MULTIVARIATE STATISTICAL OPTIMIZATION OF ENZYME IMMOBILIZATION ONTO SOLID MATRIX USING CENTRAL COMPOSITE DESIGN

A Thesis Submitted to
the Graduate School of Engineering and Sciences of
İzmir Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in Chemistry

by
Tuğba ARPAKCI

December 2013
İZMİR

We approve the thesis of **Tuğba ARPAKCI**

Examining Committee Members:

_____
Prof. Dr. Durmuş ÖZDEMİR
Department of Chemistry, İzmir Institute of Technology

_____
Prof. Dr. Şerife YALÇIN
Department of Chemistry, İzmir Institute of Technology

_____
Assoc. Prof. Dr. Figen TOKATLI
Deparment of Food Engineering, İzmir Institute of Technology

17 December 2013

_____
Prof. Dr. Durmuş ÖZDEMİR
Supervisor,
Department of Chemistry,
İzmir Instıtute of Technology

_____
Assoc. Prof. Dr. Gülşah ŞANLI
Co-Supervisor,
Department of Chemistry,
İzmir Instıtute of Technology

_____
Prof. Dr. Ahmet E. EROĞLU
Head of the Department of Chemistry

_____
Prof. Dr. R. Tuğrul SENGER
Dean of the Graduate School of
Engineering and Science

ACKNOWLEDGEMENTS

I would like to express my special appreciation to my research advisor, Prof. Dr. Durmuş Özdemir for his support, endless patience and advice through my master thesis.

I also would like to state my special thanks to my co-advisor Assoc. Prof. Dr. Gülşah ŞANLI for her guidance, her smiling face and criticism during my study.

Special thanks to my friend Yusuf SÜRMELİ for his all help, support and friendship through my master.

Next, I wish to thank all my lab mates especially Deniz ÇELİK, Esra KUDAY and İrem ANIL for their good friendship and techical supports.

Finally, I am especially grateful to my family , my mother Edibe ARPAKCI, my father Ahmet ARPAKCI and to my lovely sister, Tuğçe, for their continuous support, patience, endless love and understanding throughout my entire life.

ABSTRACT

MULTIVARIATE STATISTICAL OPTIMAZATION OF ENZYME IMMOBILIZATION ONTO SOLİD MATRIX USING CENTRAL COMPOSITE DESIGN

In recent years, scientist have been used alternative technology in order to increase enzyme stability and also reduce the cost of production of enzyme. Immobilization methods have attracted the attention of scientists due to its advantages in comparison with soluble enzyme or other methods. Immobilization process can be affected by many factors for this reason it is important to optimize the effective factors in order to enhance success of this process.

In preliminary studies, Bradford protein assay was used for determination of protein concentration. In order to increase sensitivity and accuracy of this assay, Bradford protein assay was combined with a multivariate calibration methods. Genetic Inverse Least Squares (GILS) and Partial Least Squares (PLS) were used for multivariate calibration. Calibration model was constructed for various concentration of Bovine Serum Albumin (BSA). Standard Error of Calibration (SEC) and Standard Error of Prediction (SEP) were calculated and results of multivariate calibration method were compared with univariate calibration methods and each other.

In this study, the bovine serum albumin immobilization studies were carried out. The bovine serum albumin was immobilized on chitosan nanoparticles and effective factors such as chitosan concentration, immobilization time, pH and temperature were optimized by using central composite design (CCD). Central composite design is used to investigate interaction between these parameters and to find the optimum values of effective factors.

ÖZET

MERKEZİ KOMPOZİT TASARIM KULLANILARAK ENZİMLERİN KATI FAZA SABİTLENMESİ KOŞULLARININ ÇOK DEĞİŞKENLİ İSTATİSTİKSEL OPTİMİZASYONU

Son yıllarda, bilim adamları enzim dayanıklılığını arttırmak ve üretim maliyetini azaltmak amacı ile birçok alternatif yöntem kullanmaktadırlar. İmmobilizasyon tekniği diğer yöntemlere ve çözünmüş enzime oranla sahip olduğu avantajlardan dolayı bilim adamlarının dikkatini çekmektedir. İmmobilizasyon prosesi birçok faktörden etkilenmektedir bu sebepten dolayı prosesin başarısını arttırmak için optimum koşulların bulunması önemlidir.

Protein konsantrasyon tayini Bradford yöntemi kullanılarak yapılmıştır. Bu yöntemin hassasiyetini ve doğruluğunu arttırmak için çok değişkenli kalibrasyon methodları ile birleştirilerek kullanılmıştır. Çok değişkenli kalibrasyon için Genetic Ters En Küçük Kareler (GILS) ve Kısmi En Küçük Kareler (PLS) yöntemi kullanılmıştır. Protein konsantrasyonu için kalibrasyon modeli oluşturulmuştur ve standard kalibrasyon hatası (SEC) ve standard tahmin hatası (SEP) hesaplanmıştır. Kullanılan çok değişkenli kalibrasyon yöntemi sonuçları birbirleri arasında ve tek değişkenli kalibrasyon yöntemi sonuçları karşılaştırılmıştır.

Bu çalışmada enzim immobilizasyonu sığır serum albumin (BSA) kullanılarak yapılmıştır. Sığır serum albumin kitosan nanoparçacıkları üzerine immobilize edilmiştir. Merkezi kompozit dizayn kullanılarak kitosan konsantrasyonu, sıcaklık, pH ve immobilizasyon sıcaklığı gibi faktörlerin optimum değerleri ve birbirleri ile etkileşimleri irdelenmiştir.

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

LIST OF TABLES

Table **Page**

CHAPTER 1

INTRODUCTION

1.1. Immobilization

Enzymes are biological catalyst that make a chemical reaction quickly and efficiently. They are three-dimensional natural protein molecules that are produced by all living organisms.

The use the enzymes increases because of their applications in a wide variety of processes such as fine chemistry, food chemistry, therapeutics applications, decontamination processes, protein engineering. Despite of a huge demand for enzymes, the use of enzymes has been limited their high cost of production and stabilization on storage. In order to improve the stability of enzymes, several methods, such as addition of additives, chemical modification, protein engineering, and enzyme immobilization, have been used (Costa et al., 2005). Among them, immobilized enzymes have been considerably used in a wide range of application due to their benefits in comparison with soluble enzymes or alternative technologies (Tischer and Kasche, 1999).

In general the term ‘immobilization’ refers to the act of the limiting movement or making incapable of movement. The term ‘immobilized enzymes’ refers to enzymes physically confined or localized in a certain defined region of space with retention of their catalytic activities, and which can be used repeatedly and continuously. Immobilization means associating the biocatalysts with an insoluble matrix or immobilized proteins and cells to an insoluble support. Practically, the procedure consists of mixing together the enzyme and the support material under appropriate conditions and following a period of incubation, separating the insoluble material from the soluble material by centrifugation or filtration.

Immobilized biocatalysts are not only enzymes as well cells or organelles (or combinations of these). For many industrial applications, enzymes and cells have to be immobilized,with very simple and cost-effective protocols, in order to be re-used for a long time (Meena and Raja, 2006).

1.1.1. Advantages of Immobilization

Usage of enzymes has some limitations because of their some characteristics that are not appropriate for industrial application:

- Enzymes are natively unstable
- Enzymes are easily inhibited
- They only work well on natural substrates and under physiological conditions (Bugg, 2001).

Enzyme immobilization technology has become an efficient way to increase enzymes functional properties with these advantages that are given below:

- Multiple and repetitive usage of catalyst are provided
- Greater control of the catalytic process
- Increased stability of enzyme
- Effluent problems are decreased
- Enzyme can easily be separated from the reaction
- Product is not infected by the enzyme
- Immobilized biocatalysts allows development of continuous process

1.1.2. Immobilization Methods

Enzyme immobilization process consists of three main components such as the enzyme, the matrix, and the mode of attachment or entrapment.

Ideal matrix must comprise characteristics like inertness, physical strength, stability, regenerability and ability to increase enzyme specificity/activity and decrease product inhibition, nonspecific adsorption and microbial contamination (Datta et al., 2013). Supports can be divided into two categories such as inorganic and organic according to their chemical composition. There are two types of organic supports such as natural and synthetic polymers (Kennedy and White, 1985).

Table 1.1. Examples of Carriers Used for Enzyme Immobilization
(Source: Kennedy and White, 1985; Guisan, 2006).

Organic
• Polysaccharides: Cellulose, agar, agarose, chitin, alginate dextrans. • Proteins: Collagen, albumin • Carbon • Polystyrene • Other polymers: Polyacrylate polymethacrylates, polyacrylamide, polyamides, vinyl, and allyl-polymers
Inorganic
Natural minerals: Bentonite, silica, sand. Processed materials: Glass (nonporous and controlled pore), metals, controlled pore Metal oxides (e.g. ZrO2, TiO2, Al2O3)

Chemical characteristics of ,enzymes , different properties of substrates and products and range of potential processes employed should be considered while selecting the immobilization methods. The most commonly used immobilization methods are shown in Figure 1.1.

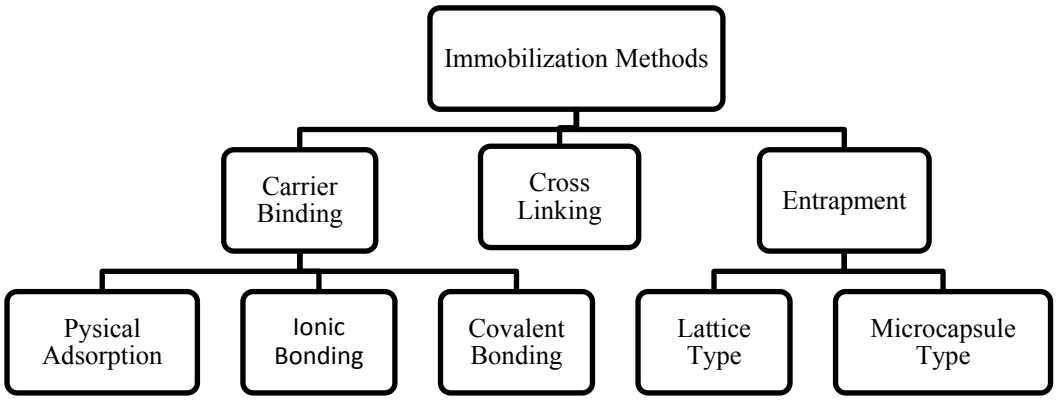


Figure 1.1. Various immobilization methods
(Source: Guisan, 2006).

These immobilization methods are categorised as chemical and physical methods. Chemical immobilization involve the formation of covalent bonds between the functional group on the enzyme and functional groups on the support material whereas

physical methods do not involve covalent bonding with the enzyme (Guisan, 2006). In this study, physical adsorption which is subheading of carrier binding is used.

1.1.2.1. Carrier Binding

The earliest immobilization technique for enzymes is the carrier binding method. Some important items have critical importance when the selection of carrier as well as the nature of enzyme. These items are given:

- Particle size
- Surface area
- Molar ratio of hydrophilic to hydrophobic groups
- Chemical composition (Dumitriu et al., 1988).

In general, higher activity of the immobilized enzymes can be enhanced by increas the ratio of hydrophilic groups and the concentration of bound enzymes. Polysaccharide derivatives such as cellulose, dextran, agarose, and polyacrylamide gel are mostly used as carriers for enzyme immobilization. According to the binding mode of the enzyme, the carrier-binding method can be further sub-classified into (Cao, 2006):

- Physical Adsorption
- Ionic Binding
- Covalent binding

1.1.2.1.1. Physical Adsorption

This method for the immobilization of an enzyme is based on the physical adsorption of enzyme protein on the surface of water-insoluble carriers. During physical adsorption, the hydrogen bonds, Van der Waals forces and hydrophobic interactions are the responsible forces for immobilization (Chen et al., 1996).

The method has some advantages such as less or no conformational change of the enzyme or destroying of its active center. This method is reversible, and this provides reuse of support material and enzymes again for different usages (Zaborsky, 1973). Another advantages of this method it is simple and cheap. Nevertheless, this method has some disadvantages that change in temperature, pH, ionic strength causes desorption of the

protein during its usage protein because of a weak binding force between the enzyme and the support material.

1.2. Natural Polymers

Natural polymers can be produced biologically and have unique functional properties which provides them to be used in different fields. Proteins such as collagen, gelatin, elastin, actin, etc.), polysaccharides (cellulose, starch, dextran, chitin, etc.) and polynucleotide (DNA and RNA) are the main natural polymers. They can be used as thickener, gel-maker, linker, distributing agent, lubricant, adhesive and biomaterial.

1.2.1. Chitin and Chitosan

Chitin is a natural polyaminosaccharides that can be synthesized and degraded in the biosphere in connection with the largest amounts of production per year (Krajewska, 2003). Chitin (Figure 1.2) is composed of 2- acetamido-2-deoxy-β-D-glucose through a β (1→4) linkage (Kumar, 2000). Chitin is the major component of the shells of crustaceans, the exoskeletons of insects and the cell walls of fungi. Chitin is a white, hard, inelastic, nitrogenous polysaccharide and the major source of surface pollution in coastal areas (Zikakis, 1984).

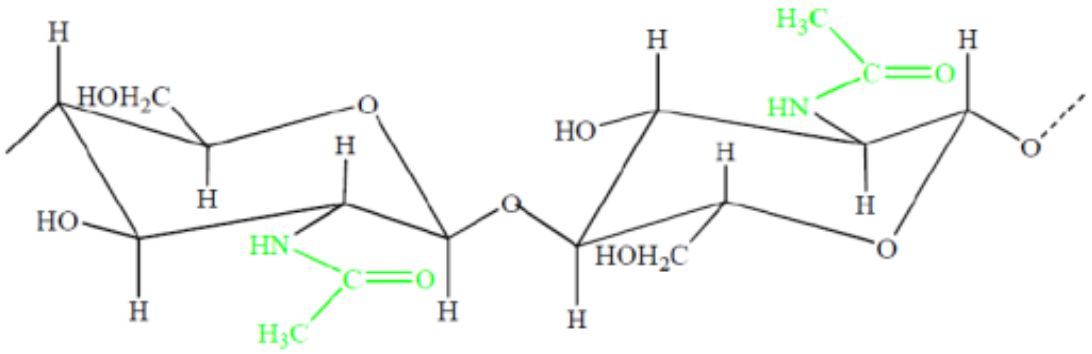


Figure 1.2. Structure of repeated units of chitin

Chitosan (Figure 1.3) is obtained by N-deacetylation of chitin which is a copolymer of glucosamine and N-acetyl glucosamine linked by β 1–4 glucosidic bonds and it has the properties of biodegradability and bio compatibility. It has a high nitrogen content (7%) which makes it as a useful chelating agent (Kurita, 2006, Tolaimate et al., 2000). Chitin and chitosan are attractive materials with unique properties of non-toxicity, film and fiber forming properties, adsorption of metal ions, coagulation of suspensions or solutes, and distinctive biological activities (Kurita, 2006).

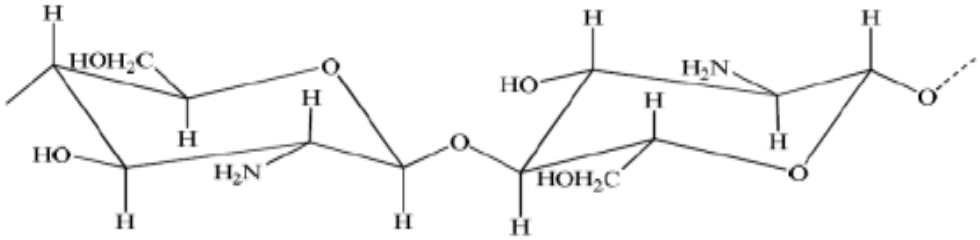


Figure 1.3. Structure of repeated units of chitosan

1.3. Determination of Protein Concentration

Accurate determination of the protein concentration is a necessary step before starting any type of protein analysis where the protein content affects the biological activity. Different methodologies can be used for the quantification of proteins, including spectroscopic methods , chemical methods and colorimetric methods (Silvério et al., 2012). When selecting a suitable method for protein concentration five issues should be concerned;

- sensitivity of the method
- clear definition of units
- interfering compounds
- removal of interfering substances before assaying samples
- correlation of information from various techniques

Colorimetric methods such as the Biuret method, the Lowry method, the bicinchoninic acid assay, the Bradford protein assay and the colloidal gold protein assay

are extensively used due to their relative simplicity and speed (Antharavally et al., 2009).

1.3.1. Bradford Protein Assay

Bradford protein assay which is also known as Coomassie Blue G dye binding assay is used to measure the concentration of proteins in a solution. Bradford protein assay is the most preferred method because of its some of advantages such as rapidity, simplicity and also the product is stable for approximately 1 h. Even though this assay has these advantages, there is a problem about linearity of this assay. Furthermore, Bradford protein assay induce precipitation of the reagent due to the its incompatibility with surfactants not only in high concentration but also in low concentrations (Lozzi et al., 2008).

Figure1.4 represents reaction schematic of Bradford Protein Assay. The principle of the Bradford Protein Assay is based on an absorbance maximum shift from 465 nm to 595 nm for Coomassie brilliant blue G-250 (CBBG) when binding to protein occurs (Lü et al., 2007). Addition of a protein sample causes formation of protein-dye complex which results in color change from green to blue.

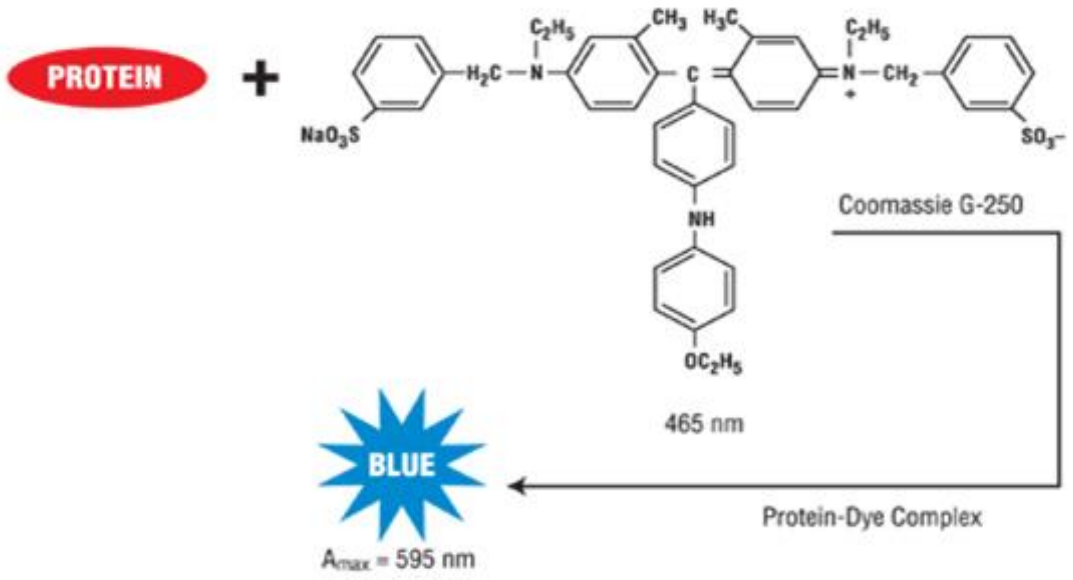


Figure 1.4. Reaction schematic Bradford Protein Assay

Starting with CBBG itself, the free dye has three protonation that are shown in Figure 1.5. They have absorption peaks at 465 nm (the cationic red dye form), 650 nm (the neutral greenish dye form), and 595 nm (the anionic blue dye form) (Wei et al., 1997).

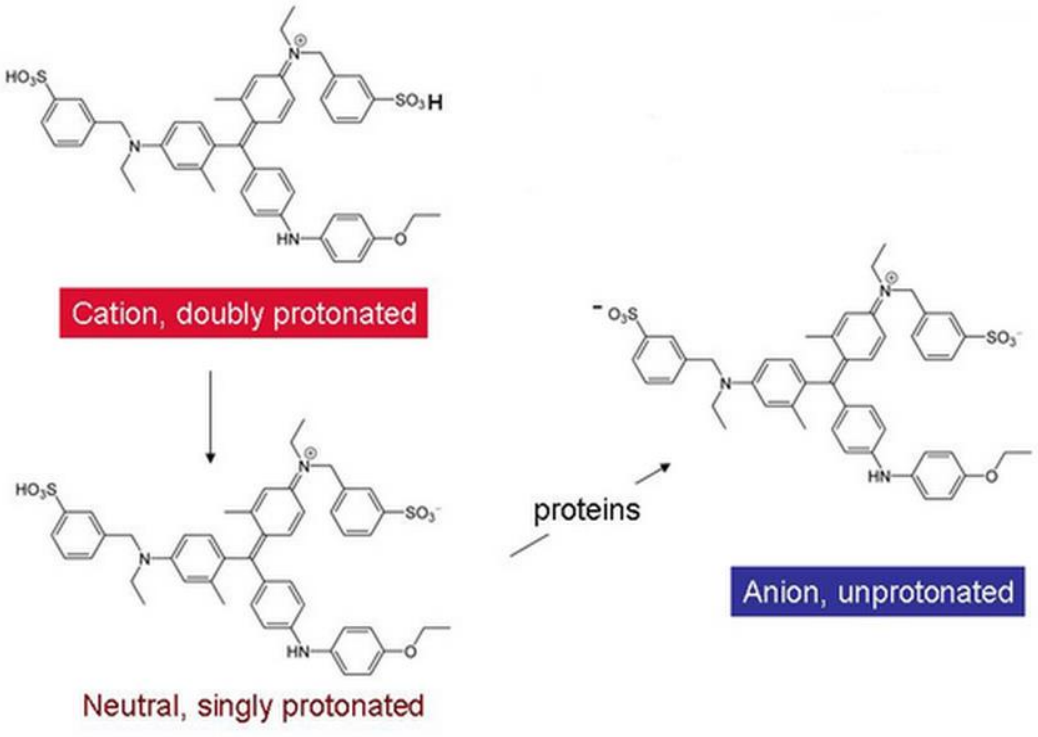


Figure 1.5. Three protonation forms of Coomassie brilliant blue G-250 (CBBG)

In general, the blue dye form (595nm) has an electrostatic attraction of the dye's sulfonic groups to arginine and lysine side chains on protein with its negative charge to constitute the complex. Besides the electrostatic interactions, nonelectrostatic interactions, such as Van der Waals forces and hydrophobic interactions, between CBBG and proteins can exist to form a complex. Therefore, it is supposed that all the three dye species can bind to protein to constitute dye-protein complexes in dye-binding scheme. All this deceptive impacts can be corrected only if the full absorption spectra are recorded (Wei et al., 1997). In addition to this, at a low protein concentration the accuracy of this assay can be affected by ignorance the absorbance at 465 nm from the CBBG dye rest which cannot to form dye-protein complexes (Lü et al., 2007). Thus, measuring the change in CBBG absorption at a single wavelength (595 nm) can be deceptive but it can be corrected only if the full absorption spectra are recorded.

CHAPTER 2

ULTRAVIOLET-VISIBLE SPECTROSCOPY

2.1. Spectroscopy

Electromagnetic radiation is a type of energy that has different forms, the electromagnetic spectrum is the distribution of electromagnetic radiation according to frequency or wavelength.

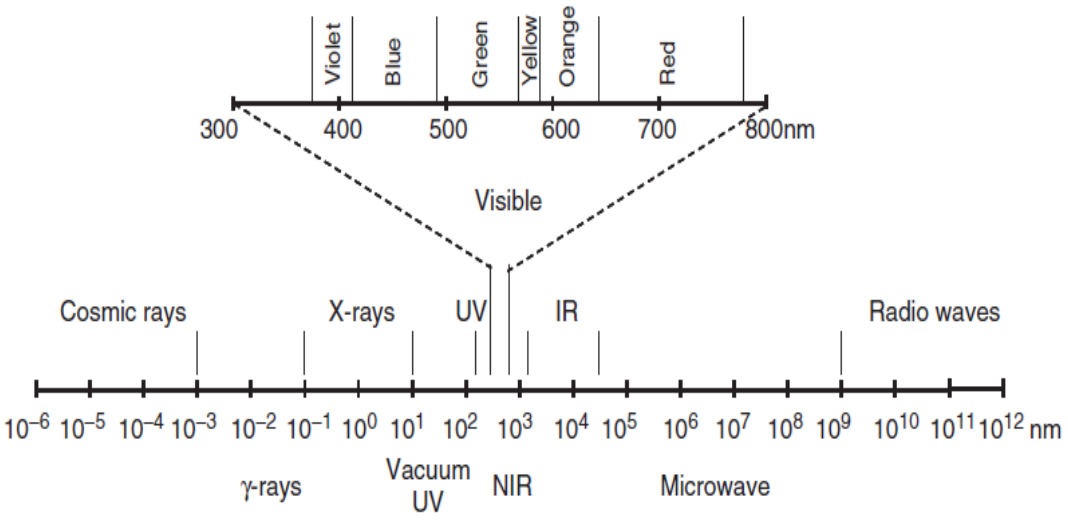


Figure 2.1. Schematic description of the electromagnetic spectrum.
(Source: Burgess, 2007)

The term *spectroscopy* refers historically to processes in which light or visible radiation is dispersed its component wavelengths for producing a spectrum (Wiberg, 2004). Spectroscopic techniques are really important in analytical chemistry due to a large number of application fields.

Adsorption is a process in which electromagnetic energy is transferred to the atoms or molecules of the sample (Wiberg, 2004). When the molecule interacts with photons of electromagnetic radiation which causes absorption of electromagnetic energy so atoms and molecules are excited to one or more higher energy level. However, the energy of an exciting photon must exactly match the energy difference between the

ground state and one of the excited states. The total potential energy of a molecule is given by:

$$E = Eelectronic + Evibrational + Erotational$$

where $E_{electronic}$ defines the electronic energy of the molecule, $E_{vibrational}$ the vibrational energy and $E_{rotational}$ the rotation energy. Figure 2.2 represents energy levels of molecules.

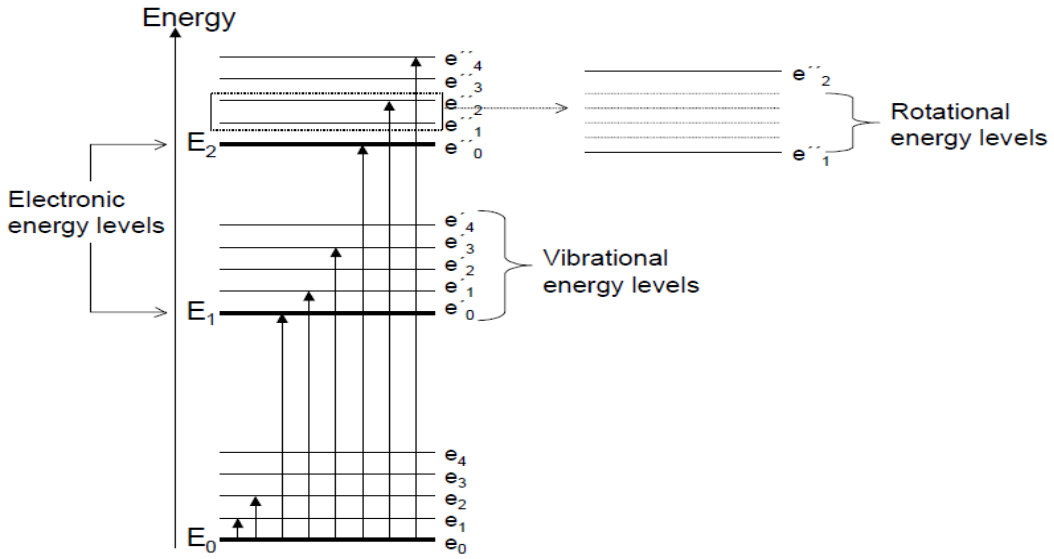


Figure 2.2. An energy level diagram for a molecule, showing electronic, vibrational and rotational energy levels (Source: Wiberg, 2004)

As shown in Figure 2.2 the number of level is different for each energy level so they have diffrent enery from each other, where:

$$E_{electronic} > E_{vibrational} > E_{rotational}$$

As shown in figure 2.2, adsorption is occurred when the light passes through a solution of a compound. P_0 is defined the incident radiant power whereas P is defined the transmitted radiant power (Skoog et al., 1998). The thickness of the solution is b cm and the concentration c (g l^{-1}).

