Monitoring of an Antigen Manufacturing Process Using Fluorescence

by

Vanessa Zavatti

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Applied Science

in

Chemical Engineering

Waterloo, Ontario, Canada, 2014

AUTHOR'S DECLARATION

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

Bordetella pertussis is one of two Gram-negative bacteria responsible for causing whooping cough in humans, a highly contagious disease that infects the human upper respiratory tract. Whole-cell and acellular vaccines have been developed but due to side-effects resulting from whole-cell vaccines, acellular vaccines are currently preferred to prevent this disease. A second bacterium known to cause whooping cough is *Bordetella parapertussis*, but since it causes less aggressive symptoms, only *B. pertussis* is utilized in the manufacture of the vaccine. One acellular vaccine is based on four virulence factors: pertussis toxin (PT), filamentous hemagglutinin (FHA), pertactin (PRN), and fimbriae (FIM).

The focus of this thesis was to explore the use of spectrofluorometry for monitoring and forecasting the performance of the upstream and downstream operations in the PRN purification process at Sanofi Pasteur. The upstream fermentation process includes a series of reactors of increasing volume where the microorganism is grown under controlled conditions. The PRN purification process involves a series of sequential steps for separating this protein from other proteins for later use in the vaccine. The PRN is precipitated in three steps with ammonium sulphate at three different concentrations. The pellet is collected by centrifugation and dissolved in a buffer solution followed by chromatographic separation. The run-through is then ultra-filtered and diafiltered in two separate steps. The resulting concentrate is dissolved in water and subjected to another chromatographic step and diafiltration. The final filtration of PRN involves a pre-filtration and sterile filtration. Finally, the samples are collected for quality control.

The objective of this work was to monitor the process at different steps of the upstream and downstream purification process by multi-wavelength fluorescence spectroscopy in combination with multi-variate statistical methods. From the spectra, it was possible to identify fluorescent compounds, such as amino acids and enzyme cofactors, without performing an additional pre-treatment or

purification step. Also, the identification of conformational changes in proteins and formation of complexes, such as NAD(P)-enzyme complex, was possible based on the shift in the emission peaks of the compounds identified. These results demonstrated the feasibility of using this tool for qualitative evaluation of the process.

Multivariate methods, such as PCA and PLS, were used to extract relevant information and compress the fluorescence data acquired. PCA was effective for correlating variability in the yield of pertactin to a particular fluorescence fingerprint. As a result of the study, it was concluded that a possible source of variability in the productivity that is observed might be a metabolic shift during the fermentation steps that leads to the accumulation of NAD(P)H (or NAD(P)H-enzyme complex) probably due to oxygen transfer limitations. This conclusion was reached after investigating changes in the dissolved oxygen, aeration, agitation, supplementation time and key metabolites (lactate, glucose, glutamine) profiles. The correlation of these parameters with low productivity it was not straightforward; however, some consistencies were observed, for example, high levels of glutamine in batches with low productivity. This fact might be related to the start of the supplementation time, which may be related to the dissolved oxygen, since the addition of the supplement is done manually when an increase of the dissolved oxygen is detected. It is believed that this factor is related to the low production of protein product, such as pertactin.

By means of PLS, it was possible to build regression models that allow for predicting the final concentration of pertactin from the fluorescence measurements. The models were built using the new variables obtained from data compression performed with PCA, and the final pertactin concentration measured by a Kjeldahl test. With this method, two regressions were constructed: (i) between NAD(P)H-enzyme complex spectra from the fermenters and pertactin concentration and (ii) between the pertactin fluorescence spectra from the last step of purification and pertactin concentration. A

third model was built using the protein content, the NAD(P)H-enzyme complex content in the fermenters and pertactin concentration.

Attempts were made to identify the possible enzyme that may bind to NAD(P)H, assumed to be a dehydrogenase. Substrates for different enzymes were employed with the objective of measuring changes in the fluorescence of the characteristic peak for this binding (Ex/Em=280/460 nm). Major changes were detected after addition of the substrates oxaloacetate, ubiquinone and succinate dehydrogenase. Since changes were detected with more than one substrate, it was not possible to unequivocally identify the enzyme; however, the results provide some insight into what may be happening at the metabolic level.

The work carried out in this thesis involved both analysis of samples provided or collected by the industrial sponsor as well as analysis of samples prepared at the University of Waterloo for measurement, interpretation and calibration. The proposed fluorescence-based method was found suitable for assessing protein quantity as well as for providing an indication of possible protein aggregation and conformational changes. Future work will be required to identify the exact source of variability in the production of pertactin, by means of monitoring the evolution of fermentation, NAD(P)H and ATP measurements, and oxidation redox potential assays.

Acknowledgements

I would like to express my deep gratitude to my supervisors Prof. Hector Budman and Prof. Raymond Legge for their huge patience, encouragement and useful critiques of this research work. I would also like to extend my thanks to Melih Tamer for his continuous support and advice.

Financial support provided by Natural Sciences and Engineering Research Council of Canada (NSERC), and Sanofi Pasteur is greatly appreciated.

I would like to offer my deeply thanks to Ramila Peiris for sharing his experience, knowledge and providing me with computer files of his previous work that gave me the start point to write my codes. I would like to show my gratitude to Kaveh Ohadi for helping me in the lab, and for his assistance in the modeling work.

I appreciate the feedback offered by Liliana Sampaleanu, Javier Menendez, George Georgiou, Dominic Cuerrier, Domenic Luciano, Mirnader Ghazali, Rob Maharaj, Lolita Champlin and Kara Ryder. Special thanks to James Roos, Vithien Nguyen and Kevin Li for helping me in collecting the samples and in the lab work.

I am also thankful to my friends and colleagues in the Department of Chemical Engineering, specially Ali Nikdel, Hengameh Mohseni and Yuncheng Du for their kind help and support.

I am truly indebted and thankful to Graciela Rahbe, Ivan Silis and Dario Bernal for their support while I stayed in Toronto. Many thanks to my friends Lorena Danisi, Katuska Martinez, Patricia De Lisa, Jennifer Charry, Alberto Betancourt and Ivan Padilla for encouraging me to write this thesis.

Finally, I would like to express my gratitude to my family, my parents Gabriele and Bruna, my siblings Elvis and Gabriela, and brother-in-law Vladimiro Mujica for their support and encouragement throughout my study.

Dedication

To Venezuela. To the brave students that are fighting for our country at this moment, to those who gave their life for freedom and no longer will have the opportunity to fulfill their dreams of becoming a professional, or finishing a master's degree as I did.

Table of Contents

List of Figures

List of Tables

Nomenclature

Å – Armstrong

AT – Autotransportes

B – Regression coefficient matrix

C – c_i vectors matrix

c_i – Loadings vector of the ith Latent variable in Y space

CO_2 – Carbon dioxide

DCA – Difteria Toxin

DiFMU – 6,8-Difluoro-4-Methylumbelliferyl

DiFMUP – 6,8-Difluoro-4-Methylumbelliferyl Phosphate

DMSO – Dimethyl sulfoxide

DTP – Diphteria tetanus and pertussis

E – Residual matrix

ELISA – Enzyme linked immunosorbent assay

Em – Emission

Ex – Excitation

F – Residual matrix

FAD – Adenine dinucleotide

FHA – Filamentous hemagglutinin

FIM – Fimbriae

FMN – Flavin mononucleotide

FORS – First order Rayleigh scattering

G – Substrate concentration (g/dm^3)

HCl – Hydrogen chloride

HPLC – High performance liquid chromatography

K – kinetic constant (h^{-1})

kDa – Kilo Dalton

LPS – Lipo-polysaccharide

MOPS – 3-(N-morpholino)propanesulfonic acid

NAD(P)H – Nicotinamide adenine dinucleotide phosphate

NADH - Nicotinamide adenine dinucleotide

O_2 – oxygen

PBS – Phosphate buffered saline

PCA – Principal component analysis

Phe – Phenylalanine

p_i – Loadings vector of the i^{th} principal component

PLS – Partial least square

PMT – Photomultiplier tube

PRN – Pertactin

PT – Pertussis Toxin

Q – q_i vectors matrix

q_i – Loadings vector of the i^{th} Latent variable in X space

RS – Raman scattering

SDS-PAGE – Sodium dodecyl sulphate polyacrylamide gels

s_i – Scores vector of i^{th} principal component

SORS – Second order Rayleigh scattering

T – t_i vectors matrix

TCA – Tetanus toxin

t_i – Scores of i^{th} Latent variable in X space

Trp – Tryptophan

Tyr – Tyrosine

U – u_i vectors matrix

u_i – Scores of i^{th} Latent variable in Y space

UV – Ultraviolet

W – Weigh matrix (relationship between U and T)

WFI – Water for Injection

x – Biomass concentration (g/dm3)

X – Fluorescence data matrix

X_{new} – New fluorescence data matrix

Y – Response data matrix

Y_p – New response data matrix

Greek Letters

α – Tryptophan peak at Ex=280 nm

α'- Tryptophan peak at Ex=230 nm

β – NAD(P)H peak at Ex=340 nm

β' – NAD(P)H peak at Ex=260 nm

γ – Amino acid emission peak

ε – Absorption coefficient

ϕ – FAD peak at Ex=450 nm

λ – Wavelength (nm)

μ – Growth rate constant (h^{-1})

Chapter 1
Introduction

1.1 Research Motivation

Whopping cough is a bacterial infection of the lungs, characterized by severe coughing fits. *Bordetella pertussis* is one of the causative agents of this condition. To prevent this disease, whole-cellular and acellular vaccines have been developed, but due to side effects associated to the whole-cellular vaccines, acellular vaccines are preferred even though are less effective (Thalen, 2008). Current acellular vaccines are based on virulence factors: pertussis toxin (PT), filamentous hemagglutinin (FHA), pertactin (PRN), and fimbriae (FIM). Not all vaccines contain the four listed virulence factors; some of them are based on just two antigens.

In this project, the upstream fermentation process and the downstream purification of PRN were monitored and evaluated using spectrofluorometry in combination with multivariate statistical based models. Process monitoring is essential for understanding the sources of variability of a biotechnological process. Having an appropriate insight into the process, allows optimizing and increasing productivity (Marose *et al.*, 1998). The use of fluorescence spectroscopy has been reported for biomass detection, characterization of reactors and bioprocess monitoring (Manrose *et at*., 1998; Lindemann *et al*., 1998; Skibsted *et al*., 2001; Hagedorn *et al*., 2002). Since a wide range of excitation-emission wavelength can be monitored using fluorescence spectroscopy, several fluorophores, such as proteins, vitamins and coenzymes, can be detected simultaneously (Lindemann *et al*., 1998). However, spectral overlap, noise, strong correlations among data and complex interferences make difficult data interpretation (Harms *et al*., 2002). For these reasons, multivariate

analysis was used to extract the most important information associated with the properties of the system being studied. First principal component analysis (PCA) was used to compress the data. The resulting compressed data for different stages of the process was analyzed for identifying sources of variability in productivity. Then, the partial least squares' approach (PLS) was used for correlating the fluorescence spectra and the final protein content of PRN as measured by a Kjeldahl test.

1.2 Description of the upstream fermentation and downstream purification process

Figure 1.1 shows a flow diagram of the upstream and the downstream purification process of pertactin (PRN) which is the antigen of interest for the current work.

The upstream process is started by thawing a frozen Component Pertussis seed. Then, the thawed culture is transferred to seed flasks containing liquid medium which are incubated for 48 to 96 hours. Following incubation, the cultures are added to two seed 20 L fermenters (initial fermentation stage) containing mainly Component Pertussis broth and growth factors. The seed fermenter cultures are grown for 22 to 25 hours. Subsequently, the cultures are transferred to two 200 L fermenters (intermediate fermentation stage) for 22 to 25 hours, and then to two 2000 L fermenters (production fermentation stage) for up to 56 hours. All fermenters work under the same operating conditions in terms of temperature (36^{o}C), dissolved oxygen (35%) and pH (7.2) levels, which are continuously monitored and controlled in closed loop. In the production fermenters, supplement is added when the initial supply of nutrients is exhausted. The depletion of nutrients is detected by a significant sudden decrease in oxygen consumption which is accompanied by a corresponding spike in dissolved oxygen.

These fermentation steps are followed by harvesting, which involves the centrifugation of the fermentation culture. The cell paste, coming from the centrifugation, is collected into a tank for further fimbria (FIM) purification, and the supernatant to be referred as centrate is filtered to ensure the complete removal of the live cells.

In the downstream process, the filtered centrate, which contains PT, FHA and PRN, is further concentrated by using 10-kDa cellulose membranes. Following this, a perlite fractionation is used in the initial separation of three antigens which are needed for the acellular vaccine formulation in addition to FIM: pertactin (PRN), filamentous haemogglutinin (FHA) and pertussis toxin (PT). PRN is collected in the runthrough of the column used in this separation step, while PT and FHA are eluted off the column.

After this step, PRN is precipitated in three consecutive precipitation steps using ammonium sulphate at three different concentrations. In the first and third precipitation, the pellet is the fraction of interest while in the second precipitation the supernatant is used. The final pellet collected by centrifugation is dissolved in 10 mM HCl Tris buffer and loaded onto a chromatography column. Finally, the product is ultrafiltered and diafiltered before being subjected to a final chromatography step. The final filtration involves a pre-filtration step and a sterile filtration (0.2 μm filter) prior to quantification.

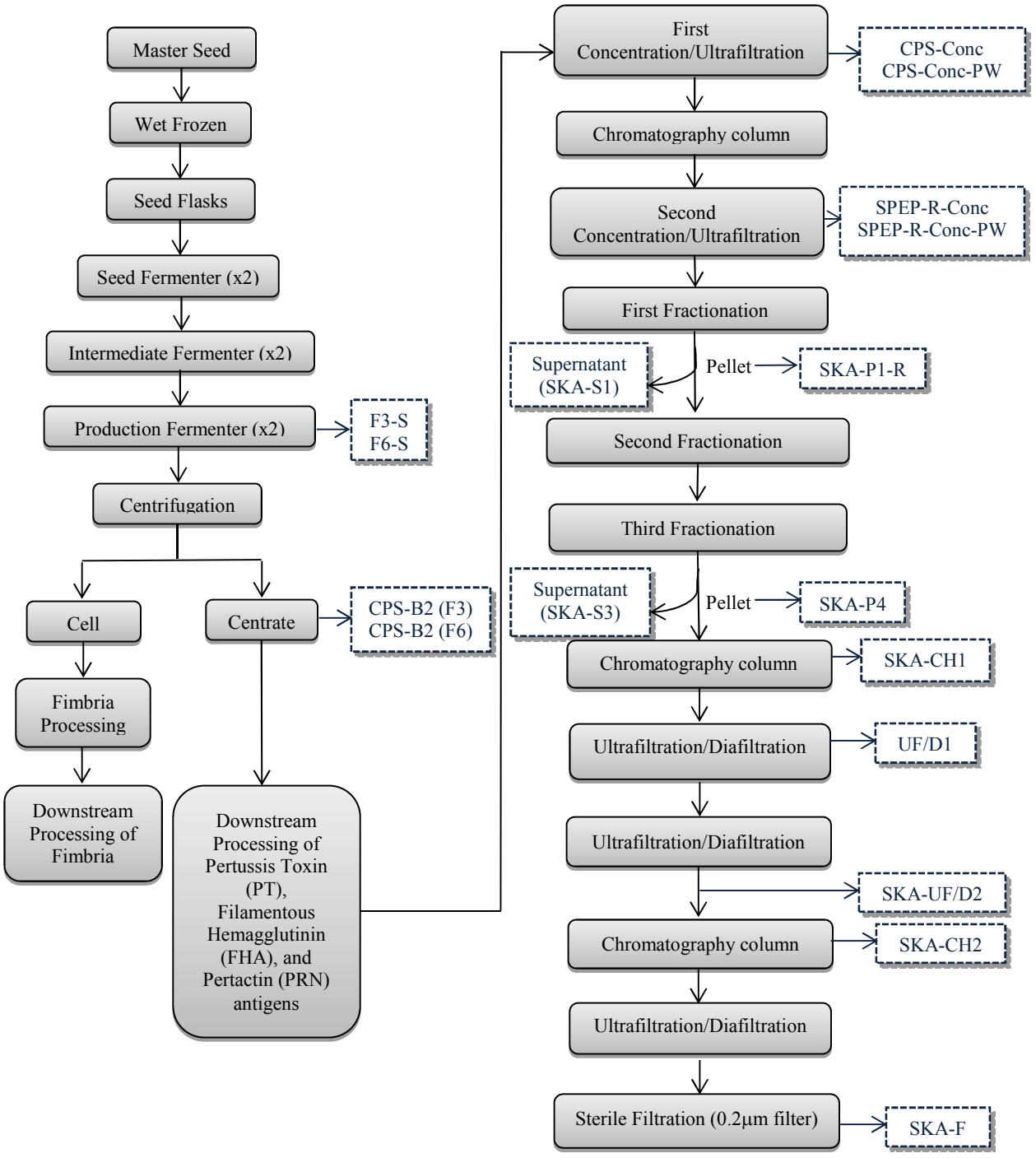

Figure 1.1 Flow-diagram of the upstream and downstream PRN purification process

After the purification process, PRN is adsorbed onto aluminum phosphate, which acts as an adjuvant. Adjuvants are compounds that serve to enhance the immune system in response to an antigen (Lindblad, 2004). In the adsorption process, an aluminum phosphate solution is added to the antigen, and mixed for 4 days at 16ºC to 24ºC. After the adsorption, PRN is stored at 2ºC to 8ºC up to 36 months.

1.3 Research Objectives

The main objective of this thesis is to monitor and evaluate the upstream and downstream purification process of PRN as a component in the manufacturing of the whooping cough vaccine at Sanofi Pasteur (Figure 1.1) by employing multi-wavelength fluorescence in combination with multi-variate statistical methods. The specific objectives are the following:

1. Obtain reproducible fluorescence spectra for different steps in the PRN manufacturing process. By acquiring fluorescence measurements, it is possible to identify fluorescent compounds present in the samples simultaneously, and to track changes of these compounds through the process.
2. Develop a fluorescence-based calibration model to identify changes in the antigen manufacturing process. Since spectra can overlap and the existence of noise can render difficult data interpretation, it is necessary to use statistical methods to extract the most important information from the measurements.
3. Assessment of parameters, such as temperature, that can influence protein degradation during storage.
4. Examine the feasibility of applying spectrofluorometry to surface phosphophilicity assessment of the adjuvants in adsorbed samples. Adjuvants adsorb in the surface of the

antigens used in vaccines formulation. Changes in the fluorescence spectra could be indicative of variation of adsorption of antigens on the adjuvant.

1.4 Thesis Structure

This thesis includes five chapters. The current chapter contains the description of the process and research objectives. Chapter 2 presents the theoretical background about whooping cough and vaccines, fluorescence, instrumentation, uses of spectrofluorometry in biological processes, and multivariate methods. Chapter 3 explains the materials and methods employed in the three main analyses that were performed in this thesis: fluorescence data acquisition and analysis of the upstream and downstream pertactin purification process, enzyme assay and surface phosphophilicity assay. Chapter 4 presents the results and discussion for data acquisition, data compression, and regression model. Chapter 5 highlights final conclusions and recommendations for future work.

Chapter 2
Literature Review

2.1 Whooping Cough

Whooping cough, or pertussis, is a contagious infection of the lungs and airways. In humans, one of the causative agents of this illness is the gram-negative bacterium *Bordetella pertussis*, which was first isolated by Bordet and Gengou in 1906. The name pertussis means “violent cough” (Hijnen, 2008). This disease happens year round everywhere in the world. Each year in Canada an average of 2,000 people fall ill from whooping cough, and worldwide, there are around 20 to 40 million cases and 400,000 deaths from this condition each year. Symptoms are initially mild, and then develop into severe coughing fits, which produce the namesake high-pitched "whoop" sound when air is inhaled after coughing. Without treatment, whooping cough can last weeks or months, and can cause brain damage or even death. It is most dangerous for children under 1 year in age, especially if they are unvaccinated or under-vaccinated (Public Health Agency of Canada, *Pertussis (whooping cough)*, Canada. Last Date Modified: 21 February 2014 <http://www.phac-aspc.gc.ca/im/vpd-mev/pertussis-eng.php>)

Aside from *Bordetella pertussis*, there are several other pathogenic species in the genus *Bordetella* that can cause upper respiratory tract infections. *B. parapertussis* causes a mild whooping cough-like syndrome, whereas *B. bronchiseptica*, an animal pathogen, is isolated rarely from humans (Locht, 1999). Since *B. parapertussis* causes less severe symptoms, all vaccines are based on *B. pertussis* (Thalen, 2008).

The creation of vaccines started immediately after the isolation of *B. pertussis*. In the early 1930s, the first mono component whole-cell pertussis vaccines were prepared (Hijnen, 2006), and in 1947 started the production of combination vaccines, where pertussis vaccines were combined with diphtheria and tetanus toxoids (DTP). Mortality was reduced in 10-fold with the introduction of these vaccines (Hijnen, 2006).

Although the whole cell vaccines were highly effective, they had drawbacks related to the side-effects (Hijnen, 2006; Thalen, 2008). Symptoms such as fever and more serious events such as convulsions were believed to be due to the lipo-polysaccharide (LPS) in the outer membrane of *B. pertussis*. For this reason, the manufacture of acellular vaccines that only consist of one or more *B. pertussis* antigens was necessary. During the 80's and 90's, the development of acellular vaccines started, which replaced the previous whole-cell vaccines (Thalen, 2008). Even though acellular vaccines are less effective than whole-cell vaccines, they are less likely to provoke side effects (Thalen, 2008).

Typically acellular pertussis vaccines contain at least two of four purified and detoxified *B. pertussis* antigens: pertussis toxin (PT), filamentous hemagglutinin (FHA), pertactin (PRN) and fimbriae (FIM2 and FIM3). The selection of these antigens was based on the strong immune response provoked in both human and animal models (Locht, 1999; Noofeli, 2007).

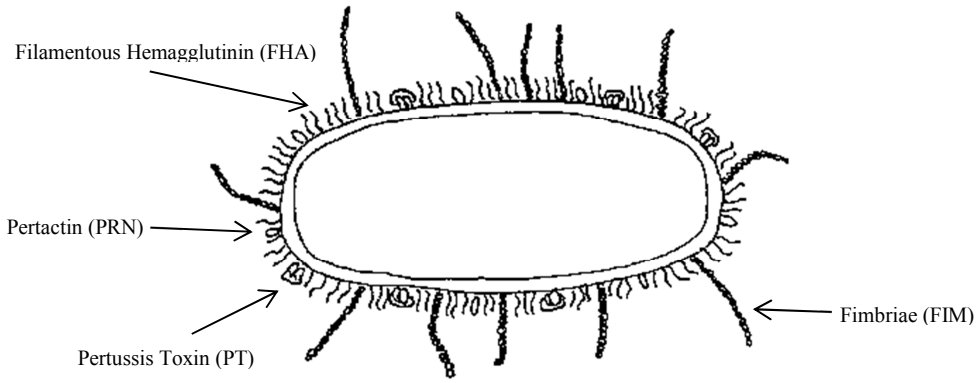

Figure 2.1 Schematic representation of B. pertussis antigens that are components of acellular whooping cough vaccines (adopted with modifications from Thalen, 2008)

Antigens or virulence factors are proteins used by a pathogen to enter the host and allow interaction with a specific target cell. Ultimately the pathogen must evade the host defenses to survive, damage the cell host and propagate either itself or its products to cause disease (Noofeli, 2007).

Many of the virulence factors characterized for the *Bordetellae* are common to many members of this genus including *B. pertussis*, *B. parapertussis* and *B. bronchiseptica*. These factors consist of adhesins (needed for the attachment to the host cells) that include filamentous hemagglutinin (FHA), tracheal colonization factor and fimbriae (FIM), pertactin (PRN), and toxins, such as adenylate cyclase haemolysin, dermonecrotic toxin and tracheal cytotoxin (Figure 2.1). However, there are other virulence factors which are expressed by just one of the species, such as pertussis toxin and serum resistance to killing protein (BrkA), which are expressed only by *B. pertussis* (Noofeli, 2007).

Pertussis toxin (PT) is one of the major protein toxins secreted by *B. pertussis*. It is released into the extracellular milieu and remains cell-bound. PT is a 105 kDa hexamer consisting of five subunits which are responsible for the interaction of the toxin with the target cell receptors (Locht, 1999).

Fimbriae are submicroscopic proteinaceous appendages that protrude from the cell surface. They are comprised of two major subunits: Fim2 (22.5 kDa) and Fim3 (22 kDa) (Noofeli, 2007). These units bind to sulphanated sugars that are present in the respiratory tract (Badu *et al*., 2001).

Filamentous hemagglutinin (FHA) is a major adhesin of *B. pertussis* and facilitates adherence of the cells to the respiratory epithelial cells. It is a filamentous structure about 2 nm wide and 50 nm long, and has a molecular weight of 220 kDa (Badu *et al*., 2001).

Pertactin (PRN) is an outer membrane protein partly responsible for adhesion of the bacterium to the host cell. This antigen was initially referred to as P69 or 69-kDa protein because of its apparent molecular weight as determined by sodium dodecyl sulphate polyacrylamide gels (SDS-PAGE). However, it was characterized by other methods such as mass spectroscopy (MS), the molecular weight was found to be approximately 60 kDa (Charles *et al*., 1994; Noofeli, 2007).

PRN belongs to the family of extracellular proteins called autotransporters (AT) found in various Gram-negative bacteria. The autotransporter secretion, or type V secretion, is one of the mechanisms developed by Gram-negative bacteria to achieve their physiological goals i.e. surface motility, host invasion (Henderson *et al*., 1998). The name autotransporter derives from an initial understanding that the protein is self-sufficient in transporting the passenger domain (stable units of protein structure that could fold autonomously) through the outer membrane present only in Gram-negative bacteria. The autotransporter structure consists of three functional domains: the amino-terminal leader

sequence (N-terminal), the secreted mature protein (passenger domain) and a carboxyl terminal (C-terminal) (Henderson *et al*., 1998) as depicted in Figure 2.2.

Figure 2.2 Autotransporter structure (adopted with modifications from Junker et al., 2006)

Virulence factors production and their interaction to the cell, as well as bacteria growth rate, are influenced by the cultivation conditions for the bacterium and the composition of the employed in the fermentation (Thalen, 2008). Small changes in the conditions, such as changes in composition, pH, temperature, and pressure can alter cell metabolism and have a significant impact in the subsequent purification steps and therefore, in productivity of the protein product (Vojinović *et al*., 2006).

In the manufacture of acellular vaccines, not only does the production and selection of antigens play an important role, but also the selection of components that improve the efficacy of vaccines, such as the vaccine adjuvants.

2.2 Vaccine adjuvants

Adjuvants are components that potentiate the immune response to an antigen. Aluminium adjuvants are the most widely used adjuvants in both human and veterinary vaccines (Lindblad, 2004). These adjuvants include: aluminum hydroxide, aluminum hydroxyphosphate sulfate, and aluminum

phosphate (Zhao *et al*., 2001). Although many other adjuvants have been proposed over the years, these have failed to be successful in humans largely because of toxicity, stability and cost (Petrovsky *et al*., 2004).

Adjuvants are used in vaccine formulations for many reasons: (i) to improve the efficacy of vaccines; (ii) to decrease the quantity of antigen needed; (iii) to modulate immune response; or (iv) as antigen delivery systems (Petrovsky *et al*., 2004).

Aluminum adjuvants are commonly employed in Diphtheria, Tetanus, Pertussis and Hepatitis A and B vaccines. It enhances antibody responses; however, its mechanism of action is still unknown. It has been proposed that alum acts as a delivery system by generating depots that trap antigens, providing slow release in order to allow a longer exposure of the antigen, and therefore the stimulation of the immune system (Tritto *et al*., 2009).

Aluminum adjuvants have a high affinity to phospho-containing species, both organic and inorganic. Since many antigens contain phosphoryl groups, they can bind to the aluminum-containing adjuvant. This ability has been denoted by some authors as "phosphophilicity", and its study has gained a broad interest in vaccine development as well as in vaccine and adjuvant manufacturing (Zhao *et al*., 2001).

2.3 Process Monitoring and Control

The monitoring and control of processes have a crucial role in all industries to maintain maximum efficiency and desired product quality. For instance, in pharmaceutical processes, e.g. vaccine production, the cultivation of cells requires among other factors, complex media composition, specialized bioreactor design and the control of various parameters in restrict limits to achieve high

productivity. Therefore, efforts in sensor design, sampling strategies, and data management have been made to improve process monitoring (Pohlscheidt *et al*., 2013).

2.3.1 Methods of Monitoring

Depending on the location where the samples are analyzed, methods for analysis can be classified in the following categories:

2.3.1.1 Off-Line Monitoring

In off-line monitoring, samples are withdrawn from the process and analyzed in a suitable device (Elshereef, 2009; Pohlscheidt *et al*., 2013). By doing sampling, it is possible to evaluate the overall state of the process at a certain time. Even though this approach allows high precision measurements, off-line analytics always has a delay between the sampling and the measurements (Vojinović *et al*., 2006). Handling and preparation of the sample can have a significant impact in the accuracy of the measurement. For instance, aseptic treatment and mixing after sampling are important aspects to be considered before analysis (Pohlscheidt *et al*., 2013).

Traditional laboratory techniques are usually implemented in off-line monitoring (Elshereef, 2009). For process control, measuring of metabolites are essential to maintain, characterize and further improve fermentation processes. Methods such as enzyme linked immunosorbent assay (ELISA) and high performance liquid chromatography (HPLC) are conventionally used in the analysis of specific proteins and substrates. Enzyme-based biosensors, e.g. bioprofiles, are usually employed to measure glucose, lactate, glutamine, and glutamate as off-line (Pohlscheidt *et al*., 2013).

2.3.1.2 *In Situ* or In-Line Monitoring

In this type of monitoring, sensors are directly positioned in the vessel or in the flow line, and no withdrawal of sample is needed. Chemical and physical parameters, such as pH, redox potential, O_2, CO_2, conductivity, and turbidity, are usually monitored *in situ*. The advantage of *in situ* monitoring over offline monitoring, is the rapid measurement and direct control (Pohlscheidt *et al*., 2013).

This type of monitoring is sometimes called on-line monitoring. Some authors make a differentiation between these two terms, naming on-line monitoring when a sample is automatically withdrawn and analyzed.

2.3.1.2.1 *In Situ* Process Monitoring Using Spectrometry

Several technologies are being developed for real-time estimation of nutrient levels. Technologies including Raman spectroscopy, infrared spectroscopy, UV spectroscopy, and fluorescence spectroscopy can monitor a wide array of species, including glucose, lactate, various amino acids, and NAD(P)H. (Pohlscheidt *et al*., 2013).

Fluorescence spectroscopy (also known as fluorometry or spectrofluorometry) is a type of electromagnetic spectroscopy which analyzes fluorescence from a sample. It involves using a beam of light, usually ultraviolet light, that excites the electrons in molecules of certain compounds and causes them to emit light.

The key characteristic of fluorescence spectrometry is its high sensitivity. Fluorometry may achieve limits of detection several orders of magnitude lower than those of most other techniques. Limits of detection of 10^{-10} M or lower are possible for intensely fluorescent molecules. Because of the low

detection limits, fluorescence is widely used for quantification of trace constituents of biological and environmental samples. Fluorometry is also used as a detection method in separation techniques, especially liquid chromatography and electrophoresis (Wehry, 1997).

Interpretation of fluorescence spectra sometimes can be difficult due to the overlapping spectra of multiple species. Methods for data compression such as principal component analysis (PCA) and partial least squares (PLS) are commonly used to identify correlations among fluorescence spectra and to reduce them into small number of orthogonal variables (Pohlscheidt *et al*., 2013).

2.4 Principles of Fluorescence Spectroscopy

Fluorescence is a type of luminescence in which molecules emit light from electronically excited states generated by a physical (e.g. absorption of light), mechanical (friction), or chemical mechanism. When luminescence is generated through excitation by ultraviolet or visible light, the phenomenon is termed photoluminescence, and is formally divided into two categories: fluorescence and phosphorescence, depending on the electronic configuration of the excited states and the emission pathways (Lakowicz, 2006; Wehry, 2007).

2.4.1 Fluorophores

Fluorophores are compounds that can re-emit light upon excitation. Fluorescent compounds have two characteristic spectra: an excitation spectrum (amount of light absorbed) and an emission spectrum (amount of light emitted).

Fluorophores are divided into two classes: intrinsic, which occur naturally, and extrinsic, which are added to the sample. In proteins, the intrinsic fluorescence originates from the three aromatic amino

acids: tryptophan (Trp), tyrosine (Tyr), and phenylalanine (Phe). The indole ring (Figure 2.3) of tryptophan is the most dominant fluorophores in proteins (Lakowicz, 2007).

