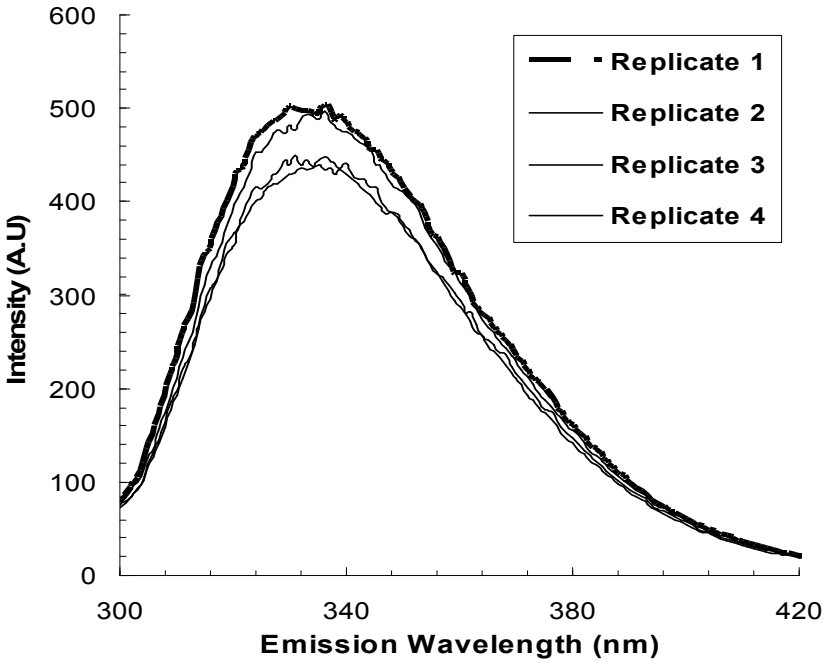


Figure 3-10.b

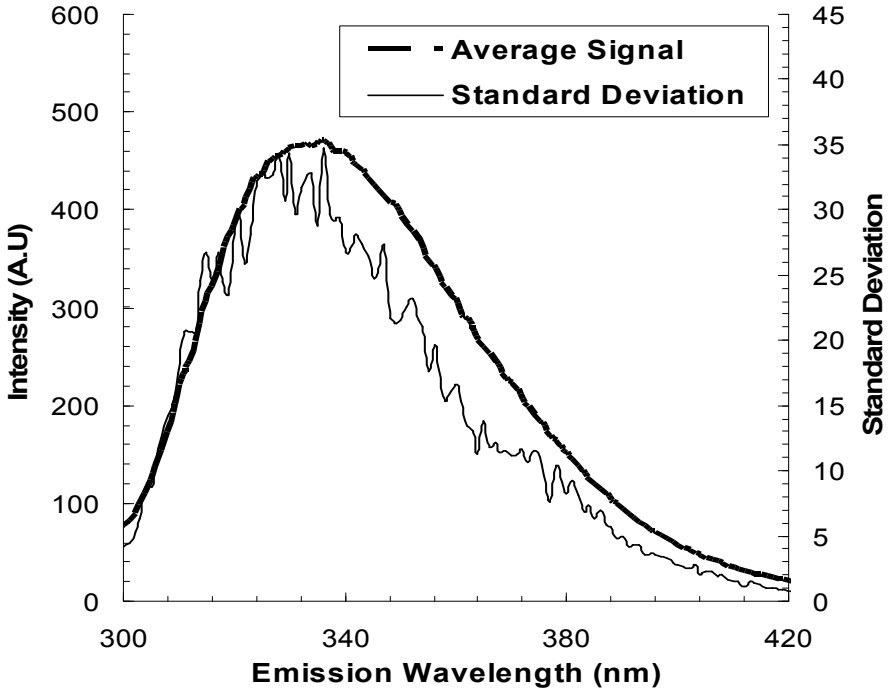


Figure 3-10. (a) Fluorescence measurements for four replicates (β-LG concentrations (pH, temperature) collected at an excitation wavelength of 285 nm. (b) The mean intensity for the four replicates (primary axis) and the variance of measurements (secondary axis) around that mean calculated at each wavelength.

Due to the heteroscedastic component of the fluorescence signal, the spectra acquired for samples with different but relatively close protein concentrations could exhibit many overlapping spectral peaks that would affect the ability to accurately determine a measurable property of the target analyte (i.e. β-LG content) in each sample.

Error measurements in the fluorescence can play a significant role in degrading the quality of results obtained from the regression model. To further reduce the prediction error, the preprocessing of the spectra prior to PLS regression was evaluated. The application of four preprocessing techniques (mean centering, columns scaling, standard normal variate and normalization) to the spectrometric data was investigated in order to improve the calibration model. A new PLS model with data from Experiment 1 was calculated and validated for each preprocessing method by testing the model on the spectra of Experiments 2 to 7 (Table 3-1). The effect of the pretreatment method on the predictive accuracy described by the Root Mean Squared Error of Prediction (RMSEP) and the Predictive Correlation Coefficient (Q^2) are presented in Table 3-2 and Table 3-3.

Table 3-2: Effect of the pretreatment method for X (spectra) and Y (β-LG concentration) on the overall predictive ability of the model as described by RMSEP and Q^2

	X: None	X: MC	X: VS	X: SNV	X: NM
RMSEP (g/L)	0.855	0.912	0.805	0.465	0.413
Q^2	0.60	0.51	0.65	0.87	0.88

Table 3-3: Effect of the pretreatment method for X (spectra) and Y (β-LG aggregation) on the overall predictive ability of the model as described by RMSEP and Q^2.

	X: None	*X: MC*	*X: VS*	*X: SNV*	*X: NM*
RMSEP (% aggregation)	9.34	10.11	8.54	5.02	4.93
Q^2.	0.53	0.41	0.66	0.85	0.86

Table 3-4: Effect of the pretreatment method for X (spectra) and Y (β-LG aggregation) on the model accuracy (RMSEP) for each validation set (all tabulated values are RMSEP expressed as β-LG aggregation (%)).

Experiment	Aggregation%	X: None	X: MC	X: VS	X: SNV	X: NM
2	0-75	10.80	11.80	7.70	4.95	4.40
3	0-30	11.23	11.93	9.55	6.34	5.09
4	0-15	6.58	7.32	6.72	3.75	4.06
5	0-12	6.19	5.44	3.32	2.93	4.42
6	0-11	6.33	6.47	7.95	3.60	3.69
7	0-11	7.69	8.87	8.00	4.93	5.00

Although mean-centering (MC) is considered to be a standard approach in PLS regression, the data preprocessing obtained from mean centering seem to be worse than when no preprocessing was used in all validation sets as seen in Table 3-4. Our results agree with those of Seasholtz *et al.* (1992), where they demonstrated that the RMSEP

will be smaller for a multivariate model made with raw input data than for a model made with mean centered data if the data, as in this case, varies with concentration.

The best results based on RMSEP and Q^2 were obtained with the use of the Standard Normal Variate (SNV) and normalization (NM) for all validation sets. It is also noticed that the performance of SNV and NM is nearly equivalent where both methods seem to improve the accuracy of prediction by about 40%. The reason for the superior performance of SNV and NM is that both were able to handle the type of measurement error occurring in the fluorescence data. Assuming that x_{ij} is the raw fluorescence signal corrupted by heteroscedastic noise it can be expressed mathematically as (Schulze, 1997):

$$x_{ij} = \varepsilon . x_{ij}^{\,o} \tag{3-9}$$

$x_{ij}^{\,\circ}$ is the noise-free fluorescence signal for a given sample that is a function of only the physical properties of the target analyte (i.e. protein content) and ε is the heteroscedastic noise. Substituting Equation 3-9 into the SNV transformation given by Equation 3-7a yields:

$$x_{ij}^{\,(SNV)} = \frac{x_{ij} - \bar{x}_i}{\sqrt{\sum_{j=1}^{p}(x_{ij} - \bar{x}_i)^2}} = \frac{\varepsilon . x_{ij}^{o} - \varepsilon . \bar{x}_i^{\,o}}{\sqrt{\sum_{j=1}^{p}(\varepsilon . x_{ij}^{o} - \varepsilon . \bar{x}_i^{\,o})^2}} \tag{3-10}$$

Note that Equation 3-10 can be simplified further by canceling out ε so the SNV-corrected signal becomes only a function of the noise-free component $x_{ij}^{\,\circ}$:

$$x_{ij}^{(SNV)} = \frac{x_{ij}^{o} - \bar{x}_{i}^{o}}{\sqrt{\sum_{j=1}^{p}(x_{ij}^{o} - \bar{x}_{i}^{o})^2}} \qquad \textbf{(3-11)}$$

Similar to SNV, Normalization (Equation 3-8) corrects the fluorescence signals by removing the heteroscedastic noise while preserving the pure noise-free component x_{ij}° that is a characteristic of the target analyte as described by Equations 3-12 and 3-13.

$$x_{ij}^{NM} = \frac{x_{ij}}{\sqrt{\sum_{j=1}^{P} x_{ij} x_{ij}}} = \frac{\varepsilon . x_{ij}^{o}}{\sqrt{\sum_{j=1}^{p}(\varepsilon . x_{ij}^{o}).(\varepsilon . x_{ij}^{o})}} \qquad \textbf{(3-12)}$$

which implies that

$$x_{ij}^{NM} = \frac{x_{ij}^{o}}{\sqrt{\sum_{j=1}^{P} x_{ij}^{o} x_{ij}^{o}}} \qquad \textbf{(3-13)}$$

The ability of SNV to distinguish between variations that are due to physical properties of the analyte from those that are non-relevant was demonstrated before by Geladi *et al.* (1985) in the pretreatment of near infrared spectra.

Model predictions of PLS made with SNV corrected spectra were compared against HPLC and dry weight measurements for three different data sets presented in Figure (3-11). It is clear from this figure that the model predicts the protein aggregation % very well over wide range of conditions (different temperatures and treatment durations).

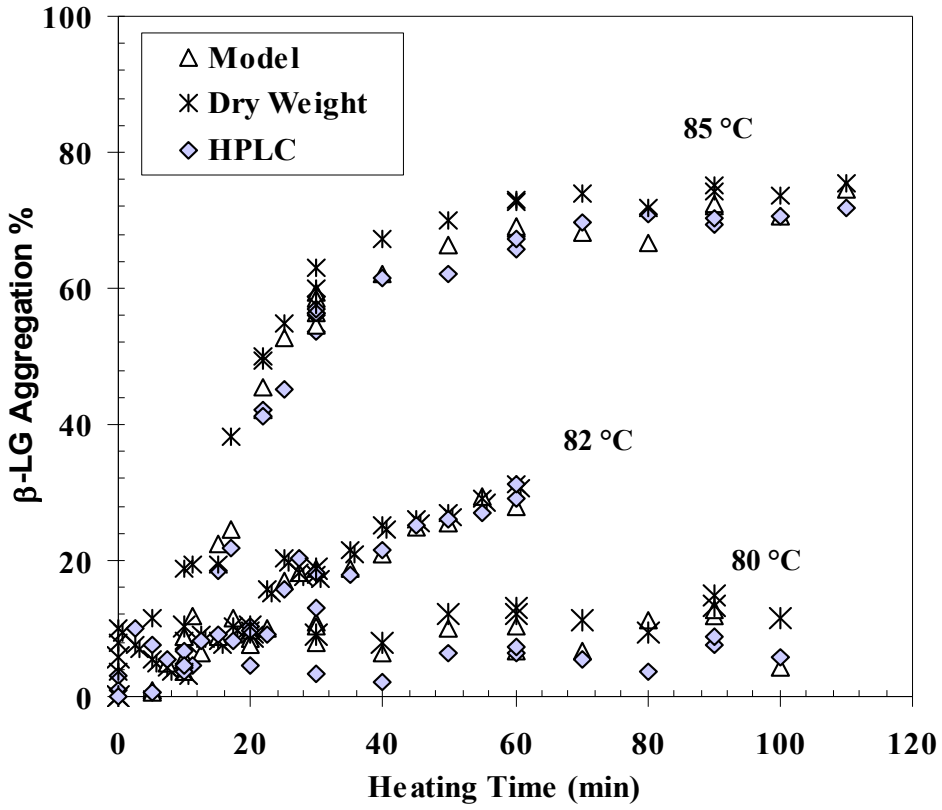


Figure 3-11: PLS model prediction for β-LG protein aggregation compared to HPLC and dry weight measurements at three different temperatures 85, 82 and 80 °C (Experiments 2, 3 and 4, respectively).

3.5 CONCLUSION

A novel fluorescence-based regression model is proposed for the prediction and quantification for a model system consisting of the protein β-LG. Results show very good agreement between the fluorescence-based predictions and measurements obtained by two analytical techniques. The usefulness of multivariate calibration tools, also known as chemometric methods, in extracting the features that are correlated to the physical properties of the target protein such as its concentration and extent of aggregation were illustrated. Standard Normal Variate, a signal preprocessing and

filtering tool, was demonstrated to play a significant role in enhancing the predictive accuracy and robustness of the sensor. Although the SNV approach is widely implemented in signal processing, its specific applicability in filtering fluorescence signals has not been studied previously.

CHAPTER 4

Fluorescence-based Soft-sensor for Monitoring β-Lactoglobulin and α-Lactalbumin Solubility during Thermal Aggregation*

A soft-sensor for monitoring solubility of native-like α-lactalbumin (α-LA) and β-lactoglobulin (β-LG) and their aggregation behavior following heat treatment of mixtures under different treatment conditions was developed using fluorescence spectroscopy data regressed with a multivariate Partial Least Squares (PLS) regression algorithm. PLS regression was used to correlate the concentrations of α-LA and β-LG to the fluorescence spectra obtained for their mixtures. Data for the calibration and validation of the soft sensor was derived from fluorescence spectra. The process of thermal induced aggregation of β-LG and α-LA protein in mixtures, which involves the disappearance of native-like proteins, was studied under various treatment conditions including different temperatures, pH, total initial protein concentration and proportions of α-LA and β-LG. It was demonstrated that the multivariate regression models used could effectively deconvolute multi-wavelength fluorescence spectra collected under a variety of process conditions and provide a fairly accurate quantification of respective native-like proteins despite the significant overlapping between their emission profiles. It was also demonstrated that a PLS model can be used as a black-box prediction tool for estimating protein aggregation when combined with simple mass balances.

---

* Elshereef R, Budman H, Moresoli C, Legge RL. (2008). Fluorescence-based soft-sensor for monitoring β-lactoglobulin and α-lactalbumin solubility during thermal aggregation. *Biotech Bioeng* **99**:567-577.

4.1 INTRODUCTION

Whey, a by-product of cheese manufacturing, is a milk fraction composed of lactose, proteins, vitamins, minerals, and fats. Whey proteins include a number of different proteins such as β-lactoglobulin (β-LG), α-lactalbumin (α-LA), bovine serum albumin, and immunoglobulin. These whey proteins have high nutritional value and can be used to manufacture different types of food products. Large amounts of whey protein are used for infant formula, yogurt, ice cream, and beverages. In the past, attempting to maximize the extraction of these proteins from whey and separating them has, and continues to be a challenging task. The impurities in whey make the extraction process relatively difficult. In addition, the similarities between α-LA and β-LG make it even more difficult to separate these proteins.

Amundson *et al*. (1982) developed a method to produce enriched fractions of β-LG and α-LA from cheese whey by concentrating the whey protein using ultrafiltration followed by pH adjustment of the concentrate. However, Muller *et al.* (2003) have shown that separation of proteins using selective thermal precipitation is more promising than using the ultrafiltration route, provided that proper conditions including initial protein concentration, pH, and length of precipitation time are maintained (Bramaud *et al.,* 1997; Tolkach *et al.,* 2005). Pearce (1987) has shown that the tendency of α-LA to aggregate is higher under specific conditions including at a pH values near the isoelectric point (pH 4.2-4.6) and in a temperature range of 50°C -65°C. They also observed that the tendency to aggregate increased with protein concentration. Bramaud *et al.* (1997) studied the effect of citrate on the precipitation of whey proteins

and they observed that the addition of citrate leads to α-LA rich fractions at temperatures around 35°C.

Research has shown that the two major whey proteins, α-LA and β-LG, become unstable at temperatures above 65°C (Zhu & Damodaran, 1994). When heated above this temperature, protein denaturation occurs resulting in protein aggregation and precipitation. The response to thermal treatment varies with different types of proteins, which results in different proteins precipitating out of the solution in different proportions making separation possible (Bramaud *et al.* 1997; Tolkach *et al.,* 2005). Therefore, heat-induced aggregation and precipitation is an important treatment process in the manufacture of many dairy products, and is used to modify functional properties with the goal of ensuring food safety of the product. Functional, physical and chemical properties of milk such as texture, heat stability, foaming properties and rheological behavior are all affected by the heat treatment (Newstead *et al*. 1975; Morr, 1985; Kessler and Beyer, 1991; Zhu and Damodaran , 1994; Luecy *et al*. 1998, Elshereef *et al*., 2006).

There were several objectives in this study. First, there was an interest in investigating the effect of different conditions on the heat treatment-based separation process of milk proteins. The rate at which whey proteins aggregate is controlled by process conditions such as protein concentration, pH and temperature and the presence of other components (Bertrand-Harb *et al.,* 2002). Our objective was to use data acquired from this approach to calibrate and validate a soft- sensor developed to monitor the separation process. Using the measurements collected under varied process conditions, it would be possible to then demonstrate the sensitivity and robustness of

this soft sensor. Soft-sensors are inferential mathematical models that use readily available process measurements or/and physical equations to estimate difficult-to-measure variables (James *et al*., 2002; Hagedorn *et al.,* 2004).

As a preliminary step, the current study focused on different solutions of β-LG and α-LA. These proteins are the predominant proteins that make up of about 70% of all the proteins in whey and are key to the functional properties and characteristics of whey. The experimental approach involved the analysis of the thermal aggregation of these two proteins at different pH values, temperatures, and protein concentrations. The soft-sensor proposed was designed by combining fluorescence spectroscopic data acquired for β-LG/α-LA solutions subjected to different heat conditions and Partial Least Squares (PLS) modelling for monitoring solubility of α-LA and β-LG in their mixtures and their aggregation behavior during heat treatment under different treatment conditions. Partial Least Squares regression is a well-known chemometric tool for developing a calibration model which correlates the set of known measurements represented by multi-wavelength fluorescence data to the desired property to be predicted (Herbert *et al.,* 2000; Becker *et al.,* 2003; Hagedorn *et al.,* 2004; Elshereef *et al*., 2006). This is because the PLS method is capable of handling data sets with large numbers of highly-correlated variables such as the fluorescence spectral measurements and with few observations (Elshereef *et al*., 2006).

4.2. MATERIALS AND METHODS

4.2.1 Materials and Sample Preparation

β-LG and α-LA in their powder form (lot JE 007-3-921 and JE 003-3-922) were of 95% purity and donated by Davisco Foods International (LeSueur, MN). All other chemicals were of analytical grade. Solutions of β-LG and α-LA were prepared in two different buffers: 0.1 M acetate buffer (pH 4.5) and 0.1 M citrate buffer (pH 3.5-6.0) as described below.

4.2.2. Thermal Treatment

Protein solutions were dispensed into 20 mL open plastic test tubes and placed in a pre-equilibrated temperature controlled water bath (GCA Precision Water Bath, Model 183) at the desired temperature and duration for the heat treatment process.

4.2.3 Centrifugation and Gravimetric Analysis of the Precipitate

After heat treatment the samples were quickly placed into an ice bath for 15 to 20 min. Samples were then centrifuged at 22000 $\times g$ in a Beckman L7 Ultracentrifuge for thirty min. The supernatant was decanted and analyzed for the final protein *(Cf)* content by HPLC and the protein precipitate (pellet) was recovered for dry weight estimation. Fluorescence measurements were performed on the supernatant prior to HPLC analysis. Protein aggregates were dried in an oven at 90^{o}C for approximately 17 h until they reached constant weight and the dry weight determined. Percentage of total protein aggregation was calculated using Equation 4-1 where W_S is the amount of

protein in a given sample prior to heat treatment and W_P is the dry weight of protein aggregate formed after heat treatment.

$$aggregate\% = 100 * \left(\frac{W_P}{W_S} \right) \quad \textbf{(4-1)}$$

4.2.4 HPLC Analysis of Soluble β-LG and α-LA in the Supernatant

The initial (C_i) and final protein (C_f) content for α-LA and β-LG in all samples were determined using High Pressure Liquid Chromatography (HPLC). The chromatography system consisted of a Waters 600 E systems controller, Waters 700 Satellite WISP, and a Waters 486 Tunable Absorbance Detector set a 280 nm.

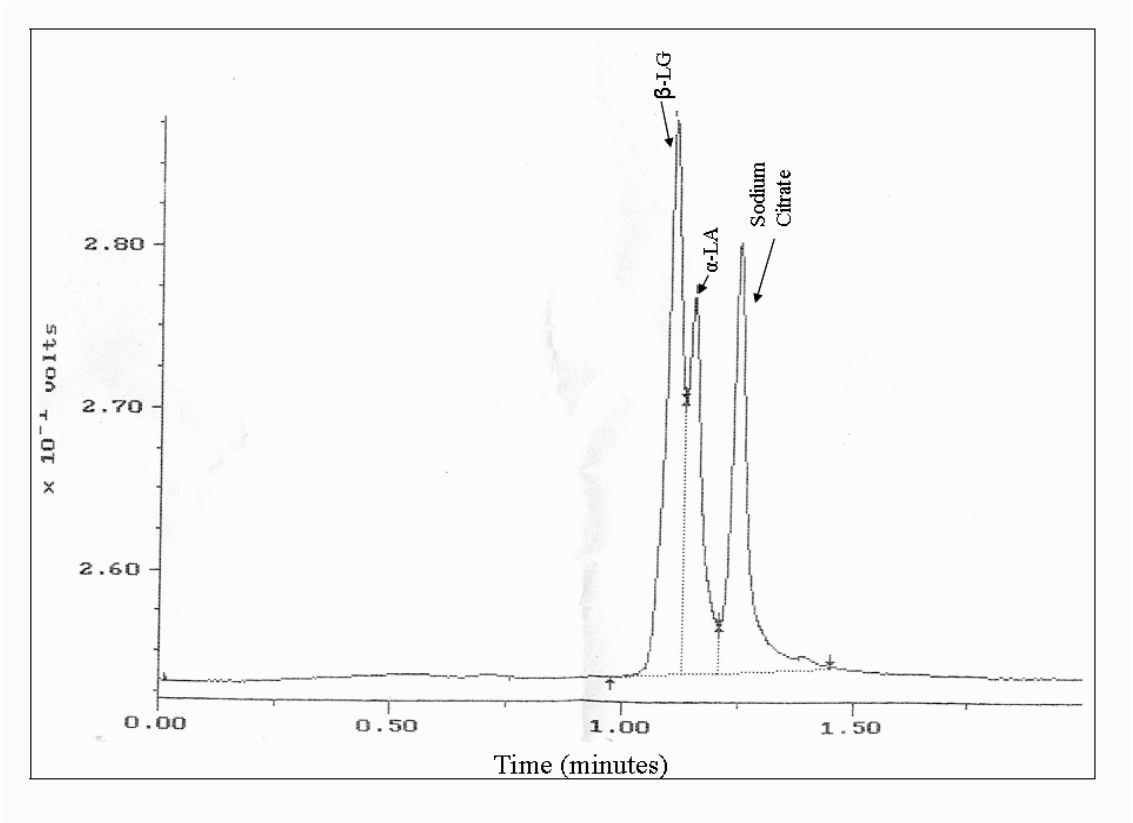


Figure 4-1. Sample HPLC chromatogram of a thermally treated β-LG/α-LA mixture.

Ten μL of supernatant was injected onto a Zorbax GF-250 (9.4×250 mm) analytical column and eluted with 200 mM phosphate buffer at pH of 7; separation the mixture was on the basis of size exclusion. An example chromatogram is given in Figure 4-1.

Calibration curves were prepared using different concentrations of pure untreated β-LG and α-LA and the initial protein concentration based on HPLC analysis. The protein concentrations were estimated from the peak height. Percentage of protein aggregation for each protein was calculated as:

$$Aggregate\% = 100 * \left(\frac{C_i - C_f}{C_i} \right) \quad \textbf{(4-2)}$$

4.2.5 Fluorescence Analysis

Fluorescence measurements were conducted in a 1.0-cm cuvette using a Varian Cary Eclipse Fluorescence Spectrofluorometer (Palo Alto, CA). Excitation and emission slit widths were both set to 10 nm. Excitation was conducted over a wavelength range from 260 to 350 nm at a scan interval of 10 nm; the emission spectra were recorded in the region 280-450 nm with a resolution of 1 nm producing a 10×170 excitation-emission wavelength pairs generating a matrix of data consisting of 1700 fluorescence intensity data points for any given sample.

4.3 CHEMOMETRIC MODELLING

The experimental data in this study were divided into two categories: input data, the fluorescence spectrometric measurements, and output data consisting of β-LG and

α-LA concentration values obtained by HPLC analysis. The matrix that contains the outputs to be predicted, i.e. HPLC measurements of β-LG and α-LA concentrations, were arranged in an output matrix to be referred to heretofore as matrix Y whereas the fluorescence measurements were arranged into an input data (X-matrix). The rows in the X matrix correspond to different samples while the columns correspond to fluorescence intensities at different excitation-emission wavelength pairs (Elshereef *et al*., 2006). The input data is considered to be multivariate in nature since it consists of many measurements (1700 fluorescence intensity data points) for many samples. Partial Least Squares regression (PLS) is a well-known chemometric tool for developing a calibration model which can be used to correlate the set of known measurements represented by the fluorescence data to the desired property to be predicted represented by the β-LG and α-LA concentrations. PLS is able to handle problems associated with noise and collinearity in multivariate data sets with large numbers of highly correlated variables (Geladi and Kowalski, 1986, Qin and McAvoy, 1992). Collinearity is very high among the different fluorescence readings obtained at different combinations of emission and excitation wavelengths (Elshereef et al., 2006). Hence, the original input variables *x* in the matrix X are replaced by a smaller set of underlying new variables that are uncorrelated, mutually independent (orthogonal) and are mathematically represented by linear combinations of the original descriptors. These calculated linear combinations, referred to as latent variables (LVs) or principal components, are calculated to both provide good representation of the X-matrix and maximize the relationship between the input and the output (Qin and McAvoy, 1992). The optimum number of latent variables (LVs) and the goodness of prediction (Q^2) are

determined by cross-validation algorithm (Geladi and Kowalski, 1986; Qin and McAvoy, 1992; Elshereef *et al.,* 2006). The data obtained from heat treatment experiments were used for calibrating and testing the PLS soft-sensor as will be shown in the following sections. All computations were carried out using MATLAB 5.3 (MathWorks, Natick, MA) along with the PLS Toolbox 3.5 (Eigenvector Research Inc., Manson, WA).

4.4 RESULTS AND DISCUSSION

The first objective of this work involved a comprehensive study of the effects of different experimental parameters on the solubility, aggregation and precipitation behavior of proteins during heat treatment. The parameters that were studied were: total protein concentration, α-LA to β-LG ratio, pH, heat treatment duration and temperature. The results are presented separately in the following subsections. Some of these effects have been partially studied and reported in the literature (Newstead *et al*., 1975; Kessler and Beyer, 1991; Luecy *et al*., 1998). The current work expands on previous findings by investigating the effects over a wider range of operating conditions and also adds new results on the effects of pH, total protein concentration and α-LA to β-LG ratio. This comprehensive study was necessary in order to calibrate the soft sensor and to validate over a wide range of process operating conditions.

4.4.1 Effect of Total Protein Concentration on Aggregation

In the majority of industrial processes involving the production of whey, thermal treatment is preceded by a preliminary concentration step to minimize energy

requirements. Hence, understanding and modeling the effect of total protein concentration prior to heat treatment on the solubility of major whey proteins as well as their aggregation behavior is of industrial interest (De la Fuente *et al*., 2002; Law & Leaver, 1997). Previous experimental studies carried out using calorimetry (Qi *et al*., 1995) or Fourier transform infrared spectroscopy (Lefevre & Subirade, 1999) and light scattering (Le Bon *et al.,* 1999) showed that the initial protein concentration has a marked effect on the thermal denaturation and aggregation of β-LG. However, such experimental studies were done with pure β-LG solutions. As far as we are aware, no investigators have reported the effect of total protein concentration on the thermal denaturation and aggregation of β-LG in the presence of α-LA only.

Table 4-1: Summary of the heat treatment experiments for β-LG and α-LA protein mixtures performed under different conditions.

Exp#	Number of Samples	pH	T[°C]	β-LG to α-LA ratio	Total Protein
1	8	3.9	75	Constant 3:1	Variable 3 -16 g/L
2	16	3.9	85	Constant 3:1	Variable 1.5 -18 g/L
3	10	3.9	75	Variable	Constant 12 g/L
4	8	3.5	85	Constant 3:1	Variable 1.5 -14 g/L
5	10	3.7	85	Constant 3:1	Variable 1.5 -16 g/L
6	9	3.9	85	Constant 3:1	Variable 1.5 -16 g/L
7	9	4.5	85	Constant 3:1	Variable 1.5 -16 g/L
8	10	5.0	85	Constant 3:1	Variable 1.5 -16 g/L
9	8	6.0	85	Constant 3:1	Variable 1.5 -14 g/L

In the first experiment summarized in Table 4-1, different amounts of α-LA and β-LG in their powdered form were mixed with 0.1 M citrate buffer at pH 3.9 resulting in 8 different protein mixtures of different total protein concentration (ranging from 3 g/L to 16 g/L) but with a constant α-LA to β-LG ratio of 1:3. The mixtures were then subjected to heat treatment at 75°C for 90 minutes.

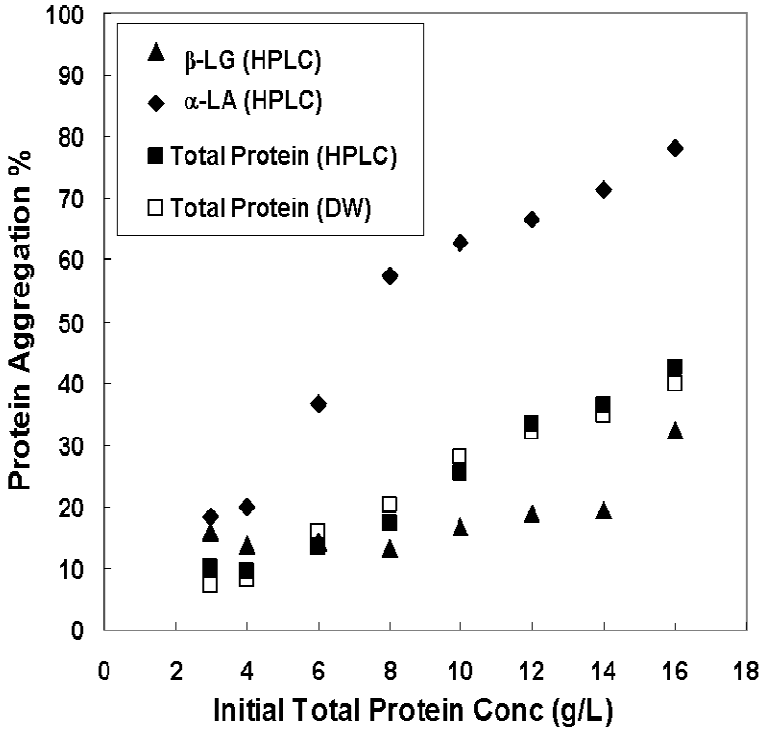


Figure 4-2. α-Lactalbumin and β-lactoglobulin aggregation at pH 3.9 for different total protein concentrations determined by HPLC (Experiment 1 in Table 4.1: heat treatment duration is 90 min; temperature 75ºC).

Figure 4-2 is a plot of the percentage of total protein precipitated based on dry weight and HPLC measurements against the initial total protein concentration. In general the amount of protein which precipitates (i.e. aggregates) increases with an increase in the initial protein concentration. To test the accuracy of the HPLC and dry weight

measurements, these were compared for different total concentrations. Figure 4-2 shows very good agreement between the total protein aggregated as measured by HPLC and dry weight with a maximum error of approximately 5% between the two measurements.

In general, the discrepancy between the two determinations becomes greater for samples with low protein precipitation that corresponds to a lower signal to noise ratio. One of the major advantages of HPLC analysis over gravimetric analysis (dry weight) is that the former method allows estimating the percentage of aggregation of α-LA and β-LG separately from the amounts of respective proteins before and after heat treatment according to Equation 4-2. As seen from Figure 4-2, the percent aggregation of α-LA and β-LG increases with increased total protein concentration. Results show that there is a marked difference in the aggregation behavior of the two proteins where the degree of aggregation observed for α-LA is significantly higher than that observed for β-LG at all protein concentrations.

4.4.2 Effect of α-LA to β-LG Ratio on Protein Denaturation and Aggregation

Appropriate amounts of α-LA and β-LG were mixed with 0.1 M citrate buffer solution at pH 3.9 resulting in 10 protein mixtures of different α-LA to β-LG ratios but with constant total protein concentration (12 g/l) (Experiment 3, Table 4-1). The mixtures were then subjected to heat treatment at 75°C for 90 minutes. Figure 4-3 shows the percentage of α-LA aggregation against the initial fraction of α-LA in the mixture before heat treatment was started. A five-fold increase in the percentage of α-LA

aggregation occurs with increasing proportion of α-LA in the initial fraction over the range of 0.1 to 0.2. A smaller percentage increase is seen for initial α-LA fractions beyond 0.2.

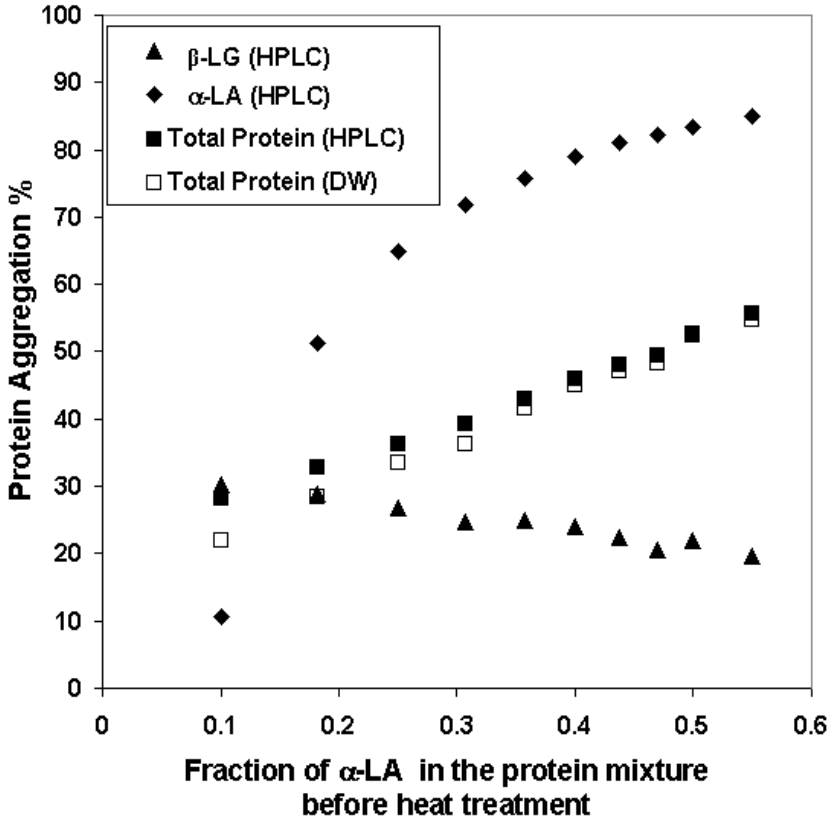


Figure 4-3. α-Lactalbumin and β-lactoglobulin aggregation at pH 3.9 for different α-LA mass fractions in the protein mixture determined by HPLC (Experiment 3 in Table 4-1: heat treatment duration is 90 min; temperature 75ºC).

Figure 4-3 also illustrates that the amount of β-LG that aggregates decreases with increasing initial α-LA fraction, implying that β-LG aggregation decreases with decreasing proportions of native β-LG present in the solution prior to heat treatment. These results agree well with those observed in Figure 4-2. The percentage of aggregation of each protein is proportional to the initial concentration of that protein in

the mixture before undergoing heat treatment with α-LA being more sensitive to heat treatment than β-LG.

4.4.3 Effect of pH During Heat Treatment

The kinetics of protein denaturation, the aggregation mechanism and the nature of the stabilizing forces involved in the formation of aggregates are affected by the solution pH during heat treatment (Xiong *et al*., 1993; Hunt and Dalgleish, 1994; Hoffmann and van Mil, 1999; Bertrand-Harb *et al*., 2002). Therefore, in order to demonstrate the robustness of a PLS based soft-sensor for monitoring aggregation behavior of these proteins under different conditions, mixtures of α-LA and β-LG were heat-treated at different pH values and the fluorescence spectra acquired from the supernatant and used to test the PLS soft sensor.

The pH over which this study was conducted (pH 3.5 to 6.0) represents the range over which β-LG exhibits different self-association behavior. At room temperature and at pH values below 4 and above 5.2, β-LG exists predominantly as monomers and dimers (Sawyer *et al*., 1999) and at pH values around 4.7, larger oligomeric structures are formed (Verheul *et al.,* 1999). In the case of α-LA, there is a conformational change in the range of pH selected for this study (Muller *et al.* 2003). Results in Figure 4-4 shows that aggregation of both proteins is more rapid near their isoelectric points (4.2 for α-LA and 5.2 for β-LG). It is also clear that the two proteins have different aggregation behaviors at all pH values and at all initial protein concentrations. Furthermore, the propensity of α-LA to aggregate is greater than that of β-LG at pH values below 4.5 as

seen in Fig. 4-4. This difference in the aggregation behavior can be partially attributed to the higher thermal stability of β-LG over this pH range.

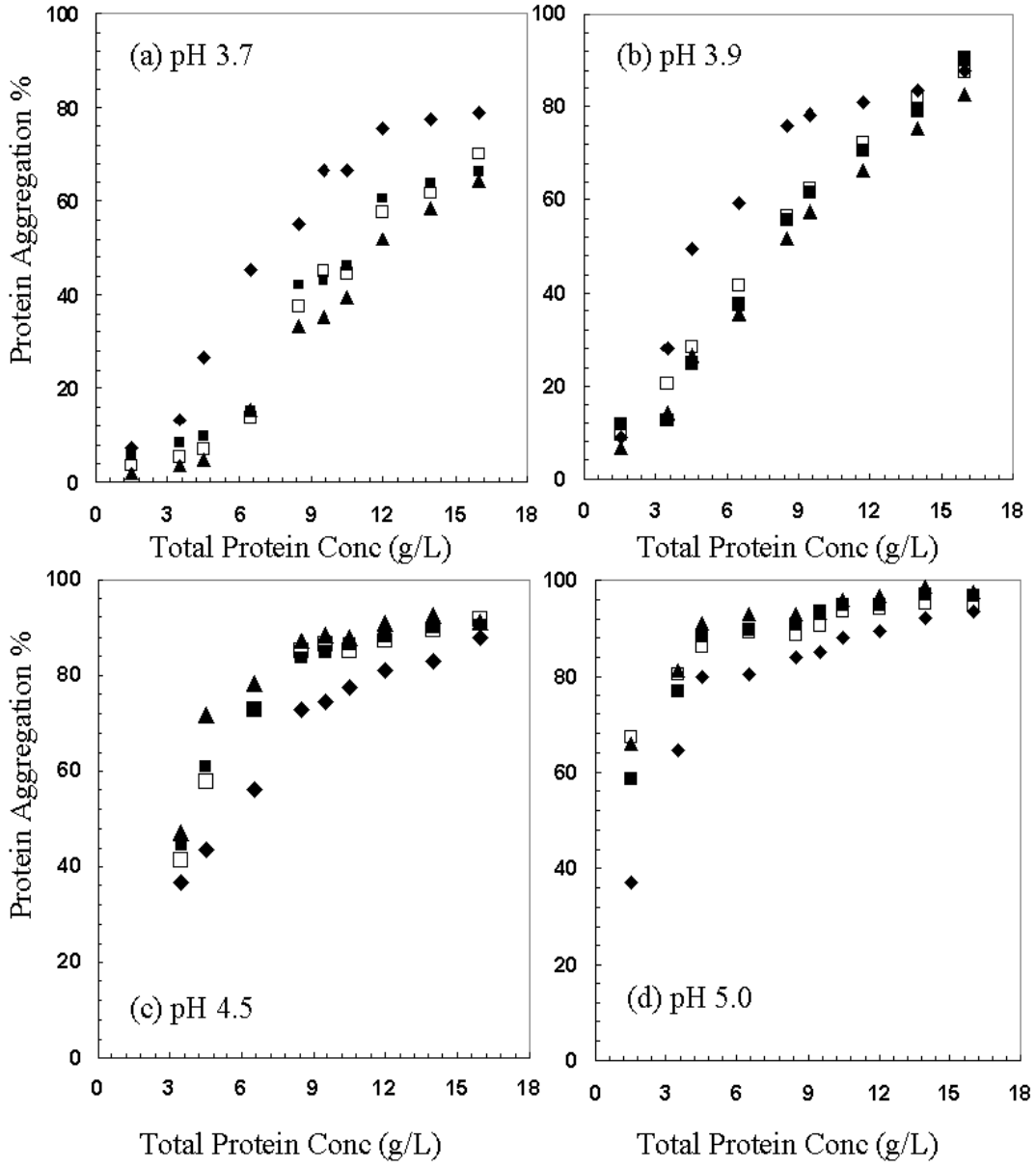


Figure 4-4. α-Lactalbumin and β-lactoglobulin aggregation at different pH values (3.7, 3.9, 4.5 and 5) for different total protein concentrations determined by HPLC (Experiments 5, 6, 7 and 8 in Table I; heat treatment duration is 90 min, temperatures 85ºC at pH 3.7, 3.9, 4.5 and 5 respectively). ♦ = α-LA; ▲ = β-LG; □ = total protein determined by dry weight; ■ = total protein determined by HPLC.

Previous studies of thermal stability of β-LG in the presence of α-LA using differential scanning calorimetry (Boye and Alli, 1999) found that β-LG retains its increased thermal stability at low pH in spite of the presence of α-LA. At higher pH values, β-LG is known to be thermally unstable (Boye and Alli, 1999). Given the residual amounts of respective proteins in the supernatant that were determined by HPLC analysis, the weight ratio of native β-LG to native α-LA under different conditions was calculated using Equation 4-3.

$$\beta - LG/\alpha - LA = \left(\frac{C_{\beta - LG}}{C_{\alpha - LA}} \right) \qquad \textbf{(4-3)}$$

The weight ratio of β-LG/α-LA was plotted as a function of the initial total protein concentration in the original protein mixture before heat treatment at four different pH values (Fig. 4-5).

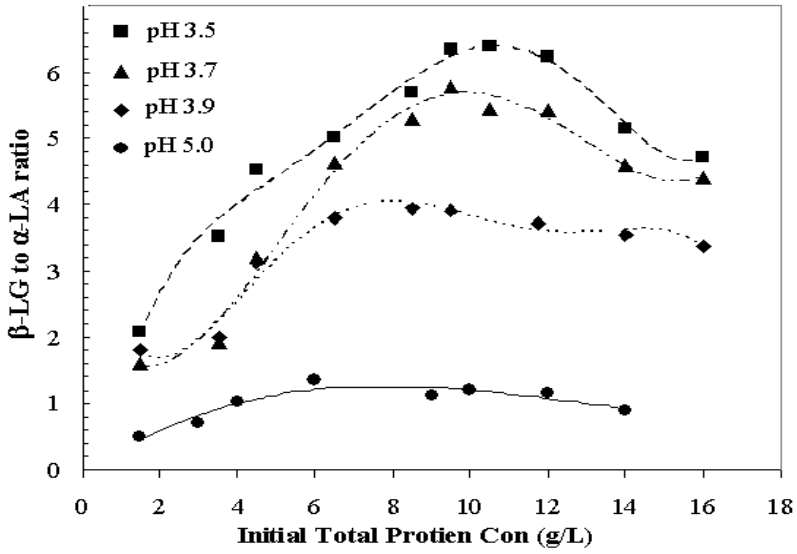


Figure 4-5. Effect of pH and initial total protein concentration on the weight ratio of β-LG to α-LA in the supernatant (heat treatment duration 90 min; temperature 85°C).

It can be seen from Fig. 4-5 that β–LG enrichment by thermal treatment is favored at lower pH values. At higher pH values, the recovery of α-LA by thermal precipitation (i.e. aggregation) is reduced because more β-LG precipitation is favored under these conditions (Figure 4-4), leading to a decrease of soluble β-LG to α-LA weight ratio in the supernatant as shown in Figure 4-5. Similar results have been previously reported by Muller *et al.* (2003) who found that pH 3.9 is the optimal pH value for recovery of α-LA. Recovery of α-LA by thermal precipitation was lower at pH values higher than 3.9.

4.4.4 Effect of Temperature on Protein Aggregation

Experiments were carried out at 75 and 85°C for 90 min (data not shown). The total protein concentration was varied between 1.5 g/l to 12 g/l at a constant α-LA:β-LG ratio of 1:3 in 0.1 M acetate buffer at pH 3.9. As expected, more protein aggregated at higher temperatures (Figure 4-6a). It is also observed that the aggregation behaviors of α-LA and β-LG are different at 75 and 85°C as seen in Figures 4-6b and 4-6c. Results of the effect of temperature on the aggregation behavior of β-LG and α-LA are consistent with those reported by Zhu & Damodaran (1994). HPLC analysis likely over-estimates protein aggregation because additional protein aggregation can occur prior to the analysis. Although HPLC analysis gives higher aggregation levels, the largest recorded difference between HPLC and DW values is about 10%.

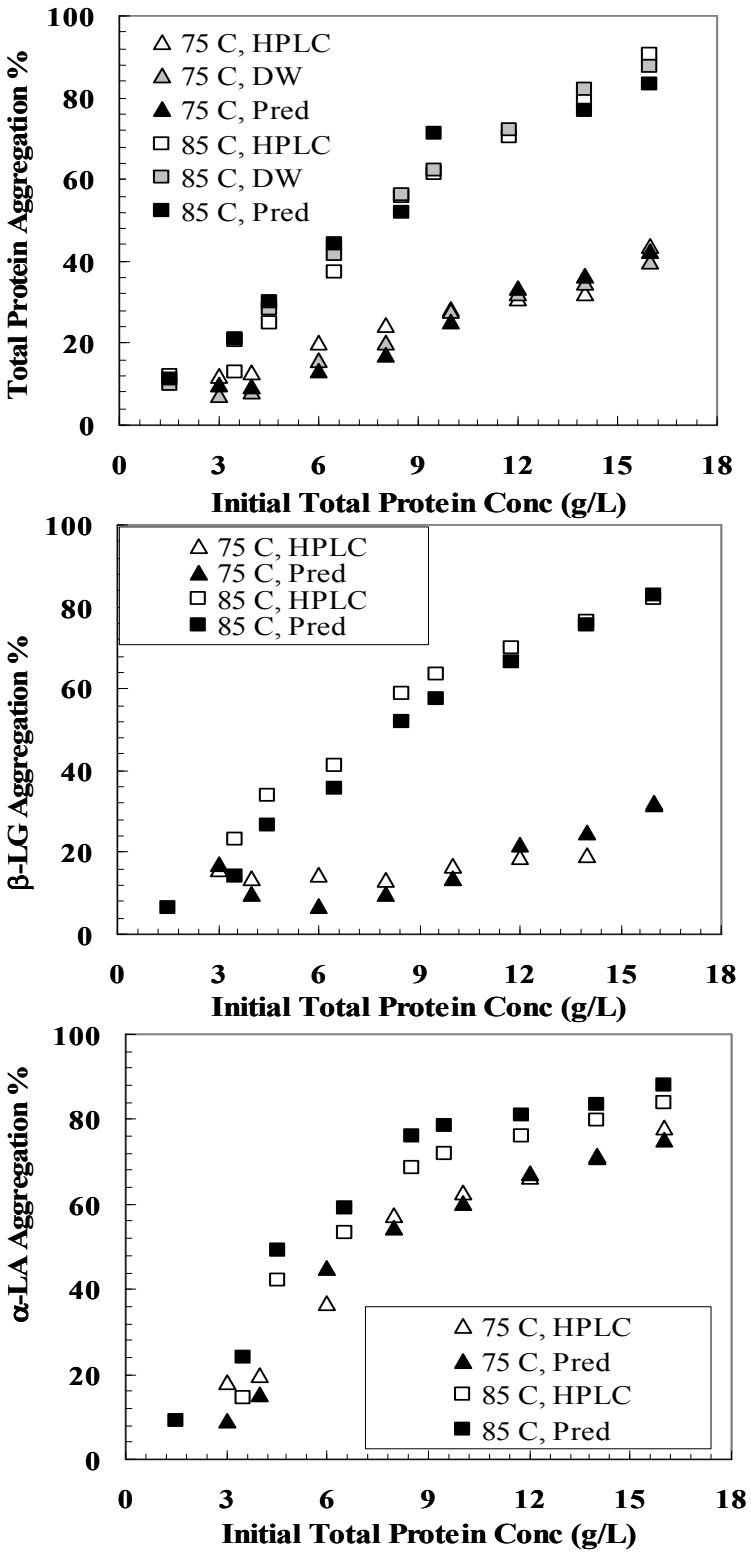


Figure 4-6. HPLC, dry weight measurements and model predictions of (a) total protein aggregation (b) β-LG aggregation and (c) α-LA aggregation at 75 & 85°C (Experiments 1 and 6 in Table 4.1).

4.4.5 PLS Soft Sensor

In practice it would be useful to be able to monitor β-LG and α-LA concentrations and ratios on line as this would permit real-time manipulation of operating conditions to modify and optimize the production process. The soft sensor proposed here is based on the use of fluorescence excitation-emission data acquired from the supernatant following centrifugation. These data were used to calibrate a model by using a Partial Least Squares (PLS) algorithm to predict the residual amount of α-LA and β-LG in solution after thermal treatment. The PLS regression model was calibrated by regressing the concentration of α-LA and β-LG in a sample where concentrations were determined by HPLC analysis against the fluorescence spectra of that sample. As explained in the Materials and Methods, the experimental data collected for 88 different β-LG/α-LA samples corresponding to different thermal treatments processes (one fluorescence spectrum per case) was divided into two sets, one was used for determining the PLS model (calibration set) and the other for testing the model (testing set). The first dataset (Table 4-1, Experiment 2) consisting of spectra acquired for 16 supernatant solutions of protein mixtures heat-treated at 85 ºC and covering a wide range of β-LG and α-LA concentrations was used to develop and calibrate the PLS soft-sensor model. The X matrix for the calibration set can be expressed mathematically as 16x1700 where each row is an individual spectrum and the 1700 columns represent fluorescence intensity measurements for different combinations of excitation wavelengths between 260-350 nm and emission wavelengths between 280-450 nm. The 16 rows are individual fluorescence spectra acquired for different supernatant solutions. The output Y matrix for the calibration

contains two response variables, i.e. the concentration of β-LG and the concentration of α-LA determined by HPLC. The PLS regression was implemented to correlate the concentrations of α-LA and β-LG to the fluorescence spectra. When applied to the calibration set, the PLS algorithm yielded three latent variables or PLS components that are statistically significant, with a goodness of prediction by cross validation (Q^2) of about 94%. These PLS components capture 94% of the variance in the X matrix (fluorescence intensity) with the rest of the variability assumed to be due mostly to experimental error and instrumental noise. The first PLS component is the most significant and it accounts for 80% of the variance in the X-matrix and it shows a strong linear correlation with the individual protein concentrations determined by HPLC determinations. The second PLS component accounts for 15% of the variance in the X-matrix.

The physical relevance of the current PLS model can be interpreted by examining the weight vectors (i.e. the weight vector of the first latent variable and the weight vector for the second latent variable). Plotting each individual weight vector against wavelength provides the so-called weight spectra (Figs. 4-7 and 4-8) which help in identifying important spectral regions with significant contribution to the prediction ability of the PLS model (Elshereef *et al.,* 2006). The loading weight spectrum of the first PLS latent variable (Fig. 4-7) reveals one peak with an emission maximum around 330-335 nm. Such a peak can be ascribed to the tryptophan residue in a relatively hydrophobic environment (Reshetnyak & Burstein, 2001).

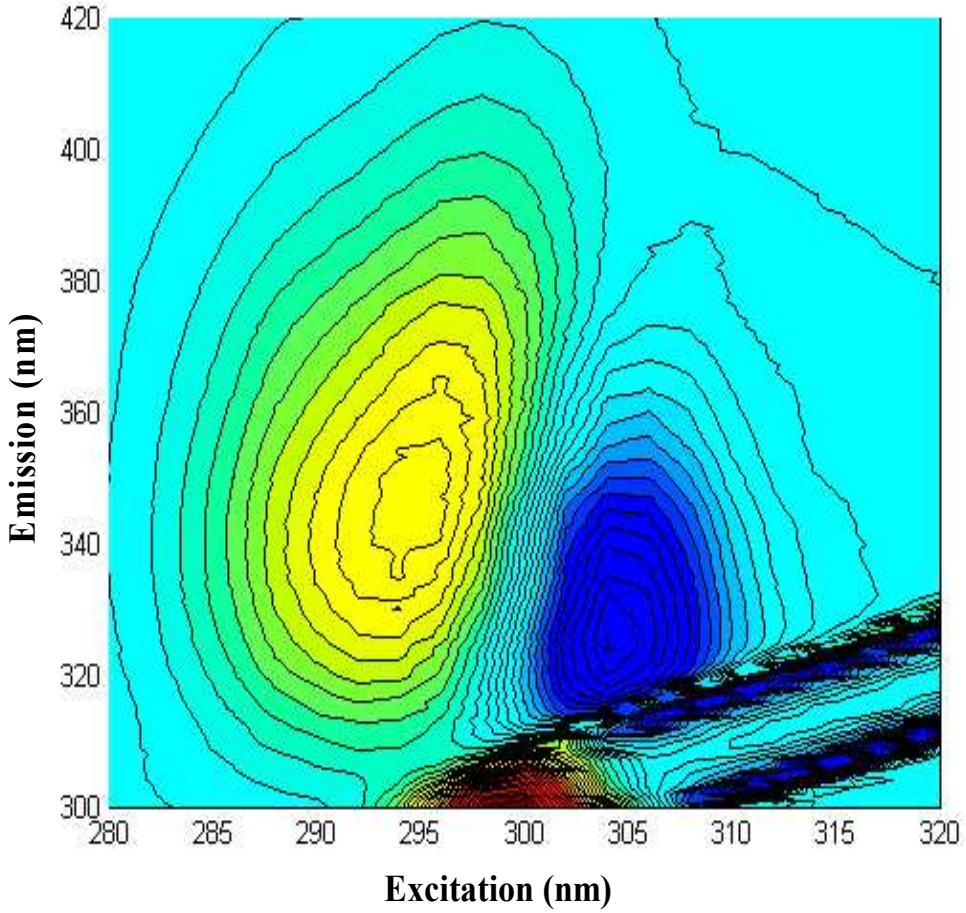


Figure 4-7. Weight spectrum of the first PLS component.

