Molecular Characterization of Phytoene desaturase (*CrtI*) Gene from *Paracoccus bogoriensis*

Kariuki Hellen Wambui (B.Sc. Biomedical Technology),
Reg. NO: I56/60069/2010
Center for Biotechnology and Bioinformatics,
University of Nairobi.

A thesis submitted in partial fulfillment of the requirement for the Degree of Master of Science in Biotechnology (Health and Environmental Option) In the University of Nairobi.

2014

DECLARATION

This thesis is my original work and has not been presented for a degree in any other university

Signature: …… Date: ……

Kariuki Hellen Wambui

This thesis has been submitted for examination with our approval as university supervisors.

Prof. Francis Mulaa
Department of Biochemistry
University of Nairobi

Signature: ……. Date: ……

Dr. George O. Osanjo
School of Pharmacy
University of Nairobi

Signature: ……. Date: ……

Dedication

To my family, daughter and Alexander,
I affectionately dedicate this work to you. Your support has been a never ending spring of hope, your belief in me is a constant source of strength.

ACKNOWLEDGEMENTS

To the following people whose belief in me has been the greatest incentive, for which I sincerely thank:

- **Professor Francis Mulaa**, for giving me a chance to carry out my research work in his laboratory. His guidance, interest and helpful suggestions have been a great contribution to completion of my research.

- **Dr. George O. Osanjo**, he has guided me through the "Scientific World". His skills and expertise, guidance, advice and encouragement throughout the investigations and thesis preparation were invaluable for my work.

- My colleagues in the laboratory; **Mr. Nelson Khan** and **David Kiliuku** for their help and encouragement during the study

- Technical staff in the Department of Biochemistry for their support, help and encouragement during the study

Table of Contents

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

Ala - Alanine

AD - Alzheimer's disease

A**MD** - Age related Macular Degenerate disease

ARM - Age-related maculopathy

CATH - Class Architecture Topology Homologous superfamily classification

CM - Comparative Modelling

CrtE - GGDP synthase

CrtB - phytoene synthase

Crt*N* - *squalene synthase*

CrtP - Phytoene desaturase

CrtI – Phytoene desaturase

***CrtQ*, Zds** - - ک Carotene desaturase

***CrtH*, *CrtI*SO** - cis-Carotene isomerase

CrtZ - Carotene hydroxylase gene

CrtW - β-carotene ketolase

CTAB - Cetyltrimethyl Ammonium Bromide

FAD - Flavin Adenosine Dinucleotide

FDP (**C15PP**) - Farnesyldiphosphate

FPP - Farnesyl pyrophosphate

GGDP (**C_{20}PP**) – geranylgeranyldiphosphate

Gly - Glycine

Glu - Glutamic acid (Glutamate)

His - Histidine

IPTG - Isopropyl β-D-1-Thiogalactopyranoside

I-TASSER - iterative implementation of the Threading ASSEmbly Refinement (TASSER) program

LB - Luria-Bertani

LDL - Low Density Lipoprotein

NAD	-	Nicotinamide Adenine Dinucleotide
NADP	-	Nicotinamide Adenine Dinucleotide Phosphate
NaCl	–	sodium chloride
Na_2CO_3	-	Sodium carbonate
NMR	-	Nuclear Magnetic Resonance
NO	-	Nitric Oxide
PDB	-	Protein Data Base
PDS	-	Phytoene related desaturases
PLOOHs	-	Phospholipid hydroperoxides
Pro	-	proline
RMSD	-	Root Mean Square Deviation
SCOP	-	Structural Classification of Proteins
SDS	-	Sodium Dodecyl Sulphate
VAD	-	vitamin A deficiency
Xaa	-	any amino acid
X-gal	-	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
3D	-	three dimensional structure

ABSTRACT

Carotenoids are natural fat soluble isoprenoid pigments that occur widely in micro-organisms and plants. Increasing use of carotenoids as food colourant, health supplements, cosmetic additives and animal feeds has led to high demands in the global market. Commercially available carotenoids including β-carotene, astaxanthin and canthaxanthin, are produced by chemical synthesis, isolation from natural sources or by microbial fermentation. However these methods have limits, and there is therefore need to develop alternate methods for caroteinoid production. The elucidation of the carotenoid biosynthetic pathway at molecular level holds the promise of providing a toolbox of carotenogenic genes that can be used to engineer micro-organisms for carotenoid production. The biosynthesis of carotenoids involves several steps catalysed by enzymes encoded by carotenogenic genes. Conversion of phytoene to lycopene is the rate limiting step in carotenoid biosynthesis. It is catalyzed by phytoene desaturases (*CrtI*). Lycopene, a key intermediate in synthesis of xanthophylls such as astaxanthin, is formed from phytoene in four desaturation steps. This study thus aimed at characterizing the phytoene desaturase (*CrtI*) gene from *Paracoccus bogoriensis,* an essential gene in the astaxanthin biosynthetic pathway.

The *CrtI* gene was isolated by extracting DNA from a 16 hour culture of *Paracoccus bogoriensis*. PCR was carried out using *CrtI* specific primers, the amplicon cloned in pGMET Easy vector and sequenced using Big dye chain termination method. A partial sequence of 1221 base pairs was sequenced. The sequence was translated to a protein sequence in molecular toolkit. Both sequences were then analysed by Phylogenetic tree building using MrBayes program and protein modeling using I-TASSER and COFACTOR to assess the evolution of this gene and to identify *CrtI* binding sites.

Phytoene desaturase from *Paracoccus bogoriensis* was shown to have evolved from a common ancestor with other Phytoene desaturases. *P. bogoriensis* phytoene desaturase have highly conserved regions which are putative dinucleotide binding motif βαβ fold in the N-terminal and a signature at the C terminus. *P. bogoriensis CrtI* amino acid sequence in the C-terminus appears

to have diverged slightly from other carotenoid producing bacteria, but perform a similar function. Superimposition of the 3D structure from *P. bogoriensis CrtI* on homologues from *Paracoccus N81106, Xanthobacter autrophicus* and *Bradyrhizobium* sp. ORS278 indicated that the *CrtI* structure from *Paracoccus bogoriensis* was different from its homologues from other genus of bacteria. However, *Paracoccus* N81106, *Xanthobacter autrophicus* and *Bradyrhizobium* sp.ORS278 *CrtI* 3D protein structures revealed structures with similar binding sites, indicating the 3 genus have a common *CrtI* type for desaturation, even though they produce different xanthophylls.

CHAPTER ONE: INTRODUCTION

Carotenoids are natural fat soluble pigments from a subfamily of the isoprenoid that are found widely in microorganisms and plants, and they are responsible for many of the colourful hues of plants leaves fruits and flowers. Carotenoids are among the most widely spread natural pigment derived from C_{30} and C_{40} pathway of the carotenoid biosynthetic pathway. Plants and micro-organisms use these pigments for photosynthesis and, especially, in protection against destructive photo oxidation (Krinsky, 1994).

Their use as as food colorant, health supplements, cosmetic additives and animal feeds has led to increased demand in the global market. These have led to improvement in their productivity through biotechnology by introduction of the biosynthetic enzymes gene cluster encoding these carotenoids in plants and micro-organisms to provide an effective, efficient and sustainable method to meet the market demand. Most of the commercially used carotenoids such as β-carotene, astaxanthin and canthaxanthin are produced by chemical synthesis (Johnson & Schroeder, 1995). Commercially available carotenoids are produced by chemical synthesis, isolated from natural sources or produced by microbial fermentation however not all can be produced by these methods. Elucidation of a number of carotenoids biosynthetic pathways at a molecular level has provided a toolbox of carotenogenic (*crt*) genes that can be used to engineer micro-organisms for carotenoid production (Dannert, 2000).

Though there are various commercial sources of carotenoids they are currently not sustainable. At present more than 150 genes encoding 24 different phytoene desaturase (*CrtI*) enzymes have been isolated from bacteria, plant, algae and fungi (Hirschberg *et al*., 1997). These genes can be used to engineer a variety of diverse carotenoids in recombinant micro-organisms and a number of them have been cloned from micro-organisms and plants and expressed in *Escherichia coli* (*E.coli*). From the biotechnological point of view, the ketocarotenoid astaxanthin (3 3'-dihydroxy-β, β-carotene-4, 4'-dione) is the most important of all the carotenoids used as food supplements for human (Boussiba, *et al*., 1992; Lorenz & Cysewski, 2000). Its supplementation is hypothesized to manage human health due to its neuro-protective potential, its immune-modulating potential and its antioxidant potential. Its global demand has increased and

most of the product found in the market has been synthesized using chemical method and sold at $2500 per kilogram (Johnson & Schroeder, 1996). The high price and increased demand provides a good opportunity for naturally produced astaxanthin. Naturally occurring astaxanthin has been observed in few micro-organisms, which include *Xanthophyllomyces dendrorhous* and *Haematococcus pluvialis* which accumulate the highest level of astaxanthin in its encysted cells. This study aimed at characterizing phytoene desaturase (*CrtI*) gene from *Paracoccus bogoriensis* isolated from Lake Bogoria in Rift valley Kenya. *CrtI* performs four desaturation steps resulting to synthesis of lycopene and three desaturation steps resulting to synthesis of neurosporene. *CrtI is* the key to the formation of oxgenated carotenoids (Xanthophylls).

1.1 Statement of Problem

Global market of carotenoids will increase upto US$1.2 billion by year 2015 (San Jose CA, 2010) and the current sources of these natural products are not sustainable. Demand for carotenoid is growing due to increase of carotenoids use as colourants, health supplements, additives in cosmetics and feeds. Therefore there is need for effective and sustainable method for production of this important natural pigment. A lot has been done in Biotechnology in plants by manipulating the isoprenoid pathway for the production of carotenoids. Rice has been incorporated with β-ketolase to accumulate β –carotene (Ye, et al., 2000) in the endosperm to curb vitamin A deficiency in children and pregnant mothers in African and South Asia where rice is the main staple food. However these sources do not meet the global demands.

1.2 Justification

Micro-organisms can be manipulated to produce carotenoids examples include yeast and bacteria which have been used to engineer novel carotenoids. Understanding the *CrtI* gene can enable metabolic engineering which has the potential to offer an effective, efficient and sustainable way of producing desirable carotenoid abundantly by using an appropriate micro-organism with superior yields and productivity. Indigenous micro-organisms from Kenyan soda lakes that produce carotenoid like *Paracoccus bogoriensis* can be exploited to produce these desirable pigments. The project aimed at characterizing carotenoid

biosynthetic enzymegene phytoene desaturase (*CrtI*) of *Paracoccus bogoriensis* which is the enzyme at the crucial step in formation of oxygenated carotenoid (Astaxanthin).

1.3 General objective

To characterize Phytoene desaturase (*CrtI*) gene from *Paracoccus bogoriensis* isolated from Lake Bogoria in Rift valley Kenya.

1.4 Specific objectives

- To amplify and sequence the carotenoid biosynthetic gene phytoene desaturase (*CrtI*) gene from *P. bogoriensis*
- To carry out phylogenetic analysis of the phytoene desaturase (*CrtI*) gene
- To carry out genetic analysis of phytoene desaturases (CrtI) by predicting the protein structure and comparing with other *CrtI* genes

CHAPTER TWO: LITERATURE REVIEW

2.0 Carotenoid and their derivatives

Carotenoid are a subfamily of the isoprenoids and are among the most widespread of all natural products. There are over 600 carotenoids in nature (Olson & Krinksy, 1995). Examples of carotenoids include β-Carotene, α-carotene, lycopene, lutein, zeaxanthin, astaxanthin and β -cryptoxanthin. Plants and micro-organisms use these pigments for photosynthesis, acting in light-harvesting and, especially, in protection against destructive photooxidation (Goodwin, 1980; Britton, 1983).

2.1 Carotenoid biosynthesis

Carotenoid derived from the C_{40} pathways are the most widely distributed, the pathway is shared by thousands of plant and microbial species. The conversion of two molecules of geranylgeranyl pyrophosphate to phytoene, a compound common to all C_{40} carotenogenic organism constitute the first reaction unique to the carotenoid branch of isoprenoid metabolism (Goodwin, 1980; George, 1993). In this pathway, two molecules of geranylgeranyldiphosphate (GGDP) (C_{20}PP) are condensed to form phytoene. The second, C_{30} pathway is known in only a few bacterial species, such as those of *Staphylococcus* and *Heliobacterium* (Taylor & Davies, 1976; Taylor F. R., 1984; Takaichi *et al.*, 1997). Here, two molecules of farnesyldiphosphate (FDP) (C_{15}PP) undergo condensation to form 4,4'-diapophytoene (dehydrosqualene). Homo (>C_{40})- and apo (<C_{40})-carotenoids are also known; they derive from C_{40} carotenoid precursors (Maoka *et al*., 2001; Krubasik *et al*., 2001; Mendez *et al*.,2002; Arrach *et al*., 2002). From the common intermediate, sequential dehydrogenation from phytoene through phytofluene, ک-carotene and neurosporene to lycopene variations occur in the biosynthetic pathway. Ring cyclizations and insertion of oxygen-containing functional groups, generate the tremendous diversity of catotenoid species observed in nature (Britton, G. 1993).

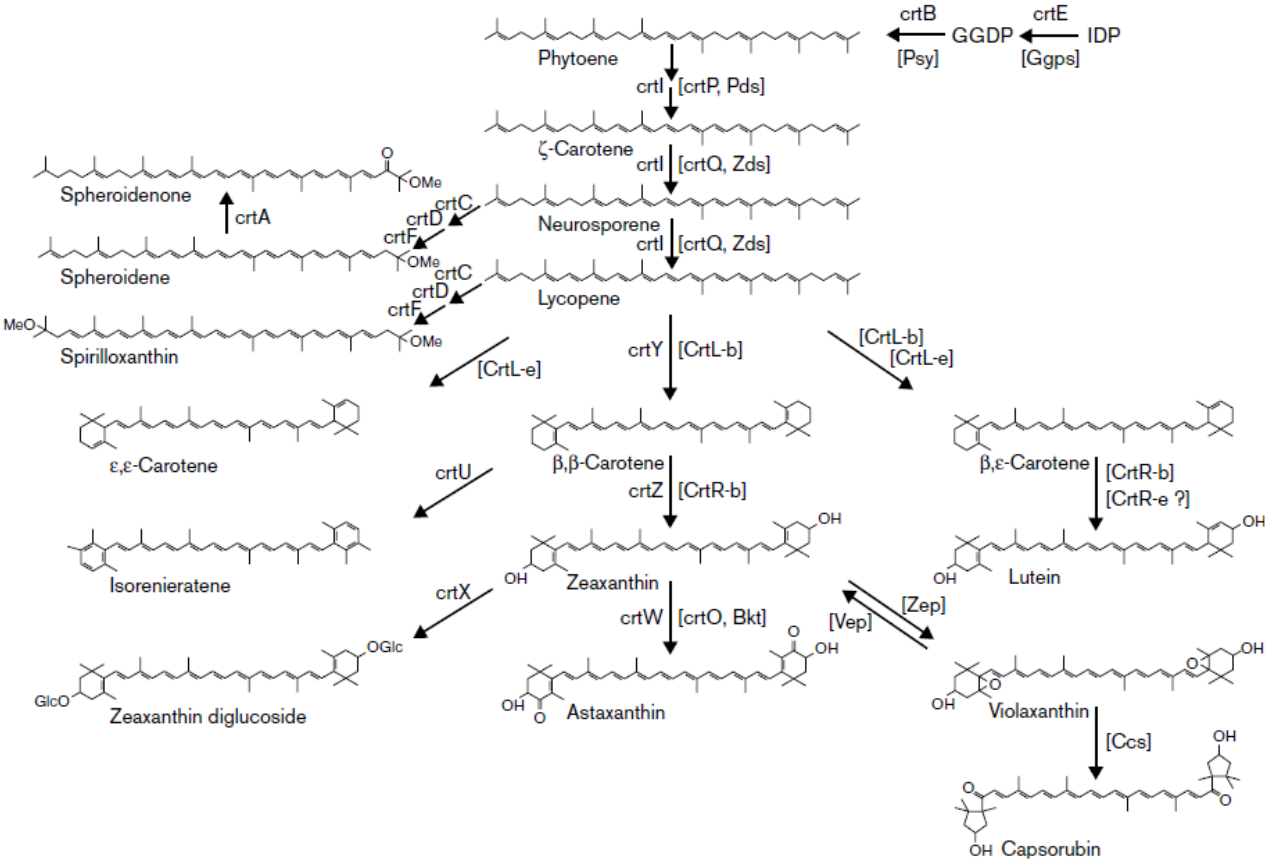


Figure 2.1: Elucidation of the main routes for the synthesis of acyclic and cyclic carotenoid at molecular level (Schmidt-Dannet, 2000)

In the last 20 years there has been increasing demand for carotenoids as food, health supplements, cosmetic additives and animal feeds. Due to the increased demand improvements have been made in the production of carotenoids for example by transforming bacteria, yeast and plant with gene cluster encoding carotenoid biosynthetic enzymes (Misawa & Shimada, 1997; Salim & Peter, 2005). *Esherichia coli* (*E.coli*) which is a non-carotenogenic has widely been used as a host for improved carotenoid production through transformation with appropriate gene cluster. Methods have been developed for increasing biosynthesis of isopentenly pyrophoshate and geranylgeranyl pyrophosphate (GGPP) as precursor in *E.coli* (Wang, *et al*., 1999; Kristala, *et al*., 2000; Seon-Won *et al*., 2001). Carotenoid biosynthesis in a noncarotenogenic microorganism such as *E. coli* requires extension of the general terpenoid pathway with the genes for geranylgeranyldiphosphate (GGDP) synthase (*CrtB*) and phytoene synthase (*CrtE*) for the production of the C_{40} carotenoids. Subsequent desaturation by phytoene desaturase (*CrtI*)

and then further modifications catalyzed by e.g. cyclases, hydroxylases, and ketolases, result in the production of different carotenoids (Britton, 1998). To date, most of the carotenogenic genes employed in recombinant biosynthesis are derived from either *Rhodobacter* or *Erwinia* species (Amstrong & Hearst, 1996; Sandman, 1994). Global carotenoid market is projected to reach US$1.2 billion by the year 2015 driven by rising consumer awareness about health benefits offered by the various carotenoids and the shift toward healthy and natural food products (San Jose CA, 2010).

Pure hydrocarbon carotenoids, such as β, β –carotene, α-carotene and lycopene, are termed carotenes, whereas their oxygenated derivatives, such as zeaxanthin and lutein, are called xanthophylls (Demming & Adams, 2002).

2.2 Sources of Carotenoids

Animals are unable to biosynthesize carotenoids, making the diet the primary source (Fraser & Bramley, 2004). Beta-carotene and alpha-carotene are found in citrus plants including carrots, sweet potatoes and pumpkins whereas astaxanthin is commercially sourced from algae, and lycopene reside in red plants such as tomatoes and grapefruits, as well as in some fish (salmon) and crustaceans. Lutein and zeaxanthin can be sourced from dark, leafy greens (spinach and kale) as well as some flowers. There are some "colorless" carotenoids (phytoene and phytofluene) that absorb only UV light and are found in algae, fungi and various plants (Myers, 2011). Few organism are capable of synthesizing astaxanthin, they include some marine eubacteria (Yokoyama & Miki, 1995; Yokoyama, *et al*., 1996) the yeast *Xanthophyllomyces dendrorhous* and the green algae *Haematococcus pluvialis* (Boussiba & Vonshak, 1991).

2.3 Uses of Carotenoids

The use of carotenoids in formulations is propelled by the science behind their health benefits, which focuses on antioxidant protection in various body systems. Studies have reported important functions played by natural carotenoids in regulating immunity and disease etiology (Park *et al*., 1998; Chew *et al*., 2009). Astaxanthin (3, 3′-dihydroxy-β, β′-carotene-4, 4′-dione) is one of the most effective carotenoids with antioxidant activity 10 times stronger than those of any other carotenoids such as zeaxanthin, lutein, canthaxanthin and β-carotene and up to 500 times stronger than vitamin E (Shimidzu, *et al*., 1996). β-carotene supplements are widely used as so-called oral sun protectants'

making carotenoids a sought-after ingredient in cosmeceuticals, either as beauty from within or in topical creams and lotions. Utra- violet radiation irradiated skin is exposed to photo-oxidative damage induced by the formation of reactive oxygen species such as singlet molecular oxygen (O_2) superoxidation radical anion O_2 and peroxyl radicals. Photo-oxidative damage affects cellular lipids, proteins and DNA and is considered to be involved in the patho-biochemistry of erythema, premature aging of the skin photodermatoses and skin cancer (Taylor *et al*., 1990). β-Carotene, other carotenoids and tocopherols are efficient scavergers of reactive oxygen species (Sies & Stahl, 1995). α-tocopherol is less active as a quencher of O_2 but occur as a corresponding lipid soluble inhibitor of lipid peroxidation in human blood. Tocopherol and carotenoids interact synergistically in scavenging process (Palozza & Krinsky, 1992; Bohm *et al*., 1997). This has led to the increased demand of these carotenoids in the cosmetic industry as skin care products.

