

CHAPTER 2

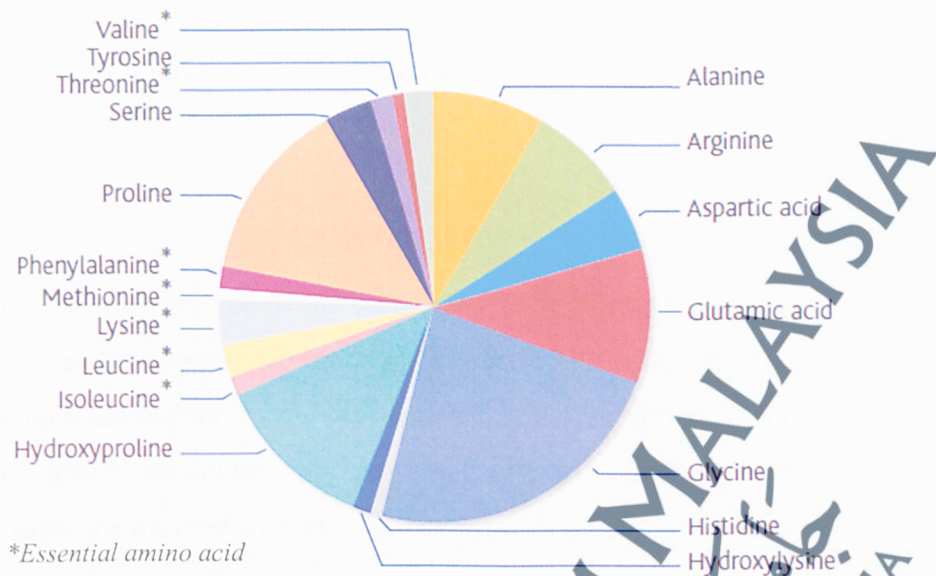
LITERATURE REVIEW

2.1 Gelatin

The animal kingdom has the structural pledge and most common protein and called by gelatin that is a primarily pure protein food ingredient and obtained by the thermals denaturation of collagen (Bailey and Paul, 1998). Water is the major component in gelatin structure so the gelatin is water soluble and solubilization involve the destruction of the tertiary, secondary and to some extent the primary structure of native collagens (Fernandez-Diaz et al., 2001), specifically by the partial hydrolysis of collagen derived from the skin, bones, hides skin, and connective tissue of animal (Morrison et al., 1999). Gelatin has benefits as gelling and thickening agent and it is used in a wide range of food products. Gelatin is composed of various elements such as carbon (50.5%), hydrogen (6.8%), nitrogen (17%), and oxygen (25.2%) (GMIA, 2012). The main component of gelatin is similar to the complex polypeptides collagen so the gelatin has a high molecular weight polypeptide and an important hydrocolloid. The hydrocolloid system in gelatin differs from other hydrocolloids because most of them are polysaccharide, where gelatin is a digestible protein containing all the essential amino acids (except tryptophan) (Ladislaus et al., 2007).

The source of gelatin can be found in many parts of animal body. The parts of animal that usually contain gelatin are bones, skin and hides. Gelatin could not be derived from the horns, hooves, and other non-collagenous part of vertebrate animals. In the 17th century, the commercial gelatin was produced in large scale. In the early 19th century, the commercial production method was gradually developed to obtain the extract of collagen with high molecular weight and a good quality of gelatin gel. Lately, many studies have been designed to find an alternative source of gelatin (Abdelfadeel, 2012).

FIGURE 2.1: Composition of Amino Acid in Gelatin (GMIA, 2012)



Gelatin is prepared by the thermal denaturation and physical and chemical degradation of collagen. Gelatin in a dry form consists of 98-99% protein. The molecular weight of these large protein structures typically ranges between 20,000 and 250,000 g/mol, with some aggregates weighting in the millions. The chemical structure of gelatin is described by a linear sequence of amino acids. It is always written from the -NH₂ end to the -COOH end. The predominant amino acids are *glycine*, *proline* and *hydroxyproline*. As a result, gelatin contains relatively high levels of these amino acids: *glycine* 26-34%; *proline* 10-18% and *hydroxyproline* 7-15%. Other significant amino acids include: *alanine* 8-11%; *arginine* 8-9%; *aspartic acid* 6-7% and *glutamic acid* 10-12%. The water content will vary between 6-9% and the ash content varies between 0.1% and 3.25% (Figure 2.1) (GMIA, 2012).

2.1.1 Gelatin Sources

Gelatins can be produced from different animal source. The principal of gelatin manufacture is that the source must be rich with collagen. It is in fact one of the most abundant sources present in mammals, including humans. About 25% of the total amount of protein in the mammalian body contains collagen which can be found in skin and bone of animals.

Animals that are typically used as the sources of gelatin are pig, beef, fish, insect and other vertebrate animals. There are alternative sources of gelatin from plant. But this is not gelatin because there is no evidence of the chemical relationship between gelatin and other plant based materials that are referred to as vegetable gelatin, such as seaweed extracts (GMIA, 2012).

1. Mammalian gelatin

Mammalian gelatin is derived from collagen which is the principal constituent of connective tissues and bones of vertebrate animals. The often chosen animal as the sources of gelatin are buffalo, bow, sheep, goat and pig but the most common animals used for the gelatin production are cow and pig. In terms of the manufacturing process, gelatins are derived into two types. Type A gelatin is usually derived from acidic process while Type B gelatin is from alkaline process bases. The structure of bovine and porcine gelatin is affected by the correlation between the average molecular weight and the gel strength which determine their strengths based on the isoelectric and melting point. From the previous study, gelatin from cow and pig had shown high strength (Lim and Mohammad, 2011). Therefore, mammalian gelatins (porcine and bovine) are the most popular and widely used because the raw material needed for the manufacture of gelatin from pigs and cattle are very easily to produce (Karim and Bhat, 2009).

2. Fish gelatin

The study of gelatin extracted from fish, has been reported for hake (Montero et al., 1999), megrim (Montero and Gómez-Guillén, 2000), and tilapia (Grossman and Bergman, 1992; Jamilah and Harvinder, 2002). The main difference between fish gelatin and mammal gelatin such as pigs and cows are: Fish gelatin has gel strengths and gelling point that is lower, but has a relatively higher viscosity than mammalian gelatin. Fish gelatin has gel strengths and melting temperatures that are associated with the place the fish lived. Generally, collagen derived from fish living in a lower temperature environment has lower content of amino acids (*proline* and *hydroxyproline*) than the species that live at higher temperatures.

Thus, gelatin produced from collagen of low temperatures fishes have lower number of hydrogen bonds in water solution and lower melting points than the gelatin made from mammals (Montero and Gómez-Guillén, 2000).

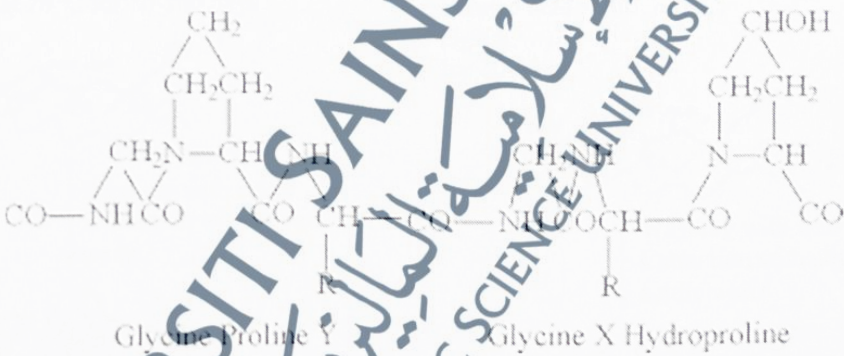
3. Insect gelatin

The use of insect as raw material to produce gelatin is relatively unknown for many people. Infact, not many countries produce the gelatin from insect, because the source of insect is limited. In Sudan, studies on gelatin extracted from insect, have been done. From the study, the types of insect such as *A. viduatus* and *A. pubescens* where used as raw materials to produce the gelatin. There are three different methods to extract gelatin from *A. viduatus* and *A. pubescens* namely, mild acid and distilled water extraction method; distilled water extraction method; and extraction with hot water. From the results obtained, extraction of insect gelatin using hot water gave higher yield of up to 3.0% followed by mild acid extraction (1.5%) and distilled water extraction (1.0%), respectively. The yields of gelatin extracted from *A. viduatus* and *A. pubescens* using distilled water with NaOH pretreatment at concentrations of 0.1, 0.2 and 0.5 mol/L was low compared to the results obtained for gelatin extracted using mild acid and distilled water (Abdelfadeel, 2012). Another study by Mariod et al., 2011 reported that, the alkaline and acid treatments during insect gelatin extraction showed positive effects on removing non-collagenous proteins with minimum collagen loss. In the study, gelatin obtained from insects was characterized using FTIR spectroscopy, and the spectra seemed to be similar to commercial gelatin. From the analysis, both insects showed 8.3% and 7.6% of moisture, 27.0% and 28.2% of crude protein, 54.2% and 57.3% fat and 3.5% and 2.5% of ash in dry condition, respectively. The two bugs proteins contained 16 essential amino acids when compared to the amino acid profile recommended by FAO/ WHO, the bug protein was of medium quality due to its moderate content of essential amino acids (Mariod et al., 2011).

