CHAPTER 1
GENERAL INTRODUCTION

Global food colouring consumption reached as high as 7.6 million kilograms just in the 3rd Quarter of Fiscal Year 2011, from 1 October 2010 until 30 June 2011 (FDA) and its production is estimated to increase by another 20% in the next fiscal year. The food colouring industry is a vital element in effectively marketing food and beverage products globally. However there are concerns that synthetic colouring may be linked to the onset of behaviour disorder in children and other diseases (Silbergeld and Anderson, 1982).

The many factors which makes synthetic colouring the ultimate choice in food production is that it is cheaper, easier to make, longer shelf-life and easily applied (Downham and Collins, 2000). This has caused the lack of effort and drive towards investing into research in natural colourant source for better and safer food products for a long time. There is also a constant struggle and debate between balancing the economics of food and health, where natural colouring is very much more expensive to manufacture and there is a need to find a consistent and reliable source to the natural colourant. For example: Corcin pigment from saffron and gardenia fruit is an expensive source of caretenoid compared to synthetic FD&C Yellow No. 6 Sunset Yellow FCF; Carminic acid from female cochineal insect are definitely price sensitive because approximately 70,000 handpicked insects are needed to produce 500 g of 50% strength pigment; and lycopene pigments from tomatoes has poor stability and high cost is incurred to stabilize the pigment during commercialization (Downham and Collins, 2000). This is where biotechnology coupled with biochemistry understanding provides a head start and platform to venture into natural colourant and to explore potential sources as starting material.

According to Wissgott and Bortlik (1996), 70% of all the plants worldwide have not been explored and only 0.5% of plants have been thoroughly investigated. These numbers indicate that the journey and exploration for natural food colour source is only at its infancy and with the assistance of better technologies, there is a huge potential in this industry.

As early as 1982, Silbergeld and Anderson reported that the increasing usage in synthetic food colouring has proven to be detrimental to children and has been linked to hyperactivity, hyperkinesis, minimal brain (or cerebral) dysfunction or attention deficit disorder like in the artificial color tartrazine (FD&C Yellow No. 5).

Wide usage of synthetic food colouring has caused deteriorating health quality like sinus, asthma, sensitivity and allergies. These concerns and studies have led to the removal and withdrawal of Red nr 40 in many European countries and Green nr 3 is illegal throughout the European Union. Additionally, E131 (Patent Blue V) and E142239 (Green S), are coal-tar colours banned in the United States. This also started the hostility against synthetic colouring amongst health conscious consumers and thus creating a niche market of natural colourants.

There are many arguments for and against using existing food crop as natural food colouring source and utilizing them for downstream applications rather than as food source. One of the major concerns is the quality of the natural colourant extracted from the source. Even though the pigment is of natural and plant origin, there are still safety issues which must be considered as contaminants and naturally occurring toxic primary and secondary metabolites can be dangerous to consumers.

For example, the red pigment betanin from *Beta vulgaris* (beetroot) which is widely used for food colouring has been linked to contain nitrosamine and geosamine, both cancer causing agents (Esquivel *et al*., 2007), from its interaction with the soil environment during planting. This leaves the food and beverage industry looking for suitable alternative with the same pigment but offers safer attributes.

There is an increased awareness in healthcare, accompanied by a revolution which is changing the way consumers choose their food intakes and steered towards healthier options with nutritional and protective elements (Nicoli *et al*., 1999).

In the wave of this revolution, the interest and search for antioxidant compounds in plants (Connor *et al*., 2002) has always been an important agenda. Antioxidants are usually found in plant phytochemicals such as flavonoids (Asplund, 2002), polyphenols, pigments and vitamins (Lako *et al*., 2007). Chemically, antioxidants are radical scavenging species (Aruoma, 1998) which have the capacity to terminate the chain reaction of highly unstable free-radicals in living systems (Lee and Gould, 2002).

If free radicals like reactive oxygen species such as hydroxyl (OH•), peroxyl radicals (ROO•) and the superoxide anion (O2•) exceeds the antioxidant capacity in organisms, the radicals will initiate attacks on lipids, proteins, and DNA, thus damaging structural integrity and function of cell membranes, enzymes, and genetic material causing cardiovascular disease, arthritis, various cancers, and Alzheimer's disease (Wang *et al*., 1999).

Combining all the factors above: Dangers of synthetic food additives; demands from health conscious consumers; the food and beverage industry's intention to satisfy the consumers and the huge potential in plant phytochemicals, this thesis looked to the red dragon fruit (*Hylocereus polyrhizus*) as a prospective candidate to fulfill the criterions as mentioned.

Hylocereus polyrhizus is a member of the Cactaceae family (Elobeidy, 2004) and one of the 13 families in the plant kingdom which are made up of betacyanins instead of anthocyanins as their pigments. *Hylocereus* is the more famous member in the family compared to *Stenocereus*, *Cereus* and *Selenicereus* because they have higher value in terms of aesthetic and economic reasons such as high antioxidant properties (Wybraniec and Mizrahi, 2002); high in protein and minerals content (Wu and Chen, 1997); and rich in polysaccharides (Ramírez-Truque *et al*., 2011).

Plant biochemistry has always been the forerunner in understanding the postharvest behaviour, nutritional characteristics and molecular approaches towards one crop. The knowledge on plant biochemistry has widen into secondary metabolites as early as 1891 (Kossel, 1891) where the term secondary is used because different sets of metabolite occur in different plant species. This critical differentiation between primary and secondary metabolite has led to extensive chemotaxonomic and phylogenetic studies (Harborne, 1988); ecophysiological traits, genetics and developmental studies (Ackerly *et al*., 2000) in plants. In a study by Markham (1990), he used the distribution of flavonoids to chemotaxonomically characterize bryophytes and lower non-seed-bearing vascular plants.

Ackerly *et al*., (2000) further elaborated that secondary metabolites can influence resource uptake and utilization, gas exchange, leaf structure and function, nutrient and biomass allocation, reproduction, defense and stress tolerance where all these are important factors in determining adaptive evolution in plants. Some examples of economically important secondary metabolites are: Alkaloids, glycosides, flavonoids, volatile oils, tannins and resins. Hence, the knowledge of the pathway, molecular behaviour, genetic control and expression in the regulation of secondary metabolite production in plants is of high importance.

To obtain a pure sample or to remove impurities entirely, crystallization is the choice method. Crystals are grown by bringing a choice solution into a supersaturated state and progressively return the solution's environment to equilibrium, the solute molecules are forced into a solid state, forming crystals (McPherson, 1990).

Crystals are then studied using the X-ray crystallography which principle was spearheaded by Max von Laue in 1912 (Ewald, 1960) where the method lends to a better understanding of chemical bonds and interactions. As early as 1913, Bragg and Bragg (1913) reported the structure of diamond using X-ray crystallography, followed by reports on organic crystals (Bragg, 1921) and non-organic crystals (Wyckoff and Posnjak, 1921).

Further, the novel discovery of the DNA double helix structure by Watson and Crick (1953) was assisted by the earlier established knowledge by Franklin and Gosling (1953) where they used X-ray diffractions to study the pure crystalline A-form and B-form DNA. Since then, crystallization of protein, secondary metabolites, carbohydrates and even virus structures were carried out.

These crystals make detection and identification of specific substance easier and through the X-ray analysis, massive amount of data and structural characterization are obtained.

With the increase in air, land and water pollution levels, there is also an increased concern if the contaminants are making its way into the food chain and ultimately affecting human health. Food products in the market are usually monitored against excessive microorganism contamination, heavy metal accumulation levels and pesticides concentration. Some example of major food contamination cases are: Milk contaminated by *Salmonella typhimurium* in USA (1985) affecting 170,000 people; Grapes contaminated with cyanide in Chile (1989) and an estimated economic cost of $100 million; Apple juice contaminated with *Escherichia coli* O157:H7 in USA (1996) causing $14 million economic loses and 1 fatality (Manning *et al*., 2005). The prevalent nature of these contaminants after a long chain of processes makes toxicology studies very essential in ensuring the safety of consumers.

According to Astbury (1961), molecular biology is an approach to understand basic sciences or classical biology by uncovering the corresponding molecular plan where the three-dimensional and structural forms of biological molecules are concerned while at the same time discovering its genesis and function in an organism. Molecular biology has since then been incorporated with computer science resulting in abundant bioinformatics data encompassing studies in gene structures and functions, taxonomy, phylogenetics and molecular genetics. These advancements and vast molecular/genetic level knowledge is vital in supporting subsequent studies in organisms and future experiments are designed to be more targeted, faster and save cost.

This thesis reports on studies to isolate, purify and characterize pigments extracted from *Hylocereus polyrhizus* (red dragon fruit) to establish its potential as a natural red dye source. The *Hylocereus polyrhizus* is gaining extensive attention and increased cultivation locally due to its exotic appearance and host of numerous nutritional values. A detailed review of the current status of *Hylocereus polyrhizus*, its taxonomic origin, pigment constituency and some studies already carried out on *Hylocereus polyrhizus* is described in Chapter 2 before proceeding with the experimental chapters.

The objectives in this thesis are:

1. To study the optimum extraction method, stability of pigments and viability in different storage (Chapter 3). Pigments were extracted using different extraction methods and different temperatures to establish the optimum extraction conditions.

2. To investigate antioxidant properties in pigments (Chapter 4) using four selected assays: Total polyphenolic content determination, reducing power assay, vanillin-HCl assay and $DPPH^{\bullet}$ radical scavenging activity determination.

3. To identify crystals isolated (Chapter 5) using four different techniques to resolve its structure and identity: X-Ray Crystallography, High Performance Liquid Chromatography (HPLC), Liquid Chromatography Tandem Mass Spectrometry (LC-MS/MS) and Nuclear Magenetic Resonance (NMR).

4. To establish the toxicology profile on extracted pigments for safety assessment (Chapter 6) through microorganism analysis, heavy metal analysis and pesticide screening.

5. To isolate and characterise selected chemotaxonomically important and pigment related genes (Chapter 7).

The results show that the extracted pigments exhibit considerable shelf life in different extraction methods and storage conditions. The pigments also exhibited high antioxidant capacity compared to many common economically important food crops. The crystals obtained in this study were successfully identified as *myo*-inositol in Chapter 5 and this is the first time it is being reported in *Hylocereus polyrhizus*. The toxicology analysis indicates that the pigments are safe for utilization in the food and beverage industry. The final experimental Chapter 7 cloned a chemotaxonomically important gene in the betacyanin bearing families. Bioinformatic tools were employed to study the genes in detail.

CHAPTER 2
LITERATURE REVIEW

2.1 INTRODUCTION TO DRAGON FRUIT

Dragon fruits are members of the Cactaceae family, one of the largest families of succulents which are commonly known as cactus plants (Elobeidy, 2004). Members of the Cactaceae family are mainly appreciated for their ornamental qualities but there are at least 250 cultivated species of fruit-bearing and industrial crop in this drought resistant family (Le Bellec *et al*., 2006).

The dragon fruit is known by many names which differ within local communities, from country to country. For example, in Chinese, dragon fruit is called Zunlongguo; in English, it is commonly known as strawberry pear or red pitaya; in Indonesia and Malaysia, it is called buah naga; in Israel, the fruit is referred to as pitaya; in Hawaii, the fruit is fondly known as päniniokapunahou or päpipi pua; and in Vietnam, it is locally known as thanh long (Gunasena *et al*., 2007; Zee *et al*., 2004). All the vernacular names given to dragon fruit translate as 'snake scales' which refers to the complex vine branching system (Figure 2.1), scaly and thorny appearance of the entire plant (Figure 2.2).

The generic term dragon fruit or pitaya includes several different species grouped into four main genera: *Stenocereus* Britton & Rose, *Cereus* Mill., *Selenicereus* (A.Berger) Riccob and *Hylocereus* Britton & Rose (Le Bellec *et al*., 2006; Mizrahi *et al*., 1997). Currently, the genus *Hylocereus* is the preferred choice for cultivation because its members are highly valued for aesthetic characteristics and economic reasons (Merten, 2004).

The fruits of *Hylocereus polyrhizus* the most popular red pitaya and member of the genus has been widely propagated, not only for its attractive red-purple pulp and trade value as food product, but also for its value added antioxidative properties (Wybraniec and Mizrahi, 2002).

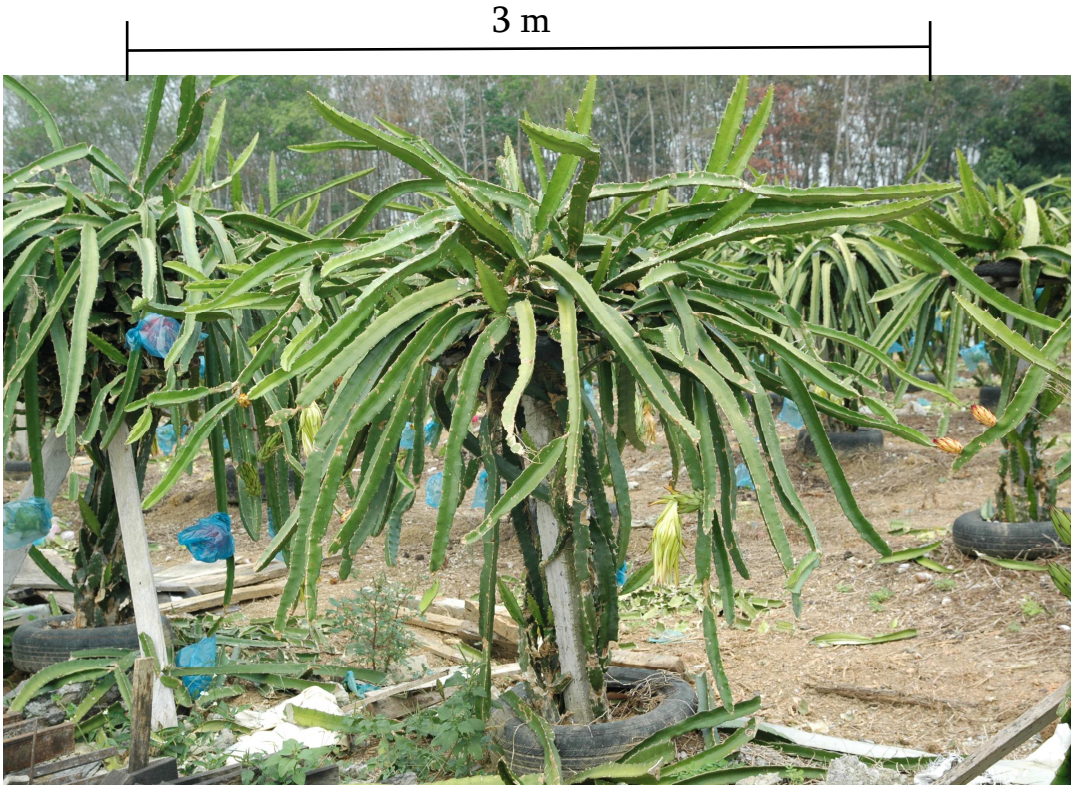


Figure 2.1: *Hylocereus polyrhizus* plant with complex vine branching system

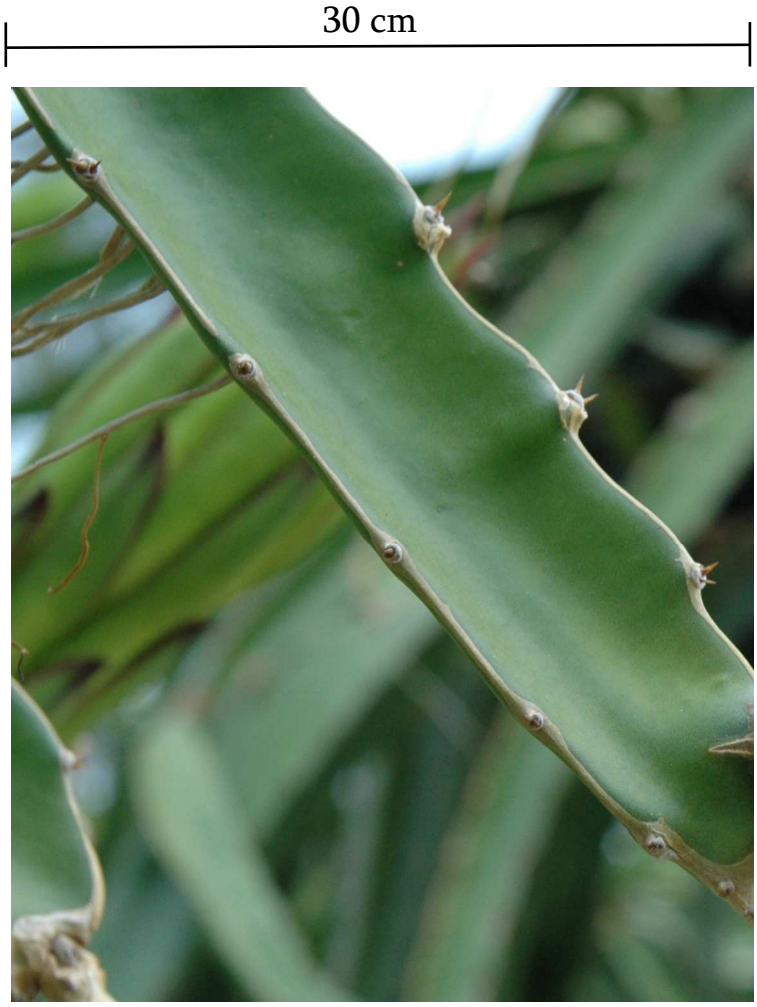


Figure 2.2: Scaly and thorny appearance on every branch of the entire *Hylocereus polyrhizus* plant

2.3 BOTANICAL DESCRIPTION OF *HYLOCEREUS POLYRHIZUS*

2.3.1 Origin

The name pitaya which means scaly fruit comes from the local lingo in the region of Antilles and the ancient Mexicans called the fruit Coapetilla which means thick serpent, referring to the appearance of the cactus plant stems (Eulogio *et al*., 1997). The plant which is commonly cultivated throughout the tropical lowlands is native to southern Mexico, the Pacific side of Guatemala and Costa Rica, and El Salvador which are all countries within the Latin American region (Blench, 2007; Zee *et al*., 2004; Stintzing *et al*., 2002b).

2.3.2 Taxonomy

The genus *Hylocereus* (Berger) Britton & Rose consists of about 16 tropical species, whose ornamental value is due to their exquisite large flowers and all members are vine (climbing with aerial roots or epiphytic) cacti (Le Bellec *et al*., 2006). The nomenclature of dragon fruit is as follows (Gunasena *et al*., 2007; Britton and Rose, 1963):

Kingdom Plantae

Sub Kingdom Tracheobionta (Vascular plants)

Super division Spermatophyta (Seed plants)

Division Magnoliophyta (Flowering plants)

Class Magnoliopsida (Dicotyledons)

Order Caryophyllales

Family Cactaceae (Cactus family)

Subfamily Cactoideae

Tribe Hylocereae

Genus *Hylocereus* (Berger) Britt & Rose

Species *Hylocereus polyrhizus* (Haw) Britt &Rose

2.3.3 The Plant

The *Hylocereus polyrhizus* is a fast growing, perennial, epiphytic cacti plant which fruits only two to three years after planting reaches its full production capacity after five years. The many-branched, green, fleshy and triangular stems can grow up to 5m within a year and a mature plant can weigh more than 100kg and consists up to 130 branches (Gunasena *et al*., 2007; Crane and Balerdi, 2005; Elobeidy, 2004).

Aerial roots grow from the corneous margins with 1-3 spines, underside of the stems which enable the plant to adhere and climb around the trellis. The reason why the plant is highly tolerant to drought and dry habitats is because the stomata are sunken in the epidermis, the stem is rich with parenchyma and they have a wax layer, are all which evolved physiological and morphological adaptation to prevent water loss (Merten, 2004; Mizrahi and Nerd, 1999). These features make the *Hylocereus* plant robust and easily grown in multifaceted land conditions.

2.3.4 The Flower

The flowers of the *Hylocereus* plant are nocturnal and are considered very attractive within the Cactaceae family, thus the nicknames, “Noble Woman” and “Queen of the Night” (Gunasena *et al*., 2007) (Figure 2.3). The flowers are large, up to 30cm across in diameter; hermaphrodite; bell-shaped; white in colour; very fragrant; and the stamen and lobed stigmas are cream coloured (Wu *et al*., 2006; Tel-Zur *et al*., 2005). The night blooming flowers which only open for one night start to unravel its petals late in the evening and a full bloom occurs around 10pm.

If the natural pollinators, like bats and hawk moths successfully pollinate the flower, the flower closes at about 2am, begins to wilt and petals are closed completely by dawn (Merten, 2004). Based on the behaviour of the flower, a guaranteed pollination can only be achieved by manual hand pollination especially in countries where they lack natural pollinators. Furthermore, there are only two to three waves of flowering season in a year and the plant is known to be a long day plant.

Thus, to ensure fruit set and a profitable yield, human intervention which include introducing artificial lighting and manual pollination is necessary (Mizrahi *et al*., 2004; Zee *et al*., 2004). One advantage that *Hylocereus polyrhizus* has compared to the other members of the family is that the plant is not over reliant on the day length and it is able to continuously produce fruit throughout the year (Luders and McMahon, 2006).

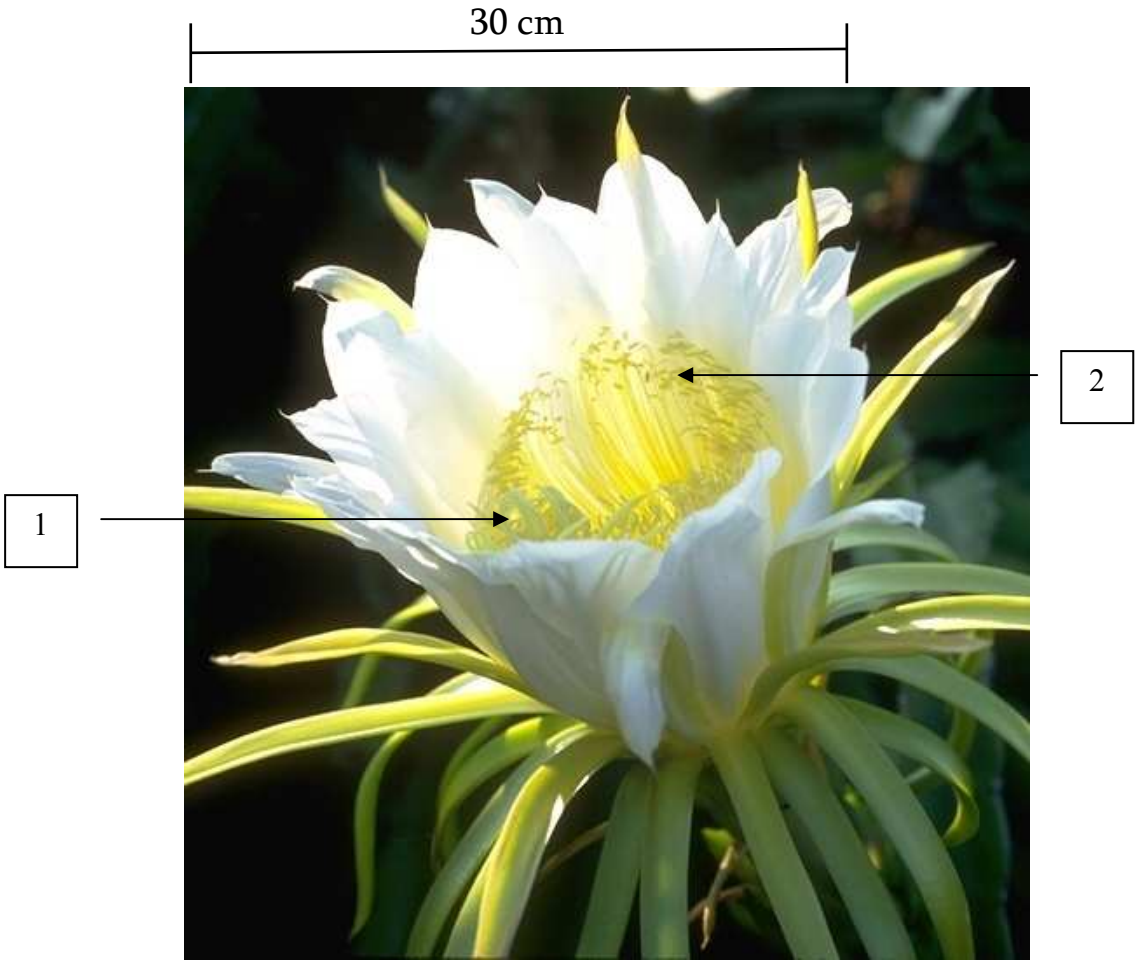


Figure 2.3: A *Hylocereus polyrhizus* flower
1 – Anther; 2 – Stigma

2.3.5

2.3.6 The Fruit

The *Hylocereus polyrhizus* fruit which is popularly produced commercially for its glowing deep red-purple flesh is an oblong shaped red skin berry (about 10-12cm in length); with large scales; and weighs between 130-350g (Gunasena *et al*., 2007; Wybraniec *et al*., 2007; Mizrahi and Nerd, 1999) (Figure 2.3). The pulp which represents 60-80% of the mature fruit weight has a pleasant texture with good taste and contains numerous small, edible black seeds (Crane and Balerdi, 2005).

The fruit is non-climateric and the harvesting time is crucial to obtain a good quality fruit because the physico-chemical characteristics and sensory properties start to decline after harvest. It has been generally accepted that the best time to harvest is 30-50 days after fruit set (pollination) or when 85% of the skin has turned pink to obtain a good quality yield and an extended shelf life (Le Bellec *et al*., 2006; Luders and McMahon, 2006; Zee *et al*., 2004). For example, fruits harvested 1–2 days before full coloration of the skin (40 days after pollination) can be preserved for up to 1 month at 10°C without any observed decrease in quality and if the fruit is harvested 5–8 days after full coloration (longer than 50 days after pollination) the fruit becomes sweeter with a blackberry flavour and gets heavier but it will experience a severely decreased shelf life.

A typical mature fruit's total soluable solids content ranges between 7-11 gL^{-1} which mainly consist of reducing sugars and more specifically glucose and fructose. The acidity of the fruit pulp is generally low and the main organic acids present are citric acid and L-lactic acid (Stintzing *et al*., 2003). Protein content varies from 0.3-1.5 percent where the main amino acid appears to be proline (1.1 to 1.6 gL^{-1}) (Le Bellec *et al*., 2006). This significant high amount of protein in cactus fruits, ranging between 0.16g to 1.25g per 100g of pulp, depending on species, could possibly be due to the presence of betalain, the nitrogen-containing pigment which contributes to the deep purple-red coloured pulp (Crane & Balerdi, 2005; Vaillant *et al*., 2005; Stintzing *et al*., 2003; Stintzing *et al*., 2001).

This is a direct comparison to protein in some common fruits like lemon (0.9g/100g); orange (0.8g/100g); grapes (0.7g/100g); jambu (0.6g/100g); and pineapple (0.5g/100g) (Paul and Shaha, 2004). Other minerals include potassium, magnesium, calcium and vitamin C. The dragon fruit has been reported by Vaillant *et al*., (2005) to have high antiradical activities and contains other phenolic compounds but characterization has not been reported yet. The polysaccharides which represent one percent of the mesocarp weight are responsible for the mucilaginous texture and they are mainly hemicelluloses; cellulose and pectin.

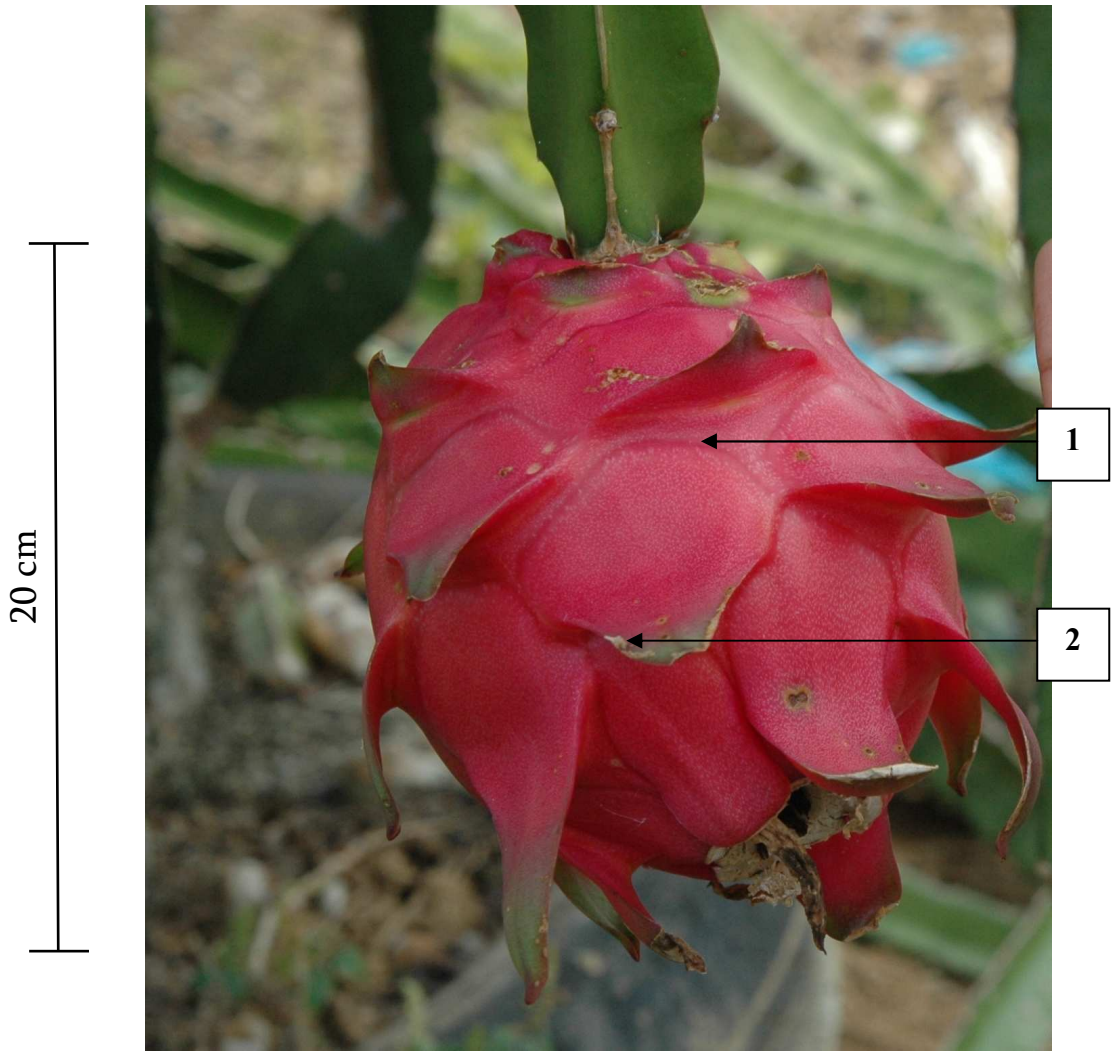


Figure 2.4: A *Hylocereus polyrhizus* fruit
1 – Peel (Skin); 2 – Bract

2.3 ECONOMIC IMPORTANCE

Although the dragon fruit is native to the tropical forest regions of Mexico and the Latin America region, it is now widely cultivated in all continents except Antartica (Moβhammer *et al*., 2006a). The fruit has been cultivated in Vietnam for almost 100 years but was never perceived as a commercially profitable crop because it was grown exclusively for the royal family coupled with the lack of attention from the scientists, until recently when studies revealed that this cactus fruit is high in nutritive values and well sought after as an alternative functional ingredient for the food and beverage industry (Nordin *et al*., 2008; Castellar *et al*., 2003; Eulogio *et al*., 1997). This has instantaneously stimulated the dragon fruit industry in many countries including Australia, China, Colombia, Japan, Malaysia, Nicaragua, North America, Sri Lanka, Taiwan and Thailand (Gunasena *et al*., 2007; Luders and McMahon, 2006; Wu *et al*., 2006; Merten, 2004). The potential yield of this fruit has been estimated at an average of 11 to 27.5 tonnes per hectare in most countries (Crane and Balerdi, 2005).

Vietnam being the biggest producer of dragon fruit had an estimated export value of $8.9 million in 2004 and this figure is expected to increase to $35 million in 2010 (338,000 tonnes of fruit), which is 10 percent of the export revenue targeted by Vietnam's fruit industry. At present, the area dedicated for dragon fruit production in Vietnam is about 13,500 hectare with a total production of an estimated 211,000 tonnes (Nguyen *et al*., 2008).

The popularity of dragon fruit in China is also notable with a total of 3 percent from China's $510 million import value of fresh fruit in 2006. In Taiwan, there is about 2000 acres of land planting dragon fruit and production was estimated at 6.7 to 11 tonnes per acre (Zee *et al*., 2004).

In Australia, main production areas for dragon fruit includes the Northern Territory, North and Central Queensland and Northern New South Wales with an estimated total production of 500 tonnes in 2006 (Onley *et al*., 2006). Thailand which produces the fruit all year round has emerged as a tough competitor to Vietnam as the main supplier status to the EU with a market share of 17 percent in 2004 (Le Bellec *et al*., 2006).

Besides being laden with widely propagated nutritional properties, this relatively new fruit crop has been looked upon as an important economic asset because there is a great demand for drought resistant crop to address the need for production in arid and semiarid environment where water supply and rainfall is scarce (Elobeidy, 2004).

The dragon fruit has been historically a food source to the Latin American community where it originated but is no longer limited only in its fresh form today or for ornamental reasons but the downstream products now available are extensive. Dragon fruit has been used traditionally for medicinal activities, cosmetics purposes and as a source of natural colour. Downstream products from dragon fruit now being commercialized include alcoholic beverages, puree, juice, juice concentrate, ice cream, yoghurt, cereal bars, and chocolates (Crane and Balerdi, 2005; Piga, 2004; Lee *et al*., 2000). The wide application of dragon fruit in other industries makes this fruit an important trade commodity and possess an enormous potential in global economy.

2.4 THE MALAYSIAN SCENARIO

Dragon fruit is a relatively new commodity introduced to Malaysia in the late 1990s and only cultivated in small scale by individual growers. Only in the last 5 years, dragon fruit was cultivated in a larger volume due to demand that exist within the local and international market. Dragon fruit in Malaysia is commercially sold to consumer in the form of fresh fruit, cut fruit, powdered juice, concentrated juice and wine (Gunasena *et al*., 2007).

Despite being a relatively new country in the dragon fruit industry, Malaysia had a total production of dragon fruit valued at RM12.6 million by the end of 2006. The area for growing dragon fruit increased from 524 hectare of land in 2005 to 964 hectare of land in 2008 and there is an estimated production of 7,713 tonnes of fruits by the end of 2008 (Department of Agriculture, 2008). The largest production of dragon fruit comes from the state of Johor where there is about 326 hactare of dragon fruit farm, followed by Negeri Sembilan with 139 hectares.

In addition, Malaysia also imports 20 percent of Vietnam's yearly total production of dragon fruit to meet the demands from consumers and for exporting purposes. Even though the dragon fruit cultivation in Malaysia right now is the smallest in terms of land and production volume compared to other national crops, the industry is expected to experience a rapid growth within the next 5 years (Department of Agriculture, 2008).

This growth is first and foremost due to market demand and secondly, the versatility that this crop offers to a wide palette of application where its downstream products promise high returns as well. Furthermore, the health benefits and antioxidant properties that come with the attractive fruit increases its popularity amongst consumers.

2.5 SECONDARY METABOLITES

The ability to manufacture thousands of structurally diverse natural products is an archetypal feature of microorganisms and plants where the metabolites produced are classified into two major groups based on their functions: Primary metabolites are essential for growth and universally used, whereas secondary metabolites are highly diverse and variable and plays a role for the survival of the producing organism (Bode and Muller, 2003; Kutchan, 2001; Kossel, 1891).

The term secondary is used only because different sets of metabolite occur in different plant species while primary metabolites are collectively found in all plant species like carbohydrates, lipids and amino acids (Namdeo, 2007; Demain, 1999).

In addition to their importance for the plant itself, numerous secondary metabolites are also of high value to humans because of their interesting biological activities and as an important historical source of medicines (Zhang *et al*., 2004). Many of these low molecular weight compounds are distinct and difficult to produce because they accumulate at extremely low levels in selected plant species.

The primary target for metabolic engineering has always been to increase the production rate of a particular compound in the biosynthetic pathway but it is important to note that low productivity may be an indication of possible intrinsic toxicity of the plant secondary metabolites, even to the host plant cell itself (Goossens *et al*., 2003).

Secondary products derived from plants accounts for almost 50% of the most important medication in the healthcare industry and some properties that these natural products have includes antibacterials; antivirals; antitumor; products with immunosuppressant; antihypertensive; antidiabetic; antimalarial; and antihypercholesterolemic (Demain, 1999).

2.5.2 General Secondary Metabolite Pathway

The shikimic acid pathway is present in plants, fungi, and bacteria but is not found in animals. Animals have no way to synthesize the three aromatic amino acids—phenylalanine, tyrosine, and tryptophan—which are therefore essential nutrients in animal diets. The shikimic acid pathway converts simple carbohydrate precursors derived from glycolysis and the pentose phosphate pathway to the aromatic amino acids. One of the pathway intermediates is shikimic acid, which has given its name to this whole sequence of reactions (Jensen, 1986)

The condensation of phosphoenolpyruvate (PEP) from glycolysis and ethyrose-4-phosphate from the pentose phosphate pathway is catalyzed by 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) synthase.

Through a series of reactions, shikimic acid is produced, which is converted to chorismate, the final precursor for the synthesis of phenylalanine, tyrosine and tryptophan (Petersen *et al*., 2006). The shikimic acid pathway is the most important pathway when it comes to the synthesis of other secondary metabolites because these secondary products can be generally divided into the terpenoids, polyketides, phenylpropanoids and alkaloids based on their biosynthetic origin from the shikimic acid pathway (Ainasoja, 2008). For example, the flavanoids which are responsible for array of plant colours is derived from the phenylalanine (Lewis *et al*., 1998) while the alkaloids which are important therapeutic agents can be derived from the tryptophan or tyrosine (Herrmann, 1995).

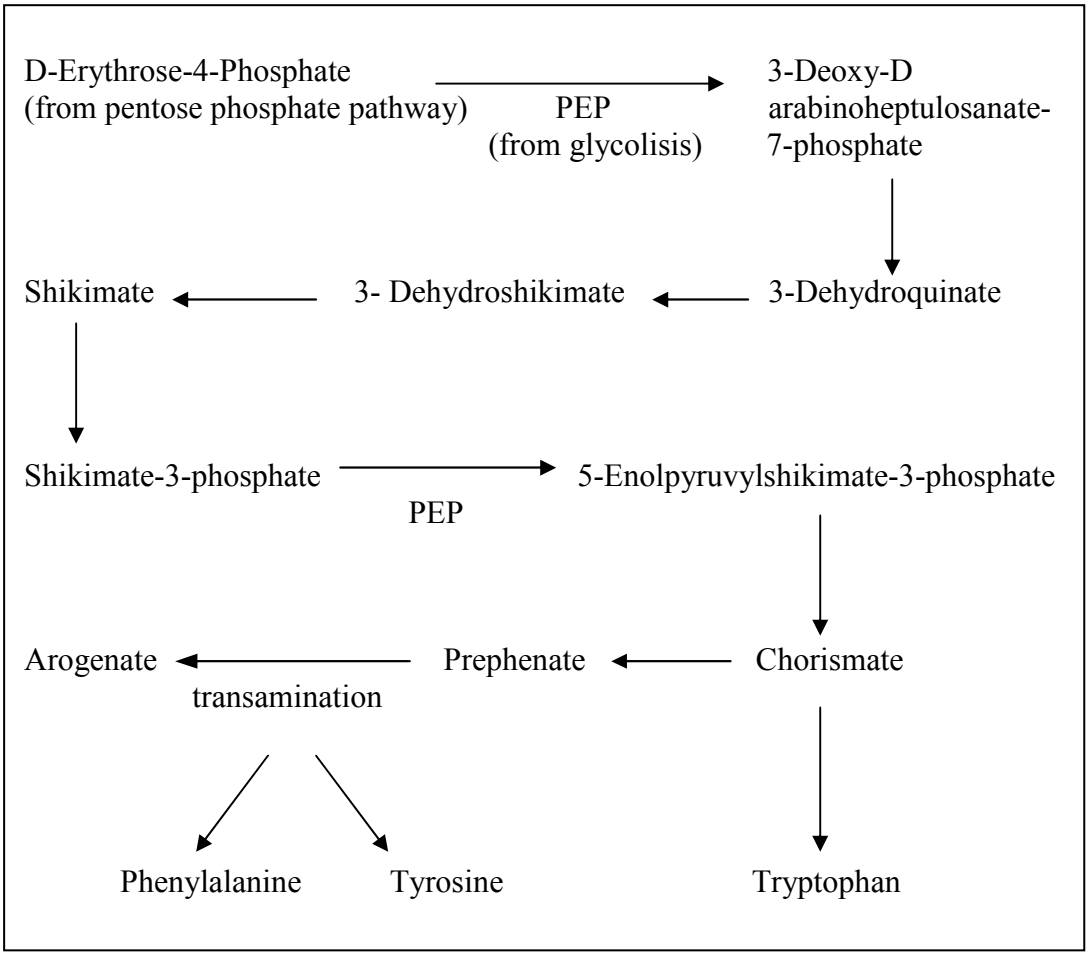


Figure 2.5: Shikimate pathway in plants (Petersen *et al*., 2006)

2.5.3 Role of Secondary Metabolite in Plants

Secondary metabolites which include enzymes and proteins are often synthesized in response to external stimulus and mostly take part in regulatory functions through a cascade of physiological and metabolic reactions (Benbrook, 2005). These compounds are not only critical in plant development but also crucial to help a plant to cope with stresses from its biotic and abiotic environment (Kutchan, 2001). Alternatively, secondary metabolites are also thought to be produced when cells are not operating under optimum conditions and are synthesized for a finite period.

von Roepenack-Lahaye *et al*., (2004) reported that secondary metabolites with antimicrobial properties help the plant defend against pathogens when damages occur in plant tissues and deterring herbivores. For example, some plant defence system repairs an injured area of its leaf or fruit tissue by formation of pigments. Other than that, plants are protected against abiotic stresses like ultra-violet (UV) rays, drought, insufficient nutrients, low temperature and high salinity through secondary metabolites (Bozack, 2006). These secondary compounds also function as internal signals or used to communicate with a symbiont which is essential in the survival of a particular plant. Ainasoja (2008) suggested that secondary metabolites which are low molecular weight compounds can act as allelochemicals to ensure plants do not grow too close to each other and also involved in other processes such as pollination, fruit/seed dispersal and plant flower pigmentation.

