Metabolic engineering of the photosynthetic bacterium *Rhodospirillum rubrum* to produce industrially interesting plant carotenoids at high level and low cost

Von der Fakultät Energie-, Verfahrens- und Biotechnik der Universität Stuttgart
zur Erlangung der Würde einer Doktorin der
Naturwissenschaften (Dr. rer. nat.) genehmigte Abhandlung

Vorgelegt von

Guoshu Wang

aus ZiBo, Shandong (China)

Hauptberichter: Prof. Dr. Robin Ghosh
Mitberichter: Prof. Dr. Arnd G. Heyer

Tag der mündlichen Prüfung: 17. Juli 2013

Biologisches Institut der Universität Stuttgart
Abteilung Bioenergetik
Pfaffenwaldring 57, 70569 Stuttgart

2013

Die vorliegende Arbeit wurde von mir selbständig und unter ausschließlicher Benutzung der angegebenen Hilfsmittel angefertigt.

I hereby assure that I performed the presented work independently and without further help or any other materials than stated therein.

Stuttgart, den 13. Mai 2013

Guoshu Wang

Declaration

The PhD work I present here, contains some data which are contributed by others.

The SLYC18 mutant was constructed by Prof. Dr. Robin Ghosh, and the HPLC-MS analysis was performed by Prof. Dr. Hartmut Grammel (Biberach University of Applied Science).

The initial β-carotene *in trans* expression in *R. rubrum* was performed by Prof. Dr. Arnd G. Heyer (pRK10665 vector construction), Prof. Dr. Robin Ghosh (*R. rubrum* mutant generation and selection) and Prof. Dr. Hartmut Grammel (HPLC analysis).

The *crtIB* promoter and the structural genes from SWGK46Y and SWGK46G$_R$ were obtained (by PCR) and sequenced by Dr. Caroline Autenrieth.

I. Menu

II. Abbreviations

Amp$_{200}$	200 µg/ml ampicillin concentration
A$_{660}$	Absorption at 660 nm with a 1 cm path-length
A$_{882}$/A$_{660}$	Ratio between the absorption at 882 nm and 660 nm
BChla	Bacteriochlorophyll a
βOG	n-octyl-β-D-glucoside
BSA	Bovine serum albumin
CD	Circular dichroism
CFU	Colony forming unit
CIP	Calf intestinal alkaline phosphatase
CM	Cytoplasmic membrane
crt	Carotenoid
CRY	Cryptochrome
Cyt c$_2$	Cytochrome c$_2$
DCM	Dichloromethane
EM	Electron microscopy
EtOH	Ethanol
FRP	Fluorescence recovery protein
HPLC	High performance liquid chromatography
ICM	Intracytoplasmic membrane
Kan$_{50}$	50 µl/ml Kanamycin concentration
Kan^{R}Tet^S	Kanamycin resistant and tetracycline sensitive
LB-Medium	Luria-Bertani Medium
LDAO	Lauryldimethylamine-N-oxide
LH1	Light-harvesting complex 1
LH2	Light-harvesting complex 2
LOV protein	Light-oxygen-voltage protein
M-Medium	Minimal-medium for Rhodospirillaceae containing 20 mM succinate
MeOH	Methanol
MTHF	Methenyltetrahydrofolate
M2S	Modified M-medium containing 40 mM succinate
M2SF	M2S medium + 0.3% (w/v) fructose
MS	Mass spectroscopy
OCP	Orange carotenoid protein
PMF	Proton motive force

PSI	Photosystem I
PSII	Photosystem II
RC	Reaction centre
RT	Room temperature
TAE	Tris-acetate EDTA: $1\times$TAE: 40 mM Tris-acetate; 1 mM EDTA
TE	Tris EDTA: $10\times$TE: 100 mM Tris-HCl, pH 8.0; 10 mM EDTA, pH 8.0
Tet_4	4 µg/ml Tetracycline concentration
Tet^R	Tetracycline-resistant
Tet^S	Tetracycline-sensitive
UQ	Ubiquinone
UQH_2	Ubiquinol

III. Abstract (English)

The purple non-sulphur photosynthetic bacterium *Rhodospirillum rubrum* has been genetically engineered to express carotenoids, including plant-derived ones, at high level. Initially, a lycopene-producing *R. rubrum* strain, SLYC18, was constructed by chromosomal replacement of the late genes of the carotenoid biosynthesis, *crtCD*, with a kanamycin cassette. SLYC18 showed a longer lag phase than the wild-type, followed by normal growth under both photoheterotrophic and chemoheterotrophic conditions. Absorption spectroscopy and mass spectrometry of extracted carotenoids showed that SLYC18 produced lycopene almost exclusively at high levels (2 mg lycopene/g dry weight cells) under semi-aerobic, dark conditions in a high cell density medium. Using biochemical and spectroscopic analysis showed that lycopene was bound exclusively to the light-harvesting (LH) 1 complexes and reaction centers. SLYC18 exhibited also wild-type levels of LH1 complexes and intracytoplasmic membrane (ICM). In a similar strategy, the *crtCD* region was replaced with an *Arabidopsis thaliana* lycopene β-cyclase (*crtL*) gene flanked by a kanamycin resistance gene, to yield the β-carotene producing strain SWGK46. SWGK46 showed a higher sensitivity to oxidative stress compared to SLYC18, but still could achieve high growth rates in the exponential growth phase and yield wild-type cell densities in the stationary phase when grown both photo- and chemoheterotrophically, respectively. SWGK46 produced β-carotene at a high level of 4.4 mg/g dry weight cells in a high cell density medium. SWGK46 expressed LH1 complexes with a unique Q_y near-infrared absorption maximum at 877 nm, corresponding neither to that of LH1 complexes with bound (882 nm) or no carotenoid (874 nm), respectively, indicating that β-carotene is assembled into the LH1 complexes which exist in an altered conformation. In contrast to all other carotenoid-containing strains of *R. rubrum* observed so far, the β-carotene content of the ICM exceeded that of the bound LH1 complex, indicating that β-carotene has also been released into the ICM phase, implying that CrtL is able to activate the endogenous carotenoid biosynthesis enzymes. Continuous growth passages of SWGK46 yield three secondary mutant phenotypes. Under dark, oxidative chemoheterotrophic conditions a brown secondary mutant (designated SWGK46B) was obtained, which exhibited strongly depressed levels of carotenoid and ICM, and a new absorption maximum at 420 nm, due to an early precursor (a protoporphyrin IX derivative) of bacteriochlorophyll biosynthesis. This phenotype is characteristic of the response of *R. rubrum* to extreme oxidative stress. Under photoheterotrophic conditions, two green secondary mutants (SWGK46G$_R$ and SWGK46G$_{IRR}$, respectively) were observed. SWGK46G$_R$ exhibited strong carotenoid down-regulation when grown photoheterotrophically, but β-carotene production was resumed when the cultures were transferred to chemoheterotrophic conditions. The SWGK46G$_{IRR}$ strain showed the same phenotype under light conditions but did not resume carotenoid production when grown chemoheterotrophically in the dark. The induction of the green

secondary mutants was shown to be blue light-dependent. These phenotypes have never been observed in any other *R. rubrum* strain so far, and probably arise from a protein-protein interaction between CrtL and the *R. rubrum* carotenoid biosynthesis enzymes. DNA sequence analysis indicated that the green phenotypes are due to mutation of the early genes of carotenoid biosynthesis, *crtIB*. Bioinformatics analysis of the *A. thaliana* flavoprotein CrtL indicated domains with possible sequence and structural homologies to known blue light sensors, in particular to plant flavoprotein cryptochromes, as well as to the cyanobacterial 3'-hydroxyechinenone-containing orange carotenoid protein. Possibly, the response of the lycopene β-cyclase to blue-green light may play a role in the putative interaction with the CrtIB, and thereby stimulate the production of secondary mutations.

IV. Zusammenfassung (deutsch)

Das photosynthetische Nicht-Schwefel Bakterium *Rhodospirillum rubrum* wurde mit Hilfe molekularbiologischer Methoden für die Überexpression verschiedener, auch pflanzlicher, Carotinoide optimiert. Zunächst wurden die Gene *crtCD*, die für die späten Carotinoid-Biosynthese-Enzyme kodieren, im Chromosom mit einer Kanymycin-Resistenzkassette ersetzt, und so der Lycopin produzierende *R. rubrum* Stamm SLYC18 erzeugt. SLYC18 zeigte eine längere Lag Phase als der Wildtyp, wuchs aber normal unter sowohl photoheterotrophen als auch chemoheterotrophen Wachstumsbedingungen. Mittels Absorptionsspektroskopie und Massenspektrometrie extrahierter Carotinoide wurde gezeigt, dass SLYC18 unter semi-aeroben Wachstumsbedingungen in einem Hochzelldichte Medium im Dunkeln fast ausschließlich Lycopin in großen Mengen (**2 mg Lycopin / g Trockengewicht Zellen**) produzierte. Biochemische und spektroskopische Analysen zeigten, dass Lycopin ausschließlich an die Lichtsammelkomplexe 1 (LH1) und Reaktionszentren gebunden war. Von SLYC18 wurden Wildtyp Mengen an LH1 Komplexen und intracytoplasmatischen Membranen (ICM) erreicht. Mittels einer ähnlichen Strategie wurde die chromosomale *crtCD* Region durch ein *Arabidopsis thaliana* Lycopin β-Zyklase Gen (*crtL*) zusammen mit einem flankierenden Kanamycin Resistenzgen ersetzt und so der β-Carotin produzierende Stamm SWGK46 erzeugt. SWGK46 zeigte im Vergleich zu SLYC18 eine erhöhte Empfindlichkeit gegenüber oxidativem Stress, konnte aber dennoch sowohl bei photo- als auch chemoheterotropher Inkubation große Wachstumsraten in der exponentiellen Wachstumsphase und Wildtyp Zelldichten in der stationären Phase erreichen. SWGK46 exprimierte LH1 Komplexe mit einem besonderen Q_y Nah-Infrarot Absorptionsmaximum bei 877 nm, das weder zu dem von carotinoidhaltigen LH1 Komplexen (882 nm) noch carotinoidlosen Komplexen (874 nm) passte, was darauf hindeutete, dass β-Carotin in LH1 Komplexe eingebaut war, die in einer veränderten Konformation vorlagen. Im Gegensatz zu allen bislang beobachteten *R. rubrum* Stämmen, war der β-Carotin Gehalt der ICM höher als der durch Carotinoid Bindung am LH1 Komplex erreichbare Gehalt, was darauf hinwies, dass β-Carotin auch in die ICM Phase freigesetzt wurde, woraus geschlussfolgert werden konnte, dass CrtL die endogenen Carotinoid-Biosynthese-Enzyme aktivieren kann. Nach mehrfachen Überimpfungspassagen von SWGK46 traten drei Sekundärmutanten mit verschiedenen Phänotypen auf. Unter oxidativen chemoheterothrophen Wachstumsbedingungen im Dunkeln wurde eine braune Sekundärmutante (SWGK46B) erhalten, die extrem erniedrigte Carotinoid und ICM Level zeigte und deren Absorptionsspektrum ein neues Maximum bei 420 nm aufwies, das von einer frühen Bakteriochlorophyll Vorstufe (ein Protoporphyrin IX Derivat) herrührte. Dieser Phänotyp ist charakteristisch für die Antwort von *R. rubrum* auf extremen oxidativen Stress. Unter photoheterotrophen Wachstumsbedingungen traten zwei grüne Sekundärmutanten (SWGK46G$_R$ und SWGK46G$_{IRR}$) auf. Bei SWGK46G$_R$ wurde die Carotinoidproduktion während

photoheterotropher Inkubation stark herunterreguliert, doch nach Transfer zu chemoheterotrophen Bedingungen setzte die β-Carotin Produktion wieder ein. SWGK46G$_{IRR}$ zeigte im Licht denselben Phänotyp, konnte jedoch im Dunkeln bei chemoheterotrophem Wachstum die Carotinoidproduktion nicht wieder aufnehmen. Es konnte gezeigt werden, dass die Induktion der grünen Sekundärmutanten Blaulicht-abhängig war. Diese Phänotypen sind bislang in keinem anderen *R. rubrum* Stamm beobachtet worden und treten eventuell aufgrund einer Protein-Protein Wechselwirkung zwischen CrtL und den *R. rubrum* Carotinoid-Biosynthese-Enzymen auf. Mittels DNA Sequenzanalyse konnte geschlossen werden, dass die grünen Phänotypen durch Mutationen der frühen Gene der Carotinoid-Biosynthese, *crtIB*, verursacht sind. Bioinformatische Analyse des *A. thaliana* Flavoproteins CrtL zeigte Domänen mit möglichen Sequenz- und Struktur-Homologien zu bekannten Blaulicht Sensoren insbesondere den pflanzlichen Flavoprotein Cryptochromen, sowie dem cyanobakteriellen 3-Hydroxyechinenon-enthaltenden *Orange Carotenoid Protein*. Die Antwort der Lycopin β-Zyklase auf blau-grünes Licht könnte eine Rolle bei der möglichen Interaktion mit CrtIB spielen und somit die Bildung von Sekundärmutationen stimulieren.

Chapter 1   Introduction

Phototrophic prokaryotes are generally separated into three groups: purple, green and cyanobacteria (blue-green algae) (Clayton and Sistrom, 1978). The phototrophic purple and green bacteria carry out anoxygenic photosynthesis, using only one photosystem and reduced sulfur compounds, molecular hydrogen, or simple organic compounds as electron donors which have a lower redox potential than water. Correspondingly, the final oxidized products are sulfate, protons, and organic compounds and CO_2, respectively. The photosynthetic pigments are bacteriochlorophylls a (BChla), b, c, d and e (purple bacteria contain only BChla and b whereas green bacteria contain BChl a, c and d or BChla and e) and a great variety of carotenoids. Cyanobacteria, like the eukaryotic phototrophs, carry out oxygenic photosynthesis during which electrons are donated by water and oxygen is ultimately oxidized, with two photosystems involved. In cyanobacteria, the role of light harvesting is carried out by phycobilisome which is composed of phycobiliproteins. Phycobilisomes are attached to the reaction centers (RCs) via baseplate that contains chlorophyll a as the major pigment, and β-carotene and zeaxanthin as the most common carotenoids (Bryant *et al.*, 1979).

Based upon the ability of usage of sulphur as an electron donor for phototrophic CO_2 assimilation, purple bacteria are categorized into two groups: purple sulphur bacteria (Chromatiaceae, formerly named as Thiorhodaceae) and purple non-sulphur bacteria, (Rhodospirillaceae, formerly named as Athiorhodaceae) (Molisch, 1907). The most heavily studied Rhodospirillaceae species are *Rhodobacter sphaeroides, Rhodobacter capsulatus,* and *Rhodospirillum rubrum.*

1.1 The subcellular architecture of purple non-sulphur photosynthetic bacteria

My PhD work has focused on the purple non-sulphur gram-negative photosynthetic bacterium, *R. rubrum*. *R. rubrum* has a spiral cell shape with a cell size ranging from 1.7 μm to 5 μm (**Fig 1.1**). Most purple bacteria are flagellated and exhibit photo- and chemotactic responses to their environmental conditions. For photoheterotrophic cells, the reserve material depends on the exogenous carbon source. The C_4 dicarboxylic acids and propionate contribute to the deposits as glycogen, however, poly-β-hydroxybutyric acid (PHB) is also deposited when acetate is supplied (Stanier *et al.*, 1959). There are two modes of cell division in Rhodospirillaceae: binary fission (*Rhodospirillum fulvum, Rhodospirillum molischianum, Rhodospirillum photometricum* (Cohen-Bazire and Sistrom, 1966)) and budding (*Rhodomicrobium vannielii* (Boatman and Douglas, 1961), *Rhodopseudomonas palustris* (Tauschel and Drews, 1967), and *Rhodopseudomonas viridis* (Whittenbury and McLee, 1967)).

Generally the structure of the cells of purple non-sulphur photosynthetic bacteria is as follows: the outer membrane (7-8 nm), the cell wall (7-8 nm) and the cytoplasmic membrane (CM). When the

oxygen tension decreases below a species-specific threshold value, the CM invaginates to form intracytoplasmic membrane (ICM) which is continuous with CM in vesicles (*R. rubrum* (Vatter and Wolfe, 1958) and *Rb. sphaeroides* (Cohen-Bazire, 1963)) or lamellar stacks (other species in Rhodospirillaceae, and can be isolated as chromatophores (**Fig 1.1**).
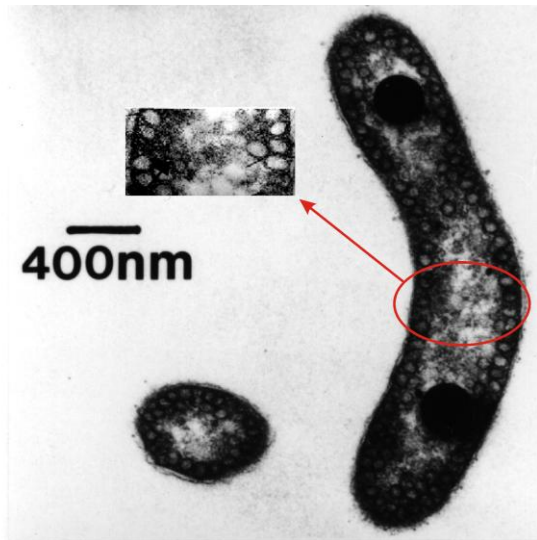
Fig 1.1 The electron microscopy image of photoheterotropic *R. rubrum* cell. The chromatophores are enlarged and indicated in the red circle and arrow (Ghosh, unpublished).

1.2 The localization of the photosynthetic complexes

The intense studies on the procedures to isolate pure chromatophores in early days (Schachman *et al.*, 1952; Frenkel and Hickman, 1959; Cohen-Bazire and Kunisawa, 1960) enabled a large progress in the determination of composition and the enzymatic characterization of the chromatophores membrane. The photosynthetic pigment-protein complexes were revealed to be localized in the ICM by studies using chromatophores isolated by French-pressure-cell extracts. The enzyme activity studies revealed that the ICM contains the components specific for photosynthetic growth, i.g. the photosynthetic apparatus (light-harvesting (LH) complexes and reaction centre (RC)) and other components for the cyclic photochemical electron transport chain (cytochrome c_2, ATPase) (Clayton and Haselkorn, 1972; Fraker and Kaplan, 1972; Takemoto and Lascelles, 1973; Niederman, 1974; Collins and Niederman, 1976), therefore was called photosynthetic membrane (PM) (**Fig 1.2**)
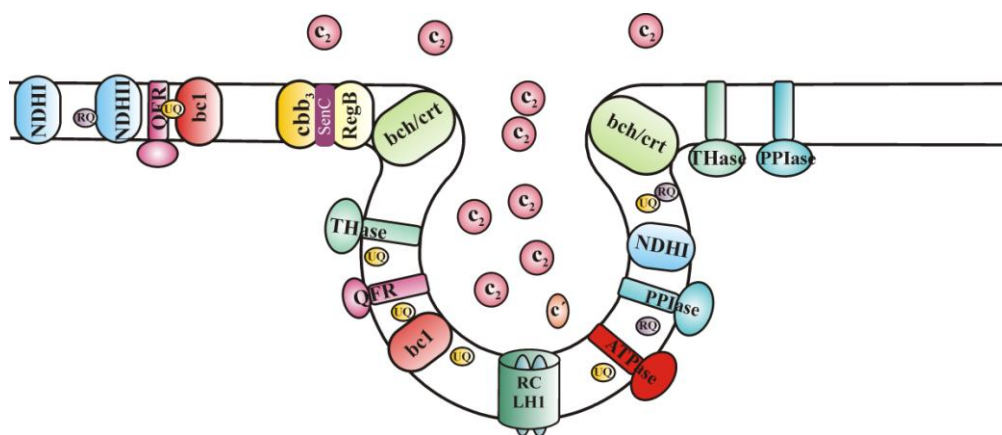
Fig 1.2 A scheme of the ICM organization in *R. rubrum*. In this scheme, the photosynthetic enzymes localizing in the ICM are indicated as follows: the LH1-RC complexes (RC-LH1), the cytochrome bc_1 complexes (bc1), the ATP synthase (ATPase), the pyrophosphate synthase (PPIase), a quinone-fumarate reductase (QFR) which also functions as a succinate dehydrogenase (SDH). Further proteins, such as two types of NADH dehydrogenases (NDHI and NDHII), transhydrogenase (THase), cbb_3 oxidase (cbb_3), histidine kinase RegB (RegB), a photosynthesis gene expression regulator SenC, and the BChla and carotenoid (crt) biosynthesis enzymes (bch/crt), are also required for photosynthesis function but are localized in the CM or at the branch point to the ICM (Ghosh *et al.*, unpublished data).

1.3 The architecture and the assembly of the photosynthetic apparatus

The photosynthetic apparatus of purple non-sulphur photosynthetic bacteria is composed of LH complexes and RC. The LH complexes are responsible for light energy collection and transfer to the RC. *R. rubrum* contains only one type of LH complex, designated as LH1. The in-plane structural order of chromatophores from *R. rubrum* and *Rb. sphaeroides* was initially discovered by Ueki *et al.* (1976) using X-ray diffraction, but this order only extended in a limited area of the chromatophores. Then it was reported by Kataoka *et al*. (1984) that the photosynthetic unit in the chromatophores from *R. rubrum, Rb. sphaeroides, Rhodospirillum molischianum* and *Rhodopseudomonas palustris* is highly organized in the plane of the membrane, using X-ray diffraction in combination with infrared absorption spectroscopy analysis. The tetragonal/hexagonal packing of the photosynthetic membrane of *Rhodopseudomonas viridis* was revealed by Miller and Jacob, 1983, and Stark *et al.*, 1984, respectively, using electron microscopy of negatively staining techniques. Later on, a structural study of the natural chromatophores membrane from *R. rubrum* G9 was performed by Ghosh *et al*. (1984) who demonstrated that the packing of the photosynthetic apparatus in the native PM is regular and compact as well as that of the artificial LH1 2D crystals by using freeze fracture techniques (**Fig 1.3**). The LH1 complexes exhibited as particles on the PM with 12 nm diameter.
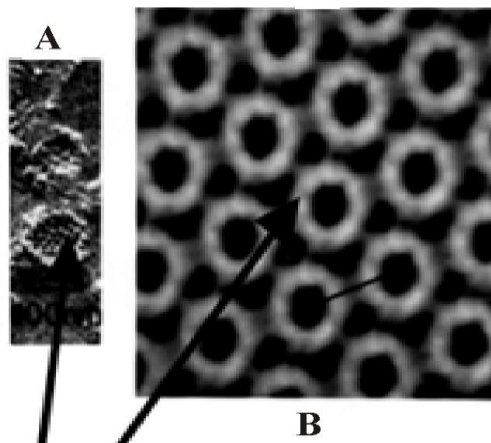
Fig 1.3 The architecture of the chromatophore membrane of _R. rubrum_. The native chromatophore membrane from _R. rubrum_ G9 (A) is compared to the artificial 2D LH1 crystals (B) (Walz and Ghosh, 1997).

To investigate the organization and assembly of bacterial antenna complexes _in vitro_, electron cryomicroscopy (cryoEM) was employed to analyze the 2D array crystals of the reconstituted LH1 complexes from _R. rubrum_. A 8.5 Å projection map showed a _R. rubrum_ LH1 complex in a closed ring of 16 (αβ) heterodimers, with the β polypeptide in the outer ring (diameter: 11.6 nm) and the α polypeptide in the inner ring (diameter: 6.5 nm) (Karrasch _et al._, 1995). This revealed that the space in the LH1 ring is large enough to accommodate the RC. The RC localization within the closed LH1 complex ring was confirmed by the projection map at 8.5 Å of the 2D crystals of the LH1-RC complexes isolated from _R. rubrum_ (Jamieson _et al._, 2002) (**Fig 1.4**). The dimensions of the LH1 ring stayed the same when the 2D crystals of the LH1-RC complex was analyzed using the same procedure described by Karrasch _et al._ (1995).
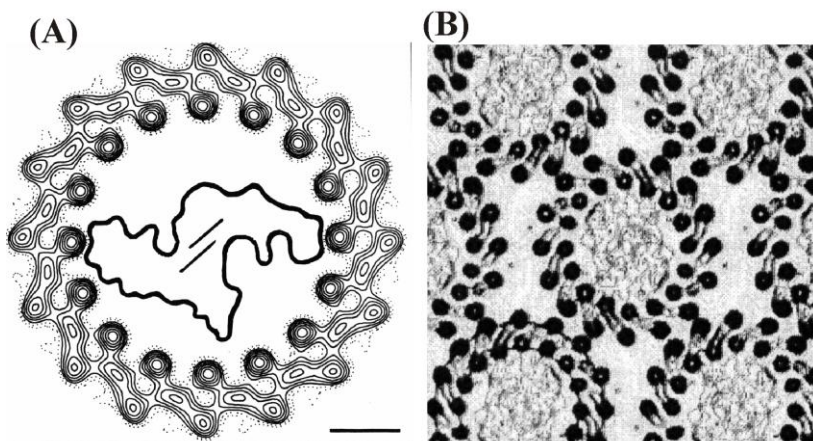
Fig 1.4 The organization of _R. rubrum_ photosynthetic apparatus. (A) The LH1 complexes was shown in a closed ring of 16 (αβ) heterodimers and the RC was hypothetically placed in the middle of the ring (taken from Karrasch _et al._, 1995). (B) The RC was confirmed to localize in the ring of the LH1 complexes (taken from Jamieson _et al._, 2002).