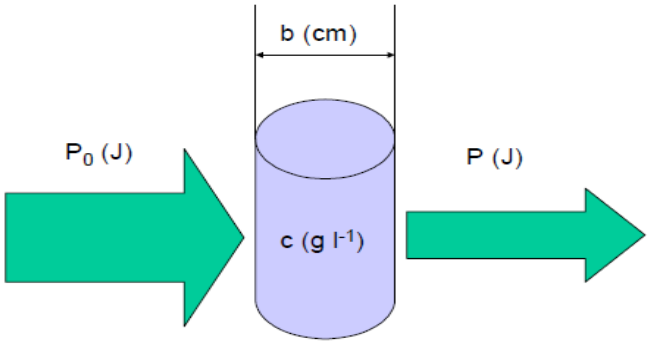


Figure 2.3. Illustration of the attenuation of a beam of radiation by an absorbing solution (Source**:** Wiberg, 2004)

Transmittance (T) is the fraction of incident radiation which is transmitted by solution and given by:

$$T = P/P_0 \tag{2.1}$$

The logarithm of the transmittance is called the *absorbance, A*:

$$A = -log10\ T \tag{2.2}$$

The amount of light absorbed is expressed as either transmittance or absorbance. Transmittance is usually stated as a percentage:

$$\%T = 100(P/P0) \tag{2.3}$$

while absorption is given by *absorbance units (AU).* Absorbance is directly is the function of both concentration and path length so the case, relationship between absorbance value and both concentration and path length is linear and this relationship is called the Lambert-Beer law:

$$A = \varepsilon * b * c \tag{2.4}$$

where ε is the *molar absorptivity*. There is a limitation of this law that this relationship linear only if solutions have an absorbance of about <1.5 AU. Thus, with

too high a concentration there is no longer a linear relationship between the absorption and the concentration.

2.1.1. Ultraviolet-Visible Absorption Spectroscopy

Electromagnetic radiation between 190 nm and 800 nm, only small part of electromagnetic radiation spectrum shown in Figure 2.1, is used in this absorption spectroscopy. Electromagnetic radiation between 190 nm and 800 nm is divided into the ultraviolet (UV, 190-400 nm) and visible (VIS, 400-800 nm) regions. UV-Vis absorption spectroscopy can simplistically be defined as the spectroscopy consisting of the electronic energy levels of a molecule, as shown in Figure 2.2. Therefore, it is sometimes called *electronic spectroscopy* because the absorption of radiation causes to transitions among the electronic energy levels of the molecule. The absorption of UV or visible radiation by a molecular compound M can be reckoned with a two-step process. First an electronic excitation occurs:

$$M + h\nu \rightarrow M^*$$

where *h*ν represents the *photon* and *M** is the electronically excited molecule. The most common type of relaxation involves conversion of excitation energy to heat:

$$M^* \rightarrow M + heat$$

The amount of thermal energy which is not usually detectable is not high enough to disturb the system under study. Other types of relaxation consist of *fluorescent* or *phosphorescent* re-emission. *Bonding* and *non-bonding* are two main types of electrons which are contained in UV-Vis absorption spectroscopy of organic molecules. Bonding electrons take part in the formation of bonds among atoms, whereas non-bonding electrons are unshared outer electrons not contained in any chemical bond like electrons in oxygen, nitrogen, sulphur and the halogens. Functional groups can be identified (*chromophores*) as a result of excitation of bonding electrons when these electrons absorb the UV-Vis radiation. Furthermore, aromatic hydrocarbons in UV-Vis absorption spectroscopy generally demonstrate strong absorption. Thus, multiple bonds

or aromatic conjunctions within molecules are easily determined by using UV-visible absorption spectroscopy due to its sensitivity. In an organic molecule involving chromophores the electronic spectrum is often complicated in comparison with that of a single atom. This can be explained by vibrational and rotational energy levels are superimposed on the electronic energy levels. This result in a broad band of absorption is obtained.

UV-Vis absorption spectrum of a sample can be affected some factors such as the solvent, pH and temperature. Some factors should have taken into account while choosing the solvent. These factors are:

- solvent should be transparent
- solvent not have an absorbance maxima intervening with the analyte.

Typical solvents used are water, ethanol or cyclohexane. UV-Vis spectrum can totally be changed by the changing in pH. However, this impact can often be controlled by the use of a buffer. Also, temperature can be an effective on UV-Vis spectrum however a thermostatted cell holder can be used to control the temperature.

2.1.1.1. Instrumentation of Ultraviolet-Visible Spectroscopy

An UV-Vıs instrument contains a source of UV-Vis radiation, a sample container which should be UV-Vis transparent, a wavelength selecting device, a detector and a signal processor, consecutively. Two sources are mostly used in UV-Vis spectrophotometers. Tungsten-Halogen lamp is used for the visible region of the spectrum (350 –800 nm) and Deuterium lamp is used for the ultraviolet region of the spectrum (160 –350 nm). Most spectrophotometers used to measure the UV-Vis range contain both types of lamps. Xenon lamp is an alternate light source which can be used both UV and visible regions. Wavelength selector is used to select a particular wavelength of light from a continuous source. Prisms disperse visible light into different wavelengths and colours. Dispersion visible light with a prism has some disadvantages since dispersion is angularly nonlinear and temperature-sensitive. Interference or absorption filters are the simplest form of wavelength selection but they have wide spectral bandwidths which causes deviations from Beer's law. Grating is an another wavelength selector and has provide better dispersion over prisms and filters. A diffraction grating consists of a series of parallel grooves (lines) on a reflecting surface.

Light falling on the grating is reflected at different angles, depending on the wavelength. Monochromators are used for spectral scanning which consists of an entrance slit, a dispersion device, and an exit slit. The output from a monochromator is band instead of monochromatic light. Quartz, glass and plastic (disposable) cuvettes can be used as a sample holder relying on the type of application since they are transparent certain sub-regions of Uv-Vis region. The two most common types of detectors used in UV-Vis spectroscopy are a photomultiplier tube or a photodiode detector. They are used to convert the light signal into an electrical signal.

There are various configurations of UV-Vis absorption spectrophotometers such as single-beam, double-beam in space, double-beam in time, multichannel (Skoog et al., 1998). Figure 2.4, Figure 2.5, Figure 2.6, Figure 2.7 represents instrumental designs for UV-visiblephotometers or spectrophotometers.

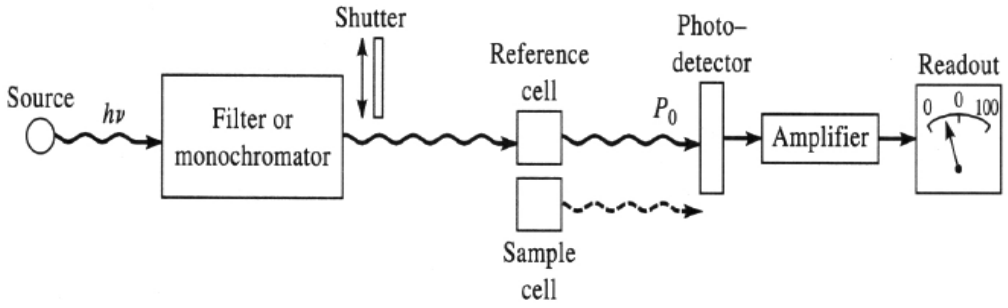


Figure 2.4. Schematic representation of single beam instrument
(Source: Skoog et al., 1998)

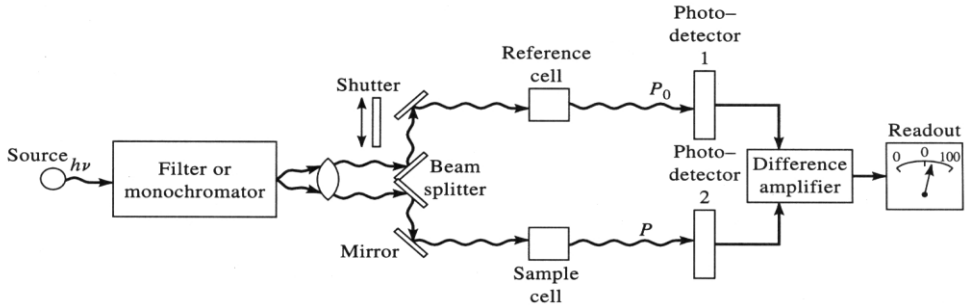


Figure 2.5. Schematic representation of double-beam in space instrument
(Source: Skoog et al., 1998)

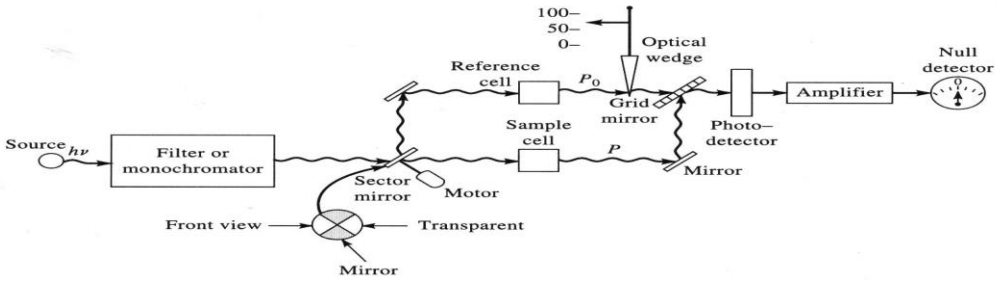


Figure 2.6. Schematic representation of Double-beam in time instrument
(Source: Skoog et al., 1998)

,

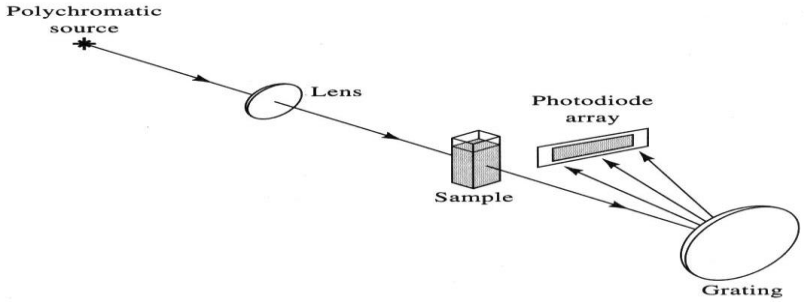


Figure 2.7. Schematic representation of Multichannel instruments
(Source: Skoog et al., 1998)

Dual-beam instruments have some advantages over a single-beam instrument such as they compensate for changes in lamp intensity between measurements of sample and blank. Howe ever, they need to add optical components that cause reducing sensitivity and sample throughput. Multichannel instruments have been widely used due to speed at which spectra can be acquired likewise their applicability to simultaneous multicomponent determinations.

CHAPTER 3

MULTIVARIATE ANALYSIS METHODS

Chemometrics which is chemical disipline provides higher chemical information and to relate quality parameters or physical properties to analytical instrument data by using mathematical and statistical methods according to The International Chemometrics Society (ICS) (Wise et al., 2002). Optimization of experimental parameters, design of experiments, calibration, signal processing are used for collecting good data and statistics, pattern recognition, principal component analysis are employed for getting information from these data in chemometrics. In this chapter, calibration and experimental design which are used in this study is centred on.

3.1. Calibration Method

3.1.1. Overview

Determination of relation between instrumental response and features of samples is obtained by constructing a model which is called calibration. Prediction is identified as a process in which the calibration model is used to predict the features ,in terms of instrument response, of a sample. Instrument responses and concentration levels of analyte are used to build the model. Then, concentration of an unknown sample can be predicted by using this model (Beebe et al., 1998).

In general, calibration methods are subdivided into two subsets such as univariate and multivariate calibration methods. Univariate calibration method used one single wavelength to detect the concentration of a single compound. On the other hand, multivariate calibration method used all or several of the wavelengths to determine the concentration of a multi-component mixture.

3.1.2. Univariate Calibration

Univariate calibration is based on using of single measurement from an instrument that is related to the analyte of interest to construct a model. In this method, Lambert Beer´s law is used to define the relationship between the concentration of an analyte and the instrumental response. When the relationship between instrument response and analyte concentration is taken into consideration as a linear, there are two options :

- Classical calibration
- Inverse calibration

3.1.2.1. Classical Univariate Calibration

This type of calibration models, concentration is modelled with the absorbance corresponding to one wavelength or data point in a spectrum. The general formula of classical calibration is:

$$\mathbf{a} \approx \mathbf{c} \cdot s \tag{3.1}$$

where **a** is the vector of absorbance at one wavelength for a number of samples and **c** is the vector of corresponding concentrations. The scalar coefficient *s* can be determined according to the following formula:

$$s \approx (\mathbf{c}' \cdot \mathbf{c})^{-1} \cdot \mathbf{c}' \cdot \mathbf{a} \tag{3.2}$$

where the $\mathbf{c}'$ is the transpose of the concentration vector.

Once the equation is solved for the *s*, the prediction model for an unknown can be performed easily by using *s*:

$$\hat{\boldsymbol{c}} \approx \hat{\boldsymbol{a}} / \boldsymbol{s} \tag{3.3}$$

where scalars *a* and *c* with hat refers to predictions.

The difference between the observed and predicted concentration values are residuals or errors. Prediction model's quality can be controlled by using residuals or errors.

$$\boldsymbol{e} = \boldsymbol{c} - \hat{\boldsymbol{c}} \tag{3.4}$$

If the residuals are less, better model is constructed (Brereton, 2000).

3.1.2.2. Inverse Univariate Calibration

Classical calibration is mostly preferred in analytical chemistry however it is not always the most proper approach due to two reasons. First, the prediction of concentration is obtained by using instrumental response in the classical univariate calibration. Therefore it is impossible to do inverse of this approach. The second reason relates to the error distributions. Besides, the response errors are originated from instrumental performance, however, determination of concentration values are mostly obtained gravimetrically, which leads to increase of the ratio of reliability of instruments. Thus, errors are mostly arised from concentration which is larger than instrumental error. Figure 3.1 represents the difference between errors derive from instrument and concentration.

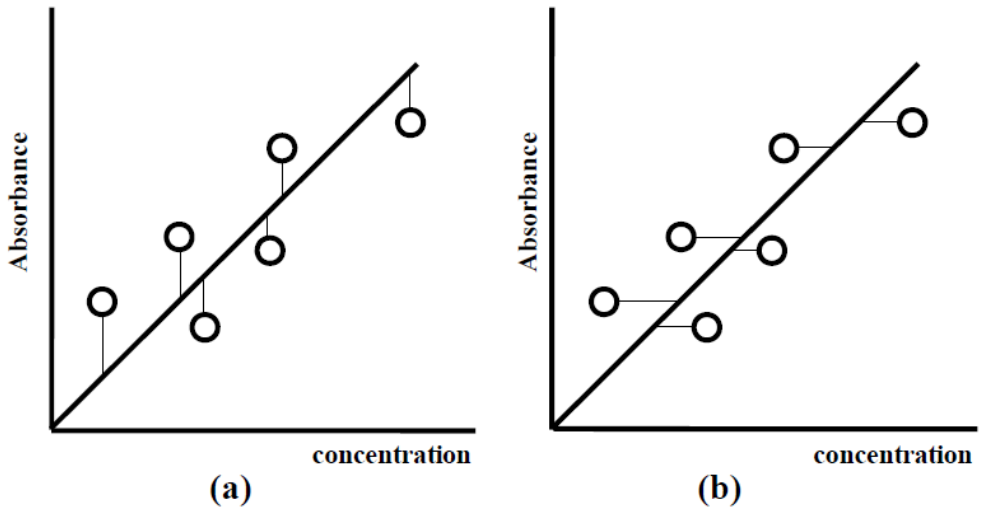


Figure 3.1. Error distributions in (a) classical and (b) inverse calibration models

Inverse calibration can be modelled as:

$$\mathbf{c} \approx \mathbf{a} \cdot b \tag{3.5}$$

where b is a scalar coefficient and is approximately inverse of s because each model makes assumptions on errors in a different way. b can be determined according to the following formula:

$$b \approx (\mathbf{a}' \cdot \mathbf{a})^{-1} \cdot \mathbf{a}' \cdot \mathbf{c} \tag{3.6}$$

and prediction of an unknown sample is constructed as:

$$\hat{c} \approx \hat{a} \cdot b \tag{3.7}$$

3.1.3. Multivariate Calibration

Multivariate calibration is an useful tool detecting all components of mixtures and for several instrument type therefore it can provide development of new analytical instrument. Besides this , analytical capacity and reliability of traditional instrument can be increased by multivariate calibration.(Martens and Naes, 2004)

Multivariate calibration has some advantages over univariate calibration.

1) Multivariate calibration can reduce time consuming since it can allow simultaneous analysis of multiple components in a sample (Beebe et al., 1998).
2) Repeating a measurement and calculating the mean is used to obtain precision in the prediction. These are outcome of minimization in the standard deviation of the mean which is referred as signal averaging (Beebe et al., 1998).
3) Multivariate calibration can cope with unknown interferences since it has fault-detection capabilities. This is not possible with univariate calibration so prediction of concentration of analyte may obtain wrong due to the presence of interferences. Solution of this problem is that physical separation of analyte from interfering material or using selective measurements however they causes to need more effort. Figure 3.2

demonstrates how the calibration curve is affected by the interferences. In multivariate calibration, choosing more variable results in eliminating nonlinearities due to the interferences. In addition, it provides to have higher chance to obtain better calibration curve (Öztürk, 2003).

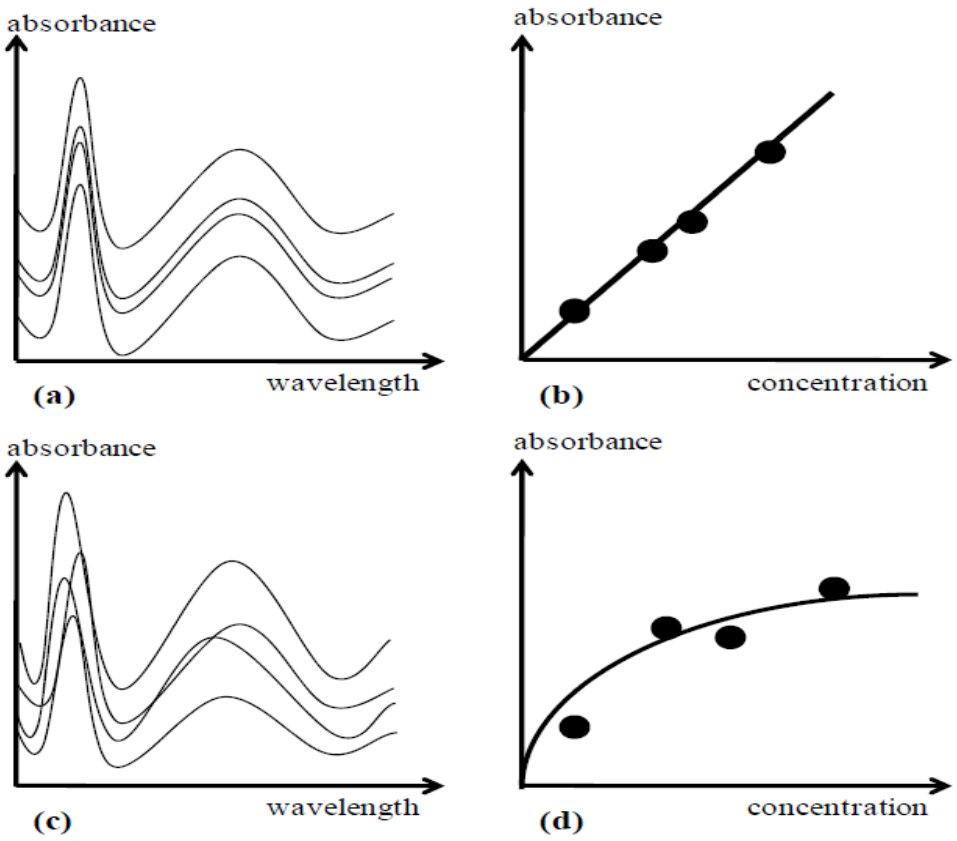


Figure 3.2. (a) Spectra of a sample in different concentrations which has no interference and its calibration curve (b) by univariate calibration; (c) spectra of a sample in different concentrations which has interfering materials and its calibration curve (d) by univariate calibration

In multivariate calibration, the equations can be improved in two ways such as classical calibration case which absorbance is directly proportional of concentration and inverse calibration which concentration is directly proportional absorbance. In addition to this, the absorbance of full spectral data is used by multivariate calibration. Therefore, more than one component can be used where concentration vector becomes a matrix. The multivariate calibration methods are the classical least squares (CLS), inverse least squares (ILS), principle component analysis (PCA), principle component regression and partial least squares (PCR and PLS), genetic regression (GR), genetic classical least squares (GCLS), genetic inverse least squares (GILS), and genetic partial

least squares (GPLS). In order to construct the best calibration model, the selection of the most appropriate calibration method is very significant (Massart et al., 1988, Brereton, 2003).

3.1.3.1. Classical Least Squares (CLS)

Classical least squares (CLS) method is based on the classical Beer's law in which the absorbance at each wavelength is modelled as a function of concentrations of an analyte. This method is modelled by the following equation:

$$\mathbf{A} = \mathbf{C} \times \mathbf{K} + \mathbf{E} \tag{3.8}$$

where A is an rn x n matrix structured of the absorbance spectra of rn calibration samples at n wavelengths, C is the rn x l concentration matrix corresponding to the concentrations of each of the l components in the rn calibration samples. E is the rn x n matrix of random errors for each calibration samples spectrum at each wavelengths. K is the l x n matrix of absorptivity-pathlength constants which represents the matrix of pure component spectra at unit concentration and unit pathlength. The method of least squares is used for calculating K matrix and given by:

$$\mathbf{K} = (\mathbf{C}' \cdot \mathbf{C})^{-1} \cdot \mathbf{C}' \cdot \mathbf{A} \tag{3.9}$$

Once the equation is solved for the **K** matrix, it can be used to predict concentrations of unknown samples from its spectrum by:

$$\hat{\mathbf{c}} = (\mathbf{K} \cdot \mathbf{K}')^{-1} \cdot \mathbf{K}' \cdot \hat{\mathbf{a}} \tag{3.10}$$

where $\hat{\mathbf{a}}$ is the spectrum of the unknown sample and $\hat{\mathbf{c}}$ is the vector of the predicted component concentrations**.**

CLS method is able to use whole spectrum to build the calibration model where as univariate methods and some other multiple linear regression methods are not. Furthermore, this method is mostly prefered since it supplies simultaneous fitting of spectral baselines and estimating pure component spectra along with the residuals.

Despite of these advantages, this technique has one major drawback. All interfering chemical components in a given spectral range and included in the calibration step is needed to known. In real life samples , ıt is not possible to know concentrations of all species, so the instrument response due to this interfering species cannot be put in the calibration model which causes a large error. This requirement can be reduced by using Inverse least squares (ILS) method.

3.1.3.2. Inverse Least Squares (ILS)

It is hard to know concentrations of all species in practice that makes CLS inapplicable. For obviating the drawback of Classical Least Squares, the Inverse Least Squares (ILS) model ,as the name suggests, is described by the inverse of Beer's Law . In this case, concentrations of an analyte are modelled as a function of absorbance. The ILS model for m calibration samples with n wavelengths for each spectrum is written in matrix form :

$$\mathbf{C} = \mathbf{AP} + \mathbf{E}_{\mathbf{C}} \tag{3.11}$$

where C and A are the same as in CLS, P is the *n x l* matrix of the unknown calibration coefficients relating *l* component concentrations to the spectral intensities and E_C is the *m x l* matrix of errors in the concentrations not fit by the model. The ILS model can be reduced for the analysis of one component at a time. The reduced model is given as:

$$\mathbf{c} = \mathbf{Ap} + \mathbf{e_c} \tag{3.12}$$

where c is the m x 1 vector of concentrations for the analyte that is being analyzed, p is n x 1 vector of calibration coefficients only for that particular analyte that are being modelled and e_c is the m x 1 vector of concentration residuals not fit by the model. During the calibration step, the least-squares estimated of p vector symbolized $\hat{p}$ as can be calculated as:

$$\hat{p} = (\mathbf{A} \cdot \mathbf{A}')^{-1}\ \mathbf{A}' \cdot C \tag{3.13}$$

Once $\hat{p}$ is calculated then the concentration of the analyte of interest can be predicted with the equation below:

$$\hat{\boldsymbol{c}} = \mathbf{a}' \cdot \hat{p} \tag{3.14}$$

where $\hat{\boldsymbol{c}}$ is the scalar estimated concentration and a is the spectrum of the unknown sample. ILS is one of the most preferable calibration method since it is able to predict one component at a time without requirement of knowing the concentrations of interfering species (Özdemir, 2006).

Even ILS has this advantages, there is a problem about dimensionality of the matrix. The problem is that in equation 3.13 where A matrix that have to inverse has much larger dimensions ,in terms of data points, compared to the number of samples in the calibration concentration vector in c. For that reason, generally, all fitted model results due to colinearity improved in the absorbance spectra of information. In addition to this adding more wavelengths to the model causes overfitting. Due to this effect calibration model would not produce reasonable predictions.

3.1.3.3. Partial Least Squares (PLS)

PLS method is used variation spectra, illustrates the changes in the absorbances at all the wavelengths in the spectra, instead of raw data to construct a calibration model Spectrum of sample could be rebuild by using variation spectra with multiplying each one with a different constant scaling factor and put in the results together. This process is end up when unknown spectrum gets similar the new spectrum.

PLS has major advantage over other multivariate calibration methods. It can be modelled one component at a time without requirement knowing all components in a given sample with avoiding wavelength selection problem. PLS does not consider only spectral errors but also errors from concentration estimates are taken into account in this model. Better calibration models and prediction can be enhanced due to these properties.

The model equation for PLS is described as:

$$\boldsymbol{A} = \boldsymbol{TB} + \boldsymbol{E_A} \tag{3.15}$$

where *A* is *mxn* matrix of spectral absorbance, B is a hxn matrix of loading spectra. T is an mxh matrix of scores defined by the h loading vectors. E_A is now the mxn matrix errors not fit by the model. As mentioned before, the loading vectors in B are not original component spectra but they are linear combinations of the original calibration spectra. The number of basis vectors, h, to illustrate an original calibration spectrum which is obtained by an algorithm throughout the calibration step.