Lycopene protects against lipid peroxidation with a resulting anti-atherosclerotic benefit. Lycopene 25mg/d or greater have been noted to reduce 10% in total serum cholesterol and low density lipoprotein (LDL) cholesterol comparable to statins whereas lower dosages did not produce such significant results (Reid & Fakler, 2010). Further, lycopene significantly reduced systolic blood pressure. Additionally, research has shown lycopene can preserve myocardial antioxidant status and significantly inhibit lipid peroxidation resulting from myocardial ischemia-reperfusion injury (Pankaj *et al*., 2006). A number of studies reported lycopene inhibited the proliferation of several prostate cancer cell lines (Ford *et al*., 2011; Paola, et al., 2010) thus the lycopene being promoted as a men's health powerhouse due to its researched protective benefits in the prostate.

Astaxanthin has been shown to protects against myocardial ischemia and reperfusion injury. In addition to reducing markers of oxidative stress, it also inhibits cancer cell growth and inflammation, and improve blood flow (Fassett & Coombes, 2009). Astaxanthin can also reduce plasma levels of nitric oxide (NO) end products and lipid peroxidation resulting in improved arterial structural and function in hypertensive subjects (Hussein *et al*., 2006). It has been confirmed by (Lee, *et al*., 2003) that supplementation with astaxanthin is practical and a beneficial strategy for management of human health due to its neuroprotective, its immunological and antioxidant potential.

Lutein has been shown to be protective against peroxidation of phospholipids, which are important components of the brain, eye and liver. Supplemented lutein incorporates into erythrocytes (human red blood cells) where it decreases phospholipid hydroperoxides (PLOOHs), which are increased in dementia patients (Nakagawa *et al*., 2009). Lutein appeared to act on PLOOHs in erythrocyte membranes, but not plasma. However, lutein's biggest role may be in the eye, specifically the retina, where it pairs with zeaxanthin as the primary macular pigments. People with the highest dietary intakes of both carotenoids have a reduced risk of Age related Macular Degenerate disease (AMD) compared with those with the lowest intakes. From another perspective, people with high plasma zeaxanthin can have as much as a 93-percent reduced risk of age-related maculopathy (ARM), while high plasma levels of lutein can result in about a 69-percent reduction in risk of ARM. In one trial, lutein protected against tissue damage caused by hepatotoxins, a benefit credited to antioxidant mechanism (Nakagawa *et al*., 2009). Lipid peroxidation is not limited to cardiovascular health. In the brain, levels of antioxidant carotenoids including lutein, zeaxanthin, beta-cryptoxanthin, lycopene and beta-/alpha-carotene are lower in patients with dementia or Alzheimer's disease (AD) (Nakagawa *et al*., 2009).

"Pro-vitamin A" Carotenoids, such as β-carotene and α-carotene, provide the primary dietary sources of vitamin A and act as the precursor of Vitamin A. Vitamin A, or retinol, is an essential nutrient for man and all mammalian species since it cannot be synthesized within the body (E-Siong, 1995). Deficiency of the vitamin results in adverse effects on growth, reproduction and resistance to infection. The most important manifestation of severe vitamin A deficiency (VAD) is xerophthalmia, and irreversible blindness may eventually occur in one or both eyes. VAD is still an important micronutrient deficiency problem in many developing countries, afflicting large numbers of pre-school children. It is often associated with protein-energy malnutrition, parasitic infestation and diarrheal disease (E-Siong, 1995). The deficiency of vitamin A is one of the most noticeable nutritional problems in many parts of the world and affect an estimated 250 million children less than 5 years of age (World Health Organisation, 2011).

These hasled to an uprising in the global market for demand of carotenoid around artificial colour and colourant to many manufactured foods, drinks and animal feeds, either in the form of natural extracts or as pure compounds manufactured by chemical synthesis. Astaxanthin, for example has been used as natural food colourant or feed additive in

aquaculture that gives the salmon its yellow colour. Sales of astaxanthin as pigment source in salmon aquaculture in 2000 in United States was estimated to be U.S. $200 million per year (Lorenz & Cysewski, 2000).The production of carotenoids by biotechnology is therefore of increasing interest, especially those from microbial sources. Fuji a chemical industry in Gustavberg Sweeden has a project to increase the production of astaxanthin as the demand is projected to rise to $200 million in 2015 most of which will be used as pigment to enhance the pink colouration of fish such as salmon. The human use market is projected to be $35-60 million (Fuji Chemical Industry Co Ltd, 2014).

2.4 *Paracoccus bogoriensis* bacteria

Paracoccus bogoriensis cells are Gram-negative, motile, aerobic, non-spore forming cocci to short rods 0.99 - 1.11 × 1.1 - 1.5 mm in size. Colonies on MH (modified Horikoshi) agar are circular, smooth, convex and orange to red in colour due to the accumulation of astaxanthin. Growth occurs between 30 and 45 °C with an optimum at 40°C and at pH 7.5 - 10.5 with an optimum at pH 9.5. Six other species of Paracoccus: *P. marcusii, P. zeaxanthinifaciens, P. carotinifaciens, P. homiensis, P. haeundaensis and P. aestuarii* have also been shown to produce carotenoids. However, the spectrum of carotenoids produced by these bacteria differs (Osanjo, *et al*., 2009) . *Paracoccus bogoriensis* (Strain BOG6T) produces carotenoids, with astaxanthin as the major carotenoid (0.4 mg/g of cells) Osanjo, *et al*., 2009.

2.5 Phytoene desaturase enyzymes

Phytoene related desaturases (PDS) are the key enzymes in the carotenoid biosynthetic pathway. The enzyme play a key role in regulating of carotenoid biosynthesis and has been implicated as a rate limiting step in the pathway of most t plant species (Chamovitz, *et al*., 1993, Lopez, *et al*., 2007). Most pharmaceutical companies target PDS in the formulation of herbicides such as norflurazon to inhibit it catalytic activity resulting in accumulation of phytoene with concomitant bleaching of plants (Boeger & Sandmann, 1998). Phytoene desaturases from cynobacteria algae and higher plants convert phytoene into ک-carotene through a 2-step desaturation process whereas functionally equivalent enzyme from bacteria and fungi in general catalyze a 4-step desaturation process leading to the maximally saturated lycopene. *Rhodobacter* genus contains a 3-step desaturation where Phytoene is converted to neurosporene (Sadmann, *et al*., 1992). Pecker, *et al*., 1992

compared different desaturases from various sources, which revealed strong similarities between all the bacterial and fungal enzymes but little conservation in the 2-steps desaturases which is commonly found in organism with oxygenic photosynthesis.

2.6 Steps involved in lycopene formation

Carotenoids get their colour due to polyene chromophore in the Carotenoid pathway leading to either C_{40} or C_{30} carotenoids. Double bonds are formed as the next step after Phytoene or diapophytoene as basic carotene molecule from two molecules of geranylgeranyl pyrophosphate (GGPP) (Goodwin, 1980) or farnesyl pyrophosphate (FPP), respectively. During these condensation reactions, a central conjugated core of three double bonds is formed. In addition, isolated double bonds (six in phytoene and four in diapophytoene) remain from the prenyl pyrophosphate precursors at positions 1, 5, 7, 1′, 5′ and 7′ (Fig. 2.3). The double bonds are therefore inserted next to the central polyenes which extends the conjugated double bond system by two. A maximum of six desaturations (=dehydrogenations) can be reached starting from phytoene and ending in the formation of 3,4,3′,4′-tetradehydrolycopene. Diapolycopene has four dehydrogenation steps which generate the maximally desaturated diapolycopene (Fig.2.3). This makes the carotenoid exist as different geometrical isomers. Each double bond can exist in a *trans* or *cis* configuration. Phytoene is synthesized as 15-cis isomer but the other double bonds are arranged in *trans* position (Goodwin, 1980). Most products of carotenoid pathways like β-carotene, zeaxanthin or violaxanthin dominate as all-trans isomers. In general, a desaturation reaction is not spontaneous but requires an energizing component. Desaturation of phytoene and other carotenes is thermodynamically favored due to the resulting extension of the conjugated double bond system which is an exergonic reaction.

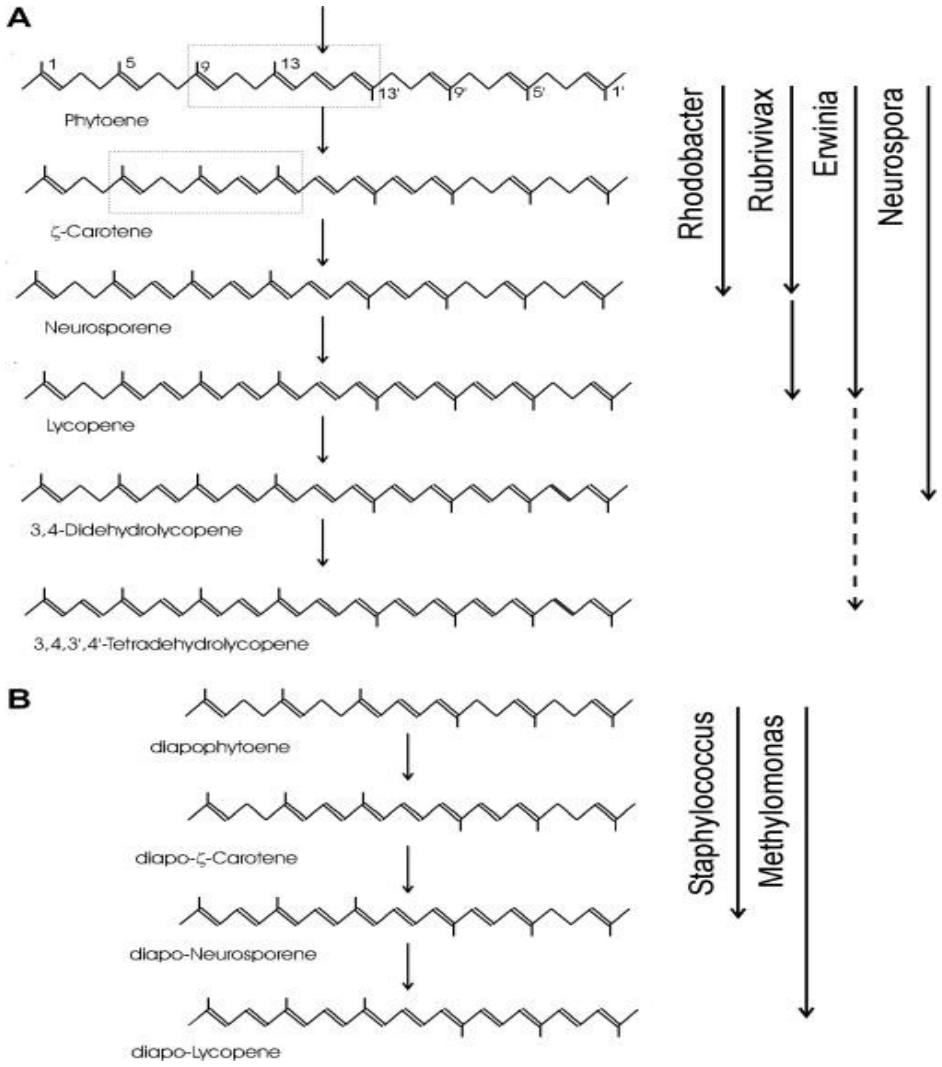


Figure 2.2: Reaction sequences of CrtI-type phytoene (A) and diapophytoene (B) desaturases from different bacteria and fungi. Arrows cover all the desaturation steps catalyzed by the individual enzymes. The proposed recognition sites for the desaturase are boxed (Sandmann, 2009).

2.7 *CrtI*-type phytoene desaturases

The first bacterial desaturase genes *CrtI* were cloned from *Rhodobacter capsulatus* (Bartley & Scolnik, 1989) and with a carotenogenic pathway to neurosporene in a three-step desaturation of phytoene and from *Erwinia herbicola* (also named as *Pantoea stewartii*) with four desaturation steps to lycopene (Fig. 2.3). *CrtI* was later cloned in *Neurospora crassa* which encodes five steps with 3,4-dehydrolycopene as the end product (Hausmann & Sandmann, 2000) . *Staphylococcus aureus CrtI* has a three-step reaction (Raisig & Sandmann, 1999) and the enzyme from a Methylomonas species a three-step desaturation. They all belong to a *CrtI* gene family together with the *CrtN* genes which

encodes diapophytoene desaturases in a C_{30} carotenoid pathway. Phytoene desaturase enzymes from different organisms catalyze the formation of different numbers of double bonds, thus leading to different reaction (Linden, *et al*., 1991).. In Eubacteria, phytoene desaturase can be divided into two structurally distinct groups *CrtI* type which include *CrtI* and CarC and *CrtP* type. The two type of *Crt* are homologous to *Neurospora crassa* (Bartely, *et al*., 1990: Schmidhauser, *et al*., 1990). The two classes of enzyme classes are thought to have evolved independently (Pecker I., *et al*., 1992). They differ with respect to their specificity for their substrates, product and chemical inhibitors. *CrtI* type of enzyme synthesizes neurosporene or lycopene from phytoene but they do not accept ک-carotene as a substrate (Bartely, *et al*., 1990: Chamovitz, (Ph.D thesis)., 1993: Fraser, *et al*., 1992). *CrtP* type desaturase on the other hand produce ک-carotene from phytoene (Bartley, *et al*., 1991: Hugeueney, *et al*., 1992: Pecker I., *et al*., 1992).The differential inhibition in these two types of phytoene desaturases enzyme has been exploited to engineer a herbicide resistant tobacco by introduction of the gene encoding norflurazon insensitive from *Erwinia uredovora CrtI* (Misawa, *et al*., 1993).

Oxygenic phototrophs require three enzymes: phytoene desaturase (*CrtP*, Pds), ک-carotene desaturase (*CrtQ*, Zds) and cis-carotene isomerase (*CrtH*, *CrtI*SO). The first two steps of desaturation are carried out by *CrtP* from phytoene to ک-carotene through phytofluene. Two additional steps are catalysed by *CrtQ* fromک -carotene to lycopene through neurosporene. Neurosporene and lycopene are isomerized to poly-cis forms during *CrtQ* desaturation step then *CrtH* isomerises all *trans* form (Masomoto, *et al*., 2001). Bacterial phytoene desaturases uses one enzyme *CrtI* to convert phytoene to lycopene. Primitive cyanobacterium *Gleobacter violaceus* PCC 7421(*G. violaceus*) uses this type of *CrtI* and the homologous gene *CrtP*, *CrtQ* and *CrtH* are not found in its genome (Steiger, *et al*., 2005: Tsuchiya, *et al*., 2005). This made *G. violaceus* the oxygenic phototroph the first primitive cyanobacteria shown to use this type of *CrtI*. The observation lead to the following evolutionary scheme for this finding: The desaturation of phytoene was initially carried out by *CrtI* in ancestral cyanobacteria, *CrtP* and related desaturase gene were acquired and ultimately there was replacement of *CrtI* by *CrtP* (Tsuchiya, *et al*., 2005). Among anoxygenic phototrophs, purple bacteria, green filamentous bacteria and *Helibacteria* use *CrtI* whereas green sulfur bacterium uses *CrtP*, *CrtQ* and *CrtH* (Takaichi, 2009).

2.8 Protein conservation in Phytoene desaturases

Enzymology study done on different carotene desaturases indicates a dehydrogenation mechanism in which protons and electrons are transferred to an appropriate acceptor (Sandmann G., 1994). In *Erwinia* phytoene desaturases, Flavin Adenosine Dinucleotide (FAD) is involved as a coenzyme in the desaturation reaction (Fraser, *et al*., 1992) whereas the phytoene desaturase from *Synechoccus* is dependant either on Nicotinamide adenine dinucleotide (NAD) or Nicotinamide adenine dinucleotide phosphate (NADP). Glenn, *et al*., 1990 identified conserved residues with possible functional homology and they identified two members GRase and LipDH of the highly homologous disulfide oxireductase enzyme family. They found a characteristic dinucleotide (ADP) binding domain with sequence Gly-Xaa-Gly-Xaa- Ala/Gly- Xaa-Ala-Xaa. Glycine was universally conserved among FAD and NAD (P) binding enzymes (Hanukoglu & Gutfinger, 1989: Rice, *et al*., 1984: Carothers, *et al*., 1989) and is present at the N-terminal of the carotenoid desaturases. Greatest homology is found at the C terminal ends of the β-strands, reflecting the location of FAD- binding pockets. The C- terminal has His, Pro and Glu residues. Amino acids in the GRase and LipDH β-sheet are mainly hydrophobic except for GRase Glu and LipDH, which interact with the adenine ribose hydroxyl of FAD (Glenn, *et al*., 1990). Similar studies done by Zhenjian, *et al*., 2007 on *Deinococcus radiodurans* with comparison to phytoene desaturase from the cyanobacterium *Gleobacter violaceus* exhibited about 68 % global identity and nearly 80% similarity in the carotenoid desaturase conserved amino acid protein sequence. Two highly conserved regions were detected in DR0861 and DR0810 genes i.e the putative dinucleotide-binding motiff (βαβ fold) in the N-terminal region and the bacterial-type phytoene desaturase signature at the C terminus. Mutation in this region was found to destroy the activity of *Rhodobacter capsulatus CrtD* (Alberti, *et al*., 1995; Armstrong, *et al*., 1990).

2.9 Protein Modelling Using I-TASSER in generation of 3D structure

Determination and prediction of protein three dimensional structures (3D) can be classified on the basis of the predominant information that is used to calculate the model. The experimental methods used include X-ray crystallography (Blundell & Johnson, 1976) and multi-dimensional nuclear magnetic resonance (NMR) techniques (Bax A., 1989). Two approaches are used in protein, modelling theoretical approaches and empirical methods. Physical prediction methods are based on interactions between atoms and

include molecular dynamics and energy minimization (Brooks III, *et al*., 1988, whereas the empirical methods depend on the protein structure that have been determined by experiment. They include combinatorial (Cohen & Kuntz, 1989) and comparative modelling (Blundell, *et al.*, 1987; Sali, *et al.*, 1990; Swindells & Thronton, 1991). The two methods used have a root mean square deviation (r.m.s.d) error of approximately 1A for sequence that have sufficiently similar homology with known 3D structure (Tophan, *et al.*, 1991) in contrasts physical prediction methods and combinatorial modelling calculate structure with r.m.s.d errors of approximately 3.5A for small protein (Cohen & Kuntz, 1989; Wilson & Doniach, 1989). X-ray crystallography and NMR which (Clore & Groneborn, 1991 are restricted to sequences with closely related proteins with known 3D structures.

Computational methods for predicting three dimensional (3D) protein structures have been divided into three categories; Based on the Protein Data Base (PDB) template structures; Comparative Modelling (CM) where evolutionary related homologous template are identified by sequence or sequence comparisons (Altshul, *et al.*, 1997) and high-resolution models can be generated by simply copying the framework of the template or by satisfying the spatial restraints collected from the structures; *ab initio* modelling where protein are built from the beginning due to lack of structurally related protein the PDB library. It is the hardest and the success is limited to small proteins with less than 120 amino acids (Zhang, 2008; Jauch, *et al.*, 2007).

Significant advantages have been shown by using composite approach in structure prediction, which combines various techniques such as threading, *ab* initio modelling and atomic-level structure refinement approaches (Zhang, 2008; Das *et al*., 2007; Zhang Y, 2007; Zhou Y. *et al*., 2007). I-TASSER (iterative threading assembly refinement) is one of the composite approaches and has been ranked as the best method for the automated protein structure prediction (Battey, *et al*., 2007; Zhang, 2009; Cozzetto, *et al*., 2009). Biological usefulness of a predicted protein model depends on the accuracy of the structure prediction (Zhang, 2009). High resolution models have been seen to have a root mean square deviation (RMSD) values in the range of 1-2Â, generated by CM using close homologous templates. This usually meet the highest structural requirement and are sometimes suitable for computational ligand binding studies and virtual compound screening (Ekins, *et al.*, 2007; Becker, 2006; Brylinski & Skolnick, 2008). Model that

have medium resolution are roughly in the PMSD range of 2-5Â and typically generated by threading and CM from distantly homologous templates. They can be used for identifying the spatial locations of functionally important residues such as active sites and the site of disease mutation (Arakaki, *et al.*, 2004; Yue & Moult, 2006; Boyd, *et al*., 2008).