The LH1 α- and β-polypeptides contain 52 and 54 amino acid residues, respectively (Cuendet and Zuber, 1977, Picorel *et al.*, 1983, Brunisholz *et al.*, 1984). The N-termini and C-termini of both polypeptides are localized in the cytoplasm and periplasm, respectively. Photosynthetic pigments are very important for the photosynthetic apparatus in the light-harvesting, energy transfer, photoprotection, as well as structural point of view because the LH1 ring is largely stabilized by pigment-protein interaction. Each αβ heterodimer contains two molecules of BChla and a single carotenoid (crt) molecule, spirilloxanthin (Cogdell *et al.*, 1982; Picorel *et al.*, 1983). Two histidine residues (αHis29 and βHis37) which are conserved between species have been implicated to bind BChla. Based upon the structural data described above, an LH1 assembly atomic model was proposed by the Ghosh group (unpublished) (**Fig 1.5**). In this model, BChla localizes in the space between the α- and the β- polypeptides, however, the position of the carotenoid in the LH1 complex is still debated.
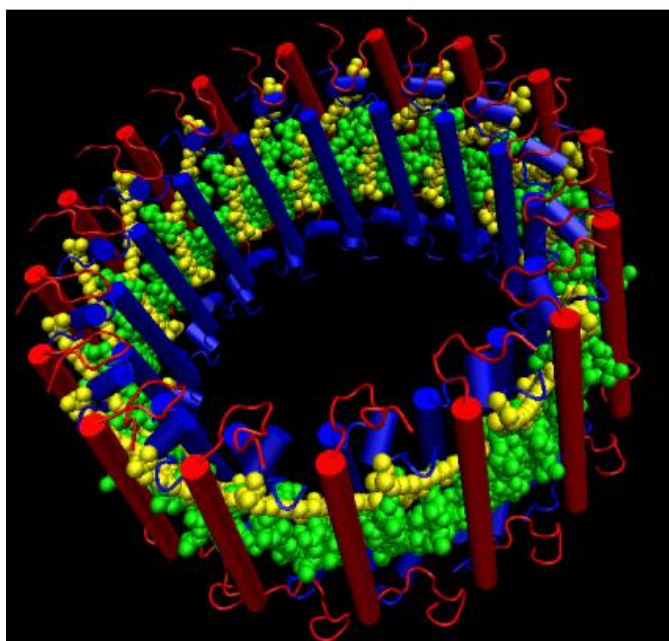
Fig 1.5 The atomic assembly model of the LH1 complexes of *R. rubrum* (Ghosh *et al.*, unpublished data). The α and β polypeptides of the LH1 complexes are indicated in blue and red. BChla and carotenoid are indicated in green and yellow.

Rb. capsulatus and *Rb. sphaeroides* have two LH complexes, LH1 and LH2. The RC is envisaged as being largely surrounded by the LH1 which is in turn surrounded by the LH2 complexes (Cogdell *et al.*, 1999). Each of the LH complexes has two types of polypeptide chains, known as α and β. A RC-LH1-PufX dimer complex is formed, whereas the LH1 is composed of $\alpha_{15}\beta_{15}$ in an incomplete cylinder so as to allow the access of ubiquinone-ubiquinol to/from RC (Siebert *et al.*, 2004). PufX is a subunit which is involved in the dimerization of the photosynthetic core complex in *Rb. sphaeroides,* but is absent in *R. rubrum*. The LH1 contains 30 molecules of BChl and 15 crt molecules, while the LH2 of known structure has 27 molecules of BChla and 9 crt molecules. Purified RCs from *Rb. sphaeroides* comprise three polypeptide chains, H, L, and M, together with four molecules of BChla, two molecules of bacteriopheophytin a (BPha), two molecules of

ubiquinol (UQ) and one molecule of non-haem iron. This subunit structure is identical to that of the *R. rubrum* RC.

1.4 The photochemical electron transport chain in purple photosynthetic bacteria

The photochemical electron transport chain of purple nonsulfur bacteria is cyclic, which is composed of two multimeric transmembrane protein complexes (the RC and the cytochrome bc_1 complex) and two diffusible redox carriers (UQ) in the membrane and the cytochrome c_2 in the periplasmic space) (Jackson, 1988). Light absorption induces charge separation in the special pair of BChla molecules in RC, which is the original electron donor. Then, the electron is transferred to a tightly-bound ubiquinone through a BPha intermediate. Through the UQ diffusion in the membrane and the Q pool ($UQ+UQH_2$), the electrons are transferred to the cytochrome bc_1 complex and two net protons are pumped into the periplasm from the cytoplasm, thereby creating a proton gradient across the photosynthetic membrane. The proton gradient is used to produce ATP through ATP synthase and the electrons cycle back to the RC from the cytochrome bc_1 complex via the periplasmic cytochrome c_2 (Cramer and Crofts, 1982; Dutton and Prince, 1978; Ferguson *et al*., 1987) (**Fig 1.6**). Thus, the RC plays a role as light-driven ferrocytochrome c-ubiquinone oxidoreductase whereas the cytochrome bc_1 complexes functions as a ubiquinol-ferrocytochrome c oxidoreductase (Feher *et al*., 1989; Gennis *et al*., 1993).

Cytochrome c_2 is a 13 kDa soluble protein which is synthesized as a preprotein containing an N-terminal leader signal peptide and secreted into the periplasm. Cytochrome c_2 functions as the electron donor for the oxidized RC and is reduced in turn by the electrons from the cytochrome bc_1 complexes upon photooxidation (Dus *et al.*, 1968; Self *et al.*, 1990). A novel c-type cytochrome (cytochrome c_y), which is membrane-associated, was discovered in *Rb. sphaeroides* and *Rb. capsulatus* to mediate photosynthetic cyclic electron transport when cytochrome c_2 was absent (Jenney and Daldal, 1993). However, cytochrome c_y is absent in *R. rubrum.*
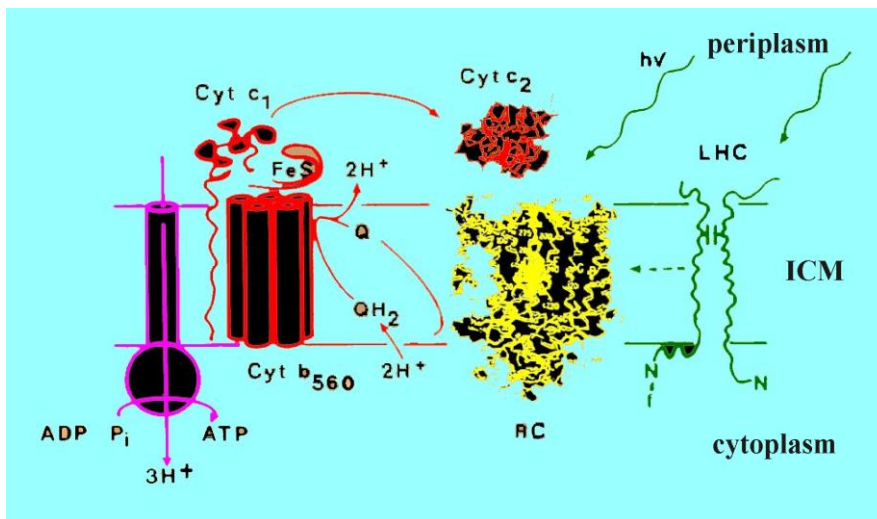
Fig 1.6 The cyclic photochemical electron transport chain of *R. rubrum* (Ghosh, unpublished). All the components for the cyclic photosynthetic electron transport are indicated in this scheme, including the LH1 complex (here shown as LHC), the RC, the cytochrome bc_1 complex (Cyt b_{560}+FeS+Cyt c_1), the ATPase, the cytochrome c_2 (Cyt c_2) and ubiquinone/ubiquinol.

1.5 The overlap between the photochemical electron transport chain and the respiratory transport chain

Purple, non-sulfur bacteria (*R. rubrum*, *Rb. sphaeroides* and *Rb. capsulatus*) are able to generate energy which is required for growth by anaerobic photosynthesis as well as aerobic respiration (Scolnik and Marrs, 1987). Typically, the electron transfer complexes for aerobic respiration are contained in the CM (Irschik and Oelze, 1973). The CM harbours the components of the respiratory electron transport chain, e.g. NADH dehydrogenase, succinate dehydrogenase, Q pool, cytochrome bc_1 complex, cytochrome c_2 and cytochrome oxidase. The electrons from NADH dehydrogenase and succinate dehydrogenase feed into the Q pool from which point the respiratory electron transport chain overlaps with the photochemical electron transport chain. Then, via the Q pool the electrons are transferred to the cytochrome bc_1 complex from where the electron pathway branches. From the cytochrome bc_1 complex, the electrons go to the cytochrome oxidase via cytochrome c_2 for respiration, whereas the electrons go to the RC via the cytochrome c_2 in the cyclic electron transport chain in *R. rubrum* (**Fig 1.7**) (Ferguson *et al*., 1987; Klamt *et al.*, 2008).

In *R. rubrum*, there are three types of NADH dehydrogenase which were schematically compressed into one species in **Fig 1.7**. Under aerobic conditions, the NADH dehydrogenase catalyzes a two-electron transfer from NADH to ubiquinone and pumps protons across membrane from the cytoplasm to the periplasm, thereby contributing to the generation of the proton-motive force (pmf) (Yagi *et al*., 1998; Albracht and Hedderich, 2000; Nicholls and Ferguson, 2001). Under

phototrophic conditions, NADH dehydrogenase can function in the reverse direction to produce NADH, driven by the proton-motive force across the membrane, and transhydrogenase catalyzes the production of NADPH using NADH (Grammel *et al*., 2003) which is important for balancing the cellular redox potential. Succinate dehydrogenase, catalyzing the electron transfer from succinate to ubiquinone and producing fumarate without contribution to the pmf, plays an important role in coupling the central metabolism with the membrane electron transport chain (Iverson *et al.*, 1999; Lancaster *et al.*, 1999; Ohnishi *et al.*, 2000). In *R. rubrum*, fumarate reductase is probably the same enzyme as succinate dehydrogenase, since only a single gene exists in the genome. Fumarate reductase functions in the reversible direction, using rhodoquinol (RQH_2) as redox input, which is very important for redox regulation when the TCA cycle is operating in the reductive direction (see **1.7**). There are at least two oxidases in *R. rubrum*, cytochrome oxidase cbb_3 (Zannoni, 1995) and ubiquinol oxidase (Swem and Bauer, 2002), both of which use O_2 as electron acceptor and pump protons across the membrane. Ubiquinol oxidase is very important for removing excess redox equivalents when the Q pool is over-reduced under semi-aerobic conditions. In *Rb. sphaeroides* and *Rb. capsulatus*, an additional cytochrome oxidase, cytochrome oxidase aa_3, is present.
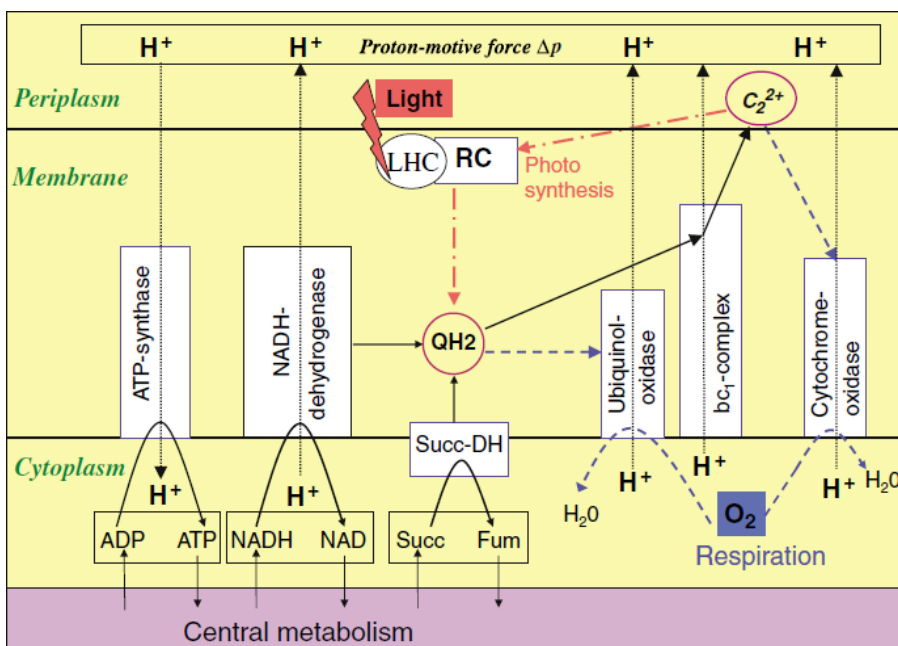
Fig 1.7 The scheme of the photosynthetic and chemoheterotrophic growth modes of *R. rubrum*. The cyclic photochemical electron transport chain overlaps with that of respiration (taken from Klamt *et al.*, 2008). In this scheme, three NADH dehydrogenases are compressed into one, and two oxidases are indicated as one. Taken from literature, the LH complex is abbreviated as LHC in this scheme.

The plant and algal antenna systems localize in the thylakoid membrane even though the two photosystems, and especially photosystem I, have much antenna chlorophyll associated with them. There are distinct polypeptide complexes, associated with the two photosystems I and II, with non-covalently attached chlorophylls and carotenoids, known as LHC I and LHC II. Photosystems I and II are responsible for the linear photochemical electron transport chain (**Fig 1.8**), ultimately resulting in the oxidation of H_2O and yielding O_2. Recent structural data have shown that both of the two photosystems are highly organized multimeric enzyme complexes, containing both LHC type proteins and photochemical reaction centres (Jordan *et al*., 2001; Zouni *et al*., 2001). The cytochrome b_6f complex is analogous in function to the cytochrome bc_1 complex in bacteria, even though the details of the individual structure differ. In plant chloroplasts, the mobile quinone species is plastoquinone (PQ).
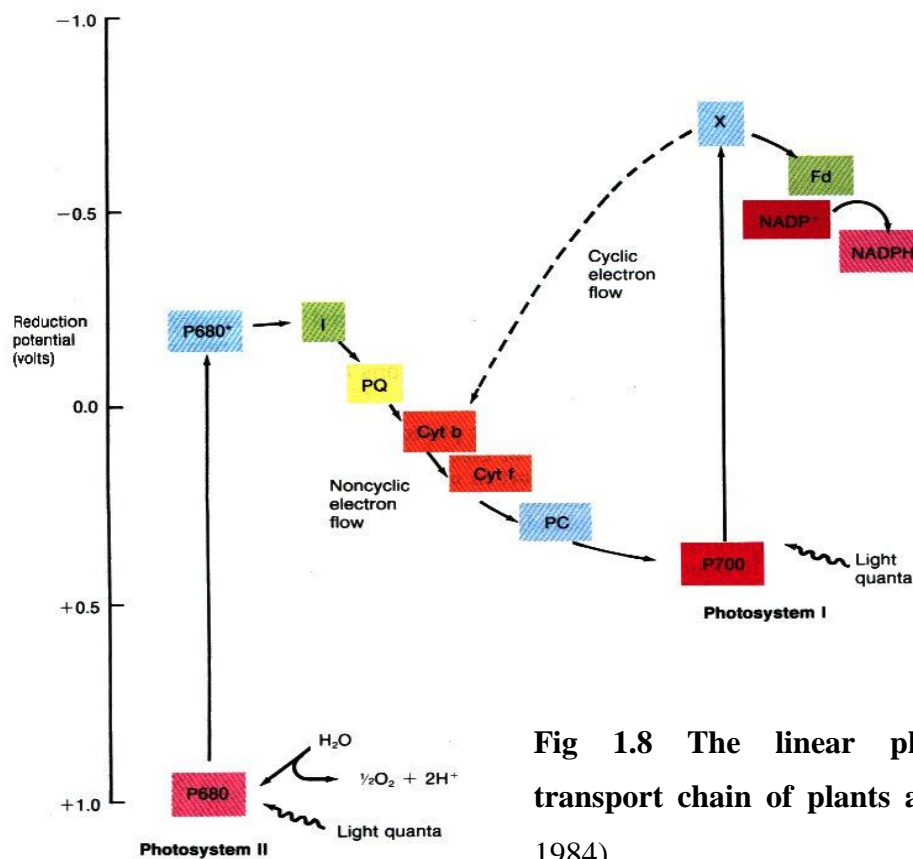
Fig 1.8 The linear photochemical electron transport chain of plants and algae (Brock *et al*., 1984).

1.6 Regulation of the expression of the photosynthetic genes

The expression of the photosynthetic genes is regulated by complex regulatory circuits by linking genes into superoperons to contain overlapping transcripts, control of the initiation of the transcription and the control of mRNA decay rate upon the effects of the environmental factors, like light and O_2 (Bauer, 1995). Here will discuss the photosynthetic gene cluster and the *trans*-acting regulatory circuits in detail.

1.6.1 The photosynthetic gene cluster

In *Rb. sphaeroides* and *Rb. capsulatus*, all of the genes which encode the enzymes required for the biosynthesis of the photosynthetic apparatus including LH1 and RC, and those for BChla and crt biosynthesis, are localized within the 40-50 kb so-called "photosynthetic gene cluster" (PGC). However, the organization of the PGC differs between photosynthetic organisms. Two subclusters of the PGC are conserved (*crtCDEF-bchCXYZ-pufQABLM* and *bchFNBHLM-LhaA-puh*), even though the sequence of some genes differ in some phototrophic proteobacteria (**Fig 1.9**). In *R. rubrum,* the photosynthetic genes are separated into three subclusters in the PGC (Swingley *et al.*, 2007).
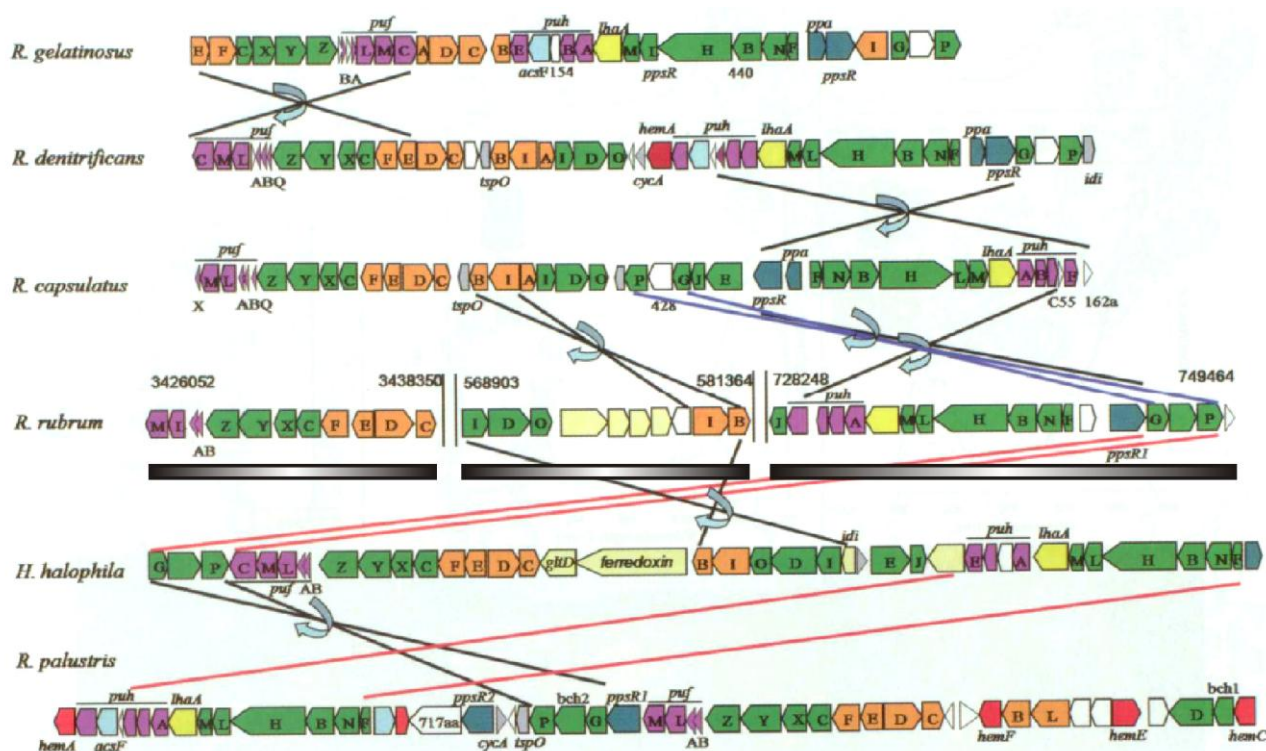
Fig 1.9 The organization of the PGC in different purple bacteira. The genes encoding for light-harvesting complexes and RC, BChl biosynthesis, crt biosynthesis, and regulatory proteins are indicated in purple, green, orange and blue. Other unique genes are indicated in other colors (taken from Swingley *et al.*, 2007). The three subclusters in the PGC of *R. rubrum* are indicated with three grey bars.

1.6.2 The *trans*-acting regulatory circuits

The *trans*-acting regulatory circuits regulate the expression of genes of the photosynthetic apparatus in response to environmental factors, such as light and O_2. Three regulatory systems are involved as follows: (1) The RegB/RegA (PrrA/PrrB) signal transduction system plays an important role in the anaerobic activation of the expression of photosynthesis genes. In both *Rb. sphaeroides* and *Rb. capsulatus*, the Cbb_3 oxidase has been demonstrated to be an oxygen sensor which can sense the

electron flow through its smallest subunit CcoQ and generate an inhibiting signal to RegA/RegB system under aerobic conditions. Under these conditions, the expression of the photosynthetic genes is repressed (Oh and Kaplan, 2000). When O_2 concentration goes down, the inhibiting signal is removed from RegB histidine kinase which is then autophosphorylated at a conserved histidine residue. Subsequently, the phosphoryl group is transferred to a highly conserved aspartate residue of the transcriptional regulator RegA, which is then activated and up-regulates the expression of photosynthesis genes (Mosley *et al.*, 1994). (2) The AppA-PpsR system is the aerobic antirepressor/repressor of *bch, crt* and *puc* operon transcription (Gomelsky and Kaplan, 1997; Oh *et al.*, 2000) (**Fig 1.10**). PpsR, the photosynthetic gene repressor, binds to its target photosynthesis promoter under aerobic conditions (Elsen *et al*., 2005). PpsR might have the ability of sensing redox changes through the formation of disulfide bonds, regarding the triggered signal, but the actual input signal is still unknown. AppA, influenced by redox signals as well as blue light, behaves as an antirepressor. Alterations in redox potential are transferred by AppA to PpsR repressor either directly or indirectly in *Rb. sphaeroides* (Gomelsky and Kaplan, 1997). As a result, PpsR repressor activity decreases, i.e., the affinity of PpsR for the TGT-N12-ACA motif located upstream of *puc* decreases, as well as a subset of *bch* and *crt* genes and operons, and of other genes presumably controlled by PpsR. This gives rise to increased expression of the corresponding genes and operons whose products are involved in development of the PS apparatus. (3) Under anaerobic conditions, with an unknown mechanism, FnrL dimerizes to its active form and enhances the expression of cytochrome c_2, cbb_3 and some photosynthetic genes including those required for LH2 formation (Zeilstra-Ryalls and Kaplan, 1998; Oh and Kaplan, 2001).
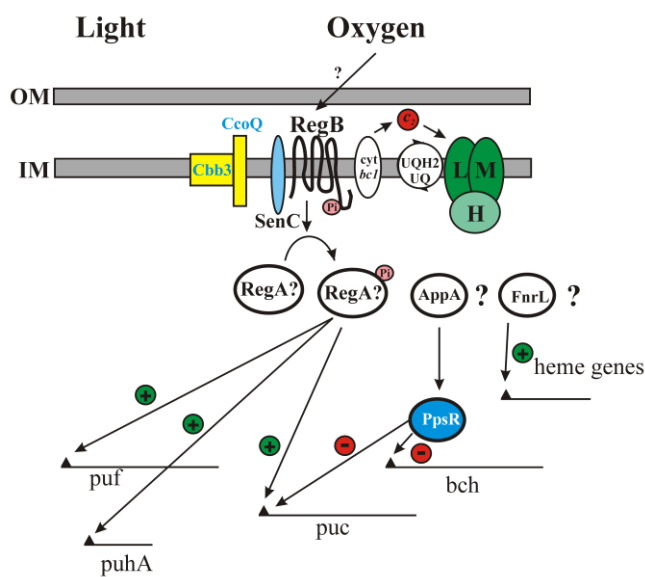
Fig 1.10 The proposed regulatory mechanism of the photosynthetic genes in response to oxygen in *Rb. capsulatus* and *Rb. sphaeroides*. The gene operons encoding for the LH1/RC (*puf* and *puhA*), LH2 (*puc*), and photosynthesis pigments are indicated (R. Ghosh, unpublished).

By analogy to the *Rb. sphaeroides* system, our hypothesis about the regulation network for *R. rubrum* is as follows. When *R. rubrum* grows aerobically, Cbb_3 is saturated by O_2, giving a negative signal to the membrane sensor kinase RegB via SenC. When pO_2 is below 0.5%, the negative signal from Cbb_3 is removed, causing RegB to be autophosphorylated at a conserved histidine residue, what subsequently activates the transcriptional activator RegA through phosphoryl transfer. RegA is the global transcriptional regulator and up-regulates the photosynthetic gene expression, including the LH1 and the RC. The activity of RegB is switched on by low oxygen, but its high activity still needs high levels of ubiquinol/ubiquinone (Swem *et al.*, 2006).