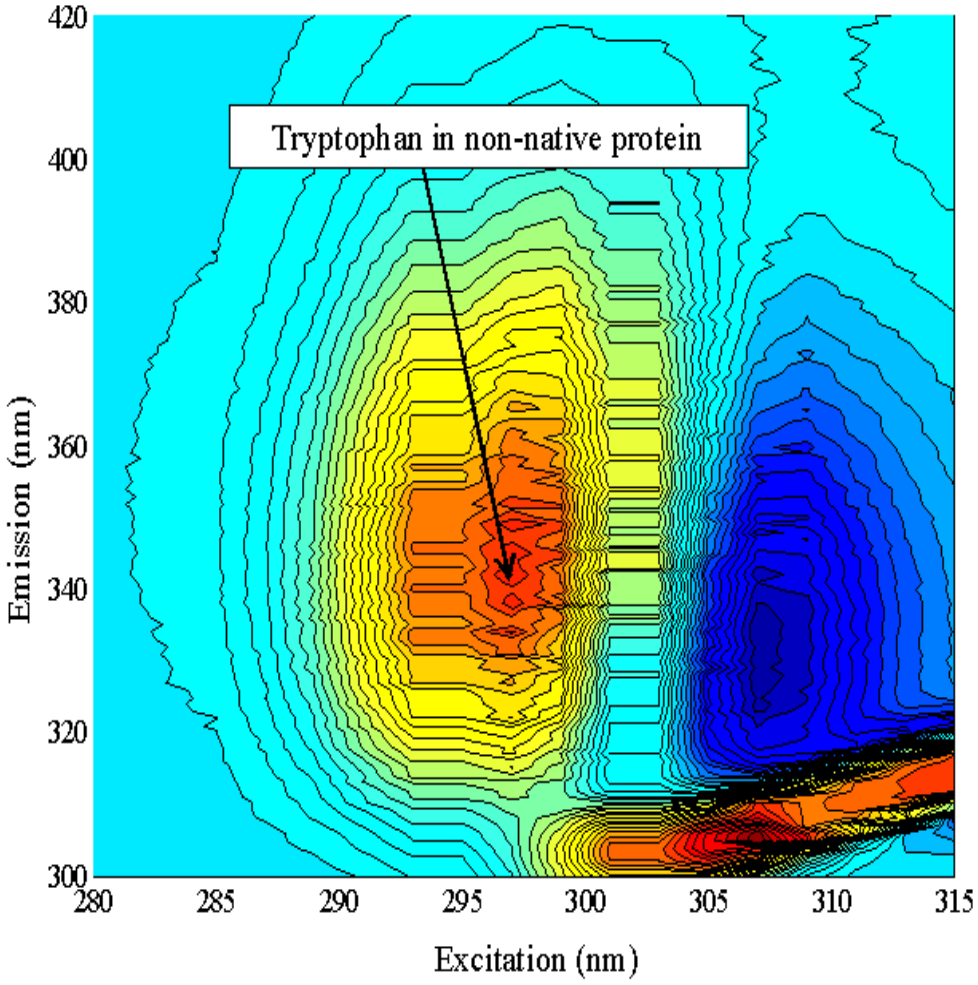


Figure 4-8. Weight spectrum of the second PLS component.

Fluorescence spectra of pure solutions of β-LG and α-LA at room temperature and pH 3.9 (data not shown) show emission intensity maxima at 330 nm and 335 nm, respectively which is consistent with published literature for these proteins under native conditions (Renard *et al.,* 1998; Svensson *et al.,* 1999). In this context it can be argued that the first PLS component that is strongly correlated with the soluble amount of β-LG and α-LA in the supernatant (Fig. 4-9) captures the fluorescence spectral region that is most likely due to the native tryptophan content of the proteins. The two different relationships observed between the protein concentration and the first PLS component for β-LG and α-LA (Fig. 4-9) may be explained by the differences in their tryptophan content.

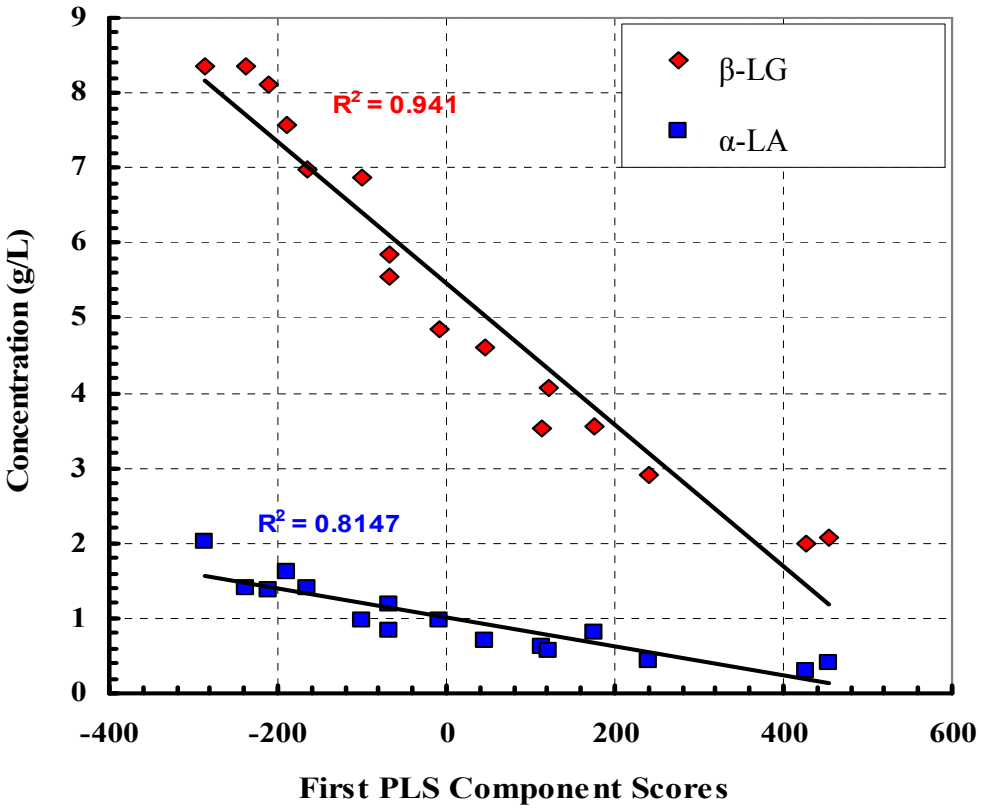


Figure 4-9. Measured protein concentrations of α-LA and β-LG versus scores of the first PLS component.

The PLS weight spectrum for the second PLS component provides two different spectral signatures possibly of proteins with a structural conformation (Fig. 4-8) that is different from that of the native form. The second PLS component may reflect the presence of non-native forms of β-LG and α-LA and their aggregates that were still soluble in the supernatant solution. The calibration model was then tested for accuracy using the remaining data sets (Table 4-1) consisting of fluorescence spectra acquired for 72 supernatant solutions. These fluorescence data were fed to the PLS model and the corresponding amounts of protein in the supernatant after aggregation were predicted. Model predictions for both β-LG and α-LA protein concentrations were compared with HPLC measurements. All results show very good agreement between PLS model predictions and HPLC measurements as seen in Fig. 4-10 indicating the robustness of the model at least for the range of conditions in this study. The strength of the model is especially evident in that only 16 samples were used for calibration resulting in a model that provides very good prediction for 72 other samples that were not used during the model calibration step. As expected, Fig. 4-10(b) shows that the error increases as the concentration of β-LG increases. The maximum errors are of the order of 0.1 g/L for α-LA and 0.9 g/L for β-LG, which correspond to approximately 5% and 10% of the total variation in concentration values of the two species.

Generally the fluorescence spectra of different species may overlap making it more difficult to predict their individual contribution to the spectra when these species are present in a mixture. For example, the typical emission profiles of α-LA and β-LG are presented in Figure 4-11. However, the results in the current study illustrate that

multivariate models can efficiently deconvolute multi-wavelength fluorescence spectra collected for a protein mixture and thereby provide a fairly accurate quantification of respective proteins despite the significant overlap of their emission profiles.

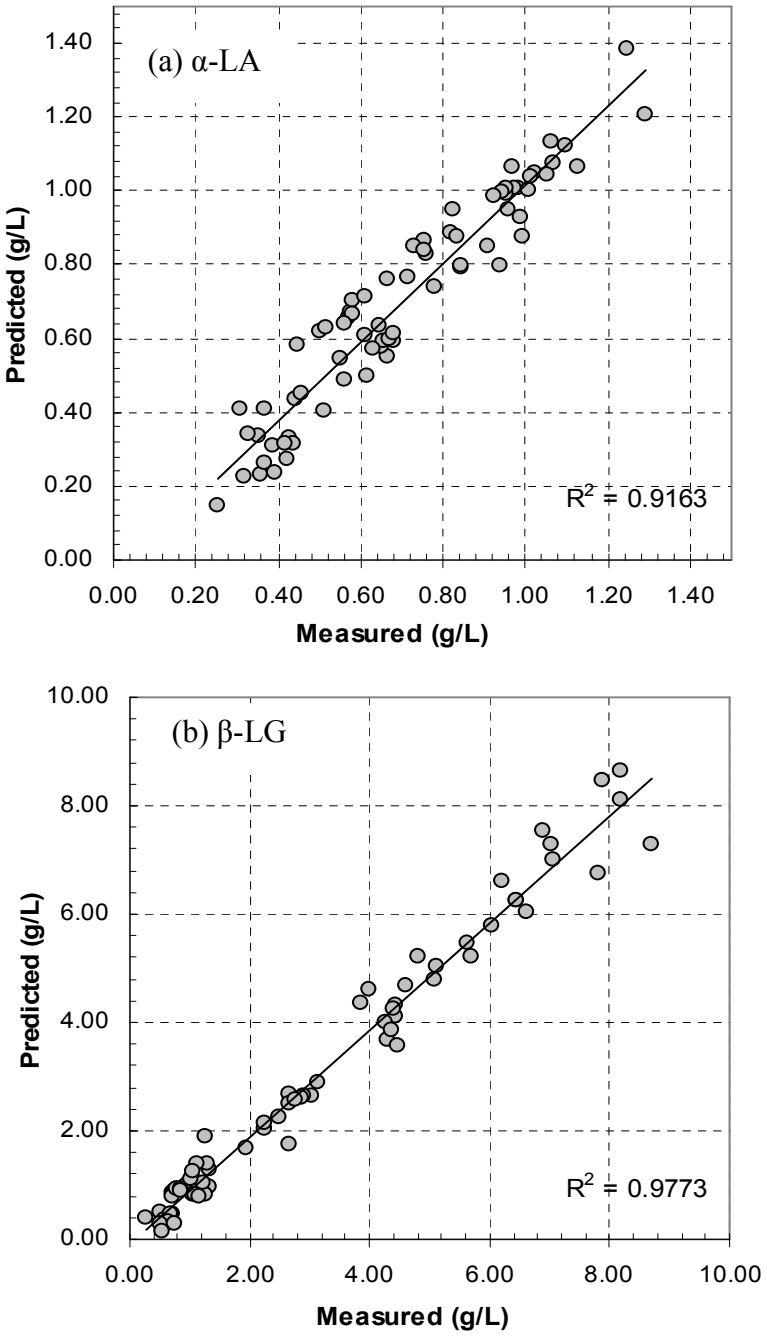


Figure 4-10. A comparison between PLS model predictions (line) and HPLC measurements for (a) α-LA and (b) β-LG (b) for 72 protein mixtures.

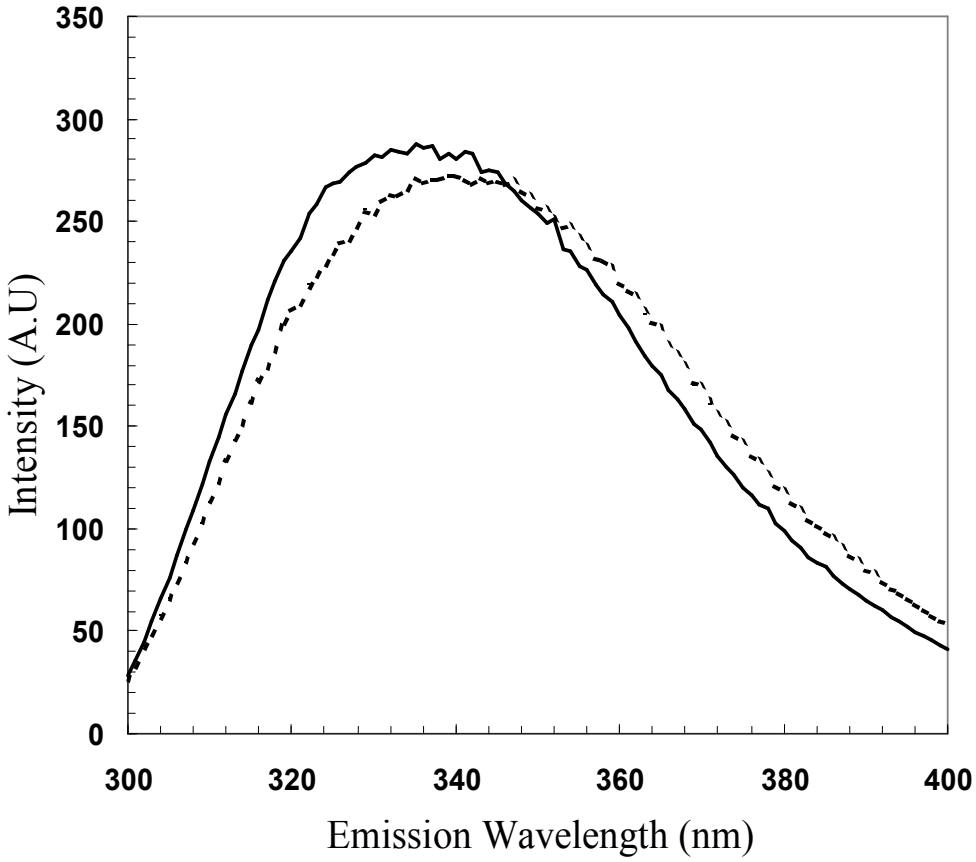


Figure 4-11. Emission spectra for β-LG (—)(3 g/L) and α-LA (-----) (7 g/L) at pH = 5.0 in citrate buffer with an excitation wavelength of 295 nm.

4.4.6 PLS/mass Balance Algorithm for Prediction of α-LA and β-LG Aggregation Behavior

The amount of aggregated protein can also be predicted by using the PLS-soft sensor predictions of the soluble amounts of protein before and after heat treatment. This can be done by calculating the difference between the inferred concentration before and after treatment and then, the amount of each aggregated protein can be inferred from a simple mass balance equation given by Equation 4-2. The result of the calculation of the aggregates by using this combination of the soft sensor predictions and the mass balance is illustrated by Figs. 4-6 & 4-12. As can be seen, model predictions for

protein aggregation show very good agreement with HPLC and dry weight measurements.

In summary, protein solubility and aggregation were monitored for a wide range of process conditions defined by different protein ratios, total protein amounts, different pH and different heat treatment temperatures. Given that inline centrifugation is likely untenable for most practical production applications, fluorescence spectroscopy was shown to be a suitable approach for the development of a chemometric-based in-process assay of protein concentrations that would have to be run off-line for the range of conditions considered in this work. Also, it was determined that the individual protein aggregation behavior during thermal treatment under different conditions can be effectively estimated by combining fluorescence data collected from supernatant with a simple mass balance approach.

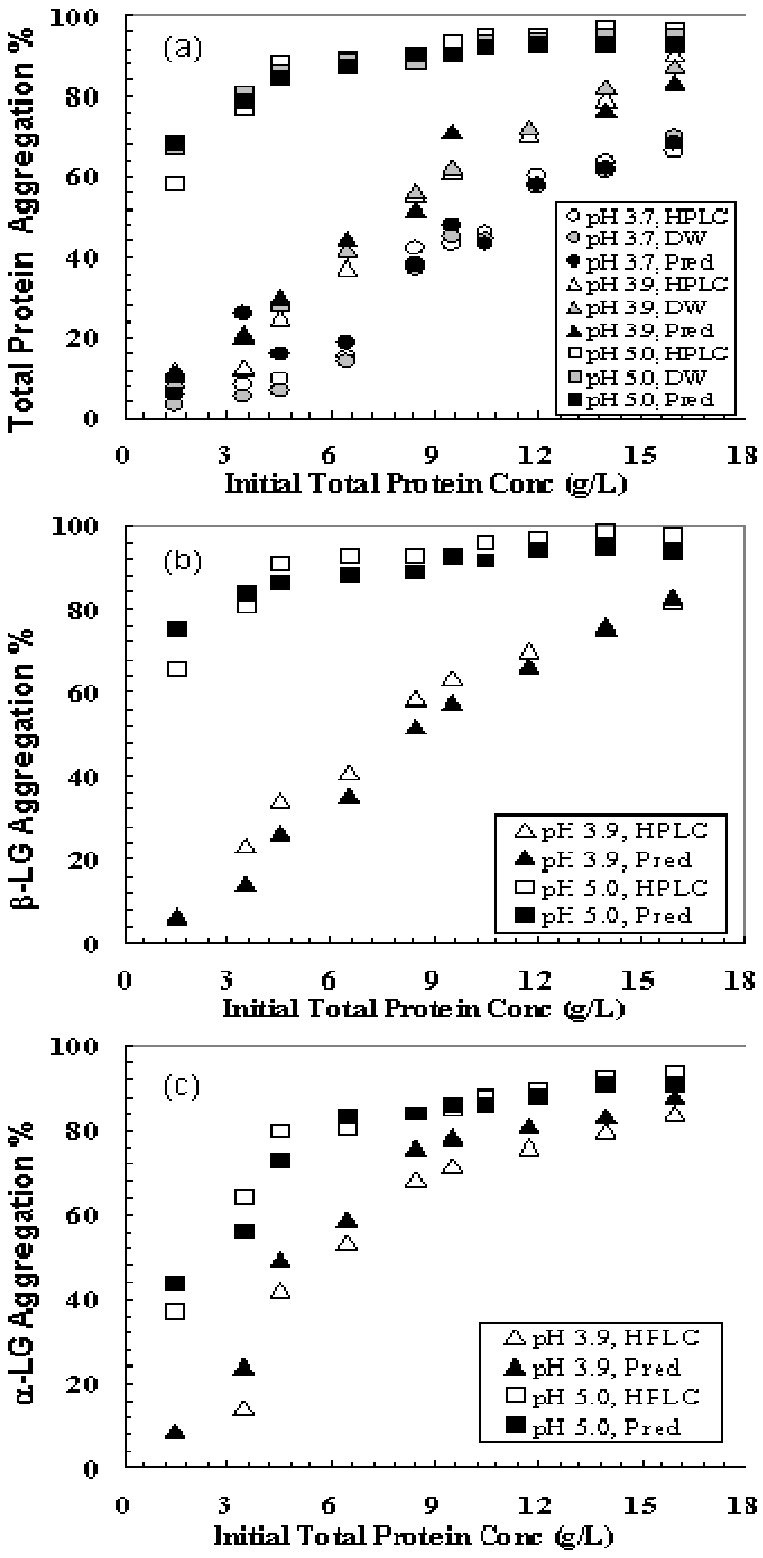


Figure 4-12. HPLC, dry weight measurements and model predictions of (a) total protein aggregation (b) β-LG aggregation and (c) α-LA aggregation at different pHs (Experiments 5, 6 & 8 in Table 4.1).

CHAPTER 5

Monitoring the Fractionation of a Whey Protein Isolate during Dead-end Membrane Filtration using Fluorescence and Chemometric Methods*

Protein fractionation using membrane-based ultrafiltration (UF) is a separation process commonly used in a broad spectrum of industries. During membrane-based separation of proteins, changes in protein concentration of the permeate and retentate streams occurs over time. The current work proposes a new approach for monitoring the changes in concentrations of proteins in both permeate and retentate by making use of data collected using fluorescence spectroscopy and intrinsic protein fluorescence analyzed by multivariate statistical techniques. Whey protein isolate consists mainly of α-lactalbumin (α-LA), β-lactoglobulin (β-LG) and small proportion of bovine serum albumin (BSA) and was used as a model system in this study. A fiber optic probe (FOP) was used to acquire multi-wavelength fluorescence spectra for permeate and retentate streams at different times during UF-based separation of the components from a multi-component solution. Multivariate regression models were developed for predicting the concentrations of α-LA, β-LG and BSA by establishing a calibration model between data acquired using the FOP and the corresponding protein concentration levels measured by size-exclusion chromatography.

---

* Elshereef R, Budman H, Moresoli C, Legge R. Submitted to Biotechnology Progress Journal.

The model was validated using FOP data that were not previously used for calibration of the regression models. This comparison showed that concentrations of α-LA, β-LG and BSA could be predicted directly from FOP data within reasonable accuracy by making use of multivariate calibration tools. This approach has several attractive features including that it is non-destructive, fast and relatively simple to perform. This technique has potential practical applications as it could offer the opportunity for *in situ* monitoring of membrane filtration processes by tracking individual protein transmission, selectivity of fractionation, protein accumulation on the membrane surface and the product yield and could be used for product quality control by monitoring product purity for different batches to ensure minimal batch to batch variation.

5.1 INTRODUCTION

Fractions enriched in α-LA and β-LG can be produced from whey using membrane-based protein ultrafiltration (UF). Protein fractionation using ultrafiltration has many uses in a broad spectrum of industries. During the UF of protein mixtures, changes in permeate and the retentate composition and protein transmission behavior (ratio of permeate to retentate concentrations) occur with time and may change with different pH, ionic strength and transmembrane pressures conditions (Huisman *et al*., 2000). Such changes are influenced by several effects that simultaneously occur on the membrane surface and inside the pores. Amongst these effects are size of the protein molecules and their aggregation behavior, protein-membrane interactions and the properties of the fouling layer (Huisman *et al*., 2000). The relative importance of these phenomena is determined by the transmembrane pressure and the physicochemical conditions related to pH and ionic strength. Key filtration performance indicators such as protein transmission, product recovery and separation efficiency can be expressed mathematically as a function of the concentration of proteins in the permeate and retentate (Ghosh, 2003). Therefore, fast and accurate quantification of the individual proteins in the retentate and permeate streams is essential for continuous monitoring of an ultrafiltration process. In earlier studies, classic qualitative and quantitative analytical techniques including UV/Vis spectrophotometery and size exclusion chromatography were used for the identification and quantification of different proteins in permeate and retentate, and based on these measurements, membrane separation performance was evaluated in terms of protein transmission, product yield and selectivity (Ghosh, 2003). However, these techniques suffer from some practical

limitations that make them difficult to apply on-line. Size exclusion chromatography, which has the advantage of providing detailed analysis of the composition of the retentate and the permeate streams, is considered to be time-consuming, labor intensive and requires preparation of chemical reagents. In addition, UV-absorbance based detection may not be sensitive enough to monitor dilute protein components. Such limitations for UV detection were implicitly reported by Wan *et al.* (2006). To overcome this limitation Wan *et al.* (2006) analyzed feed, retentate and permeate compositions using FPLC by concentrating the permeate samples prior to analysis with centrifugal filter devices. Fluorescence spectroscopy is considered to be far more sensitive than UV-Vis spectroscopy and can easily detect concentrations of the species of interest as low as 10^{-10} to 10^{-12} M while UV-Vis spectroscopy requires at least 10^{-8} M (Deshpande, 2001). Another advantage of fluorescence spectroscopy is that data is multidimensional consisting of two spectra, that is excitation and emission spectra (Deshpande, 2001) so fluorescence has higher potential information content for resolving mixtures than UV-VIS absorbance (Baker, 1991). Despite the fact that fluorescence spectroscopy provides some of the most sensitive and selective methods of analysis of many compounds, it has not been widely employed for monitoring protein transmission for membrane-based separation processes or for identifying the compositional changes of retentate and permeate during ultrafiltration. The earliest work that utilized fluorescence spectroscopy for continuous monitoring of protein fractionation is that of Crespo *et al.* (1999). Crespo *et al.* (1999) developed an on-line fluorescence detection technique for monitoring the transmission of β-LG and γ-globulin using fluorescent probes. The two proteins were labeled with different

fluorescent probes which had unique absorption and emission spectral profiles that did not overlap. Thus on-line detection of protein-fluorescent labeled conjugates was performed with a fluorescence detector that was programmed at appropriate excitation and emission wavelengths. The transient transmission behaviors of β-LG and γ-globulin through the membrane were identified by the transmission of the corresponding protein-fluorescent label conjugate. Crespo *et al*. (1999) indicated that this technique would allow off-line and continuous on-line monitoring of protein transmission. Crespo *et al.* (1999) realized that the intrinsic fluorescence of different proteins is similar resulting in significant overlap in absorption and emission spectra and consequently, to resolve this information overlap, they labeled their proteins with different fluorescent probes. The drawback of any sort of protein labeling is that it may introduce changes to the protein surface chemistry and to the overall protein charge which may alter protein folding properties and its aggregation behavior (Crespo *et al.*, 1999). In addition, the technique has some practical limitations since it requires removal of the fluorescent label downstream. Hence, there is a strong motivation to avoid the use of fluorescent labeling in order to preserve the native state of the protein product. What is proposed here is to use intrinsic protein fluorescence. Intrinsic protein fluorescence originates from the presence of three amino acid residues: tryptophan, tyrosine, phenylalanine in the protein polypeptide chain (Guilbault, 1973). The intrinsic fluorescence originating from tryptophan residues accounts for more than 80% of the total intrinsic fluorescence emissions of a native protein (Lacowicz, 1999). The number of these amino acids can vary greatly from one protein to another. It is not only the number of amino acids in the protein polypeptide chain that determines the

intrinsic protein fluorescence characteristics but it is also the microenvironment surrounding these amino acids that does play an important role (Guilbault, 1973; Lacowicz, 1999). Despite the significant spectral overlap, contribution of individual fluorophores to the overall fluorescence can be deciphered by using fluorescence measurements at multiple excitation and emission wavelengths. For instance, upon excitation above 295 nm, the intrinsic fluorescence emission of a native protein is only due to tryptophan residues. Meanwhile, both tyrosine and tryptophan contribute to fluorescence emission spectra upon excitation in the range (280- 295 nm) (Mycek and Pogue, 2003). Low-wavelength excitation (220–230 nm) was also found to provide useful spectral information that is not of less importance than the high-wavelength excitation (280-295 nm) usually reported in the literature (Mayer *et al.*, 1999). Therefore, quantitatively detecting constituents of the protein mixture would require collecting an EEM fluorescence landscape that covers the spectral signatures of the three intrinsic fluorophores. To handle the complex fluorescence signals obtained when analyzing multicomponent protein solutions and to resolve the issue of overlapping information, multivariate regression was successfully applied to excitation-emission matrix (EEM) fluorescence landscape. Elshereef *et al.* (2007) demonstrated that intrinsic fluorescence spectra of α-LA and β-LG can be effectively de-convoluted by utilizing multivariate regression modeling and that the respective protein concentrations can be successfully estimated from two-dimensional fluorescence spectra of protein mixtures with reasonable accuracy. The current work is distinct from earlier work (Elshereef *et al.*, 2007) in three key aspects. First, the fluorescence data of the whey protein isolate consisting of α-LA, β-LG and BSA were acquired in the

synchronous scan mode which is much faster to obtain than the full two dimensional spectra used in previous studies and it is consequently more amenable for potential on-line applications. Secondly, the current study used a fiber optic probe (FOP) instead of cuvettes to evaluate the potential application of this technique for possible on-line monitoring of protein compositions in mixtures. Finally, the focus of the current application was to evaluate the feasibility of tracking performance of an UF separation process using fluorescence-based estimates of protein composition.

5.2. MATERIALS AND METHODS

5.2.1 Materials and Preparation of Protein Stock Solutions

β-Lactoglobulin (β-LG) and α-lactalbumin (α-LA) were in their powdered form (lot JE 007-3-921 and JE 003-3-922), were of 95% purity, and were donated by Davisco Foods International (LeSueur, USA). Bovine serum albumin (95% purity) was from Sigma Aldrich International. BiPRO is a whey protein isolate (WPI) consisting mainly of β-lactoglobulin and α-lactalbumin and was donated by Davisco Foods International (Le Sueur, USA). The composition of Bipro as provided by the manufacturer was (w/w) 88.1% protein (N x 6.38), 9.89% moisture, 0.3% fat and 1.84% ash (0.66% Na^+, 0.075% K^+, 0.0086% Mg^{2+}, and 0.094 % Ca^{2+}). The protein content of Bipro was: 14.9% α-LA, 74.9%, β-LG, 3.2% immunoglobulin and 1.5% bovine serum albumin (BSA) (Weinbreck et al., 2004). All other chemicals were of analytical grade and from Sigma Chemical Corp. (St. Louis, MO). Ultrapure water was used for the preparation of all samples and had a resistivity greater than 17.6 MΩ. Buffers were prepared using

ultrapure water and were micro filtered and degassed under vacuum using a 47 mm 0.45-μm Nylon membrane (PALL Corporation, Michigan, USA). Stock solutions of α-LA, β-LG and BSA containing 10 g/L total solids were prepared by dissolving the protein powder in 50 mM sodium phosphate buffer at the desired pH. A stock solution of 10 g/l of WPI was prepared by dissolving the appropriate amount of WPI in 50 mM sodium phosphate buffer at the desired pH.

5.2.2. Preparation of Calibration Samples

The purpose of this work was to quantify the concentrations of α-LA, β-LG and BSA in the permeate and retentate streams during ultrafiltration by making use of the intrinsic protein fluorescence. The first step for achieving this objective was to calibrate a fluorescence-based predictive model against a set of reference samples with known protein composition (calibration set). Sixty-four mixtures of β-LG, BSA and α-LA whose composition was randomly defined according to a 4^3 full factorial design were prepared and the fluorescence spectra for those samples acquired using a FOP.

5.2.3 Preparation of Filtration Feed Solution

The feed solution for all filtration experiments was prepared by adding appropriate volumes of α-LA and BSA stock solutions into a WPI stock solution at the desired pH to obtain final concentrations of 75 % w/w β-LG, 20 % w/w α-LA and 5 % w/w BSA.

5.2.4. Experimental Setup of the Ultrafiltration Apparatus

All filtration experiments were carried out as batch filtrations using a 75 mm diameter

Amicon ultrafiltration stir cell (Model 8010, Amicon Corp., Beverly, MA). The apparatus consisted of a cylindrical chamber with a capacity of 400 ml and a membrane area of 41.8 cm^2. The stirred cell was sandwiched between two identical flat Plexiglas plates, constructed in-house, and clamped via four steel bars (3/4"). The top plate housed:

(i) A sampling port fitted with a septum cap connected to a 3/6" stainless steel fitting with a 1/8" stainless steel tube that was positioned just above the membrane. The stainless steel tube allowed sampling from the retentate solution using a 5 ml gas-tight syringe

(ii) A stainless steel pressure release valve

(iii) A gas inlet port that was used to pressurize the cell from a high pressure nitrogen cylinder

The entire assembly was placed on a magnetic stir plate. The cell was pressurized with compressed nitrogen up to 200 kPa. The permeate (filtrate) was collected from the permeate port on a mass basis of 10 g fractions in test tubes supported in a custom test-tube rack on a digital mass balance (Scout Pro Balance, Ohaus Corp., Pine Brook, NJ) that was interfaced with a computer for on-line data collection using Labview (Labview 7.0, National Instruments, Mississauga, ON). The balance had an accuracy of 1 mg.

5.2.5. Filtration Experiments

All filtration experiments were carried out batch-wise starting with an initial feed volume of 350 mL. Filtration experiments were conducted using a 30 kDa molecular

weight cut-off composite regenerated cellulose membrane (Millipore Corp., Bedford, MA). The membranes were flushed with deionized distilled water prior to use to remove any residual chemical agents. The stir cell was filled with protein solution of known protein concentration, pH, and ionic strength. The contents of the stir cell were kept well stirred using a magnetic stir bar. Once 10 g of permeate had been collected, the permeate port was transferred to the next collection tube and the FOP was inserted into the permeate solution for data collection. After acquiring the spectra the probe was removed and inserted into the next sample and so on. At various time intervals, samples of 4-10 ml of retentate solution were withdrawn from the stir cell through the sampling port using a gas-tight syringe. Retentate samples were labeled and then stored at 4°C until the experiment was completed (about 3 h). Filtration experiments were stopped when 50-ml of unfiltered solution remained on the retentate side. After completing the filtration experiment, retentate samples were then removed from storage for FOP and HPLC analysis. Retentate samples were diluted 10-fold for FOP analysis to minimize spectral shifts. Fluorescence data were also collected for the retentate samples using the FOP. To assess reproducibility a few ultrafiltration experiments with different combinations of experimental conditions (pH, transmembrane pressure) were repeated three times. The reproducibility for permeate mass and concentrations of permeate and retentate was reasonably good with a standard deviation of 10%.

5.2.6 HPLC Analysis

The composition of α-LA, β-LG and BSA in feed, permeate, and retentate streams

were analyzed using an HPLC system equipped with the size exclusion column TSKgel G2000SWXL (7.8 x 30mm) (TOSOH Bioscience, Montgomeryville, PA). The mobile phase was 20 mM sodium phosphate buffer containing 100 mM NaCl (pH 6.8). The flow rate was fixed at 1.0 mL/min and good separation between α-LA, β-LG and BSA peaks was observed. Total elution time in all cases was 15 min. No peak broadening was observed in the chromatograms indicating that the proteins were stable within the pH range of 5.0-7.0. Some peak broadening was observed for α-LA in the retentate samples at pH 2.8 which may indicate some degradation of α-LA. A calibration curve based on peak height measurements was generated for each individual protein by injecting standards at several known protein concentrations. The concentrations of different proteins in the initial feed, permeate, and retentate solutions were determined using the calibration curves. Prior to each HPLC analysis a new calibration curve was generated using new standards.

5.2.7. Fluorescence Measurements

The fluorescence spectra were acquired using a Varian fiber optic probe (FOP) assembly that was connected to a Varian Cary Eclipse Spectrofluorometer equipped with a Xenon flash lamp as the light source. The spectra were acquired using a synchronous scan mode at a scan speed of 600 nm/min, excitation and emission slit widths both set to 5 nm and PMT voltage of 800 V. Synchronous fluorescence spectra were recorded in the excitation range from 200 to 350 nm, wherein the wavelength interval between the emission and excitation wavelength was varied in the range 0 nm - 100 nm. Multi-wavelength fluorescence scans produced a 10×150 excitation-emission

wavelength pairs generating a matrix of data consisting of 1500 fluorescence intensity data points for any given sample. A higher PMT voltage than in previous work (Elshereef *et al.*, 2006; 2007) was used to compensate for the attenuation of the signal by the fiber optic bundle. Preliminary fluorescence measurements revealed that the fluorescence signal acquired with the probe is about 10-fold lower than the signals acquired using cuvettes. The time required to collect a full excitation-emission matrix scan was 3 min and 20 sec. The data were recovered in a manner which allowed direct computer processing. All spectra were corrected for background contribution by subtracting appropriate blanks containing only buffer. The FOP spectral measurements for standard ternary protein mixtures of α-LA, β-LG and BSA were calibrated against their respective concentration data from HPLC analysis for modeling a fluorescence-based PLS model. FOP spectral measurements of permeate and retentate samples with unknown protein composition were introduced directly as input to the fluorescence based PLS model and thus the concentrations of these proteins were estimated.

5.2.7 Evaluation of the Membrane Resistance

A resistance-in-series model was used to estimate fouling for this ultrafiltration approach to whey protein separation. The total membrane resistance R_{tot} (m^{-1}) was estimated using the following equation:

$$R_{tot} = R_m + R_f = \frac{\Delta P}{\mu_v J_v} \qquad \textbf{(5-1)}$$

where R_m (m^{-1}) is the specific membrane resistance of the clean membrane, R_f (m^{-1}) is

the added resistance due to fouling, J_v ($m^3 \cdot m^{-2} \cdot s^{-1}$) is the flux of the protein solution with time, μ (Pa·sec) is the permeate viscosity and ΔP (Pa) is the transmembrane pressure. The permeate viscosity was assumed to be equal to that of pure water at room temperature ($1x10^{-3}$ Pa·s). The flux of the protein solution J_v in the time interval Δt is given by Equation 5-2.

$$J_v = \left(\frac{1}{A_m} \right) \frac{\Delta V}{\Delta t} \qquad \textbf{(5-2)}$$

Where A_m is the effective membrane area (41.3 cm^2) and ΔV is the volume of the filtrate solution that was collected during the time interval Δt. Total resistance values were normalized by the specific membrane resistance R_m (m^{-1}) of the clean membrane and then the normalized resistance for different experiments was plotted versus filtration time. R_m (m^{-1}) was evaluated from clean water flux measurements at different transmembrane pressures.

5.3. CHEMOMETRIC MODELLING

Partial Least Squares Regression

For the purpose of calibration of a regression model, the experimental data in this study was divided into two categories: input data corresponding to the fluorescence spectrometric measurements, and output data consisting of β-LG, α-LA and BSA concentration values obtained by HPLC analysis. The matrix that contains the outputs to be predicted, i.e. HPLC measurements of β-LG, α-LA and BSA, were arranged in an output matrix to be referred to heretofore as matrix Y whereas the fluorescence

measurements were arranged into an input data matrix to be referred to as matrix X. The rows in matrix X correspond to different samples while the columns correspond to fluorescence intensities measured at different excitation-emission wavelength pairs (Elshereef *et al*., 2006; 2007) for each one of the samples. The input data is considered to be multivariate in nature since it consists, as mentioned above, of 1500 fluorescence intensity data points for each sample. Partial Least Squares regression (PLS) is a well-known chemometric tool for designing a calibration model which can be used to correlate the set of known measurements represented by the fluorescence data to the property to be predicted, i.e. the β-LG, α-LA and BSA concentrations. The PLS regression method has been chosen in this work to design a fluorescence-based predictive model since it provides lower sensitivity to noise for multivariate data sets with large numbers of highly correlated variables (Geladi and Kowalski, 1986; Qin and McAvoy, 1992). Collinearity is very high among the different fluorescence readings obtained at different combinations of emission and excitation wavelengths (Lemberge and Van Espen., 1999; Elshereef *et al.*, 2006). Hence, the original input variables *x* in matrix X are replaced by a smaller set of underlying new variables that are uncorrelated, mutually independent (orthogonal) and are mathematically represented by linear combinations of the original descriptors. These calculated linear combinations, referred to as latent variables (LVs) or principal components, are calculated to both provide good representation of the X-matrix and maximize the correlation between the input and the output (Qin and McAvoy, 1992). The optimum number of latent variables (LVs) and the goodness of prediction (Q^2) are determined by a cross-validation algorithm (Geladi and Kowalski, 1986; Qin and McAvoy, 1992;

Elshereef *et al.,* 2006). The data obtained from different samples were used for calibrating and testing the PLS predictive model as will be shown in the following sections. All computations were carried out using MATLAB 7.0 (MathWorks, Natick, MA) along with the PLS Toolbox 3.5 (Eigenvector Research Inc., Manson, WA). To develop the fluorescence based PLS model the samples were divided into two different sets, a calibration set consisting of the samples used for calibrating the PLS model and a validation set consisting of samples that were used to independently test the predictive accuracy of the calibrated model.

5.4. RESULTS AND DISCUSSION

5.4.1. Development of Fluorescence-based Model for Simultaneous Determination of α-LA, β-LG and BSA in a WPI mixture

The challenge in using intrinsic protein fluorescence is that the intrinsic fluorescence spectra for proteins can significantly overlap. Thus, detailed spectral analysis, combined with chemometric methods to isolate spectral features, can certainly improve the success of distinguishing between different proteins. This would potentially require collecting EEM landscapes with high spectral resolution in both dimensions (excitation and emission wavelength). Preliminary fluorescence measurements were performed using the FOP to determine the ranges of excitation and emission wavelength of interest. It was found that all spectral information are contained in the excitation range of 200–350 nm and emission range of 250–450 nm (Figure 2-10). Scanning in such a broad spectrum range using 1-nm increment for both excitation and emission ranges will generate an EEM that contains 30,000 intensity measurements. However, not all of

excitation-emission pairs are relevant to the intrinsic fluorescence of the sample. The intrinsic fluorescence of the sample is located in the region where the emission wavelength is longer than the excitation wavelength. Meanwhile, other areas in the EEM landscape (i.e. the triangular-shaped region in the EEM where the emission wavelength is less than the excitation wavelength as well as the Rayleigh scatter lines) are not correlated to the intrinsic fluorescence of the sample (Ohno and Bro, 2006). Therefore, with the goal of reducing the time for data acquisition and filtering out the non-relevant data, multiple synchronous fluorescence scans were collected instead of multiple emission scans. Synchronous fluorescence scanning mode involves the simultaneous scanning of the excitation and emission monochromators at a constant wavelength difference $\Delta\lambda = \lambda_{emission} - \lambda_{excitation}$ and thus both excitation and emission wavelengths are varied according to the formula: $\lambda_{em} = \lambda_{ex} + \Delta\lambda$. Multiple synchronous spectra were collected by varying $\Delta\lambda$. Multiple synchronous spectra in the excitation range 200-350 nm were recorded at the following wavelength intervals ($\Delta\lambda$): 10, 20, 30, 40, 50, 60, 70, 80, 90 and 100 nm.

The advantage of such an approach is that it selects a subset of the EEM landscape with less data as compared to the full excitation-emission spectra while retaining relevant fluorescence spectral information; sections of the spectra that are not sensitive enough to the concentration changes to be monitored are excluded.

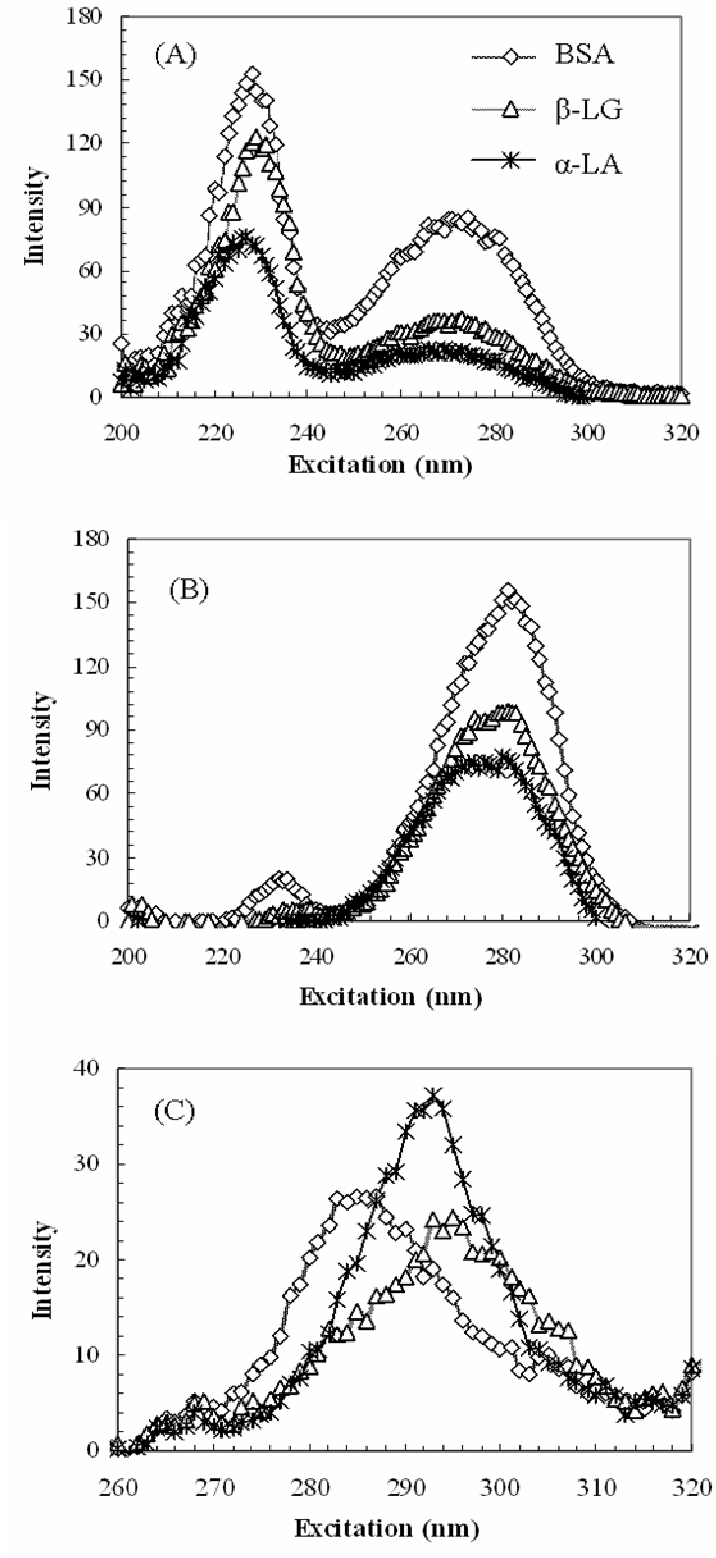


Figure 5-1. Synchronous fluorescence spectra at Δλ=100 nm (A), Δλ=60 nm (B) and Δλ=10 nm (C) for the individual whey proteins (α-LA, β-LG and BSA).

The use of the synchronous mode resulted in a reduction of the size of the fluorescence data set to be used for prediction by approximately 20-fold as compared to the case where the full spectrum is used. Figure 5-1 shows synchronous spectra acquired for protein solutions (α-LA, β-LG and BSA) at Δλ=10, 60 and 100 nm. When Δλ=10 nm, a spectrum characteristic of the fluorescence of the protein's tyrosine residues is observed whereas for Δλ=60 nm, a spectrum characteristic of the fluorescence of the protein's tryptophan residues is observed (Ma *et al*., 1999; Tan *et al*., 2005). Upon comparison of synchronous spectra at Δλ=100 nm acquired for protein solutions α-LA, β-LG and BSA, two characteristic peak maxima are observed at wavelengths 225 and 275 nm. The peak to peak ratio at excitation wavelengths 225 and 275 nm is of great interest for discriminating between different proteins. From the foregoing it can be seen that during the scan, depending on the scanning interval Δλ and the Stoke's shift, every fluorophore in a given sample will presumably contribute to fluorescence at different positions, leading to improved resolution and to a specific signature even for cases where the fluorophores overlap and/or interact significantly (Rao, 1991).

Accurate quantification of α-LA, β-LG and BSA in the permeate and retentate relies on a robust partial least squares (PLS) based regression between fluorescence spectra of the samples and their corresponding concentrations measured by HPLC. The first step in the simultaneous determination of different proteins in a mixture by PLS methodology involved constructing the calibration matrix for the ternary protein mixture. Sixty four synthetic ternary mixtures containing the individual whey proteins in different proportions were randomly designed and used to develop the calibration models. A cross-validation method that consisted of leaving out one sample at a time,

was employed in order to select the number of factors in the PLS algorithm. For the 64 calibration spectra, PLS calibration was performed on 63 calibration spectra, and the calibration model was then used to evaluate the concentration of the sample left out during the calibration process. This process was repeated 64 times until each calibration sample had been left out once. The concentration of each sample was then predicted and compared with the known concentration of this reference sample, and the prediction residual sum of squares (PRESS) was calculated. From a plot of the PRESS against the number of factors for each individual component, it was concluded that good PLS-based regression models for each of the proteins considered in this study could be obtained with three principal components. Three principal components is the optimal number of principal components at which the smallest error (PRESS) occurs. The PLS model was tested using two different types of validation data consisting of samples that were not included in the calibration as follows: (i) 12 synthetic ternary protein mixtures of β-LG, BSA and α-LA; (ii) 10 real samples of the whey WPI solution in phosphate buffer spiked with different amounts of β-LG, BSA and α-LA

5.4.2. Validation of the PLS Model on Retentate and Permeate Samples

As a first step, the synchronous fluorescence spectra for the retentate and the permeate samples were acquired using the FOP and examined. From a qualitative perspective the scans of the retentate and the permeate reveal that the retentate spectra have certain distinct visual characteristics that makes them different from permeate spectra (Figure 5-2).

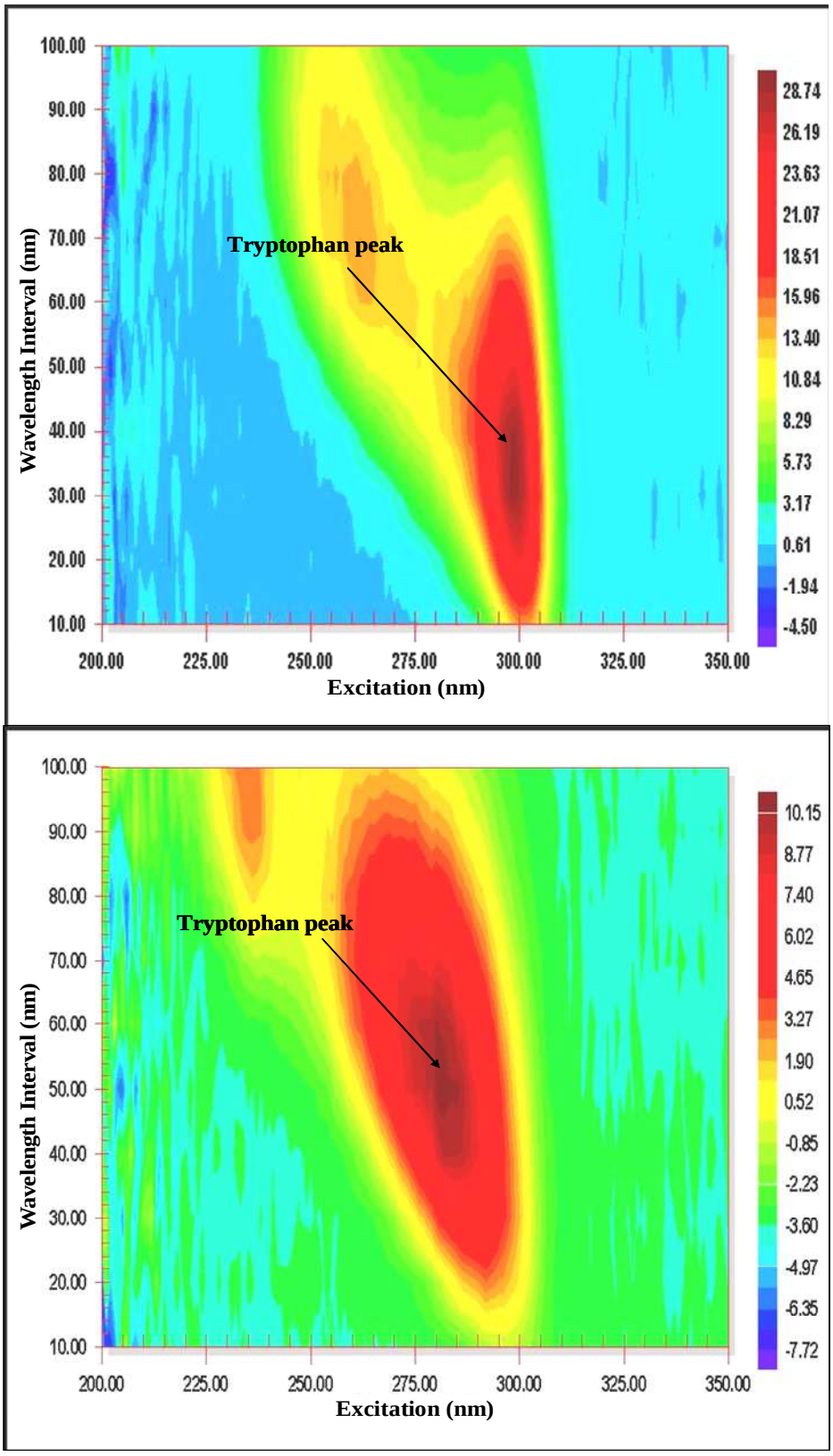


Figure 5-2: Spectra for retentate (top) and permeate (bottom) for ultrafiltration of a WPI solution.

Excitation spectra of the final permeate showed that the maximum excitation peak appeared at 285 nm, whereas the maximum excitation for retentate were *red-shifted* by 15 nm. Such spectral differences between the retentate with high protein content and the permeate with low protein content is most likely due to the differences in the molecular environment of the fluorophore (tryptophan) is experiencing. Fluorescence spectra are influenced by several phenomena (resonance energy transfer and quenching) that are related to the concentration, intra- and intermolecular interactions and the local environment of the inherent fluorophores (Christensen *et al.*, 2006). Spectral differences were observed between permeate and retentate streams, where the latter was characterized by significant fluorescence quenching and excitation red-shift (Figure 5-2). Although these two phenomena may impart a distinctive uniqueness to the fluorescence spectrum of the retentate stream due to the high protein content, they were found to induce nonlinear dependencies between the fluorescence data and the corresponding protein concentration levels which are poorly handled by the linear PLS model. For instance, using the retentate spectra as input to the current PLS model for prediction of the individual protein concentrations resulted in poor predictions (data not shown for brevity). This is likely because PLS is well calibrated over the concentration range where the relationship is predominantly linear between the fluorescence signal and the concentration of each protein but the retentate protein concentrations lie mostly within the range of values where the relation between fluorescence to protein concentration is nonlinear. To handle the non-linearity in the fluorescence data matrix would require combining PLS with non-linear tools such as Artificial Neural Networks. For the sake of simplicity, the objective of this study was to resolve the

spectra using a linear PLS methodology. This required collection of fluorescence spectra that were devoid of significant red-shifts and fluorescence quenching effects in the high concentration ranges. Therefore, prior to fluorescence measurements, samples collected from the retentate stream were diluted 10-fold using phosphate buffer solution at the same pH value of the initial protein solution prior to FOP analysis. The permeate was not diluted because the protein concentration in these samples was relatively low and the fluorescence spectra did not display significant excitation red-shifts. Concentrations of α-LA, β-LG and BSA predicted from the diluted retentate samples were multiplied by the dilution factor and then compared to values determined by HPLC analysis. Very good agreement between predictions and measured values indicated the applicability of the proposed method for simultaneous determination of α-LA, β-LG and BSA. The correlation coefficients for α-LA, β-LG and BSA were 0.99, 0.98 and 0.88, respectively.

5.4.3. Protein Concentration Profiles

The feed concentration used in this set of experiments was 4.0 kgm^3 WPI in the appropriate buffer. Figure 5-3 is a typical HPLC chromatograms for retentate and permeate samples obtained at pH 5.6 after a total of 300 ml of permeate had been collected. A comparison of the chromatograms for retentate and permeate reveal that both β-LG and α-LA are transmitted through the membrane, whereas the amount of BSA transmitted is negligible.