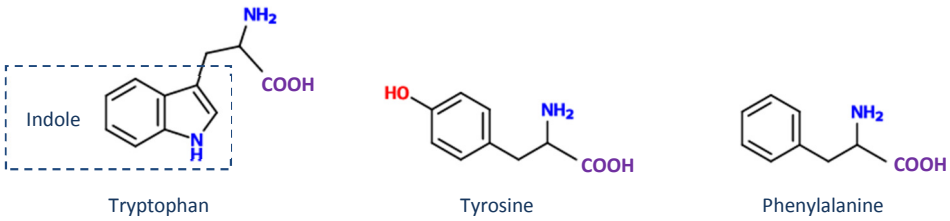

Figure 2.3 Fluorescent amino acids (adopted with modifications from Lakowicz, 2007)

Tryptophan residues inside the protein can be found in different locations (buried in nonpolar regions or in the surface), therefore they are not influenced by the environment in the same manner (Burstein *et al*., 1973). Depending on the location, tryptophan residues can emit from 308 to 352 nm. When the Trp residues are buried in non-polar regions of the protein, less energy is lost to the environment and a blue-shift in the emission spectra is observed. As the residues are located closer to the surface and more exposed to the environment, the emission spectra is shifted to higher wavelengths (red-shift), reaching the maximum emission when the protein is unfolded (Burstein *et al*., 1973; Lakowicz, 2007).

The emission wavelengths of phenylalanine overlap with the absorption wavelengths of tyrosine. In turn, the emission wavelengths of tyrosine overlap with the emission of tryptophan (Devlin, 2006). Because of these overlaps, energy absorbed by phenylalanine and tyrosine is transferred to tryptophan, and only the emission of tryptophan is observed in proteins (Devlin, 2006; Lakowicz, 2007).

Energy transfer occurs when distance between donor and acceptor are up to 80 Å. When a protein is denatured, this distance increases and the energy transfer among fluorophores is inefficient (Devlin, 2006).

Tryptophan, tyrosine and phenylalanine absorb UV light at two different wavelengths. The peak at excitation next to 280 nm is the result of the absorbance by the aromatic ring portion of their structure. The peak at lower wavelengths is caused by absorbance of peptide and carboxylic acid moieties in the compounds (Figure 2.3) (Held, 2013).

In protein fluorescence analysis, the intensity values under excitation (Ex) 250 nm, usually are not taken into account if the experiment is done in aerial conditions. Since oxygen absorbs at this range, dissolved oxygen should be removed before any fluorescence measurement by passing inert gas through the sample.

In Table 2.1, excitation-emission values of aromatic amino acids in water are shown:

Amino acid	**λ_{ex} (nm)**	**λ_{em} (nm)**
Phenylalanine	260	282
Tyrosine	275	304
Tryptophan	295	353

Table 2.1 Excitation-emission values of aromatic amino acids in water at neutral pH (Lakowicz, 2007)

Tryptophan can also have its maximum fluorescence peak at Ex=280 nm (Lakowicz, 2007).

Among biological molecules, fluorescence is observed in co-factors such as reduced nicotinamide adenine dinucleotide (NADH), from oxidized flavins (FAD, the adenine dinucleotide, and FMN, the mononucleotide) (Lakowicz, 2007). NADH absorbs light at two excitation wavelengths: 260 nm and 340 nm, with an emission wavelength at 460 nm. The excitation at 260 nm is due to the adenine group and the peak at 340 nm corresponds to the absorption by the pyridine ring (Figure 2.7) (Rover *et al*., 1997). Also, NADPH has the same Ex/Em as NADH (Lakowicz, 2007).

Flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD) absorb light at Ex~450 nm and emit at Em~525 nm (Lakowicz, 2007).

The fluorescence properties of these cofactors have been employed to study their binding to proteins. When bound to an enzyme, NAD(P)H is in an extended conformation, which prevents quenching of the pyridine ring by the adenine group (Figure 2.4). As a result, the emission intensity is higher in this case compared when free in solution. Also, due to the binding, there is a quenching of tryptophan present in the protein (Lakowicz, 2007). A binding enzyme-cofactor also results in a shift in the emission of NAD(P)H from 260 nm to 280 nm (Gazzotti *et al*., 1974). In contrast to NAD(P)H, emissions from FAD are usually quenched due to binding to proteins (Lakowicz, 2007).

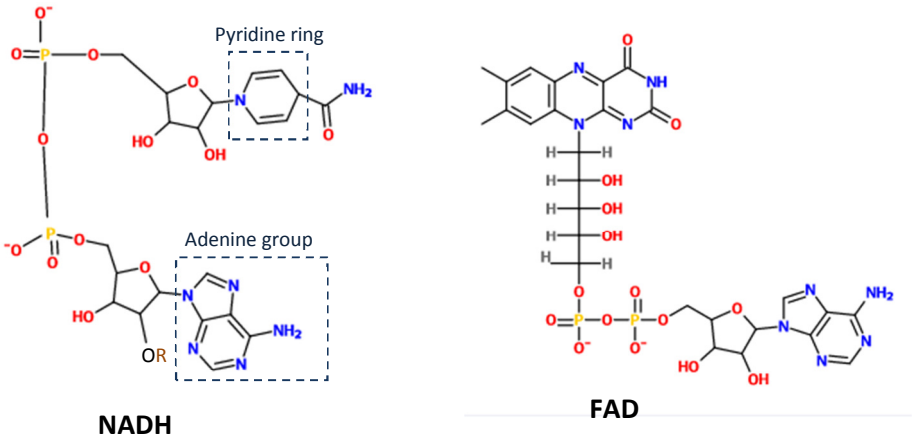

Figure 2.4 Biochemical fluorophores. R is a hydrogen in NADH, and a phosphate group in NADPH (adopted with modifications from Lakowicz, 2007)

Extrinsic fluorophores can be used to measure enzyme activities. The simplest approaches are those in which a fluorescent substrate is converted into a nonfluorescent product, or a nonfluorescent substrate is converted into a fluorescent product after enzymatic cleavage. The enzyme activity is monitored by following the decrease or the increase in fluorescence intensities respectively (Deshpande, 2001).

2.4.2 Theory of Fluorescence

When a fluorophore is irradiated by a light source with a wide spectrum of wavelengths, it will absorb light and reach different vibrational and electronic states. The probability, in which these states can

occur, depends on the electronic configuration of the molecule. This configuration is unique to each fluorophore, which made each fluorescence spectra also unique.

The process, in which the absorption and emission of light can occur, is illustrated by a Jablonski diagram. A simplified Jablonski diagram is shown in Figure 2.5-A. The ground, first and second electronic states are represented by S0, S1 and S2, respectively. The vibrational energy levels, in which a fluorophore might exist, are indicated by 0, 1, 2, 3, and 4. The vertical lines represent the transitions among states during absorption (green lines) and fluorescence emission (blue lines).

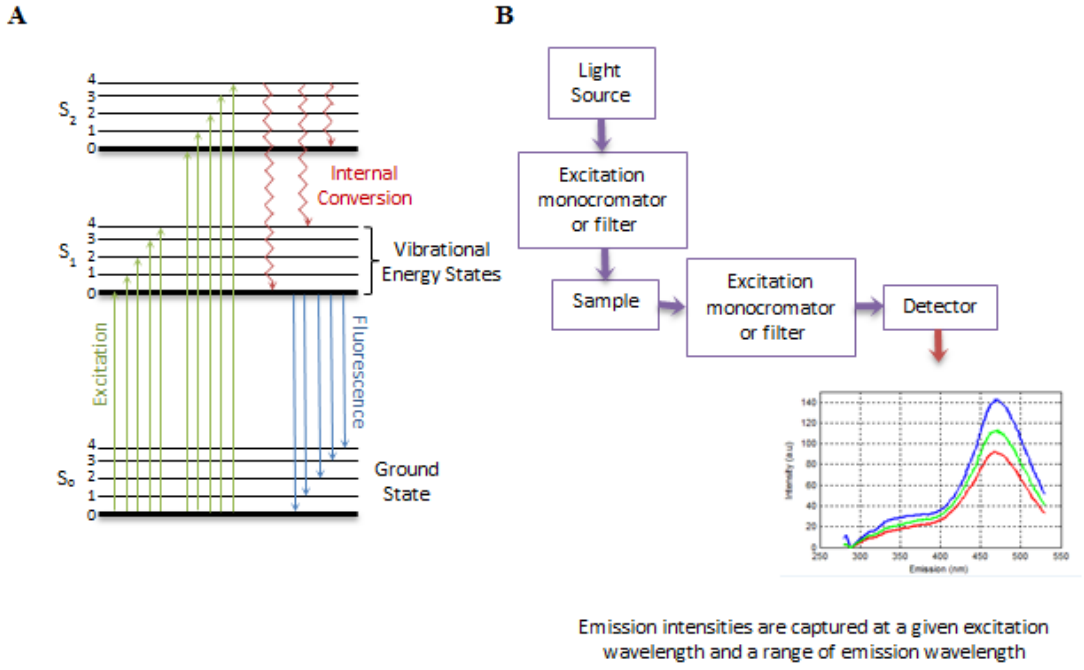

Figure 2.5 Jablonski diagram; B: Simplified operating diagram for a spectrofluorometer (adopted with modifications from Peiris, 2010).

A simplified operating diagram of a spectrofluorometer is shown in Figure 2.5-B. A spectrofluorometer is an instrument that allows measuring quantitatively fluorescence from a sample. It consists of a light source coupled to an excitation monochromator that irradiates the sample in a cuvette. Light emitted by the sample is then collected by an emission monochromator which illuminates a detector (Longworth, 1968).

2.4.3 Variables of Fluorescence Measurements

Fluorescence intensities are typically directly proportional to concentration. There are, however, variables that affect this linearity such as quenching, the molecular environment of fluorophores and light scattering.

2.4.3.1 Quenching

This term refers to factors that reduce intensities of fluorescence:

- High concentration. When concentration is too high, the light cannot pass through the sample to cause excitation; thus very high concentrations can have very low fluorescence.
- Collisions with other molecules. It results in the loss of excitation energy as heat instead of as emitted light.
- Formation of a complex between the fluorophore and the quencher (static quenching). A special case of static quenching is self-quenching, where fluorophore and quencher are the same species. Self-quenching is particularly evident in concentrated solutions.
- Resonance energy transfer. In this case, the two participating molecules do not have to collide (it can occur over distances of 100Å). When a molecule is excited, this excitation is

transferred to another molecule, without emitting light. As a result, the acceptor remains excited while the donor stays in the ground state.

2.4.3.2 Molecular Environment

The fluorescent amino acids present in protein consist of hydrophobic amino acid residues, which are sensitive to the local environment. Due to this, changes in the emission can be detected depending on the exposure of the amino acid residues to the environment, and therefore information about denaturation or protein conformation can be obtained from fluorescence measurements (Lakowicz, 2007)

2.4.3.3 Light Scattering

The light scattering effects are called Rayleigh scattering (first and second order) and Raman scattering. The first order Rayleigh scattering is observed when the emission wavelength equals the excitation wavelength ($\lambda_{em}=\lambda_{ex}$). In second order Rayleigh scattering, the emission wavelength equals twice the excitation wavelength ($\lambda_{em}=2\lambda_{ex}$), and the Raman scattering dependent upon the solvent of the sample (Rinnan *et al*., 2005).

Rayleigh scattering might interference in the fluorescence emission of fluorophores (Christensen *et al*., 2006). Since it is difficult to eliminate such as interference, mathematical corrections can be made to analyse the fluorescence data. Due to its weak contribution, Raman scattering is dismissed (Christensen et al., 2006).

2.5 Multivariate Analysis of Fluorescence Excitation-Emission Matrices

Multivariate analysis is the application of multivariate statistics to analyze data that arises from more than one variable. The most common data analytical approaches are Principal Component Analysis (PCA), used for exploring, visualizing and reducing data dimension, and Partial Least Square regression (PLS), used to construct models that allow predicting a certain variable from a measured spectra (Christensen *et al*., 2006).

2.5.1 Pre-processing of data

Prior to performing any multivariate analysis, the raw data must be pre-processed to ensure that the data are symmetrically distributed. The common methods to pre-process data are the following:

- Mean centering: calculates the mean of each variable in a set of data and subtracts the mean from each variable to make the new mean equal to zero.
- Median centering: similar to mean centering. In this case, the median is used instead of the mean.
- Auto-scaling: variables are mean centered and scaled to unit variance

2.5.2 Principal Component Analysis (PCA)

PCA is one of the methods employed for data compression and extraction of the most important information from a data set (Abdi *et al*., 2010). PCA reduces the original variables into a lower number of orthogonal (non-correlated) variables (PC's).

These PC's are linear combinations of the original variables (e.g. fluorescence intensities) that capture the variability in the data describe major trends in the original data. Only a subset of the PC's that

describe a significant percentage of the variability is kept for further analysis whereas the remaining PC's are discarded on the assumption that they correspond to noise.

A graphic example of how PCA works is shown in Figure 2.9. Having a 3D space with variables X_1, X_2 and X_3, the goal is to find a vector that captures the maximum variability (or minimizes the loss of information) in the cloud of points (e.g. **u** in Figure 2.9)

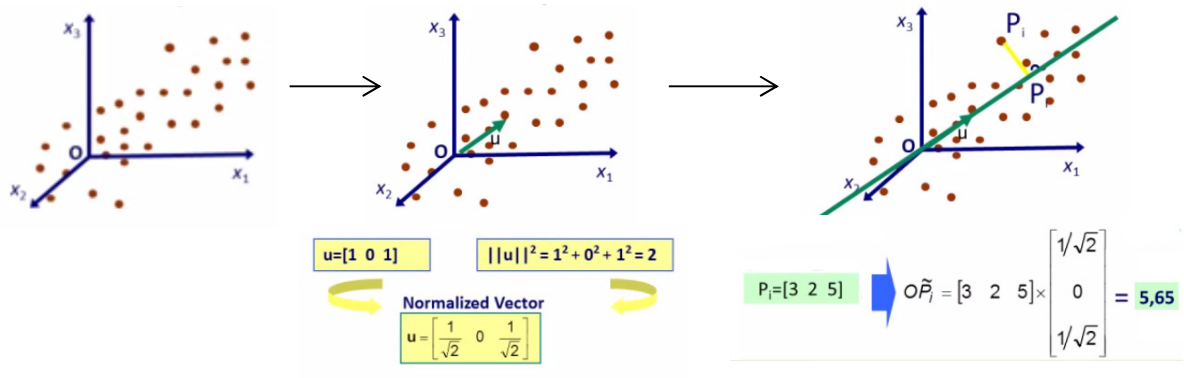

Figure 2.6 Projection of a point on an axis

One can project any point on the vector (e.g. P_i) by multiplying the three coordinates of the point by the normalized vector **u**. In this sense, a point of three dimensions is reduced to one dimension. In PCA, the goal is to maintain the distance between the points and the new vector (PC) as small as possible:

$$\sum_i \left(P_i - \tilde{P}_i\right)^2 = min \tag{2.1}$$

PCA is an iterative process. When the first PC is found, the remaining variability should be explained by another PC. As a constraint, the PC's have to be orthogonal (Figure 2.10)

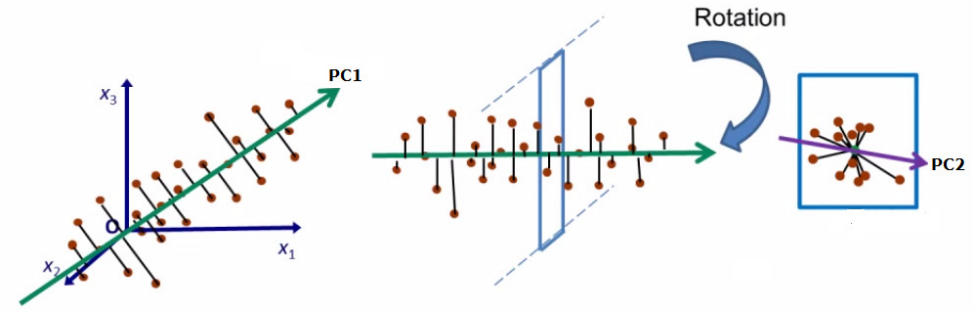

Figure 2.7 Plane rotations in PCA

For finding the second PC, the method rotates the axes of the first PC to find an orthogonal plane, and in this manner find the second PC that captures the remaining variability.

PCA decomposes a set of data (matrix **X**) as the sum of the products of vectors $\mathbf{s_i}$ and $\mathbf{p_i}$, plus a residual matrix E (Equation 3.1).

$$X = \sum_{i=1}^{n} s_i * p_i + E \tag{2.2}$$

Where **n** is the number of samples. The $\mathbf{s_i}$ vectors are known as ***scores*** and contain information on how the ***samples*** relate to the principal components, and can be interpreted as the projections of the original variables on the new space. The $\mathbf{p_i}$ vectors are known as ***loadings*** and contain information on how the original ***variables*** relate to each other along the principal components.

While PCA performs data compression on the input data only, i.e. fluorescence intensities, other multivariate statistical methods are needed to correlate the input data to other measured data. Partial Least Squares (PLS) is a multivariate regression method that correlates the information in one data matrix (X) to the information in another matrix (Y). PLS is used to obtain a regression model between X to Y so that newly acquired values of X could be used to predict Y.

2.5.2 Partial Least Square (PLS)

PLS performs data compression and extracts a small number of latent (not directly measured) variables (LVs) that describes the observed or measured data. In contrast to PCA, PLS not only finds the major variations in X and Y, but identifies the latent variables that correlate the scores in the X matrix with the scores in the Y matrix to have maximum covariance.

In this sense, PLS decomposes X and Y matrix as follows:

$$X = TQ + E = \sum_{i=1}^{m} t_i * q_i + E \tag{2.3}$$

$$Y = UC + F = \sum_{i=1}^{m} u_i * c_i + F \tag{2.4}$$

Where **m** is the number of latent variables. The $\mathbf{t_i}$ and $\mathbf{u_i}$ vectors are the ***scores*** and contain information on how the ***samples*** relate to the latent variables. Scores in **T** are the projections of the input data **X** to a LV in the X-space, and scores in **U** are the projections of the response data to a LV in the Y-space. The $\mathbf{q_i}$ and $\mathbf{c_i}$ vectors are known as ***loadings*** and represent the relationship between T and X-space, and U and Y-space respectively. **E** and **F** contain the variables left unexplained by the model.

The inner relation between **U** and **T** is given by the following equation:

$$U = WT \tag{2.5}$$

Where **W** is the weight matrix. From equation 2.5, it is possible to predict the scores in Y from scores in X, and then the new response data from equation 2.4.

Chapter 3
Materials and Methods

In this chapter, the materials and methods are organized according to the three main analyses that were performed in this thesis: fluorescence data acquisition and analysis various steps of the upstream and downstream pertactin purification process, enzyme assay and surface phosphophilicity assay.

3.1 Fluorescence analysis of the upstream and downstream pertactin purification process

3.1.1 Samples

Samples were provided by Sanofi Pasteur (Toronto, Ontario, Canada) from upstream and downstream steps for different batches conducted during the period shown in Table 3.2. For the upstream part of the process, there are two parallel fermenters (F3 and F6 in Figure 1.1) that are combined together before downstream processing. For the fluorescence analysis, a total of four samples were collected for each upstream fermentation step (F3-S, F6-S, CPS-B2-(F3), CPS-B2-(F6)) corresponding to samples from each fermenter and their corresponding centrifuged samples. Samples were collected twice per week since fermentations are conducted every 3.5 days. Thus a total of 8 samples were collected every week.

As shown in the process flow diagram (Figure 1.1), CPS-B2-(F3) and CPS-B2-(F6) samples are combined and filtered. The two resulting batches of broth are combined together and sent to the downstream purification process. From the ultrafiltration step, there are four samples (two for CPS-Conc and two for CPS-Conc-PW (permeate waste)). From the perlite chromatography column, four samples are collected, two for SPEP-R-Conc and two for SPEP-R-Conc-PW (permeate waste) (Figure

1.1). After the first ammonium sulphate precipitation (two SKA-S1 samples and two SKA-P1-R samples), the two pellets (SKA-P1-R) are combined and precipitated again with ammonium sulphate at 10% w/v. After the third ammonium sulphate precipitation at 20% w/v, the samples analyzed are: SKA-S3, SKA-P4, SKA-CH1, UF/D1, SKA-UF/D2, SKA-CH2. In total, 25 samples for each batch were analyzed as summarized in Table 3.1.

Step	Step ID	Number of Samples
Production Fermentation	F3-S	2
	F6-S	2
Centrate	CPS-B2 (F3)	2
	CPS-B2 (F6)	2
First Concentration/Ultrafiltration	CPS-Conc	2
	CPS-Conc-PW	2
Second Concentration/Ultrafiltration	SPEP-R-Conc	2
	SPEP-R-Conc-PW	2
First Fractionation	SKA-S1	2
	SKA-P1-R	1
Third Fractionation	SKA-S3	1
	SKA-P4	1
Chromatography Column	SKA-CH1	1
First Ultrafiltration/Diafiltration	UF/D1	1
Second Ultrafiltration/Diafiltration	SKA-UF/D2	1
Chromatography Column	SKA-CH2	1
	Total	**25**

Table 3.1 Steps in the upstream and downstream process along with Step ID and number of samples per batch

A flow diagram of the samples analyzed per batch is provided in Figure 3.1.

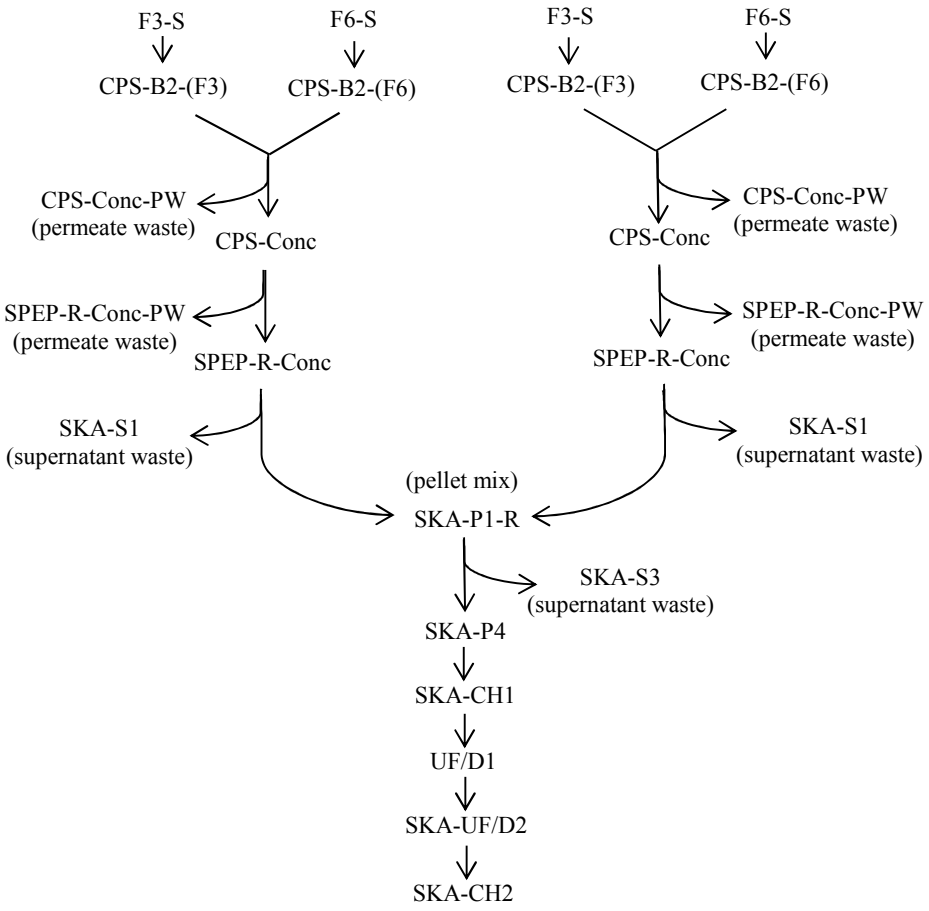

Figure 3.1 Flow diagram of samples analyzed per batch

Seven different batches were analyzed: A, B, C, D, E, F, G (Table 3.2). These samples were initially stored in a freezer at -20ºC. After thawing, all samples were diluted prior to fluorescence analysis. The upstream fermentation samples (CPS-Conc, CPS-Conc-PW), were diluted with 0.9% saline

buffer (0.15 mM NaCl). SPEP-R-Conc and SPEP-R-Conc-PW were diluted with 50 mM Tris-HCl buffer, and the rest of the samples with 10 mM Tris-HCl buffer. Tris-HCl buffer was prepared with Trizma base (Tris-(hydroxymethyl)-aminomethane) with pH adjusted to 8.0 using HCl. The dilution factor was 0.05.

Table 3.2 shows final Kjeldahl values for each batch in the last step of the PRN purification (SKA-F). The "PRN batch start year" is the year when the batch was obtained. The "Pass-thru" are samples from CPS-Conc to SKA-S1 (Table 3.1).

Batch	PRN batch year	Fermenters	Pass-thru	Kjeldahl (g N/batch) in SKA-F
A	2013	A1	A3	0.43
		A2	A4	
B	2013	B1	B3	0.49
		B2	B4	
C	2012	C1	C3	0.56
		C2	C4	
D	2013	D1	D3	0.70
		D2	D4	
E	2012	E1	E3	0.73
		E2	E4	
F	2013	F1	F3	0.89
		F2	F4	
G	2013	G1	G3	1.00
		G2	G4	

Table 3.2 Kjeldahl values per batch

The Kjeldahl method is a procedure for quantitative determination of proteins based on the nitrogen content present in a sample. This method is composed of three main steps: digestion, distillation and titration. The digestion consists of the decomposition of the nitrogen in the sample by boiling with an

excess of concentrated sulfuric acid solution. In this manner, the organic nitrogen is converted to ammonium salt. The digestion ends when the sample, initially dark-colored, becomes clear and colorless. The sample is then distilled with an excess of a concentrated sodium hydroxide solution, which converts the ammonium salt into ammonia. The latter is condensed and reacted with an excess of boric acid. As a result, ammonium borate is formed. The last step consists of the titration of the ammonium borate with hydrochloric acid. The concentration of hydrogen ions determined in this step equals the concentration of nitrogen in the sample.

The Kjeldahl analysis used to determine the values shown in Table 3.2, were conducted at Sanofi.

3.1.2 Acquisition of Fluorescence Spectra

Fluorescence excitation-emission spectra were acquired with a Cary Eclipse Fluorescence Spectrophotometer from Agilent Technologies. The samples were analyzed using quartz cuvettes using a slit width of 5 nm and PMT of 600 V. These parameters were chosen based on the spectra acquired from samples at different dilutions and at different values of slit widths and PMT. Good signal, reproducibility, inexistence of quenching due to concentration were the aspects taken into account for selecting those parameters. The excitation and emission ranges were 210-470 nm and 260-550 nm respectively. A wide range of Ex/Em values was used to ensure that all fluorophores were captured.

3.1.3 PCA and PLS Analysis

The fluorescence spectra acquired for each sample consisted of 8148 intensities obtained for different excitation/emission combinations (28 x 291 data matrix). The intensity values for each sample were rearranged into one single row, as describe in Figure 3.3. Thus, each row corresponds to the spectra of

one sample. The rows were stacked together as described in Figure 3.3 to generate one data matrix for PCA and PLS analysis.

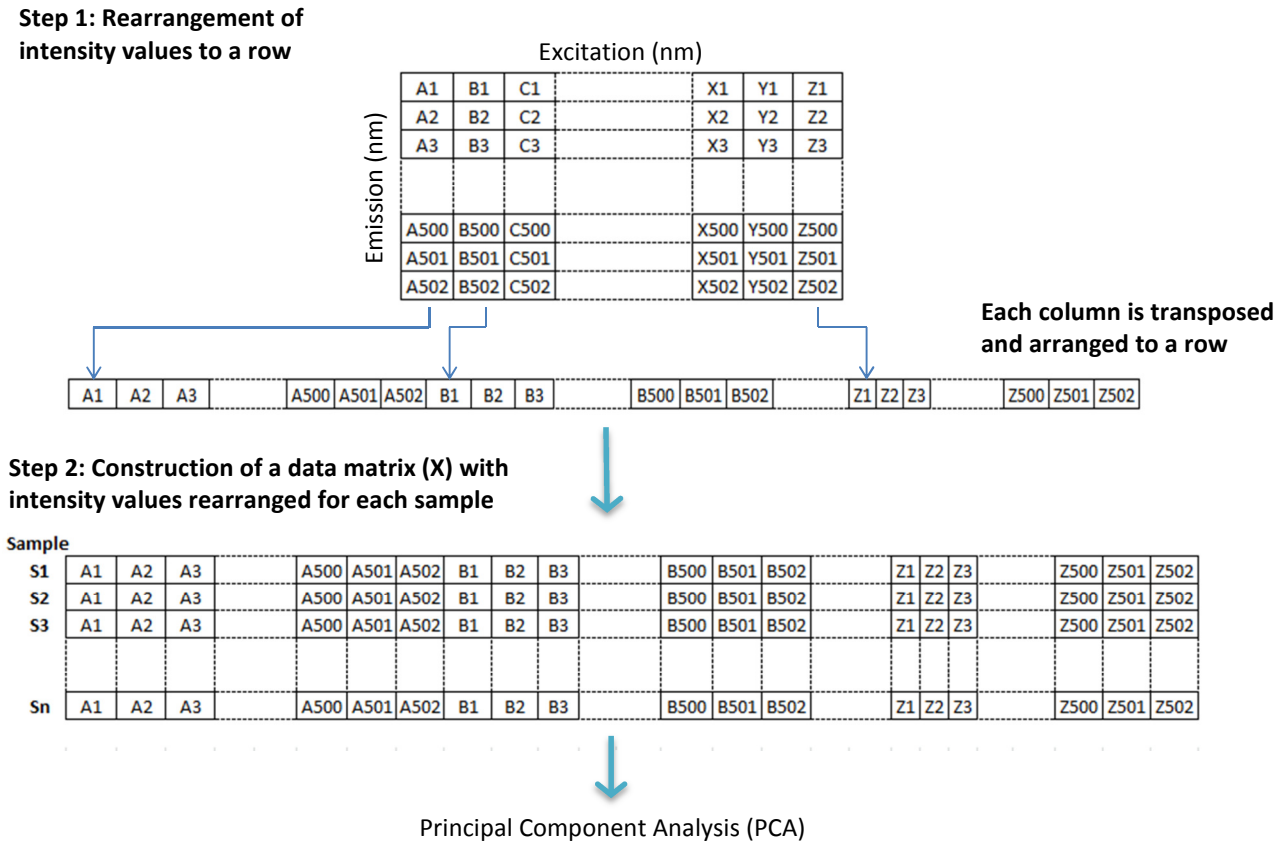

Figure 3.2 Steps involved in generating data rows of intensity values for PCA (adopted with modifications from Peiris, 2010)

To identify whether the variability in productivity originates from a specific step or a combination of steps of the manufacturing process, four main data matrices were generated as follows: (i)- Upstream (with samples from CPS-F3, CPS-F6, CPS-B2-(F3), CPS-B2-(F6), CPS-Conc, CPS-Conc-PW), (ii)-

Downstream 1 (with samples from SPEP-R-Conc, SPEP-R-Conc-PW, SKA-S1, SKA-P1-R, SKA-S3, SKA-P4, SKA-CH1, UF/D1, SKA-UF/D2, SKA-CH2), (iii)- Downstream 2 (with samples from SPEP-R-Conc, SKA-P1-R, SKA-P4, SKA-CH1, UF/D1, SKA-UF/D2, SKA-CH2) and (iv)- Downstream 3 (with samples from SPEP-R-Conc-PW, SKA-S1, SKA-S3). Downstream 1 matrix, which contains all the spectra obtained from samples in the downstream, was separated into two matrices: Downstream 2, which excludes waste samples, and Downstream 3 which includes waste samples.

Even though CPS-Conc and CPS-Conc-PW belong to the downstream process as per the company classification, they were considered together with samples from the upstream data matrix for PCA. They were considered together since up to the centrifugation step, from which CPS samples are obtained, there is a significant amount of antigens and co-enzymes in the samples. Thus, it was expected that by tracking samples where proteins are abundant, higher sensitivity to variability in fermentations outcomes may be obtained.

The four data matrices defined above were analyzed using principal component analysis (PCA). PCA is a multivariate statistical method that reduces the original variables into a lower number of orthogonal (non-correlated) principal variables referred to as principal components (PC) (Wise *et al*., 2006). These PC's are linear combinations of the original variables (fluorescence intensities) that capture the variability in the data. Only a subset of the PC's that describe a significant percentage of the variability is kept for further analysis whereas the remaining PC's are discarded on the assumption that they correspond to noise.

PCA decomposes a set of data (matrix **X)** as the sum of the products of vectors $\mathbf{s_i}$ and $\mathbf{p_i}$, plus a residual matrix E (Equation 3.1).

$$X = \sum_{i=1}^{n} s_i * p_i + E \quad (3.1)$$

Where **n** is the number of samples. The $\mathbf{s_i}$ vectors are known as ***scores*** and contain information on how the ***samples*** relate to the principal components. The $\mathbf{p_i}$ vectors are known as ***loadings*** and contain information on how the original ***variables*** relate to each other along the principal components.

Before doing this multivariate analysis, it is necessary to pre-treat the data by performing mean centering and scaling. By performing mean centering, the average is subtracted from every column, so the new mean of each column is zero. The data is scaled to have unit variance by dividing each column by the standard deviation of the values in that particular column. Without normalization by the variance, the PCA model could wrongly focus on few variables with the largest absolute variances (Wise *et al*., 2006)

To choose the number of components to be kept for further analysis cross-validation was done with the random subset method (Wise *et al*., 2006). This method consists of dividing the data into a number of subsets (N) randomly selected. Then, the model is built based on N-1 subsets, leaving one out. The model is then used to predict the subset left out. The resulting prediction error is then plotted against the number of principal components (PC's), and the number with less error is chosen. Generally, the error is presented as the Root Mean Square Error of Cross-Validation (RMSECV), which is defined as:

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (3.2)$$

Where y_i is the measured value, $\hat{y}_i$ is the predicted variable and n is the number of calibration samples (Wise *et al.,* 2007).