1.7 Metabolic modulation to regulate photosynthetic genes expression under semi-aerobic and dark conditions

High levels of ICM expression need anaerobic and low light conditions. The classical Sistrom medium (Sistrom, 1960) does not yield large quantities of ICM and shows problems like substrate exhaustion concomitant with a rapid pH rise and low cell density (Ghosh *et al.*, 1994). M2SF medium, with NH_4-succinate and fructose as carbon source, was developed in the Ghosh lab, and could maintain the pH at 6.8 due to the uptake of NH_4^+ to obtain high cell density with a high level of ICM in the dark. Growth of *R. rubrum* semi-aerobically in the dark with M2SF medium allows high cell densities and "photosynthetic" levels of ICM to be attained.

The mechanism of how this medium functions is explained as follows: In the initial growth phase when O_2 concentration is high, the TCA cycle operates in the oxidative direction and succinate consumption dominates fructose fermentation (around 20 h) (**Fig 1.11 (A)**) (Grammel *et al.*, 2003). Isocitrate dehydrogenase, α-ketoglutarate dehydrogenase (αKGDH) and malic enzyme provide sufficient NADH, which feeds into UQH_2. Succinate dehydrogenase also feeds electrons into the Q pool and provides a high level of UQH_2 which in turn transfers electron to the cytochrome cbb_3 oxidase via the cytochrome bc_1 complexes and cytochrome c_2. To prevent the Q pool over-reduction, ubiquinol oxidase accepts electrons from UQH_2 and reduces O_2 to H_2O. As mentioned before, the ICM expression level is regulated by low O_2 and high redox level, therefore in the initial phase, the ICM level is low due to too much O_2. However, when the pO_2 drops below 0.5%, then enzymes of anaerobic growth are induced. Under these conditions, the TCA cycle functions in the reductive direction, but the αKGDH still functions weakly (**Fig 1.11(B)**). Fructose fermentation starts to dominate the consumption of succinate. High levels of NADH are provided through glycolysis and the weak αKGDH activity, all of which contributes to the level of UQH_2. To remove excess redox equivalents, fumarate reductase utilizes RQH_2 to reduce fumarate and the transhydrogenase utilizes NADH to produce NADPH. Finally, both the *prr* and light signaling

pathways are activated by low pO_2 and high levels of UQH_2 respectively, and a high level of ICM is achieved (Grammel *et al.,* 2003).
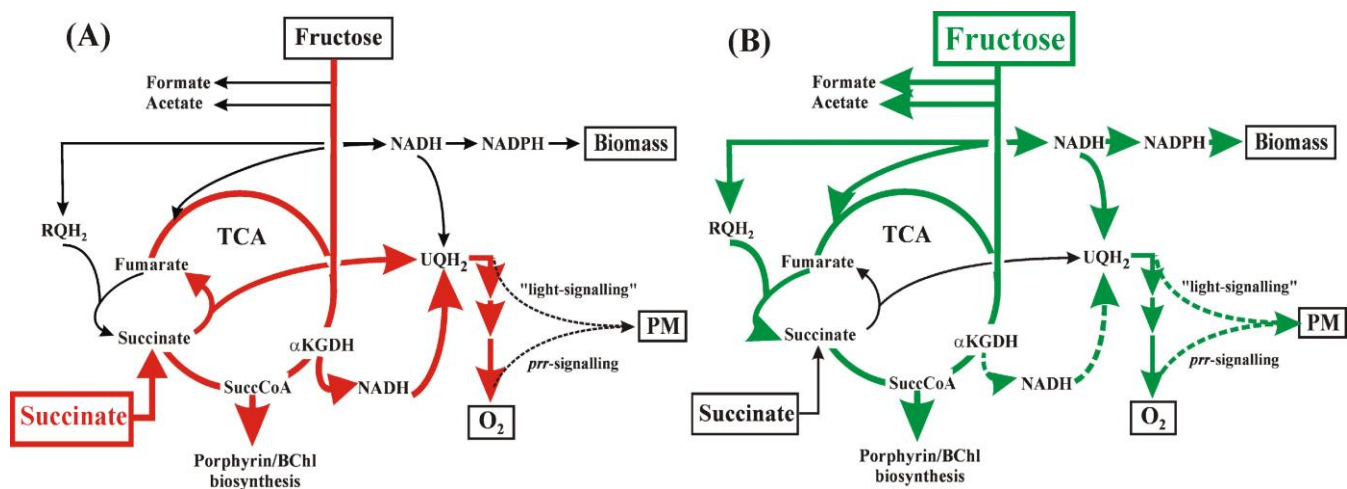
Fig 1.11 *R. rubrum* metabolic modulation with M2SF medium. The metabolism of *R. rubrum* growing under aerobic and semi-aerobic conditions is indicated in red lines (A) and green lines (B).

1.8 The BChla biosynthesis pathway in *R. rubrum*

In *R. rubrum*, the biosynthesis pathway of BChla starts from protoporphyrin IX (PP IX), which is followed by Mg^{2+} addition, catalyzed by magnesium chelatase (BchDIH) to produce Mg-PP IX. This step is a critical and regulatory point, because the production of Mg-PP IX decides the flux going to the Chl/BChl direction. Then Mg-PP IX undergoes methylation and ring closure in two steps, catalyzed by BchM and BchE, respectively, yielding Mg-divinyl-protochlorophyllide a (Mg-DV-PChllide a). The subsequent modification of the Mg-DV-PChllide a is then catalyzed by BchJ and BchBNL, giving rise to MV-protochlorophyllide and chlorophyllide, respectively. Subsequently, chlorophyllide can be catalyzed by BchWXYZ and BchF or BchF and BchWXYZ to produce the same product, 2-desacetyl-2-hydroxyethyl-BChllide a, which is followed by an oxidization step catalyzed by BchC, yielding BChllide a. Finally, BchP catalyzes the addition of a geranylgeranyl pyrophosphate (GGPP) chain to BChllide a, yielding BChla in *R. rubrum*. However, the BPha incorporated into the RC in *R. rubrum* is the phytyl product (BPha$_p$). In *Rb. capsulatus* and *Rb. sphaeroides*, the BChla is also a phytyl product (**Fig 1.12**) (Addlesee and Hunter, 1999; Addlesee and Hunter, 2002).

Protoporphyrin IX (PP IX)

bchDIH ↓

Mg-protoporphyrin IX

bchM ↓

Mg-PP IX monomethylester

bchE ↓

Mg-Divinyl-protochlorophyllide a

bchJ ↓

Mg-Monovinyl-protochlorophyllide a

bchBNL ↓

Chlorophyllide a

bchWXYZ ↙   ↘ *bchF*

2-Desacetyl-2-vinyl-   2-Desvinyl-2-hydroxyethyl-
bacteriochlorophyllide a   bacteriochlorophyllide a

bchF ↘   ↙ *bchWXYZ*

2-Desacetyl-2-hydroxyethyl-
bacteriochlorophyllide a

bchC ↓

Bacteriochlorophyllide a

bchP ↓

Geranylgeranyl-bacteriochlorophyllide a

Fig 1.12 The BChla biosynthesis pathway in *R. rubrum*.

1.9 The carotenoid biosynthesis pathway

The carotenoid biosynthesis pathway was elucidated by creating mutants and analyzing the accumulated precursors. Mutations in four loci within the *crtE, crtB, crtI* and *crtJ* regions of *Rb. capsulatus* resulted in a blue-green phenotype, due to a lack of accumulated colored carotenoids and the presence of BChla (Zsebo and Hearst, 1984). Giuliano *et al*. (1986) indicated that phytoene was converted into colored carotenoids by CrtI in those blue-green mutants and assigned the functions for *crtI, crtC, crtD, crtF* and *crtA* in *Rb. capsulatus*. Later, the PGC in *Rb. sphaeroides* was characterized by Coomber *et al*. (1990) using localized transposon Tn*5* mutagenesis. A total of 87 independent insertions were generated, and *puhA*, nine genes for BChl biosynthesis and six genes for carotenoid biosynthesis were assigned. Based upon the work performed by Coomber *et al.* (1990) it became clear that the blue/green phenotype may either arise from no carotenoid (*crtBE⁻*) or the accumulation of the colorless carotenoid phytoene (*crtI⁻*). The brown phenotype may arise from the accumulation of neurosporene (*crtC⁻*) or its methoxy and hydroxy derivatives (*crtD⁻*). Finally, an

inability to catalyse C-2 oxygenation of spheroidene (brown) to spheroidenone (red) created the *crtA⁻* phenotype.

In *R. rubrum*, the carotenoid biosynthesis pathway starts with several steps shared with the terpenoid biosynthesis pathway until the production of farnesyl pyrophosphate (FPP). As in *Rb. capsulatus* and *Rb. sphaeroides*, phytoene is produced from FPP by two steps, catalyzed by CrtE and CrtB, respectively. Subsequently, phytoene goes through four steps of dehydrogenation, catalyzed by CrtI in *R. rubrum*, yielding lycopene as the first colored carotenoid. Lycopene then undergoes hydration, dehydrogenation and methylation on both ends symmetrically, catalyzed by the enzymes CrtC, CrtD and CrtF (**Fig 1.13**).
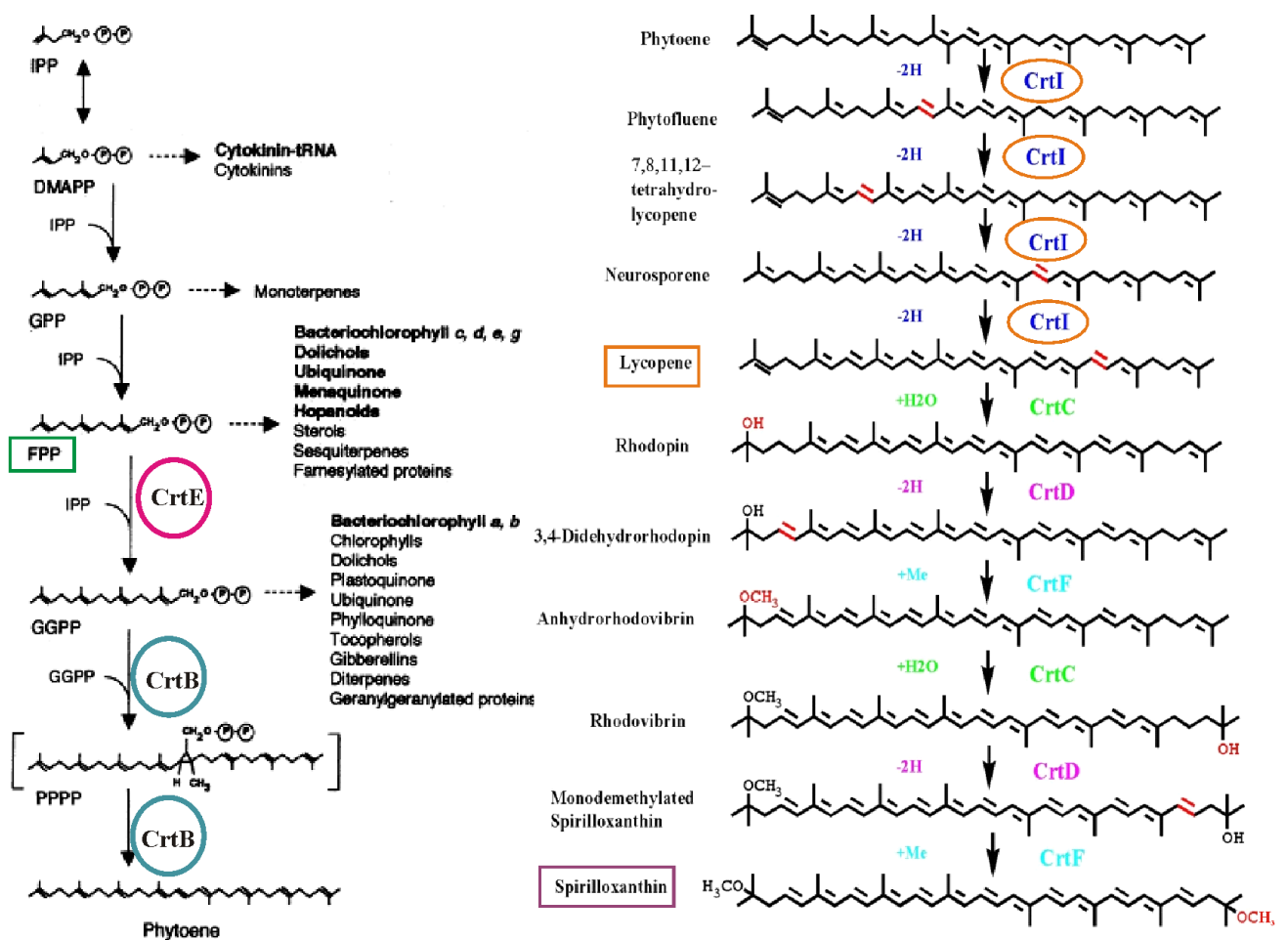
Fig 1.13 The carotenoid biosynthetic pathway in *R. rubrum*. The important intermediates FPP, lycopene and spirilloxanthin are indicated in green, orange and magenta boxes. The enzymes (CrtE, CrtB, CrtI) required for biosynthesis of lycopene are indicated in purple and sea blue and orange cirlces. CrtC, CrtD and CrtF are indicated in light green, pink and light blue.

1.10 The progress of carotenoid expression in the past 20 years

Apart from having light-harvesting, photoprotective and structural functions in photosystems, carotenoids also have many important applications in diet for health benefits, animal feed supplement, and as a natural colorant in industry (Krinsky, 1989; Sandmann *et al.*, 1999; Lee and Schmidt-Dannert, 2002). The sources of the carotenoids used for the nutritional and animal feed supplement are commonly from extraction from bacteria, fungi and algae, or by chemical synthesis from petroleum by-products (Johnson and Schroeder, 1996; Nelis and DeLeenheer, 1991). The problem of extraction from natural sources lies in the relative low carotenoid content in many organisms and that leads to an expensive cost for obtaining large quantities of pigments (Ravanello *et al.*, 2003). Thus the primary source of carotenoids in health supplements is still from chemical synthesis. A possible alternative to synthetic sources of carotenoids are through microbial source in combination with genetic engineering techniques. This has been the major driving force for the active studies in carotenoid expression in the past 20 years.

The organisms being used for carotenoid expression are: natural carotenoid producers (photosystems: plants, photosynthetic bacteria, algae) and microorganisms with the ability to produce carotenoid precursors (*Escherichia coli* and yeast) when recombinant carotenoid enzymes are introduced. The beauty of metabolic engineering is to re-engineer the appropriate productive hosts to obtain large quantities of desirable products regardless of their natural ability.

In the past 20 years, *E. coli* has been established as an excellent organism for carotenoid expression at very high levels. Several stages of development can be recognized:

(1) **Starting stage.** Extensive knowledge of *E. coli* genetics facilitates its application in biosynthesis. *E. coli* does not contain the genes encoding for carotenoid enzymes, however, it can produce the isoprenoid precursor FPP. Strategies of rational design of novel pathways by introduction of carotenogenic genes from *Pantoea* species and marine bacteria *A. aurantiacum* as well as *Alcaligenes sp.* were applied in *E. coli* to enable it to produce lycopene, β-carotene, zeaxanthin (Misawa *et al.*, 1991; Ausich *et al.*, 1991), astaxanthin and canthaxanthin (Misawa *et al.*, 1995) at low level of **0.2-0.5 mg/g dry weight (DW) cells**.

(2) **Regulatory enzymes exploration stage.** Significant efforts have been put into the identification of regulatory enzymes of the carotenoid production. Isopentenyl pyrophosphate (IPP) isomerase (*idi*) was introduced into *E. coli* from different sources (*Haematococcus pluvialis*, *Saccharomyces cerevisiae*, *Phaffia rhodozyma*) and increased the levels of lycopene and β-carotene by 1.5-4.5 fold (Kajiwara, *et al.*, 1997), reaching a yield of about **1 mg/g DW** for both of the carotenoids.

(3) **Metabolic engineering approaches**. More gene targets were identified with direct or indirect effects upon the flux going to carotenoid production by a combination of systematic approaches (stoichiometric modeling) and combinatorial approaches (transposon library searches following transposon mutagenesis (Alper *et al.*, 2005) as well as shotgun mutagenesis (Kang *et al.*, 2005)). Further, gene knock-outs in combination with overexpression of identified gene targets led to a lycopene level of **16 mg/g DW** (Jin and Stephanopoulos, 2007). To increase the flux going to IPP, either the 2-C-methyl-D-erythritol-4-phosphate (MEP) pathway was optimized (Matthew and Wurtzel, 2000) or a heterologous mevalonate (MVA) pathway was introduced into *E. coli* (Yoon *et al.*, 2006, 2008) which enabled high levels of β-carotene of **503 mg/L** or **49.3 mg/ g DW** to be obtained (for more details, see the Introduction of Chapter 4).

Yeasts (especially *S. cerevisiae* and *Candida utilis*) have also been employed as organisms for carotenoid expression because they are able to produce high levels of sterols but are unable to produce carotenoids. The development of carotenoid production in yeast also went through several stages as follows:

(1) Introduction and expression of carotenogenic genes on plasmids *in trans* (Ausich *et al.*, 1991; Yamano *et al.*, 1994). The carotenoid level obtained at this stage was low, yielding lycopene and β-carotene at the levels of **0.113 mg/g DW** and **0.103 mg/g DW**.

(2) Optimization of the carotenoid expression by codon usage modification increased the carotenoid level, for example, lycopene, β-carotene and astaxanthin to levels of **1.1 mg/g DW, 0.4 mg/g DW** and **0.4 mg/g DW** were obtained (Miura *et al.*, 1998).

(3) Metabolic engineering. 3-hydroxy-3-methyl-glutaryl-CoA (HMG-CoA) reductase overexpression in combination with squalene synthase (ERG9) gene disruption led to lycopene production at high levels of **7.8 mg/g DW** in *C. utilis* (Shimada *et al.*, 1998). So far, however, the highest β-carotene level achieved in *S. cerevisiae* was **6.29 mg/g** with HMG-CoA reductase overexpression and ergosterol biosynthesis inhibitors (Yan *et al.*, 2012).

Plants as carotenoid producers, especially tomato (*Lycopersicon esculentum*) and pepper (*Capsicum annum*), have the advantages of direct consumption without carotenoid extraction. Probably due to the complicated eukaryotic genetics and the long growth phase, plants have not been employed as mainstream carotenoid producers at high levels which could meet the industrial needs. However, from the research and knowledge point of view, it is still very important to carry on research in

carotenoid expression and regulation in plants. Many studies have showed that the carotenoid biosynthesis pathway in higher plants was modified to change carotenoid species in tissues through genetic engineering (Römer *et al.*, 2000; Kato *et al.*, 2004; Ralley *et al.*, 2004; Suzuki *et al.*, 2007; Hasunuma *et al.*, 2008).

1.11 The expression of photosynthetic genes regulated by photoreceptors

During the β-carotene expression work, the lycopene β-cyclase from *A. thaliana* was inserted into the chromosomal DNA of *R. rubrum*. It was discovered in my PhD work that *A. thaliana* lycopene β-cyclase may respond to blue-green light as a "light sensor" which enabled us to observe several unique phenomena in purple non-sulphur bacteria so far (see Chapter 5). Below I give a brief summary of the known flavin and carotenoid-containing photoreceptors involved in signal transduction that have been reported so far.

Photoreceptors have the absorption properties from the near UV (350 nm) through the blue to the red/far red (750 nm), which are due to the covalently or noncovalently bound organic chromophores serving as the primary sites of photon absorption. The photoreceptors have modular structures. One or more modules serve as sensor or input domain by binding the chromophore and absorbing light; another plays the role of dimerization or association with another protein, small molecules, or the membrane; the function of the light-dependent catalytic activity or DNA binding is carried by the effector or output domain, which also interacts with other components in the downstream signaling pathway (Möglich *et al.*, 2010).

So far there are six classes of known photoreceptors which are classified by the chemical nature and photochemistry of their chromophores: light-oxygen-voltage (LOV) sensors, blue-light sensors, cryptochromes, xanthopsins, phytochromes and rhodopsins. All of them are water-soluble proteins except rhodopsin which is an integral membrane protein. The chromophores for xanthopsins, phytochromes and rhodopsins are 4-hydroxycinamic acid, tetrapyrrole bilin and retinal, respectively. LOV sensors, blue-light sensors and cryptochromes are all flavoproteins, employing flavin mononucleotide (FMN) or flavin dinucleotide (FAD) as cofactors (van der Horst and Hellingwerf, 2004). Since the plant lycopene cyclase has been shown to be a flavoprotein (Mialoundama *et al.*, 2010), the bioinformatics summary below focuses on known flavoproteins which are also blue light sensors, e.g. LOV sensors, blue-light sensors, cryptochromes and orange carotenoid protein (OCP).

1.11.1 LOV sensor proteins

(a) LOV sensor proteins introduction

LOV protein domains are a subset of the large Per-ARNT-Sim (PAS) domain superfamily and mediate the effects by absorption of UV-A/blue light (320-500 nm). There are five major subsets of LOV proteins, including phototropins, proteins regulating circadian rhythms, LOV histidine kinases, LOV STAS proteins and LOV phosphodiesterases (Crosson *et al.*, 2003). Phototropins are the first LOV domain proteins to be discovered in two plant photoreceptor kinases which control phototropic bending, light-induced stomatal opening and chloroplast movement in response to light intensity changes (Huala *et al.*, 1997, Jarillo *et al.*, 2001). In *Arabidopsis*, blue light receptors which regulate the circadian clock function and phototropic-dependent flowering were reported (Banerjee and Batschauer, 2005). In bacteria, LOV domains coupled to histidine kinases were discovered to act as photoreceptors for phototaxis, or as controllers of the photosynthetic machinary and DNA repair (West and Stock, 2001). The LOV-STAS protein was identified as YtvA as an antisigma-factor antagonist to the environmental stress sigma-factor δ^B by genetic analysis in *Bacillus subtilis* (Akbar *et al.*, 2001). In cyanobacteria and proteobacteria families, some LOV domains are coupled to tandem GGDEF and EAL domains which are implicated in phosphodiester cleavage of cyclic nucleotides (Galperin *et al*., 2001, Pei and Grishin, 2001).

(b) LOV secondary structure

The core of the monomer of LOV2 domain (plant photoreceptor phy3) is composed of **five antiparallel β-sheets, with two α-helices on each side** (**Fig 1.14**). The FMN is non-covalently bound in the interior of the LOV2 domain by a network of hydrogen bonds, van der Waals and electrostatic interactions. The α'A helix and the β-strands C, D, and E make the majority of the contacts with flavin (Crosson *et al.*, 2003). A conserved α-helix (Jα, corresponding to α'A helix) was identified by NMR studies to be important in association with the surface of LOV2, which is located on the C-terminus of LOV2 fragments (oat phot1) docking onto the core of the β-sheet strands. A surface salt bridge, which is formed between E960 and K1001, joins two separate segments of the LOV2 domain, the αB-α′A helices and the βC-βD loop. This surface salt bridge is neutral from the energetic point of view, but is selected because of its role in the light-responsive signaling module.
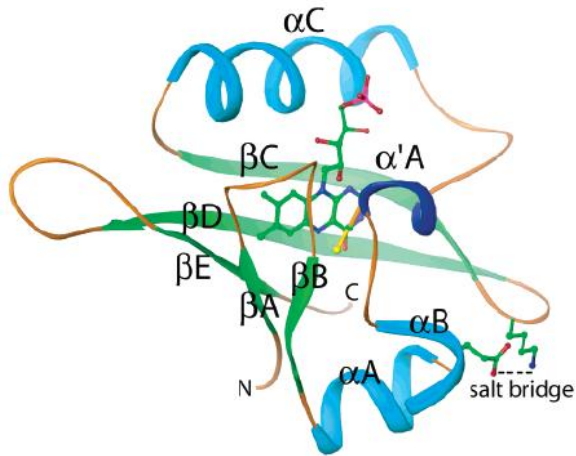
Fig 1.14 The overall fold of the phy3 LOV2 domain from *Adiantum* (taken from Crosson *et al.*, 2003). The α'A helix is indicated in dark blue.

(c) The photochemistry of the LOV domain

In darkness, the LOV domain binds to FMN non-covalently, forming a spectral species which was designated as LOV_{447} and absorbs maximally at 447 nm. Initial absorption of blue light by the FMN chromophore results in the formation of an excited singlet state which then decays into a flavin triplet state (LOV_{660}) with an absorption maximum at 660 nm. LOV_{660} is followed by decay and formation of the FMN-cysteinyl adduct with an absorption maximum at 390 nm, and is designated $LOV_{390.}$ LOV_{390} is the active signaling state which leads to the activation of the photoreceptor. The photoreaction process is fully reversible in darkness (**Fig 1.15 (A)**).