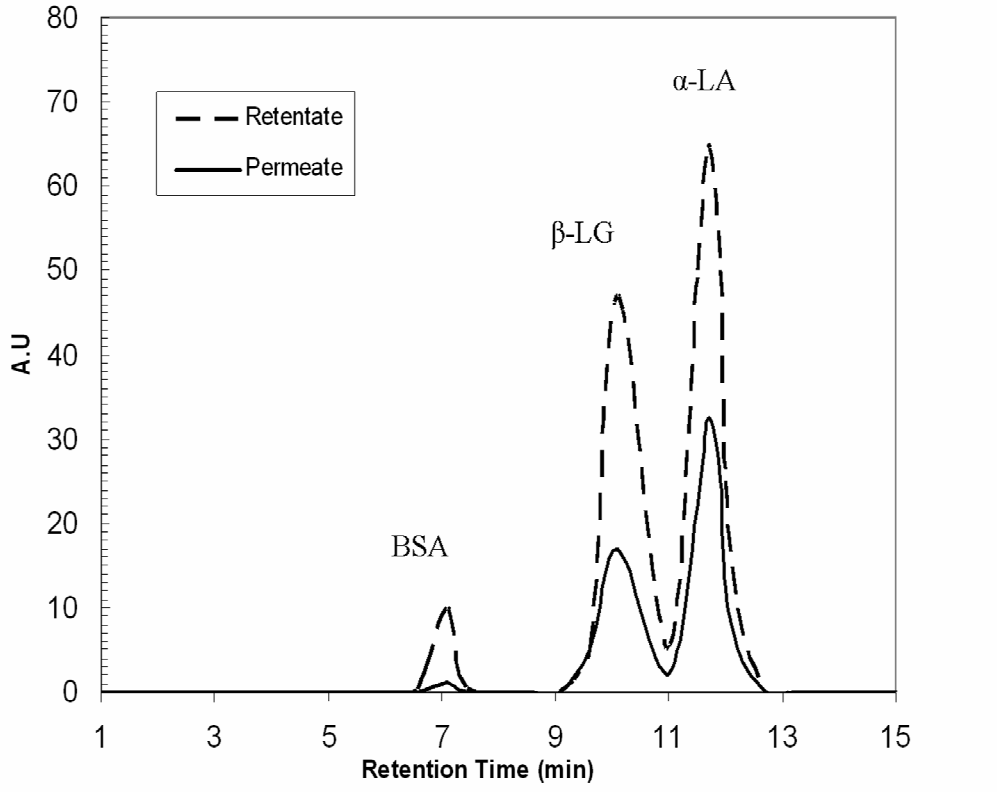
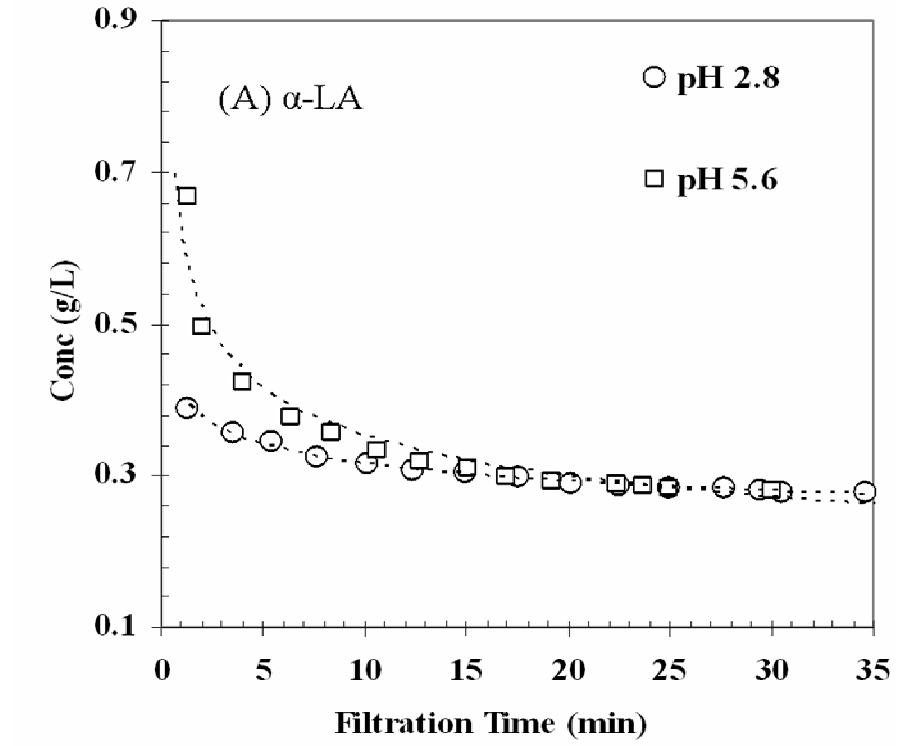


Figure 5-3. HPLC chromatograms for final retentate and permeate samples obtained at pH 5.6. In both chromatograms peaks at retention times of 7.1, 10.2, and 11.7 min correspond to BSA, β-LG and α-LA, respectively.

The concentration of β-LG and α-LA in the permeate were determined by accounting for the total protein collected and the total volume of permeate. Protein concentration profiles for the permeate stream were presented by plotting the concentration of β-LG and α-LA in the permeate streams as a function of filtration time for two ultrafiltration experiments performed at two different pHs: pH 5.6 and pH 2.8 (Fig 5-4). Figure 5-4 shows that there is a decrease in the concentration of transmitted protein, both the β-LG and α-LA, over the course of the filtration and that the rate of change is affected by the pH of the feed solution.



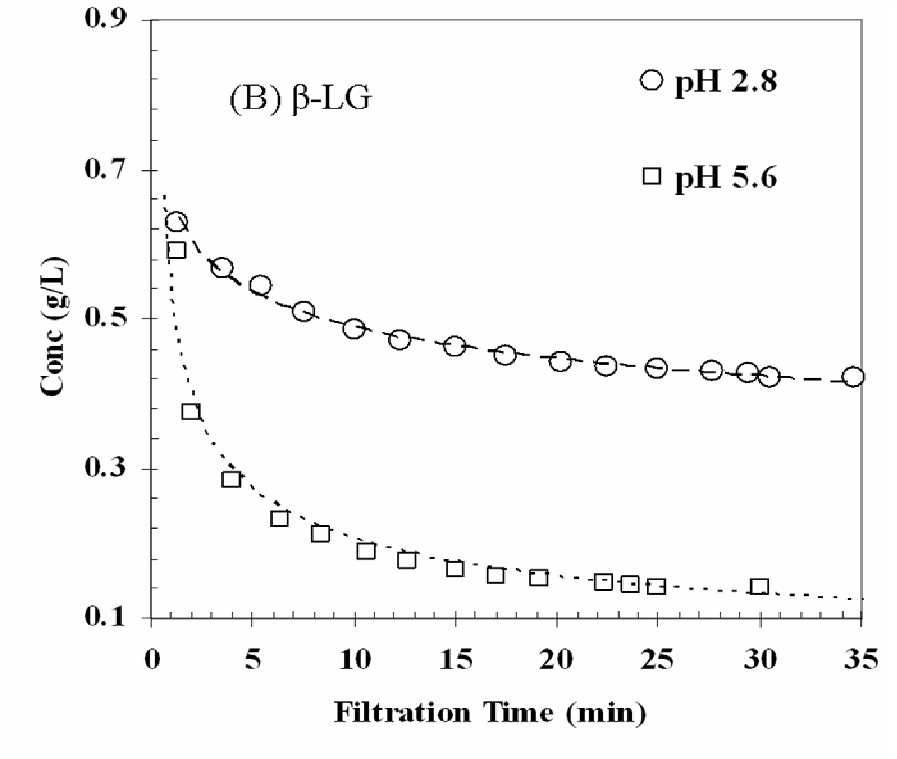


Figure 5-4. Protein concentration profiles for α-LA (A) and β-LG (B) in the permeate stream at two different pHs.

The rapid initial decrease in β-LG protein concentration seen at pH 5.6 may be attributed to the higher tendency of the protein to aggregate at pH values close to the isoelectric point thus limiting the transmission of β-LG as well as resulting in a decline in flux during the initial stage of the UF process. The amount of protein that permeated through the membrane continued to decrease with time until it reached steady state levels after approximately 10 min of operation. While the protein concentration in the permeate decreased gradually with time, the protein concentration in the retentate stream remained essentially constant during the first 15 min (Fig. 5-5).

As the filtration proceeded, the retentate gradually became more concentrated due to continuous removal of permeate. It was also observed that changing the pH of the feed solution from 2.8 to 5.6 affected the protein composition in the permeate. Ultrafiltration at pH 5.6 yielded a permeate with a β-LG content that ranged between 50%-55%, whereas ultrafiltration at pH 2.8 yielded a permeate with a β-LG content of 68-72%.

Bhattacharjee *et al.* (2006) studied protein fractionation from casein whey using 30 a kDa polyethersulfone membrane in a stir cell module comparable to the one employed in this study. Despite differences in the experimental conditions (membrane material, transmembrane pressure and the whey protein source) a similar effect of pH on β-LG purity in the permeate was observed. Figure 5-5 and 5-6 present comparisons between the FOP predictions against HPLC-based determinations for different pH values. Excellent agreement was obtained over the course of the filtration.

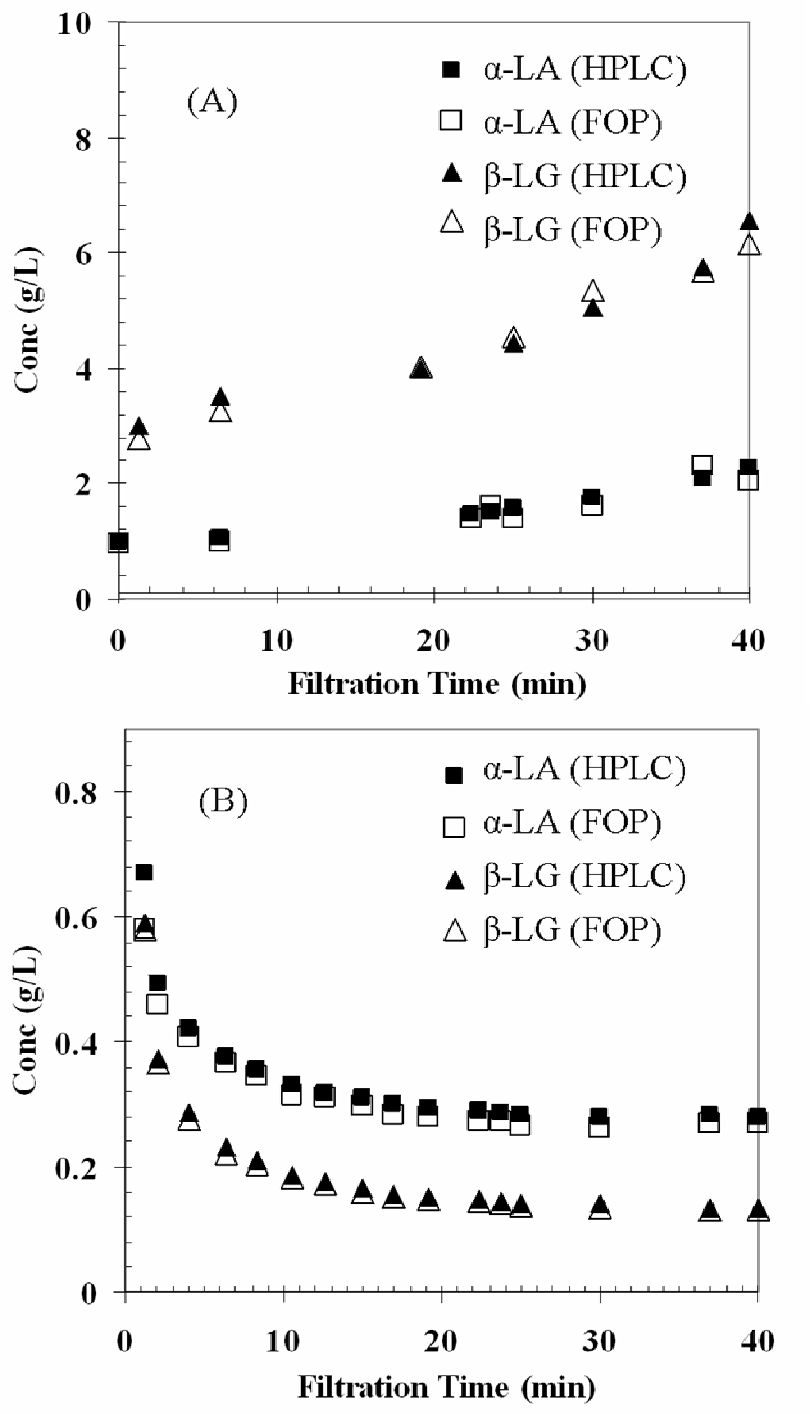


Figure 5-5. Concentrations of α-LA and β-LG in the retentate stream (A) and permeate stream (B) as estimated from HPLC and fiber optic probe (FOP) data. Ultrafiltration at pH 5.6.

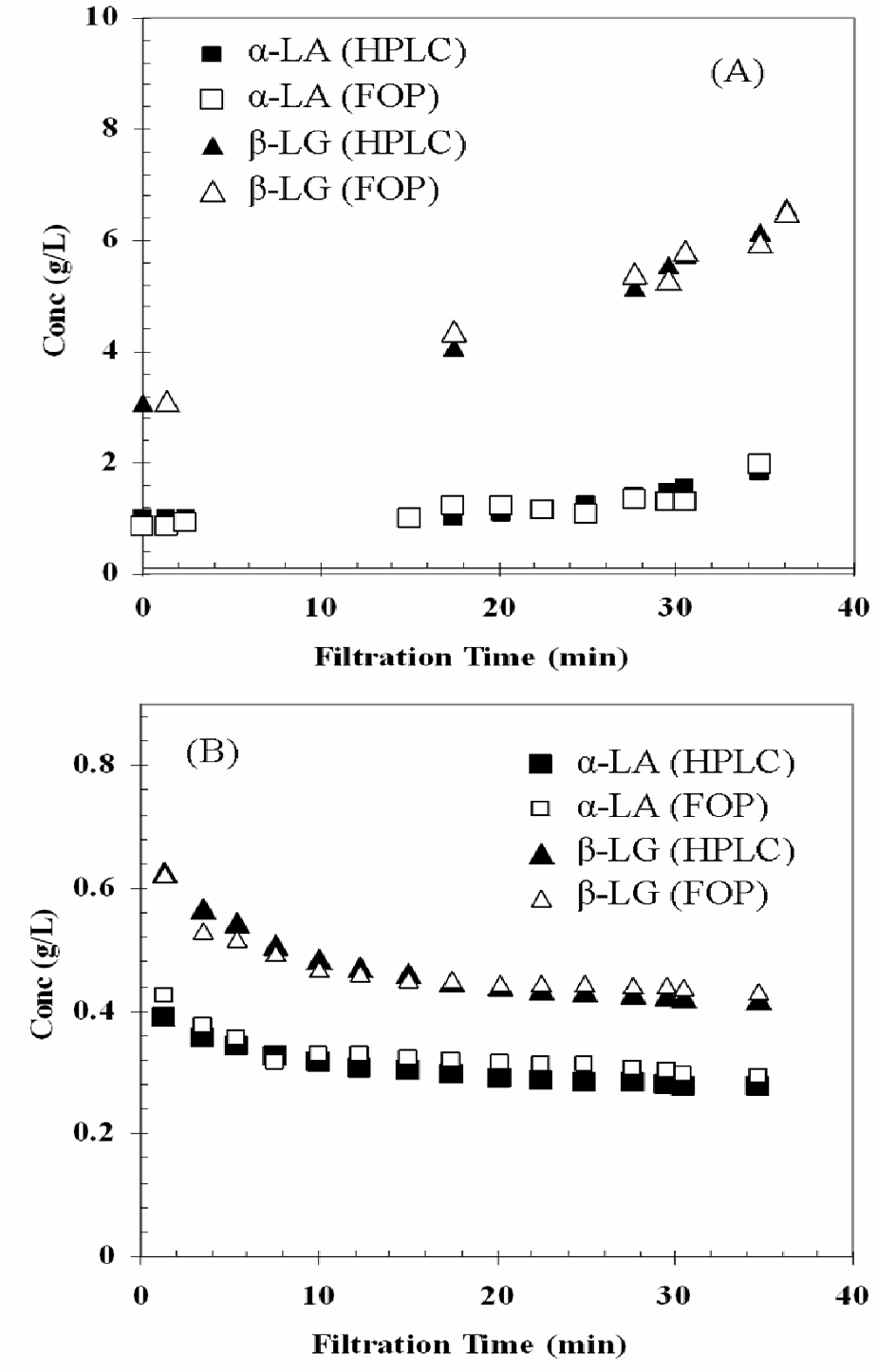


Figure 5-6. Concentrations of α-LA and β-LG in the retentate stream (A) and permeate stream (B) as estimated from HPLC and the fiber optic probe (FOP) data. Ultrafiltration at pH 2.8.

5.4.4. Incorporation of FOP-based Predictions into Membrane Process Monitoring

Monitoring the decline in membrane flux can serve as one method for on-line tracking of ultrafiltration process performance; however, the decline in flux does not give an indication of the process separation performance. Here we evaluate an approach that combines the FOP-based predictions with membrane separation theory to predict the operational membrane performance under different operating conditions. Protein transmission coefficients and effective membrane selectivity were selected as indicators of filtration process performance.

5.4.4.1. Transmission Coefficient and Membrane Selectivity

The transmission coefficient, which is equal to the ratio of the concentration of a component in the permeate (C_{pi}) to the retentate (C_{ri}), is an important quantity for monitoring membrane performance (Ghosh, 2003). It is described by the following:

$$\tau_{ob} = \frac{C_{pi}}{C_{ri}} \qquad \textbf{(5-3)}$$

where C_{pi} and C_{ri} are the permeate and the bulk concentrations, respectively of a given protein i.

Figure 5-7 presents protein transmission values for β-LG and α-LA that were determined from FOP-based estimates. Three ultrafiltration experiments were performed for WPI solutions at different pH values and transmembrane pressures. The transmission of α-LA and β-LG was observed to decrease significantly with increasing

transmembrane pressure since increasing the transmembrane pressure likely induces more fouling and more protein deposition on the membranes. The effect of pH and transmembrane pressure on the time profile of β-LG transmission appears to follow the fouling resistance (Fig. 5-8) calculated using Equation (5-1)

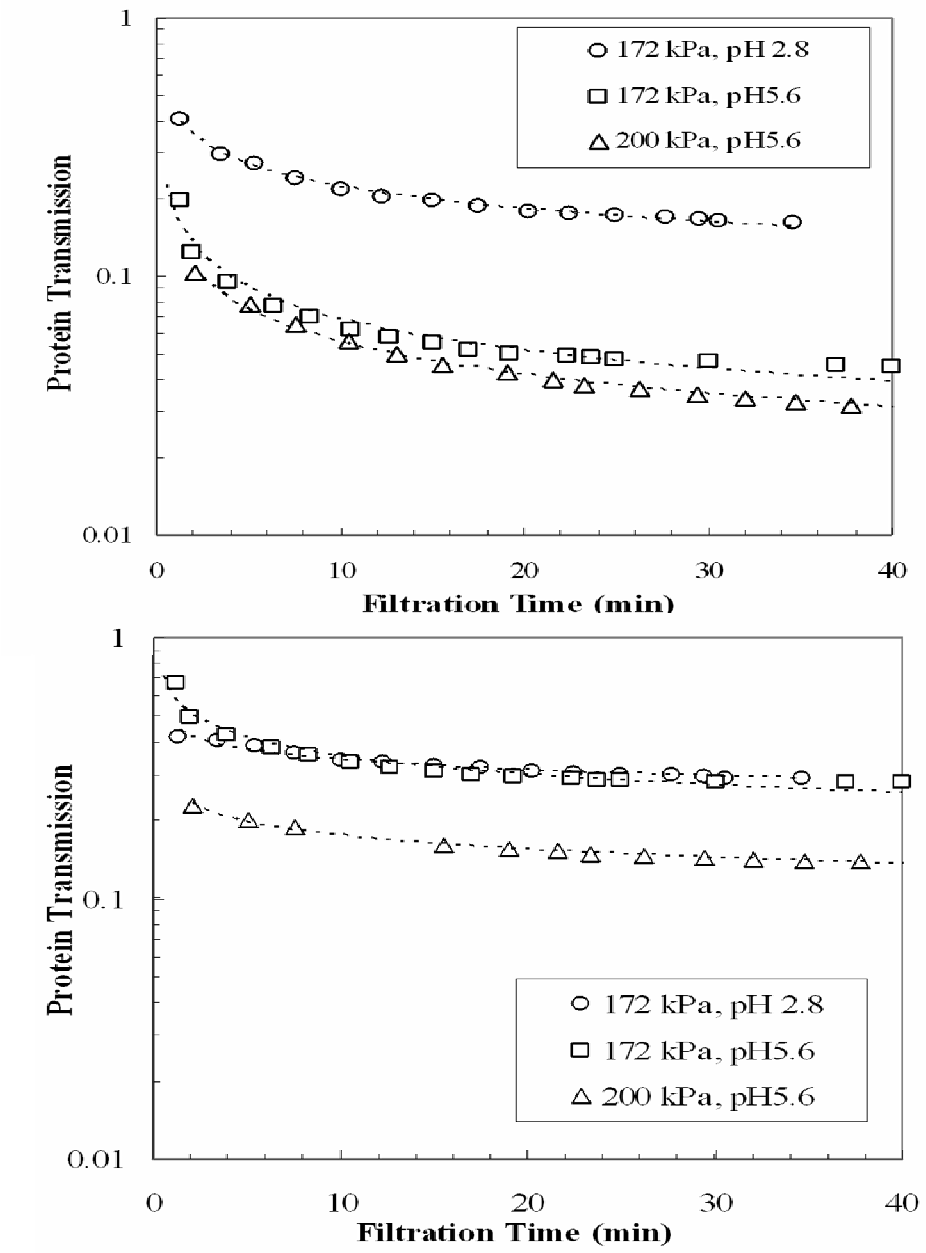


Figure 5-7. Time profiles of β-LG transmission (top) and α-LA transmission (bottom) determined from FOP-based estimates at two different transmembrane pressures (172 and 200 kPa) and two different pHs (2.8 and 5.6).

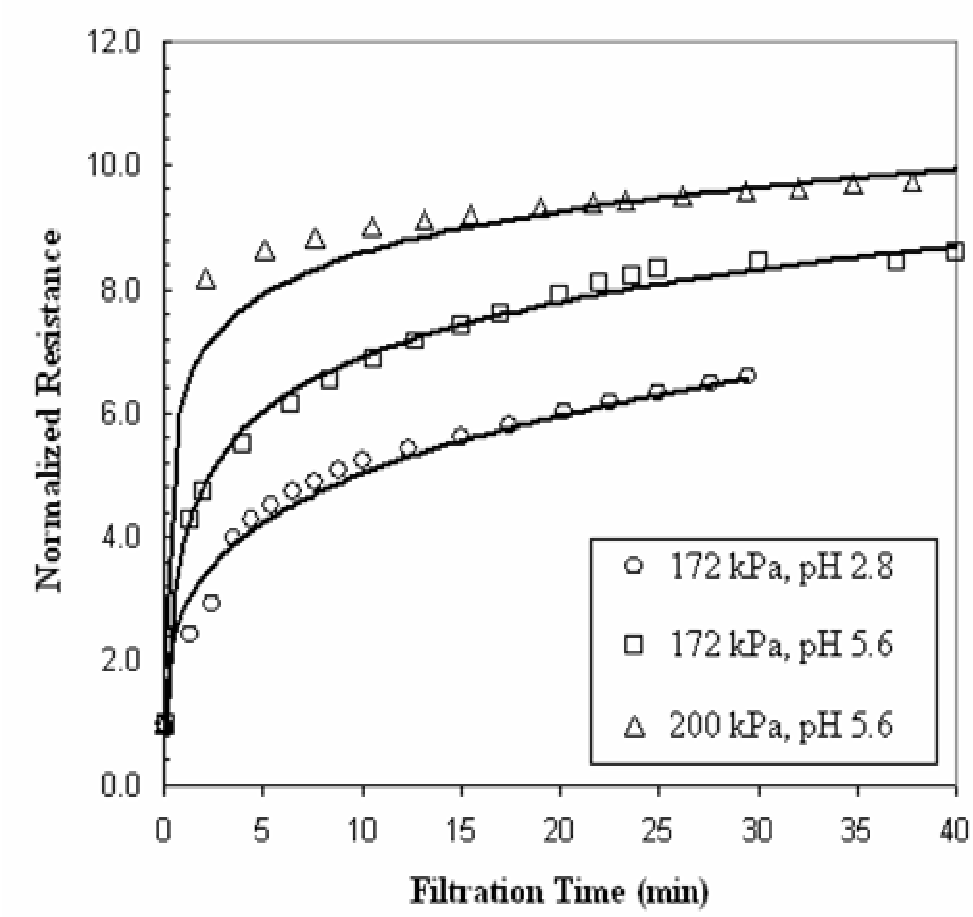


Figure 5-8. Time profile of membrane fouling resistance of the 30-kDa RC membrane evaluated at two different transmembrane pressures (172 and 200 kPa) and two different pHs (2.8 and 5.6).

These results also show that apart from the preferential permeation of α-LA, β-LG and BSA were also transmitted to different extents in the permeate. A binary selectivity value based solely on transmission of α-LA and β-LG would not be sufficiently informative for monitoring the efficiency of protein fractionation during process monitoring. Therefore, an additional parameter referred to as effective selectivity, has

been used to characterize the efficiency of protein fractionation involving complex protein mixtures and is defined as follows (Ghosh, 2003):

$$\psi_i = \frac{\tau_i}{\sum \tau_{j \neq i}} \qquad \textbf{(5-4)}$$

The numerator is the apparent transmission coefficient of a given protein *i* and the denominator consists of the summation of apparent transmission coefficients of all other proteins. For example, according to equation (5-4), the selectivity of α-LA in comparison to β-LG and BSA is:

$$\psi_{\alpha-LA} = \frac{\tau_{\alpha-LA}}{\tau_{BSA} + \tau_{\beta-LG}} \qquad \textbf{(5-5)}$$

The FOP-based estimates were combined with Equation 5-4 to estimate the time profile of the effective selectivity coefficient. Figure 5-9 shows that the selectivity coefficient for α-LA with respect to β-LG and BSA varies with time. Solution pH and transmembrane pressure are shown to have an effect on the selectivity coefficient as shown in Figure 5-9. A higher selectivity was obtained with a feed pH of 5.6 compared to that at pH of 2.8 at a fixed transmembrane pressure of 172 kPa (Fig. 5-9), which can be explained in terms of the effect of pH on the monomer–dimer equilibrium of β-LG (Bhattacharjee *et al.*,2006). The lower transmission of β-LG observed at pH 5.6 compared to that at pH 2.8 is most likely due to the tendency of β-LG to form dimers at that pH (Bhattacharjee *et al.*,2006). Results here demonstrate that quantifying time-dependent protein transmission by a combination of EEM

fluorescence and PLS data analysis may be of great value for monitoring and potentially optimizing protein fractionation by membrane ultrafiltration.

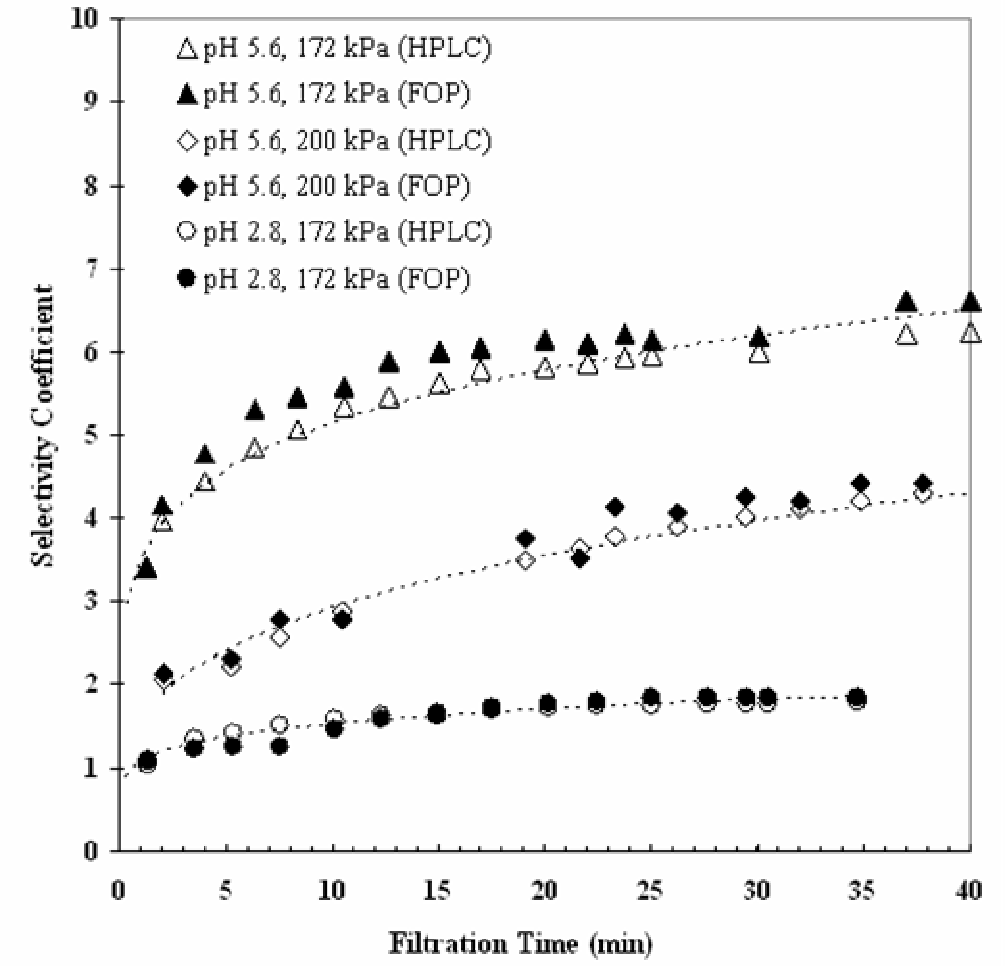


Figure 5-9. Effect of transmembrane pressure on the effective selectivity coefficient profile for α-LA as estimated from HPLC and fiber optic probe (FOP) data.

5.5. Conclusion

Performance of protein fractionation using ultrafiltration is strongly influenced by the physiochemical characteristics of the protein solution and the operating conditions. Precise tuning of these processes is necessary to achieve desired levels of protein fractionation (Ghosh and Cui, 1998; 2000a; b). The intrinsic fluorescence of the proteins has been effectively used to construct a predictive model for estimating protein concentrations during a filtration process. This novel methodology makes use of a partial least squares-based regression model (PLS) for estimating protein concentrations based on synchronous fluorescence spectra acquired using a FOP. The advantage of this approach is that with development it may allow in industrial applications inference of relative accurate estimates of protein concentration within a shorter time period than would be possible with various analytical techniques, for example chromatography-based approaches. It is also clear that synchronous spectra can provide good predictions and consequently the use of the full spectrum may not be necessary for monitoring with corresponding savings in acquisition time. Separation performance variables that are difficult to measure, such as individual protein transmission, membrane selectivity and product yield can be estimated directly from fluorescence-based predictions of protein concentrations in the retentate and permeate streams. The proposed approach showed good predictions for different pH values and transmembrane pressure values. The only practical obstacle that may hamper the current approach from becoming a workable soft-sensor for *in situ* and real-time monitoring of UF processes lies in the need to dilute the retentate samples during the UF. This is because FOP estimates were not as accurate as the absolute protein

concentration increased. However, this drawback can be overcome if an automatic dilution strategy is implemented inline or alternate approach using a path length correction in a microplate system was implemented. The major highlight of the current work is that it demonstrates the potential applicability of fluorescence-based sensors for on-line monitoring with possible use in industrial UF processes.

CHAPTER 6

Use of Multiwavelength Rayleigh Scattering Data for the Characterization of Protein Aggregation and Membrane Fouling Phenomena

One of the primary problems in membrane-based protein separation is membrane fouling. A variety of phenomena contribute to fouling and aside from changes in flux and transmembrane pressure, there are few tools to add to the arsenal for controlling membrane fouling and its consequences. In this study we explored the feasibility of employing Rayleigh light scattering data from fluorescence studies combined with chemometric techniques to determine if a correlation could be established with membrane fouling phenomena. Membrane flux was measured in a laboratory scale dead-end UF filtration system and the effect of protein solution properties on the overall membrane resistance was systematically investigated by regulating ionic strength and pH. A variety of milk proteins were used as a test case in this study. In parallel, the colloidal (aggregation) behavior of the protein solutions was assessed by employing multi-wavelength Rayleigh scattering measurements. Aggregation behavior was inferred based on published data for these same proteins and under identical solution conditions where techniques other than Rayleigh scattering had been used. Using this approach good agreement was observed between scattering data and aggregation behavior. To test the hypothesis that a high degree of aggregation will lead to increased membrane fouling, filtration data was used to find whether the

Rayleigh scattering intensity correlated with permeate flux and membrane resistance changes. It was found that for protein solutions which were stable and did not aggregate, fouling was reduced and these solutions exhibited reduced Rayleigh scattering. When the aggregation behavior of the solution was enhanced, significant flux declines occurred and were correlated with increased Rayleigh scattering. It is proposed that this methodology may be suitable for tracking physico-chemical changes in protein solutions and that this may have applications in areas such as membrane-based protein separation.

6.1. Introduction

The fouling behavior of proteins in membrane-based separation processes is strongly influenced by their stability in solution (Malmsten, 1998). Protein stability is a consequence of a delicate balance of forces, including electrostatic, hydrogen bonding, Van der Waals forces and hydrophobic interactions. Changes in the environment surrounding the protein including pH, ionic strength, temperature and shear will affect their stability (Malmsten, 1998). The consequence of these changes can be aggregation/association of protein which has an important influence on a variety of phenomena including the performance of membrane-based separation processes such as microfiltration and ultrafiltration. The formation of protein aggregates can easily plug membrane pores or/and serve as nucleation sites for the propagation of aggregation on existing deposits. These aggregates can result from non-covalent bonding between native proteins such as Van der Waals or hydrophobic interactions, or covalent bonding (Malmsten, 1998).

Microfiltration (MF) is widely used for the separation of proteins and MF membranes are specifically designed for this purpose (Kuberkar and Davis, 1999). In MF it is important that the protein be in solution so that it is free to permeate the membrane to minimized product losses (Bowen *et al.*, 1998). Efficient recovery of proteins requires that the protein be small relative to the membrane pore size so that the protein can pass through the membrane and that the processing conditions are such that protein aggregation is minimized, as protein aggregates may be retained by the membrane. Thus, the size of the protein in solution as well as its tendency to aggregate near the

membrane surface play a critical role in determining the degree of product recovery and the extent of membrane fouling (Bowen *et al.* 1998). In an ultrafiltration study of soy protein extracts Mondor *et al.* (2004) found that the average molecular size and the molecular size distribution vary with pH and salt content and that this could be attributed to aggregation–disaggregation behavior of different protein fractions. The effect of pH and salt content can affect protein transport into the concentration region. This information is fundamental to understanding what contributes to membrane fouling resistance when it can be primarily attributed to the concentrated and cake layer together (Mondor *et al.*, 2004).

There is a need for a rapid measurement technique to characterize protein-protein interactions and to identify solution conditions that will lead to membrane fouling. Light scattering represents a powerful technique for characterizing protein-protein interactions and the presence and formation of aggregates. Aggregation phenomena in protein solutions have so far been characterized primarily by methods based on radiation scattering or transport properties. Photon correlation spectroscopy, also known as dynamic light scattering, is a good method for determining the degree of protein aggregation in solutions of different ionic strengths and pHs. One of its advantages over size exclusion chromatography is that it allows measurements to be made directly from the solutions used for filtration; however, photon correlation spectroscopy has several severe practical limitations. First, it is not possible to make direct measurements of the concentration of discrete aggregate species present in a polydisperse solution. Also, photon correlation spectroscopy experiments are costly

both in terms of equipment and time that is required to clean and dilute protein solutions to obtain accurate results. The other challenge for ultrafiltration is that highly concentrated protein solutions are often involved which by their nature are very unstable and subject to particulate contamination. This is an impediment for rapid monitoring of protein aggregation and assessing its fouling potential.

There is a need for an independent non-invasive and rapid method for assessment of aggregation behavior in a way that is useful for process operation and prediction. Such a method would also provide some qualitative information about the fouling potential of protein solutions. A novel element of the present work is that the proposed measurement is based on the entire scattering spectrum to be used to retrieve information about protein aggregation behavior. The approach proposed is to use Rayleigh scattering collected spectrophotmetrically to characterize protein-protein interactions. By coupling and scanning simultaneously the excitation and the emission monochromators of a common spectrofluorometer, light scattering signals were detected at an angle of 180^0 in the wavelength range (200-700 nm) by using a fiber optic probe. Santos *et al.* (1999) reported that spectrofluorometry offers some potential advantages that are not available with classical laser light scattering spectroscopy, namely, the availability of broader wavelength ranges. Changing the wavelength changes the magnitude of the scattering vector which has the same effect as a change in the measurement angle in classical laser light scattering spectroscopy. Thus, it is possible to overcome angle dependency by accounting for wavelength dependency.

This Chapter is organized into three main parts:

i. An analysis of the general features of Rayleigh scattering spectra collected at an angle of 180^0 for different protein solutions (β-lactoglobulin, α-lactalbumin, Bovine Serum Albumin and Lysozyme) is presented. In this analysis, the effects of the average molecular weight, protein concentration, pH and ionic strength were considered.

ii. The potential for use of multiwavelength light scattering to determine the propensity of protein solutions to foul membranes is presented.

iii. The value of multiwavelength light scattering to provide quantitative measurements of the average molecular weights of proteins in solution is presented.

6.2. Materials and Methods

6.2.1. Protein Solution Preparation

β-Lactoglobulin (β-LG), α-lactalbumin (α-LA) and Bipro were used in their powdered form and were donated by Davisco Foods International (LeSueur, USA). The β-LG and α-LG were of 95% purity. Bovine serum albumin (BSA - 95% purity) and Lysozyme were purchased from Sigma Aldrich International. Bipro is a whey protein isolate (WPI) consisting mainly of β-lactoglobulin and α-lactalbumin with a content of 74.9% β-LG, 14.9% α-LA, 3.2% immunoglobulin and 1.5% BSA, (Weinbreck *et al.*, 2004). All other chemicals used were of analytical grade and obtained from Sigma Chemical Corp. (St. Louis, MO). A Millipore (Canada) water purification unit was to produce ultrapure water with a resistivity greater than 17.6 MΩ. Buffers were prepared

using ultrapure water and then micro filtered through a 0.45-μm Nylon membrane (PALL Corporation, Michigan, USA). Stock solutions (10 g/L) of the proteins were prepared by dissolving their powdered forms in 20 mM sodium phosphate buffer at the desired pH and at the desired salt concentration (e.g. NaCl). Different protein solutions were then prepared by diluting aliquots of stock solutions to the desired concentrations. Once prepared, protein solutions were allowed to stand for 5 minutes at room temperature before acquiring light scattering measurements. In order to achieve consistency and minimize the experimental error, the time interval between sample preparation and light scattering measurement was always fixed to 5 minutes. Twenty five β-LG protein solutions were prepared at different conditions (pHs, salt concentrations) identical to those used by Verheul *et al.* (1999) as shown in Table 6.1. The average molecular weights corresponding to these samples (Table 6.1) were obtained from the same source (Verheul *et al.*, 1999). Light scattering spectra were acquired for these samples as explained in the next subsection.

Table 6.1: Twenty five β-LG protein solutions prepared at different conditions (pHs, salt concentrations) identical to those used by Verheul *et al.* (1999)

Sample ID	pH	NaCl(M)	C(g/l)	MW(Da)
1	2	0	2	16.84
2	2	0	5	16.66
3	2	0	10	17.02
4	2	0.1	2	28.00
5	2	0.1	5	28.37
6	2	0.1	10	28.55
7	6.9	0	2	29.47
8	6.9	0	5	32.58
9	6.9	0	10	34.04
10	6.9	0.1	2	31.12
11	6.9	0.1	5	33.49
12	6.9	0.1	10	32.95
13	8	0	2	23.43
14	8	0	5	24.53
15	8	0	10	25.44
16	8	0.1	2	26.91
17	8	0.1	5	28.55
18	8	0.1	10	31.48
19	5.4	0	10	41.18
20	4.7	0	1	32.76
21	4.7	0	5	45.94
22	4.7	0	10	61.86
23	4.7	0.1	2	36.61
24	4.7	0.1	5	37.89
25	4.7	0.1	10	43.93

6.2.2. Multiwavelength Light Scattering Measurements

Multiwavelength light scattering measurements on protein solutions were performed using a fiber optic probe (FOP) connected to a steady state fluorescence system with a pulsed xenon flash lamp as the light source (Varian Cary Eclipse, Mississauga, ON, Canada). Scattered light intensity (Rayleigh) was recorded at wavelengths ranging

from 200 to 700 nm by setting the excitation wavelength equal to the emission wavelength. Measurements were taken with a scan speed of 120 nm/min, a slit width of 5 nm for both excitation and emission and 2 nm smoothing. The scattering spectral profile of the buffer solution alone, i.e. in the absence of protein, was measured and subtracted from the collected spectra. All samples were initially examined in 4-7 min intervals to ensure that the light scattering profile did not change with respect to time. The effect of pH, salt concentration and protein concentration on the light scattering profile was investigated. The ability of these multiwavelength measurements to provide insight into the protein aggregation behavior and the propensity of protein solutions to foul membranes is presented.

6.2.3. Ultra filtration Experiments

All filtration experiments were carried out in batch mode using a 75-mm diameter Amicon stirred ultrafiltration cell (Model 8010, Amicon Corp., Beverly, MA) in dead-end flow configuration, starting each time with an initial feed volume of 350 mL. Filtration experiments were conducted using 30 kDa molecular weight cut-off composite regenerated cellulose membranes, obtained from Millipore Corp. (Bedford, MA). The membranes were flushed with deionized distilled water prior to use to remove any possible residual storage agents. The filtration cell was then filled with protein solution of known concentration, pH, and salt concentration. The contents of the filtration cell were kept well-stirred using a magnetic stir bar. Permeate was collected in 10-ml tubes and the time was recorded for every 10 ml. of additional volume of permeate collected.

6.2.4 Evaluation of the Overall Membrane Resistance

A resistance-in-series model was used to estimate membrane fouling for ultrafiltration experiments. The total membrane resistance R_{tot} (m^{-1}) was estimated using the following equation:

$$R_{tot} = R_m + R_f = \frac{\Delta P}{\mu_v J_v} \quad \textbf{(6-1)}$$

where R_m (m^{-1}) is the specific membrane resistance of the clean membrane, R_f (m^{-1}) is the added resistance due to fouling, J_v ($m^3 \cdot m^{-2} \cdot s^{-1}$) is the flux of the protein solution with time, μ (Pa·sec) is the permeate viscosity and ΔP (Pa) is the transmembrane pressure. The permeate viscosity was assumed to be equal to that of pure water at room temperature ($1x10^{-3}$ Pa·s). The flux of the protein solution J_v in the time interval Δt is given by Equation 6-2.

$$J_v = \left(\frac{1}{A_m}\right)\frac{\Delta V}{\Delta t} \quad \textbf{(6-2)}$$

Where A_m is the effective membrane area (41.3 cm^2) and ΔV is the volume of the filtrate solution that was collected during the time interval Δt. Total resistance values were normalized by the specific membrane resistance R_m (m^{-1}) of the clean membrane and then the normalized resistance for different experiments was plotted versus filtration time. R_m (m^{-1}) was evaluated from clean water flux measurements at different transmembrane pressures. The normalized overall resistance was analyzed to determine the effect of protein solutions characteristics (i.e. salt concentration, pH, protein concentration) on membrane fouling.

6.3. Chemometric Modeling

6.3.1. Partial Least Squares Regression

For the purpose of inferring quantitative information about protein aggregation behavior from light scattering spectra, the data used in this study were divided into two categories: input data corresponding to the light scattering intensities detected in the wavelength range (200 to 700 nm), and output data consisting of the average molecular weight of β-LG protein solutions. The matrix that contains the outputs to be predicted, i.e. the average molecular weight, were arranged in an output matrix to be referred to heretofore as matrix Y whereas the light scattering measurements were arranged into an input data matrix to be referred to as matrix X. The rows in matrix X correspond to different samples while the columns correspond to scattering intensities detected at different excitation wavelengths for each one of the samples. The input data is considered to be multivariate in nature since it consists of 800 scattering intensity data points for each sample. Partial least squares regression (PLS) is a well-known chemo-metric tool for designing a calibration model which can be used to correlate the set of known measurements represented by the scattering data to the property to be predicted, i.e. the average molecular weight. Regression models with the ability to predict certain properties that are difficult-to-measure such as the average molecular weight from easy-to-measure data such as light scattering are referred to as *soft sensors* (James *et al.*, 2002). The PLS regression was chosen here since it is able to handle problems associated with high sensitivity noise for multivariate data sets with large numbers of highly correlated variables (Geladi and Kowalski, 1986; Qin and McAvoy, 1992). Collinearity is very high among the different light scattering readings obtained at

different excitation wavelengths. Hence, the original input variables *x* in matrix X are replaced by a smaller set of underlying new variables that are uncorrelated, mutually independent (orthogonal) and are mathematically represented by linear combinations of the original descriptors. These calculated linear combinations, referred to as latent variables (LVs) or principal components, are calculated to both provide good representation of matrix X and maximize the correlation between the input and the output (Qin and McAvoy, 1992). The optimum number of LVs and the goodness of prediction (Q^2) are determined using a cross-validation algorithm (Geladi and Kowalski, 1986; Qin and McAvoy, 1992; Elshereef *et al.*, 2006). The data obtained from different samples were used for calibrating and testing the PLS model as will be shown in the following sections. All computations were carried out using MATLAB 7.0 (MathWorks, Natick, MA) along with the PLS Toolbox 3.5 (Eigenvector Research Inc., Manson, WA).

To develop the PLS model, the samples were divided into two different sets, a calibration set consisting of the samples used for calibrating the PLS model and a validation set consisting of samples that were used to independently test the predictive ability of the calibrated model. The calibration and validation sets consisted of β-LG solutions prepared at different protein concentrations, pHs, and ionic strengths (see Table 6.1). The PLS model was tested using light scattering data acquired for β-LG protein solutions that were not included in the calibration set.

6.4. Results and Discussion

6.4.1. Qualitative Analysis of Multi-wavelength Rayleigh Light Scattering Spectra for Bulk Solutions

6.4.1.1. General Features of Multi-wavelength Scattering Spectra

Preliminary experiments for collecting Rayleigh scattering data were performed on very dilute protein solutions. Rayleigh scattering spectra in the excitation range 200-700 nm were compared for different dilute protein solutions of 1.0 μM (α-LA, lysozyme, and BSA) as seen in Figure 6.1. It is clear that light scattering intensity is wavelength–dependent with the maximum scattering intensity located around 310-330 nm. The scattering intensity acquired in the long-wavelength range, i.e. red-edge, decreases with increasing wavelength following the Rayleigh scattering regime where the intensity of the scattered light varies inversely with the fourth power of the wavelength. The Rayleigh scattering regime corresponds to particle radii that are much smaller than the incident wavelength. The inverse wavelength dependence makes short wavelength light scatter greater than long wavelength light, and the strong power dependence explains why the scattered energy increases rapidly as the wavelength decreases. From Figure 6.1, BSA (MW 67 kDa) solutions yielded the highest scattering spectra, followed by α-LA (MW 15 kDa) and lysozyme (14.3 kDa) which were comparable. This may indicate that the scattering signal response basically depends on the molecular weight of the various proteins. The analyses in this work were made on the entire multiwavelength scattering spectra since it was found to provide more information with respect to the protein solution properties. For example, although the scattering intensities at an excitation of 350 nm for lysozyme and α-LA are very close

in magnitude due to their similar molecular weight values, there were subtle differences between their spectra.

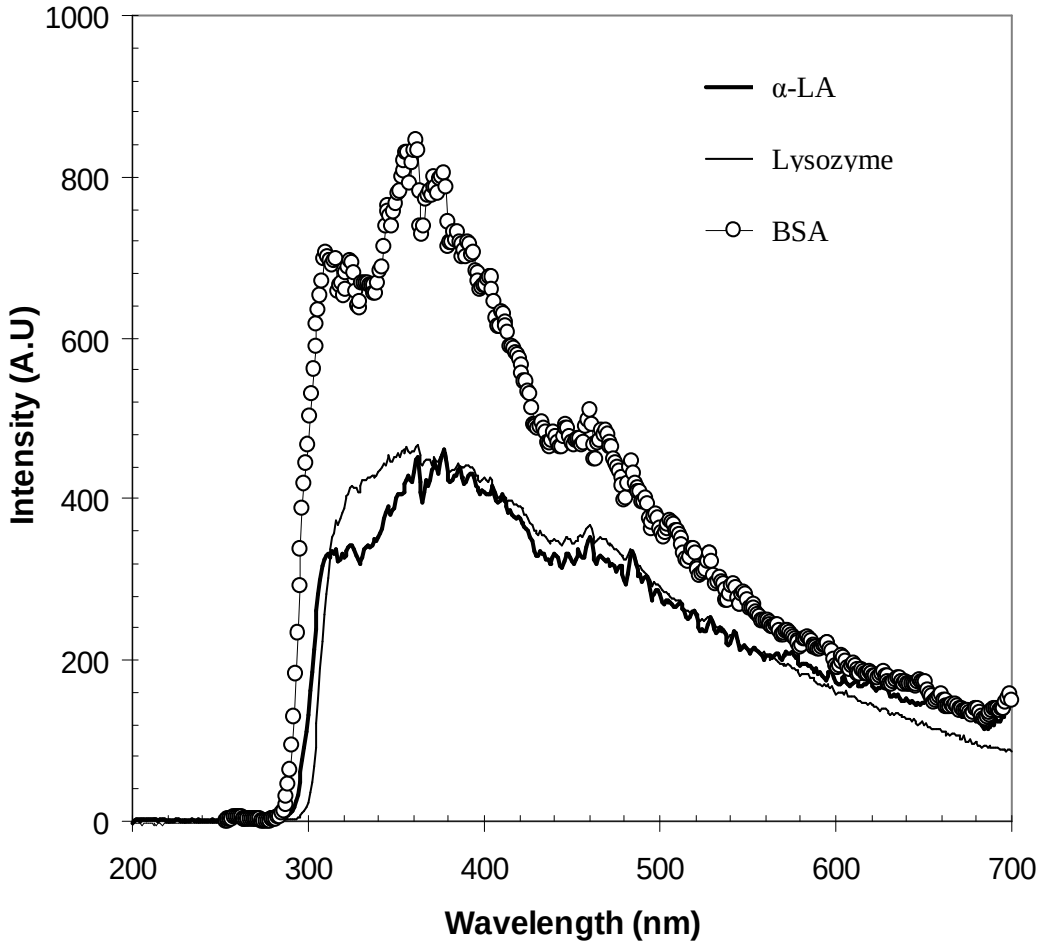


Figure 6.1: Multiwavelngth Scattering spectra for three different proteins acquired using FOP.

It will be noticed in Figure 6.1 that the spectra are not smooth and they are punctuated by distinct peaks at particular wavelengths. For example, the peak at about 360 nm is visible in all spectra but it is strongest in the spectra acquired for BSA. Such subtle differences can be deciphered further by applying the first derivative of the light

scattering spectra $dFl/d\lambda$ as shown in Figure 6.2. Figure 6.2 reveals wavelength-dependent fluctuations in the intensity of the scattered light which could provide a signature for each protein. These results imply that the utilization of the entire multi-wavelength scattering spectrum is more useful than light scattering measurements at a single wavelength since it may provide more complete information about the protein's properties in solution.

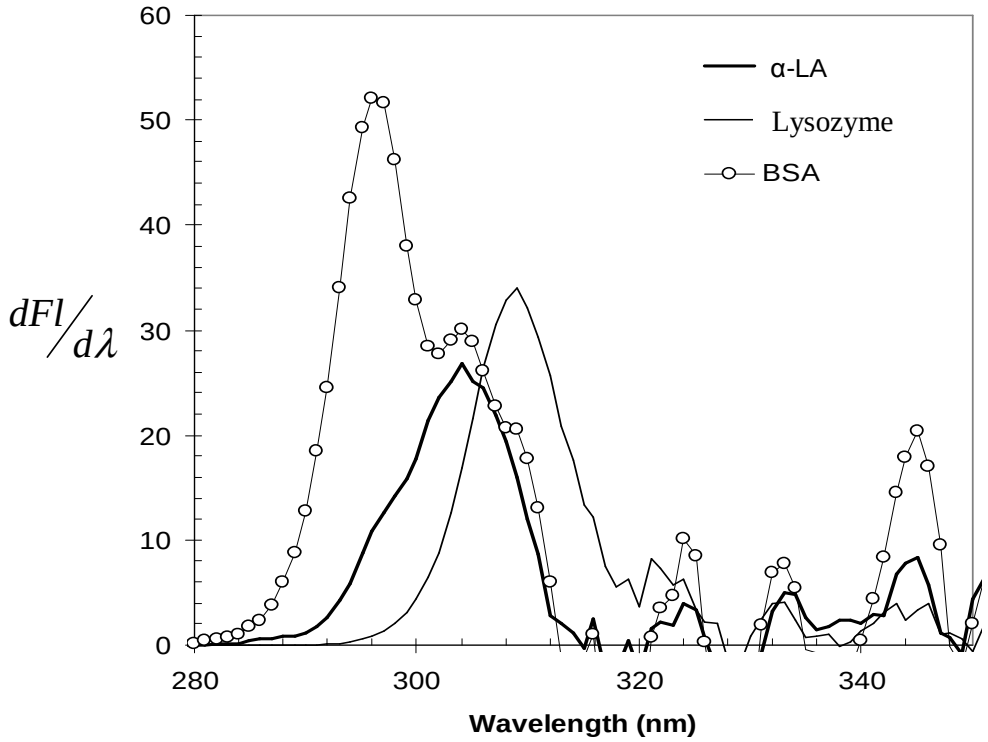


Figure 6-2: First derivative multiwavelngth scattering spectra $dFl/d\lambda$ in the wavelength range (280-350 nm) for three different proteins acquired using FOP.

6.4.1.2 Effect of pH and Ionic Strength

pH and ionic strength have a major effect on protein-protein interactions for proteins in solution. The maximal RLS scattering intensity was found to occur at an

excitation wavelength of 350 nm and so the intensity at this wavelength for different protein solutions (β-lactoglobulin, α-lactalbumin, lysozyme and hemoglobin) was plotted versus pH. For all the proteins studied, except for β-lactoglobulin, the maximum intensity as a function of pH was found to occur near the isoelectric point for each protein (Figure 6.3). This would seem reasonable as this is where attractive electrostatic interactions exist contributing to the lowest solubility and thereby to the highest turbidity and highest scattering.

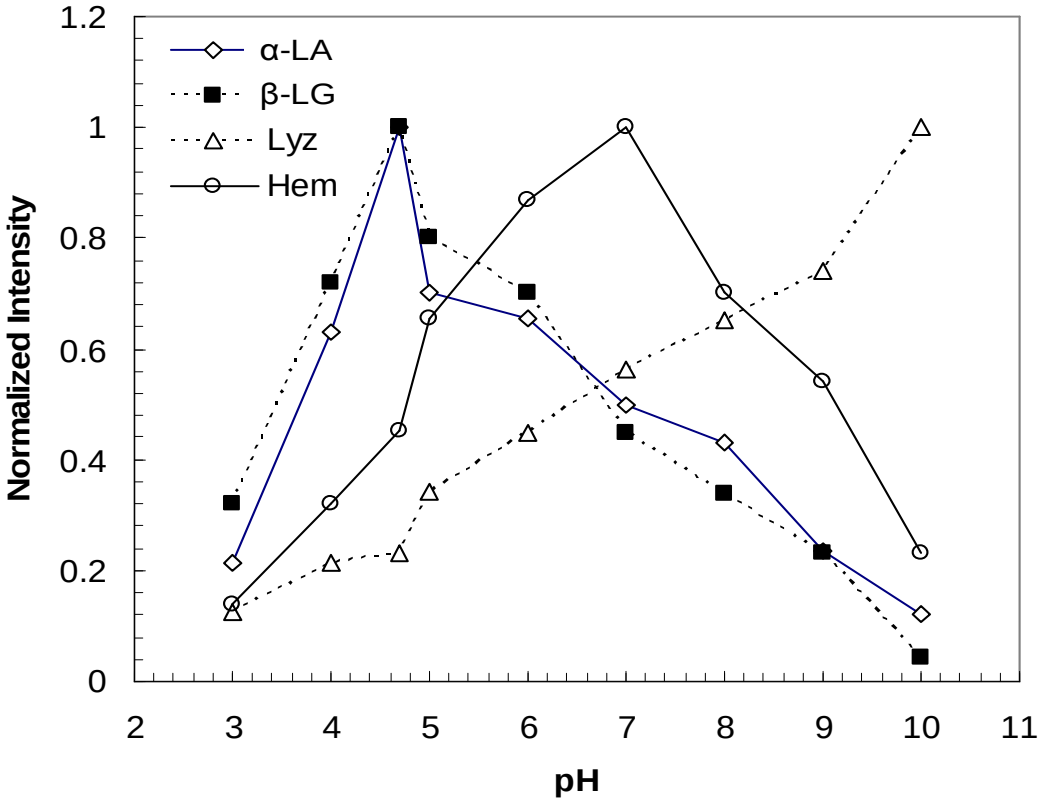


Figure 6-3: Effect of pH on the Maximum Backscattering Intensity at 350 nm for α-LA, β-LG, Lysozyme (Lyz) and Hemoglobin (Hem).