2.5.3 Phytochemicals

Phytochemicals are bioactive substances of plants that have been associated in the protection of human health and the major groups of phytochemicals that may contribute to the antioxidant capacity of plant foods include polyphenols, carotenoids and the traditional antioxidants vitamins such as vitamin C and vitamin E (Lako *et al*., 2007). It is widely believed that antioxidant phytonutrients can inhibit the propagation of free radical reactions that may ultimately lead to the development of diseases, especially those which are related to aging, cancer and cardiovascular complications (Bae and Suh, 2007).

The neutraceutical benefits which is associated with a lowered risk of cancer, heart disease, hypertension and stroke has promoted frequent consumption and more attention channelled to food crops as they are the main source of supplement intake in the human diet (Pozo-Insfran *et al*., 2006).

Many phytochemicals are polyphenol antioxidants that impart bright colours to flowers, fruits and vegetables. Lutein makes corn yellow, lycopene makes tomatoes red, carotene makes carrots orange, flavonols makes onions white, betalains makes beetroot red, and anthocyanin makes blueberries blue (Grotewold, 2006; Stintzing and Carle, 2004)

2.7.4 Phenolics and Polyphenolics

Phenolic substances are the most common phytochemicals and polyphenols are known as potent antioxidants and natural antagonists of plant pathogens. The term phenolics is used here to describe the large family of plant derived secondary metabolites which contain at least one benzene ring with a phenolic hydroxyl group attached.

In general, the phenolics can be separated into 2 main groups, the flavonoids and non-flavonoids (Prenesti *et al*., 2007; Montealegre *et al*., 2006). Phenolic compounds contribute to sensory properties in fruits and vegetable, in particularly colour, flavour, bitterness and astringency (Valentao *et al*., 2007).

Two basic pathways are involved in the biosynthesis of plant phenolics: The shikimic acid pathway and the malonic acid pathway. The shikimic acid pathway participates in the biosynthesis of most plant phenolics. The malonic acid pathway, although an important source of phenolic secondary products in fungi and bacteria, is of less significance in higher plants.

Polyphenols are widely available in higher plants including medicinal and edible plants where they constitute an integral part of human diet (Montoro *et al*., 2006; Makris *et al*., 2003). Polyphenols belong to a heterogenous class of compounds possessing a variety of effects towards antioxidant behaviour. These polyphenols are found to have the character of quenching oxygen-derived free radicals by donating a hydrogen atom or an electron to the free radical (Prenesti *et al*., 2007; Yuting *et al*., 1990).

Some examples of foods that are major sources of a particular polyphenolic compound combination include resveratrol in red wine grapes, catechins in tea, naringenin in citrus products, lycopene in tomatoes, and isoflavones in soya products (Benbrook, 2005).

2.7.5 Significance to Human Health

Secondary metabolites are important source of chemical for the production of modern day drugs, food flavours, fragrances, insecticides and dyes and thus hold an eminent status in the economy (Ainasoja, 2008). Pharmaceutically significant secondary metabolites include alkaloids, glycosides, flavonoids, volatile oils, tannins and resins (Namdeo, 2007).

Some prominent secondary compound that have been in the health industry's interest and commercially produced includes: morphine for sedative effects; quinine for malaria; L-DOPA for Parkinson's disease treatment; and vinblastine, atropine and digoxin for cardiovascular treatment (Vanisree *et al*., 2004; Verpoorte *et al*., 1994).

Our human body naturally produces radical-scavenging acids and enzymes in a very low concentration and polyphenols that may offer benefits by acting as antioxidants can be augmented through dietary intake (Bozack, 2006). The consumption of polyphenol-rich food has been associated with a decrease in cardiovascular complication, some types of cancer, asthma and diabetes (Kenkt *et al*., 2002). Polyphenols are suggested to increase the body metabolic capacity and neutralizes carcinogens, toxins and drugs (Benbrook, 2005).

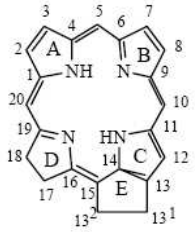
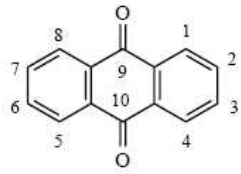
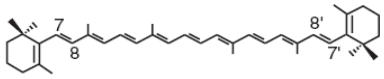
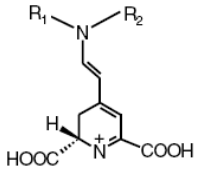
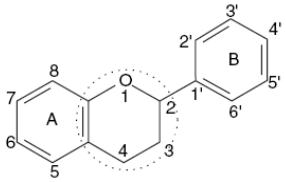
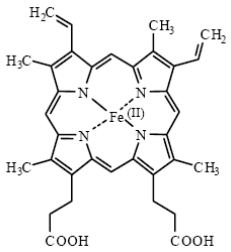
2.6 PLANT PIGMENTS

Pigments which are usually associated to the colour perceived by human eye in plant tissues play significant roles other than being appreciated for its aesthetic values alone. The aforementioned roles include: photosynthesis, protection and reproduction (Delgado-Vargas and Paredes-Lopez, 2003). In order to perform this gamut of functions, plant pigments has to occur in a huge diversity of compounds in nature and hence the abundance of pigments known thus far.

Some example of common plant pigments are: cholorophylls, flavonoids, anthocyanins, carotenoids, betalains and quinones which are responsible for imparting orange, red, purple, green, yellow and bluish hues to flowers, grasses, fruits, vegetables and grains (Stintzing and Carle, 2004).

Although plant pigments can be classified into many groups, all biological pigments are generally grouped into six structures as their core building blocks: tetrapyrroles; quinones; isoprenoids; *N*-heterocyclic compounds; benzopyrans; metalloproteins (Velisek *et al*., 2007; Delgado-Vargas and Paredes-Lopez, 2003). Table 2.1 illustrates the characteristics of the six structures and the common pigments found in each group.

Table 2.1: The six main groups of pigments

Group	Common pigments	Predominant Colour	Structure
Tetrapyrroles	Chlorophylls	Green	
Quinones	Anthraquinone	Red-Purple	
Isoprenoids	Carotenoids	Yellow-Red	
N-heterocyclic compounds	Betalains	Yellow-Red	
Benzopyrans	Flavonoids	Blue-red, Yellow-White,	
Metalloproteins	Hemoglobin	Red	

***Source*:** Adapted from Delgado-Vargas and Paredes-Lopez (2003), Delgado-Vargas *et al*., (2000) and Hendry (1996).

Each pigment has its own characteristics in terms of its chemical structure and its compartmentalization within a plant cell which further contributes to the plant's phenotypic appearance. Pigment synthesis in plants have been associated to exogenic stress and changing environments which explains the multifaceted plant pigments to maintain a balanced physiological state in plant tissues (Stintzing and Carle, 2004; Lee and Collins, 2001; Mazza, 2000).

2.7 BETALAINS

Betalains were known as the "nitrogen containing anthocyanins" until the new description and characterization of the pigment were made by Mabry and Dreiding in 1968. This new name was coined from the latin name of red beet (*Beta vulgaris*) which was the model plant in understanding the chemical structure and production pathway of betanin, the major colour pigment in the plant (Wissgott and Bortlik, 1996). The nitrogen-containing betalain group is made up of the betacyanins, which are the red-violet chromophores and the betaxanthins, which are the yellow chomophores with maximum absorptivity at λ_{max} about 535nm and λ_{max} about 480nm respectively.

These pigments are water-soluble and are usually accumulated in the vacuole of plants from the order Caryophyllales where ten families were identified to be betalain-producing: Aizoacea, Amaranthaceae, Basellaceae, Cactaceae, Chenopodiaceae, Didieraceae, Holophytaceae, Nyctaginaceae, Phytolaccaceae and Portulacaceae. (Stintzing and Carle, 2007; Herbach *et al*., 2006a; Stintzing and Carle, 2004; Strack *et al*., 2003; Delgado-Vargas and Paredes-Lopez, 2003).

Betalains which are found mutually exclusive from the anthocyanins, are also detected in some fungi like *Amanita muscaria* unlike the anthocyanins which only occurs in the Angiosperms. These special features of betalains are due to the inability of betalain-producing plants to convert flavan-3,4-diols to anthocyanidin, instead, converting tyrosine via dihydroxyphenylalanine (DOPA) and a chain of transitional metabolites into betacyanins or betaxanthins (Wang and Liu, 2006).

Thus, betalains are deemed to be important chemotaxonomic markers and there are increasing number of studies done on the biochemistry and genetics of betalains (Gandia-Herrero *et al*., 2005; Christinet *et al*., 2004; Vanisree *et al*., 2004; Trezzini and Zryd, 1990; Ootani and Hagiwara, 1969). Examples of betalain-producing plants and are usually studied on includes: *Beta vulgaris, Portulaca grandiflora, Amaranthus tricolor, Gomphrena globosa, Mirabilis jalapa* and *Hylocereus polyrhizus*. The origin of the alternative pigment biosynthetic pathway of betalain during the plant evolution is still unknown and remains a mystery.

2.7.1 Betalain structures

The common feature of all betalains is the presence of a chromophore, betalamic acid (Figure 2.5a) and the general formula is based on a protonated 1,2,4,7,7-pentasubstituted-1,7-diazaheptamethin system (Figure 2.5b) (Delgado-Vargas and Paredes-Lopez, 2003). The betaxanthins are products of condensation of betalamic acid via the aldehyde group with amino acids or amines which gives rise to a more intense yellow colour with an absorbance maximum at 470 – 484nm (Figure 2.6a).

On the other hand, betacyanins are characterized by the cyclo-dihydroxyphenylalanine (cyclo-DOPA) structure as condensed amine where the conjugation is extended at the substituted aromatic ring giving rise to a red-violet colour with an absorbance maximum at 534 – 552nm (Figure 2.7b) (Stintzing and Carle, 2004). In general, most betacyanins are glycosylated and are acylated via an ester linkage to the sugar moiety.

Due to the many possibilities of conjugation, acylation, isomerisation, glycosylation, hydrolysis, carboxylation and other structural conformation in both betaxanthins and betacyanins, the betalains exists in a great variety of structures (Table 2.2) which contributes to colours such as yellow, red and violet in flowers like *Bougainvillea, Celosia, Gomphrena, Portulaca, Mirabilis* and cactus flowers and fruits (Velisek *et al*., 2007; Herbach *et al*., 2006b; Stintzing and Carle, 2004).

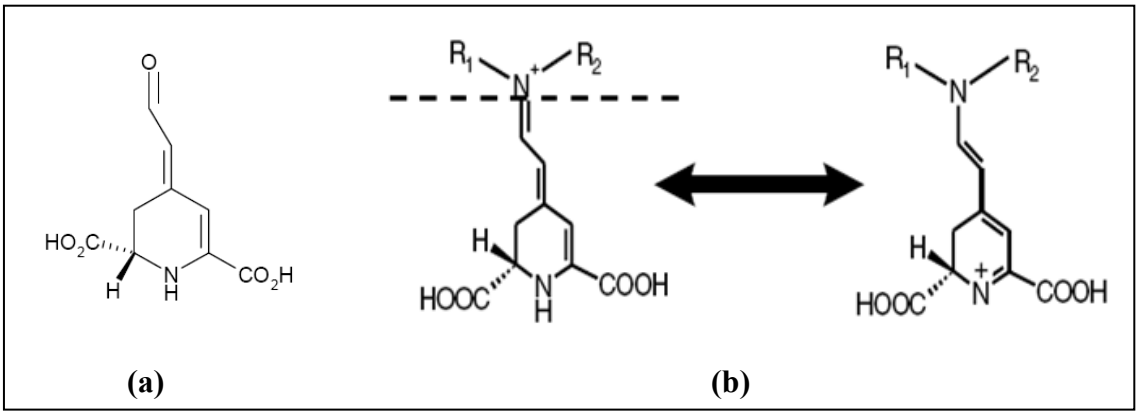


Figure 2.6: (a) Chemical structure of betalamic acid (Strack *et al*., 2003) **(b) Betalain and the resonating double bonds in the 1,2,4,7,7-pentasubstituted-1,7-diazaheptamethin system which attributes to the vast betalain colours.** The structure below the dashed line is the betalamic acid structure (Delgado-Vargas and Paredes-Lopez, 2003).

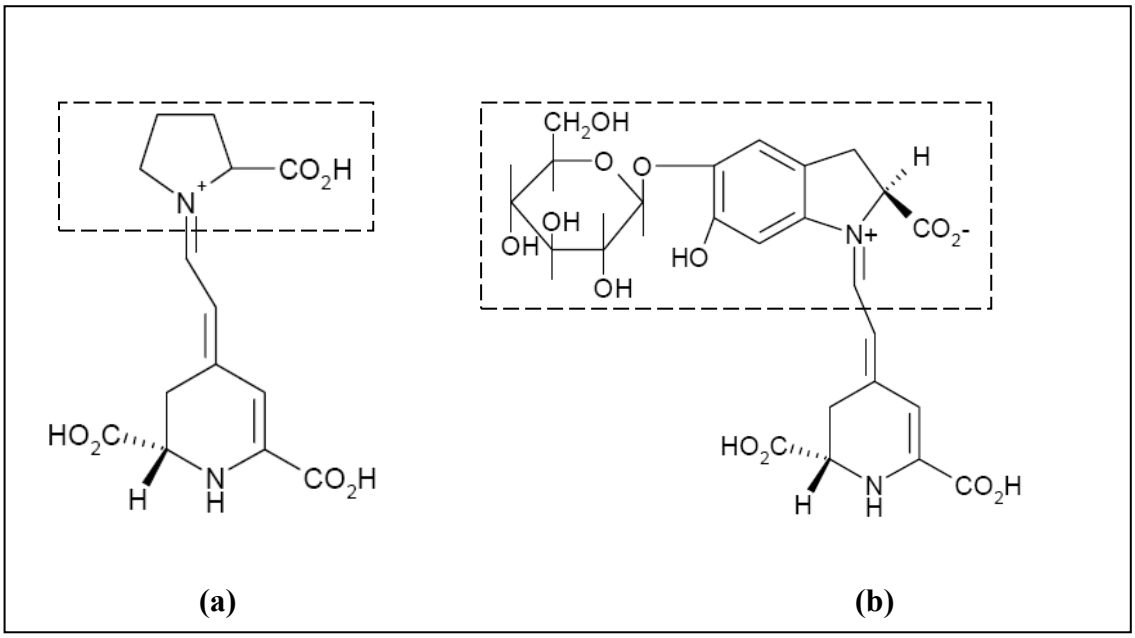


Figure 2.7: (a) An example of a betaxanthin: Indicaxanthin from *Opuntia ficus-indica*. This structure is present in all betaxanthins associated to an amino acid or an imino compound (Delgado-Vargas and Paredes-Lopez, 2003). This structure was first isolated by Piattelli *et al*., (1964) **(b) An example of a betacyanin: Betanin from *Beta vulgaris*.** This structure is present in all betacyanins associated to a cyclo-DOPA ring (Stintzing and Carle, 2004). This structure was first elucidated by Wyler *et al*., (1963).

Table 2.2: Examples of betalains

Betaxanthin			
Name	**R_1**	**R_2**	**Botanical source**
Indicaxanthin	Both groups together from proline		*Opuntia ficus-indica*
Portulacaxanthin - I	Both groups together from hydroxyproline		*Portulaca grandiflora*
Vulgaxanthin – I	H	Glutamine	*Beta vulgaris*
Vulgaxanthin – II	H	Glutamic acid	*Beta vulgaris*
Dopaxanthin	H	L-DOPA	*Glottiphyllum longum*

Betacyanins			
Name	**R_1**	**R_2**	**Botanical source**
Betanin	β-glucose	H	*Beta vulgaris*
Phyllocactin	6'-*O*-(malonyl)-β-glucose	H	*Phyllocactus hybridus*
Lamparanthin	6'-*O*-*p*-coumaroyl-β-glucose	H	*Lamparanthus* spp.
Amaranthin	2'-*O*-(β-glucoronic acid)-b-glucose	H	*Amaranthus tricolor*
Celosianin - II	2'-*O*-[O-(trans-feruloyl)-β-glucoronic acid]-β-glucose	H	*Celosia cristata* L.

Source: Adapted from Jackman and Smith (1996) and Delgado-Vargas *et al.,* (2000)

2.7.2 Betalain biosynthesis

The genesis of the betalains biosynthesis starts at arogenate which originates from the shikimate pathway and thus, the betalains are also considered secondary metabolites. The initial biosynthetic pathway consists of several enzymatic reaction steps and spontaneous chemical reaction steps as shown in Figure 2.9 (Tanaka *et al*., 2008). Piatelli (1976) has established that arogenate is converted to tyrosine, an amino acid, via arogenate dehydrogenase.

The basic dihydropyridine structure in all betalains is synthesized from two molecule of tyrosine; thus, two molecules of L-5,6-dihydrophenylalanine (L-DOPA) are formed (Mueller *et al*., 1997). Mueller *et al*., 1996 have characterized a tyrosinase from *Amanita muscaria*, an enzyme located only in the coloured parts of the fungi and it catalyzes the reaction of tyrosine hydroxylation to L-DOPA, confirming the involvement of tyrosinase in the biosynthesis of betalains. The pathway is then mediated by a DOPA-oxygenase, which is a monoxygenase, to produce cyclo-DOPA with a transitional DOPA-quinone using one of the L-DOPA molecules. The other L-DOPA molecule goes through a 4,5-extradiol oxidative cleavage, to 4,5-*seco*-DOPA, and recyclization to produce betalamic acid and this reaction is catalyzed by a DOPA-4,5-dioxygenase. It is suggested that conversion of DOPA-quinone to *cyclo*-DOPA, as well as the condensation of *cyclo*-DOPA with betalamic acid in this pathway, is a spontaneous reaction (Velisek *et al*., 2007; Gandia-Herrero *et al*., 2005; Christinet *et al*., 2004). The betalamic acid may condense with the imino-group of *cyclo*-DOPA to produce the red-purple betacyanins (Figure 2.10) or with other imino or amino group of amino acids to give the yellow betaxanthins (Figure 2.11) (Moreno *et al*., 2008; Grotewold, 2006; Steiner *et al*., 1999).

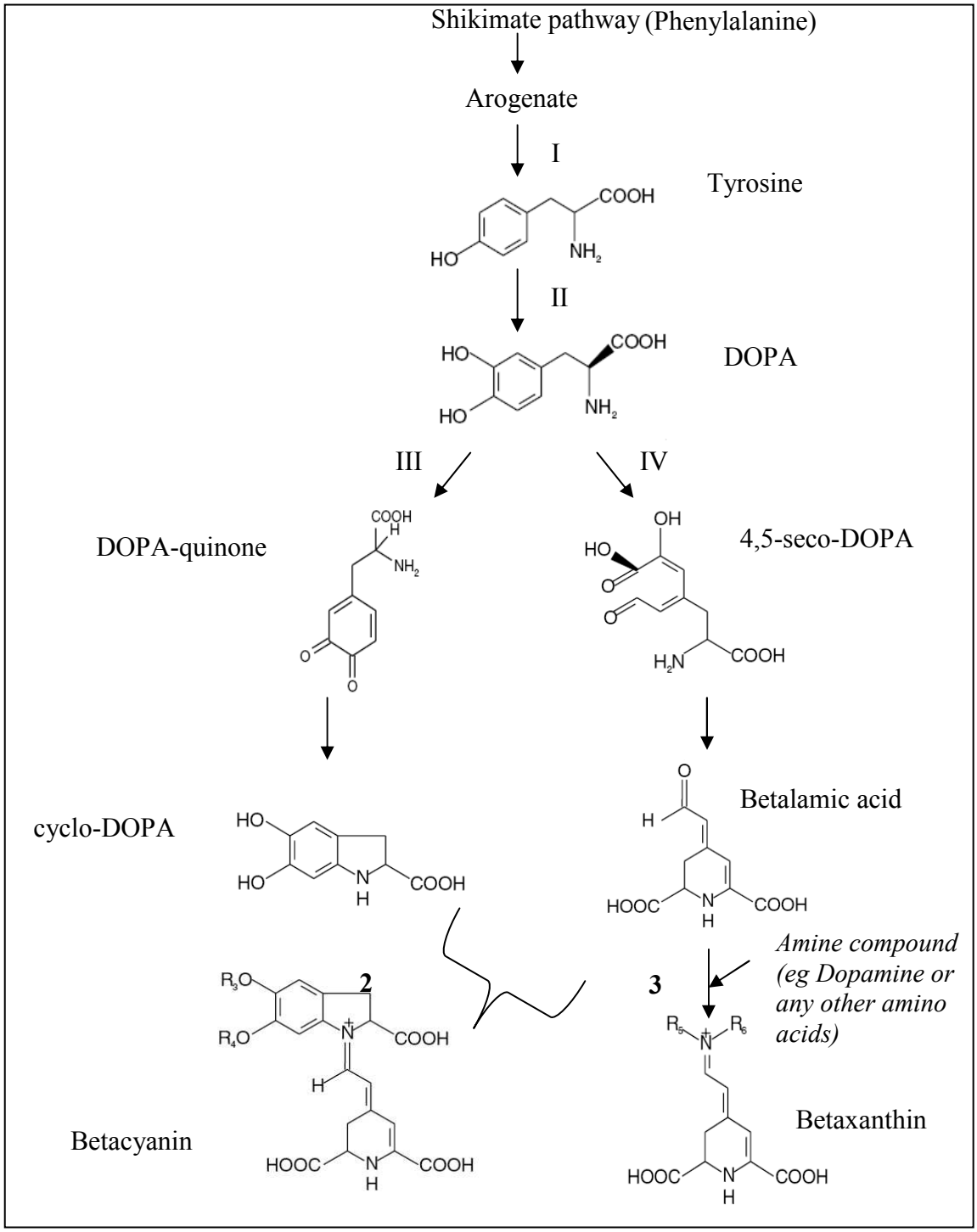


Figure 2.8: The initial biosynthetic pathway of betalains. Enzymes involved in the reaction are (I) arogenate dehydrogenase; (II) tyrosinase; (III) DOPA-oxygenase; (IV) DOPA-4,5-dioxygenase. DOPA = 5,6-dihydroxyphenylalanine. **2. Biosynthesis of betacyanins.** Betacyanins may be glycosylated or acylglycosylated at position R_3 and R_4. **3. Biosynthesis of betaxanthins.** R5 and R6 represent lateral chains of amine compounds. *Source:* Adapted from Tanaka *et al*., 2008; Velisek *et al*., 2007; Gandia-Herrero *et al*., 2005 and Delgado-Vargas and Paredes-Lopez, 2003.

2.7.3 Betalains in *Hylocereus polyrhizus*

The interest and elucidation of the pigments from *Hylocereus polyrhizus* has only started the past few years (Herbach *et al*., 2007; Wybraniec and Mizrahi, 2005; Herbach *et al*., 2004; Strack *et al*., 2003; Stintzing *et al*., 2002a; Wybraniec *et al*., 2001). Wybraniec and Mizrahi, (2002) elucidated that there are at least three betacyanins present in *Hylocereus polyrhizus* and their respective 15*R*-isoforms. All six structures are usually observed on the typical high-performance liquid chromatography (HPLC) profile and expressed as percentage of the total HPLC peak area. The betacyanins are: Betanin (18.9%); Phyllocactin (36.1%); and Hylocerenin (11.7%), with their isoforms, Isobetanin (7.2%); Isophyllocactin (19.2%); and Isohylocerenin (5.8%). A separate study carried out by Stintzing *et al*., (2002b), established that pigments elucidated from *Hylocereus polyrhizus* is totally devoid of betaxanthins. In the same study, the authors had separated ten betacyanin structures using two different HPLC systems with different conditions compared to what had been reported earlier by Wybraniec and Mizrahi, (2002) whom only used one HPLC system. The discovery of pigments and structures which are present in *Hylocereus polyrhizus* increases as other authors employ different HPLC systems (Esquivel *et al*., 2007), using liquid chromatography-mass spectrophotometry/mass spectrophotometry (LC-MS/MS) analysis (Wybraniec and Mizrahi, 2005), liquid chromatography-nuclear magnetic resonance spectroscopy (LC NMR) and 2 dimensional nuclear magnetic resonance spectroscopy (2D NMR) (Stintzing *et al*., 2004). The abundance of structures found are due to the ability of betacyanins to form stereoisomers under different conditions and going through structural changes like decarboxylation and dehydrogenation (Herbach *et al*., 2006a; Wybraniec and Mizrahi, 2005).

2.8 THE ARTIFICIAL COLOURING INDUSTRY

In 2005, global trade for dyes, pigments and dye intermediates were totalled at around $23 billion and the production volume was estimated at around 34 million tons. Asia has been a global dye manufacturer for the past 20 years, over taking the domineering European nations due to lower costs of production in the Asian region and its growing prominence as a global industrial hub. These figures and data retrieved from the Dyes and Pigments website was the only available latest information on the global dye trade due to the inexistence of such details from Food and Agriculture Organization of the United Nations (FAO), The Food and Drug Administration (FDA) nor World Heatlh Organization (WHO).

It took the first synthetic coal-tar derived dye a year before it was translated into an industrial merchandise and science has not stopped developing new dyes ever since. There were two factors which determined the rapid market penetration and replacement of natural dyes by the synthetic dyes. The first was the continuous falling prices of these dyes due to competition between synthetic dye producing countries, which made it possible to colour the common cotton and wool for wider consumer base. Secondly, the continual invention of new classes of synthetic dyes play key role in the substitution process.

The first aniline dye by William Perkins was only able to replace a few natural dyes but the gradual discovery of other synthetic dyes replaced multitude of natural dyes over the years (Murmann and Homburg, 2001). An article by the USFDA entitled “Food Color Facts” reported that in the early 1900s, there were already 80 chemically made colour dyes available for use in foods and annually the FDA receives over 100 new synthetic dye applications to be approved and utilized in foods.

2.9 DANGERS OF SYNTHETIC DYES

The dangers of the dye industry were highlighted in the 19th century by Michaels (1988) when cases of uncommon bladder cancer occurred in workers of dye factories. These cases of occupational bladder cancer had painted the extreme magnitude of risk associated with exposure to dye intermediates which are highly unstable and reactive.

In the food industry, to increase or sustain the commercial values of many foods and pharmaceutical products, multiple substances like additive and preservatives are added. These extra materials play crucial roles, from being just a colourant or flavouring agent to topping up the nutrition value or even having antimicrobial properties. All these are important aspects of maximizing the shelf life of a particular food or pharmaceutical product but many of these additives have been reported to cause adverse health complication such as urticaria, anaphylaxis and some even linked these additive to be mutagenic or carcinogenic (Ozturk *et al*., 2005; Siva, 2007).

There are many concerns and urges for discontinuing the usage of synthetic dyes that have been proven dangerous like amaranth (E-123) and erythrosine (E-127) but the debate has no closure because different countries have different regulations on synthetic food colours (Tuormaa, 1994; Perez-Urquiza and Beltran, 2000). For example, amaranth has been labelled carcinogenic by FDA but many other studies have tried to dispute the findings by creating a cloud of doubt on its safety. The United States banned amaranth due to the unresolved health threat but amaranth remained the most widely used red colourant in the world.

Some examples of banned food additives: AF-2 was an antimicrobial preservative used in Japan until its removal from the market in 1965 after being proven to be carcinogenic; and Butter Yellow, an azo dye was banned in the United States and soon after Europe followed suite after it was proven to be genotoxic (Macioszek and Kononowicz, 2004).

In a study carried out by Sasaki *et al*., 2002, synthetic dyes like amaranth, allura red, new coccine, tartrazine, erythrosine, phloxine, and rose bengal were found to be most genotoxic of all the six categories of additives studied on mouse organs. The abovementioned synthetic dyes induced DNA damage in the gastrointestinal organs even at a low dose (10 or 100 mg/kg). Further, amaranth, allura red, new coccine, and tartrazine were found to induce DNA damage in the colon at close to the acceptable daily intakes (ADI) of these dyes. This proved that there is a pressing need to have more extensive evaluation of food additives that are currently produced and marketed to consumers.

An article retrieved from the Building Antibodies and Neutralizing Toxins (BANT) Practitioners website, reported that a typical American child can ingest as much as 600 mg of food dye on a daily basis from cakes, ice creams, drinks, candies, cereals and pharmaceutical products. These dyes consumed include Red 40 (Allura Red), Blue 1 (Brilliant Blue FCF), Yellow 5 (Tartrazine) and Yellow 6 (Quinoline Yellow). The reported amount of food dye ingested is troubling and considerably worrisome when compared to the previously mentioned report by Sasaki *et al*., 2002, where Allura Red and Tartrazine induced DNA damage in the gastrointestinal mouse organs even at a low dose (10 or 100 mg/kg). Such reports are always belittled by many parties especially the food producers and the government due to economic importance; political pressure; to avoid panic; and the responsibility to review the usage of synthetic food dyes.

In reality, such reports should not be taken lightly especially when it comes to consumer's health and it is no surprise that the global health care system is being burdened with more cases of cancer, coronary failure, obesity and acute allergy reaction.

Scientists face the difficulty in proving the dangers of food additives because most assessments and toxicology tests are carried out *in vitro* and *in vivo*. Many have argued that these results are not reflective and not accurate to conclude it has that same extent of impact on human health. For example, saccharin is widely used as sweeteners and sodium saccharin is reported to be carcinogenic in experimental animals but there is no confirmation that it is in humans. Thus, all the epidemiological studies carried out on synthetic dyes can only reflect the compound's absorption, distribution, excretion, and metabolism *in vitro*.

2.10 FOOD COLOURING

Food colouring is a common pre-requisite to compensate for process-related losses to improve overall appearance of food and this factor is important to meet modern day customers' expectations (Bridle and Timberlake, 1997). Colour does not only help sell products but it also extends shelf life by covering spoilt areas in foodstuff like meat or painting the unripe fruits to deceive consumers. Following the rapid industrial revolution, the food industry employed the convenient usage of mineral and metal based compounds to disguise low quality and adulterated foods (Downham and Collins, 2000).

Even though colour is a vital element in food which determines its ultimate acceptance, the issue on usage of synthetic dyes has been widely discussed (Oser, 1966; Murmann and Homburg, 2001; Sasaki *et al*., 2002; Das and Mukherjee, 2004; Shubik, 1975). Chemically synthesized food dye are very much preferred by all food producers despite being a potential health threat because they are found to be easier to make, cost effective, mixes well with almost all foods, the colours created are easily controlled and maintained and they do not alter the flavour of the foods (Downham and Collins, 2000).

These synthetic materials are critically analysed and evaluated by the manufacturers, Food and Agricultural Organization (FAO) and World Health Organization (WHO). The FAO established an Acceptable Daily Intake (ADI) which suggests or recommends the limit of certain synthetic dyes that may be consumed before it posses any health threats.

To accentuate the depth of this matter, 700 synthetic dyes used globally have been removed from the national and international lists of permitted food colouring to only seven in the FDA list. Table 2.3 lists the currently approved synthetic food colourant in the European Union and in the USFDA according to Delgado-Vargas *et al*. (2000).

Table 2.3: Approved Colours for Food Industry in the European Union (Central and Eastern Europe, CEE) and in the Food and Drug Administration (FDA) of USA

Colours	CEE	FDA	ADI (mg/kg/day)
Red Allure AC	No	Yes (Red #40)	7.0
Brilliant Blue	No	Yes (Blue #1)	10.0
Carmosine	Yes (E122)	No	4.0
Erythrosine	Yes (E127)	Yes (Red #3)	0.1
Fast Green FCF	No	Yes (Green #3)	5.0
Indigotine	Yes (E132)	Yes (Blue #2)	5.0
Ponceau 4R	Yes (E124)	No	4.0
Sunset Yellow FCF	Yes (E110)	Yes (Yellow #6)	2.5
Tartrazine	Yes (E110)	Yes (Yellow #5)	7.5

Source: Adapted from Delgado-Vargas *et al*. (2000)

2.10.1 Natural food dyes from plants

Long before the discovery of synthetic food dyes, natural food dyes has always been an integral part of the human diet. Unknowingly, human do consume a significant amount of natural food dye when they consume pigment rich foods like green leafy vegetables, carrots, tomatoes, colourful berries and multi-coloured fruits. Unfortunately, this diet regime is quickly fading due to the increased dependence on processed and fast food in society (Tuormaa, 1994).

Due to the recent increased awareness, natural food dye is gaining back its popularity especially in the food and beverage sectors due to strong demand for more natural products by health-conscious consumers (Herbach *et al*., 2007).

Some examples of natural food colorants which have already been used in the food industry include anthocyanin, curcumin, beetroot red, caramel, lycopene, paprika extract and chlorophyll. Furthermore, natural colours and pigment from fruits and vegetables may contribute additional nutritional values to food coloured as was observed in cactus fruits (Moβhammer *et al*. 2005). In Europe, there are 13 permitted natural food colourant while there are 26 naturally derived colours which are exempted from certification in US (Downham and Collins, 2000). Table 2.4 lists the most common natural colour currently used in the food industry.

Table 2.4: List of permitted and the most common naturally derived pigments from plants utilized in the current food industry

Pigment	Sources	Colour
Curcumin	Turmeric	Bright lemon yellow
Lutein	Marigold and Alfalfa	Golden yellow
Carotenes	Palm Oil	Golden yellow to orange
Bixin/Norbixin	*Bixa orellana*	Orange
Capsanthin/Capsorubin	Paprika	Red-orange
Lycopene	Tomatoes	Orange-red
Carminic acid	Cochineal insect	Orange to red
Carmine	Cochineal insect	Pink to red
Betanin	Red table beetroot	Pink to red
Anthocyanin	Black grapeskin, elderberries, black carrots, red cabbage	Pink/red to mauve/blue
Chlorophyll	Grass	Olive green
Copper	Grass	Bluish green
Carbon black	Vegetable material	Grey to black
Crocin	Saffron	Yellow
Titanium dioxide	Anatase	White

Source: Adapted from Downham and Collins, 2000.

The most widespread pigments which exist in almost all families in the plant kingdom are the anthocyanins or the flavonoids. Anthocyanins are not only commonly found in the flower, fruit or leaf parts but also in roots, tubers, bulbs, legumes and cereals (Bridle and Timberlake, 1997). The commercially available anthocyanin food colourant are usually derived from grape (*Vitis* sp.), elderberry (*Sumbucus nigra*), red cabbage *(Braca oleracea)* and roselle *(Hibiscus sabdarifa)*.

Besides the common anthocyanins, betalains are the next most prominent pigment derived for its pink to red shade in the food industry. Some advantages that betalains have over anthocyanins includes, betalains are more water soluble, it exhibit a tinctorial strength up to three times higher than anthocyanins and it has a wider pH stability range from pH 3 to 7 making it suitable for application in a broad palette of low-acid and neutral food (Stintzing and Carle, 2007). Currently, betalains are utilized in several applications in foods such as listed in Table 2.5. The amount of pure pigment needed to colour most of these food does not exceed 50 ppm of betalains to achieve the desired hue or shade. The beet root is the main commercial source of betalains in the form of concentrates or powders (produced by freeze or spray drying) (Delgado-Vargas *et al*., 2000) and it is mainly composed of the red-purple betanin and the C15-isomer isobetanin. However, because of the unfavourable earthy flavour caused by geosmin and pyrazine derivatives, as well as high nitrate concentrations associated with the formation of carcinogenic nitrosamines, there is a demand for alternative compounds (Esquivel *et al*., 2007). Fruits from the Cactaceae family have been proposed as a promising betalain source being devoid of the abovementioned drawbacks.

Table 2.5: Application of beet root powder as natural colour in food products

Food Products	Shade
Dairy Products	
- Strawberry yogurt	Rose-pink
- Ice creams	Pink; Rose-pink
Meat Products	
- Sausages	Pink
- Cooked Ham	Pink; Brown
Dry powder beverages	Strawberry; Raspberry; Blackcurrant
Water Ices	Strawberry-red; Raspberry
Marzipan	Pastel-red; Bluish-red
Baked Goods	Pink-brown
Biscuit Creams	Pink; Brown
Hard Candies	Pink
Jellies	Raspberry-red
Fruit Cocktails	Raspberry-red

Source: Adapted from Delgado-Vargas *et al*. (2000).

2.11 POTENTIAL DYE FROM *HYLOCEREUS POLYRHIZUS*

The main generas within Cactacease which produces edible betalain-bearing fruits are the *Opuntia* sp., *Hylocereus* sp., and *Mamillaria* sp. (Stintzing *et al*., 2003). There has been substantial amount of work carried out in cactus fruits as a potential new source for betalain (Stintzing *et al*., 2002a; Strack *et al*., 2003; Sreekanth *et al*., 2007; Stintzing and Carle, 2007). For example, Moβhammer *et al.,* (2006b) carried out works on *Opuntia ficus-indica* where they evaluated different methods for optimization of the juice concentrate and fruit powders including pasteurisation, freeze drying, cross-flow microfiltratioin and spray drying led to the initiation to study the purple dragon fruits (*Cereus, Hylocereus* and *Selenicereus*) as a betalain source (Stintzing and Carle, 2004). Garambullo (*Myrtillocactus geometrizans*) which is easily propagated in Mexico is another purple fruit in the Cactacea family that also contains betalains (Reynoso *et al*., 1997).

Betalains are regarded as thermolabile compound (Stintzing and Carle, 2004) and this criterion is one of the advantages in terms of producing a natural food colourant. In dragon fruit (*Hylocereus polyrhizus*) alone, there are at least seven known betalain namely; betanin, isobetanin, phyllocactin, isophyllocactin, betanidin, isobetanidin and bougainvillein-r-I (Stintzing *et al*., 2002b) all of which have identical absorption spectra (λ_{max}) that contribute to the deep-purple colour observed in the fruit pulp.

In Israel, pigments from *Hylocereus polyrhizus* are used commercially for the production of red-violet ice creams and also found to be suitable for colouring low-temperature dairy drinks and frozen products (Wybraniec and Mizrahi, 2002; Strack *et al*., 2003).

Besides being thermolabile and are not as susceptible to hydrolytic cleavage as the anthocyanins, betalains in *Hylocereus polyrhizus* also have the ability to regenerate. Their stability is further enhanced by the addition of common natural food stabilizers such as citric acid, ascorbic acid and isoascorbic acid (Herbach *et al*., 2006b). The addition of 1% ascorbic acid significantly reduced betacyanin degration and after 6 months of storage, 70% of the initial betacyanins in *Hylecereus polyrhizus* remained intact with ascorbic acid added into juice samples (Herbach *et al.*, 2007). Another interesting feature of the pigments in *Hylocereus polyrhizus* is that it produces both acylated and non-acylated pigments at a ratio of 25:75 and is considered a viable natural colourant (Wybraniec and Mizrahi, 2002; Stintzing and Carle, 2004; Stintzing and Carle, 2007). Hence, *Hylocereus polyrhizus* pigment processing for a potential food dye is considered feasible if optimization of extraction method can be established to enhance the overall pigment yield.

CHAPTER 3
EXTRACTION EFFICIENCY AND STABILITY OF DRAGON FRUIT (*Hylocereus polyrhizus*) PIGMENTS

3.1 INTRODUCTION

One of the crucial parameter when it comes to evaluation of food products is colour which largely affects human evaluation and preference (Baeyens *et al*., 1996). In this chapter, the prominent attribute of the *Hylocereus polyrhizus* pulp which is the intense red-purple pigments will be evaluated with parameters related to developing the pigment as potential natural dye.

Plant pigments are usually quantified using the spectrophotometer where every pigment has their own absorption spectra. For example, anthocyanins are quantified at 700nm (Guisti and Wrolstad, 2001); chlorophyll and carotenoids at 647 and 446 nm respectively (Raveh *et al*., 1998). In this study, the pigments which are known as betacyanins from the *Hylocereus polyrhizus* pulp will be quantified using the similar technique. Previous studies which also quantified betacyanins using spectrophotometer includes Stintzing and Carle (2007), Herbach *et al*., (2006c) and Moβhammer *et al*., (2005).