LOV2 can be viewed as a molecular light switch that controls the activity of the C-terminal kinase domain. Blue light induced the C966 of the Jα-helix to rotate by 100º to allow an adduct between C966 and the FMN ring C(4a) to be formed. Subsequently, the ring of the FMN is tilted by 8º, as shown in **Fig 1.15 (B)**. Through the interactions between the FMN and the conserved residues which are involved from the nearby of FMN to the surface salt bridge, the light-induced conformational changes may decrease the structural mobility of the αB-α′A helices and the βC-βD loop, thus affecting LOV-partner interactions (**Fig 1.15 (C)**).
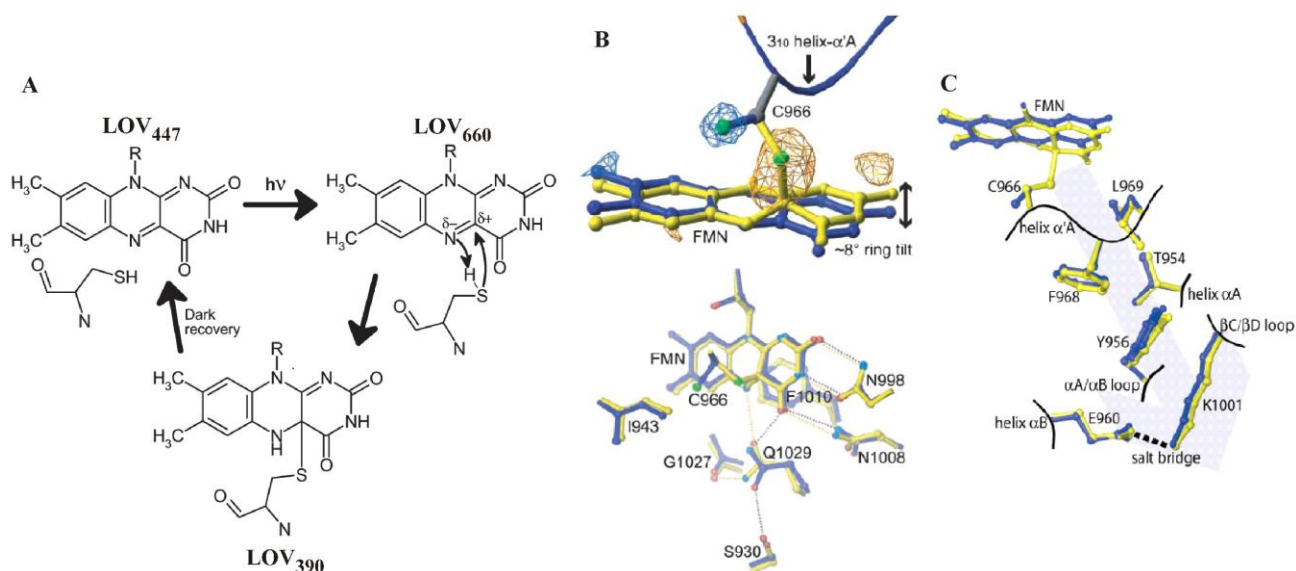
Fig 1.15 The photochemistry and conformational changes of the LOV domain upon blue light induction. The dark state, the light-induced triplet state and the active signaling state of the LOV domain are indicated as LOV_{447}, LOV_{660}, and LOV_{390} (A). The conserved residues in contact between the α′A helix and the FMN isoalloxazine are shown in (B) (taken from Crosson and Moffat, 2002). The conserved residues from the FMN to the salt bridge are also shown (C).

1.11.2 BLUF proteins

(a) The overview of the BLUF proteins

BLUF domains are a distinct family of photoreceptors containing FAD as cofactor, which absorbs blue light to induce the interactions between the FAD isoalloxanthin ring and the residues to be changed (Gauden *et al.*, 2006). BLUF domains predominately occur in prokaryotes, and are also found in eukaryotes (euglenozoa (Iseki *et al.*, 2002) and fungi (Gomesky and Klug, 2002)). The first BLUF domain discovered was AppA from *Rb. sphaeroides* involved in repression of photosynthesis genes (Masuda and Bauer, 2002). So far, no BLUF proteins have been identified in plants.

(b) The secondary structure of the BLUF domain

Tll0078 protein from *Thermosynerchococcus elongatus* BP-1 was the first example of a BLUF protein to be solved by X-ray crystallography (Kita *et al.*, 2005). This protein was shown to be a decamer, assembled by two pentamers. The monomer of Tll0078 is composed of the N-terminal BLUF domain which displays **an α/β fold (β1α1β2β3α2β4β5) with a central domain composed of five β-sheets and two parallel α-helices on one side**, and the C-terminal domain which consists of two α-helices (α3α4) running perpendicular to the β-sheets of the BLUF domain. FAD is located between the α1 and α2 helices (**Fig 1.16**).
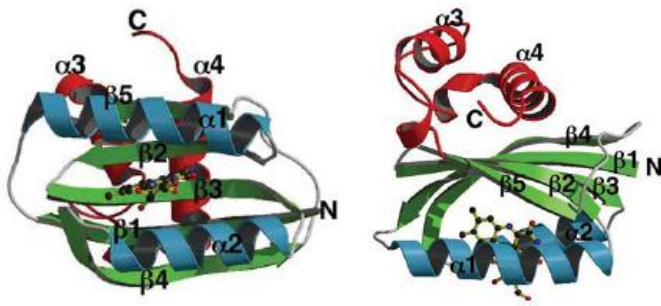
Fig 1.16 The overall fold of the Tll0078 monomer. The BLUF domain is shown in green and blue and the C-terminal domain is shown in red (taken from Kita *et al.*, 2005).

(c) The photochemistry of BLUF proteins

Upon absorption of blue light, the FAD cofactor of the BLUF domain undergoes minimal conformational changes. An electron is transferred from a conserved tyrosine residue (Tyr8) to the FAD ring and a short-lived radical pair is formed (**Fig 1.17**). The side chain of a conserved glutamine residue (Gln50) nearby is assumed to rotate, followed by the electron back transfer and the formation of the signaling state, P_{red}, which differs from the ground state in the hydrogen bonds that the flavin ring forms and shows a slightly red-shifted absorption spectrum by approximately 10 nm (Gauden *et al.*, 2006).
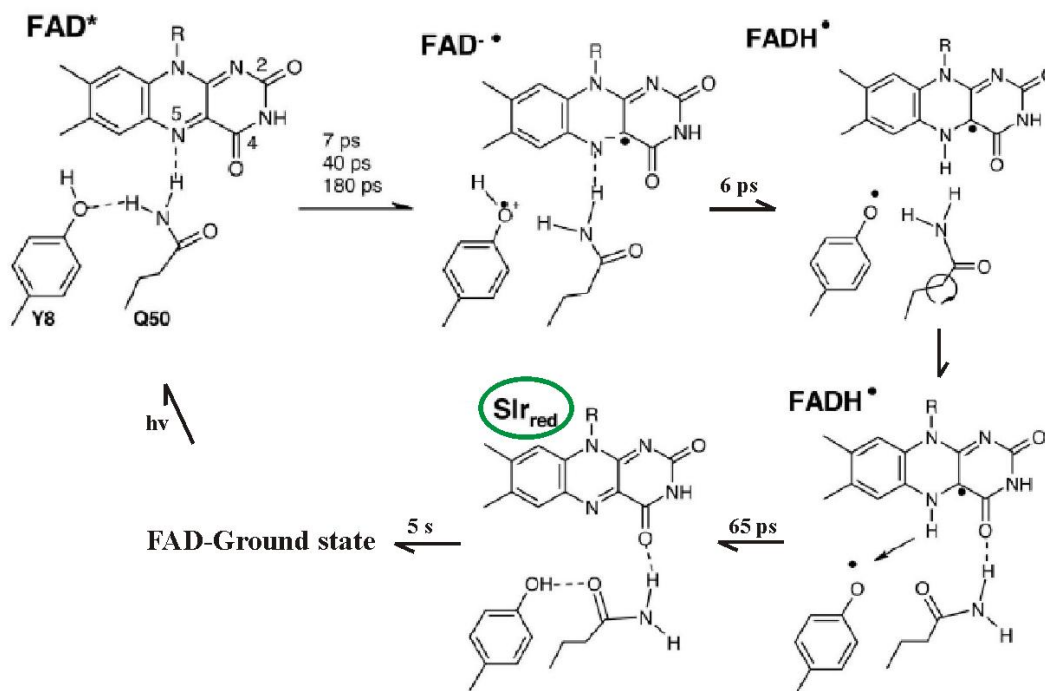
Fig 1.17 The proposed reaction mechanism for photoinduced hydrogen switching in the *Synechocystis* **Slr1694 BLUF domain** (taken from Gauden *et al.*, 2006). The active signalling state FAD (Slr_{red}) is indicated in a green circle.

(d) The residues effected upon light illumination

Upon light illumination, the hydrogen bonds between the FAD isoalloxanthin ring and the interacting residues undergo rearrangement. The Gln63 rotates by 180° and flips out of the flavin-binding pocket. Then, the interaction between the Gln63 and the Met106, and the interaction between the Asn45 and the His105 are changed, which enables Trp104 to move into the void. Finally, His44 transmits the conformational change of β5 to the BLUF domain, serving as a reporter of the active site changes (**Fig 1.18**) (Jung *et al.*, 2006).
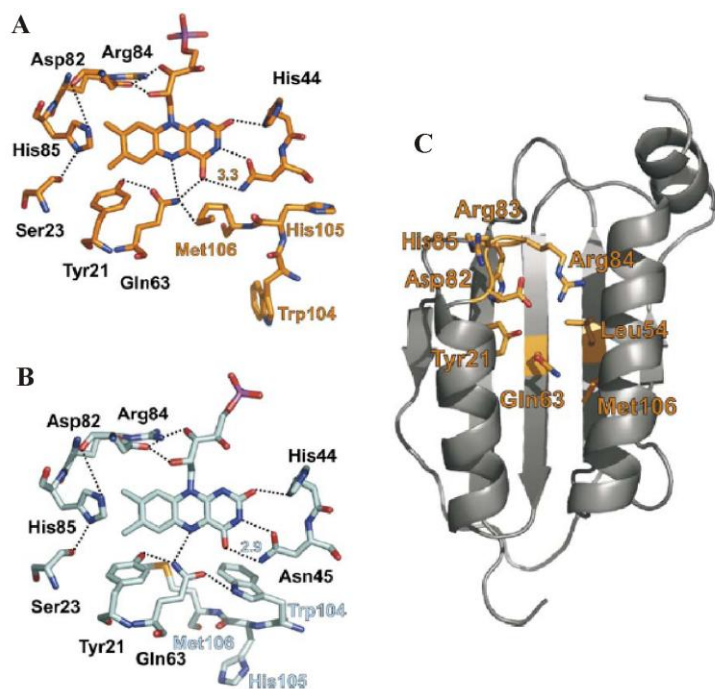
Fig 1.18 The interacting residues during the proposed mechanism of photochemistry of BLUF domains. The interactions between the conserved residues and the flavin ring are indicated before (A) and after (B) blue light illumination. The conserved residues in the secondary structure (taken from Jung *et al.*, 2006) are indicated (C).

1.11.3 Cryptochromes

(a) Introduction of cryptochromes

Cryptochromes, together with the photolyases, which are enzymes using blue light to repair UV-induced DNA damage, make up the photolyase/cryptochrome superfamily. Although cryptochromes show sequence similarities to photolyases, however, they possess no photolyase activity. Cryptochromes are flavoproteins, functioning as blue-light photoreceptors, are distributed into four subfamilies: plant cryptochromes, animal cryptochromes, cryptochrome-DASH (Cry-DASH) family and CPD photolyases (Brudler *et al.*, 2003, Lin and Todo, 2005). *A. thaliana* cryptochrome 1 and 2 (AtCry1 and AtCry2), which are nuclear proteins, and mediate light control of stem elongation, leaf expansion, phototropic flowering and circadian clock by interacting with phytochromes, COP1, clock proteins, chromatin and DNA (Yang *et al.*, 2001, Somers *et al.*, 1998). The *A. thaliana* cryptochrome 3 (AtCry3) belongs to a different class of Cry-DASH proteins. In *Synechocystis*, Cry-DASH was demonstrated to be a transcriptional repressor by structural and functional studies (Daiyasu *et al.*, 2004, Brudler *et al.*, 2003). AtCry1 and AtCry2 are composed of

an N-terminal photolyase homology region (PHR) and a C-terminal extension (CCT); AtCry3 contains a shorter N-terminal extension of the PHR domain (**Fig 1.19**).
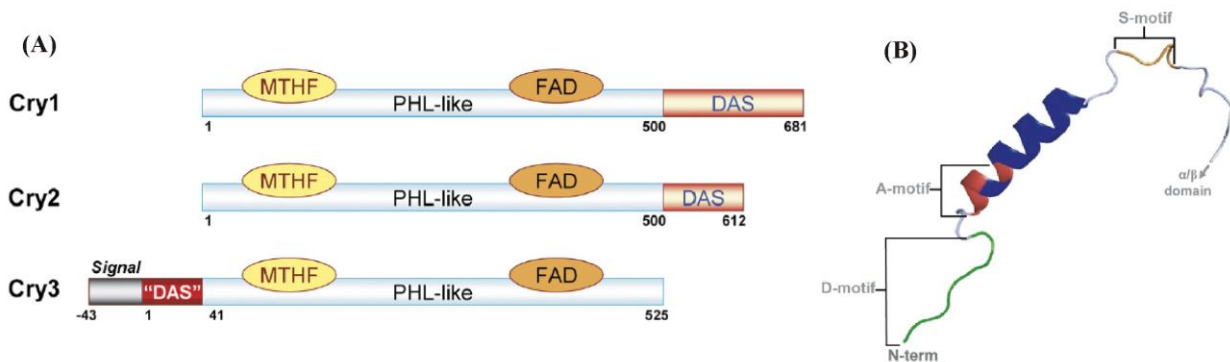
Fig 1.19 The scheme of structural comparison among the three types of cryptochromes from
A. thaliana (taken from Klar *et al.*, 2007). The N- and C-terminal extensions in the three types of cryptochromes are shown in (A) and the scheme of DAS motif in Cry1 is shown in (B) (taken from Klar *et al.,* 2007).

(b) The secondary structure of cryptochromes

The photosensory PHR domains in AtCry1 and AtCry3 share the same architecture with DNA photolyases, which are composed of an N-terminal α/β domain and a C-terminal α-helical domain. The α/β domain adopts a dinucleotide-binding fold consisting of **a five-stranded parallel β sheet surrounded by five helices**. FAD binds non-covalently to the protein through an access cavity which is in a U-shaped conformation and formed by the two lobes of the helical domain (**Fig 1.20**).
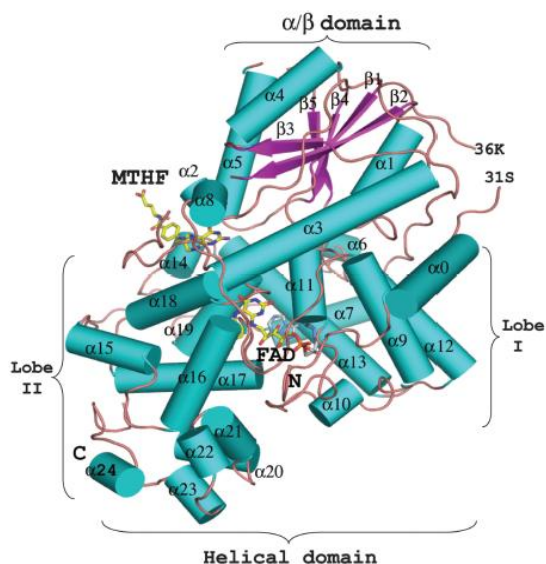
Fig 1.20 The overall fold of a CRY-DASH protein of *Synechosystis sp.* PCC6803 cryptochrome (taken from Huang *et al.*, 2006).

(c) The photochemistry of cryptochromes and the signal transduction of cryptochromes

Plant and animal cryptochromes exhibit a complicated photochemistry upon absorption of blue light. In AtCry1 and AtCry2, a radical intermediate state (semiquinone), FADH·, is induced and accumulates in the activated signaling state. Absorption of green light by FADH· causes further reduction to FADH⁻, which abolishes signaling. The FADH⁻ in turn is reoxidized to the fully oxidized form during dark reversion.

It has been shown that *Arabidopsis* cryptochromes are phosphorylated in response to blue light and this is associated with its function and regulation (Shalitin *et al.*, 2002 and 2003). Upon blue light illumination, the FADH· may trigger the phosphotransfer from ATP to serine residues on the CCT. The phosphorylated CCT is therefore negatively charged and will be repelled from the negatively charged PHR surface which is interacting with CCT, leading to the conformational changes in cryptochrome. It is still to be verified whether the photo-excited FAD can trigger a phosphotransfer from ATP to the CCT. Proteolysis experiments on full-length AtCry1 in the dark and after illumination suggest that the CCT domain undergoes a light-induced order-to-disorder transition particularly in the serine-rich region of the DAS motif which is conserved in the N-terminus of Cry3 and C-terminus of Cry1 and Cry2 (**Fig 1.21**)
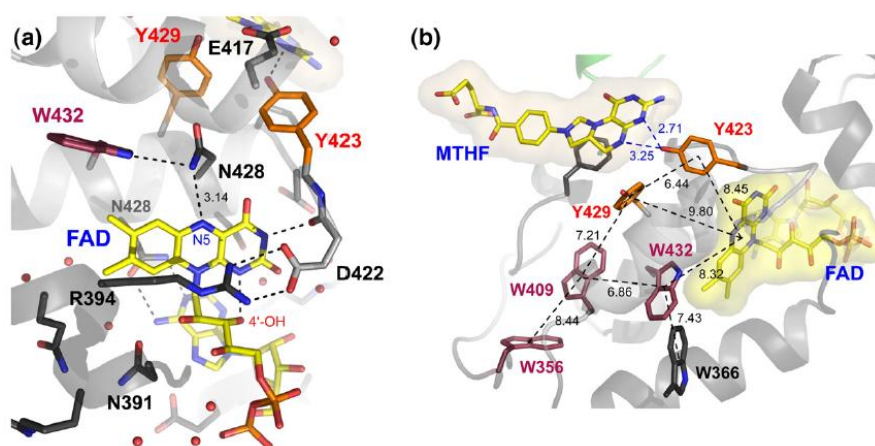
Fig 1.21 The putative electron transfer pathway leading from and to the FAD binding site upon blue light illumination (taken from Klar *et al.*, 2007).

1.11.4 Orange carotenoid protein (OCP) and fluorescence recovery protein (FRP)

(a) Discovery of OCP

In response to the variations of light quality and intensity, photosynthetic organisms have developed protective mechanisms for adaptation and survival including rapid (the conformational changes and reorganization of the photosynthetic apparatus) and slow processes (synthesis and/or degradation of protein complexes). In plants and green algae, LHCII is composed of six hydrophobic membrane-intrinsic polypeptides of the LHC family, with non-covalently bound chlorophylls and carotenoids (Ballottari *et al.,* 2012). Carotenoids play important roles in light harvesting, energy transfer as well as photoprotection by interacting with the triplet state of chlorophyll (Chl^T) or the triplet state of

BChla (BChl^T) to prevent singlet oxygen to be formed. In cyanobacteria and red algae, the phycobilisome, composed of a core and rods surrounding the core, attaches to the outer surface of the thylakoid membranes and plays the role of light harvesting.

A 35 kDa soluble orange carotenoid protein (OCP) was discovered in *Arthrospira maxima*, *Microcystis aeruginosa*, and *Aphanizomenon flos-aquae* in 1981 (Holt and Krogmann) and then purified from cyanobacterium *Synechocystis* PCC 6803 (Wu and Krogmann, 1997). The OCP was shown to play an important role in photoprotection when light irradiance exceeds what can be utilized in photosynthesis. The overall fold of the OCP is composed of two domains, **an all-helical N-terminal domain and an α/β C-terminal domain** (**Fig 1.22**). The carotenoid (3'-hydroxyechinenone, (3'-hECN)) spans both domains, embedded inside in *all-trans* configuration with its keto terminus nestled within the C-terminus of OCP. The gene that encodes for OCP in *Synechocystis* is *slr1963*, which is constitutively expressed. The amino acid sequence of the OCP homologs is highly conserved. The marine *Synechococcus* OCP sequences are very similar to each other, with 77%-95% identity.
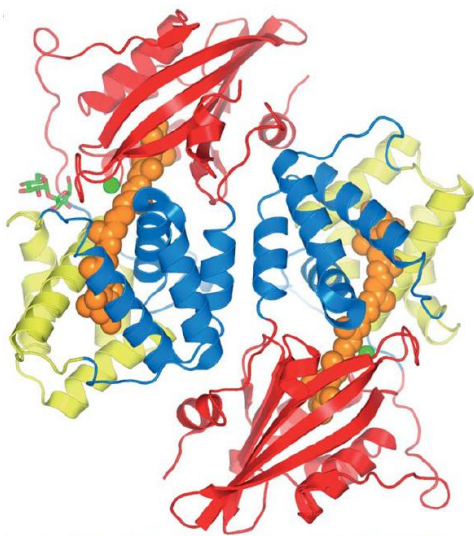
Fig 1.22 The dimer of OCP. For each monomer, the two α-bundles of the N-terminal domain are indicated in yellow and blue, and the C-terminal domain is indicated in red (taken from Kerfeld, 2004).

Upon high blue light irradiance, the OCP was discovered to be in a red form (RCP) which is a proteolytic fragment of OCP lacking the first 15 amino acids at the N-terminus and the last 160 amino acids at the C-terminus (Kerfeld, 2004). Spectroscopically, RCP shows a 30 nm bathochromic shift and plays an important role in photoprotection (**Fig 1.23**).
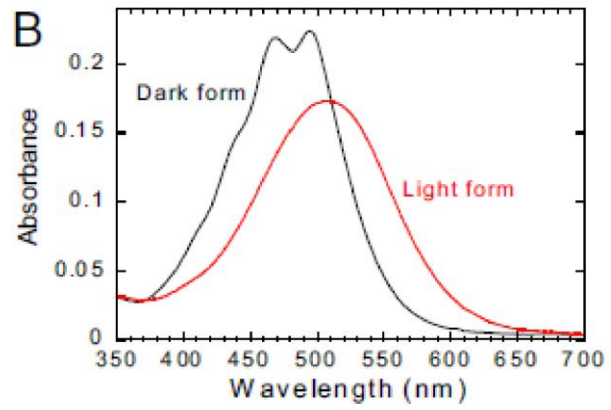
Fig 1.23 3-hECN in organic solvent, OCP and RCP are shown in (A) (taken from Kerfeld, 2004). The spectra of OCP (dark form) and RCP (light form) are also shown (B) (taken from Wilson *et al.*, 2008).

(b) The conserved residues in OCP

Light induces the conformational changes of the carotenoid (3'-hECN) in the OCP which leads to the conjugated length of 3'-hECN to be increased by about one conjugated bond and a distorted but more planar structure (**Fig 1.24**) (Polívka *et al.*, 2005).
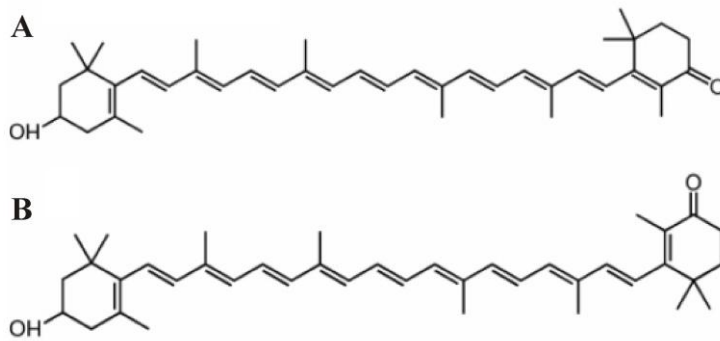
Fig 1.24 3-hECN in OCP (A) and in solution (B) (taken from Polívka *et al.*, 2005).

The conformational changes in the hECN lead to the conformational changes in the OCP through the conserved interacting residues. Some hydrogen bonds between the carotenoid and the protein are conserved. The conserved residues W41, Y44, and W110 at the N-terminus interact with the hydroxyl ring of the 3'-hECN, whereas the conserved residues W288 and Y201 at the C-terminal domain are deep within the hydrophobic pocket, making hydrogen bonds with the keto oxygen atom of the 3'-hECN (**Fig 1.25**) (Wilson *et al.*, 2011).
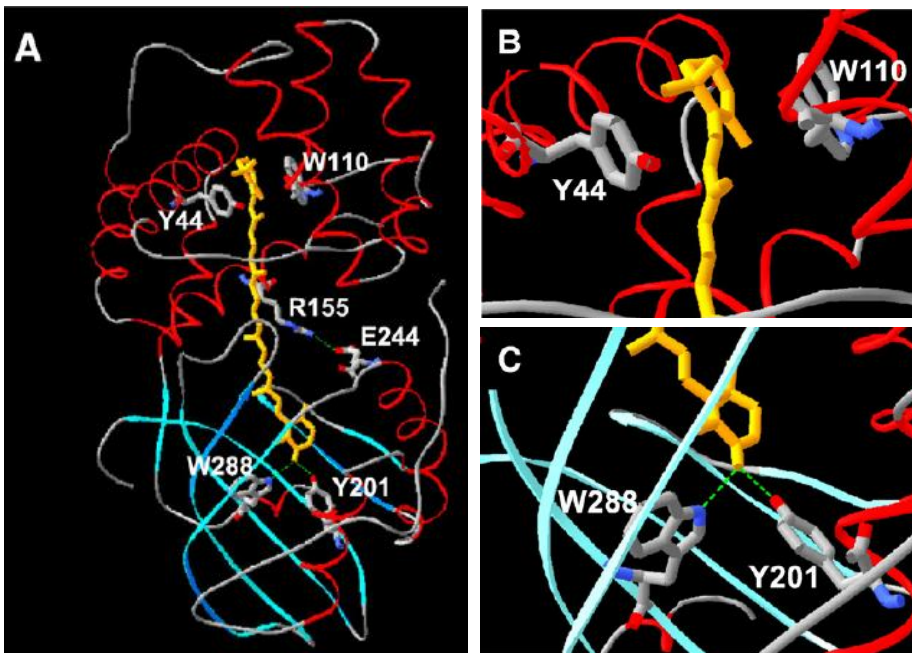
Fig 1.25 The conserved residues in OCP interacting with the carotenoid (taken from Wilson *et al.*, 2011).