At pH values outside the range of the isoelectric point it was observed that the RLS scattering intensity decreased for all proteins except for β-lactoglobulin which can be interpreted as an indication that increased electrostatic repulsive interactions

resulted in enhanced protein solubility. In contrast, β-LG showed a more distinct pH dependent behavior (Fig. 6.4) where the highest RLS profile occurred at a pH of 4.6 which is below the isoelectric point of 5.2 for β-LG. This behavior is most likely due to the pH dependent non-covalent self-association of β-LG (Malmsten, 1998). Malmsten showed that the formation of higher form oligomers, such as octamers, occurred in the pH range of 3.7 to 5.2 with maximal association at pH 4.6.

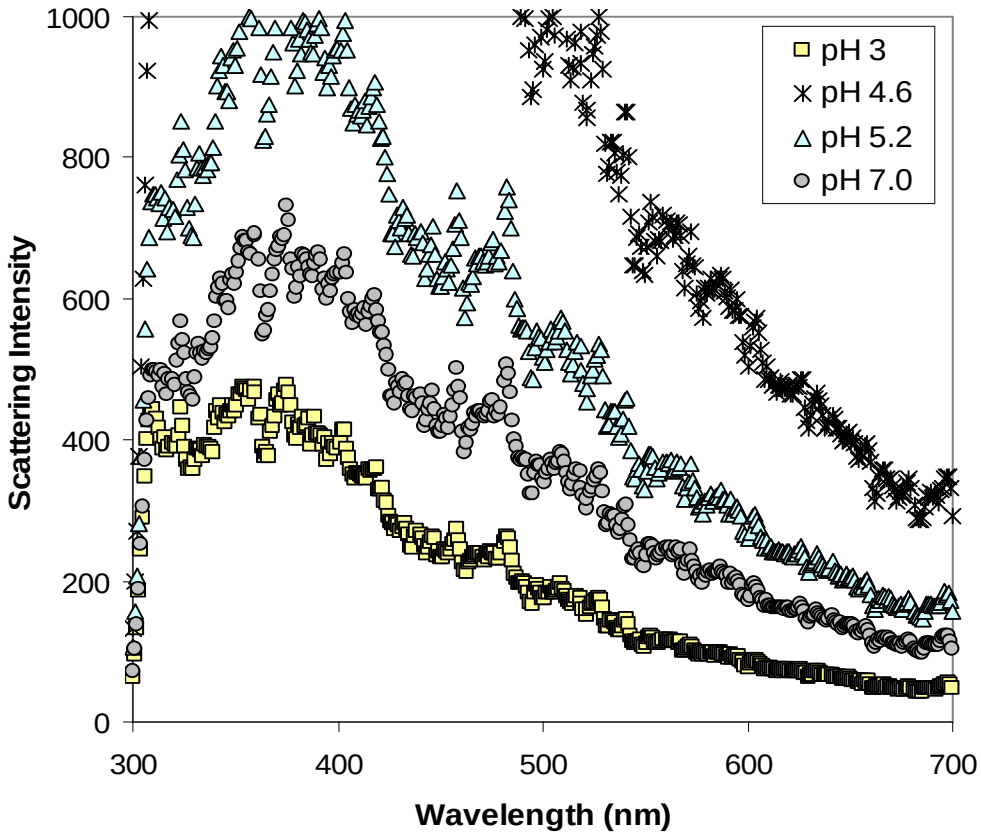


Figure 6.4: Effect of pH on FOP Backscattering Profile for 10 g/L β-LG solution.

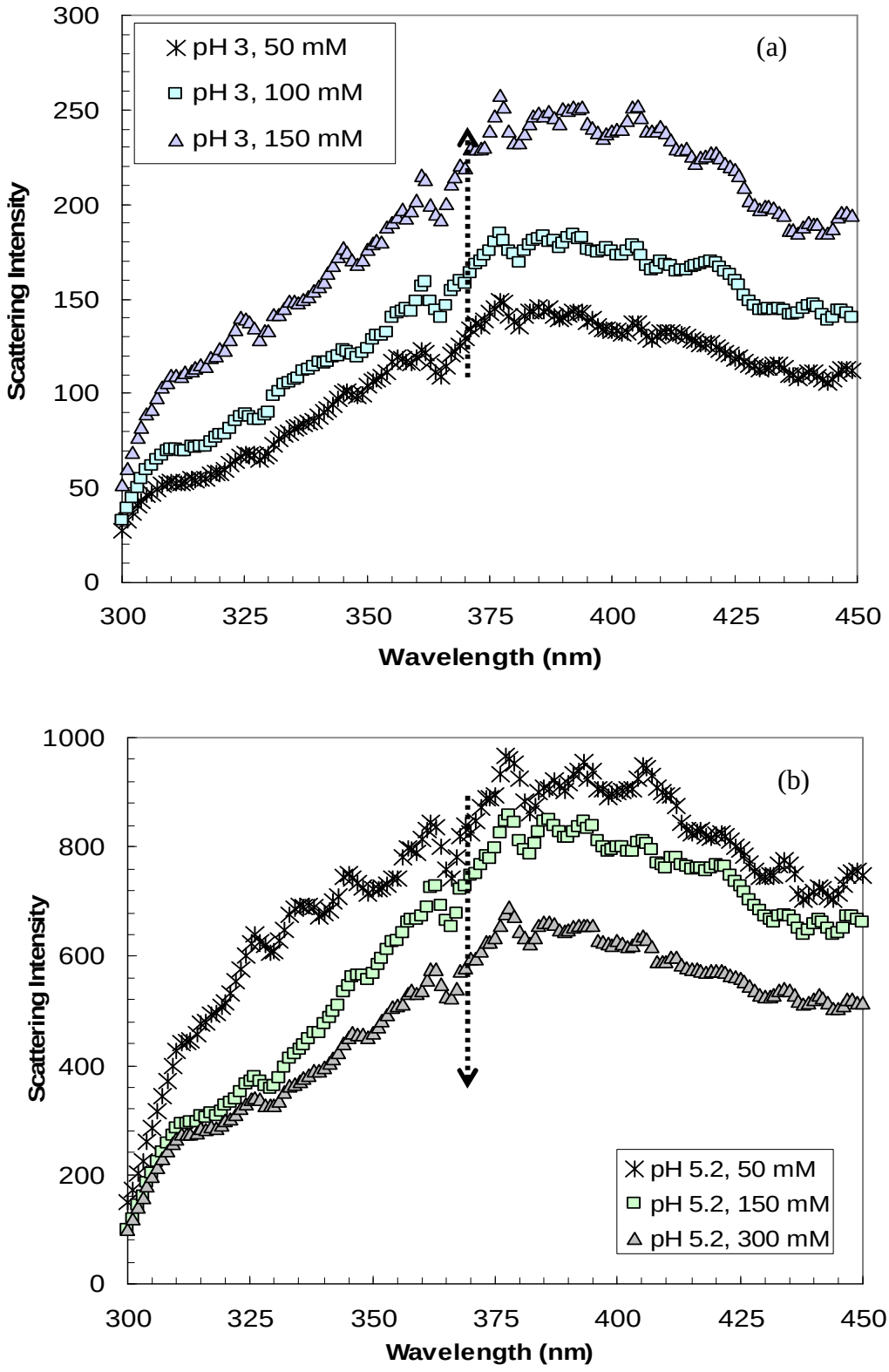


Figure 6.5: Effect of NaCl concentration in the FOP backscattering for 3 g/l β-LG solutions at pH 3 (a) and pH 5.2 (b)

The RLS profile for β-LG at pH 3 was found to increase with NaCl addition as seen in Figure 6.5a. A possible explanation of this behavior is that salt can have a shielding effect on the repulsive electrostatic interactions resulting in increased attractive interactions and dimerization. This observation agrees well with previous observations made using small-angle neutron scattering (Verheul *et al.*, 1999) and X-ray scattering (Baldini *et al.*, 1999) that indicated that β-LG at pH 3 exists as monomer but it forms dimers with increasing salt concentration. Salt addition does not always enhance the RLS and its effect seems to depend on other physico-chemical conditions of the protein solution. For instance, the pH effect on the RLS profile of β-LG solutions was maximal for β-LG solutions without salt at pH values closer to the isoelectric point as compared to the pH conditions of the maxima for solutions containing salt. It can be seen in Fig. 5b that the addition of salt results in suppression of the backscattering intensities for β-LG solutions at pH a close to the isoelectric point likely resulting in a decrease in attractive electrostatic interactions.

The sensitivity of backscattering measurements arising due to subtle changes in solution conditions and protein aggregation behaviour was demonstrated for lysozyme. Lysozyme is a highly basic protein of low molecular weight, with an isoelectric point located between pH 10.5 to 11.5 (Abdellatif *et al.*, 2004). Backscattering spectra were collected for lysozyme solutions at pH 4.0 at different salt concentrations. As seen in Figure 6.6 the backscattering profile intensity for solutions of lysozyme at pH 4 increased significantly when the NaCl concentration was increased which could be an indication of salt-induced aggregation. Salt-induced aggregation of lysozyme at around pH 4.0 has been well studied by using numerous experimental techniques

including X-ray diffraction (Narayanan & Liu, 2003), covalent cross-linking followed by SDS-PAGE (Wang *et al.*, 1996), NMR (Poznański *et al.*, 2005a), and calorimetry (Georgalis *et al.*, 1997; Poznański *et al.*, 2004; 2005b).

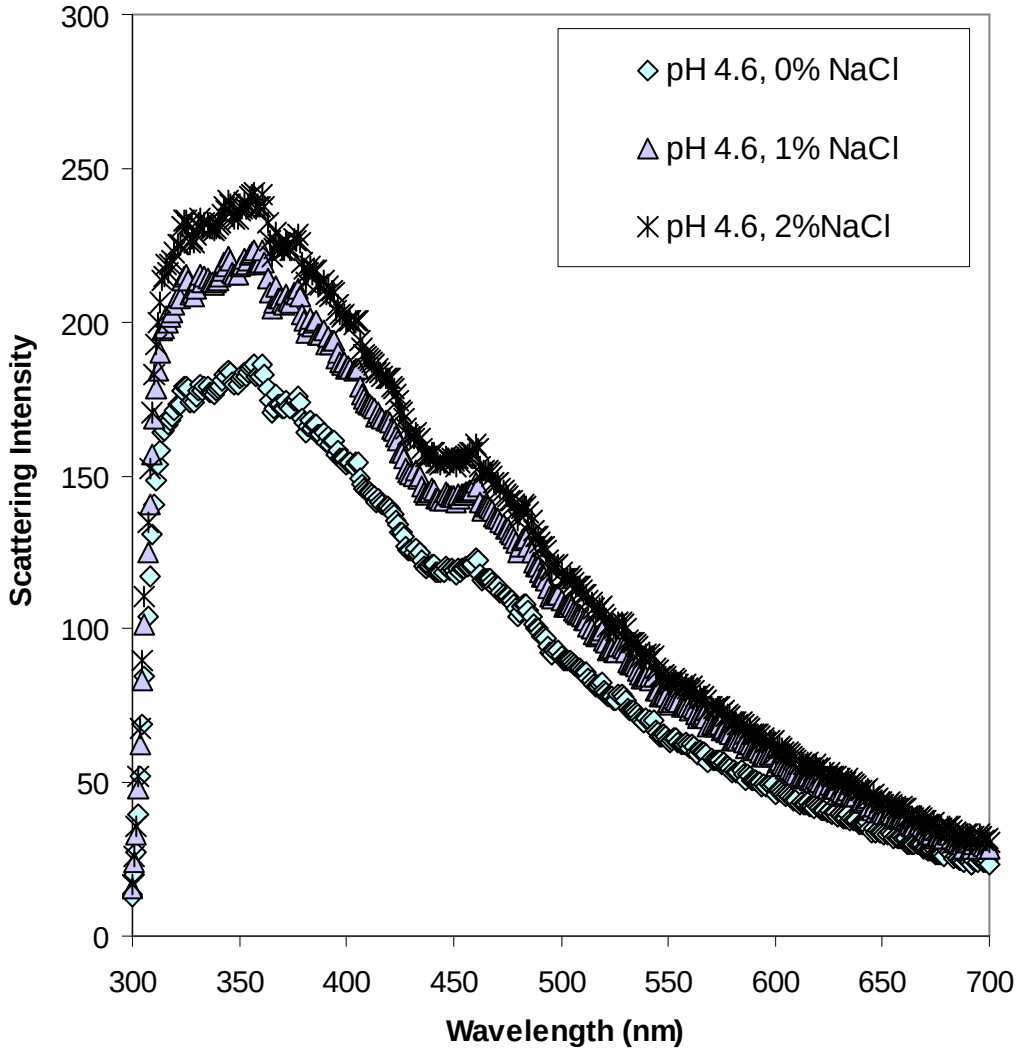


Figure 6.6: Effect of NaCl salt concentration on the FOP backscattering for 10 g/l Lyzosyme solution at pH 4.6.

The results of all these studies support that lysozyme exists as monomer in aqueous solution at pH 4.0 without salt. There is an observed salt-induced aggregation of protein which results from the screening of electrostatic interactions by nonspecific

binding of electrolyte counter ions to the charged solvent-exposed protein functional groups, which in turn reduces intermolecular repulsion forces (Ries-Kautt & Ducruix, 1989; Retailleau *et al.*, 2002). In addition to the salt effects, pH effects on backscattering were also studied. Figure 6.7 shows that the backscattering profile intensities also increased by increasing the pH (pH 4-10) which could be related to the monomer-to-dimer transition reported to occur in this pH range (Georgalis *et al.*, 1999).

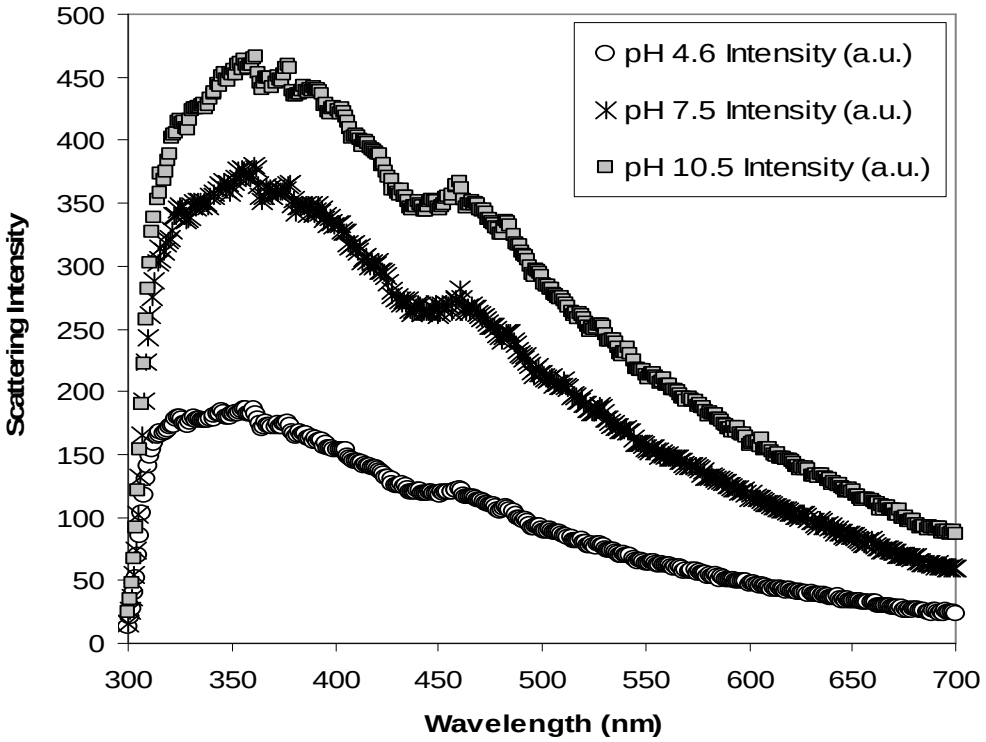


Figure 6.7: Effect of pH on the FOP backscattering for 10 g/l lysozyme solution.

6.4.1.3 Effect of Protein Concentration

In addition to being dependent on molecular weight, light scattering also has dependent on protein concentration. The concentration dependence of light scattering

was studied at different pHs using β-LG as a model protein system. As seen in Figure 6.8, increasing protein concentration results in an increase in the light scattering intensity. Light scattering was found to linearly increase with respect to β-LG concentration in the low concentration range (0-4 g/l). In the high protein concentration range (5-10 g/l), the deviations from linearity were significant, especially for β-LG solutions at pHs near the protein's isoelectric point (pH 4.5-5.0). The pattern of increase for light scattering is most likely the result of protein surface electrostatic properties and β-LG's self-association behavior.

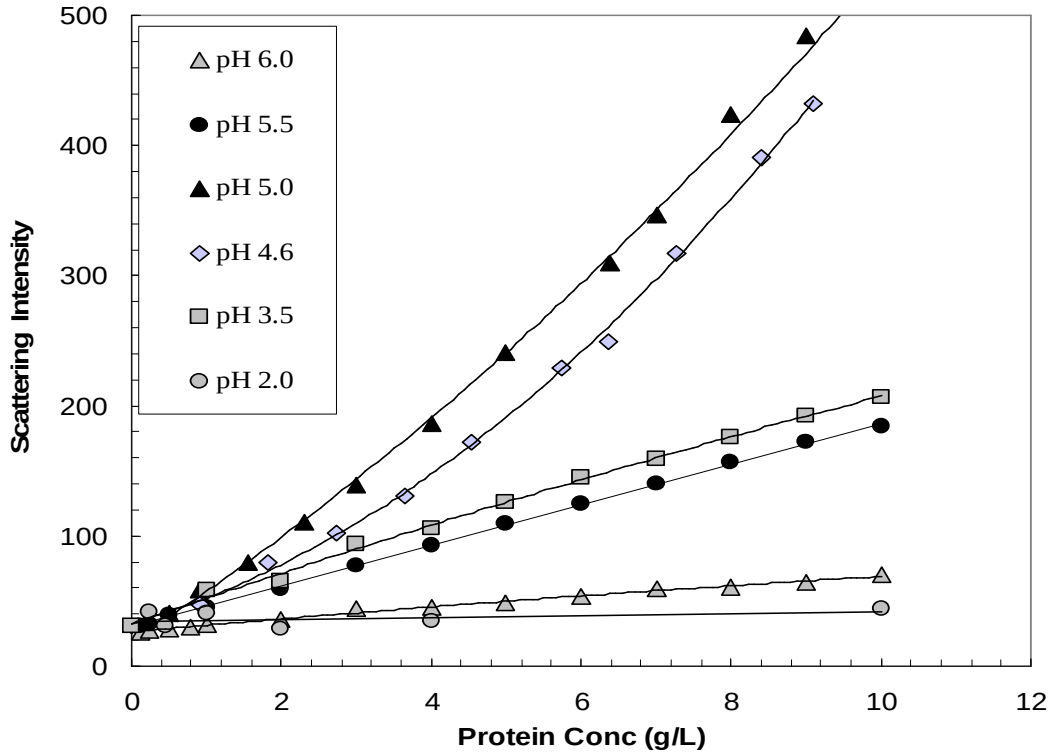


Figure 6-8: Effect of protein concentration on the FOP backscattering for β-LG solutions

The increase in slope with pH values close to the isoelectric point reflects a decrease in charge on the protein surface and the resulting increase in electrostatic attraction. The same conclusion can be made when the concentration-dependent curves

for lyzosyme are analyzed at different pHs (data not shown). It was found that the steepest curve was the one that corresponded to pH 11.0, which is close to lysozyme's isoelectric point. In general, for macromolecular solutes in an ideal solution, the magnitude of light scattering intensity *(I)* at a given wavelength (λ) is proportional to the product of the solute concentration (c) and its average molecular weight (M_w) as described by the following equation:

$$I_s = KM_w c \qquad \textbf{(6-1)}$$

K is an optical constant that depends only on the solvent properties, excitation wavelength and the angle between the incident light and the scattered light. *K* is therefore a system constant that is independent of the concentration of the solution and the average molecular weight of the macromolecule. It is also observed that the protein concentration does not only affect the magnitude of the scattering intensity, but it also influences the overall shape of the scattering spectrum. While light scattering intensity in the 300-360 nm range increased with increasing protein concentration, light scattering near the UV-visible wavelength range (230-280 nm) was reduced. Such attenuation of light scattering in the UV-visible range might be related to re-absorption effects that arise in systems with absorbing particles (Quinten *et al.*, 1995). Quinten *et al.* examined the scattering of colloidal systems containing strongly absorbing spherical silver nanoparticles nanometers in the wavelength range from the near UV to the near IR. They concluded that re-absorption of the scattered light by absorbing neighboring spheres or aggregates alter the measured light scattering spectra of these systems. According to observations reported in the literature, it is possible to propose a physical explanation of the results. Since UV light (200-300 nm) is strongly absorbed by

proteins, it might be possible that the scattered light in the UV range was partially reabsorbed by proteins in the solution, leading to the observed attenuation of the scattered light in the UV range upon increasing protein concentration. The contribution of re-absorption effects becomes more significant at higher protein concentrations. Interference of light re-absorption has been encountered with other optical spectroscopic techniques such as Raman resonance spectroscopy (Biswas and Umapathy, 1998) and fluorescence spectroscopy (Lacowics, 1999). In summary, it can be concluded that the backscattering pattern for protein solutions could be influenced by many factors such as electrostatic protein-protein interactions, protein concentration and average molecular weight, non-covalent self-association behavior and the presence of covalently formed aggregates.

6.4.2. Scattering of Bulk Solutions in Relation to Protein Membrane Fouling

As fouling is primarily driven by protein-protein interactions as well as interactions between depositing protein and the membrane surface (Malmsten., 1998), it is of interest to find a relationship between membrane fouling and the protein aggregation behavior in the feed solution. Such a relationship may be obtained from FOP measurements as previously described. FOP scattering measurements were made to determine the degree of aggregation for different protein solutions as a function of pH and ionic strength. Ultrafiltration experiments were then conducted using these protein solutions and the flux behavior estimated.

6.4.2.1. Scattering Behavior of Bipro Whey Protein Isolate

Scattering measurements were made to determine the degree of aggregation of Bipro whey protein isolate as a function of pH and concentration. According to FOP measurements, the degree of aggregation in the whey protein isolate was the highest at pH 4.5, which is close to the isoelectric point of the major whey protein constituents (β-LG, α-LA and BSA). Lower backscattering was observed for solutions at pH values higher or lower than pH 4.5 (data not shown). Backscattering data would predict that the worst solution conditions for the ultrafiltration of the protein isolate would correspond to a buffer concentration of 0.02 M at pH 4.5. Figures 6.9a and 6.9b show the time course of membrane resistance for different whey conditions and the corresponding scattering measurements. The data suggest that the higher light scattering intensity occurred at pH 4.5 for which membrane fouling would be considered more significant. This graph clearly suggests that the protein solution properties, as probed by FOP scattering, are most likely correlated with membrane fouling. The propensity of proteins to aggregate over the membrane surface is the highest at pH 4.5.

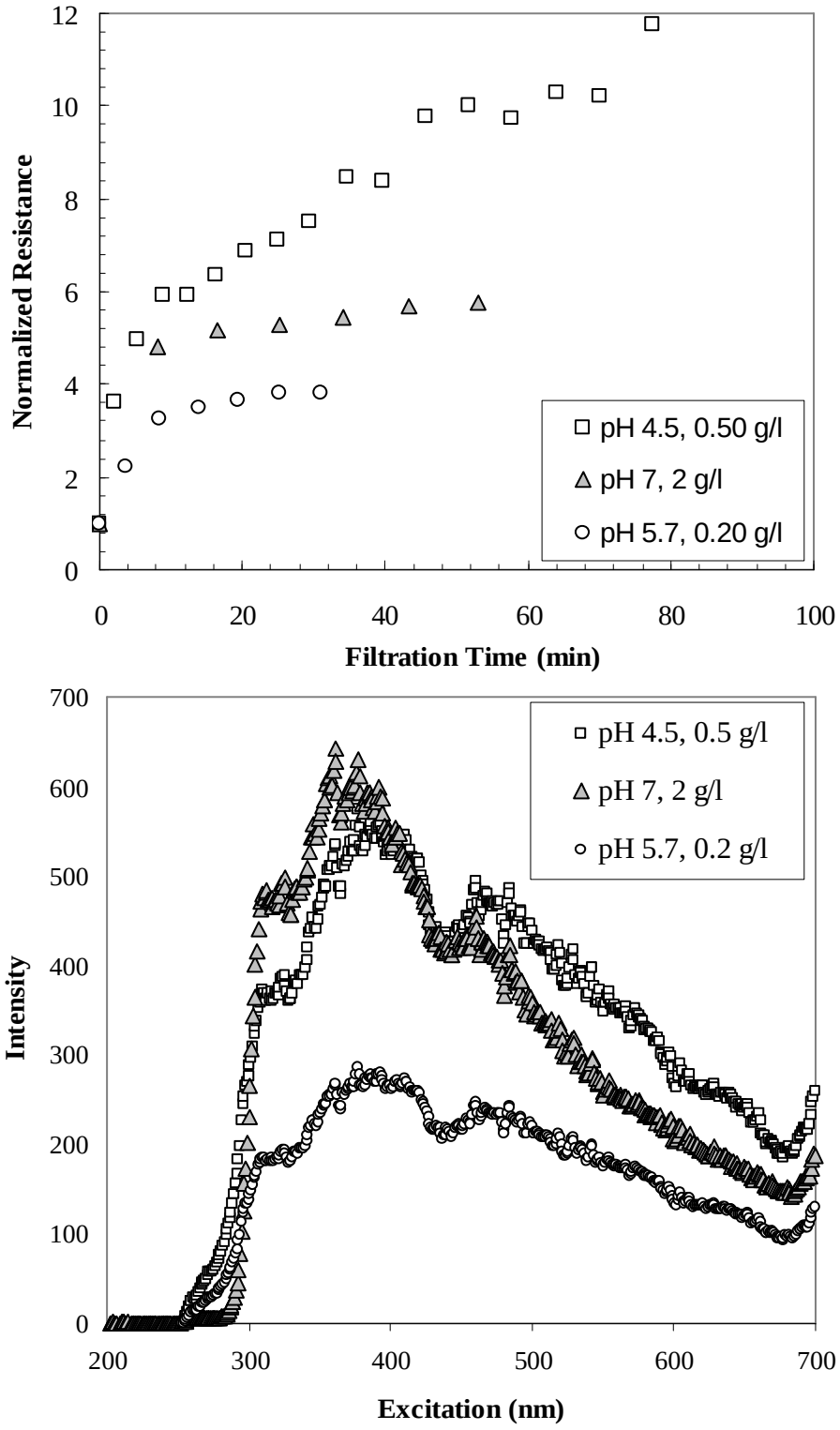


Figure 6.9: The time course of membrane total resistance for different whey protein isolate solutions (top) and its relevance to FOP scattering measurement (bottom)

6.4.3 Prediction of the Average Molecular Weight of Protein in Solution using Multiwavelength Light Scattering Spectral Data.

In the previous section, a qualitative study was performed where backscattering data was used as a probe of protein-protein interactions and aggregation behavior for lysozyme and β-lactoglobulin. It was also shown that backscattering data for whey protein isolate correlated well with their membrane fouling potential under different conditions. Analysis of the light scattering data in the previous sections did not include quantitative predictions on average molecular weight or particle size. In general it is expected that the light scattering will depend on several colloidal properties that will affect protein aggregation behavior including the molecular size, protein concentration and electrostatic charge as modulated by physicochemical conditions (pH and ionic strength). If quantitative predictions are desired it is necessary to assess whether the increase in light scattering is mainly due to protein concentration effects or increase in aggregate size. In this section, the possibility of inferring quantitative information about the protein aggregation behavior in the bulk solution from light scattering data was addressed by using multivariate regression modeling. It was hypothesized that a PLS prediction model based on multi-wavelength scattering measurements could be more useful than conventional models based on single scattering measurements. This is justified by the following:

I. Multiwavelength scattering measurements will contain more information. As stated above, light scattering intensity is a function of many parameters including molecular weight, concentration, size and shape of the aggregates

(Santos *et al.*, 1999). Information related to the effects of all of these parameters is expected to be imbedded in the entire multiwavelength scattering spectra.

II. The scattering spectra can be affected other optical phenomena that are not relevant to the property of interest. In section 3.1.3, the interference of bulk re-absorption effects of the scattered light was discussed. Upon increasing the protein concentration above 2 g/L, it was observed that the enhancement of light scattering intensity in the range 300-360 nm range was accompanied by attenuation of light scattering near the UV-visible wavelength range (230-280 nm). It was hypothesized that the scattered light in the UV range was partially reabsorbed by the protein in the solution, leading to the observed attenuation of the scattered light in the UV range. It is expected that by using the scattering data in the 200–800 nm range as input to the PLS model will result in a more robust predictive model with less sensitivity to bulk re-absorption effects at higher protein concentrations.

As protein aggregation behavior is usually quantified in terms of the average molecular weight, this property was chosen as the one to be predicted. The calibration of the PLS model was done based on average molecular weight data measured for β-LG by Verheul *et al.* (1999) who used small angle neutron scattering . The first step in the determination of the protein aggregation behavior by PLS methodology involved constructing the calibration matrix containing β-LG protein solutions prepared at different conditions (Table 6.1) identical to those used by Verheul *et al.* Under these

conditions, the calibration models were obtained. PLS calibration model was performed using Rayleigh scattering data acquired for samples (1-12) in Table 6.1 To select the number of factors in the PLS algorithm, a cross-validation method, leaving out one sample at a time, was employed. From a plot of the PRESS against the number of factors for each individual component, it was concluded that the optimal number of principal components yielding the smallest error (PRESS) was 3. Such PLS model was then tested on Rayleigh scattering spectra obtained for the remaining samples in Table 6.1 (samples 13-25) that were not used in the calibration. The PLS methodology was repeated using samples (13-25) as the calibration set and samples (1-12) as the testing set. The average molecular weight predicted from Rayleigh scattering data were then compared to values measured by Verheul *et al.* (1999) as given in Figures 6.10, 6.11 and 6.12.

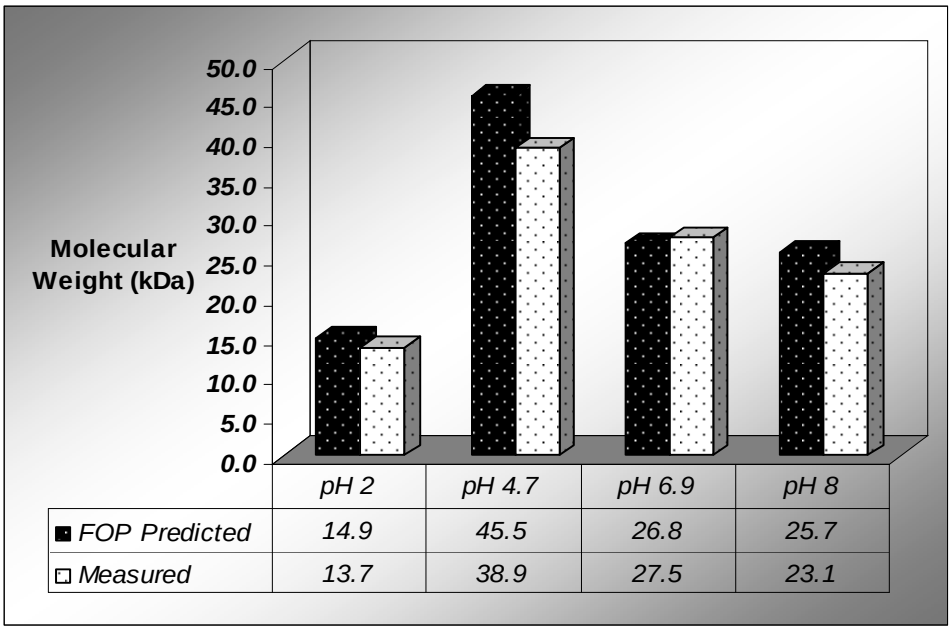


Figure 6-10: The average molecular weight predicted from Rayleigh scattering data compared to values measured by Verheul *et al.* (1999) for protein solutions at different pHs

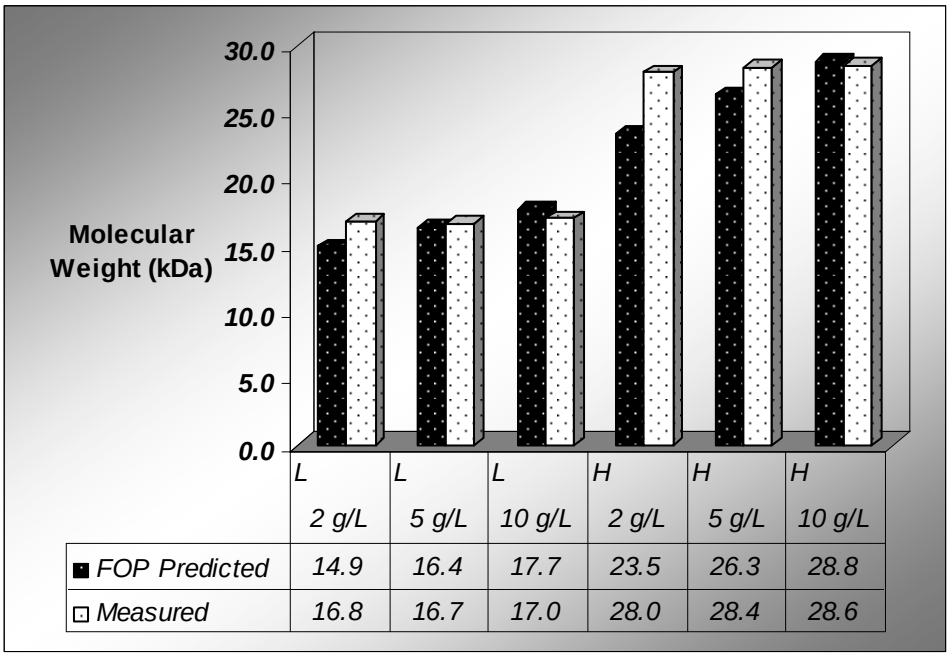


Figure 6-11: The average molecular weight predicted from Rayleigh scattering data compared to values measured by Verheul *et al.* (1999) for protein solutions at different protein concentrations, pH 2 and at low salt concentration (L), high salt concentration (H).

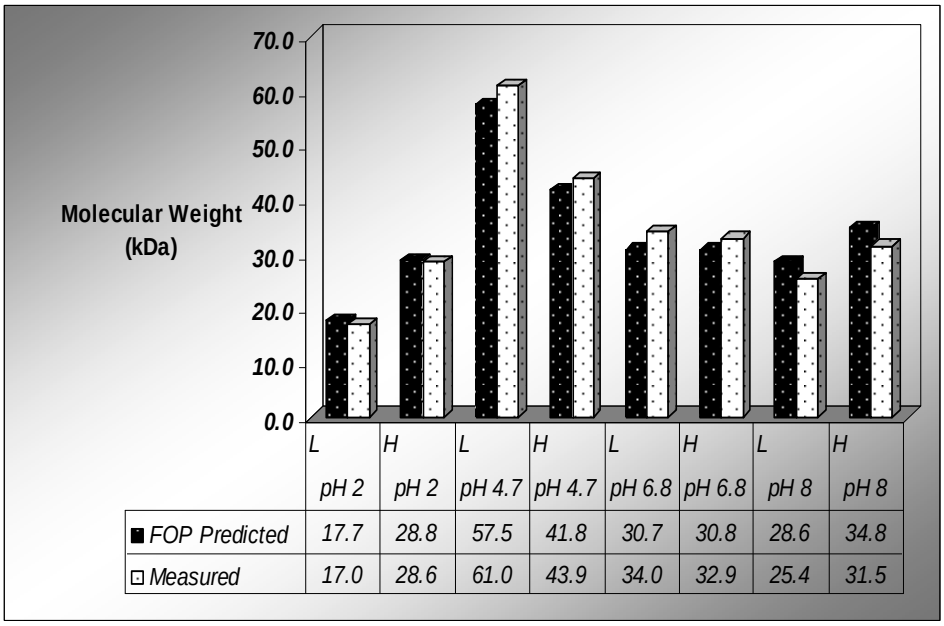


Figure 6-12: The average molecular weight predicted from Rayleigh scattering data compared to values measured by Verheul *et al.* (1999) for protein solutions at different pHs, low salt concentration (L) and high salt concentration (H).

Very good agreement between these results and those obtained by Verheul *et al.* (1999) support the applicability of the proposed method for detecting the changes in the protein aggregation behavior. With PLS modeling of the backscattering data, the average molecular weight of protein solutions under different conditions can be easily determined. The time required for obtaining one scattering spectral scan in the range 200–800 nm with the necessary scan quality was in the range of 20 sec but this time can be significantly shortened down to 10 seconds when using faster scanning speed. Thus, it is believed that this multiwavelength scattering measurements combined with PLS modeling may permit fast independent information on the aggregates size and the average molecular weight for more complex protein solutions.

6.5. Conclusions

The use of backscattering measurements for inference of fouling potential was examined in this study. It was hypothesized that by using this methodology, one could track changes in the aggregation behavior of proteins in solution as salt solution conditions were modified, affecting protein aggregation and ultimately their fouling behavior. It was found that protein aggregation behavior is strongly dependent on the conditions of the media and this behavior can be inferred from FOP backscattering measurements. Since the occurrence of aggregation has a major effect on membrane fouling, the prediction of these aggregation-phenomena by backscattering measurements could be effectively used to infer the membrane fouling potential of protein in solution.

CHAPTER 7

Conclusions and Recommendations

The focus of the research described in this thesis is the use of fluorescence spectroscopy data in combination with multivariate calibration tools for assessing the performance of membrane filtration processes. The discussion that follows will present the main conclusions of this work emphasizing the new techniques developed as compared to what had been previously reported in the literature. Advantages and drawbacks of the methods and sources of error will be summarized.

7.1 Practical Issues Related to Multiwavelength Fluorescence

Building a predictive model using intrinsic fluorescence spectra is not a straightforward process. Many factors affecting the fluorescence such as concentration quenching, protein-protein interactions, spectral overlapping and instrumental noise considerably complicate the model building process and the prediction accuracy of the obtained models. Chapter 3 and 4 presented a preliminary study on the capability of PLS-based regression models to handle the sensitivity of fluorescence measurements in the face of possible changes in the surrounding conditions. In Chapter 3, a model protein, β-lactoglobulin was used as a first simple case scenario consisting of predicting the residual concentration of β-LG after heat-treatment by using fluorescence spectroscopy data regressed with multivariate statistical techniques.

Assessment of aggregation and residual concentration following heat treatment was possible for a large range of protein concentration up to 12 mg/ml even though the intrinsic fluorescence of such relatively high concentration protein solutions is expected to be influenced by concentration quenching, light scattering, protein-protein interactions, heat treatment effects and instrumental noise. A signal preprocessing and filtering tool referred to as Standard Normal Variate (SNV) method was found to enhance the predictive accuracy and robustness of the sensor. Although the SNV approach is widely implemented in signal processing, the applicability of this method for filtering fluorescence signals has not been studied previously and is novel for this particular application. In Chapter 4, a two-component system consisting of the proteins α-LA and β-LG was used as a model system. The effect of factors including pH, temperature, total protein concentration, protein-protein interactions, and heat treatment time on the amount of aggregation was studied. The fluorescence spectra collected for a two-protein solution of α-lactalbumin and β-lactoglobulin served as inputs to the chemometric predictive model and the residual amounts of respective proteins after aggregation were estimated. The results of Chapter 4 demonstrate that multivariate models could be used to efficiently deconvolute the multi-wavelength fluorescence spectra collected for a high protein concentration solution and provide a fairly accurate quantification of respective proteins despite the significant overlapping between their emission profiles. Prediction was reasonably good in the high concentration range (2-10 g/l) because the fluorescence intensity is approximately linear with concentration. However, the prediction was found to be in error of more than 15% for samples whose actual residual protein concentration of α-LA and β-LG

was below 2 g/l. Such poor prediction in the low concentration range could be due to non-linearity in the fluorescence data that is not accounted for by the linear PLS regression. In fact, it was found that fluorescence intensity is locally linear with protein concentration in two regions: in the low concentration range (0-0.5 g/l), where fluorescence intensity increased approximately linearly with increasing protein concentration whereas fluorescence intensity decreased linearly with concentration in the high concentration range (2-10 g/l). Such concentration-dependence in the fluorescence signal will make building a single linear PLS model impossible due to the observed non-linearity over the entire concentration range. One of the possible causes of the non-linearity in the fluorescence measurements is the interference caused by phenomena such as concentration quenching, protein-protein interactions and light scattering. For instance, the first principal component of the PLS regression model was linear with respect to concentrations of α-LA and β-LG. However, the second principal component was found to be nonlinear with respect to concentration of α-LA and β-LG. Interferences due to fluorescence quenching were also observed in the spectra collected for the retentate side where a fiber optic probe (FOP) was used to acquire multi-wavelength fluorescence spectra of permeate and retentate at different times during the ultrafiltration of α-LA and β-LG protein solutions. The intrinsic fluorescence spectra of the protein solution in the retentate were observed to have unique spectral features that are different from that on the permeate side. Reliable and accurate estimation of the individual protein concentrations in the permeate and retentate directly from fluorescence spectra using a single PLS regression model is feasible provided that the following three conditions are satisfied:

1) The concentration range of individual proteins must be at the appropriate level where the relationship between a given protein concentration and fluorescence intensity is linear or approximately linear and this must apply to all proteins in the mixture.
2) The fluorescence signals from each of the given proteins in the mixture must be independent of each other.
3) The interferences from quenching phenomena and re-absorption effects are small compared to the fluorescence signal such that the spectral fingerprints for the inherent fluorophores are preserved.

Generally, such conditions are difficult to fulfill in a real protein ultrafiltration system where protein concentration in the retentate side increases progressively with filtration resulting in the occurrence of concentration-dependent phenomena that interfere with the fluorescence measurements. In addition to concentration quenching of the fluorescence signals, another important phenomenon that becomes more significant with increasing concentration in the retentate is energy transfer arising from protein-protein interactions. This phenomenon is possibly responsible for the concentration-dependent red shift of the fluorescence excitation and emission maxima and for the distortion of the spectral fingerprints for the mixture components. Under these conditions, fluorescence signals from individual proteins at high concentration levels may not be independent from each other and therefore the fluorescence signal for a protein mixture is not the sum of the individual contributions from the mixture components. By implementing a dilution strategy for the retentate side, concentrations

of α-LA and β-LG in permeate and retentate could be directly predicted from FOP fluorescence data within a reasonable accuracy using a single PLS model that encompassed a low concentration range where fluorescence was not expected to be significantly influenced by concentration-dependent interferences. However, there might be some practical limitations associated with dilution, e.g. imprecision introduced by dilution can be significant if excessive dilution is employed. PLS models are expected to be accurate if they are based on calibration samples that have acceptable upper and lower concentration limits within which the concentration of each protein is linearly related to the fluorescence intensity and the concentration-dependent interferences are insignificant compared to the fluorescence signal. However, a situation may occur where the concentration of a given protein in the diluted samples could lie within the suitable range of calibration while the concentration of other proteins could be outside the calibration range, i.e. the concentration may be either below or above the lower and the upper concentration limits respectively. To avoid this situation an optimal dilution ratio should be sought for which the concentration of *all individual protein* in the diluted samples lies within the suitable calibration range of the PLS model. Such a condition must apply to all proteins in the mixture. Establishing the optimal dilution ratio requires a preliminary knowledge of the individual protein concentrations in the retentate in order to decide how much dilution is required. One possible solution to this problem is to use two local PLS models in sequence for predicting the retentate concentration: one local PLS regression is constructed for the high concentration range where the protein concentration is negatively linear with the fluorescence signal while the other is

applicable for the low concentration range where the protein concentration is positively linear with the fluorescence signal. In Chapter 4 it was demonstrated that α-LA and β-LG protein concentrations could be predicted well from the excitation emission spectra in the concentration 2-10 g/l range.

In Chapter 5 a local PLS model in the low concentration range was used to predict concentrations of α-LA, β-LG and BSA in the retentate samples after dilution. Results in Chapter 4 and 5 showed that the PLS model constructed for the high concentration range (Chapter 4) is less accurate as compared to the PLS model constructed in the low concentration range (Chapter 5). The PLS model in the high concentration range could be used to provide preliminary estimates of individual protein concentrations in the retentate-undiluted samples. Preliminary estimates provide some orders of magnitude rather than precise and accurate prediction as a worst-case scenario. In summary, the need to use an optimal dilution ratio requires that *a priori* knowledge be incorporated into the modeling methodology for more accurate estimation of the proteins concentration time-profiles.

7.2 Multiwavelength fluorometry for monitoring membrane filtration processes

Monitoring, controlling and optimizing membrane-based filtration of complex biological fluids is difficult to achieve in practice because all the components of a complex biological fluid interfere with the performance of the filtration process (Darnon *et al*., 2002). During filtration of a complex biological fluid, tracking transient

changes in the separation efficiency requires information about the transport of the various feed components through the membrane. It was demonstrated that such information can be extracted from multi-wavelength fluorescence spectra collected for the feed, permeate and retentate. Multiwavelength fluorometry in combination with chemometric techniques has several attractive features: it is non-destructive, fast and relatively simple to perform. Chemometric tools have been found to be essential for extracting information from fluorescence spectra. It was demonstrated that two informative spectral zones in the two-dimensional fluorescence excitation-emission maps could be used for monitoring protein solutions during membrane-based filtration. These spectral zones are: (i) the intrinsic protein fluorescence zone and (ii) the Rayleigh scattering zone. The use of the information of these two spectral zones provided valuable insight into the performance of ultrafiltration as shown in Chapter 5 and 6. In Chapter 5, it was demonstrated that the analysis of the spectral zone corresponding to the intrinsic protein fluorescence can yield information about the concentrations of individual proteins in both permeate and retentate streams and thus transient changes in protein transmission and membrane selectivity in response to changes in pH and the transmembrane pressure could be estimated. In Chapter 6 it was shown that the Rayleigh scattering region of the excitation-emission matrix measures optical scattering, induced by protein aggregates and multimers, could be correlated with the fouling potential of protein solutions. Results show that light scattering profiles vary significantly with respect to pH and ionic strength. Feed solutions with higher light scattering properties also exhibited higher fouling potential. Thus, multiwavelength light scattering spectra, acquired using the fiber optic probe, was

shown to be a useful indicator for the protein self-association behavior. Since the protein aggregation behavior is concentration dependent, it is expected that an increase in protein concentration in the retentate side during ultrafiltration will be accompanied by a greater tendency for proteins to form aggregates. Having a rapid and sensitive method for monitoring protein aggregation during processing is important because the protein aggregation behavior is expected to change with respect to time. A predictive PLS model for estimating protein aggregation from Rayleigh scattering measurements was developed. At this point such a model can not be fully validated on samples collected from the retentate side due to the unavailability of a reliable independent measurement method for protein aggregation. Instead, such a predictive model was partially tested by using molecular weight experimental values obtained from literature. Although this comparison was only partial due to the limited amount of molecular weight data available, the findings verified the possibility of estimating the aggregate size from multiwavelength Rayleigh scattering spectra acquired using a conventional spectrofluorometer. In classical light scattering the characterization of protein solutions is inferred from scattering measurements that are performed at a number of different angles to the incident light. This allows the root-mean-square (rms) of the molecular radius to be calculated in addition to the molecular weight of the macromolecular solution. A conventional spectrofluorometer, such as the one used in this work, does not have the capability for detecting light scattering at different angles. In this thesis this limitation was overcome by determining light scattering measurements over a broader range of excitation wavelengths than those available with other classical light scattering techniques. Scattering measurements obtained by

varying wavelengths at a fixed scattering angle to the incident light have been reported to be equivalent to those obtained at different angles to the incident light for a fixed wavelength (Santos *et al.*, 1999).

Thus, the results in Chapters 5 and 6 imply that both intrinsic fluorescence and light scattering multiwavelength measurements provide complementary information about the process.

7.3 Future Work

7.3.1. Non-linear multivariate methods:

To correctly model non-linear dependence relationships between fluorescence and protein concentrations, some nonlinear PLS extensions would be useful. One of the approaches for developing non-linear PLS model would be to pretreat the input matrix (X) containing fluorescence measurements by including non-linear combinations of the original input measurements (i.e. logarithms, squared terms and cross terms) before performing linear PLS. A more structured approach to the development of a non-linear PLS model is to introduce a non-linear functional relationship that would relate the output latent variable scores **u** to the input latent variable scores **t**, without modifying the input and output variables.

7.3.2. Characterizing protein aggregates using multiwavelength Rayleigh scattering:

Protein-protein interactions and the protein aggregation state are influenced by pH and ionic strength and affect the transport behavior of most soluble components across the

filtration membrane (Sirkar and Prasad, 1987). This might provide more insight into the different types of aggregates. The PLS model based on Rayleigh scattering measurements needs to be validated using an independent method of aggregate determination such as small angle X-ray scattering. A method for estimating molecular weight distribution or particle size distribution from the multiwavelegth Rayleigh scattering spectra should be investigated further. One possibility is to infer the distribution shape parameters (aggregate mean size and standard deviation) from the spectra. This can be done by regressing the spectra (i.e. input data) against the distribution shape parameters (response data). Another possibility is to discretize the size distribution into different portions that can be used as multiple y responses and then perform PLS regression with the spectral data.

7.3.3. Monitoring product quality using fluorescence spectroscopy

This thesis focused on the use of fluorescence spectroscopy for monitoring the quantitative aspects of ultrafiltration protein fractionation. Meanwhile, Lilly (1992) emphasized the importance of product quality and not just the amount of product produced during ultrafiltration process. Product functionality and quality could be adversely affected during processing in ultrafiltration systems because the protein product can undergo conformational changes and subsequent denaturation during its passage through membrane pores possibly resulting in loss of its biological value (Sadana, 1998). More studies are required that clearly examine the use of fluorescence spectroscopy for monitoring the structure, functionality and biological activity of different proteins during the process. This could be beneficial for industry (Geisow,

1992; Sadana, 1998) because it would allow manipulating the processing conditions such that undesirable losses of the valuable biological product can be either prevented or at least minimized (Dunnill, 1983; Sadana, 1998). The multidimensional analysis for permeate and retentate can be extended further to include fluorescence anisotropy, life-time measurements and time-resolved fluorescence. All of these methods are well-known for providing a more detailed knowledge of conformational changes.

REFERENCES

1. Abdellatif A M, Rayas-Duarte P , Sanghoon. K (2004). Effect of starch on the thermal kinetics and transmittance properties of lysozyme. *J. Sci. Food Agric*. **85**: 450-458.

2. Amundson CH, Watanawanichakorn S, Hill CG. (1982). Production of enriched protein fraction of β-lactoglobulin and α-lactalbumin from cheese whey. *J Food Proc Pres,* **6**:55-71.

3. Arora, N.; Davis, R.H. (1994). Yeast cake layers as secondary membranes in dead-end microfiltration of bovine serum albumin. *Journal of Membrane Science* **92**: 247-256.

4. Babu, R.P.; Gaikar, V.G. (2001). Membrane characteristics as determinant in fouling of UF membranes. *Sep. Purif. Technol* **24**: 23–34.

5. Baker A. (1991). Organic Substances and Sediments in Water, Volume II: Processes and Analytical Robert A. Baker, Lewis Publishers, 1991, ISBN 0-87371-528-4, 560.

6. Balakrishnan, M; Agarwal, G.P. (1996). Protein fractionation in a vortex flow filter. I: Effect of system hydrodynamics and solution environment on single protein transmission. *Journal of Membrane Science* **112**: 47-74.

7. Baldini G, Beretta S, Chirico G, Franz H, Maccioni E, Mariani P, Spinozzi F. (1999). Salt-induced association of beta-lactoglobulin by light and X-ray scattering. *Macromolecules*. **32**(19): 6128-6138.

8. Becker, E.M.; Christensen, J.; Frederiksen, C.S.; Haugaard, V.K. (2003). Front-face fluorescence spectroscopy and chemometrics in analysis of yogurt: rapid analysis of riboflavin. *J. Dairy Sci.* **86**: 2508-2515.

9. Belfort, G.; Davis, R.H.; Zydney, A. (1994). The behavior of suspensions and macromolecular solutions in crossflow microfiltration. *Journal of Membrane Science* **96**: 1-58.

10. Bertrand-Harb C, Baday A, Dalgalarrondo M, Chobert JM, Haertle T. (2002). Thermal modifications of structure and co-denaturation of α-lactalbumin and β-lactoglobulin induce changes of solubility and susceptibility to proteases. *Nahrung* **46**:283-9.

11. Bhattacharjee S, Ghosh S, Datta S and Bhattacharjee C. (2006). Studies on ultrafiltration of casein whey using a rotating disk module: effects of pH and membrane disk rotation, *Desalination*. **195**(1-3): 95 -108.

12. Biswas N and Umapathy S (1998). Resonance Raman spectroscopy and ultrafast chemical dynamics. *In: Current Science*, **74** (4): 328-340.

13. Boehl D, Solle D, Hitzmann B and Scheper T. (2003). Chemometric modeling with two-dimensional fluorescence data for *Claviceps purpurea* bioprocess characterization. *J.Biotechnol,* **105**: 179-188.

14. Bowen W.R, Hall N.J, Pan LC, Sharif AO, Williams PA (1998). The relevance of particle size and zeta-potential in protein processing. *Nature Biotechnology* **16**, 785–787.

15. Boye JI, Alli I. (2000). Thermal denaturation of mixtures of α-lactalbumin and β-lactoglobulin: a differential scanning calorimetric study. *Food Res Int* **33**:673-682.

16. Bramaud C, Aimar P, Daufin G.(1997). Whey protein fractionation: Isoelectric precipitation of α-lactalbumin under gentle heat treatment. *Biotechnol Bioeng* **56**:391-397.

17. Bro R, Sidiropoulos N.D, Smilde A.K. (2002). Maximum likelihood fitting using ordinary least squares algorithms, *J. Chemom*. **16:** 387-400.

18. Bro, R.; Smilde, A.K. (2003). Centering and scaling in component analysis. *J. Chemo.* **17**:16-33.

19. Chan, R. (2002). Fouling mechanisms in the membrane filtration of single and binary protein solutions. PhD Thesis, School of Chemical Engineering and Industrial Chemistry, The University of New South Wales. Sydney, Australia.

20. Cheang B and Zydney AL. (2004). A two-stage ultrafiltration process for fractionation of whey protein isolate. *J Membr Sci* **231**:159–167.

21. Cheang, B., and Zydney, A.L. (2003). Separation of α-lactalbumin and β-lactoglobulin using membrane ultrafiltration. *Biotech. Bioeng* **83:** 201–209.