While PCA performs data compression on the input data only, i.e. fluorescence intensities, other multivariate statistical methods are needed to correlate the input data to other measured data. Partial Least Squares (PLS) is a multivariate regression method that correlates the information in one data matrix (X) to the information in another matrix (Y). PLS is used to obtain a regression model between X to Y so that newly acquired values of X could be used to predict Y.

In this project, PLS was used to obtain a regression model between matrix X, corresponding to the fluorescence spectra of pure PRN corresponding to samples of SKA-CH2 (see Fig. 1.1) , and matrix Y corresponding to the final protein content measured by a Kjeldahl test. PLS also requires autoscaling and cross-validation as explained above for the PCA procedure.

PLS performs data compression and extracts a small number of latent (not directly measured) variables (LVs) that describes the observed or measured data. In contrast to PCA, PLS not only finds the major variations in X and Y, but identifies the latent variables that correlate the scores in the X matrix with the scores in the Y matrix to have maximum covariance.

In this sense, PLS decomposes X and Y matrix as follows:

$$X = TQ + E = \sum_{i=1}^{m} t_i * q_i + E \tag{3.3}$$

$$Y = UC + F = \sum_{i=1}^{m} u_i * c_i + F \tag{3.4}$$

Where **m** is the number of latent variables. The $\mathbf{t_i}$ and $\mathbf{u_i}$ vectors are the ***scores*** and contain information on how the ***samples*** relate to the latent variables. Scores in **T** are the projections of the input data **X**, i.e. fluorescence intensities to a LV in the X-space, and scores in **U** are the projections of the response data, i.e. protein content to a LV in the Y-space. The $\mathbf{q_i}$ and $\mathbf{c_i}$ vectors are known as

loadings and represent the relationship between T and X-space, and U and Y-space respectively. **E** and **F** contain the variables left unexplained by the model.

The inner relation between **U** and **T** is given by the following equation

$$U = WT \tag{3.5}$$

Where **W** is the weight matrix. From equation 3.5, it is possible to predict the scores in Y from scores in X, and then the new response data from equation 3.4.

To predict the response (Y_p) from a new set of data (X_{new}), the T score matrix can be calculated by multiplying X_{new} by the known loadings (Q) from the calibration model. Subsequently, U can be calculated using Equation 3.5. The new response (Y_p) is then calculated by multiplying U by the known loading value C (Equation 3.4).

Another way to obtain the new Y response is by multiplying the known values Q, C and W from the calibration model to obtain one matrix (B) called the regression coefficient matrix. In this manner, Y can be predicted by multiplying the new data (X_{new}) by matrix B:

$$Y = XB \tag{3.6}$$

In this project, PCA and PLS analysis were performed using the PLS_ Matlab Toolbox (Eigenvector Research Inc., Manson,WA). Data collection, analysis and plotting were conducted using MATLAB R2012b software (The Mathworks Inc., Natick, MA).

3.2 Enzyme assay

3.2.1 Materials

The following substrates were obtained from Sigma-Aldrich Co. LLC.: β-nicotinamide adenine dinucleotide, reduced dipotassium salt, sodium pyruvate solution, maleate standard for IC, oxaloacetic acid, α-ketoglutaric acid, iodonitrotetrazolium chloride, CB 1954 or 5-(1-Aziridinyl)-2,4-dinitrobenzamide (substrate for ubiquinone oxidoreductase) . The analyzed samples were provided by Sanofi Pasteur (Toronto, Ontario, Canada) CPS-Conc-PW from batch A (Table 3.2).

The samples were diluted in 0.9% saline buffer (0.15 mM NaCl). A stock solution for each substrate (4 μM) was prepared. The dilution factor for the sample was 0.05. The samples were spiked with the substrates using three different dilution factors: 0.007, 0.013, and 0.033.

3.2.2 Fluorescence Spectra

Fluorescence excitation-emission spectra were acquired with a Cary Eclipse Fluorescence Spectrophotometer from Agilent Technologies. The samples were analyzed in quartz cuvettes. A Slit width of 5 nm and PMT of 600 V were used. The excitation and emission ranges were 210-470 nm and 260-550 nm, respectively. These wide ranges were taken in order to cover the Ex/Em values for all the fluorophores.

3.3 Assessment of the effect of temperature on fluorescence

3.3.1 Materials

SPER-R-Conc samples from batches A3 and A4, CPS-B2-(F3) samples from batches G1 and G2, and a F3 sample from batch J were provided by Sanofi. 0.9% saline buffer (0.15 mM NaCl) was employed for the dilution of the samples. The dilution factor was 0.05.

3.3.2 Temperature Assessment and Fluorescence Measurements

In the first assessment, SPEP-R-Conc samples were left at -30ºC for 24 hrs, and then brought to 20ºC by rapid thawing. Another sample was left at 20ºC for 24 hrs and a third sample was heated at 90ºC for 1 min. Fluorescence spectra were acquired in the three cases. In the second assay, fluorescence spectra for CPS-B2-(F3) samples were measured after cooling at 4ºC and raising the temperature to 25ºC, repeatedly for three days. In the third assay, a F3 sample was subjected to changes of 5ºC for three times (25ºC, 30ºC, 35ºC). Fluorescence spectra were measured in each case. In all cases, the fluorescence signals were acquired as explained in section 3.2.2.

3.4 Centrifugation of Samples

3.4.1 Materials

F3 and F6 samples from batches G1 and G2 were provided by Sanofi. 0.9% saline buffer (0.15 mM NaCl) was employed for the dilution of the samples. The dilution factor was 0.05.

3.4.2 Centrifugation and Fluorescence Measurements

The samples were centrifuged for 5 min at 23,000 x*g*. Fluorescence spectra were acquired before and after centrifugation as explained in section 3.2.2.

3.5 Surface phosphophilicity assay

3.5.1 Materials

The following adsorbed samples onto aluminum phosphate adjuvant were provided by Sanofi Pasteur (Toronto, Ontario, Canada): pertactin (PRN) from batch K (February 15, 2011), difteria toxin (DCA) from batch L (September 12, 2011), tetanus toxin (TCA) from batch M (August 30, 2011).

In the process of adsorption, aluminum phosphate solution is added to the antigens, and then stirred for 4 days, at 16ºC to 24ºC.

Two types of buffer were used: 0.9% saline buffer (0.15 mM NaCl), and PBS (phosphate buffered saline).

Fluorinated fluorogenic substrate 6,8-difluoro-4-methylumbelliferyl phosphate was purchased from Molecular Probes, Inc. Dimethyl sulfoxide DMSO and MOPS buffer were purchased from VWR International LLC.

A 10 mM DiFMUP stock solution was made by dissolving the entire content of the vial (5 mg) in 1.71 mL of DMSO (dimethyl sulfoxide). From this solution, a 500 mM DiFMUP solution in 100 mM Mops buffer, pH 7.0, was prepared.

3.5.2 Fluorescence Spectra

A Synergy 4 Multidetection Microplate Reader from Biotek Instruments, Inc was used to read fluorescence intensities.

Five wells in the microplate were used and filled as follows:

- Well A: 100 μL saline buffer – 20 μL DiFMUP
- Well B: 100 μL PBS – 20 μL DiFMUP
- Well C: 100 μL PRN – 20 μL DiFMUP
- Well D: 100 μL DCA – 20 μL DiFMUP
- Well E: 100 μL TCA – 20 μL DiFMUP

The excitation and emission wavelengths were 358 nm and 450 nm, respectively, corresponding to the characteristic fluorescence values for 6,8-difluoro-4-methylumbelliferone (DiFMU), resulting from DiFMUP hydrolysis.

Chapter 4
Results and Discussion

Proteins consist of one or more chains of amino acid residues. Some amino acids, such as tryptophan, tyrosine and phenylalanine are intrinsically fluorescent. Tryptophan is the most useful in fluorescence analysis since, due to its high quantum yield and high sensitivity to its local environment, can be used for inferring conformational transitions, protein-protein interactions and/or denaturation (Royer *et al*., 2006; Lakowicz *et al*., 2007). Enzymes cofactors, such as NAD(P)H and FAD, are usually fluorescent. Since their quantum yields change upon binding to proteins, fluorescence from these two fluorophores has been extensively used to study their interaction with enzymes (Lakowicz *et al*., 2007).

To monitor the PRN manufacturing process, reproducible fluorescence spectra were obtained in order to track changes of the aforementioned fluorophores that could elucidate possible causes of variability in the final product yield.

4.1 Fluorescence Analysis

The main information contained in typical fluorescence EEMs measured for samples analyzed for each batch (Table 3.1) are qualitatively interpreted in this section. Explanations are given for the spectra obtained at different intermediate steps of the manufacturing process to elucidate possible sources of process variability and correlations of the observations with the protein concentration measured and collected at the end of the manufacturing process. The different steps of the manufacturing process are referred to by the notation in Figure 1.1 which describes the process.

In general, the spectra for all samples analyzed exhibited two peaks (α, α') related to tryptophan (Trp) at excitation wavelengths of (Ex) 230 nm and 280 nm, and emission wavelengths (Em) from 330 nm to 350 nm. The wavelengths at which tryptophan exhibits significant emission intensities depend on the location of the amino acid residue in the protein. When the Trp residues are buried in non-polar regions of the protein, a blue-shift is observed. As the residues are located closer to the surface and more exposed to the environment, the emission intensities occur at higher wavelengths (red-shift) (Burstein *et al*., 1973). Emissions for free tryptophan in solution are observed around 350-360 nm. Excitation of Trp at 230 nm is caused by absorbance of its carboxylic moiety and excitation at 280 nm is the result of absorbance by the indole ring. Another two peaks (β, β') associated with NAD(P)H at excitations 280 nm and 340 nm, and emission 460 nm are observed in the measured spectra. Typical excitations wavelengths for NAD(P)H are occurring at 260 nm and 340-360 nm, corresponding to the absorption by the adenine and pyridine rings, respectively (Rover *et al*., 1998). Binding of NAD(P)H to other proteins can cause a shift in its excitation wavelength from 260 nm to 280 nm. This shift is observed in the spectra obtained from early steps in the manufacturing process, and it is believed to be due to the binding of NAD(P)H to a dehydrogenase (Gazzotti *et al*., 1974; Lakowicz *et al*., 2007).

In the fluorescence spectrum for component pertussis (CP) broth, a peak (γ) attributed to amino acids appears at Ex/Em~330/410 nm, and strong peaks (δ, δ') at Ex/Em~230-275/310 appear as the apparent emission by tyrosine (Tyr).

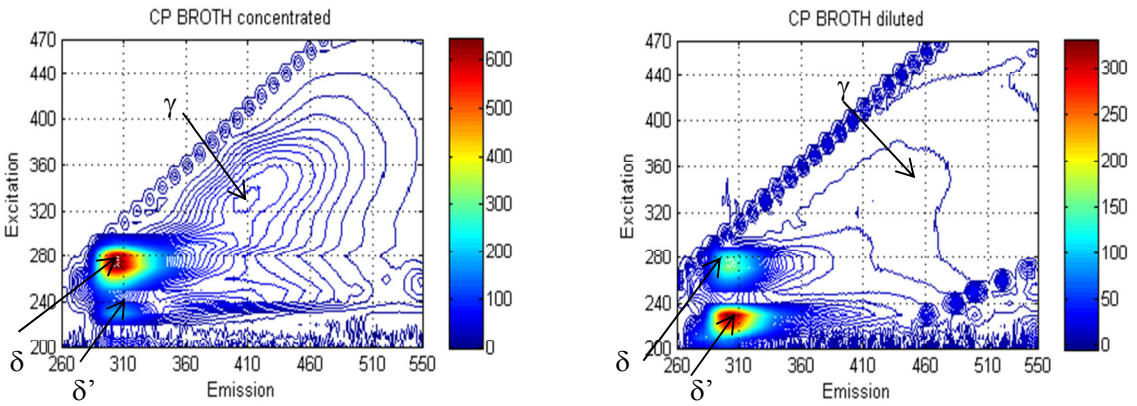

Figure 4.1 EEMs for CP broth (concentrated and diluted).

There are two spectra (Fig. 4.1) for the same broth, but one for concentrated broth and the other for diluted (the dilution factor was 0.05). In the spectrum for concentrate broth, the tyrosine peak at Ex~275 nm is stronger than the peak at Ex~230 nm, even though the coefficient of absorption (ε) is higher at lower wavelengths (ε~9000 M^{-1} cm^{-1} at Ex~220 nm and ε~1400 M^{-1} cm^{-1} at Ex~275 nm in water) (Creed, 1983). This is likely due to energy transfer that occurs at higher concentrations between molecules resulting in higher emission at Ex~275 nm. On the other hand, in the spectra shown for the diluted broth, the tyrosine peak at Ex~230 nm appears with higher intensity. In this case, energy transfer is not efficient since the sample is diluted and contact between molecules is reduced. These differences in the emission also can be observed for tryptophan which shows a principal absorption maxima at 220 nm (ε~36000 M^{-1} cm^{-1}) and 280 nm (ε~5500 M^{-1} cm^{-1}) (Creed, 1983).

First order Rayleigh scattering (FORS) and second order Rayleigh scattering (SORS) can be indicative of protein aggregation. Raman scattering (RS) emerges from the interaction of light with solvent molecules, and it is noticeable at low concentrations of fluorophores (Deshpande, 2001).

Similar patterns have been observed in the spectra among the seven batches analyzed in this work, but there are differences among steps and among batches with different final protein content as determined by Kjeldahl analysis. Such differences are observed in the emissions wavelengths, in the intensities at a particular excitation wavelengths and shifts in significant excitation wavelengths. The latter shifts are for example related to the fact that tryptophan can emit at an excitation wavelength of 230 nm and 280 nm (peaks for both excitations are observed in all steps). For native PRN, emission of tryptophan arises at 335 nm, but when in solution it occurs at 350 nm (Junker *et al*., 2006).

For the upstream fermentation steps (F3-S, F6-S, CPS-B2-(F3), CPS-B2-(F6)), the intensities were relatively low despite the expected abundance of proteins in the samples, since at this point of the manufacturing process all antigens are present. The low fluorescence intensity is believed to be due to quenching resulting from the presence of suspended particles. A shift in the emission of Trp to ~340 nm is observed suggesting that Trp residues may be exposed on the surface due to unfolding of the protein (α,α'). Figure 4.2 show the spectra obtained for fermenters F3-S and F6-S, and Figure 4.3 show the spectra for CPS-B2-(F3) and CPS-B2-(F6) for batch A.

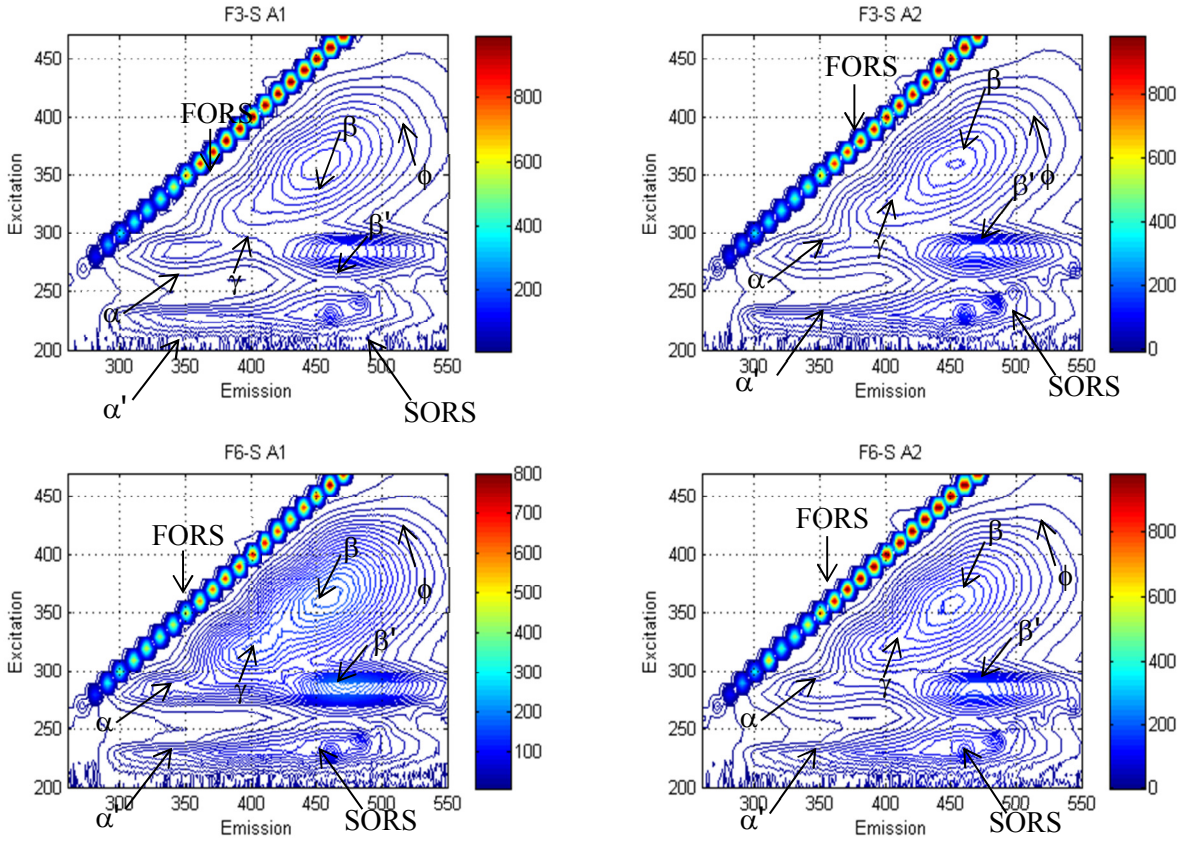

Figure 4.2 EEMs for fermenters F3-S and F6-S for batch A.

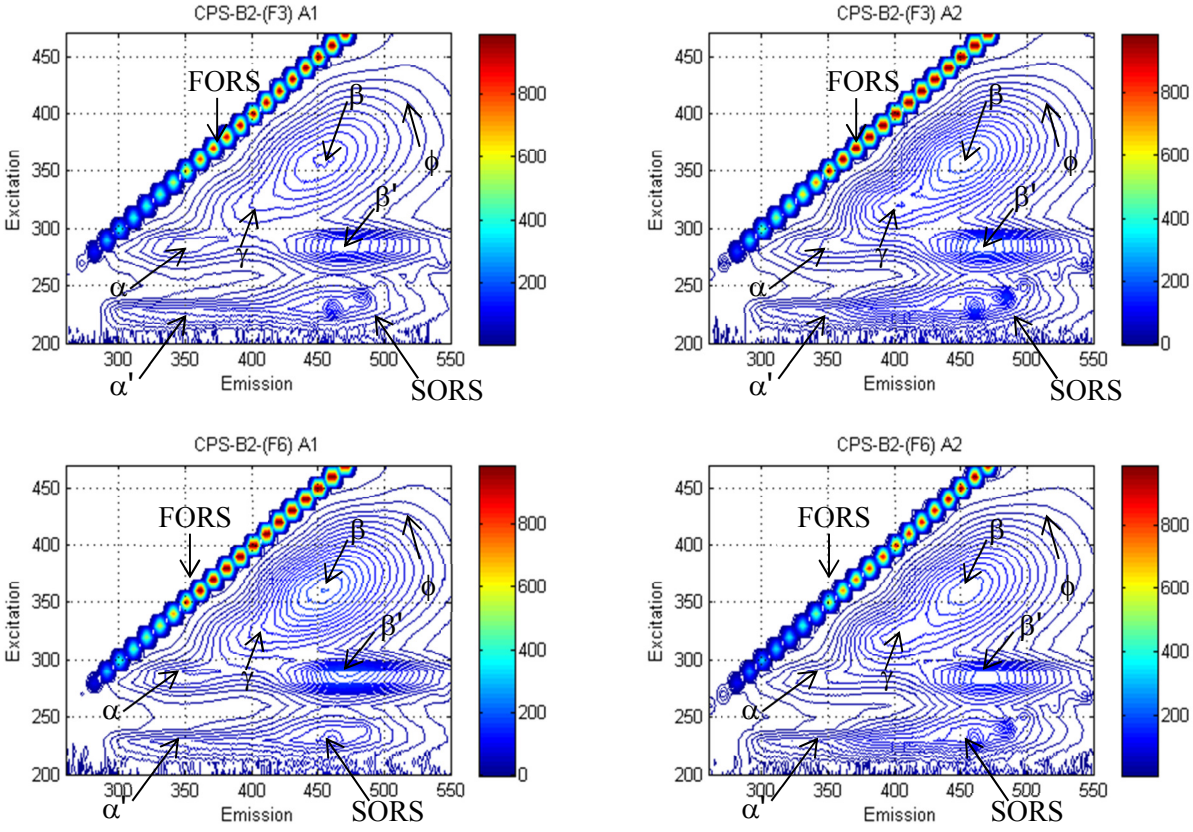

Figure 4.3 EEMs for CPS-B2-(F3) and CPS-B2-(F6) for batch A.

As stated before, an excitation shift from 260 nm to 280 nm for NAD(P)H is observed in the spectra (β'), suggesting a binding with a dehydrogenase. The typical emission of this cofactor at excitation 340 nm is also observed (β). A small shoulder around Ex/Em 330/410 nm (γ) also appears possibly due to the emission by amino acids in the media.

Another peak (ϕ) generally associated to FAD occurs as a shoulder at Ex/Em=450/530 nm. FAD is a product of the TCA cycle for which activity is expected to be significant for aerobic fermentations. At this point, it is not possible to determine with fluorescence whether the emissions originate solely from tryptophan contained in PRN since PT, FHA, FIM among other proteins are also present in the samples. Since FIM does not contain tryptophan so fluorescence from this protein would originate mainly from tyrosine (Crowley-Luke *et al*., 2001).

In the fluorescence spectra of CPS-Conc samples, a shift in the emission of Trp is still observed (α) while emissions for the remaining fluorophores appear without a significant variation. On the other hand, in the EEM for the permeate waste (CPS-Conc-PW), a prominent emission for NAD(P)H-dehydrogenase complex is observed. This peak appears with higher intensity in batches where the final protein content for PRN, measured by Kjeldahl, is low thus possibly suggesting a correlation between the abundance of this complex in the samples collected from the fermentations and the final productivity of pertactin (Figure 4.4).

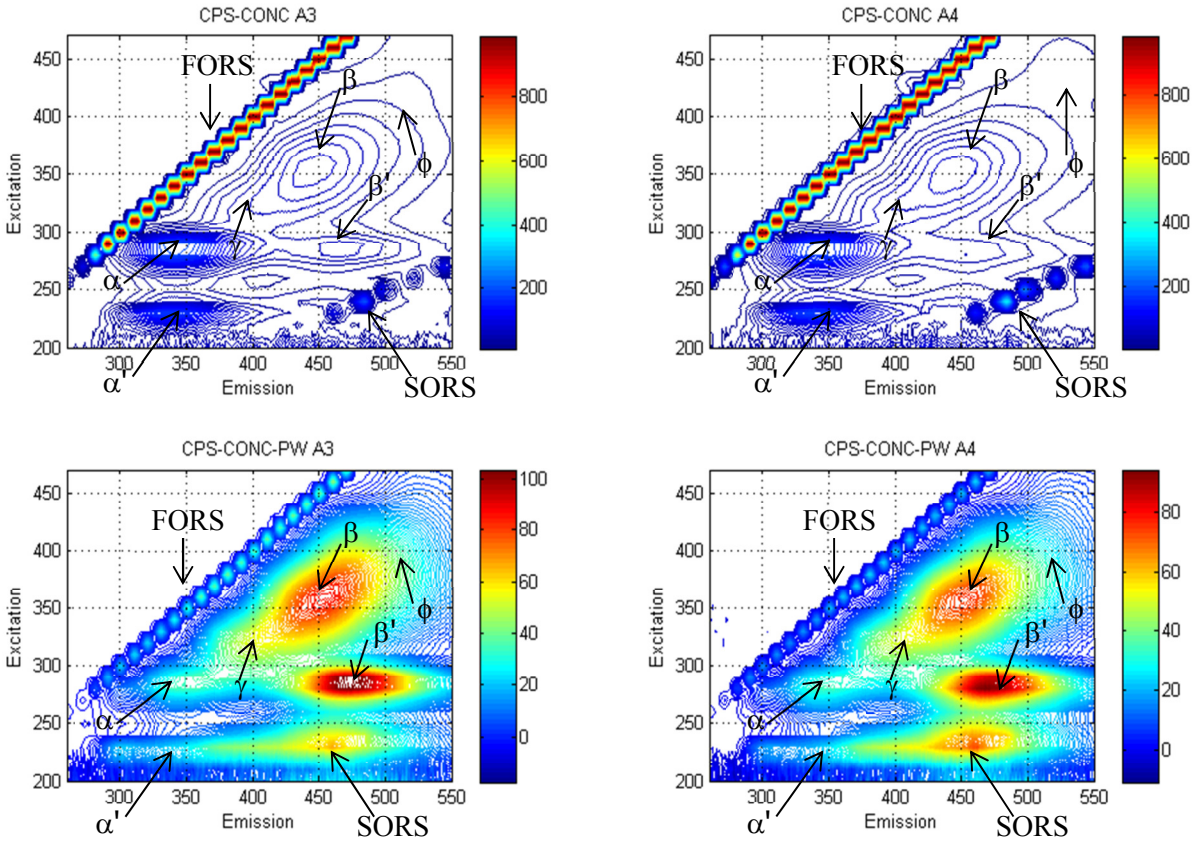

Figure 4.4 EEMs for CPS-Conc and CPS-Conc-PW for batch A.

After the perlite column, samples from the ultrafiltration, SPEP-R-Conc and SPEP-R-Conc-PW, show the following spectra:

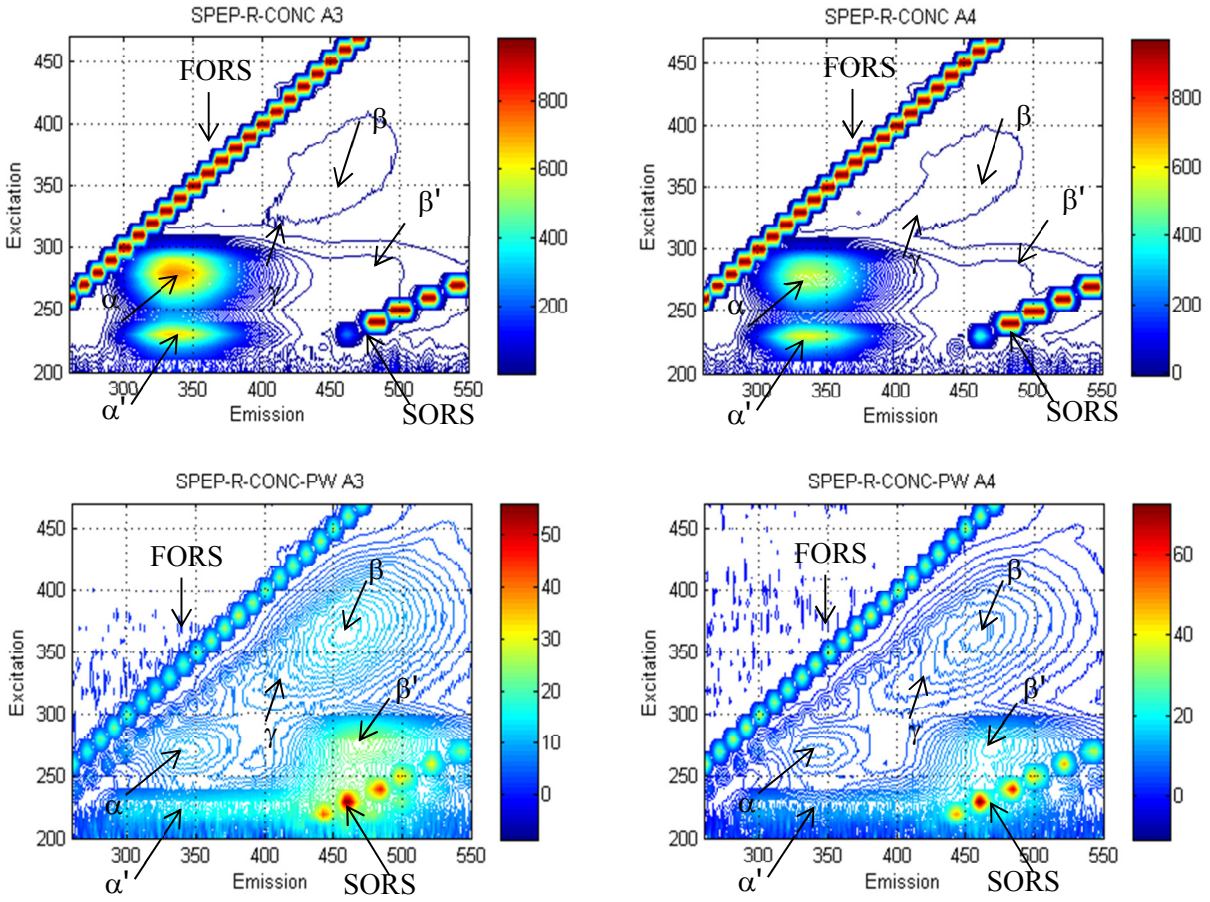

Figure 4.5 EEMs for SPEP-R-Conc and SPEP-R-Conc-PW for batch A.

In the SPEP-R-Conc EEM, the emission peaks for fluorophores other than tryptophan start to become less evident . On the other hand, fluorescence signals corresponding to NAD(P)H, FAD and tryptophan in a lesser amount, appear in the permeate waste (SPEP-R-Conc-PW). As the purification

proceeds, the peak of Trp becomes more intense possibly correlating with the expected progressive increase in concentration of the protein in the sample.

For EEM of the sample referred to as SKA-R-P1, the most intense peak as expected is the one corresponding to tryptophan since the sample contains mostly protein. At this point, the two parallel manufacturing lines of the process are combined together to obtain a higher concentration of protein, as compared to the protein levels observed in samples prior to being combined. In contrast, lower intensities of Trp are observed for the SKA-S1 waste indicating a relatively small loss of protein from the system. The Trp could be related to the presence of other antigens and to some wasted PRN. Also, some traces of NAD(P)H and/or NAD(P)H-dehydrogenase complex are also present in these two samples.

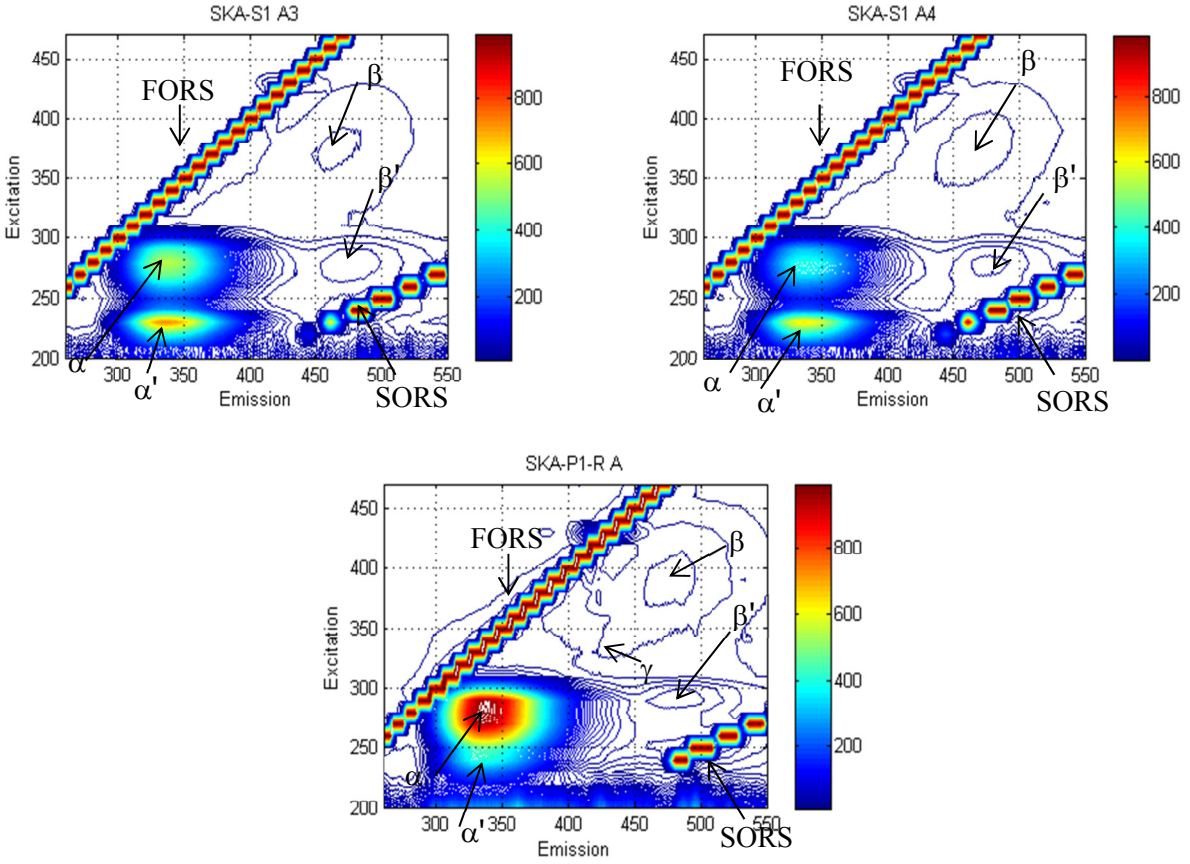

Figure 4.6 EEMs for SKA-S1 and SKA-P1-R for batch A.