Under the regulations of Food and Drug Administration (FDA), which is the global food safety governing body, natural colour additives are exempted from certification but the stability of the pigment would be first tested. In this study, the stability and potential of the *Hylocereus polyrhizus* pigment as natural food colourant will be assessed following selected parameters suggested by the FAO in the Combined Compendium of Food Addictive Specification:

1. Total betalain concentration (To investigate the strength and potency of the pigment)
2. Water extraction (To investigate the extractability and stability of pigment in water matrices)
3. Temperature (To investigate the temperature tolerance of pigment and optimum temperature for higher pigment yield)
4. Storage conditions (To investigate the optimum storage condition)
5. pH (To investigate the pH stability of the pigment throughout the experiments)

The above selected parameters were chosen based on previous studies by Mahmoud *et al*., (2007), Pardo and Forchiassin (1999), Sims and Morris (1984), and Savolainen and Kuusi (1978), who used different parameters to carry out stability test on different dye and food related products. The aim of this chapter is to investigate the pigment extraction efficiency using water and heat, and stability of the pigment after storage under various conditions.

3.2 MATERIALS AND METHODS

3.2.1 Plant Material

Red-pulped dragon fruits (*Hylocereus polyrhizus*) were obtained from Multi Rich farm in Nilai, Negeri Sembilan, located about 80 km from the Postharvest Biotechnology laboratory, University of Malaya. All fruits were freshly harvested at Stage 5 (25 ± 2 days after anthesis) and transported to the laboratory for analysis. Fruits were washed thoroughly under running water, treated with Benomyl 0.05 % and air dried overnight. Fruit pulp was cut into small cubes, frozen under liquid nitrogen and stored in -20 ºC until used.

3.2.2 Pigment Extraction

Frozen *Hylocereus polyrhizus* samples were thawed 30 minutes prior to extraction. All extracts were filtered using Whatman paper No.1 (9cm) to separate the filtrate. Total betalain concentration and pH were measured for all extracts. Experiments were carried out in triplicates.

3.2.3 Sample Measurements

3.2.3.1 Determination of total betalain concentration in samples

Absorbance of all samples was measured using a spectrophotometer (Pharmacia, Ultrospec II) at 538 nm against a sterile distilled water (SDW) blank. Total betalain concentration was calculated using the following formula (Herbach *et al*., 2007):

$$\text{BC (mg/L)} = \frac{A \times MW \times 1000 \times DF}{\varepsilon \times \iota}$$

where,

A = absorbance
DF = dilution factor
MW = molecular weight of betanin = 550 g/mol
ε = molar extinction coefficients = 60,000 L/mol cm in H_2O
ι = pathlength of cuvette = 1 cm

3.2.3.2 pH measurement

pH for all samples was measured using a Hanna pH meter. The pH meter probe was rinsed with distilled water and dried with a clean C-fold hand towel. The probe was gently placed into test tubes containing samples and the pH was recorded once the reading on the display stabilizes.

3.2.4 Efficiency of Water Volume to Extract Pigment

Approximately 5 g of pulp were immersed in 5, 10, 15 and 25 ml of sterilized distilled water (SDW) for 10 minutes to determine best weight:volume ratio for highest pigment yield. Total betalain concentration and pH were measured for all extracts. The best weight:volume ratio was then used for subsequent experiment.

3.2.5 Efficiency of Temperature to Extract Pigment

Approximately 5 g of pulp were immersed in 5 ml SDW and pigments were extracted at room temperature (RT) (22 ± 3 ºC), 40 ºC, 60 ºC, 80 ºC and 100 ºC for 10 minutes to determine best temperature for highest pigment yield. The best temperature for highest pigment yield was then used for experiments in Section 3.2.6 and Section 3.2.7.

3.2.6 Stability of Pigments Extracted with Water

3.2.6.1 Sample preparation

Approximately 5 g of pulp in 5 ml of SDW were extracted at RT and 100ºC for 10 minutes.

3.2.6.2 Stability test samples

Approximately 5 ml of samples prepared according to Section 3.2.6.1 were aliquoted into test tubes and stored under the following light and temperature condition: RT, 4 ºC and -20 ºC. Samples stored at RT were subjected to light and dark. Samples for dark storage were wrapped with aluminium foil. All samples were stored for one week following which the pH and total betalain concentration were determined.

3.2.7 Stability of Pigments from Juice Concentrate

3.2.7.1 Sample preparation

Approximately 500 g of fruit pulp were pressed using a commercial sieve to obtain juice concentrate. 5 ml of the juice was poured into test tubes and extracted at RT and 100 ºC for 10 minutes. The pH and total betalain concentration for each sample was determined.

3.2.7.2 Stability test samples

Approximately 5 ml of samples prepared according to Section 3.2.7.1 were aliquoted into test tubes and stored under the following temperature condition: RT, 4 ºC and -20 ºC. Samples stored at RT were subjected to light and dark. Samples for dark storage were wrapped with aluminium foil. All samples were stored for one week following which the pH and total betalain concentration were determined.

3.3 RESULTS

3.3.1 Efficiency of Water Volume to Extract Pigment

The total betalain concentration obtained from 5 g of pulp in 5, 10, 15 and 25 ml of SDW is summarised in Table 3.1 below. In this experiment, the most consistent and highest yield was obtained when 5 g of pulp was extracted in 5 ml of SDW. When 10 ml of SDW was used to extract sample, the pigment yield was comparable to 5 ml of SDW used but the deviation of each replicate was greater. When samples were extracted in 15 and 25 ml of SDW, the average pigment yield were slightly lower and samples showed highest deviation when extracted in 25 ml of SDW. In terms of choosing and deciding the most efficient water volume to extract pigment, 5 ml of SDW was chosen for subsequent experiments based on producing highest average pigment yield and lower deviation between replicates. Hence, the best weight:volume to extract pigment is 1:1. Table 3.2 shows that the pH for all samples was stable even though extracted at different volumes of SDW.

Table 3.1: Total betalain concentration of samples in different volumes of SDW. 5 ml of SDW was chosen as the most efficient water volume to extract 5 g of pulp as it showed highest average pigment yield and lower standard deviation.

SDW (ml)	Total Betalain Concentration (mg/L)
5	25.59 ± 0.15
10	25.21 ± 0.64
15	24.99 ± 0.06
25	24.64 ± 0.72

Table 3.2: pH measurement of samples in different volumes of SDW. The pH of all samples was stable in different volumes of SDW.

SDW (ml)	pH
5	5.43 ± 0.07
10	5.54 ± 0.05
15	5.43 ± 0.07
25	5.47 ± 0.05

3.3.2 Efficiency of Temperature to Extract Pigment

The total betalain concentration obtained from 5 g of pulp extracted in 5 ml of SDW at RT, 40 ºC, 60 ºC, 80 ºC and 100 ºC is summarised in Table 3.3 below. In this experiment, RT and 100 ºC were chosen as the two preferred heat treatment as the average pigment yield was higher than other temperatures. On top of having a very similar pigment yield despite the vast temperature difference, 100 ºC was selected for subsequent pigment stability tests because the pigment colour of samples treated at this temperature changed to scarlet-red from red-purple at RT (Figure 3.1). This alteration of pigment colour at 100 ºC is of great experimental interest. When samples were extracted at 40 ºC, 60 ºC and 80 ºC, the pigment colour of samples remained purple. Table 3.4 shows that the pH for all samples was stable even though extracted at different temperatures.

Table 3.3: Total betalain concentration of samples at different temperatures. RT and 100°C showed similar and higher average pigment yield compared to the other temperature.

Temperature (°C)	Total Betalain Concentration (mg/L)
RT	26.22 ± 0.25
40	25.62 ± 0.14
60	26.05 ± 0.32
80	25.75 ± 0.47
100	26.23 ± 0.32

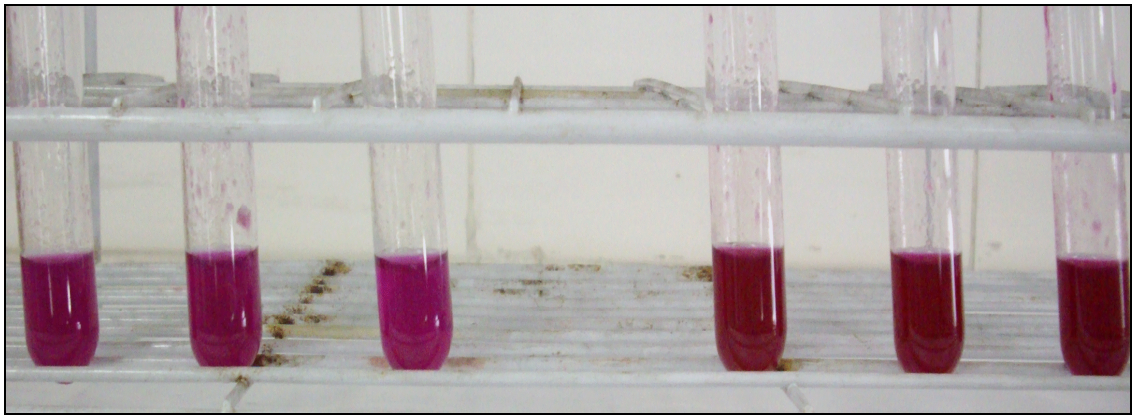


Figure 3.1: Pigment colour change extracted at RT and 100°C. When samples were extracted at RT, the pigment yield was red-purple in colour (3 test tubes on the left). At 100°C, samples changed colour to scarlet red (3 test tubes on the right)**.**

Table 3.4: pH measurement of samples at different temperatures. The pH of all samples was stable at different temperatures

Temperature (°C)	pH
RT	5.35 ± 0.07
40	5.27 ± 0.05
60	5.25 ± 0.05
80	5.20 ± 0.03
100	5.28 ± 0.02

3.3.3 Stability of Pigments Extracted with Water

3.3.3.1 Samples extracted at room temperature

The total betalain concentration at Day 0 (before samples were stored) and Day 7 (after one week of storage) is summarised in Table 3.5 below. All samples showed an increase in total betalain concentration after one week of storage. The highest increase was observed in samples that were stored at 4 ºC, from 26.20 mg/L at Day 0 to 94.78 mg/L at Day 7 while the lowest increase, thus the most stable, was observed in samples that were stored at -20 ºC, from 26.32 mg/L at Day 0 to 63.15 mg/L at Day 7. Table 3.6 shows that the pH for all samples was stable even though stored under various conditions after one week.

Table 3.5: Total betalain concentration of 5g of *Hylocereus polyrhizus* pulp extracted in 5ml of SDW at room temperature stored under various conditions.

Storage	Exposure	Total betalain concentration (mg/L)	
		Day 0	Day 7
RT	Light	26.43 ± 0.78	67.53 ± 22.2
	Dark	25.99 ± 0.26	68.90 ± 20.57
4 °C		26.20 ± 0.49	94.78 ± 11.10
-20 °C		26.32 ± 0.50	63.15 ± 9.74

Table 3.6: pH measurement of samples stored under different conditions.

Storage	Exposure	pH	
		Day 0	Day 7
RT	Light	5.94 ± 0.26	5.87 ± 0.26
	Dark	5.97 ± 0.34	5.84 ± 0.24
4 °C		5.69 ± 0.17	5.75 ± 0.18
-20 °C		5.49 ± 0.06	5.57 ± 0.02

3.3.3.2 Samples extracted at 100 ºC

The total betalain concentration at Day 0 (before samples were stored) and Day 7 (after one week of storage) is summarised in Table 3.7 below. All samples showed an increase in total betalain concentration after one week of storage. The highest increase of total betalain concentration was observed in samples that were stored at RT exposed to light, from 25.60 mg/L at Day 0 to 129.07 mg/L at Day 7. The lowest increase, thus most stable, was observed in samples that were stored at RT in the dark from 25.80 mg/L at Day 0 to 67.73 mg/L at Day 7. Table 3.8 shows that the pH for all samples was stable even though stored under various conditions after one week.

Table 3.7: Total betalain concentration of 5g of *Hylocereus polyrhizus* pulp extracted in 5ml of SDW at 100 ºC stored under various conditions.

Storage	Exposure	Total betalain concentration (mg/L)	
		Day 0	Day 7
RT	Light	25.60 ± 0.35	129.07 ± 13.90
	Dark	25.80 ± 0.19	67.73 ± 22.19
4 °C		25.65 ± 0.22	94.65 ± 9.19
-20 °C		25.57 ± 0.26	102.75 ± 9.10

Table 3.8: pH measurement of samples stored under different conditions.

Storage	Exposure	pH	
		Day 0	Day 7
RT	Light	5.25 ± 0.03	6.02 ± 0.54
	Dark	5.30 ± 0.03	5.71 ± 0.47
4 °C		5.38 ± 0.01	5.55 ± 0.02
-20 °C		5.34 ± 0.01	5.43 ± 0.02

3.3.4 Stability of Pigments from Juice Concentrate

3.3.4.1 Juice concentrate extracted at room temperature

The total betalain concentration at Day 0 (before samples were stored) and Day 7 (after one week of storage) is summarised in Table 3.9 below. All samples showed a decrease in total betalain concentration after one week storage except for samples stored at -20 ºC which showed an increase in total betalain concentration from 236.69 mg/L at Day 0 to 252.45 mg/L at Day 7. The highest decrease was observed in samples stored at RT exposed to light, from 246.20 mg/L at Day 0 to 67.73 mg/L at Day 7. Table 3.10 shows that the pH for all samples was stable even though stored under various conditions after one week.

Table 3.9: Total betalain concentration of 5 ml of *Hylocereus polyrhizus* juice concentrate extracted at RT stored under various conditions.

Storage	Exposure	Total betalain concentration (mg/L)	
		Day 0	Day 7
RT	Light	246.20 ± 17.55	67.73 ± 8.81
	Dark	246.30 ± 8.24	219.85 ± 11.01
4 °C		261.96 ± 5.07	227.65 ± 11.51
-20 °C		236.69 ± 11.58	252.45 ± 17.21

Table 3.10: pH measurement of samples stored under different conditions

Storage	Exposure	pH	
		Day 0	Day 7
RT	Light	4.84 ± 0.03	4.94 ± 0.09
	Dark	4.88 ± 0.01	4.65 ± 0.15
4 °C		4.89 ± 0.01	5.06 ± 0.01
-20 °C		4.89 ± 0.01	4.99 ± 0.02

3.3.4.2 Juice concentrate extracted at 100ºC

The total betalain concentration at Day 0 (before samples were stored) and Day 7 (after one week of storage) is summarised in Table 3.11 below. The results showed that there was a general decrease in total betalain concentration in samples after one week of storage.

The highest decrease was observed in samples stored at RT exposed to light, from 259.51 mg/L at Day 0 to 211.38 mg/L at Day 7 while the lowest decrease was observed in samples that were stored at -20 ºC, from 244.35 mg/L at Day 0 to 204.22 mg/L at Day 7. Table 3.12 shows that the pH for all samples was stable even though stored under various conditions after one week.

Table 3.11: Total betalain concentration of 5 ml of *Hylocereus polyrhizus* juice concentrate extracted at 100 °C stored under various conditions.

Storage	Exposure	Total betalain concentration (mg/L)	
		Day 0	Day 7
RT	Light	259.51 ± 8.58	211.38 ± 13.31
	Dark	249.53 ± 25.51	203.65 ± 23.16
4 °C		254.10 ± 20.56	212.02 ± 9.67
-20 °C		244.35 ± 16.97	204.22 ± 13.72

Table 3.12: pH measurement of samples stored under different conditions

Storage	Exposure	pH	
		Day 0	Day 7
RT	Light	4.66 ± 0.02	4.78 ± 0.17
	Dark	4.63 ± 0.01	4.76 ± 0.1
4 °C		4.65 ± 0.02	5.04 ± 0.01
-20 °C		4.69 ± 0.01	5.08 ± 0.03

3.4 DISCUSSION

From the results in Section 3.3.1, the best weight:volume ratio to obtain optimum pigment yield from *Hylocereus polyrhizus* pulp is by using the ratio of 1:1. As mentioned earlier in the section, other volumes of SDW used to extract the pigment also gave comparable results. In this experiment which was designed to investigate the best volume, the 1:1 ratio was chosen based on the highest average results and lower deviation between replicates. The results indicate that the pigments have high extractability level using water and pigments have high potency even subjected to extraction with high volumes of water. The results support the claim that betacyanins are water soluble, tinctorial strength up to three times higher than other plant pigments and it has a wider pH stability range from pH 3 to 7 making it suitable for application in a broad palette of low-acid and neutral food (Stintzing and Carle, 2007). The results from this section also show that it is possible that the pigment from *Hylocereus polyrhizus* is extractable with larger volumes of water for more economical applications at industrial levels.

When pigments were subjected to heat treatment, the pH for each temperature was stable and the pigment yield was also satisfactory. Samples extracted at RT and 100 ºC showed highest and almost identical pigment yield compared to other temperatures. At 100 ºC, the colour appearance of the pigment showed a remarkable change from the initial red-purple colour to scarlet red which remained even after storage for one week. The specific main pigment known in *Hylocereus polyrhizus* which is betanin gives the red-purple colour.

It is possible that the occurring pigment which is scarlet red after 100 ºC thermal treatment is isobetanin, the isomer of betanin as the samples gave identical absorbance at 538nm using the spectrophotometer. According to Stintzing *et al*., (2002b), betanin and isobetanin exist in the *Hylocereus polyrhizus* pulp at a specific ratio and this ratio can change depending on external stimuli or changes in its natural matrices. For example, before heat treatment, the betanin: isobetanin ratio can be at 9:1 and after treatment the ratio may be 1:9. In order to have a similar absorbance spectrum, pigments in samples which were heated up to 100 ºC must have all underwent structural change from betanin to isobetanin displaying a different colour portfolio. Further, as established earlier by Herbach *et al*., (2006a), as betacyanins undergo thermal treatment, it is known that the betalains (major pigment class in *Hylocereus polyrhizus*) will experience degradation and fluctuating chromatic stability

All samples extracted with water at RT and 100 ºC showed a significant increase in total betalain concentration after one week of storage in the different temperatures. The increase in samples extracted at RT and 100 ºC can be explained by the fact that betanin has the ability to regenerate by recondensation of hydrolysis products associated with a colour regain (Stintzing and Carle, 2007). It is also possible that when betacyanins are extracted with water, it is drawn out from its protective matrics where the condition in the pulp is the stable environment for betacyanins. Thus, the main pigment may undergo multiple structural adjustment/regeneration to stabilize itself in the water as oppose to pigments used directly from the juice concentrate without any alteration of physical condition.

This regeneration is encouraged when pH value is close to pH6 with the presence of the basic building blocks of the betacyanins cyclo-DOPA ring and betalamic acid, which forms the betacyanin chromophores.

The results show that betacyanins have great pigment retention ability even after being heated and stored in adverse conditions and exposed to illumination. The pH for all samples was stable after one week of storage. Samples which exhibited minimal total betalain change were samples extracted at RT and stored at -20 ºC. The pH for all samples was stable after one week of storage.

Studies on the stability and the regeneration of betacyanins in *Hylocereus polyrhizus* are still at an early stage because this is a relatively new crop being focused on as a natural food dye source. A noteworthy behaviour of betalain (the major pigment in *Hylocereus polyrhizus*) that should be taken into consideration in future experiments is that, it undergoes multiple processes like decarboxylation, deglycosylation, hydrolysis, isomerization, dehydrogenation and others when it is subjected to these factors : water activity, pH change, antioxidants present, chelating agents, temperature (heat), illumination and oxygen/nitrogen atmosphere (Herbach *et al*., 2006b). The structural alterations that occur will give a different pigment configuration but all these betacyanins still gives a purple-red colour which is detected at 538 nm using the spectrophotometry method.

From the results obtained, it was observed that samples extracted at RT showed minimal pH and total betalain content changes as compared to samples extracted at 100 ºC. This shows that pigments subjected to heat during extraction exhibit lower pigment stability and support the findings that heat is an important stability factor as previously mentioned in Section 3.1.

The pH value was stable for juice concentrate samples extracted at RT and 100 ºC after one week of storage but there was a general decrease in total betalain concentration. Comparing water extract and juice concentrate, the juice concentrate samples exhibited lower or less changes in total betalain concentration.

Only the juice concentrate samples stored at RT (Light) showed a tremendous drop in total betalain concentration but all other samples were stable and showed minimal changes. This shows that pigments subjected to illumination during storage exhibit lower pigment stability and support the findings that light is an important stability factor. Nevertheless, the total betalain content is much higher after one week of storage compared to results obtained in previous section. According to Herbach *et al*., (2006b), one of the factors that affect betalain stability is water activity where lower water activity (a_w) improves betalain stability and it is possible that betacyanins in their natural matrices have superior stability compared to purified solutions. Plant constituents like sugars, acids and pectic substance will lower the a_w value, thereby stabilizing betalainic pigments from the start of the experiment. This suggests that for high total betalain content, juice concentrate is a preferred choice.

Looking generally at the results in this section, it was observed that all samples showed stable pH reading throughout the experiments. Samples extracted at RT showed minimal total betalain concentration changes as compared to samples extracted at 100 ºC. This shows that pigments subjected to heat during extraction exhibit lower pigment stability and support the findings that heat is an important stability factor as previously mentioned. Samples stored at -20 ºC in both extraction methods showed minimal change compared to other samples stored at RT and 4 ºC which suggests that lower temperature stabilizes the pigments more effectively.

CHAPTER 4
IDENTIFICATION AND ANTIOXIDANT PROPERTIES OF DRAGON FRUIT (*Hylocereus polyrhizus*) PIGMENT

4.1 INTRODUCTION

High Performance Liquid Chromatography (HPLC) is used to separate, identify/qualify and quantify organic compounds. This chromatographic method is applicable in many areas including pharmaceutical, food and beverage, forensics, environment and industries. The essential components of a HPLC includes: A column, mobile phase(s), an oven, degasser and a detector. The HPLC technique was introduced as early as 1979 by Abayashi and Riley to analyze phytoplankton pigments (Schluter *et al*., 2000). Prior to that, there were very few reliable chromatographic methods commercially available to laboratory scientists and by 1980s, HPLC was commonly used for separation and identification of chemical compounds. Since then, major improvements in terms of column, reproducibility and incorporation of computers and automations to HPLC has made it the most common method to confirm the identity of an unknown compound by comparing it with a chosen known standard. HPLC provides a unique retention time of a compound for qualification and is more sensitive compared to other chromatographic methods as it can detect low concentration of compounds, as low as nanogrammes (Ferreira *et al*., 2004).

Previous studies which utilized HPLC to identify pigments or compounds includes: Stintzing *et al*., (2006) on anthocyanins and betacyanins; Yang *et al*, (2006) on phenols; Garcia-Falcon and Simal-Gandara (2005) on synthetic food dyes; and Perkins-Veazie and Collins (2004) on lycopenes.

Antioxidants are by-products from the photosynthesis process in plants along side many other outputs like oxygen, sugars, fuel (energy) and thousands of other phytochemicals. Antioxidants function as protection in plants against oxidative damages which may be a result of the complex photosynthesis process (Demmig-Adams and Adams, 2002). This function of the antioxidants in plants has been extended to deal with the oxidative stress which is the key element that causes diseases in human. In this Chapter, four out of the many possible assays to assess antioxidant capacity were selected to represent the antioxidant levels in *Hylocereus polyrhizus*.

Total polyphenolic content determination is one of the preferred assessments when it comes to evaluating the nutrition value of a certain food source. This is because polyphenols are the most abundant antioxidants in the everyday human diet (Scalbert *et al*., 2005) and its intake is 10 times higher than vitamin C and 100 times higher than vitamin E. The main sources for polyphenols are fruits, vegetables, cereals and plant derived beverages like red wine and tea (Bae and Suh, 2007; Lako *et al*., 2007; Kumaran and Karunakaran, 2006). The reducing power assay is carried out to investigate the presence of reductones in *Hylocereus polyrhizus* to stop the free radical chain reaction initiated by oxidative reagents.

The Vanillin-HCl assay measures the amount of condensed tannins which represents all polymeric flavanoids which is a diverse group of metabolites with heterocyclic ring system derived from the phenylalanine and polyketide biosynthesis. The $DPPH^{\bullet}$ radical scavenging activity determination is aimed at estimating the scavenging level of antioxidants in *Hylocereus polyrhizus* to quench the $DPPH^{\bullet}$ radicals.

All four assays are chosen based on the different representation of the diverse function of antioxidants and its manifestation in different oxidative reagents and environment. These assay has been carried out in many previous studies with different types of pigments: Anthocyanins (Rebecca *et al*., 2008), Betacyanins (Wu *et al*., 2006), Lycopenes (Kun *et al*., 2006); Crops: Mulberries (Bae and Suh, 2007), Grapes (Nakamura *et al*., 2003), Amaranthaceae (Cai *et al*., 2003). This variation shows that the selected four antioxidant assays are versatile and preferred to represent the antioxidant capabilities in a particular source.

4.2 MATERIALS AND METHODS

4.2.1 Plant Material

Red-pulped dragon fruits (*Hylocereus polyrhizus*) were obtained from Multi Rich farm in Nilai, Negeri Sembilan, located about 80 km from the Postharvest Biotechnology laboratory, University of Malaya. All fruits were freshly harvested at Stage 5 (25 ± 2 days after anthesis) and transported to the laboratory for analysis. Fruits were washed thoroughly under running water, treated with Benomyl 0.05 % and air dried overnight. Fruit pulp was cut into small cubes, frozen under liquid nitrogen and stored in -20 ºC until used.

4.2.2 Qualification of Betacyanins using HPLC Method

4.2.2.1 Reagent Preparation

96% ethanol
768 ml of ethanol [Fluka] was added into 32 ml of sterile distilled water (SDW).

Aqueous trifluoroaceteic acid (0.5% TFA) in acetonitrile
Reagent I
5 ml of TFA [BDH] was added into 995 ml of SDW.
Reagent II
100 ml of acetonitrile [Merck]

The 0.5% TFA in acetonitrile was obtained by adding 900 ml of *Reagent I* with 100 ml of *Reagent II*

Sephadex-G25 [Sigma]
2 g of Sephadex-G25 was slowly added into 50 ml of SDW in a beaker while gently stirring with a glass rod. The beaker was sealed with a parafilm after all the Sephadex-G25 powder has been added and wetted. The Spehadex-G25 was left overnight at room temperature before using.

4.2.2.2 Sample Preparation

Frozen *Hylocereus polyrhizus* pulp was thawed 30 minutes prior to experiment. The fruit pulp was squeezed manually through a commercial sieve and the resulting juice was filtered using mira cloth. The pectic substances in the filtered juice were precipitated with 96 % ethanol at a ratio of 2 ml of ethanol to 1 ml of juice. Precipitates were removed using mira cloth and the filtrate was rotary evaporated at 30 °C for 90 minutes. 1 ml of sample for HPLC was purified on Sephadex G-25 in a glass column before analysis. The fractions were collected and freeze dried. Samples were re-diluted with SDW before running it in the HPLC for analysis. All experiments were carried out in triplicates.

4.2.2.3 HPLC

HPLC was carried out with modifications according to Wybraniec *et al*. (2001) by using a Shimadzu Class VP series (LC-10AT-VP HPLC System) with a UV/VIS detector (SPD-10A-VP) equipped with a LiChroCart Purospher Star RP-18 column (id. 250 mm x 4.6 mm x 5 µm) (MERCK). An aqueous 0.5% TFA in acetonitrile was used as the mobile phase. The detection was set at 537nm and 10 µl of sample was allowed to elute through the system for 25 minutes at a flow rate of 1.0 ml/min and column temperature was set to 30 °C. The betanin standard was purchased from ABCR GmbH & Co. KG (Karlsruhe, Germany).

4.2.3 Determination of Antioxidant Properties

4.2.3.1 Reagent preparation

4.2.3.1.1 Reagents for total polyphenol determination

Sodium carbonate [BDH]
10 g of sodium carbonate was dissolved in 40 ml of distilled water. Volume of solution was made up to 50 ml with distilled water.

Gallic acid [Sigma]
0.02 g, 0.04 g, 0.06 g, 0.08 g and 0.10 g of gallic acid was dissolved in 90 ml of distilled water respectively. Volume of each solution was made up to 100 ml with distilled water.

4.2.3.1.2 Reagents for reducing power assay

Phosphate buffer pH 6.5
Reagent I
4.82 g of sodium phosphate monobasic [Sigma] was dissolved in 200 ml of deionized water. Volume of solution was made up to 250 ml with deionized water.
Reagent II
4.04 g of sodium phosphate dibasic heptahydrate [Sigma] was dissolved in 200 ml of deionized water. Volume of solution was made up to 250 ml with deionized water.

The phosphate buffer pH 6.5 was obtained by adding 205.5 ml of *Reagent I* with 94.5 ml of *Reagent II*.

Potassium ferricyanide [BDH]
1 g of potassium ferricyanide was dissolved in 90 ml of deionized water. Volume of solution was made up to 100 ml with deionized water.

Trichloroacetic acid [BDH]
10 g of trichloroacetic acid was dissolved in 90 ml of deionized water. Volume of solution was made up to 100 ml with deionized water.

Ferric chloride [Sigma]
0.1 g of ferric chloride was dissolved in 90 ml of deionized water. Volume of solution was made up to 100 ml with deionized water.

4.2.3.1.3 Reagents for Vanillin-HCl assay

Vanillin Reagent
Reagent I
1 g of vanillin [Sigma] was dissolved in 90ml of deionized water. Volume of solution was made up to 100 ml with deionized water.
Reagent II
8 ml of HCl conc. [Systerm] was added into 90 ml of methanol []. Volume of solution was made up to 100 ml with deionized water.

The vanillin reagent was obtained by adding 50 ml of *Reagent I* with 50 ml of *Reagent II.*

4 % HCl [Systerm]
4 ml of HCl conc. was added into 90 ml of deionized water. Volume of solution was made up to 100 ml with deionized water.

Catechin [Sigma] standard

0.625 mg, 1.25 mg, 2.5 mg, 5.0 mg, 10 mg, 20 mg and 40 mg of catechin was dissolved in 0.5 ml of deionized water. Volume of each solution was made up to 1 ml with deionized water.

4.2.3.1.4 Reagents for 1,1-diphenyl-2-picrylhydrazyl (DPPH•) Radical Scavenging Activity

80 % ethanol

96 ml of ethanol [Fluka] was added into 20 ml of distilled water. Volume of solution was made up to 120 ml with distilled water.

80 % ethanolic 0.6 mM DPPH• solution

0.0372 g of DPPH• [Sigma] was dissolved in 40 ml of 80 % ethanol. Volume of solution was made up to 50 ml with 80 % ethanol.

Vitamin C [Sigma] standard

Reagent I (25mM of vitamin C)

1.1 g of vitamin C was dissolved in 200 ml of distilled water. Volume of solution was made up to 250 ml with distilled water.

Reagent II (12.5mM of vitamin C)

Add 100 ml of *Reagent I* with 100 ml of distilled water.

Reagent III (6.25mM of vitamin C)

Add 100 ml of *Reagent II* with 100 ml of distilled water.

Reagent IV (3.125mM of vitamin C)

Add 100 ml of *Reagent III* with 100 ml of distilled water.

Reagent V (1.56mM of vitamin C)

Add 100 ml of *Reagent IV* with 100 ml of distilled water.

Reagent VI (0.78mM of vitamin C)

Add 100 ml of *Reagent V* with 100 ml of distilled water.

Reagent VII (0.39mM of vitamin C)

Add 100 ml of *Reagent VI* with 100 ml of distilled water.

Reagent VIII (0.195mM of vitamin C)

Add 100 ml of *Reagent VII* with 100 ml of distilled water.

4.2.3.2 Sample preparation

Frozen *Hylocereus polyrhizus* pulp was thawed 30 minutes prior to experiment. The fruit pulp was squeezed manually through a commercial sieve and the resulting juice was filtered using mira cloth. The pectic substances in the filtered juice were precipitated with 96 % ethanol at a ratio of 2 ml of ethanol to 1 ml of juice. Precipitates were removed using mira cloth and the filtrate was rotary evaporated at 30 °C for 90 minutes. The extract was freeze-dried and frozen at -20ºC prior to analysis.

4.2.3.3 Determination of total phenolic contents and reducing power assay

The total phenolic contents of *Hylocereus polyrhizus* pulp were determined using the Folin-Ciocalteu [Sigma] method according to Bae and Suh, (2007) by calculating the polyphenol concentration from a calibration curve (r^2 = 0.9794) using gallic acid as standard with sample detection at 750 nm. The reducing power assay was also carried out according to Bae and Suh, (2007) with sample detection set at 700 nm.

4.2.3.4 Determination of flavonoid content (Vanillin-HCl assay)

The vanillin-HCl assay which measures the amount of condensed tannins was carried out with modification according to Nakamura *et al*., (2003) from a calibration curve (r^2 = 0.9792). 1 ml of sample was dispensed into a test tube and 5 ml of Vanillin reagent (8 % HCl in methanol/1 % vanillin in methanol, 1:1, v/v) was added to the sample and incubated in water bath for 20 minutes at 30 °C. Samples were measured at 500 nm and the condensed tannins content was expressed as catechin equivalents in mg. Absorbance of samples were calculated according to the following formula:

$$\mathbf{A = (A_s - A_b) - (A_c - A_o)}$$

A_s = sample (1 ml) + vanillin reagent (5 ml)

A_b = methanol (1 ml) + vanillin reagent (5 ml)

A_c = sample (1 ml) + methanol (2.5 ml) + 8 % HCl (2.5 ml)

A_o = methanol (1 ml) + methanol (2.5 ml) + 8 % HCl (2.5 ml)

4.2.3.5 1,1-diphenyl-2-picrylhydrazyl (DPPH$^{\bullet}$) Radical Scavenging Activity

The DPPH$^{\bullet}$ radical scavenging assay was carried out according to Cai *et al*., (2003) with minor modifications. 0.1 ml of sample was reacted with 3.9 ml of 80 % ethanolic 0.1 mM DPPH$^{\bullet}$ solution in a test tube. The test tube was vortexed for 15 s and solution was allowed to stand at room temperature (25 ± 2 °C) for 180 minutes. Absorbance was measured at 515 nm. Antioxidant activity was expressed by (i) calculating the radical scavenging activity: Median effective concentration (EC_{50}) = concentration of sample required to decrease 50 % in absorbance of DPPH$^{\bullet}$ radicals and (ii) inhibition (%) of DPPH$^{\bullet}$ absorbance = ($A_{control} - A_{test}$) X $100/A_{control}$. Ethanol (80%) was used as blank and DPPH$^{\bullet}$ solution without test sample was used as control. A dose-response curve (% inhibition of DPPH$^{\bullet}$ vs concentration of sample) was established and the EC_{50} was determined (r^2 = 0.9809) using vitamin C as a standard. Results were expressed as vitamin C equivalents.

4.3 RESULTS

4.3.1 Qualification of Betacyanins using HPLC Method

In the pigment qualification, only one peak could be confidently identified in the *Hylocereus polyrhizus* pulp sample by comparing the results with the available betanin standard. The peak was observed at a retention of 11.37 ± 0.17 minute with a peak area of 215703.3 ± 28735.2 µV*s This observation corresponds with the results from the betanin standard which gave a similar single peak at retention time 11.73 ± 0.06 minute with a peak area of 78167 ± 20504.8 µV*s (Table 1).

Table 4.1: Retention time and peak area of *Hylocereus polyrhizus* sample and betanin standard

Sample	Retention time (min)	Peak area (µV*s)
Hylocereus polyrhizus	11.37 ± 0.17	215703.3 ± 28735.2
Betanin standard	11.73 ± 0.06	78167 ± 20504.75

4.3.2 Antioxidant Properties

The total polyphenol assay which expresses gallic acid as equivalent showed that there was 101.9 ± 6.27 mg of total polyphenolic compound in 0.50g of dry *Hylocereus polyrhizus* extract (Table 2). The reducing power assay showed that the reducing capability of antioxidants in the *Hylocereus polyrhizus* extract increased from 0.18 ± 0.02 in 0.03g extract to 2.37 ± 0.18 in 0.50g extract (Figure 4.1).

The Vanillin-HCl assay showed that the sample had an equivalent of 2.3 ± 0.2 mg catechin/g dried extract (Table 2). The DPPH• radical scavenging activity determination showed that the effective concentration (EC_{50}) for *Hylocereus polyrhizus* was 2.9 ± 0.4mM vitamin C equivalents/g dried extract (Table 4.2).

Table 4.2: Total polyphenolic content, flavonoid content and DPPH• radical scavenging activity

Sample	Total polyphenolic content (mg/g gallic acid)	Total flavonoid content (mg/g catechin)	DPPH• radical scavenging activity (EC50) (mM vitamin C)
Hylocereus polyrhizus	101.9 ± 6.27	2.30 ± 0.20	2.90 ± 0.40

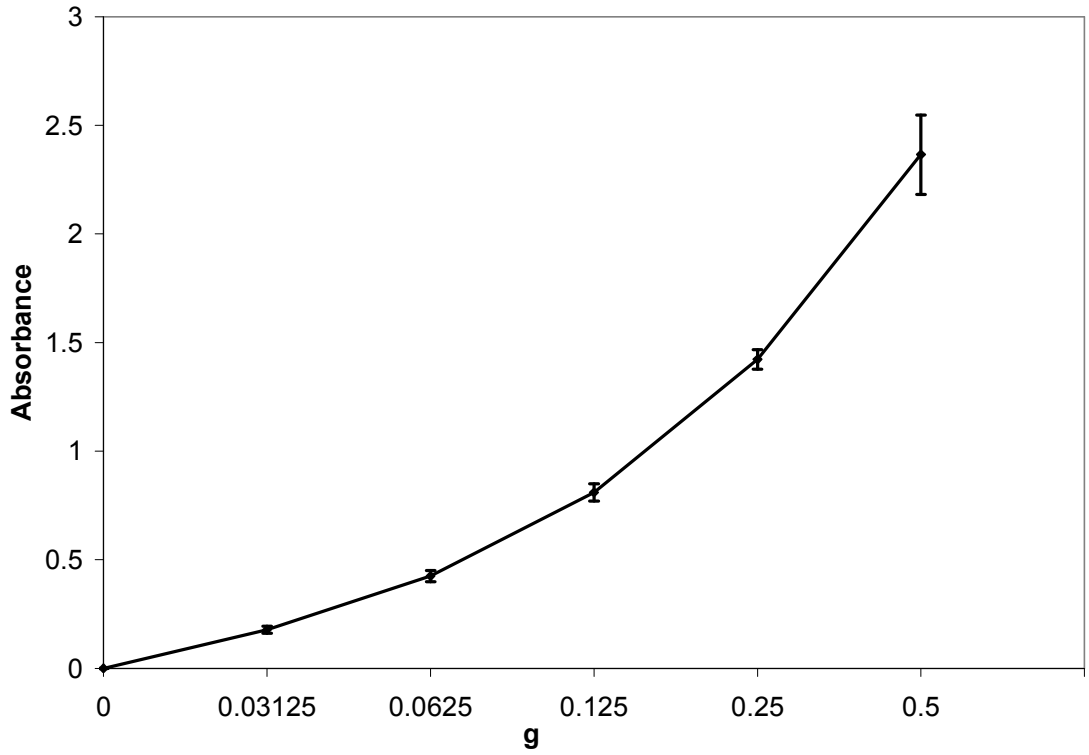


Figure 4.1: Reducing power in *Hylocereus polyrhizus* pulp extract

4.4 DISCUSSION

4.4.1 Qualification of Betacyanin using HPLC Method

Betanin was the first described betacyanin in red beetroot (Gandia-Herrero *et al*., 2005; Wyler and Dreiding, 1957) and is one of the main pigments present in the *Hylocereus* genus. To date, betacyanin identification employs a two-way strategy which is first, using the HPLC to establish the retention time of sample and then secondly, using the electrospray MS-MS and ^{1}H NMR techniques to further elucidate their structure (Cai *et al*., 2006; Wybraniec and Mizrahi, 2002). Wu *et al*., (2006) and Cai *et al*., (2003) have managed to identify three and more peaks in *Hylocereus polyrhizus* using a comparison analysis of pre-determined HPLC retention time and available literatures from many other betalain producing plant like *Amaranthus* sp. and *Beta vulgaris*.

Looking at the results in this chapter, the single peak from the purified *Hylocereus polyrhizus* sample corresponded to the peak eluted from the betanin standard. This direct comparison in HPLC analysis between sample and a specific standard is the best technique when designating or confirming sample's identity using retention time. The attempt to elucidate other betacyanin was not carried out due to the inexistence of other commercially available betacyanin standards except for the betanin standard.

The betanin standard which was imported all the way from Germany had to be purified also using Sephadex G-25 prior to the HPLC analysis to ensure the purity of the standard and it will not contaminate the HPLC column.

In the *Hylocereus* genus, the three common pigments reported are betanin, phyllocactin and hylocerenin along with their 15*R*-isoforms. In these reports, the ratio for betanin : phyllocactin : hylocerenin in *Hylocereus polyrhizus* was reported at 2 : 3 : 1 (Wybarniec and Mizrahi, 2002) and 2 : 2 : 1 (Wybraniec *et al*., 2001). The ratio and results of the betacyanin pigments differs because the HPLC chromatogram and retention time of sample is highly dependable on the HPLC elution conditions and column used. In this chapter, the HPLC system and column used was different from what the abovementioned studies and even though the run time for this HPLC analysis was extended to 25 minutes, there was only one peak observed in all three replicate of the *Hylocereus polyrhizus* sample. The possible explanation is that the *Hylocereus polyrhizus* sample which was purified and separated earlier using the Sephadex G-25 had elucidated only betanin and other lower molecular weight pigments may have been suspended and retained within the Sephadex G-25 matrixs.