Concentration of the analyte which is related to the ILS model given by:

$$\mathbf{c} = \boldsymbol{T}\boldsymbol{v} + \mathbf{e_c} \tag{3.16}$$

where c is the mxl vector of component concentrations, *v* is the hxl vector of coefficients which relate spectral intensities to the component concentration and e_c is the mxl vector of errors in reference values of the component that is being modelled.

The least-squares estimated of *v* vector that has similar solution to the equation (3.13) in ILS can be calculated as:

$$\hat{\boldsymbol{v}}_{\boldsymbol{h}} = (\boldsymbol{T}^{\boldsymbol{T}}\,\boldsymbol{T})^{-1}\,\boldsymbol{T}^{\boldsymbol{T}}\,\boldsymbol{c} \tag{3.17}$$

where $\hat{\boldsymbol{v}}_{\boldsymbol{h}}$ is the least-squares estimate of *v*. The T and B matrices are calculated in a stepwise manner (one vector at a time) till the desired model has been obtained.

There are two types of PLS methods that are present to analyze complex chemical mixtures. These are called PLS1 and PLS2 methods. In the PLS1 method, only one component is used in the model building step. This is widely used form the PLS method and it is assumed that the PLS1 predictions are better that those determined PLS2. It is proposed that PLS2 algorithm is more likely suitable for using qualitative application.

PLS1 algorithm starts with the calculation of the estimated first weighed loading vector, $\hat{\boldsymbol{w}}_{\boldsymbol{h}}$, by setting h to 1. The method of least squares is used for calculating estimated first weighed loading vector, $\hat{\boldsymbol{w}}_{\boldsymbol{h}}$ and is given by:

$$\hat{\boldsymbol{w}}_{\boldsymbol{h}} = \mathrm{A}^{\mathrm{T}}\ \mathrm{c}\ \ (\mathrm{c}^{\mathrm{T}}\ \mathrm{c})^{-1} \tag{3.18}$$

where $\widehat{\boldsymbol{w}}_{\boldsymbol{h}}$ is an *nxl* vector representing the first order approximation of the pure component spectra for the component that is being analyzed. After calculating weighted loading vector, ıt is used to obtain the score vector $\hat{t}_h$, with an ILS prediction model. The first estimated $\hat{t}_h$ vector is estimated by:

$$\hat{\boldsymbol{t}}_{\boldsymbol{h}} = \mathbf{A}\, \widehat{\boldsymbol{w}}_{\boldsymbol{h}} \tag{3.19}$$

Component concentrations are related this score vector by a linear least-squares regression. The scalar regression coefficient, $\widehat{\boldsymbol{v}}_{\boldsymbol{h}}$, as given:

$$\widehat{\mathbf{V}}_{\boldsymbol{h}} = \hat{\boldsymbol{t}}_{\boldsymbol{h}}^{\boldsymbol{T}}\, \boldsymbol{c}\, (\hat{\boldsymbol{t}}_{\boldsymbol{h}}^{\boldsymbol{T}}\, \hat{\boldsymbol{t}}_{\boldsymbol{h}})^{-1} \tag{3.20}$$

Afterwards, concentration residuals is obtained by using the this least-square estimated regression coefficient. The PLS loading vector $\hat{b}_h$ is calculated by a new model for A to reduce collinearty problems. In order to obtain estimated b vector, the method of least squares is used with the equation below:

$$\widehat{\boldsymbol{b}}_{\boldsymbol{h}} = \hat{\boldsymbol{t}}_{\boldsymbol{h}}^{\boldsymbol{T}}\, \mathbf{A}\, (\hat{\boldsymbol{t}}_{\boldsymbol{h}}^{\boldsymbol{T}}\, \hat{\boldsymbol{t}}_{\boldsymbol{h}})^{-1} \tag{3.21}$$

where $\widehat{b}_h$ is an nx1 vector. It is now possible to calculate the first PLS approximation to the calibration spectra by multiplying the score vector ($\hat{t}_h$) with transpose of PLS loading vector ($\hat{b}_h^T$).

Final calibration coefficients, b_f, that have the dimension of an original spectrum is obtained in the prediction step of PLS1. Once the bf is calculated, it is possible to calculate the concentration of a new sample using the average concentration of the analyte and its spectra. The prediction step in PLS1 is defined by the following formula:

$$\mathbf{b_f} = \widehat{\boldsymbol{W}}\, (\widehat{\boldsymbol{B}}\, \widehat{\boldsymbol{W}}^{\boldsymbol{T}})^{-1}\, \hat{\mathbf{v}} \tag{3.22}$$

where $\widehat{W}$ and $\hat{B}$ contains individual $\widehat{w}_h$ and $\hat{b}_h$ vectors, respectively and $v\hat{}$ is formed from individual regression coefficients ($\hat{v}_h$) The final prediction equation is then given as:

$$\hat{c} = \mathbf{a}^T \mathbf{bf} + \mathbf{c_0} \quad (3.23)$$

where $\hat{c}$ is the predicted unknown sample c, a is the spectrum of that sample and c_0 is the average concentration of calibration samples.

The optimal number of PLS factors can be determined in a different way which based on an algorithm. One of the methods for this is the cross-validation. In this method, validating the model is done by using left out spectrum. For this reason, PLS algorithm is performed on m-1 spectra for m calibration spectra. This process is finished when each spectrum is left out once in the calibration set. After that, the predicted concentration for each left out sample is checked with their original values and the prediction error sum of the squares (PRESS) is calculated for each added factor. The PRESS is a measure of how well a particular model fits the calibration data and given by:

$$\text{PRESS} = \sum_{i=1}^{m} (\hat{c} - c_i)^2 \quad (3.24)$$

where c_i is the reference (known) concentration of the ith sample and concentration is the predicted concentration of the i^{th} sample for m calibration standard.

3.1.3.4. Genetic Inverse Least Squares (GILS)

Genetic inverse least squares (GILS) model as understood from its name, is the combination of Genetic algorithms (GA) and ILS. Genetic algorithms (GA), global search and optimization methods, are used to eliminate wavelength selection problems from a large spectrum of data. GA is based on natural evolution and selection as developed by Darwin (Wang et al., 1991). Individuals can generate their offspring as a result of breeding only if they fit better and adapt in their environment. However, who are not fit and adapt in their environment will be eliminated from the population. Better solutions to problems can be enhanced only if the generations fit better to their environment. GA includes five basic steps as shown in Figure 3.3.

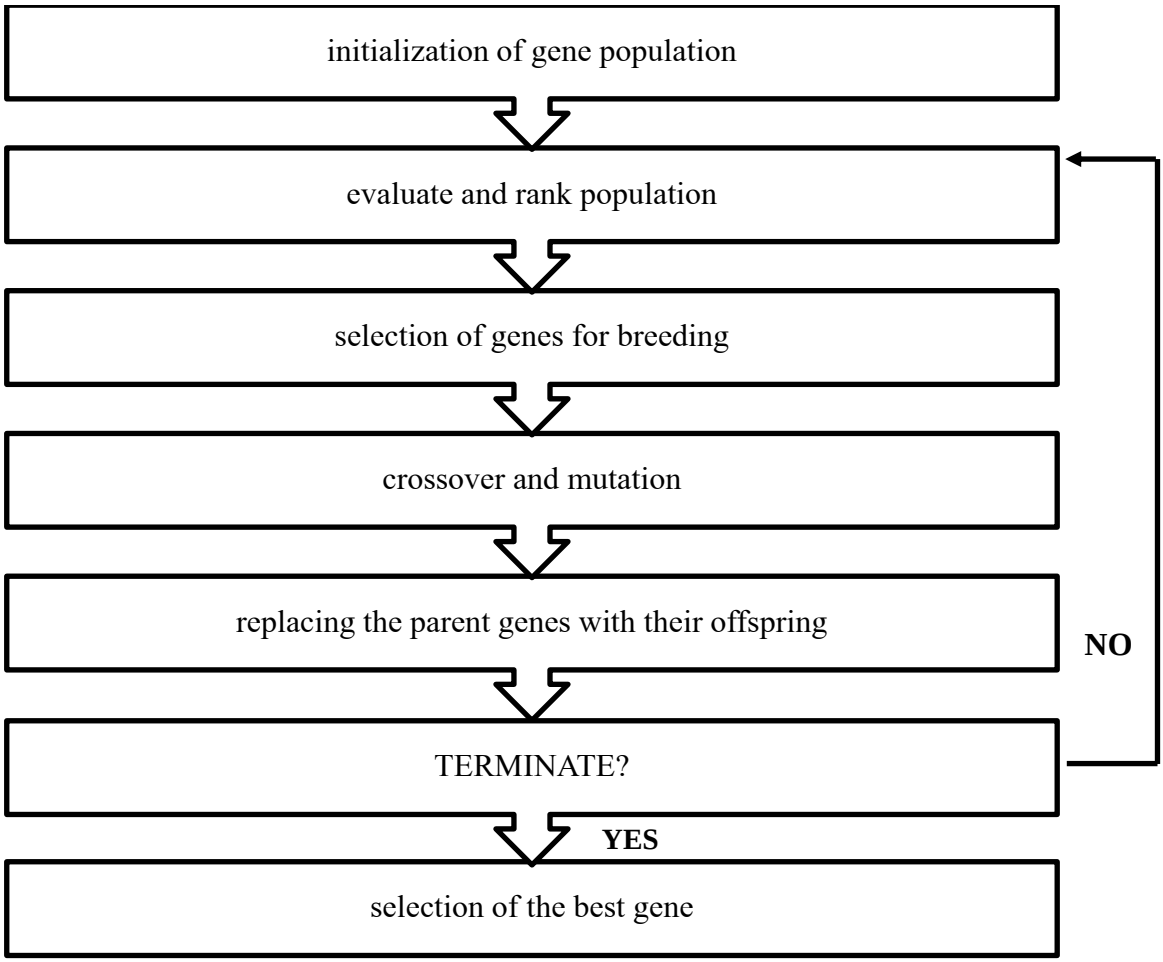


Figure 3.3. Flow chart of general genetic algorithm used in GILS

These steps consist of initialization of a gene population, evaluation of the population, selection of the parent genes for breading and mating, crossover and replacing parents with their offspring. The name of these steps originates in the biological feature of the genetic algorithm.

3.1.3.4.1. Initialization

A gene is a potential solution of given problems which changes from application to application. In the GILS method, the term ‘gene’ is referred as the collection of instrumental response at the wavelength range of the data set. The term ‘population’ is referred as the collection of individual genes in the current generation.

The first generation of genes is generated randomly with a fixed population size in initialization step. The number of the gene pool size is important because it

determines the estimating time. The number of the gene pool size is defined by user which permits breeding of each gene in the population. If the population size is large, ıt requires longer estimating time. Each gene consists of the number of instrumental responses which is obtained randomly in the range of fixed low limit and high limit. The lower limit was set to 2 in order to allow single-point crossover whereas the higher limit was set to reduce overfitting problems and reduce the estimating time.

3.1.3.4.2. Evaluate and Rank the Population

This step includes the evaluation of the genes with the use of fitness function. Besides, each gene's success for the calibration model can be obtained by the value of the fitness function. The value of the fitness function is found by the inverse of the standard error of calibration (SEC) :

$$\text{Fitness} = 1/\text{SEC} \tag{3.25}$$

SEC is calculated from the ILS model in which absorbance values from the selected wavelengths are used to construct the model. SEC is calculated from the following equation:

$$SEC = \sqrt{\frac{\sum_{i=1}^{m}\left(c_i - \hat{c}_i\right)^2}{m-2}} \tag{3.26}$$

where c_i is the reference and $\hat{c_i}$ is the predicted values of concentration of i^{th} sample and m is the number of samples. Two parameters are extracted from the sample number while calculating standard error of calibration. They are the slope of the actual vs. reference concentration plot and the intercept. In each step, the aim is that decreasing in standart error of calibration value.

3.1.3.4.3. Selection of Genes for Breeding

This step involves the selection of the parent genes from the present population for breeding. The goal is to generate best performing genes with higher fitness value and these genes will be able to pass their information to future generations. Thus, the genes which appropriate for the problem will generate better off-spring. The genes with low fitness values will be given lower chance to breed and hence most of them will be unable to survive.

Parent selection can be done by various methods. (Wang et al., 1991). Among these methods, top down selection method is the simplest one where the genes are permited to mate following ranking in the current gene pool, in a way that the first gene mates with the second gene, third one with the forth one and so on. This process is end up when all genes of the current population got a chance to breed. In GILS, roulette wheel selection method is used in which the chance of selecting gene is obtained according its fitness value. In this method, each segment in the roulette wheel represents a gene. The gene with the highest fitness value has the largest segment and the gene with the lowest fitness has the smallest segment. It was expected that a gene with high fitness has a higher chance of selection than for a gene with a low fitness when the wheel is rotated. There will be also the genes, which are chosen more then once in a certain period of time while some of them will not be chosen at all and will be eliminated from the gene pool. After all the main genes are chosen they are permited to mate top-down, wherewith the first gene (S_1) mates with the second gene (S_2), S_3 with S_4 and so on until all the genes mate. There is no ranking for the genes selected by roulette wheel so the genes with low fitness have a chance to mate with better performing genes which means that increasing the possibility of recombination.

3.1.3.4.4. Crossover and Mutation

In this step, genes are broken at random points and cross-coupling them as represented in the following example:

S_1 and S_2 are parent genes; S_3 and S_4 are their corresponding off-springs.

$S_1 = [\ A_{4255}\, A_{5732} \oplus A_{9237}\, A_{4890}\ ]$

$S_2 = [A_{5123}\, A_{8457}\, A_{9743}\, A_{7832} \oplus A_{8922}]$

$S_3 = [A_{4255}\, A_{5732}\, A_{8922}]$

$S_4 = [A_{5123}\, A_{8457}\, A_{9743}\, A_{7832}\, A_{9237}\, A_{4890}]$

Here, the first part of S_1 is combined with the second part of S_2 to give S_3 likewise the second part of S_1 with the first part of S_2 to give S_4. In this procedure using the single point crossover which is called in GILS. The symbol $\oplus$ is used to indicate the separation of the genes and the place where crossover takes place. Two point crossover and uniform crossover are also other types of crossover methods. In the uniform case, more as a result of a process where each gene is broken every step of many combinations are possible and mating. However, it may be disturb good genes. Single point crossover will not generate different off-spring if two parent genes have similar information that may occur in the choice of the roulette wheel selection, broken at the same point. In order to eliminate this problem, each gene is broken in two points and recombined can be used which is called two point crossover. In general, good genes are not destroyed via single point crossover however it supplies as many recombinations as other types of crossover schemes. It can also increase or decrease the number of base pairs in the off-spring on the mating.

3.1.3.4.5. Replacing the Parent Genes by Their Off-springs

After crossover, the parent genes are replaced by their off-springs. Following the evolution step, the ranking process is done according to their fitness values. Then the selection for breeding/mating starts again. This is concluded when a predefined number of iterations are finished.

Eventually, the gene with the lowest SEC (highest fitness) which means with the highest fitness value is selected to construct model. The concentrations of component that are being modelled in the validation set are predicted by this model. The success of the model in the prediction of the validation set is utilized using standard error of prediction (SEP) which is calculated as:

$$SEP = \sqrt{\frac{\sum_{i=1}^{m} \left(\hat{c}_i - c_i\right)^2}{m}} \tag{3.27}$$

where *m* is the number of validation samples in this case.

3.1.3.4.6. Termination

The termination of the algorithm is done by setting predefined iteration number for the number of breeding/mating cycles. However no extensive statistical test has been done to optimize it, though it can also be optimized. Since the random processes are heavily involved in the GILS, the program is set to run predefined number of times for each component in a given multi-component mixture. The run which have the lowest SEC for the calibration set and at the same time generating SEP for the validation set that is agreeable with SEC is the best run, is selected for evaluation and further analysis.

GILS has some major advantages over the classical univariate and multivariate calibration methods. First of all, in the model building and prediction steps involve quite simple mathematics than the other methods. Also, it has the advantages of the multivariate calibration methods by using reduced data set since the full spectrum is used to take genes. It is applicable to reduce nonlinearities that might be present in the full spectral region since it selects a subset of instrument response.

3.2. Experimental Design

Even though all chemist acceptance need to be skilful to design laboratory based experiments, formal statistical (or chemometric) rules have still been seen within the scope of mainstream chemistry. Generally, in real world experiments are time consuming and have high cost so chemist should have good assessment of the fundamentals of design. Due to several reasons such as mentioned above, chemist can be more productive only if they comprehend the principle of design, involving the following four main areas:

- Screening

- Optimisation
- Saving time
- Quantitative modelling (Brereton, 2003)

There are several statistical design have been widely used in literature but it is important to choose most suitable one for improving product performance and reliability, process capability and yield in our experiments.

3.2.1. Factorial Designs

If experiment includes a large number of factors, Factorial design is really useful due to its simplify. Although it has some limitations, factorial design is mostly preferred since it is easy to understand.

3.2.1.1. Full Factorial Designs

Two level full factorial design is used to obtain the influence of a number of effects on a response and to eliminate insignificance factors. If there is no need to detail predictions, the information from factorial designs is enough, especially qualitative (Brereton, 2003).

The following stages are used to construct the design and interpret the results.

1) The first step includes choosing a high and low level for each factor.

2) In order to consruct a standard design, the value of each factor is usually coded as − (low) or + (high).

3) Next, perform the experiments and obtain the response. Figure 3.4 shows an example of design matrix.

Intercept	Temperature	pH	Temp*pH
1	30	4	120
1	30	6	180
1	60	4	240
1	60	6	360

b0	b1	b2	b12
+	-	-	-
+	-	+	-
+	+	-	+
+	+	+	+

Figure 3.4. Design matrix of Full Factorial Designs

4) The next step is to analyse the data by setting up a design matrix. Interactions must be taken into account and set up a design matrix as given in Table 2.17 based on a model of the form $y = b_0 + b_1X_1 + b_2X_2 + b_{11}X_1X_2$.These are the possible four coefficients that can be obtained from the four experiments.

5) Calculate the coefficients. It is not necessary to employ specialist statistical software for this. In matrix terms, the response can be given by $y = \mathbf{D.b}$, where **b** is a vector of the four coefficients and **D** is the degrees of freedom. Simply use the matrix inverse so that $b = D^{-1}.y$. Note that there are no replicates and the model will exactly fit the data.

6) Finally, commentate the coefficient such as significance factor and interactions.

A major advantage of this design is that it allows finding significance or importance of factors and their interactions by directly the values of the b parameters. However, two level factorial designs have some disadvantages like they cannot consider quadratic terms since the experiments are performed only at two levels. Furthermore, there is no replicate information and they only enable a prediction within the experimental range.

3.2.1.2. Fractional Factorial Designs

Full factorial designs require large number of experiments which make them impracticable. A large number of factors can be important and have to be analysed but doing so many experiment is inefficient. Thus, it is important to reduce the number of experiment. Two level fractional factorial designs are used to reduce the number of experiments by 1/2, 1/4, 1/8 and so on. There are some rules which have been enhanced to produce these fractional factorial designs obtained only if taking the correct subset of the original experiments (Brereton, 2003). But some features should be taken into account:

- every column in the experimental matrix is different;
- in each column, there are an equal number of − and + levels;
- for each experiment at level + for factor 1, there are equal number of experiments for factors 2 and 3 and so on which are at levels + and −, and the columns are orthogonal.

Two level fractional factorial designs only exist when the number of experiments equals a power of 2. A half factorial design involves reducing the experiments from 2^k to 2^{k-1}. In more complex situations, such as 10 factor experiments, it is unlikely that there will be any physical meaning attached to higher order interactions, or at least that these interactions are not measurable. Therefore, it is possible to select specific interactions that are unlikely to be of interest and consciously reduce the experiments in a systematic manner by confounding these with lower order interactions.

Two level fractional factorial designs have some disadvantages:

- there are no quadratic terms since the experiments are performed only at two levels;
- there are no replicates
- the number of experiments must be a power of two

3.2.1.3. Central Composite Designs

More detailed model of a system is often needed to optimize the process and to obtain relation between response and the values of various factors (Brereton, 2003). Replicate information is not be provided by most exploratory designs not only any information on squared but also interaction terms. Also, the degrees of freedom for the lack-of-fit for the model (*D*) are often zero. More informative models reduce the volume of experimentation. Figure 3.5 represents such designs for a three factor experiment.

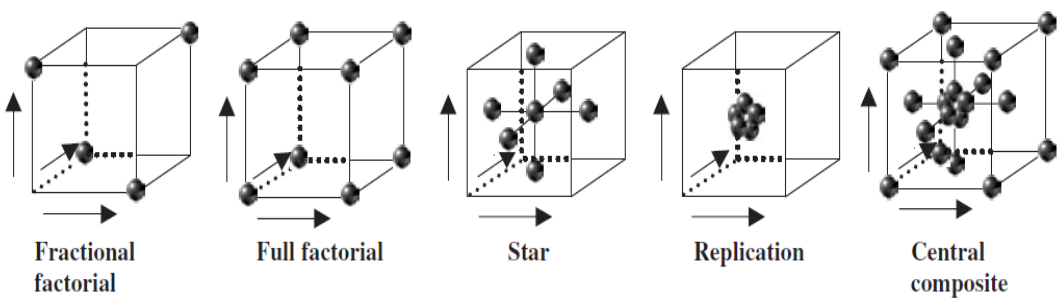


Figure 3.5. Construction of a three factor central composite design
(Source: Brereton, 2003)

1) In order to estimate three linear terms and interactions, a minimal three factorial design which includes four experiments is used. However, estimates of the interactions, replicates or squared terms are not provided by this design.
2) In order to estimate three linear terms and interactions, a minimal three factorial design which includes four experiments is used. However, estimates of the interactions, replicates or squared terms are not provided by this design.
3) Estimates of all interaction terms can be enhanced by extending this to eight experiments. When represented by a cube, these experiments are placed on the eight corners of the cube.
4) Another type of design, often indicated as a star design, can be used to estimate the squared terms. In order to do this, at least three levels are required for each factor, often indicated by +1, 0, and −1, with level '0' being in the centre. The reason for this is that there must be at least three points to fit a quadratic. For three factors, a star design consists of the centre point, and a point in the middle of each of the six faces of the cube.
5) Estimating the error is really significant so and this is typically performed by repeating the experiment in the centre of the design five times.
6) Performing a full factorial design, a star design and five replicates, results in twenty experiments. This design is often called a central composite design.

These twenty experiments can be divided as 10 parameters in the model, 5 degrees of freedom to determine replication error, degrees of freedom for the lack-of-fit shown in Figure 3.6.

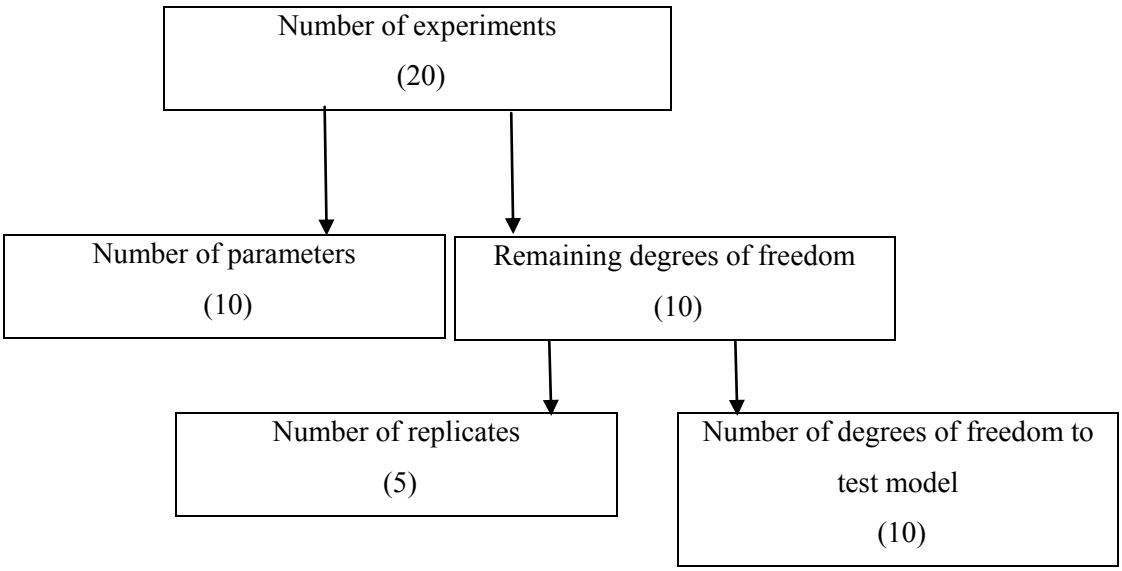


Figure 3.6 Degree of freedom tree for a three factor central composite design (Source: Brereton, 2003)

For statistical reasons, the position of star poins is determined at $\alpha = \pm\sqrt[4]{2^f}$ in which f is the number of factors. Thus, star points 1.41 for two factors, 1.68 for three factors and 2 for four factors. These designs are often termed to as rotatable central composite designs as all the points except the central points lie approximately on a circle or sphere or equivalent multidimensional surface, and are at equal distance from the origin.

After performing the design, the values of the terms are calculated by using regression and design matrices or almost any standard statistical software including Excel and obtain the significance of each term using ANOVA.

It is important to choice of the position of the axial (or star) points and how this relates to the number of replicates in the centre. Rotatability implies that the confidence in the predictions depends only on the distance from the centre of the design. Orthogonality implies that all the terms (linear, squared and two factor interactions) are orthogonal to each other in the design matrix, i.e. the correlation coefficient between any two terms (apart from the zero order term where it is not defined) equals 0.

CHAPTER 4

EXPERIMENTATION & INSTRUMENTATION

4.1. Protein Concentration Determination

Bradford protein assay method (Bradford, 1976) was modified in order to evaluate nonlinearity problem and to improve accuracy and sensitivity of this assay. In this method, Coomassie Brilliant Blue G-250 dye was used. In the classical Bradford method, an acidic solution of Coomassie is added to a protein solution, and the absorbance of the resulting mixture is measured at 595 nm (Bradford, 1976). In this study, Bradford protein assay was combined with multivariate calibration method that used all spectra in contrast to classical Bradford method to build up a calibration model by using the genetic algorithms based genetic inverse least squares (GILS).

4.1.1. Preparation of Bradford Reagent

Coomassie Brilliant Blue G-250 (CBB) was purchased from Sigma–Aldrich. The bradford assay reagent (1.17×10^{-4} M) was prepared by dissolving 10 mg of Coomassie Blue G250 in 5 mL of 95% ethanol. The solution was then mixed with 10 mL of 85% phosphoric acid and made up to 100 mL with distilled water . The reagent should be filtered through filter paper. Then stored in an amber bottle at 4 °C.

4.1.2. Preparation of Standard Protein Solution

Bovine serum albumin (BSA) was purchased from Sigma–Aldrich. BSA at a concentration of 0.02 mg/mL in distilled water was used as a stock solution. Standart protein solution was stored at -20^{o}C.