A large Trp peak is present in the supernatant waste originating from the third ammonium sulphate precipitation (SKA-S3). The Em at ~340 nm suggests that the Trp residues may be exposed to the solvent due to unfolding, but it is not possible to ascertain whether this emission comes from PRN since other antigens could still be present in the sample. For the resulting pellet from this step (SKA-P4), only the Trp peak is present, and the dominant emission wavelength starts to resemble the characteristic value for PRN at 335 nm as reported in literature (Junker *et al*., 2006).

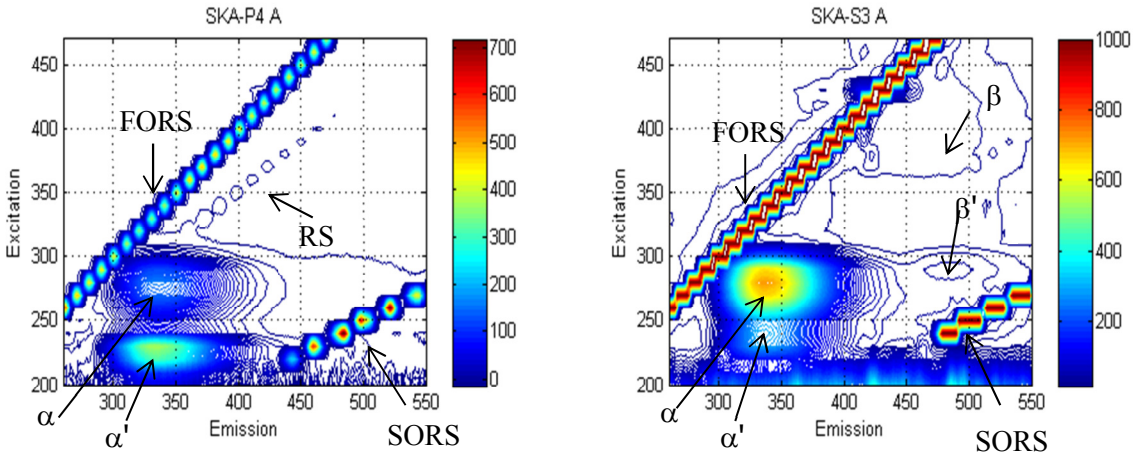

Figure 4.7 EEMs for SKA-P4 and SKA-S3 for batch A.

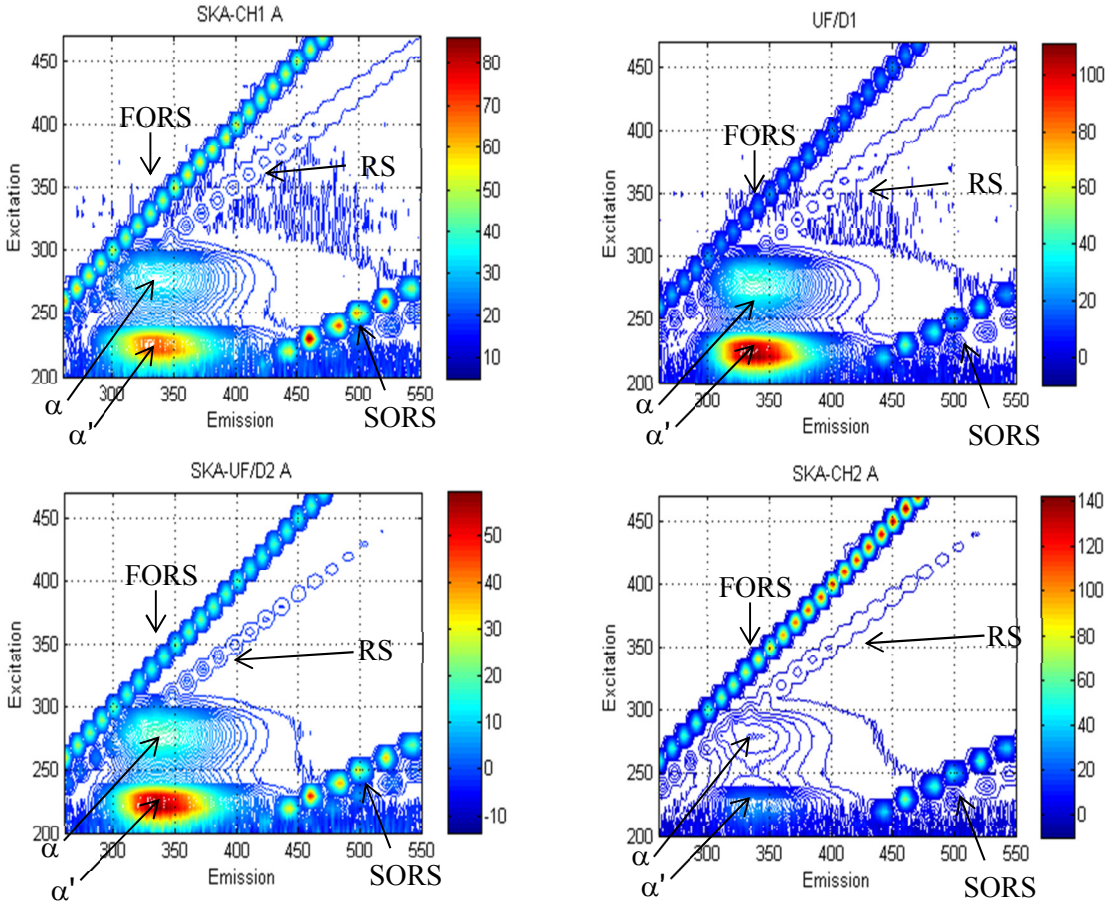

Figure 4.8 EEMs for SKA-CH1, UF/D1, SKA-UF/D2 and SKA-CH2 for batch A.

From this point on in the process until the last step (SKA-CH2 sample), only the Trp peak is observed in the EEM. There is a decrease in the intensities which may be a result of the elimination of traces of other antigens during the later stages of the purification.

4.2 PCA Analysis

As described in Section 3.2.3, four data matrices associated with different sections of the manufacturing process were analyzed using PCA in order to obtain fewer variables to describe major correlations in the original data. After performing cross-validation, and calculating the RMSECV, the number of components and the variance captured for each data matrix was obtained and is presented in Table 4.1.

	Upstream		**Downstream 1**		**Downstream 2**		**Downstream 3**	
PC	Variance %	Cumulative Variance %	Variance %	Cumulative Variance %	Variance %	Cumulative Variance %	Variance %	Cumulative Variance %
1	52.55	52.55	80.17	80.17	80.98	80.98	80.77	80.77
2	23.85	76.39	14.12	94.29	15.75	96.73	13.11	93.88
3	12.69	89.09	-	-	-	-	-	-

Table 4.1 Variance captured by each principal component in upstream and downstream samples.

For the four data sets, the cumulative variance captured for the PC's was approximately 93% and the percentage leftover, is related to unaccounted effects and noise.

The variance and the cumulative variance for a data set containing samples from the whole process are given in Table 4.2.

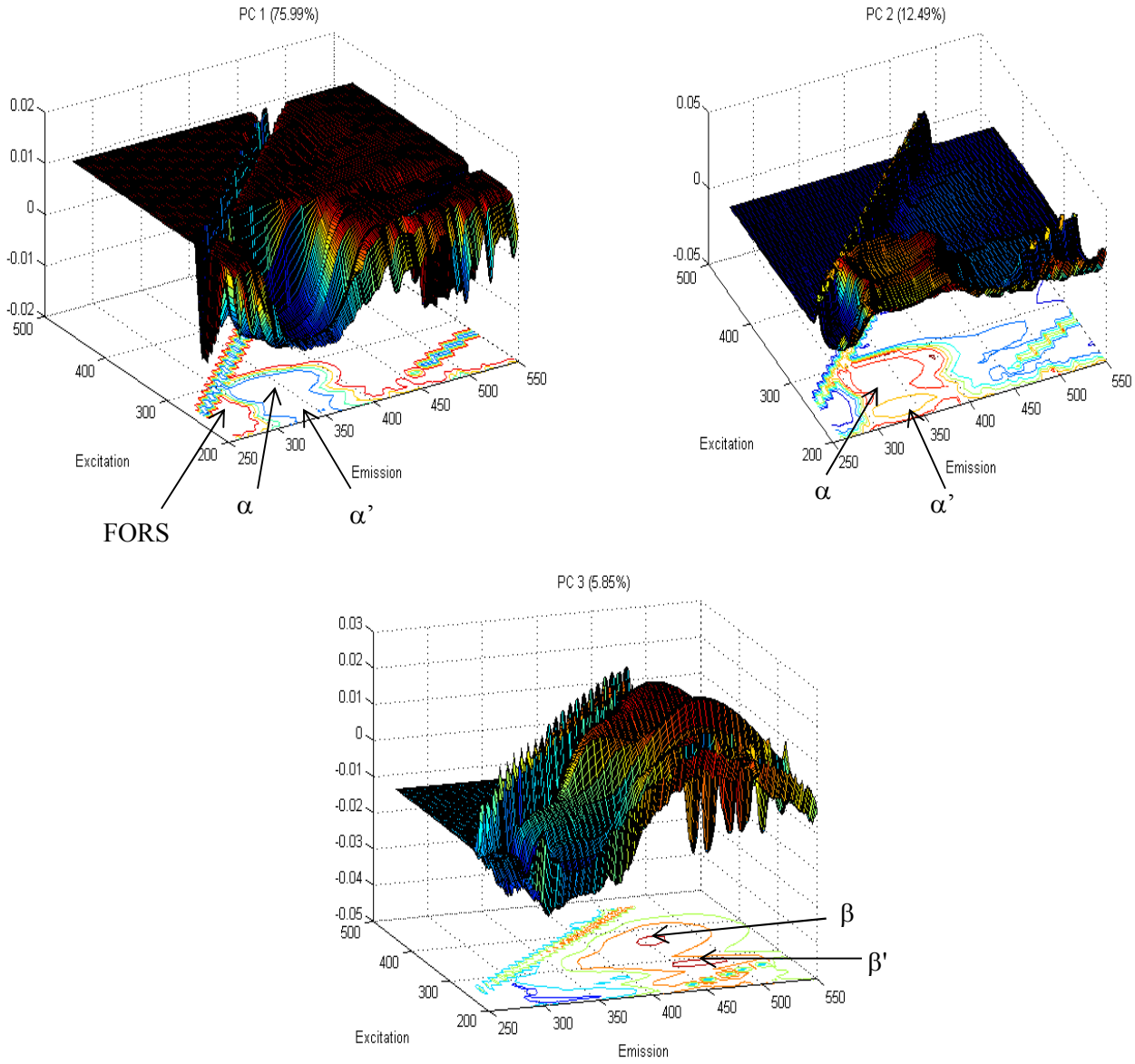

Figure 4.9 PC1, PC2 and PC3 loadings for upstream and downstream samples.

For PC1, the scattering peaks (FORS) and Trp (α, α') present the highest intensities, and for this reason, it is assumed that this component is primarily correlated with the protein content. The negative values indicate that there is an inverse correlation: when protein content is high, the score values are negative. On the other hand, the loadings for PC2 show a peak in the region where the emission of Trp is characteristic (α, α') so It is also assumed that PC2 is related to protein. Unlike in PC1, the loadings for PC2 are positive, meaning that there is a direct proportional correlation. For PC3, the loading plot presents higher positive intensities for the peak related to NAD(P)H-enzyme complex (β'), indicating a direct correlation with this fluorophore.

To identify possible differences between the upstream and downstream phases, the loadings from the spectra measured for the upstream and downstream samples were individually calculated. The loadings for the first 3 principal components PC1, PC2 and PC3 for the upstream data matrix are shown in Figure 4.10.

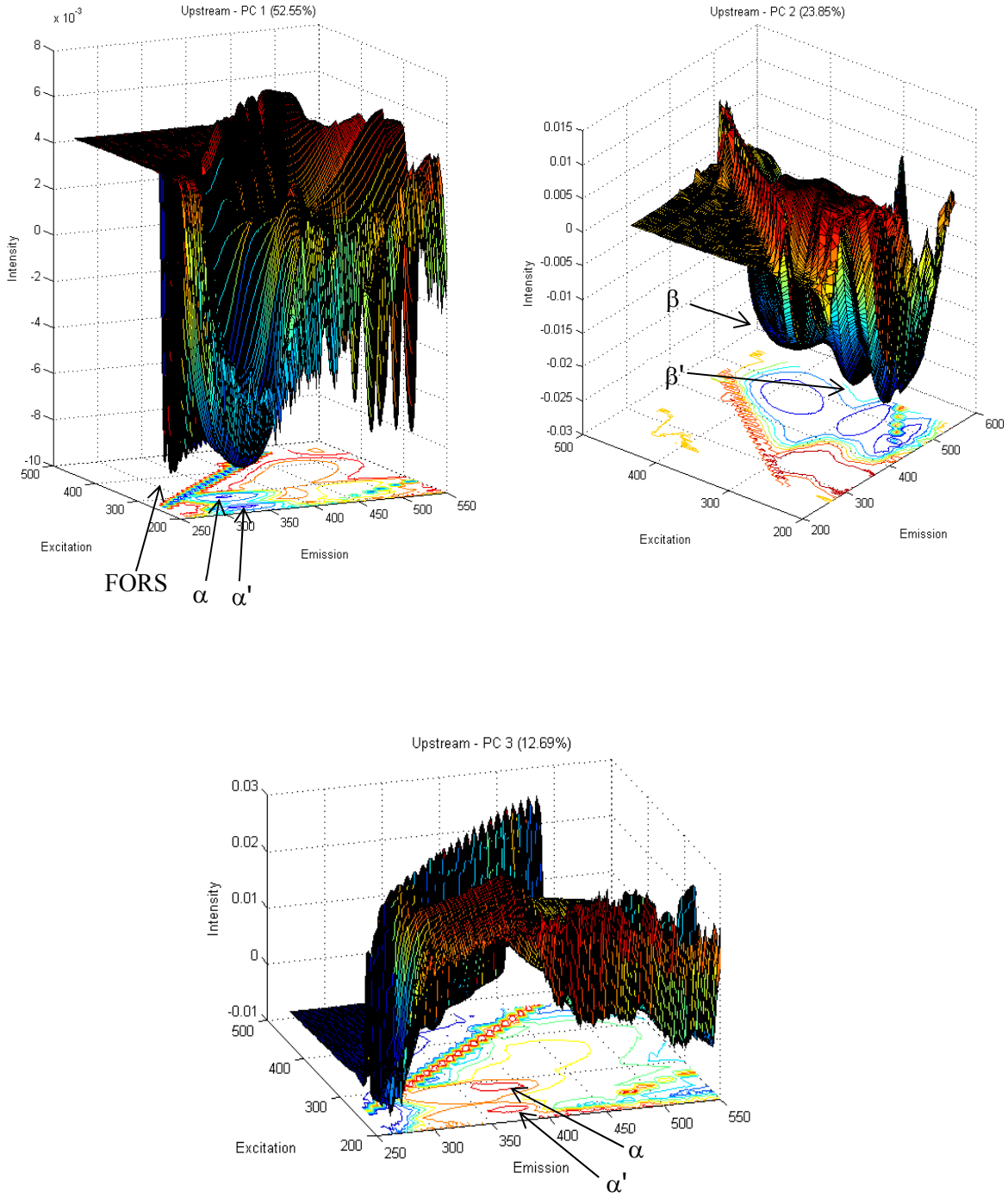

Figure 4.10 PC1, PC2 and PC3 loadings for upstream sample.

In the loading plot for PC1, the intensities with higher values are those corresponding to Trp (α, α') and scattering (FORS). Because intensity is large for what is attributed toTrp, it is hypothesized that PC1 is most likely related to protein. Since the actual values of the loading vector are negative, there is an inverse correlation: when the score value is negative, protein content is high, and *vice versa*. It should be noticed that although the fluorescence intensities are positive, negative PCA scores occur because of the mean centering and scaling procedure applied to the data before building the PCA model. On the other hand, the loading plot for PC2 possesses higher negative intensities for the peak attributed to the NAD(P)H-enzyme complex (β'), indicating an inverse correlation with this fluorophore. For PC3, the EEM shows a shoulder with positive intensity values where Trp emission is characteristic (α, α'), but the maximum emission presents a red-shift. This could be indicative of conformational changes in the proteins, since when tryptophan is more exposed to the environment, the emission is shifted to higher wavelength (Burstein *et al*., 1973; Junker *et al*., 2006; Seabrook *et al*., 1991).

For the three data matrices constructed from samples collected from steps further downstream, the loading plots related to PC1 and PC2 show an inverse correlation to scattering (FORS) and a proportional correlation to protein content, respectively. Loading plots for these principal components are shown in Figures 4.11, 4.12 and 4.13.

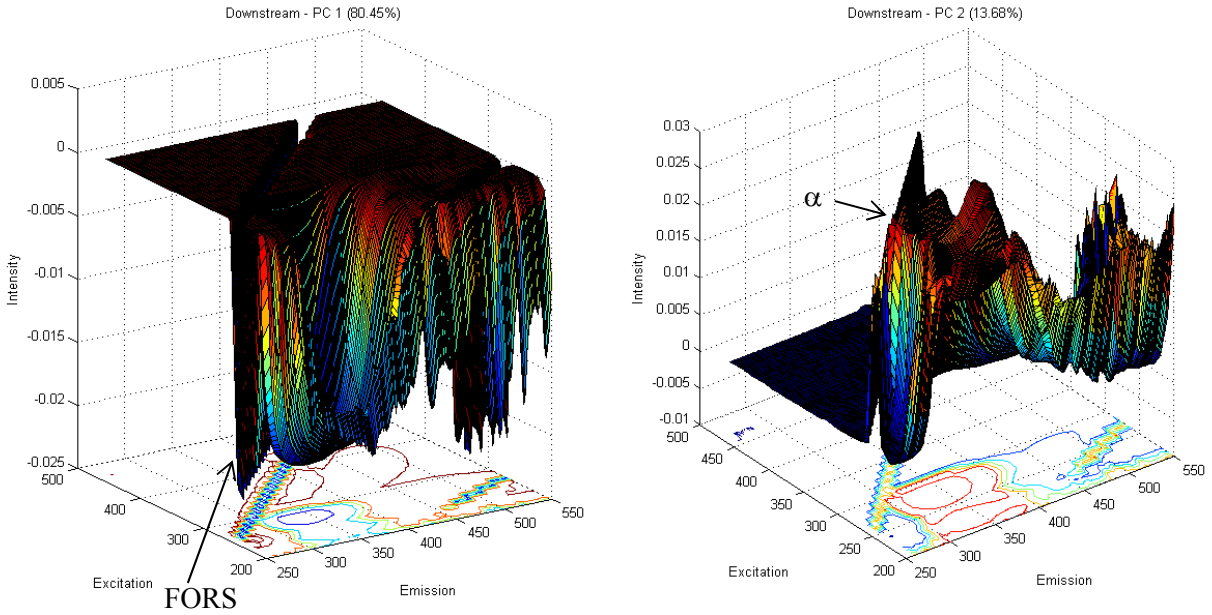

Figure 4.11 PC1 and PC2 loadings for downstream 1.

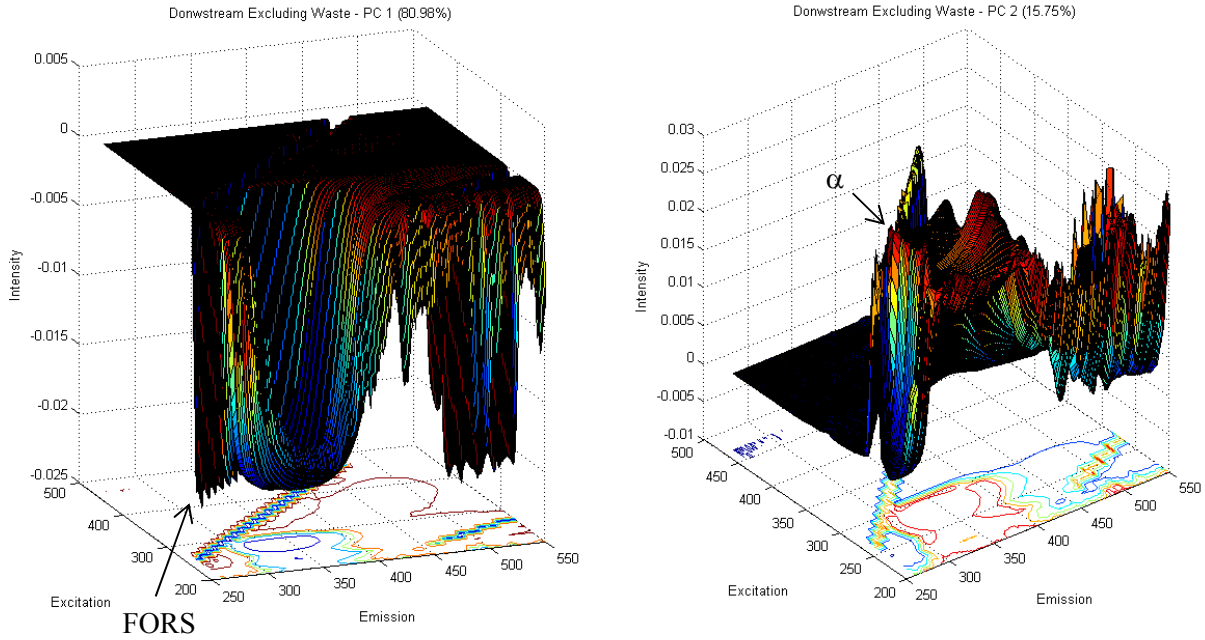

Figure 4.12 PC1 and PC2 loadings for downstream 2 (downstream excluding waste).

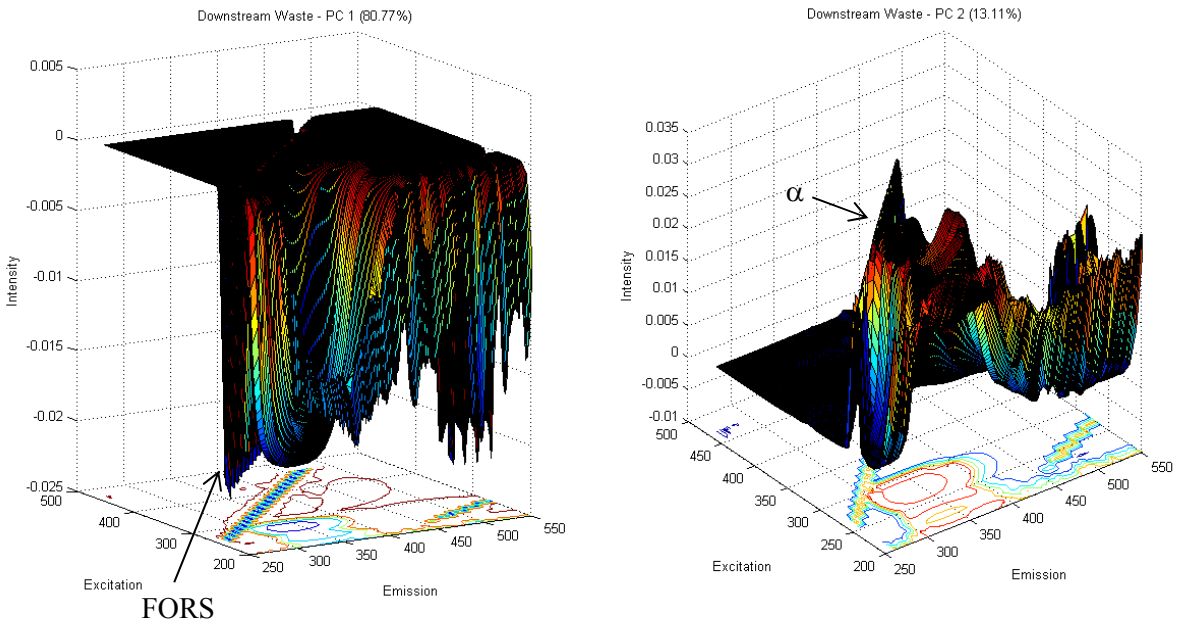

Figure 4.13 PC1 and PC2 loadings for downstream 3 (downstream waste).

Comparing the loadings calculated for the downstream samples with those calculated for the upstream samples, the NAD(P)H-enzyme binding peak is not significant or is not present at all in the downstream loadings. When comparing the loadings from the PC's of the entire process to the loadings from the upstream and the downstream steps, the most appreciable difference is the NADH-enzyme complex peak. In the analysis of the data set containing all samples, PC3 which is dominated by the NADH-enzyme complex fluorescence, captures only 5.85% of the variance in the data. However, in the data set with samples from the upstream step, the variance captured by PC2, which is dominated by the NADH-enzyme complex for the upstream samples, is 23.85%. This suggests that

this NADH-enzyme complex only has a significant influence in the variability of the process related to the upstream steps, and not to the entire process.

The relationship among different samples can be extracted from the score plots (Wise *et al.,* 2007). By comparing the score of each sample for each principal component, similarities and differences can be estimated among the original samples (Persson *et al*., 2001).

Score plots for the data matrix of the whole process are shown in Figures 4.14 and 4.15, and the score plots for the upstream data matrix are shown in Figures 4.16 and 4.17.

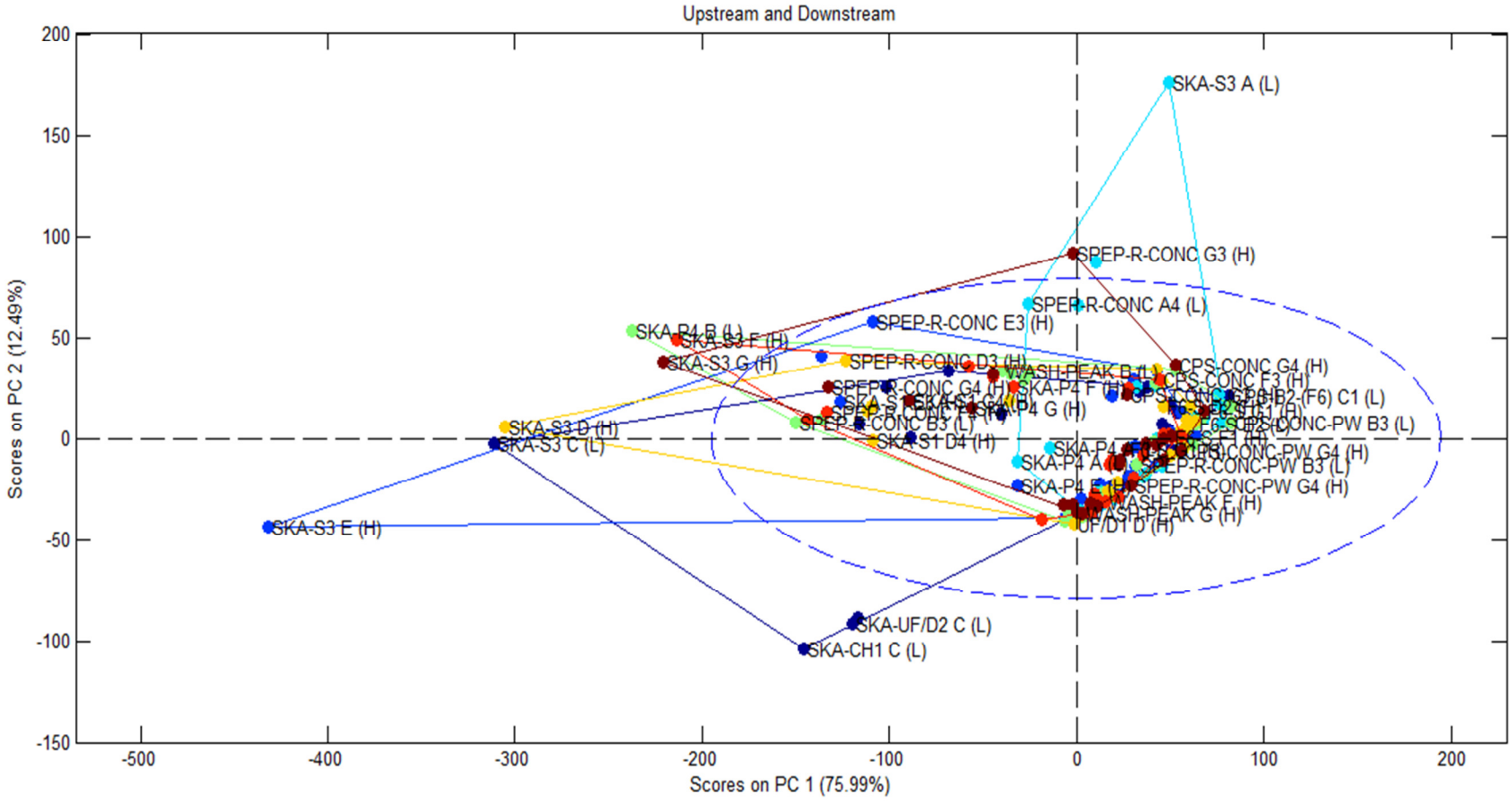

Figure 4.14 PC1-PC2 scores plot for data matrixes for the whole process.

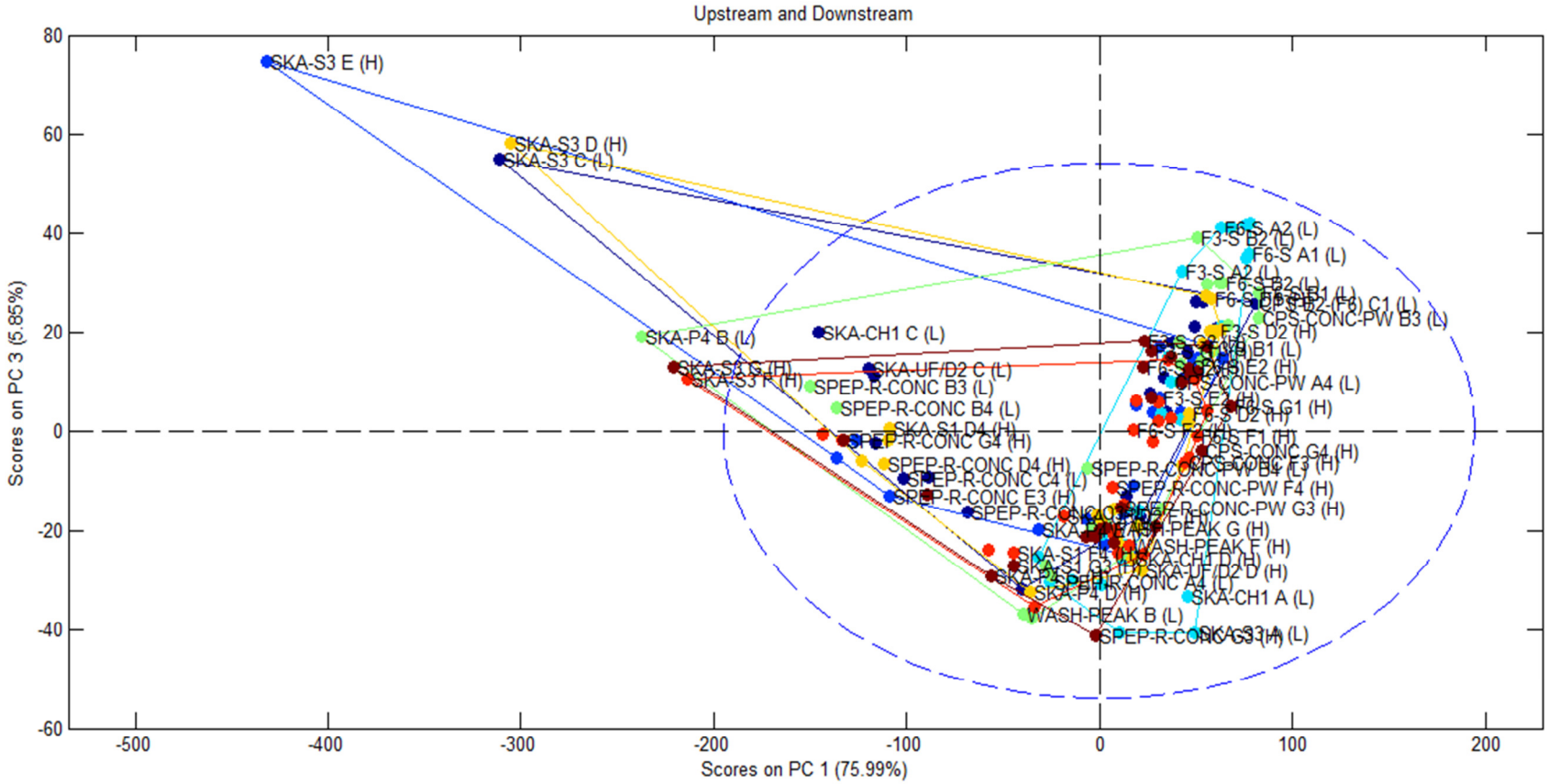

Figure 4.15 PC1-PC3 score plot for data matrices for the whole process.