As mentioned in the discussion of Chapter 3, betanin alongside with other betacyanin has the ability to regenerate under optimum and suitable conditions. Herbach *et al* (2006c) reported that the betanin were more stable than the acylated structures (phyllocactin and hylocerenin) and betanin regeneration capacity is higher because its degradation mechanism is the least complex making its regeneration most feasible. Betanin regeneration involves resynthesis from its hydrolysis products where condensation of cyclo-DOPA and betalamic acid forms betanin rapidly (Huang and Elbe, 1985). Hence, it can be concluded that, betanin is the most important betacyanin in *Hylocereus polyrhizus* due to its abundance and its superior stability to other betacyanins.

This characterization, identification and isolation of betanin using HPLC analysis is pivotal to future development of *Hylocereus polyrhizus* pigment as potential food dye, in support of the previous experimental chapter, or as a source of antioxidant, which will be discussed in this chapter. Results obtained from this chapter identified the presence of betanin in *Hylocereus polyrhizus* and as the main contributing pigment in the deep purple coloured pulp.

4.4.2 Antioxidant Properties

Bae and Suh (2007) reported that active oxygen species such as hydroxyl (OH•) are thought to be agents that causes oxidative damage and much attention has been focused on active oxygen scavenging agents such as natural phenolics to prevent cell damage. One of the important constituents in *Hylocereus polyrhizus* is the betacyanins which are strong antioxidants (Wu *et al*., 2006; Vaillant *et al*., 2005; Stintzing *et al*., 2003). One of the key structural features responsible for radical-scavenging activity and antioxidative capacity in a compound is the phenolic hydroxyl group (-OH).

In a compound, an –OH group is able to donate the hydrogen atom to free radicals which will then stop the chain reaction of oxidation by the free radicals (Pinedo *et al*., 2007). Looking at the betanin structure in Section 2.7.1, the betanin pigment contains many functional side groups including the –OH groups which makes it potential antioxidant.

According to Somers and Evans, (1977) there are no methods to precisely quantify total phenolics because of its diverse chemical structures. In this study, the Folin-Ciocalteu method was used because it is a generally preferred analytical method for determination of total polyphenol using gallic acid as an arbitrary standard (Lako *et al*., 2008; Rebecca *et al*., 2008b; Wu *et al*., 2006; Cai *et al*., 2004). In this experiment, the Folin-Ciocalteu which consists of sodium tungstate will react with the phenolic compounds giving rise to tungsten blue which is measurable at 750nm. This metal binding reaction is not only used for total polyphenol estimation but also for protein estimation (Folin and Ciocalteu, 1927) and potassium quantification (Abul-Fadl, 1948). Gallic acid was used as a standard because it is a widely distributed phenolic compound in the plant kingdom and has been proposed to be a natural antioxidant (Faried *et al*., 2007) and anticancer agent (Elvira *et al*., 2006). Hence, using gallic acid as a comparison/standard to study the total polyphenol in a sample would be ideal. Lako *et al*., (2007 and 2008) reported total polyphenol contents of common fruits in human diet where there was 110 μg/g in *Musa* sp. (Banana), 150 μg/g in *Ananas comosus* (Pineapple), 260 μg/g in *Carica papaya* (Papaya), 350 μg/g in tomatoes, 670 μg/g in cherries and 3180 μg/g in blueberries. Results from this study showed that the total phenolic content from *Hylocereus polyrhizus* were a significant and good source of polyphenol to be integrated into the human diet.

The reducing power method employs ferric chloride ($FeCl_3$) as an oxidant and ferrous ions are produced from the redox reaction which forms the Perl's Prussion blue complex with trichloroacetic acid, measured at 700nm. The higher the absorbance, the higher reducing power due to the concentration of ferric (II) in the sample.

According to Kumaran and Karunakaran (2006), reducing capabilities of materials are usually associated with the presence of reductones which exhibit antioxidant action by stopping the free radical chain reaction by donating a hydrogen atom.

Looking at the betanin structure and the total polyphenol results from the previous experiment, the *Hylocereus polyrhizus* sample would be a good reducing agent due to the abundance of the –OH group in its structure. Other studies which have employed this method to investigate the reducing power of sample includes: Barros *et al*., (2007) on mushrooms; Sikder *et al*., (2010) on *Spilanthes calva* (Indian medical plant); Karawita *et al*., (2005) on seaweed; and Benzie and Szeto (1999) on tea leaves. In this experiment, no standard was used because the aim of this method was to evaluate and reflect the reduction capability of the sample. The increasing absorbance reading observed in the results justifies the basis of the reducing power assay which is the increase in absorbance reading with the increase of reductones forming ferrous (III) complex in the sample.

Hylocereus polyrhizus samples showed reduction capability even at low concentrations, indicating that antioxidant activities were present and the reducing capability increases with increasing sample concentration.

The Vanillin-HCl assay which is based on the metal-complexing properties and the high affinity of tannins to form protein-tannin complexes is used to detect the presence of condensed tannins and is preferred because of its sensitivity and simplicity. Condensed tannins consist of two or more flavan-3-ol like catechin, epicatechin or gallocatechin.

In the Vanillin-HCl assay, vanillin is protonated in an acidic solution which gives a weak electrophilic carbocation that will react with available hydrogen atoms (Nakamura *et al*., 2003).

The multiple phenolic hydroxyl groups in tannin make it possible to form complexes with proteins (Hagerman *et al*., 1998), metal ions (Foo *et al*., 1997) and macromolecules like polysaccharides (Mueller-Harvey and McAllan, 1992). These formations of complexes in biological functions play a significant role especially in nutrition and the physiology aspect. Moreover, tannins have been reported to be potent antioxidants with protective role to human health and possess properties against heart disease by reducing lipid oxidation (Santos-Buelga and Scalbert, 2000; Ricarda Da Silva *et al*., 1991). The chosen standard for comparison in this experiment is catechin which is the most important member in the tannin family and is ubiquitously distributed in the plant kingdom, for example, in tea, cacao and legumes (Chobot *et al*., 2009; Arunachalam *et al*., 2003).

The vanillin-HCl assay has been used to evaluate the amount of tannins available in: mature beach pea seed (117mg/g catechin equivalent) by Chavan *et al*., (2001); barley (1.95mg/g catechin equivalent) by Aastrup *et al*., (1984) and plums (2.2mg/g catechin equivalent) by Cinquanta *et al*., (2002). When comparing the above studies and findings to the results in this experiment, the amount of catechin equivalent (2.30 ± 0.20mg/g) in the *Hylocereus polyrhizus* sample is significant. This result shows that the *Hylocereus polyrhizus* sample potentially contains abundant tannin structures which are biologically and nutritionally important to human health.

The DPPH• radical scavenging activity is determined by the decrease in absorbance induced by antioxidant, reducing the purple colour of DPPH• radical to a yellow diphenylpicryhydrazine. In both Vanillin-HCl and DPPH• assay, the results indicated that the *Hylocereus polyrhizus* extract contained phenolic contents comparable to standard antioxidant agents like Vitamin C and catechin.

CHAPTER 5
ISOLATION AND IDENTIFICATION OF *MYO*-INOSITOL CRYSTALS FROM DRAGON FRUIT (*Hylocereus polyrhizus*)

5.1 INTRODUCTION

By definition, a crystal is a three dimensional atomic, ionic, or molecular structure consisting of periodically repeated, identically constituted and congruent unit cells. In order to isolate or grow crystals from a solution, the strategy would be to bring the solution into a supersaturated state and when the solution environment returns to equilibrium, the solute molecules are forced into a solid state, forming crystals (McPherson, 1990).

This is possible because when there is a thermodynamic requirement to minimize free energy in a supersaturated solution (which is returning to equilibrium), the most logical strategy for the forced out solute molecules would be to form greatest number of bonds and interactions to minimize mobility and prevent further loss of energy. This result in same solute molecules assembling into fixed lattice and forming its most stable/pure state.

In chemistry, crystallization is the choice technique to purify a product or to ensure the purity of a starting material/sample. This technique was employed as early as 1940 by Scott-Moncrieff and Sturgess to isolate anthocyanin pigments from *Verbena hybrid* and in 1942, by Cohen to isolate and crystallize plant viruses and protein marcomolecules. Other more recent studies which have utilized this technique to identify compounds or proteins (enzymes) include: Chu *et al*., (2010) on adenylate isopentenyltransferase in *Humulus lupulus*; Hofmann *et al*., (2006) on allene oxide cyclase in *Arabidopsis thaliana*; and Schrader *et al.,* (2003) on plant sulfite oxidase.

In this chapter, crystals will be obtained based on the basic steps of crystallization: choose a solvent, dissolve desired solute, filter, cool, wash and dry. After crystallization, the process of identifying the obtained crystal will take place. To obtain a detailed and precise description of a compound in mathematical terms, the only technique which does this is the X-ray Crystallography (McPherson, 1990). This technique allows for the determination of vast macromolecular structures and its functions are practical, reliable and rapid. The principle of this technique is based on the interference pattern produced as x-rays passes through the three dimensional (3-D) grid/lattice of a crystal and defining three axis and all the angles between them. The results will give an accurate identification of the entire structure and every elements present in a particular sample. The only pre-requisite to utilize this technique is that the sample must be crystallized first and good quality crystals are needed to obtain high resolution X-ray diffractions.

As mentioned in Chapter 4.1, one of the choice methods to identify a compound is by using the High Performance Liquid Chromatography (HPLC) which provides a unique retention time of a compound for qualification.

Another modern chromatographic technique which is useful in identifying compounds is the Liquid Chromatography Tandem Mass Spectrometry (LC-MS/MS) which has become a very powerful analytical tool (Thomas, 2001). With the introduction of electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI) as the atmospheric pressure ionization (API) interfaces, LC-MS/MS can detect trace amounts of target sample in a complex mixture due to its sensitivity, selectivity and specificity (Dams and Huestis, 2003).

For example, a study carried out by Eloff and Rosemann (2009) used LC-MS/MS to distinguish very low levels of toxic and non-toxic pyrrolizidine alkaloids in plants and Takatsuki *et al*., (2003) reported levels of acrylamide in many different types of processed foods using LC-MS/MS.

The Nuclear Magenetic Resonance (NMR) Spectroscopy is regarded as a tool for screening a compound without separating or breaking down the structure to determine its stereochemical purity by observing the chemical shifts resulting from the resonance and nuclear spin of each atom present (Bartos and Görög, 2008). The first two NMR spectra observed was of H_2O by Bolch, Hansen and Packard; and of paraffin by Purcell, Torray and Pound in 1945 through proton signals coming from the nuclei of hydrogen atoms (Rigden, 1986).

The basic NMR technique detects signals from ^{1}H and ^{13}C isotopes when placed in a magnetic field and there are several types of two-dimensional (2D) NMR techniques which include: Correlation spectroscopy (COSY), J-spectroscopy, exchange spectroscopy (EXSY), nuclear overhauser effect spectroscopy (NOESY) and total correlation spectroscopy (TOCSY).

In this chapter, the abovementioned four identification techniques were employed to investigate the crystals which were successfully isolated from the *Hylocereus polyrhizus* extract. All four techniques are important to confirm the identity of the crystals and each individual result generated from each technique is then used to support and correlate with the other results obtained.

5.2 MATERIALS AND METHODS

5.2.1 Sample preparation

Fruits from the five weeks after anthesis stage were obtained from a local farm situated 20 km from the laboratory. Fruits were halved and peeled manually. 450g of fruit pulp was immersed in 500 ml of 60% methanol and left to stand at room temperature (24 ± 2°C) for 2-3 hours. Sample was warmed at 60-70°C for a final 20 minutes and the *Hylocereus polyrhizus* pulp was removed by filtering sample through one layer of mira cloth.

5.2.2 Purification

500ml of the filtrate was concentrated *in vacuo* at 40-50°C using a rotary evaporator (BUCHI ROTAVAPOUR R-210) for about 150 minutes into a slurry syrup. The slurry was thoroughly washed with 100ml of ether and then 100ml of ethanol. The solvent was decanted off carefully and the insoluble portion was dissolved with 30ml of warm water. The sample was filtered using mira cloth and 30ml of ethanol was added into the filtrate. Sample was kept at 4 - 6°C for 3-4 hours and filtered using mira cloth to remove any impurities. The filtrate obtained was evaporated to dryness *in vacuo* at 40-50°C.

5.2.3 Crystallization

The resulting mass was dissolved with 25ml of cold water and 50ml of ethanol was added. The sample was left at 4 - 6°C for 7 days for crystallization. The resulting crystals were removed from solution by filtering through one layer of mira cloth and air dried before stored in a glass vial.

5.2.4 Crystal Structure Determination using X-Ray Crystallography

Data for the crystal structure was collected on a Bruker SMART APEX diffractometer by using a SADABS - Bruker Nonius detector. For all measurements, graphite-monochromated Mo Kα radiation was used. The crystal was cooled down by a cold dry nitrogen gas stream and the temperature stability was within 100K. The structure was solved by direct methods (Sheldrick, 1990) and refined (Sheldrick, 1997) by full-matrix least squares on F^2 (all data) by using the SHELXL-97 program package. All non-hydrogen atoms were refined anisotropically. Atomic scattering factors had values incorporated in the computer programs. The H atoms bound to C aromatic atoms were refined using a riding model with C–H = 0.93 Å $U_{iso}(H) = 1.2\ U_{eq}(C)$. The remaining H atoms were located from different Fourier map and refined isotropically with $U_{iso}(H) = 1.2\ U_{eq}$ (C or O). Crystallographic data (excluding structure factors) for the structure in this paper have been deposited with the Cambridge Crystallographic Data Centre (CCDC) and allocated the deposition number CCDC-784857.

5.2.5 Qualification of Crystal Purity Using High Performance Liquid Chromatography

HPLC was carried out by using LC-10A UFLC system with a SIL-HT automatic sample injector (Shimadzu, Kyoto Japan) with a UV/VIS detector (SPD-20A) equipped with a 2.1 X 150mm Zorbax Eclipse XDB C18 3.5µm column**.** Mobile Phase A was deionised water and mobile phase B was 0.1% formic acid in methanol running with gradient mode. The gradient program began with 95% B then ramped to 5% B at 3.00 minutes and then return to 95% B at 3.10 minutes and this condition was held for further 9.00 minutes.The detection was set at 546nm and 10 µl of sample was allowed to elute through the system for 9.00 minutes at a flow rate of 0.3 ml/min and column temperature was set to 40 °C.

5.2.6 Liquid Chromatography Tandem Mass Spectrometry (LC-MS/MS) Analysis

The LC-MS /MS system consisted of an LC-10A UFLC system with a SIL-HT automatic sample injector (Shimadzu, Kyoto Japan) and an API 5500 Q-Trap LC-MS/MS system (Applied Biosystems, Lincoln Centre Drive, Foster City, CA, USA). The separations were performed on a 2.1 X 150mm Zorbax Eclipse XDB C18 3.5µm column. A gradient elution was employed on the column at 0.25ml/min with mobile phase A (deionised water) and mobile phase B (0.1% formic acid + HPLC grade methanol) running a linear gradient in which the percent of mobile phase B was held at 10% for 0.10 minutes, then linearly increased to 70% during the next 1.90min.

The mobile phase B was held at 90% for 0.50 minutes and then immediately returned to 10%. The analysis time was 5.0 minutes per sample. The typical injection volume was 1.0µl. All the mass spectra were collected on a AB Sciex 1024945-AC linear ion trap quadrupole LC/MS/MS mass spectrometer. The Turbo Ion Spray (TIS) interface was operated in positive ion mode and temperature was maintained at 350°C while the voltage was set at 4500 V. The curtain gas was set at 20 and the declustering potential (DP) at 140 V while the nebulizer (GS1) and TIS (GS2) gases were set at 40 psi and 35 psi respectively. The Collision Induced Dissociation (CID) gas was set at medium and the collision energy was set at -20eV for the Neutral Loss scan experiments. Unit mass resolution was used in all the experiments. LC-MS/MS data were acquired using Analyst Software Version 1.5.1.

5.2.7 Nuclear Magentic Resonance (NMR) Analysis

The obtained crystal was dissolved in 0.75 ml of $^{2}H_2O$ (99.996% ^{2}H; MERCK) and the solutions were filled into 5mm diameter NMR tubes. All spectra were accumulated at room temperature (24 ± 2°C) using a JEOL LAMBDA 400 NMR spectrometer equipped with a Delta software. The resonance was 399.8MHz for ^{1}H and 100.5MHz for ^{13}C. For one-dimensional ^{1}H spectra, excitation pulses of 45° were employed and the spectral width was set to 7993.6 Hz which yields a digital resolution of 0.26 Hz/point.

The one-dimensional ^{13}C spectrum was obtained with 30° excitation pulses and spectral width of 27100.3 Hz Hz which yields a digital resolution of 0.83 Hz/point. For the 2-Dimensional (2D) NMR $^{1}H/^{1}H$ analysis also know as Correlation Spectroscopy (COSY), the correlated spectra was acquired using resonance of 399.8MHz and spectral width of 7993.6 Hz which gave a digital resolution of 15.61 Hz/point.

5.3 RESULTS

5.3.1 Purification and Crystallization

After the purification and crystallization process, the *Hylocereus polyrhizus* sample which was kept at 6 ± 2°C gave yield to approximately 2.0g of crystals which were tetragonal and the crystals formed prism or double pyramid like structures. Figure 5.1 shows the aerial view of the air-dried crystals before storage and Figure 5.2 show a 10X view under a microscope.

5.3.2 X-Ray Crystallograhy Analysis

Unit-cell parameters, as measured with a Bruker SMART APEX diffractometer were: a = 6.6226 (3) Å, b = 12.0462 (5) Å, c = 18.8942 (8) Å, α = 90.00, β = 93.98, δ = 90.00. The crystal density was reported as 1.592 cm^{-3} while the crystal volume was recorded at 1503.69(11) $Å^3$. These parameters are in good agreement with those reported by Rabinowitz and Kraut (1964), indicating that the crystals isolated from *Hylocereus polyrhizus* pulp are *myo*-inositols. The chemical formula and molecular weight of the crystals were $C_{12}H_{24}O_{12}$ and 360.31, showing that there are two units of *myo*-inositol per asymmetrical unit. This is further supported by Figures 5.3, 5.4 and 5.5 which were generated using the enCIFer Software from CCDC. From the X-Ray analysis, the crystals were reported have monoclinic cell setting with $P2_1/n$ space group while the unit cell dimension is 0.35 X 0.25 X 0.15. All the cell parameters results were obtained from 18816 reflections with wavelength set at Mo $K\alpha$ radiation, λ = 0.71073 Å using a Graphite monochromameter.

Figure 5.1: Example of crystals obtained after storage at 4 – 6°C for 7 days.

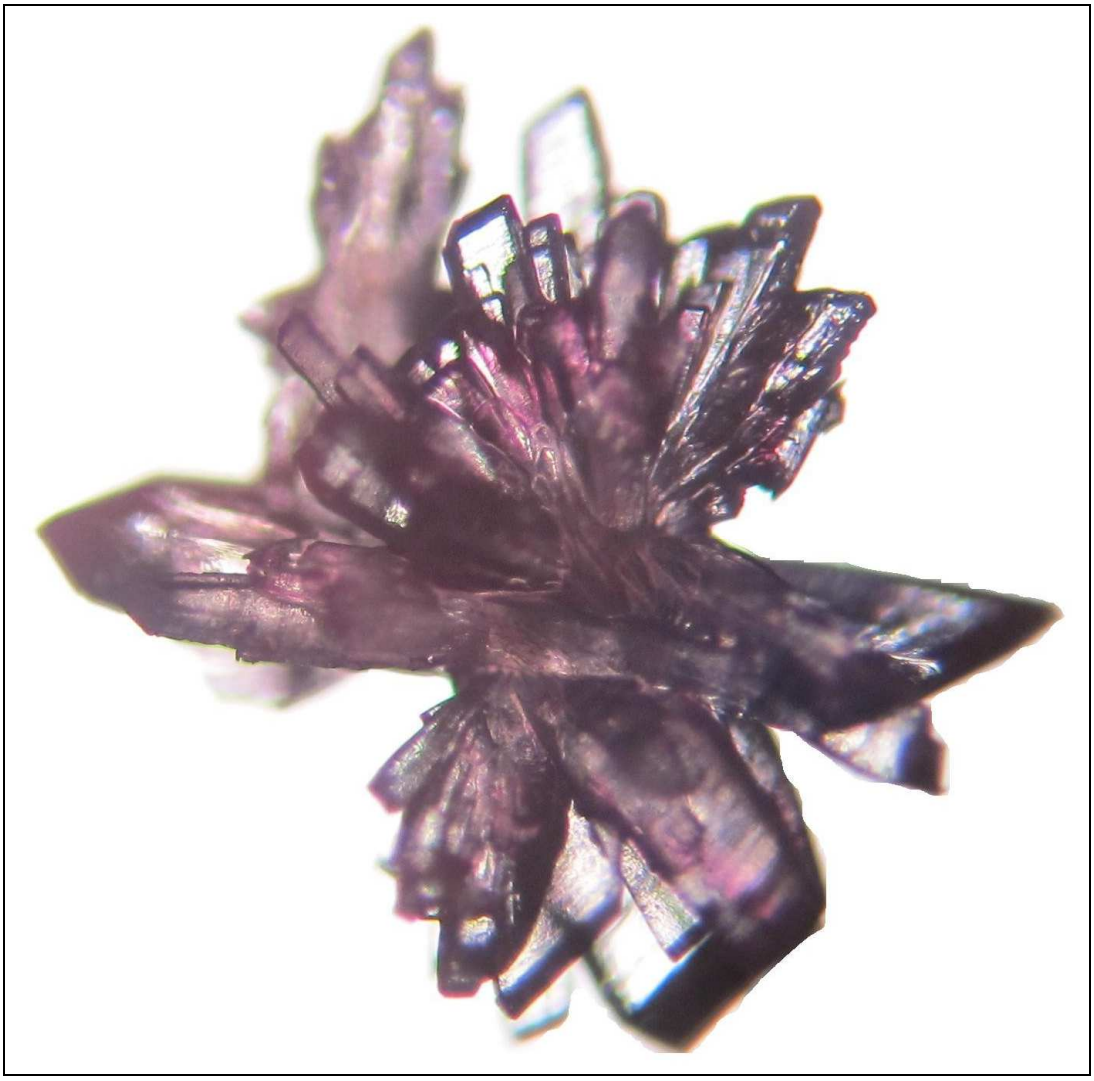


Figure 5.2: View of crystal under a microscope at 10X magnification

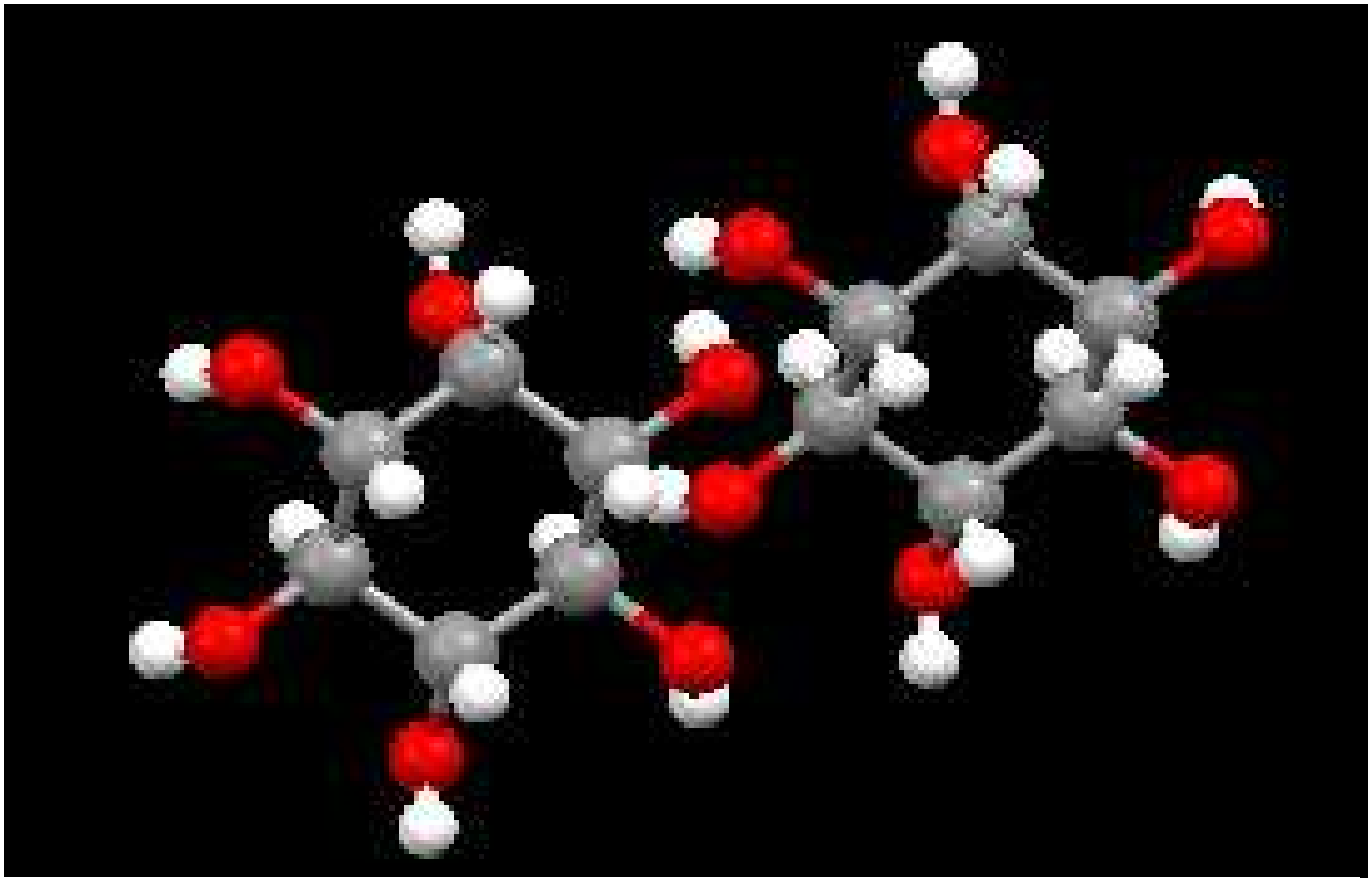


Figure 5.3: Orthographic projection of the asymmetric unit of *myo*-inostiol along the *c* axis

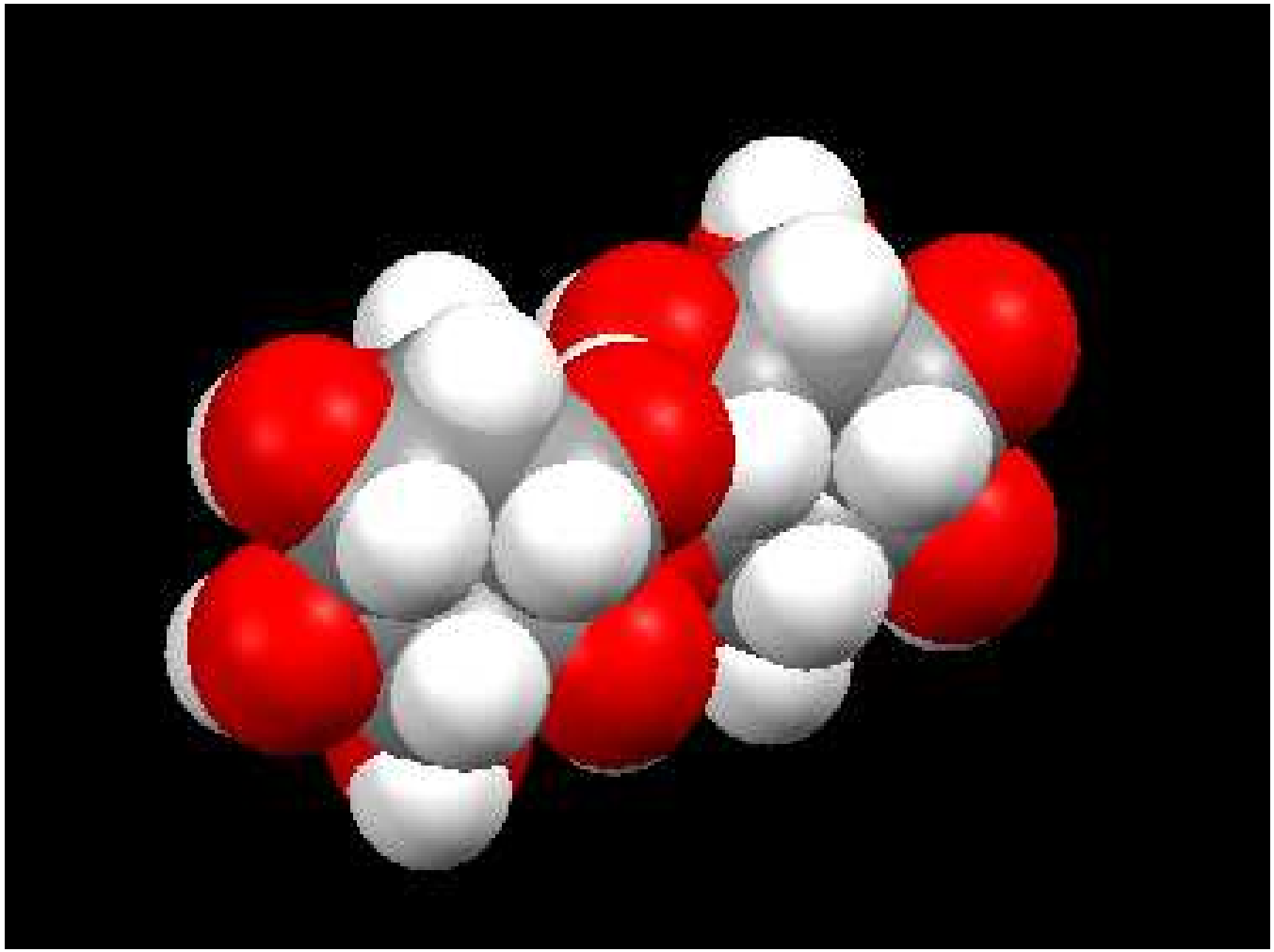


Figure 5.4: A spacefilled orthographic projection of the asymmetric unit of *myo*-inostiol along the *c* axis

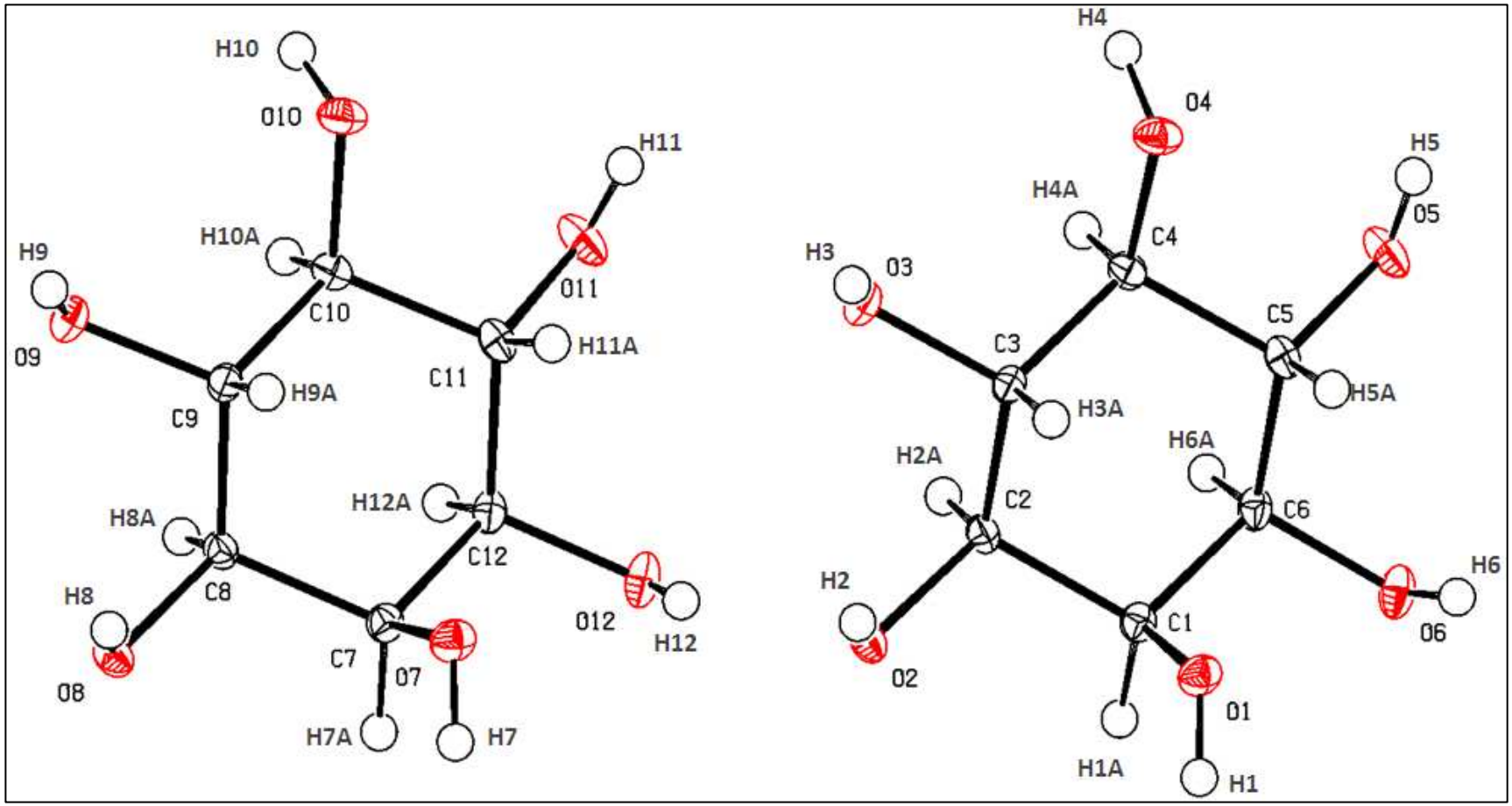


Figure 5.5: A complete labelled ellipsoid plot of the two units of the asymmetric unit *myo*-inositol ($C_{12}H_{24}O_{12}$)

5.3.3 Qualification of Crystal Purity using High Performance Liquid Chromatography

In the crystal purity qualification using HPLC, only one peak was detected in the HPLC chromatogram indicating that the isolated *myo*-inositol crystal does not contain any other compounds or suffers any contamination throughout the experiment. The sample peak was observed at 4.8 minute with a peak area of 41232 μV*s (Figure 5.6). The result from this purity check was a qualification to proceed later on to using Liquid Chromatography Tandem Mass Spectrometry (LC-MS/MS) for a more accurate analysis on the crystal.

5.3.4 Liquid Chromatography Tandem Mass Spectrometry (LC-MS/MS) Analysis

Using the the LC-MS/MS analysis, the retention time, precursor ion and product ion of the crystal sample was determined (Figure 5.7 and Table 5.1) alongside the *myo*-inositol standard sample (Figure 5.8 and Table 5.1). The Multichannel Analyser (MCA) scan on the *myo*-inositol sample shows clearly the precursor ion and product ion which confirms the identity of the crystal sample.

Table 5.1: LC-MS/MS analysis on *myo*-inositol standard and crystal sample

Analyte Peak Name	***Analyte RT***	***Precursor ion (M/Z)***	***Product Ion (M/Z)***
Myo-inositol Standard	1.24	179.00	87.00
Crystal Sample	1.24	179.00	87.00

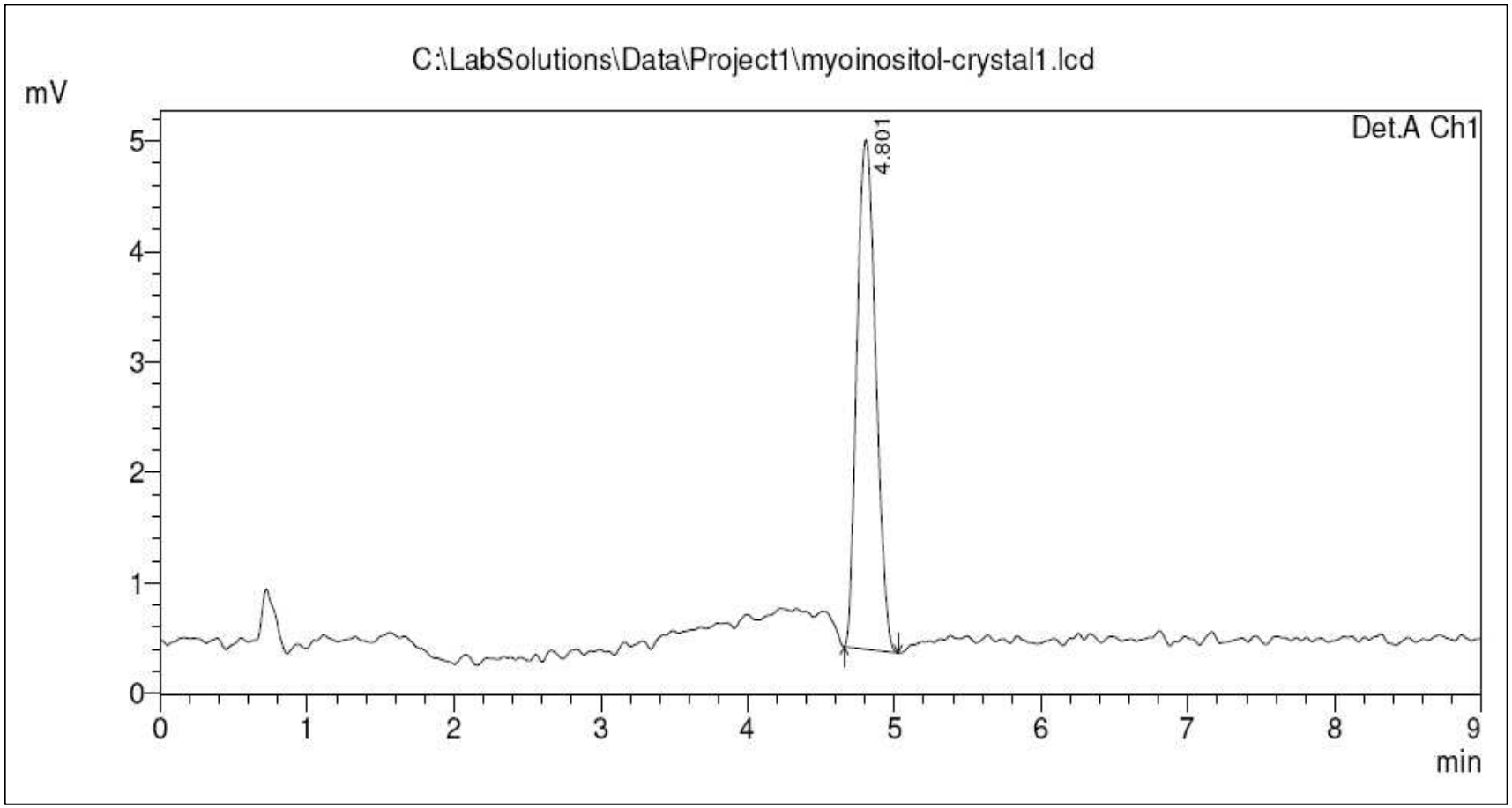


Figure 5.6: Crystal purity check using HPLC where the peak and retention time of the *myo*-inositol sample was observed at 4.8 minute

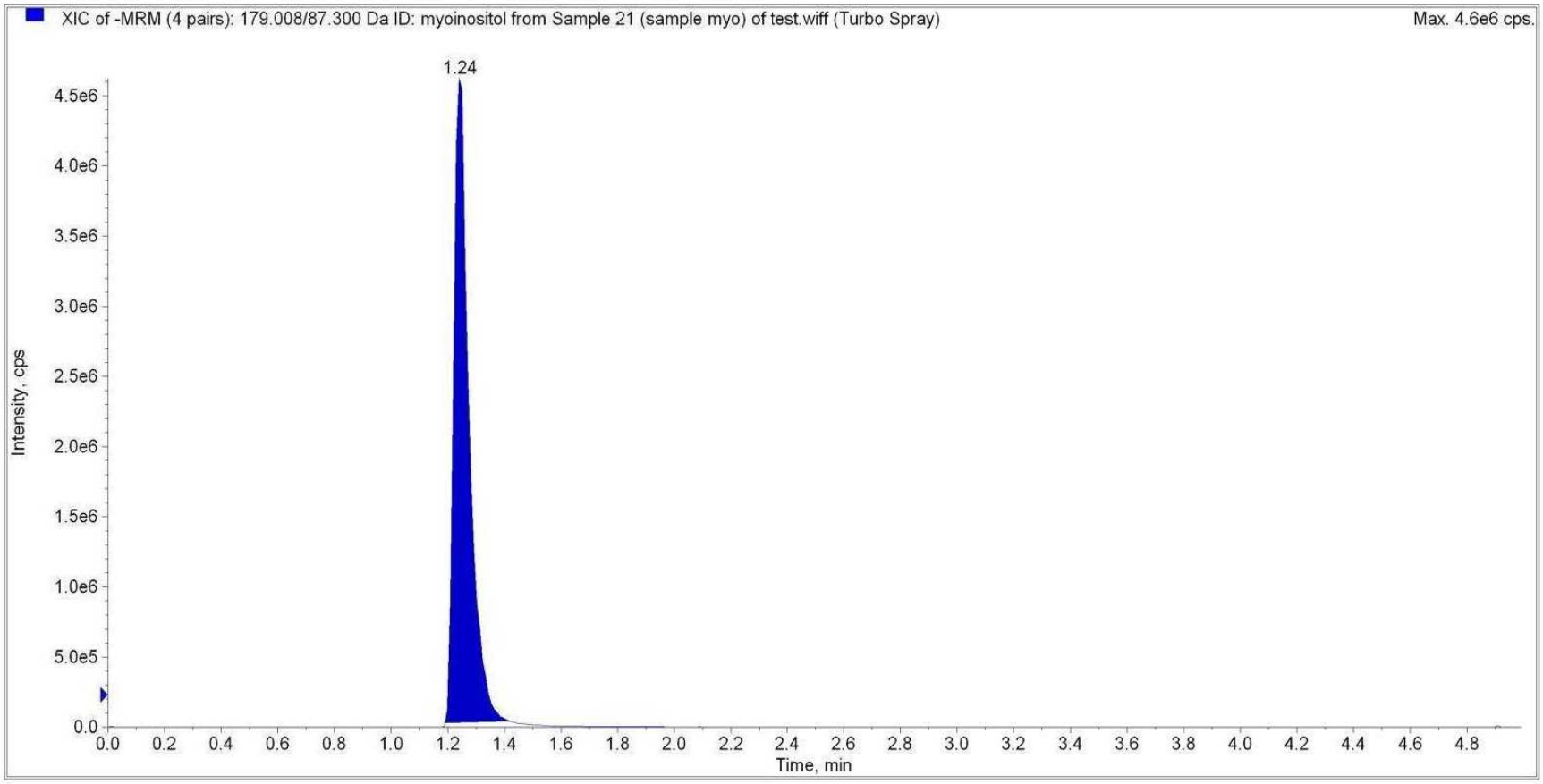


Figure 5.7: Retention time of crystal sample observed using LC-MS/MS. One clean peak was obtained at 1.24 minute