2.1.2 Gelatin Structure

Gelatin consisted of amino acids and so the composition of amino acids will affect the chemical structure of gelatin. The effect is also influenced by animal species and their connective tissues (Zhou and Regenstein, 2006). Gelatin contains *glycine*, *proline* and *4-hydroxyproline* residues. Gelatin is an amphoteric protein with iso-ionic point between 5 and 9 depending on raw material and method of manufacture. The only other animal product containing *hydroxyproline* is elastin and then at a very much lower concentration, so *hydroxyproline* is used to determine the collagen or gelatin content of foods. In brief, the protein is made up of peptide triplets, *glycine* - X - Y, where X and Y can be any one of the amino acids but *proline* has a preference for the X position and *hydroxyproline* the Y position. Approximately 1050 amino acids produce an alpha-chain with the left-handed proline helix conformation. Gelatin has a definite molecular weight distribution pattern and is not completely poly dispersed, which correspond to the α -chain and the oligomers (Rbii et al., 2011) (Figure 2.2)

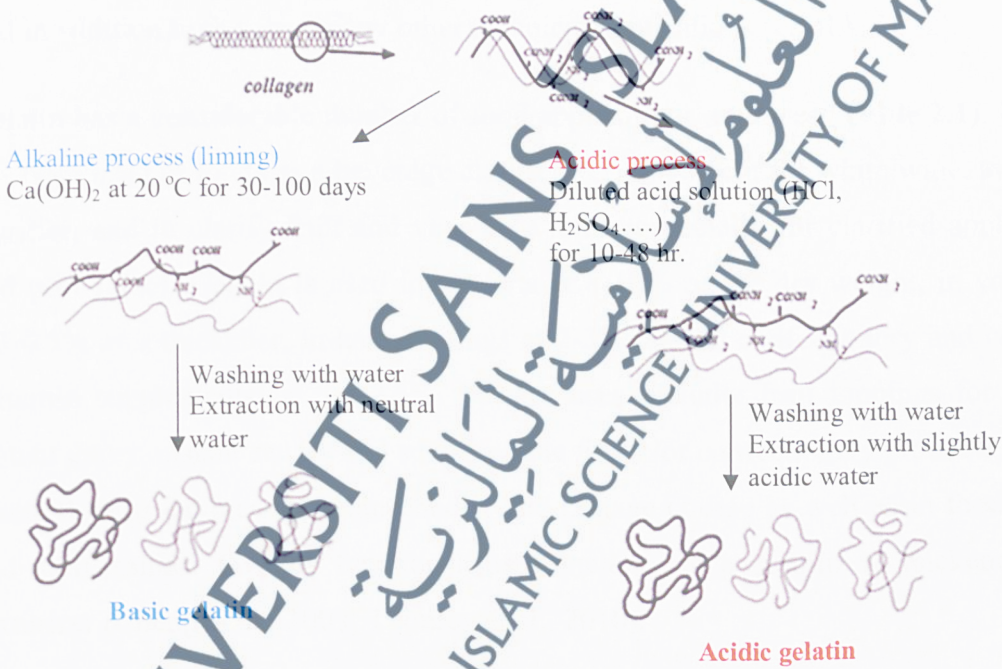
FIGURE 2.2: The Chemical Structure of Gelatin (Remawati, 2016)



Gelatin is the product of denaturation or disintegration of collagen. Initially the alpha-chains of collagen are held together with several different but easily reducible cross-links. As the collagen matures, so the cross-links become stabilized (Baily and Light, 1989). Then as time progresses the eta-amino groups of lysine become linked to arginine by glucose molecules (*maillard* reaction) to form the *pentosidine* type cross-links which are extremely stable.

Hence when the alkaline processing is used on young animal skin the alkali breaks one of the initial (*pyridinoline*) cross-links and as a result, on heating, the collagen releases, mainly, denatured alpha-chains into solution. Once the *pentosidine* cross-links of the mature animal have formed in the collagen, the main process of denaturation has to be thermal hydrolysis of peptide bonds resulting in protein fragments of various molecular weights i.e. polydisperse protein fragments. With the acid process, the collagen denaturation is limited to the thermal hydrolysis of peptide bonds with a small amount of alpha-chain material from acid soluble collagen in evidence (Cole and Roberts, 1996; Cole and Roberts, 1997). Nutritionally, gelatin is not a complete protein food because the essential amino acid *tryptophan* is missing and *methionine* is present only at a low level (**Figure 2.3**) (Ofori, 1999).

FIGURE 2.3: Process of Collagen Conversion to Gelatin (Ofori, 1999)



The functional properties of gelatin are related to their chemical characteristics. The gel strength, viscosity, setting behavior and melting point of gelatin depend on their molecular weight distribution and the amino acid composition. It is generally recognized that the amino acids like *proline* and *hydroxyproline* are important in the renaturation of gelatin subunits during gelling.

As a result, gelatin with high levels of amino acids tends to have higher gel strength and melting point. The molecular weight distribution is also important in determining the gelling behavior of gelatin. The sum of intact α and β fractions together with their peptides is proportional to the gel strength while the viscosity, setting rate and melting point increase with increase in the amount of the high molecular weight fraction (Johnston-Banks, 1990).

2.1.3 Gelatin Uses and Applicants in Food and Pharmaceutical

Gelatin has the capacity to stabilize colloidal solutions. The protein molecule moves towards the interface of the two phases and forms a monomolecular film around the micelles. Moreover, it increases the viscosity of the aqueous phase and promotes the formation and stability of suspensions or emulsions. In terms of gelling and stabilizing properties, gelatin plays a key role in food, pharmaceutical and photographic products and in addition to this, has many other technical applications (GMIA, 2012).

Gelatin has a considerable number of food applications and uses (Table 2.1). Gelatin has been used in foods as a beverage clarifier, a fining agent for white wine, as a beer clarifier, and to clarify fruit and vegetable juice (especially for clarified apple juice and pear juice). Gelatin is used in desserts at 8-10% of the dry weight, in yogurt at 0.3-0.5% as a thickener, in ham coatings at 2-3%, and in confectionery and capsules (vitamin supplements) at 1.5-2.5%. Further uses include: fruit toppings for pastry, instant gravy, instant sauces and soups, edible films for confectionery products, and as a stabilizer in ice cream, cream cheese, and cottage cheese as well as in food foams and fruit salads. Overall functional uses include as a stabilizer, thickener, and texturizer (Cheng et al., 2009; Tyburcy et al., 2010).

The largest proportion of gelatin procured by the pharmaceutical industry is used mainly for hard and soft gelatin capsules and for binding tablets, where it helps prevent oxidation and makes the preparation more palatable. The capsules are formed on mould pins on the surface of which carry a lubricant to facilitate the subsequent removal of the capsule (GMIA, 2012).

Gelatins with bloom in the range of 0-140 are offered for the microencapsulation of vitamins A, D and E. High Bloom gelatin (between 200 and 260 Bloom) is used for the production of hard gelatin capsules. Bloom is also known as the gel strength and is one of the most important properties of gelatin. The need to commercially characterize gelatin gels has resulted in the concept of gel strength. Commercial gelatins vary from 50 to 300 blooms. Hydrolysis of gelatin gels can be initiated by numerous factors, including pH, temperatures, enzymes, acids, bases as well as bacteria. These cause a reduction in the gelling properties of the gelatin (GMIA, 2012).

TABLE 2.1: The Function and Uses of Gelatin in Food

Functions	Area Used
Gel former	Gelled desserts, lunch meats, confectionery, pâté, consommé, aspics
Whipping agent	Marshmallows, nougats, mousses, soufflés, chiffons, whipped cream
Protective colloid	Confectionery, icings, ice creams, frozen desserts and confections
Binding agent	Meat rolls, canned meats, confectionery, cheeses, dairy products
Clarifying agent	Beer, wine, fruit juices, vinegar
Film former	Coating for fruits, meats, deli items
Thickener	Powdered drink mixes, bouillon, gravies, sauces, soups, puddings, jellies, syrups, dairy products
Process aid	Microencapsulation of colors, flavors, oil and vitamins
Emulsifier	Cream soups, sauces, flavorings, meat pastes, whipped cream, confectionery, dairy products
Stabilizer	Cream cheese, chocolate milk, yogurt, icings, cream fillings, frozen desserts
Adhesive agent	Affix nonpareils, coconut and other items to confections, to bond layered confections together, to bind frostings to baked goods, to bind seasonings to meat products.

(Turner, 1988)

Another use of gelatin is the percipients in pharmaceutical formulations, including vaccines, and is used as a binder for tablets and for excipients (a material other than the active substance that is added in the formulation of a preparation for various purposes or functions) and may be originating from quite distinct sources, including gelatin (Muyonga et al., 2004a).

2.2 Analytical Method for Detection of DNA in Gelatin Capsule

A reliable analytical method needs to be developed to detect the absence or presence of certain DNA in gelatin products. Several methods have been reported so far, based on chemisorption, chromatographic, immunochemical, mass spectrometric, spectroscopic and molecular techniques.