4.2. Instrumentation and Data Processing

Ultraviolet-Visible spectroscopic analyses were performed with Shimadzu 2550 UV-VIS spectrometer. This spectrometer was fitted out with 50W halogen and deuterium lamp as a source, Single monochromator as a wavelength selector, and Photomultiplier as a detector. Uv-Visible spectroscopic analyses of calibration standards and immobilization samples were done between 300 to 800 nm with using 10 mm path length disposable plastic sample holder. Duplicate measurements were done for each sample against two different types of blank namely Coomassie Blue G250 reagent and pure water. The collected spectra were transferred as text file format to set up a calibration model for the prediction. Calibration and validation sets were prepared as text files by using Microsoft Excel (MS Office 2007, Microsoft Corporation) program, that are required for the multivariate calibration method used in this study. The genetic algorithm based genetic inverse least squares (GILS) multivariate calibration method was written in MATLAB programming language using Matlab 7 (MathWorks Inc., Natick, MA). Partial least square (PLS) analysis were done by using Minitab 15 software (Minitab Inc., Coventry).

4.3. Design of the Data Sets

The first step in the development of a calibration model is the design of calibration set. In the design of calibration set it is important to choose the samples that have maximum and minimum concentration values. In addition, the success of model in prediction can be tested by independent validation (prediction) set. In order to build up a calibration model, 41 samples were prepared. Table 4.1 represents that concentrations of 41 Bovine serum albumin samples. Each sample mixture was scaled down to 5 mL final volume using different volume of BSA protein sample and pure water with a costant volume of CBB reagent (1 mL). The order of mixing reagents is, water, BSA and lastly CBB solution. Then mixture of these were incubated at room temperature for 5 minutes. Spectrum of each sample was taken at two different blanks namely Coomassie Blue G250 reagent and pure water that means two different calibration models were built up. After the analysis of GILS and PLS method, comparision was

done between these two methods in order to select suitable method for the Bovine serum albumin immobilization analysis.

Table 4.1 Concentration profile of 41 BSA protein samples

Sample No	BSA Concentration (µg/mL)	Sample No	BSA Concentration (µg/mL)	Sample No	BSA Concentration (µg/mL)
1	0.00	15	2.20	29	4.00
2	0.40	16	3.00	30	5.00
3	0.80	17	3.40	31	6.00
4	0.12	18	3.80	32	7.00
5	0.16	19	0.00	33	8.00
6	2.00	20	0.40	34	9.00
7	2.40	21	0.40	35	10.00
8	2.80	22	0.80	36	11.00
9	3.20	23	1.20	37	12.00
10	3.60	24	1.60	38	13.00
11	4.00	25	2.00	39	14.00
12	0.60	26	2.40	40	15.00
13	1.00	27	3.20	41	16.00
14	1.40	28	3.60		

4.4. Optimization of Conditions for Bovine Serum Albumin Immobilization on Chitosan Nanoparticles

4.4.1. Preparation of Chitosan Nanoparticles

For the purpose of chitosan nanoparticle preparation, chitosan and sodium tripolyphosphate pentabasic (TPP) were purchased from Sigma–Aldrich. Chitosan nanoparticles were prepared using the method readily constituted by Jiayin and Jianmin (Zhao and Wu 2006) with some modifications. In the procedure, 0.5 g of chitosan was dissolved in 100.0 mL 1.0% (v/v) acetic acid glacial. Then 6M NaOH was used to adjust the pH of the solution to 4.7. Chitosan nanoparticles were formed by the addition

of 1.0 mL of 0.25% (w/v) TPP to the chitosan aqueous solution dropwise by automatic micropipette under magnetic stirring at room temperature. After one hour incubation, the solution was centrifuged at 13,500 rpm for 30 min and chitosan nanoparticles were pelleted as transparent gel. The pellet of chitosan nanoparticles was dried by a freeze dryer.

4.4.2. Immobilization of Bovine Serine Albumin on Chitosan Nanoparticles

In immobilization procedure, Bovine serum albumin was immobilized by physical adsorption onto Chitosan nanoparticles. The BSA enzyme concentration of stock solution was 0.2 mg/mL. Immobilization on chitosan nanoparticles was carried out at different chitosan concentration (0.01-1.00 mg/mL), pH (5.0–11.0), immobilization time (5-180 min.) and temperature (15-60 ^{o}C) which were defined based on the previous literature studies. During the preparation of immobilization samples, 0.10 mL (0.2 mg/mL) enzyme stock solution is taken for each sample and to this solution chitosan nanoparticles (70 mg/mL) from 0.52 μL to 52 μL was added in order to provide a concentration ragen from 0.01 to 1.00 mg/mL for chitosan nanoparticles. To complete the final volume to 1.00 mL, aproximately 0.75 mL corresponding buffer solutionto was added to each sample. The mixtures were incubated at different temperature for different immobilization time with shaking (100 rpm). Immobilized enzyme on chitosan nanoparticles were centrifuged at 3000 rpm for 20 min. The supernatant solution was used to estimate the residual amount of bovine serum albumin. The amount of bovine serum albumin was determined in clear supernatant by the method of Bradford Protein Assay which was combined with GILS using supernatant of nonloaded nanoparticles as a basic correction. Immobilization efficiency was given by percent of yield. The yield of bovine serum albümin immobilization was calculated according to the equation given below:

$$\%Yield = ((Total\ BSA - Free\ BSA) / Total\ BSA) \times 100 \quad (4.1)$$

4.4.3. Experimental Design and Data Analysis

Central composite design (CCD) was employed to determine optimum conditions for the maximum immobilization yield and to investigate importance and interaction of the factors affecting on immobilization. This optimaziton process includes three major steps:

- performing the statistically designed experiments,
- fitting experimentally determined response data into a quadratic model
- estimating the coefficients in a mathematical model, and predicting the response and checking the adequacy of the model (Tanyıldızı et al., 2006).

The factors and their values were chitosan concentration from 0.01 to 1 mg/mL, immobilization time from 5 to 180 minutes, temperature from 15 to 60^{o}C and pH from 5 to 11. Each factor was coded at five levels: −2, −1, 0 +1, and +2. The factor were coded according to following equation:

$$x_i = \frac{X_i - X_0}{\Delta X_i} \tag{4.2}$$

where x_i is the dimensionless coded value of an independent variable, X_i is the real value of an independent variable, X_0 is the real value of an independent variable (X_i) at the center point and ΔX_i is the step change value which is sum of the axial points. The relationship between the coded and uncoded (actual) values is shown in Table 4.2.

Table 4.2. Range of coded and uncoded values for central composite design

Variables	Symbol coded	Range and level				
		-2	-1	0	1	2
Immobilization time (minute)	X_1	5.0	49.0	92.5	136.0	180.0
Temperature(°C)	X_2	15.0	26.0	37.5	49.0	60.0
pH	X_3	5.0	7.0	8.0	9.0	11.0
Chitosan concentration (mg/mL)	X_4	0.01	0.30	0.51	0.71	1.00

5-level-4-factor central composite design leading to 30 runs that composes of 16 factorial points, 8 axial points and 6 replicates at the center points was carried out. Experimental run was performed in a random order to reduce effect of uncontrolled factors. The corresponding central composite design and their values were shown in Table 4.3

Table 4.3. Five-level and four-factor central composite design with actual values, coded values and the response of (immobilization yield) the experiments.

Experiment	Immobilization Time (minute) (X_1)	Temperature (°C) (X_2)	pH (X_3)	Chitosan Concentration (mg/mL) (X_4)	X_1	X_2	X_3	X_4	Yield (%)
1	49.0	26.0	7.0	0.30	-1	-1	-1	-1	24.75
2	49.0	26.0	7.0	0.71	-1	-1	-1	1	30.66
3	49.0	26.0	9.0	0.30	-1	-1	1	-1	18.26
4	49.0	26.0	9.0	0.71	-1	-1	1	1	8.35
5	49.0	49.0	7.0	0.30	-1	1	-1	-1	35.48
6	49.0	49.0	7.0	0.71	-1	1	-1	1	22.23
7	49.0	49.0	9.0	0.30	-1	1	1	-1	52.50
8	49.0	49.0	9.0	0.71	-1	1	1	1	3.16
9	136.0	26.0	7.0	0.30	1	-1	-1	-1	44.83
10	136.0	26.0	7.0	0.71	1	-1	-1	1	37.40
11	136.0	26.0	9.0	0.30	1	-1	1	-1	58.60
12	136.0	26.0	9.0	0.71	1	-1	1	1	9.56
13	136.0	49.0	7.0	0.30	1	1	-1	-1	53.51
14	136.0	49.0	7.0	0.71	1	1	-1	1	17.61
15	136.0	49.0	9.0	0.30	1	1	1	-1	88.08
16	136.0	49.0	9.0	0.71	1	1	1	1	25.41
17	5.0	37.5	8.0	0.51	-2	0	0	0	3.54
18	180.0	37.5	8.0	0.51	2	0	0	0	46.86
19	92.5	15.0	8.0	0.51	0	-2	0	0	27.41
20	92.5	60.0	8.0	0.51	0	2	0	0	31.46
21	92.5	37.5	5.0	0.51	0	0	-2	0	45.77
22	92.5	37.5	11.0	0.51	0	0	2	0	57.48
23	92.5	37.5	8.0	0.01	0	0	0	-2	56.77
24	92.5	37.5	8.0	1.00	0	0	0	2	23.63
25	92.5	37.5	8.0	0.51	0	0	0	0	22.44
26	92.5	37.5	8.0	0.51	0	0	0	0	13.40
27	92.5	37.5	8.0	0.51	0	0	0	0	20.73
28	92.5	37.5	8.0	0.51	0	0	0	0	22.65
29	92.5	37.5	8.0	0.51	0	0	0	0	25.51
30	92.5	37.5	8.0	0.51	0	0	0	0	21.95

A second order full quadratic model was constructed with CCD design data giveb in Table 4.3. Predicted responses were obtained with the following second order polynomial equation:

$$y = b_0 + \sum_{i=1}^{4} b_i x_i + \sum_{i=1}^{4} b_{ii} x_i^{\,2} + \sum_{i=j}^{3} \sum_{j=i+1}^{4} b_{ij} x_{ij} \qquad (4.3)$$

where y is the response (immobilization yield), b_0 is the intercept term, b_i is the linear effects, b_{ii} is the quadratic effects, b_{ij} are the interaction effects and x_i are independent factors.

Experimental results were analyzed using the regression analysis. The polynomial equation for the response was validated by ANOVA (analysis of variance) to determine the significance of each term in the equation and also to estimate the goodness of fit in each case. The 3-D response surface and contour plot were obtained by using Matlab 7 (MathWorks Inc., Natick, MA). In addition to three dimensional surface plots contour plots were also plotted to establish the optimum conditions for specific activity.

CHAPTER 5

RESULTS AND DISCUSSION

5.1. Calibration Results

5.1.1. Ultraviolet-Visible Absorption Spectroscopy

Figure 5.1 represents the spectra of BSA, CBB and BSA-CBB complex against water blank. Figure 5.1 shows that BSA does not absorb UV-Vis light at any wavelength. Figure 5.2 shows the spectra of BSA, CBB and BSA-CBB complex against water blank with secondary axis for BSA-CBB complex. Secondary axis was used for BSA-CBB complex in order to see its spectrum with enlarged scale. The spectra of 41 standard samples given in Table 4.1 against water blank are shown in Figure 5.3. As it can be seen from these figures protein complex absorbs the UV-Vis light not only 595 nm but also 465 nm and 620 nm. This result can be expressed by the properties of dye. CBB has three ionic species which have pH dependent absorbance spectra so other forms of CBB can bind the protein, without non-electrostatic interactions, to form a complex. In addition, if protein are processed by mixing, turbidity occurs. When protein denature under shear stress, they tend to form aggregates that increase solution turbidity which reults in a shift in the spectra to the lower wavelenght.

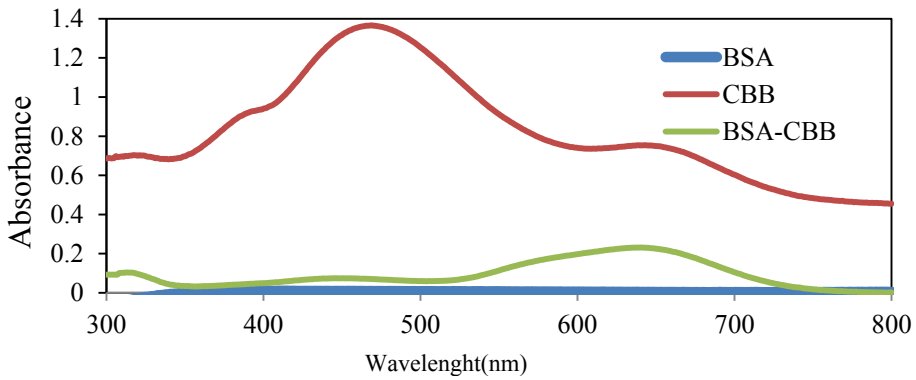


Figure 5.1. UV-Vis spectra of BSA standart, CBB and BSA-CBB complex against water blank

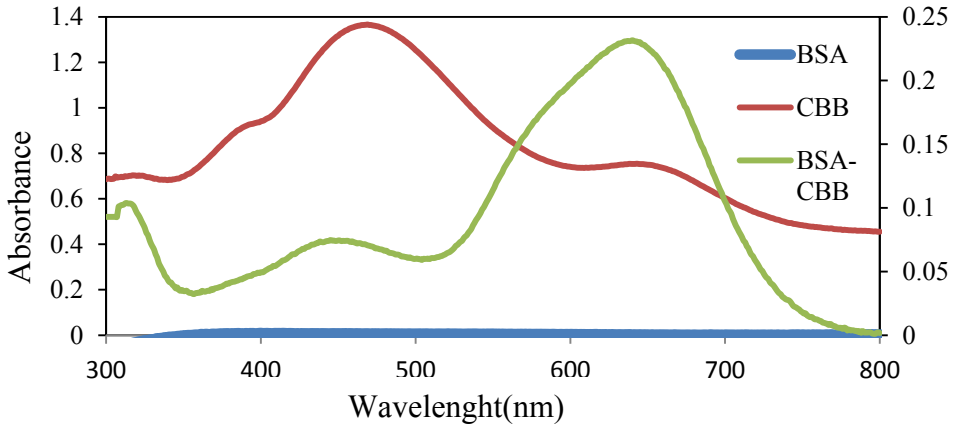


Figure 5.2. UV-Vis spectra of BSA standart, CBB and BSA-CBB complex against water blank with secondary axis for BSA-CBB complex

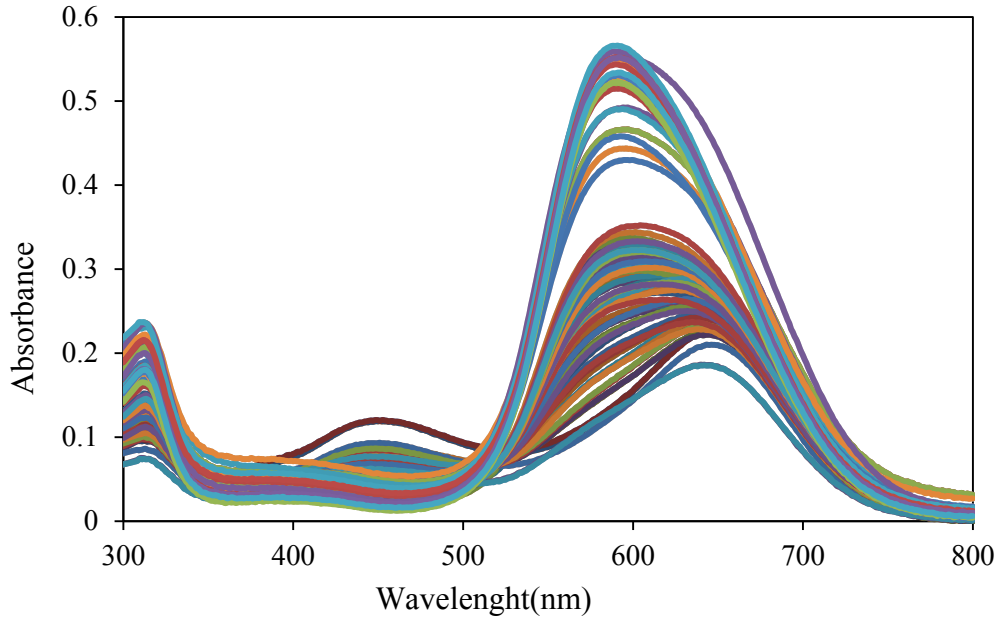


Figure 5.3. Uv-Vis spectra of 41 standard samples of BSA-CBB complex against water blank

Figure 5.4 illustrates spectra of BSA and BSA-CBBcomplex against Coomassie Blue G250 reagent (CBB) blank. Figure 5.5 represents spectra of BSA standart and protein-dye complex against CBB blank with secondary axis for BSA-CBB complex and the spectra of 41 samples of BSA-CBB complex against Coomassie Blue G250 reagent (CBB) blank are shown in Figure 5.6. When CBB reagent was used as blank, shifts in the spectra was lower than the water blank. While prepearing standart samples of BSA at a various concentration, excess dye was used so that concentration of free dye is more higher than the BSA concentration in every sample. After substracting the absorbance of blank sample, absorbance of the free dye became lower therefore it was

possible to measure only protein-dye comlex absorption. Also, there are negative peaks in these spectra due to subtracting the value of the blank sample from the value of BSA standart samples because dye concentration in the blank sample was higher than the free dye concentration in the samples.

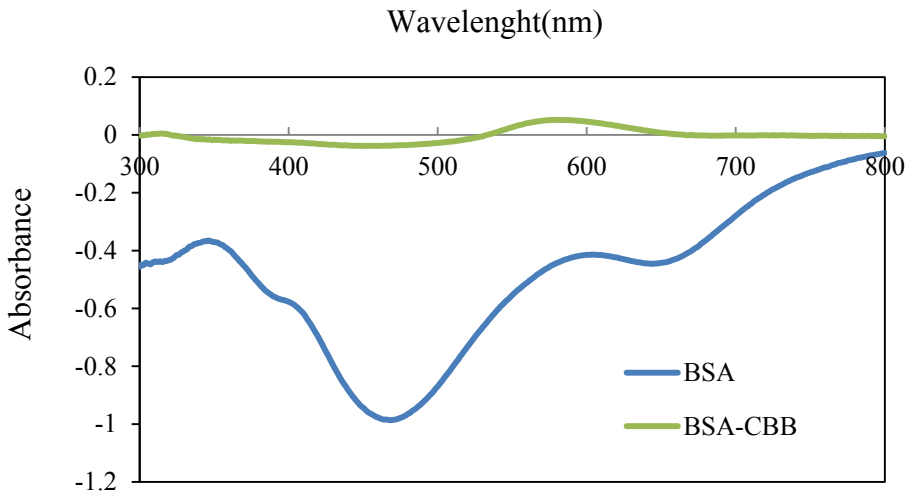


Figure 5.4. UV-Vis spectra of BSA standart and BSA-CBB complex against CBB blank

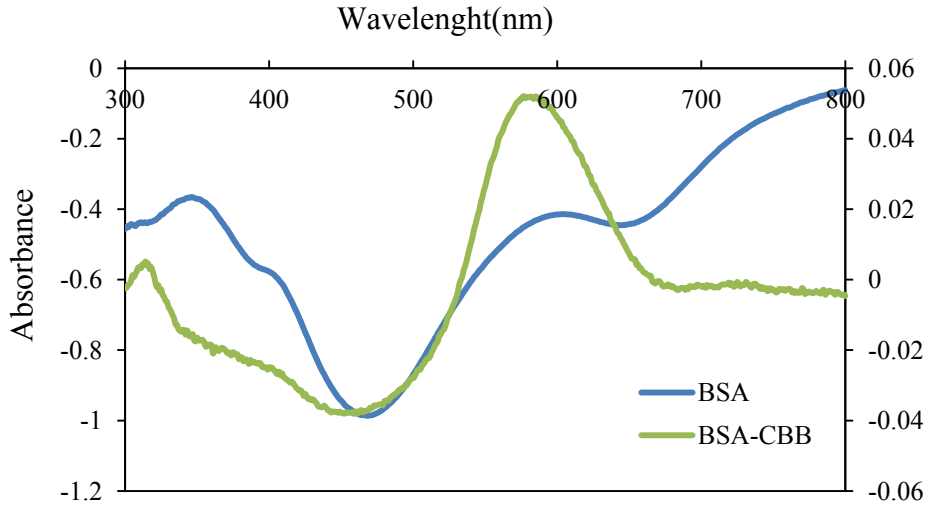


Figure 5.5. UV-Vis spectra of BSA standart and BSA-CBB complex against CBB blank with secondary axis for BSA-CBB complex

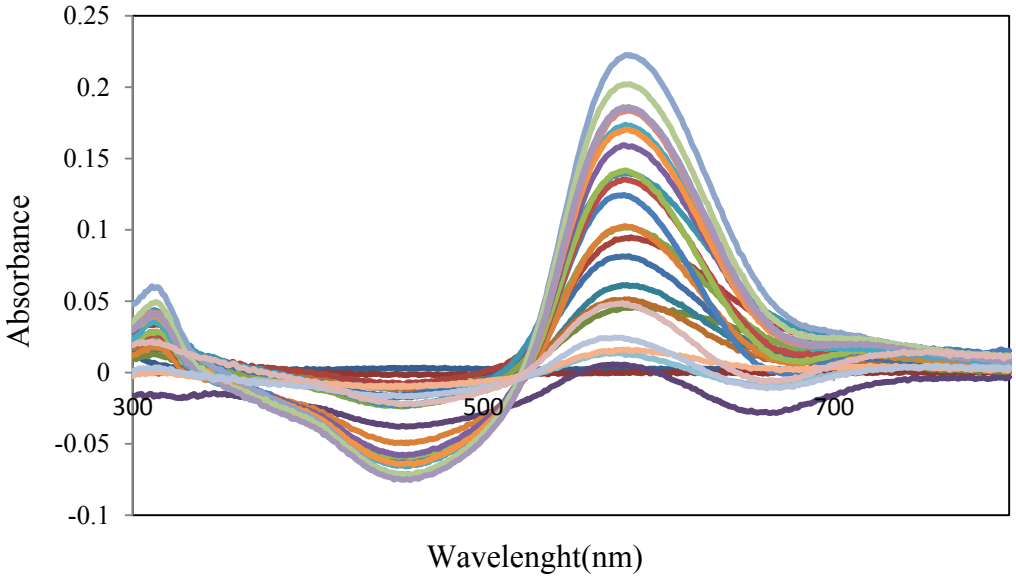


Figure 5.6. UV-Vis spectra of 41 samples of BSA-CBB complex against CBB blank

As is known, protein-dye complex stability is pH dependent. Thus, the effect of pH in UV-Vis spectra of BSA-CBB complex should be evaluated given that the extent of the Bradford reaction can be drastically affected when protein determinations are carried out under different pH conditions. In this way, BSA standards were buffered at pH 4, pH 7 and pH 10 in order to see their effect on the spectra. The absorption spectra of BSA-dye complex at three different concentrations (4µg/mL, 8µg/mL and 12µg/mL) in three type of buffer (pH 4, pH 7 and pH 10) against corresponding buffer blank are shown in Figure 5.7.

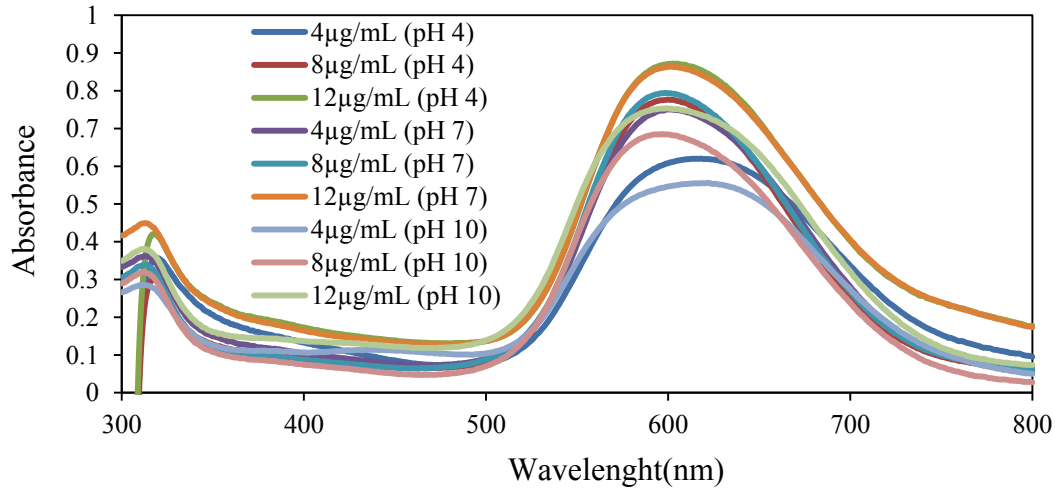


Figure 5.7. UV-Vis spectra of BSA-CBB complexs in buffer solutions against buffer corresponding blank. Protein concentrations were 4µg/mL, 8µg/mL and 12µg/mL BSA.

Figure 5.8 show spectra of BSA-dye complex at 8 µg/mL against the three buffer blanks. Here, 8 µg/mL BSA concentration was the one which is used in the immobilization studies. It is seen that spectral intensity at pH 4 and pH 7 were higher than pH 10. And also, effect of pH on the spectral shift was not significant. In fact, one of the factor in the immobilization study in CCD given in Table 4.3 was pH and therefore all the solution were carried out in pure water instead on bufferd medium since the effect of buffer on the spectral shift was negligible.

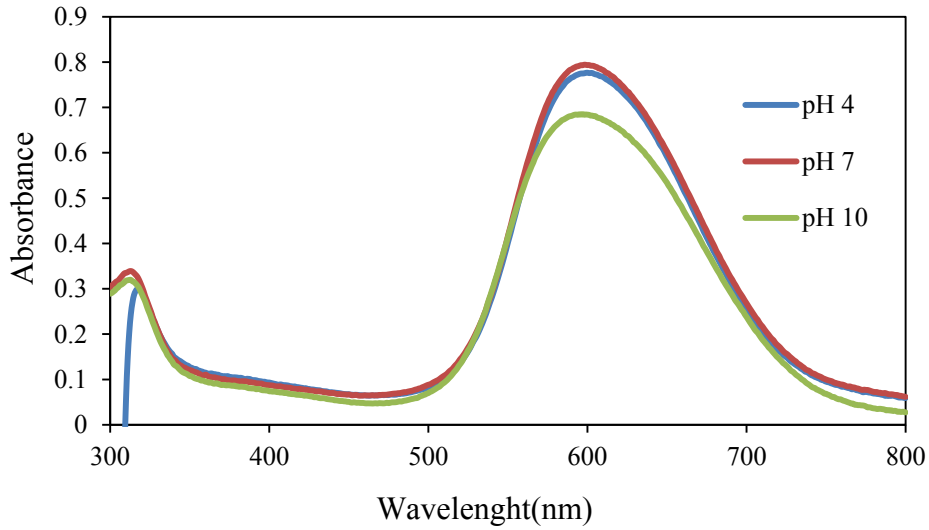


Figure 5.8. UV-Vis spectra of BSA-CBB complex in buffer solutions against buffer blank. Protein concentration was 8µg/mL BSA

The effect of solvent on BSA-CBB complex absorption when buffered CBB taken as blank was investigated using different type of buffer against CBB blank. The BSA concentrations were chosen 4µg/mL, 8µg/mL and 12µg/mL. Figure 5.9 show the absorption spectra of BSA-dye complex at a different concentrations (4µg/mL, 8µg/mL and 12µg/mL) in three type of buffer (pH 4, pH 7 and pH 10) against CBB blank. Figure 5.10 show spectra of BSA-dye complex at 8 µg/mL against CBB blank in order to investigate pH effects in detailed. It was observed that shift in the spectra was independent of pH at 8 µg/mL BSA concentration. A similar spectral intensity with CBB blank was observed in the samples containing pH 4 and pH 7 buffers. From the results obtained, it is possible to conclude that the pectral features were not significantly diffrerent among the BSA-dye complex spectra at three diffrent pH conditions studied.