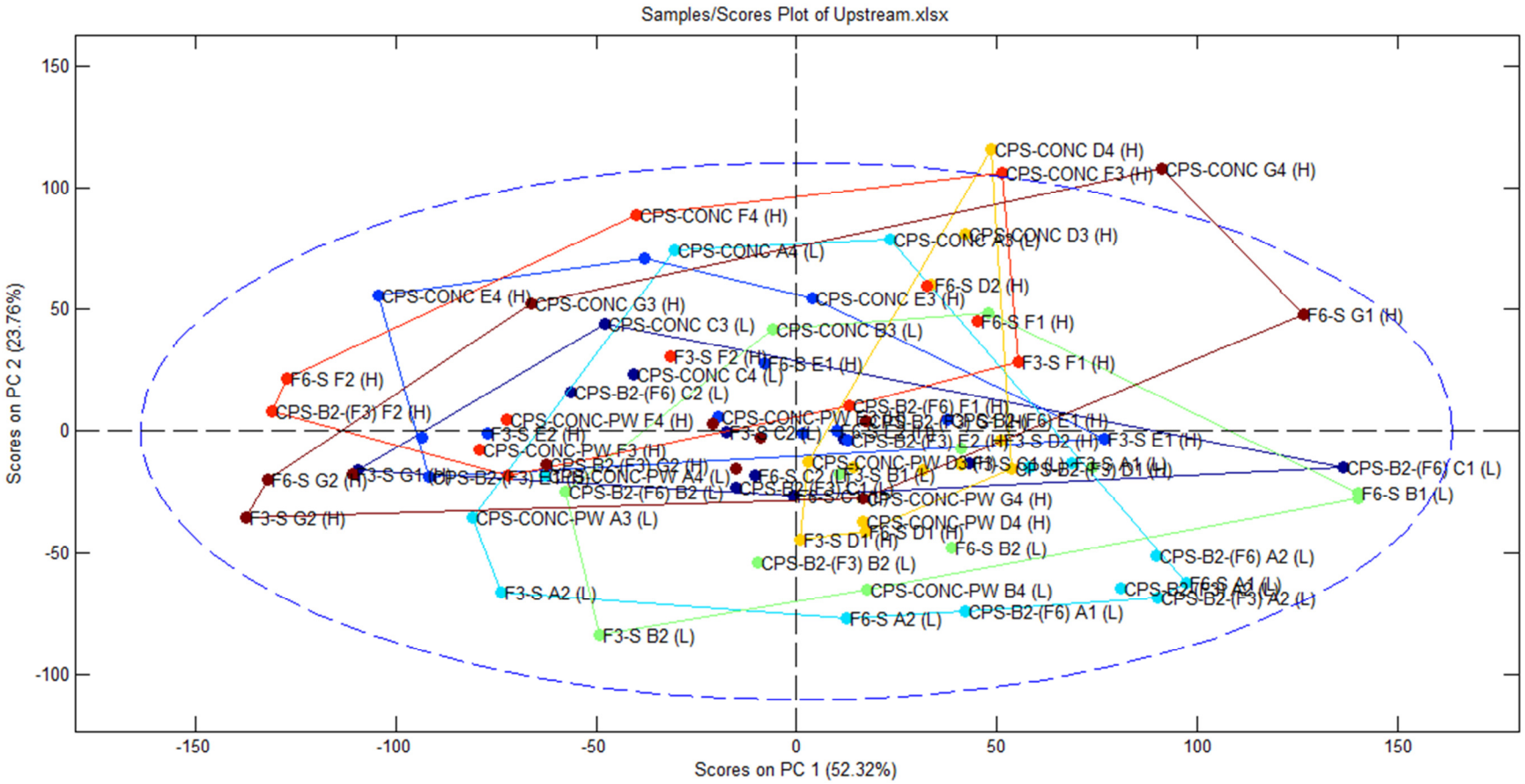

Figure 4.16 PC1-PC2 score plot for upstream.

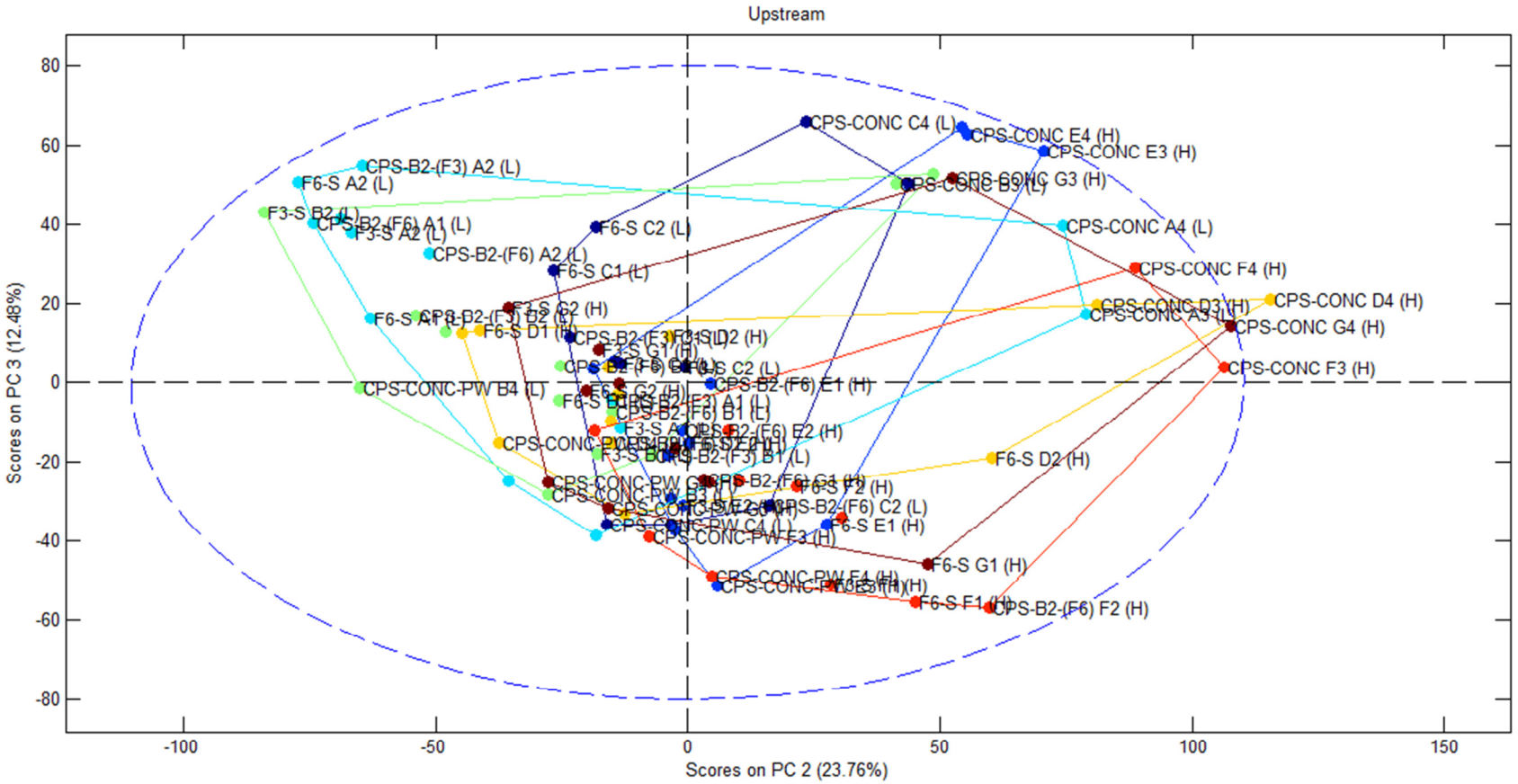

Figure 4.17 PC2-PC3 scores plot for upstream.

Each point in the score plots represents the score for a particular sample. The lines between scores correspond to samples for a particular batch to allow for easier visualization of the correlation between fluorescence and productivity according to batch. The batches that resulted in low or high final Kjeldahl values (as indicated in Table 3.1) are denoted by an L or H, respectively next to the sample number. For easy visualization, the batches with low final Kjeldahl (L) values are shown with blue lines in all the score plots.

As previously discussed, for the upstream data PC1 and PC2 are inversely correlated to protein content and NAD(P)H-enzyme complex concentration respectively, and PC3 in directly correlated to protein content. The main objective of this analysis was to find a possible correlation between the fluorescence corresponding to different steps of the production process and the final PRN concentration in the final stage of the purification in order to identify possible causes of low productivity for this product.

For the PC1-PC2 score plot, the batches with low final Kjeldahl (Table 3.1) have the largest negative values for PC2, indicating that the quantity of NAD(P)H-enzyme complex in the samples is higher in the batches with low final Kjeldahl as compared to those lots with high final Kjeldahl. Also most of the score points for the batches with low productivity have positive values for PC1. The fact that there is a high NAD(P)H-enzyme complex concentration and low protein content for these batches hint at the possibility that some metabolic steps related to the consumption of NADH are proceeding at different rates in different fermentations and these differences maybe a source of variability in the final yield of PRN.

During oxidative phosphorylation, co-enzymes such as NADH and $FADH_2$ act as electron donors, and O_2 as acceptor resulting in ATP formation. In the absence of oxygen or in cases where there may be oxygen transfer limitations, NADH cannot be oxidized sufficiently fast and therefore it may accumulate (Marose *et al*., 1998).

Some score points for batches with high productivity also fall in the low protein content area of the plot. This suggests that for these batches, not all the four fermentations proceeded to completion.

Score plots for the groups of downstream data are shown in Figures 4.18, 4.19 and 4.20.

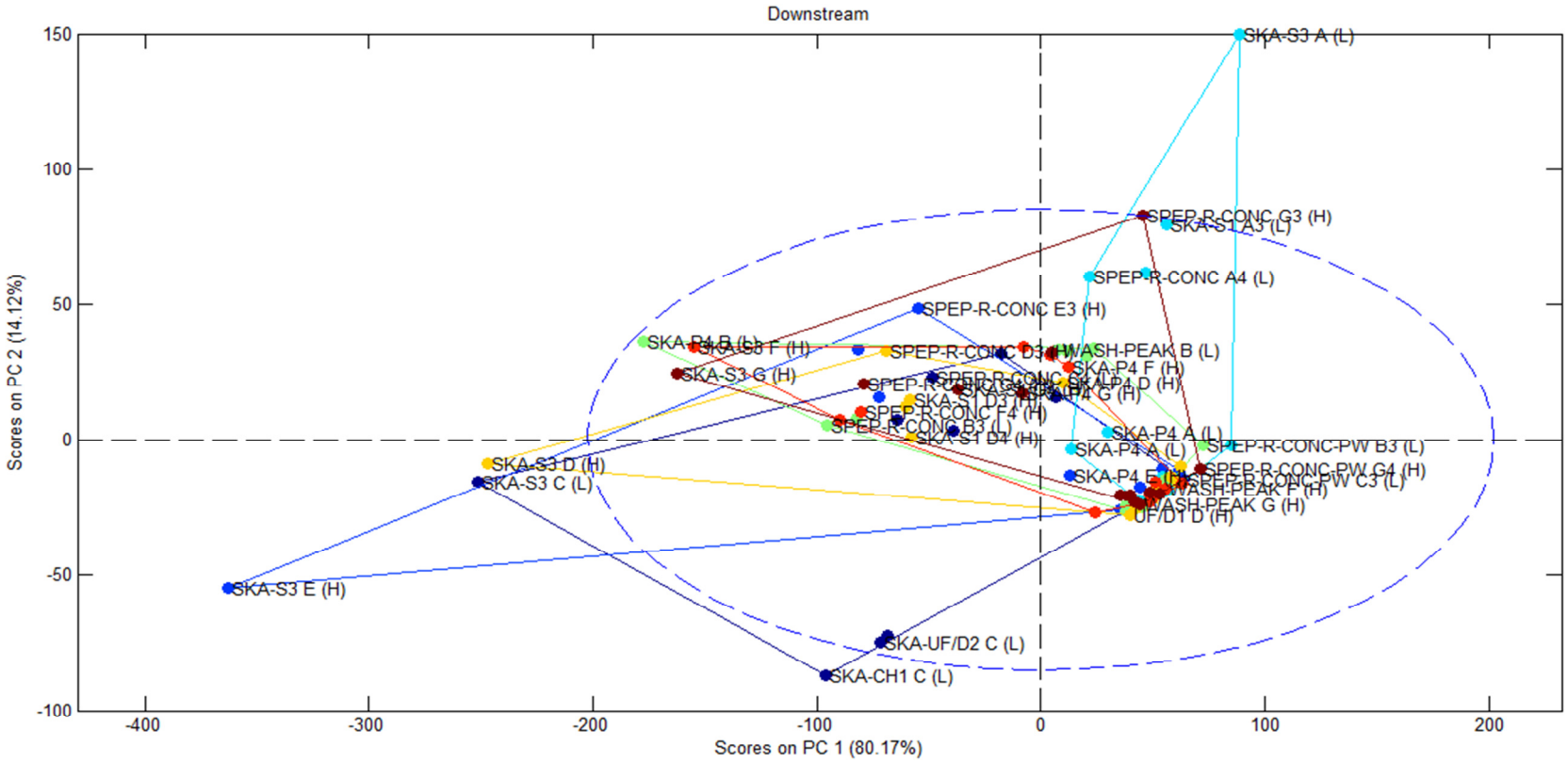

Figure 4.18 PC1-PC2 score plot for downstream 1.

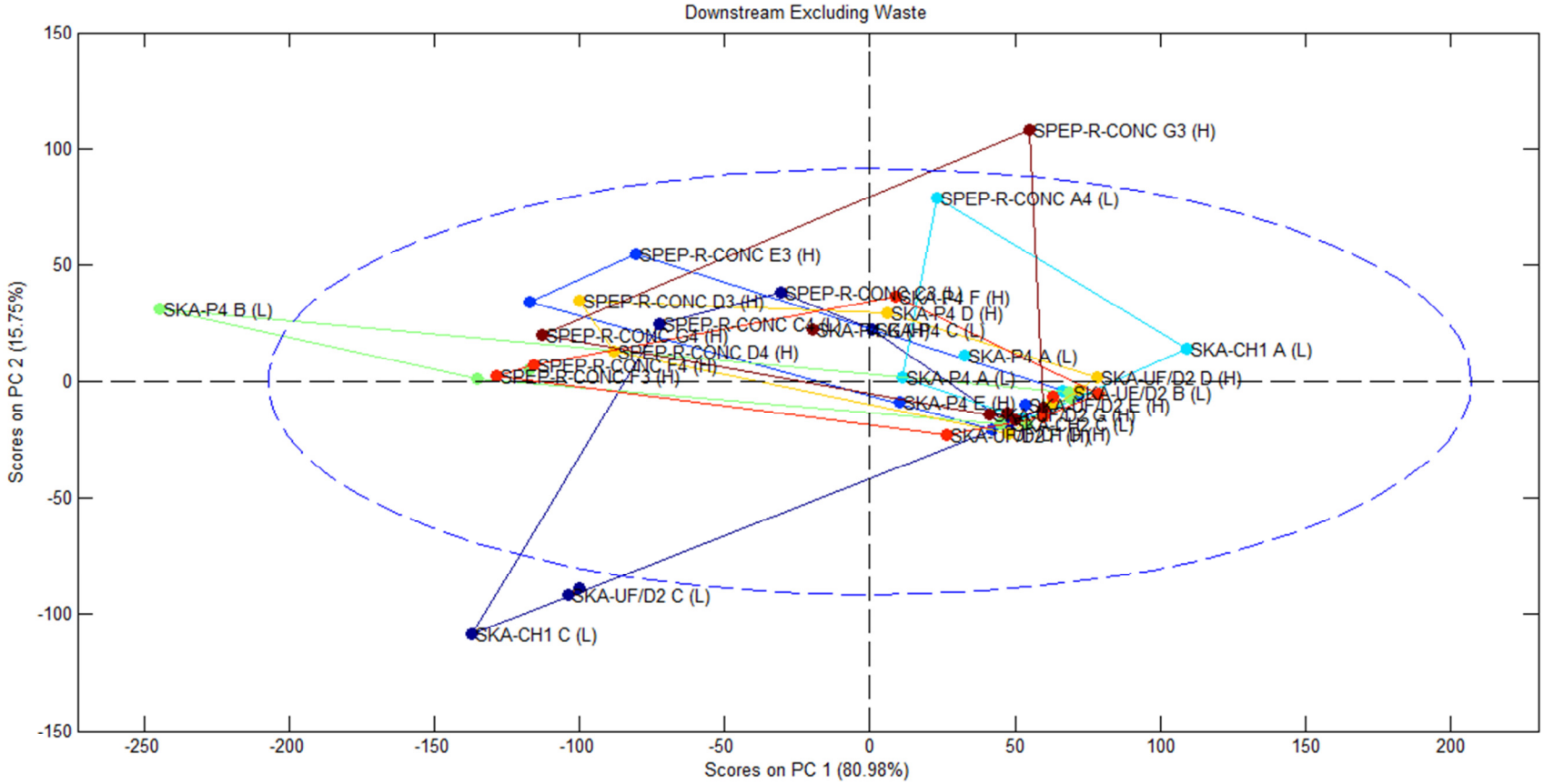

Figure 4.19 PC1-PC2 score plot for downstream 2.

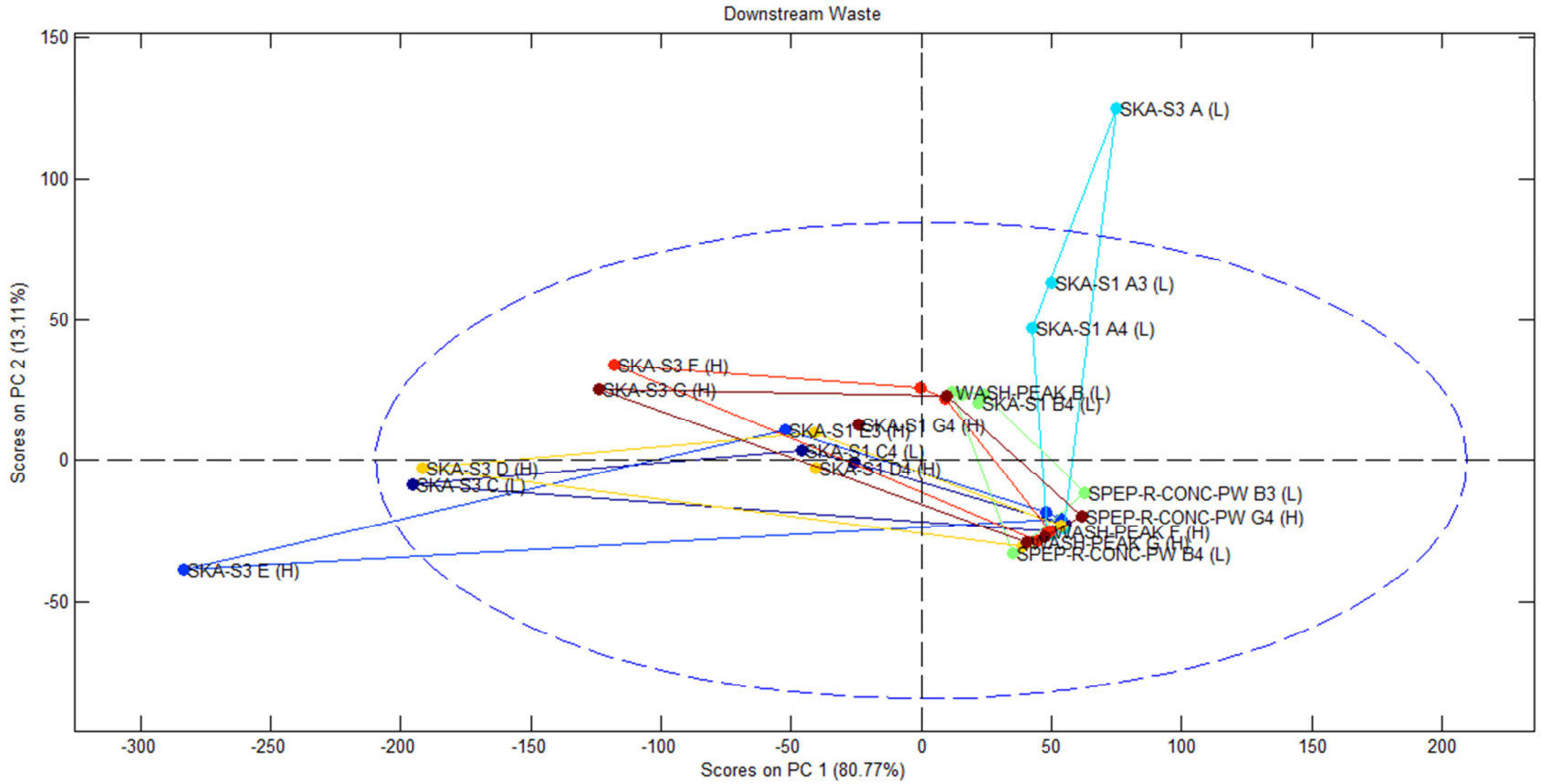

Figure 4.20 PC1-PC2 score plot for downstream 3.

From these plots, it is not possible to identify any pattern that differentiates batches with low final Kjeldahl from batches with high final Kjeldahl. Thus, samples that belong either to batches with high Kjeldahl or low Kjeldahl are dispersed in the plots thus exhibiting a lack of correlation between scores and productivity. For this reason, downstream purification protocols were dismissed as a cause of low productivity.

4.3 Investigation of the correlation between productivity and dissolved oxygen, aeration and agitation profiles recorded during the fermentation.

With the observation that NADH accumulates in different amounts for different runs and that this accumulation maybe related to oxygen transfer limitations, dissolved oxygen and aeration rates were investigated. These parameters are recorded on-line during the fermentations and the results are stored in a database.

Dissolved oxygen profiles for each batch are provided in Figure 4.21.

No significant differences between the profiles of dissolved oxygen for fermentations that produced higher or lower levels of pertactin were observed. Thus, there seems to be no appreciable relationship between dissolved oxygen levels in the fermenters and the variability observed following the downstream processing. It should be noted that dissolved oxygen is controlled in closed loop during the fermentation by manipulating the aeration rate. Thus, the dissolved oxygen is generally fairly constant at its set point and may not vary significantly as a result of changes in oxygen consumption by the cells.

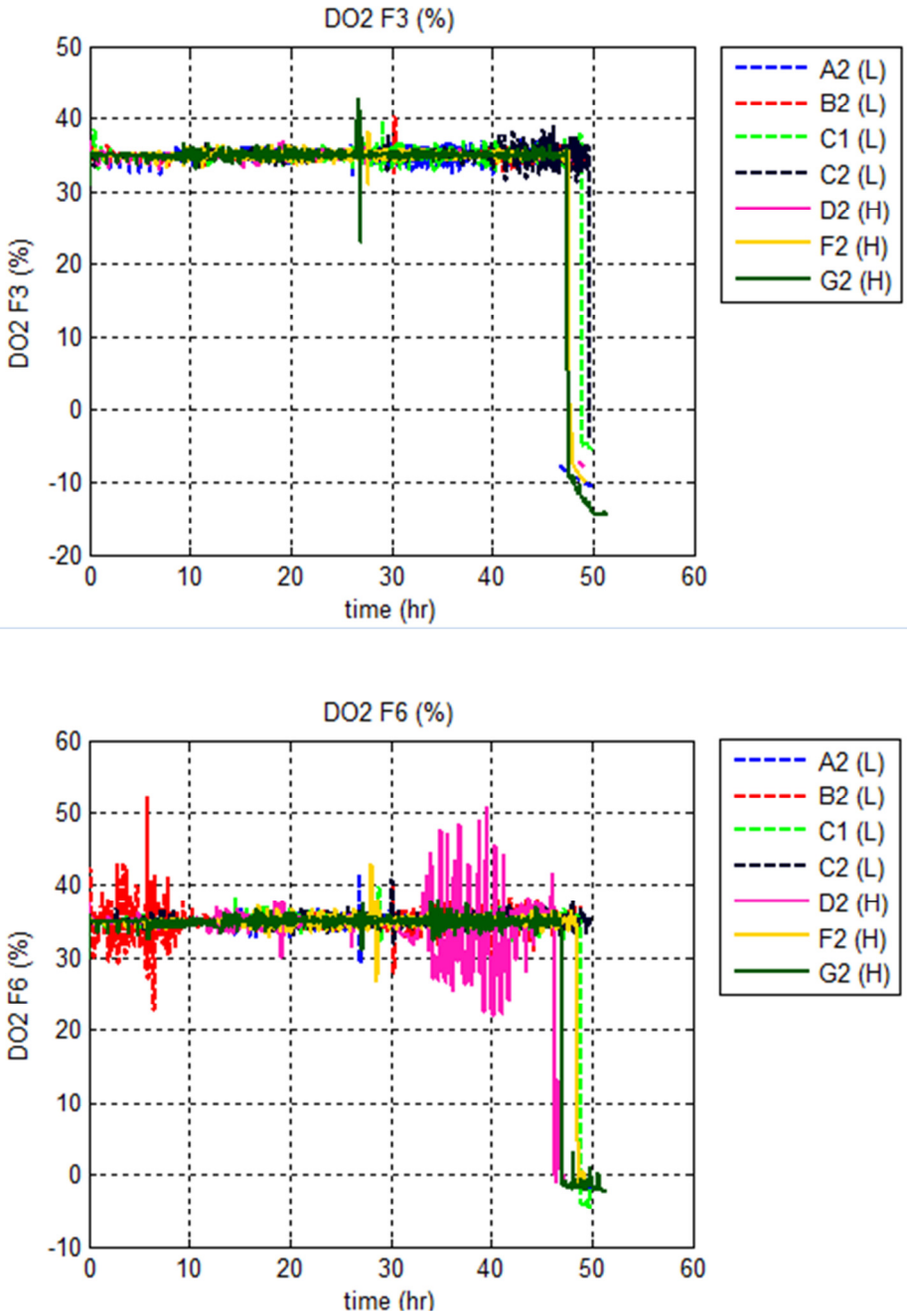

Figure 4.21 Dissolved oxygen in fermenters F3 and F6. Batches with high final Kjeldahl are marked with solid lines, and the ones with low final Kjeldahl with dashed line.

An additional possible indicator of changes to oxygen rates is the aeration rate time profile since the aeration rate is manipulated to control the dissolved oxygen at a preset set point value. In the airflow profiles, there are some differences among batches, but the relationship between the aeration levels and variability in protein concentration is not straightforward. Since the final protein concentration was found to be negatively correlated to the scores of PC1 and PC2, the aerations rates were compared in different batches to these scores. For example, in a high productivity batch D2 (green line in Figure 4.16), the airflow in the first fermenter F3 was low compared to the other batches, while the scores for this sample (Figure 4.16) were positive for PC1, implying low protein content, and negative for PC2 implying high NAD(P)H content. The airflow in the second fermenter F6 for this batch was high, and the scores were positive for PC1, as in F3, but positive for PC2 thus implying a low NAD(P)H content. Comparing to a batch with low productivity, such as A2 (dark blue line in Figure 4.16), the F3 airflow was higher than for batch D2, but it had a higher quantity of NAD(P)H indicated by its high negative PC2 value.

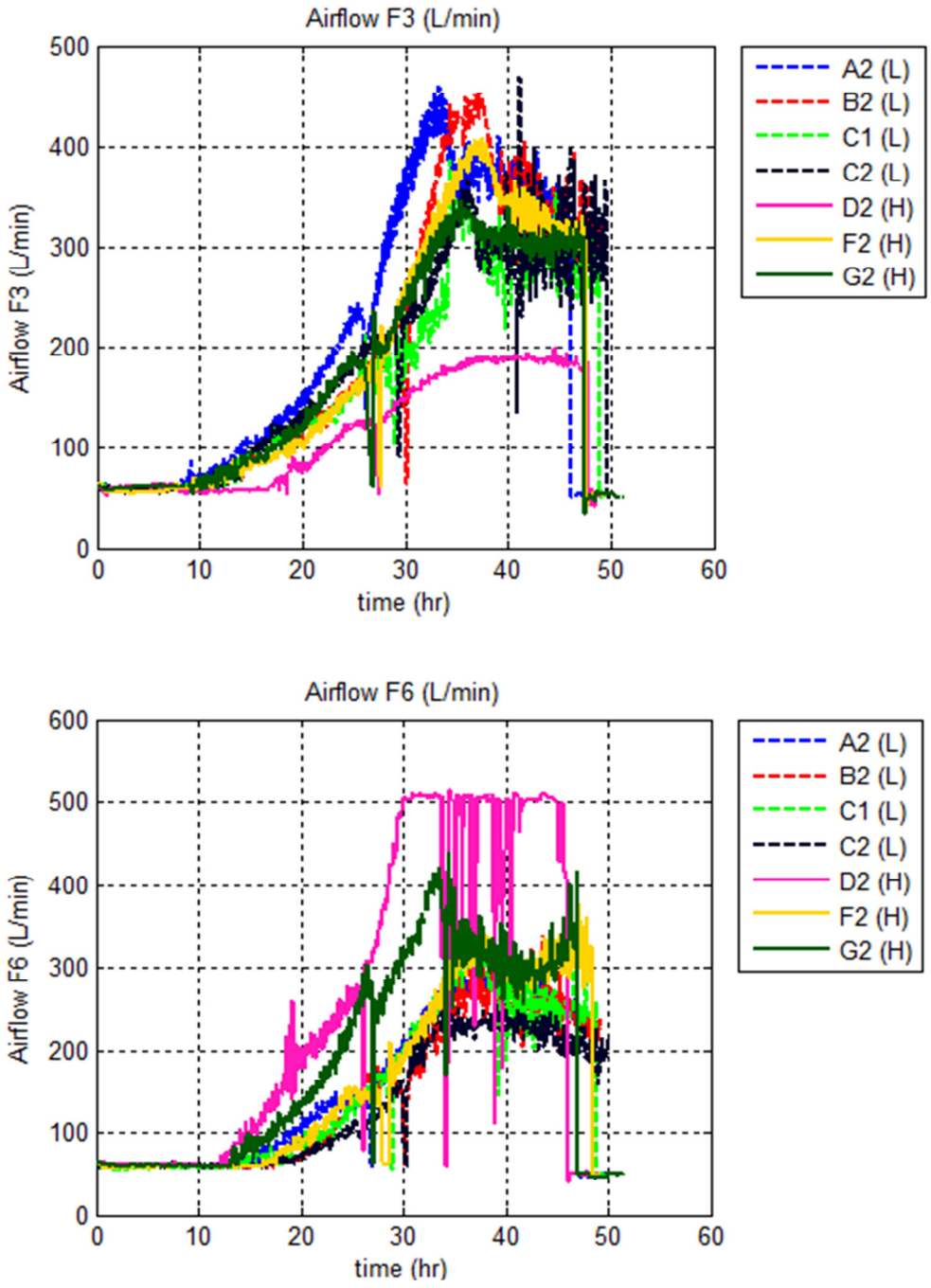

Figure 4.22 Airflow in fermenters F3 and F6. Batches with high final Kjeldahl are marked with solid lines, and the ones with low final Kjeldahl with dashed line.

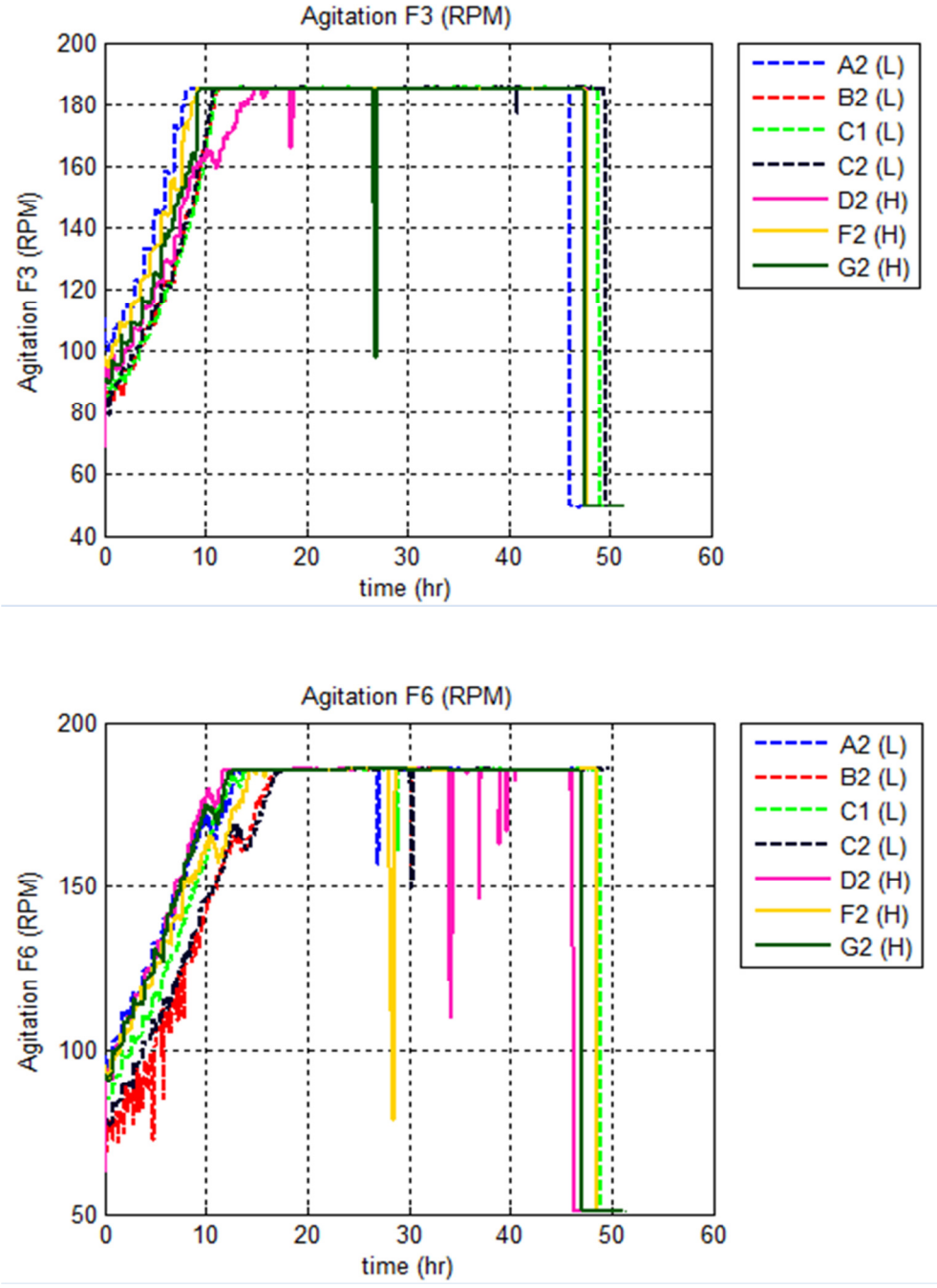

Figure 4.23 Agitation in fermenters F3 and F6. Batches with high final Kjeldahl are marked with solid lines, and the ones with low final Kjeldahl with dashed line.