1. Chemisorption Method

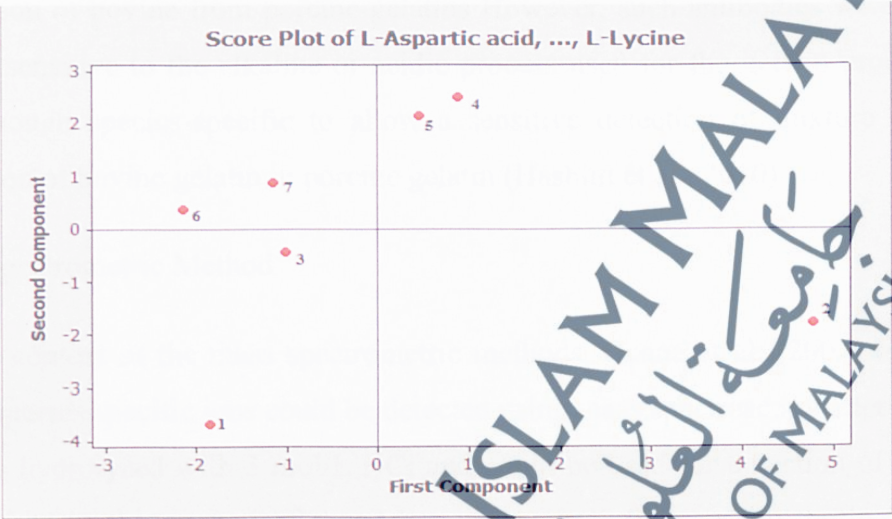
In the chemisorption method, it was reported by Hunter et al. (1986) that collagen, gelatin, and agarose gels promote the formation of hydroxyapatite from amorphous calcium phosphate. Hydroxyapatite is a compound which is formed by chemisorption of gelatin. Termine et al. (1970) reported that the presence of collagen or gelatin enhanced ACP formation, suggesting that the interaction between gelatin and calcium phosphate precipitation may be used to analyze the source of gelatins. Based on this fact, Hidaka and Liu (2003) presented a new method distinguishing bovine bone gelatin from porcine skin gelatin using the in vitro formation of calcium phosphate precipitates. In their study, the reaction of calcium phosphate precipitation was found to be useful in distinguishing bovine gelatin from other gelatin products. However, it was stated that further study was needed to clarify these effects.

2. Chromatographic Methods

In the use of chromatographic methods, Raraswati et al. (2012) carried out HPLC-based separation and determination of amino acids in gelatin using principal component analysis. In their study, 7 samples of porcine and bovine gelatins purchased from Sigma Aldrich, USA and 5 laboratory prepared soft candy gelatins were examined using the analytical procedure following a complete hydrolysis of samples by conventional acid hydrolysis to promote release of amino acid residues. Separation and determination of amino acids was achieved by reversed-phase HPLC (RP-HPLC) following pre-column derivatization. The 16 peaks resulted from HPLC analysis concluded that one sample was very typical of bovine gelatin (Raraswati et al., 2012). In this study, PCA was used to extract the significant variables from parameters of peak height percentage for each amino acid.

The result of PCA is the principal component (PC), which contains information to a certain amount of data variability. PC1 (first principal component) accounts for the most variation among data, while PC2 explains for the next largest variation and so on (Raraswati et al., 2012).

FIGURE 2.4: The Score Plot of Soft Candy Samples (Raraswati et al., 2012)



PCA score plot for classification of porcine and bovine gelatins and laboratory prepared soft candy from bovine and porcine gelatins. 1= porcine gelatin; 2 = bovine gelatin; 3= laboratory prepared soft candy from porcine gelatin; 4 and 5 = laboratory prepared soft candy from bovine gelatin; 6 and 7 = laboratory prepared soft candy from porcine and bovine gelatins.

Figure 2.4 shows the PCA score plot of porcine and bovine gelatins coming from Sigma and laboratory prepared soft candy. The horizontal axis is the scores for the first PC, and the vertical axis for the second PC. Bovine and porcine gelatins, both in standard or in soft candy, were clearly separated. However, the figure could not give the classification of porcine and bovine. The aforementioned physicochemical methods based on HPLC and calcium phosphate precipitation have not been proved yet to be able to detect bovine and porcine in commercial food product.

3. Immunochemical Methods

As for the immunochemical methods, Hofmann et al. (1999) examined an ELISA kit for its applicability to species identification in gelatin and gelatin-containing products, such as gum confectionery, processed turkey and chicken.

However, results were influenced by gelatin type, gelatin quality and concentration used and, in some cases, led to false negative or positive readings. The very high homology between collagen sequences of mammals' makes their immunochemical differentiation difficult when polyclonal antibodies raised against the whole molecule are used (Venien and Levieux, 2005). To overcome the challenge, Venien and Levieux (2005) produced highly specific antibodies, immunized rabbits against putative species-specific sequences of the bovine collagen alpha 1 (I) chain for differentiation of bovine from porcine gelatins. However, such antibodies were found to be very sensitive to the alkaline or acidic process used for the gelatin production and not enough species-specific to allow a sensitive detection of mixture of low concentration of bovine gelatin in porcine gelatin (Hashim et al., 2010).

4. Mass Spectrometric Method

Within the context of the mass spectrometric methods, Ocana et al. (2004) reported that some species specific ions could be detected using mass spectroscopy after bovine gelatin was hydrolyzed with 3 mol/L HCl and could be used for detection of bovine gelatin. However, the content of target ions might be influenced by the hydrolysis time and temperature. In this respect, Zhang et al. (2009) have developed a new method to differentiate bovine and porcine gelatin based on detection and identification of marker peptides in digested gelatins. In their study, the gelatins were digested by *trypsin*, and the resulting peptides were analyzed by high performance liquid chromatography/tandem mass spectroscopy (HPLC-MS/MS). Using this methodology, they could successfully detect the marker peptides specific for bovine and porcine in the digested bovine and porcine gelatin, respectively. Another important result that was reached by this study was that the peptide identification was remarkably influenced by *proline* hydroxylation. However, it was also reached to a conclusion that it was necessary to manually verify the sequence for peptides (GPPGSAGSPGK and GPPGSAGAPGK detected in digested bovine and porcine gelatin, respectively) since there might be a risk to confuse the *proline* hydroxylation with the mass difference between Ser and Ala residues. They concluded that detection of marker peptides in the digested gelatin sample using HPLC-MS/MS was an effective method to differentiate between bovine and porcine gelatin. However, the method also has some drawbacks (Zhang et al., 2009).

During MS/MS data processing, the threshold for specific peptide identification might be different from one species to another (Li Chen et al., 2007; Shadforth et al., 2005). More clearly, it is very difficult to detect marker peptides in the digested collagens because of the very high homology between the collagen sequences of mammals using this method. Furthermore, *proline* hydroxylation is another challenge making identification of peptide more difficult than for that of most proteins (Zhang et al., 2009).

5. Spectroscopic Method

In the spectroscopy method, the FTIR spectroscopy together with attenuated total reflectance (ATR) or transmission accessories has been used to determine gelatin and intermolecular cross-linking study of collagen and proteins (Cao and Xu, 2008; Muyonga et al., 2004). Hashim et al. (2010) have used the Fourier Transform Infrared (FTIR) spectroscopy for differentiation of bovine and porcine gelatins. In their study, in order to determine unknown gelatin sources, deformation of N-H bonds found in the range 3290–3280 and 1660–1200 cm^{-1} within infrared spectra of all gelatin samples were analyzed using discriminant analysis. These regions were found to give information about the origin of the gelatin. However, this method was concluded to need repeated results (Hashim et al., 2010).

6. Molecular technique

DNA techniques or molecular technique have become very important and are widely used nowadays compared to protein techniques. Advantages of DNA analyzing methods are manifold: DNA is a relatively stable molecule easily allowing analysis of processed and heat treated food products. Due to the ubiquity of DNA in every type of cells, all kinds of tissue can be analyzed. Early methods based on hybridization of specific probes (Chikuni et al., 1990; Wintero et al., 1990) were complicated, time consuming and insufficient for complex matrices. The polymerase chain reaction (PCR) proved to be an adequate technique for detection of small amounts of DNA, specifically amplifying a target region of template DNA in a rapid and sensitive manner (Chikuni et al., 1990).

PCR systems used today only allow a qualitative detection of animal species. Because even trace amounts of less than 0.1% of one species lead to positive PCR results (Behrens et al., 1999), it is necessary to differentiate between technically unavoidable contamination or intentional admixture. Therefore, adequate quantitative detection methods for small amounts of DNA must be available. PCR analysis can be used as a quantitative method if internal standards (competitors) are co-amplified with the target DNA (Section 2.4).

Cai et al. (2011) have used real-time PCR for detection and quantization of porcine and bovine DNA in gelatin mixtures and gelatin capsules. In this study, two species-specific qPCR assays were developed based upon repetitive elements. They allowed sensitive detection of porcine and bovine DNA at as low as 1 pg/mL. The lack of cross-reactivity when the sets were used to amplify DNA from the other species indicates high specificity of the assay. When binary gelatin blends containing various amounts of porcine and bovine gelatin were prepared and analyzed by the qPCR assays, the determined ratios of porcine material to bovine material were very close to their theoretical values, and a contamination level as low as 1% of the other species in the gelatin blends could be determined. When evaluated in gelatin capsules, although significantly less DNA was detected, determination of porcine and bovine species identities and estimation of the relative abundance of each species was possible. Therefore, the porcine and bovine species-specific qPCR assays described here represent simple, reliable and sensitive DNA-based tests for determination and quantization of the species of origin from highly processed products (Cai et al., 2011).

2.3 DNA Isolation Methods for Gelatin

The isolation of genomic DNA is a fundamental requirement for many analytical procedures. Although proper collection and stabilization of the sample is crucial, purification of the DNA is often the key step for success in downstream analysis. Ideally, in an analytical environment, an effective DNA extraction procedure should be as simple, safe and cost and time efficient as possible. It should also reproducibly provide DNA of a sufficient quality and yield to allow subsequent analysis (Ricci et al., 2013).