(c) Discovery of the fluorescence recovery protein (FRP) and the proposed photoprotective mechanism mediated by OCP and FRP

A 13-14 kDa protein, encoded by *slr1964*, is essential for the conversion from OCP^R to OCP^O and the fluorescence recovery in the dark or at low light irradiation (Boulay *et al.*, 2010). This protein was called fluorescence recovery protein (FRP) because the *Synechocystis* cells cannot recover the lost fluorescence without this protein under decreased light irradiation.

The conformational changes in the N- and C-terminus of OCP make the protein more accessible to proteolysis and result in a red, active form (OCP^R) (Kerfeld, 2004). The C-terminal modification in the OCP^R is necessary for binding to phycobilisome and fluorescence quenching (**Fig 1.26**). The N-terminal modification in the OCP^R enables the binding of FRP at higher affinity, which is followed by the release of OCP^R from the phycobilisome, and the conversion from OCP^R to OCP^O. Subsequently, the fluorescence quenching due to OCP^R is recovered (Boulay *et al.*, 2010).
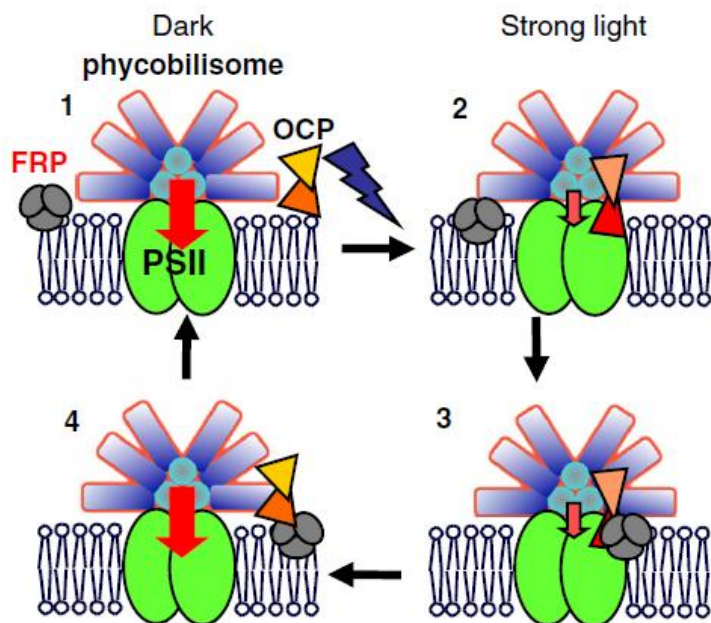
Fig 1.26 The photoprotective mechanism mediated by OCP and FRP (taken from Kirilovsky and Kerfeld, 2012).

In the following, a lycopene-producing *R. rubrum* mutant SLYC18 has been constructed and characterized. I also describe the construction and characterization of the β-carotene-producing mutant SWGK46Y. Further, I describe a secondary regulatory mutation of SWGK46Y which causes a green phenotype SWGK46G to be observed. The reversible SWGK46Y to SWGK46G transition is also examined, which might relate to a putative blue-light induced interaction of the lycopene β-cyclase with carotenoid biosynthesis enzyme, causing down-regulation of the carotenoid production. Below, my PhD work will be presented in the following order: Materials and Methods (Chapter 2), lycopene expression in *R. rubrum* (Chapter 3), β-carotene expression in *R. rubrum* (Chapter 4), the SWGK46 yellow-to-green phenotype transition (Chapter 5), bioinformatic analysis of the *A. thaliana* lycopene β-cyclase and known flavoproteins (Chapter 6) and Discussion (Chapter 7).

Chapter 2   Materials and methods

In this chapter, **2.14-2.16** are taken from Wang *et al.* (2012).

2.1. Media

2.1.1. M-medium (Sistrom medium A) (Sistrom, 1960)

Composition		Conc. g/litre or ml/litre	/litre M-medium	/litre M2S-medium	/litre M2SF-medium
M1	KH_2PO_4/ K_2HPO_4	148 222	10 ml	20 ml	20 ml
M2	Succinate KOH	236 224	10 ml	–	–
	Succinate Adjust pH with 25% NH_4OH	236	–	20 ml	20 ml
M3	NH_4Cl	214	2 ml	–	–
M4	Nitrilotriacetic acid	20	10 ml	10 ml	10 ml
	Aspartic acid	4			
	Glutamic acid	10			
	KOH	22			
	$MgSO_4$	28.8			
	$FeSO_4 \cdot 7H_2O$	0.2			
	Nicotinic acid	0.1			
	Thiamine	0.05			
	Biotin	0.002			
	$CaCl_2 \cdot 2H_2O$	6.6			
	Hutner´s supplement	20 ml			
	2M HEPES (adjust with NH_4OH to pH 6.8)	–	–	20 ml	20 ml
M5	NaCl	120	1ml	1ml	1ml
M6	p-Aminobenzoic acid	2.85	1ml	1ml	1ml
	30% Fructose (sterile)	–	–	–	10 ml

Hutner´s supplement for M-medium

Chemical	Conc. g/litre
EDTA	50
$ZnSO_4·7H_2O$	22
H_3BO_3	11.4
$MnCl_3·4H_2O$	5.1
$FeSO_4·7H_2O$	5
$CoCl_2·6H_2O$	1.6
$CuSO_4·5H_2O$	1.1
$(NH_4)_6Mo_7O_{24}·4H_2O$	1.1

2.1.2. Luria-Bertani (LB) medium (Sambrook and Russell, 2001)

10g Tryptone

5g Yeast extract

10g NaCl

Dissolved in 1 litre H_2O obtained from the Millipore® water purification operator, autoclave at 121°C for 20 min.

2.2. Chemicals

Chemicals used in this study were from Fluka or Sigma, analytical grade. Organic solvents were HPLC grade.

2.3. Plates and slopes

1-litre M-or LB medium was made according to the protocols mentioned above, then 500 ml aliquots were distributed into two 1-litre Schott-bottles. For M-agar, 7.7 g agar was added and autoclaved at 121°C for 15 min (others for 20 min). Antibiotics were added after the autoclaved medium cooled to around 40 °C.

2.4. Growth conditions

E. coli cultures were grown in LB medium at 37°C (Sambrook and Russell, 2001). Antibiotics were added as required at the following concentrations: ampicillin (sodium salt), 100 μg/ml; kanamycin sulphate, 50 μg/ml; tetracycline-HCl, 10 μg/ml. *R. rubrum* strains were inoculated into M-medium (Sistrom, 1960) and cultivated phototrophically in closed bottles (Pyrex) at 30°C. For semi-aerobic growth, *R. rubrum* strains were cultivated in modified M-medium: either M2S medium (containing 40 mM NH_4^+-succinate as carbon source), M2SF medium (containing 40 mM NH_4^+-succinate and 16.7 mM (0.3%) fructose as carbon source) or M2SF$^+$ medium (containing 60 mM NH_4^+-succinate and 111.3 mM (2%) fructose as carbon source) (Ghosh *et al*., 1994). Small-scale growth experiments were performed in 100-ml cultures whereas medium-scale experiments were

performed using 3-litre cultures. For growth curves using either M2SF or M2SF$^+$ media, *R. rubrum* strains were grown chemoheterotrophically in the dark in 250-ml baffled Erlenmeyer flasks containing 100-ml medium with a 4-ml inoculum (obtained from anaerobically grown late exponential phase cultures) at 30°C, with shaking at 150 rpm (2-cm throw). The optical densities, A_{660} and A_{882}, were measured using a single-beam spectrophotometer with a 4-mm path-length cuvette. In a control experiment, we ascertained that the optical densities measured in a 4-mm cuvette correlated linearly with dry weight up to a value of 1.4. Therefore, cell cultures showing a high OD (> 1.4) were diluted prior to the measurement, so as not to exceed this value.

2.5. Strain purification

Strains were plated out as a 3-fold streak several times until a clear pure single colony was obtained. In general, replated cultures were used to inoculate a slope. After usually three days, the slope was filled up, allowed to grow either photosynthetically or semiaerobically and reinoculated to obtain a 100 ml culture.

2.6. Strain storage

Glycerol cultures were used to store strains for long periods. 10 ml fresh culture was pelleted by centrifugation at 4000 rpm (1,940×g) and 4°C for 10 min. The supernatant was removed and the pellet was resuspended with 1 ml M-medium (*R. rubrum*) or 1 ml LB medium (*E. coli*). The suspended pellet was transferred to a cryotube and 1 ml sterile 87% glycerol was added. The glycerol was mixed thoroughly, and the glycerol culture was left in ice for 10 min. Finally, the cryotube was frozen with liquid N_2 and transferred to the -80°C freezer.

2.7. Vector cloning

2.7.1. Plasmid preparation

For minipreps, the "High Pure Plasmid Isolation Kit" (Roche) was used to isolate plasmid DNA. For large amounts of DNA or for low copy plasmids, Q-20 tips (Qiagen) were used to purify the plasmid DNA.

2.7.2. Restriction enzyme digestion

Restriction enzyme digestion was performed in a 20 μl volume, with 1× restriction digestion buffer, 10-20 units of enzyme (1-2 μl) at 37°C (few exceptions, according to the New England Biolabs (NEB) catalog). To inactivate the enzyme, the digested DNA sample was either heated at 70°C for 10 min or at 85°C for 30 min, according to the NEB recommendation.

2.7.3. Agarose gel electrophoresis

During vector cloning, 0.8% agarose analytical gel electrophoresis was used for obtaining fragment sizes after digestion with specific restriction enzymes.

2.7.4. λ-DNA

λ-DNA digested by HindIII was used as DNA marker for size and mass. Table 1.1 shows the bands and their corresponding size and mass weight.

λ/HindIII (kb)	23.13	9.42	6.56	4.36	2.32	2.03	0.56
Mass weight (ng)	152.23	62.15	43.28	28.77	15.31	13.39	3.69
Distance from slot to band (cm)	1.55	2.05	2.35	2.75	3.40	3.55	4.75

Table 2.1 λ/HindIII DNA marker. Seven fragments were obtained after λ-DNA was digested with HindIII, which were used as DNA marker for size and mass.

2.7.5. Preparative enzyme digestion and preparative gels.

1.5-2 µg DNA was digested in a 100 µl volume, with 1× restriction digestion buffer complemented with 1× bovine serum albumin (BSA). Digestion was started with 30-60 U restriction enzyme and lasted for 1 h at 37°C. Another 20-40 U enzyme was added for another 1 h digestion at 37°C. Usually 20-60 ng DNA was removed for analytical gel electrophoresis. If the DNA has been well digested, 20 µl loading buffer was added to the rest of the well-digested DNA. The mixture was then loaded onto a preparative gel which ran at 15V for 12h at 4°C. If the analytical gel showed insufficient digestion, another 1-2 µl enzyme were added to the sample to incubate at 37°C overnight and reanalyzed. The desired band with the right size was cut out at the end of preparative gel and was put into a 2 ml Eppendorf tube and weighed. Usually the "QIAquick gel extraction Kit" (Qiagen) was used for DNA extraction.

2.7.6. DNA precipitation

To precipitate DNA, 1/10 volume (vol) 3M sodium acetate (NaOAc) and 2.5 vol 100% EtOH were added to 1 vol DNA, mixed and left at -20°C overnight. DNA was pelleted by centrifugation at 13,000 rpm (16,700×g) and 4°C for 16 min. The supernatant was removed and the DNA pellet was washed with 200 µl 70% EtOH, then recentrifuged again at 13,000 rpm (16,700×g) and 4°C for 12 min. Finally, after the removal of the supernatant, the DNA pellet was dried in a dessicator for 3 min. Generally, DNA was dissolved in TE for further use or in H_2O for DNA sequencing.

2.7.7. Phenol/chloroform extraction.

1 vol of phenol/dichloromethane (DCM)/TE was added to 1 vol of DNA. After thoroughly vortexing, the mixture was centrifuged at 13,000 rpm (16,700×g) and RT for 10 min, and the DNA upper phase was washed with 1 vol of chloroform/TE. After centrifugation, the DNA upper phase was precipitated as described before.

2.7.8. Blunting and polishing.

5' restriction enzyme overhangs were blunted with any buffer with $1\times$ dNTPs (0.2 mM), 5 U Klenow (NEB) for 20 min at 21°C. 3' restriction enzyme overhangs were polished using 3 U T4 DNA polymerase (NEB) in any buffer with $1\times$ dNTPs for 15 min at 12 °C.

2.7.9. Dephosphorylation

Generally, phenol/chloroform extraction was performed prior to dephosphorylation. DNA was precipitated as described before and the DNA pellet was dried for 3 min and redissolved in 20 μl TE. DNA dephosphorylation was performed in a 50 μl volume with $1\times$dephosphorylation buffer, 2 U calf intestinal phosphatase (CIP (NEB)) at 37°C for 30 min which was followed by another 30 min at 55°C with the addition of another 2 U CIP.

2.7.10. Linker concentration determination

The linker concentration was calculated according to its measured OD_{260} and molar absorbance at 260 nm. For one example:

Xho-Mlu-Pst-Sal linkers 5´ TCGAGACGCGTCCTGCAG 3´ (linker 1)

3´CTGCGCAGGACGTCAGCT 5´ (linker2)

The molar extinction coefficients for the bases are listed as follows: $(cm^{-1}M^{-1})$

A: 44.52×10^3, T: 27.72×10^3, G: 73.68×10^3, C: 47.40×10^3

The absorbance per mole of the linker1, which contains 3A, 3T, 6G and 6C, was calculated as:

$(3\times44.52+3\times27.72+6\times73.68+6\times47.4)\times10^3=1.93\times10^5$

Thus the absorbance of 1 μM of this linker at 260 nm is 0.19. Practically, the OD_{260} of 5 μl oligo diluted with 1 ml TE buffer was measured.

2.7.11. Linker phosphorylation.

500 fmol linker was phosphorylated in a 30 μl volume using $1\times$ T4 DNA ligase buffer (NEB), 10 U of phosphonucleotide kinase (NEB) at 37°C for 15 min, which was subsequently inactivated at 70°C for 10 min. Then, the sample was cooled in ice for immediate use.

2.7.12. Ligation.

Ligation using 26-30 ng DNA was performed in a 20 μl volume with $1\times$ T4 DNA ligase buffer, 20 U T4 DNA ligase (NEB). For blunt-end ligation, the molar vector-to-insert ratio ranged from 1:1 to 4:1, and the ligation was performed at 23°C overnight. For sticky-end ligation, the molar vector-to-insert ratio ranged from 1:1 to 1:4, and the ligation was performed at 17°C overnight.

2.7.13. Site-directed mutagenesis.

The "QuikChange® II site-directed mutagenesis kit" (Stratagene) was used for removal of the PstI site of pBsWGE5 by site-directed mutagenesis. The Sel-ScaI-HpaI mutagenesis oligo (29 bp, 5' TGACTGGTGAGTAGTTAACCAAGTCATTC 3') was used as selection primer (changed ScaI to HpaI) and the WGE5-PstI oligo (27 bp, 5' CTTGGCCGATCTGTAGGAATTCGATAT 3') was

used as mutagenesis primer. Both primers were phosphorylated prior to the annealing step which was performed in a 20 µl volume containing 360 ng double stranded DNA, 100 ng Sel-ScaI-HpaI primer (in 2 µl), 100 ng WGE5-PstI primer (in 2 µl) and 4.1 µl buffer 1, 100°C for 3 min, cooled immediately at 0 °C (ice with EtOH), and followed by 70°C for 1 h. To continue the synthesis of mutant DNA strand, the 20µl annealing system was enlarged to a final 50 µl system by adding 5 µl MgCl$_2$ solution, 5 µl 10× dNTPs, 1 µl T4 DNA polymerase (NEB), 1 µl T4 DNA ligase (NEB), 1 µl T4 gene 32 protein and H$_2$O, 37°C for 1 h. After DNA synthesis, restriction enzyme digestion with ScaI and analytical gel electrophoresis were performed to check whether there was supercoiled DNA which was not cut by ScaI and might be the candidate. Phenol/chloroform extraction was performed and the DNA was redigested with ScaI to remove the background. The digested DNA was transformed into *E. coli* strain Top10*mut*S (which is a *mut*S derivative of Top10) (Stratagene), and plated out. 12 colonies were picked for miniprep and restriction digestion to check whether they were real candidates.

2.8. Competent cells.

Competent cells of the *E. coli* strains XL1 MR, RR28, and Top10mutS were made by the CaCl$_2$ method. The appropriate strain was plated out directly from the glycerol culture and grew at 37°C overnight. The colonies from the plate were then used to inoculate a 10ml pre-culture which was grown at 37°C for 12h shaking at 340 rpm. The pre-culture was inoculated into two 100 ml flasks containing 50 ml LB medium and grown at 37°C shaking at 290 rpm. After about 2 h, the cultures achieved a OD around 0.5 at 420 nm (path-length: 1cm), 20 ml aliquots were distributed into four sterile polycarbonate tubes. After centrifugation at 1,940×g and 4°C for 10 min, the supernatant was removed and the cell pellet was resuspended with 5 ml cold and sterile 100 mM CaCl$_2$. After 2 h in ice, the resuspended cells were recentrifuged at 1,940×g and 4°C for 10 min. The supernatant was removed and the cell pellet was resuspended in 2 ml cold and sterile 50 mM CaCl$_2$ containing 20% glycerol. 200 µl aliquots of the suspended culture were distributed into cold 1.5 ml eppendorf tubes and then left in ice overnight and transferred to the -80°C freezer. The overnight incubation allows a somewhat higher competence to be achieved.

2.9. Transformation.

Competent cells were removed from the -80°C freezer and left in ice for 10 min to thaw out. After mixing and resuspending gently, 20 µl of the plasmid DNA was added and the cells were left in ice for 30 min, after which a 3 min heat shock was performed at 42°C. The samples were cooled at RT for 10 min and 1 ml LB medium was added and the cells were then incubated at 37 °C for 1 h. Finally, samples were plated out. The frequency of competence was measured using 1µl (5.3 ng) pUC19 DNA according to the procedure described above. Finally, 50 µl, 200 µl, and the rest of the

sample were plated out on three LB agar plates containing 200 µg/ml ampicillin. The number of colonies on a 50 µl plate was counted and the frequency of competence was calculated.

2.10. Conjugation.

The OD_{660} of fresh *E. coli* and *R. rubrum* culture were measured to calculate how many ml would give 5×10^8 cfu for conjugation. It had been determined previously (Sägesser, 1992) that 5×10^8 cfu/ml equals to A_{660} (1cm) = 1.4. To remove the antibiotics in the medium, *E. coli* and *R. rubrum* strains were washed twice respectively with 1 ml 0.9% NaCl and 1 ml M-medium. After centrifugation at 13,000 rpm (16,700×g) for 2 min, the supernatant was removed and the cells were resuspended in 1 ml 0.9% NaCl. For triparental mating, pRK2013/RR28 strain and the plasmid strain were mixed with *R. rubrum* cells which were adjusted to 5×10^8 cfu, the ratio of *R. rubrum* to *E. coli* was 100:1. After mixing and centrifugation at 13,000 rpm (16,700×g) and RT for 2 min, the supernatant was removed and the cell pellet was resuspended in 17 µl 0.9% NaCl. The resuspended cells were pipetted onto a nitrocellulose filter on a LB plate and left in the dark to dry for 15 min at RT, and then transferred to the 30°C oven. After 5 h incubation, the filter was transferred to a 1.5 ml eppendorf tube and washed with1 ml 0.9% NaCl. After removal of the filter, 200 µl of the resuspended cells were replated out on plates with appropriate antibiotics. The plates were placed in an anaerobic jar after a short period of drying and incubation as described above.

2.11. Preparation of samples for DNA sequencing.

DNA samples were provided at a concentration of 30-100 ng/ml, according to the instructions of the GATC company. Usually, 1.5 µg DNA was precipitated and dissolved in 30 µl H_2O, mixed thoroughly. DNA sequencing primers were provided at the concentration of 10 pmol/µl.

2.12. PCR.

In this study, the "Phusion high fidelity PCR master mix" was used for PCR. A 50 µl volume was employed containing 120 ng DNA template, 2 µl primer (10 pmol), 25 µl 2× Phusion Master Mix, 2 µl 50% DMSO and H_2O. The mixture was overlaid with 50 µl sterile paraffin oil prior to PCR.

The PCR cycle was designed as follows:

Procedure	T (°C)	Time	Cycles
Initial denaturation	98	30 sec	1
Denaturation	98	5-10 sec	
Annealing	56	10-30 sec	30
Extension	72	30 sec/kb	
Final extension	72	8 min	
Holding	4	99 h	

A thermometer was used to equilibrate the PCR thermocycler, and the PCR cycle was programmed as follows to achieve the designed cycles.

Program 1: 99°C, 2 min, link to program 3

Program 3: 99°C, 30 sec

       57.7 °C, 1 min 10 sec

       73.9 °C, 1 min 20 sec

30 cycles, link to program 7

Program 7: 73.9 °C, 8 min, link to program 99

Program 99: 4°C, 99 h

After PCR, the "QIAquick PCR purification Kit" (Qiagen) was used to purify the DNA fragments for DNA sequencing.

2.13. Hybridization (The compositions of the solutions used in hybridization are listed in the Appendix 1)

(a) Chromosomal DNA isolation

70ml fresh culture was pelleted down by centrifugation at 2450×g, 4°C for 10 min. The supernatant was removed and the pellet was resuspended in 7 ml solution 1 (50 mM Tris-HCl pH 8, 20 mM EDTA), after which 0.35 ml lysozyme (20 mg/ml) was added and incubated at 37°C for 50 min. Subsequently, 20 µl RNAase (DNAase free) was added and incubated at 37°C for 20 min, followed by incubation at 37°C for one hour after addition of 0.35 ml pronase E (20 mg/ml). The pronase E had been pre-incubated at 37°C for 1 h to remove traces of DNAase. 1.05 ml of 10% SDS was added to the sample, which was then incubated at 37°C for 15 min. To separate the chromosomal DNA from proteins and other contaminants, 1 vol of phenol/TE (8 ml) was added and mixed gently for 20 min. The DNA/phenol mixture was then centrifuged at 2,450×g and 4°C for 15 min. The supernatant was removed and extracted by 1 vol (8 ml) phenol/TE. After the centrifugation at 2,450×g and 4°C for 15 min, the upper DNA/H$_2$O phase mixed with 1 vol (8 ml) chloroform/TE

gently, and was centrifuged at 2,450×g and 4°C for 15 min. The chromosomal DNA (supernatant) was precipitated and dissolved in 1ml fresh TE.

(b) Determination of the chromosomal DNA concentration

Chromosomal DNA obtained as above was diluted 200-fold before the measurement of the OD_{260}. The concentration of DNA was calculated based upon conversion factor that an OD_{260} (1 cm)=1 is given by 50 µg/ml dsDNA.

(c) Chromosomal DNA digestion prior to hybridization

Prior to hybridization, chromosomal DNA (4 µg) was digested by several restriction enzymes in a 100 µl volume as described above. A 10 µl aliquot of the digested DNA was loaded onto an analytical gel to check whether the digestion was complete. After restriction digestion, the DNA was extracted with phenol/chloroform, and then precipitated with EtOH as above. After drying, the DNA was dissolved in 40 µl TE, vortexed thoroughly. Half of the DNA was loaded onto a gel which was then run at 4 °C and 15 V for 12 h. λDNA digested with HindIII was used as a size marker, labeled with digoxigenin (DIG, Roche) and used as a marker. The method of DNA labeling will be mentioned in "hybridization probe preparation" in detail.

(d) Processes before DNA transfer

The hybridization gel was cut above the slots and soaked once in 200 ml 0.25 M HCl for 15 min for depurination and then rinsed with sterile H_2O. Subsequently, denaturation was performed twice by 10 min incubation in 200 ml 1.5 M NaCl/0.5 M NaOH, and then the gel was rinsed with sterile H_2O. The gel was neutralized by incubating twice in 200 ml 1.5 M NaCl/0.5 M TrisHCl (pH 7.4) for 15 min.

(e) Transfer of DNA from gel to nylon membrane (Hybond-N, Amersham Biosciences)

DNA transfer was carried out on a clean glass plate which was placed on a clean plastic tray. A piece of 3 MM chromatography paper (Whatman) was placed on the top of the glass plate and was wetted by blotting buffer (10×SSC (Appendix 1)) without immersing the glass plate. Another 3 pieces of 3 MM chromatography paper were saturated with 10×SSC and placed on the top. Finally, the gel was placed on with the DNA side facing upwards. A wet nylon membrane, which was saturated with 2×SSC beforehand, was placed on top of the gel. 3 pieces of 3 MM chromatography paper, which were pre-saturated with 2×SSC, were placed onto the membrane, followed by a stack of gel-blotting paper (GB004, Schleicher & Schuell) and a stack of paper towels (2.5 cm). A light weight was placed on the top to keep all layers compressed.