Similar to the dissolved oxygen profiles, there is an overlap among agitation profiles belonging to batches with high and low productivity, showing no clear correlation between this parameter and variability in the final protein content. Agitation rates are shown in Figure 4.23. In these profiles, there is an apparent relationship between agitation rate and productivity but this is not consistent. For instance, in batches C1, C2 and B2, which had low productivity, agitation ramped slowly from the minimum to the maximum rpm value required to maintain the dissolved oxygen set point. However, in F3 for batch D2, the agitation ramped slowly compared to F6 for the same batch, even though the final batch for this fermenter had high productivity. The PC2 score for F3-S D2 in Figure 4.16 has a negative value suggesting high NAD(P)H content which corresponds to lower oxygen consumption and a corresponding slower ramp in aeration needed to maintain the desired oxygen set point as observed. Thus, the fluorescence scores appear to be more correlated to productivity than agitation rate.

It should be noted that to calculate the actual consumption of oxygen by the cells it would be necessary to measure, in addition to dissolved oxygen and aeration rates, the oxygen leaving the fermenters, which is not possible with the current set-up. Thus, measuring the NADH using fluorescence offers an advantage for identifying differences in metabolism, in particular as related to oxygen consumption.

4.4 Investigation of supplementation time as indicator of productivity

An attempt was made to correlate the variability in productivity to the start time for fed-batch supplementation. For example, it was observed that the supplementation time, i.e. the time at which fed-batch operation is started, varies among the different runs. The supplementation times as obtained from Figure 4.23 which show the change in the level of the supplementation tank from the beginning of the fermentation. During the first period that corresponds to batch operation and no supplement is fed to the fermenter, the level in the supplementation tank is constant at its initial value. Thus, the supplementation time corresponds to the instant at which the level in the supplementation tank starts to decrease (Figure 4.23). It is important to describe how the start of the fed-batch supplementation is initiated at Sanofi. In

the standard current operating procedure it is assumed that a sudden increase in dissolved oxygen correlates with the complete consumption of glutamine thus necessitating supplementation. Comparing the time at which the spike in dissolved oxygen occurred and the supplementation time, there are not significant differences and the supplement was added at the moment when the spike appeared. On the other hand, it is seen from Figure 4.23 that some batches that had low productivity (C1, C2 and B2) exhibit relatively late supplementation time (~30 hrs) compared to other batches (~25 hrs), suggesting that the supplementation time may be correlated to the final product yield. However, this is not consistent for all batches since there are batches with low supplementation time that resulted in low pertactin productivity. It should be pointed out that although the need for supplementation is established, as explained above using dissolved oxygen measurements, the start of supplementation is manual. Thus for the actual process the supplementation may not be implemented exactly after the observed increase in DO leading to additional variability and exposure of the culture to physiological stress.

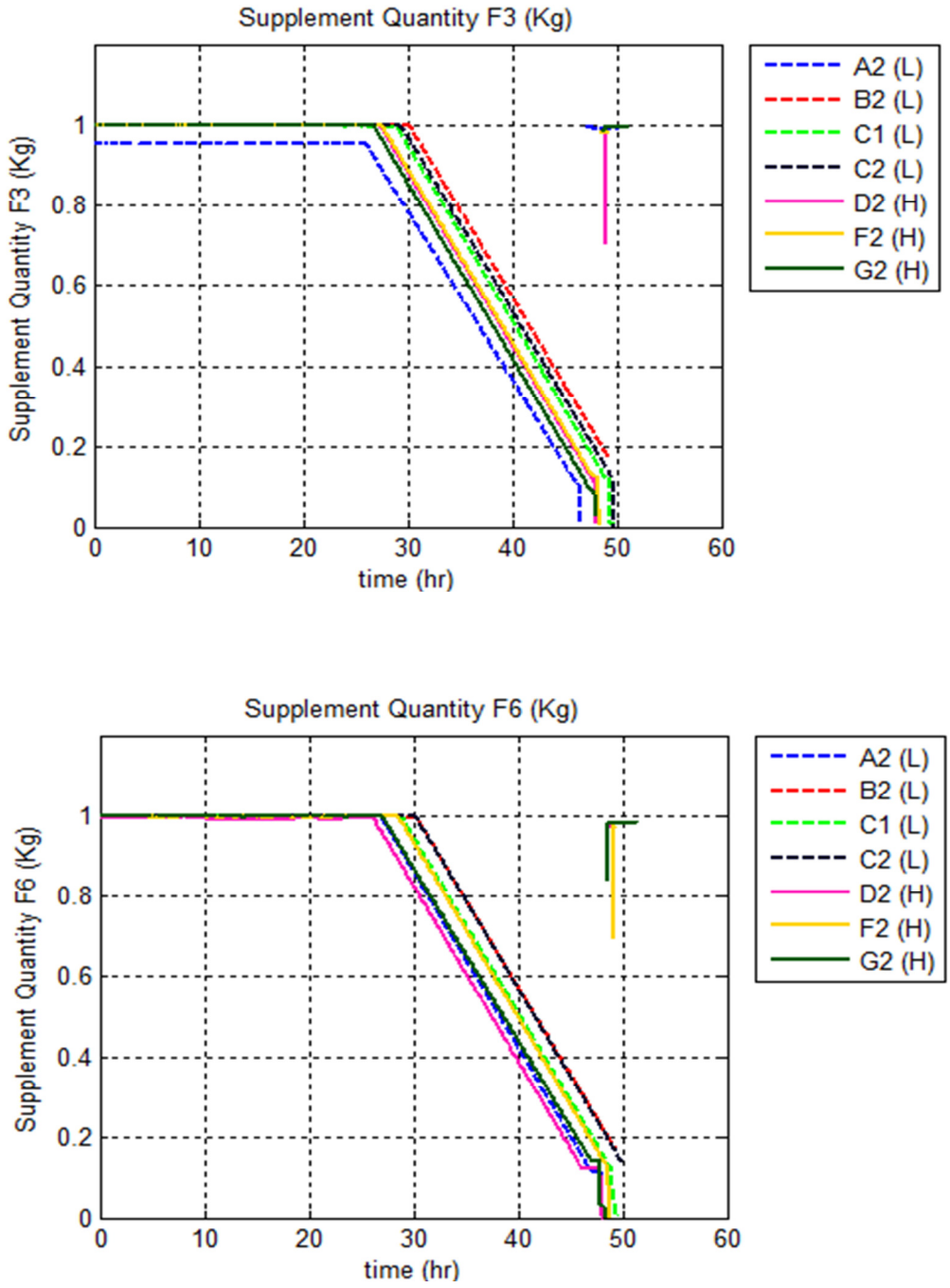

Figure 4.24 Supplement Quantity in fermenters F3 and F6. Batches with high final Kjeldahl are marked with solid lines, and the ones with low final Kjeldahl with dashed line.

4.5 Investigation of key metabolites' levels

Besides the parameters monitored during the process, key components of cell metabolism, such as lactate, glutamate and glutamine, were also monitored. It was hypothesized that changes in NADH will be associated to changes in the consumption or production of key metabolites. Figures 4.25, 4.26 and 4.27 show the measured concentrations for lactate, glutamine and glutamate for different batches conducted in fermenters F3 and F6 as a function of the corresponding final Kjeldahl pertactin concentration. The samples were obtained at the end of the fermentations.

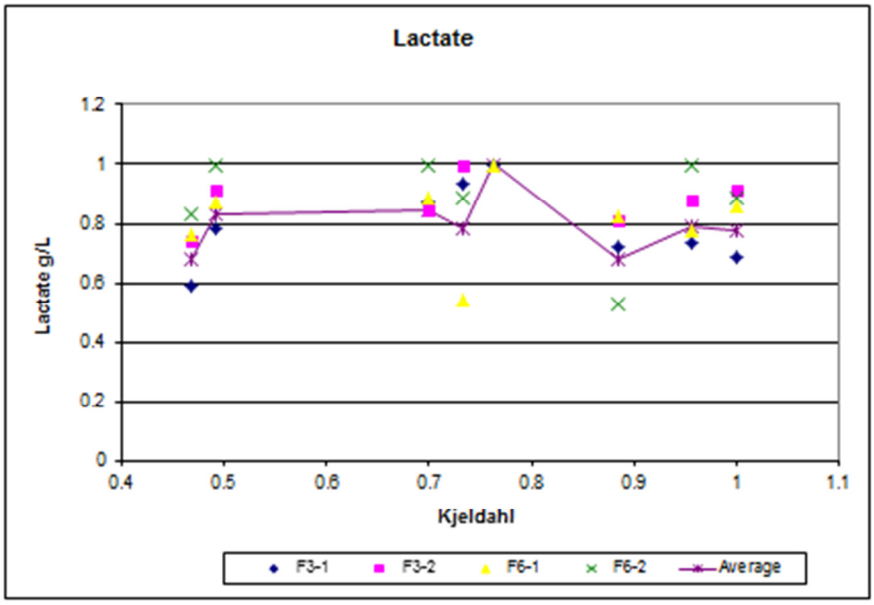

Figure 4.25 Lactate concentrations vs. Kjeldahl.

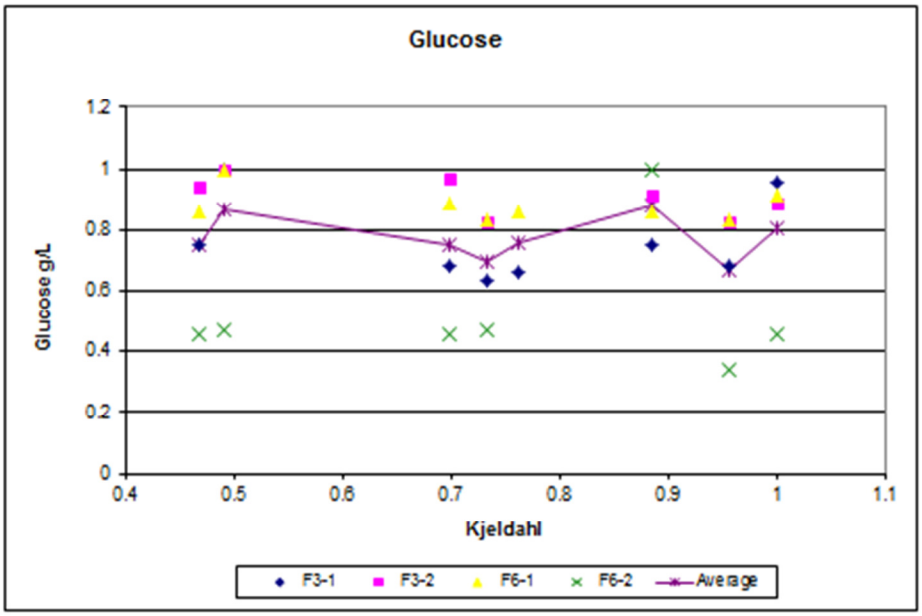

Figure 4.26 Glucose concentrations vs. Kjeldahl.

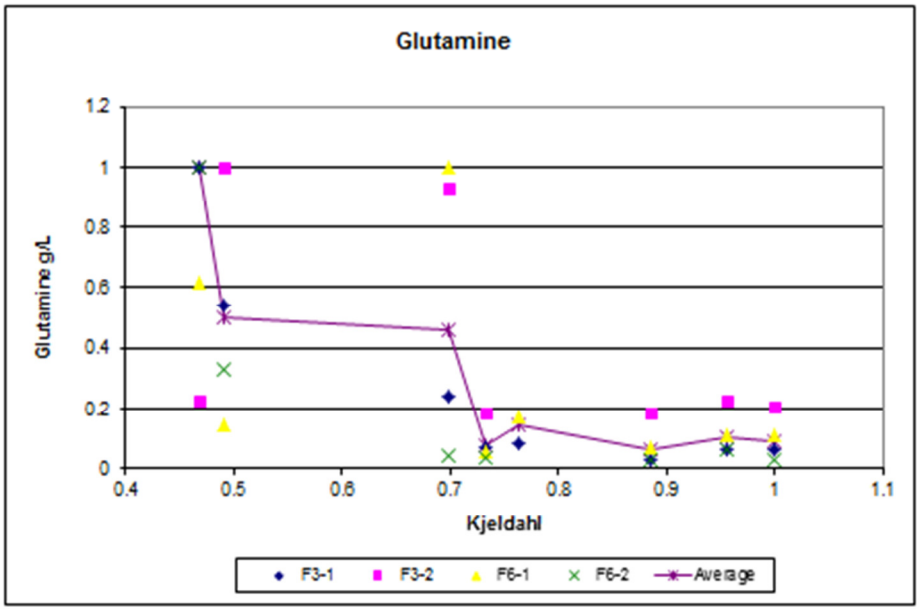

Figure 4.27 Glutamine concentrations vs. Kjeldahl.

For lactate and glucose, there are not pronounced differences among the batches with high and low productivity. However, for the glutamine profile, there are some batches with low final Kjeldahl that show high concentration of this metabolite respect to the others. This could reinforce the hypothesis that low pertactin productivity is associated with oxygen consumption.. Following such limitations glutamine consumption may be lower as compared to fermentations where such limitations do not occur. It should be noted that the analysis performed here was conducted at the end of the fermentation. A more detailed assessment of metabolic changes would require measuring metabolite levels during the course of the fermentation process. This would be a very challenging task since such measurements cannot be done currently on-line.

4.6 Simple growth model to explain the observed variability

A possible explanation to the variability observed in the fermenters can be proposed based on a simplistic model of the fermentation process. The biomass and substrate mass balances in batch are given as follows:

$$\frac{dx}{dt} = \mu x \tag{4.1}$$

$$\frac{dG}{dt} = KGx \tag{4.2}$$

where x is the biomass concentration, μ is the growth rate constant, G is the substrate concentration and K is a kinetic constant. A graphical solution of these equations for two different initial biomass concentrations is given in Figure 4.28.

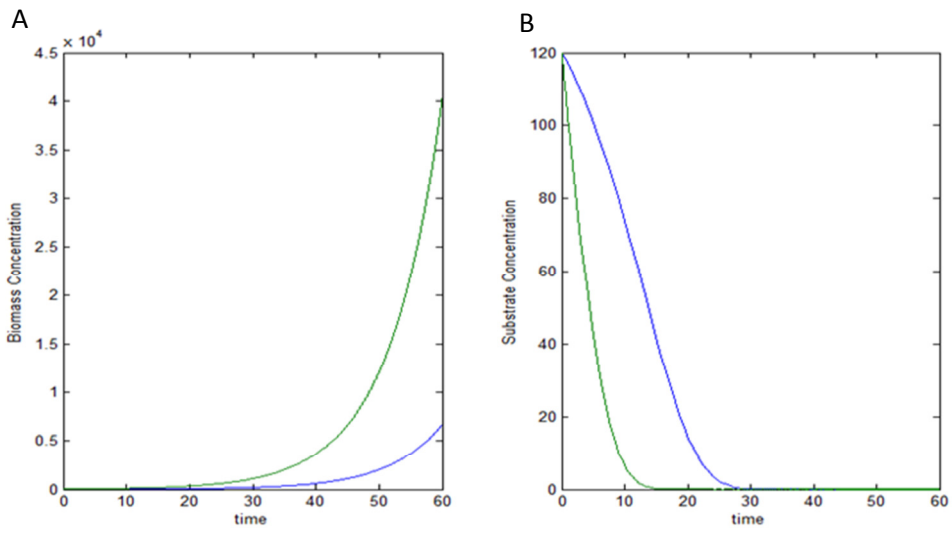

Figure 4.28 A: Biomass concentration vs. time; **B:** Substrate concentration vs. time.

Based on these profiles, it can be observed that the smaller the inoculum (small x_o), the start of the supplementation time will occur later as observed in the process. Also, slower consumption of glutamine for a smaller initial biomass concentration will explain the observed lower aeration and stirring speed rates observed for the different runs. Thus, variability in the inoculums could explain some of the observed operational differences. In general, it was observed that low productivity was associated with slower oxygen consumption rates and later supplementation times. Lower oxygen consumption may explain the higher levels of NADH observed by fluorescence. Also, such lower consumption would be correlated to lower levels of ATP with resulting lower productivity. However, these correlations were not completely consistent for all investigated batches.

4.7 Regression models between PCA scores obtained from fermentation samples and pertactin productivity

This section describes the formulation of a regression model among the averages of PC1 scores and PC2 scores obtained for samples collected from the fermenters, and final Kjeldahl values. This model was done for assessing whether there is a statistically significant correlation between fluorescence fingerprints in the fermenters and final pertactin productivity (Figure 4.29).

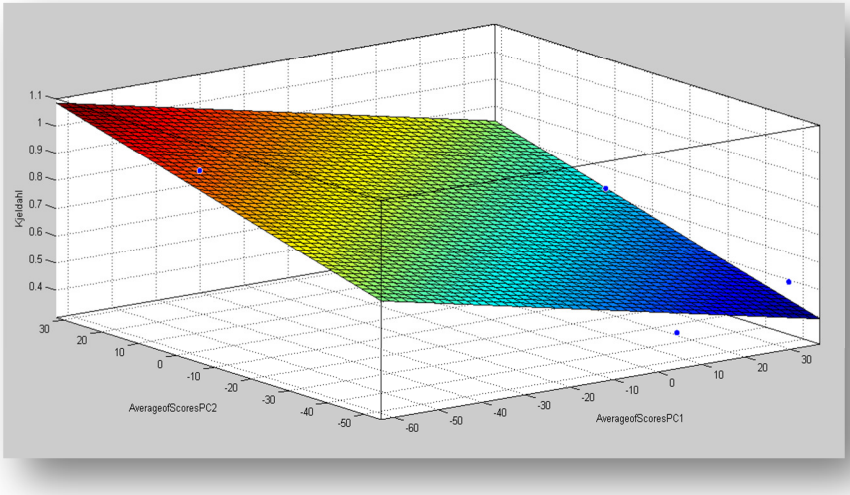

Figure 4.29 Regression among PC1 scores average and PC2 score average in the fermenters, and final Kjeldahl values.

The resulting equation is:

$$f(x,y) = a - b^2 \cdot x + c \cdot y \tag{4.3}$$

R^2: 0.8753

Where:

a = 4.2 (3.592, 4.808)

b = 0.1393 (0.06899, 0.2095)

c = 0.02326 (0.001802, 0.04472)

(Coefficients with 95% confidence bounds)

x = Average of PC1 Scores (protein content in the fermenters)

y = Average of PC2 Scores (NAD(P)H-enzyme content in the fermenters)

$f(x,y)$ = Protein content measured by Kjeldahl (PRN in the final purification step)

From the regression model, one can conclude that when the protein content is high, x (average of PC1 scores) is negative, therefore $f(x,y)$ increases. On the other hand, when NAD(P)H content is high, y (average of PC2 scores) is negative, therefore $f(x,y)$ decreases.

The coefficient b was chosen to be squared to avoid zero as a possible value. When b is not squared in Equation 4.3, the 95% confident bounds for this coefficient are -0.03896 and 0.0001768, allowing zero as a possible value, which would mean that in some cases the protein content measured ($f(x,y)$) can be independent of x, i.e. protein content in fermenters, which is not valid.

A PLS model was also built using the average for PC2 scores related to NADH complex and Kjeldahl values. First, the regression was performed using the average of the scores for seven batches, resulting in a R^2=0.482 (Figure 4.30).

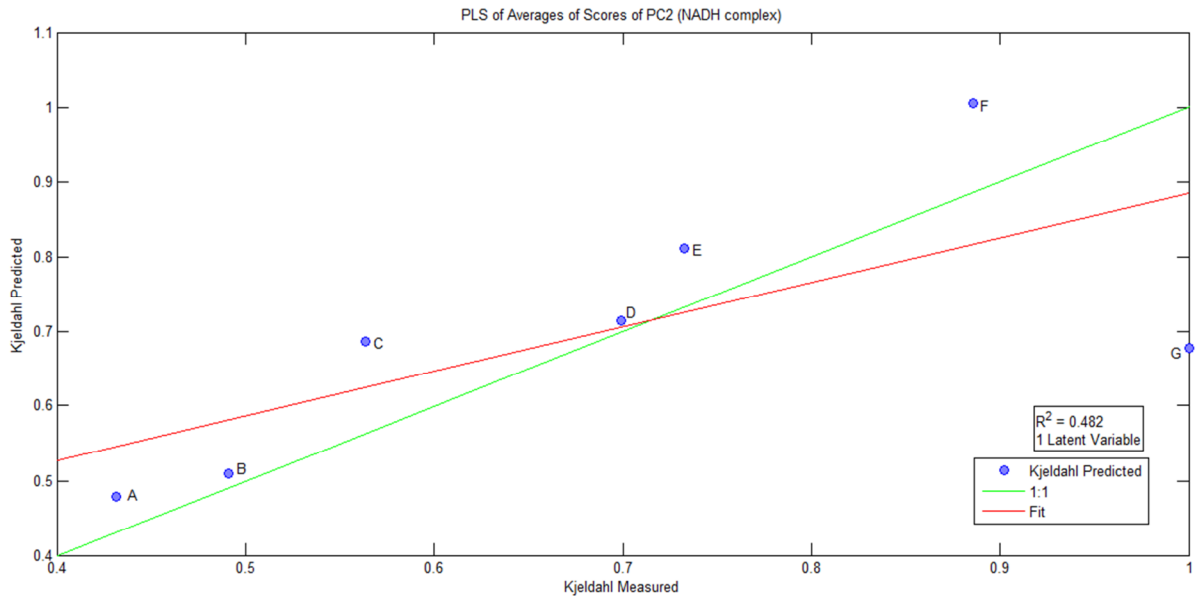

Figure 4.30 PLS regression between PC2 scores (NADH complex) and Kjeldahl values for 7 samples.

Since the model did not fit the data properly, a second model was constructed with six batches (eliminating batch G) which resulted in a goodness of fit of 0.94.

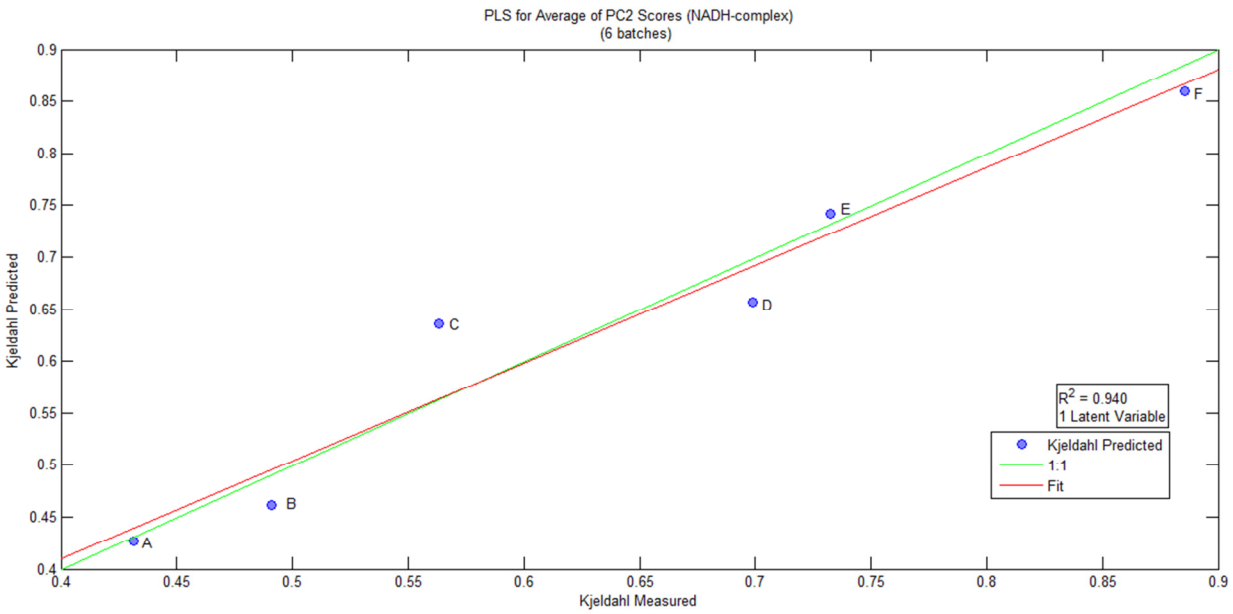

Figure 4.31 PLS regression between PC2 scores (NADH complex) and Kjeldahl values for 6 samples.

4.8 PLS based regression for predicting pertactin

An additional regression model was developed using Partial Least Square (PLS) for predicting the final pertactin concentration based on fluorescence measurements obtained from samples collected in the final step of the purification process. The objective was to see whether fluorescence can be used to predict the pertactin concentration when there are no additional proteins or species in the sample which may interfere with the pertactin-related fluorescence.

PLS generally provides better prediction accuracy for regression models that correlate noisy input data (fluorescence) to noisy output data (Kjeldahl). Besides capturing the variance, PLS tries to describe the data with a small number of latent variables (LV) that maximize covariance to achieve a correlation between two sets of data (Wise *et al*., 2007). Specifically, PLS was used to find a correlation between SKA-CH2 samples (matrix X) and final Kjeldahl values (matrix Y). The PLS model for this set of data

was expressed as a linear combination of three latent variables (LV). In this model, two more samples were added (batches K and L).

The R-squared value was 0.868 showing a positive linear association, between fluorescence intensities and PRN content. This model could be used to predict Kjeldahl values from fluorescence measurements.

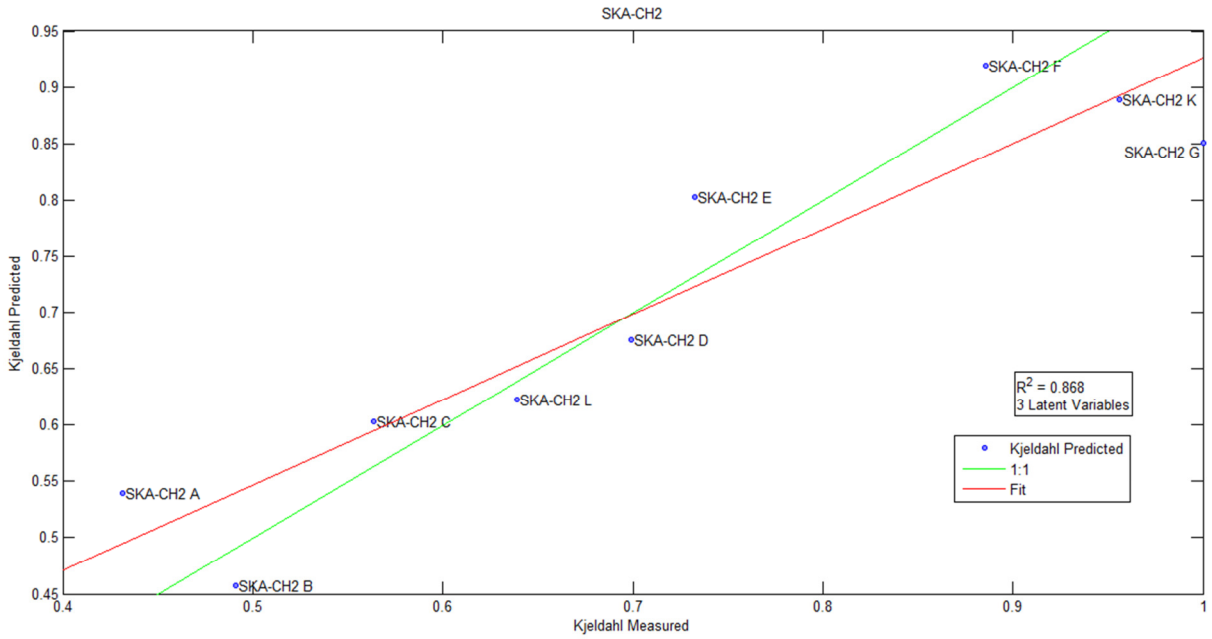

Figure 4.32 PLS regression between SKA-CH2 and Kjeldahl values.

4.9 Dynamic fluorescence analysis using samples collected along the fermentation process

In order to further elucidate the sources of variability between the fermentations and the effect on productivity, the occurrence of fluorophores were monitored for different time instances along the fermentation. Specifically fluorescence spectra were collected at different times for fermentation batches H and I. In general, all the spectra show the same pattern. The spectra obtained for batch H are shown in Figures 4.33 and 4.34, and for batch I are shown in Figures 4.35 and 4.36.

The spectra measured for samples taken before the addition of the supplement (t~29 hrs) show emissions related to the broth components (δ, δ', γ) such as tyrosine and other amino acids, and FAD (ϕ) (Figure 4.31). The peak characteristic of NAD(P)H-dehydrogenase complex (β') appears only after the supplement addition (t = ~29 hrs). Even though this peak appears in fermentation batches with low and high productivity, it was hypothesized that it might be an indicator of the existence of reactions triggered at the beginning of the fed-batch process that could be affecting the production of antigens, and therefore the final yield of product.

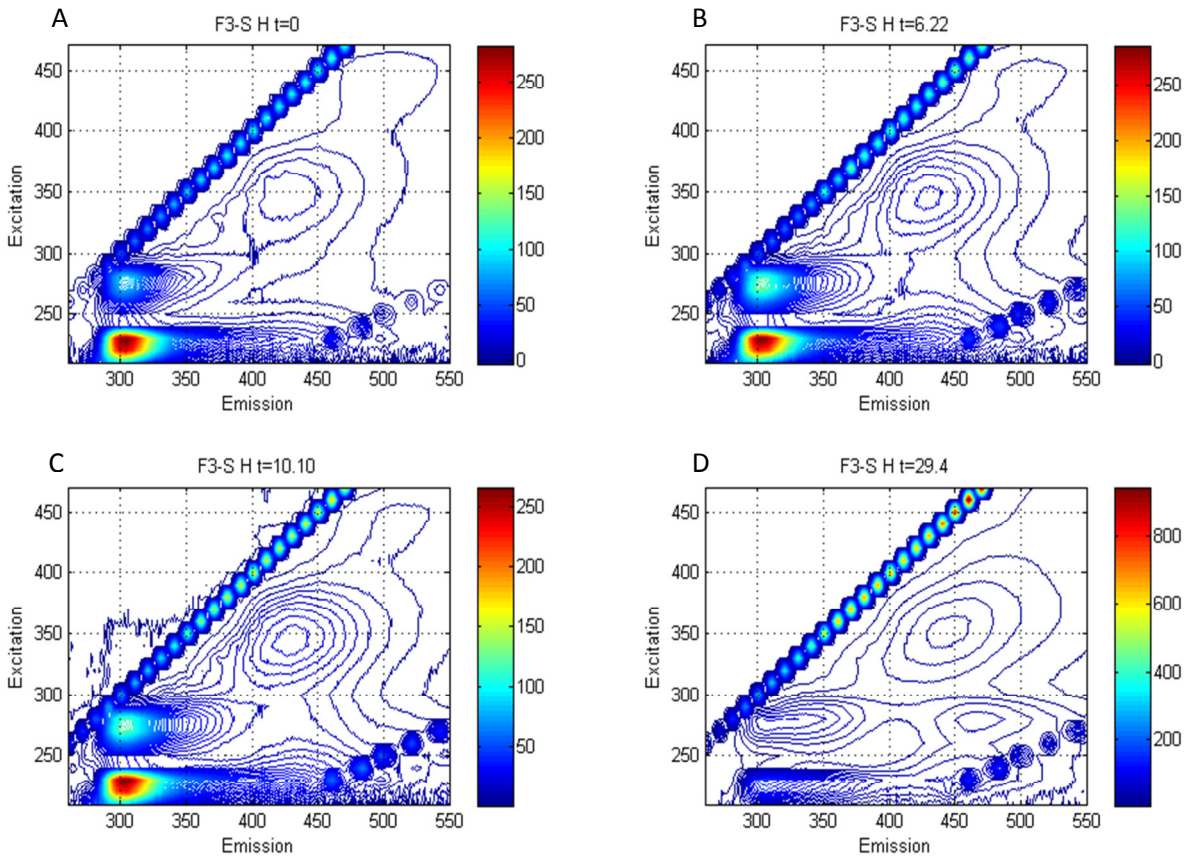

Figure 4.33 EEMs for fermenter F3 (batch H) at different fermentation times: A) 0 hrs; B) 6.22 hrs; C) 10.10 hrs; D) 29.4 hrs.