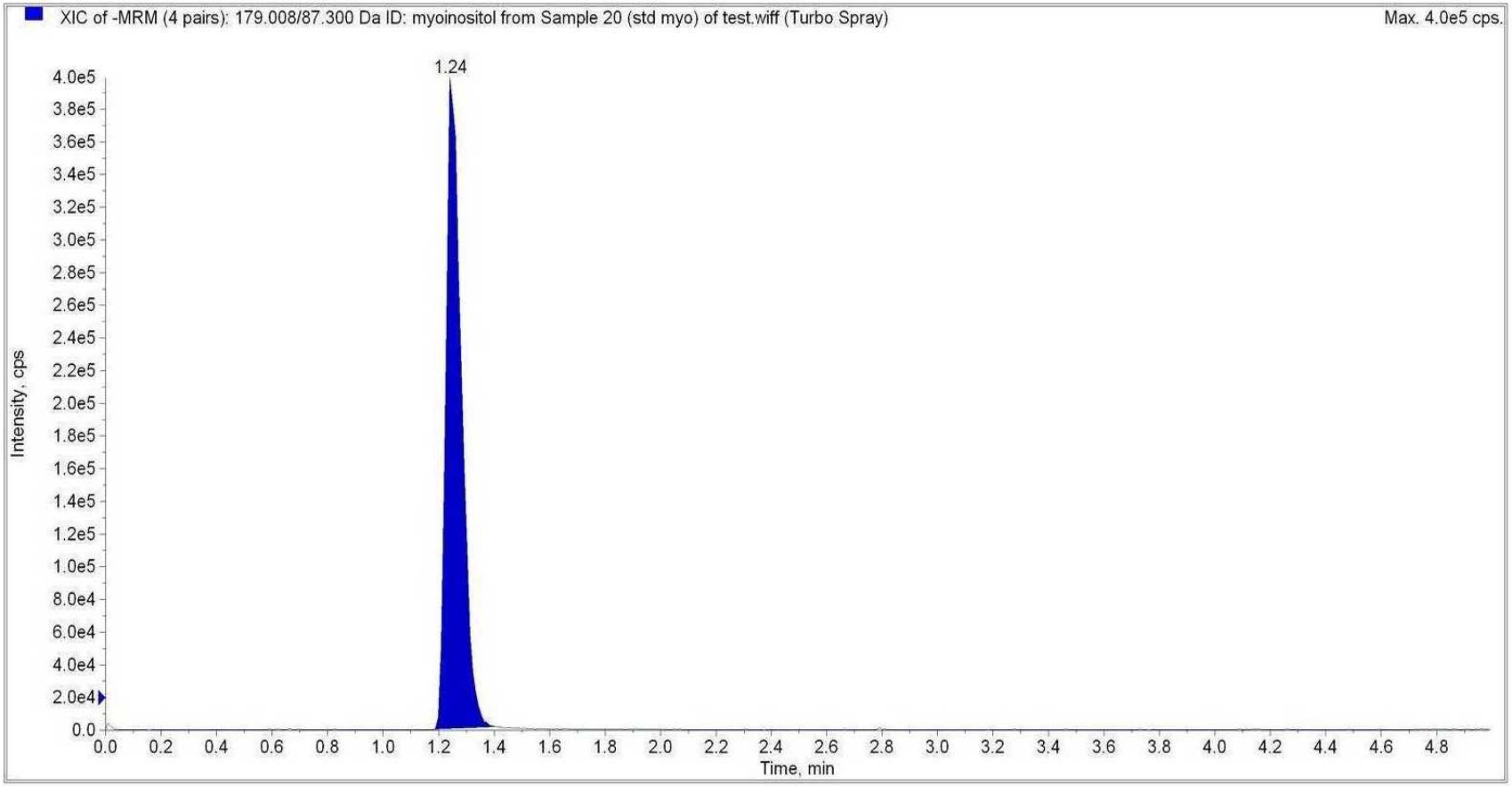


Figure 5.8: Retention time of *myo*-inostiol standard observed using LC-MS/MS. One clean peak was obtained at 1.24 minute

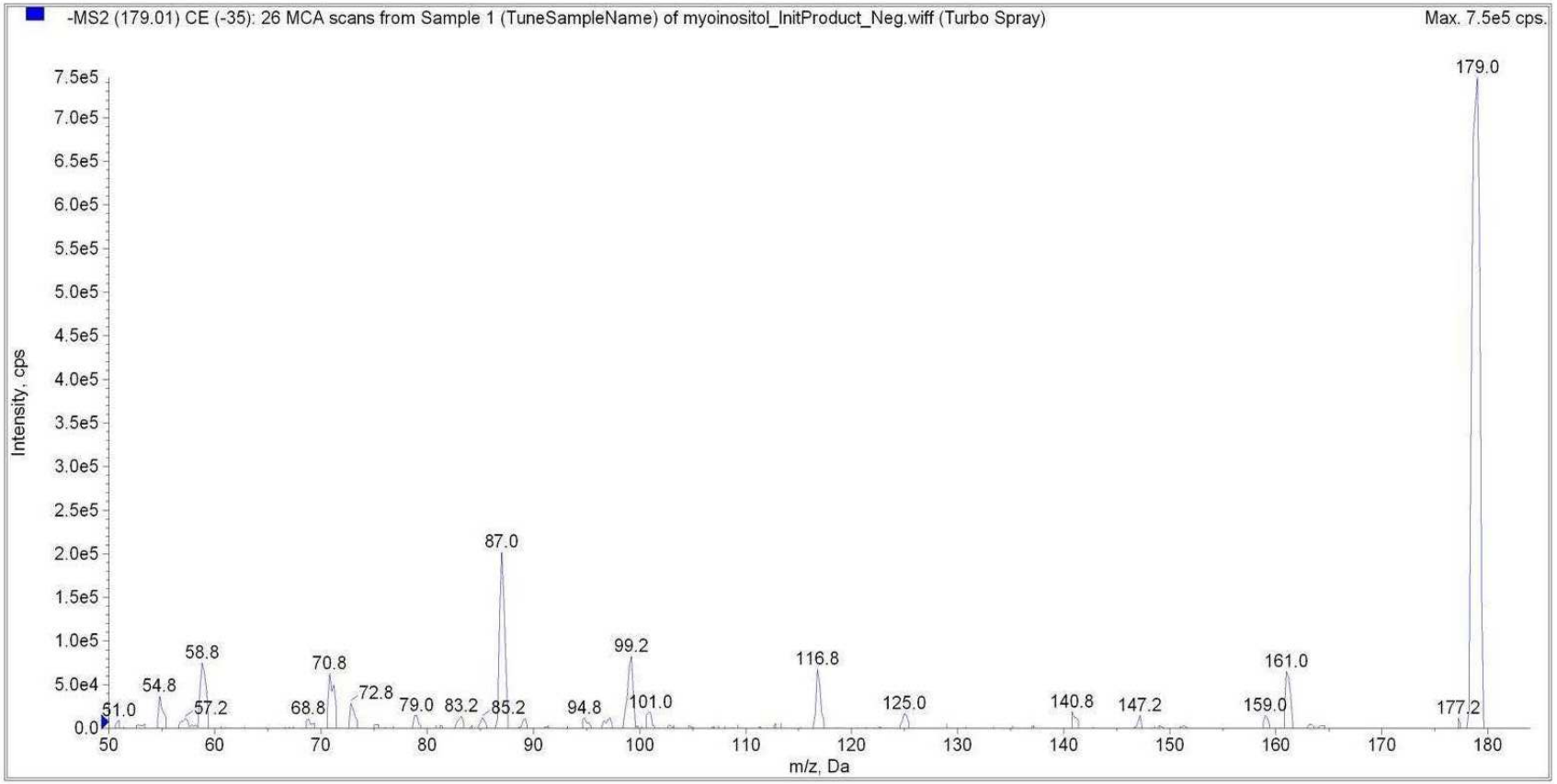


Figure 5.9: The Multichannel Analyser (MCA) scan from *myo*-inositol standard showing the precursor ion at 179.0 *m/z* and the product ion at 87.0 *m/z*

5.3.5 NMR Analysis

In the one-dimensional (1D) ^{1}H spectra analysis, the proton separation on the chromatogram was at a ratio of 1 : 2 : 2 : 1 (Figure 5.10) and the one-dimensional (1D) ^{13}C spectra analysis showed a ratio of C at 1 : 2 : 1 : 2 (Figure 5.11). The two-dimensional (2D) $^{1}H/^{1}H$ spectra using COSY showed the correlation and relativity of each proton to each other in the sample (Figure 5.12).

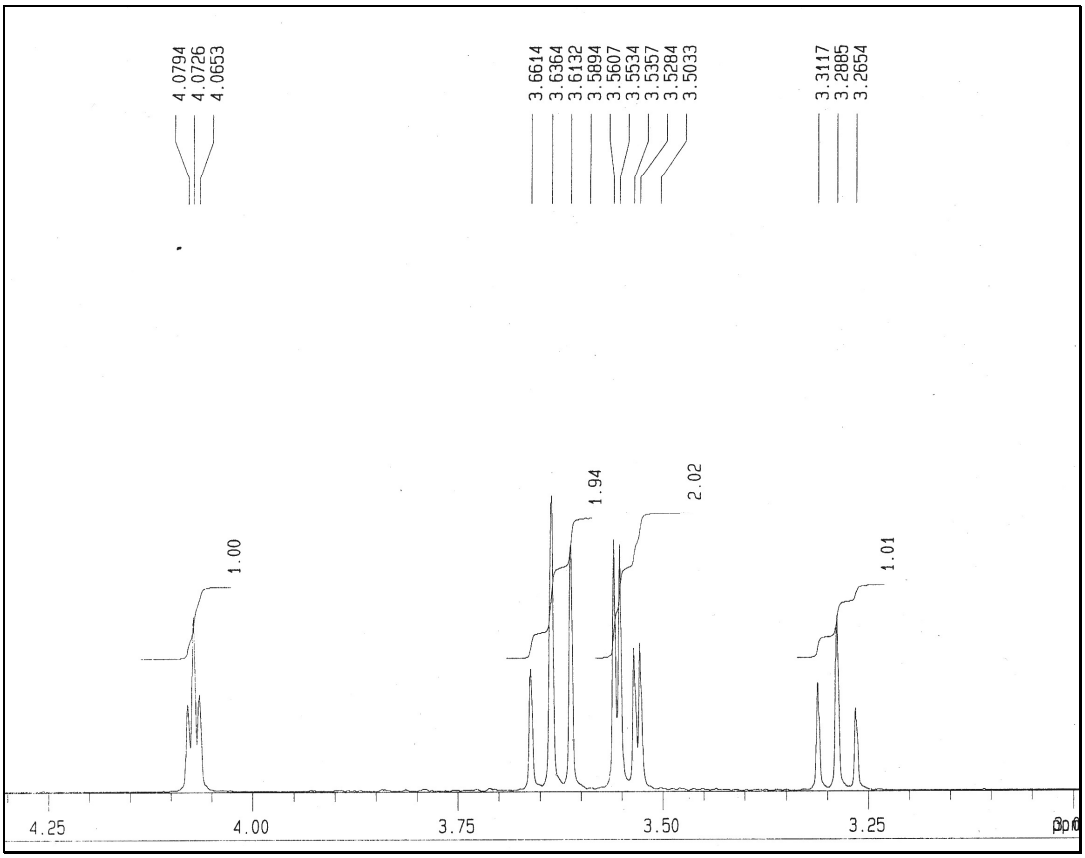


Figure 5.10: One-dimensional ^{1}H spectra of crystal sample

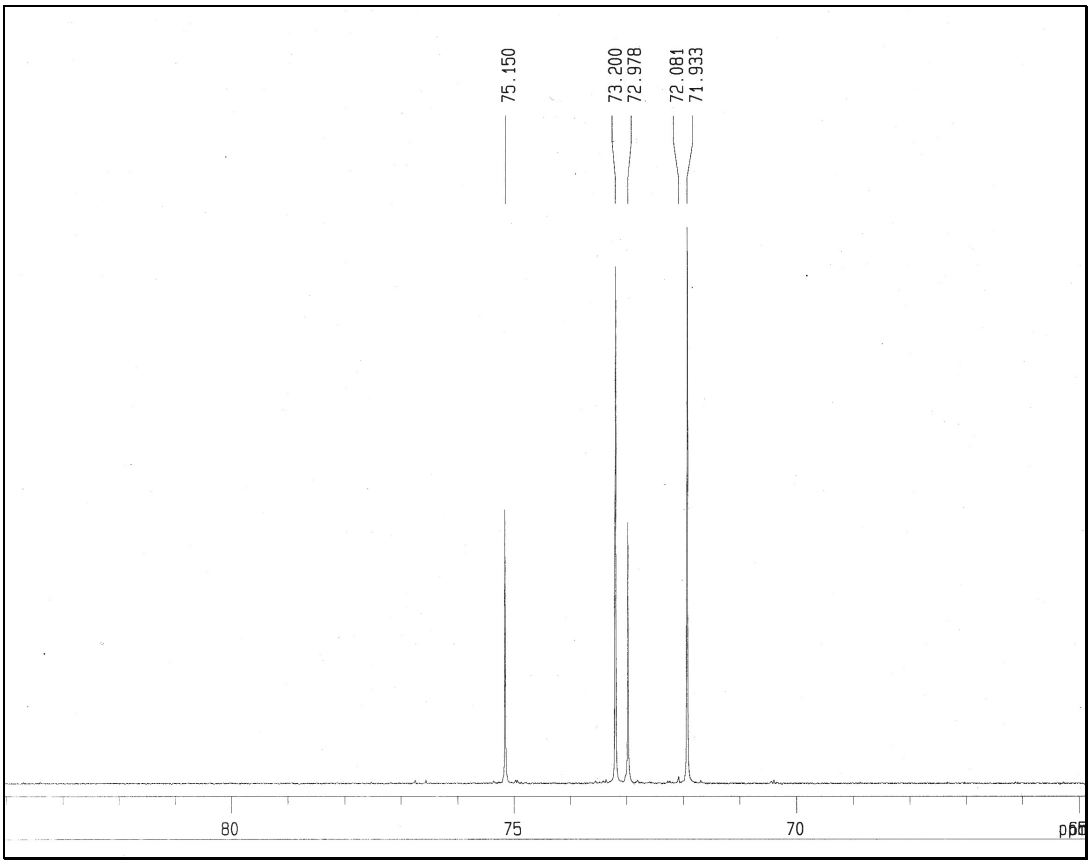


Figure 5.11: One-dimensional ^{13}C spectra of crystal sample

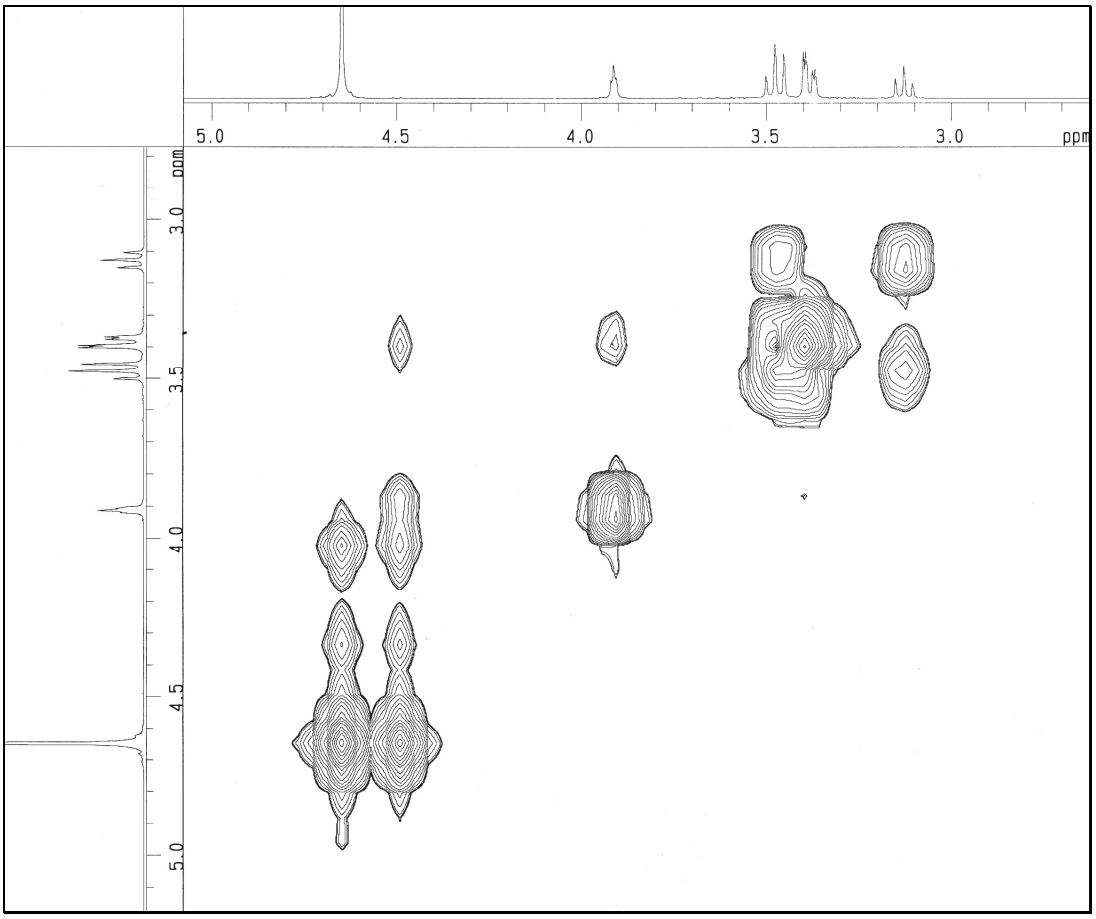


Figure 5.12: Two-dimensional $^{1}H/^{1}H$ spectra using COSY analysis on crystal sample

5.4 DISCUSSION

Generally, the pinkish crystals which were obtained had individual long slender tetragonal shape where each crystal needle has four sides with its length longer than its width along one axis. The crystals were aggregated together to form pyramid like structures with prisms sticking out from an almost round base. Since there were no previous reports or studies indicating any crystals were isolated from *Hylocereus polyrhizus*, the crystals needed to be identified as there are numerous possibilities of what the crystals actually are based on its exterior description only. The crystallization and purification methods carried out in this chapter were modified and combined from works of Saito *et al*., (1964) and Pucher *et al*., (1937) on the principles of forcing solute molecules to form a solid state (crystals) in a supersaturated solution.

The crystallization and purification methods carried out in this chapter were modified and combined from works which isolated red anthocyanin pigments from cornflower (Saito *et al*., 1964) and red betanin crystals from *Beta vulgaris* (Pucher *et al*., 1937) on the principles of forcing solute molecules to form a solid state (crystals) in a supersaturated solution. The red anthocyanin crystals isolated by Saito *et al*., (1964) were in form of individual needles while the red betanin pigments obtained by Pucher *et al*., (1937) were in crystalline powder form. Due to the very different description and characteristics of the crystals obtained in this study compared to the two references used, it is of utmost interest to determine the identity of the abundant crystals which may or may not be pigment crystals.

This difference can be due to several reasons which include that the method adopted from Saito *et al*., (1964) was targeting at anthocyanins which are structurally different from betacyanins, hence, the method was not suitable to isolate betanin crystals from *Hylocereus polyrhizus*; while Pucher *et al*., (1937) was working on *Beta vulgaris* which contains both betacyanins and betaxanthins, hence, the method was not suitable to isolate betanin crystals from *Hylocereus polyrhizus* which only contains betacyanin.

The closest example of crystals isolated from member of the Cactaceae family was the work carried out by Saenz *et al*., (1998) where they isolated glucazones, fructosazones and sugar crystals from the pulp *Opuntia ficus indica* L. Other works on the members of the same family were done on other parts of the plant: Silicon dioxide from the epidermis and hypodermis of *Stenocereus* sp. (Loza-Cornejo and Terrazas, 2003); Calcium oxalate from the leaves of *Opuntia* sp. (McConn and Nakata, 2004); and alkaloids from the branches of *Trichocereus terscheckii* (Reti and Castrillón, 1951). Since there are very limited works done on obtaining crystal of any forms or from any parts from the members of the Cactaceae family, there is a very strong incentive in solving the classification of the crystals isolated from the fruit pulp of *Hylocereus polyrhizus*.

There is no possible way of guessing what type of crystal were extracted from the procedure as it could be a type of protein, amino acid, carbohydrate complex, polysaccharide or pigment. The diversity of crystals which have been isolated from other fruits and plants can be shown in the following examples: Pelargonidin 3-monoglucoside, an anthocyanin, from strawberry (Meschter, 1953); Jacalin, a plant base lectin, from jack fruit (Bourne *et al*., 2002); Pectin methylesterase, a cell wall degradation enzyme, from carrots (Johansson *et al*., 2002); Prolycopene and pro-γ-carotene, isomers of carotenoid, from firethorn plant (Zechmeister and Schroeder, 1942); and potassium chloride from spiderwort (Brizuela *et al*., 2007). Looking at the copious probability and there is a potential that the isolated crystals are valuable, the following identification procedure and process are necessary.

The choicest technique as the first step in identifying the crystals from this study is definitely the X-ray crystallography which is able to determine the unique atoms arrangement and chemical bonds within a crystal through the diffraction of X-ray beams which produces a three dimensional representation based on electron density. As such, X-ray crystallography remains the chief method for characterizing atomic structure of many compounds and resolving new materials which is not resolved by other methods. From the initial X-ray crystallography analysis, the unknown crystals obtained through the experiment were identified as *myo*-inositol through a unit-cell parameter match to the report by Rabinowitz and Kraut (1964).

Myo-inositol crystals have a very distinct lattice motif (non-equivalent axes; *a, b*, and *c*; and one non-equivalent oblique angel at β) translated from its single unit-cell parameter repeated many times to form three dimensional arrangements which then generates other supporting data like crystal density, volume, dimension and space group.

All these data are unique and exclusive to individual crystals, making subsequent identification methods easier and targeted. The results generated from this analysis also supports the earlier observation made on the typical appearance of *myo*-inositol which have slender and long prism, a typical characteristic of crystals from the monoclinic system.

Myo-inositol, a six carbon cyclic polyalcohol (Alam, 1971), is highly involved in plant metabolism and the following list is just to name a few processes which it is involved in: seed dessication, osmo-regulation (Nelson *et al*., 1999), nutrient storage, membrane biogenesis, senescence, auxin physiology, fertilization as well as synthesis of cell wall uronosyl and pentose units (Loewus and Murthy, 2000). Other than that, Lorence *et al*., (2004) reported that *myo*-inositol is one of the important precursors in the biosynthetic pathway of L-ascorbic acid which is the major antioxidant in plant cells and a cofactor for many enzymes involved in secondary metabolites production.

Due to its significance, *myo*-inositol has been reported as early as 1980 by Clements and Darnell where it is commonly found in citrus fruits, beans, grains and nuts. In their study on 487 different types of food, the highest amount of *myo*-inositol were in 4.07 mg/g of *myo*-inositol in *Prunus domestica*, 3.55 mg/g in *Cucumis melo*, 3.07 mg/g in *Citrus sinensis*, 4.40 mg/g in *Phaseolus vulgaris,* 2.83 mg/g in *Pisum sativum*, 2.74 mg/g in bran, 0.42 mg/g in oatmeal, 2.78 mg/g in *Prunus dulcis* and 1.34 mg/g in *Arachis hypogaea.* In this study, 2 g of *myo*-inositol was recovered from 450 g of *H. polyrhizus* pulp and this translates to about 4 mg/g of *myo*-inositol in one fruit, which is comparable to its content in *Phaseolus vulgaris*, a very important global agricultural produce.

While *myo*-inositol is a crucial compound in regulating plant biochemistry and physiology, it is also found concentrated in cerebrospinal fluid in the range of 100 – 500 μM and it increases to 10 mM or more in brain cells (Harwood, 2005). The significance of *myo*-inositol to animals was documented in 1979 where Whiting *et al*., discovered that a decrease in peripheral motor-nerve-conduction velocity is associated with a decrease in nerve *myo*-inositol content in diabetic rats. This discovery prompted many studies to establish the relations between *myo*-inositol and its effects to human health.

The functions and roles of *myo*-inositol in human has been linked to bipolar disorder (Deranieh and Greenberg, 2009), production of L-chiro-inositol and D-chiro-inositol in insulin action (Larner, 2002), multiple sclerosis (Brex *et al*., 1999), Alzheimer's disease (McLaurin *et al*., 1998) and regulation of the sorbitol pathway in diabetic patients (Thomas *et al*., 1994). The probable reason given to this phenomenon is that *myo*-inositol actively participates in the synthesis of membrane phospholipids which affects neuronal plasticity and synapse formation in neuron cell (Deranieh and Greenberg, 2009). This is where Shimon *et al*., (1997) and Silverstone *et al*., (2005) reported the possible connection between inositol depletion in the frontal temporal lobes and patients with bipolar disorder. Other roles of inositol and its common derivative *myo*-inositol, include as potent regulators for a large number of hormones, growth factors and neurotransmitter.

As a result its significance, *myo*-inositol is used as part of treatment for diabetes mellitus (Clements and Reynertson, 1977); obsessive-compulsive disorder (Fux *et al*., 1996); status epilepticus (Solomonia *et al*., 2010); psoriasis and eczema; and recently a report that *myo*-inositol may improve metabolic syndromes in postmenopausal women (Giordano *et al*., 2011).

This finding of *myo*-inostiol, an essential compound for many metabolic processes in organisms puts *Hylocereus polyrhizus* in a position as a new alternative for many health/diet recommendations and potential source for pharmaceutical products. Subsequent analyses were essential to ensure the purity of the crystals obtained and to confirm its identity.

The High Performance Liquid Chromatography (HPLC) is typically used to separate, identify/qualify and quantify organic compounds and this method is applicable in many areas including pharmaceutical, food and beverage, forensics, environment as well as industries. Since HPLC is more sensitive compared to other chromatographic methods as it can detect low concentration of compounds, as low as nanogrammes (Ferreira *et al*., 2004), this method is employed in this study to confirm the purity of the crystals from *H. polyrhizus* before putting the sample through LC-MS/MS with a highly sensitive system to avoid possible column contamination or confusion in identifying the sample. One of the most crucial factors which can affect the ability for crystals to form is the presence of trace impurities during the growth process.

From the result (Figure 5.6), the sample gave a prominent single peak which gives a confident resolution that the crystals were perfectly isolated and other compounds were not attached to its crystal formation even when the crystals were pinkish in colour.

When the crystals were dissolved in deionised water prior to the HPLC analysis, the solution was colourless. The only explanation available to how a *myo*-inositol crystal which is supposed to be colourless retained a pinkish hue is that in the crystallization process, the *myo*-inositol prisms trapped a microscopic amount of betacyanin pigment which is red in colour, during its aggregation into pyramid like structures.

Due to the angel and manifold reflection caused by numerous individual *myo*-inositol prisms in one unit of aggregated *myo*-inositol crystal (Figure 5.1), the entire unit appears to be pink.

Theoretical, if indeed the *myo*-inositol crystal trapped some pigments in its formation, it will severely affect its purity and stability because in any crystallization process, any contamination or addition of substance to a crystal will affect its molecular structure and also its stability in storage. However this was not observed in the HPLC chromatogram and the crystals which were kept in room temperature remained stable with the pinkish hue even after 6 months. This could indicate that the amount of possible pigments trapped inside the crystal formation is negligible and did not compromise the level of purity and stability of the *myo*-inositol crystals. In order to ascertain this claim, the next technique used is the Liquid Chromatography Tandem Mass Spectrometry (LC-MS/MS) and a final molecular conformation using the Nuclear Magnetic Resonance (NMR).

The Liquid Chromatography Tandem Mass Spectrometry (LC-MS/MS) is becoming an indispensible technique in laboratories analysis as it combines the physical separation capabilities of HPLC with high analytical sensitivity, specificity and accuracy of mass spectrometric detection (de Jong *et al*., 2011). In this study a *myo*-inositol standard was used to establish the mass spectrum (Figure 5.9) and to obtain the distinctive precursor of 179 m/z and product ion of 89 m/z for this compound. Results (Figure 5.7 and Figure 5.8) showed that the crystal sample was an exact match to the standard used where the ion peak was observed at 1.24 min with same precursor and product ion mass as well. This can very well assure that the method employed in obtaining *myo*-inositol crystals from the pulp of *Hylocereus polyrhizus* is viable and the crystal is of high level of purity. This leads to the final qualification technique, the Nuclear Magnetic Resonance (NMR) analysis.

An important aspect derived from the ^{1}H and ^{13}C NMR analysis in this study is the purity and qualitative information of molecular formula on the *myo*-inositol crystals from the pulp of *Hylocereus polyrhizus.* In terms of quantitative and exact assignment of each signal on the spectrums, validation by comparison with other techniques is needed but in this study, the purpose is only to identify the crystals. The crystals were confirmed as *myo*-inositol by looking at the NMR spectrums obtained which shows six protons and six carbons per unit cell.

CHAPTER 6
TOXICOLOGY ANALYSIS OF BETACYANIN EXTRACTED FROM DRAGON FRUIT (*Hylocereus polyrhizus*)

6.1 INTRODUCTION

In the food and beverage industry, it is of utmost importance to have stringent regulations on quality and safety because any contaminated product which reaches the food network causes serious consequences from the producer, investors, brand and most significantly to the consumers. On top of only carrying out biochemical properties analysis only for the food supply chain, agricultural and chemical residues and microbiological contaminants are monitored regularly (Lester, 1994).

Under the United States Food and Drug Administration (USFDA), food safety encompasses food borne illnesses, contaminants and adulteration. The usual food borne pathogens which are targeted in food safety analysis are the common yeast and mold (Ferrati *et al*., 2005), coliforms (Patrick, 1951), *Escherichia coli* (Lateef *et al*., 2004) and *Salmonella* sp (Tasnim *et al*., 2010). In terms of food contaminants and adulteration, the substances which are usually targeted for monitoring includes organochlorines insecticides, organophosphorus insecticides (Lambropoulou and Albanis, 2002) and other heavy metal contaminants like arsenic, lead (Tufuor *et al*., 2011), mercury, cadmium (Bingöl *et al*., 2010), tin and antimony.

These analysis and test parameters are essential because there is a complex and extensive processing of any raw material before becoming a finished product in the food industry. There are many production steps such as planting, harvesting, post-harvest management, sorting, washing, transporting, filtration, fermentation, sterilization and many other possible processes involved in finishing a product to make it marketable and storable for consumers. In each step and process, there are possible entry points where microorganisms and contaminants can affect the quality and safety of a food product.

When it comes to standards and regulations in food safety, the World Health Organization (WHO) and Food and Agriculture Organization (FAO) developed the Codex Alimentarius Commission in 1963 which develop food standards, guidelines and related texts such as codes of practice. This guideline is followed closely by countries which have their individual food safety protocols and standards which are compliant to WHO/FOA standards. For example, International Organization for Standardisation (ISO) by Switzerland, Australian Standards (AS), British Standards (BS) and the Malaysian Standards (MS).

Even though different standards exist with slightly different methodology differing from country to country, the food safety parameters and acceptable limits are universal. For example, a method used in this chapter, the total bacterial count by the Australian Standard (AS 1766.1.3, 1991) uses 37°C while ISO 4833:2003 uses 30°C instead but both methods are aimed at estimating the number of colony-forming units (CFUs) in a food sample.

In this chapter, all the methods for the microorganism analysis follows recommended standards accepted in Malaysia which are the Australian Standards (AS), British Standards (BS) and International Organization for Standardisation (ISO).

To ensure the industry's quality and safety requirements, it is important to utilise reliable, accurate and precise instruments to execute the methods recommended by the various standards. In the heavy metal analysis, the Inductively Coupled Plasma-Optical Emission Spectrometry (ICP-OES), Atomic Absorption Spectrometry - Cold Vapour (AAS - Cold Vapour) and Atomic Absorption Spectrometry – Hydride Generation (AAS - Hydride Generation) will be used while the pesticides analysis will be utilizing the Gas Chromatography – Mass Spectroscopy (GC – MS).

The Inductively Coupled Plasma-Optical Emission Spectrometry (ICP-OES) is an analytical technique used for the detection of trace elements with excellent detection power which provides reliable results and rapid determination of targeted analytes (Thiel and Danzer, 1997). The versatility of this instrument with very sensitive detection ability has been widely applied to detect low level of non-metals in waste oils (Krengel-Rothensee *et al*., 1999), heavy metals in soils and sediments (Bettinelli *et al*., 2000), cadmium in drinking water (Cerutti *et al*., 2003) and major and minor elements in milk and infant formula (McKinstry *et al*., 1999).

Atomic Absorption Spectroscopy (AAS) is a technique for determining the concentration of a particular metal element in a sample. The sample would first need to be turned into an atomic gas to emit a unique spectrum for detection. The Cold Vapour AAS is usually used for detection of mercury while the Hydride Generation AAS is usually employed for arsenic and selenium detection (Welz and Melcher, 1985). The difference between the two techniques is the manner in which the samples are generated into atomic gas.

The Cold Vapour AAS reduces mercury to a free atomic state by reacting the sample with a strong reducing agent in a closed system while the Hydride Generation AAS heats samples in air or acetylene flame to create free atomic state in an external system. Both techniques have been used for detection of heavy metal in various food products: Mushrooms (Demirbas, 2000); cereal, fruit juices, milk, meat products (Mindak and Dolan, 1999); and a variety of seafood (Plessi *et al*., 2001; Ubillus *et al*., 2000).

The Gas Chromatography – Mass Spectroscopy (GC – MS) identifies different and multiple substances or trace elements in a test sample. In food and environment safety, this method is widely utilized for the detection of various types of herbicides and pesticides in water (Thurman, 1992); soil (Richter *et al*., 2003); fresh fruits, vegetables, milk (Sheridan and Meola, 1999); and fruit juices (Albero *et al*., 2003). Pesticides have been given special attention in the food industry because it is easily transported over long distance through the atmosphere, soil and water into non-targeted host due to its volatility and resistant to degradation (Whiteaker and Prather, 2003). The prevalence of pesticides in the environment is as high as 99.7% as only 0.3% reached it targeted pest (van der Werf, 1996) and its residue can also be found in processed food product (Sannino *et al*., 2004) and even deposited in animal and human tissue (Egan and Roburn, 1965).

In this chapter, the toxicology studies on the pigments extracted from *Hylocereus polyrhizus* will include the microorganism analysis where the aforementioned common food borne pathogens and also a general analysis of total bacterial count; heavy metal analysis and the pesticide screening. All the analyses are important to ensure that the purified extract from *Hylocereus polyrhizus* is safe and will determine its quality as natural food colourant for the food industry.

6.2 MATERIALS AND METHODS

6.2.1 Plant material

Fruits from the five weeks after anthesis stage were obtained from a local farm situated 20 km from the laboratory. Fruits were halved and peeled manually. 800g of fruit pulp was squeezed manually through a commercial sieve and the resulting juice was filtered using mira cloth. The filtrate was then centrifuged at 14,000g at 4°C for 15 minutes and the supernatant was used for the subsequent filtration steps.

6.2.2 Filtration

The supernatant was filtered using a FAVORIT filtration assembly and a Rocker 300 vacuum pump through a series of filters as described in Figure 6.1. The final volume of pure and clarified *H. polyrhizus* extract was 200ml. The extract was kept at 4°C for 24 hours before sent for further analysis.

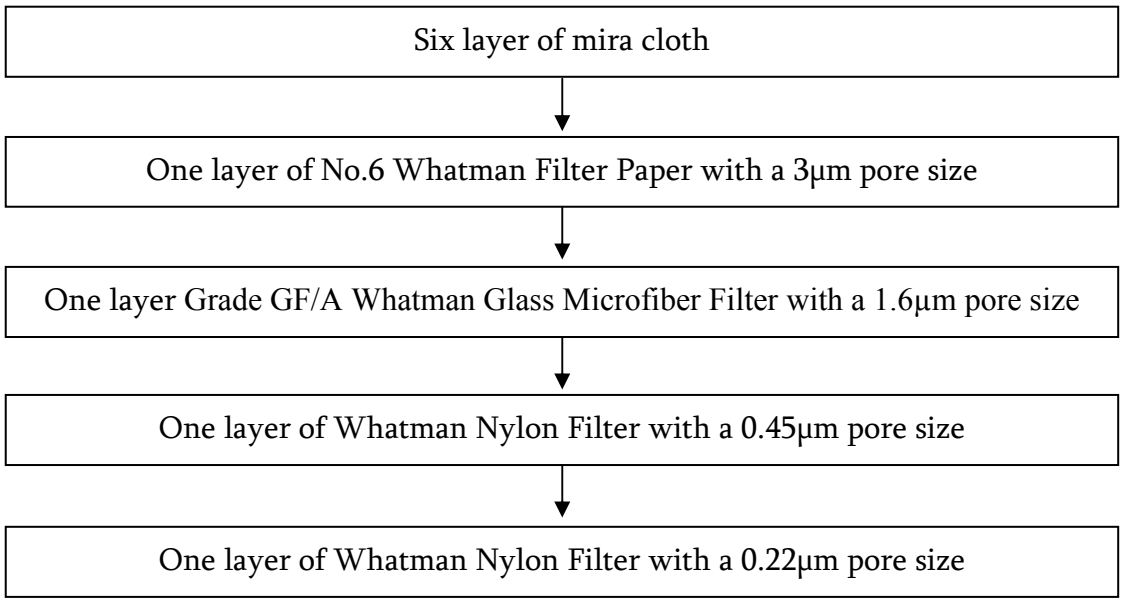


Figure 6.1 A customized filtration method to obtain a pure and clarified *H. polyrhizus* extract

6.2.3 Toxicology Analysis

The 200ml of pure *Hylocereus polyrhizus* extract was sent to Consolidated Laboratory (M) Sdn Bhd for subsequent toxicology analysis. The choice of analysis was based on the recommended assessments for natural product by USFDA before the product is commercialised and marketed.

6.2.3.1 Total bacterial count

The total bacterial count was carried out according to the Australian Standard (AS 1766.1.3, 1991) as described earlier in Chapter 6.1 where CFUs were estimated using the pour plate technique and samples were incubated for 48 hours at 37˚C.

6.2.3.2 Yeast and mould colony count

The colony count of yeasts and mould was carried out according to the Australian Standard (AS 1766.2.2) where it estimates the number of colony forming units (CFUs) in a sample using dichloran Rose Bengal chloramphenicol agar (DRBC) and dichloran 18% glycerol agar (DG18).

6.2.3.3 General coliform group detection

The coliform detection was carried out according to the British Standard (BS 5763: Part 3) which is similar to the ISO 4831 (International Standard). The general method is to inoculate the sample into liquid media with decimal dilutions, incubated and number of coliforms was calculated per millilitre of sample.

6.2.3.4 Escherichia coli detection

The *Escherichia coli* (a member of the colliform bacteria group) detection was carried out according to the British Standard (BS 5763: Part 3) and similar to the methods in the ISO 7251 (International Standard). The standard aims to detect *Escherichia coli* using the liquid-medium culture technique and calculated after incubation at 37°C and then 44°C.

6.2.3.5 Salmonella sp. detection

The *Salmonella* sp. detection was carried out according to the British Standard (BS 5763: Part 4) and similar to the methods in the ISO 6579 (International Standard) where it specifies a horizontal method for the detection of *Salmonella*, including *Salmonella* Typhi and *Salmonella* Paratyphi. Generally, the sample was pre-enriched in a non-selective liquid medium and then incubated. The pre-enriched sample was subcultured into two different selective enrichment media and incubated. The enriched samples were inoculated onto two selective agar media and incubated. *Salmonella* colonies were confirmed via the biochemical confirmation and serological confirmation.

6.2.3.6 Heavy metal analysis

The detection of Lead (Pb), Cadmium (Cd), Tin (Sn) and Antimony (Sb) were carried out using the Consolidated Laboratory in-house method, CL-TM-01-020, based on AOAC 985.01 and AOAC 922.02.2005: AACC 40-70, Volume II, 9th Edition using the Inductively Coupled Plasma-Optical Emission Spectrometry (ICP-OES).