The most common motivation for predicting the protein structure is to use the structural information to gain insight into the protein biological function. Convenient approach used in the structure based functional assignment involves global structural comparison of protein pairs for fold recognition and family assignment which in many cases can be directly used to infer function (Zhang, 2009; Brylinski & Skolnick, 2008). I-TASSER uses combination of local and global structural similarities in annotation of the biological function of known protein. Its uniqueness is in the significant accuracy and reliability of full length structure prediction for targets of varying difficulty and the comprehensive structure based function predictions (Ambrish, *et al.*, 2010).

2.9.1 I-TASSER

I –TASSER is a hierarchical protein structure modeling approach based on the secondary-structure enhanced Profile-Profile threading Alignment (PPA) and the iterative implementation of the Threading ASSEmbly Refinement (TASSER) program. Threading refers to a bioinformatic procedure for identifying template proteins from unsolved structure databases that have similar structure or similar structural motif as the query protein sequence. A query sequence is matched against a nonredundant sequence database by position-specific iterated BLAST (PSI-BLAST) (Altshul, *et al*., 1997) to identify evolutionary relatives. A sequence profile is then created based on the multiple alignment of the homologs. This is used to predict the secondary structure using PSIPRED (Jones, 1999). Sequence profile and the predicted secondary structure assist in the threading through a representative PDB structure library using LOMETS (Wu & Zhang, 2007). LOMETS is a locally installed meta- threading server combining seven state of the art threading programs (FUGUE, HHSEARCH, MUSTER, PROSPECT, PPA, SP3 and SPARKS). If they are individual threading programs, the template are ranked by a variety of sequence-based and structure based scores. Quality of the template alignment judged based on the statistical significance of the best threading alignment (Z-score). The Z-score is defined as the energy score in standard deviation units relative to the statistical mean of all alignments. Structural assembly is the second stage where continuous fragments in

threading alignment are excised from the template structures. They are used to assemble structural confirmations of the section aligned well, with the unaligned regions built by an initio modelling (Wu, *et al.*, 2007; Zhang, *et al.*, 2005). The fragments assembly is performed using a modified replica exchange Monte Carlo simulation technique (Zhang, *et al.*, 2004) which implement several simulations in parallel at different temperatures. Overall simulation is guided by a composite knowledge based force field which include; General statistical terms derived from PDB (Cα/side-chain correlations, H- bonds and hydrophobicity); Spatial restraints from threading templates; Sequence-based contact predictions from SVMEQ. The conformations generated in the low-temperature replicas during the refinement stimulatin are clustered by SPICKER (Zhang & Skolnick, 2004) with the purpose of identifying low free-energy states. Cluster entroids are then obtained by averaging the 3D coordinates all the clustered structural decoys. Third stage the fragment assembly simulation is performed again starting from the selected cluster centroids. The purpose of the second iteration is to remove steric clashes and to refine the global topology of the cluster centroids. Decoys generated during this second simulation are clustered again and the lowest energy structures are selected as input for REMO (Li & Zhang, 2009). This generate the final structural models by building all-atom models from Cα traces through the optimization of hydrogen bonding networks. The fourth stage of the query protein is inferred by structurally matching the predicted 3D models against the proteins of known structure and function in the PDB. The libraries have been constructed independently and biweekly updated; at present these library of 5,798 nonredundant entries with known EC mumbers (Barret, 1997), a library of 26,045 non-redundant entries with known GO terms (Ashburner *et al.*, 2000) and library of 19,658 nonredundant entries with known ligand-binding sites. GO library analogs of the query protein are matched based on the global topology using TM-align (Zhang & Skolnick, 2005) and consensus is derived based on the frequency of occurrence of the GO terms. The structural analogs in th EC and binding site are matched based on global and local structural similarity (Ambrish, *et al*., 2010). Fuctional analogs from the global search are ranked based on the conserved structural patterns present in the model, measured using a scoring scheme that combines template modelling score (TM-score) 66 RNSD,sequence identity and coverage of the structure alignment (Ambrish, *et al*., 2010). TM-score is used to assess the topological similarity of protein structure pairs with a value range of 0,1 a higher score indicates better structural match. The local similarity search looks for conserved spatial motifs in the predicted I-TASSER model. Candidates are ranked based on their structure and sequence

similarity to functional cavities (binding pockets) in known structures. Finally results from the two search (global and local) are combined to present a comprehensive list of functional analogs.

Estimation of accuracy of the structure predicted a confidence score named *C*-score is defined based on the quality of the threading alignment and the convergence of the I-TASSER structural assembly refinement simulations i.e.

$$\text{C-score} = \ln\left[\frac{M}{M_{tot}} \times \frac{1}{[\text{RMSD}]} \times \frac{1}{7}\sum_{i=1}^{7}\frac{Z(i)}{Z_0(i)}\right]$$

Whereby M is the number of structure decoys in the cluster and M_{tot} is the total number of decoys generated during the I-TASSER simulations. [RMSD] is the average RMSD of the decoys to the cluster centroid. Z (*i*) is the Z-score of the best template generated by *i* threading in the seven LOMETS programs and Z_0 (*i*) is a program specified Z-score cutoff for distinguishing between good and bad templates. Functions prediction the confidence score is defined based on the *C*-score of the structure prediction and the global/local similarity between the predicted models and their structural analogs in the PDB (Ambrish, *et al*., 2010).

2.9.2 Cofactor prediction for visualization of binding site

COFACTOR combines both local and global structural comparison algorithms to deduce the biological function of protein starting from the 3D structure obtained during protein modeling using I-TASSER. After structural assembly the ligand binding information is then derived from the known protein in a comprehensive protein-ligand complex library. In the library the templates are identified using both global and structure comparisons between the target and template proteins. Test having a sequence identity >30% to the target are removed from both the structure and function libraries. To examine the ligand-binding details the binding pocket of the template are scanned through the target structure to identify the best geometric and sequence matches. The binding pose the ligand in the target structure is predicted based on the local alignment of predicted and template binding site residues. The global structure match is performed by TM-align (Zhang & Slolnick, 2005) which identifies the best alignment between the target and the template structure. A

heuristic dynamic programming iteration using TM-score rotation matrix is used. TM-score (Zhang & Skolnick, 2004) with the value in (0, 1) is reported to assess the global structural similarity. All templates with a nonrandom structural similarity (TM-score >0.3) to the target structure (Xu & Zhang, 2010) are selected for further processing.

Local match between the target and the template proteins is conducted in two steps: First is to identify a set of conserved residues in the target that are used as the seed to local structure comparisons. Multiple sequence alignment (MSA) of the query target sequence is constructed by PSI-BLAST (Altschul, *et al*., 1997) through the NCBI non-redundant (NR) sequence database. Conserved residue are identified from MSA based on their Jensen-Shannon divergence score (Capra & Singh, 2007). Triplets of these conserved residues along with their two flanking residues are used for generating initial candidate binding site motifs. Second step involves superimposition of the candidate binding site motiff on the template motif. The rotation and translation matrix acquired from this local superimposition is used to bring the complete structure of query and template protein together. The local similarities between query and the template protein is used to L_{sim} Equation (Schmidt, *et al*., 2011). The best local match with all the probable set (highest L_{sim}) is selected. Viewing of the 3D structure was done using RasMol. It allows one to rotate protein structure, zoom on them, and render them in different ways and using various colouring schemes, label atoms and residues.

CHAPTER THREE: MATERIALS AND METHODS

3.1 Isolation of *Paracoccus bogoriensis*

The bacteria, *Paracoccus bogoriensis* was revived from a glycerol stock maintained at –20º C in the department of biochemistry, college of health sciences, University of Nairobi. A loop full of the glycerol stock was taken and streaked on LB agar plates. The LB medium was prepared using NaCl 0.5 g/l (Sigma, Germany), bacto-tryptone 0.5g/l (Sigma, Germany), yeast extract 0.25 g/l (Sigma, Germany) and bacto-agar 0.75 g/l (Sigma, Germany). The medium was sterilized using an autoclave (75X, All America , USA) at 121ºC for 30 min. Sterile Na_2CO_3 solution was then prepared by dissolving 15 g of anhydrous Na_2CO_3(Sigma, Germany) in 100 ml distilled water. The Na_2CO_3 solution was filter sterilized using a 0.20 µm disposable cellulose acetate syringe filter (Iwaki, Japan) and used to adjust the pH of the sterile LB medium to 9.2. The medium was then poured out on plates, and allowed to solidify. The bacterium, *Paracoccus bogoriensis*, was streaked onto the solidified LB agar plates and incubated at 37º C for 16 h. A single 16 h colony was used to inoculate sterile 150 ml LB broth medium (NaCl 1.5 g/l (Sigma, Germany), bacto-tryptone 1.5g/l (Sigma, Germany), yeast extract 0.75 g/l (Sigma, Germany)) at pH 9.2 (adjusted using sterile 15% Na_2CO_3 (Sigma, Germany) solution) and incubated for 16 h with continuous shaking at 130 rpm at 37^0C in an Orbital Shaker. Glycerol stocks of 40 % sterile glycerol and 60 % *Paracoccus bogoriensis* liquid culture in a 1.5 ml eppendorf tube was prepared and stored at -20ºC.

3.2 Genomic DNA Extraction

The procedure described by Kate Wilson (1987) was used to harvest total genomic DNA from the bacterial isolates. In each case, 5ml of an overnight culture was subjected to centrifugation (Eppendorf centrifuge) at 8720.4g for 5 minutes and the pellet resuspended in547µl of Tris-EDTA (TE) buffer. Into this, 233µl of cell lysis buffer containing, lysozyme, sodium dodecyl sulphate (SDS), cetyltrimethyl ammonium bromide (CTAB) and proteinase K was added and the mixture incubated for 1hr at 37º C. DNA was then extracted sequentially with Phenol/Chloroform/Isoamyl alcohol (25:24:1) and Chloroform/Isoamyl alcohol (24:1). DNA was then precipitated with 0.6 volumes isopropanol alcohol and the pellet resuspended in 50µl of TE buffer (pH 8) containing

RNAse to get rid of RNA. DNA was quantified using a nanodrop 1000 (Thermo Scientific, USA) and its quality checked on a 1% agarose gel before storage at -20ºC.

3.3 DNA analysis by TAE/agarose/EtBr gel electrophoresis

The quality of genomic DNA was analysed on 1% (w/v) agarose (Sigma,USA) gel in 1× TAE (Tris-Acetate-EDTA) buffer. A 1% TAE /agarose/Ethidium Bromide gel was prepared by heating up 0.35 g agarose in 35 ml of 1× TAE to boiling. The hot agarose solution was cooled down to 60ºC before Ethidium bromide was added (final concentration: 0.5µg/ml). The solution was then poured into the gel casting chamber and appropriate combs placed in position. The gel was left to polymerise for 30 minutes then transferred to the electrophoresis chamber with the slots facing the cathode and covered with 1× TAE buffer (running buffer). 6× orange DNA loading dye (Fermentas, USA) was premixed with the DNA samples in the ratio 1:6µl of sample (final concentration:1×) prior to loading of the samples into the sample wells in the gel.

The samples were electrophoresed at 5V/cm (80V) for genomic DNA and 10V/cm (120V) for PCR products for 45 minutes -1 hr. The current was supplied by an electrophoresis power supply (Consort EV265). A 1kb DNA ladder (fermentas, USA) was run alongside DNA samples. The DNA bands were visualised under a UV transilluminator (Herolab E.A.S.Y 442K, Germany).

3.4 PCR amplification of Phytoene desaturase (*CrtI*) gene

For a 50 µl polymerase chain reaction (PCR) reaction, the PCR reagents were mixed in a thin-walled PCR reaction tube (Simport, Canada) according to table (1). The primers, *CrtI* Forward 5'*CTATGTCTGGCACGATCAGG* 3'and *CrtI* Reverse 5'*GCCCACCAGATAGAAGTTGC* 3' from IDT Belgium were used in this reaction.

Table 3.1 PCR Master mix reaction

Reagent	Volume	Final concentration
10× Dream Taq Buffer	5µl	
Nuclease free PCR water	36.25	
dNTPs (10mM)	5µl	2mM
Forward primer (50µM)	1µl	1µM
Reverse primer (50µM)	1µl	1µM
DNA template	1µl	
$MgCl_2$		2mM
Dream Taq DNA polymerase	1.25	
Final volume	50µl	

The reaction was run in a TProfessional thermocycler (Biometra, Germany) according to the cycling conditions in table (2).

Table 3.2 CrtI PCR cycling conditions

	Temperature	**Time** (min)	Cycling steps		
Step 1	94ºC	5 min	Initial denaturation		
Step 2	94ºC	1 min	Denaturation		
Step 3	60ºC	1 min	Annealing	3	35 times
Step 4	72ºC	3.30 min	Extension		
Step 5	72ºC	7 min	Final extension		
Step 6	4ºC	Pause			

The quality of the PCR products was checked on a 1% agarose gel as described in section 3.3. Purification of the product for sequencing was done as described section 3.5.

3.5 Gel purification of PCR products

MinElute Gel Extraction Kit (QIAGEN) was used to purify PCR products from the gel. The PCR products were electrophoresed on a 1 % agarose gel as described in section **3.3**.Using a sterile scalpel and under UV light, agarose bands containing the sample of interest were excised from the gel and weighed. To the gel slices of about 0.3 g, 300 µl of binding buffer QG was added in micro-centrifuge tubes and incubated at 60º C until the agarose slices were dissolved. 300µl Isopropanol was added to the solution to precipitate

the DNA. The sample was applied on a MinElute column assembled with a collection tube and centrifuged for 1minute at 17091.984g.The flow through was discarded and the MinElute column was placed back in the same collection tube. 750 µl of PE buffer was added to the column and centrifuged for 1 minute, and the flow through discarded. The MinElute tube was centrifuged again for 1 minute. The column was placed in 1.5ml micro-centrifuge tube and 50 µl of nuclease free water was added for placed at room temperature for 20 minutes before being centrifuged at 14,000rpm for 1minute.The purified sample was run on agarose gel electrophoresis. The purified samples were used for cloning

3.6 Preparation of competent cells

BL21 strain of *Escherichia coli* was used and cells were prepared according to the method of Christine *et al.*, (1997). An aliquot of 100 µl of frozen BL21 cells were inoculated into 500 ml of sterile LB pH 7.2 and cultured overnight. 200 mlof the overnight culture was sub-cultured into sterile LB broth for 4hours until the Optical Density (OD) was 600nm.100ml of the culture was pippeted into a sterile falcon tube and centrifuged (Jouan tabletop centrifuge) at 2180.1 g for 10 minutes at 4^0 C(the culture was on ice).The supernatant was discarded and the pellet was resuspended in 12.5 ml of ice cold 0.1M $MgCl_2$ for 5 minutes. The suspension was centrifuged at 4˚C for 10 minutes at 4000rpms. The pellet was resuspended in 500ml ice cold 0.1M $CaCl_2$ and kept on ice for 20 minutes. The suspension was centrifuged at 1395.264 g for 10 minutes and the supernatant discarded. The pellet was finally resuspended in 200 ml ice cold $CaCl_2$ and 85 ml of 15 % glycerin. Suspension was then dispensed in 25 µl aliquots and frozen at -80^0C.

3.6.1 Assessing cell competence

5 µl of puC 18 was mixed with 100 µl of thawed competent cells in 1.5 ml eppendorf tubes gently. The cells were placed on ice for 30 minutes. The mixture was heat shocked for 90 seconds and immediately placed in ice for 10 minutes. 600 ml of super optimal broth with catabolite repression (SOC) solution was added and cells placed in water bath at 37^{0C} for 90 minutes. Meanwhile 8 µl of Isopropyl β-D-1- thiogalactopyranoside (IPTG) and 200 µl of 5-bromo-4-chloro-3-indolyi-β-D galactopyranoside (X-gal) were evenly spread on LB /ampicillin plate and allowed to dry. After the incubation period 200 µl of transformed cells were spread on the LB/ ampicillin/IPTG/X-gal plate and allowed to dry.

The plates were then sealed and incubated at 37^0C for 16 hrs for colony development. Competence was confirmed by development of blue colonies.

3.6.2 Transformation of Phytoene desaturase (*Crt I*) gene

A ligation mixture was prepared as shown in table 3

Table 3.3 Ligation reaction mixture

Item	Quantity
Crt I DNA insert	4µl
Double distiled water	3.5µl
T4 buffer 10	1µl
T4 ligase	0.5µ
PGEM-T vector	1µl
Total volume	10µl

Transformation was done according to the protocol in section 3.6.1. The success of transformation was confirmed by development of blue and white colonies.

3.6.3 Blue/White colonies' Broth preparation

The white colony was inoculated in 100ml LB treated with 100µl ampicillin and so was the blue colony. The two broths were placed in an incubator at 37^0 C for 16hours and growth was determined by development of turbidity and the cultures were used for plasmid extraction.

3.7 Plasmid extraction

1ml of bacteria culture, blue (control) and white (test) colony was pipette into a sterilized 1.5ml eppendorf tubes and centrifuged at 17091.984 g for 5 minutes. The supernatant was discarded and the pellet resuspended in 100µl of solution 1(50mM glucose, 2.5mM TrisHCl, 10mM EDTA). 200 µl of solution 2(0.2M NaOH and 1% SDS) was added to the two tubes and inverted gently several times to mix the solution. 150µl of solution 3 sodium acetate (3M NaOAc) was added to the two tubes and the tubes inverted repeatedly. The two samples were centrifuged at high speed for five minutes and the supernatant was carefully removed and each sample placed in a sterile eppendorf tube.0.5ml of

phenol/chloroform was added and vortexed for 30 second followed by centrifugation at maximum speed for 3 minutes.

The aqueous upper layer was pippeted into a sterile eppendorf tube and the procedure repeated until there was no material at the interface. 0.5 ml of Chloroform was added to the final aqueous phase and centrifuged briefly to remove the remaining phenol. The aqueous solutions was transferred to a sterile eppendorf tube and 2.5 vol of ice cold Isopropanol was added and placed at -20 for 2 hours to hasten precipitation of the plasmid. The samples were centrifuged at high speed for 10 minute and the supernatant discarded.400 µl of 70% ethanol ice cold was added to the samples and centrifuged at 17091.984 g for 3 minutes the supernatant discarded. The procedure was repeated using 95 % ice cold ethanol and the tube left to air dry for 5 minutes. The pellet was then resuspended with 30µl double distilled nuclease free water and 1µl of RNAse added. A 1% agarose gel was run to determine the presence plasmid DNA.

3.7.1 Amplification of cloned *CrtI* gene

For a 50µl PCR (polymerase chain reaction) reaction, the PCR reagents were mixed in a thin-walled PCR reaction tube (Simport, Canada) according to table (4). The primers, T7 and Sp6 were used in this reaction.

Table 3.4 PCR reagents added to the recombinant plasmid PCR reaction mixture.

Reagent	Volume	Final concentration
Water PCR Grade	36.5µl	
10× PCR buffer	5µl	1×
MgCl2 (25mM)	4µl	2mM
dNTPs (10mM)	1µl	200µM
T_7 (50µM) (Forward)_	1µl	1µM
Sp6 (50µM) (Reverse)	1µl	1µM
plasmid template	1µl	
Taq DNA polymerase (5U/µl)	0.5µl	1.25U
Final volume	50µl	

The reaction was run in a TProfessional thermocycler (Biometra, Germany) according to the cycling conditions in table 3.2 except step 4 which was 2 minutes.

The quality of the PCR products was checked on a 1% agarose gel as described in section 3.3. The products were then purified from the gel before they were sent for sequencing at Segolip (ILRI).