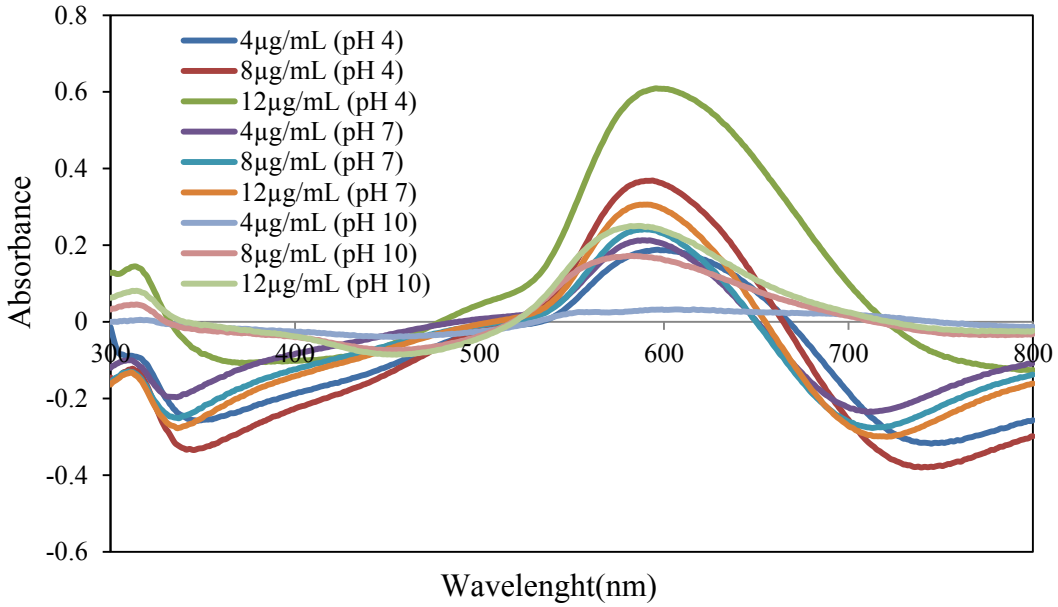


Figure 5.9. UV-Vis spectra of BSA-CBB complexs in buffer solution against CBB blank prepare in corresponding buffer. Protein concentrations were 4µg/mL, 8µg/mL and 12µg/mL BSA

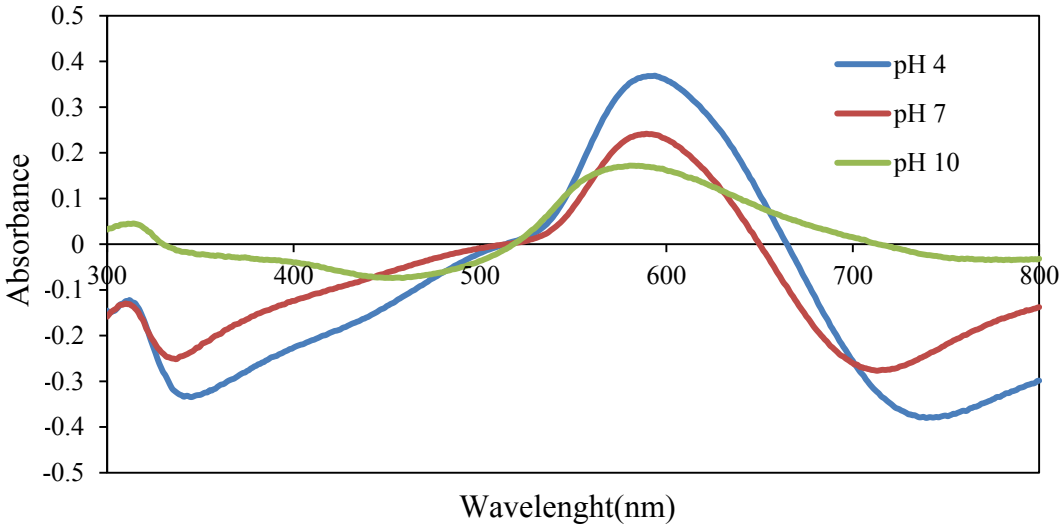


Figure 5.10. UV-Vis spectra of BSA-CBB complex in a buffer solution against CBB blank prepare in corresponding buffer. Protein concentration was 8µg/mL BSA

It is obvious that determining protein concentration at a single wavelenght was not ensure to obtain accurate results due to these spectral problems. Therefore, full absorption spectra were recorded and used for the analysis in this study.

5.1.1.1. Univariate Calibration Results For Coomassie Blue G250 Reagent (CBB) Blank

Univariate calibration model was composed of 59 BSA standard samples for calibration set shown in Table 5.1 and 19 BSA samples for validation set shown in Table 5.2.

Table 5.1. Concentration profile of the calibration samples against CBB blank

Sample No	Concentration (µg/mL)	Sample No	Concentration (µg/mL)	Sample No	Concentration (µg/mL)
1	0.40	21	2.20	41	4.00
2	0.40	22	3.00	42	5.00
3	0.80	23	3.00	43	6.00
4	1.20	24	3.40	44	6.00
5	1.20	25	3.80	45	7.00
6	1.60	26	3.80	46	8.00
7	2.00	27	0.40	47	8.00
8	2.00	28	0.80	48	9.00
9	2.40	29	0.80	49	10.00
10	2.80	30	1.20	50	10.00
11	2.80	31	1.60	51	11.00
12	3.20	32	1.60	52	12.00
13	3.60	33	2.00	53	12.00
14	3.60	34	2.40	54	13.00
15	4.00	35	2.40	55	14.00
16	0.60	36	2.80	56	14.00
17	0.60	37	3.20	57	15.00
18	1.00	38	3.20	58	16.00
19	1.40	39	3.60	59	16.00
20	1.40	40	4.00		

Table 5.2. Concentration profile of the validation samples against CBB blank

Sample No	Concentration (µg/mL)	Sample No	Concentration (µg/mL)	Sample No	Concentration (µg/mL)
1	0.80	8	3.40	15	7.00
2	1.60	9	0.40	16	9.00
3	2.40	10	1.20	17	11.00
4	3.20	11	2.00	18	13.00
5	4.00	12	2.80	19	15.00
6	1.00	13	3.60		
7	2.20	14	5.00		

Figure 5.11 represents calibration graphs of Bradford protein assay at 595 nm. As seen in these figures, there is a significant nonlinearity in the response pattern after 8.0 µg/mL BSA concentration. Calibration graph shows distinct curvature in the range of 0.0–16.0 µg/mL BSA. In order to eliminate nonlinearity problem of this assay, concentration range was reduced to 0.0-8.0 µg/mL. The correlation coefficient, R^2, is increased from 0.8408 to 0.8720 by reducing concentration range. However, while this reduction causes a decrease in the dynamic range of this assay the improvemet in the calibratin quality is still not sufficient.

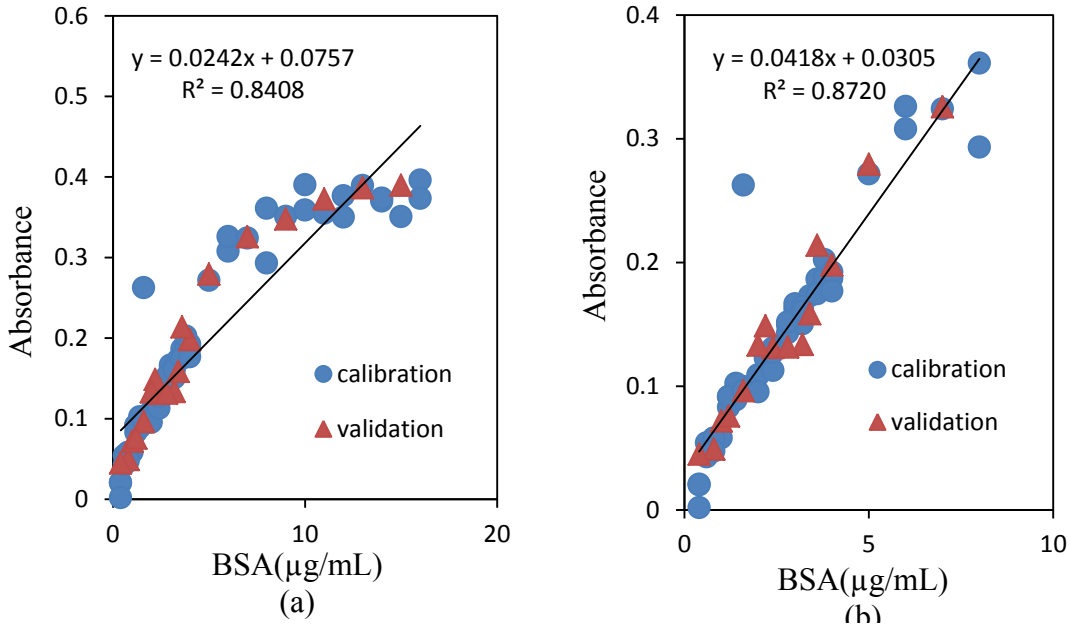


Figure 5.11. Calibration graphs of Bradford protein assay at 595 nm against CBB blank a)concentration range between 0.0-16.0 µg/mL BSA and b) concentration range between 0.0-8.0 µg/mL BSA

When these impacts are considered, univariate calibration method is not suitable to determine the protien concentration at a single wavelength. For this reason, a genetic algorithm, effective to solve wavelength selection problems from a large spectrum of data, based multivariate calibration method is needed. GILS method is a genetic algorithm based multivariate calibration technique, it was expected that it could select certain combination of wavelengths which had maximum correlation with the protein concentration in sample. And also another multivariate calibration method, Partial Least Square (PLS), is used to eliminate problems of Bradford protein assay since PLS is a full-spectrum methods so that it was expected that it will reduce wavelength shift problem of this assay.

5.1.1.2. GILS Results For Coomassie Blue G250 Reagent (CBB) Blank

In order to construct calibration model 41 samples were prepared and duplicate measurements were done. The calibration set composed of 59 samples, and validation set composed of 19 samples which are shown in Table 5.3 and Table 5.4, respectively, along with the GILS predicted BSA concentrations. In the design of calibration set, samples were randomly selected with having minimum and maximum concentration values.

Table 5.3. Actual versus genetic inverse least squares (GILS) predicted protein concentration for calibration samples against CBB blank

Sample No	Concentration (µg/mL)	Predicted Concentration(µg/mL)	Sample No	Concentration (µg/mL)	Predicted Concentration(µg/mL)
1	0.40	0.19	31	1.60	1.51
2	0.40	0.35	32	1.60	1.83
3	0.80	0.79	33	2.00	1.82
4	1.20	0.93	34	2.40	2.21
5	1.20	1.48	35	2.40	2.34
6	1.60	1.57	36	2.80	2.88
7	2.00	1.91	37	3.20	3.10
8	2.00	2.14	38	3.20	3.00
9	2.40	2.41	39	3.60	4.12
10	2.80	2.83	40	4.00	3.78
11	2.80	2.99	41	4.00	3.79

(cont. on next page)

Table 5.3 (cont.)

Sample No	Concentration (µg/mL)	Predicted Concentration(µg/mL)	Sample No	Concentration (µg/mL)	Predicted Concentration(µg/mL)
12	3.20	3.11	42	5.00	5.54
13	3.60	3.94	43	6.00	6.17
14	3.60	3.64	44	6.00	6.03
15	4.00	4.22	45	7.00	6.45
16	0.60	0.72	46	8.00	8.39
17	0.60	0.65	47	8.00	8.14
18	1.00	0.73	48	9.00	9.67
19	1.40	1.16	49	10.00	8.77
20	1.40	1.85	50	10.00	10.45
21	2.20	2.27	51	11.00	11.06
22	3.00	2.64	52	12.00	12.15
23	3.00	3.13	53	12.00	11.96
24	3.40	3.83	54	13.00	12.88
25	3.80	3.87	55	14.00	13.92
26	3.80	4.02	56	14.00	13.77
27	0.40	0.20	57	15.00	14.60
28	0.80	0.73	58	16.00	15.95
29	0.80	0.40	59	16.00	15.72
30	1.20	1.52			

Table 5.4. Actual versus genetic inverse least squares (GILS) predicted protein concentration for validation samples against CBB blank

Sample No	Concentration (µg/mL)	Predicted Concentration(µg/mL)	Sample No	Concentration (µg/mL)	Predicted Concentration(µg/mL)
1	0.80	0.76	11	2.00	1.98
2	1.60	1.60	12	2.80	2.23
3	2.40	2.38	13	3.60	4.44
4	3.20	3.08	14	5.00	5.56
5	4.00	4.57	15	7.00	6.20
6	1.00	1.08	16	9.00	9.77
7	2.20	2.61	17	11.00	10.72
8	3.40	2.99	18	13.00	12.81
9	0.40	-0.20	19	15.00	14.72
10	1.20	1.41			

Actual BSA concentration versus predicted values based on UV-VIS spectra using GILS method are shown in Figure 5.12. Calibration models for protein concentration determination gave standard error of cross validation (SECV) and standard error of prediction (SEP) values as 0.30 µg/mL and 0.45 µg/mL for calibration and independent test sets, respectively. The R^2 value of regression lines for BSA concentration was 0.9954.

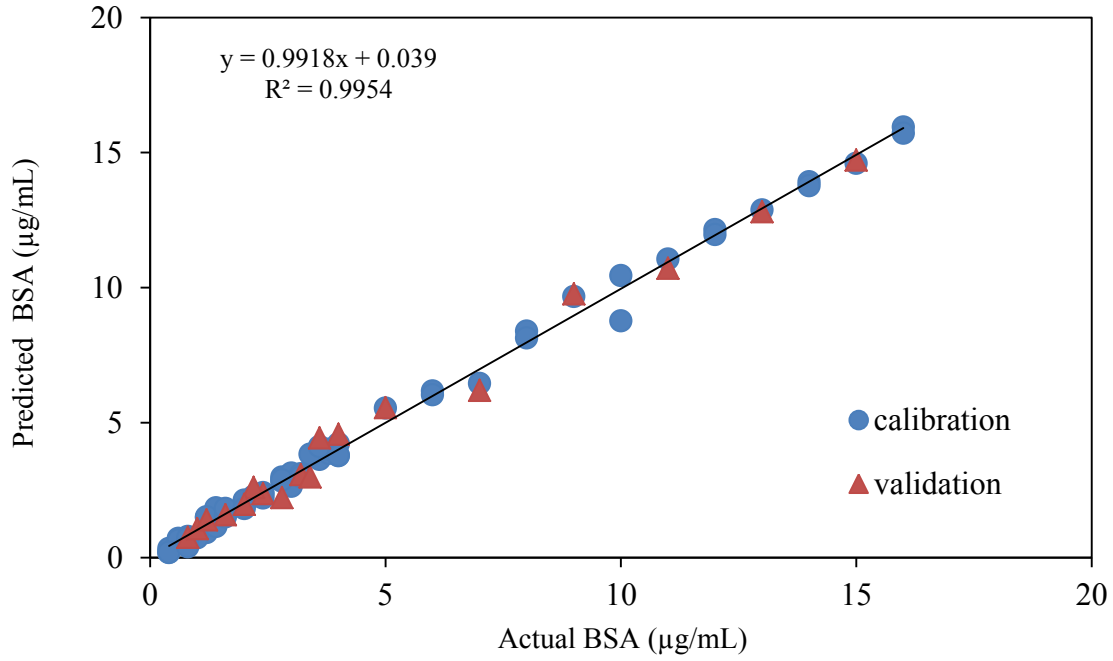


Figure 5.12. Actual versus genetic inverse least squares (GILS)-predicted protein concentration against CBB blank

When these SECV and SEP values are examined, it is seen that these values are compatible with each other, which illustrates a good prediction for protein concentration determination. When the overall calibration performance of the models is examined, it is possible to state that GILS provides a significant improvement in linearity of Bradford protein assay over the univariate calibration methods which are shown in Figure 5.11. Here, cross-validation procedure was carried out where the GILS algorithm is performed on m-1 spectra and the left out spectrum is used to validate the model for m calibration spectra. This process is repeated until each spectrum is left out once in the calibration set. GILS algorithm uses the optimum number of wavelengths in order to achive better performance in prediction.

Since GILS is a method which depends on wavelength selection, the distribution of selected wavelengths in multiple runs over the entire full spectral region would be

usuful to observe selectivity of GILS over the full spectral range. The frequency distributions of selected wavelengths in 250 runs with 30 genes and 100 iterations were plotted against wavelength range for BSA concentration in Figure 5.13.

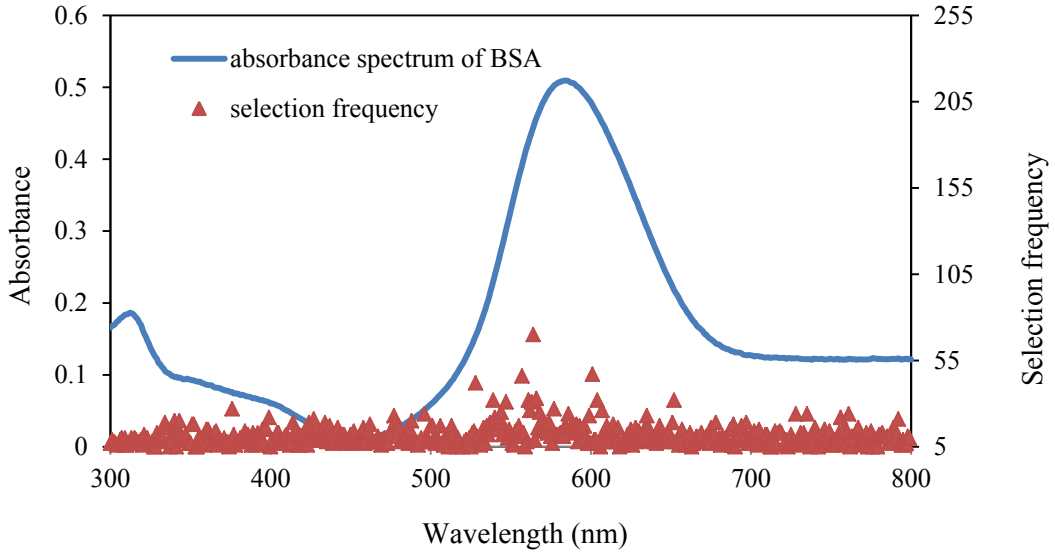


Figure 5.13. Frequency distribution of GILS selected UV-Vis wavelengths for BSA concentration against CBB blank

As can be seen from the figures, the frequency of the selected wavelength is significantly higher in the left shoulder of the peak around 590 nm for the BSA-CBB complex. Here, it is evident that highest selected wavlengths are now shifts away from the peak maxima as a resut of the shif in the BSA-CBB complex. This indicates that the GILS method selects wavelengths, where the most concentration related information is contained. As a result of this, it can be said that GILS method can be used for the determination of BSA concentration over a much larger concentration range without any distortion from the wavelength shift of the BSA-CBB complex as in the case of univariate calibration.

5.1.1.3. PLS Results For Coomassie Blue G250 Reagent (CBB) Blank

Partial least squares (PLS) is a popular multivariate calibration method for quantitative analysis of spectral data, and PLS performs the calibration using the full-spectrum information to construct a regression model to determine the property of interest.

As in the univariate calibration and GILS, the same calibration and validation sets were used with PLS in order to compare the performance of the three methods. Table 5.5 and Table 5.6. show the actual and PLS predicted concentration values of BSA.

Table 5.5. Actual versus partial least squares (PLS) predicted protein concentration for calibration samples against CBB blank

Sample No	Concentration (µg/mL)	Predicted Concentration(µg/mL)	Sample No	Concentration (µg/mL)	Predicted Concentration(µg/mL)
1	0.40	-0.47	31	1.60	1.58
2	0.40	-0.37	32	1.60	5.59
3	0.80	0.48	33	2.00	2.94
4	1.20	1.77	34	2.40	1.29
5	1.20	1.41	35	2.40	2.91
6	1.60	2.04	36	2.80	2.90
7	2.00	1.91	37	3.20	3.30
8	2.00	1.73	38	3.20	3.25
9	2.40	2.17	39	3.60	3.72
10	2.80	2.99	40	4.00	3.82
11	2.80	3.30	41	4.00	3.27
12	3.20	3.37	42	5.00	5.78
13	3.60	4.49	43	6.00	6.64
14	3.60	4.34	44	6.00	6.56
15	4.00	4.02	45	7.00	5.93
16	0.60	0.45	46	8.00	8.60
17	0.60	0.66	47	8.00	9.19
18	1.00	0.63	48	9.00	9.88
19	1.40	1.37	49	10.00	5.73
20	1.40	1.57	50	10.00	11.52
21	2.20	2.42	51	11.00	11.97
22	3.00	2.55	52	12.00	11.89
23	3.00	2.49	53	12.00	12.17
24	3.40	3.95	54	13.00	12.13
25	3.80	3.78	55	14.00	14.34
26	3.80	4.13	56	14.00	13.89

(cont. on next page)

Table 5.5 (cont.)

Sample No	Concentration (µg/mL)	Predicted Concentration(µg/mL)	Sample No	Concentration (µg/mL)	Predicted Concentration(µg/mL)
27	0.40	-0.21	57	15.00	14.05
28	0.80	0.99	58	16.00	15.39
29	0.80	-0.14	59	16.00	14.41
30	1.20	1.73			

Table 5.6. Actual versus partial least squares (PLS) predicted protein concentration for validation samples against CBB blank

Sample No	Concentration (µg/mL)	Predicted Concentration(µg/mL)	Sample No	Concentration (µg/mL)	Predicted Concentration(µg/mL)
1	0.80	0.52	11	2.00	1.71
2	1.60	2.17	12	2.80	1.73
3	2.40	2.83	13	3.60	3.75
4	3.20	3.02	14	5.00	6.33
5	4.00	4.80	15	7.00	5.95
6	1.00	1.02	16	9.00	10.02
7	2.20	2.78	17	11.00	12.16
8	3.40	2.89	18	13.00	12.71
9	0.40	-0.27	19	15.00	14.28
10	1.20	1.31			

Actual BSA concentration versus predicted values based on UV-Vis spectra using PLS method are shown in Figure 5.14. Calibration models for protein concentration determination gave the standard error of cross validation (SECV) was found 0.98 µg/mL and standard error of prediction (SEP) was found 0.70 µg/mL. The R^2 value of regression lines for BSA concentration was 0.9732.

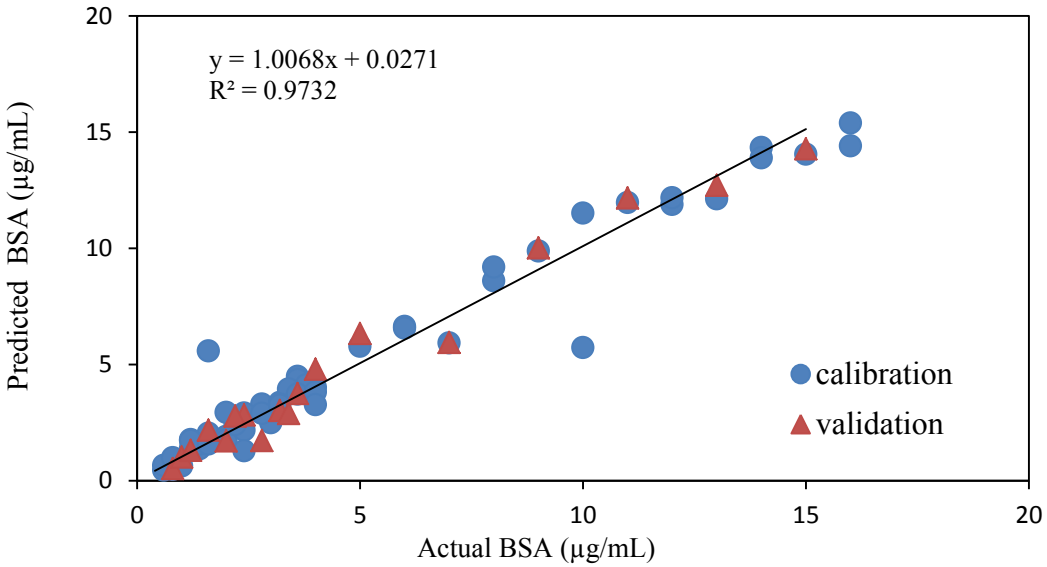


Figure 5.14. Actual versus partial least squares (PLS) predicted protein concentration against CBB blank

PLS model represents much better prediction ability compared to univariate valibration as is evident from R^2, SECV and SEP values. This is expected since PLS is a full-spectrum method so it eliminates the wavelenght shift problem of Bradford protein assay. Here, PLS algorithm uses the optimum number of factors in order to succeed better performance in prediction. The cross-validation procedure was applied, consisting of removing one of the training samples in turn, and using only the remaining ones for construction of the latent factors and regression. The optimal number of PLS factors were found 17 according to full cross-validation procedure.

The PLS algorithm was carried out to regulate regression model depend on the full-spectrum information (300-800 nm, include 501 variables). However, among the 501 variables, there can be a great number of collinear and irrelevant variables that were not releated to BSA concentration. Both these collinear and irrelevant variables were called unwanted information. If PLS model includes too much unwanted information, the performance of PLS model become weaker. In order to eliminate this problem, a second PLS run was performed with the best gene produced by GILS in the 250 runs reported above. Table 5.7and Figure 5.15 show these wavelngths as numbers and as plot for the best gene used in both GILS and PLS. As seen, there were 22 wavelengths selected on the spectra which are created by GILS program.