Generally, there are three important factors that affect the yield of the result of suitable analytical method of isolated DNA as an analyte for a given technique, namely: (i) Concentration of the DNA (ng/ μ l), (ii) Purity of the DNA (1.8-2.0) and (iii) Stability. The three factors can be influenced with how the technique is engaged and, in turn, in subsequent analysis can impact upon the validity of techniques applied (Ricci et al., 2013).

The result of yields and purity of DNA give any different if the extraction method is engaged differently. From the previous study extraction methods have been systematically evaluated for specific applications with the different sample such as soil and sediment (Collen, 2011), human microbiome (Yuan et al., 2012), and fecal (Claassen et al., 2013; Kennedy et al., 2014; Peng et al., 2013) and the study had been concluded that the use of modern method such as silica based technologies, magnetic separation, anion exchange technologies give the optimum result if we look based on the purity and the yield of the DNA.

Silica based technologies with micro spin tube is already the method that have been used for many isolation kits. Qiagene, Kogene and Agilent also use the method to isolate the DNA from samples. The basis for most of the products related to nucleic acid purification is the unique properties of silica matrices for selective DNA binding. Types of silica materials including glass particles, such as glass powder, silica particles, and glass microfibers prepared by grinding glass fiber filter papers, and including diatomaceous earth. Hydrated silica matrix, which was prepared by refluxing silicon dioxide in sodium hydroxide or potassium hydroxide at a molar ratio of about 2:1 to 10:1 for at least about 48 hours, had been introduced in DNA purification (Padhye et al., 1997; Woodard et al., 1994).

The principle of silica matrices purification is based on the high affinity of the negatively charged DNA backbone towards the positively charged silica particles. Sodium plays a role as a cation bridge that attracts the negatively charged oxygen in the phosphate backbone of nucleic acid. Sodium cations break the hydrogen bonds between the hydrogen in water and the negatively charged oxygen ions in silica under high salt conditions ($\text{pH} \leq 7$). The DNA is tightly bound, and extensive washing removes all contaminations (Arnolds et al., 2005).

The purified DNA molecules can be eluted under low ionic strength ($\text{pH} \geq 7$) later by using TE buffer or distilled water (Esser et al., 2006). Besides silica matrices, nitrocellulose and polyamide membranes such as nylon matrices are also known to bind with nucleic acids, but with less specificity (Arnold et al., 2005).

The first stage in DNA isolation is the process of destruction the membrane and cell wall. The destruction cell (lysis) is a stage of initial isolation of DNA which aims to remove the contents of the cell (Holme and Hazel, 1998). The destruction of cells or tissues have several method namely physically, chemically and enzymatic. Techniques destruction of cell or tissue in this research is enzymatic technique, which as many as 220 μl proteinase K working solution (200 μl of proteinase K digestion buffer and 20 μl of proteinase K) put in sample and incubated at 65°C for 1 hour. Extraction using Proteinase K working solution causes protein to lose the solubility and undergo further the precipitation that could be separated from the DNA by centrifugation (Karp, 2008). Bettelheim and Landesberg (2007) stated that after centrifugation would be formed two phases which is the organic phase on the bottom layer and the phase aqueous phase (water) on the top layer DNA would be in phase aqueous after centrifugation, protein that was denatured would be in inter-phase and lipids would be in the organic phase.

After extraction process, the DNA had been obtained was concentrated by precipitation stage. In general, the solution used at this stage is ethanol or isopropanol. The two solutions would precipitate DNA in aqueous phase so that DNA clumping to form fiber structure and after it is formed pellet (centrifuge) (Switzer and Liam, 1999). In this precipitation stage, DNA had been separated from the residues of RNA and protein that remained. At the time of ethanol or isopropanol had been discarded and the pellet was dried in a tube, the pellets was remaining in the tube is concentrated DNA. The process of re-precipitation with ethanol or isopropanol before the dried pellets could increase the degree of purity of DNA isolated (Bettelheim and Landesberg, 2007). Keller and Mark (1989) explained that washing back the pelleted precipitated using ethanol or propanol was intended to eliminate the residues of salt remaining (Keller and Mark, 1989).

The salts that was involved in the extraction process is less soluble in isopropanol so could be precipitated along with the DNA, and therefore is required ethanol after precipitation with isopropanol to remove residual salt (Ausubel et al., 2003). After precipitation and washing with ethanol, the ethanol solution which has been used was discarded and the pellet was dried by centrifuged the pellet without any treatment. The stage was aimed to remove residual ethanol from pellet DNA. After the DNA pellet had been dried, the next step was the addition of the elution buffer into tubes that contained pellets and then stored in freezer with temperatures at around -20°C . Verkuil et al. (2008) suggested that the elution buffer and temperature storage at -20°C was intended the DNA that had been extracted could be kept up for weeks (Ausubel et al., 2003).

Traditional DNA extraction methodologies employing chemicals such as SDS, proteinase K and phenol are now reasonably well established (Sambrook and Russell, 2001). These methods tend, however, to be time consuming and involve multiple liquid transfer operations. Alternative commercial kits often offer reduced hands-on time and cleaner approaches to extractions; however they can be more expensive and limited to very specific applications. In an analytical environment neither of the approaches mentioned may be ideal and the situation can be further complicated by the sample matrix composition. Whilst simple approaches may be preferable, complex matrices and non-ideal samples may demand additional clean-up procedures. Measurement of DNA yield in itself is not sufficient to determine the suitability of an extraction methodology (Sambrook and Russell, 2001).

The quality, encompassing purity and integrity, of the DNA analyte can also be important in ensuring suitability for downstream measurement. An inappropriate choice or a sub-optimal extraction methodology could have significant consequences for subsequent analyses, which may have to be repeated, or produce false negative results. Automation of the entire extraction procedure increase throughput and reduces analyst errors, often resulting in a more accurate and cost-effective DNA isolation step in the laboratory workflow. Validation of sampling procedures, sample storage, sample preparation and DNA extraction should all be considered vital to the production of quality data in subsequent analyses whether these are accomplished manually or on automated platforms (Sambrook, 2001).

2.4 Analysis of Gelatin Capsule

2.4.1 Polymerase Chain Reaction

PCR (polymerase chain reaction) is one of the techniques to learn about molecular biology and it an enzymatic method that was used to multiply exponentially a nucleotide sequence specific in vitro way. The method is very sensitive, so it can be used to multiply the DNA molecule (Rosenthal, 1992).

The original concept of PCR technology requires that certain parts of the DNA sequence to be multiplied should be known before the multiplication process can be done. The known sequence is important to provide a primer, which the short oligonucleotide sequence that the function initiate synthesis of DNA in the polymerase chain reaction. Further development of PCR method allows the multiplication of a DNA fragment of the unknown sequence, for example with Alu-PCR method (Rosenthal, 1992). Alu is a DNA sequence (length approximately 300 bp), which is widely available throughout the human genome (repetitive DNA sequences). Alu-PCR is a PCR method that utilizes the alu sequences as a basis for primers to multiply a DNA fragment of unknown sequence that contained between two Alu sequences (Rosenthal, 1992).

Target PCR is a nucleic acid (DNA) double strand that was extracted from the cells and denatured into single-stranded nucleic acids. The components of PCR reaction consists of a form of specific oligonucleotide primers for the target genes selected, enzymes (generally Taq polymerase, thermostable and thermoactive enzyme derived from *Thermus aquaticus*) and triphosphat deoxynucleoside (dNTP) that is used to amplify the target gene replicated exponentially with multiply the initial target. The reaction is carried out in a heating machine that is programmed automatically and called a thermocycler. The machine provides thermal conditions that necessary for amplification process (Nollet and Toldrá, 2011). The process that occurs in the PCR machine includes three main stages, namely is denaturation (the separation of double stranded DNA), annealing and extension (primer elongation). The process that is started from denaturation, annealing and extensions was referred to as one cycle. The PCR products can be directly visualized through a process of electrophoresis and used for further analysis (Weissensteiner et al., 2004).

PCR techniques began to develop after the discovery of DNA polymerase enzyme, which can replicate DNA. The technique was initially developed using Klenow fragment of DNA polymerase I that was derived from *Escherichia coli*. Klenow fragment is DNA polymerase that has eliminated the exonuklease activities. The enzyme has some weaknesses in the highest temperature, the medium of polymerization medium, and the process is low (Ali et al., 2012).

The low process of this reaction indicates a low ability of the enzyme to incorporate nucleotide polymerase with a primer continuously without experiencing dissociation of a complex of primer-template DNA. PCR method was first developed in 1985 by Kary B. Mullis, the researcher at the company Cetus Corporation. In the early development of PCR method is only used to multiply the DNA molecule, but then it was developed further so that it could also used to multiply and perform quantization of mRNA molecules. This time PCR method has been widely used for a variety of genetic manipulation and analysis (Ali et al., 2012).

PCR process involves four main components: (1) DNA template, the DNA fragment to be duplicated, (2) oligonucleotide primer, which is a short sequences of oligonucleotide (15-25 nucleotide bases) that initiate the synthesis of DNA chains, (3) deoxyribonucleotides triphosphate (dNTPs) that is consisting of dATP, dCTP, dGTP, dTTP, and (4) DNA polymerase enzyme, which is a reaction catalyst synthesis of DNA chains. PCR process consists of three stages: denaturation, annealing and amplification. At the stage of denaturation, a DNA fragment (double strand) was heated at 95 °C for 1-2 minutes so it will separate into single chain (single-strand). The annealing stage is started at a temperature of 55 °C for 1-2 minutes, the oligonucleotide primer attached to the template DNA that complementary with the primer sequences. After the annealing process, the temperature was raised to 72 °C for 30 seconds. At this temperature, the DNA polymerase enzyme will do the poses polymerase, the DNA chain that will form a hydrogen bridge with the template DNA. The process is called amplification (Yuwono, 2006).