3.8 Phylogenetic Analysis of the sequences

The *CrtI* gene of *Paracoccus bogoriensis* was sequenced and the gene sequence analyzed with standard nucleotide BLAST (Basic Local Alignment Search) search at NCBI (National Center for Biotechnology Information) obtained from; http://www.ncbi.nlm.nih.gov/BLAST. The deduced sequences were matched with known *CrtI* genes at NCBI using the BLAST algorithm. All the sequences, including the one retrieved from the database were then aligned with MUSCLE (Edgar, 2004) and phylogenetic trees constructed based on the nucleotide sequences with the Bayesian phylogenetic method which uses Monte Carlo Markov Chain Model in building phylogenetic trees. MrBayes software was obtained at http://mrbayes.net. The trees were then visualized using fig tree software obtained at http://tree.bio.ed.ac.uk. Similar process was followed in the construction of 16srDNA with bacteria that produce carotenoids.

3.9 Protein Modelling

The DNA sequence of the *CrtI* gene was translated into protein by using of Emboss Transeq at http://www.ebi.ac.uk/tool/st/emboss-transeq which translates nucleic acid into their corresponding peptide sequences using the standard universal genetic code. The sequence of *CrtI* gene from *Paracoccus bogoriensis*, *Paracoccus* sp. N81106, *Bradyrhizobium* sp. ORS278, *Xanthobacter autrophicus* Py2 and *Haematococcus pluvialis* which is an algae that produce astanthanthin were used for the protein modeling. The translated protein sequences were modelled using I-Tasser. It is a hierarchical protein structure modeling approach based on the secondary-structure enhanced Profile-Profile threading Alignment (PPA) and the iterative implementation of the Threading ASSEmbly Refinement (TASSER) program. I-TASSER uses combinatorial approach, employing all three conventional methods for structure modelling; comparative modelling, threading and *ab intio* modelling (Roy *et.al.*, 2010). The 3D models were built based on multiple threading alignments by LOMET (Local Meta Threading Server) which generate 3D models by collecting high scoring targets to template from 10 locally installed threading

programs and iterative template fragment assembly simulations. The target sequence is first threaded through a representative Protein Data Bank (PDB) library by LOMETS, a locally installed meta-threading program (Wu & Zhang, 2007). The matching fragments were then excised from LOMET alignments and used to reassemble the structure by replica-exchange Monte Carlo simulations (Wu & Zhang, 2007).

The simulation trajectories were then clustered by SPICKER (Zhang & Skolnick, 2004) and were used as the starting state of the second round I-TASSER assembly simulation. Finally, the structures of the lowest energy are selected, which are then refined by a fragment-guided molecular dynamic procedure, with the purpose of optimizing the hydrogen-binding network and removing steric clashes. The matching of the function insight of the 3D structure derived was done matching the 3D with Biolip Protein Function Database. The structure superimposition was done using TM-score algorithm to calculate the similarity of the two proteins modeled (*Paracoccus bogoriensis* and *Paracoccus* sp. N81106, *Paracoccus* sp. N81106 and *Bradyrhizobium* sp. ORS278, *Paracoccus* sp. N81106 and *Xanthobacter autrophicus, Paracoccus bogoriensis* and *Bradyrhizobium* sp, *Paracoccus bogoriensis* and *Xanthobacter autrophicus*, and *Paracoccus bogoriensis* and *Haematococcus pluvialis*. Topology fold is more sensitive than Root-Mean-Square-Deviation (RMSD).TM-score values were 0,1 based on statistics ,<0.17 corresponding to a random similarity and a TM-score >0.5 generally corresponding s to the same fold in SCOP/CATH. Viewing of the structure was done using RasMol(Roger & Milner-White, 1995; Bernstein, 2000) which is a graphical program for visualization of proteins, nucleic acids and small molecules.

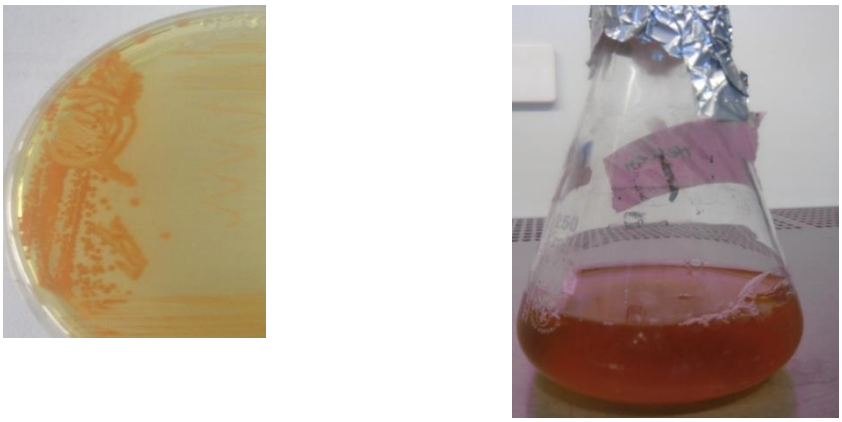
Binding site for the protein was done using COFACTOR at http://www.Zhanglab.ccmb.umich.edu/COFACTOR/. COFACTOR combines both local and global structural comparison algorithms to deduce the biological function of protein starting from the 3D structure that was obtained during protein modeling. Global alignment was used to identify template of similar fold and topology; by matching the query structure with all proteins in the three newly developed representative functional libraries which have known protein ligand binding information; Protein-ligand binding interactions, Enzyme Commission number or Gene Ontology terms (Roy, A. *et al*., 2012). Local functional site identification was done by first constructing a multiple sequence alignment and evolutionary conserved residues in the query sequence, were identified based on their

Jensen Shannon divergence (JSD) score. The confirmations of various triplet residues from conserved residue pool were excised from the query structure to construct a set of local 3D structure motif. Each of the local query motifs was then superimposed onto the known functional residues of the template protein. Translation and rotational matrices was used to refine the local structure match. TM was used to do the alignment and the scoring matrix was repeated repeatedly. The local similarities between query and the template protein were then evaluated using L_{sim} Equation(Schmidt, *et al*., 2011).The bestlocal match was with all the probable set (highest L_{sim}) is selected

CHAPTER FOUR: RESULTS

4.1 Cultivation of *Paracoccus bogoriensis*

Paracoccus bogoriensis was cultured on LB media at pH of 9.2 at 37 ^{0}C for 16 hrs. Colonies on LB agar appeared circular, smooth, convex and orange- red. Figure 4.1(Panel A) shows a culture of *Paracoccus bogoriensis* on LB Agar Plate. The 16 hour culture was used for DNA extraction. The colour developed to orange-red due to accumulation of carotenoids.



A **B**

Figure 4.1: *Paracoccus bogoriensis* colonies grown on Luria Bertani (LB) media. A: Growth of *Paracoccus bogoriensis* colonies grown on Luria Bertani (LB) on Agar. B : Growth of *Paracoccus bogoriensis* 200ml culture grown in Luria Bertani (LB) broth.

4.2 Isolation of genomic DNA

Genomic DNA was isolated from 16 hrs culture using CTAB method. The quality of genomic DNA extracted from the bacteria was analysed on 1% agarose gel (Figure 4.2).

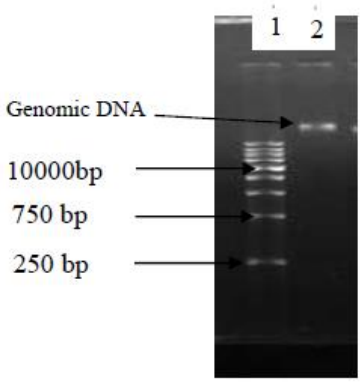


Figure 4.2 Agarose gel analysis of genomic DNA isolated from 16 hour culture of *Paracoccus bogoriensis*. Lane 1: Molecular weight marker 250 – 10000 bp, lane 2: genomic DNA. The concentration of the genomic DNA was approximated to be 40 ng/µl by comparing with the nano reading on the DNA ladder.

4.3 Amplification of Phytoene desaturase (*CrtI*) gene

CrtI gene was amplified using the primer pair Fwd 5'- *CTATGTCTGGCACGATCAGG*-3' and Rev 5'*GCCGACCAGATAGAAGTTGC*-3'. A fragment of approximately 1500 bps Figure 4.3 was amplified and analysed on 1% agarose gel.

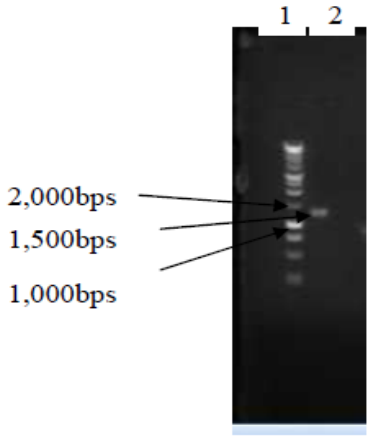


Figure 4.3: agarose gel analysis of amplified PCR product of *CrtI* gene. Lane 1: Molecular weight marker 1kb from Promega and lane 2 *CrtI* gene.

4.4 Assessing the competency of cells

BL21 strain of *Escherichia coli* was used and the cells were prepared according to the method of Christine *et al*., (1997).

Fig 4.4

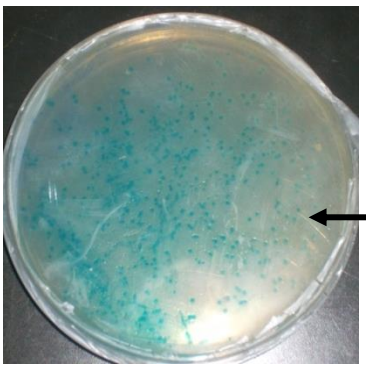


Fig 4.5

Figure 4.4: Liquid Broth of BL21 strain of *Eschericha coli* after 24 hr incubation.

Figure 4.5: Transformed BL21 cells with a puC18 insert. Blue colonies formed on LB/ampicillin/ IPTG/X-gal agar plate. The colour blue indicates presence of intact lac Z gene.

Blue colonies formed onLB/ampicillin/IPTG/X-gal agar plate. The colour blue indicates presence of an intact lack Z gene hence non transformant.

4.6 Cloning of the *CrtI* gene

The amplicon was cloned on PGEM-T easy vector and transformed using BL21 cells from *Escherichia coli* (*E.coli*). Transformed cells were cultured in LB medium containing LB/ ampicillin/IPTG/X-gal. Lac Z gene was used as selection marker of the recombinant cells. Blue colonies indicated recombinant with no insert (Figure 4.6) and white colonies indicated the recombinant cell with an insert (*CrtI* gene) Figure 4.6 C. Cloning efficiency was 88% for the negative control and 44% for the test. Screening for recombinant colonies was depicted by the following reaction.

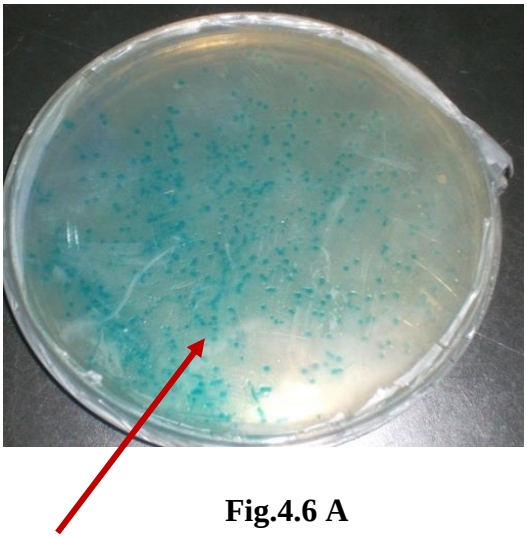


Fig.4.6 A

Blue colony (Control)

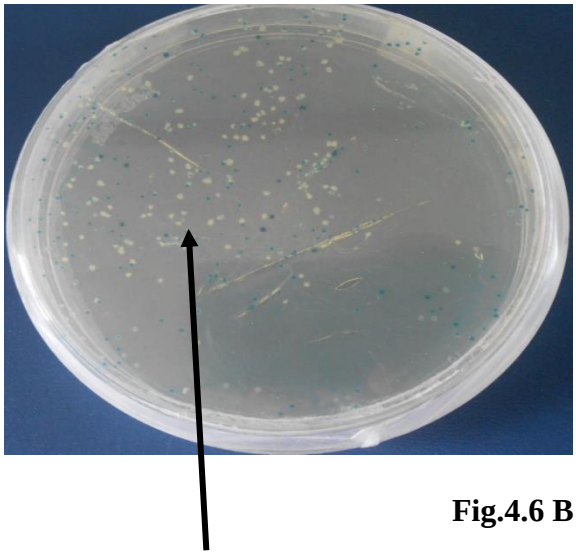


Fig.4.6 B

White colony (with an insert)

Figure 4.6: A 16 hrs culture of transformed BL21 strain of *E.coli*.

A: Transformed cells with no insert, B: Transformed cells with a purified PCR product of *CrtI* gene on ampicillin/agar /IPTG LB plate.

4.7 Plasmid purification pGEM®-T Easy Vector

Plasmid extraction was done according to section 3.7. Analysis of the plasmid was done using 1% agarose gel electrophoresis as (Figure 4.7A). Re-amplification was done using T7 and SP 6 primers to confirm the presence of the insert DNA (Figure 4.7B). The plasmids were then purified using gel purification method (3.5) and the product was sequenced.

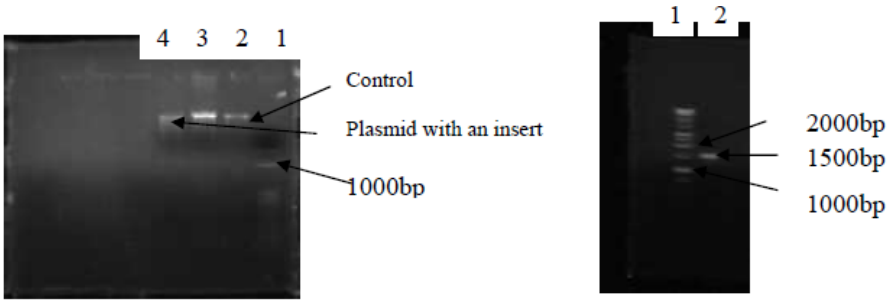


Fig.4.7A Fig.4.7B

Figure 4.7: Agarose gel electrophoresis. A recombinant plasmid with a *CrtI* gene insert. Lane 1: Molecular weight marker is 1kb (USA): Lane 2: Plasmid with no insert (Control) Lane 3: Control, Lane 4: Plasmid with an insert (*CrtI* gene). Figure 4.7B: Agarose gel electrophoresis of re-amplified *CrtI* gene using T7 and SP6 primers. Lane 1: Molecular weight marker is 1kb (USA): Lane 2: *CrtI* gene amplicon

4.8 *CrtI* sequence

Purified plasmid pGEM®-T was sequenced at International Livestock Research Institute (ILRI) using Big Dye termination method. 1222 bp was able to be sequenced and DNA Baser was used for the sequence assembly of the Fwd and Rev Primer sequences (Appendix 2).

4.9 Phylogenetic analysis of phytoene desaturase (*CrtI*)

Phylogenetic tree were constructed using Bayesian Phylogenetic method using MrBayes software and alignment done using MUSCLE. The trees were based on *CrtI* and 16s rDNA gene sequences based on the similarity between the bacterial species production of Carotenoids (Blast sequences) and their respective 16s rDNA. Topography robustness of the tree was evaluated by a bootstrap analysis involving 1000 replications. Phylogenetic tree for *CrtI* and 16s rDNA gene respectively indicating that *CrtI* genes are homologous (Orthologs) in the different type of bacteria with *CrtI* gene from other *Paracoccus* species evolving from *Paracoccus bogoriensis* through speciation event (Figure 4.9A and 4.9B respectively).

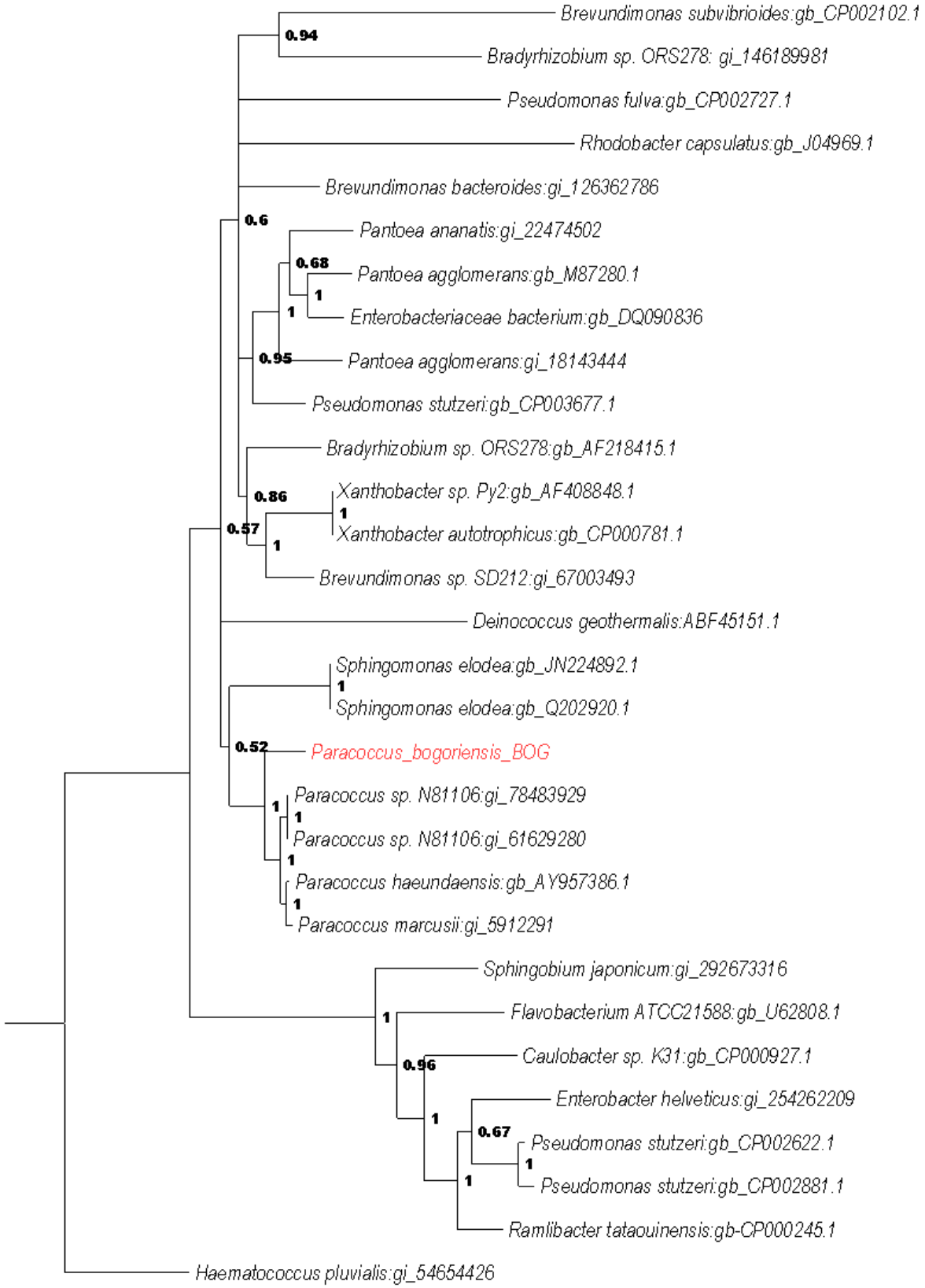


Figure 4.9: A rooted ptree of *Paracoccus bogoriensis CrtI* gene and other related *CrtI* genes from other bacteria's.

The tree was constructed in MrBayes, a program for the Bayesian inference of phylogeny that is based on the Markov Chain Monte Carlo (MCMC) method. Numbers at the nodes show percentage of posterior probabilities indicating topological robustness of the Phylogenetic tree. The tree is rooted by a *CrtI* sequence from *Hematococcus pluvialis*.