22. Chirino, A.J.; Mire-Sluis, A. (2004). Characterizing biological products and assessing comparability following manufacturing changes. *Nature Biotechnology* **22**:1383-1391.

23. Christensen J, Norgaard L, Bro R, Engelsen SB. (2006). Multivariate autofluorescence of intact food systems. *Chemical Reviews* **106**:1979-1993.

24. Cowie J. M. G. (1991). Polymer: Chemistry and Physics of Modern Materials, 2nd. Blackie: London.

25. Crespo JPSG, Trotin M, Hough D, Howell JA. (1999). Use of fluorescence labeling to monitor protein fractionation by ultrafiltration under controlled permeate flux. *J Membr Sci* 155:209-230.

26. Cromwell ME, Hilario E, Jacobson F. (2006). Protein Aggregation and Bioprocessing. *AAPS Journal*. **8(3)**: 572-579.

27. Darnon E., Lafitte L., Belleville MP and Rios GM (2002). A global approach of ultrafiltration of complex biological solutions. *Separation and Purification Technology*. **26**:283-293.

28. Davis, R.H., (1992) Modeling of Fouling of Crossflow Microfiltration Membranes *Separation and Purification Methods,* **21**: 75-126.

29. De la Fuente MA, Singh H, Hemar Y. (2002). Recent advances in the characterization of heat-induced aggregates and intermediates of whey proteins. *Trends Food Sci Technol* **13**: 262–274.

30. De Rham O, Chanton S. (1984). Role of ionic environment in insolubilization of whey protein during heat treatment of whey products. *J Dairy Sci* **67**: 939-949.

31. Deshpande SS. (2001). Principles and applications of luminescence spectroscopy. *Crit Rev Food Sci Nutr* **41**(3):155-224.

32. Elshereef R, Budman H, Moresoli C, Legge R. (2006). Fluorescence spectroscopy as a tool for monitoring solubility and aggregation behavior of β-lactoglobulin after heat treatment. *Biotech Bioeng* **95**:863-874.

33. Elshereef R, Budman H, Moresoli C, Legge RL. (2008). Fluorescence-based soft-sensor for monitoring β-lactoglobulin and α-lactalbumin solubility during thermal aggregation. *Biotech Bioeng* **99**:567-577.

34. Eriksson L., Johansson E., and Kettaneh-Wold N. (200l). *Multi- and Megavariate Data Analysis. Principles and Applications*. Umetrics Academy, Umeå.

35. Euston SR, Finnigan SR, Hirst RL. (2001). Aggregation kinetics in heated whey protein stabilized emulsions II: Effect of low-molecular weight emulsifiers. *Food Hydrocolloids* **15**: 253-262.

36. Fang Y, Dalgleish DG. (1998). The conformation of beta-lactoglobulin studied by FTIR: effect of pH, temperature and hydrophobic surfaces. *J. Colloid Interface Sci* **196**: 292-298.

37. Folta-Stogniew, E.; Williams, K.R. (1999). Determination of molecular masses of proteins in solution: Implementation of an HPLC size exclusion chromatography

and laser light scattering service in a core laboratory. *J. Biomol. Technol.*, **10**, 51–63.

38. Franken A.C.M., Sluys J.T.M., Chen V., Fane A.G., and Fell C.J.D. (1989). Role of protein conformation on membrane characteristics, *Proc. 5th World Filtration Congress.*

39. Geisow M.J. (1992). *TIBTECH, 10,* 230.

40. Geladi P, MacDougall D, Martens H. (1985). Linearization and scatter-correction for NIR reflectance spectra of meat. *Appl Spectrosc* **39**: 491-500.

41. Geladi, P.; Kowalski, B. (1986). Partial least-squares regression: a tutorial. *Anal Chim Acta* **185**: 1-17.

42. Georgalis Y, Umbach P, Zielenkiewicz A, Utzig E, Zielenkiewicz W, Zielenkiewicz P, Saenger W (1997) Microcalorimetric and small-angle light scattering studies on nucleating lysozyme solutions. *J Am Chem Soc* **119:** 11959-11965.

43. Ghosh R, Cui ZF. (1998). Fractionation of BSA and lysozyme using ultrafiltration: effect of pH and membrane surface pretreatment. *J Membr Sci* **139**:17-28.

44. Ghosh R, Cui ZF. (2000[a]). Purification of lysozyme using ultrafiltration: *Biotechnol Bioeng* **68**:191-202.

45. Ghosh R, Cui ZF. (2000b). Analysis of protein transport and polarization through membrane using pulsed sample injection technique. *J Membr Sci* **175**:75-84.

46. Ghosh R. (2003) Protein Bioseparation Using Ultrafiltration: Theory, Applications and New Developments, Imperial College Press, London, UK. 166 p.

47. Güell, C.; Czekaj, P.; Davis, R.H. (1998). Membrane Fouling During Microfiltration of a Protein Mixture and a Yeast Suspension. *13th Int. Cong. Chem. Proc. Engr.*.

48. Guettler B. (2006). Analysis of the fouling process during dead-end ultrafiltration of two different produced soy protein extracts. Diploma Thesis, University of Applied Sciences Berlin.

49. Guilbault GG. (1973). Practical fluorescence. Theory, methods and techniques, New York: Marcel Dekker Inc. pp. 498-502.

50. Hagedorn, A.; Legge, R.L.; Budman, H. (2003). Evaluation of spectrofluorometry as a tool for estimation in fed-batch fermentations. *Biotech. Bioeng*. **83**:104-111.

51. Harrington H. J. (1987). Poor Quality Cost. New York, NY: Marcel Dekker-ASQC Press.

52. Hasmukh, A.; Patel, Harjinder Singh, Palatasa Havea, Thérèse Considine, and Lawrence K. Creamer (2005), "Pressure-Induced Unfolding and Aggregation of the Proteins in Whey Protein Concentrate Solutions". *J. Agric. Food Chem., **53***: 9590 – 9601.

53. Heath C.A and Belfort G. (2006). Synthetic membranes in biotechnology: Realities and possibilities.

54. Herbert S, Rioua NM, Devaux MF, Riaublanca A, Bouchet B, Gallantb DJ, Dufourc E. (2000). Monitoring the identity and the structure of soft cheeses by fluorescence spectroscopy. *EDP Sciences* **80**: 621-634.

55. Hiemenz P.C. (1984). Polymer Chemistry - The Basic Concepts. *Marcel Dekke, p 659*.

56. Hilmer J.M and Scheper T. (1996). Application of scanning fluorometry for monitoring of a fermentation process. *Biotechnol Prog* **12:** 126–131.

57. Hisiger,S.; Jolicoeur,M. (2005) Multiwavelength fluorescence probe: Is one probe capable for on-line monitoring of recombinant protein production and biomass activity? J.Biotechnol. pp 117, **4**: 325-336.

58. Hoffmann MA, van Mil PJ. (1999). Heat-induced aggregation of β-lactoglobulin as a function of pH. *J Agric Food Chem* **47**:1898-905.

59. Howell, J.A., Sanchez, V., and Field, R.W. (1993). Membranes in Bioprocessing: Theory and Applications, Chapman & Hall, London.

60. Huisman IH, Pradanos P, Hernandez A. (2000). The effect of protein-protein and protein-membrane interactions on membrane fouling in ultrafiltration. *J Membr Sci* **179**:79-90.

61. Hunt JA, Dalgleish DG. (1994). The effect of pH on the stability and surface composition of emulsions made with whey protein isolate. *J. Agric. Food Chem*. **42**: 2131-2135.

62. Hunt CD, Tanis D, Bruce E, Taylor M. (2007). Optical signatures of seawater and potential use for verification of mid-ocean ballast water exchange. *Mar Ecol Prog Ser* **331**:35-47.

63. Iametti S, De Gregori B, Vecchio G, Bonomi F. (1996). Modifications occur at different structural levels during heat denaturation of β-lactoglobulin. *Eur J Biochem* **237**: 106-112.

64. Ingham K.C., Busby T.F., Sahlestrom Y and Castino F (1979). "Separation of macromolecules by ultrafiltration: Influence of protein adsorption, protein-protein interactions and concentration polarization. In: A.R. Cooper, Editor, Ultrafiltration membranes and application, Plenum Press, New York & London, p. 141.

65. James SC, Legge RL, Budman H. (2002). Comparative study of black-box and hybrid estimation methods in fed-batch fermentation. *J Process Control* **12**:113-121.

66. Kelly ST, Zydney AL. (1997). Protein fouling during microfiltration: comparative behavior of different model proteins. *Biotech Bioeng* **55**: 91-100.

67. Kelly, B. D and Peterson, J.T (2001) Virus removal by tangential-flow filtration for proteins therapeutics, in Membrane Separations in Biotechnology, Marcel Dekker, New York, pp. 351–392.

68. Kessler HG, Beyer HJ. (1992). Thermal denaturation of whey proteins and its effect in dairy technology. *Int J Biol Macromol* **13**:165-173.

69. Kuberkar V.T., Davis R.H. (1999). Effects of added yeast on protein transmission and flux in cross-flow membrane microfiltration. *Biotechnol. Prog.,* **15** (3): 472 - 479.

70. Lakowicz, RJ. (1999). Principles of Fluorescence Spectroscopy. 2nd Edition. Kluwer Academic Publishers and Plenum Press. NewYork. p 342–381**.**

71. Law AIJ, Leaver J. (1997). Effect of protein concentration on rates of thermal denaturation of whey proteins in milk. *J Agric Food Chem* **45**: 4255-4261.

72. Le Bon C, Nicolai T, Durand D. (1999). Kinetics of aggregation and gelation of globular proteins after heat-induced denaturation. *Macromolecules* **32:**6120–6127.

73. Lefèvre T, Subirade M. (1999). Structural and interaction properties of β-lactoglobulin as studied by FTIR spectroscopy. *Int J Food Sci & Technol* **34**:419-428.

74. Lemberge P, Van Espen PJ (1999). Quantitative energy-dispersive X-ray fluorescence analysis of liquids using partial least-squares regression. *X Ray Spectrom*. **28** (2): 77-85.

75. Lilly M.D. (1992). Trans. Inst.Chem.Eng., 70, Part C.

76. Lucey JA, Munro PA, Singh H. (1998). Rheological properties and microstructure of acid milk gels as affected by fat content and heat treatment *J Food Sci*.**63**:660-664.

77. Ma CQ, Li KA, Zhao FL, Tong SY.(1999). A study on the reaction mechanism between chrome-azurol S and bovine serum albumin. *Acta Chimica Sinica.* **57**(2):389.

78. Malmsten, M. (1998). Biopolymers at Interfaces, Marcel Dekker, Inc., New York.

79. Marshall, A. D.; Munro, P. A.; Trågårdh, G. (1997). Influence of permeate flux on fouling during the microfiltration of β-Lactoglobulin solutions under cross.flow conditions. *J. Membrane Sci*., **130**: 23-30.

80. Mayer LM, Schick LL, Loder TC. (1999). Dissolved protein fluorescence in two Marine estuaries. *Mar Chem* **64**:171–179.

81. McSwiney M, Singh H, Camapanella O, Creamer L.K. (1994). Thermal gelation and denaturation of bovine β-lactoglobulin A and B. *J Dairy Res* **61**: 221-232.

82. Meltzer T.H and Jornitz M.W. (2003). Filtration in the biopharmaceutical industry. 2nd Edition.

83. Mondor M, Ippersiel D, Lamarche F.O, Boye J.I. (2004). Effect of electroacidification treatment and ionic environment on soy protein extract particle size distribution and ultrafiltration permeate flux. *J.Membr.Sci*., **231:** 169.

84. Morr CV, (1985). Composition, physicochemical and functional properties of reference whey protein concentrates. *J Food Sci* **50**:1406-1411.

85. Mukai Y., Iritani E., and Murase T. (1998). Fractionation characteristics of binary protein mixtures by ultrafiltration. *Separation Sci. and Tech*. 33(2): 169-185.

86. Mulder, J. (1996). Basic Principles of Membrane Technology, 2nd edition.

87. Muller A, Chaufer B, Merin U, Duafin G (2003). Purification of α-lactalbumin from a prepurified acid whey: ultrafiltration or precipitation. *Lait* **83**:439-451.

88. Muller A, Daufin G, Chaufer B. (1999). Ultrafiltration modes of operation for the separation of α-lactalbumin from acid casein whey. *J Membr Sci.* **153**:9–21.

89. Mulvihill DM, Donovan M. (1987). Whey proteins and their thermal denaturation. A review. *Irish. J Food Sci Technol* **11**: 43-75.

90. Mycek MA, Pogue BW. (2003). Handbook of Biomedical Fluorescence, Marcel Dekker, New York, Basel.

91. Nakagaki M (1980). Turbidity of dispersions of nonspherical particles calculated in the Stevenson-Heller approximation. *J. Phys. Chem.* **84** (12): 1587-1591.

92. Newstead DF, Sandersom WB, Baucke AG. (1975). The effects of heat treatment and pH on heat stability of recombined evaporated milk. *New Zealand J Dairy Sci & Tech*. **10**:113-118.

93. Ofran,Y.; Rost, B. (2002). Analysing six types of protein-protein interfaces. *J Mol Biol* **325**: 377-387.

94. Ohno T, Bro R. (2006). Dissolved organic matter characterization using multi-way spectral decomposition of fluorescence landscapes. *Soil Sci Soc Am J* **70**:2028-2037.

95. Orchard B. (2006). Analyzing and monitoring: the importance of investigating turbidity. *Filtration & separation.* **43** (2):22.

96. Palacio, L., Ho, C.C.; Pradanos, P.; Hernandez, A.; Zydney, A.L. (2003). Fouling with Protein Mixtures in Microfiltration: BSA – Lysozyme and BSA - Pepsin", *J. Membrane Sci.***222**: 41-51.

97. Palecek, S. P.; Zydney, A.L.(1994). Hydraulic Permeability of Protein Deposits Formed During Microfiltration: Effect of Solution pH and Ionic Strength. *J. Membrane Sci.* **95**: 71-82.

98. Patel MT, Kilara A, Huffman LM, Hewitt SA, Houlihan AV (1990), Studies on Whey Protein Concentrates.1.Compositional and Thermal Properties. *J.Dairy Sci*.**73**:1439-1449.

99. Patra and Mishra AK (2002). Total synchronous fluorescence scan spectra of petroleum products, *Analytical and Bioanalytical Chemistry* **373**: 304-309.

100. Pearce RJ. (1987). Fractionation of whey proteins. In: Trends in whey utilization. *Bull Int Dairy Fed.* **212**:150-153.

101. Poznanski J, Georgalis Y, Wehr L, Saenger W, Zielenkiewicz P (2003) Comparison of two different lysozyme types under native and crystallization conditions using two-dimensional NMR and dynamic light scattering. *Biophys Chem* **104:** 605-616.

102. Poznanski J, Szymanski J, Basinska T, Slomkowski S, Zielenkiewicz W (2005a) Aggregation of aqueous lysozyme solutions followed by dynamic light scattering and 1H NMR spectroscopy. *J Mol Liq* **121:** 21-26.

103. Poznanski J, Wszelaka-Rylik M, Zielenkiewicz W (2004) Concentration dependencies of NaCl salting of lysozyme by calorimetric methods *Thermochim Acta* **409:** 25-32.

104. Poznanski J, Wszelaka-Rylik M, Zielenkiewicz W (2005b) HEW lysozyme salting by high-concentration NaCl solutions followed by titration calorimetry. *Biophys Chem* **113:** 137-144.

105. Prabakaran S, Damodaran S. (1997). Thermal unfolding of β-lactoglobulin: characterization of initial unfolding events responsible for heat-induced aggregation. *J Agric Food Chem* **45**: 4303-4308.

106. Qi XL, Brownlow S, Holt C, Sellers P. (1995). Effect of concentration on the denaturation of β-lactoglobulin near neutral pH. *Biochim Biophys Acta* **1248**:43-49.

107. Qin, S.J; McAvoy, T.J. (1992). Nonlinear PLS modeling using neural networks. *Comp Chem Eng* **16**: 379-397.

108. Quinten. M; Stier. J.1995. Absorption of scattered light in colloidal systems of aggregated particles, Colloid Polym Sci 273:233-241.

109. Ramanujam, N. (2000). Fluorescence spectroscopy in vivo. In *Encyclopedia of Analytical Chemistry*, John Willey & Sons, Ltd., Chichester, NY. pp. 20–56.

110. Rao CM. (1991). Synchronous scan fluorescence spectroscopy of proteins and human eye lenses. *Biochem Biophys Res Com*. **176**(3):1351-1357.

111. Redkar, S.G. and Davis, R.H. (1993). Crossflow Microfiltration of Yeast Suspensions in Tubular Filters *Biotechnology Progress* **9**: 625-634.

112. Renard D, Lefebvre J, Griffin MC, Griffin WG. (1998). Effects of pH and salt environment on the association of β-lactoglobulin revealed by intrinsic fluorescence studies. *Int J Biol Macromol* **22**:41-49.

113. Reshetnyak YK, Burstein EA. (2001). Decomposition of protein tryptophan fluorescence spectra into log-normal components: II. The statistical proof of discreteness of tryptophan classes in proteins. *Biophys J.* **81**: 1710-1734.

114. Retailleau P, Ducruix A, Ries-Kautt M (2002) Importance of the nature of anions in lysozyme crystallisation correlated with protein net charge variation. *Acta Crystallogr D* **58:** 1576-1581.

115. Ries-Kautt MM, Ducruix AF (1989) Relative effectiveness of various ions on the solubility and crystal growth of lysozyme. *J Biol Chem* **264**: 745-748.

116. Roland K Appel *et al* 2002, Enhanced Raman spectroscopy using collection optics designed for continuously tunable excitation, *Meas. Sci. Technol.* **13** 411-420.

117. SA, Houlihan AV (1990), Studies on Whey Protein Concentrates.1.Compositional and Thermal Properties. *J.Dairy Sci.*73:1439-1449.

118. Sadana A (1998). Bioseparation of Proteins: Unfolding/Folding and Validation, in: S. Ahuja (Ed.). *Academic Press*, New York. pp 4-10.

119. Saksena, S.; Zydney, A.L (1997). Influence of Protein-Protein Interactions on Bulk Mass Transport in Ultrafiltration. *J. Membrane Sci*. **125**: 93-108.

120. Santos N.C, Fernandes M.X.J.J.,. Castanho M.A.R.B (1999). Overcoming angular dependency when teaching light scattering using a spectrofluorometer. The molecular weights of latex beads, *J. Chem. Educ.*, **76**, 1259.

121. Sawyer L, Kontopidis G, Wu SY. (1999). β-Lactoglobulin – a three-dimensional perspective. *Int J Food Sci & Technol* **34**:409–418.

122. Schall CA, Arnold E, Wiencek JM (1996) Enthalpy of crystallization of hen egg-white lysozyme *J Crystal Growth* **165:**293-298.

123. Schokker EP, Singh H, Pinder DN, Norris GE, Creamer LK. (1999). Characterization of intermediates formed during heat-induced aggregation of β-lactoglobulin AB at neutral pH. *Int Dairy J*. **9**: 791-800.

124. Schokker, E.P.; Singh, H; Creamer, L. K. (2001). Heat-induced aggregation of ß-lactoglobulin A and B with ¤-lactalbumin. *International Dairy Journal* **10**:843–853.

125. Schreyer SK, Bidinosti M, Wentzell PD. (2002). Application of maximum likelihood principal components regression to fluorescence spectra. *Appl Spectrosc*, **56**: 789-796.

126. Schulze MA. (1997). An edge-enhancing nonlinear filter for reducing multiplicative noise. In: Dougherty ER, Astola J, editors. *Nonlinear Image Processing VIII, Proc. SPIE*, **3026**: 46-56.

127. Seasholtz MB, Kowalski BR. (1992). The effect of mean centering on prediction in multivariate calibration. *J Chemom*. **6**:103-111.

128. Shea, J.J. (1997). *Handbook of Instrumental Techniques for Analytical Chemistry*. (Settle , F.A., Ed.; 1997) Chapter 26.

129. Sirkar, KK and Prasad, R. (1987). Protein ultrafiltration - some neglected considerations. In Membrane Separations in Biotechnology, *McGregor*, *WC*, (ed.), *Marcel Dekker*, pp. 37-59.

130. Skořepová. J. (2007). Effect of electroacidification on ultrafiltration performance and physicochemical properties of soy protein extracts. PhD thesis, University of Waterloo.

131. Sulaiman, H.N., and Aroua, M.K. (2002). Cake layer reduction by gas sparging cross flow ultrafiltration of skim latex serum. *Songklanakarin J. Sci. Technol.* **24**: 947-953.

132. Svensson M, Sabharwal H, Håkansson A, Mossberg AK, Lipniunas P, Leffler H, Svanborg C, Linse S. (1999). Molecular characterization of α-lactalbumin folding variants that induce apoptosis in tumor cells. Jo *Biol Chem* **274**: 6388-6396.

133. Tan F, Guo M, Yu Q (2005). Studies on interaction between gatifloxacin and human serum albumin as well as effect of copper (II) on the reaction. *Spectrochim Acta A Mol Biomol Spectrosc.* **61**(13-14).

134. Tanaka S, Yamamoto M, Kawashima K, Ito K, Hayakawa R, Ataka M (1996). Kinetic study on the early stage of the crystallization process of two forms of lysozyme crystals by photon correlation spectroscopy. *J Crystal Growth* **168:** 44-49.

135. Tartakovsky, B., Lishman, L.A. and Legge, R.L.(1996). Application of multi-wavelength fluorometry for monitoring wastewater treatment dynamics. *Water Research* **30**: 2941-2948.

136. Tolkach A, Steinle S and Kulozik U. (2005). Optimization of thermal pretreatment conditions for the separation of native α-lactalbumin from whey protein concentrates by means of selective denaturation of β-lactoglobulin. *J Food Sci* **70**:557-566.

137. Torgrip, J. (2003). Chemometric Analysis of First order Chemical Data, PhD Thesis, Stockholm University. Sweden.

138. Tracey, E.M.; Davis, R.H. (1994). BSA Fouling of Track-etched Polycarbonate Microfiltration Membranes", Journal of Colloid and Interface Science, **167**: 104-116.

139. Tutunjian R.S. (2006). Ultrafiltration Processes in Biotechnology Scientific Systems Division Amicon Corporation 17 Cherry Hill Drive Danvers, Massachusetts 01923 **413**: 238 – 253.

140. Van Reis, R., and A. L. Zydney, Bioprocess membrane technology, *J. Membrane Sci.*, 297, 16-50 (2007).

141. Verheul, M.; Pedersen, J.S.; Roefs, S.P.F.M.; Kruif, K.G. de (1999), Association behavior of native beta-lactoglobulin. *Biopolymers*. **49**: 11-20.

142. Vetri V, Militello V. (2005). Thermal induced conformational changes involved in the aggregation pathways of β-lactoglobulin. *Biophys Chem* **113**: 83-91.

143. Vojinovic V, Cabral JMS, Fonseca LP. (2006). Real-time bioprocess monitoring Part I: In situ sensors. *Sensors and Actuators* B **114**:1083-1091.

144. Wan H, Lu JR and Cui Z.F. (2006). Separation of lysozyme from chicken egg white using ultrafiltration, *Sep Purif Technol.* **48**:133–142.

145. Wang F, Hayter J, Wilson LJ (1996). Salt-induced aggregation of lysozyme studied by cross-linking with glutaraldehyde: implications for crystal growth. *Acta Crystallogr D Biol Crystallogr.* **52:**901–908.

146. Wang, W.K., Ed., Membrane Separations in Biotechnology, 2nd Edition, Marcel Dekker, New York, 2001.

147. Wang,T.; Lucey, J.A. (2003). Use of Multi-Angle Laser Light Scattering and Size-Exclusion Chromatography to Characterize the Molecular Weight and Types of Aggregates Present in Commercial Whey Protein Products. *J. Dairy Sci*. **86**: 3090-3101.

148. Weinbreck F, Wientjes, RHW, Nieuwenhuijse H, Robijn GW. (2004). Rheological properties of whey protein/gum arabic coacervates. *J Rheology* **48**(6):1215-1228.

149. Wiberg, K. (2004). Multivariate spectroscopic methods for the analysis of solutions Doctoral thesis, Stockholm University, Faculty of Science, Department of Analytical Chemistry.

150. Xiong YL, Dawson KA, Wan L. (1993). Thermal aggregation of β-lactoglobulin: effect of pH, ionic environment, and thiol reagent. *J Dairy Sci* **76**:70-77.

151. Zeman, L.J and Zydney, A.L, (1996). Microfiltration and Ultrafiltration: Principles and Applications. Marcel Dekker, Inc., New York.

152. Zhu H., Damodaran S., (1994). Heat-induced conformational changes in whey protein isolate and its relation to foaming properties. *J Agric Food Chem* **42**:846-855.

APPENDICES

APPENDIX A: Fluorescence Spectral Differences between Food Proteins

Identification of proteins in a multi-flourophore mixture can be easily done using multiple excitation emission landscapes. Proteins can be easily distinguished from other fluorophores by looking at the Excitation and emission maxima of fluorescence spectra, in solution rather than for the identification of specific proteins, as can be seen in Figure A-1.

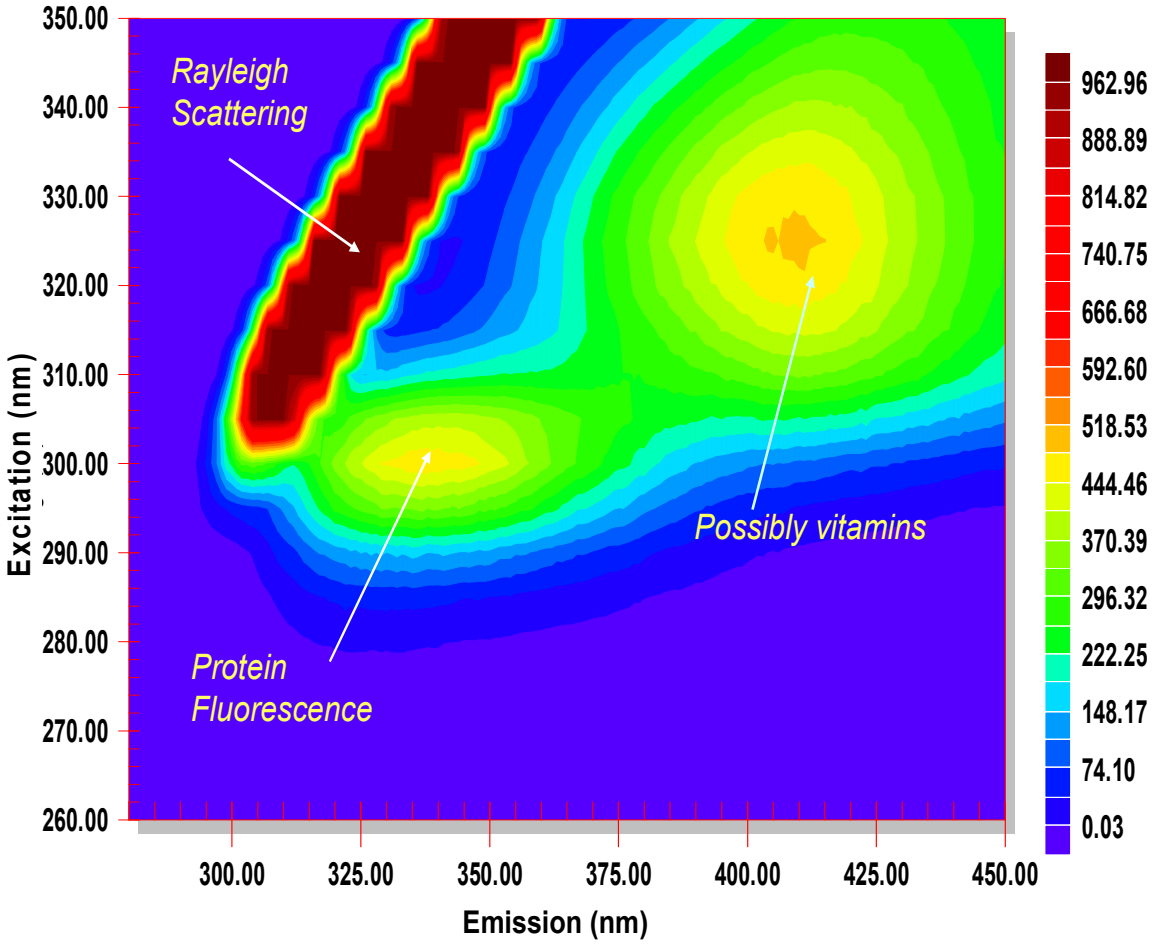


Figure A-1: Excitation-emission matrix for 6 g/l whey protein concentrate solution.

For instance, it is easy to discriminate between proteins and other components by looking at the fluorescence landscape acquired for milk where there is a significant difference in the peak positions of Tryptophan (the main fluorophore in proteins) and other biological components such as Vitamin B12 (Figure A-1).

Although EEM allows detecting the presence or the absence of proteins in a biological mixture, discriminating between different proteins or identification of the type of protein in a biological mixture is hard since there are quite subtle differences between their fluorescence spectra. The first objective of this work is to demonstrate spectral features at multiple combinations of excitation-emission wavelengths and to illustrate the usefulness of EEM for discriminating between proteins in biological mixtures when combined with chemometric methods.

APPENDIX B: Single Protein PLS Model

Appendix A presents the raw data for the validation of the Single Protein Model on b-LG solutions that were subjected to heat treatment at different conditions. Fluorescence spectra of these solutions are affected by fluorescence quenching, light scattering and instrumental noise. Four preprocessing methods, mean centering (MC), variable scaling (VS), standard normal variate (SNV) and normalization (NM), have been compared in terms of squared prediction error. The total sum of squared error (TSE) and the mean sum of squared error (MSE) were calculated for each set of samples.

Table B-1: Effect of data preprocessing methods on model prediction of beta-lactoglobulin concentrations after heat treatment of beta-lactoglobulin solutions at 85 C, Ph 4.5, 0.10 M sodium citrate acetate buffer.

Heating time (min)	Measured (g/l)	Predicted (g/l)				
		None	MC	VS	SNV	NM
0.00	9.32	6.95	6.83	6.85	8.20	8.51
0.00	9.43	8.49	8.47	8.01	9.04	9.08
5.00	9.25	9.25	9.34	9.09	8.73	8.64
10.00	8.88	8.89	9.05	8.78	8.07	7.89
15.00	7.59	7.97	8.06	7.34	7.54	7.55
15.00	7.28	8.22	8.40	7.71	7.27	7.10
20.00	5.39	6.21	6.31	5.67	5.83	5.76
25.00	5.09	5.42	5.58	5.06	4.88	4.76
30.00	4.30	5.10	5.24	4.75	4.39	4.22
30.00	4.09	4.53	4.64	4.14	4.07	4.01

Table B-2: Effect of data preprocessing methods on model prediction of beta-lactoglobulin concentrations after heat treatment of beta-lactoglobulin solutions at 85 C, Ph 4.5, 0.10 M sodium citrate acetate buffer.

Heating time (min)	Measured (g/l)	Squared Prediction Error				
		None	MC	VS	SNV	NM
0.00	9.32	6.70	7.34	7.23	1.79	1.50
0.00	9.43	0.22	0.25	0.91	0.01	0.00
5.00	9.25	0.47	0.59	0.27	0.03	0.01
10.00	8.88	0.44	0.67	0.31	0.02	0.22
15.00	7.59	0.55	0.70	0.01	0.09	0.06
15.00	7.28	0.52	0.82	0.04	0.05	0.25
20.00	5.39	0.90	1.11	0.17	0.32	0.31
25.00	5.09	0.70	1.00	0.23	0.09	0.02
30.00	4.30	1.11	1.43	0.50	0.12	0.02
30.00	4.09	0.34	0.48	0.04	0.02	0.00
	TTSE	***5.26***	***7.04***	***2.48***	***0.75***	***0.90***
	MSE	***0.76***	***0.88***	***0.52***	***0.29***	***0.32***

TTSE: Total sum of squared error
MSE: Mean Squared Error

Table B-3: Effect of data preprocessing methods on model prediction of beta-lactoglobulin concentrations after heat treatment of beta-lactoglobulin solutions at 82 C, Ph 4.5, 0.10 M sodium citrate acetate buffer.

Heating time (min)	Measured (g/l)	Predicted (g/l)				
		None	MC	VS	SNV	NM
15.00	9.39	5.90	5.73	6.26	7.82	8.52
15.0b	9.21	6.87	6.77	7.26	7.92	8.23
15.10	8.36	9.14	9.16	8.99	9.47	9.37
15.20	8.58	8.77	8.77	8.45	9.11	9.13
15.30	8.79	9.41	9.53	9.35	9.01	8.62
15.40	8.92	9.22	9.35	9.14	8.69	8.20
15.4b	8.95	7.92	7.92	7.93	8.66	8.38
15.50	8.53	8.22	8.24	8.17	8.62	8.41
15.60	8.45	8.05	8.07	7.90	8.51	8.14
15.70	8.54	8.42	8.50	8.29	8.31	8.02
15.80	8.34	7.94	7.95	7.91	8.38	8.40
15.8b	8.39	8.12	8.16	8.01	8.34	8.28
15.90	8.45	8.10	8.13	7.88	8.28	8.15
15.10	7.83	8.10	8.20	8.06	7.84	7.59
15.11	7.41	7.14	7.14	6.99	7.49	7.83
15.12	7.62	6.72	6.66	6.49	7.47	7.87
15.12b	7.63	7.89	7.96	7.60	7.60	7.75
15.13	7.64	7.63	7.68	7.34	7.54	7.54
15.14	7.31	7.15	7.18	6.83	7.38	7.28
15.15	6.95	7.71	7.83	7.29	7.22	6.91
15.16	6.89	7.75	7.87	7.37	7.21	6.79
15.17	6.80	7.43	7.58	7.18	6.82	6.45
15.18	6.59	7.63	7.69	7.40	7.52	7.29
15.18b	6.41	7.62	7.71	7.33	7.19	6.91

Table B-4: Effect of data preprocessing methods on model prediction of beta-lactoglobulin concentrations after heat treatment of beta-lactoglobulin solutions at 82 C, Ph 4.5, 0.10 M sodium citrate acetate buffer.

Heating time (min)	Measured (g/l)	Squared Prediction Error				
		None	MC	VS	SNV	NM
15.00	9.39	9.775	10.874	7.636	1.458	0.258
15.0b	9.21	4.631	5.064	3.112	1.215	0.625
15.10	8.36	0.034	0.039	0.001	0.261	0.171
15.20	8.58	0.002	0.002	0.135	0.085	0.100
15.30	8.79	0.698	0.907	0.605	0.193	0.002
15.40	8.92	0.527	0.739	0.417	0.040	0.083
15.4b	8.95	0.405	0.403	0.393	0.011	0.029
15.50	8.53	0.022	0.015	0.037	0.065	0.002
15.60	8.45	0.093	0.078	0.205	0.026	0.045
15.70	8.54	0.053	0.099	0.010	0.015	0.030
15.80	8.34	0.080	0.087	0.064	0.525	0.558
15.8b	8.39	0.142	0.173	0.073	0.355	0.291
15.90	8.45	0.128	0.152	0.021	0.292	0.165
15.10	7.83	0.237	0.342	0.203	0.052	0.001
15.11	7.41	0.014	0.014	0.001	0.215	0.647
15.12	7.62	0.074	0.108	0.251	0.234	0.768
15.12b	7.63	1.102	1.254	0.582	0.584	0.834
15.13	7.64	0.714	0.808	0.310	0.569	0.566
15.14	7.31	0.388	0.429	0.090	0.728	0.572
15.15	6.95	1.848	2.177	0.885	0.761	0.309
15.16	6.89	2.728	3.139	1.626	1.234	0.487
15.17	6.80	1.959	2.396	1.302	0.613	0.171
15.18	6.59	2.276	2.443	1.637	1.958	1.366
15.18b	6.41	2.179	2.455	1.406	1.083	0.581
	TTSE	20.335	23.326	13.365	11.114	8.404
	MSE	0.940	1.007	0.762	0.695	0.604

Table B-5: Effect of data preprocessing methods on model prediction of beta-lactoglobulin concentrations after heat treatment of beta-lactoglobulin solutions at 80 C, Ph 4.5, 0.10 M sodium citrate acetate buffer.

Heating time (min)	Measured (g/l)	Predicted (g/l)				
		None	MC	VS	SNV	NM
0	8.95	7.43	7.38	7.42	8.19	8.33
0	8.95	8.79	8.87	8.72	8.51	8.39
10	8.34	9.02	9.12	8.87	8.45	8.35
20	8.12	8.72	8.84	8.46	8.14	8.07
30	8.29	8.33	8.41	8.00	7.90	7.93
40	8.13	8.60	8.77	8.48	7.81	7.68
40	8.18	8.83	8.94	8.92	8.35	8.26
50	8.05	8.41	8.49	8.39	8.14	8.11
60	7.92	8.43	8.55	8.45	7.92	7.84
60	7.94	7.36	7.32	7.33	7.31	7.40
70	7.84	7.03	6.93	6.87	8.22	8.50
80	7.82	8.48	8.57	8.25	8.02	8.03
90	7.88	8.29	8.37	8.04	7.97	8.00
90	7.77	7.48	7.46	7.17	7.88	8.06
100	7.95	7.80	7.80	7.57	8.20	8.29
110	7.85	7.74	7.71	7.31	8.25	8.30

Table B-6: Effect of data preprocessing methods on model prediction of beta-lactoglobulin concentrations after heat treatment of beta-lactoglobulin solutions at 80 C, Ph 4.5, 0.10 M sodium citrate acetate buffer.

Heating time (min)	Measured (g/l)	Squared Prediction Error				
		None	**MC**	**VS**	**SNV**	**NM**
0	8.95	2.314	2.469	2.344	0.588	0.388
0	8.95	0.025	0.006	0.054	0.190	0.313
10	8.34	0.502	0.651	0.307	0.018	0.001
20	8.12	0.405	0.568	0.141	0.003	0.000
30	8.29	0.004	0.022	0.069	0.136	0.112
40	8.13	0.255	0.446	0.148	0.085	0.173
40	8.18	0.466	0.621	0.590	0.039	0.013
50	8.05	0.149	0.220	0.134	0.014	0.008
60	7.92	0.303	0.450	0.336	0.002	0.001
60	7.94	0.293	0.339	0.333	0.356	0.259
70	7.84	0.589	0.741	0.860	0.183	0.497
80	7.82	0.492	0.635	0.225	0.060	0.063
90	7.88	0.211	0.289	0.041	0.019	0.026
90	7.77	0.060	0.066	0.306	0.025	0.113
100	7.95	0.013	0.013	0.119	0.080	0.142
110	7.85	0.005	0.009	0.246	0.192	0.237
	TSSE	***3.770***	***5.074***	***3.908***	***1.403***	***1.958***
	MSE	***0.501***	***0.687***	***0.625***	***0.306***	***0.361***

Table B-7: Effect of data preprocessing methods on model prediction of beta-lactoglobulin concentrations after heat treatment of beta-lactoglobulin solutions at 75 C, Ph 4.5, 0.10 M sodium citrate acetate buffer.

Heating time (min)	Measured (g/l)	Predicted (g/l)				
		None	MC	VS	SNV	NM
0.00	8.62	7.89	7.85	7.87	8.69	8.76
0.00	8.76	7.80	7.74	7.81	8.82	8.89
10.00	8.35	9.00	9.07	8.88	8.69	8.67
20.00	8.34	8.10	8.13	7.67	8.20	8.24
30.00	8.30	8.71	8.79	8.37	8.38	8.37
30.00	8.25	8.31	8.38	7.93	8.09	8.07
40.00	8.56	8.14	8.18	8.02	8.29	8.28
50.00	8.18	8.71	8.80	8.43	8.30	8.26
60.00	8.18	8.45	8.52	8.19	8.30	8.27
60.00	8.09	7.91	7.90	7.55	8.31	8.40
70.00	7.94	8.04	8.04	7.80	8.43	8.44
80.00	8.27	7.88	7.90	7.63	8.18	8.22
90.00	8.06	7.73	7.73	7.41	8.10	8.20
90.00	7.93	8.63	8.78	8.61	7.96	7.78
100.00	8.23	7.84	7.83	7.92	8.35	8.36

Table B-8: Effect of data preprocessing methods on model prediction of beta-lactoglobulin concentrations after heat treatment of beta-lactoglobulin solutions at 75 C, Ph 4.5, 0.10 M sodium citrate acetate buffer.

Heating time (min)	Measured (g/l)	Squared Prediction Error				
		None	MC	VS	SNV	NM
0.00	8.62	1.178	1.278	1.228	0.085	0.048
0.00	8.76	1.687	1.847	1.643	0.078	0.043
10.00	8.35	0.011	0.001	0.051	0.169	0.185
20.00	8.34	0.875	0.834	1.868	0.710	0.640
30.00	8.30	0.041	0.017	0.296	0.292	0.300
30.00	8.25	0.356	0.273	0.959	0.660	0.705
40.00	8.56	0.377	0.327	0.539	0.215	0.220
50.00	8.18	0.149	0.086	0.439	0.636	0.703
60.00	8.18	0.322	0.247	0.686	0.514	0.556
60.00	8.09	1.102	1.112	1.971	0.427	0.317
70.00	7.94	0.602	0.598	1.025	0.148	0.139
80.00	8.27	0.836	0.792	1.352	0.380	0.327
90.00	8.06	1.755	1.739	2.685	0.911	0.727
90.00	7.93	0.195	0.086	0.213	1.231	1.658
100.00	8.23	1.592	1.610	1.387	0.572	0.555
	TSSE	11.079	10.847	16.342	7.029	7.122
	RMSE	0.859	0.850	1.044	0.685	0.689

Table B-9: Effect of data preprocessing methods on model prediction of beta-lactoglobulin concentrations after heat treatment of beta-lactoglobulin solutions at 65 C, Ph 4.5, 0.10 M sodium citrate acetate buffer.

Heating time (min)	Measured (g/l)	Predicted (g/l)				
		None	MC	VS	SNV	NM
0.00	9.38	7.80	10.04	9.97	9.38	9.35
20.00	9.38	9.00	9.53	9.44	9.16	9.17
20.00	9.14	8.10	9.70	9.55	9.47	9.46
40.00	8.76	8.71	9.30	8.96	9.41	9.45
60.00	8.52	8.31	9.10	8.65	9.19	9.28
60.00	8.36	8.14	9.53	9.12	9.04	9.06
80.00	8.41	8.71	9.64	9.63	9.12	9.04
100.00	8.52	8.45	9.72	9.70	8.75	8.63
120.00	8.48	7.91	9.83	9.88	8.77	8.63
120.00	8.55	8.04	9.71	9.57	8.79	8.72
140.00	8.41	7.88	8.53	8.51	9.26	9.33
160.00	8.62	7.73	9.38	9.35	8.71	8.56
180.00	9.13	8.63	8.94	8.70	8.96	8.94
180.00	8.99	7.84	9.14	8.94	8.97	8.94

Table B-10: Effect of data preprocessing methods on model prediction of beta-lactoglobulin concentrations after heat treatment of beta-lactoglobulin solutions at 65 C, Ph 4.5, 0.10 M sodium citrate acetate buffer.

Heating time (min)	Measured (g/l)	Squared Prediction Error				
		None	MC	VS	SNV	NM
0.00	9.38	2.515	0.429	0.344	0.000	0.001
20.00	9.38	0.149	0.021	0.004	0.051	0.048
20.00	9.14	1.080	0.314	0.161	0.109	0.098
40.00	8.76	0.002	0.287	0.038	0.422	0.472
60.00	8.52	0.046	0.334	0.016	0.442	0.566
60.00	8.36	0.048	1.371	0.586	0.468	0.496
80.00	8.41	0.086	1.504	1.467	0.496	0.394
100.00	8.52	0.005	1.451	1.394	0.053	0.012
120.00	8.48	0.324	1.821	1.959	0.084	0.024
120.00	8.55	0.264	1.345	1.041	0.058	0.029
140.00	8.41	0.279	0.014	0.010	0.725	0.853
160.00	8.62	0.790	0.585	0.535	0.008	0.003
180.00	9.13	0.245	0.033	0.185	0.029	0.034
180.00	8.99	1.330	0.021	0.003	0.000	0.002
	TSS	***7.161***	***9.530***	***7.743***	***2.945***	***3.032***
	RMSE	***0.715***	***0.825***	***0.744***	***0.459***	***0.465***

Table B-11: Effect of data preprocessing methods on model prediction of beta-lactoglobulin concentrations after heat treatment of beta-lactoglobulin solutions at 45 C, Ph 4.5, 0.10 M sodium citrate acetate buffer.

Heating time (min)	Measured (g/l)	Predicted (g/l)				
		None	MC	VS	SNV	NM
0.00	8.82	9.18	9.26	9.10	8.72	8.61
20.00	8.39	9.13	9.26	8.95	8.35	8.22
20.00	8.21	8.74	8.87	8.42	7.93	7.79
40.00	8.38	8.51	8.62	8.18	7.88	7.76
60.00	8.24	8.71	8.87	8.52	7.87	7.70
60.00	8.21	7.95	7.96	7.66	8.25	8.24
80.00	8.26	8.63	8.76	8.36	7.92	7.79
100.00	7.92	8.61	8.75	8.50	7.90	7.73
120.00	8.01	8.34	8.43	7.97	7.99	7.91
120.00	8.22	8.64	8.72	8.29	8.24	8.15
140.00	8.13	8.38	8.50	8.21	7.87	7.70
160.00	8.15	8.62	8.74	8.28	7.98	7.84

Table B-12: Effect of data preprocessing methods on model prediction of beta-lactoglobulin concentrations after heat treatment of beta-lactoglobulin solutions at 45 C, Ph 4.5, 0.10 M sodium citrate acetate buffer.

Heating time (min)	Measured (g/l)	Predicted (g/l)				
		None	MC	VS	SNV	NM
0.00	8.82	0.000	0.000	0.000	0.000	0.000
20.00	8.39	0.128	0.198	0.080	0.010	0.043
20.00	8.21	0.547	0.747	0.303	0.002	0.030
40.00	8.38	0.278	0.435	0.043	0.081	0.176
60.00	8.24	0.017	0.059	0.040	0.246	0.385
60.00	8.21	0.225	0.400	0.076	0.133	0.289
80.00	8.26	0.063	0.060	0.300	0.002	0.001
100.00	7.92	0.140	0.254	0.011	0.115	0.214
120.00	8.01	0.475	0.684	0.326	0.001	0.038
120.00	8.22	0.106	0.173	0.001	0.000	0.011
140.00	8.13	0.174	0.249	0.005	0.000	0.005
160.00	8.15	0.063	0.137	0.008	0.066	0.181
	TSS	0.220	0.347	0.018	0.028	0.096
	RME	2.436	3.744	1.210	0.684	1.469

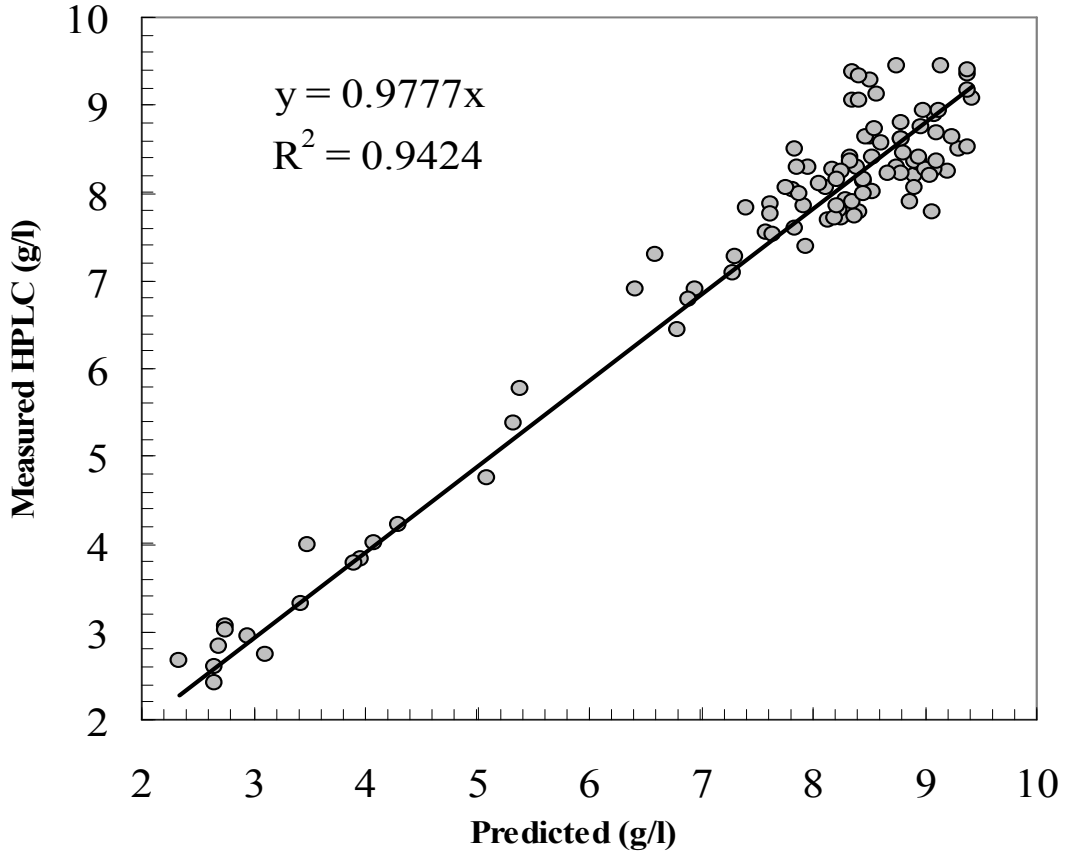


Figure B-1: Comparison between PLS model Predictions and HPLC measurements of β-LG concentrations for 110 samples that were subjected to heat treatment at different conditions

Figure B-1 suggests that the model prediction error increases for highly concentrated protein solutions which is most likely due to non-linearity in the fluorescence spectra that increases with increasing protein concentration (fluorescence quenching). Non-linearity is not totally accounted for in the PLS model development.

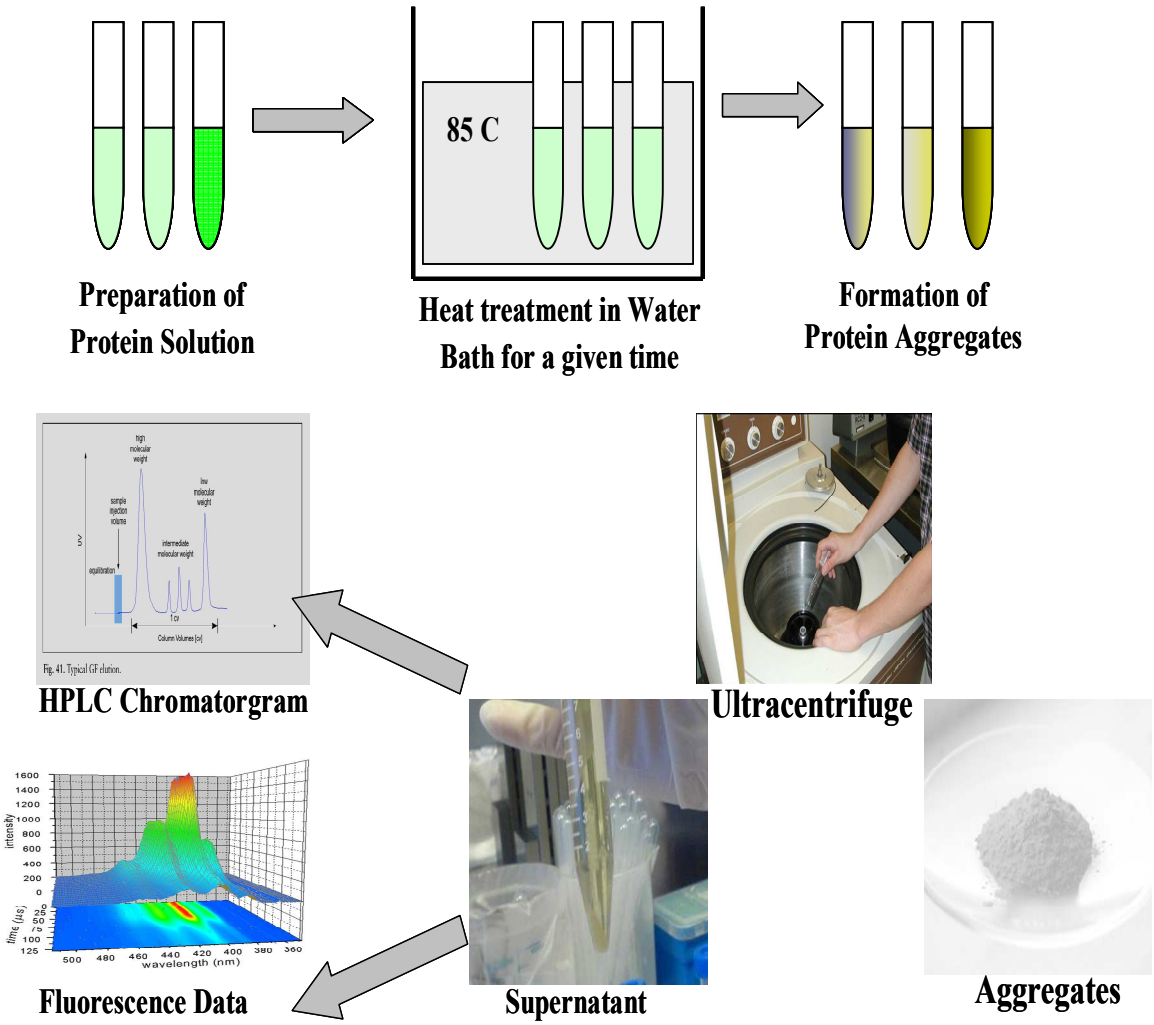


Figure B-2: Schematic diagram for the experimental procedure of thermal treatment of protein solutions at different conditions.

APPENDIX C: Binary Protein (α-LA/β-LG) PLS Model

Appendix C presents the raw data for the validation of the Binary Protein (α-LA/β-LG) PLS Model on (α-LA/β-LG) protein solutions that were subjected to heat treatment at different conditions. Fluorescence spectra of these solutions are affected by fluorescence quenching, light scattering and instrumental noise.

Table C-1: Model prediction of α-LA concentrations in the supernatant for samples of α-LA/β-LG solutions that were subjected to heat treatment at different conditions (pH 3.5, pH 3.7, pH 3.9).

Initial Conditions				*Heat treatment Conditions*	*α-LA concentration in the supernatant (g/l)*	
Sample	**Proportion of α-LA**	**Tot Con (g/l)**	**pH**	**T(C)**	**Measured**	**Predicted**
1.0	0.25	1.50	3.5	85.0	0.51	0.41
2.0	0.25	3.50	3.5	85.0	0.61	0.50
3.0	0.25	4.50	3.5	85.0	0.68	0.59
4.0	0.25	6.50	3.5	85.0	1.10	1.12
5.0	0.25	8.50	3.5	85.0	1.06	1.13
6.0	0.25	9.50	3.5	85.0	1.13	1.07
7.0	0.25	10.50	3.5	85.0	1.29	1.20
8.0	0.25	12.00	3.5	85.0	1.25	1.38
9.0	0.25	1.50	3.7	85.0	0.31	0.41
10.0	0.25	3.50	3.7	85.0	0.76	0.86
11.0	0.25	4.50	3.7	85.0	0.76	0.83
12.0	0.25	6.50	3.7	85.0	0.82	0.89
13.0	0.25	8.50	3.7	85.0	0.96	0.95
14.0	0.25	9.50	3.7	85.0	0.94	0.80
15.0	0.25	10.50	3.7	85.0	0.83	0.88
16.0	0.25	12.00	3.7	85.0	0.78	0.74
17.0	0.25	14.00	3.7	85.0	0.84	0.79
18.0	0.25	16.00	3.7	85.0	0.73	0.85
19.0	0.25	1.50	3.9	85.0	0.55	0.55
20.0	0.25	3.50	3.9	85.0	0.76	0.84
21.0	0.25	4.50	3.9	85.0	0.57	0.65
22.0	0.25	6.50	3.9	85.0	0.66	0.76
23.0	0.25	8.50	3.9	85.0	0.58	0.67
24.0	0.25	9.50	3.9	85.0	0.58	0.67
25.0	0.25	14.00	3.9	85.0	0.58	0.70
26.0	0.25	16.00	3.9	85.0	0.56	0.64

Table C-2: Model prediction of α-LA concentrations in the supernatant for samples of α-LA/β-LG solutions that were subjected to heat treatment at different conditions (pH 4.5, pH 5, pH 6).