(f) Hybridization probes preparation

The DNA probes for hybridization were labeled with DIG as follows. Prior to labeling, 100 ng probe DNA was heated at 95°C for 10 min and cooled in ice for 5 min. Labeling was performed in a 20 µl reaction volume containing 100 ng probe DNA, 2 µl hexanucleotide mix (Roche), 2 µl dNTPs

labeling mix (10×, which contains 1 mM dATP, 1 mM dGTP, 1 mM dCTP, 0.65 mM dTTP, 0.35 mM Dig-11-dUTP (Roche)), 1 µl specially pure Klenow fragment (Roche). The mixture was incubated at 37°C for 1 hour which was then followed by overnight incubation at 37°C. Afterwards, 2 µl 0.2 M EDTA pH 8 and 2 µl 4 M LiCl were added and mixed, which was followed by addition of 60 µl EtOH and incubation at 70°C for 30 min or -20°C for 2 h. Labeled probes were pelleted and washed with 200 µl 70% EtOH. After drying, the labeled probes were redissolved in 60 µl fresh TE.

(g) Prehybridization and hybridization

The filter-bound DNA was immobilized by exposure to high intensity UV light for 3 min. Subsequently, the membrane filter was transferred into a hybridization tube, 20 ml prehybridization solution (Appendix 1) was added and then incubated at 68-70°C for 1 h. The prehybridization solution was retrieved and a 3 ml of pre-warmed prehybridization solution, which contained 30 ng labeled DNA probe (pre-heated at 95°C for 10 min), was placed into the tube and hybridized at 70°C for 16 h.

(h) Washing and antibody-AP labeling

The membrane filter was transferred to a clean dish and then washed twice at low stringency with 50 ml 2×SSC/0.1% SDS at RT for 5 min which was followed by twice washing at high stringency with 50 ml 0.1×SSC/0.1% SDS at 70-75°C for 5 min. The membrane filter was then washed twice with 50 ml washing buffer (Appendix 1) at RT for 2 min prior to be washed with 50 ml Buffer 2 (Appendix 1) at RT for 30 min. Subsequently, the filter was transferred to a clean dish which contained 10-20 ml diluted alkaline phosphatase anti-DIG-Fab conjugate solution (Anti-Dig fab fragment, Roche) (Appendix 1), and labeled at RT for 30 min. The filter was washed twice with 50 ml washing buffer at RT for 15 min and then washed once with 10 ml buffer 3 (Appendix 1) at RT for 5 min. Finally, the filter was transferred to a hybridization bag and incubated with 2 ml CDP-Star solution (Roche) (Appendix 1) at RT for 5 min which was followed by 30 sec blotting on 3MM paper prior to be sealed in a new hybridization bag.

(I) Exposure and development

The filter was placed in an autoradiography cassette (CAWO) with intensifying screens and exposed to X-ray film (Super RX Fujifilm, 18×24 cm, Tokyo). The exposure time started with 5 min and then the film was developed by an automatic film developing machine (CP1000, AGFA).

(J) Removal of the AP-Fab conjugate and DIG-labeled probe prior to filter hybridization

The membrane filter was rinsed briefly in sterile H_2O and then washed twice with 0.2 M NaOH/ 0.1% SDS at 37°C for 15 min to remove the DIG-labeled probe. After being rinsed thoroughly with 2 ×SSC for 5 min, the filter could be used for a further hybridization with a second probe.

(To note: methods **2.14**, **2.15**, and **2.16** are taken from my published paper (Wang *et al.*, 2012)

2.14. Absorption spectroscopy

The absorption spectra of intact cells and also extracted carotenoids were determined by using 2-mm path-length cuvettes with a Jasco V-560 UV/VIS spectrophotometer equipped with a photodiode detector for turbid samples. Intact cells were measured after suspension in M- medium containing 80% glycerol. For comparisons of relative amounts of crts in cells, the suspensions to be measured were adjusted to the same optical density (usually A_{660} (2mm path-length cuvette) = 0.5) at 660 nm.

2.15. Biochemical methods

2.15.1. Dry weight determination. Dry weights were determined from 1-2-ml aliquots taken from growing cultures, by first filtering through pre-weighed filters (Versapor®-200, (PALL, USA)), which were then dried in an oven at 80°C for 24 h to achieve constant weight.

2.15.2. Protein determination. The total protein of cell culture aliquots was determined by a modified Lowry procedure using bovine serum albumin as a standard. Generally, 100 μl or 50 μl aliquots of cell cultures were used for protein determination.

2.15.3. Carotenoid extraction. 2 ml of a stationary phase cell culture were centrifuged briefly and the cell pellets were extracted twice with 1 ml MeOH to remove the bacteriochlorophyll a (BChla). Three sequential hexane extractions were then performed to obtain the carotenoids. The entire extraction process was performed at low light conditions under N_2 to prevent photobleaching and degradation. The MeOH and hexane extracts were initially evaporated under vacuum.

2.15.4. Thin-layer chromatography (TLC). 100 μl samples were dried under N_2, then redissolved in 100 μl pentane/ethanol (100:2 (v/v)) and applied to silica gel plates (DC-Fertigplatten SIL G-25, Merck), which were subsequently developed using pentane/ethanol (100:2 (v/v)) as running solvent.

2.16. Purification of the LH1 complex from the SLYC18 mutant. LH1 complexes were isolated by established procedures (Picorel *et al.*, 1983; Ghosh *et al.*, 1988). Briefly, the ICM fraction, isolated from a crude cell extract by differential centrifugation, was extracted three times using 0.3% (w/v) lauryldimethylamine-N-oxide (LDAO). The first two extractions primarily solubilize the reaction center (RC), and the third extract is highly enriched in the LH1 complex. The LDAO fraction was diluted to 0.1% LDAO and then dialyzed extensively, followed by DEAE-cellulose chromatography using 50 mM TrisHCl as the initial equilibration buffer. Fractions eluted with the same buffer in the presence of 0.2% and 0.3% (w/v) n-octyl-β-D-glucoside (βOG), respectively, in the presence of 250 mM NaCl contained highly purified LH1 complexes. The purified LH1 complexes were dialyzed extensively to remove detergent, pelleted by ultracentrifugation, and then stored at -85°C. For the determination of the total pigments, the purified LH1 complexes (1 vol.) were extracted with 3 vol. methylene chloride/methanol (MeOH) (1:2 (v/v)) for 20 min, and then

phase separation was induced by the further addition of H_2O/MeOH (1.8 vol./1.8 vol.). Phase separation was completed after a short centrifugal step (3000xg). The total organic solvent-soluble pigments were obtained from the lower fraction and then dried under N_2. The procedure was repeated three times and all fractions were pooled. The dried extract was then dissolved in diethyl ether prior to spectral measurement.

Chapter 3

HIGH-LEVEL PRODUCTION OF THE INDUSTRIAL PRODUCT, LYCOPENE, USING THE PHOTOSYNTHETIC BACTERIUM, *RHODOSPIRILLUM RUBRUM*

by

Guo-Shu Wang[1], Hartmut Grammel[2,3], Khaled Abou-Aisha[1,4], Rudolf Sägesser[1], and Robin Ghosh[1*]

[1]Department of Bioenergetics, Institute of Biology, University of Stuttgart, Stuttgart, Germany, D-70569 Stuttgart, Germany

[2]Max-Planck-Institute for the Dynamics of Complex Technical Systems, Sandtorstr.1, D-39016 Magdeburg, Germany

[3]Present address: Biberach University of Applied Science, Hubertus-Liebrecht-Str. 35, D-88400 Biberach, Germany

[4]Present Address: Department of Microbiology and Biotechnology, The German University in Cairo, Cairo, Egypt

*Corresponding author: Robin Ghosh

Mailing address: Department of Bioenergetics, Institute of Biology, University of Stuttgart, Pfaffenwaldring 57, D-70569 Stuttgart, Germany.

Phone: +49-711-68565047; Fax: +49-711-68565096; Email: robin.ghosh@bio.uni-stuttgart.de

Running title: Production of lycopene in *Rhodospirillum rubrum*

Abstract

The biosynthesis of the major carotenoid, spirilloxanthin, of the purple non-sulphur bacterium, *Rhodospirillum rubrum*, is thought to occur via a linear pathway proceeding through phytoene and later lycopene as intermediates. This assumption is based solely on early chemical evidence (**Davies, B. H.** 1970. Biochem. J. **116**:93-99). In most purple bacteria, the desaturation of phytoene, catalyzed by the enzyme phytoene desaturase (CrtI) leads to neurosporene, involving only three dehydrogenation steps, and not four as in the case of lycopene. Here we show that the chromosomal insertion of a kanamycin resistance cassette into the *crtC-crtD* region of the partial carotenoid gene cluster, whose gene products are responsible for the downstream processing of lycopene, leads to the accumulation of the latter as the major carotenoid. We provide spectroscopic and biochemical evidence that *in vivo*, lycopene is incorporated into the light-harvesting complex 1 as efficiently as the methoxylated carotenoids spirilloxanthin (in the wild-type) and 3,4,3',4'-tetrahydrospirilloxanthin (in a *crtD⁻* mutant), both under semi-aerobic, chemoheterotrophic, as well as photosynthetic, anaerobic conditions. Quantitative growth experiments under dark, semi-aerobic conditions, using a growth medium for high cell density/high intracellular membrane levels, suitable for conventional industrial production in the absence of light, yielded values for lycopene of up to 2 mg/g cellular dry weight, or up to 15 mg/liter culture. These values are comparable to those of many previously described *Escherichia coli* strains engineered for lycopene production. The study provides the first genetic proof that the *R. rubrum* CrtI produces lycopene exclusively as an end product.

Introduction

The production of the carotenoid (crt) lycopene, which is used industrially as a natural colorant and also as a food additive, using microbial sources, is presently of great interest (29, 35, 46). Until now, most microbial production strategies have mostly employed *Escherichia coli* as a host strain (1, 2, 10, 13, 24, 26, 30, 33-35, 43, 45, 47, 56-58, 64). However, a few studies employing yeasts have also been reported (36, 49, 62). In early studies (46) employing *E. coli* as a host strain, lycopene levels of up to about 0.5 mg/g dry weight (DW) cells have been reported, whereas in *Candida utilis*, after extensive re-engineering of the ergosterol pathway, up to 7.8 mg lycopene/g DW has been achieved (49). More recently, using a combination of rational systems biological design and random screening approaches, lycopene levels of up to 18 mg/g dry weight (DW) have been successfully achieved in non-linearly batch-fed *E. coli* cultures (2). In another approach, the Stephanopoulos group has introduced up to 40 consecutive copies of heterologous carotenoid biosynthesis pathway genes into the *E. coli* chromosome, which allowed up to 14 mg lycopene/g DW to be produced (56). In an alternative approach, Farmer and Liao (10) demonstrated that the carbon flow to lycopene in *E. coli* could be increased by the introduction of a synthetic regulatory circuit derived from the Ntr regulon, which senses an intracellular signalling molecule, acetyl phosphate. However, even these latter, very successful examples of lycopene production in *E. coli* have required extensive re-engineering of the organism. In addition, to date, these latter strategies have only been reported for small shake cultures (about 50-250-ml) or small-scale bioreactors.

The implicit rationale for employing *E. coli* as a production strain for a highly hydrophobic terpenoid such as lycopene is not completely clear. Although *E. coli* is the organism of choice due to the wealth of genetic detail that is available for this organism, together with its high growth rate and general robustness, the production of hydrophobic terpenoids in this strain poses several severe problems. First, lycopene biosynthesis requires three unique enzymes, geranylgeranyl pyrophosphate (GGPP) synthase (CrtE), phytoene synthase (CrtB), and phytoene desaturase (CrtI) which are not present in *E. coli* and must be therefore maintained either *in trans* or introduced chromosomally. These types of constructions often perform well at a small- to medium-scale but are frequently prone to stability problems when adapted to the industrial scale. Secondly, GGPP synthase uses the substrate farnesyl pyrophosphate (FPP) to produce GGPP, which is an essential intermediate in the quinone (both ubiquinone and menaquinone) pathways. Thus, a common observation is that increased flux to a recombinant phytoene synthase causes the GGPP pool to be depleted, thus leading to growth problems (37). This problem must be alleviated by complicated genetic strategies such as an extensive re-design of the pyruvate/melavonate pathways using rational and random screening techniques (2, 56), or by employing a synthetic feedback loop using

the metabolite acetyl phosphate as a sensor of energy metabolism (10). The latter approach has the potential difficulty that acetyl phosphate is also a global regulator of metabolism, where many of its roles, particularly in anaerobically operative pathways, have yet to be defined (see (61) for a review). However, the highest yields reported (2, 56) have usually required a fed-batch strategy, which is often difficult to apply during up-scaling. Thirdly, it is probable that the maximal attainable levels of terpenoid (in this case lycopene) are limited to the available volume of the cytoplasmic membrane. In fact, the localization of the carotenoid in *E. coli* has never been studied in detail. This factor is probably much more important than often expected, since a considerable perturbation of the natural membrane would be expected to affect electron transport, and thus viability.

An interesting alternative strategy would be to utilize purple photosynthetic bacteria, which naturally overproduce large amounts of carotenoids in specific compartments (e.g. the photosynthetic membrane), for the production of industrially interesting carotenoids. However, only a single attempt with this strategy (see below) has been reported so far (22). The reason for the low activity in this area is probably that a considerable degree of genetic engineering is still mandatory, since, as noted recently by Takaichi (55), the natural carotenoid products in photosynthetic purple bacteria are generally distinct from those of plants, and that processes requiring controlled light conditions are not easily accessible for industrial production systems.

In a recent review, Takaichi (55) has classified the carotenoid biosynthesis pathways in purple bacteria into two types: (a) the spirilloxanthin (spx) pathway, which encompasses both the "normal"(Fig. 1) and "unusual" pathways (see(55)) which lead either to spx (e.g. found in *R. rubrum*), or spheroidene (e.g. in *Rhodobacter capsulatus* (3, 4, 63) and *Rhodobacter sphaeroides* (22)), and (b) the okenone pathway (e.g. in *Allochromatium oekenii* (48)), which leads to okenone or other ketocarotenoids as final products. For both pathways, the first step unique to carotenoid biosynthesis is the oxidation of phytoene, catalyzed by the enzyme phytoene desaturase (CrtI, see Fig. 1). However, the final product of this reaction varies depending upon the source. Thus, the CrtIs of *Rb. capsulatus* and *Rb. sphaeroides* have been shown to produce exclusively neurosporene as a product (5, 28), whereas the CrtI of *Rubrivivax gelatinosus* oxidizes phytoene to 90% neurosporene and 10% lycopene (21), and the CrtI from the non-photosynthetic *Pantoea* (formerly called *Erwinia*) yields exclusively lycopene (13).

In the single study employing *Rb. sphaeroides* as a host organism (22), the gene encoding CrtI from *Erwinia* was introduced into *Rb. sphaeroides* on a conjugable plasmid *in trans* and the

transconjugant was shown to produce lycopene in the intracytoplasmic membrane (ICM), which contains the photosynthetic apparatus. However, the levels of lycopene reported for semi-aerobic cultures appeared to be extremely low (about 20% of those obtainable with the wild-type strain).

In *R. rubrum*, the product of the CrtI reaction is assumed to be lycopene (55), based upon chemical analysis of carotenoid intermediates performed in early studies (8). So far, however, no genetic or functional proof has been provided to confirm this suggestion. In this exploratory study, we show that a $crtC^- crtD^-$ deletion mutant of *R. rubrum* indeed produces lycopene exclusively as a final product, thus confirming the early proposal of Davies (8). We also show that the lycopene levels attainable in *R. rubrum* are equivalent to those reported for several highly engineered *E. coli* strains reported as "starting points" for "superproducing" *E. coli* strains (2, 10, 56). We also provide biochemical evidence, using purified light-harvesting (LH) 1 complexes, which are localized exclusively in the ICM in vivo, that the lycopene produced is specifically bound to this protein.

Materials and Methods

Growth conditions. Bacterial strains and plasmids are listed in Table 1. *E. coli* cultures were grown in Luria-Bertani medium at 37°C (44). Antibiotics were added as required at the following concentrations: ampicillin (sodium salt), 100 μg/ml; kanamycin sulphate, 50 μg/ml. Initially, *R. rubrum* strains were inoculated into Sistrom medium A (here designated M medium) (51) containing 20 mM potassium succinate as carbon source, and cultivated phototrophically in closed bottles (Pyrex) at 30°C. For semi-aerobic growth, *R. rubrum* strains were cultivated in modified M medium: either M2S medium (containing 40 mM NH_4^+-succinate as carbon source), M2SF medium (containing 40 mM NH_4^+-succinate and 16.7 mM (0.3%) fructose as carbon source) or $M2SF^+$ medium (containing 60 mM NH_4^+-succinate and 111.3 mM (2%) fructose as carbon source) (18). Small-scale growth experiments were performed in 100-ml medium whereas medium-scale experiments were performed using 3-liter cultures.

For growth curves using either M2SF or $M2SF^+$ media, *R. rubrum* strains were grown chemoheterotrophically in the dark in 250-ml baffled Erlenmeyer flasks containing 100-ml medium with a 4-ml inoculum (obtained from anaerobically grown late exponential phase cultures) at 30°C, with shaking at 150 rpm (2-cm throw). The optical densities, A_{660} and A_{882}, were measured using a single-beam spectrophotometer with a 4-mm path-length cuvette. In a control experiment, we ascertained that the optical densities measured in a 4-mm cuvette correlate linearly with dry weight up to a value of 1.4. Therefore, cell cultures showing a high OD (> 1.4) were diluted prior to the measurement, so as not to exceed this value.

Construction of the lycopene-producing mutant, SLYC18. A 6.37 kb EcoRV fragment, containing *crtC*, *crtD*, and *crtE*, was obtained from the cosmid pSC4 (Fig. 2), which was isolated from a cosmid gene bank containing chromosomal DNA from the *R. rubrum* wild-type S1, and which encodes about half of the photosynthetic gene cluster. The EcoRV fragment was subcloned into the SmaI site of pBluescript (pBs) KSII+ to yield the plasmid pBsSGE5 (Fig. 2). Subsequently, a 1.2 kb BstEII fragment was replaced with a 1.5 kb HindIII/SalI blunt-ended fragment, derived from Tn*5*, which contains the *npt* gene encoding neomycin phosphotransferase (32), yielding plasmids pBsSGE5K1 and pBsSGE5K2, with the *npt* gene in the opposite and same orientation as the *crtD* promoter, respectively. The BstEII fragment deletion simultaneously eliminates large regions of both the *crtC* and *crtD* genes, thus leading to their functional arrest. Finally, the inserts of pBsSGE5K1 and pBsSGE5K2 were subcloned into the conjugable suicide vector pSUP202 (50) to yield the plasmids pSUPSGE5K1 and pSUPSGE5K2, respectively, which were subsequently triparentally conjugated with *R. rubrum* S1 (12, 32). Selection of doubly recombinant *npt* interposon mutants was determined by their Kan^{r}Tet^s phenotype. The chromosomal localization of the *npt* gene was confirmed using Southern hybridization (44).

Absorption spectroscopy. The absorption spectra of intact cells and also extracted carotenoids were determined by using 2-mm path-length cuvettes with a Jasco V-560 UV/VIS spectrophotometer equipped with a photodiode detector for turbid samples. Intact cells were measured after suspension in M medium containing 80% glycerol. For comparisons of relative amounts of crts in cells, the suspensions to be measured were adjusted to the same optical density (usually A_{660} (2mm path-length cuvette) = 0.5) at 660 nm.

Biochemical methods

Dry weight determination. Dry weights were determined from 1-2-ml aliquots taken from growing cultures, by first filtering through pre-weighed filters (Versapor®-200, (PALL, USA)), which were then dried in an oven at 80°C for 24 h to achieve constant weight.

Protein determination. The total protein of cell culture aliquots was determined by a modified Lowry procedure (40) using bovine serum albumin as a standard. Generally, 100 μl or 50 μl aliquots of cell cultures were used for protein determination.

Purification of the LH1 complex. LH1 complexes were isolated by established procedures (41). Briefly, the ICM fraction, isolated from a crude cell extract by differential centrifugation, was extracted three times using 0.3% (w/v) lauryldimethylamine-N-oxide (LDAO). The first two extractions primarily solubilize the reaction center (RC), and the third extraction is highly enriched in the LH1 complex. The LDAO fraction was diluted to 0.1% LDAO and then dialyzed extensively, followed by DEAE-cellulose chromatography using 50 mM TrisHCl as the initial equilibration

buffer. Fractions eluted with the same buffer in the presence of 0.2% and 0.3% (w/v) n-octyl-β-D-glucoside (βOG), respectively, in the presence of 250 mM NaCl contained highly purified LH1 complexes. The purified LH1 complexes were dialyzed extensively to remove detergent, pelleted by ultracentrifugation, and then stored at -85°C. For the determination of the total pigments, the purified LH1 complexes (1 vol.) were extracted with 3 vol. methylene chloride/methanol (MeOH) (1:2 (v/v)) for 20 min, and then phase separation was induced by the further addition of H_2O/MeOH (1.8 vol./1.8 vol.). Phase separation was completed after a short centrifugal step (3000xg). The total organic solvent-soluble pigments were obtained from the lower fraction and then dried under N_2. The procedure was repeated three times and all fractions were pooled. The dried extract was then dissolved in diethyl ether prior to spectral measurement.

Carotenoid extraction. 2 ml of a stationary phase cell culture were centrifuged briefly and the cell pellets were extracted twice with 1 ml MeOH to remove the bacteriochlorophyll a (BChla). Three sequential hexane extractions were then performed to obtain the carotenoids (see Supplementary Fig. 1 for comparative spectra in hexane). The entire extraction process was performed at low light conditions under N_2 to prevent photobleaching and degradation. The MeOH and hexane extracts were initially evaporated under vacuum. In some cases, crt extinction coefficients not reported in the literature were determined in this study (see Supplementary Table 1). The extinction coefficients used were $\varepsilon_{1cm}^{488nm} = 140$ mM^{-1}cm^{-1} (spx in petroleum ether), $\varepsilon_{1cm}^{502nm} = 184$ mM^{-1}cm^{-1} (lycopene in diethyl ether) and $\varepsilon_{1cm}^{502nm} = 172$ mM^{-1}cm^{-1} (lycopene in hexane) (see Supplementary Table 1). For BChla, an extinction coefficient of 90 mM^{-1}cm^{-1} (in diethyl ether (52)) was used.

Thin-layer chromatography (TLC). 100 μl samples were dried under N_2, then redissolved in 100 μl pentane/ethanol (100:2 (v/v)) and applied to silica gel plates (DC-Fertigplatten SIL G-25, Merck), which were subsequently developed using pentane/ethanol (100:2 (v/v)) as running solvent.

High performance liquid chromatography-mass spectroscopy (HPLC-MS). Extraction of carotenoid pigments was performed essentially as described above, except that the final carotenoid fraction, following the two initial MeOH extractions, was obtained by a further two sequential extractions with methylene chloride. The determination and identification of carotenoid pigments in cell extracts were conducted using an Agilent 1100 quaternary HPLC system with two detectors - a diode array detector (DAD) and a single-quadrupole mass spectrometer Agilent MSD SL (Agilent, Palo Alto, CA) - coupled in series. For data acquisition and analysis, the Agilent LC/MSD ChemStation software (version Rev.B02.01-SR2) was applied. HPLC separation was achieved on a Grom-Sil 300 ODS 150 × 4 mm I.D., 5 μm column (Alltech Grom GmbH, Rottenburg-Hailfingen, Germany) using solvents A (acetonitrile), B (MeOH:2-propanol (4:1 (v/v)), ammonium acetate (10 mM)), C (H_2O), and D (tetrahydrofuran). The applied elution profile was as follows: starting with

30% A, 20% B, 50% C, a linear gradient to 30% A, 50% B, and 20% D was run within 20 min at a flow rate of 1 ml/min. This condition was maintained for 20 min for eluting all carotenoid pigments. Column temperature was 20°C, controlled by a column thermostat. After passing the DAD flow cell, the HPLC mobile phase was directly introduced into the mass spectrometer via an atmospheric pressure chemical ionization (APCI) ion source. A nitrogen gas generator was used for supplying nitrogen from pressurized air as a drying gas. Nebulizer pressure was 60 psi. The drying gas flow was 5 liter/min and gas temperature was 300°C (vaporizer temperature 325°C). The capillary voltage was 4000 V in both positive and negative modes. A corona current of positive 10 μA, and negative 15 μA was applied. Ions were monitored in the scan mode, covering a mass range from 200 - 1200 mass units with negative polarity. The fragmentor voltage was 100 V. The identity of lycopene was confirmed with a reference standard from Sigma (Sigma-Aldrich, St. Louis, MO).

Results

Isolation of the lycopene-producing strain SLYC18. Transconjugants were only obtained using pSUPSGE5K2, where the *npt* gene is in the same orientation as *crtD*. Approximately 400 transconjugants were isolated using an initial kanamycin selection on Sistrom (M) agar under dark, aerobic conditions. Both purple and brown colonies were observed. An initial spectral analysis (Fig. 3, a detailed analysis is given below) showed the purple colonies to contain spx, whereas the brown colonies exhibited absorption spectra resembling those previously reported for LH1-bound 3,4,3',4'-tetrahydrospirilloxanthin (thspx (14, 27)) in the carotenoid region. All the purple colonies examined had the Kan^{r}Tet^r phenotype, and were therefore probably single recombinants. Double recombinant brown colonies were identified by their Kan^{r}Tet^s phenotype, and one of these, designated SLYC18, was chosen for further analysis.