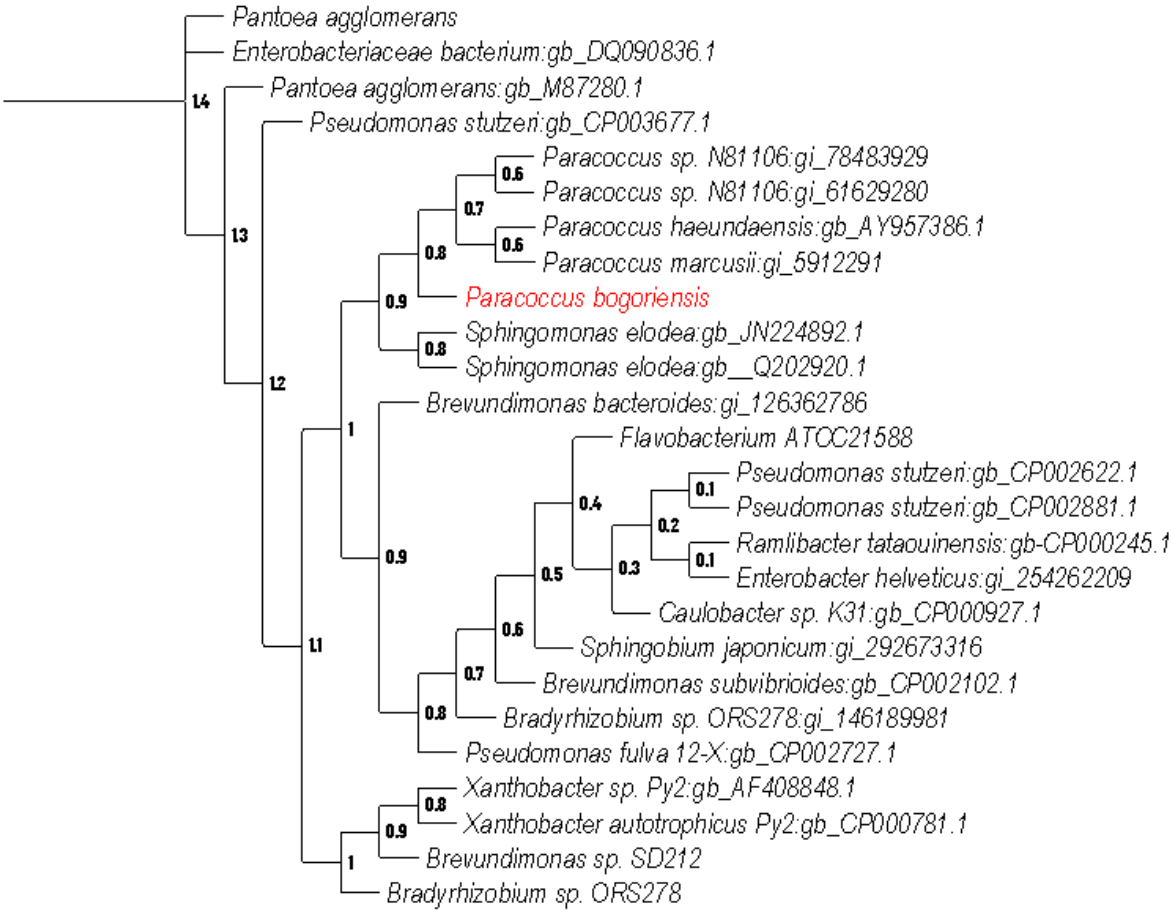


Figure 4.9A: unrooted phylogenetic tree of *Paracoccus bogoriensis CrtI* gene and other *CrtI* genes as obtained from the blast sequences.

The tree was constructed in MrBayes, a program for the Bayesian inference of phylogeny that is based on the Markov Chain Monte Carlo (MCMC) method. Numbers at the nodes show percentage of posterior probabilities indicating topological robustness of the Phylogenetic tree.

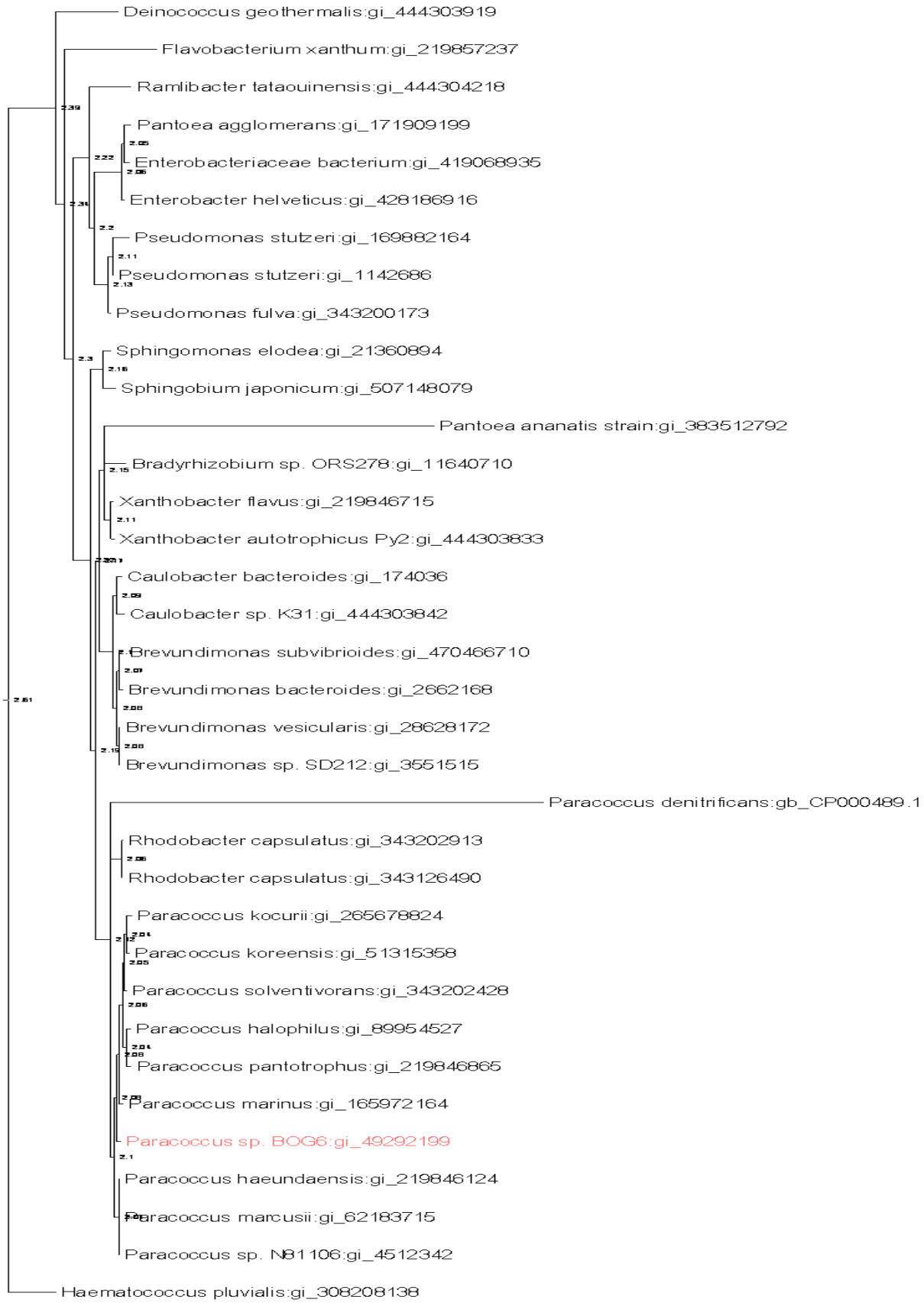


Figure 4.9B: A Phylogenetic tree based on 16s rDNA gene sequences. The tree was constructed in MrBayes, a program for the Bayesian inference of phylogeny that is based on the Markov Chain Monte Carlo (MCMC) method. Numbers at the nodes show

percentage of posterior probabilities indicating topological robustness of the Phylogenetic tree. The tree is rooted by a 16SrDNA sequence from *Hematococcus pluvialis.*

4.10 *CrtI* Sequence analysis

The DNA sequence of the *CrtI* gene was translated into protein by using of Emboss Transeq at http://www.ebi.ac.uk/tool/st/emboss-transeq which translates nucleic acid to their corresponding peptide sequences using the standard universal genetic code. Multiple sequence alignment was done using Jalview (Java Bioinformatics Analysis Web Services version 2.8). The DNA sequence of the *CrtI* gene was translated to protein and alignment of different cluster of organism was done: *Paracoccus* species (Figure 4.10A), different species (Figure 4.10B) species that perform the same desaturation activity but give different end product (Figure 4.10C). The colour scheme emulated for alignment was ClustalX, a graphical interface for clustalW using the default ClustalX. Colouring scheme conserved A, C, F, H, I, L, M,V,W,Y or P residues are coloured blue, while columns with conserved S or T residues are in green. Conserved positively charged residues and negatively charged residues are coloured red and magenta, respectively. White regions with a black character indicate low scoring regions. The numbers following the locus tags indicated the start position of the domain in the complete protein sequence. Asterisks indicated the conserved catalytic residues Ser, Asp and His of these alpha/beta hydrolases.

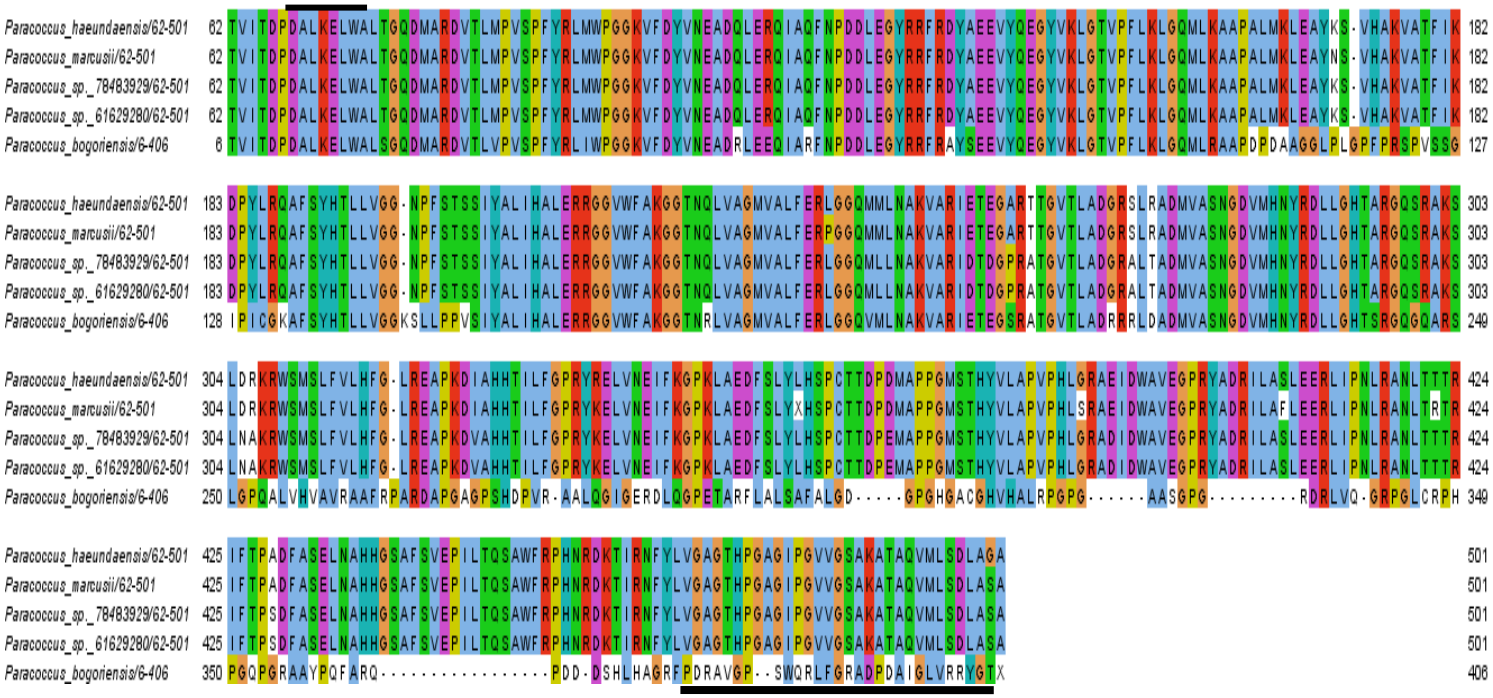


Figure 4.10A: Multiple sequence alignment Phytoene desaturase gene protein sequence from some *Paracoccus species*.

Alignment was done using Jalview 2.8 provided by Java Bioinformatics Analysis Web Services. The binding site of the enzyme is predicted to be at position 62 upto to 300 of the amino acid sequence as given by COFACTOR prediction which is the conserved region across the species. The 80% similarity indicated by the phylogenetic tree also indicate that the other species have evolved thus the differences in the conserved regions as shown in Figure 4.10A. Highly conserved dinucleotide binding motif (βαβ fold) and the consensus signature is underlined.

Figure 4.10B: Multiple sequences alignment of Phytoene desaturase gene, protein sequence from *Paracoccus haeundansis*, *Paracoccus marcusii*, *Paracoccus* sp. gi-78483929, *Paracoccus* sp. gi-61629280, *Pantoea ananatis*, *Enterobacteriaceae bacterium*, *Bradyrhizobium* sp. ORS278, *Paracoccus bogoriensis*, *Deinococcus geothermalis* and *Haematococcus pluvialis*. Alignment was done using Jalview 2.8

provided by Java Bioinformatics Analysis Web Services. The binding for the enzyme is predicted to be at position 62 upto to 300 of the amino acid sequence as given by COFACTOR prediction which is the conserved region across the species including 330-368 and 445 to 510 for *Hematococcus pluvialis.* Highly conserved dinucleotide binding motif (βαβ fold) and the consensus signature is underlined.

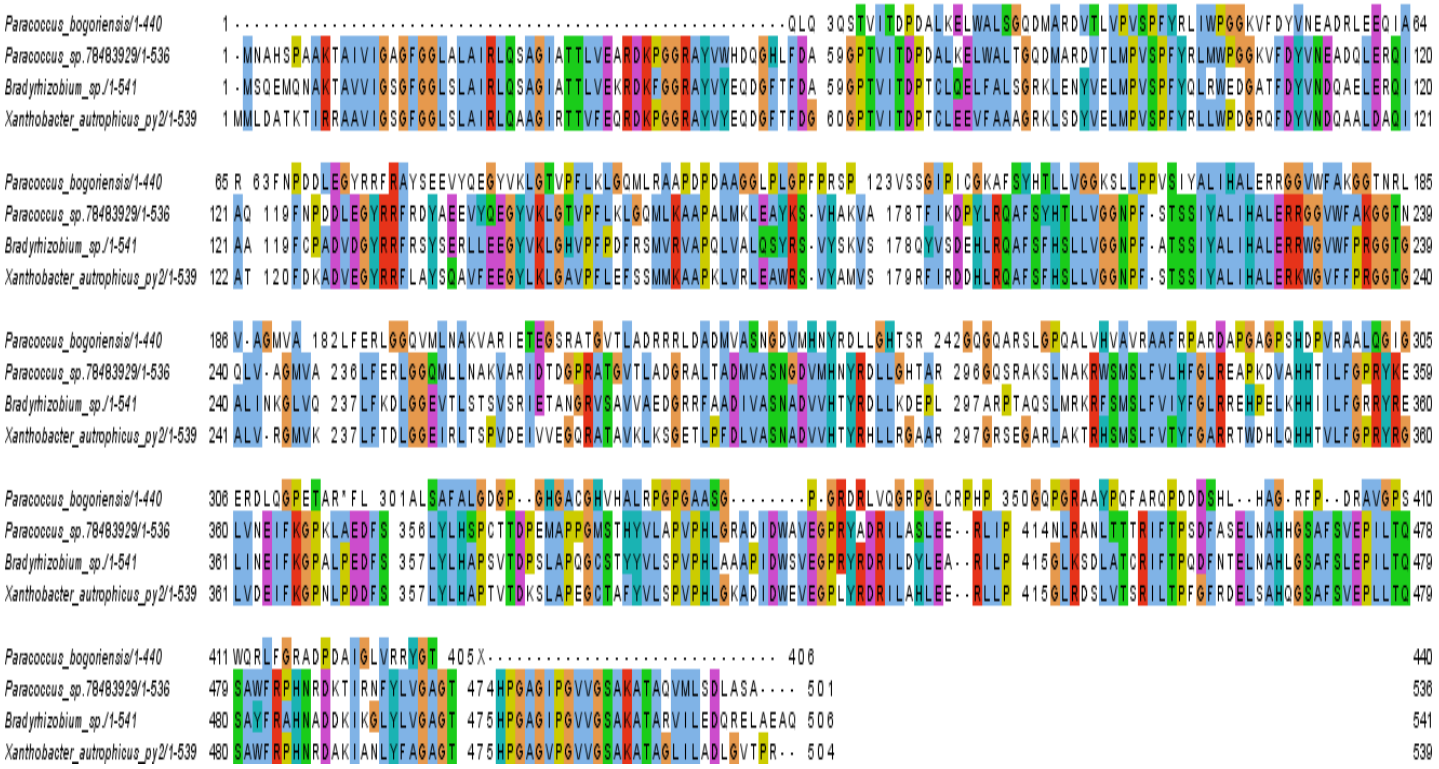


Figure 4.10C: Multiple sequences alignment of Phytoene desaturase gene, protein sequence from *Paracoccus bogoriensis, Paracoccus* sp. gi-78483929, *Paracoccus* sp. gi-61629280, *Bradyrhizobium* sp. ORS278, and *Xanthobacter autotrophicus*. Alignment was done using Jalview 2.8 provided by Java Bioinformatics Analysis Web Services. The binding for this gene is predicted to be at position 62 upto to 300 of the amino acid sequence as given by COFACTOR prediction which is the conserved region across the species though *Paracoccus bogoriensis* shows slight difference with the other species.

4.11 Protein Modeling

The sequence of *CrtI* gene from *Paracoccus bogoriensis*, *Paracoccus* sp. N81106, *Bradyrhizobium* sp. ORS278, *Xanthobacter autrophicus* Py2 and *Haematococcus pluvialis* which is an algae that produce astanthanthin were used for protein modeling. The translated protein sequences were modelled using I-Tasser. Viewing of the structure was done using RasMol (Roger & Milner-White, 1995; Bernstein, 2000) which is a graphical program for visualization of proteins, nucleic acids and small molecules.

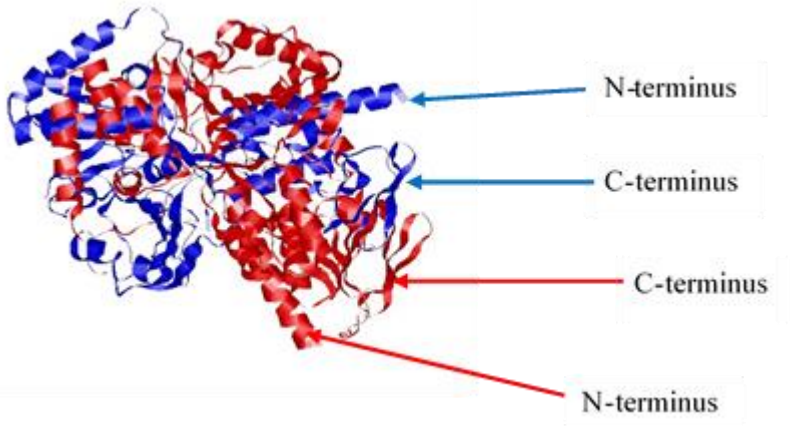


Figure 4.11A: 3D structure of *Paracoccus* sp. N81106 and *Paracoccus bogoriensis CrtI* gene modeled using I-Tasser. Protein code 4DGK crystal of Phytoene desaturase (*CrtI*) from *Pantoea annatis* and Protein code 3KA7A crystal structure of an oxireductase from *Methanosarcina mazei* were used as the templates for modeling.

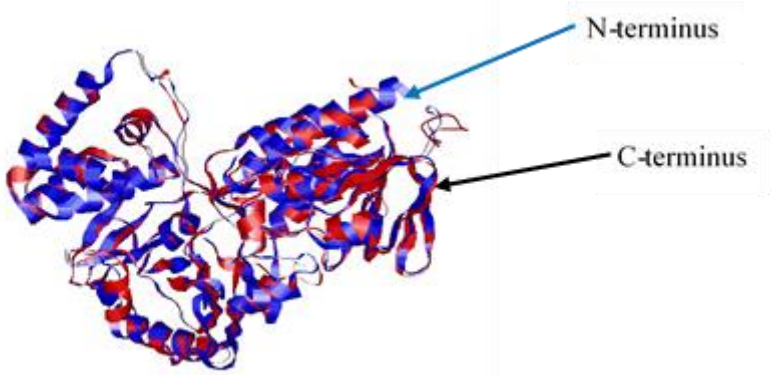


Figure 4.11B: 3D structure of *Paracoccus* sp. N81106 and *Xanthobacter autrophicus CrtI* gene modeled using I-Tasser. Protein code 4DGK crystal of Phytoenedesaturase (*CrtI*) from *Pantoea annatis* and Protein code 3KA7A crystal structure of an oxireductase from *Methanosarcina mazei* were used as the templates for modeling.