Table 5.7. The distributions of selected UV-Vis wavelengths by GILS for a single best gene against CBB blank

Order	Wavelength (nm)	Order	Wavelength (nm)
1	360	12	351
2	458	13	316
3	342	14	528
4	563	15	541
5	493	16	522
6	649	17	376
7	448	18	501
8	493	19	601
9	429	20	490
10	795	21	705
11	590	22	705

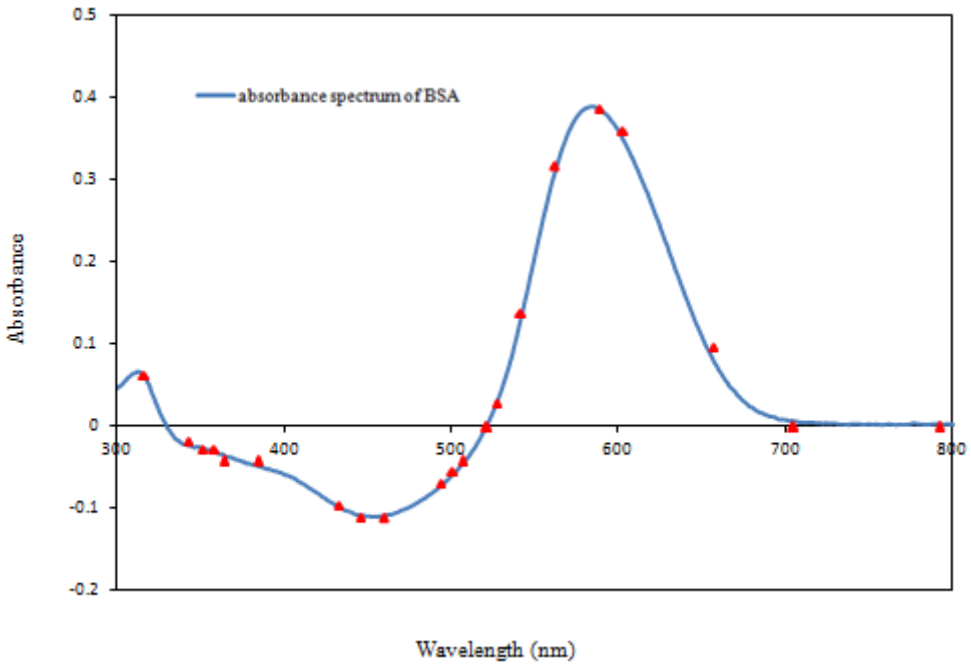


Figure 5.15. The distributions of selected UV-Vis wavelengths for a single best gene that are used in both GILS and PLS on the spectrum of BSA against CBB as blank.

Here, the second PLS modelling was done by using these 22 variables. Actual BSA concentration versus predicted values based on UV-Vis spectra using PLS method with selected wavelength are shown in Figure 5.16. Calibration models for

determination of protein concentration gave the standard error of cross validation (SECV) as 0.63 µg/mL and standard error of prediction (SEP) as 0.44 µg/mL. The R^2 value of regression lines for BSA concentration was 0.9799.

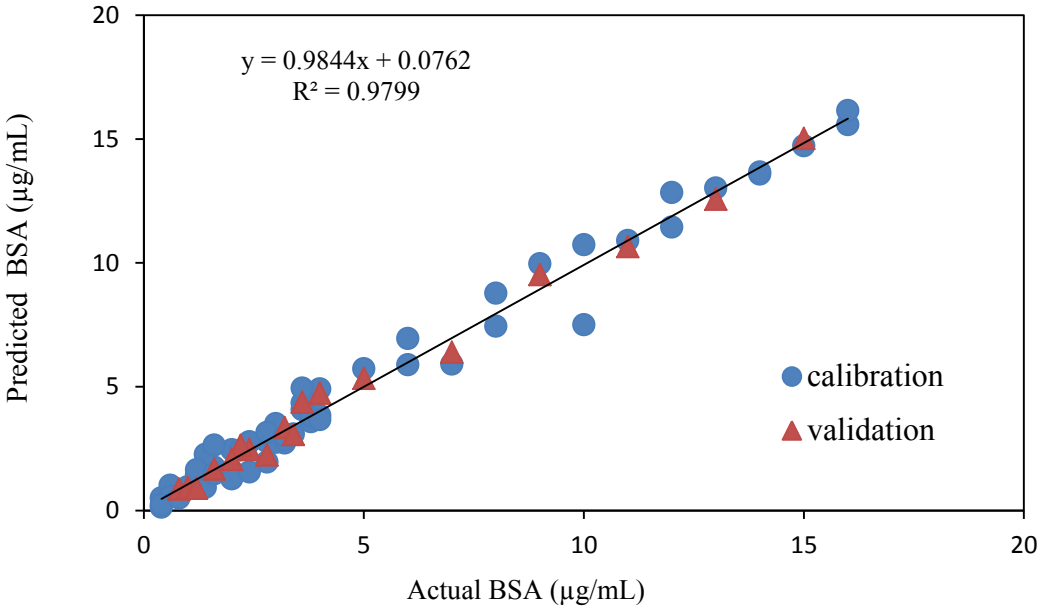


Figure 5.16. Actual versus partial least squares (PLS)-predicted protein concentration with selected wavelength against CBB blank

According to the SECV, SEP and R^2 values, this PLS model could be considered as adequate for the prediction of BSA concentration. The optimal number of PLS factors were found 17 according to full cross-validation procedure.

5.1.1.4. Comparison of GILS and PLS for CBB Blank

According to the Table 5.8 PLS results are suffering from lower R^2 and higher SECV and SEP values for the BSA concentration when the full spectral range was used without any wavelength selection. On the other hand, GILS provided more successful models with a better prediction ability for the concentrations of BSA samples. However, when the PLS model was constructed with use of selected wavelengths that was obtained in GILS, the predeiction ability of PLS was improved significantly and the SEP values for the independent validation set was almost the same with GILS. As a result, it is possible to conclude that both of this methods have an equal prediction ability in terms of prediction of unknown samples. These results demonstrated that

Genetic Algorithms (GA) are effective to solve complex problem such as wavelength selection.

Table 5.8. The SECV, SEP and R^2 results GILS, PLS and PLS* methods for Bradford protein assay against CBB blank

Name of Method	SECV	SEP	R^2	Factors
GILS	0.30	0.45	0.9954	
PLS	0.70	0.98	0.9732	17
PLS*	0.63	0.44	0.9799	17

PLS* Model was costructed with GILS selected wavelength

5.1.1.5. Univariate Calibration Results For Water Blank

Univariate calibration model was composed of 63 BSA standard samples for calibration set shown in Table 5.9 and 19 BSA samples for validation set shown in Table 5.10. As seen from Table 5.9, the number of calibration sample in the calibration set is now 63 which contain 4 addition sample compared to the calibration set that were used in CBB blank cases given above sections. These 4 samples are the ones that do not contain any BSA but just CBB.

Table 5.9. Concentration profile of the calibration samples against water blank

Sample No	Concentration (µg/mL)	Sample No	Concentration (µg/mL)	Sample No	Concentration (µg/mL)
1	0.00	22	2.20	43	3.60
2	0.00	23	2.20	44	4.00
3	0.40	24	3.00	45	4.00
4	0.80	25	3.40	46	5.00
5	0.80	26	3.40	47	6.00
6	1.20	27	3.80	48	6.00
7	1.60	28	0.00	49	7.00
8	1.60	29	0.40	50	8.00
9	2.00	30	0.40	51	8.00
10	2.40	31	0.40	52	9.00

(cont. on next page)

Table 5.9 (cont.)

Sample No	Concentration (µg/mL)	Sample No	Concentration (µg/mL)	Sample No	Concentration (µg/mL)
11	2.40	32	0.80	53	10.00
12	2.80	33	0.80	54	10.00
13	3.20	34	1.20	55	11.00
14	3.20	35	1.60	56	12.00
15	3.60	36	1.60	57	12.00
16	4.00	37	2.00	58	13.00
17	4.00	38	2.40	59	14.00
18	0.00	39	2.40	60	14.00
19	1.00	40	2.80	61	15.00
20	1.00	41	3.20	62	16.00
21	1.40	42	3.20	63	16.00

Table 5.10. Concentration profile of the validation samples against water blank

Sample No	Concentration (µg/mL)	Sample No	Concentration (µg/mL)	Sample No	Concentration (µg/mL)
1	0.40	8	3.00	15	7.00
2	1.20	9	3.80	16	9.00
3	2.00	10	0.40	17	11.00
4	2.80	11	1.20	18	13.00
5	3.60	12	2.00	19	15.00
6	0.60	13	2.80		
7	1.40	14	5.00		

Figure 5.17 shows standard curve of BSA. As can be seen from these figures, outside the narrow range of protein concentration causes linearity problem. Over a broad range of protein concentrations (0.0-16.0 µg/mL) the degree of curvature is quite large; therefore, only a range of relatively high protein concentrations, 0.0–6.0 mg/mL BSA, can be used for assay and construction of the calibration graph. However, this solution for nonlinearity problem makes the minimum detected protein quantity lower. Besides, there is no notably improvemet in the calibratin quality.

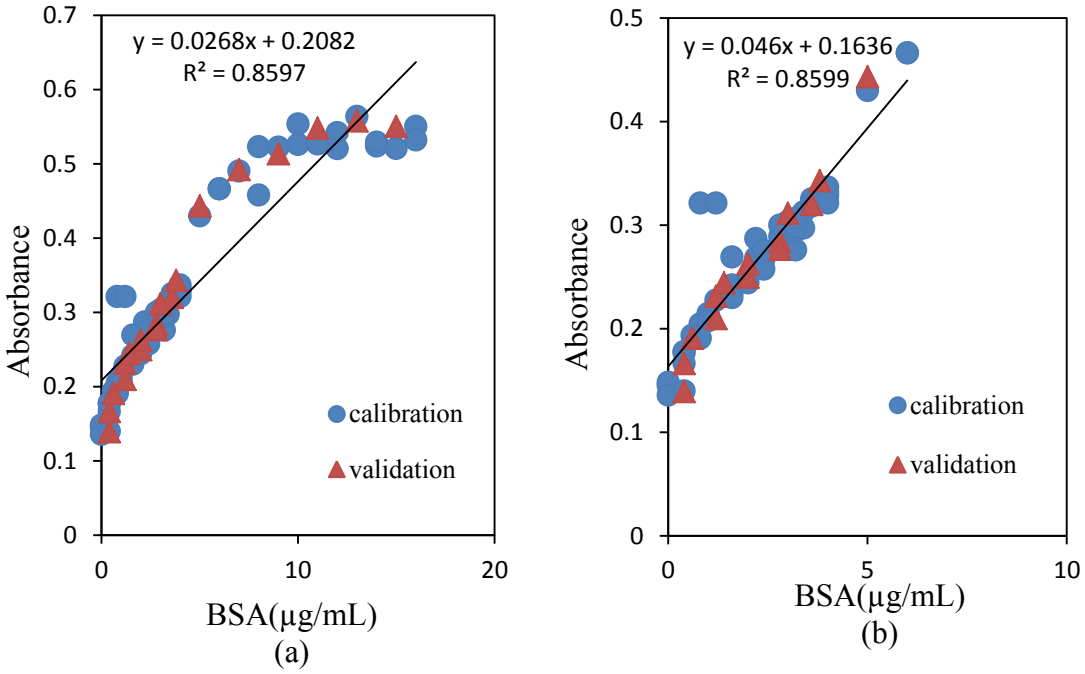


Figure 5.17. Calibration graphs of Bradford protein assay at 595 nm against water blank a) concentration range between 0-16 µg/mL BSA and b) concentration range between 0-6 µg/mL BSA

Univariate calibration method is not capable of determining BSA concentration with an accurate result. For this reason, multivariate calibration methods are employed in order to increase calibration quality of this assay.

5.1.1.6. GILS Results For Water Blank

The calibration set were generated from 41 samples with duplicate measurements. Here, 63 of them were randomly selected with the samples having minimum and maximum BSA concentration and these samples were assigned as calibration set shown in Table 5.11. The remaining 19 samples were reserved for independent test samples shown in Table 5.12.

Table 5.11. Actual versus genetic inverse least square (GILS) predicted protein concentration for calibration samples against water blank

Sample No	Concentration (µg/mL)	Predicted Concentration(µg/mL)	Sample No	Concentration (µg/mL)	Predicted Concentration(µg/mL)
1	0.00	-0.19	33	0.80	0.67
2	0.00	-0.02	34	1.20	1.31
3	0.40	0.47	35	1.60	1.68
4	0.80	0.67	36	1.60	1.73
5	0.80	0.89	37	2.00	1.95
6	1.20	1.32	38	2.40	2.42
7	1.60	1.33	39	2.40	1.91
8	1.60	1.64	40	2.80	2.60
9	2.00	1.89	41	3.20	3.35
10	2.40	2.59	42	3.20	2.69
11	2.40	2.09	43	3.60	3.58
12	2.80	3.07	44	4.00	3.59
13	3.20	3.29	45	4.00	3.47
14	3.20	3.01	46	5.00	5.00
15	3.60	3.45	47	6.00	6.16
16	4.00	4.30	48	6.00	6.16
17	4.00	3.48	49	7.00	6.68
18	0.00	0.73	50	8.00	8.17
19	1.00	1.20	51	8.00	8.00
20	1.00	0.86	52	9.00	9.54
21	1.40	1.31	53	10.00	9.71
22	2.20	2.09	54	10.00	10.44
23	2.20	1.94	55	11.00	11.11
24	3.00	2.75	56	12.00	12.13
25	3.40	3.38	57	12.00	12.34
26	3.40	3.46	58	13.00	12.75
27	3.80	3.75	59	14.00	13.78
28	0.00	-0.02	60	14.00	13.94
29	0.40	0.60	61	15.00	14.56
30	0.40	0.70	62	16.00	15.96
31	0.40	0.08	63	16.00	15.58
32	0.80	1.04			

Table 5.12. Actual versus genetic inverse least square (GILS)predicted protein concentration for validation samples against water blank

Sample No	Concentration (µg/mL)	Predicted Concentration(µg/mL)	Sample No	Concentration (µg/mL)	Predicted Concentration(µg/mL)
1	0.40	0.11	11	1.20	1.74
2	1.20	1.32	12	2.00	2.01
3	2.00	1.81	13	2.80	2.29
4	2.80	2.65	14	5.00	5.90
5	3.60	3.52	15	7.00	7.26
6	0.60	0.64	16	9.00	9.58
7	1.40	1.39	17	11.00	10.90
8	3.00	2.73	18	13.00	12.88
9	3.80	4.04	19	15.00	14.95
10	0.40	-0.09			

Actual BSA concentration versus predicted values based on UV-VIS spectra using GILS method are shown in Figure 5.18. Calibration models for protein concentration determination gave standard error of cross validation (SECV) and standard error of prediction (SEP) values as 0.35 µg/mL and 0.43 µg/mL for calibration and independent test sets, respectively. The R^2 value of regression lines for BSA concentration was 0.9972.

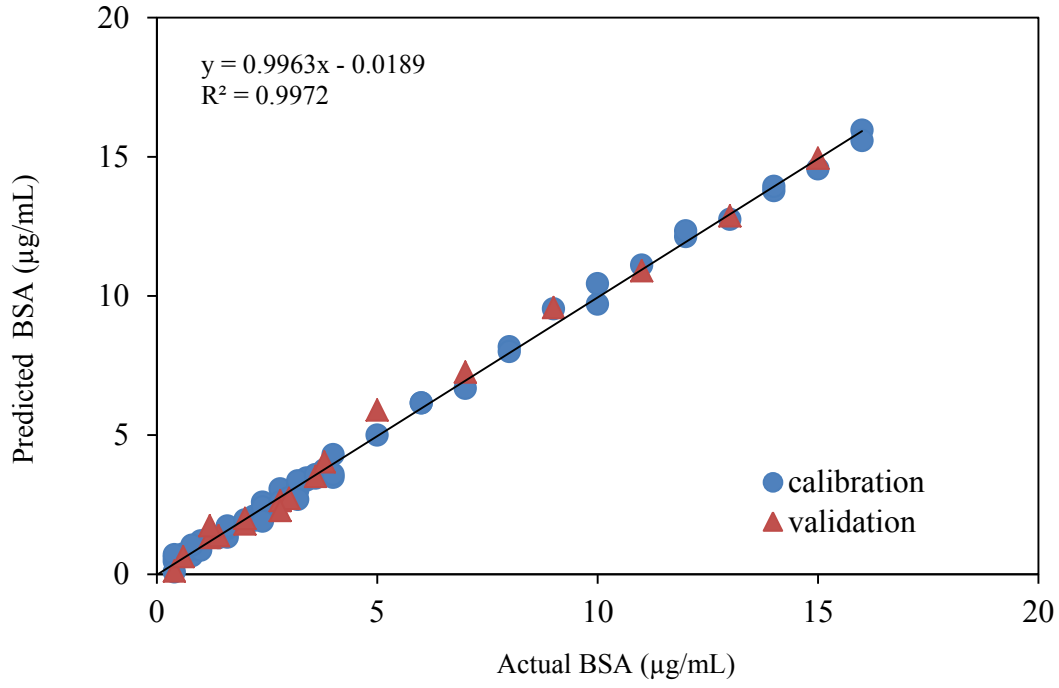


Figure 5.18. Actual versus genetic inverse least squares (GILS)-predicted protein concentration against water blank

When these SECV and SEP values are examined, it is seen that these values are agreeable with each other, which represent a good prediction for protein concentration determination. When the overall calibration performance of the models is examined, it is possible to state that GILS provides better solution for the nonlinearity problem of Bradford protein assay. Dynamic range of this assay is about twice of the univariate values. Thus, accuracy and sensitivity of this assay were improved.

GILS is a wavelength selection based method, it is possible to observe the distribution of selected wavelengths in multiple runs over the entire full spectral region. Figure 5.19 illustrates the frequency distribution of selected wavelengths in 250 runs with 30 genes and 100 iterations for BSA concentration.

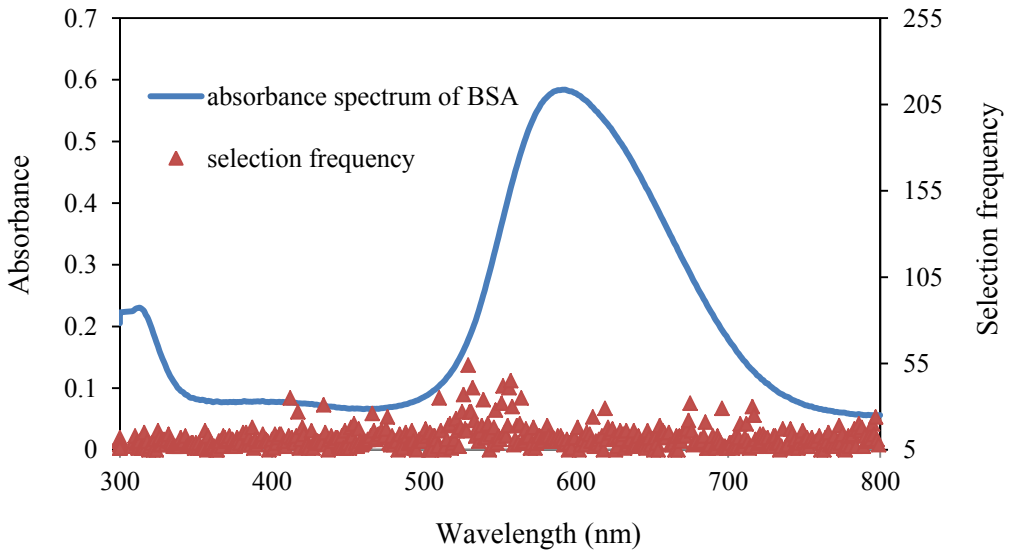


Figure 5.19. Frequency distribution of GILS selected UV-Vis wavelengths for BSA concentration against water blank

As can be seen from Figure 5.6 there are a number of regions where selection frequencies are very high compared to the rest of the spectrum. The wavelength region around 550 and 620 nm for BSA-CBB complex indicates a strong tendency for GILS method. This result arise from the GILS algorithm which focus on wavelengths where the most concentration related information is contained.

5.1.1.7. PLS Results For Water Blank

The calibration set were generated from 78 samples which 59 of them were randomly selected with the samples having minimum and maximum BSA concentration and these samples were assigned as calibration set which is shown in Table 5.13. The remaining 19 samples were reserved for external validation samples that are shown in Table 5.14.

Table 5.13. Actual versus partial least squares (PLS)-predicted protein concentration for calibration samples against water blank

Sample No	Concentration (µg/mL)	Predicted Concentration(µg/mL)	Sample No	Concentration (µg/mL)	Predicted Concentration(µg/mL)
1	0.00	-1.00	33	0.80	3.18
2	0.00	0.55	34	1.20	3.09
3	0.40	-0.44	35	1.60	1.68
4	0.80	0.39	36	1.60	2.68
5	0.80	0.36	37	2.00	2.82
6	1.20	2.00	38	2.40	1.30
7	1.60	1.56	39	2.40	2.35
8	1.60	1.83	40	2.80	2.75
9	2.00	1.95	41	3.20	3.64
10	2.40	2.10	42	3.20	2.53
11	2.40	2.28	43	3.60	3.31
12	2.80	3.29	44	4.00	3.59
13	3.20	3.49	45	4.00	2.38
14	3.20	2.12	46	5.00	6.05
15	3.60	3.38	47	6.00	6.05
16	4.00	4.91	48	6.00	6.05
17	4.00	2.83	49	7.00	6.27
18	0.00	0.69	50	8.00	8.90
19	1.00	1.67	51	8.00	7.54
20	1.00	0.86	52	9.00	10.18
21	1.40	1.44	53	10.00	5.69
22	2.20	2.56	54	10.00	11.73
23	2.20	1.82	55	11.00	13.50
24	3.00	2.09	56	12.00	12.00

(cont. on next page)

Table 5.13 (cont.)

Sample No	Concentration (µg/mL)	Predicted Concentration(µg/mL)	Sample No	Concentration (µg/mL)	Predicted Concentration(µg/mL)
25	3.40	2.33	57	12.00	12.27
26	3.40	3.18	58	13.00	11.52
27	3.80	4.46	59	14.00	13.21
28	0.00	-0.80	60	14.00	13.52
29	0.40	0.63	61	15.00	14.80
30	0.40	0.55	62	16.00	15.07
31	0.40	-0.71	63	16.00	14.25
32	0.80	2.03			

Table 5.14. Actual versus partial least squares (PLS)-predicted protein concentration for validation samples against water blank

Sample No	Concentration (µg/mL)	Predicted Concentration(µg/mL)	Sample No	Concentration (µg/mL)	Predicted Concentration(µg/mL)
1	0.40	-0.67	11	1.20	1.28
2	1.20	1.38	12	2.00	3.22
3	2.00	1.74	13	2.80	2.33
4	2.80	2.54	14	5.00	7.41
5	3.60	2.72	15	7.00	7.70
6	0.60	0.61	16	9.00	11.21
7	1.40	1.03	17	11.00	11.39
8	3.00	1.90	18	13.00	14.44
9	3.80	5.20	19	15.00	13.52
10	0.40	-0.40			

Actual BSA concentration versus predicted values based on UV-VIS spectra using PLS method are shown in Figure 5.20. Calibration models for protein concentration determination gave the standard error of cross validation (SECV) was found 1.04 µg/mL between and standard error of prediction (SEP) was found 1.11 µg/mL. The R^2 value of regression lines for BSA concentration was 0.9484.

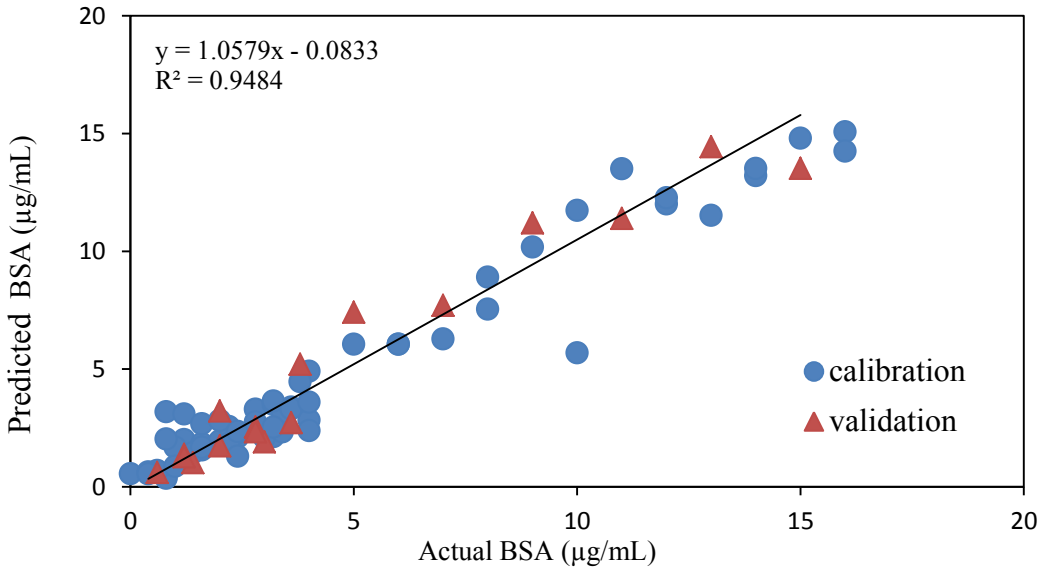


Figure 5.20. Actual versus partial least squares (PLS)-predicted protein concentration against water blank

When the SECV, SEP and R^2 value of regression line are examined it is possible state that PLS are able to predict BSA concentration with an accurate results. By using PLS, linearity of Bradford protein assay is increased with a high sensitivity and accuracy. Dynamic range of this assay is twice about the univariate calibration model. Cross-validation was performed by leaving out one sample at a time to determine the optimal number of PLS components for obtaining a model with good predictive power. The optimal number of PLS factors were found 8 according to full cross-validation procedure.

PLS is full-spectrum method so that model involves all variables. Therefore, collinear and irrelevant variables can be exist in model which causes to make the performance of PLS model weaker. In order to avoid this problem, second PLS was constructed by using 23 best gene which were obtained by GILS. Table 5.15 and Figure 5.21 show these wavelengths as numbers and as plot for the best gene used in both GILS and PLS. These 23 wavelengths are the most sensitive spectral variables in 250 runs.

Table 5.15. The distributions of selected UV-Vis wavelengths by GILS for a single best gene against water blank

4

Order	Wavelength (nm)	Order	Wavelength (nm)
1	555	13	708
2	517	14	501
3	758	15	566
4	701	16	779
5	414	17	598
6	397	18	343
7	791	19	308
8	377	20	686
9	549	21	454
10	797	22	631
11	797	23	334
12	343		

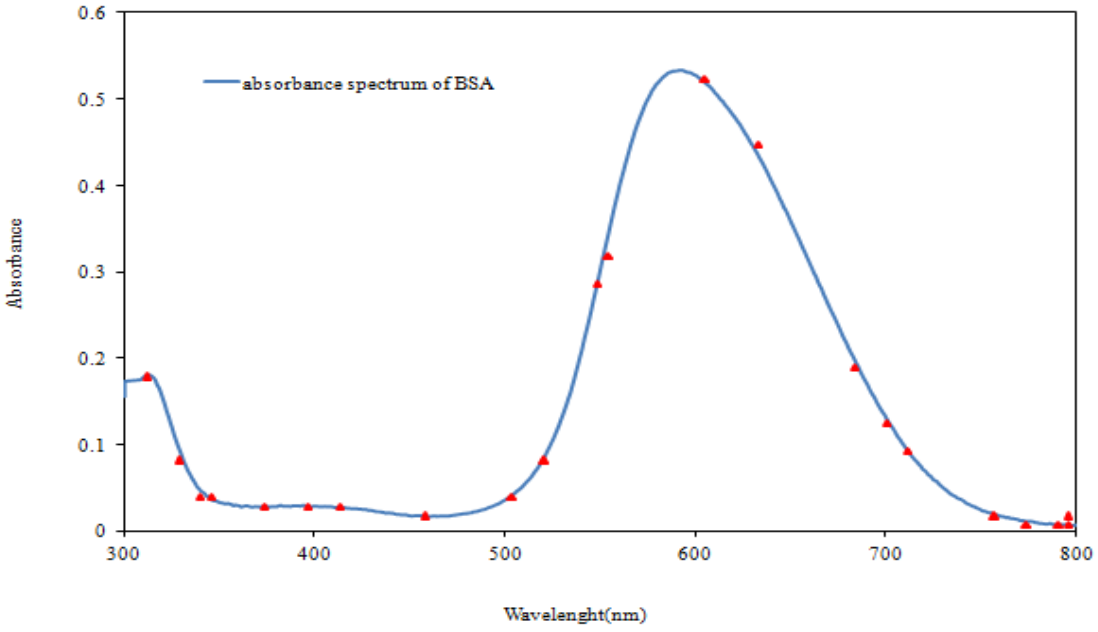


Figure 5.21. The distributions of selected UV-Vis wavelengths by GILS for a single best gene on the spectrum against water blank

Then PLS are modeled by using these 23 spectral variables. Actual BSA concentration versus predicted values based on UV-Vis spectra using PLS method are shown in Figure 5.22. Calibration models for protein concentration determination gave the standard error of cross validation (SECV) was found 0.68 µg/mL and standard error of prediction (SEP) was found 0.42 µg/mL. The R^2 value of regression lines for BSA concentration was 0.9768.