Initial Conditions				*Heat treatment Conditions*	*α-LA concentration in the supernatant (g/l)*	
Sample	**Proportion of α-LA**	**Tot Con (g/l)**	**pH**	**T©**	**Measured**	**Predicted**
27.0	0.25	1.50	4.5	85.0	0.43	0.33
28.0	0.25	3.50	4.5	85.0	0.66	0.55
29.0	0.25	4.50	4.5	85.0	0.65	0.63
30.0	0.25	6.50	4.5	85.0	0.61	0.72
31.0	0.25	8.50	4.5	85.0	0.65	0.58
32.0	0.25	9.50	4.5	85.0	0.61	0.61
33.0	0.25	10.50	4.5	85.0	0.65	0.59
34.0	0.25	12.00	4.5	85.0	0.63	0.57
35.0	0.25	14.00	4.5	85.0	0.67	0.60
36.0	0.25	16.00	4.5	85.0	0.56	0.49
37.0	0.25	1.50	5.0	85.0	0.39	0.24
38.0	0.25	3.50	5.0	85.0	0.39	0.31
39.0	0.25	4.50	5.0	85.0	0.32	0.23
40.0	0.25	6.50	5.0	85.0	0.44	0.32
41.0	0.25	8.50	5.0	85.0	0.35	0.34
42.0	0.25	9.50	5.0	85.0	0.25	0.15
43.0	0.25	10.50	5.0	85.0	0.36	0.23
44.0	0.25	12.00	5.0	85.0	0.42	0.31
45.0	0.25	14.00	5.0	85.0	0.42	0.27
46.0	0.25	16.00	5.0	85.0	0.37	0.26
47.0	0.25	1.50	6.0	85.0	0.33	0.34
48.0	0.25	3.00	6.0	85.0	0.44	0.44
49.0	0.25	4.00	6.0	85.0	0.37	0.41
50.0	0.25	6.00	6.0	85.0	0.45	0.45
51.0	0.25	9.00	6.0	85.0	0.50	0.62
52.0	0.25	10.00	6.0	85.0	0.45	0.58
53.0	0.25	12.00	6.0	85.0	0.52	0.63
54.0	0.25	14.00	6.0	85.0	0.72	0.77

Table C-3: Model prediction of α-LA concentrations in the supernatant for samples of α-LA/β-LG solutions that were subjected to heat treatment at different conditions (proportion of α-LA in the original protein solution).

Initial Conditions				*Heat treatment Conditions*	*α-LA concentration in the supernatant (g/l)*	
Sample	**Proportion of α-LA**	**Tot Con (g/l)**	**pH**	**T(C)**	**Measured**	**Predicted**
55.0	0.25	3.00	3.9	75.0	0.68	0.61
56.0	0.25	4.00	3.9	75.0	0.85	0.80
57.0	0.25	6.00	3.9	75.0	0.82	0.95
58.0	0.25	8.00	3.9	75.0	0.91	0.85
59.0	0.25	10.00	3.9	75.0	0.99	0.93
60.0	0.25	12.00	3.9	75.0	0.98	1.01
61.0	0.25	14.00	3.9	75.0	1.01	1.00
62.0	0.25	16.00	3.9	75.0	0.99	0.88
63.0	0.10	12.00	3.9	75.0	1.07	1.07
64.0	0.18	12.00	3.9	75.0	0.97	1.06
65.0	0.25	12.00	3.9	75.0	1.02	1.05
66.0	0.31	12.00	3.9	75.0	1.05	1.04
67.0	0.36	12.00	3.9	75.0	1.01	1.04
68.0	0.40	12.00	3.9	75.0	0.97	1.01
69.0	0.44	12.00	3.9	75.0	0.95	0.99
70.0	0.47	12.00	3.9	75.0	0.96	1.01
71.0	0.50	12.00	3.9	75.0	0.95	1.00
72.0	0.55	12.00	3.9	75.0	0.93	0.98

Table C-4: Model prediction of β-LG concentrations in the supernatant for samples of α-LA/β-LG solutions that were subjected to heat treatment at different conditions (pH 3.5, pH 3.7, pH 3.9).

Initial Conditions				*Heat treatment Conditions*	*β-LG concentration in the supernatant (g/l)*	
Sample	**Proportion of α-LA**	**Tot Con (g/l)**	**pH**	**T(C)**	**Measured**	**Predicted**
1.0	0.25	1.50	3.5	85.0	1.18	0.84
2.0	0.25	3.50	3.5	85.0	2.67	1.75
3.0	0.25	4.50	3.5	85.0	2.67	2.67
4.0	0.25	6.50	3.5	85.0	4.44	4.32
5.0	0.25	8.50	3.5	85.0	6.62	6.03
6.0	0.25	9.50	3.5	85.0	7.83	6.76
7.0	0.25	10.50	3.5	85.0	8.70	7.29
8.0	0.25	12.00	3.5	85.0	8.20	8.63
9.0	0.25	1.50	3.7	85.0	1.10	0.92
10.0	0.25	3.50	3.7	85.0	1.93	1.66
11.0	0.25	4.50	3.7	85.0	3.03	2.64
12.0	0.25	6.50	3.7	85.0	4.45	4.12
13.0	0.25	8.50	3.7	85.0	4.29	3.69
14.0	0.25	9.50	3.7	85.0	3.98	4.60
15.0	0.25	10.50	3.7	85.0	5.10	4.78
16.0	0.25	12.00	3.7	85.0	4.26	4.00
17.0	0.25	14.00	3.7	85.0	4.46	3.59
18.0	0.25	16.00	3.7	85.0	4.39	4.27
19.0	0.25	1.50	3.9	85.0	1.31	1.27
20.0	0.25	3.50	3.9	85.0	2.25	2.02
21.0	0.25	4.50	3.9	85.0	2.48	2.23
22.0	0.25	6.50	3.9	85.0	3.14	2.88
23.0	0.25	8.50	3.9	85.0	2.90	2.64
24.0	0.25	9.50	3.9	85.0	2.85	2.60
25.0	0.25	14.00	3.9	85.0	2.65	2.50
26.0	0.25	16.00	3.9	85.0	2.26	2.16

Table C-5: Model prediction of β-LG concentrations in the supernatant for samples of α-LA/β-LG solutions that were subjected to heat treatment at different conditions (pH 4.5, pH 5, pH 6).

Initial Conditions				*Heat treatment Conditions*	*β-LG concentration in the supernatant (g/l)*	
Sample	**Proportion of α-LA**	**Tot Con (g/l)**	**pH**	**T©**	**Measured**	**Predicted**
27.0	0.25	1.50	4.5	85.0	0.95	0.99
28.0	0.25	3.50	4.5	85.0	1.31	1.38
29.0	0.25	4.50	4.5	85.0	1.32	0.96
30.0	0.25	6.50	4.5	85.0	1.08	1.06
31.0	0.25	8.50	4.5	85.0	1.04	0.82
32.0	0.25	9.50	4.5	85.0	1.08	0.81
33.0	0.25	10.50	4.5	85.0	1.14	1.39
34.0	0.25	12.00	4.5	85.0	1.26	0.84
35.0	0.25	14.00	4.5	85.0	1.18	0.78
36.0	0.25	16.00	4.5	85.0	1.24	1.05
37.0	0.25	1.50	5.0	85.0	0.28	0.39
38.0	0.25	3.50	5.0	85.0	0.52	0.49
39.0	0.25	4.50	5.0	85.0	0.63	0.30
40.0	0.25	6.50	5.0	85.0	0.57	0.35
41.0	0.25	8.50	5.0	85.0	0.71	0.45
42.0	0.25	9.50	5.0	85.0	0.67	0.48
43.0	0.25	10.50	5.0	85.0	0.65	0.31
44.0	0.25	12.00	5.0	85.0	0.52	0.29
45.0	0.25	14.00	5.0	85.0	0.54	0.15
46.0	0.25	16.00	5.0	85.0	0.74	0.28
47.0	0.25	1.50	6.0	85.0	0.72	0.86
48.0	0.25	3.00	6.0	85.0	0.75	0.90
49.0	0.25	4.00	6.0	85.0	0.73	0.80
50.0	0.25	6.00	6.0	85.0	0.79	0.91
51.0	0.25	9.00	6.0	85.0	0.86	0.93
52.0	0.25	10.00	6.0	85.0	0.86	0.90
53.0	0.25	12.00	6.0	85.0	1.02	1.09
54.0	0.25	14.00	6.0	85.0	1.07	1.26

Table C-6: Model prediction of β-LG concentrations in the supernatant for samples of α-LA/β-LG solutions that were subjected to heat treatment at different conditions (proportion of α-LA in the original protein solution).

Initial Conditions				*Heat treatment Conditions*	*β-LG concentration in the supernatant (g/l)*	
Sample	**Proportion of α-LA**	**Tot Con (g/l)**	**pH**	**T(C)**	**Measured**	**Predicted**
55.0	0.25	3.00	3.9	75.0	1.27	1.89
56.0	0.25	4.00	3.9	75.0	2.78	2.58
57.0	0.25	6.00	3.9	75.0	4.37	3.85
58.0	0.25	8.00	3.9	75.0	5.71	5.20
59.0	0.25	10.00	3.9	75.0	6.47	6.25
60.0	0.25	12.00	3.9	75.0	7.02	7.29
61.0	0.25	14.00	3.9	75.0	7.90	8.48
62.0	0.25	16.00	3.9	75.0	8.20	8.11
63.0	0.10	12.00	3.9	75.0	6.88	7.54
64.0	0.18	12.00	3.9	75.0	7.05	6.99
65.0	0.25	12.00	3.9	75.0	6.21	6.60
66.0	0.31	12.00	3.9	75.0	6.44	6.25
67.0	0.36	12.00	3.9	75.0	6.03	5.79
68.0	0.40	12.00	3.9	75.0	5.62	5.47
69.0	0.44	12.00	3.9	75.0	4.81	5.23
70.0	0.47	12.00	3.9	75.0	5.11	5.05
71.0	0.50	12.00	3.9	75.0	4.61	4.68
72.0	0.55	12.00	3.9	75.0	3.87	4.34

Table C-7: PLS Model predictions for α-LA and β-LG aggregation behavior compared to the reference values estimated by HPLC.

Initial Conditions		Heat treatment conditions		Model Predictions (Aggregation %)		Reference Values (Aggregation %)	
α-LA Fraction	Total Conc (g/L)	T(°C)	pH	α-LA	β-LG	α-LA	β-LG
0.25	6.50	85	3.5	32.39	5.92	30.89	11.48
0.25	8.50	85	3.5	29.54	8.29	46.77	5.43
0.25	9.50	85	3.5	40.93	5.01	55.12	5.08
0.25	10.50	85	3.5	44.78	11.81	54.13	7.40
0.25	12.00	85	3.5	58.46	22.61	53.86	14.16
0.25	14.00	85	3.5	67.75	38.11	52.57	28.88
0.25	1.50	85	3.7	18.20	2.55	7.30	2.00
0.25	3.50	85	3.7	40.49	26.31	13.30	3.60
0.25	4.50	85	3.7	32.66	10.28	26.52	4.80
0.25	6.50	85	3.7	49.68	8.73	45.30	15.55
0.25	8.50	85	3.7	54.94	32.76	55.25	33.30
0.25	9.50	85	3.7	60.46	44.15	66.53	35.44
0.25	10.50	85	3.7	68.27	35.23	66.59	39.34
0.25	12.00	85	3.7	73.97	52.65	75.43	52.00
0.25	14.00	85	3.7	75.91	57.49	77.33	58.30
0.25	16.00	85	3.7	81.76	63.41	78.75	64.45
0.25	3.00	75	3.9	9.27	17.00	18.23	15.94
0.25	4.00	75	3.9	15.46	10.00	20.04	13.85
0.25	6.00	75	3.9	45.00	7.00	36.79	14.41
0.25	8.00	75	3.9	54.67	10.00	57.49	13.33
0.25	10.00	75	3.9	60.45	13.78	62.88	16.72
0.25	12.00	75	3.9	67.18	22.03	66.47	18.98
0.25	14.00	75	3.9	71.15	24.79	71.31	19.28
0.25	16.00	75	3.9	75.19	31.64	78.09	32.39
0.10	12.00	75	3.9	10.91	36.28	10.68	30.15
0.18	12.00	75	3.9	55.57	28.18	51.29	28.76
0.25	12.00	75	3.9	65.86	31.04	64.97	26.71
0.31	12.00	75	3.9	71.47	22.47	71.81	24.82
0.35	12.00	75	3.9	76.37	21.78	75.79	24.83
0.40	12.00	75	3.9	79.71	21.95	79.00	24.03
0.44	12.00	75	3.9	83.92	28.76	81.14	22.47
0.47	12.00	75	3.9	83.81	19.56	82.20	20.48
0.50	12.00	75	3.9	87.31	23.23	83.35	22.04

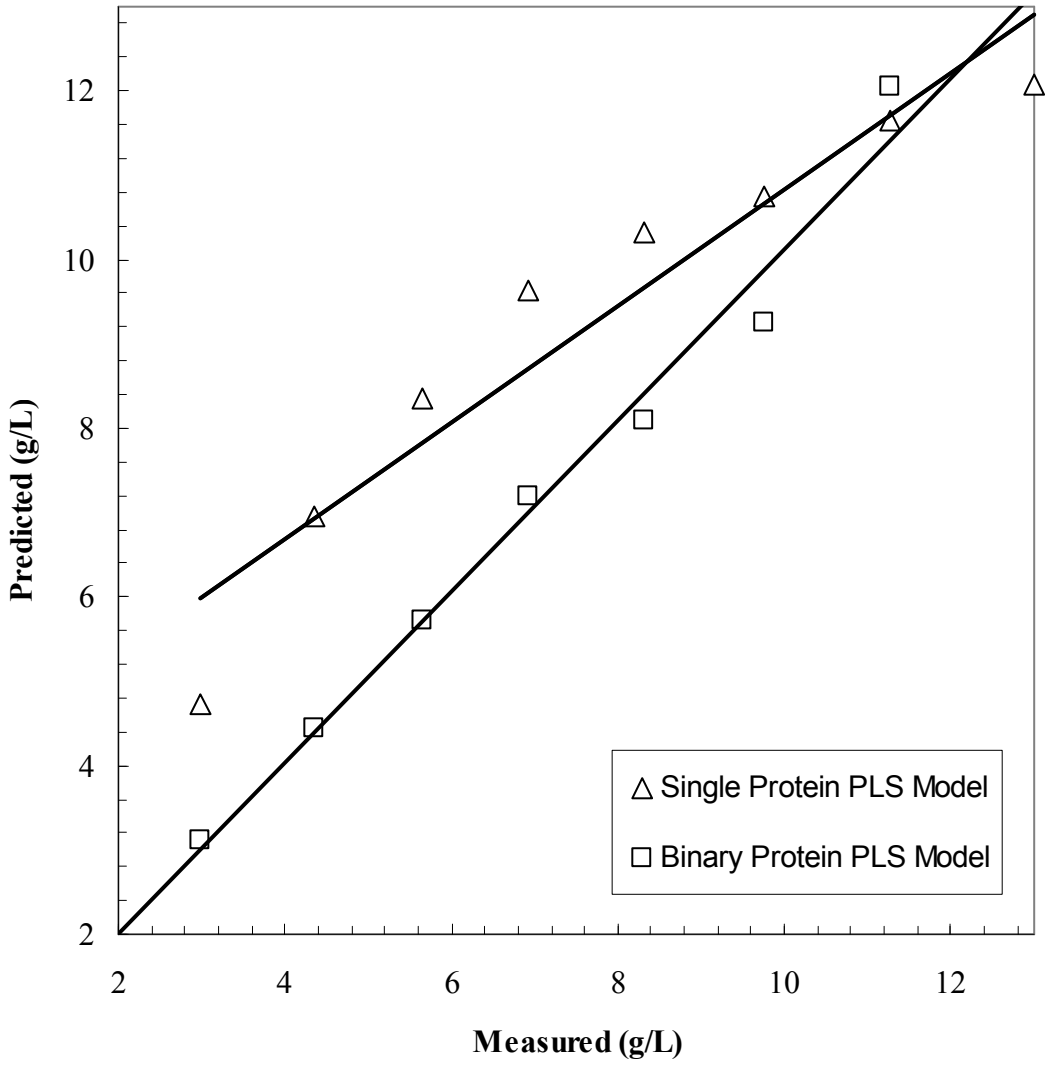


Figure C-1: Comparison between two fluorescence-based PLS models: (i) Single protein PLS model calibrated using β-LG protein solutions. (ii) Binary protein PLS model calibrated using protein mixture samples of α-LA and β-LG.

APPENDIX D: Analysis of Whey Protein Solutions Using HPLC and Fluorescence

Appendix D presents the raw data for the validation of the Binary Protein (α-LA/β-LG) PLS Model on whey protein isolate solutions. The PLS model was tested on validation set that contains whey protein concentrate solutions spiked with different amounts of α-LA and β-LG in their purified form.

Table D-1: Model prediction of β-LG and α-LA concentrations in the whey protein isolate solutions spiked with different amounts of α-LA and β-LG in their purified form.

	Predicted α-LA (g/l)	Predicted β-LG (g/l)	Actual α-LA (g/l)	Actual β-LG (g/l)
1	1.69788	6.0028	1.092	6.604
2	3.02519	5.5	2.11	6.604
3	3.58144	5.99584	3.32	6.604
4	6.20019	5.57	5.437	6.604
5	8.76169	6.14	8.502	6.604
6	6.60606	3.57	5.668	4.402
6	6.73272	6.18766	5.668	6.174
8	6.29626	7.62394	5.668	7.03
9	3.48377	9.06292	5.668	8.724
10	4.61995	7.52759	4.251	6.543
11	4.71095	8.03035	4.251	7.756
12	3.19899	5.44226	3.13	6.028

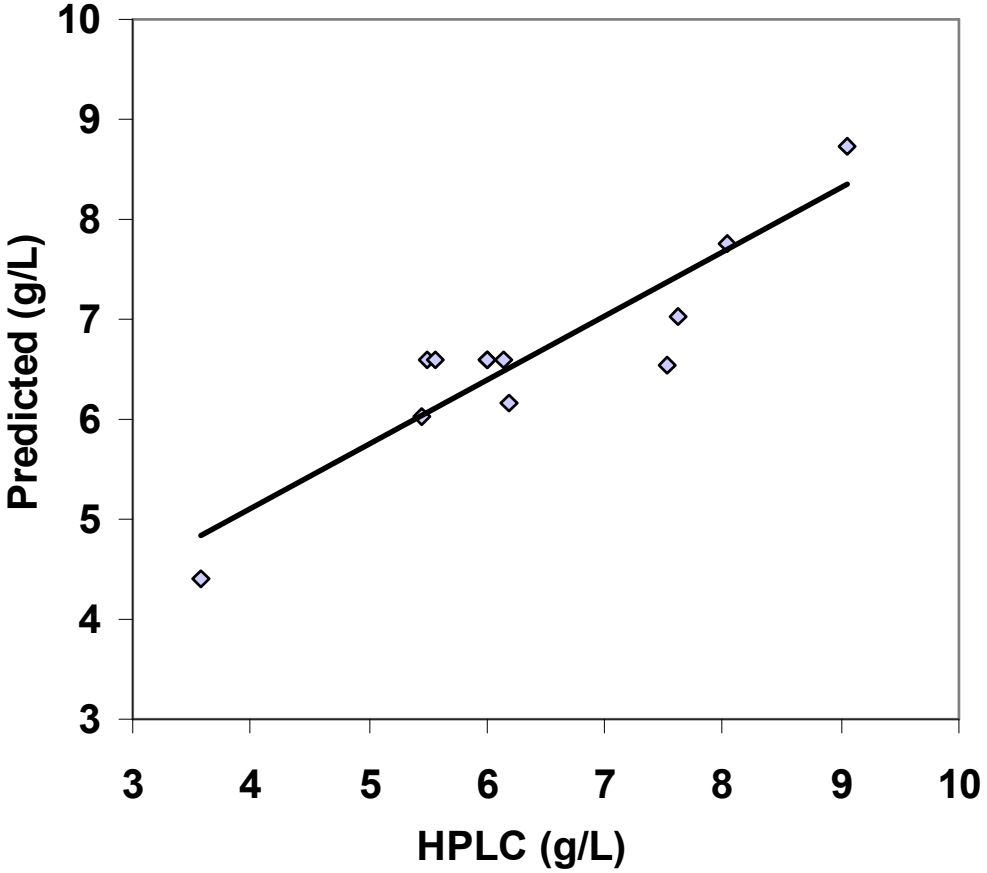


Figure D-1: Comparison between model prediction of β-LG and HPLC measurements for 12 different whey protein isolate solutions

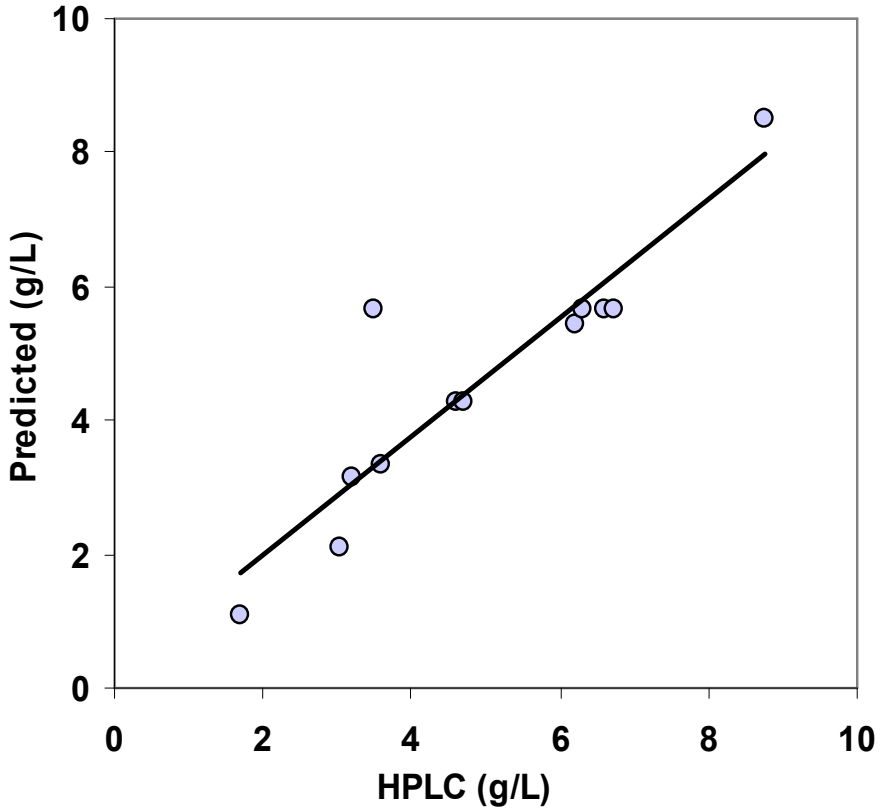


Figure D-2: Comparison between model prediction of α-LA and HPLC measurements for 12 different whey protein isolate solutions

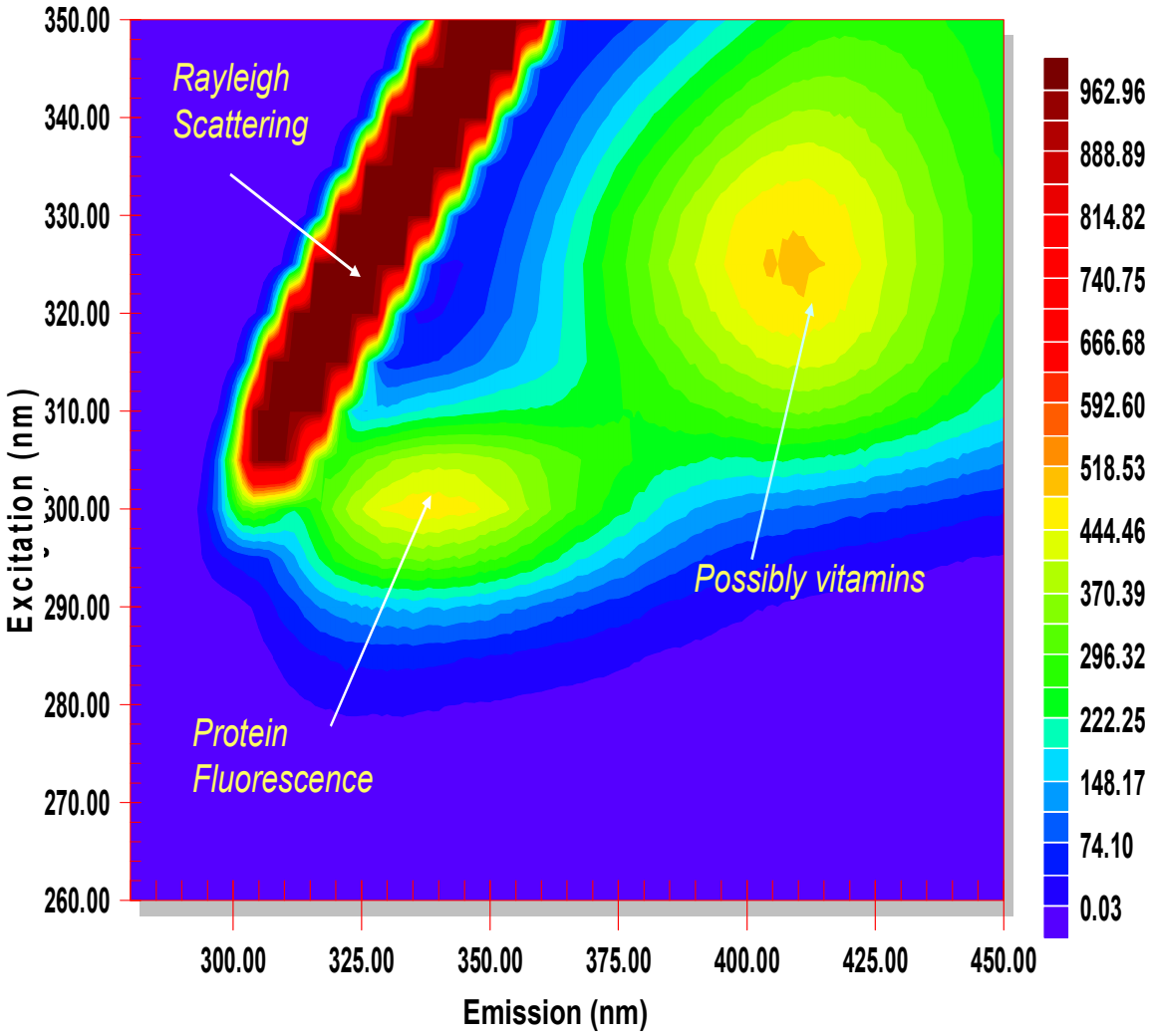


Figure D-3: Excitation-emission matrix for 6 g/l whey protein concentrate solution.

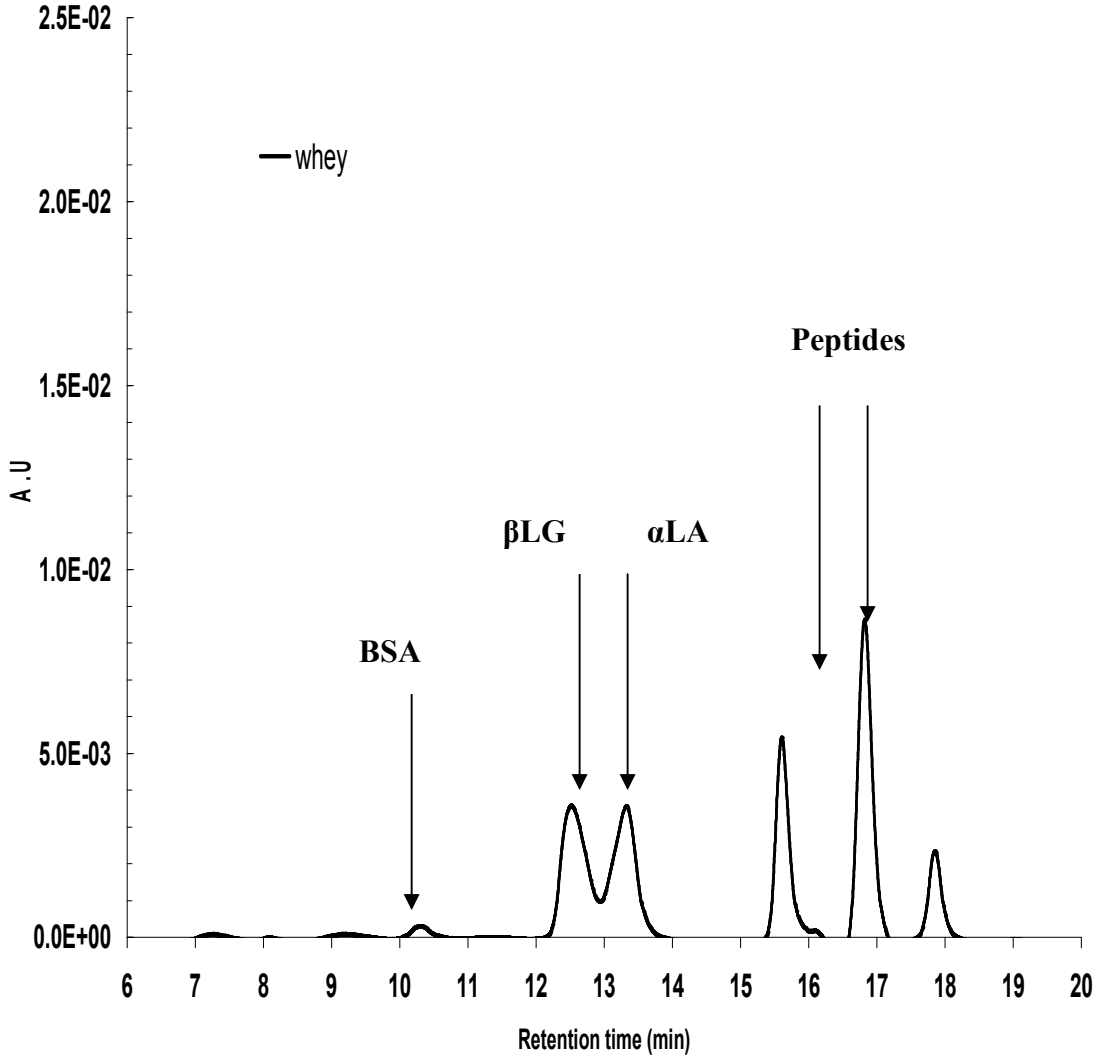


Figure D-4: HPLC Chromatogram for 2 g/l whey protein isolate solution.

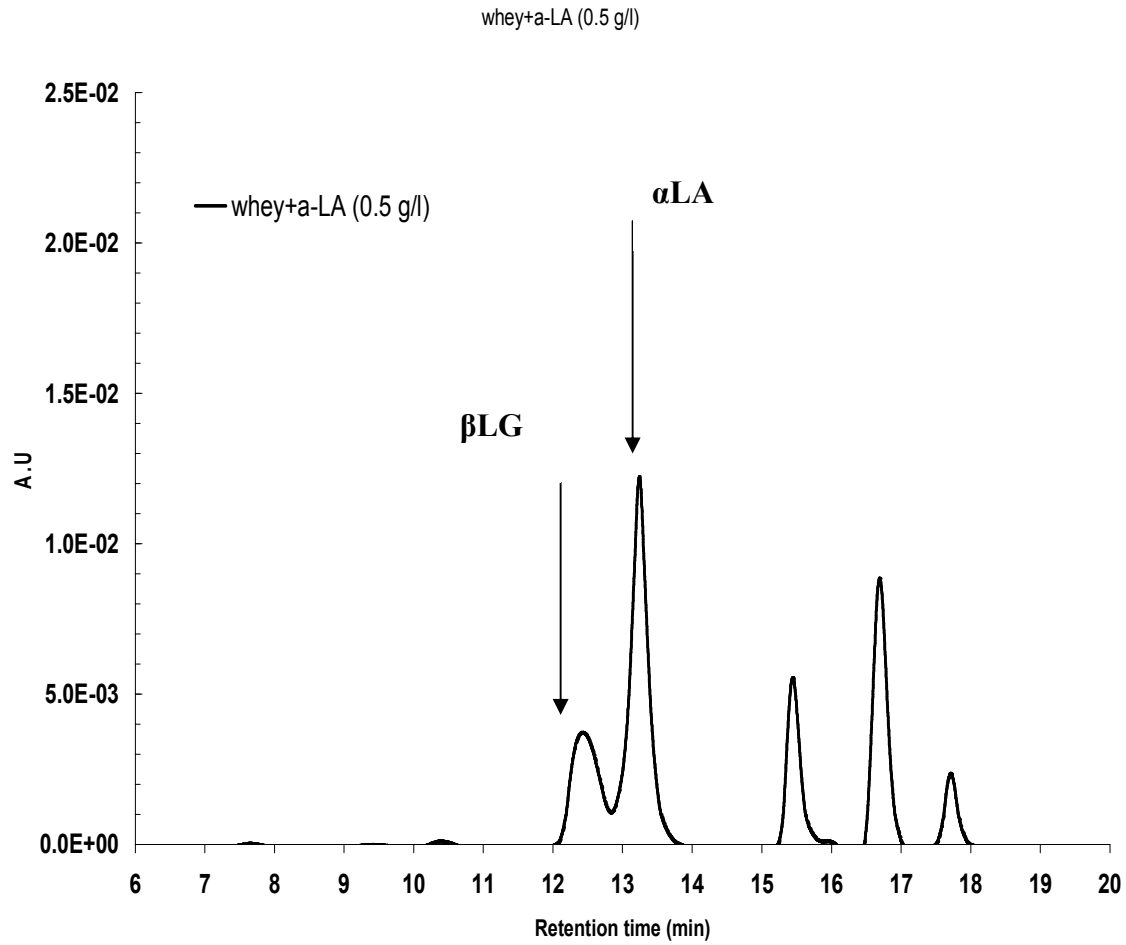


Figure D-5: HPLC Chromatogram for 2 g/l whey protein isolate solution spiked with 0.50 g/l α-LA.

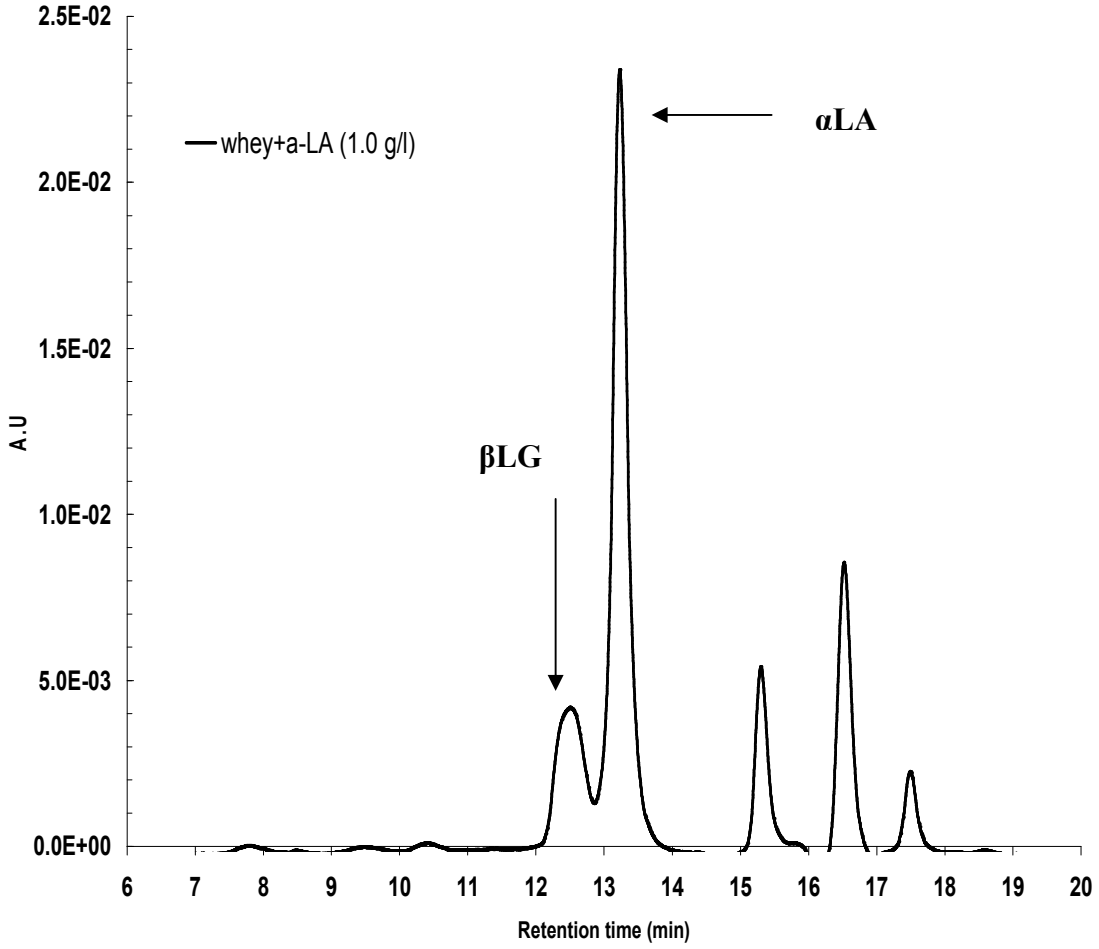


Figure D-6: HPLC Chromatogram for 2 g/l whey protein isolate solution spiked with 1 g/l α-LA**.**

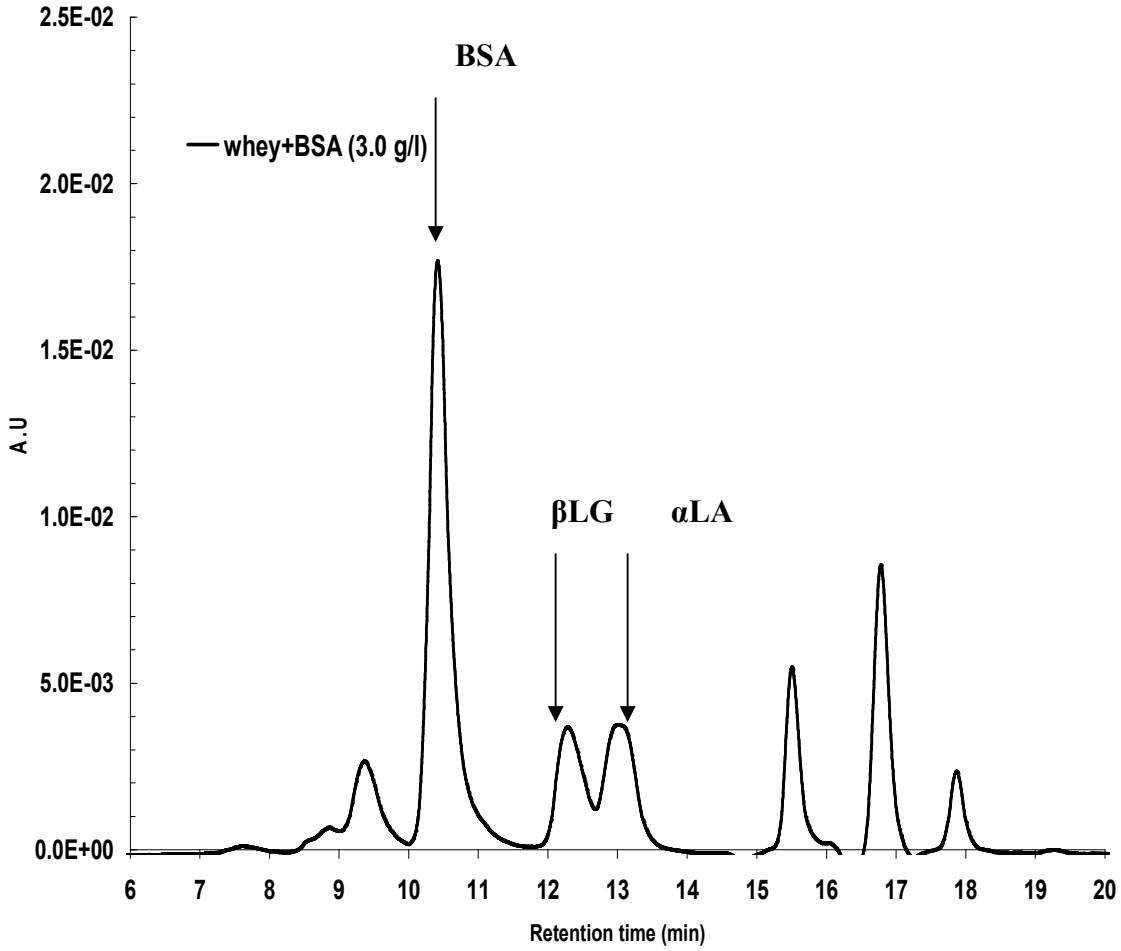


Figure D-7: HPLC Chromatogram for 2 g/l whey protein isolate solution spiked with 3 g/l BSA.

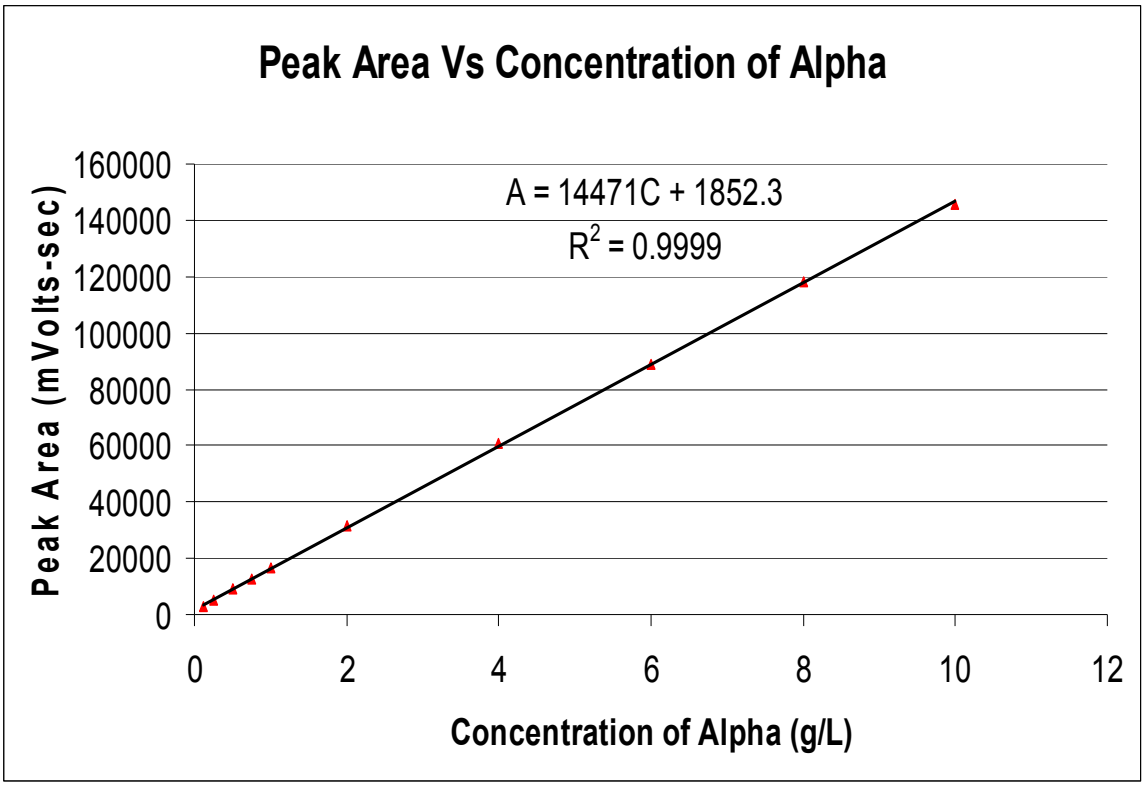


Figure D-8: Alpha-Lactalbumin's HPLC Peak Area Standards

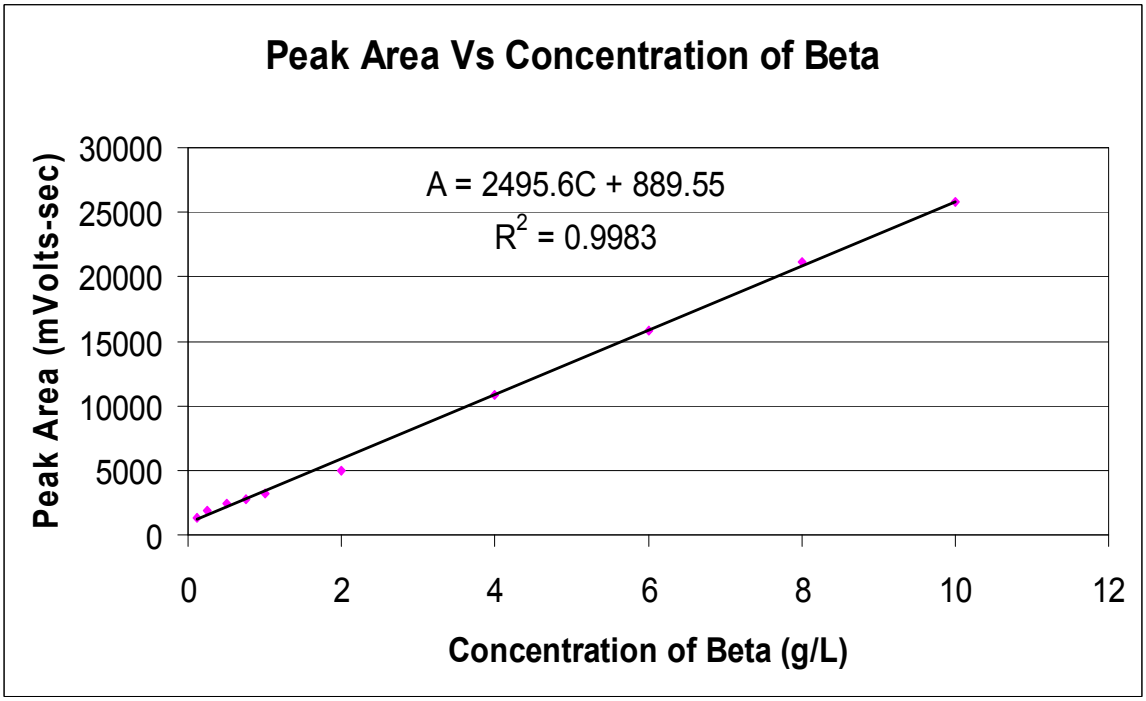
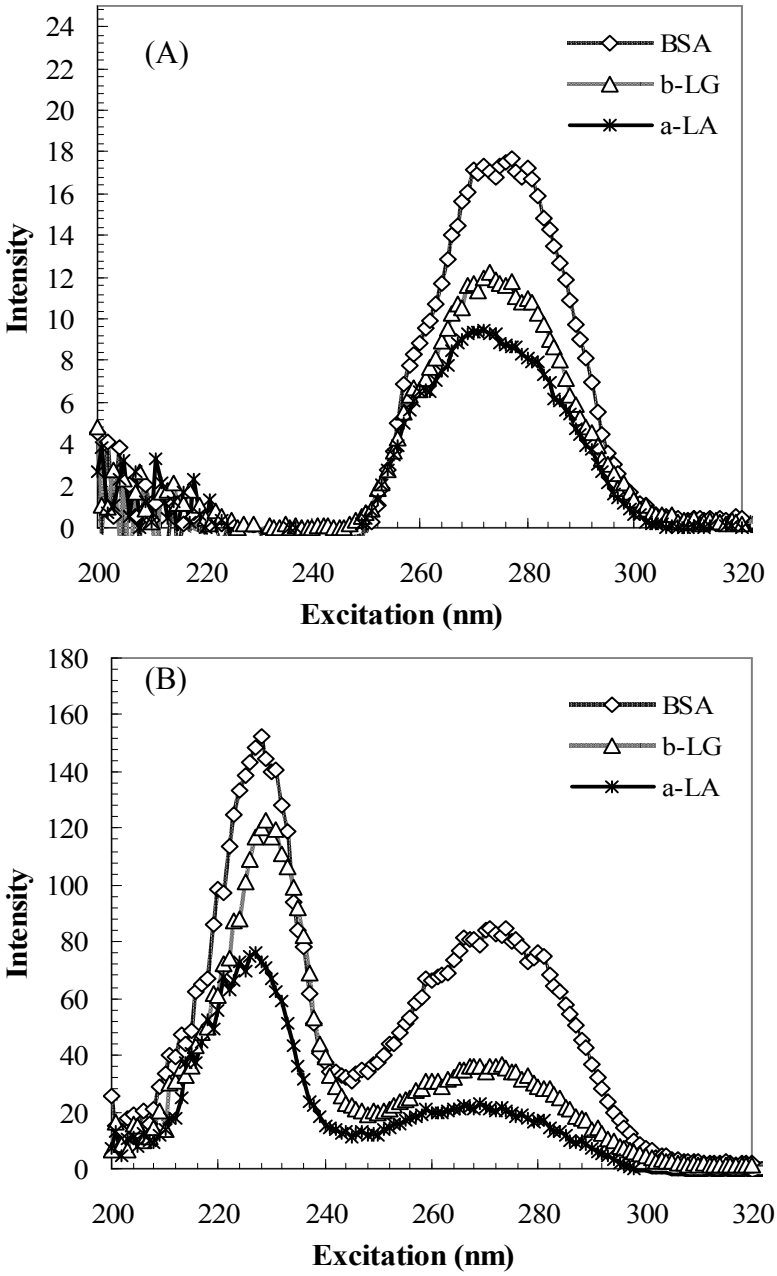


Figure D-9: Beta-Lactoglobulin's HPLC Peak Area Standards

APPENDIX E: Fluorescence measurements obtained from fiber optic probe

Protein fractionation using ultrafiltration is a membrane based separation process commonly used in the biotech, food and beverage industries. Protein mixtures can be fractionated based on their size where protein species larger than the membrane pores are retained by the membrane (i.e. in the retentate stream) while protein species smaller than the membrane pores passes through the membrane (permeate stream). During ultrafiltration of protein mixtures, changes in protein concentrations in the permeate and the retentate have been observed with time. Appendix E shows the development of a fluorescence based sensor for monitoring the changes in concentration of proteins in both the permeate and the retentate by making use of fluorescence spectroscopy and multivariate methods. A preliminary three-protein component system consisting of α-lactalbumin (α-LA), β-lactoglobulin (β-LG) and Bovine Serum Albumin (BSA) was used as a model system in this study. A fiber optic probe was used to acquire multi-wavelength fluorescence spectra of the permeate and the retentate at different times during ultrafiltration of α-LA and β-LG protein solutions. Multivariate models were developed for predicting the concentrations of α-LA and β-LG in both the permeate and the retentate by establishing a calibration model between fluorescence data acquired by the fiber optic probe and α-LA and β-LG concentrations measured by size-exclusion chromatography. Figure E-1 shows the difference between the spectra collected using the fiber optic probe and cuvette-based method. Such difference is most likely due to the fact that collection of the spectra using the fiber optic probe is based on front-face geometry, while collection of the spectra using cuvettes is based on right-angle geometry. Figures E-2, E-3, E-4, E-5 and E-6 show the plot of regression coefficients of α-LA and β-LG versus excitation wavelengths at different $\Delta\lambda$ values. The model was validated on fiber optic fluorescence data that were not used for the calibration. Results in Table E-1 show that concentrations of α-LA and β-LG can be predicted directly from fluorescence data acquired by the fiber optic probe within a reasonable accuracy by making use of multivariate calibration tools.



Figures E-1: Synchronous fluorescence spectra at Δλ=100 nm for BSA, α-LA and β-LG (a) plastic cuvettes and (b) fiber optic probe

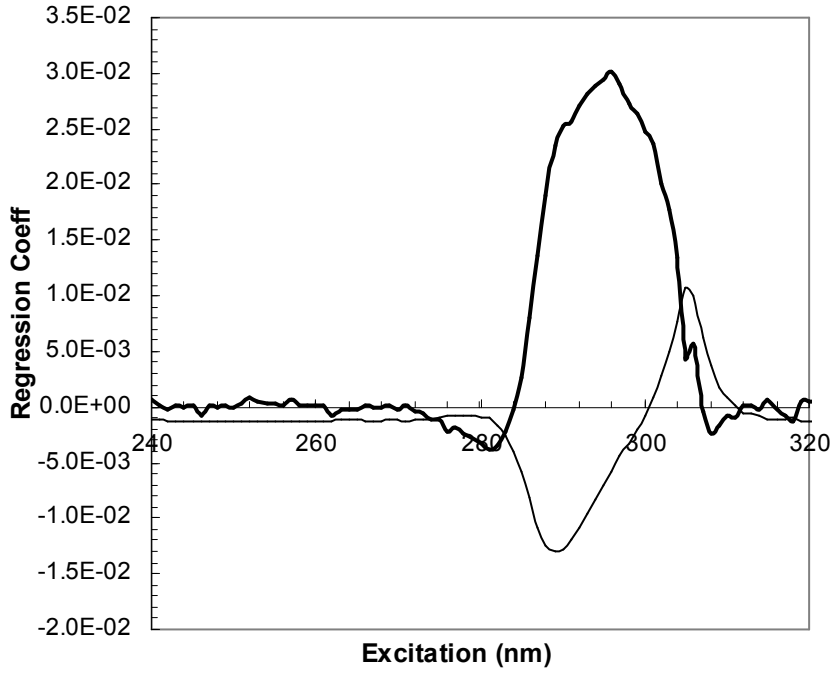


Figure E-2: PLS regression coefficients for α-LA (heavy line) and β-LG (light line) versus excitation wavelength at Δλ=10 nm.