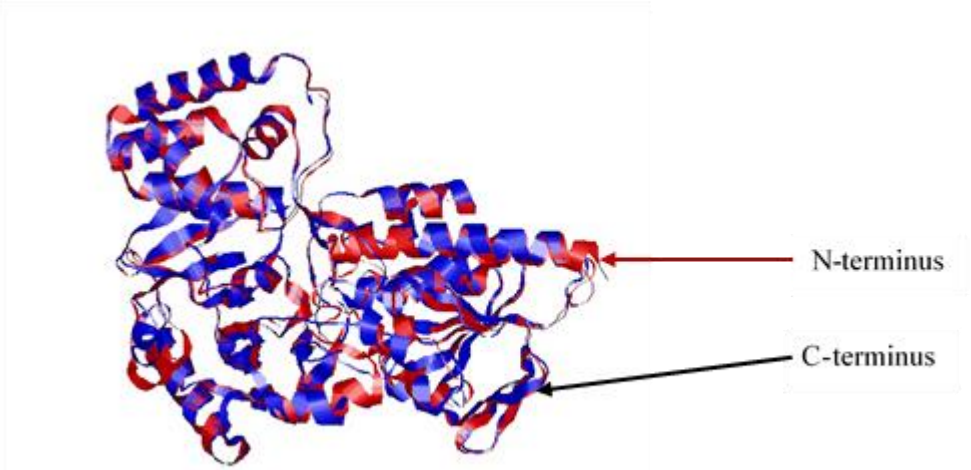


Figure 4.11C: 3D structure of *Paracoccus* sp. N81106 and *Bradyrhizobium* sp. ORS278 *CrtI* gene modeled using I-Tasser. Protein code 4DGK crystal of Phytoene desaturase (*CrtI*) from *Pantoea annatis* and Protein code 3KA7A crystal structure of an oxireductase from *Methanosarcina mazei* were used as the templates for modeling.

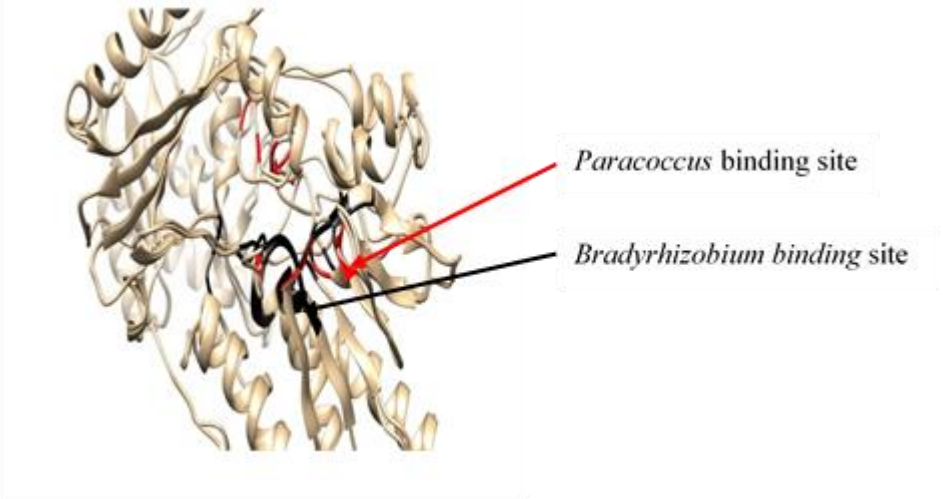


Figure 4.11D: 3D structure *Paracoccus* sp. N81106 and *Bradyrhizobium* sp. ORS278 *CrtI* gene (Figure 4.11C) binding site using COFACTOR prediction method. The binding site are at position 53,-190 of the following template 1VG0A, 1UKVG, 1ITXT, 2IVDB,1F8SG. Chain A coloured red for *Paracoccus* sp. N81106 binding site and Chain B coloured black for *Bradyrhizobium* sp. ORS278 binging site.

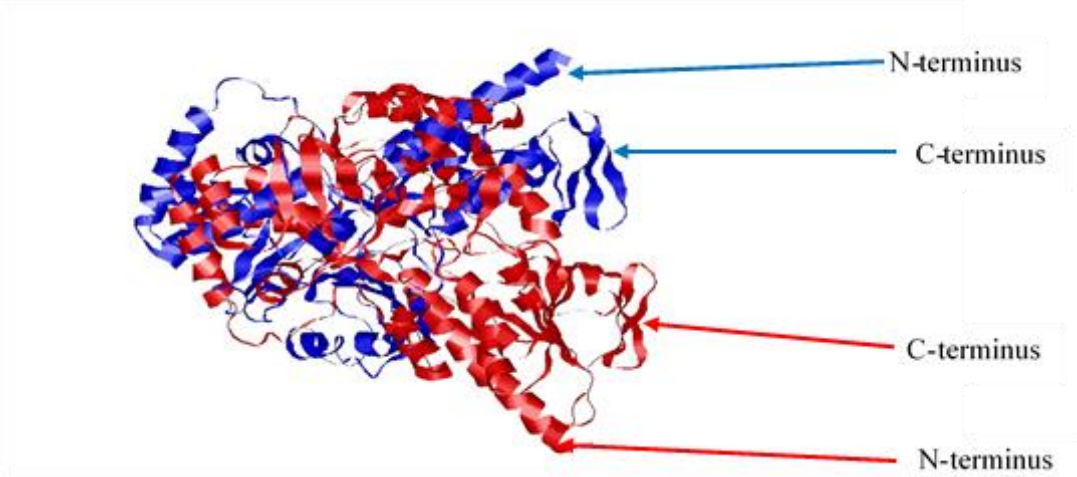


Figure 4.11E: 3D structure of *Paracoccus bogoriensis* and *Bradyrhizobium* sp. ORS278 *CrtI* gene modeled using I-Tasser. Protein code 4DGK crystal of Phytoene desaturase (*CrtI*) from *Pantoea annatis* and Protein code 3KA7A crystal structure of an oxireductase from *Methanosarcina mazei* were used as the templates for modeling.

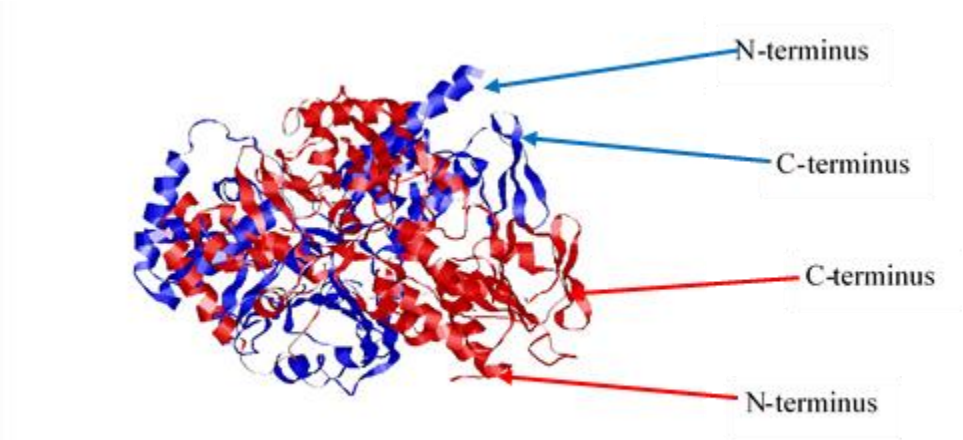


Figure 4.11F: 3D structure of *Paracoccus bogoriensis* and *Xanthobacter autrophicus* *CrtI* gene modeled using I-Tasser. Protein code 4DGK crystal of Phytoene desaturase (*CrtI*) from *Pantoea annatis* and Protein code 3KA7A crystal structure of an oxireductase from *Methanosarcina mazei* were used as the templates for modeling

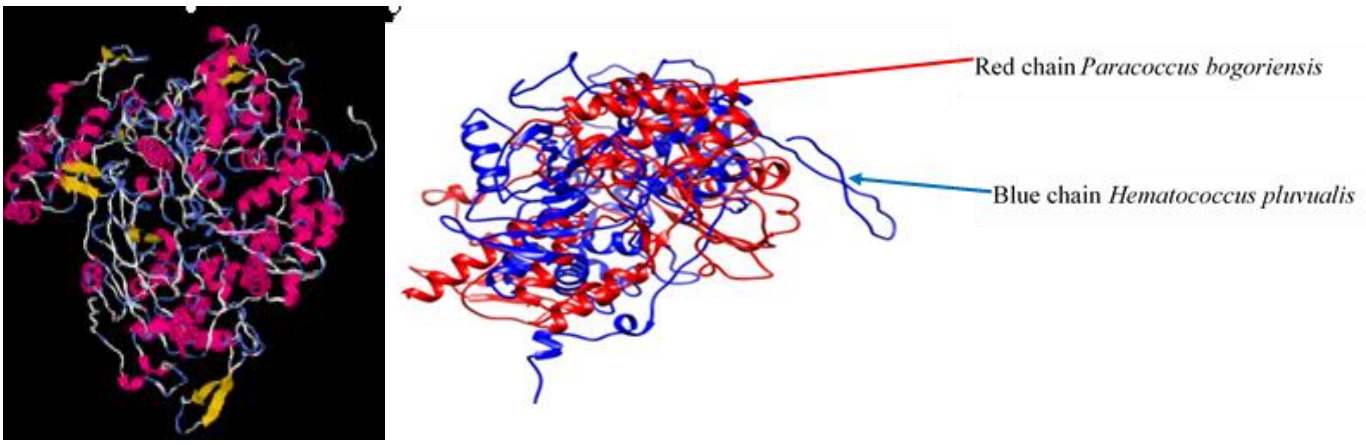


Figure 4.11G(i): 3D structure *Haematococcus pluvialis CrtI* gene modeled using I-Tasser. Protein code 2VVMA crystal of *Aspergillus niger* Monoamine oxidase, human Monoamine oxidase IGOSA crystal and histone demethylase LSD2/KDM1b crystal 4HSUA were used as the templates for modeling. Structure superimposition was done using TM-score algorithm (Figure 4.11G (ii)) that calculated the structural similarities of two protein models *Paracoccus bogoriensis* and *Haematococcus pluvialis*.

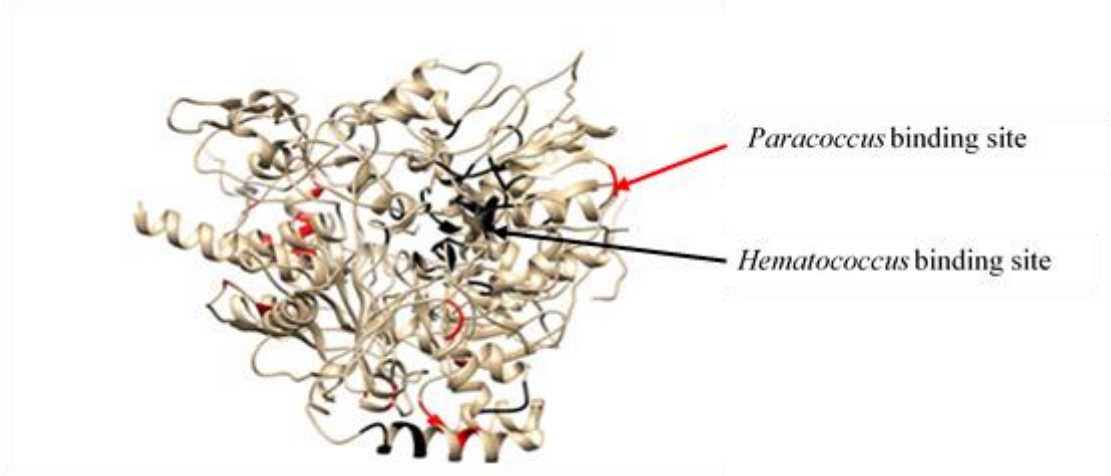


Figure 4.11H: 3D structure of *Haematococcus pluvialis* and *Paracoccus bogoriensis CrtI* gene (Figure 4.11G (ii)) with binding site modelled using COFACTOR prediction method. The binding sites are at position 78, 79, 81, 82, 83, 102, 103, 104, 109, 110, 125, 126, 127, 128, 308, 309, 339, 340, 341, 348, 461, 466, 470, 471, 496, 497, 505, 506, 507, 510 for 2WLA template and 128, 247, 276, 277, 372, 419 for 2WMA template for *Haematococcus pluvialis*. *Paracoccus bogoriensis* binding site are 53, 54, 72, 73, 135, 160, 163 for 1VG0A template, *Paracoccus bogoriensis* 1UKVG template,54, 58, 59, 66, 67, 69, 70, 73, 74 for 1ITXT template,180, 181, 183, 184, 187, 188, 190 for 2IVDB template and 158, 159, 163 for 1F8SG template. Chain A coloured red for *Paracoccus bogoriensis* and chain B coloured black for *Haematococcus pluvialis* binding sites.

CHAPTER FIVE: DISCUSSION

Phytoene desaturase from *P.bogoriensis* was estimated to be 1500bps (Figure 7). A fragment of 1221 was sequenced using big dye terminator method. The size of the 1500bps (Figure 4.3) obtained after PCR is in correlation with the other sequenced *CrtI* from other Paracoccus species e.g *Paracoccus* strain N81106 which has 1506bp (Misawa, et al., 1995). *Deinococcus radiodurans* and *Erwinia uredovora, CrtI* complete genome is 1647bp and 1489bp respectively according to Zhenjian, *et al*., (2007). *Rhodoacter capsulatus CrtI* was estimated to be 57KDa (1539bp) which has a 3 step desaturation activity with neurosporene as the end product (Axel, *et al*., 1996), unlike *CrtI* from *Paracoccus* species which has an average size of 1520bp.

The *CrtI* from *P. bogoriensis* clustered with other species of *Paracoccus*. The hypothesis that *CrtI* come from a common ancestor is confirmed by the phylogenetic analysis done using Bayesian method which uses Markov Chain Monte Carlo. *Gleobacter violaceus* (*G. violaceus* isolated from Switzerland) *CrtI* was the first *CrtI* to be identified as the first oxygenic photosynthetic organism using bacterial type *CrtI*. Tohru T. *et al.*, 2005 obeservation from evolution analysis hypothesised that following evolutionary scheme; ancestral cyanobacteria carried *CrtI* then other *CrtI* (*CrtP*, *CrtQ*) were acquired later in life by replacement of *CrtI* with other desaturases. This confirmed that all *CrtI* are derived from a common ancestor in Figure 4.9A phylogenetic analysis. The difference between *Paracoccus* species *CrtI* can be deduced from the fact that some species might be having other form of desaturases acquired through time by convergent evolution as shown in the tree. This is supported by studies done by Zhenjian, *et al*., 2007 on *Deinococcus radiodurans* which is a red-pigmented, non-photosynthetic bacterium which was seen to have the highest sequence identity to *G. violaceus*. Dinucleotide binding motif found in some Phytoene desaturases of fungi and bacteria were present in *D.radiodurans* which indicated that Phytoene desaturases come from a common ancestor.

DNA sequence from *P.bogoriensis* and other catotenoid producing *CrtI* was translated to protein using Emboss-transeq universal genetic code. *Paracoccus* species showed 100% sequence similarity apart from *P. bogoriensis* which showed 80% similarity in figure 4.10A. The first few amino of the N-terminal were missing in *P. bogoriensis* which can be

attributed to sequencing errors as mutation found in this conserved regions have been shown to cause destruction of *Rhodobacter capsulatus CrtI* (Albert, *et al*., 1996; Amstrong, *et al*., 1990). All *CrtI* have conserved region Figure 4.10B (underlined). Multiple sequence alignment of protein done with different *CrtI* from 13 different Genus of bacteria indicated two highly conserved regions as putative dinucleotide binding motif βαβ fold in the N-terminal and the bacterial phytoene desaturases signature at the C terminus (Zhenjian, *et al*., 2007). This is confirmed in Figure 4.10B that show the two regions with *Deinococcus, Hematococcus* and *Paracoccus bogoriensis* showing slight amino acid variation. Though the amino acids differ they possess the same charges e.g alanine and leucine are both non polar and neutral amino acids. Studies done on the N-terminal site indicated that the site is used as cofactor binding site for *CrtI* either FAD or NAD(P) as coenzyme in the reaction. Bacteria from different genus use either of the two e.g *Erwina* utilize FAD whereas *Synechococcus* is dependent on either NAD or NADP (Axel, *et al*., 1996). This site has been evolutionarily preserved in both photosynthetic and non-photosynthetic bacteria as putative binding site for the cofactors.

Protein modelling was done on distant related organism, *Xanthobacter autrophicus* and *Bradyrhizobium* sp. ORS278 with *Paracoccus* species, *Paracoccus* N81106 and *Paracoccus bogoriensis* (4.10C). All of the above organisms have a four step desaturation activity with their end product being zeaxanthin, canthaxanthin, adonixanthin and astaxanthin respectivly. *Paracoccus bogoriensis CrtI* had 40% sequence similarity with only the predicted N and C terminus having consensus sequences. The sequences were modelled using I-Tasser and structure superimposition done using TM-score algorithm. The structure was viewed using Rasmol a graphical interphase for viewing protein, nucleic acid and small molecules. This was done to investigate whether the protein structure was similar for the 4 species of bacteria as they all perform four desaturation steps with the end product of the *CrtI* being lycopene. Other group of genes are involved in the downstream production of the various carotenoid after synthesis of β-carotene by lycopene cyclase from lycopene shown in the below (Figure 5.4).

Paracoccus bogoriensis CrtI protein structure was compared with the protein structure of *Paracoccus* N81106 (Figure 4.11A). The two protein structures were not similar suggesting the genes are homologous (orthologs) perform similar functions but the subsequent product after the desaturation process are different hence different Carotenoid

products. This is confirmed by the protein alignment and phylogenetic analysis done on the *CrtI* genes. The two bacteria species use the same pathway as elucidated by Gregory 1997 which confirms the evolution theory that some of the bacteria might have acquired other form of desaturases *Crt* P/*CrtH*/ *CrtZ* form other organism but perform the same function as the original form found in cyanobacteria. Figure 4.11B shows phytoene desaturase protein structure for *Paracoccus* sp. N81106 and *Xanthobacter autrophicus* which has 95% similarity apart from the α-helices. The 2 *CrtI* genes are known to desaturate their substrate with lycopene being the end product. Similar result is achieved with *Paracoccus* sp N81106 and *Bradyrhizobium* sp (Figure 4.11C). *CrtI* genes from the 3 genera of bacteria are phylogeneticaly distant related though having come from a common ancestor suggesting that the three species have homologous protein sequences with similar functions but different end products.

The protein structure derived from the genus indicated similar 3D conformation which confirms the theory that DNA sequence alone cannot be used exclusively in elucidating the changes at the protein level. This is due to degeneracy of the codon usage in different species of bacteria and having two different amino acids that have the same charge thus perform the same function. To confirm if the 3 genes have the same binding site the two superimposed structure of *Paracoccus* sp. N81106 and *Bradyrhizobium* sp. were used (Figure 4.11C). The binding site were almost similar for the β plated sheet as confirmed by the protein alignment and the 3D structure but slightly different on the α- helices sheets (Figure 4.11D)

Similar test was carried out on *Paracoccus bogoriensis* and *Bradyrhizobium* and *Xanthobacter* genus (Figure 4.11E and 4.11F). The two genuses (*Bradyrhizobium* and *Xanthobacter*) differ significantly with *P.bogoriensis*. The 2 genera differ significantly too at protein alignment level (Figure 4.10C) with *P.bogoriensis* having different amino acid sequences thus different conformation as compared with *Bradyrhizobium* and *Xanthobacter* which though slightly different amino acid sequence but similar function.

Hematococcus pluvialis belongs to green algae and is known for its ability to produce astaxanthin (3,3'-dihydroxy-β β- catotene-4,4'dione) in its encysted cells (Johnson & Schroeder, 1995). *P. bogoriensis* also produce astaxanthin as the major carotenoid (0.4mg/g of cells) (Osanjo, et al., 2009).The two *CrtI* perform 4 desaturation steps and the

end product is astaxanthin. The test was carried out to investigate whether the 2 *CrtI* are similar as the end product is the same. Figure 4.11G (ii) indicated the two genes used by the two organism are different which is expected as *Hematococcus* is an algae. The two have some few conserved region unlike the other bacteria *CrtI* as shown below (Figure 5.5).