The advent of PCR during the mid 1980s enabled measurements of target DNA across a wide dynamic range, and is sensitive to as little as a few copies (Saiki et al., 1988).

Conventional PCR analysis involves end-point detection of the products formed, and as the rate of product generation is not linear over the course of the reaction then extrapolating from the final amount of product to be the initial amount of starting material is not straightforward. However by defining the sensitivity of the reaction or by the use of reference samples, comparative standards or competitive mimics it is possible to generate data ranging through qualitative to semi-quantitative and quantitative (Saiki et al., 1988).

Such strategies all involve the detection or quantifications of PCR product, using a range of methodologies including agarose gel in the presence of ethidium bromide, capillary electrophoresis and mass spectrometry. Detection methods vary in the accuracy and precision of analysis. More recently the development of quantitative real time PCR has enabled highly accurate quantification using appropriate calibration standards, and this will be detailed out in the next section (Section 2.4.2) (Higuchi et al., 1993).

2.4.2 Real-Time PCR

The sensitivity of analysis achievable with PCR has led to the technology being adopted across a range of sectors. For many applications a quantitative results is required, which has driven the development of a range of strategies to determine the amount of starting material in a sample. Approaches such as competitive PCR (Gilliland et al., 1990) and limiting dilution analysis (Levinson et al., 1994) have been used as routes to quantification although the variable nature of the PCR process and the amplification of the target to a maximal level irrespective of the starting amount of target limit the accuracy of these methods (Alvares et al., 2000).

The development of PCR method that currently used is real-time PCR. The detection in real-time PCR is underway in single stage because the accumulation of specific products is recorded continuously during the cycle and it could not be performed on a standard PCR (conventional PCR) that still relied on agarose gel electrophoresis to quantify the amplicons.

The detection of Real-time PCR products is quantitatively different from the standard PCR that the quantization of the amplicon based on the length of base or molecular weight (Dharmaraj, 2009).

The quantity of real-time PCR products are calculated based on the threshold cycle (Ct) which is the fluorescent intensity is greater than the fluorescence that was induced by noise (background fluorescence). Noise can be caused the attachment of a solution of DNA isolates along with PCR reagents in the tube wall micro-well. The advantages of real-time PCR are the detection was measured precisely when the target amplification was first detected in each cycle (exponential phase) and not in the final phase of amplification (the plateau phase) as occurs in standard PCR. The more amplification cycles that are performed, the higher the yield of the reaction, as long as the components of the reaction are not completely limiting and the enzyme retains some activity. Through the repeated denaturation cycles of the PCR the enzyme activity does become depleted, and in practice running more than 40 cycles does not appreciably increase yield. It is also important to balance speed/throughput with overall efficiency, so it may be effective to use a smaller number of amplification cycles, if this allows an additional run to be performed within the working day (Montowska and Pospiech, 2010).

According to Edwards et al. (2004), the application of real-time PCR technology reduces the time of treatment or testing and improves the accuracy of PCR quantification methods. Thus, the use of real-time PCR technique is more efficient and effective than standard PCR (Edwards et al., 2004). Quantification of real-time PCR can be performed by two techniques: (1) the use of fluorescent dye that binds to double-stranded DNA and (2) the use of a modified DNA oligonucleotide probes that emit fluorescence when hybridized with a complementary DNA.

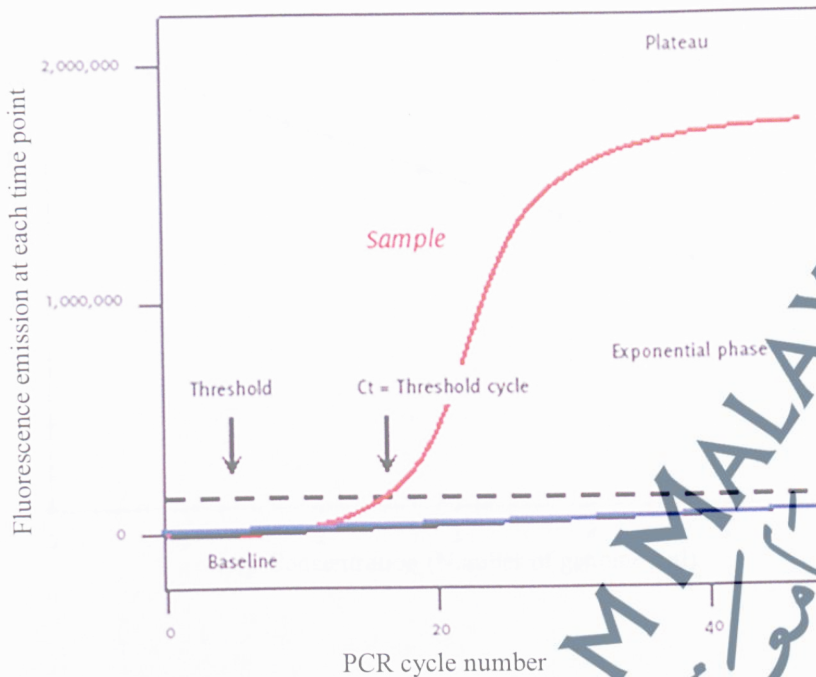
The use of fluorescent dye is also known as SYBR green method. SYBR Green-based detection is the least expensive and easiest method available for real-time PCR. Other methods (such as TaqMan) require an expensive third primer labeled with a dye and a quencher. Most real-time systems detect and accommodate SYBR Green making the method very flexible; however, some instrumentation may also require the simple addition of a reference dye to normalize the system's optics.

SYBR Green specifically binds double-stranded DNA by intercalating between base pairs, and fluoresces only when bound to DNA. Detection of the fluorescent signal occurs during the PCR cycle at the end of either the annealing or the extension step when the greatest amount of double-stranded DNA product is present (Soares et al., 2013). However, SYBR Green detects any double-stranded DNA non-specifically. Therefore, the reaction must contain a combination of primers and master mix that only generates a single gene-specific amplicon without producing any non-specific secondary products (Soares et al., 2013).

The data generated can be analyzed by computer software that is connected with a thermal cycler to calculate the number of copies of DNA or the threshold cycle (Ct) of certain pathogens in food samples. Amplification is carried out in a thermal cycler was shown in the form of a graph on a computer screen with software that applicable against thermal cycler, for example, is a software steponeplus (applied biosystem). The graphics are obtained in the form of graphs amplification, standard curves and melt curve (only for SYBR green method). The graph is used to evaluate the performance of real-time PCR amplification (Wang et al., 2010).

Graph amplification is formed since the amplification process is started. The graph serves to determine whether the thermal cycler was amplification or not. Graph amplification could be seen in Figure 2.5. Amplification is determined based on fluorescence intensity. The more the amplification product is produced the greater the accumulation of fluorescent is legible. The increase in fluorescence is characterized by the formation of graph sigmoidal (red line). Sigmoid graph would be intersected with base line threshold that has been determined automatically by the program. The intersection point between the sigmoid graph and base line threshold if it reflected in the X-axis (Cycle) is called the threshold cycle (Ct) for the samples that were amplified. The blue line will be formed if there is no amplification in a thermal cycler. Ct value is the cycle above the noise product which is the product of accumulation (2^n , n is the number of repetition cycles of amplification) was first legible in the exponential phase. Exponential phase ended into the plateau phase when PCR reagents in the reaction mixture had completely reacted (Kordo et al., 2013).

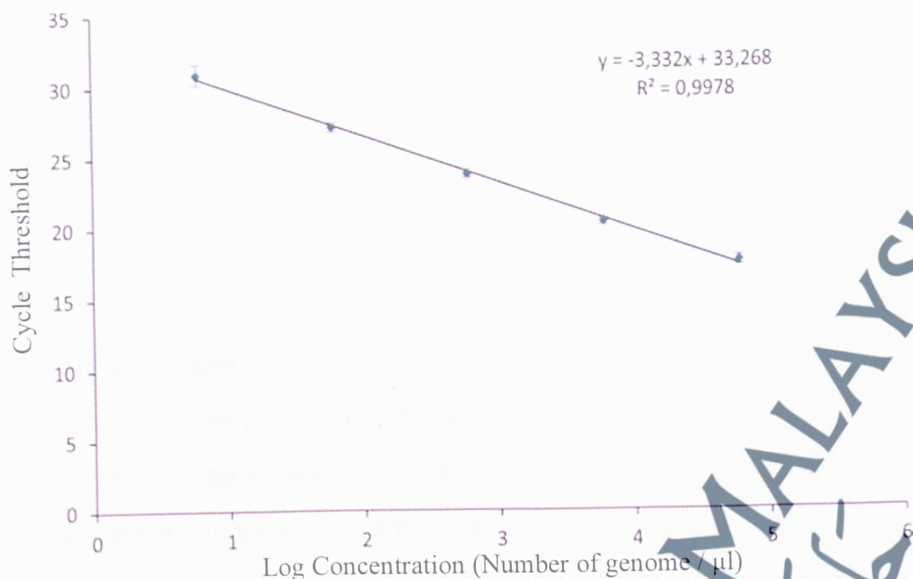
FIGURE 2.5: The graphs Amplification in *Real-Time* PCR (Kordo et al., 2013)



The model software will construct amplification plots from the extension phase fluorescent emission data collected during the PCR amplification. The standard deviation is determined from the data points collected from the base line of the amplification plot. Ct values are calculated by determining the point at which the fluorescence exceeds a threshold limit (usually 10 times the standard deviation of the base line) (Kordo et al., 2013).