The detection of Mercury (Hg) was carried out using an in-house method, CL-TM-01-033, based on AOAC 971.21, 2005 using the Atomic Absorption Spectrometry - Cold Vapour (AAS - Cold Vapour) while arsenic (As) was detected using an in-house method, CL-TM-01-012, based on AOAC 986.15 and AOAC 971.21, 2005 using the Atomic Absorption Spectrometry – Hydride Generation (AAS - Hydride Generation). The results for heavy metal analysis were expressed in mg/kg.

6.2.3.7 Pesticide screening

The sample was also subjected to pesticide screening using the Gas Chromatography – Mass Spectroscopy (GC – MS) to detect 52 different types of organochlorine insecticides and 136 types of organophosphorus insecticides which are usually used in agriculture. The lists of the insecticides are attached in Appendix 7 and Appendix 8.

6.3 RESULTS

6.3.1 Microorganisms Analysis

In the microorganism analysis, only the total bacterial count showed a result of 750 cfu/g while all the test parameters on yeast and mould colony count, general coliform group detection, *Escherichia coli* and *Salmonella* sp. did not detect the presence of the respective microorganisms.

Test Parameters	Unit	Result
Total Bacterial Count	cfu/g	750
Yeast and Mould Colony Count	cfu/g	Not Detected (< 100)
Coliform	MPN/g	Not Detected (< 3)
Escherichia coli	MPN/g	Not Detected (< 3)
Salmonella in 25g	-	Absent

Table 6.1 Microorganisms detection from the toxicology analysis carried out on *Hylocereus polyrhizus* sample

6.3.2 Heavy Metal Analysis

In the heavy metal analysis, the results showed that none of the tested heavy metals were detected in the sample.

Test Parameters	Unit	Result
Lead (Pb)	mg/kg	Not Detected (< 0.01)
Cadmium (Cd)	mg/kg	Not Detected (<0.01)
Tin (Sn)	mg/kg	Not Detected (<0.40)
Antimony (Sb)	mg/kg	Not Detected (<0.50)
Mercury (Hg)	mg/kg	Not Detected (<0.01)
Arsenic (As)	mg/kg	Not Detected (<0.05)

Table 6.2 Heavy metal analysis from the toxicology analysis carried out on *Hylocereus polyrhizus* sample

6.3.3 Pesticide Screening

In the pesticide screening, the targeted 52 organochlorine pesticides and 136 organophosphorus insecticides was not detected.

The official result for all the analysis carried out on the *Hylocereus polyrhizus* extract from Consolidated Laboratory (M) Sdn Bhd is attached as Appendix 5 and Appendix 6.

6.4 DISCUSSION

In this experimental chapter, the filtration techniques has been customised and optimised to obtain the best clarity of the *Hylocereus polyrhizus* extract. The objective is to eliminate the usage of temperature and also avoid the necessity to use any enzymatic treatment to obtain the pigment extracts. The centrifugation attempt successfully removed most of the mucilage or larger particle from the pulp, making subsequent filtration steps much easier. The six layers of mira cloth ensured that any remaining large particles were removed before using the cellulose, glass microfiber and nylon membrane filters.

Whatman No.6 filter paper with a 3µm pore size is a standard cellulose filter which is used in general laboratory filtration to remove particulates for analysis. At this point, the fouling of the cellulose filter paper did not occur and it is highly possible that majority of the complex polysaccharide was successfully removed via centrifugation at 14,000 rpm for 15 minutes. In addition, the cellulose filter paper further ensured that any other particulates were removed from the extract to facilitate its filtration through the glass microfibre and nylon membranes with finer pore sizes. Larger particles like mucilage, crystalline precipitates or fibres from pulp can increase filtration time and exposes the extract to oxidation and additional stress due to the pressure from the vacuum system.

The 1.6 µm glass microfiber filters are made from 100% borosilicate glass and was used to remove turbidity and larger microorganisms like amoeba (12.0-15.0µm), yeast cells (2.0-8.0µm) and puffball spores (3.0µm) which may be present in the extract. The final type of filter used was the 0.45µm and 0.22µm nylon membrane filters.

This type membrane is suitable for the purpose of biological preparations where solutions like tissue culture media, microbiological media and buffers requires minimal contamination and microorganisms free. Using the 0.22µm nylon membrane, most bacteria and microorganisms would have been filtered out. For example, cocci (0.5µm), bacilli (1.0µm), *Escherichia coli* (0.5µm) and colloids (0.06-0.30µm). As a result, the 200ml extract of pigment from *Hylocereus polyrhizus* was extremely clear and after storage at 4°C for 24 hours prior to analysis by Consolidated Laboratory, the extract did not show any signs of sedimentation, cloudiness or any precipitation.

The usage of membrane filters for filtration was developed as early 1935 and since then has been widely used in many industries including biotechnology, microbiology and food and beverage applications where ultimate sterility is targeted. Currently, there are several types of membrane filtration: microfiltration, ultrafiltration, nanofiltration and reverse osmosis where microfiltration and ultrafiltration are often associated with the term "membrane filtration" (USEPA, 2005). Each of these technologies utilizes a membrane barrier that allows the passage of water but removes contaminants.

Ultrafiltration which is used widely in the fruit juice industry uses temperature as high as 50°C for the initial aroma recovery step and then the juice is treated with enzymes to degrade pectin and starch completely before filtration (Stutz, 1993). The disadvantage of this technique is that the system needs constant physical cleaning and resuming an entire complex filtration system can cause pressure variation, instable temperature and inconsistent flow velocity. If the system is not halted and cleaned, the gel layer formed on the membrane will compromise protein, lipids and solid filtration (Ge´san *et al*., 1995).

In a study carried out by Merin and Shomer (1999) on orange juice, the proteins and pectins which are earlier treated with heat will coagulate and interact with the subsequent filtration membranes causing reduction in filtration efficacy and fouling of the filter membrane. Furthermore, fruit juice like yellow passion fruit are sensitive to any amount of heat exposure during processing or stabilization methods like pasteurization which causes loss in volatile and aromatic compounds (Vaillant *et al*., 1999).

Microfiltration technique of clarifying juice extract do not involve the application of heat and its product is of superior quality which can be used as juice blends, liquor, soft drinks and other downstream items. The general problem associated with this technique is that when it is applied to juices rich in pulp, fouling of the membrane filters by polysaccharides occurs frequently and this significantly reduces filtration efficiency (Yu *et al*., 1986).

The proposed solution to this problem was to increase the concentration of enzymes responsible for hydrolysation of polysaccharides (Cheryan and Alvarez, 1995) but in a report by Vaillant *et al*., (1999), their enzymatic treated fruit juice showed a decrease in aromatic strength.

Looking at the disadvantages of the existing methods available commercially for filtration of fruit juices, the techniques developed in this experimental chapter could be adopted in commercialization of *Hylocereus polyrhizus* extract without putting the extract through complex preparatory steps to ensure longer shelf life without sterilization, fermentation, pasteurization and addition of stabilisers. All the complex technique involves the employment of high temperature and chemical changes which can temper and compromise the quality of the extract.

For example, pasteurization and sterilization may remove bacterial content but the health qualities in the product are significantly decreased as reported by Herbach *et al*., (2007) where pasteurized dargon fruit juice exhibited changes in betacyanin stability and extended heat exposure led to pigment destruction (discolouration) and encouraged formation of degradation products. In a separate study on apple juice processing, Gökmen *et al*., (2001) reported that ultrafiltration and other methods using activated charcoal and polyvinylpolypyrrolidone (PVPP) caused a variation to the natural apple juice colour which is associated to reduction of phenolic compounds and organic acids.

A study carried out by He *et al*., (2007) was successful at excluding pasteurization and enzymatic treatments to effectively filter apple juice without compromising the product quality and simplifying the process. Similarly, the customized filtration method as proposed in this experimental chapter which does not involve the usage of enzymatic or heat treatment is possibly the best resolution and viable method to be considered to extract *Hylocereus polyrhizus* pigment while striking a balance in giving consumers longer shelf life and optimum health benefits from the product.

Food sources including natural food colourant, fruit juices, processed flour and other downstream products are not completely free from natural or synthetic toxins (Nasreddine and Parent-Massin, 2002) even after many levels of processes and treatment. In a review by Magkos *et al*., (2006), environmental contaminants like plant toxins, biological pesticides and pathogenic microorganisms detection in food are persistent. For example, in 2002, the Public Health Laboratory Service (PHLS) and Communicable Disease Surveillance Centre (CDSC) reported a case of *Salmonella* spp. contamination in spinach after processing and packaging.

Hence, there is a need to carry out risk assessment in food products where the probability of the occurrence of an adverse and severe health complication due to hazard in food is minimised (Wal and Pascal, 2000). The earliest step in risk assessment is hazard identification which is to know and recognize the existing threats in a specific product.

According to Morgan (1999), once heavy metals are acquired by food through soil contamination, their presences are rarely affected by the subsequent processing and food preparatory steps. Thus, even though the produce and products are organic/natural, the items still can be host to certain health risk elements. The FAO/WHO joint efforts have consistently urged that states must guarantee that any food products from any origins (whether from conventional planting, organically cultivated, natural source or genetically modified) are free from toxic chemical substances like heavy metals and pesticides residues (Scotter, 2011). This is why toxicology analysis is still relevant and urgent to be carried out in any food item which is put up in the market for consumers as contaminated food products can pose large numbers of health threats including death.

In this study, the toxicology analysis was first carried in determining the microorganism content in the extract. The certificate of analysis from Consolidated Laboratory showed impressive results where the *Hylocereus polyrhizus* extract did not contain any of the pathogens targeted as the test parameters. This is a very important validation that could save cost in future commercialization efforts and reflects the absence of hazardous microorganisms in the extract.

The total bacterial count (TBC) only enumerates microorganisms that grow aerobically at moderate warm temperatures and is used in food safety tests to give a general indication of the number of microorganisms in food. The results from TBC showed that the extract contain 750 cfu/g where the figure is well below the usual levels of < 1000 cfu/ml or < 1000 cfu/g in many other commercial fruit juices in the market like passion fruit, apple, mangosteen, orange and raspberry.

The subsequent microorganism analysis was carried out on the common food borne pathogens: yeast and mold, coliforms, *Escherichia coli* and *Salmonella* sp. Fruit juice have a suitable sugar environment for the growth of mold and yeast which are resistant to extreme conditions like low pH and elevated temperatures (Koc *et al*., 2007). This means that mold and yeast can elude and survive sterilization processes and cause spoilage during storage. For example, a study carried out by Tournas *et al*., (2006) showed that pasteurized fruit juices like apple, carrot and orange sold in supermarkets are contaminated with yeast even before their expiration date. Lateef *et al*., (2004) reported that 40 samples of 20 brands of sachet orange juice products in their study were contaminated with yeasts including *Saccharomyces sp* and *Rhodotorula sp.*

One of the proposed solutions include increasing the usage of synthetic food preservative like sodium benzoate and potassium sorbate which are not preferred by consumers. The alternative would then to remove mold and yeast even before processing to ensure the safety of the product and minimizing the chances of their occurrence.

This is where the filtration technique can be applied and in this study, 1.6 µm glass microfiber filters used would have removed many yeast cells (2.0-8.0µm) and puffball spores (3.0µm) as reflected in the results. In the results, yeast and mold were not detected as the concentration or the cell count was < 100 MPN/g (Most Probable Number).

Microbiological evaluation on three major food bacterial pathogens: coliforms, *Escherichia coli* and *Salmonella* sp also came back negative where there was no detection of any of the mentioned three microorganisms. The absence of the pathogens is a good indication to the safety of the *Hylocereus polyrhizus* extract as all three bacteria causes' food borne diseases which mainly affect the gastrointestinal tract (Tambekar *et al*., 2009) if ingested. Majority of coliforms, *Escherichia coli* and *Salmonella* sp are usually removed in fruit juices through pasteurization and sterilization but the pathogens have amazing adaptability to low pH and some can survive extreme treatments like gamma radiation and even storage at -20°C (Song *et al*., 2007; Yuk and Schneider, 2006; Cheng and Chou, 2001; Larkin *et al*., 1955). The 0.45µm and 0.22µm nylon membrane filters used in this experiment would have possibly removed any small bacteria or microorganisms as explained earlier. The microbiological analyses of this study confirmed the effectiveness of a customized membrane filtration process in obtaining a clarified *Hylocereus polyrhizus* extract while successfully removing any potential common food borne pathogens.

The heavy metal analysis on the *Hylocereus polyrhizus* extract took into account the possible presence of common heavy metals. Heavy metals like lead, mercury, cadmium and arsenic are major lethal contaminants which consistently find their way into the food chain via water channels (Farooq *et al.*, 2010). Detections of various heavy metals are reported by Tufuor *et al*., (2011) in citrus juice and Bingol *et al*., (2010) in soft drinks but at acceptable limits set by FAO/WHO. The concern is then the slow accumulation of these heavy metals in human tissue as more of these contaminants are deposited slowly but surely until it finally causes health complications.

Their toxicity levels are enhanced due to increased industrialization, non-biodegradibility and accumulation along the food chain where their effects can be seen in humans. For example, lead can cause brain damage (Low *et al*., 2000); mercury can create neurological and renal damages (Boening, 2000); cadmium is carcinogenic (Godt *et al*., 2006); and acute arsenic poisoning will result in hematuria (Jain and Ali, 2000).

Heavy metal persistence can be demonstrated in the report by Turkdogan *et al*., (2002) where cadmium present in soil was 59 folds higher than the acceptable limits and it was found to be 50 folds higher in the fruit and vegetable samples tested while Krejpcio *et al*., (2005) reported an incidence where 12% of fresh fruits and juices available in the Polish market exceeded the permissible limits for heavy metal like cadmium. A study by Tormen *et al*., (2011) correlated the presence of heavy metals in fruit juices to the soil where the plant was grown and different processing conditions.

Hence, when it comes to heavy metal analysis, early detection in food sources including fruit juice is vital in ensuring steps to treat the heavy metal present and guarantee that the final product is safe for consumers. Following this concern, the *Hylocereus polyrhizus* extract obtained was subjected to a heavy metal analysis to ensure that the fruit source was heavy metal free and eliminate the possibilities of contamination during processing.

In this study, all commonly targeted heavy metals (lead, mercury, arsenic, cadmium, tin and antimony) as the test parameter were not detected. This is an excellent indicator that the extract obtained from *Hylocereus polyrhizus* source, through the filtration procedures until the toxicology screening was carried, was of good quality and there is no need to treat the extract for heavy metal contamination.

The last and final toxicology assessment on the *Hylocereus polyrhizus* extract is the pesticide screening on organochlorine and organophosphorus insecticides. Pesticide residue determination is important in evaluating food safety as organochlorine insecticides are highly persistent in the environment while organophosphorous insecticides are highly toxic (Zhao *et al*., 2006). Widely used organochlorine compounds like Endosulfan and Heptachlor are very persistent environment pollutant and were reported as high as 90.3 g/kg and 13.0 g/kg respectively in human blood sample (Ansari *et al*., 1997). Pesticides prevalence are reported in various fruit juices like apple, orange, grape, pineapple and pear by Ravelo-Pérez *et al*., (2008) and Xiao *et al*., (2006).

Children who are exposed to organophosphorous pesticides through the environment and ingestion of fresh fruits and fruit juice (Becker *et al*., 2006) can experience problems in neurodevelopment, growth and respiratory illnesses (Eskenazi *et al*., 1999). Since diet is a significant source of pesticide ingestion through contaminated fruits, vegetables, water and other beverages (Jaga and Dharmani, 2003), there is an increased demand for stringent and stricter monitoring on the usage and waste management of these chemicals in agriculture due to the health risk they can cause.

The pesticides screening on 52 organochlorine and 136 organophosphorus insecticides in this study came back negative as none of the tested pesticides were detected. This would suggest that the *Hylocereus polyrhizus* extract is from obtained from a safe source and there is no pesticides prevalence.

As a conclusion, the toxicology analysis results on the *Hylocereus polyrhizus* extract, where it showed acceptable levels of microbiological contaminants, absence of heavy metals and pesticides, indicates that the extract was of good consumption quality and the customized membrane filtration method is a potential method for processing of *Hylocereus polyrhizus* extract.

CHAPTER 7
CLONING AND CHARACTERIZATION OF *MAT*K AND 5-GLUCOSYLTRANSFERASE GENES FROM DRAGON FRUIT (*Hylocereus polyrhizus*)

7.1 INTRODUCTION

The polymerase chain reaction (PCR) is one of the widely used techniques in molecular biology to synthesize targeted DNA fragments in large quantities (Mullis *et al*., 1986). This application utilises a pair of primers to isolate a gene or partial gene for many applications such as identifying genetically inherited diseases, cloning of genes, diagnosis of diseases and plant evolution relationship (Ma, 2005). The basic steps involved in this technique are: denaturation, annealing and extension.

The significance of this technique caused the rise of many modifications to optimize the synthesis of targeted DNA fragments under different circumstances. For example, the nested PCR developed as early as 1992 (Yourno, 1992) to reduce contamination using two sets of primers in two successive PCR runs; the inverse PCR which is used when only one internal sequence is known was developed as early as 1988 (Ochman, 1988); and the quantitative PCR which is employed to quantify small amounts of DNA was used since 1991 (Becker-Andre, 1991). Another improvement made to the PCR technique is the development of real-time PCR which was first demonstrated by Higuchi *et al*., (1992) where detection of DNA sequences were carried out by adding ethidium bromide (EtBr) to PCR reaction for visualization under UV light.

One of the most important modifications of the PCR technique apart from the ones aforementioned is the Reverse Transcriptase PCR (RT-PCR) which was first reportedly utilized by Chelly *et al*., (1988). RT-PCR is a sensitive technique for RNA detection and is used to convert RNA into cDNA. It is most commonly applied in gene mapping and to study the expression of genes.

Following the discovery of DNA by Watson and Crick in 1953, the genetic engineering era of molecular biology begun with the utilization of restriction enzymes and cloning in the 1970s (Westerhoff and Palsson, 2004). The progression then saw the beginning of genome sequencing in 1990s (Hunkapiller, 1991) where automation and technology has led to the current state of massive genomic information and large volume of data. This is where bioinformatics was developed as a model to analyse multiple types of molecular data in a systematic fashion.

According to National Institutes of Health (NIH), bioinformatics can be defined as research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioural or health data, including those to acquire, store, organise, analyse or visualise such data. This system is a combination of computer science, information technology and molecular biology. The data and information generated results in identification and assignment of genes, the functional and structural classification of proteins based on similarity of sequence and/or properties such as motifs or conserved regions (Rehm, 2001) of an unknown gene or protein. Further works carried on the obtained information can lead to construction of phylogenetic trees to study the taxonomic relatedness and degree of similarity.

The *mat*K gene was formerly known as the *orf*K gene and used by many (Hilu and Liang, 1997; Kim and Kim, 2011; Gao *et al*., 2011; Ozdilek *et al*., 2012) for molecular systematic and evolution studies. The importance of this gene was further highlighted when its presence was detected in *Epifagus virginiana* by Wolfe *et al*., (1992) where this parasitic flowering plant has lost more than 65% of its chloroplast gene, rendering it completely nonphotosynthetic but retained the *mat*K gene in its genome. More recently, *mat*K gene has been reported that it could be useful in tagging plants with medicinal values and possibly play a role in the initiation of secondary metabolite production in plants (Saslis-Lagoudakis *et al*., 2011; Cordell and Colvard, 2012).

In the biosynthetic pathway of betacyanins, there are two possible routes which lead to the synthesis of betanin. One route is the spontaneous reaction between the building blocks of betanin which are the cyclo-DOPA and betalamic acid. This spontaneous reaction results in the possible regeneration of betacyanin pigments without enzymatic assistance as discussed in Chapter 3 of this study. The other possible route is the conversion of betanidin to betanin via betanidin-5-O-glucosyltransferase which has been reported by Vogt *et al*., (1999).

The main objective in this chapter was to clone, sequence and characterize the *mat*K and 5-O-glucosyltransferase genes from *Hylocereus polyrhizus*. It is worth noting that no gene sequence information is available currently in the Gene Bank for the two genes isolated from *Hylocereus polyrhizus*.

7.2 MATERIALS AND METHODS

7.2.1 Plant Material

Fruits from the five weeks after anthesis stage were obtained from a local farm situated 20 km from the laboratory. Fruits were halved and peeled manually. Fruit pulp was diced using a knife in 1cm X 1cm cubes and was used for DNA extraction and excess material was frozen under liquid nitrogen and kept at -20˚C for future utilisation.

7.2.2 Extraction of DNA from *Hylocereus polyrhizus* Pulp Tissue

7.2.2.1 Preparation of reagents

The following reagents were prepared for total DNA extraction.

0.5M EDTA (BDH)
A total of 14.61g of Ethylenediaminetetraacetic acid (EDTA) was dissolved in sterile distilled water (SDW) and pH was adjusted to 8.0 using sodium hydroxide pellets (Merck) and volume of solution was made up to 100ml. Solution was autoclaved and stored at room temperature (24 ± 2 ˚C).

1M Tris-HCl
A total of 60.6g of Tris-base (Sigma) was dissolved in SDW and pH was adjusted to 8.0 using sodium hydroxide pellets (Merck) and volume of solution was made up to 500ml. Solution was autoclaved and stored at room temperature.

5M sodium chloride (NaCl) (Sigma)
A total of 14.61g of sodium chloride (NaCl) was dissolved in 40ml of SDW and the volume was made up to 50ml. Solution was autoclaved and stored at room temperature.

10% sodium lauroyl sarcosinate (SLS) (BioWhittaker)
A total of 5g of SLS was dissolved in 50ml of SDW and the volume was made up to 50ml.

1X TAE running buffer (Promega)
A volume of 25ml ready made 40X TAE buffer (Promega) was added to 975ml of SDW to make 1liter of 1X TAE running buffer

DNA isolation buffer

Reagent I (DNA extraction buffer)
A total of 6.38g of sorbitol (Sigma) was dissolved in 50ml of sterile distilled water (SDW) and 20ml of 1M Tris-HCl was added. A volume of 2ml of 0.5M EDTA (BDH) was added to the solution and the pH was adjusted to 7.5 using conc. HCl (Systerm). Volume of solution was made up to 200 ml with SDW. Solution was autoclaved and stored at room temperature.

Reagent II (Nuclei lysis buffer)
A total of 2g of CTAB (Sigma) was dissolved in 20ml of SDW and 40ml of 1M Tris-HCl, 20ml of 0.5M EDTA (BDH) and 80ml of 5M NaCl (Sigma) was added to the solution. Volume of solution was made up to 200 ml with SDW.

Reagent III [5% sodium lauroyl sarcosinate (SLS)]
A volume 20ml of 10% SLS stock (BioWhittaker) was added to 20ml of SDW.

The DNA isolation buffer was obtained by adding 10ml of *Reagent I*, 10ml of *Reagent II*, 4ml of *Reagent III*, 0.48g of polyvinyl pyrrolidone (PVP) (Research Organics) and 0.12g of sodium bisulfite (BDH) to make a 25ml of solution. All the components were dissolved thoroughly and stored at room temperature.

Chloroform: Isoamyl (24:1)
A volume of 48ml of chloroform (Systerm) was added to 2ml of isoamyl (Unilab).

70% ethanol
A volume of 70ml of ethanol (Fluka) was added to 30ml of SDW and kept in -20°C until use.

7.2.2.2 DNA extraction

A total of 0.1g of frozen *Hylocereus polyrhizus* pulp tissue were ground in a chilled mortar and pestle under liquid nitrogen. The fine frozen powder was quickly transferred to a 1.5ml microcentrifuge tube containing 700µl of DNA isolation buffer, mixed thoroughly and incubated at 65°C for 30 minutes using a waterbath (Memmert). Upon removal from the waterbath, 600µl of chloroform: isoamyl was added and the mixture was shaken vigorously for 2 minutes and centrifuged at 14 000 rpm for 5 minutes using a tabletop microcentrifuge (UEC Micro 4/B).

The resulting supernatant was transferred into a new microcentrifuge tube and 600µl of chloroform: isoamyl was added. The mixture was shaken vigorously for 2 minutes and centrifuged at 14 000 rpm for 5 minutes. The resulting supernatant was transferred into a new microcentrifuge tube and an equal volume of cold absolute isopropanol (Sigma) was added into the tube. The mixture was inverted gently for 10 seconds and subsequently incubated in ice (4 ± 2°C) for 10 minutes. After incubation, the tube was centrifuged at 14 000 rpm for 5 minutes. The supernatant was poured off carefully and 700µl of cold 70% ethanol was added to the microcentrifuge tube. The tube was inverted gently for 10 seconds and centrifuged at 14 000 rpm for 2 minutes. The supernatant was decanted and the tube was inverted to air dry the resulting pellet. The DNA pellet was resuspended in 50µl of SDW and stored in -20°C until use.

7.2.3 Preparation of 1% Agarose Gel

The 1% agarose gel was prepared by adding 0.5g of agarose powder (Promega) to 49ml of SDW and 1ml of 40X TAE ready made buffer (Promega). The solution was boiled until clear and then cooled to 50°C. The gel solution was added with 1.5µl of ethidium bromide (Sigma) before pouring into a gel casting tray to solidify. A volume of 4µl of DNA was added with 2µl of loading dye, loaded into the gel and was electrophorised for 45 minutes in 500ml of 1X TAE running buffer.

7.2.4 Quantification of DNA Template

Quantification of extracted DNA was carried out using a 1000X dilution by adding 2997µl of SDW to a quartz cuvette followed by 3 µl of sample. The solution was thoroughly mixed by pipetting and SDW was used as a blank. The absorbance was recorded at 260nm and 280nm. DNA concentration was calculated using the following equation:

DNA concentration (µg/µl) = OD_{260} X Dilution factor X (50µg DNA/ml)

Purity was calculated by taking the ratio of reading of OD_{260}/OD_{280}

7.2.5 PCR of *Hylocereus polyrhizus* DNA

7.2.5.1 matK gene primers

The *mat*K gene primers were designed based on several earlier works carried out by Johnson and Soltis (1994); Plunkett *et al*., (1997); Kusumi *et al*., (2000); and Wojciechowski *et al*., (2004) and commercially made by 1st Base Pte Ltd (Singapore). The sequences of the two primers are as follows:

Forward Primer (*mat*KF1)

5' CAA GGA ATT CGT CTA CAT 3'

Reverse Primer (*mat*KR1)

5' TTG AAT ACT CAA TTG ATT 3'

7.2.5.2 5-GT gene primers

The 5-GT gene primers were designed based on several earlier works carried out by Heuer *et al*., (1996); Vogt *et al*., (1999); and Masada *et al*., (2007) and commercially made by 1st Base Pte Ltd (Singapore). The sequences of the two primers are as follows:

Forward Primer (GT5F1)

5' ACC GGA ACC GGT GTC 3'

Reverse Primer (GT5R1)

5' GTC AAG AGC GGA CCT 3'

7.2.5.3 Template preparation

The frozen DNA template was thawed on ice and a volume of 2µl (about 1µg of DNA, calculated from Chapter 7.2.3) was used for the amplification. The polymerase chain reaction (PCR) was carried out in a volume of 50µl in a 0.2 ml Eppendorf tube using the GoTaq® qPCR Master Mix:

Component	Volume
2X Master Mix	25µl
Forward primer (10µM)	4µl
Reverse primer (10µM)	4µl
DNA template	2µl
Nuclease free water	15µl

The reaction mixture above was thoroughly mixed gently using a pipette. PCR was started on with amplification cycles as described below:

Step	**Temperature**	**Duration**	**Cycle(s)**
Denaturation	95°C	2 minutes	1
Annealing	95°C	30 seconds	40
	55°C	45 seconds	
	72°C	90 seconds	
Extension	72°C	5 minutes	1
Soak cycle (hold)	4°C	Overnight	

7.2.6 Purification of PCR Product

PCR product was purified using the Wizard^{R} SV gel and PCR purification Clean-up system from Promega according to the manufacturers' instruction (Appendix 9)

7.2.7 Sequencing of PCR Product

DNA sequencing of the PCR generated products was carried out commercially through 1st Base Pte Ltd (Singapore).

7.2.8 Bioinformatics

The data analysis in this chapter was performed using:

i. Basic Local Alignment Search Tool (BLAST) from http://blast.ncbi.nlm.nih.gov/
BLAST finds regions of local similarity between sequences and compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

ii. ExPASy Translate from http://web.expasy.org/translate/
Translat**e** is a tool which allows the translation of a nucleotide (DNA/RNA) sequence to a protein sequence.

iii. ORF Finder (Open Reading Frame Finder) from http://www.ncbi.nlm.nih.gov/gorf/gorf.html
The ORF Finder is a graphical analysis tool which finds all open reading frames of a selectable minimum size in a sequence.

iv. ClustalW2 - Multiple Sequence Alignment fromhttp://www.ebi.ac.uk/Tools/msa/clustalw2/
ClustalW2 is a general purpose multiple sequence alignment program for DNA or proteins and generating other additional information like Distance Matrix and phylogenetic trees.

v. PROSITE scan from http://prosite.expasy.org/

PROSITE consists of documentation entries describing protein domains, families and functional sites as well as associated patterns, motifs and profiles to identify them.

vi. ScanProsite tool from http://prosite.expasy.org/scanprosite/

ScanProsite is an advanced scan to describe the motifs within a protein sequence.

vii. Conserved Domain Database (CDD) from http://www.ncbi.nlm.nih.gov/cdd

CDD is a protein annotation resource that consists of a collection of well-annotated multiple sequence alignment models for ancient domains and full-length proteins.

viii. Protein Homology/AnalogY Recognition Engine (PHYRE) from http://www.sbg.bio.ic.ac.uk/phyre/

PHYRE is used for protein structure prediction from a submitted protein sequence.

Other sequence analysis and proteomic tools were accessed through EXPASy homepage at http://www.expasy.org/tools/ and www.molbiol.net.

7.3 RESULTS

DNA quantification showed that there was an amount of 550µg/ml of DNA in the extracted DNA sample and the purity of DNA was recorded at 1.79.

7.3.1 PCR Results with *mat*K Primers

PCR with the *mat*K primers yielded a single product with a size 2500bp and clear products were visible in the 1.0% agarose gel.

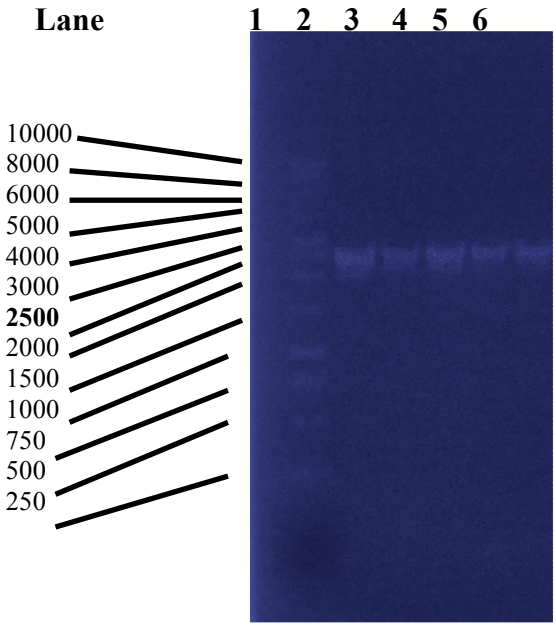


Figure 7.1: PCR results with *mat*K primers. In Lane 1 is the DNA 1kb GeneRuler Ladder (Fermentas). Lane 2 is the PCR product of about 2500bp generated using primers *mat*KF1 and *mat*KR1 for total DNA from *Hylocereus polyrhizus* analysed through 1.0% agarose gel. Lanes 3-6 were products obtained from different primer concentrations for optimization purposes.

7.3.1.1 Nucleotide sequence of the matK protein

The DNA sequence for *Hylocereus polyrhizus mat*K protein is shown in Figure 7.2.

```
1     CCCCCCTCCGGCCTTTTTAAGCAGGCAAGGATCTTTTACATATTTGGATG
51    AAGCAAGGAATTCGTCTACATCATCGGTAGAGTTTATCAGACCACGACTG
101   ATCCTGAAAGGAAATGAATGGAAAAAAGAGCATGTCGTATCAATAGAGAA
151   TTCGGAGAATATTTCATTCGTACCAAATCGGTACCAAACAAACATTATTT
201   GAATTGAACGAAATGAATTCGCAAGTTTGGTCGATTGAATAAATGGATCG
251   AGCCCTATGGTTCAAATTCTAGGGAAAGAAAGAGCAACGAGCTTATCTTC
301   GTAATTTGAATGATTACCCGATCTAATTTAACGTTAAAAAAAAATTAGTG
351   CCAGATACGGGAAAGGCTTCTCACACGAGTGAATTTTTTCGTTTTTAGTG
401   AATCCTAACTATTAGATTATCCATTTTCTATATGGAGATGAATGTGTAGA
451   AGAAACAGTATATTGATAAAGATACTTTTCCAAAATCAAAAGAGCGATTG
501   GGTTGAAAAAATAAAGGATTTCTAACCATCTTGCTTTGTTATCCTATAAA
551   AAAAAACCAATTAGATGGAATAAAATAGAGAATCCGTTGATGAATTTACC
601   TGTCTCCGAGGTACTTAATTATTTCAAAATAGAATACCTTGTTTTGACTG
651   TATCGCACTATGTATCATTTGATAATCCAAGAAACTCCCTGCTTTTTTTT
701   AGTTTTCGGTCTAATTTGAAATGGAAGAATTCCAAAGATATATAGAACTA
751   GATAGGTCTTGGCAACACAACTTTTTCTATCCACTTATCTTTCAGGAATA
801   TATTTATGGATTTGCATATGATCGTGGTTTAAATAAATCCATTTTGTTGG
851   AAAATGCAGGCGACAAGAAATCCAGTTTACTGATTGTAAAACGTTCAATT
901   AACCGAATGTATCAACAGACTCATTTGATTCTTTCTGCTAATCATTCTAA
951   CCAAAATGACTTTTTTGGGCACGAGCTCCAAGAAGAATTTGTATTATCAA
1001  AAAATACTAAAAAGGATTTGCAGTCATTTATGAAAATTCCAATTTTCTTT
1051  ATTTTTTTGGGAAAAATTGAAATTTTAGTAAAAGTCAAAAAATTTTTCTA
1101  AGGATTTTTTGGGTTATCTTAAGGCTTTTCGAAGACCCTTTCCTGCATTT
1151  TGTTTGGTTTCGAGGAAAATCCATGTCGGCTTCAAAGGGGACATCTCTTC
1201  TGTTGCATAAATGGAAATATTATCTTTTCAAATTTTGGCAATGTCATTTT
1251  TCCCTGTGGTCTCCACCAAGAAGAATCTATATCAATCGGTTATCAAAGCA
1301  TTCTCTCGACTTTATGGGTTTTTTTTCAAGTGTTCGACTCAATTCTTCAG
1351  TGGTACGGAGTCAAATGGTAGAAAATTCATTTCTAATAGATAATCCTATT
1401  AAGAAATTCGATACCATAGTTCGAATTATTCCTCTGGTTGGATCGTTGGC
1451  TAAAGCGCAATTTTGTAACGTATTAGGACATCCCATTAGTAAGTCGGTCT
1501  GGACCGATTTATTGGATTCTGATATTATTGATCGGTTTGGGCGCATATGC
1551  AGAAATCTTTCTCATTATTATAGTGGCTCTTCAAGAAAAAAGAGTTTGTA
1601  TCGAATAAAGTATATACTTCGACTTTCTTGTGCTAGAACTTTGGCTCGTA
1651  AACACAAAAGTACTGTACGCGCTTTTTTGAAAAGATTAGGTTCAGAATTT
1701  TTGGAAGAATTCTTTACAGAGGAAGAAAAAGTTCTTTCTTTGATCTTTCC
1751  AAGAGATTCTTCTATCTCACGAGGATTATATAGGGGTCCTTTTTGGTACT
1801  TGGATATTATTTGTATCCATGATTTGGCCAATGATGAATGATTGGTTATG
1851  AGACTGTCTAAATGAAATGGAAATTATACCTAAATGAATGACGGAATAAG
1901  ATTAAAAAAAAAAATGCATTCATTTCTATACTGAAATGTATATCTTTATGC
1951  TTATGCAGTAAGGGTTGAATCAATTGAGTATTCAACTTTGTTAGAGTCTC
2001  TTCTAGGAAAGGAACTTAATTTTAGATGTATACATAGGGAAAGCCGTGTG
2051  CAATGAAAAATGCAAGCACGGCTTGGGGAGGGGTCTTCTTTTTTTACCAA
2101  AAAAATG
```

Figure 7.2: Nucleotide sequence of purified *mat*K PCR product from *Hylocereus polyrhizus*

The complete nucleotide sequence of the *mat*K gene and the amino acid sequence of its predicted product is shown in Figure 7.3.

```
CCCCCCTCCGGCCTTTTTAAGCAGGCAAGGATCTTTTACATATTTGGATGAAGCAAGGAATT
   P  L  R  P  F  -  A  G  K  D  L  L  H  I  W  M  K  Q  G  I
CGTCTACATCATCGGTAGAGTTTATCAGACCACGACTGATCCTGAAAGGAAATGAATGGA
 R  L  H  H  R  -  S  L  S  D  H  D  -  S  -  K  E  M  N  G
AAAAAGAGCATGTCGTATCAATAGAGAATTCGGAGAATATTTCATTCGTACCAAATCGGT
 K  K  S  M  S  Y  Q  -  R  I  R  R  I  F  H  S  Y  Q  I  G
ACCAAACAAACATTATTTGAATTGAACGAAATGAATTCGCAAGTTTGGTCGATTGAATAA
 T  K  Q  T  L  F  E  L  N  E  M  N  S  Q  V  W  S  I  E  -
ATGGATCGAGCCCTATGGTTCAAATTCTAGGGAAAGAAAGAGCAACGAGCTTATCTTCGT
 M  D  R  A  L  W  F  K  F  -  G  K  K  E  Q  R  A  Y  L  R
AATTTGAATGATTACCCGATCTAATTTAACGTTAAAAAAAAATTAGTGCCAGATACGGGA
 N  L  N  D  Y  P  I  -  F  N  V  K  K  K  L  V  P  D  T  G
AAGGCTTCTCACACGAGTGAATTTTTTCGTTTTTAGTGAATCCTAACTATTAGATTATCC
 K  A  S  H  T  S  E  F  F  R  F  -  -  I  L  T  I  R  L  S
ATTTTCTATATGGAGATGAATGTGTAGAAGAAACAGTATATTGATAAAGATACTTTTCCA
 I  F  Y  M  E  M  N  V  -  K  K  Q  Y  I  D  K  D  T  F  P
AAATCAAAAGAGCGATTGGGTTGAAAAAATAAAGGATTTCTAACCATCTTGCTTTGTTAT
 K  S  K  E  R  L  G  -  K  N  K  G  F  L  T  I  L  L  C  Y
CCTATAAAAAAAAACCAATTAGATGGAATAAAATAGAGAATCCGTTGATGAATTTACCTG
 P  I  K  K  N  Q  L  D  G  I  K  -  R  I  R  -  -  I  Y  L
TCTCCGAGGTACTTAATTATTTCAAAATAGAATACCTTGTTTTGACTGTATCGCACTATG
 S  P  R  Y  L  I  I  S  K  -  N  T  L  F  -  L  Y  R  T  M
TATCATTTGATAATCCAAGAAACTCCCTGCTTTTTTTTAGTTTTCGGTCTAATTTGAAAT
 Y  H  L  I  I  Q  E  T  P  C  F  F  L  V  F  G  L  I  -  N
GGAAGAATTCCAAAGATATATAGAACTAGATAGGTCTTGGCAACACAACTTTTTCTATCC
 G  R  I  P  K  I  Y  R  T  R  -  V  L  A  T  Q  L  F  L  S
ACTTATCTTTCAGGAATATATTTATGGATTTGCATATGATCGTGGTTTAAATAAATCCAT
 T  Y  L  S  G  I  Y  L  W  I  C  I  -  S  W  F  K  -  I  H
TTTGTTGGAAAATGCAGGCGACAAGAAATCCAGTTTACTGATTGTAAAACGTTCAATTAA
 F  V  G  K  C  R  R  Q  E  I  Q  F  T  D  C  K  T  F  N  -
CCGAATGTATCAACAGACTCATTTGATTCTTTCTGCTAATCATTCTAACCAAAATGACTT
 P  N  V  S  T  D  S  F  D  S  F  C  -  S  F  -  P  K  -  L
TTTTGGGCACGAGCTCCAAGAAGAATTTGTATTATCAAAAAATACTAAAAAGGATTTGCA
 F  W  A  R  A  P  R  R  I  C  I  I  K  K  Y  -  K  G  F  A
GTCATTTATGAAAATTCCAATTTTCTTTATTTTTTTGGGAAAAATTGAAATTTTAGTAAA
 V  I  Y  E  N  S  N  F  L  Y  F  F  G  K  N  -  N  F  S  K
AGTCAAAAAATTTTTCTAAGGATTTTTTGGGTTATCTTAAGGCTTTTCGAAGACCCTTTC
 S  Q  K  I  F  L  R  I  F  W  V  I  L  R  L  F  E  D  P  F
CTGCATTTTGTTTGGTTTCGAGGAAAATCCATGTCGGCTTCAAAGGGGACATCTCTTCTG
 L  H  F  V  W  F  R  G  K  S  M  S  A  S  K  G  T  S  L  L
TTGCATAAATGGAAATATTATCTTTTCAAATTTTGGCAATGTCATTTTTCCCTGTGGTCT
 L  H  K  W  K  Y  Y  L  F  K  F  W  Q  C  H  F  S  L  W  S
CCACCAAGAAGAATCTATATCAATCGGTTATCAAAGCATTCTCTCGACTTTATGGGTTTT
 P  P  R  R  I  Y  I  N  R  L  S  K  H  S  L  D  F  M  G  F
TTTTCAAGTGTTCGACTCAATTCTTCAGTGGTACGGAGTCAAATGGTAGAAAATTCATTT
 F  S  S  V  R  L  N  S  S  V  V  R  S  Q  M  V  E  N  S  F
CTAATAGATAATCCTATTAAGAAATTCGATACCATAGTTCGAATTATTCCTCTGGTTGGA
 L  I  D  N  P  I  K  K  F  D  T  I  V  R  I  I  P  L  V  G
TCGTTGGCTAAAGCGCAATTTTGTAACGTATTAGGACATCCCATTAGTAAGTCGGTCTGG
 S  L  A  K  A  Q  F  C  N  V  L  G  H  P  I  S  K  S  V  W
```

```
ACCGATTTATTGGATTCTGATATTATTGATCGGTTTGGGCGCATATGCAGAAATCTTTCT
 T  D  L  L  D  S  D  I  I  D  R  F  G  R  I  C  R  N  L  S
CATTATTATAGTGGCTCTTCAAGAAAAAAGAGTTTGTATCGAATAAAGTATATACTTCGA
 H  Y  Y  S  G  S  S  R  K  K  S  L  Y  R  I  K  Y  I  L  R
CTTTCTTGTGCTAGAACTTTGGCTCGTAAACACAAAAGTACTGTACGCGCTTTTTTGAAA
 L  S  C  A  R  T  L  A  R  K  H  K  S  T  V  R  A  F  L  K
AGATTAGGTTCAGAATTTTTGGAAGAATTCTTTACAGAGGAAGAAAAAGTTCTTTCTTTG
 R  L  G  S  E  F  L  E  E  F  F  T  E  E  E  K  V  L  S  L
ATCTTTCCAAGAGATTCTTCTATCTCACGAGGATTATATAGGGGTCCTTTTTGGTACTTG
 I  F  P  R  D  S  S  I  S  R  G  L  Y  R  G  P  F  W  Y  L
GATATTATTTGTATCCATGATTTGGCCAATGATGAATGATTGGTTATGAGACTGTCTAAA
 D  I  I  C  I  H  D  L  A  N  D  E  -  L  V  M  R  L  S  K
TGAAATGGAAATTATACCTAAATGAATGACGGAATAAGATTAAAAAAAAAAATGCATTCAT
 -  N  G  N  Y  T  -  M  N  D  G  I  R  L  K  K  K  C  I  H
TTCTATACTGAAATGTATATCTTTATGCTTATGCAGTAAGGGTTGAATCAATTGAGTATT
 F  Y  T  E  M  Y  I  F  M  L  M  Q  -  G  L  N  Q  L  S  I
CAACTTTGTTAGAGTCTCTTCTAGGAAAGGAACTTAATTTTAGATGTATACATAGGGAAA
 Q  L  C  -  S  L  F  -  E  R  N  L  I  L  D  V  Y  I  G  K
GCCGTGTGCAATGAAAAATGCAAGCACGGCTTGGGGAGGGGTCTTCTTTTTTTACCAAAA
 A  V  C  N  E  K  C  K  H  G  L  G  R  G  L  L  F  L  P  K
AAATG
 K
```

Figure 7.3: The complete nucleotide sequence of *mat*K gene and the amino acid sequence of its predicted product.