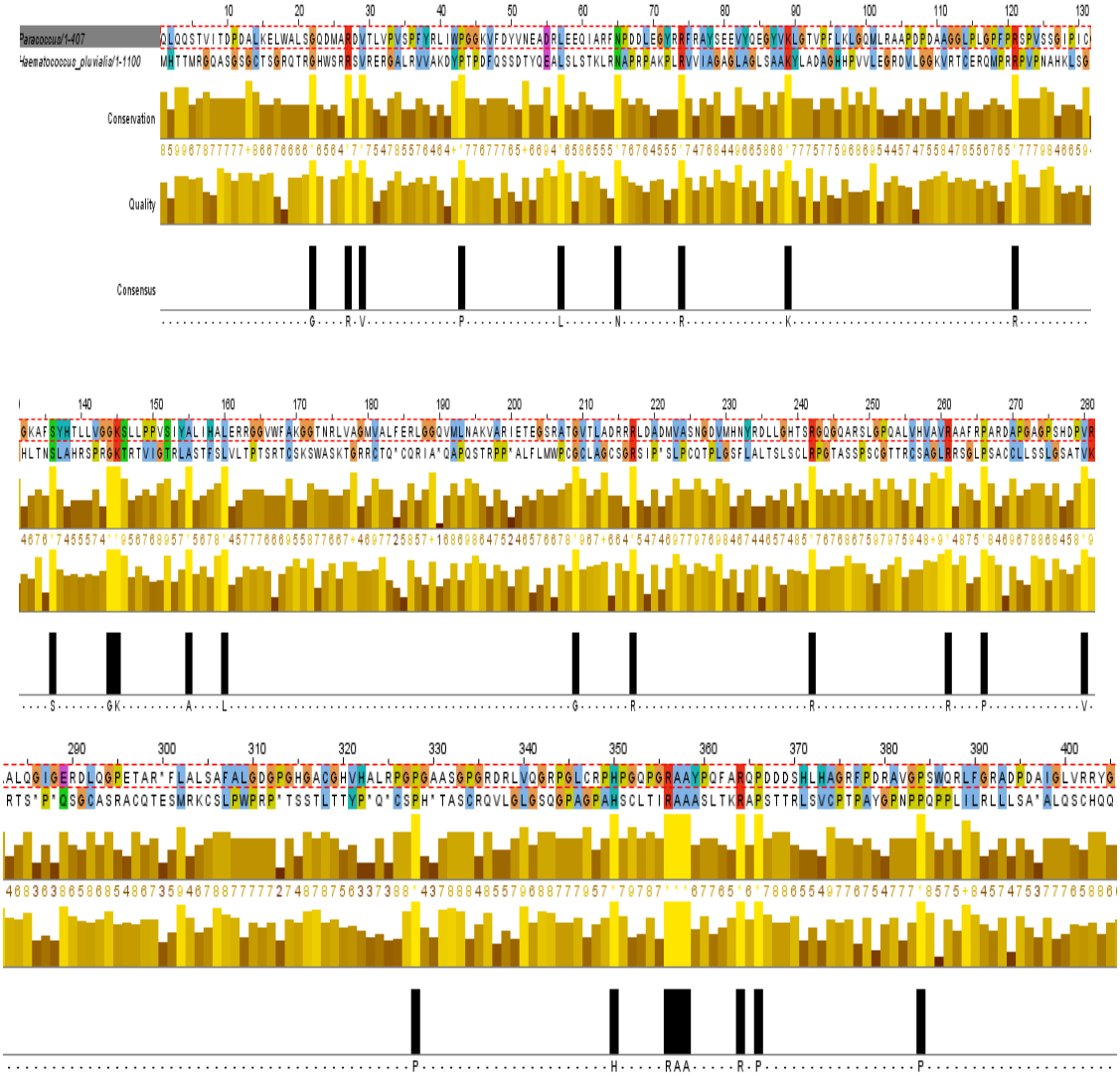


Figure 5.5 Protein alignments of *Paracoccus bogoriensis* and *Hematococcus pluvialis*. Black bars indicate the consensus sequence in the 2 phytoene desaturase gene. Both Micro-organisms end products are astaxanthin. Alignment done using Jalview 2.8 and colour by ClustalW.

To demonstrate whether the two have the same binding sites, the two structures were superimposed on each other and the binding site illustrated for *H.pluvialis* and Red *P. bogoriensis* (Figure 4.11H). The binding sites for the two enzymes were different which is confirmed by the protein alignment and the 3D structure. Studies done on desaturases

indicated that there are two types of phytoene desaturases which can be divided structurally and functionally, into two distinct groups: *CrtI* and *CrtP* type. They are both homologue to Al-I in *Neurospora crassa* and Phytoene desaturases in algae and higher plants. The two enzymes are thought to have evolved independently (Pecker L. *et al*., 1992) and differ with respect to their specificity for substrate and product. The *CrtI* type synthesizes neurosporene or lycopene from phytoene but cannot accept zeta-carotene as a substrate while as the *CrtP* make zeta-Carotene from phytoene (Pecker I. *et al*., 1992). The difference between the two is attributed to type of enzymes involved and desaturation intermediates used in the formation of lycopene. *H. pluvialis* use the *CrtP* type of phytoene desaturase as confirmed by Pecker L. *et al*.,1992 while *P. bogoriensis* uses the *CrtI* type.

CHAPTER SIX: CONCLUSION AND RECOMMENDATION

6.1 Conclusion

Phytoene desaturase from *P.bogoriensis* was estimated to be 1500bps. The size of the 1500bps obtained after PCR is in correlation with the other sequenced *CrtI* from other *Paracoccus* species e.g *Paracoccus* strain N81106 which has 1506bp. The *CrtI* from *P. bogoriensis* clustered with other species of *Paracoccus* confirming the hypothesis that *CrtI* come from a common ancestor confirmed by the phylogenetic analysis done using Bayesian method. DNA sequence from *P.bogoriensis* and other carotenoid producing *CrtI* was translated to protein using Emboss-transeq universal genetic code. *Paracoccus* species showed 100% sequence similarity apart from *P. bogoriensis* which shows 80% similarity. Multiple sequence alignment of protein done with different *CrtI* from 13 different Genera of bacteria indicated two highly conserved regions as putative dinucleotide binding motif βαβ fold in the N-terminal and the bacterial phytoene desaturases signature at the C terminus. Similarly *P. bogoriensis* indicated similar results apart from the C-terminal where amino acid sequence differ but perform the same function. The N and the C-terminal have been evolutionarily preserved in both photosynthetic and nonphotosynthetic bacteria as putative binding site for the cofactors NAD, NAD (P) or FAD. Protein modelling done using I-TASSER showed no similarity of the 3D structure of *Paracoccus* N81106, *Bradyrhizobium* sp. and *Xanthobacter autrophicus* with *P. bogoriensis*. The four species *CrtI* perform four desaturation with *Paracoccus* N81106, *Bradyrhizobium* sp. and *Xanthobacter autrophicus* having similar structure and similar binding sites. A comparison done with the 3D structure of *Hematococcus pluvialis* which produce astaxanthin as its end product indicated that the *CrtI* gene used are different. *Hematococcus pluvialis* use a different *CrtI* (*CrtP* type) than *P. bogoriensis* as confirmed by the protein alignment and the 3D structure.

6.2 Recommendation

Sequencing of the other gene cluster from *Paracoccus bogoriensis* involved in the biosynthesis of astaxanthin will shed more insight in the genes involved and also future prospect of genetically engineering of the pathway either in plants or micro-organism in the production astaxanthin. *P.bogoriensis* offers a great potential in exploration of its carotenoid gene cluster for the production of this high value pigment as few micro-organism are known to produce it naturally

.

REFERENCES

Alberti, M., Burke, H. D., & Hearst, E. J. (1995). Structure and sequence of the photosynthesis gene cluster. *Advances in Photosynthesis and Respiration , 2*, 1083-1106.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res. , 25*, 3389–3402.

Ambrish, R., Alper, K., & Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocol , 5* (4), 725.

Amstrong, G. A., & Hearst, J. E. (1996). Carotenoids 2: genetics and molecular biology of carotenoid pigment biosynthesis. *FASEB J , 10* (228–237).

Amstrong, G. A., Alberti, M., & Hearst, J. E. (1990). Conserved enzymes mediate the early reactions of carotenoid biosynthesis in nonphotosynthetic and photosynthetic prokaryotes. *Proc. Natl.Acad. Sci. , 87*, 9975-9979.

Arakaki, A. K., Zhang, Y., & Scolinck, J. (2004). Large-scale assessment of the utility of low-resolution protein structure for biochemical function assignment. *Bioinformatics , 20*, 1087-1096.

Arrach, N., Shmidhauser, T. J., & Awalos, J. (2002). Mutants of the carotene cyclase domain of al-2 from Neurospora crassa. *Mol. Genet. Genomics. , 266*, 914-921.

Ashburner, M,. Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T. (2000). Gene ontology: tool for unification of biology. The Gene Ontology Consortium. *Nat Genet., 25*, 25-29

Axel, R., Glenn, B., Pablo, S., & Gerhard, S. (1996). Purification in an Active State and Properties of the 3-Step Phytoene Purification in an Active State and Properties of the 3-Step Phytoene Escherichia coli. *J. Biochem , 119*, 559-564.

Battey, J. N., Kopp J., Bordoli, L., Read R. J., Clarke, N. D., Schwede, T., (2007). Automated server predictions in CASP7. *Proteins , 69*, 68–82.

Barret, A. J. (1997). Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB). Enzyme Nomenclature. *Eur. J. Biochem. , 250*, 1-6.

Bartely, G. E., Schmidhauser, T. J., Yanofsky, C., & Scolnik, P. A. (1990). Carotenoid desaturases from Rhodobacter capsulatus and Neurospora crassa are structurally and functionally conserved and contain domains homologous to flavoprotein disulfide oxidoreductases. *J. Biol. Chem. , 265*, 16020-16024.

Bartley, G. E., & Scolnik, P. A. (1989). Carotenoid biosynthesis in microorganisms and plants. *J. Biol. Chem , 264*, 13109–13113.

Bartley, G. E., Viitanen, P. V., Pecker, I., Chamovitz, D., Hirschberg, J., & Scolnik, P. A. (1991). Molecular cloning and expression in photosynthetic bacteria of a soybean cDNA coding for phytoene desaturase, an enzyme of the carotenoid biosynthesis pathway. *Proc. Natl. Acad. Sci. , 88*, 6532-6536.

Bax, A. (1989). Two-dimensional NMR and protein structures. *Annu. Rev. Biochem. 58*, 223-256.

Becker, O. M. (2006). An integrated in silico 3D model-driven discovery of a novel, potent, and selective amidosulfonamide 5-HT1A agonist PRX-00023) for the treatment of anxiety and depression. *J. Med. Chem. , 49*, 3116-3135.

Bernstein, H. J. (2000). Recent changes to RasMol, recombining the variants. *TIBS, 25* (*9*), 453-455.

Blundell, T. L., Sibanda, B. L., Sternberg, M. J. E. & Thornton, J. M. (1987). Knowledge-based prediction of the protein structures and the design of novel molecules. *Nature* (*London*), *326*, 347-352.

Blundell, T. l., & Johnson, L. N. (1976). *Protein Crystallography*, Academic Press, New York.

Boeger, P., & Sandmann, G. (1998). Carotenoid biosynthesis inhibitor herbicides mode of action and resistance mechanisms. *Pestic Outlook , 9*, 29-35.

Bohm, F., Edge, R., Land, E. J., McGarvey, D. J., & Turscott, T. G. (1997). Carotenoids enhance vitamin E antioxidant efficiency. *J Am Chem Soc. , 119*, 621-622.

Boussiba, S., & Vonshak, A. (1991). Astaxanthin accumulation in the green alga *Haematococcus pluvialis*. *Plant Cell Physiol , 32*, 1077–1082.

Boussiba, S., Fan, I., & Vonshak, A. (1992). Enhancement and determination of astaxanthin accumulation in green alga *Haematococcus pluvialis* method. *Enzymol. , 213*, 386-391.

Boyd, A., & al, e. (2008). A random mutagenesis approach to isolate dominant-negative yeast sec1 mutants reveals a functional role for domain 3a in yeast and mammalian Sec1/Munc18 proteins. *Genetics , 180*, 165-178.

Britton, G. (1998). *Carotenoid Biosynthesis and metabolism* (Vol. 3). (G. Britton, Ed.).

Britton, G. (1983). The biochemistry of natural pigments. *Cambridge University Press, Cambridge* .

Brooks III, C. L., Karplus, M. & Pettit, B. M. (1988). Proteins: *A Theoretical Perspective of Dynamics, Structure and Thermodynamics*. John Wiley & Sons, New York.

Brylinski, M., & Skolnick, J. (2008)**a**. A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. *Proc. Natl. Acad. Sci. , 105*, 129-134.

Brylinski, M., & Skolnick, J. Q.-D. (2008)**b**. Low-resolution flexible ligand docking with pocket-specific threading restraints. *J. Comput. Chem. , 29*, 1574-1588.

Capra, J. A., & Singh, M. (2007). Predicting functionally important residues from sequence conservation. *Bioinformatics , 23*, 1875–1882.

Carothers, D. J., Pons, G., & Patel, M. S. (1989). Crystallization and preliminary X-ray crystallography *Arch. Biochem. Biophys. , 269*, 409-425.

Chamovitz, D. (1993). Ph.D. thesis. The Hebrew University of Jerusalem, Jerusalem, Israel.

Chamovitz, D., Sandmann, G., & Hirschberg, J. (1993). Molecular and biochemical characterization of herbicide-resistant mutants of cyanobacteria reveals that phytoene desaturation is a rate-limiting step in carotenoid biosynthesis. *J Exp Bot , 268*, 17348–17353.

Chew, B. P., Park, J. S., Birkhauser, P. B., Liaanen, J. S., & Pfander, H. (2009). Carotenoids Against Disease: Part C: The Immune System and Disease. Carotenoids: Nutrition and Health. *Carotenoids , 5*, pp. 368-382.

Christine, E., Seidman, K. S., Jen, S., & Timm Jessen. (1997). Introduction of plasmid DNA into cells. *Curr. Prot. Mol. Biol.*1.8.1- 1.8.10.

Claudia, S.-D. (2000,). Engineering novel carotenoids in microorganisms. *Current Opinion in Biotechnology , 11* (255–261).

Clore, G. M. & Gronenborn, A. M. (1991). Two-, three-, and four dimensional NMR methods for obtaining larger and more precise three-dimensional structures of proteins in solution. *Annu. Rev. Biophys. Chem. 20*, 29-63.

Cohen, F. E. & Kuntz, I. D. (1989). Tertiary structure prediction. *In Prediction of Protein Structure and the Principle of Protein Conformation* (Fasman, G. D., ed.), 647-705, Plenum Press, New York.

Cozzetto, D., & et at., (2009). Evaluation of template-based models in CASP8 with standard measure. *Proteins , 77 (Suppl 9)*, 18-28.

Dannert, C. S. (2000). Engineering Novel Carotenoid in Micro-organism. *Current opinion in Biotechnology , 11*, 255-261.

Das, R. E. (2007). Structure prediction for CASP7 targets using extensive all-atom refinement with Rosetta@home. *Proteins , 29*, 118-128.

Demming, A. B., & Adams, W. W. (2002). Antioxidants in photosynthesis and human nutrition. *Science , 298* (3), 2149-2153.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy throughtput. *Nucleic Acids Res. 32* (*52*), 1792-1797.

Ekins, S., Mestres, J., & Testa, B. (2007). In silico pharmacology for drug discovery: applications to targets and beyond. *Br. J. Pharmacol , 152*, 21-37.

E-Siong, T. (1995). The medical importance of vitamin A and carotenoids (with particular reference to developing countries) and their determination. *Mal J Nutr. , 1*, 179-230.

Fassett, R. G., & Coombes, J. S. (2009). Astaxanthin, oxidative stress, inflammation and cardiovascular disease. *Future Cardiology , 5* (4), 333-342.

Ford, N. A., Elsen, A. C., Zuniga, K., Lindshield, B. L., & Erdman, J. W. (2011). Lycopene and apo-12'-lycopenal reduce cell proliferation and alter cell cycle progression in human prostate cancer cells. *Nurt Cancer , 63* (2), 256-263.

Fraser, P. D., & Bramley, P. M. (2004). The biosyntheis and nutrional uses of carotenoid. *Prog. Lipid. Res. , 43*, 228-265.

Fraser, P. D., Misawa, N., Linden, H., Yamano, S., Kobayashi, K., & Sandmann, G. (1992). Expression in *Escherichia coli*, purification and reactivation of the recombinant *Erwinia uredovora* phytoene desaturase. *J. Biol. Chem. , 27*, 19891-19895.

George, B. (1993). *The biochemistry of natural pigments.* Cambridge University Press, Cambridge.

G, S. (1994). Carotenoid biosynthesis in microorganisms and plants. *Eur. J. Biochem. , 223*, 7-24.

Glenn, E. B., Thomas, J. S., Charles, Y., & Pablo, A. S. (1990). Carotenoid Desaturases from *Rhodobacter capsulatus* and *Neurospora crassa* Are Structurally and Functionally Conserved and Contain Domains Homologous to Flavoprotein Disulfide Oxidoreductases. *The J.Biological Chem, 265* (26), 16020-16024.

Goodwin, T. (1980). *The Biochemistry of the Carotenoids* (second ed. ed.). New York: Chapman and Hall, London,.

Gregory, A. A. (1997). Genetics Of Eubacterial Carotenoid Biosynthesis:A Colorful Tale. *Annu. Rev. Microbiol , 51*, 629-659.

Hanukoglu, I., & Gutfinger, T. (1989). *Eur.J.Biochem. , 180*, 479-496.

Hausmann, A., & Sandmann, G. (2000). Neurosporaxanthin Production by Neurospora and Fusarium. *Fung. Genet. Biol , 30*, 147–153.

Hirschberg, J., Cohen, M., Lotan, T., Mann, V., & Pecker, I. (1997). Molecular genetics of the carotenoid biosynthetic pathway in plants ans algae. *Pure Appl Chem. , 69*, 2151-2158.

Hugeueney, P., Romer, S., & Camara, B. (1992). Characterization and molecular cloning of a flavoprotein catalyzing the synthesis of phytofluene and c-carotene in Capsicum chromoplasts. *Eur. J.Biochem. , 209*, 399-407.

Hussein, G., Goto, H., Watanabe, H., Oda, S., Sankawa, U., & Matsumoto, K. (2006). Antihypertensive potential and mechanism of action of astaxanthin: III. Antioxidant and histopathological effects in spontaneously hypertensive rats. *Biol Pharm Bull. , 29* (4), 684-688.

Jauch, R., Yeo, H. C., Kolatkar, P. R., & Clarke, N. D. (2007). Assesment of CASP& structure prediction for template free targets. *Protein , 69*, 57-67.

Johnson, E. A., & Schroeder, W. A. (1996). Microbial carotenoids. *Adv. Biochem. Eng. Biotechnol. , 53*, 119-178.

Johnson, E. A., & Schroeder, W. A. (1995). Microbial carotenoids. *Adv. Biochem and Eng. Biotechnology , 53*, 119-178.

Jones, T. A., Taylor, W. R., & Thornton, J. M. (1999). A new approach to protein fold recognition. *Nat. (London), 358*, 86-89.

Kate, W. (1987). Preparation of genomic DNA from bacteria. *Curr. Prot. Mol. Biol.* 2.4.1-2.4.5

Krinsky, N. I. (1994). The biological properties of carotenoid. *Pure Appl. Chem. , 66*, 1003-1010.

Kristala, L. J., Seon-won, k., & Keasling, J. D. (2000). Low-Copy Plasmids can Perform as Well as or Better Than High-Copy Plasmids for Metabolic Engineering of Bacteria. *Metabolic Engineering , 2* (4), 328-338.

Krubasik, P., Kobayashi, M., & Sandman, G. (2001). Expression and functional analysis of a gene cluster involved in the synthesis of decaprenoxanthin reveals the mechanisms for C-50 carotenoid formation. *Eur. J. Biochem. , 268*, 3702-3708.