The fluorescent conversion factor can be determined by means of a calibration graph, on which the fluorescence values from a dilution series of DNA standard has been plotted against known concentration (Figure 2.6). The variations in fluorescence with concentration should be linear provided the DNA concentration is not excessively high and sufficient dye has been used to saturate all possible intercalation sites. It is important to ensure that the concentration ranges of the standards give fluorescence values that are within the dynamic range of the fluorometer or fluorescence plate reader being used for the assay. The gradient of the calibration graph can then be used as a conversion factor to determine sample DNA concentrations (Alexander et al., 2013).

FIGURE 2.6: Calibration Curve of DNA Standard (Alexandra et al., 2013)



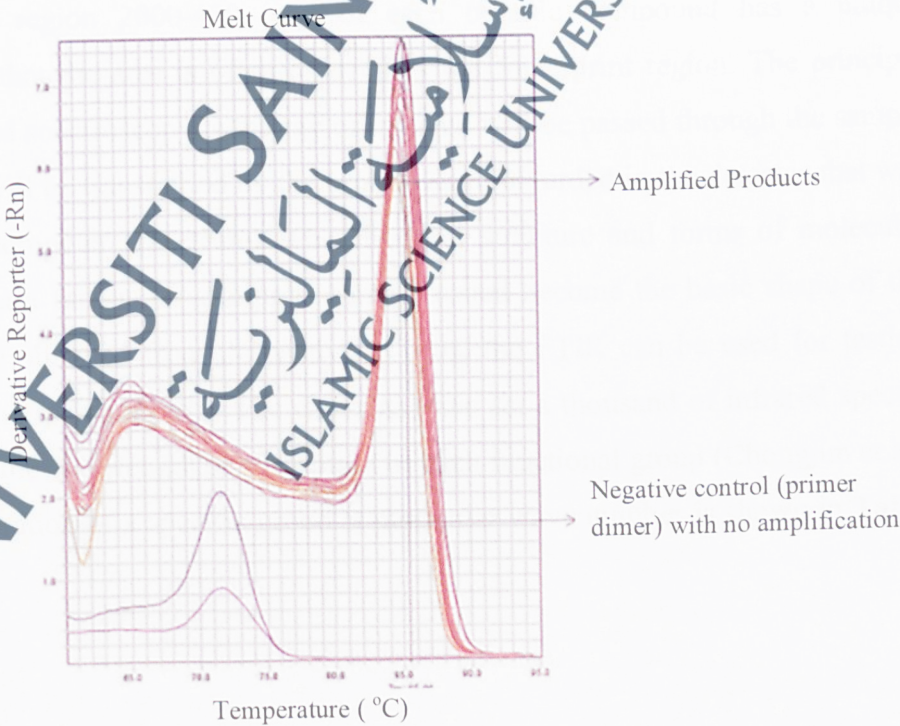
Concentration values of the standard could have been measured by another established methodology such as UV spectroscopy, or supplied by a manufacturer. The accuracy of the quantification process is significantly influenced by the quality of the standards used, and thus consistency in the source and use of standards is central to acceptable method performance. Further, as determination of absolute DNA quantities remains technically challenging, then the accuracy of any assigned standard values must be critically assessed to determine the reliability of standards used for fluorescent spectroscopy (Wilhelm et al., 2003).

In addition to measuring the increase in product at each cycle, an analysis of the products generated in the reaction may be performed at the end of the amplification process. This is termed 'melt analysis' and is compatible with both intercalating dye reporter system and those where the primers binds to the PCR product to achieve a change in fluorescence intensity. To perform the analysis the fluorescent signal is monitored as the temperature is gradually increased from around 50 to 95 °C, which results in an increase in fluorescence as the double stranded DNA or primer: product complex is dissociated. The change in fluorescence against temperature is usually plotted by instrument software, yielding peaks corresponding to the denaturation maxima of each double-stranded species present in the system (Figure 2.7).

The dissociation curve for the analysis was showing a single, sharp peak, suggesting that only a specific PCR product was generated with this set of primers. Primers-dimmers, non-specific or mismatched sequences will generally have a lower melting temperature than the specific product of the reaction, and can be distinguished by this post-PCR analysis (Cubie et al., 2001).

Recently, an extension of melt curve analysis have been developed, termed High Resolution Melt (HRM) (Corbett Life Science), enabled by improvements in real-time PCR instrument capabilities and the dye used in analysis. Rapid data collection is required, with very high precision thermal resolution (down to 0.02 °C) and dedicated analysis software. Samples are characterized based on very detailed measurement of their disassociation kinetics, and with more detailed melt profiles samples may be discriminated by length, GC content and sequence. Even single base-pair mismatched can be distinguished, allowing application of the method to detailed genotyping analysis, and thus this approach could potentially be exploited to replace the use of more complex probe reporter system (Dufresne et al., 2006).

FIGURE 2.7: Melt Peaks of DNA Amplification Process (Kumar et al., 2013)



2.4.3 Fourier Transform Infrared Spectroscopy

In infrared spectroscopy, infrared light is passed through the sample and then was measured the fraction of absorbed radiation in the wavelength range and produces a spectrum in which shows the qualitative information from protein. The chemical structure and forms a molecular bond with certain functional groups of samples tested becomes the basis of the spectrum form to be obtained from the result of analysis. Thus it can be used for testing qualitatively and quantitative. Infrared spectroscopy is a method to observe the interaction of molecules with electromagnetic radiation in the wavelength of 0.75 to 1000 μm , or in a wave numbers 13000-10 cm^{-1} (Rohman et al., 2012).

Infrared is an electromagnetic ray that the wavelength is more than the visible light and less than the microwave (between 700 nm and 1 mm). Wavelength region in infrared spectroscopy is the mid-infrared region, which is at wavelengths from 2.5 to 50 μm or at wave number 4000-200 cm^{-1} (Petibois and Deleris, 2006).

Vibration that was used for identification is the vibration bends, particularly rocking, which is located in the area of wave number 2000-4000 cm^{-1} (functional groups). Therefore, the region 2000-400 cm^{-1} of each organic compound has a unique absorption, so that the area is often referred as the fingerprint region. The principle work of infrared spectroscopy is the ray of infrared will be passed through the sample and the wave will be transmitted by the sample and forwarded in the detector that was connected in computer (spectrum). The chemical structure and forms of molecular bond with certain functional groups in sample tested become the basic shape of the spectrum to be obtained from the analysts. Thus, the FTIR can be used for testing qualitatively and quantitatively. The chemists have set a thousand of infrared spectra and determine the wavelength of absorption of each functional group (Chongjun et al., 2010). The vibration of a specific group at particular wave number is shown in **Table 2.2.**

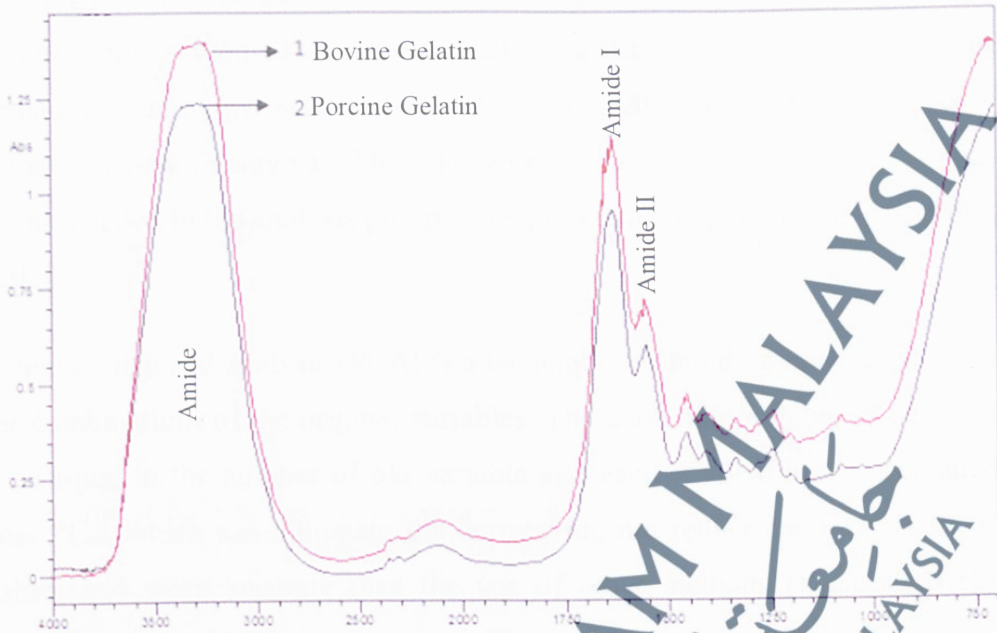
TABLE 2.2: Specific Functional Groups on a Particular Wave Numbers

Group	Specific compound	Absorption (cm^{-1})
C-H	Alkane	2960, 2850, 1350
C-H	Alkenes	3080-3020, 870-675
C-H	Aromatic	3100-3000, 870-675
C-H	Alkynes	3300
C=C	Alkenes	1680-1640
C=C	Aromatic	1600-1500
C-O	Alcohols, Ethers, Carboxylic Acids, Esters	1300-1080
C=O	Aldehydes, Ketones, Carboxylic Acids, Esters	1760-1690
O-H	Alcohol, Phenol (Monomer)	3640-3610
O-H	Alcohol, Phenol ('H' Bond)	3600-2000
O-H	Carboxylic Acid	3600-3000
N-H	Amine	3500-3310
C-N	Amine	1360-1180
NO ₂	Nitro	1560-1515, 1386-1345

(Sastrohamidjojo, 1992)

Generally, bovine gelatin and porcine gelatin have absorption peaks at almost identical wave numbers. However, if we look carefully at the curves, there are some differences between the bovine and porcine gelatins. In **Figure 2.8**, the spectrum showed that bovine gelatin have relatively higher peaks compared to the spectrum of porcine gelatin as shows in amide I and II regions ($1656\text{-}1644\text{ cm}^{-1}$ and $1560\text{-}1335\text{ cm}^{-1}$). The region between $3290\text{-}3280\text{ cm}^{-1}$ is related to the N-H bond stretching and intramolecular hydrogen bonds of amine group in the amino acid chains. The absorption corresponded to the N-H bonds, showing the interactions of hydrogen bonds in the alpha helical structures of the gelatin. The resulted peaks can be shifted to lower frequencies when the hydrogen bonds strength increases (Hafidz et al., 2011).