7.3.1.2 BLAST analysis on nucleotide sequence of the matK protein

BLAST homology search at www.ncbi.nlm.nih.gov/blast revealed the following match in the database.

1. 2028/2069 positive identities (99%) with *Selenicereus boeckmannii* (a)
2. 2028/2069 positive identities (99%) with *Hylocereus peruvians* (b)
3. 2028/2069 positive identities (98%) with *Selenicereus spinulosus* (c)
4. 2020/2068 positive identities (98%) with *Selenicereus vagans* (d)
5. 2021/2069 positive identities (98%) with *Neoraimondia arequipensis* (e)
6. 2020/2070 positive identities (98%) with *Disocactus amazonicus* (f)
7. 1731/1752 positive identities (98%) with *Hylocereus ocamponis* (g)
8. 1999/2054 positive identities (98%) with *Pfeiffera monacantha* (h)
9. 2021/2073 positive identities (98%) with *Acanthocereus pentagonus* (i)
10. 2012/2069 positive identities (98%) with *Peniocereus oaxacensis* (j)

(a) Accession AY015311 - Nyffeler (2002)
(b) Accession AY015310 - Nyffeler (2002)
(c) Accession FN997254 – Barcenas *et al*., (2010)
(d) Accession FN997113 - Barcenas *et al*., (2010)
(e) Accession AY015299 - Nyffeler (2002)
(f) Accession AY015312 - Nyffeler (2002)
(g) Accession FN997352 - Barcenas *et al*., (2010)
(h) Accession FR716766 - Korotkova *et al*., (2010)
(i) Accession AY015295 - Nyffeler (2002)
(j) Accession FN997122 - Barcenas *et al*., (2010)

7.3.1.3 ClustalW (2.1) multiple sequence alignment for matK gene

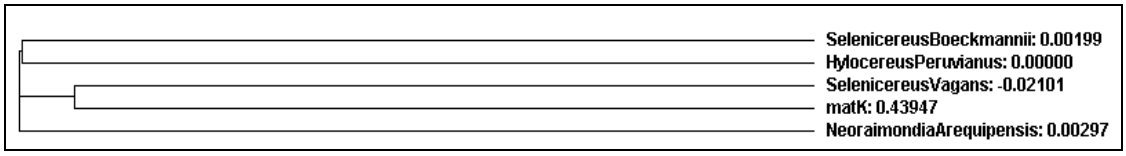
Figure 7.4: Comparison of the deduced amino acid sequence of *mat*K gene from *Hylocereus polyrhizus* and sequences encoding the maturase K coding region from *Selenicereus boeckmannii* (AY015311), *Hylocereus peruvians* (AY015310), *Selenicereus vagans* (FN997113), and *Neoraimondia arequipensis* (AY015299). Amino acids were aligned using CLUSTALW from http://www.ebi.ac.uk

```
SelenicereusBoeckmannii      ------------------MEEFQRYIELDRSWQHNFFYPLIFQEYIYGFA 32
HylocereusPeruvianus         ------------------MEEFQRYIELDRSWQHNFFYPLIFQEYIYGFA 32
NeoraimondiaArequipensis     ------------------MEEFQRYIELDRSWQHNFFYPLIFQEYIYGFA 32
SelenicereusVagans           ------------------MEEFQRYIELDRSWQHNFFYPLIFQEYIYGFA 32
matK                         PLRPFAGKDLLHIWMKQGIRLHHRSLSDHDSKEMNGKKSMSYQRIRRIFH 50
                                               :. .:* :. . * : *   .: :*.    *

SelenicereusBoeckmannii      -YDRGLNKSILLENAGDKKYSLLIVKRLINRMYQQTHLILSANHSNQNDF 81
HylocereusPeruvianus         -YDRGLNKSILLENAGDKKYSLLIVKRLINRMYQQTHLILSANHSNQNDF 81
NeoraimondiaArequipensis     -YDRGLNKSILLENAGDKKYSLLIVKRLINRMYQQTHLILSANHSNQNDF 81
SelenicereusVagans           -YDRGLNKSILLENAGDKKYSLLIVKRLINRMYQQTHLILSANHSNQKDF 81
matK                         SYQIGTKQTLFELNEMNSQVWSIEMDRALWFKFGKKEQRAYLRNLNDYPI 100
                              *: * ::::: *  :.:   : :.* :   : :..     .: *:  :

SelenicereusBoeckmannii      FGHKHKKN---LYYQIISEGFAVIMEIPFSLLLISS-----------LEA 117
HylocereusPeruvianus         FGHKHKKN---LYYQIISEGFAVIMEIPFSLLLISS-----------LEA 117
NeoraimondiaArequipensis     FGHKHKKN---LYYQIISEGFAVIMEIPFSLLLISS-----------LEA 117
SelenicereusVagans           FGHKHKKN---LYYQIISEGFAVIMEIPFSLLLICS-----------LEA 117
matK                         FNVKKKLVPDTGKASHTSEFFRFILTIRLSIFYMEMNVKKQYIDKDTFPK 150
                             *. *:*        .  ** * .*: * :*:: :            :

SelenicereusBoeckmannii      KEKKIVKSHNLRSIHSIFPFFEDKFLHLNYVLEILIP------------Y 155
HylocereusPeruvianus         KEKKIVKSHNLRSIHSIFPFFEDKFLHLNYVLEILIP------------Y 155
NeoraimondiaArequipensis     KEKKIVKSHNLRSIHSIFPFFEDKFLHLNYVLEILIP------------Y 155
SelenicereusVagans           KEKKIVKSHNLRSIHSIFPFFEDKFLHLNYVLEILIP------------Y 155
matK                         SKERLGKNKGFLTILLCYPIKKNQLDGIKRIRIYLSPRYLIISKNTLFLY 200
                             .:::: *.:.: :*   :*: ::::  :: :   * *            *

SelenicereusBoeckmannii      PIHLEILVQTL----------------RYWVKDASSLHLLRFFLYEYRNW 189
HylocereusPeruvianus         PIHLEILVQTL----------------RYWVKDASSLHLLRFFLYEYRNW 189
NeoraimondiaArequipensis     PIHLEILVQTL----------------RYWVKDASSLHLLRFFLYEYRNW 189
SelenicereusVagans           PIHLEILKE------------------------------------------ 164
matK                         RTMYHLIIQETPCFFLVFGLINGRIPKIYRTRVLATQLFLSTYLSGIYLW 250
                                 .:: :

SelenicereusBoeckmannii      NSLITPQKSISIFSKRNQRLFLFLYNFHVCEYESIFVFLCNQSSHLRSTS 239
HylocereusPeruvianus         NSLITPQKSISIFSKRNQRLFLFLYNFHVCEYESIFVFLCNQSSHLRSTS 239
NeoraimondiaArequipensis     NSLITPQKYISIFSKRNQRLFLFLYNFHVCEYESIFVFLCNQSSHLRSTS 239
SelenicereusVagans           ----------------SRLFLFLYNFHVCEYESIFVFLCNQSSHLRSTS 197
matK                         ICISWFKIHFVGKCRRQEIQFTDCKTFNPNVSTDSFDSFCSFPKLFWARA 300
                                           .  *    .*:     . *  :*. .. : : :

SelenicereusBoeckmannii      ------------FGALLERN---YFSGKLEYLVKVKTFTKDFCVILWLFK 274
HylocereusPeruvianus         ------------FGALLERN---YFSGKLEYLVKVKTFTKDFCVILWLFK 274
NeoraimondiaArequipensis     ------------FGALLERN---YFSGKLEYLVKVKTFTKDFCVILWLFK 274
SelenicereusVagans           ------------FGALLERN---YFYGKLEYLVKVKTFTKDFCVILWLFK 232
matK                         PRRICIIKKYKGFAVIYENSNFLYFFGKNNFSKSQKIFLRIFWVILRLFE 350
                                       *.: *..   ** ** ::  . * * : * *** **:

SelenicereusBoeckmannii      DPFLHYVRYRGKSILASKGTSLLMHKWKYYLFNFWQCHFSLWSPPRRIYI 324
HylocereusPeruvianus         DPFLHYVRYRGKSILASKGTSLLMHKWKYYLFNFWQCHFSLWSPPRRIYI 324
NeoraimondiaArequipensis     DPFLHYVRYRGKSILASKGTSLLMHKWKYYLFNFWQCHFSLWSPPRRIYI 324
SelenicereusVagans           DPFLHYVRYRGKSILASKGTSLLMHKWKYYLFNFWQCHFSLWSPPRRIYI 282
matK                         DPFLHFVWFRGKSMSASKGTSLLLHKWKYYLFKFWQCHFSLWSPPRRIYI 400
                             *****:* :****: ********:********:*****************
```

Distance Matrix

		1	2	3	4	5
1	*Selenicereus boeckmannii*	0.000	0.000	0.00199	0.03303	0.43947
2	*Hylocereus peruvianus*	0.000	0.000	0.00096	0.03303	0.03303
3	*Selenicereus vagans*	0.00199	0.00096	0.000	0.00096	-0.0210
4	*Neoraimondia arequipensis*	0.03303	0.03303	0.00096	0.000	0.00297
5	*mat*K	0.43947	0.03303	-0.0210	0.00297	0.000

Figure 7.5: Phylogenetic tree generated from the CLUSTALW multiple alignment programme to show the relationship between *mat*K gene and the maturase K gene from *Selenicereus boeckmannii* (AY015311), *Hylocereus peruvians* (AY015310), *Selenicereus vagans* (FN997113), and *Neoraimondia arequipensis* (AY015299).



A phylogenetic analysis of the relationship of *mat*K gene with other maturase K genes is presented in Figure 7.5. The phylogenetic tree derived from the aligned amino acid sequence showed that the *mat*K gene is closely related to *Selenicereus boeckmannii* and *Hylocereus peruvians* with closest relationship at 99% identity. The amino acid sequence of the *mat*K gene from *Hylocereus polyrhizus* determines the protein's 3D structure and is presented in Figure 7.9. This involved generating a set of structures representative of most of the possible folds for specific protein domains and then solving the structures for new proteins based on known fold-structure relationships. However, most of the domain families have to be characterized further before analysis of their function and evolution on a higher level can be carried out.

7.3.1.4 Sequence analysis of matK gene

The complete nucleotide sequence of the *mat*K gene from *Hylocereus polyrhizus* and the deduced amino acid sequence of the translation product are shown in Figure 7.6. The nucleotide is 2107 bp long and includes the signature concensus pattern *mat*K coding region **(**Consensus pattern L - Y - R - I - K - Y - I - L - R - L - S - C - A - R - T - L - A - R - K - H - K - S - T - V - R - A - F - L - K - R - L - G - S - E - F - L - E - E - F - F - T - E - E - E - K - V - L - S - L - I). This consensus pattern was obtained using the ScanProsite tool (http://prosite.expasy.org/scanprosite/) and is consistent with the *mat*K coding region in eight other deposited sequences (Table 7.1).

Using the ORF finder graphical analysis tool from http://www.ncbi.nlm.nih.gov/gorf, the ATG at position 1173 was predicted to be the translation initiation codon and the TGA at position 1841 was predicted to be the in-frame coding region termination codon. The predicted coding region for this nucleotide sequence spans 222 amino acids (Figure 7.7) from which a translation product with a predicted average molecular mass of 26048.35 Daltons and a theoretical pI of 9.86 was obtained. The amino acid analysis showed that the sequence contains two conserved domain where the super-family domains are *mat*K/*trn*K amino terminal region and Type II intron maturase as shown in Figure 7.8. The function of the *mat*K/*trn*K amino terminal region is unknown up until now while the Type II intron maturase region are involved in the activities of intron-encoded reverse transcriptase, maturase and DNA endonuclease. These two super-family domains make up the entire *mat*K domain as illustrated in Figure 7.8 and the sequences are as presented in Table 7.2.

Table 7.1: Consensus pattern of *mat*K isolated from *Hylocereus polyrhizus* consistent with eight other deposited sequences

Organism	**Sequence identifier**	**Pattern**
Austrocylindropuntia vestita (Cactus) *(Opuntia vestita)*	Q95EE0	LYRIKYILRLSCARTLARKHKSTVRAFL KRLGSEFLEEFFTEEEKVLSLI
Browningia hertlingiana	Q95EA5	LYRIKYILRLSCARTLARKHKSTVRAFL KRLGSEFLEEFFTEEEKVLSLI
Grahamia bracteata	Q95EE5	LYRIKYILRLSCARTLARKHKSTVRAFL KRLGSEFLEEFFTEEEKVLSLI
Maihuenia poeppigii (Hardy cactus)	Q95ED6	LYRIKYILRLSCARTLARKHKSTVRAFL KRLGSEFLEEFFTEEEKVLSLI
Opuntia quimilo (Cactus)	Q95ED9	LYRIKYILRLSCARTLARKHKSTVRAFL KRLGSEFLEEFFTEEEKVLSLI
Pereskia aculeata (Barbados gooseberry)	Q3MKB3	LYRIKYILRLSCARTLARKHKSTVRAFL KRLGSEFLEEFFTEEEKVLSLI
Phaulothamnus spinescens (Snake-eyes) (Devilqueen)	Q5J2W2	LYRIKYILRLSCARTLARKHKSTVRAFL KRLGSEFLEEFFTEEEKVLSLI
Schlumbergera truncata (Thanksgiving cactus)	Q95E84	LYRIKYILRLSCARTLARKHKSTVRAFL KRLGSEFLEEFFTEEEKVLSLI

```
1      CCCCCCTCCGGCCTTTTTAAGCAGGCAAGGATCTTTTACATATTTGGATGAAGCAAGGAATT
          P  L  R  P  F  -  A  G  K  D  L  L  H  I  W  M  K  Q  G  I
63     CGTCTACATCATCGGTAGAGTTTATCAGACCACGACTGATCCTGAAAGGAAATGAATGGA
        R  L  H  H  R  -  S  L  S  D  H  D  -  S  -  K  E  M  N  G
123    AAAAAGAGCATGTCGTATCAATAGAGAATTCGGAGAATATTTCATTCGTACCAAATCGGT
        K  K  S  M  S  Y  Q  -  R  I  R  R  I  F  H  S  Y  Q  I  G
184    ACCAAACAAACATTATTTGAATTGAACGAAATGAATTCGCAAGTTTGGTCGATTGAATAA
        T  K  Q  T  L  F  E  L  N  E  M  N  S  Q  V  W  S  I  E  -
244    ATGGATCGAGCCCTATGGTTCAAATTCTAGGGAAAGAAAGAGCAACGAGCTTATCTTCGT
        M  D  R  A  L  W  F  K  F  -  G  K  K  E  Q  R  A  Y  L  R
303    AATTTGAATGATTACCCGATCTAATTTAACGTTAAAAAAAAATTAGTGCCAGATACGGGA
        N  L  N  D  Y  P  I  -  F  N  V  K  K  K  L  V  P  D  T  G
363    AAGGCTTCTCACACGAGTGAATTTTTTCGTTTTTAGTGAATCCTAACTATTAGATTATCC
        K  A  S  H  T  S  E  F  F  R  F  -  -  I  L  T  I  R  L  S
423    ATTTTCTATATGGAGATGAATGTGTAGAAGAAACAGTATATTGATAAAGATACTTTTCCA
        I  F  Y  M  E  M  N  V  -  K  K  Q  Y  I  D  K  D  T  F  P
483    AAATCAAAAGAGCGATTGGGTTGAAAAAATAAAGGATTTCTAACCATCTTGCTTTGTTAT
        K  S  K  E  R  L  G  -  K  N  K  G  F  L  T  I  L  L  C  Y
543    CCTATAAAAAAAAACCAATTAGATGGAATAAAATAGAGAATCCGTTGATGAATTTACCTG
        P  I  K  K  N  Q  L  D  G  I  K  -  R  I  R  -  -  I  Y  L
603    TCTCCGAGGTACTTAATTATTTCAAAATAGAATACCTTGTTTTGACTGTATCGCACTATG
        S  P  R  Y  L  I  I  S  K  -  N  T  L  F  -  L  Y  R  T  M
663    TATCATTTGATAATCCAAGAAACTCCCTGCTTTTTTTTAGTTTTCGGTCTAATTTGAAAT
        Y  H  L  I  I  Q  E  T  P  C  F  F  L  V  F  G  L  I  -  N
723    GGAAGAATTCCAAAGATATATAGAACTAGATAGGTCTTGGCAACACAACTTTTTCTATCC
        G  R  I  P  K  I  Y  R  T  R  -  V  L  A  T  Q  L  F  L  S
783    ACTTATCTTTCAGGAATATATTTATGGATTTGCATATGATCGTGGTTTAAATAAATCCAT
        T  Y  L  S  G  I  Y  L  W  I  C  I  -  S  W  F  K  -  I  H
843    TTTGTTGGAAAATGCAGGCGACAAGAAATCCAGTTTACTGATTGTAAAACGTTCAATTAA
        F  V  G  K  C  R  R  Q  E  I  Q  F  T  D  C  K  T  F  N  -
903    CCGAATGTATCAACAGACTCATTTGATTCTTTCTGCTAATCATTCTAACCAAAATGACTT
        P  N  V  S  T  D  S  F  D  S  F  C  -  S  F  -  P  K  -  L
963    TTTTGGGCACGAGCTCCAAGAAGAATTTGTATTATCAAAAAATACTAAAAAGGATTTGCA
        F  W  A  R  A  P  R  R  I  C  I  I  K  K  Y  -  K  G  F  A
1023   GTCATTTATGAAAATTCCAATTTTCTTTATTTTTTTGGGAAAAATTGAAATTTTAGTAAA
        V  I  Y  E  N  S  N  F  L  Y  F  F  G  K  N  -  N  F  S  K
1083   AGTCAAAAAATTTTTCTAAGGATTTTTTGGGTTATCTTAAGGCTTTTCGAAGACCCTTTC
        S  Q  K  I  F  L  R  I  F  W  V  I  L  R  L  F  E  D  P  F
1143   CTGCATTTTGTTTGGTTTCGAGGAAAATCCATGTCGGCTTCAAAGGGGACATCTCTTCTG
        L  H  F  V  W  F  R  G  K  S  M  S  A  S  K  G  T  S  L  L
1203   TTGCATAAATGGAAATATTATCTTTTCAAATTTTGGCAATGTCATTTTTCCCTGTGGTCT
        L  H  K  W  K  Y  Y  L  F  K  F  W  Q  C  H  F  S  L  W  S
1263   CCACCAAGAAGAATCTATATCAATCGGTTATCAAAGCATTCTCTCGACTTTATGGGTTTT
        P  P  R  R  I  Y  I  N  R  L  S  K  H  S  L  D  F  M  G  F
1323   TTTTCAAGTGTTCGACTCAATTCTTCAGTGGTACGGAGTCAAATGGTAGAAAATTCATTT
        F  S  S  V  R  L  N  S  S  V  V  R  S  Q  M  V  E  N  S  F
1383   CTAATAGATAATCCTATTAAGAAATTCGATACCATAGTTCGAATTATTCCTCTGGTTGGA
        L  I  D  N  P  I  K  K  F  D  T  I  V  R  I  I  P  L  V  G
1443   TCGTTGGCTAAAGCGCAATTTTGTAACGTATTAGGACATCCCATTAGTAAGTCGGTCTGG
        S  L  A  K  A  Q  F  C  N  V  L  G  H  P  I  S  K  S  V  W
1503   ACCGATTTATTGGATTCTGATATTATTGATCGGTTTGGGCGCATATGCAGAAATCTTTCT
        T  D  L  L  D  S  D  I  I  D  R  F  G  R  I  C  R  N  L  S
1563   CATTATTATAGTGGCTCTTCAAGAAAAAAGAGTTTGTATCGAATAAAGTATATACTTCGA
        H  Y  Y  S  G  S  S  R  K  K  S  L  Y  R  I  K  Y  I  L  R
1623   CTTTCTTGTGCTAGAACTTTGGCTCGTAAACACAAAAGTACTGTACGCGCTTTTTTGAAA
        L  S  C  A  R  T  L  A  R  K  H  K  S  T  V  R  A  F  L  K
1683   AGATTAGGTTCAGAATTTTTGGAAGAATTCTTTACAGAGGAAGAAAAAGTTCTTTCTTTG
        R  L  G  S  E  F  L  E  E  F  F  T  E  E  E  K  V  L  S  L
1743   ATCTTTCCAAGAGATTCTTCTATCTCACGAGGATTATATAGGGGTCCTTTTTGGTACTTG
        I  F  P  R  D  S  S  I  S  R  G  L  Y  R  G  P  F  W  Y  L
```

```
1803  GATATTATTTGTATCCATGATTTGGCCAATGATGAATGATTGGTTATGAGACTGTCTAAA
       D  I  I  C  I  H  D  L  A  N  D  E  -  L  V  M  R  L  S  K
1863  TGAAATGGAAATTATACCTAAATGAATGACGGAATAAGATTAAAAAAAAAAATGCATTCAT
       -  N  G  N  Y  T  -  M  N  D  G  I  R  L  K  K  K  C  I  H
1923  TTCTATACTGAAATGTATATCTTTATGCTTATGCAGTAAGGGTTGAATCAATTGAGTATT
       F  Y  T  E  M  Y  I  F  M  L  M  Q  -  G  L  N  Q  L  S  I
1983  CAACTTTGTTAGAGTCTCTTCTAGGAAAGGAACTTAATTTTAGATGTATACATAGGGAAA
       Q  L  C  *  S  L  F  -  E  R  N  L  I  L  D  V  Y  I  G  K
2043  GCCGTGTGCAATGAAAAATGCAAGCACGGCTTGGGGAGGGGTCTTCTTTTTTTACCAAAA
       A  V  C  N  E  K  C  K  H  G  L  G  R  G  L  L  F  L  P  K
2103  AAATG
       K
```

Figure 7.6: Sequence analysis of *mat*K gene. The amino acid is given in the single letter code. The letter M (in bold, position 48) is the start codon and the asterisk denotes the stop codon for the entire sequence (TAG). The positions of the two primers used in PCRs are highlighted in bold, italics and underlined. The *mat*K/*trn*K amino terminal region is highlighted in yellow and the Type II intron maturase region is highlighted in grey. The conserved *mat*K coding regions' consensus pattern is highlighted and underlined in red, within the grey Type II intron maturase region.

```
1173 ATGTCGGCTTCAAAGGGGACATCTCTTCTGTTGCATAAATGGAAA
      M  S  A  S  K  G  T  S  L  L  L  H  K  W  K
1218 TATTATCTTTTCAAATTTTGGCAATGTCATTTTTCCCTGTGGTCT
      Y  Y  L  F  K  F  W  Q  C  H  F  S  L  W  S
1263 CCACCAAGAAGAATCTATATCAATCGGTTATCAAAGCATTCTCTC
      P  P  R  R  I  Y  I  N  R  L  S  K  H  S  L
1308 GACTTTATGGGTTTTTTTTCAAGTGTTCGACTCAATTCTTCAGTG
      D  F  M  G  F  F  S  S  V  R  L  N  S  S  V
1353 GTACGGAGTCAAATGGTAGAAAATTCATTTCTAATAGATAATCCT
      V  R  S  Q  M  V  E  N  S  F  L  I  D  N  P
1398 ATTAAGAAATTCGATACCATAGTTCGAATTATTCCTCTGGTTGGA
      I  K  K  F  D  T  I  V  R  I  I  P  L  V  G
1443 TCGTTGGCTAAAGCGCAATTTTGTAACGTATTAGGACATCCCATT
      S  L  A  K  A  Q  F  C  N  V  L  G  H  P  I
1488 AGTAAGTCGGTCTGGACCGATTTATTGGATTCTGATATTATTGAT
      S  K  S  V  W  T  D  L  L  D  S  D  I  I  D
1533 CGGTTTGGGCGCATATGCAGAAATCTTTCTCATTATTATAGTGGC
      R  F  G  R  I  C  R  N  L  S  H  Y  Y  S  G
1578 TCTTCAAGAAAAAAGAGTTTGTATCGAATAAAGTATATACTTCGA
      S  S  R  K  K  S  L  Y  R  I  K  Y  I  L  R
1623 CTTTCTTGTGCTAGAACTTTGGCTCGTAAACACAAAAGTACTGTA
      L  S  C  A  R  T  L  A  R  K  H  K  S  T  V
1668 CGCGCTTTTTTGAAAAGATTAGGTTCAGAATTTTTGGAAGAATTC
      R  A  F  L  K  R  L  G  S  E  F  L  E  E  F
1713 TTTACAGAGGAAGAAAAAGTTCTTTCTTTGATCTTTCCAAGAGAT
      F  T  E  E  E  K  V  L  S  L  I  F  P  R  D
1758 TCTTCTATCTCACGAGGATTATATAGGGGTCCTTTTTGGTACTTG
      S  S  I  S  R  G  L  Y  R  G  P  F  W  Y  L
1803 GATATTATTTGTATCCATGATTTGGCCAATGATGAATGA 1841
      D  I  I  C  I  H  D  L  A  N  D  E  *
```

Figure 7.7: The predicted coding region of *mat*K domain within the nucleotide sequence with an ORF from the ATG position at 1173 and an in frame TGA termination codon for the coding region at position 1841

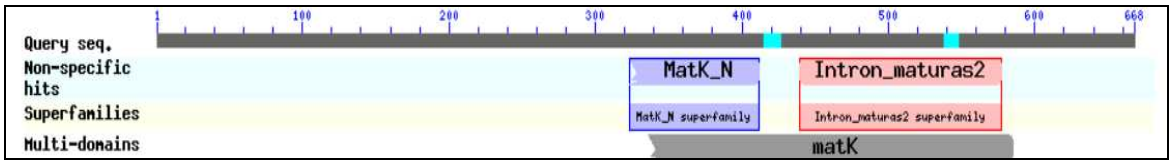


Figure 7.8: The two conserved *mat*K domain in the amino acid sequence where the domains are *mat*K/*trn*K amino terminal region and Type II intron maturase

Table 7.2: The sequences of the super-family domains of *mat*K/*trn*K amino terminal region and Type II intron maturase region

Super-family	Position	Sequence
*mat*K/*trn*K amino terminal region	L_{323}-G_{412}	LYFFGKNNFSKSQKIFLRIFWVILRL FEDPFLHFVWFRGKSMSASKGTSLLL HKWKYYLFKFWQCHFSLWSPPRRIYI NRLSKHSLDFMG
Type II intron maturase region	$I_{439} - C_{577}$	IKKFDTIVRIIPLVGSLAKAQFCNVL GHPISKSVWTDLLDSDIIDRFGRICR NLSHYYSGSSRKKSLYRIKYILRLSC ARTLARKHKSTVRAFLKRLGSEFLEE FFTEEEKVLSLIFPRDSSISRGLYRG PFWYLDIIC

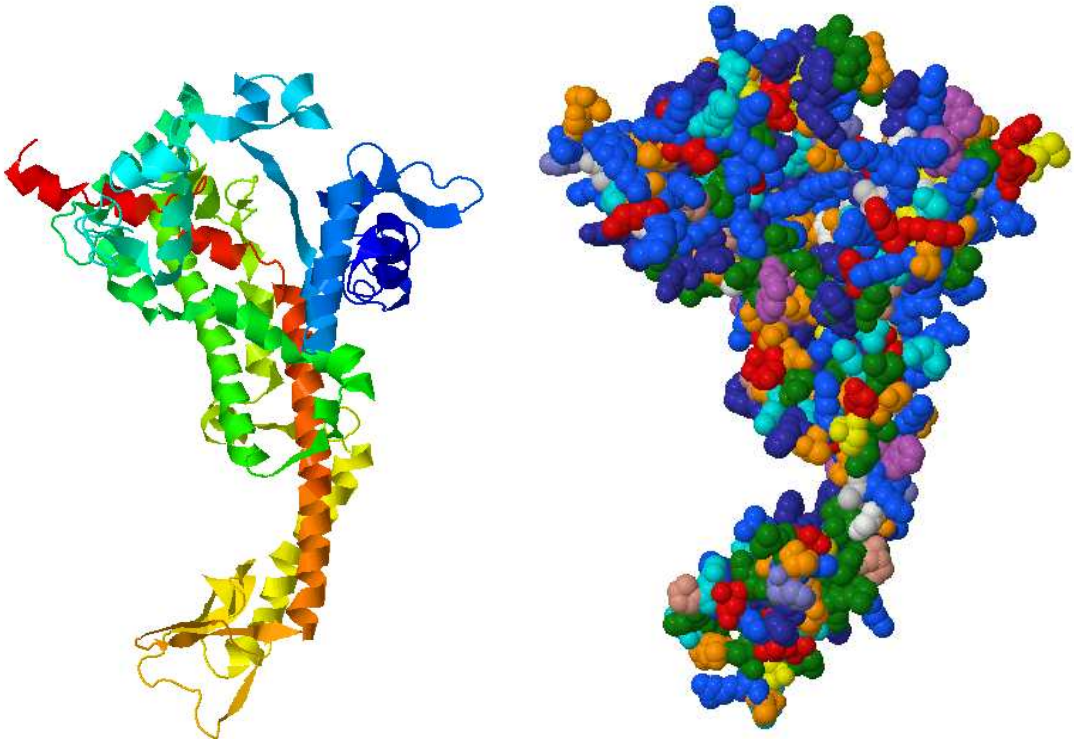
7.3.1.5 Features of the matK gene

Analysis of the *mat*K gene sequence using PROSITE scan at the http://prosite.expasy.org/ website revealed that this gene has seven distinct patterns (Table 7.3):

Table 7.3: Seven distinct patterns of *mat*K isolated from *Hylocereus polyrhizus*

Pattern	Number of site(s)	Residues
N-glycosylation	5	N_{280} – K_{283}, N_{330} – K_{333}, N_{420} – V_{423}, N_{491} – H_{494} and N_{595} – M_{598}
Protein kinase C phosphorylation	8	T_{111} – K_{113}, T_{126} – R_{128}, S_{186} – R_{188}, S_{334} – K_{336}, S_{416} – R_{418}, S_{499} – R_{501}, S_{500} – K_{502} and T_{527} – R_{529}
cAMP- and cGMP-dependent protein kinase phosphorylation	1	R_{501} – S_{504}
Casein kinase II phosphorylation	6	S_{25} – D_{28}, S_{27} – D_{30}, T_{59} – E_{62}, T_{545} – E_{548}, T_{597} – D_{600} and S_{633} – E_{636}
Leucine zipper pattern	1	L_{620} – L_{641}
Amidation site	2	N_{35} – K_{38} and F_{83} – K_{86}
N-myristoylation site	3	G_{55} – L_{60}, G_{219} – R_{224} and G_{657} – L_{662}

Figure 7.9: 3D structure of *Hylocereus polyrhizus mat*K protein generated by a 3 state prediction (Helix/Strand/Coil) by using the protein predicted tool at http://www.sbg.bio.ic.ac.uk/phyre/



ALA – 11; ARG – 36; ASN – 21; ASP – 18; CYS – 10; GLN – 14; GLU – 13; GLY – 18; HIS – 9; ILE – 42; LEU – 45; LYS – 47; MET – 8; PHE – 43; PRO – 15; SER – 44; THR – 19; TRP – 11; TYR – 21; VAL – 18; UNK - 1

7.3.2 PCR Results with GT5GENE Primers

PCR with the GT5GENE primers yielded a single product with a size of about 650bp and clear products were visible in the 1.0% agarose gel as shown in Figure 7.10.

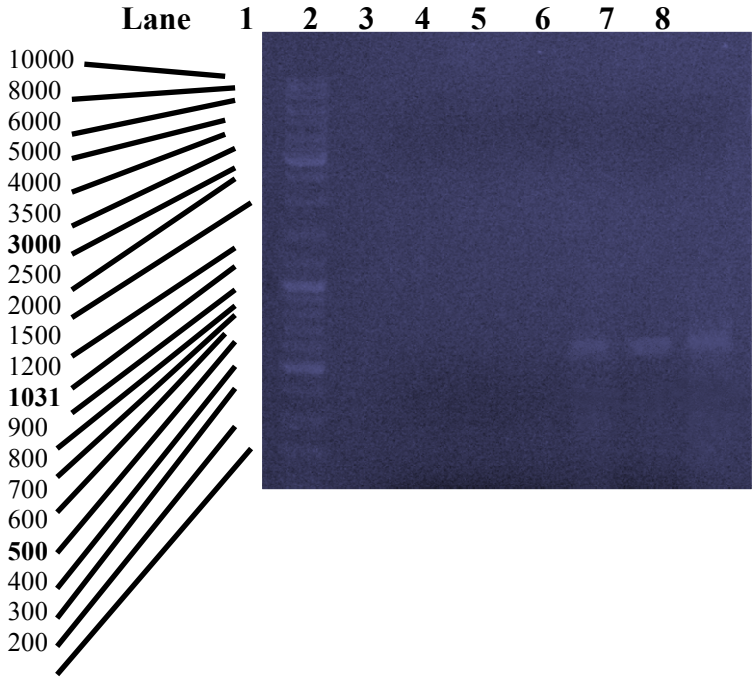


Figure 7.10: PCR results with GT5GENE primers. In Lane 1 is the DNA 1kb GeneRuler Ladder (Fermentas). Lane 2 was negative control. Lanes 3-5 were PCR products using different sets of primers which did not give any results. Lanes 6-8 were the PCR products of about 600bp generated using primers GT5GENEF1 and GT5GENER1 (at different concentrations) from total DNA from *Hylocereus polyrhizus* analysed through 1.0% agarose gel.

7.3.2.1 Nucleotide sequence of the GT5GENE protein

The DNA sequence for the *Hylocereus polyrhizus* GT5GENE protein is shown in Figure 7.11.

```
1    AACGCTTGCTCTAGAAACCCTCAAAACTCCTTCCTGGAGTGTCTCGAGGA
51   TGATCCATCTTGTGGGCGGTCCTCTTGGTCTGCAGCATGTCCACCGGAAC
101  CGGTGTCCGGACAAGTCAGAGCCTTGGCTATCGACGGACCTGGCCACCGC
151  CTAATGTCCCCACTCGCCAGAAGAAGAGGGAAGGCAACGATCGTGCCCCC
201  TGCTCCTCACCCCACCTCAGTGTGCTCTCGTGTTCATGCGTTTTAAACAG
251  CACAGCCCACGTGGGTAAAATCATAGGGTGGGCCCCACACGTCCGAGGTG
301  ACTTGCGCCACAACACAATTTTTTTGTGAACACAGGATCCATAGTGCTGT
351  CCCCCGGGCACCGTCCTCGTGTCGAACCGGAAACAGAACGGATTTTAAGG
401  ATGCGTTTGGGACTCGGTCCTCTTAACACGCTCCTTTTTATTGTCTGAAC
451  CAAAAAACGGGTGTTGGTCCTTTGCACATAAAGGGGGGTCTGGTGTCCAG
501  TTATTAGATCCAACGGCCTCGACGAACATGGTACCCGGTTCCTGTGCTCC
551  CGACCCCGACGAATTTGGCAGGTCCGCTCTTGACCCCTAGTCCTCCCCTG
601  ATCCCCGGGAACACCCTCGAGAGGGGTTTTTGAAGCCCCGGACCCGGTGG
651  GACGATTTT
```

Figure 7.11: Nucleotide partial sequence of 659bp from the purified PCR product of *Hylocereus polyrhizus* GT5GENE protein

The complete nucleotide sequence of the GT5GENE gene and the amino acid sequence of its predicted product is shown in Figure 7.12.

```
AACGCTTGCTCTAGAAACCCTCAAAACTCCTTCCTGGAGTGTCTCGAGGATGATCCATCTTG
   R  L  L  -  K  P  S  K  L  L  P  G  V  S  R  G  -  S  I  L
TGGGCGGTCCTCTTGGTCTGCAGCATGTCCACCGGAACCGGTGTCCGGACAAGTCAGAGC
 W  A  V  L  L  V  C  S  M  S  T  G  T  G  V  R  T  S  Q  S
CTTGGCTATCGACGGACCTGGCCACCGCCTAATGTCCCCACTCGCCAGAAGAAGAGGGAA
 L  G  Y  R  R  T  W  P  P  P  N  V  P  T  R  Q  K  K  R  E
GGCAACGATCGTGCCCCCTGCTCCTCACCCCACCTCAGTGTGCTCTCGTGTTCATGCGTT
 G  N  D  R  A  P  C  S  S  P  H  L  S  V  L  S  C  S  C  V
TTAAACAGCACAGCCCACGTGGGTAAAATCATAGGGTGGGCCCCACACGTCCGAGGTGAC
 L  N  S  T  A  H  V  G  K  I  I  G  W  A  P  H  V  R  G  D
TTGCGCCACAACACAATTTTTTTGTGAACACAGGATCCATAGTGCTGTCCCCCGGGCACC
 L  R  H  N  T  I  F  L  -  T  Q  D  P  -  C  C  P  P  G  T
GTCCTCGTGTCGAACCGGAAACAGAACGGATTTTAAGGATGCGTTTGGGACTCGGTCCTC
 V  L  V  S  N  R  K  Q  N  G  F  -  G  C  V  W  D  S  V  L
TTAACACGCTCCTTTTTATTGTCTGAACCAAAAAACGGGTGTTGGTCCTTTGCACATAAA
 L  T  R  S  F  L  L  S  E  P  K  N  G  C  W  S  F  A  H  K
GGGGGGTCTGGTGTCCAGTTATTAGATCCAACGGCCTCGACGAACATGGTACCCGGTTCC
 G  G  S  G  V  Q  L  L  D  P  T  A  S  T  N  M  V  P  G  S
TGTGCTCCCGACCCCGACGAATTTGGCAGGTCCGCTCTTGACCCCTAGTCCTCCCCTGAT
 C  A  P  D  P  D  E  F  G  R  S  A  L  D  P  -  S  S  P  D
CCCCGGGAACACCCTCGAGAGGGGTTTTTGAAGCCCCGGACCCGGTGGGACGATTTT
 P  R  E  H  P  R  E  G  F  L  K  P  R  T  R  W  D  D  F
```

Figure 7.12: The complete nucleotide sequence of GT5GENE gene and the amino acid sequence of its predicted product.

7.3.2.2 BLAST analysis on nucleotide sequence of the GT5GENE protein

BLAST homology search at www.ncbi.nlm.nih.gov/blast revealed the following match in the database.