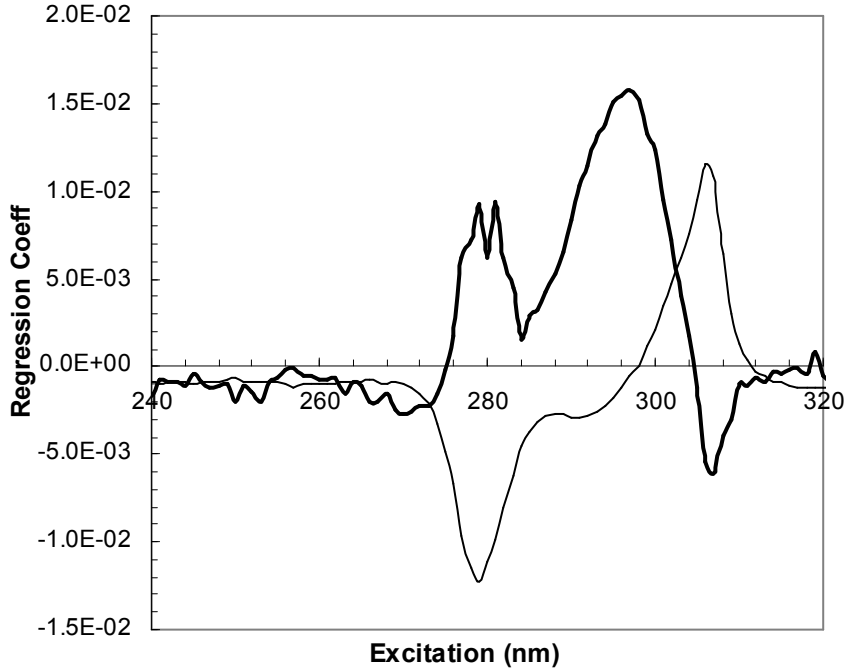


Figure E-3: PLS regression coefficients for α-LA (heavy line) and β-LG (light line) versus excitation wavelength at Δλ=20 nm.

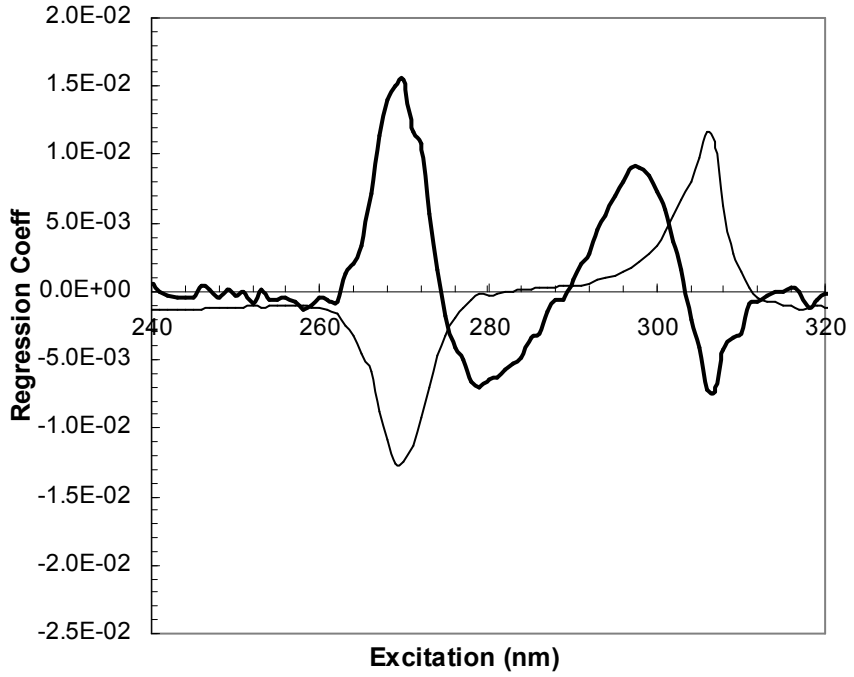


Figure E-4: PLS regression coefficients for α-LA (heavy line) and β-LG (light line) versus excitation wavelength at Δλ=30 nm.

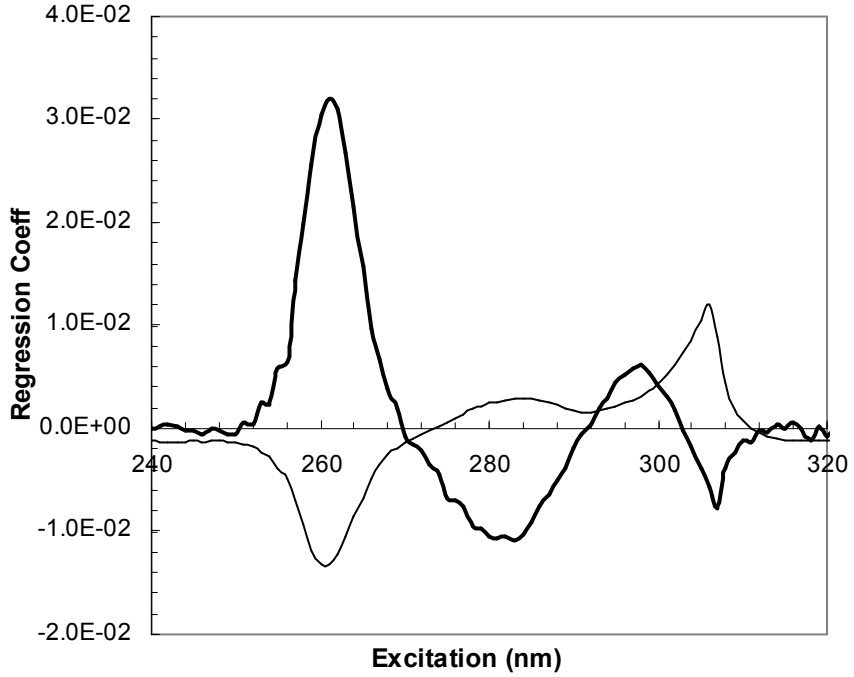


Figure E-5: PLS regression coefficients for α-LA (heavy line) and β-LG (light line) versus excitation wavelength at Δλ=40 nm.

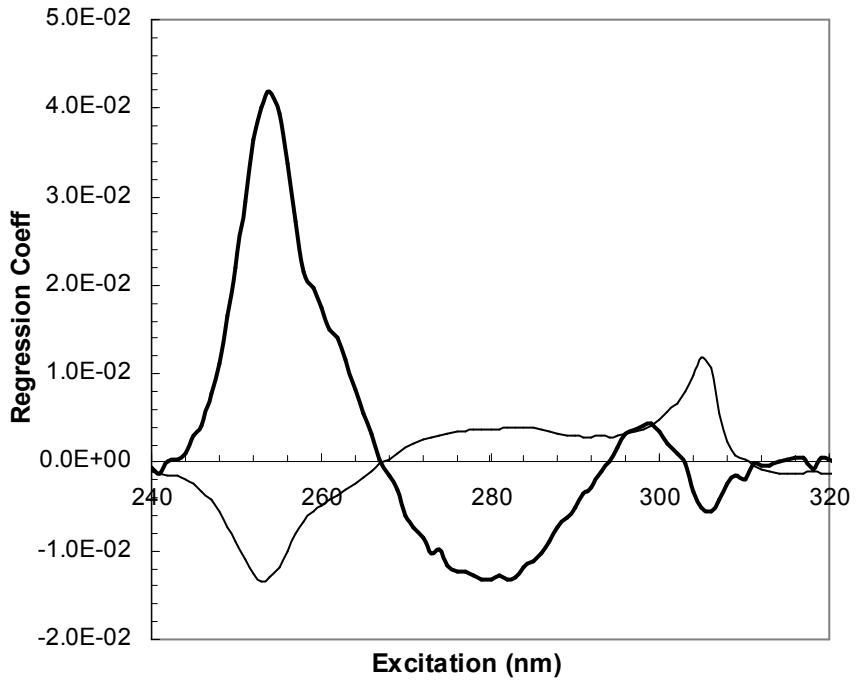


Figure E-6: PLS regression coefficients for α-LA (heavy line) and β-LG (light line) versus excitation wavelength at Δλ=50 nm.

Table E-1: PLS model validation on 17 protein mixtures of β-LG, α-LA and BSA that were not used for the calibration

	Measured (g/L)			***Predicted (g/L)***		
	β-LG	α-LA	BSA	β-LG	α-LA	BSA
pH 7	0.5	0	0	0.417	-0.007	0.001
pH 7	0.5	0	0	0.437	-0.024	0.003
pH 7	0	0.5	0	0.003	0.554	-0.004
pH 7	0	0.2	0	0.007	0.172	0.01
pH 7	0.1	0	0	0.107	-0.006	0.005
pH 5.6	0.013	0	0.013	0.014	-0.001	0.016
pH 5.6	0.025	0	0.025	0.025	0	0.027
pH 5.6	0.05	0	0.05	0.047	0.002	0.043
pH 5.6	0.075	0	0.075	0.068	0.004	0.058
pH 5.6	0.1	0	0.1	0.088	0.007	0.071
pH 5.6	0.125	0	0.125	0.109	0.008	0.08
pH 5.6	0	0.01	0.01	-0.001	0.012	0.014
pH 5.6	0	0.025	0.025	-0.002	0.028	0.025
pH 5.6	0	0.05	0.05	-0.001	0.054	0.043
pH 5.6	0	0.075	0.075	0.001	0.078	0.052
pH 5.6	0	0.1	0.1	-0.002	0.101	0.068
pH 5.6	0	0.125	0.125	-0.03	0.148	0.095

APPENDIX F: Membrane Filtration Setup

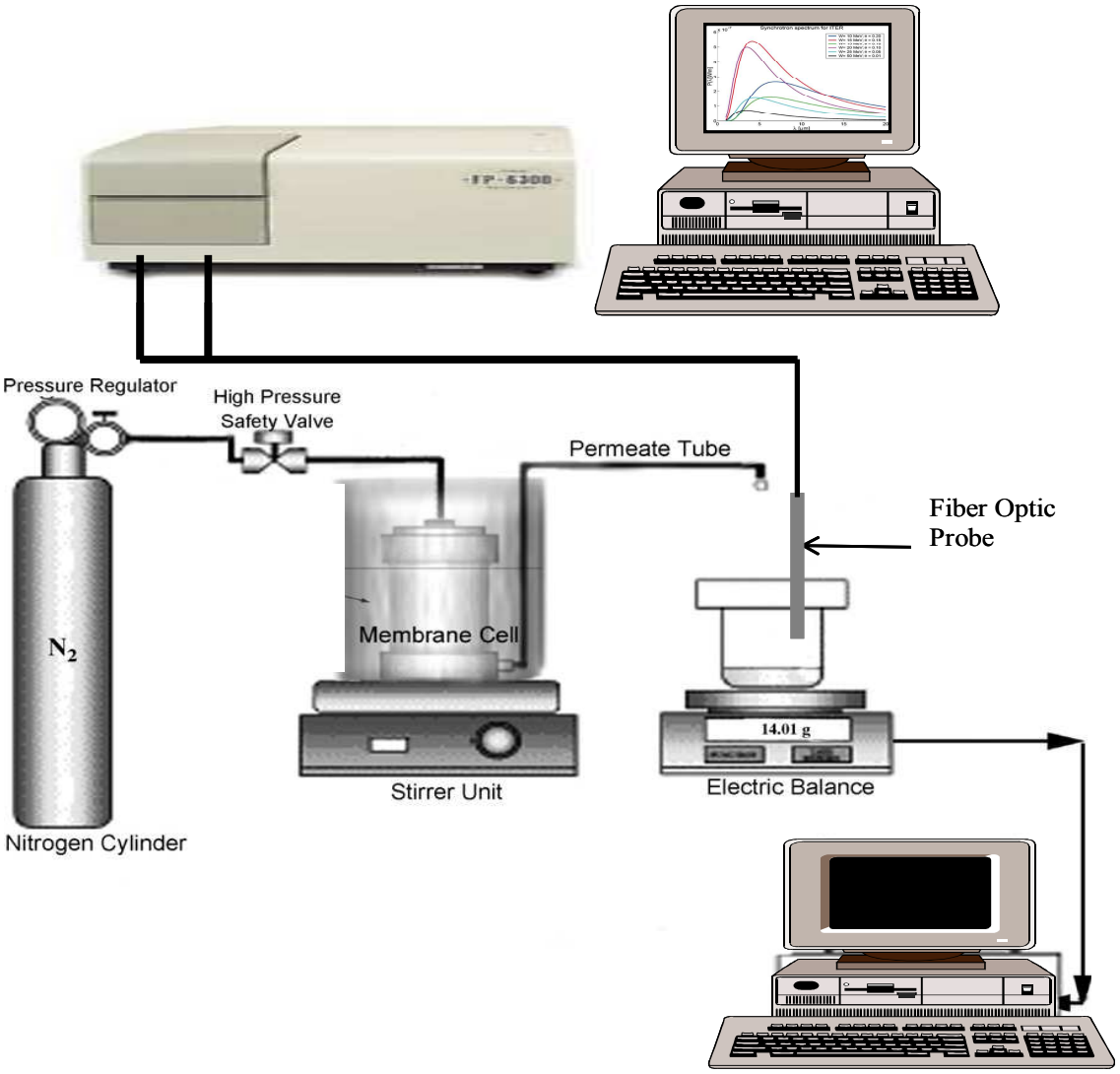


Figure F-1: Schematic diagram for the experimental membrane filtration setup and the fiber optic probe assembly.

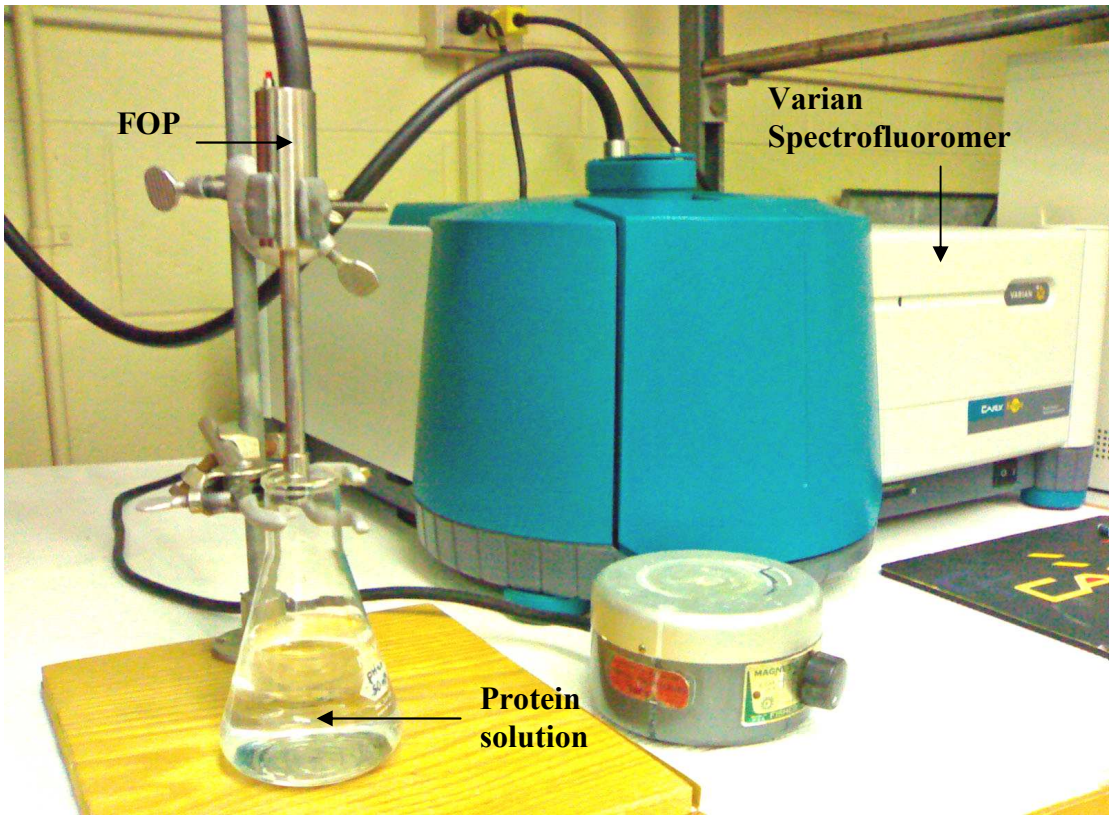


Figure F-2: Picture of the fiber-optic probe (FOP) connected to Varian Cary Eclipse spectrofuorometer and inserted into the a flask containing the protein solution.

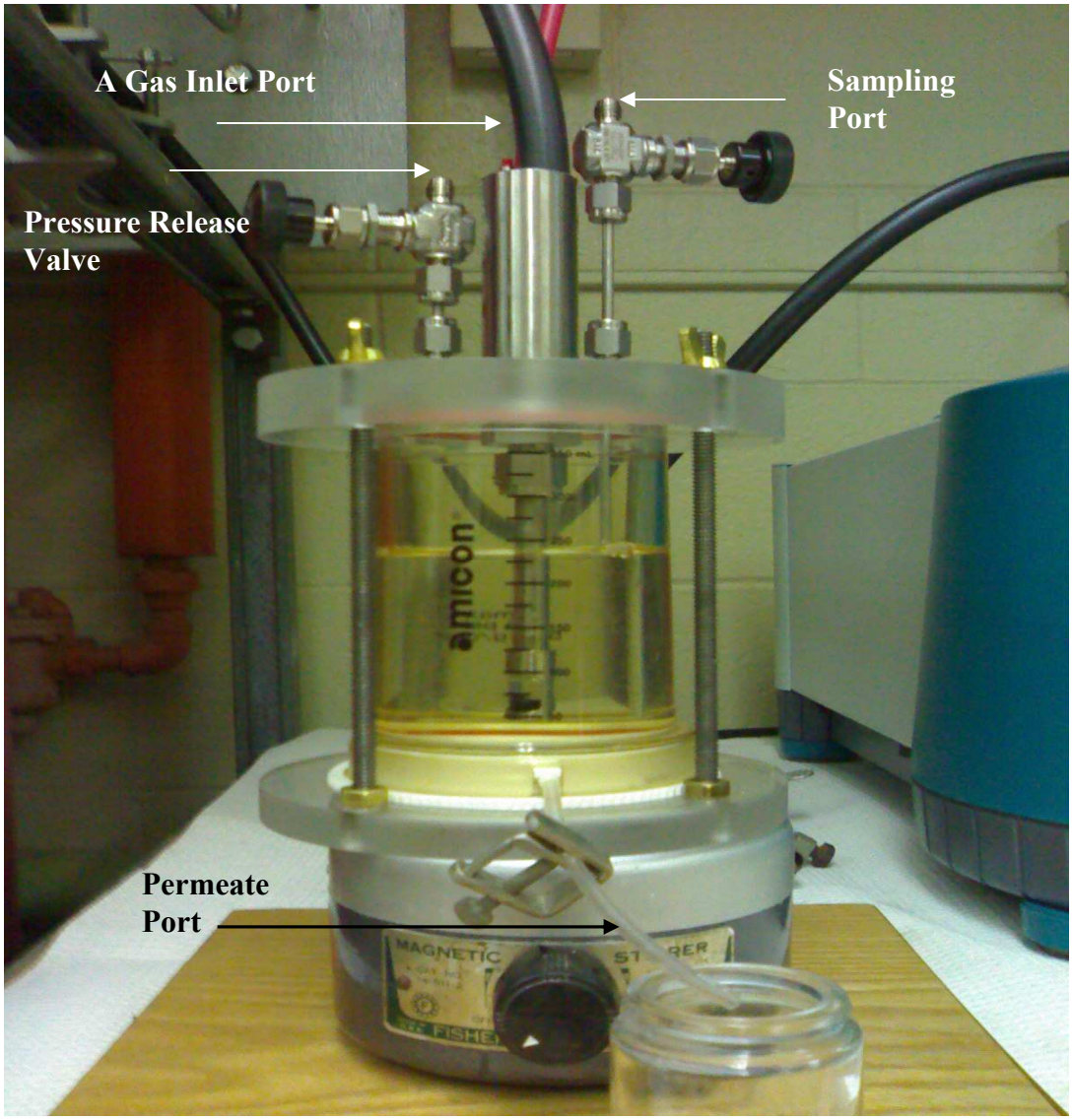


Figure F-3: Picture of the Amicon filtration stirred cell (Model 8010, Amicon Corp, Beverly, MA).

APPENDIX G: The protein Association Behavior

Table G-1: Tabulated data for the average molecular weight and radius of gyration reported in Verhuel et al., 2003 and FOP light scattering measurements for protein solutions prepared at different conditions

pH	NaCl (mM)	β-LG (g/L)	FOP Light scattering	Mw (kDa)	Rg (A)
2.0	100	2	6.37	28.00	19.9
2.0	100	5	12.06	28.37	20.3
2.0	100	10	21.96	28.55	20.1
4.7	0	2	40.64	32.76	22.6
4.7	0	5	109.88	45.94	27.9
4.7	0	10	146.25	61.86	34.5
4.7	100	2	43.32	36.61	24.8
4.7	100	5	83.49	37.89	24.8
4.7	100	10	100.68	43.93	29.3
6.9	0	2	8.04	29.47	22.1
6.9	0	5	14.40	32.58	23.2
6.9	0	10	25.74	34.04	23.6
6.9	100	2	9.38	31.12	22.6
6.9	100	5	19.74	33.49	23.2
6.9	100	10	26.02	32.95	22.6
8.0	0	2	6.48	23.43	19
8.0	0	5	13.54	24.53	20.5
8.0	0	10	23.45	25.44	22.1
8.0	100	2	7.82	26.91	21.4
8.0	100	5	14.17	28.55	22
8.0	100	10	24.89	31.48	22.3
5.4		5	44.93	41.18	30.2

Table G-2: Tabulated data for the degree of Lysozyme self-association reported in (Seth and Aswal, 2002) and FOP light scattering measurements for protein solutions prepared at different conditions pHs and NaCl concentrations

pH	NaCl (mM)	β-LG (g/L)	FOP Light scattering	Percentage of Dimers
4.6	0	2	51	0
6.2	0	5		19
7.5	0	10	132	50
8.5	0	2		55
10.5	0	5	162	65
4.6	0.5	10		11
4.6	1	2	61	20
4.6	1.5	5		27
4.6	2	10	81	31

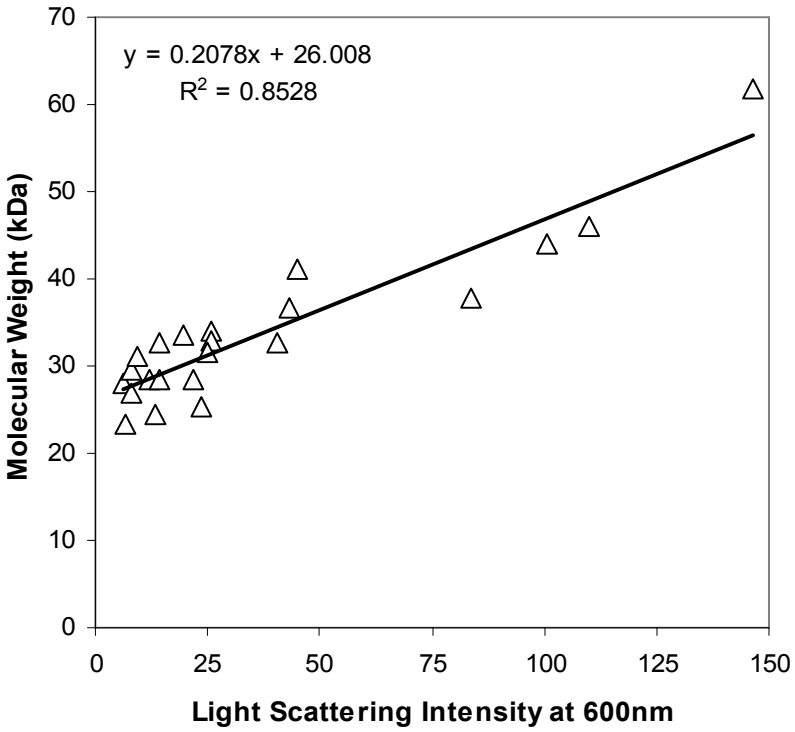


Figure G-1: The correlation between light scattering measurements at λ= 600 nm acquired using the fiber optic probe for β-LG protein solutions prepared at different conditions (pHs, protein concentrations and salt concentrations) and the average molecular weight for those solutions reported in Verheul et al.,2003.

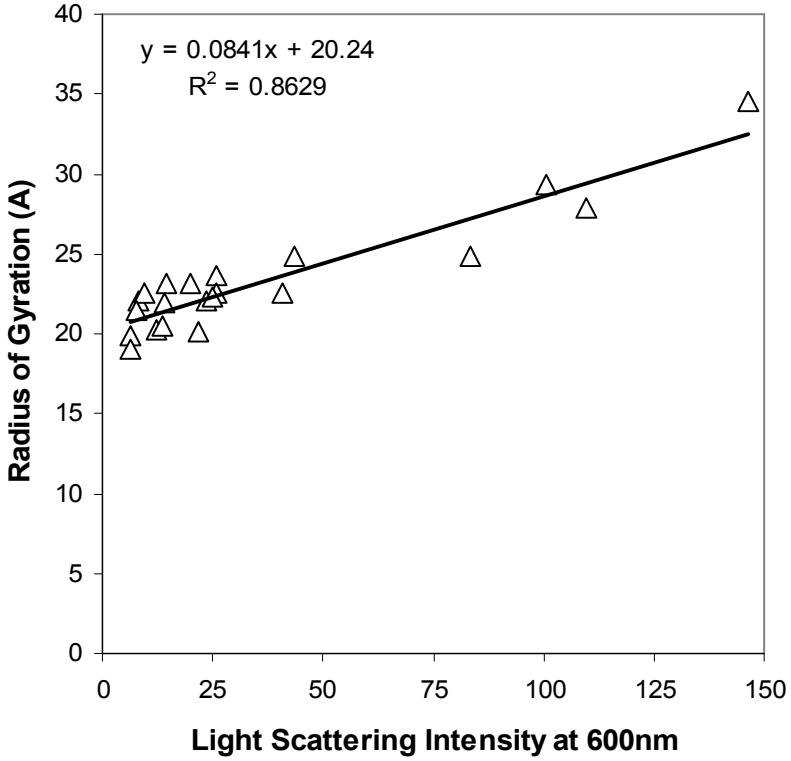


Figure G-2: The correlation between light scattering measurement at λ= 600 nm acquired using the fiber optic probe for β-LG protein solutions prepared at different conditions (pHs, protein concentrations and salt concentrations) and the average radius of gyration for those solutions reported in Verheul et al.,2003

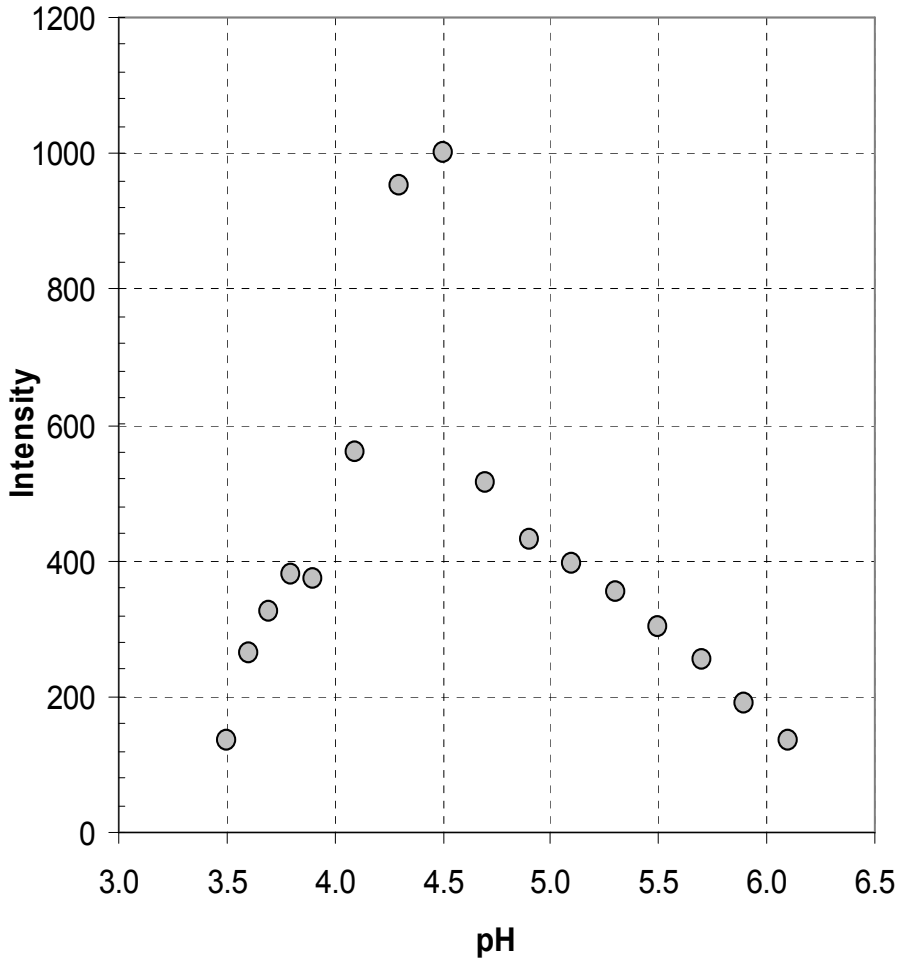


Figure G-3: The effect of pH on Rayleigh scattering intensity measured at excitation wavelength of 400 nm for whey protein isolate solution.

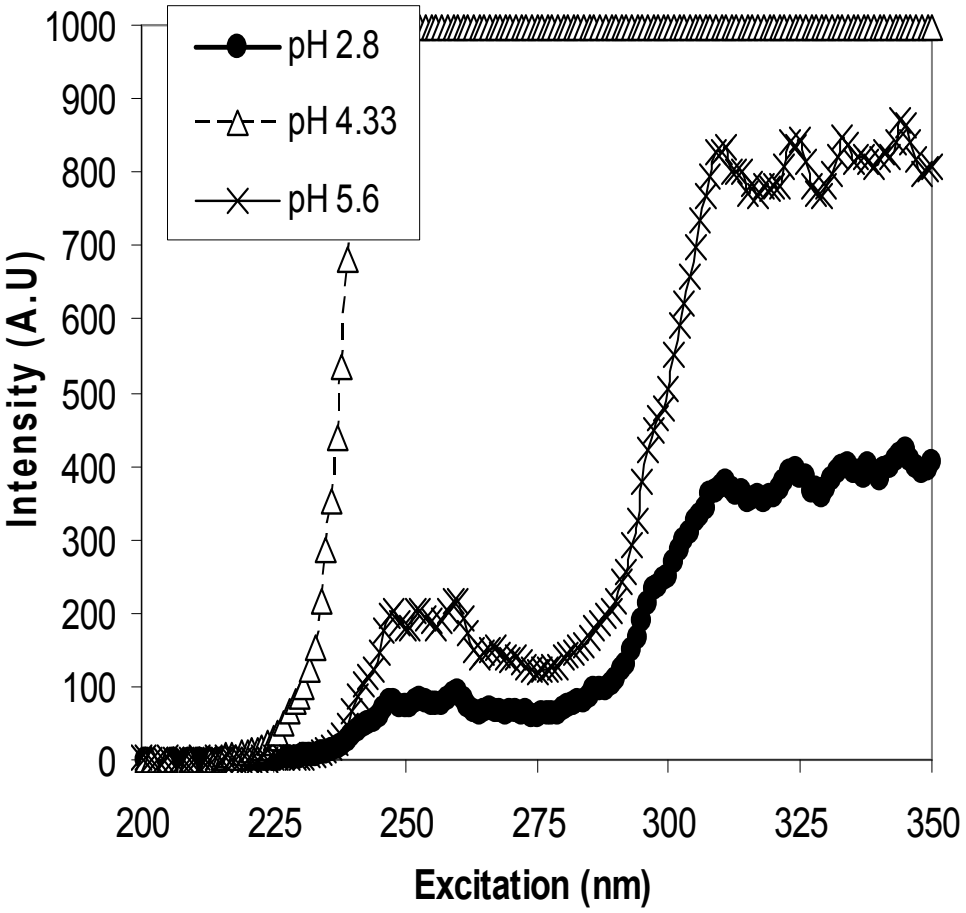


Figure G-4: Rayleigh scattering of Bipro's whey protein isolate solution at three different pHs (short wavelength range).

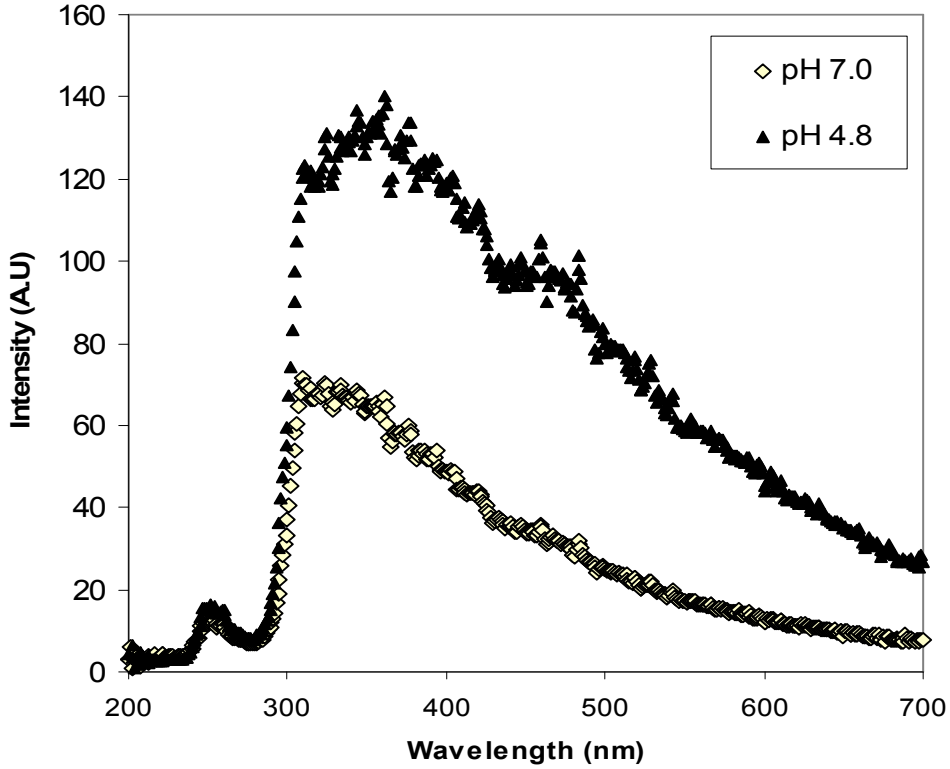


Figure G-5: Rayleigh scattering of protein mixture of 50% BSA and 50% β-LG at two different pHs.

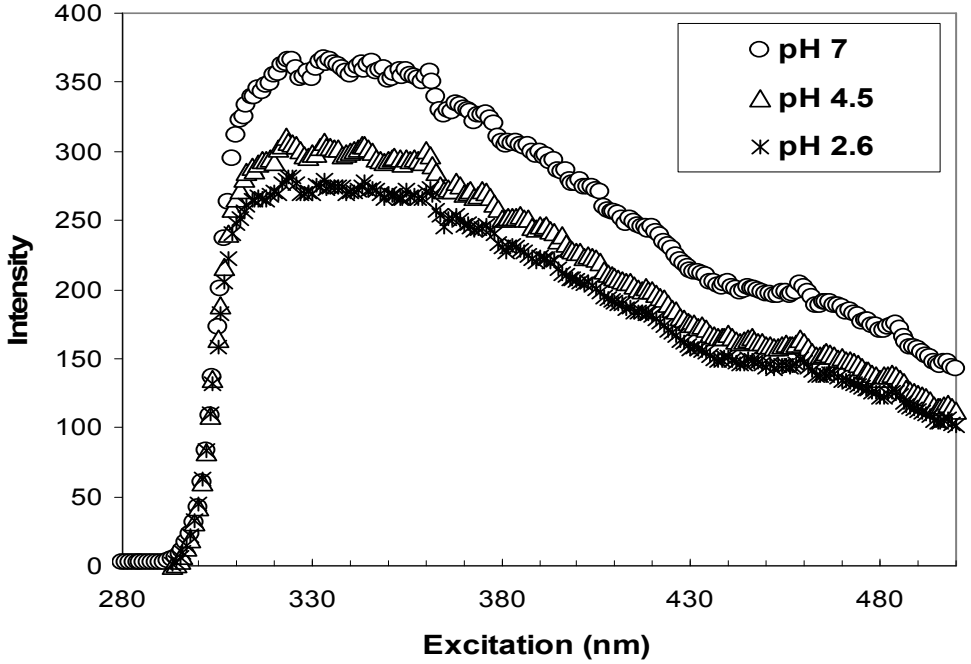


Figure G-6: Rayleigh scattering of protein mixture of 50% BSA and 50% LYS at three different pHs (short wavelnrth range).

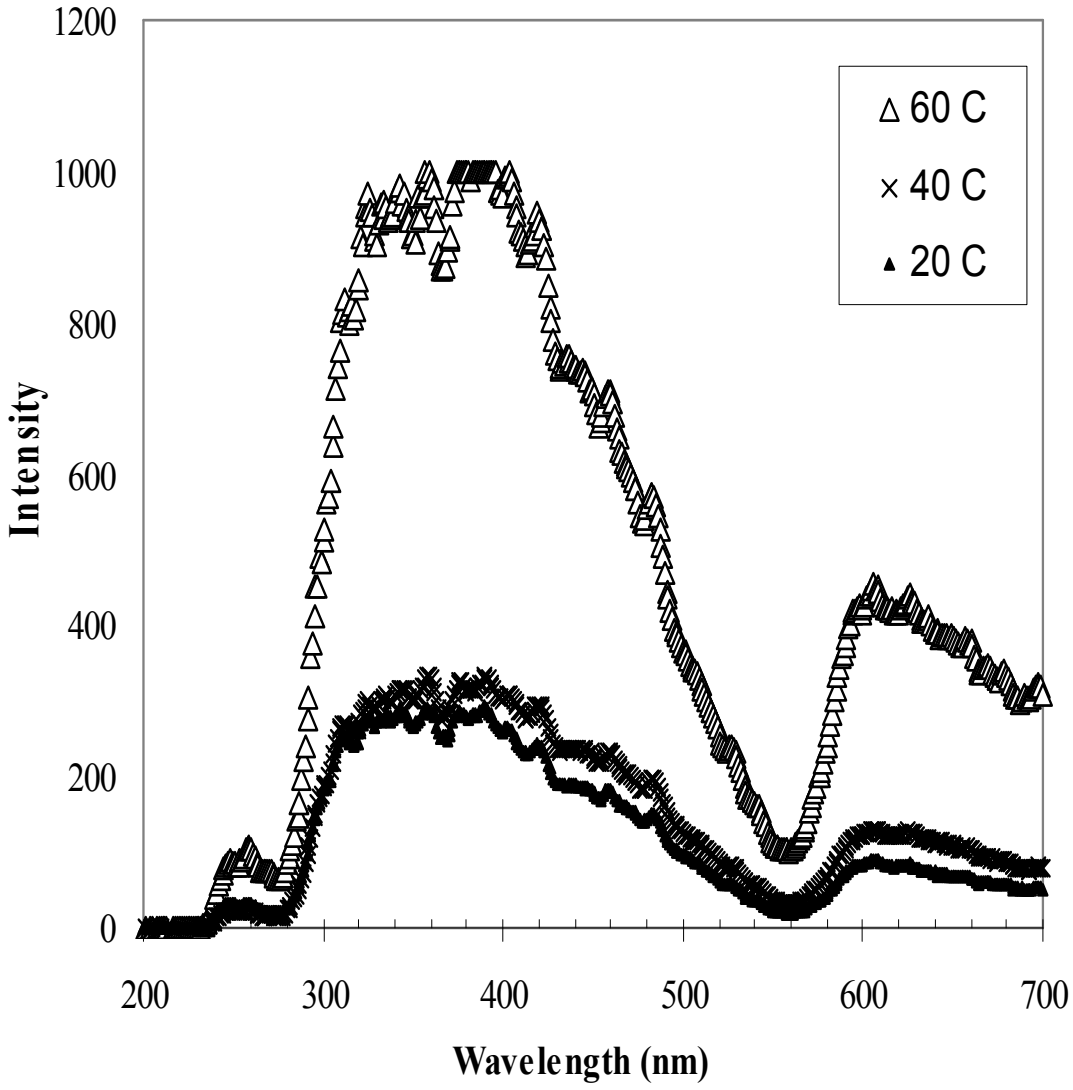


Figure G-7: Rayleigh scattering profiles for three samples of BSA in culture medium that were subjected to heat treatment at 20, 40 and 60 C.

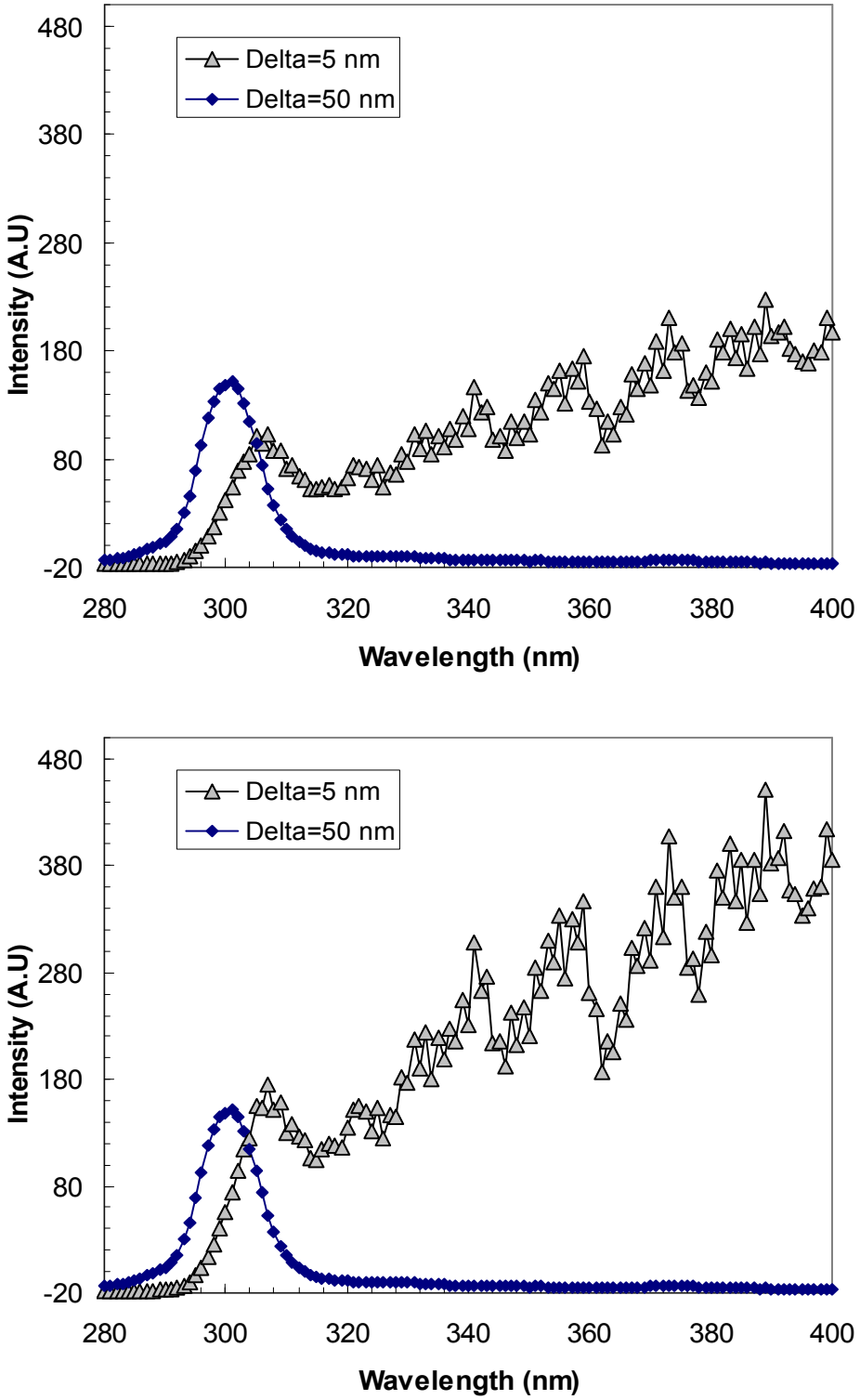


Figure G-8: Effect of salt concentration on electrostatic interactions of β-LG at pH=3.0: (top) 4 g/L β-LG, 0 mM salt, pH 3.01 (bottom) 4 g/L β-LG, 216 mM salt, pH 3.01.

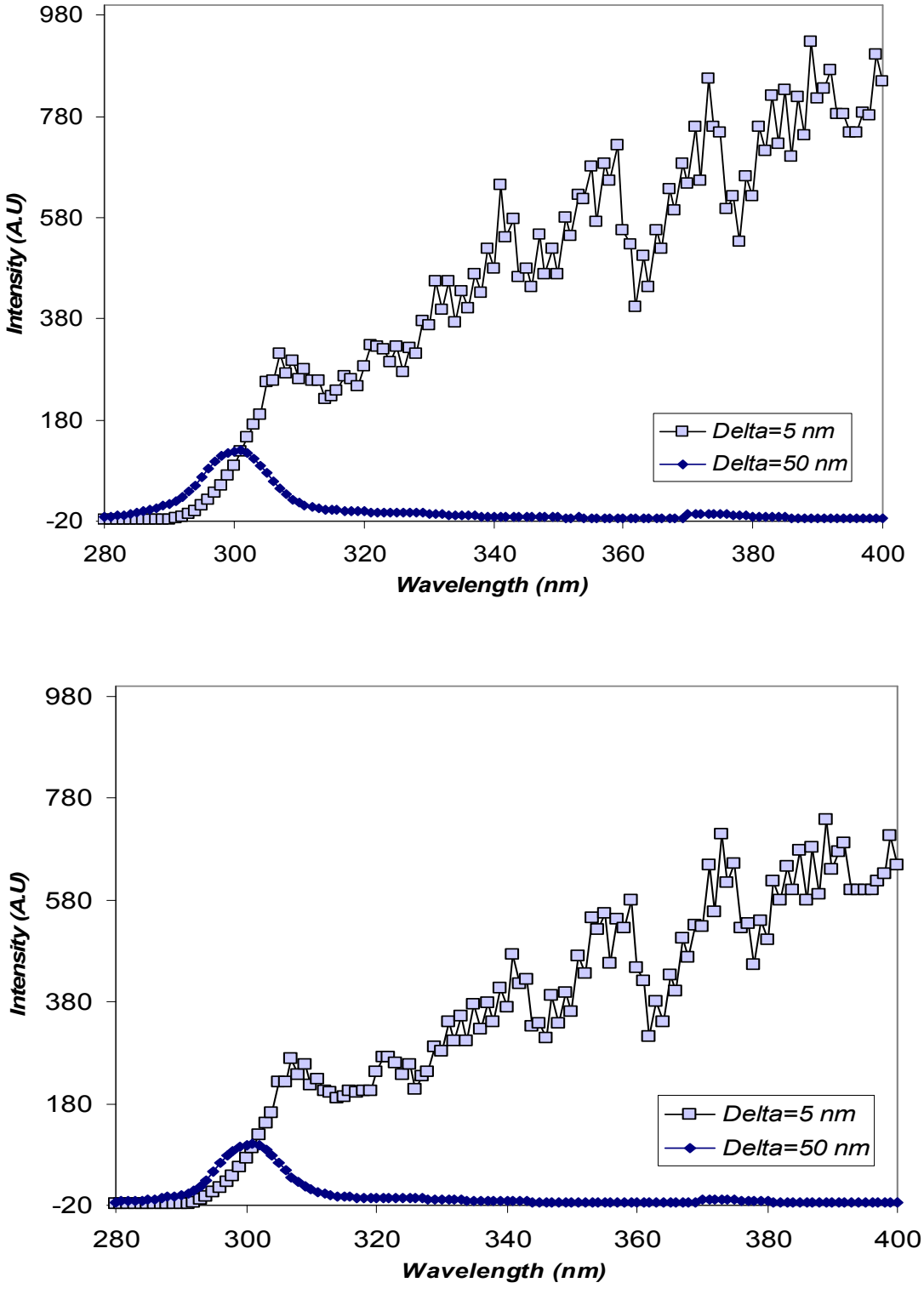


Figure G-9: Effect of salt concentration on electrostatic interactions of β-LG at pH=5.2: (top) 4 g/L β-LG, 0 mM salt, pH 5.2 (bottom) 4 g/L β-LG, 300 mM salt, pH 5.2.

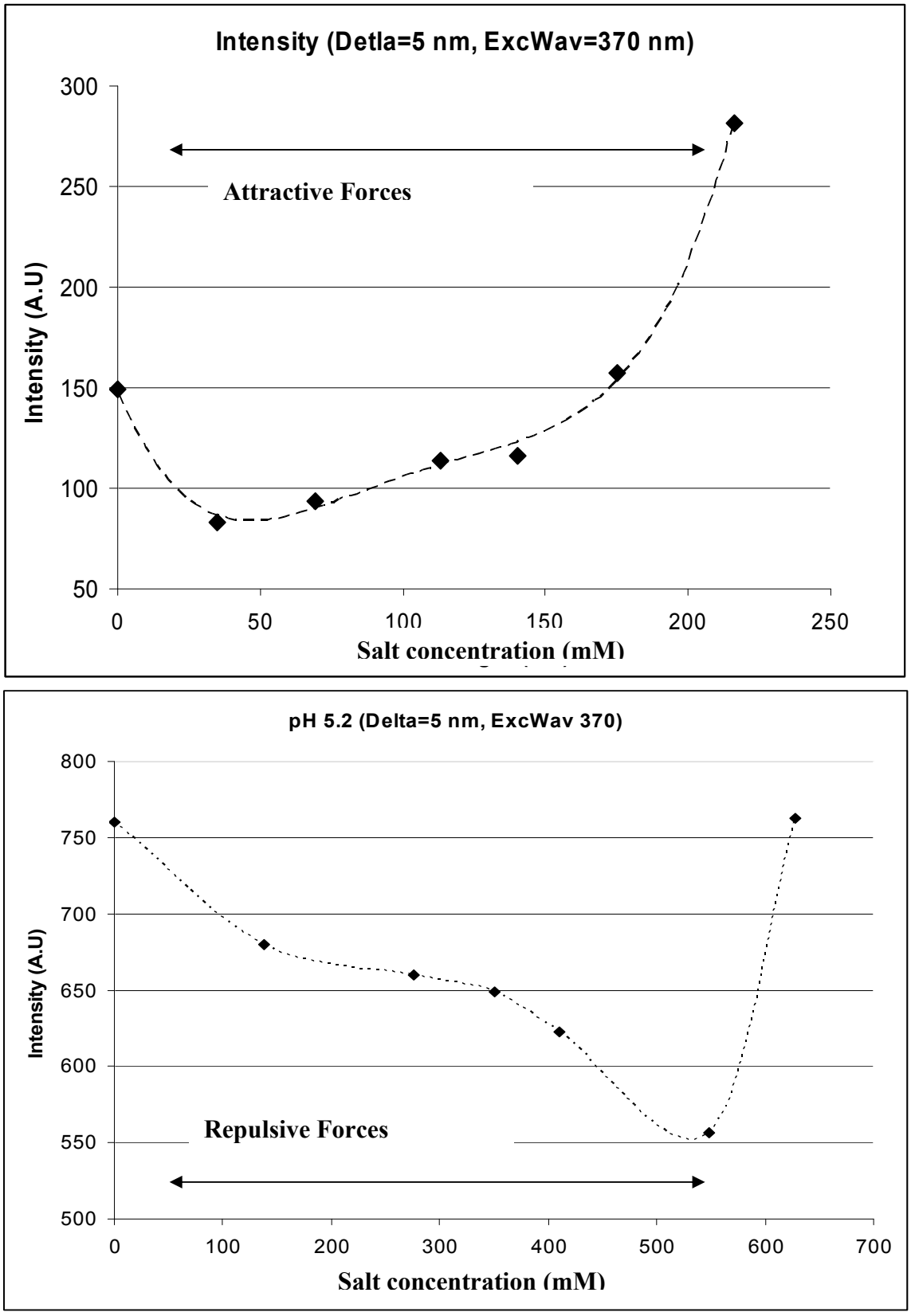


Figure G-10: Effect of salt concentration on Rayleigh scattering intensity at excitation 370 nm for 4 g/l β-LG protein solution at pH 3.0 (top) and pH 5.2 (bottom).

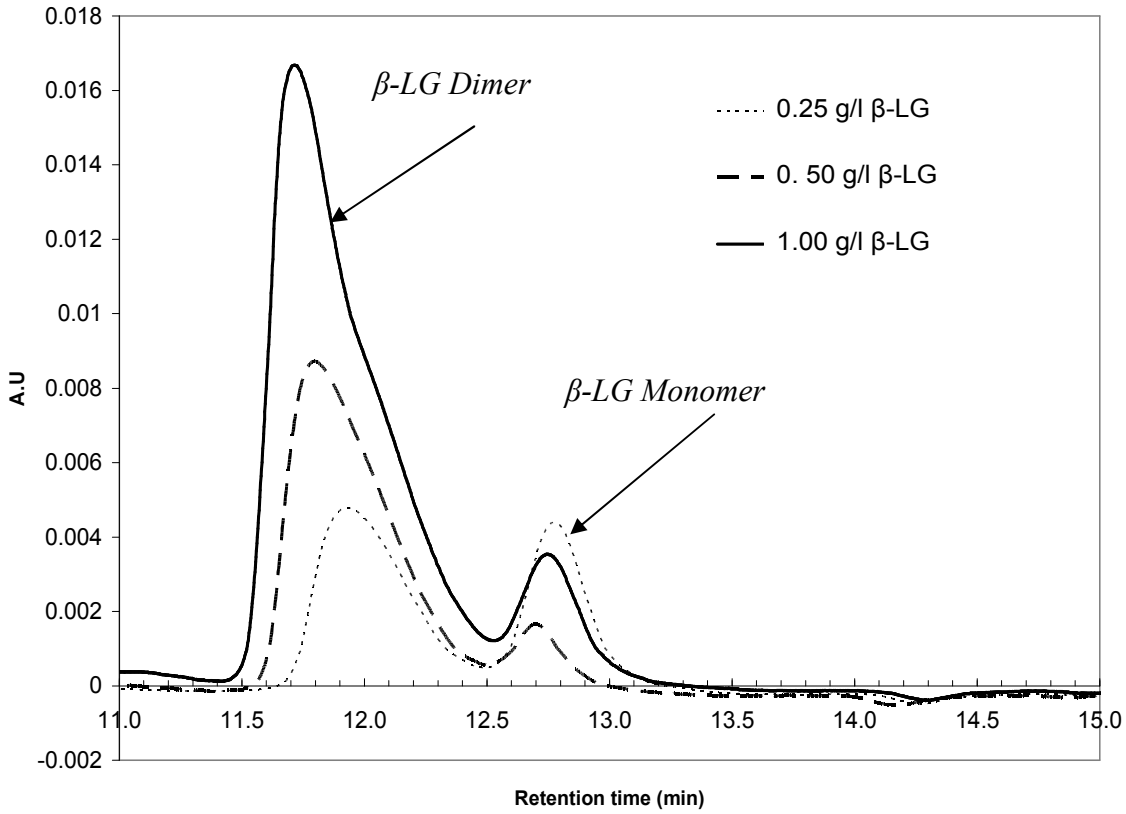


Figure G-11: HPLC Chromatogram for three beta-lactoglobulin protein solutions at different concentrations: higher dimmer fraction ocuurs at higher protein concentration.