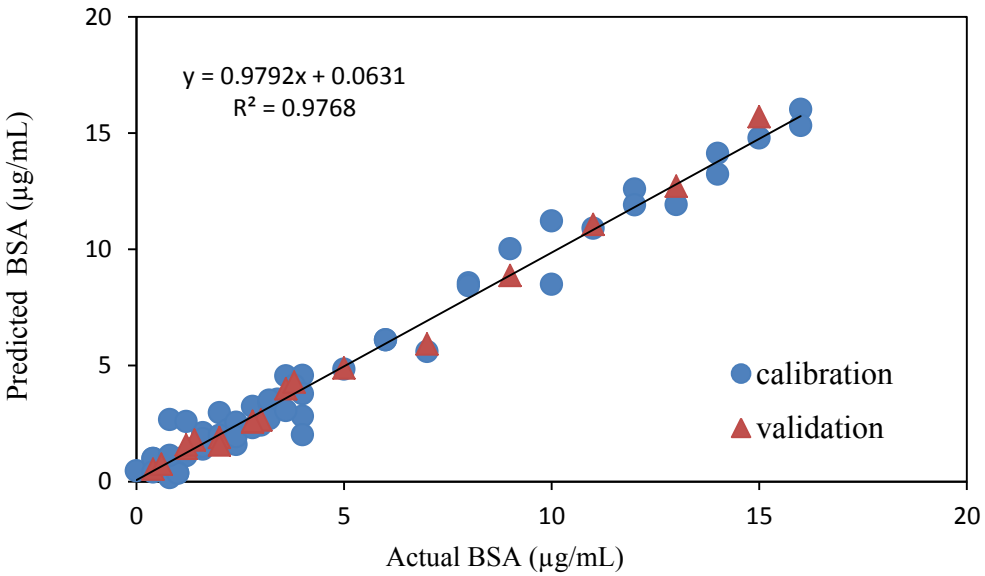


Figure 5.22. Actual versus partial least squares (PLS)-predicted protein concentration with selected wavelenght against water blank

Model indicates that the optimized PLS calibration is capable of predicting the BSA concentration. Also, it can be seen that selected data was sufficient to ensure accurate results with high correlation coefficient, R^2. The optimal number of PLS factors were found 20 according to full cross-validation procedure.

5.1.1.8. Comparison of GILS and PLS for Water Blank

Table 5.16 summarizes the standard error of cross-validation, standard error of prediction and R^2 results obtained with GILS and PLS. As can be seen, GILS model outperform PLS in terms of both standard error of cross-validation and standard error of prediction (smaller SECV and SEP and larger R^2). GILS is more robust with respect to differences between the calibration set and prediction set. It can be concluded that GILS

is able to predict BSA concentration. However, when the PLS model was constructed by using the most sensitive spectral variables that was obtained in GILS, both of this methods have similar performance for the determination of BSA concentration according to the standard error of prediction values.

Table 5.16. The SECV, SEP and R^2 results GILS, PLS and PLS* methods for Bradford protein assay against water blank

Name of Method	SECV	SEP	R^2	Factors
GILS	0.35	0.43	0.9972	
PLS	1.04	1.04	0.9484	8
PLS*	0.68	0.42	0.9768	20

PLS* Model was costructed with GILS selected wavelength

According the multivariate calibration results with CBB and water blank, GILS results with CBB blank was chosen for the immobilization analysis.

5.2. Central Composite Design

Optimization of enzyme immobilization is an important process in order to increase activity and stability of immobilized enzymes. The classical method of finding out optimum conditions by varying one independent variable while keeping the other variables constant at a specified levels has some drawbacks such as requirement more runs which means in industry higher time consumption and having an unfavourable impact on the economy, ignoring to estimate of interactions and probability of optimum values missing. (Nasirizadeh, Dehghanizadeh et al., 2012). Central composite design (CCD) which is a very useful method to reduce the number of experimental run when optimizing the effective parameters in a process. This method provides better results for obtaining the effect of interactions among the parameters that have been optimized and also CCD is suitable for fitting a quadratic surface model (Nasirizadeh, Dehghanizadeh et al., 2012). For this reason, a CCD model was used to optimize of immobilization parameters of BSA onto chitosan nanoparticles. The statistical combination of the independent variables in actual and coded values along with the experimental an predicted responses are shown in Table 5.17

Table 5.17.The statistical combination of the independent variables in coded values along with the predicted and experimental response

Experiment	immobilization time (minute) (X_1)	temperature (°C)(X_2)	pH (X_3)	chitosan concentration (X_4)	X_1	X_2	X_3	X_4	Yield (%)	Predicted Y (%)
1	49.0	26.0	7.0	0.30	-1	-1	-1	-1	24.75	22.67
2	49.0	26.0	7.0	0.71	-1	-1	-1	1	30.66	37.36
3	49.0	26.0	9.0	0.30	-1	-1	1	-1	18.26	21.65
4	49.0	26.0	9.0	0.71	-1	-1	1	1	8.35	6.27
5	49.0	49.0	7.0	0.30	-1	1	-1	-1	35.48	30.63
6	49.0	49.0	7.0	0.71	-1	1	-1	1	22.23	20.15
7	49.0	49.0	9.0	0.30	-1	1	1	-1	52.50	50.42
8	49.0	49.0	9.0	0.71	-1	1	1	1	3.16	9.86
9	136.0	26.0	7.0	0.30	1	-1	-1	-1	44.83	44.83
10	136.0	26.0	7.0	0.71	1	-1	-1	1	37.40	37.40
11	136.0	26.0	9.0	0.30	1	-1	1	-1	58.60	58.60
12	136.0	26.0	9.0	0.71	1	-1	1	1	9.56	21.10
13	136.0	49.0	7.0	0.30	1	1	-1	-1	53.51	53.51
14	136.0	49.0	7.0	0.71	1	1	-1	1	17.61	20.92
15	136.0	49.0	9.0	0.30	1	1	1	-1	88.08	88.08
16	136.0	49.0	9.0	0.71	1	1	1	1	25.41	25.41
17	5.0	37.5	8.0	0.51	-2	0	0	0	3.54	4.04
18	180.0	37.5	8.0	0.51	2	0	0	0	46.86	41.74
19	92.5	15.0	8.0	0.51	0	-2	0	0	27.41	20.99
20	92.5	60.0	8.0	0.51	0	2	0	0	31.46	33.26
21	92.5	37.5	5.0	0.51	0	0	-2	0	45.77	47.58
22	92.5	37.5	11.0	0.51	0	0	2	0	57.48	51.06
23	92.5	37.5	8.0	0.01	0	0	0	-2	56.77	61.88
24	92.5	37.5	8.0	1.00	0	0	0	2	23.63	13.90
25	92.5	37.5	8.0	0.51	0	0	0	0	22.44	21.11
26	92.5	37.5	8.0	0.51	0	0	0	0	13.40	21.11
27	92.5	37.5	8.0	0.51	0	0	0	0	20.73	21.11
28	92.5	37.5	8.0	0.51	0	0	0	0	22.65	21.11
29	92.5	37.5	8.0	0.51	0	0	0	0	25.51	21.11
30	92.5	37.5	8.0	0.51	0	0	0	0	21.95	21.11

Regression analysis was used to calculate the effect of each factor and their interactions. The model expressed by Equation (5.1) represents % immobilization yield(y) as a function of immobilization time (X_1), temperature (X_2), pH (X_3) and chitosan concentration (X_4).

$$\mathbf{y} = 21.11 + 9.43x_1 + 3.07x_2 + 0.87x_3 - 12.00x_4 + 0.44x_1^2 + 1.50x_2^2 + 7.05x_3^2 + 4.19x_4^2 + 0.18x_1x_2 + 3.70x_1x_3 - 5.53x_1x_4 + 5.20x_2x_3 - 6.29x_2x_4 - 7.52x_3x_4 \quad (5.1)$$

The statistical significance of Equation (5.1) was controlled by the analysis of variance (ANOVA) for quadratic model given in Table 5.18. The model highly significant, as is evident from the model F-values and a very low p-values (<0.0001). The coefficient of determination (R^2) was also shown in Table 5.18. This value indicates that the accuracy of the model is adequate. The lack of fit measures the failure of the model to represent data in the experimental domain at points which are not included in the regression. The F-value of lack-of-fit which is 3.12 for regression of Equation (5.1) is not significant. Non-significant lack of fits is good and indicates that the model equation was adequate for predicting the % yield of BSA immobilization under any combination of values of the variables.

Table 5.18. Analysis of variance (ANOVA) for the fitted quadratic polynomial model for optimization of immobilization parameters.

Source	Sum of squares	Degree of Freedom	Mean squares	F-value	p-value
Model	10184.50	14	727.46	18.00	0
Linear	5830.30	4	1457.58	36.06	0
Square	1675.30	4	418.82	10.36	0
Interaction	2678.90	6	446.48	11.05	0
Residual Error	606.30	15	40.42		
Lack-of-fit	522.60	10	52.26	3.12	0.111
Pure Error	83.70	5	16.75		
Total	10790.80	29			
R^2= 0.9438; Pred R^2= 0.7099 Adj R^2=0.8914					

The P-values mark the significance of coefficients and are also important for understanding the pattern of the mutual interactions between the parameters. A value of P-value less than 0.05 indicates that the model terms are significant. The responses

taken from Table 5.19 reveal that immobilization time (X_1), chitosan concentration (X_4), square of pH (X_3^2) and binary interaction of temperature and chitosan concentration (X_2X_4) are the most significant terms in the full quadratic model equation. These values suggest that immobilization time and chitosan concentration have a direct relationship with yield of immobilization. Any changes in these two factors affect BSA immobilization yiels, considerably. However, the terms X_3, X_4^2, X_1X_3,X_1X_4, X_2X_3 and X_3X_4 have less effect on the yield of BSA immobilizaton process. These results also show interactions between immobilization time (X_1), temperature (X_2), pH (X_3) and chitosan concentration (X_4) which must be taken into the account due to effects on the immobilization process. However, these effects are ignored by classical optimization process since it is not possible to evaluate the interaction effects between paramaters in classical one at a time aproach.

Table 5.19. The least-squares fit and statistical significance of regression coefficient for the estimated parameters.

	Coefficients	Standard Error	t Stat	P-value
Intercept	21.11	2.60	8.13	0.0000
X_1	9.43	1.30	7.26	0.0001
X_2	3.07	1.30	2.36	0.0320
X_3	0.87	1.30	0.67	0.5128
X_4	-12.00	1.30	-9.24	0.0001
X_1^2	0.44	1.21	0.37	0.7197
X_2^2	1.50	1.21	1.24	0.2349
X_3^2	7.05	1.21	5.81	0.0001
X_4^2	4.19	1.21	3.45	0.0035
X_1X_2	0.18	1.59	0.11	0.9113
X_1X_3	3.70	1.59	2.33	0.0345
X_1X_4	-5.53	1.59	-3.48	0.0034
X_2X_3	5.20	1.59	3.27	0.0052
X_2X_4	-6.29	1.59	-3.96	0.0013
X_3X_4	-7.52	1.59	-4.73	0.0003

The regression coefficient values were evaluated and the subsequent refined equation, including only the significant terms, were derived using the coefficients of the coded variables for BSA immobilization yield which is given in equation 5.2:

$$\mathbf{y} = 21.11 + 9.43x_1 + 3.07x_2 - 12.00x_4 + 7.05x_3^2 + 4.19x_4^2 + 3.70x_1x_3$$

$$- 5.53x_1x_4 + 5.20x_2x_3 - 6.29x_2x_4 - 7.52x_3x_4 \quad (5.2)$$

Predicted values are calculated by regression analysis. The relationship between predicted and experimental immobilization yield is shown in Figure 5.23. As can be seen, the predicted values of the response from the model are in well agreement with the observed experimental values.

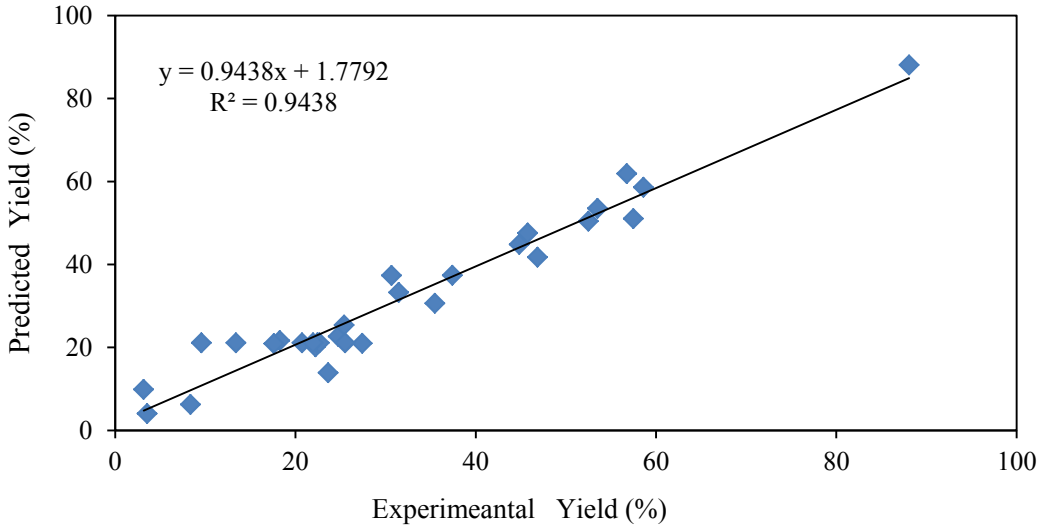


Figure 5.23. Predicted yield versus experimental immobilization yield

It is important to control the fitted model in order to assure that it provides an adequate approximation to the real system. The residuals from the least squares fit have a critical role in controlling model adequacy. Normality assumption was checked by constructing a normal probability plot of the residuals. Figure 5.24 represents approximately linear pattern for the probability, which shows that the residuals are normally distributed.

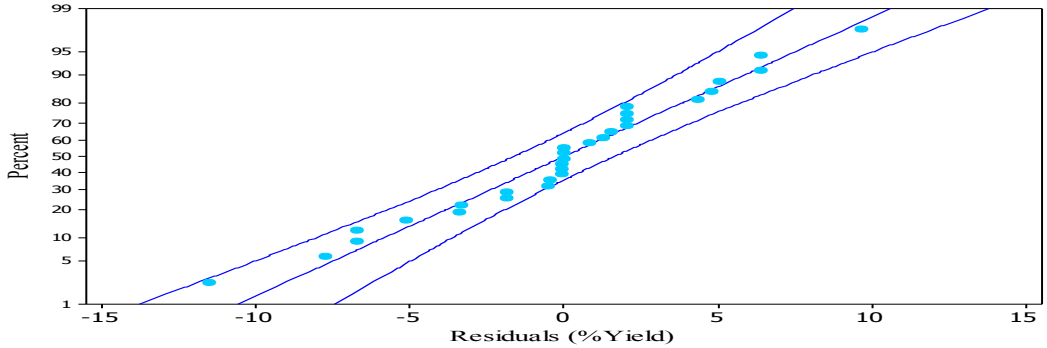


Figure 5.24. Normal probability of residual

Figure 5.25 represents a plot of residuals versus the predicted response. The residual plots were scattered randomly, indicating the variance of the original observation is constant for all values of Y. Considering both of these plots, it was concluded that the proposed full quadratic model is adequate to describe the BSA immobilization yield.

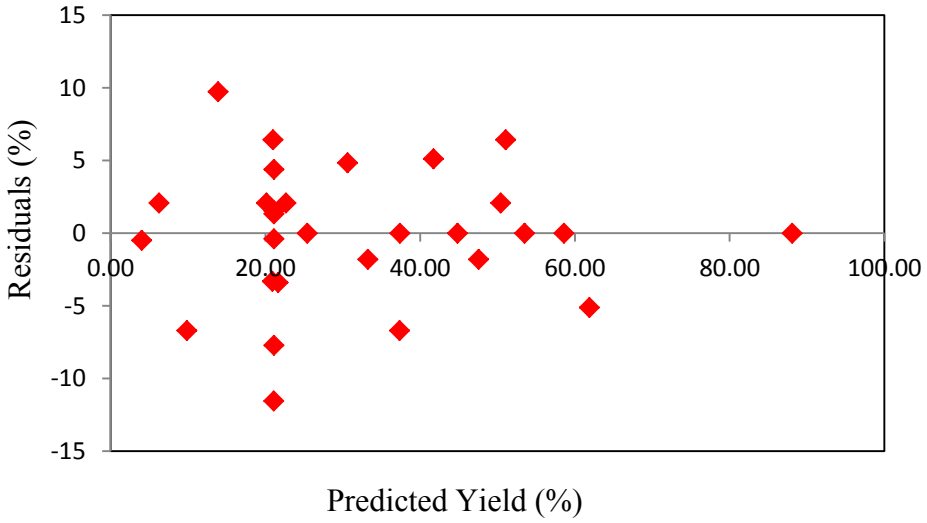


Figure 5.25. Plot of the residuals versus the predicted response

In order to achive highest possible immobilization yield, the Solver tool of Microsoft Excel (MS Office 2007, Microsoft Corporation) program was used to optimize the regression equation for optimum values of the four factor studied. The optimal values of the immobilization time (X_1), temperature (X_2), pH (X_3) and chitosan concentration (X_4) were determined in coded units as shown belove for each factor:

X_1= 1.45 X_2=0.47 X_3=0.30 X_4= -1.93

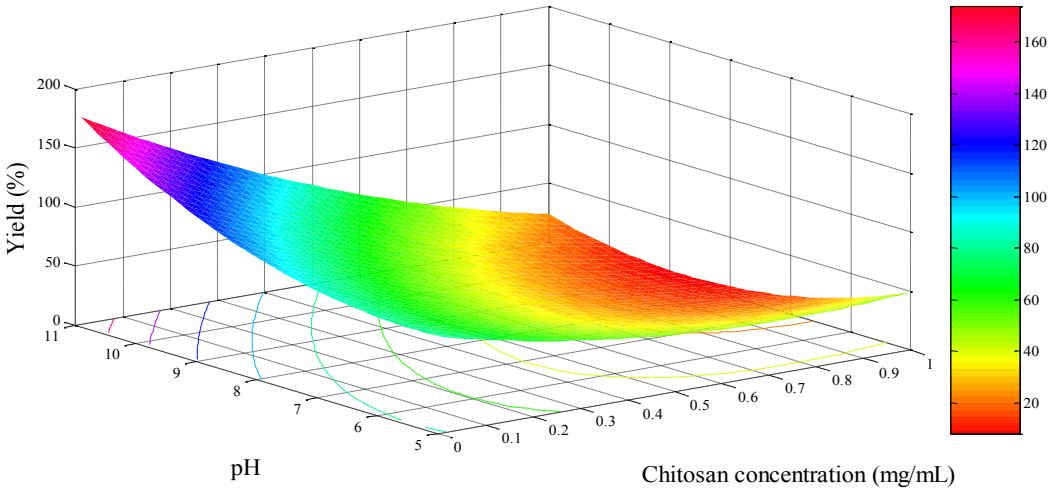
with a corresponding 99.9% immobilization yield. The actual values obtained by putting the respective values of each factor in the following equation;

$$a_i = \frac{X_i - X_0}{\Delta X_i} \qquad (5.3)$$

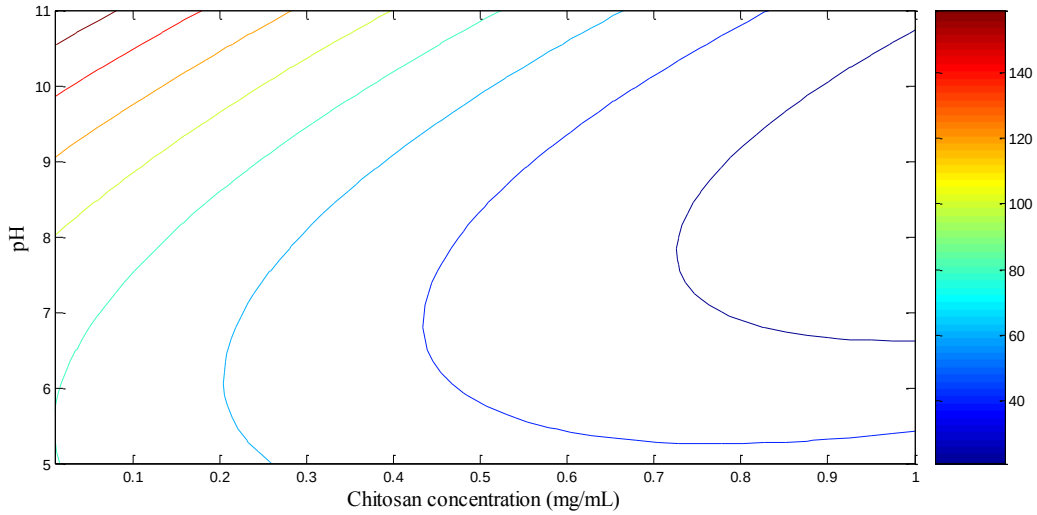
wher a_i is the actual value of the i^{th} factor, X_i is the coded value and ΔX_i is the diffrence between the highest and the lowest coded values. By using this equation, optimum conditions of BSA immobilization onto chitosan nanoparticules were found as: immobilization time 154 minutes, temperature 43°C, pH 8.45 and chitosan concentration 0.0348 mg/mL.

The 3-D response surface is used to determine the potential relationship between three variables. 3-D surface plots display the three-dimensional relationship in two dimensions, with predictor variables on the x- and y-scales, and the response (z) variable represented by a smooth surface (surface plot). And also, 3-D response surface plots are the graphical representations of the regression equation. To evaluate the effects of different process variables on BSA immobilization yield, graphical representations were made in Figure 5.26- 5.31 which demonstrate three dimensional model surface and contour plot.

Figure 5.26 depicts the 3D and 2D plots showing the effects of pH (X_3) and chitosan concentration (X_4) on BSA immobilization yield while keeping immobilization (X_1) time and temperature (X_2) at the central level (154 min) and (43°C), respectivly. The BSA immobilization yield increased slightly with the increase of pH at a low level of chitosan concentration. Relatively lower BSA immobilization yield were obtained at a lower pH value. The adsorption process seemed to be affected due to charge interactions. When pH value increased cationic value of BSA was not existed anymore, as well anionic value of BSA going strong. So repellent force between chitosan nanoparticles and BSA disappeared, on the contrary, a great interaction appeared (Li et al., 2011). The results show that under the experimental conditions examined, chitosan concentration has a greater effect on BSA immobilization yield than pH, especially at a low chitosan concentration level. Chitosan concentration has negative effect on BSA immobilization yield. Increasing the chitosan concentration decreased BSA immobilization yield since highly viscous nature of the gelation medium hinders immobilization of BSA. Relatively lower adhesiveness of chitosan with lower concentration promotes immobilization of BSA. Also, it is seen from Equation (5.1) that the signs in front of the coefficients pH and chitoan concentration are plus and minus, respectively, while the sign in front of the cofficient of the X_3*X_4 interaction is minus. This clearly indicates that the chitosan concentration has a dominant effect over the pH.



(a)

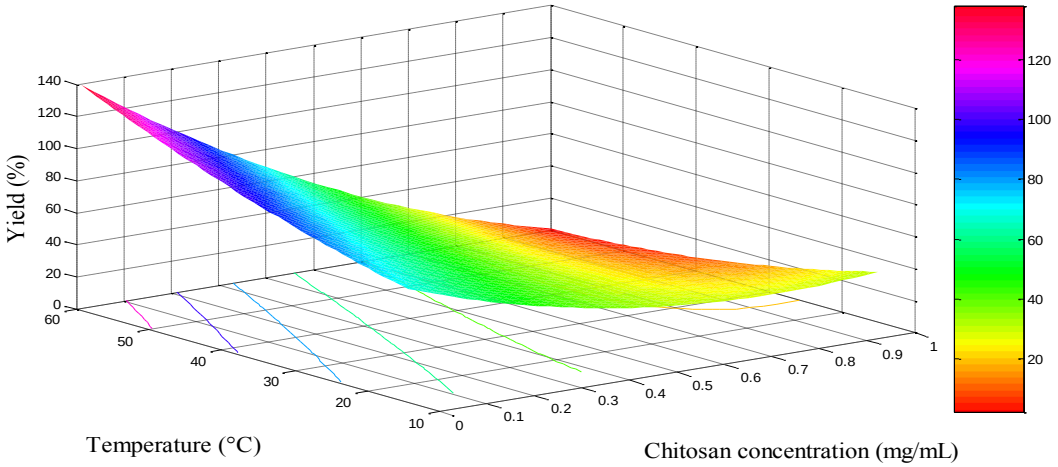


(b)

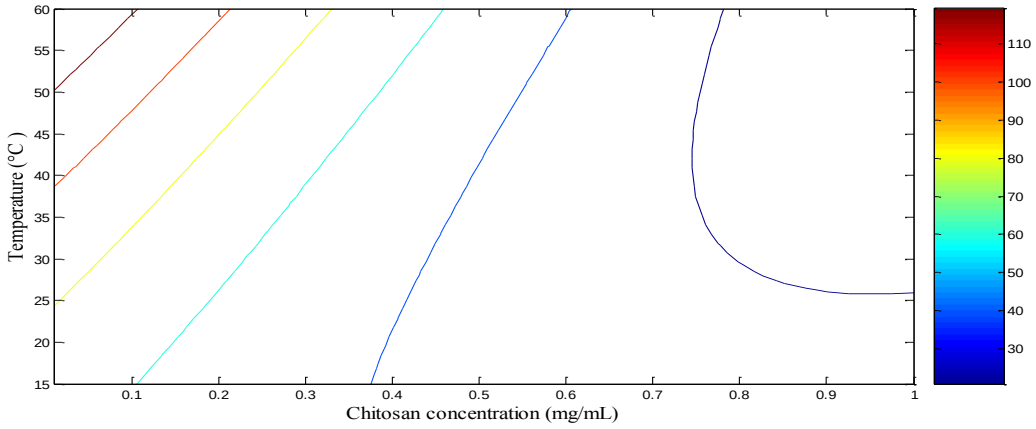
Figure 5.26. Response surface plot (a) and contour plot (b) of the showing the effect of pH and chitosan concentration on the BSA immobilization yield at a fixed temperature 43°C of and immobilization time 154 minute

The effects of temperature (X_2) and chitosan concentration (X_4) on BSA immobilization are shown in Figure 5.27 while keeping pH (X_3) and immobilization time (X_1) are at the middle point 154 minutes and 8.45, respectiv ely. As indicated in these figure by increasing tempearture at a costant chitosan concentration, the BSA immobilization is remarkably enhanced. This suggests that BSA has a structure that is much easier to make an interaction with chitosan nanoparticles at higher temperatures. This may be due to the enzyme having either a more flexible structure or a big number of potential binding sites on its surface, making it more likely to spread on the

nanoparticle surface. On the other hand, according this figure and equation (5.1) chitosan concentration has negative effect and decreased BSA immobilization yield.