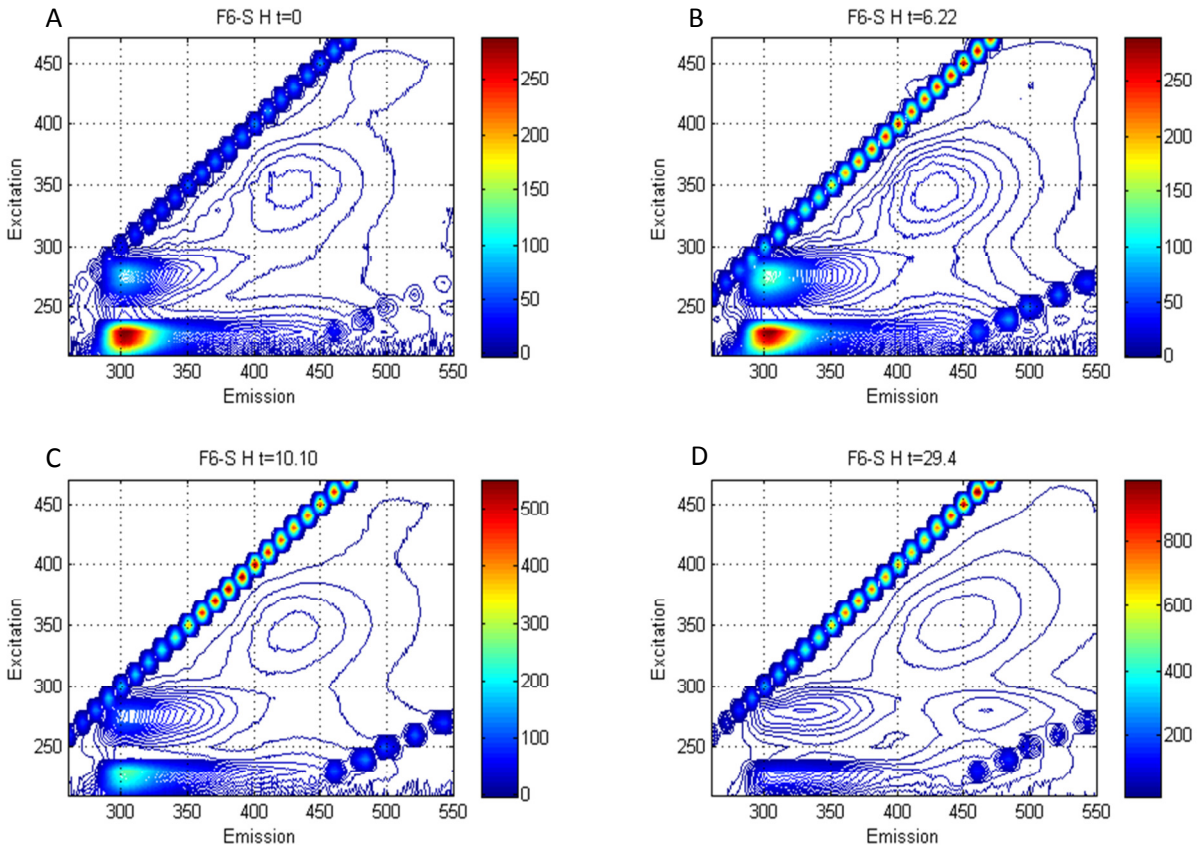

Figure 4.34 EEMs for fermenter F6 (batch H) at different fermentation times: A) 0 hrs; B) 6.22 hrs; C) 10.10 hrs; D) 29.4 hrs.

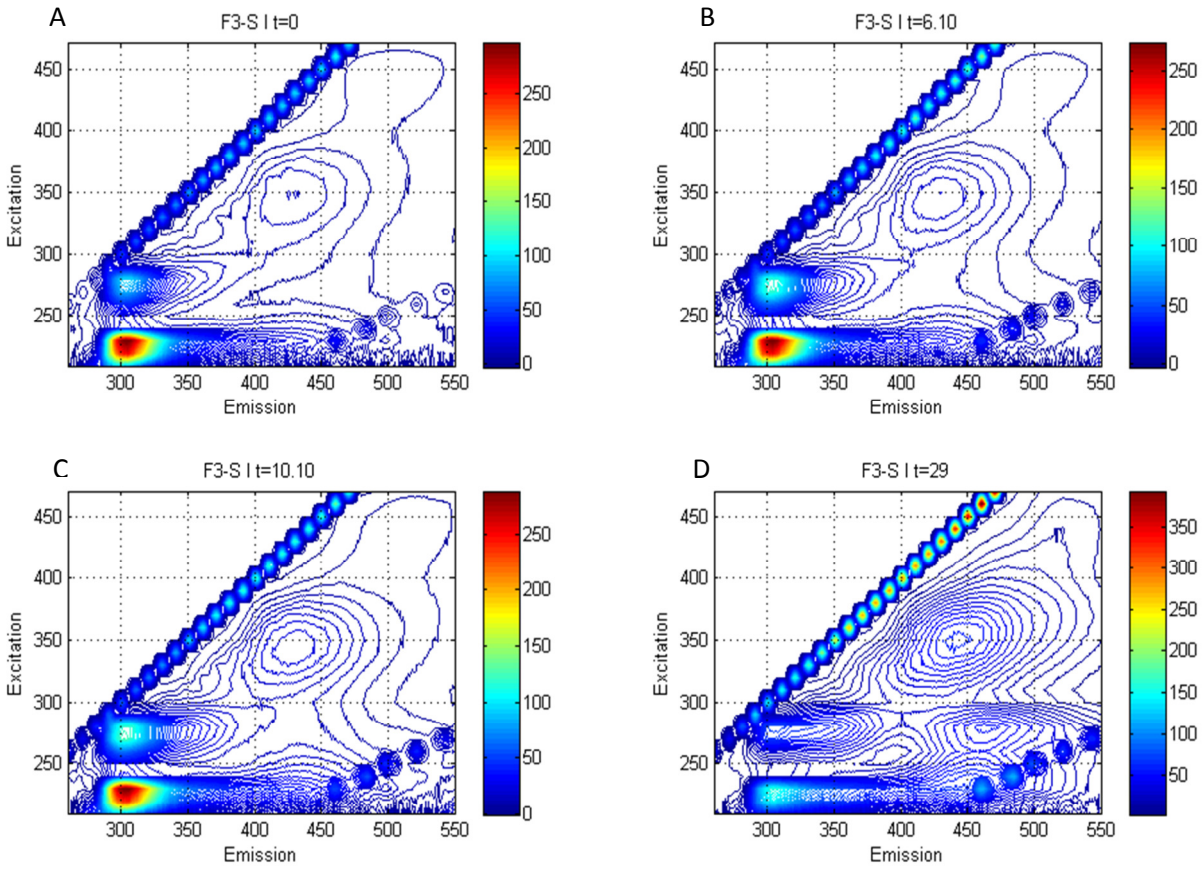

Figure 4.35 EEMs for fermenter F3 (batch I) at different fermentation times: A) 0 hrs; B) 6.10 hrs; C) 10.10 hrs; D) 29 hrs.

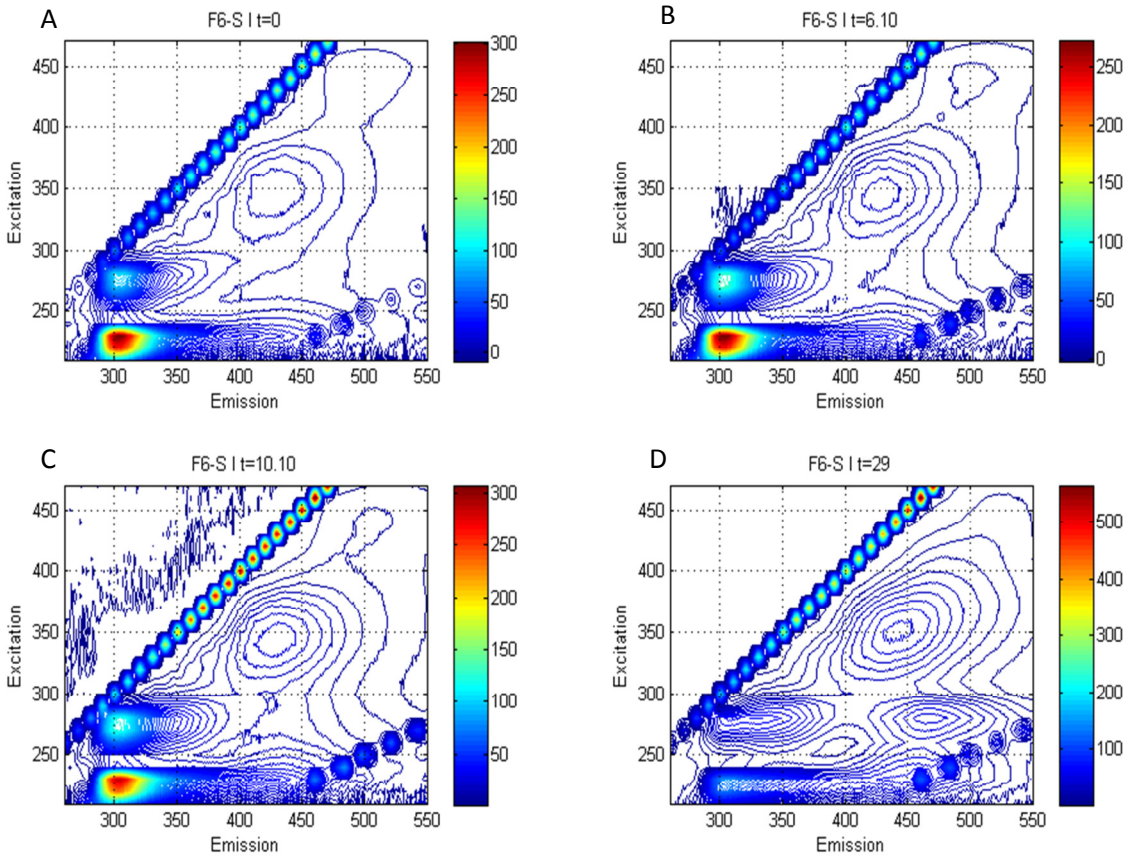

Figure 4.36 EEMs for fermenter F6 (batch I) at different fermentation times: A) 0 hrs; B) 6.10 hrs; C) 10.10 hrs; D) 29 hrs.

4.10 Enzyme Assay

In an effort to identify the type of dehydrogenase that might be binding to the NAD(P)H, substrates for different dehydrogenases were used to measure changes in the fluorescence intensities. Because of the high fluorescence intensities observed in the peak characteristic of NAD(P)H-dehydrogenase, CPS-Conc-PW was employed to conduct this analysis.

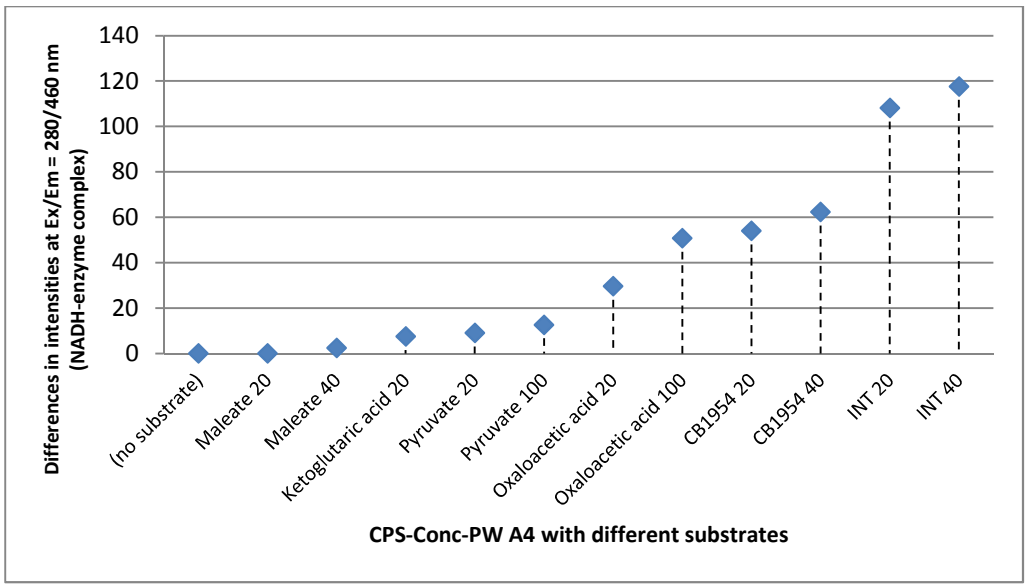

Figure 4.37 Differences in intensities at Ex/Em = 280/460 nm (NADH-enzyme complex) between the emission of the original sample (CPS-Cond-PW A4) and the sample spiked with a substrate**.**

Figure 4.37 shows the differences in the fluorescence intensities at Ex/Em = 280/460 nm between the intensity of CPS-Conc-PW A4 and the intensity for the same sample spiked with different substrates. For each enzyme substrate, three different dilution factors were used (section 3.2.1). In Figure 4.37, a dilution factor (DF) of 0.007 is indicated as 20, a DF of 0.013 is indicated as 40 and a DF of 0.033 is indicated as 100. For maleate, α-ketoglutaric acid and pyruvate no significant changes were observed. In contrast, for oxaloacetic acid, CB1954 (substrate for ubiquinone or complex I) and INT (which get reduced with some enzymes as succinate dehydrogenase) significant changes were observed (Deng *et al*., 1992).

Complex I, also named NADH:ubiquinone oxidoreductase, is an enzyme that participates in the first step of the respiratory chain of bacteria (Roos *et al*., 2000). It is a membrane-bound molecule that facilitates the transport of electrons between NADH and ubiquinone, and also the transport of protons out of bacterial cytosol (Roos *et al*., 2000). The presence of this enzyme could be explained by the existence of binding between the Complex I and cell membrane debris, which are filtered in the first ultrafiltration step of the PRN purification (CPS-Conc-PW) (Roos *et al*., 2000).

As a preliminary approach, these fluorescence results provided an idea of the enzymes that could be present and bound to NAD(P)H, but the results are not conclusive since appreciable changes in the intensities were observed for more than one substrate. Fluorescence emission of CPS-Conc-PW combined with NADH was also measured. After the NADH addition, the peak at Ex/Em~280/470 nm increased as expected. Also, there is a decrease at Em~340 nm because of the quenching effect of NADH.

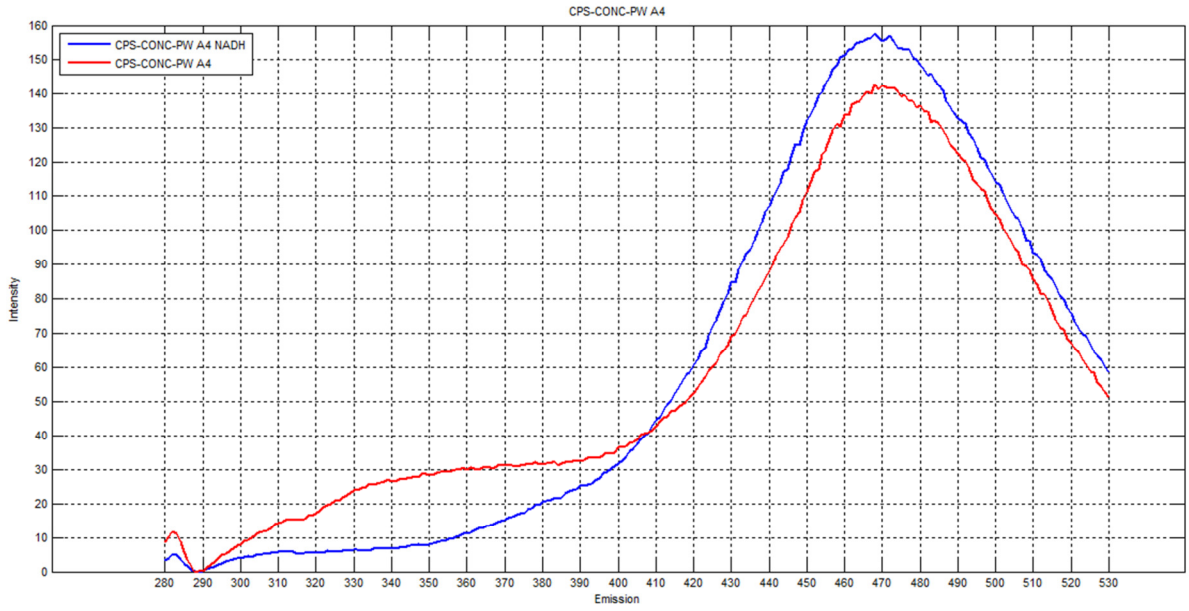

Figure 4.38 CPS-Conc-PW A4 and NADH.

To investigate the existence of binding between PRN and NADH, a sample of pure PRN (SKA-CH2) was spiked with NADH.

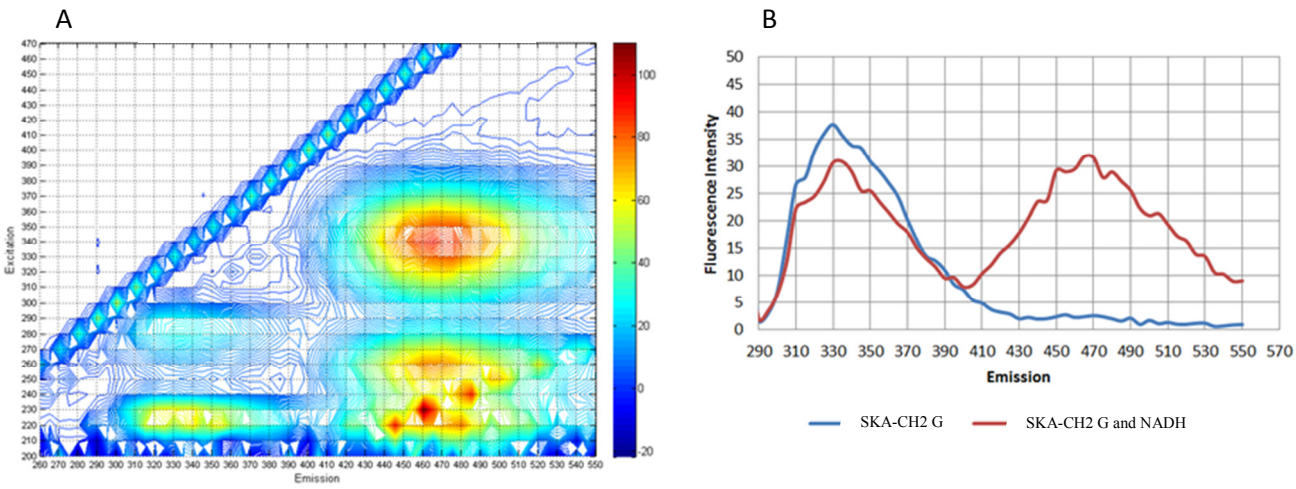

Figure 4.39 A: Spectra of SKA-CH2 G and NADH (4 μM); **B:** Spectra at Ex=280 nm of SKA-CH2 G (in solution and spiked with NADH).

Figure 4.39-A shows the Trp peak corresponding to PRN at Ex/Em = 230-280/335 nm and the NADH peak at Ex/Em = 260-340/460 nm. The NADH peak at Ex=260 nm is not shifted to Ex=280 nm, indicating that binding did not occur with this antigen. On the other hand, in Figure 4.39-B it is shown the same spectrum at Ex=280 nm for a sample of SKA-CH2 G spiked with NADH. In this case, as expected, the emission of Trp is quenched by the NADH. Thus, it is possible to conclude that NADH does not bind to pertactin but to some other species.

4.11 Assessment of the effect of temperature on fluorescence

Process samples were subjected to different changes of temperature in order to assess the influence of storage and handling conditions. In Figure 4.40, fluorescence spectra for two SPEP-R-Conc samples after changes in temperature are shown.

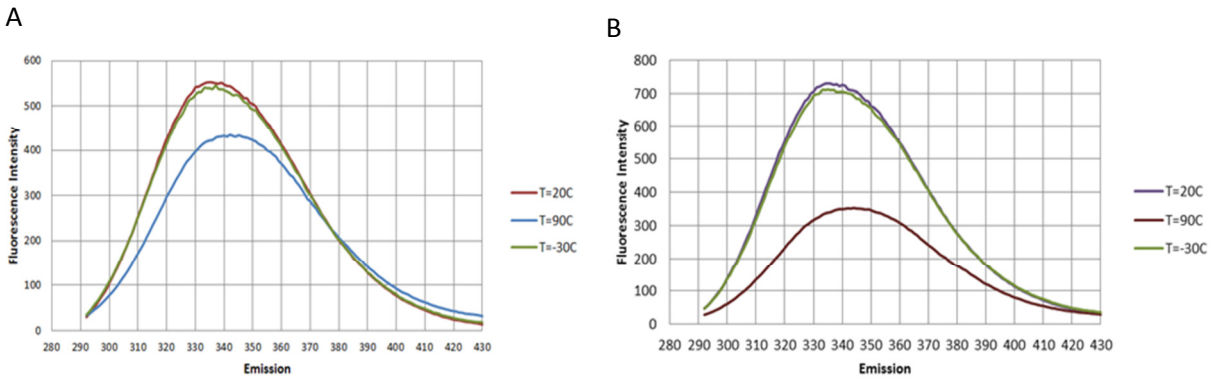

Figure 4.40 A: SPEP-R-Conc A4 at three different temperatures: -30ºC, 20ºC, 90ºC
B: SPEP-R-Conc A3 at three different temperatures: -30ºC, 20ºC, 90ºC

The samples were left at -30ºC for 24 hrs, and at 90ºC for 1 min. The sample at -30ºC was brought to 20ºC by rapid thawing. From the two plots it can be concluded that freezing does not have a major effect on the fluorescence characteristics. It is evident though that a decrease and shift of the fluorescence emission occurred at higher temperatures. These changes in the emission are possibly due to protein denaturation and unfolding of the protein (Gally *et al*., 1962).

Figure 4.41 provides the fluorescence spectra acquired for CPS-B2-(F3) samples after cooling at 4ºC and raising the temperature to 25ºC, repeatedly for three days, are shown.

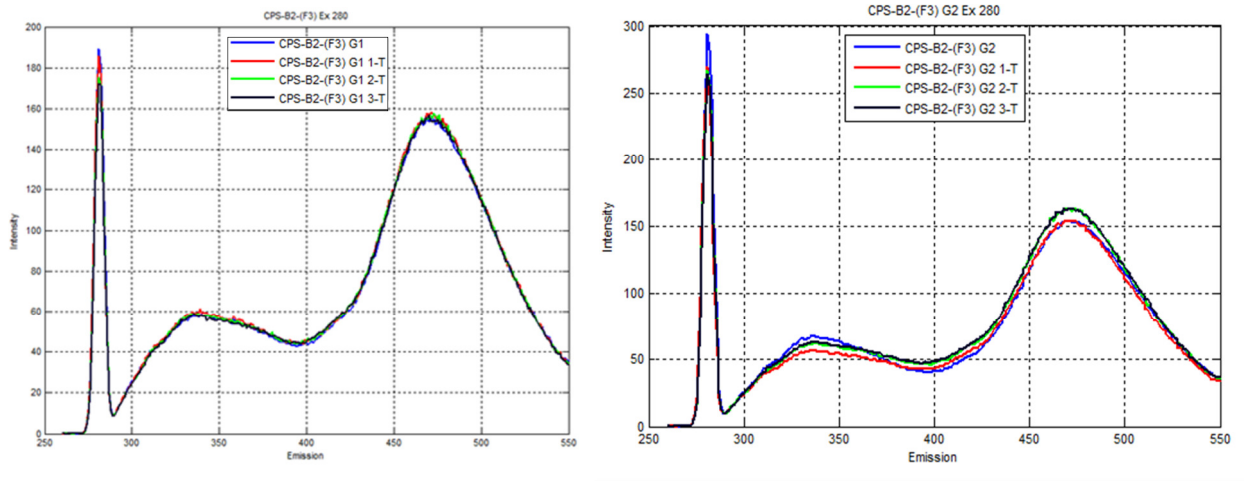

Figure 4.41 A: CPS-B2-(F3) G1 after cooling (at 4°C) and back to 25ºC three times; **B:** CPS-B2-(F3) G2 after cooling (at 4°C) and back to 25ºC three times.

In the two plots of Figure 4.41, there are no significant differences between the original sample and the samples subjected to changes of temperature. Figure 4.41-B exhibits small differences in the curves, but these are attributed to noise.

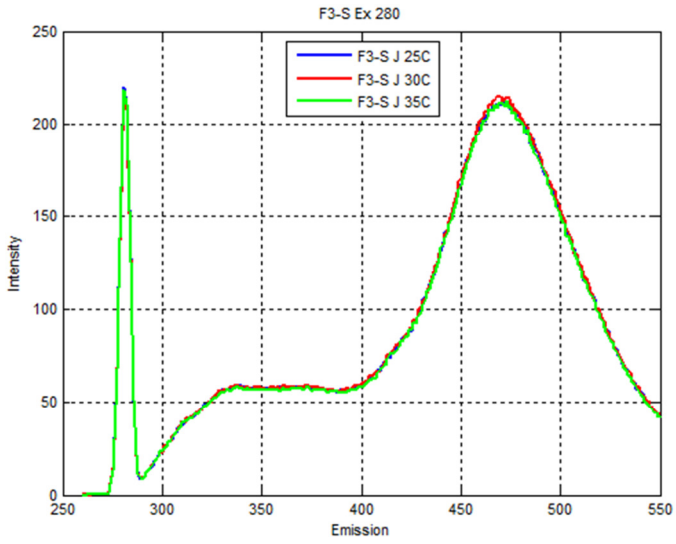

Figure 4.42 F3-S J after 5ºC changes during the same day.

Figure 4.42 shows the spectra for a F3-S J sample after undergoing changes of 5ºC in temperature. There are no appreciable differences among the curves.

4.12 Centrifugation of Samples

In Figure 4.43, fluorescence spectra of fermenter samples for batches G1 and G2, before and after centrifugation are presented. This effect was investigated to assess whether stirring of the samples during shipment could affect the fluorescence spectra.

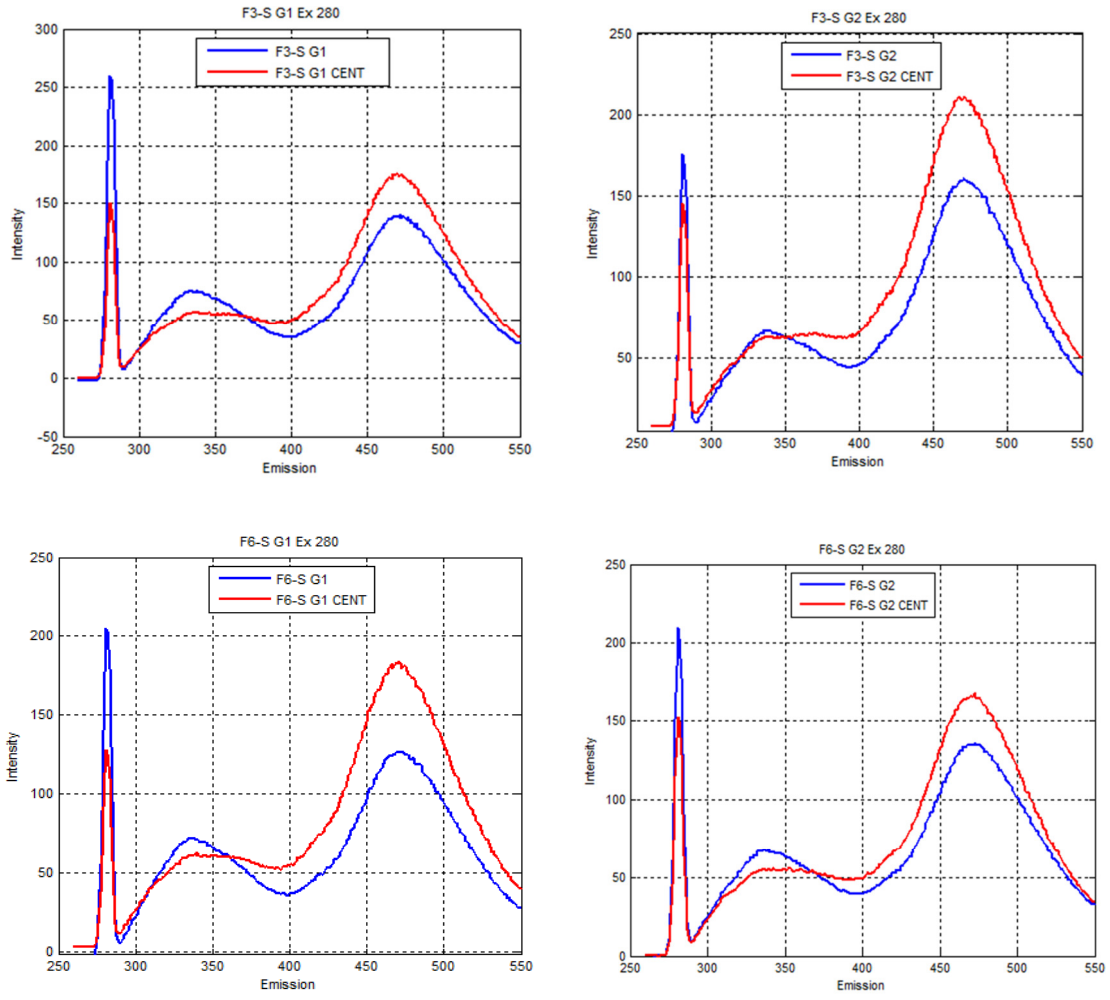

Figure 4.43 Fluorescence spectra of fermenter samples for batches G1 and G2, before and after centrifugation.

Samples that were centrifuged are shown in red, and the original samples are marked with a blue line. The samples were centrifuged for 5 min at 23,000 xg. In the four plots, it is evident that a decrease in the peak of tryptophan and an increment of the intensity in the peak of NADH occurred after centrifugation.

4.13 Surface Phosphophilicity Assessment

As discussed in Chapter 1, surface phosphophilicity was investigated to assess the possibility for monitoring adjuvant potency during storage using a fluorescence-based assay. Surface phosphophilicity was assessed using the fluorogenic substrate 6,8-difluoro-4-methylumberiferyl (DiFMU) which is the leaving group of 6,8-difluoro-4-methylumberiferyl phosphate (DiFMUP) after hydrolysis (Zhao *et al*., 2001).

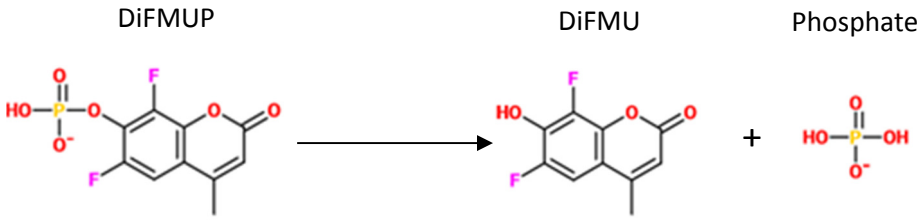

Figure 4.44 Reaction for DiFMUP hydrolysis.

This reaction is catalyzed by alum (aluminum-contained adjuvant). The assessment was done in a microplate reader with an Ex/Em=358/450 nm (characteristic excitation emission values for DiFMU) (Zhao *et al*., 2001). Three adsorbed antigens were assessed: pertactin (PRN) from batch K (February 15, 2011); difteria toxin (DCA) from batch L (September 12, 2011); tetanus toxin (TCA) from batch M (August 30, 2011), and two buffers: saline and PBS. The fluorescence signals acquired are shown in Figure 4.43.

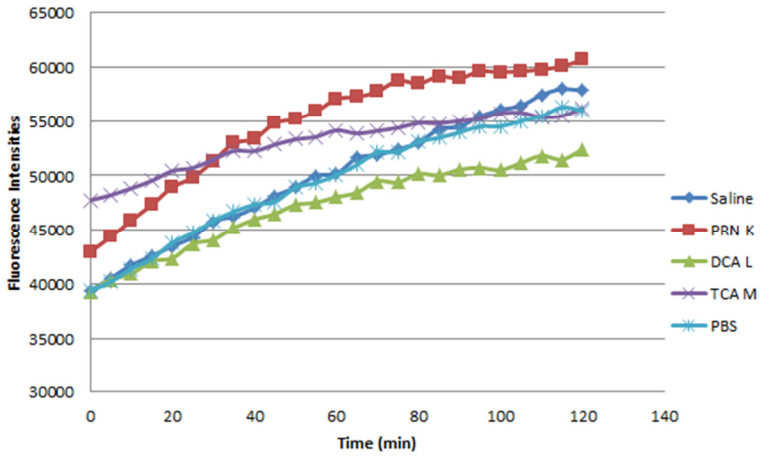

Figure 4.45 DiFMUP hydrolysis.