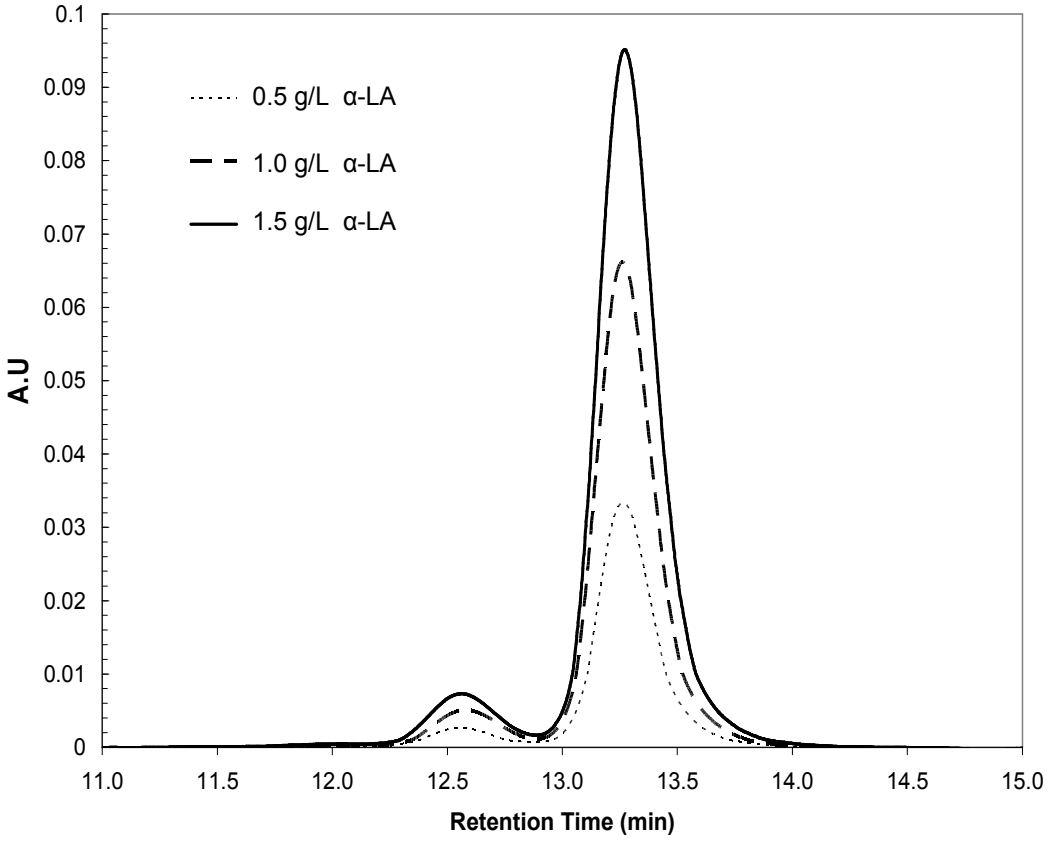


Figure G-12: HPLC Chromatogram for three alpha-lactalbumin protein solutions at different concentrations.

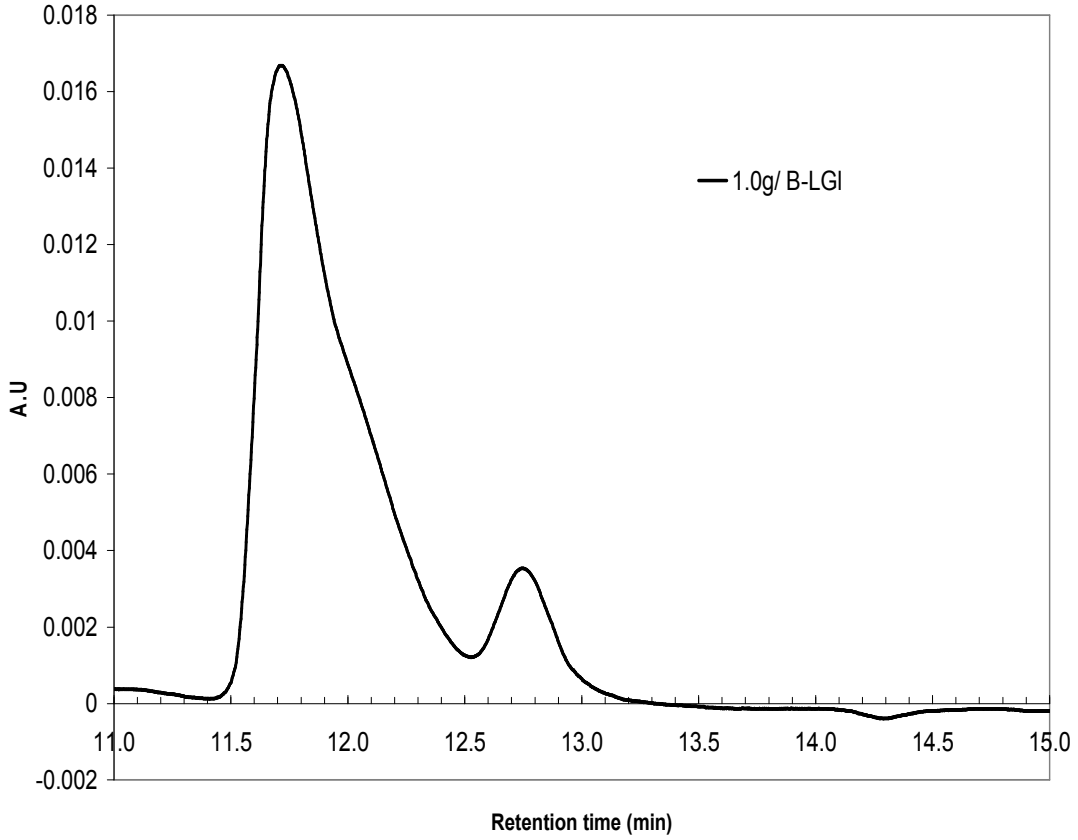


Figure G-13: HPLC Chromatogram for 1 g/l beta-lactoglobulin protein solution where the peak at retention time 12.75 min represents the monomer and the peak at retention time of 12.0 min represents the dimmer.

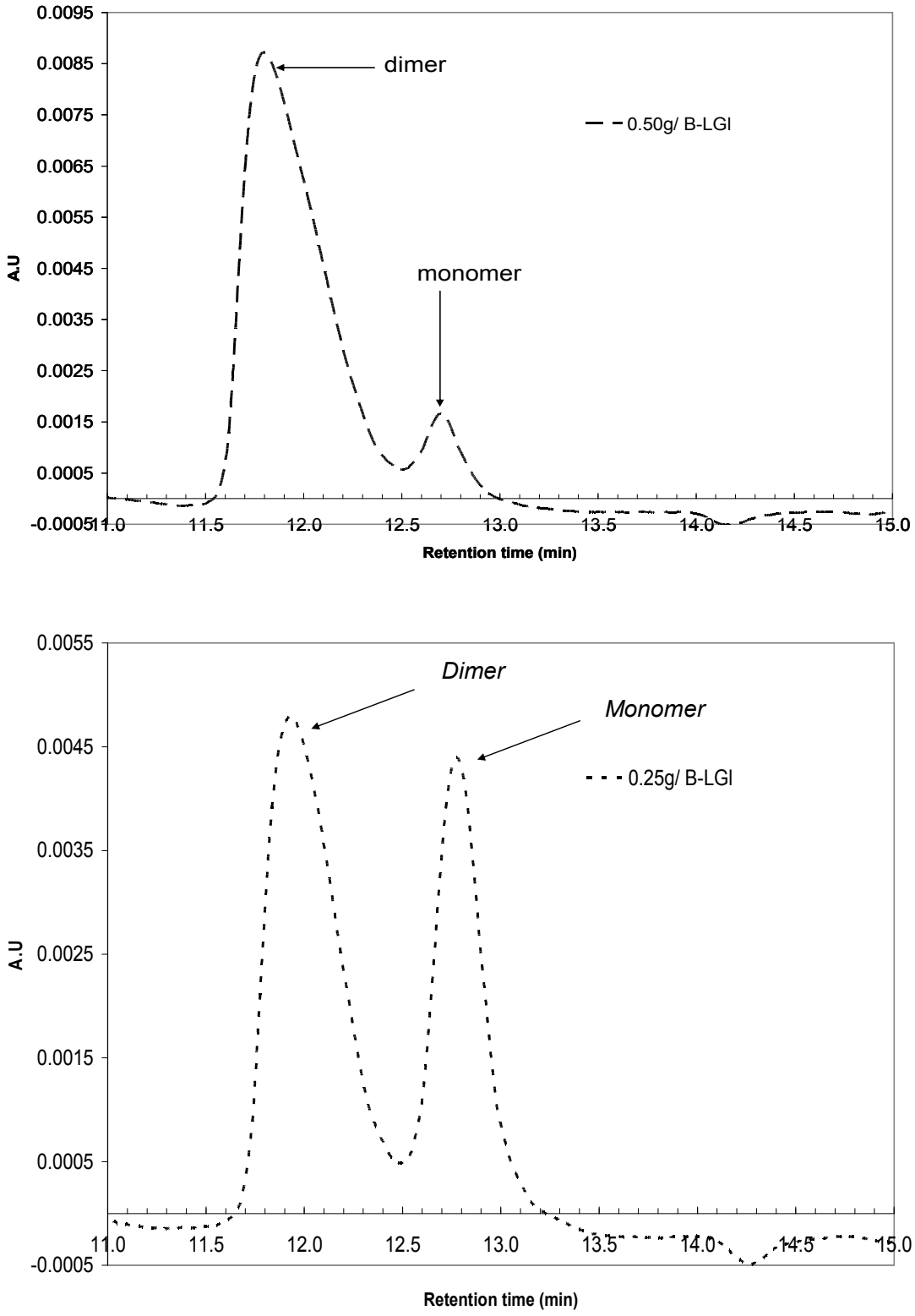


Figure G-14: HPLC Chromatogram for 0.5 g/l beta-lactoglobulin (top) and 0.25 g/l beta-lactoglobulin protein solution: the peak at retention time 12.75 min represents the monomer and the peak at retention time of 12.0 min represents the dimmer.

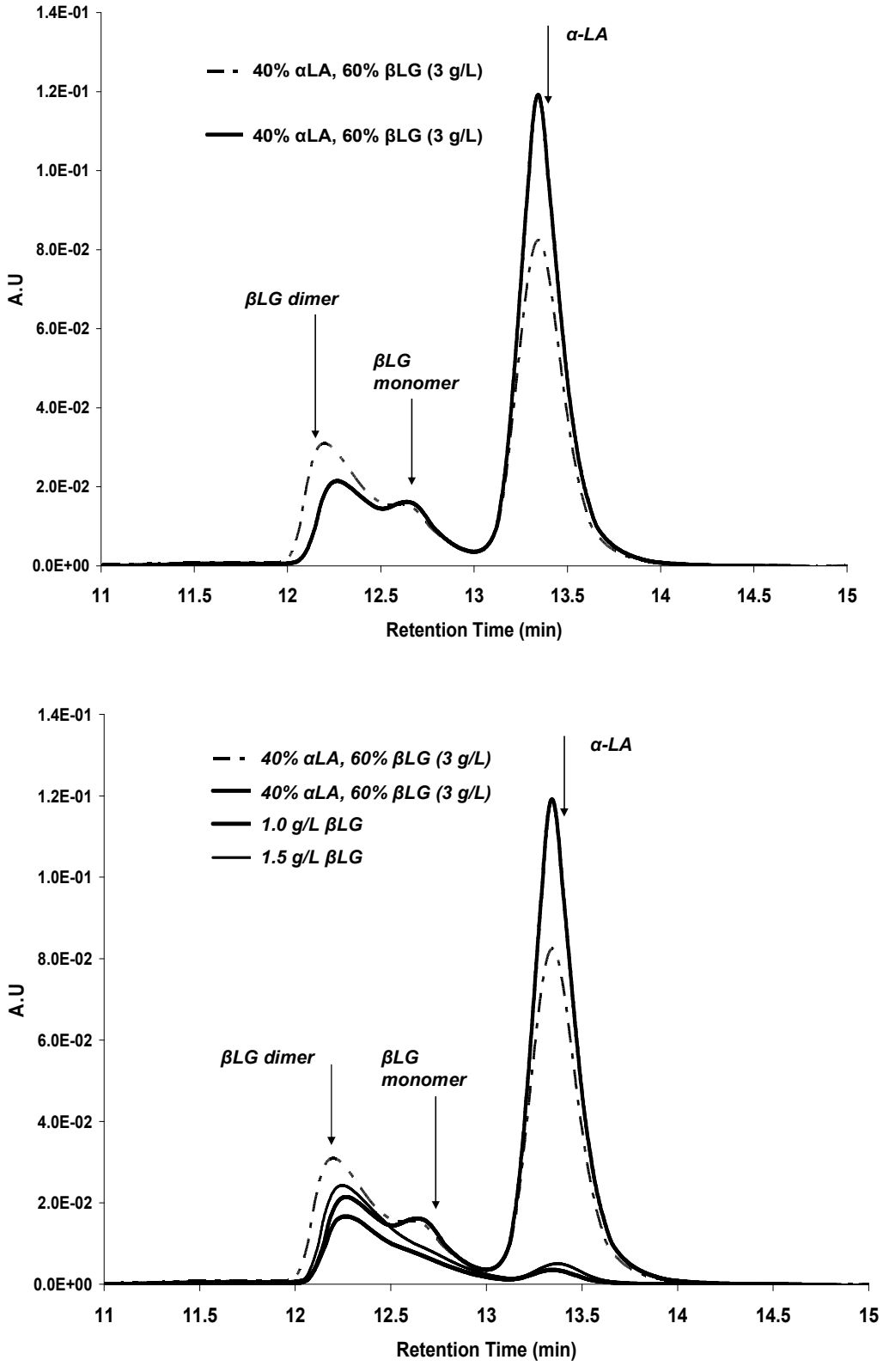
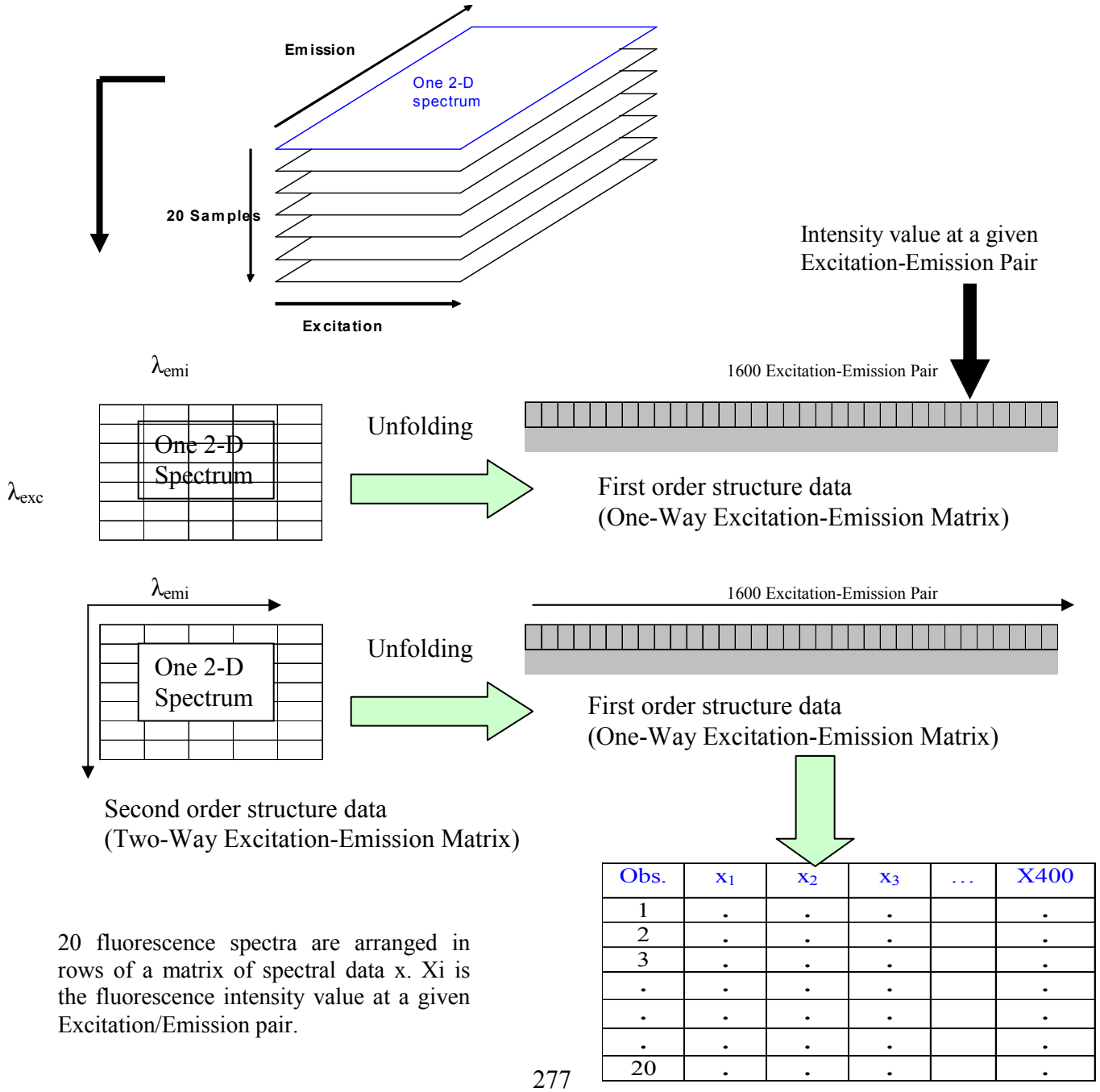


Figure G-15: HPLC Chromatograms for different β-lactoglobulin /α-lactalbumin protein solutions.

APPENDIX H: Unfolding of the Excitation-Emission Matrices for PLS Analysis

Let us take the case where 40 samples have been measured using fluorescence spectroscopy with 20 excitation wavelengths and 20 emission wavelengths making a three-way data array (40x20x20). Spectral raw data have to be transformed into a form suitable for the PLS analysis where each (20x20) excitation-emission matrix has been unfolded to (1x400) matrix as seen in figure. The unfolded excitation-emission matrices of dimension (1x400) can be subsequently arranged in one single two-way matrix of dimension (20x400). Such two-way matrix then has 20 rows (observations) and 400 columns (fluorescence intensity at excitation-emission pairs) as given below:



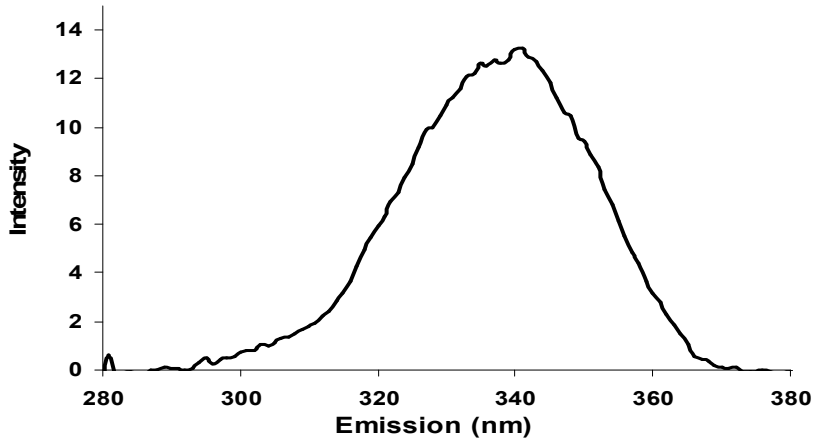
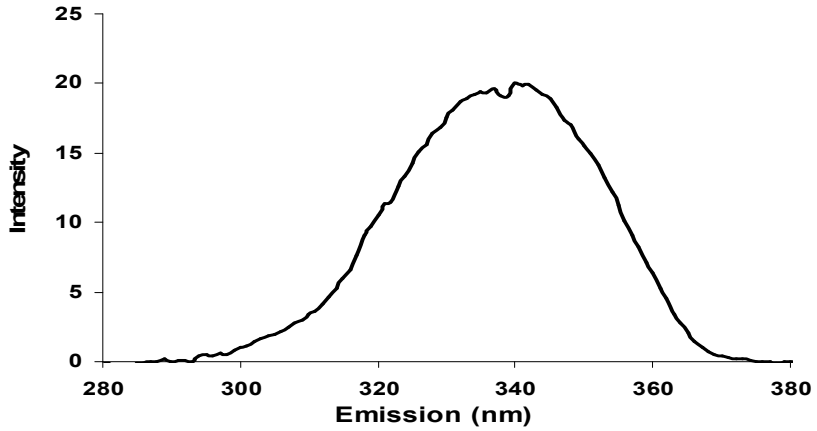
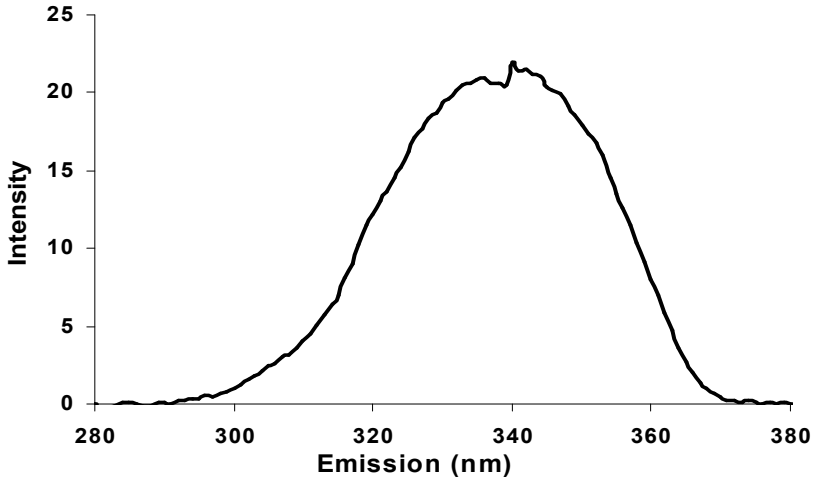
Obs.	x_1	x_2	x_3	...	X400
1	.	.	.		.
2	.	.	.		.
3	.	.	.		.
.	.	.	.		.
.	.	.	.		.
.	.	.	.		.
20	.	.	.		.

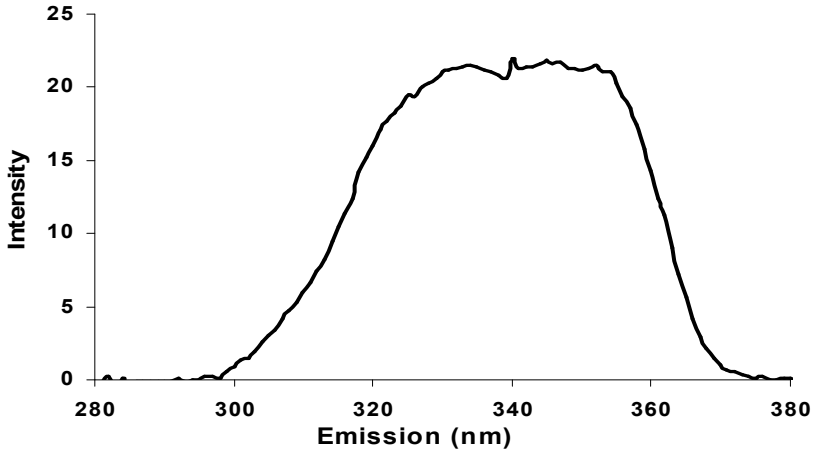
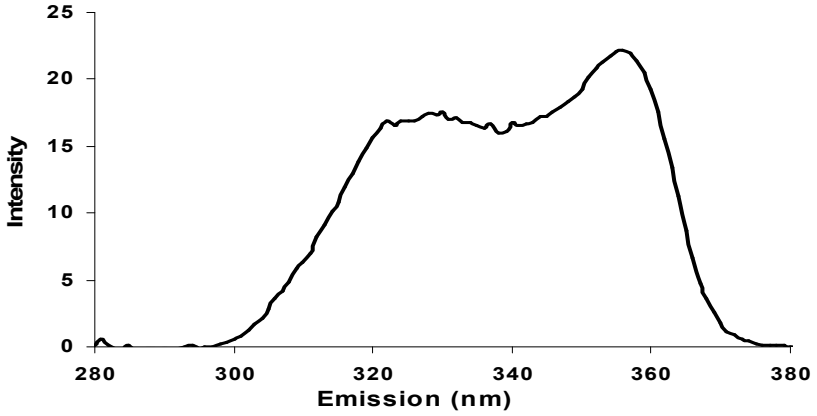
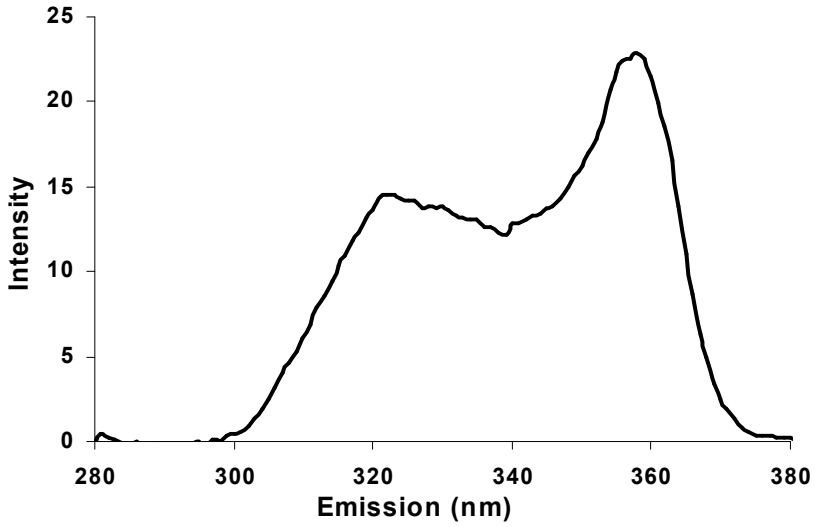
20 fluorescence spectra are arranged in rows of a matrix of spectral data x. Xi is the fluorescence intensity value at a given Excitation/Emission pair.

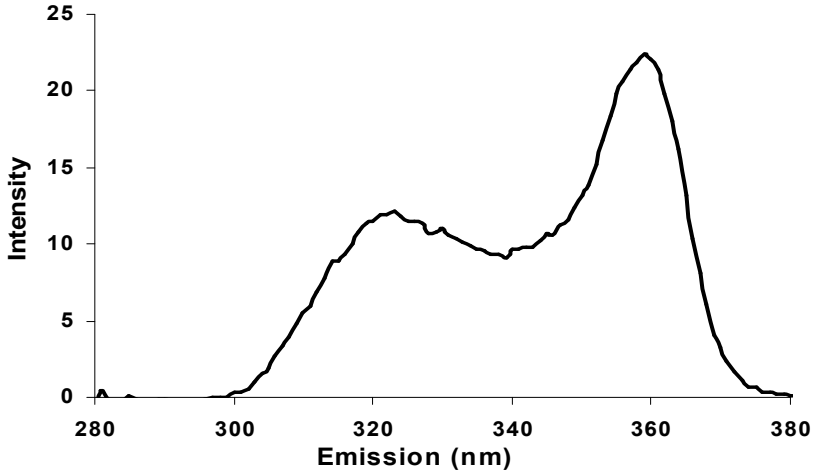
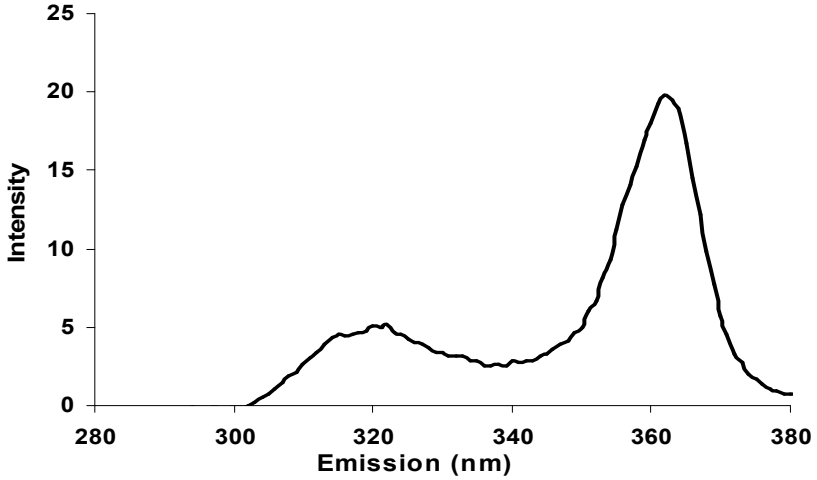
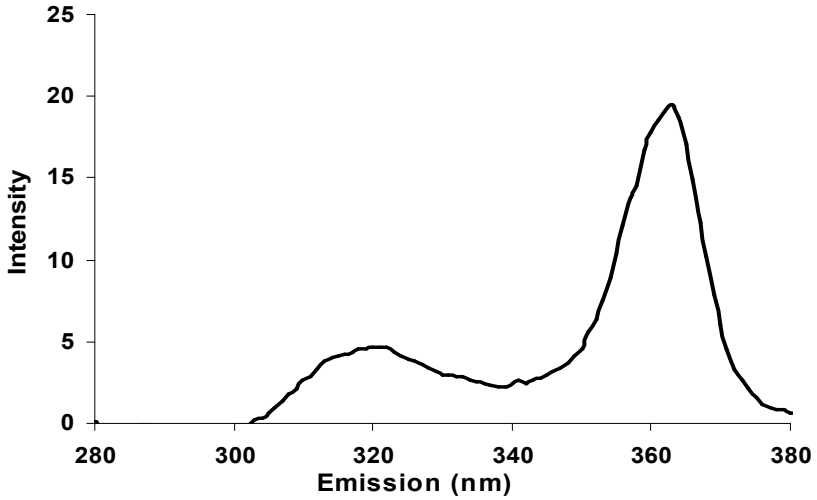
APPENDIX I: Factors Affecting the Fluorescence Signals of Proteins

Fluorescence spectroscopy is a rapidly growing science that can be used to derive significant information about biological solutions with little or no sample preparation. It may be used in applications when other methods would be too time consuming or require destruction of precious samples. It is possible to set up real-time monitoring of processes using spectroscopy, such as monitoring the ultrafiltration of protein-containing solutions. Monitoring the ultrafiltration of protein-containing solutions requires a fluorescence based assay for quantifying the foulants in the permeate and retentate during ultra filtration. However, we still need to better understand the subtle changes in fluorescence features as a result of changes commonly encountered during processing of protein-containing solutions. Good spectral databases documenting all the fluorescence features in response to changes in the processing conditions are also needed before fluorescence spectroscopy can be as widely used a tool as HPLC. Utilizing fluorescence spectroscopy in the design of a reliable monitoring system for bioprocesses that is accurate, fast, sensitive and that can be potentially implemented on-line is an extremely complex process requiring awareness of the major factors affecting the fluorescence signal. The material in the following sections outlines the sensitivity of protein fluorescence towards pH, ionic strength and other physical phenomena such as concentration quenching.

Table I-1: Effect of concentration-dependent interferences on the shape of synchronous spectra of whey protein isolate solutions acquired at Δλ=50 nm.

Concentration (g/L)	Synchronous spectra at Δλ=50 nm
0.125 g/L pH 6.3	
0.25 g/L pH 6.3	
0.5 g/L pH 6.3	

Concentration (g/L)	Synchronous spectra at Δλ=50 nm
1 g/L pH 6.3	
2 g/L Ph 6.3	
3 g/L pH 6.3	

Concentration (g/L)	Synchronous spectra at Δλ=50 nm
4.0 g/L pH 6.3	
9 g/L pH 6.3	
10 g/L pH 6.3	

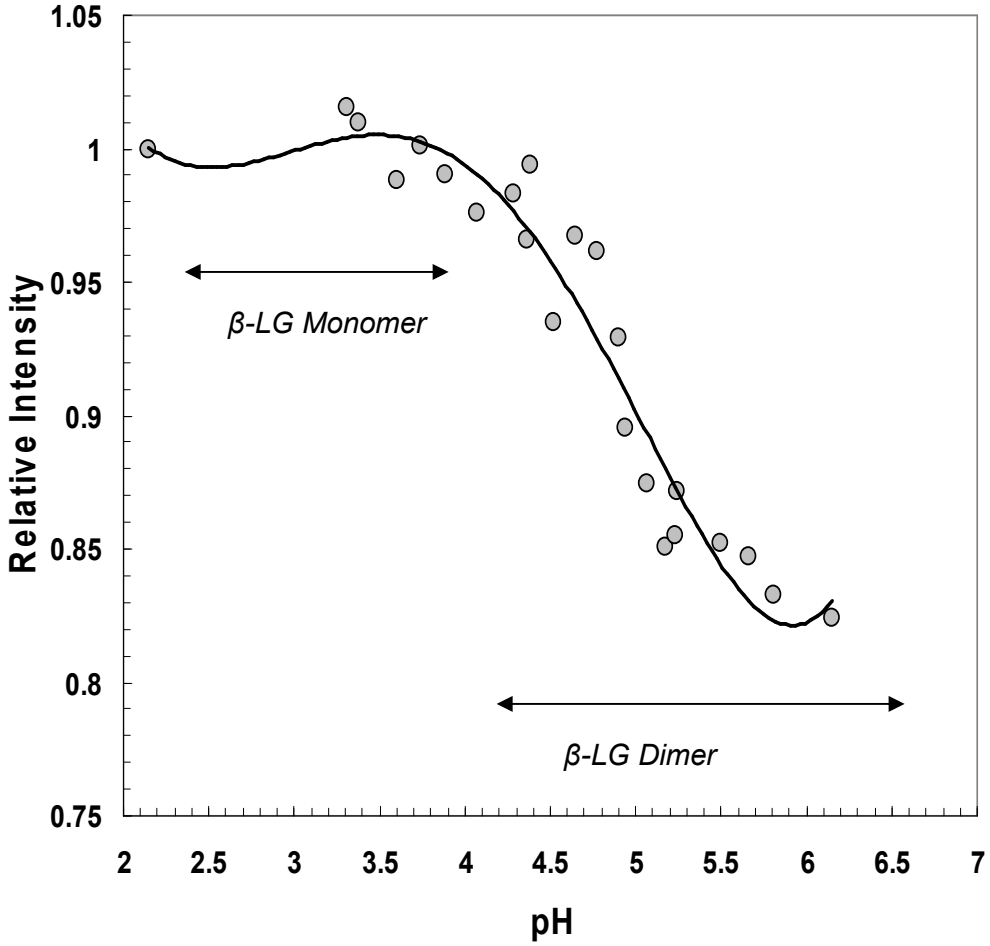


Figure I-1: Effect of pH on the emission intensity collected for 2 g/l β-LG solutions

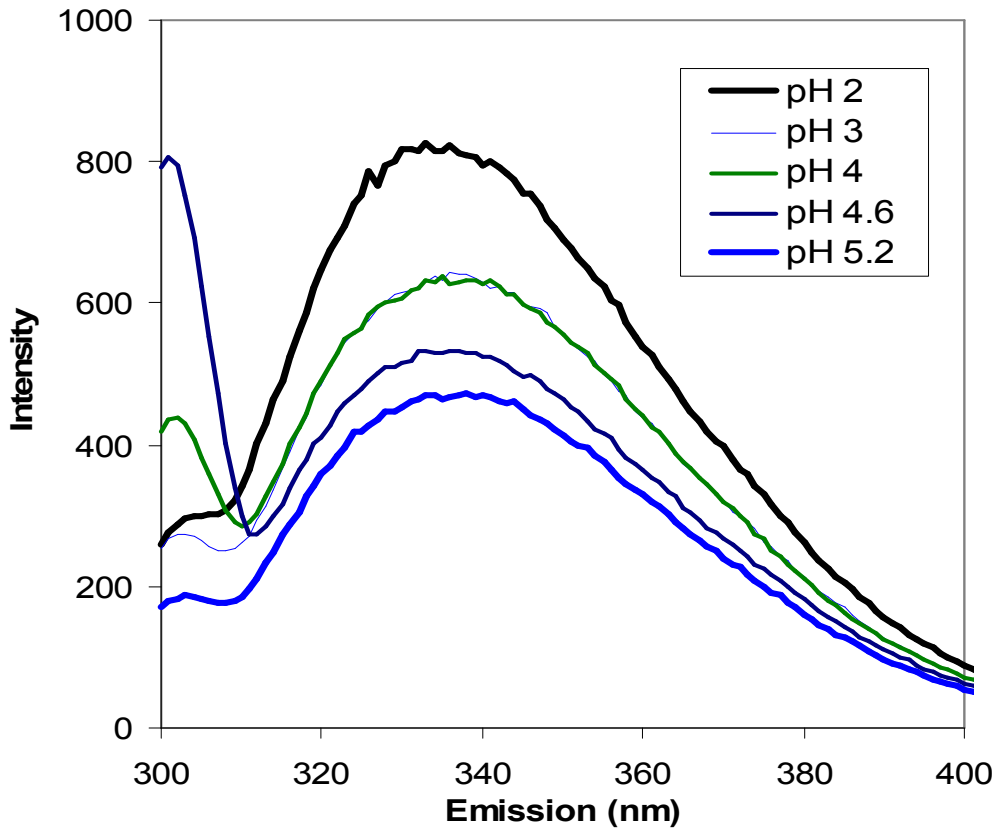


Figure I-2: Effect of pH on the emission spectra of 2 g/l β-LG solutions acquired at 270 nm excitation.

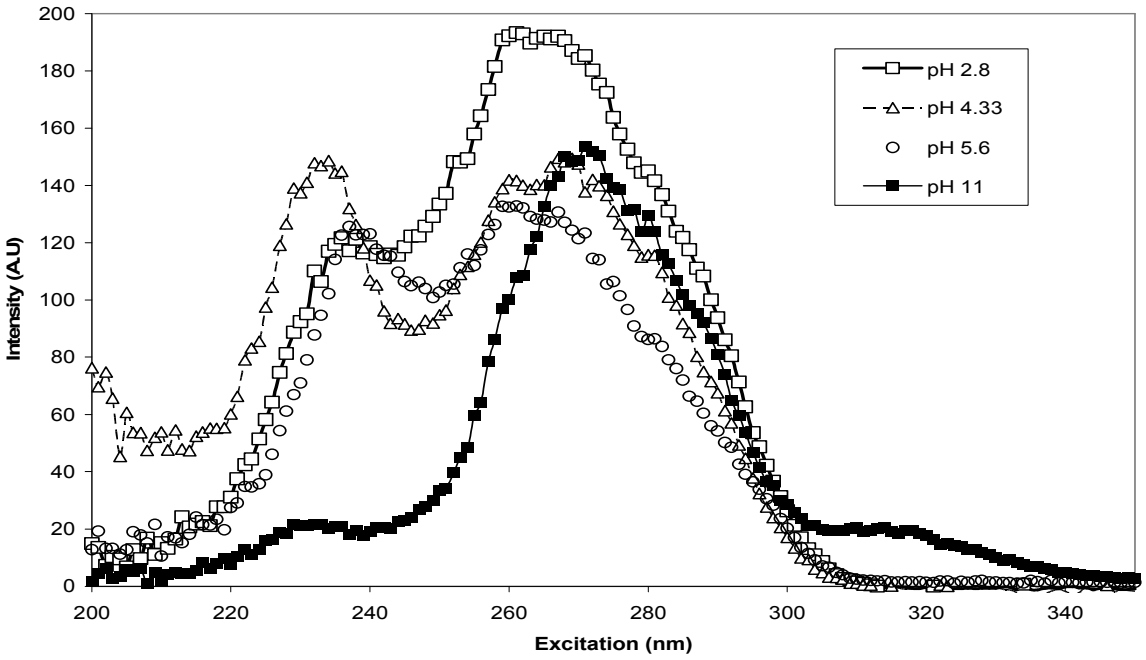


Figure I-3: Synchronous fluorescence spectra for four β-LG solutions (1 g/L) at different pHs.

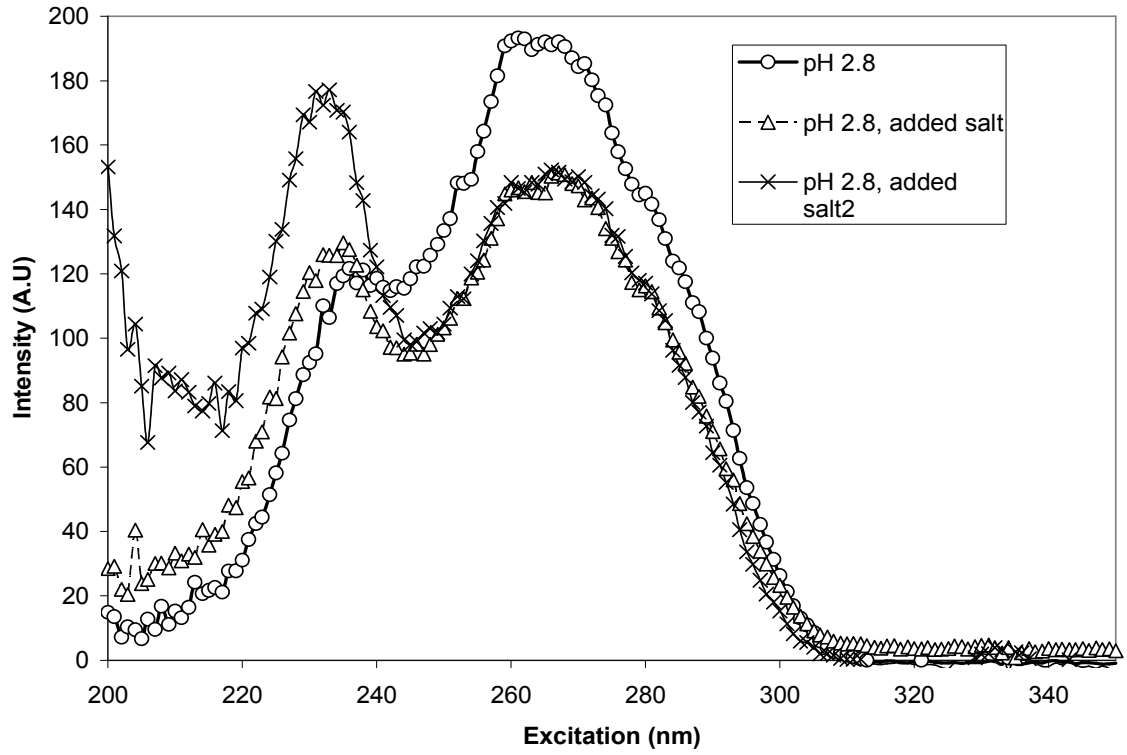


Figure I-4: Synchronous fluorescence spectra for three b-LG solutions (1 g/L) at three different ionic strengths (0, 100mM,200Mm).

APPENDIX J: Preliminary filtration experiments for single protein solutions of β-LG and two-protein solutions of α-LA and β-LG

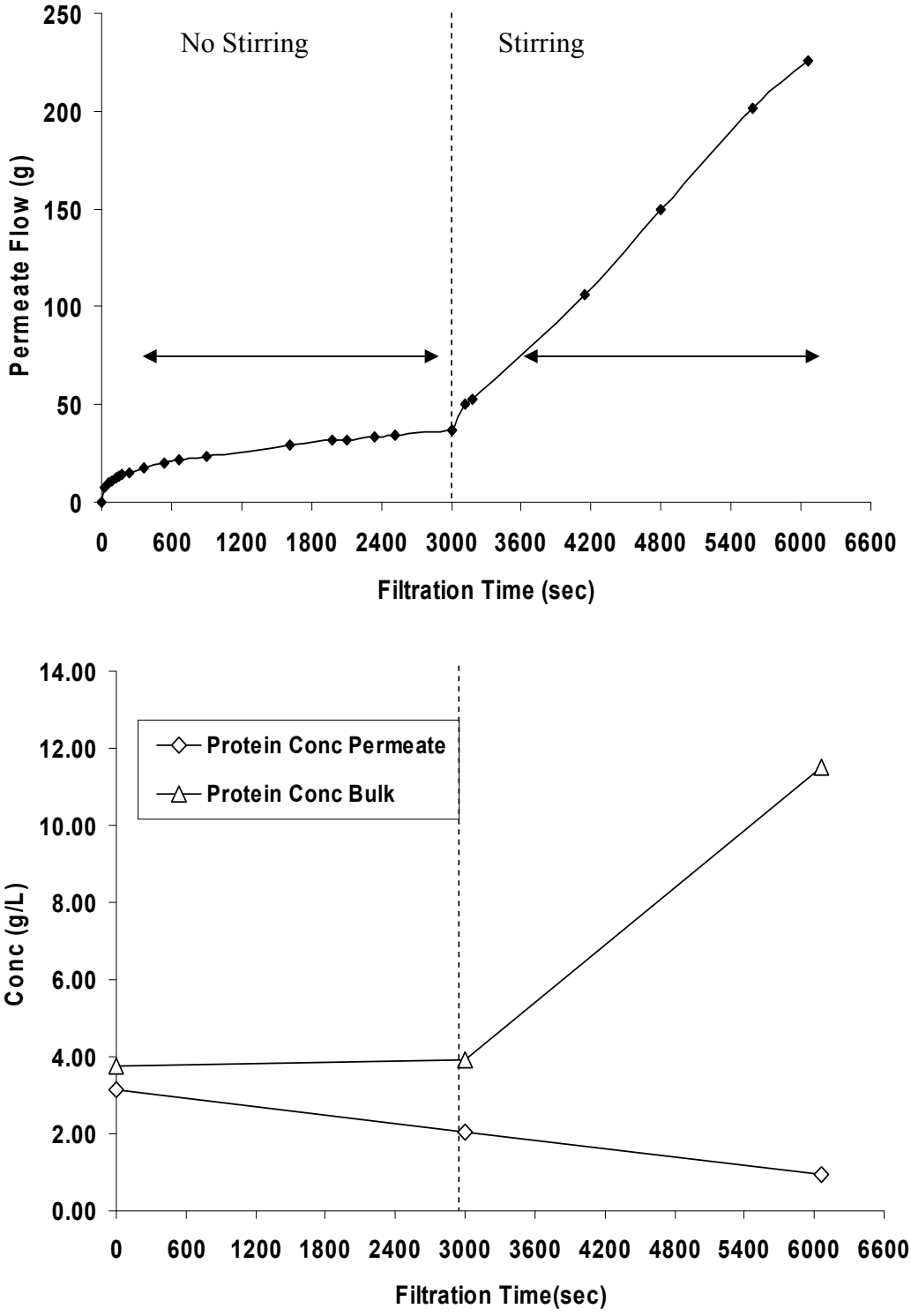


Figure J-1: The effect of stirring on the permeate flow (top) and β-LG protein concentration in the permeate and the bulk with respect to filtration time.

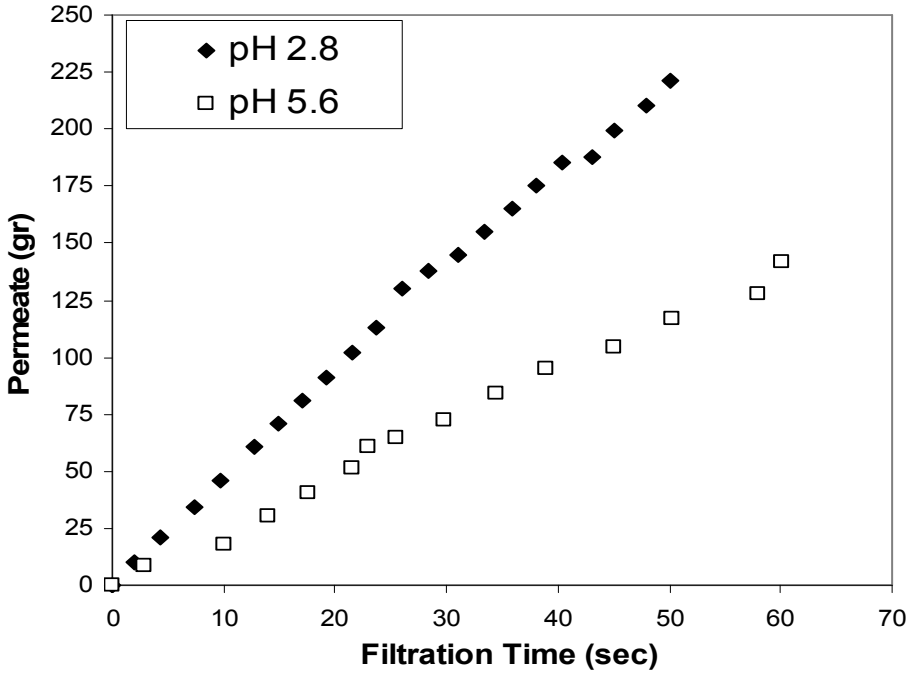


Figure J-2: Permeate mass collected for two different ultra filtration experiments (Exp 11 and Exp 12) performed at two different pHs (pressure difference 30 psi) for protein solution of α-LA and β-LG.

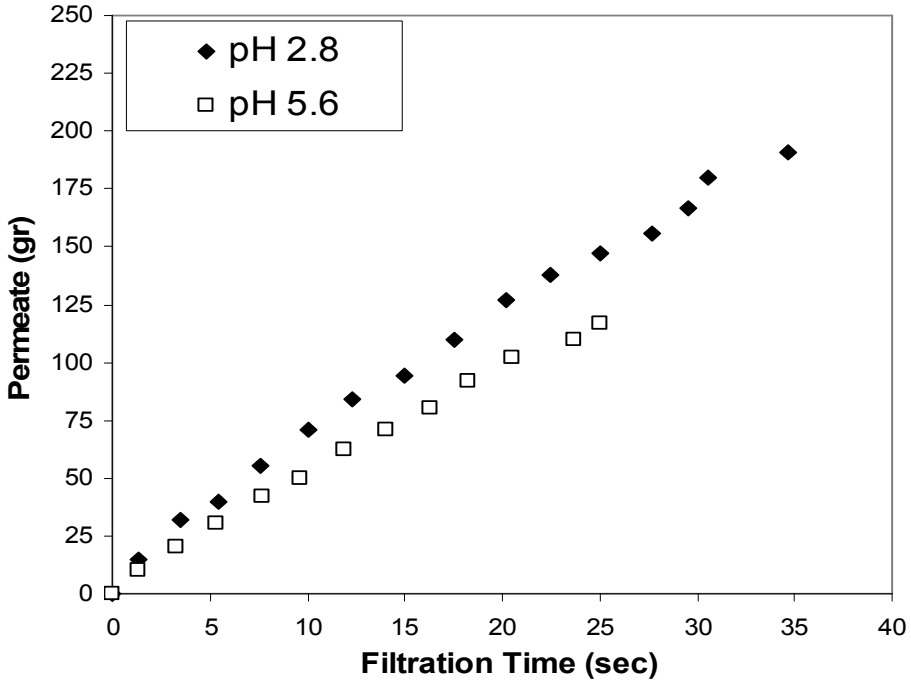
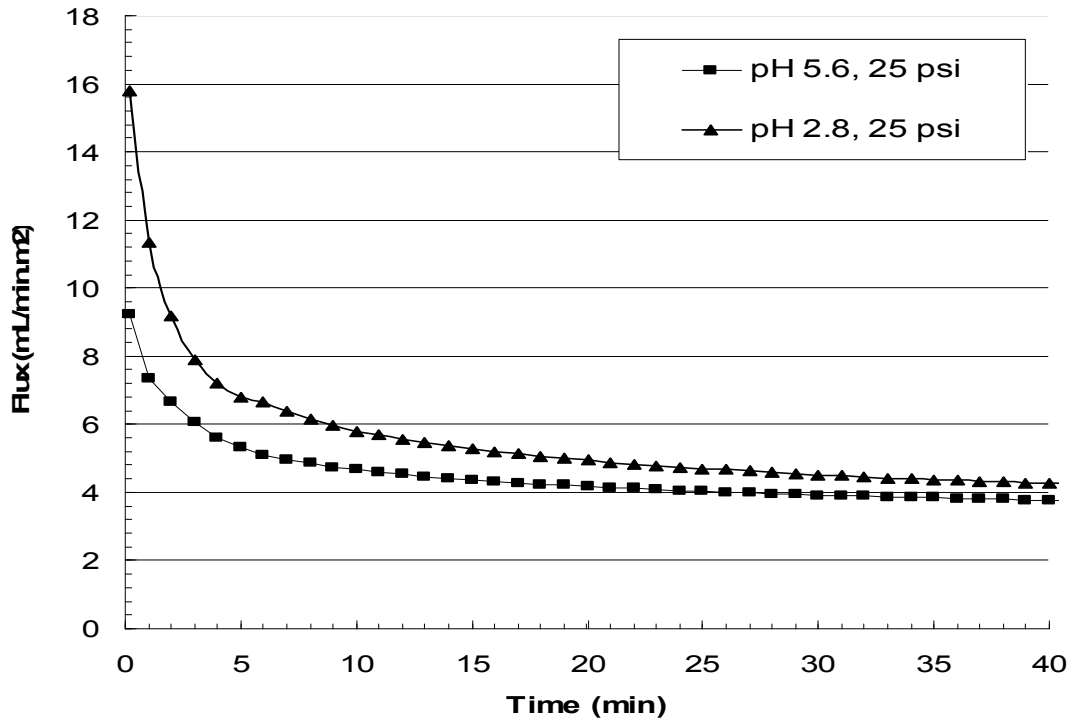


Figure J-3: Permeate mass collected for two different ultra filtration experiments (Exp 8 and Exp 10) performed at two different pHs (pressure difference 25 psi) for protein solution of α-LA and β-LG.



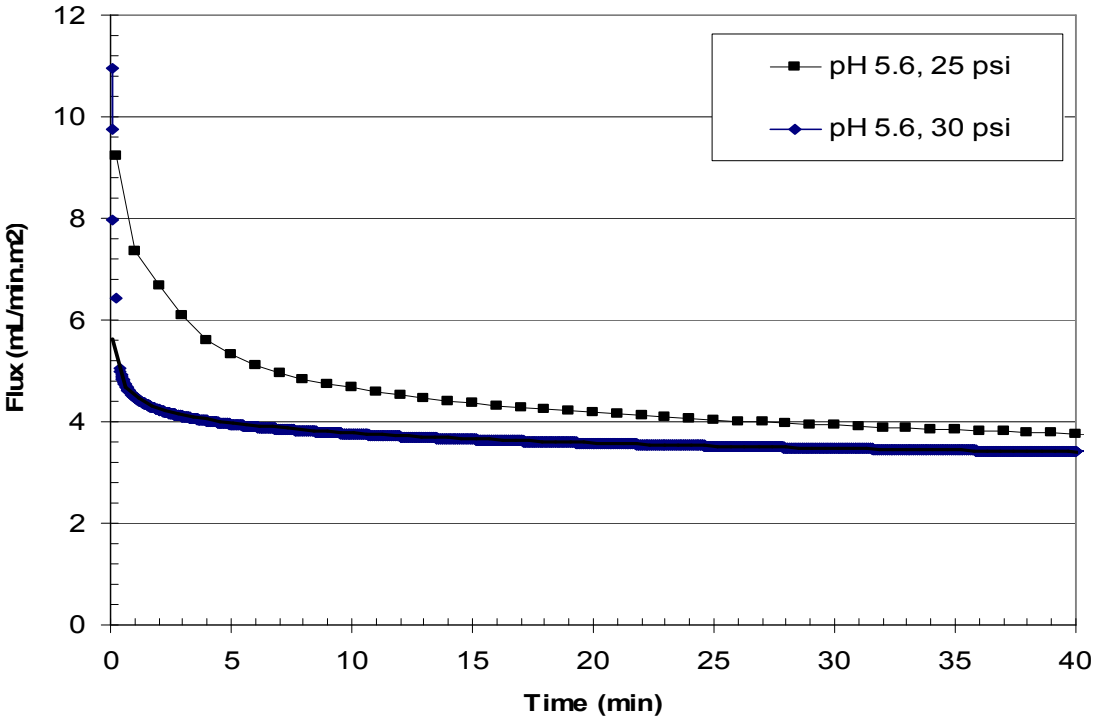


Figure J-4: Flux decline for two filtration experiments at two different pHs: pH 2.8 and pH 5.6, 25 psi (top) and flux decline data for UF pH 5.6 at two different transmembrane pressures 25 psi and 30 psi (bottom).