FIGURE 2.8: FTIR Spectra for (1) Bovine Gelatin and (2) Porcine Gelatin (Fathimah, 2010)



The stretching of the carbonyl (C=O) group, appeared in the area between 1660 to 1620 cm^{-1} which referred to the existence of amide I. Peaks in the range between 1660- 1650 cm^{-1} represents the alpha helical structures and 1640-1620 cm^{-1} represents as the beta-sheet structure. The existence of amide II is depicted by peaks in region between 1550-1520 cm^{-1} which corresponded to the NH bond deformation in the structure of alpha helices (1550-1540 cm^{-1}) and the beta-sheet structures (1525-1520 cm^{-1}) (Fischer et al., 2005). The peaks at 1500-1200 cm^{-1} are a representation of the CH_2 . This is related to the multiple hydrocarbon groups contained in macromolecular compounds such as fatty acids, proteins and polysaccharides (**Figure 2.8**) (Fatimah, 2013).

2.4.4 Chemometrics Analysis using Unscrambler

The function of the software is to assist in analyzing multivariate data and form a design of the experiment. One abilities of Unscrambler is to classify unknown samples systematically.

Classification aims to find new sample similar to the categorization of samples that have been used to create the model. If the new sample is in accordance with a model that has been created, it can be seen that sample category (Citrasari, 2015). Chemometrics is multidisciplinary involving multivariate statistical mathematical modeling and information technology. Multivariate analysis is statistical analysis that used on the data that consists of many variables, and between the correlated variables. Some of the method included in the analysis group is the principal component analysis (Rohman et al., 2012)

Principal component analysis (PCA) is a technique for building new variables that are linear combinations of the original variables. The maximum number of new variables will be equal to the number of old variable and each of them is not correlated. The Excess PCA which can eliminate the correlation, not reduce the number of original variables and more accurate than the use of other methods (Kautsar, 2012). The working principle of PCA can be seen in **Figure 2.9**.

FIGURE 2.9: PCA Principle (Kautsar, 2012)



The principle of PCA is to find the main component that is a linear combination of the original variables. The selected of the major components become the first principal component with the greatest variance in the data group, while the second main component perpendicular to the first principal component and have the next largest variant. PCA technique works on the data matrix X ($J \times I$) into two matrices T ($I \times A$) and matrix P ($J \times A$) that perpendicular to each other (**Figure 2.9**) (Kautsar, 2012) T matrix called the matrix score that describes the variation in the object, while the loading matrix describes the effect of loading on the main components. Loading matrix composed of the original data in the new coordinate system (Nurcahyo, 2015).

Error of the model forms expressed in E , while the value of A is the number of PCs that are used to create the model (Kautsar, 2012). In matrix representation, the model with a given number of components has the following equation:

$$X = TP^T + E \quad (2.1)$$

Where T is the scores matrix, P the loadings matrix and E the error matrix. The combination of scores and loadings is the structured part of the data; the part that is most informative, what remains is called error or residual, and represents the fraction of variation that cannot be modeled well. Most of the chemometrics publication is using the same terminology to discuss the data of PCA. From the unscrambler, it can be concluded that 4 terminologies can be explained namely; Scores plots, loading plots, explained variation and influence (Kong et al., 2007).

1. Scores plot

Scores also called component scores in PCA; are the scores of each sample (row) on each factor (column). To compute the factor score for a given case for a given factor, one takes the case's standardized score on each variable, multiplies by the corresponding factor loading of the variable for the given factor, and sums these products. Scores are estimated in bilinear modeling methods where information carried by several variables is concentrated onto a few underlying variables. Each sample has a score along each model component. The scores show the locations of the samples along each model component, and can be used to detect sample patterns, groupings, similarities or differences (Kong et al., 2007).

2. Loading plot

The plot shows the importance of the different variables for the two components specified. Variable with loadings to the right in the loadings plot will be variables which actually have high values for samples to the right in the score plot. Variables close to each other in the loadings plot will have a high positive correlation if the two components explain a large portion of the variance of X . The same is true for variables in the same quadrant lying close to a straight line through the origin. Variables in diagonally opposed quadrant will have a tendency to be negatively correlated.

In geometrical terms, a loading is the cosine of the angle between the variable and the current PC: the smaller the angle (i.e. the higher the link between variable and PC), the larger the loading. It also follows that loadings can range between -1 and $+1$. The correlation r between two variables (vectors), x and y , is defined as,

$$r(x,y) = \frac{\text{Cov}(x,y)}{s_x s_y} \tag{2.2}$$

Where, Cov is the covariance between x and y (Kong et al., 2007).

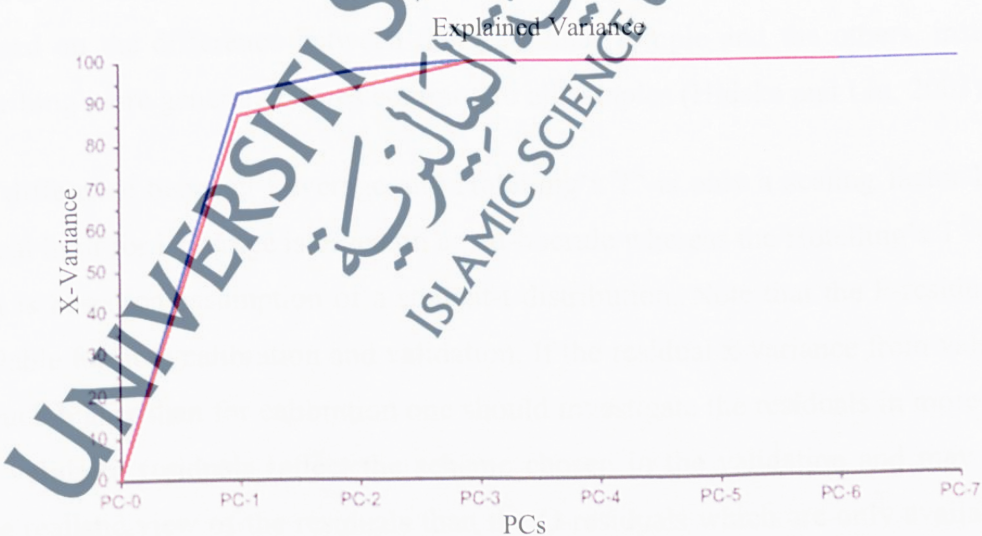
3. Explained Variances

The plot gives an indication of how much of the variation in the data is described by the different components. Total residual variants are computed as the sum of squares of the residuals for all the variables, divided by the number of degrees of freedom. Total explained variance is then computed as:

$$100 \times (\text{Initial Variance} - \text{Residual Variance}) / (\text{Initial Variance})$$

It is the percentage of the original variance in the data that is taken into account by the model.

FIGURE 2.10: Total Explained Variance Curve (Kong et al., 2007)



Calibration variance is based on fitting the calibration data to the model. Validation variance is computed by testing the model on data that were not used to build the model. Based on the **Figure 2.10**, if they differ significantly, there is good reason to question whether either the calibration data or the test data are truly representative. The figure shows a situation where the explained validation variance is much similar than the explained calibration variance. This means that the calibration data are well fitted and the models also describe the new data well. But if the explained variance are not close together the model does not represent the data (Kong et al., 2007).

4. Influence

The plot shows the Q-residual, X-variance or F-residual vs. Leverage or Hotelling's T^2 . It is most useful for detecting outliers, influential samples and dangerous outliers. Samples with high residual variance, i.e. lying to the top of the plot, are likely outliers. Sample with high leverage, i.e. lying to the right of the plot, are influential. This means that they attract the model so that it describes them better. Influential samples are not necessarily dangerous, if they obey the same model as more "average" samples. A sample with both high residual and high leverage is dangerous outlier (Syahariza et al., 2005).

It is not well described by a model which correctly describes most samples, and it distorts the model so as to be better described, which means that the model then focused on the difference between that particular sample and the others, instead of describing more general features common to all samples (Hidaka and Liu, 2003).

The difference between Leverage and Hotelling's T^2 is only a scaling factor but the critical limit for Leverage is based on an ad-hoc rule whereas the Hotelling's T^2 critical limit is based on assumption of a student-t distribution. Note that the F-residuals are available for both calibration and validation. If the residual x-variance from validation is much higher than for calibration one should investigate the residuals in more detail. The validated residuals reflect the scheme chosen in the validation and may give a more realistic view of the residuals than the Q-residuals which are only available for calibration (Kong et al., 2007).