Lee, S. J., Bai, S. K., Lee, K. S., Namkoong, S., Na, H. J., Ha, K. S., et al. (2003). Astaxanthin inhibits nitric oxide production and inflammatory gene expression by suppressing I(kappa)B kinase-dependent NF-kappaB activation. *Mol Cells , 16* (1), 97-105.

Linden, H., Misawa, N., Chamovtiz, D., Peker, I., Hirschberg, J., & Sandmann, G. (1991). Functional complementation in *Escherichia coli* of different phytoene desaturase genes and analysis of accumulated carotenes. *Z. Naturforsch , 46,* 1045-1051.

Lopez, A. B., Van, E. J., Conlin, B. J., Paolilo, D. J., O'Neil, J., & Li, L. (2007). Effect of the cauliflower Or transgene on carotenoid accumulation and chromoplast formation in transgenic potato tubers. *J. Exp Bot, 59,*213-223.

Lorenz, R. T., & Cysewski, G. R. (2000). Commercial potential for *Haematococcus* microalgae as a natural source of astaxanthin. *Trends Biotechnol , 18,* 160–167.

Fuji Chemical LTD. (2014). *Fuji double astaxanthin capacity as global demand grows.* BioReal Ab in Gustavberg Sweeden: http/www.nutraingredients.com/industry/fuji-double-astaxanthin-capacity-as-global-demand-grows.

Maoka, T., Fujiwara, T., Hashimoto, K., & Akimoto, N. (2001). Isolation of a series of apocarotenoids from the fruits of the red paprika Capsicum annuum. *L. J. Agric. Food Chem. ,* 1601-1606.

Masomoto, K., Kaneko, T., Takaichi, S., & Wada, H. (2001). Identification of a gene required for cis-to-trans carotene isomerization in carotenogenesis of the cyanobacterium Synechocystis sp. PCC 6803. *Plant Cell Physiol. , 42,* 1398–1402.

Mendez, H., & Britton, G. (2002). Involvement of NADPH in the cyclization reaction of carotenoid biosynthesis. *FEBS Lett. , 515,* 133-136.

Misawa, N., & Shimada, H. (1997). Metabolic Engineering for the production of carotenoid in non-carotenogenic bacteria and yeast. *J. Biotechnologology , 59,* 169-187.

Misawa, N., Masomoto, K., Hori, T., Ohtani, T., Boger, P., & Sandmann, G. (1994). Expression of an *Erwinia* phytoene desaturase gene not only confers multiple resistance to herbicides interfering with carotenoid biosynthesis but also alters xanthophyll metabolism in transgenic plants. *Plant J. , 6,* 481–89.

Misawa, N., Satomi, Y., Kondo, K., Yokagama, A., Kajiwara, S., Saito, T., et al. (1995). Structure and functional analysis of marine bacterial carotenoid biosynthesis gene cluster and astaxanthin biosynthetic pathway proposed at gene level. *J Bacteriol , 177* (22), 6575-6584.

Misawa, N., Yamano, S., Linden, H., de Felipe, M. R., & Lucas, M. (1993). Functional expression of the *Erwinia uredovora* carotenoid biosynthetic gene crtI in transgenic plants showing an increase in carotene biosynthesis activity and resistance to the bleaching herbicide norflurazon. *Plant J. , 4,* 883-840.

Misawa, N., Yamano, S., Linden, H., de Felipe, M. R., Lucas, M., Ikenga, H., et al. (1993). Functional expression of the *Erwinia uredovora* carotenoid biosynthetic gene crtI in transgenic

plants showing an increase in 1-carotene biosynthesis activity and resistance to the bleaching herbicide norflurazon. *Plant J. , 4*, 833-840.

Myers, S. (2011, March 21). Vibrant Carotenoids. *Heart Health, Eye Health and Immune Health.*

Nakagawa, K., Kiko, T., & Hatade, K. (2009). Antioxidant effect of lutein towards phospholipid hydroperoxidation in human erythrocytes. *Br J Nutr. , 102* (9), 1280-1284.

Olson, J. A., & Krinksy, N. I. (1995). Introduction: the colorful, fascinating world of the carotenoids: important physiologic modulators. *The FASEB Journal , 9*, 1547-1550.

Osanjo, G. O., Muthike, E. W., Tsuma, L., Okoth, M. W., Bulimo, W. D., Lünsdorf, H., et al. (2009). A salt lake extremophile, *Paracoccus bogoriensis* sp.nov., efficiently produces xanthophyll carotenoids. *African Journal of Microbiology Research , 3* (8), 426-433.

Palozza, P., & Krinsky, N. I. (1992). β-Carotene and a-tocopherol are synergistic antioxidants. *Arch Biochem Biophys. , 287*, 184-187.

Pankaj, B., Gupta, S. K., Ojha, S. K., Nandave, M., Kumari, S., & Arya, D. S. (2006). Cardioprotective effect of lycopene in the experimental model of myocardial ischemia-reperfusion injury. *Molecular and Cellular Biochemistry , 289*, (1-2) 1-9.

Paola, P., M. C., R. S., AssunCatalanota, A. B., P. L., et al. (2010). Lycopene induces cell growth inhibition by altering mevalonate pathway and Ras. *Carcinogenesis , 31* (10), 1813-1821.

Park, T. S., Chew, B. P., & Wong, T. S. (1998). Dietary lutein from marigold extract inhibits mammary tumor development in BALB/c mice. *J. Nutr , 128*, 1650-1656.

Pecker, L., Chamovitz, D., Linden, H., Sandmann, G., & Hirschberg, J. (1992). A single polypeptide catalyzing the conversion of phytoene to ;-carotene is transcriptionally regulated during tomato fruit ripening. *Proc. Natl. Acad. Sci. , 89*, 4962-4966.

Raisig, A., & Sandmann, G. (1999). Evolution of carotene desaturation:The complication of a simple pathway. *J. Bacteriol , 181*, 6184–6187.

Reid, K., & Fakler, P. (2010, December 14). Protective effect of lycopene on serum cholesterol and blood pressure: Meta-analyses of intervention trials. *Epub ahead of print* .

Report, G. S. (2010, Oct 28). Global Carotenoids Market to Reach US$1.2 Billion by 2015, Claim Analysts.

Rice, D. W., Schulz, G. E., & Guest, J. R. (1984). *J.Mol.Biol. , 174*, 483-496.

Roger, S., & Milner-White, E. J. (1995). RasMol: Biomolecular graphics for all. *TIBS, 20* (*9*), 374.

Roy, A., Yang, J., & Zhang, Y. (2012)**a**. COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Research, 40, Web Server issue* W473.

Roy, A., Kucukural, A., & Zhang, Y. (2010)**b**. I-TASSER; a unified platform for automated protein structure and function prediction. *Nat. Protoc.5,* 725-738

Sadmann, G., Fraser, P. D., & Linden, H. (1992). *Diversity of phytoene desaturating enzymes and corresponding genes in-involved in carotenoid biosynthesis of photoautotrophic prokaryotes in Research in Photosynthesis.* (N. E. Murata, Ed.) Amsterdam: Kluwer Academic Publ.

Sali, A., Overington. J. P., Johnson, M. S. & Blundell, T. L. (1990). From comparisons of protein sequences and structures to protein modelling and design. *TIBS, 15*, 235-240.

Salim, A. B., & Peter, B. (2005). Golden Rice- five yeras on the road-five years to go? *Trends Plant Science, 10, 565-575. , 10* (12), 565-575.

San, Jose. CA (Vocus). (2010). *Global Carotenoids Market to Reach US$1.2 Billion by 2015, According to a New Report by Global Industry Analysts, Inc.* California: Global Industry Analysts, Inc.

Sandmann, G. (2009). Evolution of carotene desaturation: The complication of a simple pathway. *Arch. Biochem. Biophys , 483* (2), 169-174.

Sandmann, G. (1994). Carotenoid biosynthesis in microorganisms and plants. *Eur.J.Biochem. , 223*, 7-24.

Sandmann, G., & Fraser, P. D. (1993). Differential inhibition of phytoene desaturases from diverse origin and analysis of resistant cynabacterial mutants. *Z.Natur-forsch.C.Biosci. , 48*, 307-11.

Sandman G., P.D., and Linden, H. (1992). Diversity of phytoene desaturating enzymes and corresponding genes involved in carotenoid biosynthesis of photoautrophic prokaryotes in *Research in photosynthesis* (Murata, N., ed.) pp 51-54, Kluwer Academic Publ., Amsterdam.

Seon-Won, K., & Keasling, J. D. (2001). Metabolic Engineering of the Nonmevalonate Isopentenyl Diphosphate Synthesis Pathway in Escherichia coli Enhances Lycopene Production. *Biotech. Bioeng., 72* (4), 408-415.

Schmidhauser, T. J., Lauter, F. R., Russo, V. E., & Yanofsky, C. (1990). Cloning, sequence, and photoregulation of al-i, a carotenoid biosynthetic gene of Neurospora crassa. *Mol. Cell. Biol. , 10*, 5064-5070.

Schmidt, T., Haas, J., Cassarino, T.G.,& Schweed, T. (2011). Assessment of ligand-binding residue predictions in CASP9. *Proteins, 79 (Suppl. 10)* 145-155.

Schmidt-Dannet, C. (2000). Engineering Novel Carotenoid in Micro-organisms. *Curr. Opinion. Biotech. , 11*, 255-261.

Shimidzu, N., Goto, G., & Miki, W. (1996). Carotenoids as singlet oxygen quenchers in marine organisms. *Fisheries Sci. , 62*, 134.

Sies, H., & Stahl, W. (1995). Vitamins E and C, b-carotene, and other carotenoids as antioxidants. *Am J Clin Nutr. , 62(suppl)*, 1315S-1321S.

Steiger, S., Jackisch, Y., & Sandmann, G. (2005). Carotenoid biosynthesis in Gloeobacter violaceus PCC7421 involves a single CrtI-type phytoene desaturase instead of typical cyanobacterial enzymes. *Arch. Microbiol. , 184*, 207–214.

Swindells, M. B. & Thornton, J. M. (1991). Modelling by homology. *Curr. Opin. Struct. Biol. 1*, 219-223.

Takaichi, S. (2009). Distribution and biosynthesis of carotenoids. In C. D. Hunter, *In The Purple Phototrophic Bacteria* (pp. 97–117). Dordrecht, The Netherlands: Springer.

Takaichi, S., Inoue, K., Akaike, M., Kabayashi, M., Oh-Oka, M., & Madgan, M. T. (1997). *Arch. Microbiol. , 168* (4), 277-281.

Taylor, C. R., Stern, R. S., Leyden, J. J., & Gilcherst, B. A. (1990). Photoaging, photodamage and photoprotection. *J Am Acad Dermatol. , 22*, 1-15.

Taylor, F. R. (1984). *Microbiological Reviews , 48* (3), 181-198.

Taylor, F. R., & Davies, H. B. (1976). Triterpenoid carotenoids and related lipids. Triterpenoid carotenoid aldehydes from Streptococcus faecium UNH 564P. *Biochemistry , 153* (2), 233-239.

Tohru, T., Shinichi, T., Norihiko, M., Takashi, M., Hideaki, M., & Mamoru, M. (2005). The cyanobacterium *Gloeobacter violaceus* PCC 7421 uses bacterial-type phytoene desaturase in carotenoid biosynthesis. *Federation of European Biochemical Societies , 579*, 2125–2129.

Tophan, C. M., Thomas, P., Overington, J. P., Johnson, M. S., Eisenmerger, F. & Blendell, T. L. (1991). An assessment of COMPOSER: a rule-based approach to modelling protein structure. *Biochem. Soc. Symp. 57*, 1-9.

Tsuchiya, T., Takaichi, S., Misawa, N., Maoka, T., Miyashita, H., & Mimuro, M. (2005). The cyanobacterium *Gloeobacter violaceus* PCC 7421 uses bacterial-type phytoene desaturase in carotenoid biosynthesis. *FEBS Lett , 579*, 2125–2129.

Wang, C.-W., Oh, M.-K., & James, C. L. (1999). Engineered isoprenoid pathway enhances astaxanthin production in Escherichia coli. *Biotech. Bioeng , 62* (2), 235-241.

W.H.O. (2011). Micronutrients deficiency.

Wilson, C. & Doniach, S. (1989). A computer model to dynamically simulate protein folding studies with Crambin. *Protein, 6*, 193-209.

Wu, S., & Zhang, Y. (2007)**a**. LOMETS: a local meta-threading-server for protein structure prediction. *Nucleic Acids Res., 35*, 3375-3382.

Wu, S., Skolnick, J., & Zhang, Y. (2007)**b**. *Ab initio* modeling of small proteins by iterative TASSER simulations. *BMC Biol. , 5*, 17.

Xu, J., & Zhang, Y. (2010). How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics , 26*, 889–895.

Ye, X., Al-Babili, S., Kloti, A., Zhang, J., Lucca, P., Beyer, P., et al. (2000). Engineering the Provitamin A (β-Carotene) Biosynthetic Pathway into (Carotenoid-Free) Rice Endosperm. *Science , 287* (5451), 303-305.

Yokoyama, A., & Miki, W. (1995). Composition and presumed biosynthetic pathway of carotenoids in the astaxanthin-producing bacterium *Agrobacterium aurantiacum*. *FEMS Microbial Lett , 128*, 139-144.

Yokoyama, A., Miki, W., Izumida, H., & Shizuri, Y. (1996). New trihydroxy-keto-carotenoids isolated from an astaxanthin-producing marine bacterium. *Biosci Biotech , 60*, 200-203.

Yue, P., & Moult, J. (2006). Identification and analysis of deleterious human SNPs. *J. Mol. Biol. , 356*, 1263-1274.

Zhang, Y. (2009). I-TASSER: fully automated protein structure prediction in CASP8. *Proteins , 77*, 100-113.

Zhang, Y. (2009). Protein structure prediction: when is it useful? *Curr. Opin.Struct. Biol. , 19*, 145-155.

Zhang, Y. (2008). Progress and challanges in protein structure prediction. *Curr. Opin. Struct. Biol. , 18*, 342-348.

Zhang, Y. (2007)**a**. Template-based modeling and free modeling by I-TASSER in CASP7. *Proteins , 69*, 108-177.

Zhang, H. et al., (2007)**b**. Analysis of TASSER-based CASP7 protein structure prediction results. *Proteins , (Suppl 8)*, 90-97.

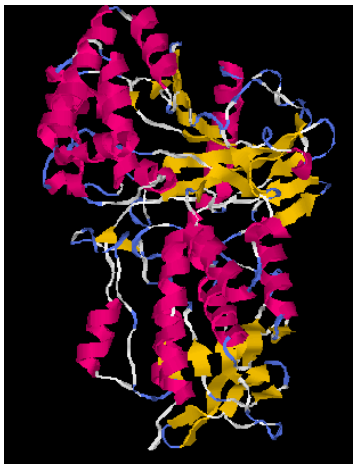
Zhang, Y., & Skolnick, J. (2005). TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res. , 33*, 2302-2309.

Zhang, Y., & Skolnick, J. (2004). Scoring function for automated assessment of protein structure template quality. *Proteins, 57*, 702–710.

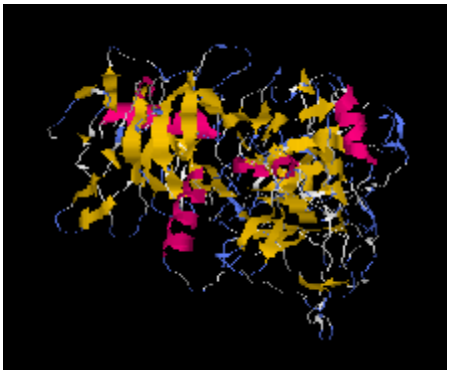
Zhenjian, X., Bing, T., Zongatao, S., Jun, L., & Yuenjin, H. (2007). Identification and functional analysis of a phytoene desaturase gene from the extremely radioresistant bacterium *Deinococcus radiodurans*. *J.Microbiology , 153*, 1642–1652.

Zhou, Y., Pandit, S. B., Lee, S. Y., Borreguero, J., Chen, H., Wroblewska, L., Skolnick, J. (2007). Analysis of TASSER-based CASP7 protein structure prediction results. *Proteins*, *69 Suppl* (*8*), 90-97.

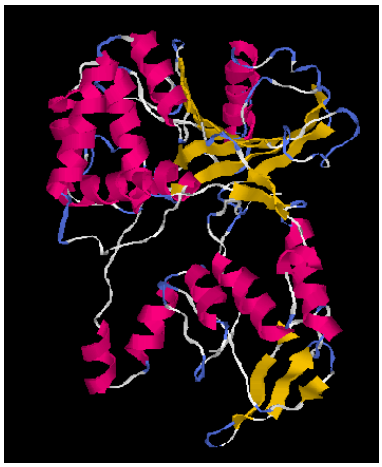
Appendix



Xanthobacter autotrophicus

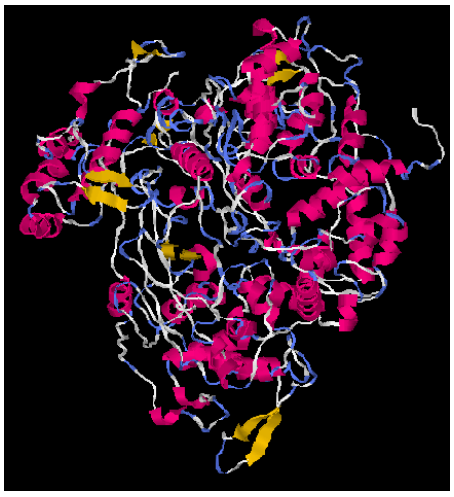


Rhodobacter capsulatus



Paracoccus bogoriensis

Paracoccus sp N81106



Hematococcus pluvialis

Bradyrhizobium sp ORS278

CrtI* sequence from *Paracoccus bogoriensis

5’CAACTTCAGCAGTCCACGGTGATCACCGATCCCGACGCGCTGAAGGAACTCTG
GGCGCTGAGCGGGCAGGACATGGCGCGCGACGTGACGCTGGTGCCGGTCTCG
CCCTTCTACCGGCTGATCTGGCCGGGCGGAAAGGTCTTTGATTATGTGAACGAG
GCCGACCGGCTGGAAGAACAGATCGCGCGCTTCAACCCCGACGATCTTGAGGG
CTATCGCCGCTTTCGCGCCTATTCCGAAGAGGTCTATCAGGAAGGCTATGTGAAG
CTGGGCACCGTGCCCTTCCTGAAACTGGGTCAGATGCTGCGCGCCGCCCCGGA
CCCTGATGCGGCTGGAGGCTTACCGCTCGGTCCATTTCCAAGGTCGCCAGTTTC
ATCAGGGATCCCTATTTGCGGCAAGGCTTTTTCCTATCACACGCTGCTGGTGGGC
GGGAAATCCCTTTTGCCACCAGTTTCGATCTATGCGCTGATCCATGCGCTGGAAC
GGCGCGGCGGCGTCTGGTTCGCCAAGGGCGGCACCAACCGGCTGGTCGCCGG
CATGGTCGCCCTGTTCGAGCGTCTGGGCGGTGATGCTGAACGCCAAGGTCGCC
CGGATCGAAACCGAGGGCAGCCGCGCCACCGGCGTCACGCTGGCCGATCGCC
GGCGCCTGGATGCCGACATGGTCGCCAGCAACGGCGACGTGATGCACAATTACC
GCGATCTTCTGGGCCACACGTCGCGCGGGCAGGGACAGGCGCGGTCGCTGGG
ACCGCAAGCGCTGGTCCATGTCGCTGTTCGTGCTGCATTTCGGCCTGCGCGAGA
TGCCCCAGGGGCTGGCCCATCACACGATCCTGTTCGGGCCGCGCTACAAGGAAT
TGGTGAACGAGATCTTCAAGGGCCCGAAACTGCCCGATGATTTCTCGCTCTATCT
GCATTCGCCCTCGGTGACGGACCCGGACATGGCGCCTGCGGGCATGTCCACGC
ATTACGTCCTGGCCCCGGTGCCGCATCTGGGCCGGGCCGAGATCGACTGGTCC
AAGGAAGGCCCGGCTTATGCCGACCGCATCCTGGCCAGCCTGGAAGAGCGGCT
TATCCCCAATTTGCGCGCCAACCTGACGACGACTCGCATCTTCACGCCGGTCGAT
TTCCAGACCGAGCTGTCGGCCCATCATGGCAGCGCCTTTTCGGTCGAGCCGATC
CTGACGCAATCGGCCTGGTCCGCCGCTACGGAACCG 3’