HPLC analysis of the carotenoids extracted from late-exponential or stationary phase cells of SLYC18 with organic solvents showed a dominant peak, which was confirmed by both spectral and MS analysis to be lycopene (Fig. 4A, B). By comparison, carotenoid extracts obtained from cells of several purple strains were shown to be dominated by spx, showing HPLC/MS profiles corresponding to the wild-type S1 (Fig. 4C, D).

Finally, we confirmed by Southern hybridization (data not shown) that only a single chromosomal insertion of the *npt* gene was present in the chromosome. The mutant could also be complemented (Supplementary Fig. 2A) by the plasmid pRKSGE5, which contains the 6.37 kb EcoRV fragment from the cosmid pSC4, containing *crtC*, *crtD*, and *crtE*. The complemented mutant was purple in color, and grew almost normally (i.e. as the wild-type S1) under anaerobic, photoheterotrophic conditions, as well as under dark, chemoheterotrophic semi-aerobic conditions (see Supplementary

Fig. 2A). The carotenoid spectrum of both total cells and also the extracted pigments of the complemented strain was identical to that observed for the wild-type (Supplementary Fig. 2B (whole cell spectrum), and Supplementary Fig. 3A (extracted pigments). Furthermore, tlc analysis of the extracted pigments of the complemented strain showed only the pink band due to spx, with no trace of the orange lycopene band observed for the extracted pigments of SLYC18 (Supplementary Fig. 3B). We conclude that the complemented cells contain identical spx levels as the wild-type strain S1.

We note, also, in passing, that there appears to be no polar effect of the interposon lesion on the downstream $cobZ_{RR}$ gene. Lesions in this gene lead to a very distinct phenotype (R. Ghosh, manuscript in preparation) - $cobZ_{RR}^-$ mutants are photosynthetically incompetent and excrete a brown BChla precursor in large quantities. By contrast, SLYC18 grows well photosynthetically and produces no brown precursor.

Detailed spectral analysis of the SLYC18 mutant. The absorption spectra of whole cells (Fig. 3) showed, as reported previously for the thspx-expressing mutant, ST4 (27), that the near-infrared (near-IR) Q_y absorption maximum of the LH1 complex is present at 882 nm, which is characteristic for crt-containing LH1 complexes (11, 60). The LH1 peak maximum at 882 nm is easily distinguishable from that of LH1 complexes lacking carotenoids, which occurs at 874 nm (11, 17, 32, 60). The absorption maxima of the observable reaction center (RC) at 750 nm and 802 nm, corresponding to the RC-bound bacteriopheophytin (BPh) and accessory BChla, respectively, are unchanged in both position and relative intensity for all three strains. Although the small absorption peak (at 870 nm) corresponding to the special pair of the RC is not observable in the cell spectrum (since the LH1 peak is dominant), we assume the RC to be intact and functional, since SLYC18 grows normally (i.e. with respect to S1) under photoheterotrophic conditions. Also, the relative intensities of the three crt peaks compared to those (Q_x, Q_y) of BChla are very similar in both the wild-type and the two crt mutants. It is now well-established (6, 17, 41), that spx present in the wild-type strain, is always protein-bound, with about 94% (16 mol spx/mol LH1 complex) bound to the LH1 complex and 6% (1 mol spx/mol RC) bound to the RC. For practical purposes, therefore, the observed spectra in Fig. 3 correspond primarily to that of the LH1 complex. In the case of ST4, we have previously shown (14) that thspx is also bound to the LH1 complex and RC, and that this generates the characteristic peak maximum of BChla at 882 nm, and relative intensity of this peak maximum to those thspx as described above. The very strong similarity between the ST4 and SLYC18 cell spectra suggests strongly that, in SLYC18, lycopene is also bound to pigment-protein complexes. Fractionation of SLYC18 cell paste after breakage with a French press-type apparatus,

followed by differential centrifugation (17, 32), showed the lycopene to be located exclusively in the membrane fraction, of which the ICM, containing the LH1 and RC, comprises about 80%.

Finally, we purified the LH1 complex from SLYC18 and analyzed its spectroscopic properties and pigment content. As shown in Fig. 5A, the purified LH1 complex shows an absorption spectrum which is essentially identical to that of total ICM (which in isolated form are called chromatophores), including the position of the LH1-Q_y maximum at 882 nm, which is characteristic of native aggregated LH1 rings (53), and the relative intensities of the peaks of LH1-bound lycopene to that of the LH1 peak maximum at 882 nm. For all of the LH1 complexes obtained from spx-, thspx-, and lycopene-containing strains, this latter ratio is essentially constant. The absorption maxima of the LH1-bound lycopene correspond perfectly to those of intact cells (see Fig. 3). Finally, the pigment stoichiometry of the LH1 complex, determined from the characteristic absorption maxima of BChla (771 nm) and lycopene (502 nm) of the total extracted pigments redissolved in diethyl ether (Fig. 5B), was calculated to be 2.27 ± 0.07, corresponding well to the expected value of 2 mol BChla/1 mol crt (41). Comparison of the biochemical and spectroscopic data therefore indicates clearly that, *in vivo*, lycopene is almost exclusively bound to LH1 complexes in the ICM.

Growth physiology of the crt mutants. In many studies employing purple photosynthetic bacteria, it is natural to focus on their growth characteristics obtained under light, anaerobic conditions. Although SLYC18 indeed grows well photosynthetically, in this study we have focussed on the dark, semi-aerobic growth condition for lycopene production, as we believe that the latter is most readily adapted to a large-scale industrial environment. These conditions are facilitated by the availability of M2SF and M2SF$^+$ medium (18) which allows exceptionally high levels of pigment production under semi-aerobic growth conditions, which has not been achieved by any other purple bacterium so far. However, this study is the first to employ this medium for the production of a "non-natural" pigment product, so we felt it useful to study the growth characteristics in detail in this case.

To compare the physiology of the wild-type and crt mutants, we grew them in two different media, M2SF and M2SF$^+$. M2SF has been described previously (18) to yield high cell densities with high levels of ICM which are normally only observed under anaerobic photoheterotrophic conditions with conventional media. M2SF$^+$ is a modification of M2SF, employing even higher concentrations of both fructose (111.3 mM) and succinate (60 mM), which leads to four-fold higher cell densities than with M2SF while maintaining the phototrophic levels of ICM/cell. A very useful monitor of semi-aerobic growth is the ratio A_{882}/A_{660}, which indicates the relative level of ICM (proportional to

the amount of LH1 complex/cell). The A_{882}/A_{660} ratio reflects the internal redox physiology (i.e. the intracellular pO_2 as well as the ratio between reduced and oxidized ubiquinone in the ICM) of the cell, which influences the signal transduction events that regulate ICM expression (20). Thus, in a typical semi-aerobic growth curve, performed using shake flasks and a high ICM-containing inoculum, the A_{882}/A_{660} ratio initially decreases (corresponding to repression of ICM synthesis at high pO_2, while maintaining cell growth) until the cell density is high enough to reduce the local pO_2 to less than 0.5% (18). At this point, ICM synthesis resumes, concomitant with a rise in the A_{882}/A_{660} ratio. During this growth phase, both "aerobic", oxidative and "anaerobic", reductive modes of metabolism are utilized (18, 20). The A_{882}/A_{660} turning point is designated here as a "trough".

Typical semi-aerobic growth curves (100-ml cultures) in both M2SF and M2SF$^+$ media, are shown for the wild-type and mutants in Fig. 6. The M2SF growth profiles indicate the major physiological differences between strains. In all cases, a lag phase is observed, although for the wild-type this phase is short (about 14.7 h) and within the normal range observed for cultures where the inoculum has been stored for about 12-24 h at 4°C before inoculation. However, lag phases exhibited by the mutants are significantly longer (ST4, 22.8 h; SLYC18, 52.5 h) than is ever observed for wild-type cells. Interestingly, the position of the A_{882}/A_{660} trough does not correlate perfectly with the cell density. Whereas both S1 and SLYC18 exhibit the trough at an A_{660} of about 0.28, the ST4 trough occurs at about twice this value. (Fig. 6A, B). The long lag phase observed for SLYC18 may indicate that the rate of O_2 consumption is lower in this strain in comparison to those of S1 and ST4, or that low level oxidation of carotenoids may be toxic when the cell density is low. However, after the appropriate time, all of the cultures reach the same final cell densities, and the ICM content in the late exponential-stationary phase is comparable. We note, in passing, that the carotenoid-dependent lag phase does not correspond to the number of reducing equivalents removed from the carotenoids during the transition from phytoene to the final product (S1, 12 [H]; SLYC18, 8 [H]; ST4, 8 [H]).

Since we believe the length of the lag phase to have a distinct physiological meaning, related to the type of crt being produced, it would be useful to precisely determine its length using an objective procedure. We found the usually applied semi-logarithmic plots to be both subjective and inaccurate, so for this reason we turned to a precise fitting procedure, which also yields reliable growth rate estimates. First, we noticed that the growth behaviour following the lag phase fits the well-known logistic equation (see (38)):

$$n = \frac{n_0 * K * e^{rt}}{K + n_0(e^{rt} - 1)} \qquad (1)$$

In equation (1), n is the number of cells at measurement time t, n_o is the initial cell number, K is the "carrying capacity" of the medium (an empirical constant which indicates the maximal cell density when the medium is exhausted), and r is the growth rate. The difficulty here is to objectively define the end of the lag phase. To do this, in a series of trials, each curve was fitted to equation (1), with a successively longer lag time as the theoretical "time zero" value. The success of the fit was judged by observing the variance (S^2) of the residual [$S^2 = \sum_{1}^{N} (n_{\exp tl} - n_{fit})^2 \Big/ (N-1)$] where n_{fit} and n_{exptl} correspond to the fitted and experimental A_{660} values at each time point, and N is the number of time points used for the fit to equation (1) (see Supplementary Fig. 4 for an example). The fitted lag phase end point was defined as the time point corresponding to the smallest variance. This treatment also allowed us to determine the exponential growth rates (r) precisely for each strain. In fact, all strains show very similar growth rates (about 0.1 $A_{660}.h^{-1}$ (M2SF) and about 0.06 $A_{660}.h^{-1}$ (M2SF$^+$)) upon entering the exponential phase (see Supplementary Table 2).

In M2SF$^+$ medium, crt mutants exhibited a similar long lag phase (about 39.5 h for SLYC18 and 46.6 h for ST4, respectively) which was about twice that of the wild-type (21.8 h) (Fig. 6C, see Supplementary Table 2), although active growth still followed equation (1), and all strains attained comparable cell densities (approximately four-fold of those reached in M2SF medium). Also, the A_{882}/A_{660} ratio in the late exponential phase was very similar to that observed in M2SF medium. The position of the A_{882}/A_{660} trough with respect to the cell density appeared to be nearly the same in all strains (Fig. 6D). In comparison to the M2SF data, the M2SF$^+$ data are relatively "noisy". This is probably due to the fact that at very high cell densities, even during the short time (generally about 2-3 min) required for manual sampling, the redox physiology can change significantly. We have noticed this effect repeatedly in other experiments utilizing very high cell densities.

Quantitation of carotenoid levels. Carotenoid levels were determined using stationary phase M2SF$^+$ grown cells (Fig. 6C), by MeOH/hexane extraction followed by spectral analysis. The wild-type S1 yielded values of 1.41 mg spx/g DW (Fig. 7A) and 2.17 mg spx/g protein (Fig. 7C), respectively. The latter value is very close to the value 2.45 mg spx/g protein, determined by Jensen (23) for photoheterotrophic cultures of S1, and corresponds to the ratio of 65% protein/g DW, which we have consistently observed for S1 cells growing either in M2SF or M2SF$^+$. The lycopene and thspx levels per g protein appear to be significantly higher (36% and 16%, respectively, Fig. 7C), but in fact, these values only reflect a corresponding lowering of the percentage total protein/g DW in SLYC18 and ST4 (53% and 46%, respectively). However, notwithstanding the above considerations, the final levels of both lycopene and thspx per g DW still appear to be about 22%

higher than those of spx in S1 (Fig. 7A), and the amount of lycopene and thspx obtainable per liter M2SF$^+$ culture appears to be about 40% higher for the mutants than for the wild-type (Fig. 7B). These large differences are inconsistent with the spectral analyses of whole cells, which when normalized to the same cell density, show identical crt levels (and also LH1 levels) in all strains. This inconsistency can be partly rationalized by the fact that for S1, the growth sampling was aborted prior to attainment of the stationary phase. In the last part of the growth curve, even relatively small differences in cell turbidity can reflect large differences in cell mass. Our decision to abort the S1 growth curves prematurely was dictated by the fact that the turbidity can also fall significantly in the stationary phase. This behaviour is due partly to loss of cell viability, but also by the optical effects due to cell elongation. However, other physiological differences may also play a significant role (see Discussion).

Discussion

In this study, we have shown unambiguously that the final product of the *R. rubrum* phytoene desaturase reaction is indeed lycopene, as indicated by early chemical studies (8). Using LC/MS techniques, we could find no trace of neurosporene, which is the final physiological product of the phytoene desaturase present in both *Rb. capsulatus* (5, 19) and *Rb. sphaeroides* (28). Another type of CrtI has been found in *Rvi. gelatinosus* (39), where the final products appear to be a mixture of lycopene (10%) and neurosporene (90%) (21), consistent with the presence of both spx and spheroidene as the end carotenoids in this organism (42). Thus, the *R. rubrum* phytoene desaturase corresponds to a third type of CrtI, with the same enzymatic function as in the non-photosynthetic *P. ananatis* (31, 34) and *P. stewartii* (15, 22). So far, only one other purple bacterial CrtI (from *Rhodopseudomonas palustris*) has been reported (see (55)), which yields exclusively lycopene as the final product.

The reasons for the different product specificities of the various CrtIs are not yet clear. Although Wang and Liao (59) were able to isolate a random mutant of the *Rb. sphaeroides* CrtI which could produce about 90% lycopene and 10% neurosporene, the amino acid changes they reported do not correlate with the corresponding amino acids in either the *Rvi. gelatinosus* or *R. rubrum* enzymes (Supplementary Fig. 5, homology set 1). Recently, Stickforth and Sandmann (54) showed that the *Rvi. gelatinosus* enzyme could also be mutated to increase the lycopene production to 80%. However, also here the amino acid changes reported bear no relation to those present at the corresponding sequence positions of the *R. rubrum* CrtI (Supplementary Fig. 5). A sequence comparison of crt enzymes with known function shows that although the N- and C-terminal domains are reasonably conserved, there is considerable sequence variation within the central domains amongst the various enzymes (Supplementary Fig. 5). Also, the lycopene-producing

bacterial CrtIs from *R. rubrum* and *Rps. palustris* are more similar to other purple bacterial enzymes than to those from the non-photosynthetic *Pantoea* species. Finally, the *Rvi. gelatinosus* CrtI appears to show more extensive sequence homologies to the CrtI enzymes from *R. rubrum* and *Rps. palustris* than to those from *Rb. sphaeroides* and *Rb. capsulatus*, despite the fact that its product specificity is much closer to those of the latter enzymes (Supplementary Fig. 5, homology set 2). Thus, the sequence differences between the various enzymes which lead to different substrate and product specificities must be subtle, as no characteristic sequence motifs can be readily detected.

The spectral analysis of whole cells and also isolated ICM indicates strongly that lycopene is incorporated quantitatively into the LH1 complex. First, the relative ratios of the carotenoid maxima to that of the near-IR absorption maximum of the LH1 complex are essentially identical to those of the wild-type, as well as those of the thspx mutant ST4. We have previously reported that the LH1 complex from ST4 is functionally intact, and contains tightly-bound thspx (16, 27). In fact, the thspx-LH1 complex shows a somewhat higher photochemical stability than that from S1 (data not shown). Secondly, the position of the near-IR peak maximum of the SLYC18 LH1 at 882 nm, which is a well-established diagnostic feature of LH1 containing bound carotenoid (11, 17, 32, 60), is identical to those of S1 and ST4 (11). The prediction that the lycopene is indeed bound specifically to LH1 complexes was confirmed unambiguously by the analysis of purified LH1 complexes obtained from SLYC18. Here, the observed ratio of approximately 2 mol BChla/ mol lycopene is completely consistent with the structural (6) and biochemical (41) data available for the LH1 complex so far. Finally, the spectrum of the isolated ICM from SLYC18, which is known to contain the LH1 and RC, is almost identical to that of the isolated LH1 complex from the same strain, as is the case for both spx- as well as thspx-containing strains. These observations contradict the conclusions of Fiedor and Koyama (11, 25), based on *in vitro* reconstitution studies, that non-hydroxylated carotenoids such as lycopene cannot be incorporated efficiently into *R. rubrum* LH1 complexes.

The crt levels reported here are comparable to early attempts to produce carotenoids in engineered strains of *E. coli* (58) or yeasts (36, 49, 62), by the introduction of plasmid-borne *crt* genes from *Pantoea* species. More recent studies, employing more sophisticated metabolic engineering techniques, have reported levels ranging from 14-22 mg crt/g DW *E. coli* cells (2, 56, 64).

The growth physiology of the *R. rubrum* crt mutants well-illustrates the potential and relative simplicity of future crt production systems using phototrophic bacteria. Our choice of *R. rubrum* as a production strain was dictated partly by its natural high-level crt production capacity, and partly

by the now well-established fact that a massive redirection of carbon flux to pigments, as well as to phospholipid biosynthesis, is caused by the lowering of the pO_2 to below 0.5%. The latter largely eliminates the necessity to re-design the early metabolic steps (as in *E. coli*) to achieve high pigment levels, and simultaneously allows the enlargement of the membrane compartment. Finally, the availability of special media (M2SF, M2SF[+]) which allow ICM (and thus crt) to be expressed semi-aerobically at levels normally observed only under anaerobic, photoheterotrophic growth conditions (18, 20), lends itself well to up-scaling for an industrial process. In this study, we have routinely used 3-liter semi-aerobic shake cultures, which yield identical crt levels to those of 100-ml shake flask cultures grown in parallel. For all strains we have also obtained 10-50-liter batch cultures, which from a comparison of their absorption spectra, appear to have identical crt levels (for all strains here) to those reported here.

The significantly different levels of crt/g DW in the three strains is partly attributable to the time point of harvesting in the stationary phase. Although *R. rubrum* is often portrayed as a large elongated spiral bacterium with many "bends", this form arises only when the cells are in the late exponential or stationary phase. Rapidly growing *R. rubrum* cultures show only a single "bend" and are quite short (about 2 μm). The physiological changes (i.e. protein, lipid content) in the elongation phase have not yet been documented, but may be significant. Possibly this factor is responsible for the very different protein levels present in this phase (Fig. 7C).

An understanding of the growth physiology of phototrophic crt-producing strains will be necessary to proceed to produce higher crt levels, as well as other, possibly oxygenated carotenoids in *R. rubrum*. Thus, a relevant question is: if the growth physiologies of all three strains examined are so similar, what determines the very different lag phases observed? At present, we favour the hypothesis that the lag phase reflects the relative sensitivity of the strains to pO_2, mediated at the crt level, in the initial growth phase. First, we do not observe significant lag phase differences when the three strains are grown photoheterotrophically, indicating that the incorporation of non-natural carotenoids (for *R. rubrum*, we consider also lycopene to be "non-natural" as only trace amounts are observed in growing cultures of wild-type strains) into the LH1 complex and RC is not fundamentally toxic for the cells. Secondly, in bioreactor experiments, performed in parallel to this study, the early reduction of the O_2 level essentially abolishes the lag phase difference (R. Feuer, J. Bona-Lovacz, O. Sawodny, and R. Ghosh, unpublished data). Possibly, the incorporation of non-natural carotenoids into the LH1 complex induces a conformational strain upon the structure, thereby making the BChla, which is normally quite well-protected in the protein environment, more accessible to oxidative degradation. It has been shown (17) that even low oxidative damage of the

71

LH1 complex leads to a significant destruction of the ICM organization, which is associated with aerobic toxicity. After a certain time, a small portion of the population is able to tolerate the toxicity so to commence cell growth, thereby lowering the pO_2 and alleviating growth inhibition. Presumably, the hydroxylated/methoxylated carotenoids (e.g. spx) lead to less structural strain than the non-hydroxylated/non-methoxylated ones (e.g. lycopene), which explains the relative efficiencies of the *in vitro* reconstitution experiments (11, 25). We have also observed long lag phases in other non-related *R. rubrum* mutants where the lesion causes the LH1 complex to be abnormally sensitive to oxidative stress (C. Autenrieth and R. Ghosh, manuscript in preparation). Future strategies might therefore focus on making the LH1 complex less sensitive to the incorporation of non-natural carotenoids.

Although we have shown that it is feasible to produce respectable quantities of lycopene in *R. rubrum*, the levels are still much lower than those observed for recently reported, highly optimized *E. coli* strains (2, 10). However, there is good reason to believe that the lycopene levels observed with *R. rubrum* correspond to *minimal* levels. In particular, the very similar levels of crt in all three strains examined, together with the spectral and biochemical evidence that only sufficient levels to saturate the LH1 complex are produced, suggest that some kind of feedback regulation at the post-translational level is occurring. This is supported by the observation that almost all mutants with lesions in BChla but *not* crt biosynthesis do not produce carotenoids (R. Ghosh, unpublished data). It is quite possible that if one could release this putative "brake" then significantly higher levels of crt may be attainable. This area is the focus of our ongoing research.

References

1. **Albermann, C.** 2011. High versus low level expression of the lycopene biosynthesis genes from *Pantoea ananatis* in *Escherichia coli*. Biotechnol. Lett. **33**:313-319.

2. **Alper, H., K. Miyaoku, and G. Stephanopoulos.** 2005. Construction of lycopene-overproducing *E. coli* strains by combining systematic and combinatorial gene knockout targets. Nat. Biotechnol. **23**:612-616.

3. **Armstrong, G. A., M. Alberti, F. Leach, and J. E. Hearst.** 1989. Nucleotide sequence, organization, and nature of the protein products of the carotenoid biosynthesis gene cluster of *Rhodobacter capsulatus*. Mol. Gen. Genet. **216**:254-268.

4. **Armstrong, G. A.** 1995. Genetic analysis and regulation of carotenoid biosynthesis. p. 1135-1157. *In* R. E. Blankenship, M. T. Madigan, and C. E. Bauer (eds), Anoxygenic Photosynthetic Bacteria. Kluwer Academic Publishers, Netherlands.

5. **Bartley, G. E., and P. A. Scolnik.** 1989. Carotenoid biosynthesis in photosynthetic bacteria. J. Biol. Chem. **264**:13109-13113.

6. **Brunisholz, R. A., H. Zuber, J. Valentine, J. G. Lindsay, K. J. Woolley, and R. J. Cogdell.** 1986. The membrane location of the B890-complex from *Rhodospirillum rubrum* and the effect of carotenoid on the conformation of its two apoproteins exposed at the cytoplasmic surface. Biochim. Biophys. Acta. **849**:295-303.

7. **Cohen-Bazire, G., W. R. Sistrom, and R. Y. Stanier.** 1956. Kinetic studies of pigment synthesis by non-sulfur purple bacteria. J. Cell. Comp. Physiol. **49**:25-68.

8. **Davies, B. H.** 1970. A novel sequence for phytoene dehydrogenation in *Rhodospirillum rubrum*. Biochem. J. **116**:93-99.

9. **Ditta, G., T. Schmidhauser, E. Yakobson, P. Lu, X.-W. Liang, D. R. Finlay, D. Guiney, and D. R. Helinski.** 1985. Plasmids related to the broad host range vector, pRK290, useful for gene cloning and for monitoring gene expression. Plasmid. **13**:149-153.

10. **Farmer, W. R., and J. C. Liao.** 2000. Improving lycopene production in *Escherichia coli* by engineering metabolic control. Nat. Biotechnol. **18**:533-537.

11. **Fiedor, L., J. Akahane, and Y. Koyama.** 2004. Carotenoid-induced cooperative formation of bacterial photosynthetic LH1 complex. Biochemistry. **43**:16487-16496.

12. **Figurski, D. H., and D. R. Helinski.** 1979. Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided *in trans*. Proc. Natl. Acad. Sci. USA. **76**:1648-1652.

13. **Fraser, P. D., N. Misawa, H. Linden, S. Yamano, K. Kobayashi, and G. Sandmann.** 1992. Expression in *Escherichia coli*, purification, and reactivation of the recombinant *Erwinia uredovora* phytoene desaturase. J. Biol. Chem. **267**:19891-19895.

14. **Gaertner, P., H. Port, M. Branschaedel, and R. Ghosh.** 2004. FS-study on energy relaxation in light-harvesting (LH1) complexes from *Rhodospirillum rubrum* with carotenoids of different conjugation length. J. Luminescence. **108**:111-116.

15. **Garcia-Asua, G., R. J. Cogdell, and C. N. Hunter.** 2002. Functional assembly of the foreign carotenoid lycopene into the photosynthetic apparatus of *Rhodobacter sphaeroides* achieved by replacement of the native 3-step phytoene desaturase with its 4-step counterpart from *Erwinia herbicola*. Mol. Microbiol. **44**:233-244.

16. **Gerken, U., F. Jelezko, B. Götze, M. Branschädel, C. Tietz, R. Ghosh, and J. Wrachtrup.** 2003. Membrane environment reduces the accessible conformational space available to an integral membrane protein. J. Phys. Chem. **107**:338-343.