2.5 Parameters for Analytical Methods Validation

Method validation is the practical process of determining the suitability of a method for providing analytical data that is fit for the intended purpose. Method validation is defined in ISO/IEC 17025:2005 as:

'confirmation by examination and the provision of objective evidence that the particular requirements for a specific intended use are fulfilled'

For any method to produce meaningful and reliable data, some performance checks should be made before the method is applied to real samples. The validation process typically entails firstly understanding the reason the measurement are being made and the performance of the method that is required to produce data that are fit for that intended purpose. Secondly experiments are planned and performed to evaluate the performance of the method. The observed performance is then compared with the required performance of the method, and relevant/specified criteria are used to determine whether the performance is adequate. The actual level of assessment and validation that is undertaken will depend on the intended use of the method and the importance of the data produced (Ali et al., 2012).

There are many performance characteristics that can potentially be investigated for a particular method, some of which are listed in Table 2.3, and described in more detail in the following sections. Various performance parameters are important depending on the type of measurement being made and the reasons for making the measurements, so choosing the characteristics to be investigated is a crucial part of the validation process. Accuracy (comprising both precision and bias) may be important for calculating absolute values of properties or analytes, for example, whereas precision is more significant in comparative studies. Random errors in the method are reflected in the precision of the results, whereas systematic errors (such as out-of-calibration instrumentation or consistently low recovery rates during sample preparation) give rise to method bias. Working range will be of some interest in most cases. For trace work, limits of detection (LoD) and quantification (LoQ) may be relevant, but for planning calibration strategies it may be more useful to know the range over which the method response is linear (Watson et al., 1987).

TABLE 2.3: Parameters and Approaches to Consider in Assessing qPCR Performance

Performance characteristic	qPCR performance	Experimental procedure
Dynamic range	Range of sample concentrations over which the assays remains linear	Dilution series from a known concentration DNA analyzed
Repeatability	Variability of result under closely controlled conditions	Same measurement on same sample repeated by same analyst
Reproducibility	Variability of result under differing conditions	Same sample measurement repeated by different analysis using different instruments, different laboratories or over time
Bias	Consistent over-or underestimation of the true result	Average measured value of a reference material compared to the assigned value
Specificity	Ability of the assay to detect the target but not other potential analytes present in the sample	Assay performed with a variety of related targets to check for false positive signals
Sensitivity (LoD / LoQ)	The lowest amount of the target that is detectable/reliably quantifiable	Assay performed with increasing dilutions of analyte to determine the limit of detection/ linearity

(ISO/IEC, 2005)

2.5.1 Selectivity (Sensitivity)

The selectivity or specificity is the ability to measure only certain substances carefully and thoroughly with the other components that may be present in the sample matrix. Selectivity can often be expressed as the degree of bias. The selectivity of the method was determined by comparing the results of the analysis of samples containing contamination, the result of dust, similar compounds, other foreign compounds or carrier placebo with sample analysis results without the addition of these materials (Vessman et al., 2001). Selectivity and specificity are often used synonymously; specificity can be termed the ultimate in selectivity as if a process is specific it is by definition wholly selective. The selectivity of an analytical method may be affected by many factors, including the presence of impurities, degraded components and possible inhibitors or enhancers of the reaction, and physical parameters such as temperature, ionic strength or pH (Brown, 2005).

Poor selectivity of a reaction indicates that other substances can interfere with the analysis, and selectivity may change as assay conditions are altered. For example, a PCR at the optimized annealing temperature will produce a single, specific amplification product but, if the annealing temperature is changed and the reaction made less stringent, multiple non-specific products may be generated. In clinical diagnostics the selectivity of an assay is often reported, and is calculated as the percentage of true negative results obtained in testing a number of known negative samples (Sambrook et al., 1989).

2.5.2 Limit of Detection (Sensitivity)

It should be noted that the biological definition of sensitivity given here varies from the ISO chemical definition. Detection limit terminology is inconsistent and confusing, so it is important to try and follow specific sector guidelines where available. In most biological analysis, sensitivity is used to describe the lowest level of an analyte that can be measured. However, in fields outside of biological analysis the term 'sensitivity' has different definitions (ISO, 2005).

For example, in chemical analysis, sensitivity is usually defined as 'the change in the response of a measuring instrument divided by the corresponding change in the stimulus (ISO, 1993). In clinical applications sensitivity again has a slightly different meaning, and is often expressed as the percentage of tests that give the correct positive result in testing a number of known positives. Often two limits are defined for a quantitative assay, firstly the limit of detection (LoD), which is the lowest amount of a target which can be reliably detected and distinguished from zero results and background, signals with confidence. The second is the limit of quantification (LoQ), which is the lowest concentration of analyte that can be quantitatively measured with an acceptable level of uncertainty (Lovatt, 2002).

Practically there are several ways of determining the sensitivity of a method. In qualitative analysis the analyte is typically diluted serially until it can no longer be detected reliably using the method (usually once the percentage detection falls below a specific level, often 95 or 100%) (Drosten et al., 2001).

A variety of approaches is used in quantitative determinations, but most approaches use the result of repeated analysis of a negative sample or zero calibrator (a sample known not to contain the analyte of interest). The zero calibrator is analysed between 10 and 20 times, and the mean and standard deviation of the data obtained. Usually the limit of detecting, or analytical sensitivity, is set as the mean signal +2 standard deviations while the limit of quantification is set as the mean +10 standard deviation. In quantitative PCR analysis, for example, where negative result do not yield a meaningful value, the definition of LoD and LoQ is more difficult. One developed approach is to define the LoD as the input analyte level giving a (for example) 95% probability of a positive PCR result, calculated using probit regression analysis of dilution series data (Drosten et al., 2001). Sensitivity can be expressed in many ways depending on the assays, for example:

1. Number of cells per mass of matrix that is detectable;
2. Percentage of adulterant in a matrix that is detectable;
3. Mass of DNA required for reliable qualitative analysis (such as STR profiling);
4. Amount or copy number of gene, genome or DNA target that is detectable, per volume or mass (Madej, 1991).

Sensitivity should be assessed across the range of analyte levels and sample masses that may be routinely used, as the ratio of detectable target within a given sample cannot always be linearly extrapolated. For example if 100 microbial cells can be detected in 1 gram of soil by PCR, 10 cells may not be detected if only 0.1 g of soil is tested. Sensitivity is often determined an indication of the lower operating limits of the test method such as a replicated sequence of experiments on low level samples, blanks and low-level spiked materials or standards. Therefore, the results of an analysis are often close to the lower operating limit to perform regular assessment of the LoD. This ensures that negative results are being interpreted and reported correctly with reference the detection or quantification limit of the method (Persing and Tenover, 2004).

2.5.3 Efficiency

In kinetic real time PCR each individual reaction position is excited and detected independently based on the fluorescence history. The former shapes of amplification curves differ in the steepness of any increasing fluorescence and in the absolute fluorescence levels at plateau depending on background fluorescence levels. The PCR efficiency has a major impact on the fluorescence history and the accuracy of the calculated expression result and is critically influenced by PCR reaction components. Efficiency evaluation is an essential marker in real-time gene quantification procedure (Liu and Saint, 2002a; 2002b; Rasmussen et al., 2001; Tichopad et al., 2003a; 2003b; 2004). The condition of amplification efficiency in all compared samples must be in constant state because it is one of crucial factors for reliable comparison between samples. The factors also become crucially important from analyzes the relationship between an unknown sample (DNA) and a standard sample, which is performed in all relative quantification models (Atlas et al., 2006).

The threshold method is most commonly used for the quantification of unknowns, and as illustrated in **Figure 2.6** utilizes information from the points at which known DNA standards reach a specified fluorescence threshold to construct a standard curve of crossing threshold against target level. As described, the reaction efficiency is determined from the slope of the standard curve (**Figure 2.6**). The threshold level should be set in the exponential phase of the amplification, and most instruments calculate an optimal level setting. Manually setting the level is possible, but is subjective and may also introduce variability between runs of the assay. Advantages are that the method is simple, and the quality of the assay may be monitored using the parameters of the standard curve. Disadvantages are that the dilution series used to construct the standard curve is prone to errors, and the assumption that the reaction efficiency is a constant in the exponential phase of the reaction is not always valid (Rasmussen et al., 2001).

To overcome the problems in using dilution series, alternative methods based on estimating the amplification efficiency from single reactions have been developed (Wittwer and Garling, 1991).

The rate of change of fluorescent signal within a single reaction may be monitored, ideally within the linear phase of signal increase, to determine the efficiency of each reaction. The second derivative maximum option in the Applied Biosystem StepOnePlus software similarly calculate the maximum rate of change of the signal in the reaction, and utilizes the peak to determine the fluorescence at the maxima, and hence the initial number of copies in the reaction (Wittwer and Garling, 1991). The PCR process generates anywhere between an average of 0 to 1 copy of each target in each reaction cycle, so that for any cycle the number of molecule is:

$$N_c = N_{c0} \times (E + 1)^c \tag{2.3}$$

Where N_c is the number of molecules at cycle N , N_{c0} is the number of molecules at cycle 0, E is efficiency of the reaction and C is the cycle number. Making assumptions that the efficiency of the amplification is constant in the early exponential stages of the reaction, and that all standards possess the same number target molecules at the point at which their signal crosses the determined threshold level, then the equation may be simplified to allow determination of the reaction efficiency (Birren et al., 1997),

$$E_s = \left(\left(10^{(-1/-slope)} \right) - 1 \right) \tag{2.4}$$

The number of molecules at the threshold point, N_t , can also be determined from the standard curve (Muller et al., 2002).

$$N_t = 10^{intercept} \tag{2.5}$$