Comparing the spectra obtained from the hydrolysis of DiFMUP in the two buffers, saline and PBS, the differences are small. Differences were appreciable in the hydrolysis carried out in presence of the adsorbed adjuvants. It was expected that, in case of an efficient adsorption of the adjuvant in the antigen, the aluminum based adjuvant would not be capable of totally catalyzing the reaction, because its active sites would be cleaved in the phosphate groups of the antigen, and therefore a decrease of emission would be noticed.

There is relatively more increase in intensities of PRN, saline and PBS compared to DCA and TCA. PRN is not entirely adsorbed whereas the adsorption of TCA and DCA are considerable. When there is free adjuvant, DiFMUP is hydrolyzed and that reflects in higher intensities. Unabsorbed antigen would not be obstructing or competing with DiFMUP. DiFMUP hydrolysis by adjuvant is dependent on the availability of free adjuvant.

Chapter 5
Conclusions and Future Work

The thesis presents a novel application of fluorescence spectroscopy in combination with multivariate analysis for diagnosing and identifying the sources of variability in a vaccine manufacturing process. The work also proposes the use of fluorescence for quantifying the amount of protein product in the final stages of the process. Attempts were made to correlate the variability in protein production with operating conditions of the process. Finally, the use of fluorescence for assessing efficacy of vaccines' adjuvants was assessed.

5.1 Fluorescence Analysis

This research shows that fluorescence analysis is effective for off-line monitoring of a vaccine manufacturing process. By the acquisition of fluorescence signals in samples collected from different stages of upstream and downstream processes of a Component Pertussis vaccine, it was possible to identify at the same time fluorescent compounds, such as amino acids present in cells and media, cofactors and other nutrients, which makes this technique a valuable tool for rapid qualitative analysis. Fluorescence was also useful in identifying possible conformational transitions in the proteins based on the fluorescence emission of tryptophan (longer wavelength emission when it is more exposed to the environment). The fluorescence signals of enzyme cofactors were helpful for identifying the existence of complexes formed by the interactions between NAD(P)H and dehydrogenases, based on changes in the NAD(P)H emissions at lower excitations (shift from Ex=260 nm to Ex=280 nm at Em=460 nm).

5.2 PCA analysis

The fluorescence-based calibration models were built using multivariate statistical modelling tools, PCA and PLS. PCA was found to be effective for recognizing correlations between conditions at different steps of the manufacturing process with the pertactin concentration at the end of the manufacturing process.

PLS was used to predict the final protein content that can be measured with Kjeldahl analysis. PCA was able to extract the most important information related to the fluorophores and to compress this information into two or three principal components that captured most of the variance. This analysis was performed by gathering the fluorescence data in different arrangements: i-one matrix containing the spectra collected from samples of the entire process; ii-one matrix containing upstream samples; iii-three different matrices with downstream samples: a matrix with all samples from downstream, a matrix excluding downstream waste samples, and a matrix with downstream waste samples. The main idea of partitioning the data according to different stages of the manufacturing process was to try to find a possible correlation between the fluorescence corresponding to the different production steps and the final PRN concentration in the final stage of the purification in order to identify possible causes of yield variability for this antigen (pertactin).

From the PCA analysis of the upstream matrix, it was noticed that the amount of NAD(P)H-enzyme complex was high in the fermenters of the batches with low productivity, suggesting that certain types of reactions are taking place in the fermentation steps that might be causing variability in the final yield of PRN. NAD(P)H acts as electron donor and oxygen acts as acceptor during oxidative phosphorylation needed for ATP formation. It is hypothesized that as a result of oxygen transfer limitations, NAD(P)H cannot be oxidized and therefore it might accumulate. Since there could be less corresponding formation of ATP, the production of antigens such as PRN might be diminished.

In the analysis of the three matrices containing downstream data, it was not possible to find any pattern that differentiates batches with low final Kjeldahl from batches with high final Kjeldahl. In contrast to the upstream analysis, the peak of NAD(P)H-enzyme binding was not significant. The score plots of these set of data exhibited a lack of correlation between scores (samples) and productivity. For this reason, downstream purification protocols were dismissed as a cause of low productivity.

When performing PCA analysis in the data matrix with samples from the entire process, it was difficult to find a correlation with productivity. Also, the principal component related to the NAD(P)H-enzyme complex captured only 5.85% of the variance, compared to the principal component for the same fluorophore in the separated upstream data, where the variance captured was 23.85%. This suggests that the NADH-enzyme complex only have a significant influence in the variability of the process related to the upstream steps, and not to the entire process.

5.3 Correlation among productivity and dissolved oxygen, aeration, agitation and supplementation time profiles recorded during fermentation

Changes in the dissolved oxygen, aeration, agitation and supplementation time profiles were investigated as a possible causative of accumulation of NAD(P)H in the fermenters. These parameters are continuously monitored by Sanofi Pasteur in all fermentations.

No correlation could be found between dissolved oxygen and low productivity. This parameter is maintained relatively constant at its set point by means of closed loop control by manipulating the agitation speed initially followed by the aeration rate. However, in the aeration (airflow) profiles, some differences among batches were found, but the relationship between the aeration levels and variability in protein concentration was not consistent, e.g. in some fermenters that had high final Kjeldahl, the aeration rate was low.

Similarly, the relationship between agitation profiles and low productivity was not consistent. However, it should be remembered that the purification process is applied to one batch that is assembled together from four fermentation batches, so it might be possible that one or more of these batches did not provide high protein production even though the final Kjeldahl, corresponding to the combination of four batches, was high.

It should be noticed that to calculate the actual consumption of oxygen by the cells it would be necessary to measure, in addition to dissolved oxygen and aeration rates, the oxygen leaving the fermenters which cannot be measured in the current operation. Thus, measuring the NADH by fluorescence offers a distinctive advantage for identifying differences in metabolism in particular as related to consumption of oxygen.

The supplementation time, or the time at which fed-batch operation starts, depends on the dissolved oxygen levels. When a sudden increased of the dissolved oxygen is observed, it is assumed that the substrate, in this case glutamine, is consumed by cells, and additional supplementation is needed. In the supplementation profiles, it was noticeable that some batches had a delay at the start of the fed-batch phase. Even though, this delay did not happen in all the fermenters with low productivity, it is important to highlight that the supplement addition is done manually, and it might not occur exactly at the moment when dissolved oxygen increases.

To further assess changes in the cell metabolism key metabolites, such as lactate, glucose, and glutamine were measured at the end of different batches. For lactate and glucose levels, no appreciable correlation could be found between their levels and PRN productivity. In contrast, high levels of glutamine generally occurred for batches with low PRN productivity, i.e. low substrate consumption, which at the same time can be associated with oxygen consumption rates.

Since all these measurements are performed at the end of the fermentation, an additional fluorescence analysis was done at different time instances along two fermentation runs. From these spectra, it was observed that the peak characteristic of NAD(P)H-enzyme complex appeared after the start of the supplement addition, i.e. fed-batch phase of cultivation. Even though this peak appears in fermentation batches with low and high productivity, it might be an indicator of the existence of reactions triggered at the beginning of the fed-batch process that could be affecting the production of antigens, and therefore the final PRN yield.

5.4 Regression models

5.4.1 Regression between PCA scores obtained from fermentation samples and pertactin concentration

The two regressions built between PCA scores obtained from fermentation samples and pertactin concentration, demonstrated that there is a statistical significant correlation between these parameters. One regression was made between the average of PC1 scores (inversely correlated to protein content), the average of PC2 scores (inversely correlated to NAD(P)H-enzyme content) in the fermenters and the final content of PRN measured by Kjeldahl analysis (in the last step of the purification). The goodness of fit was R^2=0.8753.

A second regression was built using PLS analysis. In this case, the averages of PC2 scores were employed. It was found that when using only six batches, the goodness of fit was high (R^2=0.94), but when using seven batches it was R^2=0.482.

5.4.2 PLS for predicting pertactin concentration

An additional regression model was calculated using PLS for predicting the final pertactin concentration from fluorescence measurements obtained from samples collected in the final step of the purification process. The objective was to assess whether fluorescence can predict the pertactin concentration when there are no additional proteins or species in the sample which may interfere with the pertactin related fluorescence.

A good correlation among PRN content and fluorescence intensities was obtained, with a R^2=0.868, demonstrating the feasibility of using fluorescence as a rapid method to predict protein concentration with fluorescence.

5.5 Enzyme analysis

In order to identify the enzyme dehydrogenase bound to NAD(P)H, substrates for different enzymes were used to measure changes in the fluorescence intensities. Changes after addition of substrates for oxaloacetate, ubiquinone and succinate, were the most significant. However, these results are not conclusive because changes were observed with more than one substrate.

5.6 Assessment of the effect of temperature and centrifugation in fluorescence

In the assessment of temperature, appreciable changes were observed only when the samples were heated up to 90ºC with a decreased in the fluorescence intensities possibly due to protein unfolding. When the samples where frozen, cooled and warmed, some small changes were observed attributed to noise.

It was also found that a decrease in the peak of tryptophan and an increment of the intensity in the peak of NADH occurred after centrifugation. This could be due to the exposure of Trp residues as a consequence to the unfolding of the protein, with a subsequent release of energy, and energy transfer to NADH.

5.7 Surface phosphophilicity assessment

Spectrofluorometry-based measurements to assess surface phosphophilicity of aluminum-based adjuvants in adsorbed samples were found to be sensitive. From differences in fluorescence spectra when an adsorbed sample is tested with a fluorogenic compound, e.g. DiFMU (leaving group after hydrolysis of DiFMUP), it is possible to infer the efficiency of adsorption since in case of an efficient adsorption of the adjuvant in the antigen.

5.8 Future Work

While fluorescence provided indications that the variability in PRN production is related to the fermentation process, the sources of the variability have not been identified.

It is proposed to identify the cause of the variability through a series of different studies that will attempt to map the intracellular conditions within the cells as well as the extracellular operating conditions, by using fluorescence and other methods. It is believed that the continuation of the current project may serve to develop a comprehensive approach for diagnosing variability in this and other fermentation processes.

The following future research directions are proposed:

1- Monitoring the dynamic evolution of fluorescence along the fermentation processes to identify the fermentation stage which is most responsible for the observed variability.
2- Assessment of a fiber optic probe for online monitoring for collecting fluorescence measurements.
3- Since the preliminary results revealed a correlation between large amounts of extracellular NADH in the fermenters with low PRN yield, the aim is to identify the sources of the difference in metabolic behaviour that may lead to this occurrence. As part of this study, it is proposed to investigate the following: (i) intracellular measurements of NAD^+/NADH; (ii) intracellular measurements of ATP, enzymes (e.g. dehydrogenases) and organic acids that participate in the TCA cycle and can be correlated with the observed changes in NADH; (iii) measurements of ORP (oxidation redox potential) associated to NADH and ATP intracellular levels.
4- Assessing the use of fluorescence to monitor vaccine characteristics during storage. The potential of fluorescence to assess aggregation and/or changes in stored formulated vaccine samples could be assessed by monitoring changes in the fluorescence spectra for samples stored for different lengths of time and under various conditions.

Bibliography

Abdi, H., Williams, L. J., Abdi, H., & Williams, L. J. Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics, 2*(4), 433-459.

Albani, J. (2007). New insights in the interpretation of tryptophan fluorescence. *Journal of Fluorescence, 17*(4), 406-417.

Babu, M. M., Bhargavi, J., Saund, R. S., & Singh, S. K. (2001). Virulence factors of *Bordetella pertussis*. *Current Science, 80*(12), 1512-1522.

Beechem, J. M., & Brand, L. (1985). Time-resolved fluorescence of proteins. *Annual Review of Biochemistry, 54*(1), 43-71.

Bieri, O., Wirz, J., Hellrung, B., Schutkowski, M., Drewello, M., & Kiefhaber, T. (1999). The speed limit for protein folding measured by triplet-triplet energy transfer. *Proceedings of the National Academy of Sciences of the United States of America, 96*(17), 9597-9601.

Bilski, P., Li, M., Ehrenshaft, M., Daub, M., & Chignell, C. (2000). Vitamin B6 (pyridoxine) and its derivatives are efficient singlet oxygen quenchers and potential fungal antioxidants. *Photochemistry and Photobiology, 71*(2), 129-134.

Bolgiano, B., Crane, D. T., Xing, D., Williams, L., Jones, C., & Corbel, M. J. (1999). Physico-chemical Analysis of Bordetella pertussis Antigens. *Biologicals, 27*(2), 155-162.

Braselmann, E., & Clark, P. L. (2012). Autotransporters: The Cellular Environment Reshapes a Folding Mechanism to Promote Protein Transport. *The Journal of Physical Chemistry Letters, 3*(8), 1063-1071.

Bryan, A. W.,Jr, Starner-Kreinbrink, J. L., Hosur, R., Clark, P. L., & Berger, B. (2011). Structure-based prediction reveals capping motifs that inhibit beta-helix aggregation. *Proceedings of the National Academy of Sciences of the United States of America, 108*(27), 11099-11104.

Budman, H., Patel, N., Tamer, M., & Al-Gherwi, W. (2013). A dynamic metabolic flux balance based model of fed-batch fermentation of *Bordetella pertussis*. *Biotechnology Progress, 29*(2), 520-531.

Cabiscol, E., Tamarit, J., & Ros, J. (2010). Oxidative stress in bacteria and protein damage by reactive oxygen species. *International Microbiology, 3*(1), 3-8.

Calvet, A., Li, B., & Ryder, A. G. (2012). Rapid quantification of tryptophan and tyrosine in chemically defined cell culture media using fluorescence spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis, 71*, 89-98.

Charbonneau, M., Côté, J., Haurat, M. F., Reiz, B., Crépin, S., Berthiaume, F., Mourez, M. (2012). A structural motif is the recognition site for a new family of bacterial protein O-glycosyltransferases. *Molecular Microbiology, 83*(5), 894-907.

Charles, I., Fairweather, N., Pickard, D., Beesley, J., Anderson, R., Dougan, G., & Roberts, M. (1994). Expression of the Bordetella pertussis P.69 pertactin adhesin in Escherichia coli: fate of the carboxy-terminal domain. *Microbiology (Reading, England), 140* (Pt 12), 3301-3308.

Charles, I. G., Dougan, G., Pickard, D., Chatfield, S., Smith, M., Novotny, P., Fairweather, N. F. (1989). Molecular cloning and characterization of protective outer membrane protein P.69 from *Bordetella pertussis*. *Proceedings of the National Academy of Sciences of the United States of America, 86*(10), 3554-3558.

Chattopadhyay, A., & Mukherjee, S. (1999). Red edge excitation shift of a deeply embedded membrane probe: implications in water penetration in the bilayer. *The Journal of Physical Chemistry B, 103*(38), 8180-8185.

Chattopadhyay, A., Rawat, S. S., Kelkar, D. A., Ray, S., & Chakrabarti, A. (2003). Organization and dynamics of tryptophan residues in erythroid spectrin: Novel structural features of denatured spectrin revealed by the wavelength-selective fluorescence approach. *Protein Science, 12*(11), 2389-2403.

Corbel, M. J., Xing, D. K. L., Bolgiano, B., & Hockley, D. J. (1999). Approaches to the Control of Acellular Pertussis Vaccines. *Biologicals, 27*(2), 133-141.

Creed, D. (1984). The Photophysics and Photochemistry of the Near-UV Absorbing Amino Acids–II. Tyrosine and Its Simple Derivatives. *Photochemistry and Photobiology, 39*(4), 563-575.

Crowley-Luke, A., Reddin, K., Gorringe, A., Hudson, M. J., & Robinson, A. (2001). Formulation and characterization of Bordetella pertussis fimbriae as novel carrier proteins for Hib conjugate vaccines. *Vaccine, 19*(25–26), 3399-3407.

Degli Esposti, M. (1998). Inhibitors of NADH–ubiquinone reductase: an overview. *Biochimica et Biophysica Acta (BBA)-Bioenergetics, 1364*(2), 222-235.

Deng, H., Burgner, J., & Callender, R. (1992). Raman spectroscopic studies of the effects of substrate binding on coenzymes bound to lactate dehydrogenase. *Journal of the American Chemical Society, 114*(21), 7997-8003.

Denoel, P., Poolman, J., Carletti, G., & Veitch, K. (2002). Effects of adsorption of acellular pertussis antigens onto different aluminium salts on the protective activity in an intranasal murine model of *Bordetella pertussis* infection. *Vaccine, 20*(19), 2551-2555.

Desvaux, M., Hébraud, M., Talon, R., & Henderson, I. R. (2009). Secretion and subcellular localizations of bacterial proteins: a semantic awareness issue. *Trends in Microbiology, 17*(4), 139-145.

Desvaux, M., Parham, N. J., & Henderson, I. R. (2004). The autotransporter secretion system. *Research in Microbiology, 155*(2), 53-60.

Eftink, M. R. (2000). Intrinsic fluorescence of proteins. *Topics in fluorescence spectroscopy* (pp. 1-15) Springer, N.Y.

Elshereef, R., Budman, H., Moresoli, C., & Legge, R. L. (2008). Fluorescence-based soft-sensor for monitoring β-lactoglobulin and α-lactalbumin solubility during thermal aggregation. *Biotechnology and Bioengineering, 99*(3), 567-577.

Elshereef, R. (2009). Application of multi-wavelength fluorometry to monitoring protein ultrafiltration. Doctoral dissertation. University of Waterloo, Waterloo, Canada.

Emsley, P., Charles, I. G., Fairweather, N. F., & Isaacs, N. W. (1996). Structure of *Bordetella pertussis* virulence factor P. 69 pertactin. *Nature*, 381, 90-92.

Fawcett, C. P., Ciotti, M. M., & Kaplan, N. O. (1961). Inhibition of dehydrogenase reactions by a substance formed from reduced diphosphopyridine nucleotide. *Biochimica et Biophysica Acta, 54*(1), 210-212.

Ferrán, B., Proenza, L., Marcos, E., Pérez, C., & Campos, J. (2010). Expresión extracromosomal del antígeno pertactina en *Escherichia coli* y *Bordetella pertussis*. *Revista CENIC Ciencias Biológicas, 41*(1), 53-60.

Foster, P., Dunnill, P., & Lilly, M. (1971). Salting-out of enzymes with ammonium sulphate. *Biotechnology and Bioengineering, 13*(5), 713-718.

Frazier, A. E., & Thorburn, D. R. (2012). Biochemical analyses of the electron transport chain complexes by spectrophotometry. *Mitochondrial Disorders* (pp. 49-62). Humana Press.

Gally, J., & Edelman, G. (1962). The effect of temperature on the fluorescence of some aromatic amino acids and proteins. *Biochimica et Biophysica Acta, 60*(3), 499-509.

Gazzotti, P., Bock, H., & Fleischer, S. (1974). Role of lecithin in d-β-hydroxybutyrate dehydrogenase function. *Biochemical and Biophysical Research Communications, 58*(1), 309-315.

Geladi, P. (2003). Chemometrics in spectroscopy. Part 1. Classical chemometrics. *Spectrochimica Acta Part B: Atomic Spectroscopy, 58*(5), 767-782.

Gorinstein, S., Goshev, I., Moncheva, S., Zemser, M., Weisz, M., Caspi, A.. Martín-Belloso, O. (2000). Intrinsic tryptophan fluorescence of human serum proteins and related conformational changes. *Journal of Protein Chemistry, 19*(8), 637-642.

Hagedorn, A., Legge, R. L., & Budman, H. (2003). Evaluation of spectrofluorometry as a tool for estimation in fed-batch fermentations. *Biotechnology and Bioengineering, 83*(1), 104-111.

Harms, P., Kostov, Y., & Rao, G. (2002). Bioprocess monitoring. *Current Opinion in Biotechnology, 13*(2), 124-127.

Heiner, Z., Roland, T., Léonard, J., Haacke, S., & Groma, G. (2013). Ultrafast Absorption Kinetics of NADH in Folded and Unfolded Conformations. *EPJ Web of Conferences, 41,* 07003.

Henderson, I. R., Navarro-Garcia, F., & Nataro, J. P. (1998). The great escape: structure and function of the autotransporter proteins. *Trends in Microbiology, 6*(9), 370-378.

Jackson, G., Fahim, R., Tan, L., Chong, P., Vose, J., & Klein, M. (1995). *U.S Patent No. 5,444,159.* Washington, DC: U.S. Patent and Trademark Office.

Jiang, D., Johnston, C. T., & Hem, S. L. (2003). Using rate of acid neutralization to characterize aluminum phosphate adjuvant. *Pharmaceutical Development and Technology, 8*(4), 349-356.

Jones, L. S., Peek, L. J., Power, J., Markham, A., Yazzie, B., & Middaugh, C. R. (2005). Effects of adsorption to aluminum salt adjuvants on the structure and stability of model protein antigens. *The Journal of Biological Chemistry, 280*(14), 13406-13414.

Junker, M., Besingi, R. N., & Clark, P. L. (2009). Vectorial transport and folding of an autotransporter virulence protein during outer membrane secretion. *Molecular Microbiology, 71*(5), 1323-1332.

Junker, M., & Clark, P. L. (2010). Slow formation of aggregation-resistant β-sheet folding intermediates. *Proteins: Structure, Function, and Bioinformatics, 78*(4), 812-824.

Junker, M., Schuster, C. C., McDonnell, A. V., Sorg, K. A., Finn, M. C., Berger, B., & Clark, P. L. (2006). Pertactin beta-helix folding mechanism suggests common themes for the secretion and folding of autotransporter proteins. *Proceedings of the National Academy of Sciences of the United States of America, 103*(13), 4918-4923.

Kallonen, T., & He, Q. (2009). Bordetella pertussis strain variation and evolution postvaccination. *Expert Review of Vaccines, 8*(7), 863-875.

Lakowicz, J. R. (2007). Principles of fluorescence spectroscopy. Springer, NY.

Leo, J. C., Grin, I., & Linke, D. (2012). Type V secretion: mechanism(s) of autotransport through the bacterial outer membrane. *Philosophical Transactions of the Royal Society of London.Series B, Biological Sciences, 367*(1592), 1088-1101.

Li, B., Ryan, P. W., Shanahan, M., Leister, K. J., & Ryder, A. G. (2011). Fluorescence Excitation–Emission Matrix (EEM) Spectroscopy for Rapid Identification and Quality Evaluation of Cell Culture Media Components. *Applied Spectroscopy, 65*(11), 1240-1249.

Li, J., & Humphrey, A. E. (1992). Factors affecting culture fluorescence when monitoring bioreactors. *Journal of Fermentation and Bioengineering, 74*(2), 104-111.

Lindemann, C., Marose, S., Nielsen, H., & Scheper, T. (1998). 2-Dimensional fluorescence spectroscopy for on-line bioprocess monitoring. *Sensors and Actuators B: Chemical, 51*(1), 273-277.

Locht, C., (2007). *Bordetella: molecular microbiology*. Norfolk: Horizon Bioscience.

Locht, C. (2010). Molecular aspects of *Bordetella pertussis* pathogenesis. *International Microbiology, 2*(3), 137-144.

Mach, H., Middaugh, C. R., & Lewis, R. V. (1992). Statistical determination of the average values of the extinction coefficients of tryptophan and tyrosine in native proteins. *Analytical Biochemistry, 200*(1), 74-80.

Mahler, H., Friess, W., Grauschopf, U., & Kiese, S. (2009). Protein aggregation: pathways, induction factors and analysis. *Journal of Pharmaceutical Sciences, 98*(9), 2909-2934.

Marose, S., Lindemann, C., Ulber, R., & Scheper, T. (1999). Optical sensor systems for bioprocess monitoring. *Trends in Biotechnology, 17*(1), 30-34.

Mayer, L. M., Schick, L. L., & Loder III, T. C. (1999). Dissolved protein fluorescence in two Maine estuaries. *Marine Chemistry, 64*(3), 171-179.

Miller, J. L., Le Coq, J., Hodes, A., Barbalat, R., Miller, J. F., & Ghosh, P. (2008). Selective ligand recognition by a diversity-generating retroelement variable protein. *PLoS Biology, 6*(6), e131.

Mooi, F. R., van Oirschot, H., Heuvelman, K., van der Heide, H. G., Gaastra, W., & Willems, R. J. (1998). Polymorphism in the Bordetella pertussis virulence factors P.69/pertactin and pertussis toxin in The Netherlands: temporal trends and evidence for vaccine-driven evolution. *Infection and Immunity, 66*(2), 670-675.

Morefield, G. L., HogenEsch, H., Robinson, J. P., & Hem, S. L. (2004). Distribution of adsorbed antigen in mono-valent and combination vaccines. *Vaccine, 22*(15), 1973-1984.

Nguyen, M., Westerhoff, P., Baker, L., Hu, Q., Esparza-Soto, M., & Sommerfeld, M. (2005). Characteristics and reactivity of algae-produced dissolved organic carbon. *Journal of Environmental Engineering, 131*(11), 1574-1582.

Nishimura, Y., & Tsuboi, M. (1978). Raman spectra of flavins: avoidance of interference from fluorescence. *Chemical Physics Letters, 59*(2), 210-213.

Noble, J. E., & Bailey, M. J. (2009). Quantitation of protein. *Methods in Enzymology, 463*, 73-95.

Nozaki, Y. (1972). The preparation of guanidine hydrochloride. *Methods in Enzymology, 26*, 43-50.

Otsuka, N., Han, H., Toyoizumi-Ajisaka, H., Nakamura, Y., Arakawa, Y., Shibayama, K., & Kamachi, K. (2012). Prevalence and genetic characterization of pertactin-deficient *Bordetella pertussis* in Japan. *PLoS One, 7*(2), e31985.

Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., & Gray, T. (1995). How to measure and predict the molar absorption coefficient of a protein. *Protein Science, 4*(11), 2411-2423.

Paul, R., & Schneckenburger, H. (1996). Oxygen concentration and the oxidation-reduction state of yeast: Determination of free/bound NADH and flavins by time-resolved spectroscopy. *Naturwissenschaften, 83*(1), 32-35.

Peiris, B., Budman, H., Moresoli, C., & Legge, R. (2009). Acquiring reproducible fluorescence spectra of dissolved organic matter at very low concentrations. *Water Science & Technology, 60*(6)

Reichart, O., Szakmár, K., Jozwiak, A., Felföldi, J., & Baranyai, L. (2007). Redox potential measurement as a rapid method for microbiological testing and its validation for coliform determination. *International Journal of Food Microbiology, 114*(2), 143-148.

Renn, J. P., & Clark, P. L. (2008). A conserved stable core structure in the passenger domain β-helix of autotransporter virulence proteins. *Biopolymers, 89*(5), 420-427.

Renn, J. P., Junker, M., Besingi, R. N., Braselmann, E., & Clark, P. L. (2012). ATP-independent control of autotransporter virulence protein transport via the folding properties of the secreted protein. *Chemistry & Biology, 19*(2), 287-296.

Rhee, J. I., Kang, T., Lee, K., Sohn, O., Kim, S., & Chung, S. (2006). Application of principal component analysis and self-organizing map to the analysis of 2D fluorescence spectra and the monitoring of fermentation processes. *Biotechnology and Bioprocess Engineering, 11*(5), 432-441.

Rinnan, Å., Booksh, K. S., & Bro, R. (2005). First order Rayleigh scatter as a separate component in the decomposition of fluorescence landscapes. *Analytica Chimica Acta, 537*(1), 349-358.

Robinson, J. P., Picklesimer, J. B., & Puett, D. (1975). Tetanus toxin. The effect of chemical modifications on toxicity, immunogenicity, and conformation. *The Journal of Biological Chemistry, 250*(18), 7435-7442.

Ross, D., Kepa, J. K., Winski, S. L., Beall, H. D., Anwar, A., & Siegel, D. (2000). NAD (P) H: quinone oxidoreductase 1 (NQO1): chemoprotection, bioactivation, gene regulation and genetic polymorphisms. *Chemico-Biological Interactions, 129*(1), 77-97.

Rover Jr, L., Fernandes, J. C., Neto, G. d. O., Kubota, L. T., Katekawa, E., & Serrano, S. H. (1998). Study of NADH stability using ultraviolet–visible spectrophotometric analysis and factorial design. *Analytical Biochemistry, 260*(1), 50-55.

Royer, C. A. (2006). Probing protein folding and conformational transitions with fluorescence. *Chemical Reviews, 106*(5), 1769-1784.

Ryan, P. W., Li, B., Shanahan, M., Leister, K. J., & Ryder, A. G. (2010). Prediction of cell culture media performance using fluorescence spectroscopy. *Analytical Chemistry, 82*(4), 1311-1317.

Sasiak, A. B., Bolgiano, B., Crane, D. T., Hockley, D. J., Corbel, M. J., & Sesardic, D. (2000). Comparison of in vitro and in vivo methods to study stability of PLGA microencapsulated tetanus toxoid vaccines. *Vaccine, 19*(7), 694-705.

Shifrin, S. (1964). Charge transfer and excitation-energy transfer in a model for enzyme-coenzyme interactions. *Biochimica et Biophysica Acta (BBA)-Specialized Section on Enzymological Subjects, 81*(2), 205-213.

Shirley, B. A. (1995). Urea and guanidine hydrochloride denaturation curves. *Protein Stability and Folding* (pp. 177-190) Springer, N.Y.

Shlens, J. (2005). A tutorial on principal component analysis. *Systems Neurobiology Laboratory,* University of California, San Diego. http://www.cs.cmu.edu/~elaw/papers/pca.pdf.

Sigler, K., Chaloupka, J., Brozmanova, J., Stadler, N., & Höfer, M. (1999). Oxidative stress in microorganisms—I. *Folia Microbiologica, 44*(6), 587-624.

Skibsted, E., Lindemann, C., Roca, C., & Olsson, L. (2001). On-line bioprocess monitoring with a multi-wavelength fluorescence sensor using multivariate calibration. *Journal of Biotechnology, 88*(1), 47-57.

Tamer, I. M., & Chisti, Y. (2001). Production and recovery of recombinant protease inhibitor α_1-antitrypsin. *Enzyme and Microbial Technology, 29*(10), 611-620.

Tamer, I. M., Moo-Young, M., & Chisti, Y. (1998). Disruption of *Alcaligenes latus* for recovery of poly (β-hydroxybutyric acid): comparison of high-pressure homogenization, bead milling, and chemically induced lysis. *Industrial & Engineering Chemistry Research, 37*(5), 1807-1814.

Textbook of biochemistry : with clinical correlations (2006). In Devlin T. M. (Ed.),. Hoboken, N.J.: Hoboken, N.J.: Wiley-Liss.

van Hoek, M. J., & Merks, R. M. (2012). Redox balance is key to explaining full vs. partial switching to low-yield metabolism. *BMC Systems Biology, 6*(1), 22.

Viallet, P. M., & Vo-Dinh, T. (2003). Monitoring intracellular proteins using fluorescence techniques: from protein synthesis and localization to activity. *Current Protein and Peptide Science, 4*(5), 375-388.

Vishwanath, K., & Ramanujam, N. (2000). Fluorescence spectroscopy *in vivo*. *Encyclopedia of Analytical Chemistry,* R. Meyers, Ed. New York: Wiley, 2000, pp. 20–56.

Vivian, J. T., & Callis, P. R. (2001). Mechanisms of tryptophan fluorescence shifts in proteins. *Biophysical Journal, 80*(5), 2093-2109.

Vojinović, V., Cabral, J., & Fonseca, L. (2006). Real-time bioprocess monitoring: Part I: *In situ* sensors. *Sensors and Actuators B: Chemical, 114*(2), 1083-1091.

Wehry, E. L. (1997). Molecular fluorescence and phosphorescence spectrometry. *Handbook of Instrumental Techniques for Analytical Chemistry*; Settle, F. A., Ed.; Prentice Hall: Upper Saddle River, NJ, 1997; pp 507−539.

Weinberg, R. B. (1988). Exposure and electronic interaction of tyrosine and tryptophan residues in human apolipoprotein A-IV. *Biochemistry, 27*(5), 1515-1521.

White, A. (1959). Effect of pH on fluorescence of tryosine, tryptophan and related compounds. *The Biochemical Journal, 71*(2), 217-220.

Wise, B. M., Gallagher, N., Bro, R., Shaver, J., Windig, W., & Koch, R. S. (2007). PLS-Toolbox Version 3.5 for use with MATLAB™, Manual. Eigenvector Research, Inc. Manson, WA, USA.

Wise, B. M., & Gallagher, N. B. (1996). The process chemometrics approach to process monitoring and fault detection. *Journal of Process Control, 6*(6), 329-348.

Wold, S., Esbensen, K., Geladi, P., Wold, S., Esbensen, K., & Geladi, P.Principal component analysis. *Chemometrics and Intelligent Laboratory Systems, 2*(1), 37-52.

Wold, S., Geladi, P., Esbensen, K., & Öhman, J. (1987). Multi-way principal components-and PLS-analysis. *Journal of Chemometrics, 1*(1), 41-56.

Wos, M., & Pollard, P. (2006). Sensitive and meaningful measures of bacterial metabolic activity using NADH fluorescence. *Water Research, 40*(10), 2084-2092.

Zhao, Q., & Sitrin, R. (2001). Surface phosphophilicity of aluminum-containing adjuvants probed by their efficiency for catalyzing the P-O bond cleavage with chromogenic and fluorogenic substrates. *Analytical Biochemistry, 295*(1), 76-81.