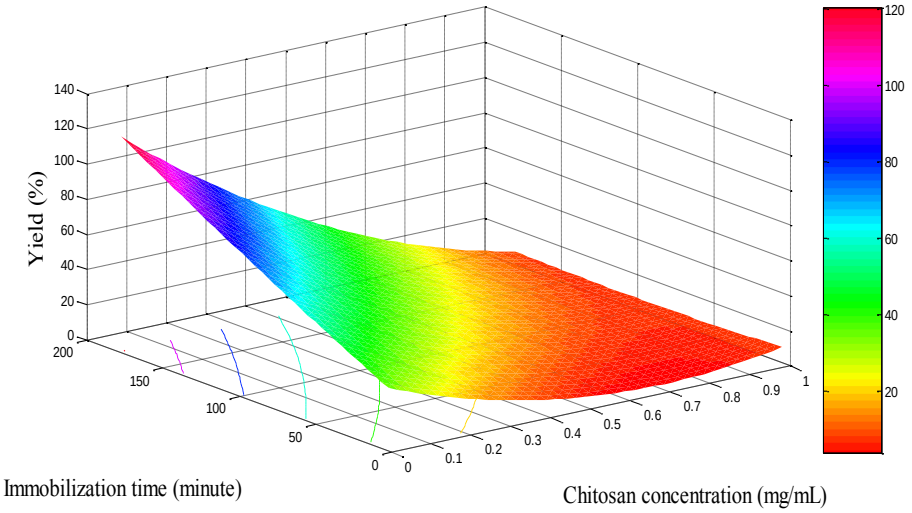
(a)



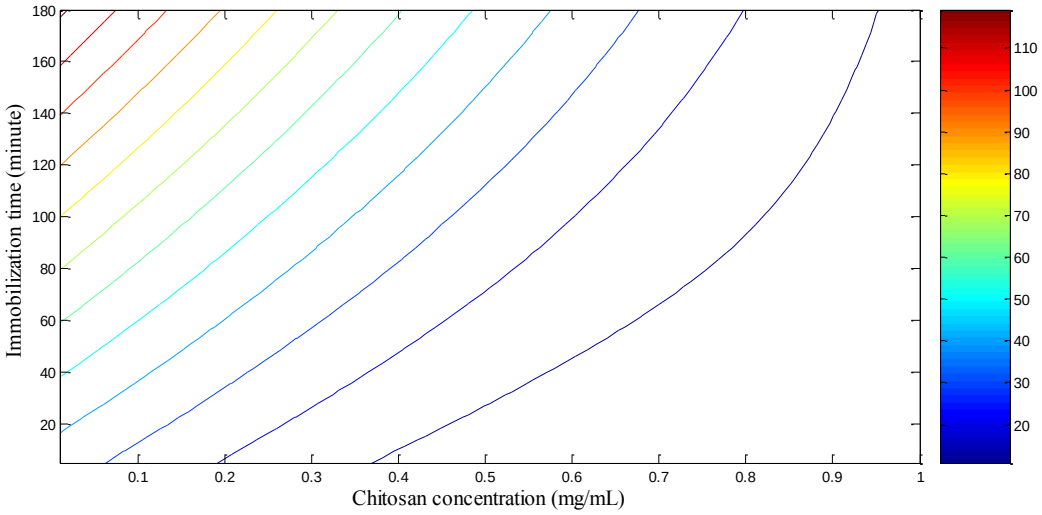
(b)

Figure 5.27. Response surface (a) and contour plot (b) showing the effect of temparature and chitosan concentration on the BSA immobilization yield at a fixed pH of 8.45 and immobilization time 154 minute

Figure 5.28 shows interaction between immobilization time (X_1) and chitosan concentration(X_4) on immobilization yield while temperature (X_2)and pH (X_3) keep costant at value of 43°C and 8.45. The maximum yield of BSA immoblization was observed with low chitosan concentration. In contrast to the low chitosan concentration, immobilization time, which was necessary for the maximum yield was obtained high immobilization time. This result indicated that the immobilization procedure was not quick between BSA and chitosan nanoparticles because of their smaller specific surface area of contact.



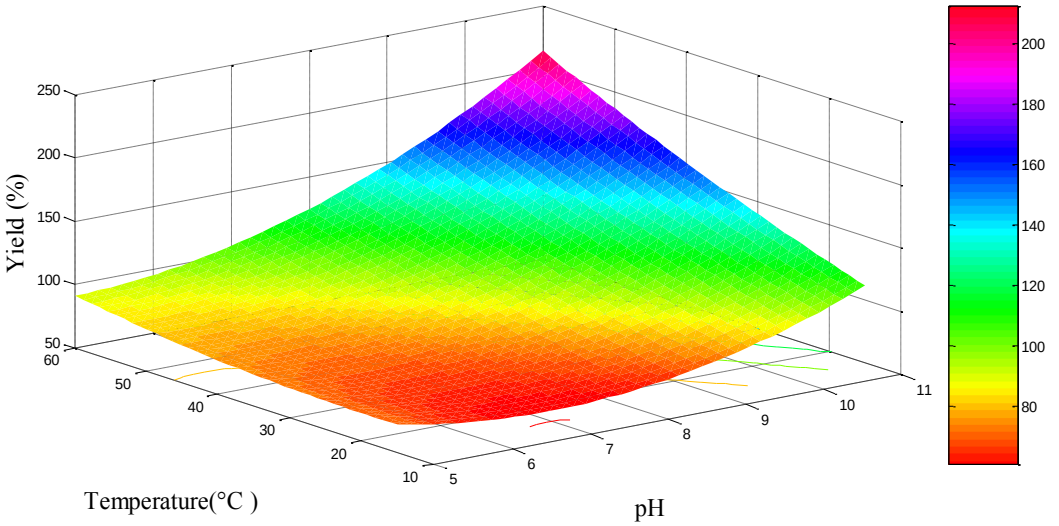
(a)



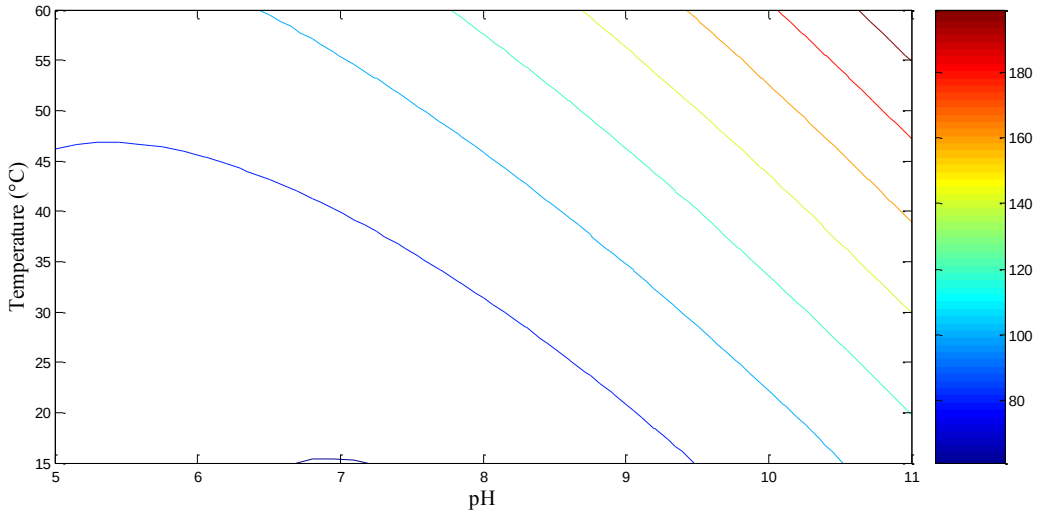
(b)

Figure 5.28. Response surface (a) and contour plot (b) showing the effect of immobilization time and chitosan concentration on the BSA immobilization yield at a fixed pH of 8.45 and temperature 43°C

The combined effects of temperature (X_2) and pH (X_3) on BSA immobilization were examined by keeping immobilization time (X_1) and chitosan concentration (X_4) at the central level 154 minutes and 0.0348 mg/mL and the result was shown in Figure 5.29. Both of these parameters have positive effect on BSA immobilization yield. The effect of temperature on BSA immobilization yield is higher than pH according to Equation (5.1) and the values reported in Table 5.19. The curvature of 3D surface in Figure 5.29 is due to the more effectiveness of immobilization time on BSA immobilization yield than the pH.



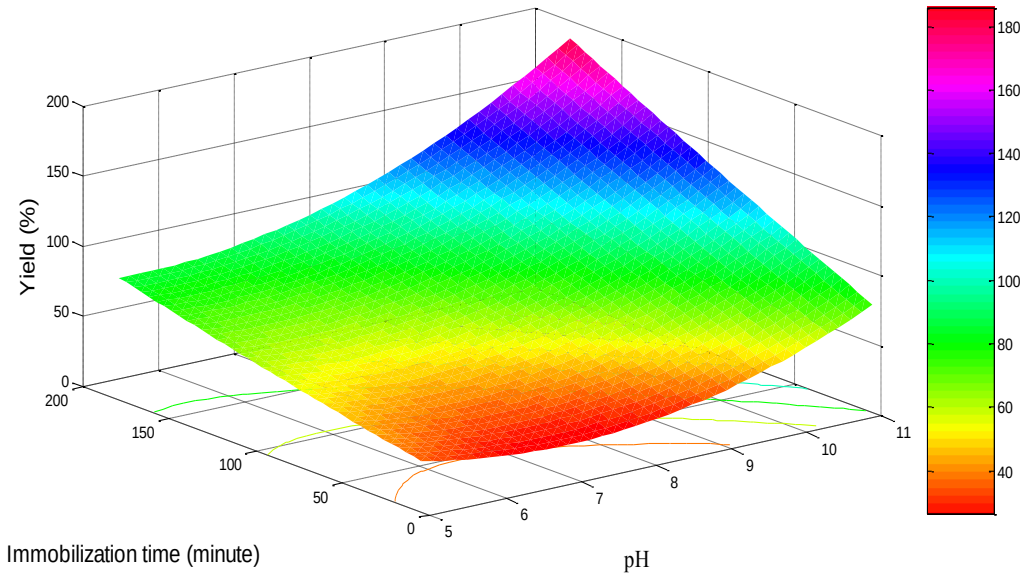
(a)



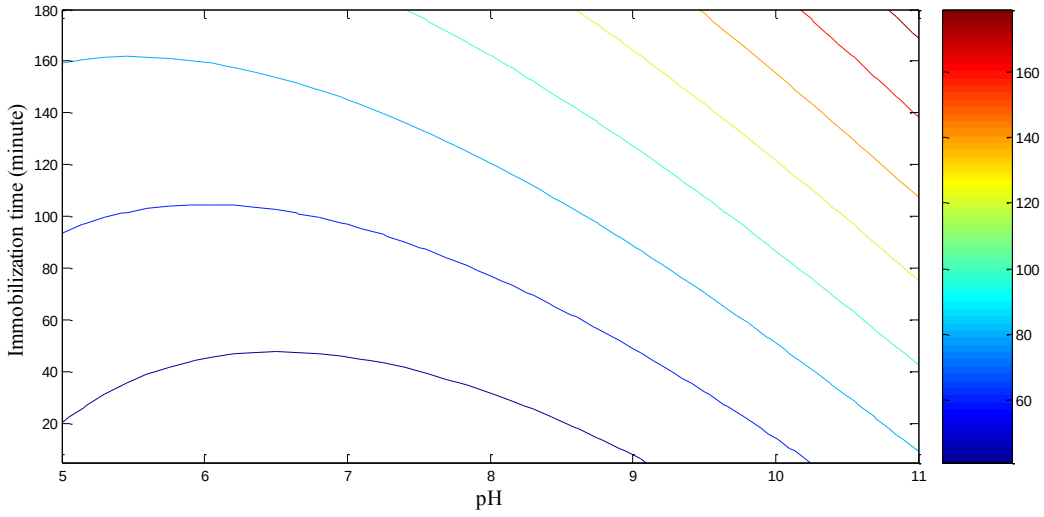
(b)

Figure 5.29. Response surface (a) and contour plot (b) showing the effect of pH and temperature on the BSA immobilization yield at a fixed chitosan concentration of 0.0348 mg/mL and immobilization time 154 minute

Figures 5.30 illustrate the 3D surface generated by immobilization time (X_1) versus pH (X_3) on BSA immobilization yield by keeping temperature (X_2) and chitosan concentration (X_4) at the central level 43°C and 0.0348 mg/mL. As indicated in these figures, immobilization time has positive effect on BSA immobilization yield similar to pH and temperature. The effect of immobilization time on BSA immobilization yield is higher than pH according to Equation (5.1) and the values reported in Table 5.19.



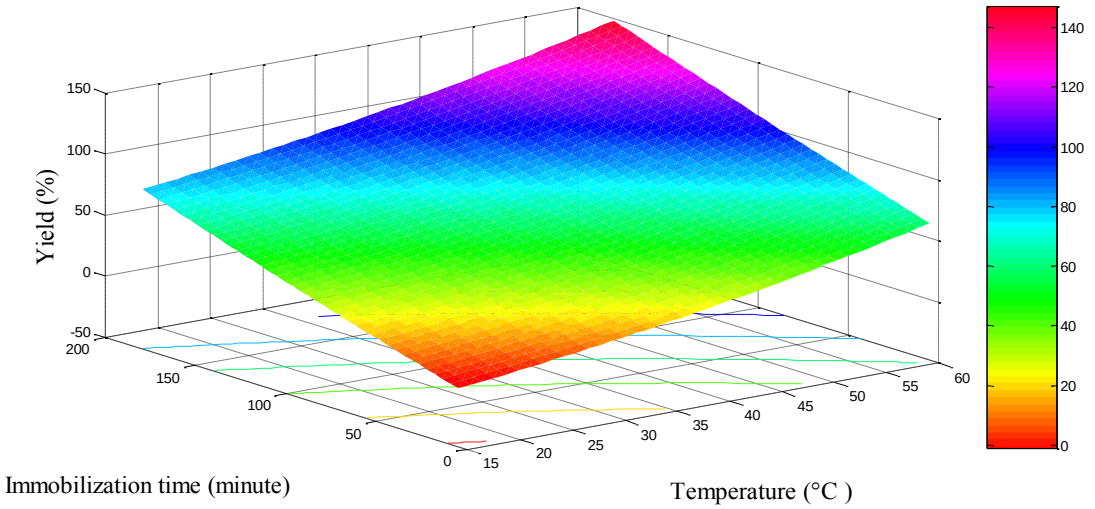
(a)



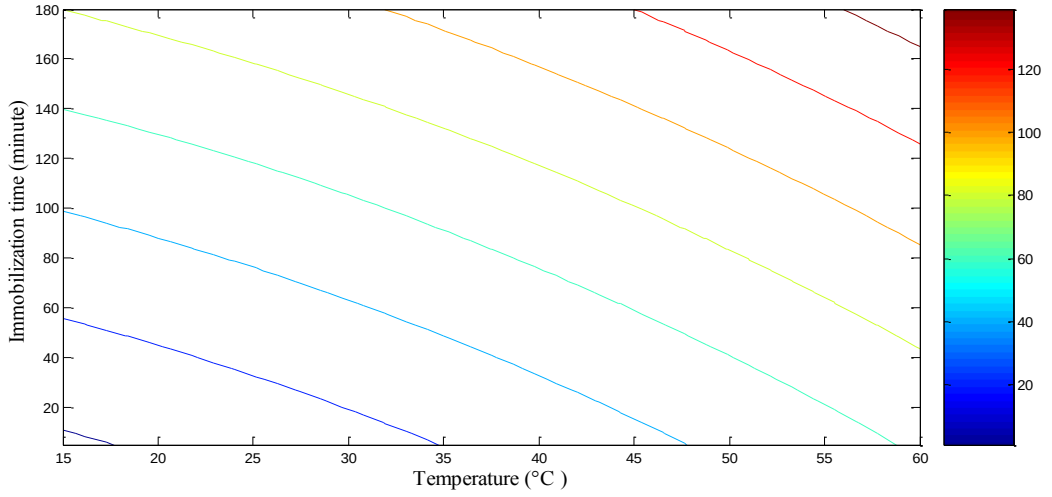
(b)

Figure 5.30. Response surface (a) and contour plot (b) showing the effect of pH and immobilization time and on the BSA immobilization yield at a fixed temperature of 43°C and chitosan concentration 0.0348 mg/mL

Figure 5.31 shows the response surface obtained by plotting immobilization time (X_1) versus temperature (X_2) with the keeping pH and chitosan concentration at the central point 8.45 and 0.0348 mg/mL, respectivly. Consequently, when immobilization time and temperature are at their maximum points, BSA immobilization yield would obtain the highest value. The effect of immobilization time on BSA immobilization yield is higher than temperature according to Equation (5.1) and the values reported in Table 5.19.



(a)



(b)

Figure 5.31. Response surface (a) and contour plot (b) showing the effect of temperature and immobilization time on the BSA immobilization yield at a fixed pH of 8.45 and chitosan concentration 0.0348 mg/mL

CHAPTER 6

CONCLUSION

In the first part of this study, Bradford protein assay was used to determine the concentration of BSA. This assay involves using the CBBG reagent which has different pronatation form. All the three dye forms (red, green, and blue) are able to combine with protein by nonelectrostatic forces. The anionic blue form of the dye has an advantage over other two forms in binding to protein by ionic attraction, which is the key point of the entire binding and color changing process. And also, there is a shift in the spectra to the lower wavelength due to the turbidity. In order to investigate effects of pH on the Bradford protein assay mechanism three different buffer solutions were used. Results indicated that variying pH does not cause significant changes in the spectral shifts as a result of increase in BSA concentration. By analyzing diffrences in the spectroscopic responses of BSA and protein binding effect of CBB in various buffer solutions, we have demonstrated that calibration models based on the full spectra as opposed to monitoring a single wavelength were much more effective for the determination of free BSA concentration. For these reasons, univariate calibration is not suitable for determining BSA concentration for large dynamic range of BSA concentration. Multivariate calibration techniques, such as Genetic Inverse Lleast Square (GILS) and Partial Least Square (PLS), were applied to Bradford protein assay. The success of the calibration models was obtained by SECV and SEP values as well as with the R^2 values from the reference vs. predicted concentration plots. These results demonstrated that successful calibration models can be constructed by using the method mentioned to provide good linearity for Bradford protein assay. When a comprasion is made between GILS and PLS, it was notably indicated that GILS models had better prediction performance than PLS models using full spectral range for determination of BSA concentration. However, when PLS is constructed by using the GILS selected spectral variables both PLS and GILS produced comparable results for the independent validation samples. These results can be explained by wavelength selection algorithm of the calibration models since GILS algorithm only focuses on the regions where the most concentration related information is contained. Also, dynamic range of Bradford protein assay is increased from 0-10 μg/ml BSA to 0-16 μg/ml BSA by multivariate calibration

method. According to the calibration results, GILS results that are obtained aginst CBB blank spectral collection were chosen for the further immobilization analysis.

Immobilization of Bovine Serum Albumin (BSA) on chitosan nanoparticles with physical adsorption was performed and the parameters were optimized by using central composite design (CCD). A second-order quadratic model was determined to explain the relationship between the immobilization yield and the parameters of chitosan concentration, pH, temperature and immobilization time. Emprical model is adequate for predicting the BSA immobilization yield. The results indicated that chitosan concentration have significant effects for enhancement of BSA immobilization. The optimized parameters were found 154 minutes, 43°C, 8.45 and 0.0348 mg/mL for immobilization time, temperature, pH for and chitosan, respectivly. The optimization of the BSA immobilization resulted that CCD provides fast and more detailed model to enhance the maximum yield.

REFERENCES

Antharavally, B. S., Mallia, K. A., Rangaraj, P., Haney, P., & Bell, P. A. (2009). Quantitation of proteins using a dye–metal-based colorimetric protein assay.*Analytical Biochemistry*, *385*(2), 342-345.

Beebe, K. R., Pell, R. J., & Seasholtz, M. B. (1998). *Chemometrics: a practical guide*. Wiley-Interscience: John Wiley & Sons, Inc.

Brereton, R. G. (2000). Introduction to multivariate calibration in analytical. *Analyst*, *125*(11), 2125-2154.

Brereton, R. G. (2003). *Chemometrics: data analysis for the laboratory and chemical plant*. England: John Wiley & Sons.

Bugg, T. D. (2001). The development of mechanistic enzymology in the 20th century. *Natural product reports*, *18*(5), 465-493.

Burgess, C. (2007). The basics of spectrophotometric measurement.*Techniques and Instrumentation in Analytical Chemistry*, *27*, 1-19.

Cao, L. (2006). *Carrier-bound immobilized enzymes: principles, application and design*. John Wiley & Sons.

Chen, S., Liu, Y. and Yu, P. (1996). Study on column reactor of chitosan immobilized. *Chemical Abstract*, 127 (4), 127-129.

Costa, S. A., Azevedo, H. S., & Reis, R. L. (2005). Enzyme immobilization in biodegradable polymers for biomedical applications.

Datta, S., Christena, L. R., & Rajaram, Y. R. S. (2013). Enzyme immobilization: an overview on techniques and support materials. *3 Biotech*,*3*(1), 1-9.

Dumitriu, S., Popa, M., & Dumitriu, M. (1989). Review: Polymeric Biomaterials As Enzyme and Drug Carriers Part IV: Polymeric Drug Carrier Systems.*Journal of bioactive and compatible polymers*, *4*(2), 151-197.

Guisan, J. M. (Ed.). (2006). Immobilization of enzymes and cells (Vol. 22). *Springer*.

Kennedy, J. F., & White, C. A. (1985). *Principles of immobilization of enzymes*. Handbook of enzyme biotechnology, 2, 147-207.

Krajewska, B. (2004). Application of chitin-and chitosan-based materials for enzyme immobilizations: a review. *Enzyme and microbial technology*, *35*(2), 126-139.

Kurita, K. (2006). Chitin and chitosan: functional biopolymers from marine crustaceans. *Marine Biotechnology*, *8*(3), 203-226.

Liu, Y., Sun, Y., Li, Y. L., Xu, S. C., & Xu, Y. X. (2011, September). Interactions Analysis in BSA-loaded Chitosan Nanoparticles at Different pH Values. In *Materials Science Forum* (Vol. 694, pp. 160-164).

Lozzi, I., Pucci, A., Pantani, O. L., D'Acqui, L. P., & Calamai, L. (2008). Interferences of suspended clay fraction in protein quantitation by several determination methods. *Analytical biochemistry*, *376*(1), 108-114.

Lü, X., Li, D., Huang, Y., & Zhang, Y. (2007). Application of a modified Coomassie brilliant blue protein assay in the study of protein adsorption on carbon thin films. *Surface and Coatings Technology*, *201*(15), 6843-6846.

Martens, H. T., Næs. 1989. *Multivariate calibration*. Wiley: John Wiley & Sons Ltd.

Massert, D. L., Vandeginste, B. G. M., Deming, S. N., Michotte, Y., & Kaufman, L. (1988). *Chemometrics: a textbook*. New York:Elsevier.

Meena, K., & Raja, T. K. (2006). Immobilization of Saccharomyces cerevisiae cells by gel entrapment using various metal alginates. *World Journal of Microbiology and Biotechnology*, *22*(6), 651-653.

Nasirizadeh, N., Dehghanizadeh, H., Yazdanshenas, M. E., Moghadam, M. R., & Karimi, A. (2012). Optimization of wool dyeing with rutin as natural dye by central composite design method. *Industrial Crops and Products*, 40, 361-366.

Ozdemir, D. (2006). Genetic multivariate calibration for near infrared spectroscopic determination of protein, moisture, dry mass, hardness and other residues of wheat. *International Journal of Food Science and Technology*, *41*, 12-21.

Özdemir, D. (2008). Near infrared spectroscopic determination of diesel fuel parameters using genetic multivariate calibration. *Petroleum Science and Technology*, *26*(1), 101-113.

Öztürk, B. (2003). Monitoring the esterification reactions of carboxylic acids with alcohols using near infrared spectroscopy and multivariate calibration methods. *Izmir Institute of Technology thesis of M.Sc.*

Silvério, S. C., Moreira, S., Milagres, A. M., Macedo, E. A., Teixeira, J. A., & Mussatto, S. I. (2012). Interference of some aqueous two-phase system phase-forming components in protein determination by the Bradford method. *Analytical Biochemistry*, *421*(2), 719-724.

Skoog, D. A., Holler, F. J., & Nieman, T. A. (1998). *Principles of instrumental analysis.* Philadelphia: Saunders College Publishing, Harcourt Brace College Publishers.

Tanyildizi, M. S., Özer, D., & Elibol, M. (2005). Optimization of α-amylase production by Bacillus amyloliquefaciens using response surface methodology. *Process Biochemistry*, 40(7), 2291-2296.

Tischer, W., & Kasche, V. (1999). Immobilized enzymes: crystals or carriers?.*Trends in Biotechnology*, *17*(8), 326-335.

Tolaimate, A., Desbrieres, J., Rhazi, M., Alagui, A., Vincendon, M., & Vottero, P. (2000). On the influence of deacetylation process on the physicochemical characteristics of chitosan from squid chitin. *Polymer*, *41*(7), 2463-2469.

Wang, Y., Veltkamp, D. J., & Kowalski, B. R. (1991). Multivariate instrument standardization. *Analytical chemistry*, *63*(23), 2750-2756.

Wei, Y. J., Li, K. A., & Tong, S. Y. (1997). A linear regression method for the study of the Coomassie brilliant blue protein assay. *Talanta*, *44*(5), 923-930.

Wiberg, K. (2004). Multivariate spectroscopic methods for the analysis of solutions (Doctoral dissertation, Stockholm).

Zaborsky, O. (1973). *Adsorption ımmobilized enzyme*, edited by Weast.

Zikakis, J. (Ed.). (1984). *Chitin, chitosan, and related enzymes*. Academic Press.

Zhao, J., & Wu, J. (2006). Preparation and characterization of the fluorescent chitosan nanoparticle probe. *Chinese Journal of Analytical Chemistry*, *34*(11), 1555-1559