(i) 28/30 positive identities (93%) with *Dianthus caryophyllus* (a)

(ii) 29/33 positive identities (88%) with *Volvox carteri* f. *nagariensis* (b)

(iii) 20/59 positive identities (34%) with *Forsythia* x *intermedia* (c)

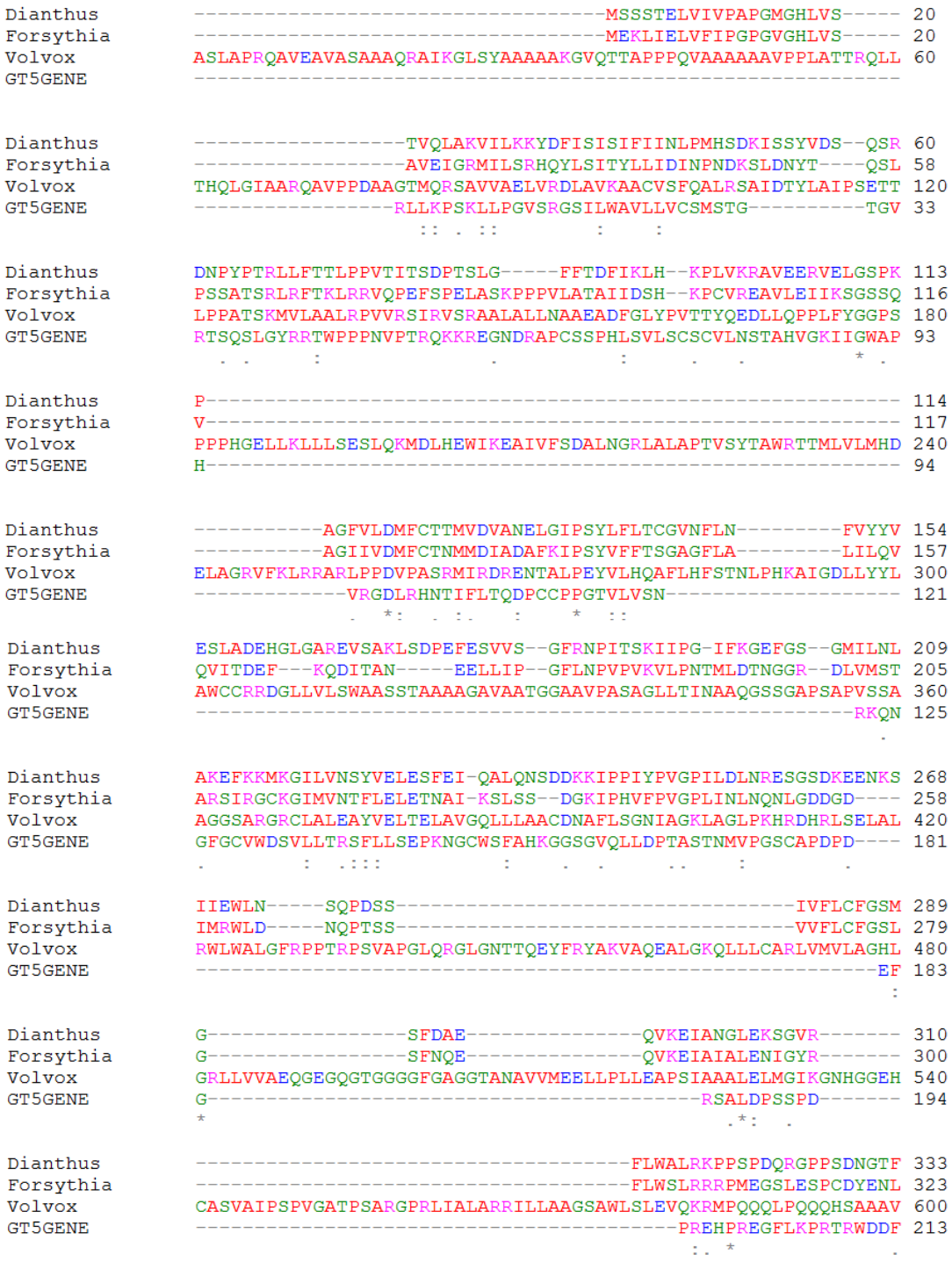
(a) Accession AB294380 - Togami *et al*., (2011)

(b) Accession XM002952324 – Prochnik *et al*.,(2010)

(c) Accession BAI65913 – Ono *et al*., (2009)

7.3.2.3 ClustalW (2.1) multiple sequence alignment for GT5GENE gene

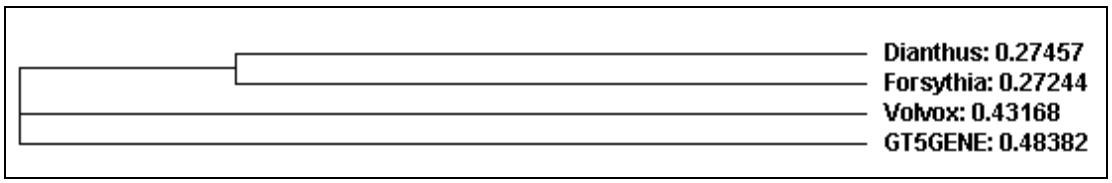
Figure 7.13: Comparison of the deduced amino acid sequence of GT5 gene from *Hylocereus polyrhizus* and sequences from *Dianthus caryophyllus* (AB294380), *Volvox carteri* f. *nagariensis* (XM002952324) and *Forsythia* x *intermedia* (BAI65913). Amino acids were aligned using CLUSTALW from http://www.ebi.ac.uk



Distance Matrix

		1	2	3	4
1	Dianthus	0.000	0.547	0.829	0.887
2	Forsythia	0.547	0.000	0.833	0.879
3	Volvox	0.829	0.833	0.000	0.915
4	GT5GENE	0.887	0.879	0.915	0.000

Figure 7.14: Phylogenetic tree generated from the CLUSTALW multiple alignment programme to show the relationship between GT5 gene from *Hylocereus polyrhizus* with *Dianthus caryophyllus* (AB294380), *Volvox carteri* f. *nagariensis* (XM002952324) and *Forsythia* x *intermedia* (BAI65913).



A phylogenetic analysis of the relationship of GT5 gene with other glucosyltransferase gene is presented in Figure 7.14. The phylogenetic tree derived from the aligned amino acid sequence showed that the GT5 gene is closely related to *Dianthus caryophyllus*, *Volvox carteri* f. *nagariensis* and *Forsythia* x *intermedia* with closest relationship at 93% identity. The amino acid sequence of the GT5 gene from *Hylocereus polyrhizus* determines the protein's 3D structure and is presented in Figure 7.18. This involved generating a set of structures representative of most of the possible folds for specific protein domains and then solving the structures for new proteins based on known fold-structure relationships. However, most of the domain families have to be characterized further before analysis of their function and evolution on a higher level can be carried out.

7.3.2.4 Sequence analysis of GT5 gene

The nucleotide sequence of the GT5 gene from *Hylocereus polyrhizus* and the deduced amino acid sequence of the translation product are shown in Figure 7.15. The partial nucleotide is 659 bp long and its translation product spans 213 amino acids. Using the ORF finder graphical analysis tool from http://www.ncbi.nlm.nih.gov/gorf, the ATG at position 87 was predicted to be the translation initiation codon and the TGA at position 329 was predicted to be the in-frame termination codon. The nucleotide sequence has a predicted average molecular mass of 23238.54 Daltons and a theoretical pI of 9.47.

The amino acid analysis showed that the sequence contains a bipartite nuclear localization signal profile within the open reading frame sequence from residues $R_{42} - R_{57}$ (Figure 7.16).

```
1       AACGCTTGCTCTAGAAACCCTCAAAACTCCTTCCTGGAGTGTCTCGAGGATGATCCATCTTG
         R  L  L  -  K  P  S  K  L  L  P  G  V  S  R  G  -  S  I  L
63      TGGGCGGTCCTCTTGGTCTGCAGCATGTCCACCGGAACCGGTGTCCGGACAAGTCAGAGC
         W  A  V  L  L  V  C  S  M  S  T  G  T  G  V  R  T  S  Q  S
123     CTTGGCTATCGACGGACCTGGCCACCGCCTAATGTCCCCACTCGCCAGAAGAAGAGGGAA
         L  G  Y  R  R  T  W  P  P  P  N  V  P  T  R  Q  K  K  R  E
183     GGCAACGATCGTGCCCCCTGCTCCTCACCCCACCTCAGTGTGCTCTCGTGTTCATGCGTT
         G  N  D  R  A  P  C  S  S  P  H  L  S  V  L  S  C  S  C  V
243     TTAAACAGCACAGCCCACGTGGGTAAAATCATAGGGTGGGCCCCACACGTCCGAGGTGAC
         L  N  S  T  A  H  V  G  K  I  I  G  W  A  P  H  V  R  G  D
303     TTGCGCCACAACACAATTTTTTTGTGAACACAGGATCCATAGTGCTGTCCCCCGGGCACC
         L  R  H  N  T  I  F  L  -  T  Q  D  P  -  C  C  P  P  G  T
363     GTCCTCGTGTCGAACCGGAAACAGAACGGATTTTAAGGATGCGTTTGGGACTCGGTCCTC
         V  L  V  S  N  R  K  Q  N  G  F  -  G  C  V  W  D  S  V  L
423     TTAACACGCTCCTTTTTATTGTCTGAACCAAAAAACGGGTGTTGGTCCTTTGCACATAAA
         L  T  R  S  F  L  L  S  E  P  K  N  G  C  W  S  F  A  H  K
483     GGGGGGTCTGGTGTCCAGTTATTAGATCCAACGGCCTCGACGAACATGGTACCCGGTTCC
         G  G  G  S  G  V  Q  L  L  D  P  T  A  S  T  N  M  V  P  G  S
543     TGTGCTCCCGACCCCGACGAATTTGGCAGGTCCGCTCTTGACCCCTAGTCCTCCCCTGAT
         C  A  P  D  P  D  E  F  G  R  S  A  L  D  P  *  S  S  P  D
603     CCCCGGGAACACCCTCGAGAGGGGTTTTTGAAGCCCCGGACCCGGTGGGACGATTTT
         P  R  E  H  P  R  E  G  F  L  K  P  R  T  R  W  D  D  F
```

Figure 7.15: Sequence analysis of GT5 gene. The amino acid is given in the single letter code. The letter M (in bold) is the start codon and the asterisk denotes the stop codon (TAG). The positions of the two primers used in PCRs are in bold, italics and underlined. A putative "PSPG-box" is shaded in grey while the bipartite nuclear localization signal profile is highlighted in red.

```
87      ATGTCCACCGGAACCGGTGTCCGGACAAGTCAGAGCCTTGGCTAT
         M  S  T  G  T  G  V  R  T  S  Q  S  L  G  Y
132     CGACGGACCTGGCCACCGCCTAATGTCCCCACTCGCCAGAAGAAG
         R  R  T  W  P  P  P  N  V  P  T  R  Q  K  K
177     AGGGAAGGCAACGATCGTGCCCCCTGCTCCTCACCCCACCTCAGT
         R  E  G  N  D  R  A  P  C  S  S  P  H  L  S
222     GTGCTCTCGTGTTCATGCGTTTTAAACAGCACAGCCCACGTGGGT
         V  L  S  C  S  C  V  L  N  S  T  A  H  V  G
267     AAAATCATAGGGTGGGCCCCACACGTCCGAGGTGACTTGCGCCAC
         K  I  I  G  W  A  P  H  V  R  G  D  L  R  H
312     AACACAATTTTTTTGTGA
         N  T  I  F  L  *
```

Figure 7.16: The predicted open reading frame within the nucleotide sequence with an ORF from the ATG position at 87 and a TGA termination codon at position 329. The bipartite nuclear localization signal profile within the open reading frame sequence from residues R_{42} – R_{57} is highlighted in red.

7.3.2.5 Features of the GT5 gene

Analysis of the GT5 gene sequence using PROSITE scan at the http://prosite.expasy.org/ website revealed that this gene has five distinct patterns (Table 7.4):

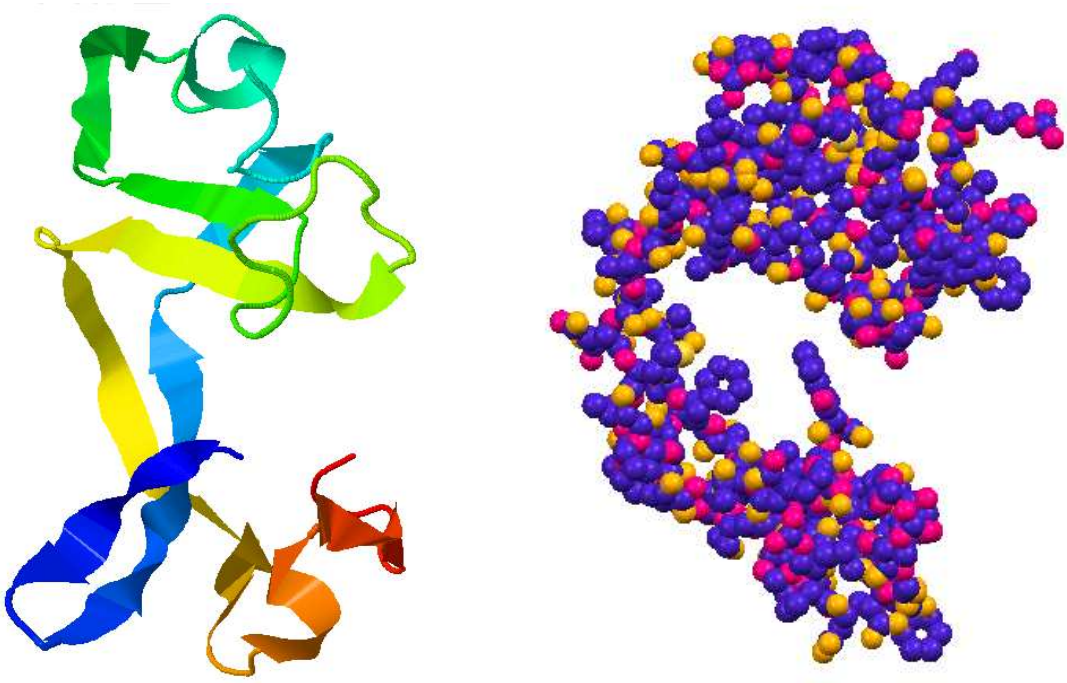
Table 7.4: Five distinct patterns of GT5 gene isolated from *Hylocereus polyrhizus*

Pattern	Number of site(s)	Residues
N-glycosylation (where its predicted feature is a carbohydrate)	1	N_{80}– A_{83}
Protein kinase C phosphorylation (where its predicted feature is a phosphoserine)	1	S_{120} – R_{122}
Cell attachment sequence site	1	R_{96} – D_{98}
Casein kinase II phosphorylation	3	S_{186} – D_{189}, S_{191} – D_{194} and T_{208} – D_{211}
N-myristoylation site	2	G_{11} – S_{16} and G_{32} – Q_{37}

Figure 7.17: Alignment of the putative 5-O-glucosyltransferase (GT5) gene from *Hylocereus polyrhizus* with CaUGT2 (Curcumin glucosyltransferase), NtGT1b (a phenolic glucosyltransferase from tobacco) and B5GT (betanidin 5-O-glucosyltransferase).

```
CaUGT2          WAPQT--------LILEHE---SVGAFVTH--------CGWNSTL--EGICAGVP-LVTW 38
B5GT            WAPQV--------LILEHE---ATGGFLTH--------CGWNSAL--EGISAGVP-MVTW 38
NtGT1b          WAPQL--------AILSHP---SVGGFVSH--------CGWNSTL--ESVRSGVP-IATW 38
GT5             WAPHVRGDLRHNTIFLTQDPCCPPGTVLVSNRKQNGFGCVWDSVLLTRSFLLSEPKNGCW 60
                ***:          :* :    . * .:          * *:*.*  ...  . *    *

CaUGT2          PFFAEQ 44
B5GT            PTFAEQ 44
NtGT1b          PLYAEQ 44
GT5             SFAHKG 66
                .   :
```

Figure 7.18 3D structure of *Hylocereus polyrhizus* GT5 protein generated by a 3 state prediction (Helix/Strand/Coil) by using the protein predicted tool at http://www.sbg.bio.ic.ac.uk/phyre/



ALA – 5; ARG – 6; ASN – 5; ASP – 8; CYS – 5; GLN – 3; GLU – 3; GLY – 11; HIS – 3; ILE – 1; LEU – 10; LYS – 3; MET – 1; PHE – 5; PRO – 12; SER – 11; THR – 6; TRP – 3; VAL – 7

7.4 DISCUSSION

Two genes, a maturaseK (*mat*K) genomic sequence and a partial RNA sequence 5-O-glucosyltransferase were isolated using the PCR approach from *Hylocereus polyrhizus*. Sequence similarity search using BLAST showed that the *mat*K was most likely a splicing Group II introns gene. The *mat*K gene cloned showed a similar coding region to the Intron Maturas2 Superfamily where type II introns uses intron-encoded reverse transcriptase, maturase and DNA endonuclease activities for site-specific insertion into DNA (Moran *et al*., 1994); and contains a *mat*K_N Superfamily region which is a *Mat*K/*Trn*K amino terminal region (Mohr *et al*., 1993) but the exact function of this region in plants is still generally unknown at the point where this study was carried out. Maturase proteins are required to enable in vivo self splicing in this type of intron and the specific region which encodes maturase protein is conserved in type II introns.

The cloned *mat*K gene from the *Hylocereus polyrhizus* displayed a consensus pattern which is highly identical to eight other deposited sequences and appears to be highly conserved where all eight organisms belong to the Cactaceae family. The *mat*K nucleotide sequence contains an open reading frame (ORF) encoding putative proteins around 102 amino acids longs and these lengths are comparable to the splicing Group II introns genes isolated from other plants such as *Hylocereus undatus* and *Selenicereus boeckmannii.* Although *mat*K have been found in many members of the Cactaceae family, the *mat*K which is a Group II intron has never previously been reported in *Hylocereus polyrhizus*.

The primary role of a Group II intron is to self-direct its extrication from gene transcripts prior to translation of the mRNA into a protein and Group II introns comprise the commonly sequenced regions *trnK/matK*, the *rps16* intron, and the *rpl16* intron (Kelchner, 2002). Group II introns are restricted to plant and fungi organelles and certain prokaryotes of cyanobacterial and proteobacterial lineages. Looking at its role in highly regulating and determining the end product of a translation process, it is fair to conclude that any changes in an intron DNA sequence would affect its RNA counterpart and this is where Group II introns are increasingly popular for sequence analysis and phylogeny construction. Examples where the Group II introns have been used for intrafamilial phylogeny estimation are: Malvaceae (Pfeil *et al*., 2002), Caryophyllales (Oxelman *et al*., 1997), and Rubiaceae (Bremer and Manen, 2000). Cactaceae is amongst 17 families in the higher plant lineages which are reported to experience the absence of the common chloroplast Group II intron *rpl2* due to the high mutation rate in Group II introns throughout the evolution process (Downie *et al*., 1991).

Mutation patterns in nucleic acid are closely related to the diversification/variation of molecular functions and different structural requirements attained during evolution (Learn *et al*., 1992; Downie *et al*., 1996) where plants adapt to various physical structures for survival and selection in preserving their most important functional purposes. This may explain why only 13 families in the plant kingdom are made of betacyanins instead of anthocyanins as their pigments, including members from the Cactaceae family. Since *rpl2* is not available in members of the Cactaceae, this study targeted the *mat*K Group II intron to phylogenetically link the *Hylocereus polyrhizus* to other betacyanin producing plants.

This was successful as the conservation of the active sites in the *mat*K coding region in *Hylocereus polyrhizus* (L - Y - R - I - K - Y - I - L - R - L - S - C - A - R - T - L - A - R - K - H - K - S - T - V - R - A - F - L - K - R - L - G - S - E - F - L - E - E - F - F - T - E - E - E - K - V - L - S - L - I) was identical (100%) to eight cactus plants using http://prosite.expasy.org/ (Advanced Scan). In addition to this conserved coding region site from $L_{505} - I_{554}$, comparison of the sequence against the Prosite and Pfam databases of the highly conserved protein motifs and patterns using the program Pfscan, revealed the presence of a protein kinase C phosphorylation sites (at residues $T_{527} - R_{529}$) and one casein kinase II phosphorylation sites (at residues $T_{545} - E_{548}$). Chloroplast Group II introns are believed to have lost their ability to move which may be due to the degradation of their ORF and in the case of *mat*K, evolved reverse transcriptase domains (Kelchner, 2002). These unique characteristics increased the probability that the Group II introns like *mat*K have evolved by vertical descent within the current host genes, making them excellent phylogenetic indicators to identify mutation patterns in sequences.

Jacquier (1996) estimated that any Group II introns would be made up of at least 600 nucleotides to maintain all essential structural features to enable proper splicing and carrying out its function. This is consistent with the results generated in this study where the length of the sequence obtained was 2107bp and total amino acid was 667 where the protein prediction using http://www.predictprotein.org/ revealed that the 3D protein structure consists of 50.22% helices, 13.94% strands and 35.83% loops where the protein appears as a compact globular domain, consistent with earlier reports on the general structure of a Group II intron (Michel *et al*., 1989; Toor *et al*., 2001).

According to Cue'Noud *et al*., (2002), the DNA region in *mat*K contains enough variable sites to resolve many phylogenetics relationships within Caryophyllales which hosts the betalain-producing families like Cactaceae and Amaranthaceae. In the same study, the authors reported that within the core Caryophyllales, there are anthocyanin-producing taxa like the Molluginaceae and Caryophyllaceae nested within the betalain-producing clades.

This shows that there is a need to fully understand and document how plants within the same clade as mentioned above can end up producing completely different sets of pigments. Nyffeler (2002) included *mat*K in his study to resolve phylogenetic relationship in the Cactaceae but did not discuss the relationship with regards to pigment synthesis within the cactus family.

Ehrendorfer (1976) provided a hypothesis to explain the presence of betalains in Caryophyllales where he reported that the ancestral taxa of the group may have evolved in dry environments and was more susceptible to wind pollination; having lost the need to attract pollinators, they would have also lost the ability to synthesize anthocyanin pigments. If Ehrendorfer (1976) was correct, then these plants would have been forced into starting betalain synthesis to produce alternative pigments when there was a reversal back to dependence on attracting pollinators or zoochory throughout the evolution. As agreed by many authors including all those cited in this chapter, we are far from fully understanding the peculiarities and uniqueness of Caryophyllales especially in terms of how, where and why betalains are produced selectively by 13 families only.

As mentioned earlier in the Chapter 7.1, *mat*K gene has been reported by Saslis-Lagoudakis *et al*., (2011) and Cordell and Colvard (2012) that the gene may prove to be more than just taxonomically important for phylogenetic studies but possibly evolved overtime with specific mutations to adapt to the changing environment including affecting genes which are involved in secondary metabolite production like pigments. Looking at how conserved and unique the *mat*K coding region in *Hylocereus polyrhizus* alongside only eight other documented cactus plants, this study is important to establish the relationship between a mutated *mat*K sequences in cactus plants and how this could have affected the changes in secondary metabolites pathways including production of betacyanins in future studies for production in a commercial scale.

In the biosynthetic pathway of betacyanin pigments, the 5-O-glucosyltransferase (5-GT) catalyses the conversion of betanidin to betanin by transferring a glucose moiety from uridine diphosphate glucose (UDP-glucose) to the 5-hydroxyl group of the betanidin where the by-product is a UDP molecule (Vogt *et al*., 1999). Using the BLAST analysis, the putative 5-GT gene isolated in this study displayed homology to the flavonoid-5-O-glucosyltransferase of *Dianthus caryophyllus*, *Volvox carteri* f. *nagariensis* and *Forsythia* x *intermedia.* This was a rather unexpected result all three plants are not betalain-producing plants.

The *Volvox carteri* f. *nagariensis* is a multicellular green algae which produces chlorophyll as its pigment (Prochnik *et al*., 2010). Nevertheless, there is a possibility that the 5-GT gene is conserved within the genome of the *Volvox carteri* f. *nagariensis* where the pigment-coding gene has evolved from other plants but retained a degree of the 5-GT DNA sequence.

According to Baldauf (2003), cellular differentiation has evolved independently in diverse lineages, including green and red algae, animals, fungi and plants yet the genetic changes that underlie these transitions remain poorly understood. It is also not impossible to find a 5-GT gene in a green algae because the betalain has been reported in a fungi, *Amanita muscaria* (Fly agaric) when at the same time, only 13 families in the plant kingdom exclusively produce betalains (Yamamoto *et al*., 2001).

Since the shikimic acid pathway is the key biosynthetic pathway in generating the basic building blocks C_6-C_3 units in many chemical structures including chlorophyll, anthocyanins, betacyanins, amino acids, other secondary metabolites like tannins and coumarins, it is possible the 5-GT in an alage is phylogenetically linked to a 5-GT gene found in a cactus plant, both being involved in pigment production and shikimic acid pathway.

In the same manner, the *Forsythia intermedia* which is a yellow flowering, drought resistant shrub, produces carotenoids as its pigment and completely lack in anthocyanins or betalains in its floral structures (Rosati *et al*., 1998) is found to have similar DNA sequence to the 5-GT cloned in the *Hylocereus polyrhizus*. The *Forsythia intermedia* which is a drought resistant shrub may have an evolved pigment-producing gene, in the same fashion where cactus plants (Caryophyllales) modified their pigment producing capabilities in their desert/drought surrounding to depending on wind for seed dispersal (energy conservation) but reverted back to producing betalains when their surrounding changed (according to the theory put forth by Ehrendorfer, 1976). This would explain the similarities between the two pigment-producing genes of *Forsythia intermedia* and *Hylocereus polyrhizus* where it is modified to resist dry environments and then possibly modified again to produce pigment in a new environment throughout years of evolution.

The closest identity of the 5-GT sequence from *Hylocereus polyrhizus* against the three matches obtained from the BLAST analysis is the *Dianthus caryophyllus* (Family: Caryophyllaceae) which belongs to the same order, Caryophyllales, where there is still much unresolved phylogenetics between members in the order. The common name for *Dianthus caryophyllus* is carnations, a perennial drought tolerant plant, making them excellent long-lasting cut flowers, producing anthocyanins as its floral pigments. The probability of the pigment-producing gene, 5-GT being highly identical (93%) is not surprising at all as the possibility of the evolved gene still retained some genetic code within the same order.

Even the isolated and characterized betanidin-5-O-glucosyltransferase by Vogt *et al*., (1999) from *Dorotheanthus bellidiformis* (Order: Caryophyllales; Family: Aizoceae) displayed homology to glucosyltransferase sequences of tobacco and tomato which are from the Order Solanales that does not produce betalains as their pigments.

Currently, only one report has successfully isolated the betanidin-5-O-glucosyltransferase which is from *Dorotheanthus bellidiformis* (Vogt *et al*., 1999) and another report on a putative betanidin-5-O-glucosyltransferase which from *Opuntia ficus indica* (Stintzing *et al*., 2005) but the sequence obtained in this study did not show any similarities to both the reports. This could be due to the reasons that *Hylocereus polyrhizus* is a cactus and *Dorotheanthus bellidiformis* is a low-growing daisy-like flower while *Opuntia ficus indica* is a cactus, it produces betaxanthins and not betacyanins in its plant. A notable behavior of *Dorotheanthus bellidiformis* is that it thrives in open sunny areas and tolerant to dry and poor soil, just like the cactuses.

This goes to show that plants which are tailored by nature to resist dry environment and drought conditions, may display similar evolution in its pigment-producing genes as well.

Up to date, various glucosyltransferase (GT) genes have been reported and most are flavonoid related glucosyltransferase (Hughes and Hughes, 1994; Gong *et al*., 1997; Ford *et al.*, 1998; Ford and Hoj, 1998; Kaminaga *et al*., 2004). Several works were carried out by Heuer and Strack (1992); Heuer *et al*., (1996) and Vogt *et al*., (1999) to characterize GT genes which are possibly involved in betacyanin production including UDP-glucose: betanidin-5-O- and 6-O-glucosyltransferases.

A proposed similarity or restricted feature which can be expected in both flavonoid GTs and betacyanin GTs is the presence of a plant secondary product glucosyltransferase box (PSPG box) sequence in the GT gene.

In this study, the putative 5-GT gene isolated contains a similar PSPG box and the sequence was closely related to the PSPG box sequences reported earlier by Vogt *et al*., (1999) on betanidin 5-O-glucosyltransferase (B5GT) from *Dorotheanthus bellidiformis*; Masada *et al*., (2007) on curcumin glucosyltransferase (CaUGT2) isolated from *Catharanthus roseus* and Taguchi *et al*., (2003) on NTGT1b from *Nicotiana tabacum*. The similarities observed between the isolated 5-GT from *Hylocereus polyrhizus* to flavonoid GT shows that the origin of betalain biosynthesis can be related to flavonoid pathway where the precursors are from the shikimic acid pathway. Up until now, it is not known whether it was possible that anthocyanins and betalains co-existed and it is still a mystery how or when did plants chose to synthesize betalains instead.

Nevertheless, increasing works carried out on characterizing betalain production related genes can soon provide enough information on the divergence of the GT genes which are crucial in the pigment production pathways and conserved sequences like the PSPG box can be used to assist the engineering of commercially important enzymes for natural pigment production. It would be of high interest and benefits to further pursue the understanding of this 5-GT gene and other pigment-producing genes in *Hylocereus polyrhizus*.

The bottom line is that, the starting point where all pigments are produced, whether they are cholorophylls, flavonoids, anthocyanins, carotenoids, betalains or quinones are all traced back to the shikimic acid pathway where the enzymes and genes would be expected to have evolved through time, adapting to different conditions but retain some percentage of the original DNA sequences in their genes. As shown in this study, the 5-GT gene from a betacyanin producing *Hylocereus polyrhizus* displays similarities to the pigment genes of an algae which produces chlorophylls, a flowering shrub which produces carotenoids and an anthocyanin producing carnation while a conserved PSPG box within the 5-GT sequence can be useful for future studies.

In conclusion, the *mat*K and 5-GT genes characterized in this chapter further propagate new knowledge on betalain biosynthetic pathway and the information generated will also help in future plans of application of the betalain genes into technologies like genetic engineering of pigment production and up scaling strategies without being dependent on crop availability. Both sequences were successfully deposited in the GenBank (NCBI) and were assigned the following accession numbers: JQ770196 and JQ770197.

CHAPTER 8
GENERAL DISCUSSION

This study was undertaken to address several issues pertaining to the development of a natural red food colourant as a viable alternative to the existing sources. The aspects which determine its viability include extractability, yield, stability, safety and possible by-products. The results obtained from this study showed that *Hylocereus polyrhizus* has the potential to an alternative to beetroot which is currently the sole source for red natural colourant.

Hylocereus polyrhizus crop which is widely cultivated in many countries including Malaysia is sold locally in the form of fresh or cut fruit, powdered juice, concentrate and wine. The Malaysian *Hylocereus polyrhizus* industry is expected to experience a rapid growth within the next 5 years (Department of Agriculture, 2008) due to high demands and the versatility that this crop offers in terms of industrial application where its downstream products promises high returns.

In this study, the red pigment was extracted and stored under different conditions to investigate the ability of the pigment to withstand some common commercialization processes and also its ability to regenerate. The study was carried out without addition of any stabilizers or additives to understand the individual stability of betacyanins unlike previous reports where most of the authors employed different types and concentration of stabilizers in their studies (Harivaindaran *et al*., 2008; Herbach *et al*., 2007; Herbach *et al*., 2006a; Moβhammer *et al*., 2005; Stintzing *et al*., 2003)

It was found that the pigments are highly extractable using water and the colour remained intense even when subjected to extraction with high volumes of water. This is a very good indication that the pigments are economically viable for applications at industrial levels. The stable betacyanin concentration and pH readings support previous claims that betacyanins are water soluble, tinctorial strength up to three times higher than other plant pigments and it has a wide pH stability range from pH 3 to 7 (Stintzing and Carle, 2007). When pigments were subjected to heat treatment to test its stability against heat application, pH for all samples was stable and pigment yield was satisfactory at room temperature and at 100ºC.

A notable discovery in this study was that at 100ºC, the pigment changed colour from its initial purple-red to scarlet red which remained even after storage for one week and gave identical absorbance at 538nm. This suggests that the betacyanins may not be susceptible to hydrolytic cleavage like other pigments when exposed to high temperature. Another factor which could contribute to its stability is the absence of betaxanthins which causes browning when processed. This is why *Hylocereus polyrhizus* should be preferred over other betalain producing plants like beetroot, spinach and bougainvillea which contain both betacyanin and betaxanthins; low abundance of betacyanin and difficulty in extracting the pigments.

Results from this study showed that samples extracted with water at room temperature and 100 ºC showed a significant increase in total betalain concentration after one week of storage under different temperature conditions. This supports the theory that betacyanins has the ability to regenerate by recondensation of hydrolysis products associated with a colour regain and encouraged when pH value is close to pH6 (Herbach *et al*., 2006c).

Overall, it can be concluded that water extraction and heat are viable methods to obtain high concentration of betacyanins and these pigments have a great stability towards factors that are important when it comes to food colouring. These findings were published in Journal of Biological Sciences, 2008 (Appendix 1) and was also awarded a Silver Medal in the University of Malaya Research Expo 2009 (Appendix 2).

Secondary metabolites are major contributors to medicinal extracts and bioactive substances with antioxidant and antibacterial properties (Dai and Mumper, 2010). Betacyanins, anthocyanins and carotenoids which are common pigments used in food colouring, have been reported to be significantly beneficial to health as antioxidants (Meyer, 2005) and protective materials (Cotelle, 2001). These additional benefits in a pigment other than just providing colour to food makes them highly prized and of great demand. Antioxidants are able to retard or halt free radicals' chain of reactions which are produced as a result of complex metabolic processes in human body and these free radicals can be detrimental to health as they can initiate development of diseases related to aging and cancer (Bae and Suh, 2007).

Antioxidant assay results in this study points to the existence of antioxidant capacities in the betacyanin pigment from *Hylocereus polyrhizus* and this link was confirmed by the detection of a single peak via HPLC qualification which was carried out using betalain standard.

Antioxidants in pigment forms are abundant in the plant kingdom and serves to function as protective mechanism in plant physiology against oxidative damages which are results of complex photosynthesis and other biological processes. This very function of the antioxidants in plants has been extended to be applied in human health to combat oxidative stress which will eventually generate free radicals.

Results in this study showed that the betacyanin pigments extracted from the *Hylocereus polyrhizus* pulp contain antioxidant properties that could be potentially exploited. What this does is that it could elevate the desirability and the value of this pigment specifically in the food and beverage industry, and possibly to other related industries like the cosmetics and pharmaceutical industries where colour and antioxidants are important elements to their businesses. The results from this chapter were published as a research paper in African Journal of Biotechnology (Appendix 3).

To further establish other value added properties in the *Hylocereus polyrhizus* extract, *myo*-inositol crystals, which is a six carbon cyclic polyalcohol (Alam, 1971) that is highly involved in plant metabolisms was successfully isolated in this study.

Myo-inositol was also reported to play an active role in human health but unfortunately, this important chemical structure in plants is yet to receive due attention due to the existence of synthetic *myo*-inositol made available to industries. *Myo*-inositols are commonly found in high concentrations in citrus fruits, beans, grains, nuts but thus far, there are no reports on its isolation from *Hylocereus polyrhizus* or any members of the Cactaceae family.

The *myo*-inositol crystals yield from this study is comparable to the common sources where it is usually found in *Prunus domestica*, *Cucumis melo*, *Citrus sinensis*, *Phaseolus vulgaris, Pisum sativum*, *Prunus dulcis* and *Arachis hypogaea* (Clements and Darnell, 1980). This finding of *myo*-inostiol within a betacyanin pigment extract from *Hylocereus polyrhizus* puts the crop in a position where it is highly regarded as not only potential as food colourant but also host to many natural health related compounds including antioxidants and *myo*-inositols. The results from this chapter were published as a research paper in Molecules (Appendix 4).

The safety of many natural plant extracts is questionable due to increased pollution over time. Hence, a comprehensive safety and toxicology study of the pigment extract using guidelines as recommended by the WHO and FAO as well as employing reliable and precise analytical instruments, would reflect the pigment's quality for application.

This thesis measured the important aspects of food safety against possible microbiological contamination, heavy metal accumulation and agricultural residues in the betacyanin pigment extracted using a customized filtration method. The comprehensive toxicology screening included foodborne pathogens like common yeast and mold, coliforms, *Escherichia coli* and *Salmonella* sp; heavy metals like arsenic, lead, mercury, cadmium, tin and antimony and agricultural chemicals like the organochlorines insecticides and organophosphorus insecticides.

A novel customized filtration method to remove a majority of mucilage, polysaccharides and possible microorganisms from the *Hylocereus polyrhizus* juice was designed and developed to obtain pure pigment extract with good clarity without using heat and enzymatic reactions where volatile and aromatic compounds can be destroyed. This is a better alternative compared to all previous methods reported where *Hylocereus polyrhizus* pigments were pasteurized and the procedure led to the degradation of betacyanins and compromised the colour of the extract.

The quality of the filtration method was also reflected in the safety and toxicology test results where the only positive result was the total bacterial count and no other pathogens, heavy metal or pesticides were detected. The total bacterial count (TBC) should not be of any alarming concern because its result of 750cfu/g is well below the usual levels of < 1000 cfu/g standard in many other commercial fruit juices currently in the market.

This is a good indication that the pigment extract obtained through a customized membrane filtration method came from a safe source and led to an excellent toxicology analysis result. The methods and results generated from this study can serve as an initial prototype and good base knowledge for future steps in the commercialization of betacyanin pigment extract as food colourant for the food and beverage industry. It is important to remember that even natural products like natural food colourants must be safe and adhere to the food quality standards. The results and approaches adopted this study with regards to the safety of the betacyanin pigment extract from *Hylocereus polyrhizus* is the first of its kind in establishing the potential of the source as an excellent and safe food colourant.

Before commercialization of any natural product for the market, there is a pre-requisite that the biochemistry, biosynthetic pathway, additional data and safety must be documented. This is even more necessary when it comes to applying up scaling strategies for commercial purposes and the basic understanding of the natural extract serves as the baseline for future commercial development. This basic understanding also can be used as a reference for any intended manipulation or advanced development such as genetic manipulation, gene selection and expression studies.

The molecular chapter in this thesis is a novel extension and a comprehensive addition to the already existing knowledge on the extracts from *Hylocereus polyrhizus*.

Currently, there are no taxonomic or genetic reports on *Hylocereus polyrhizus* in any database and the two sequences, a maturaseK (*mat*K) and a 5-O-glucosyltransferase isolated in this study are the first to be documented in the GenBank. Using molecular tools to analyse these two sequences, this study successfully linked *Hylocereus polyrhizus* to the other 13 families which all exclusively contain betacyanins instead of anthocyanins and additional knowledge in the pigment pathway production.

The cloned *mat*K gene from the *Hylocereus polyrhizus* displayed a consensus pattern which is identical to only eight other deposited sequences and appears to be highly conserved where all eight organisms belong to the Cactaceae family where its coding region is exclusively unique compared to other *mat*K genes which have been cloned from various sources.

This highly reflect that the cactus, within the Caryophyllales, has evolved in many ways due to environmental stress and its functional genes including the *mat*K, a Group II intron which is involved in regulating and determining the end product of translation processes. This supports many reports (Kelchner, 2002; Popp and Oxelman, 2001; Zhang, 2000) which claim that Group II introns are increasingly popular for sequence analysis and phylogeny construction as described in this study.

The 5-O-glucosyltransferase (5-GT), which catalyses the conversion of betanidin to betanin, displayed homology to the flavonoid-5-O-glucosyltransferase of *Dianthus caryophyllus*, *Volvox carteri* f. *nagariensis* and *Forsythia* x *intermedia.*

Literature shows that all pigments would have to be traced back to originating from the shikimic acid pathway where it would have one or another form of 5-GT gene to facilitate the biosynthesis of the respective pigments. Furthermore, a restricted feature which can be expected in flavonoid GTs and betacyanin GTs is the presence of a plant secondary product glucosyltransferase box (PSPG box) which is found in the putative 5-GT gene isolated in this study. According to Baldauf (2003), microorganisms and organisms may have evolved independently overtime but there is a high rate of over-lapping genetic materials which are conserved throughout these transitions. Hence, it was not impossible to find a betacyanin related 5-GT gene in both non-betacyanin-producing plants and betacyanin-producing plants like *Hylocereus polyrhizus.*

Both sequences were successfully deposited in the GeneBank (NCBI) and were assigned the following accession numbers: JQ770196 and JQ770197.

The data generated from this study on stability, value added characteristics, safety assessments and molecular approaches on *Hylocereus polyrhizus* pigments are most pertinent to function as a prototype and used to facilitate the next methodologies for future plans of up scaling or industrial scale applications. The results generated in this thesis also act a comprehensive study and provide the essential additional knowledge on *Hylocereus polyrhizus.*

In conclusion, all the results combined point towards the incredible potential of the betacyanin pigment from *Hylocereus polyrhizus* to be used as a food colourant. This thesis was dedicated to study the basic capacity and comprehensive study before pursuing any commercialization ambitions and including more biotechnological applications like fermentation for effective production. Future investigations should be focused on the molecular approaches to further understand the biosynthetic pathway of the betacyanin pigment and also utilization of biotechnological tool like fermentation or batch up-scaling using enzymes for enhanced production without relying on raw material supply.