17. **Ghosh, R., R. Bachofen, and H. Hauser.** 1985. Structural changes accompanying the irreversible oxidation of the chromatophores membrane from *Rhodospirillum rubrum* G9. Biochim. Biophys. Acta. **765**:97-105.

18. **Ghosh, R., A. Hardmeyer, I. Thoenen, and R. Bachofen.** 1994. Optimization of the Sistrom culture medium for large-scale batch cultivation of *Rhodospirillum rubrum* under semi-aerobic conditions with maximal yield of photosynthetic membranes. Appl. Environ. Microbiol. **60**:1698-1700.

19. **Giuliano, G., D. Pollock, and P. A. Scolnik.** 1986. The gene *crtI* mediates the conversion of phytoene into colored carotenoids in *Rhodopseudomonas capsulata*. J. Biol. Chem. **261**:12925-12929.

20. **Grammel, H., E.-D. Gilles, and R. Ghosh.** 2003. Microaerophilic cooperation of reductive and oxidative pathways allows maximal photosynthetic membrane biosynthesis in *Rhodospirillum rubrum*. Appl. Environ. Microbiol. **69**:6577-6586.

21. **Harada, J., K. V. P. Nagashima, S. Takaichi, N. Misawa, K. Matsuura, and K. Shimada.** 2001. Phytoene desaturase, CrtI, of the purple photosynthetic bacterium, *Rubrivivax gelatinosus*, produces both neurosporene and lycopene. Plant Cell Physiol. **42**:1112-1118.

22. **Hunter, C. N., B. S. Hundle, J. E. Hearst, H. P. Lang, A. T. Gardiner, S. Takaichi, and R. J. Cogdell.** 1994. Introduction of new carotenoids into the bacterial photosynthetic apparatus by combining the carotenoid biosynthetic pathways of *Erwinia herbicola* and *Rhodobacter sphaeroides*. J. Bacteriol. **176**:3692-3697.

23. **Jensen, S. L., G. Cohen-Bazire, T. O. M. Nakayama, and R. Y. Stanier.** 1958. The path of carotenoid synthesis in a photosynthetic bacterium. Biochim. Biophys. Acta. **29**:477-498.

24. **Kajiwara, S., P. D. Fraser, K. Kondo, and N. Misawa.** 1997. Expression of an exogeneous isopentenyl diphosphate isomerase gene enhances isoprenoid biosynthesis in *Escherichia coli*. Biochem. J. **324**:421-426.

25. **Kakitani, Y., J. Akahane, H. Ishii, H. Sogabe, H. Nagae, and Y. Koyama.** 2007. Conjugation-length dependence of the T_1 lifetimes of carotenoids free in solution and incorporated into the LH2, LH1, RC, and RC-LH1 complexes: possible mechanisms of triplet-energy dissipation. Biochemistry. **46**:2181-2197.

26. **Kang, M. J., Y. M. Lee, S. H. Yoon, J. H. Kim, S. W. Ock, K. H. Jung, Y. C. Shin, J. D. Keasling, and S. W. Kim.** 2005. Identification of genes affecting lycopene accumulation in *Escherichia coli* using a shot-gun method. Biotechnol. Bioeng. **91**:636-642.

27. **Komori, M., R. Ghosh, S. Takaichi, Y. Hu, T. Mizoguchi, Y. Koyama, and M. Kuki.** 1998. A null lesion in the rhodopin 3,4-desaturase of *Rhodospirillum rubrum* unmasks a cryptic branch of the carotenoid biosynthetic pathway. Biochemistry. **37**:8987-8994.

28. **Lang, H. P., R. J. Cogdell, A. T. Gardiner, and C. N. Hunter.** 1994. Early steps in carotenoid biosynthesis: Sequences and transcriptional analysis of the *crtI* and *crtB* genes of *Rhodobacter sphaeroides* and overexpression and reactivation of *crtI* in *Escherichia coli* and *Rhodobacter sphaeroides*. J. Bacteriol. **176**:3859-3869.

29. **Lee, P. C., and C. Schmidt-Dannert.** 2002. Metabolic engineering towards biotechnological production of carotenoids in microorganisms. Appl. Microbiol. Biotechnol. **60**:1-11.

30. **Lee , P. C., B. N. Mijts, and C. Schmidt-Dannert.** 2004. Investigation of factors influencing production of the monocyclic carotenoid torulene in metabolically engineered *Escherichia coli*. Appl. Microbiol. Biotechnol. **65**:538-546.

31. **Linden, H., N. Misawa, D. Chamovitz, I. Pecker, J. Hirschberg, and G. Sandmann.** 1991. Functional complementation in *Escherichia coli* of different phytoene desaturase genes and analysis of accumulated carotenes. Z. Naturforsch. **46**:1045-1051.

32. **Lupo, D., and R. Ghosh.** 2004. The reaction center H subunit is not required for high levels of light-harvesting complex 1 in *Rhodospirillum rubrum* mutants. J. Bacteriol. **186**:5585-5595.

33. **Matthews, P. D., and E. T. Wurtzel.** 2000. Metabolic engineering of carotenoid accumulation in *Escherichia coli* by modulation of the isoprenoid precursor pool with expression of deoxyxylulose phosphate synthase. Appl. Environ. Microbiol. **53**:396-400.

34. **Misawa, N., M. Nakagawa, K. Kobayashima, S. Yamamo, Y. Izawa, K. Nakamura, and K. Harashima.** 1990. Elucidation of the *Erwinia uredovora* carotenoid biosynthetic pathway by functional analysis of gene products expressed in *Escherichia coli*. J. Bacteriol. **172**:6704-6712.

35. **Misawa, N., and H. Shimada.** 1998. Metabolic engineering for the production of carotenoids in non-carotenogenic bacteria and yeasts. J. Bacteriol. **59**:169-181.

36. **Miura, Y., K. Kondo, T. Saito, H. Shimada, P. D. Fraser, and N. Misawa.** 1998. Production of the carotenoids lycopene, β-carotene, and astaxanthin in the food yeast *Candida utilis*. Appl. Environ. Microbiol. **64**:1226-1229.

37. **Neudert, U., I. M. Martinez-Ferez, P. D. Fraser, and G. Sandmann.** 1998. Expression of an active phytoene synthase from *Erwinia uredovora* and biochemical properties of the enzyme. Biochim. Biophys. Acta. **1392**:51-58.

38. **Novak, M. A.** 2006. Evolutionary dynamics: exploring the equations of life. p. 13. Harvard University Press.

39. **Ouchane, S., M. Picaud, C. Vernotte, F. Reiss-Husson, and C. Astier.** 1997. Pleiotropic effects of *puf* interposon mutagenesis on carotenoid biosynthesis in *Rubrivivax gelatinosus*. J. Biol. Chem. **272**:1670-1676.

40. **Peterson, G. L.** 1977. A simplification of the protein assay method of Lowry *et al*. which is more generally applicable. Anal. Biochem. **83**:346-356.

41. **Picorel, R., G. Bélanger, and G. Gingras.** 1983. Antenna holochrome B880 of *Rhodospirillum rubrum* S1. Pigment, phospholipid, and polypeptide composition. Biochemistry. **22**:2491-2497.

42. **Pinta, V., S. Ouchane, M. Picaud, S. Takaichi, C. Astier, and F. Reiss-Husson.** 2003. Characterization of unusual hydroxyl- and ketocarotenoids in *Rubrivivax gelatinosus*: Involvement of enzyme CrtF or CrtA. Arch. Microbiol. **179**:354-362.

43. **Ruther, A., N. Misawa, P. Böger, and G. Sandmann.** 1997. Production of zeaxanthin in *Escherichia coli* transformed with different carotenogenic plasmids. Appl. Microbiol. Biotechnol. **48**:162-167.

44. **Sambrook, J., and D. W. Russell.** 2001. Molecular cloning: a laboratory manual, 3rd ed. Cold Spring Harbor Press, Cold Spring Harbor, N.Y.

45. **Sandmann, G., W. S. Woods, and R. W. Tuveson.** 1990. Identification of carotenoids in *Erwinia herbicola* and in a transformed *Escherichia coli* strain. FEMS Microbiol. Lett. **71**:77-82.

46. **Sandmann, G., M. Albrecht, G. Schnurr, O. Knörzer, and P. Böger.** 1999. The biotechnological potential and design of novel carotenoids by gene combination in *Escherichia coli*. TIBTECH. **19**:233-237.

47. **Schmidt-Dannert, C., D. Umeno, and F. H. Arnold.** 2000. Molecular breeding of carotenoid biosynthetic pathways. Nat. Biotechnol. **18**:750-753.

48. **Schmidt, K., S. Liaan-Jensen, and H. G. Schlegel.** 1963. Die Carotinoide der Thiorhodaceae: I. Okenon als Hauptcarotinoid von *Chromatium okenii* Perty. Arch. Microbiol. **46**:117-126.

49. **Shimada, H., K. Kondo, P. D. Fraser, Y. Miura, T. Saito, and N. Misawa.** 1998. Increased carotenoid production by the food yeast *Candida utilis* through metabolic engineering of the isoprenoid pathway. Appl. Environ. Microbiol. **64**:2676-2680.

50. **Simon, R., U. Priefer, and A. Pühler.** 1983. A broad host range mobilization system for *in vivo* genetic engineering: transposon mutagenesis in gram negative bacteria. Nat. Biotechnol. **1**:784-791.

51. **Sistrom, W. R.** 1960. A requirement for sodium in the growth of *Rhodopseudomonas sphaeroides*. J. Gen. Microbiol. **22**:778-785.

52. **Smith, J. H. C., and A. Benitez.** 1955. Chlorophyll: analysis in plant materials. p. 142. *In* Paech, K., and M. V. Tracy (eds.), Modern methods of plant analysis, vol. 4. Springer-Verlag, Berlin.

53. **Stahlberg, H., J. Dubochet, H. Vogel, and R. Ghosh.** 1998. Are the light-harvesting I complexes from *Rhodospirillum rubrum* arranged around the reaction centre in a square geometry? J. Mol. Biol. **282**:819-831.

54. **Stickforth, P., and G. Sandmann.** 2007. Kinetic variations determine the product pattern of phytoene desaturase from *Rubrivivax gelatinosus*. Arch. Bioch. Biophys. **461**:235-241.

55. **Takaichi, S.** 2009. Distribution and biosynthesis of carotenoids. p. 97-117. *In* C. N. Hunter, F. Daldal, M. C. Thurnauer, and J. T. Beatty (eds.), The Purple Phototrophic Bacteria, Springer Science + Business Media.

56. **Tyo, K. E. J., P. K. Ajikumar, and G. Stephanopoulos.** 2009. Stabilized gene duplication enables long-term selection-free heterologous pathway expression. Nat. Biotechnol. **27**:760-765.

57. **Vadali, R. V., Y. Fu, G. N. Bennett, and K. Y. San.** 2005. Enhanced lycopene productivity by manipulation of carbon flow to isopentenyl diphosphate in *Escherichia coli*. Biotechnol. Prog. **21**:1558-1561.

58. **Wang, C.-W., M.-K. Oh, and J. C. Liao.** 1999. Engineered isoprenoid pathway enhances astaxanthin production in *Escherichia coli*. Biotech. Bioeng. **62**:235-241.

59. **Wang, C.-W., and J. C. Liao.** 2001. Alteration of the product specificity of *Rhodobacter sphaeroides* phytoene deaturase by directed evolution. J. Biol. Chem. **276**:41161-41164.

60. **Wiggli, M., L. Cornacchia, R. Saegesser, R. Bachofen, and R. Ghosh.** 1996. Characterization of *Rhodospirillum rubrum* ST2. A new Tn5-induced carotenoid-less mutant for functional studies. Microbiol. Res. **151**:57-62.

61. **Wolfe, A. J.** 2005. The acetate switch. Microbiol. Mol. Biol. Rev. **69**:12-50.

62. **Yamono, S., T. Ishii, M. Nakagawa, H. Ikenaga, and N. Misawa.** 1994. Metabolic engineering for production of β-carotene and lycopene in *Saccharomyces cerevisiae*. Biosci. Biotechnol. Biochem. **58**:1112-1114.

63. **Yen, H.-C., and B. Marrs.** 1976. Map of genes for carotenoid and bacteriochlorophyll biosynthesis in *Rhodopseudomonas capsulata*. J. Bacteriol. **126**:619-629.

64. **Yoon, S.-H., Y.-M. Lee, J.-E. Kim, S.-H. Lee, J.-H. Lee, J.-Y. Kim, K.-H. Jung, Y.-C. Shin, J. D. Keasling, and S.-W. Kim.** 2006. Enhanced lycopene production in *Escherichia coli* engineered to synthesize isopentenyl diphosphate and dimethylallyl diphosphate from mevalonate. Biotechnol. Bioeng. **94**:1025-1032.
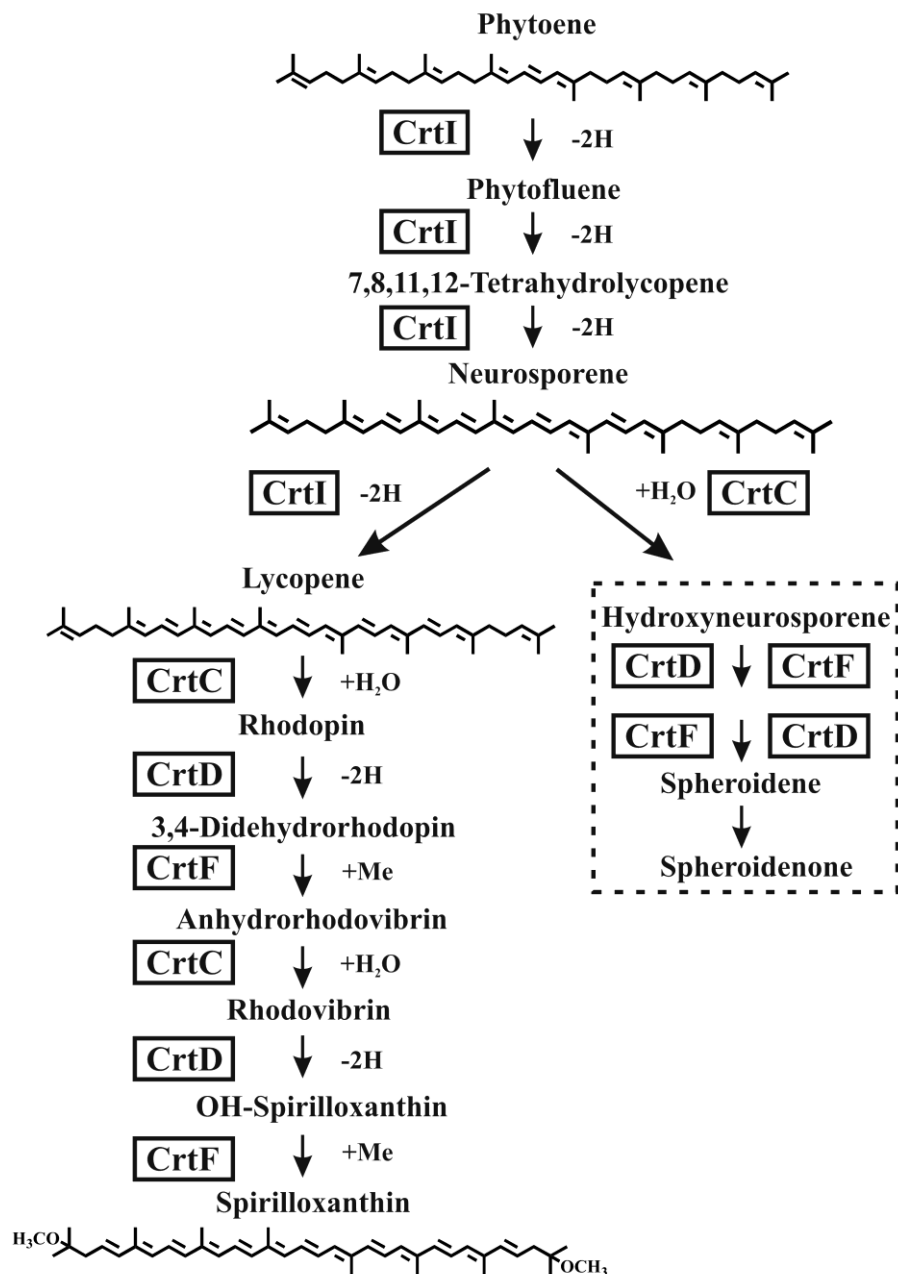
Fig. 1. The carotenoid biosynthesis pathway in *R. rubrum* (end product, spx). The known pathway to spheroidene/spheroidenone in *Rb. capsulatus* and *Rb. sphaeroides*, which branches at the third CrtI-mediated oxidation step, is also shown (dashed box) for comparison. For the latter pathway, several intermediate steps have been omitted for convenience. The gene products which catalyze the various steps are also indicated (small solid-line boxes).
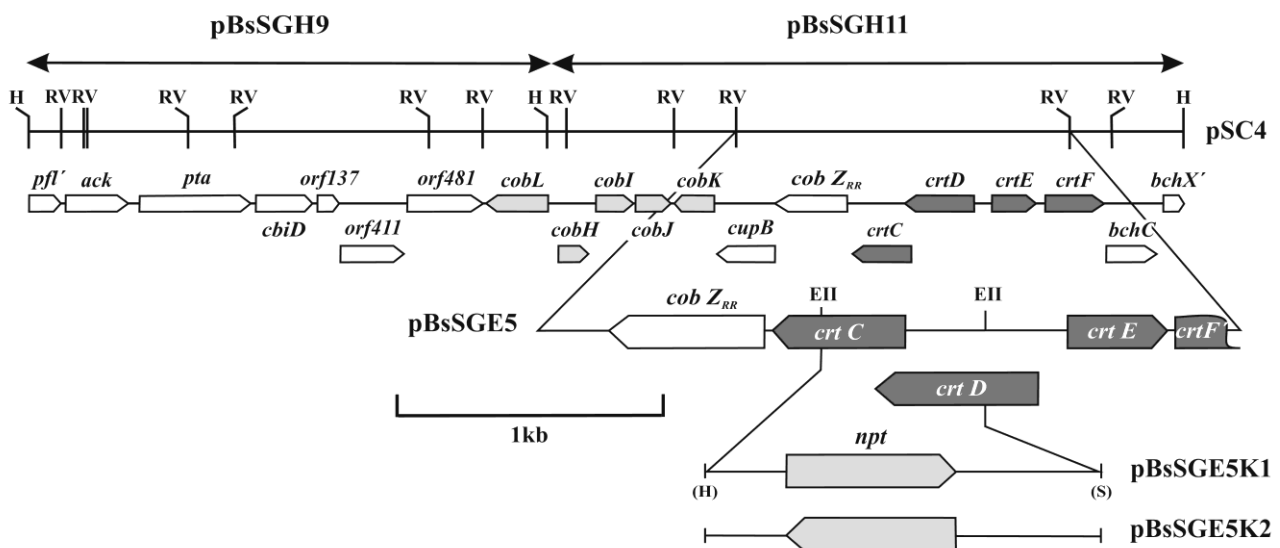
Fig. 2. Construction of the plasmids pBsSGE5K1 and pBsSGE5K2. The gene organization of the cosmid initially isolated, pSC4, which complemented the *crtD*⁻ mutant ST4, is indicated. The partial *crt* cluster (*crtCDEF*) is indicated (dark grey) as is the pSC4-derived EcoRV fragment used to construct the plasmid pBsSGE5. Site-directed deletion of the *crtC* and *crtD* genes was performed by replacing a BstEII fragment overlapping both genes with the *npt* cassette. Both *npt* orientations relative to the *crt* cluster were obtained (indicated).
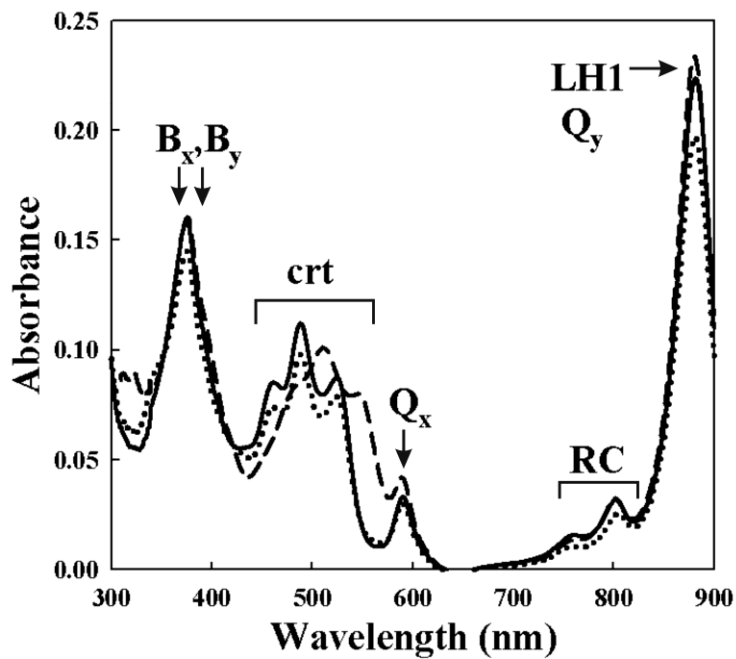
Fig. 3. Whole cell spectra of strains SLYC18 (solid line), ST4 (dotted line) and the wild-type S1 (dashed line), containing different carotenoids. The cell spectra were normalized to the same A_{660} for comparison. Indicated are the absorption peaks corresponding to crt, as well as the BChla near-IR maxima (Q_x, Q_y) and Soret bands (B_x, B_y) of the LH1 and RC.
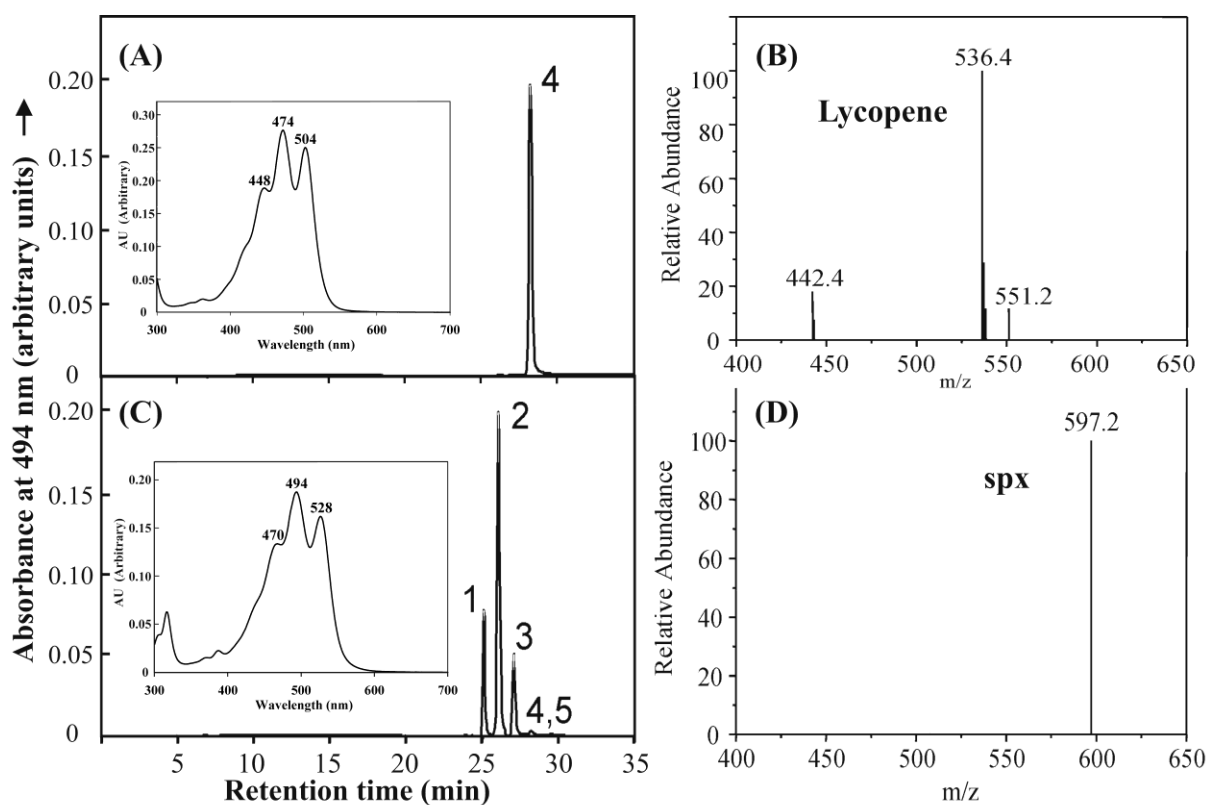
Fig. 4. HPLC and MS analysis of extracted carotenoids from SLYC18 (A, B) and S1 (C, D).
The inserts show the absorption spectra of the dominant peaks (peak 4 (in (A)) and peak 2 (in (C)), respectively). The masses (M^- mode) shown in (B) correspond to lycopene (m/z, 536.4), 3,4-didehydrorhodopin (m/z, 551.2), and an unassigned species (m/z, 442.4). (D) shows the mass spectrum ($M+H^+$ mode) of the dominant peak 2 (spx, m/z, 597.2) in (C). The remaining peaks in (C) were also assigned from their mass spectra (data not shown): peak 1, OH-spx (m/z, 568.8); peak 3, anhydrorhodovibrin (m/z, 567.2); peak 4, lycopene; peak 5, 3,4-didehydrorhodopin (m/z, 553.4).
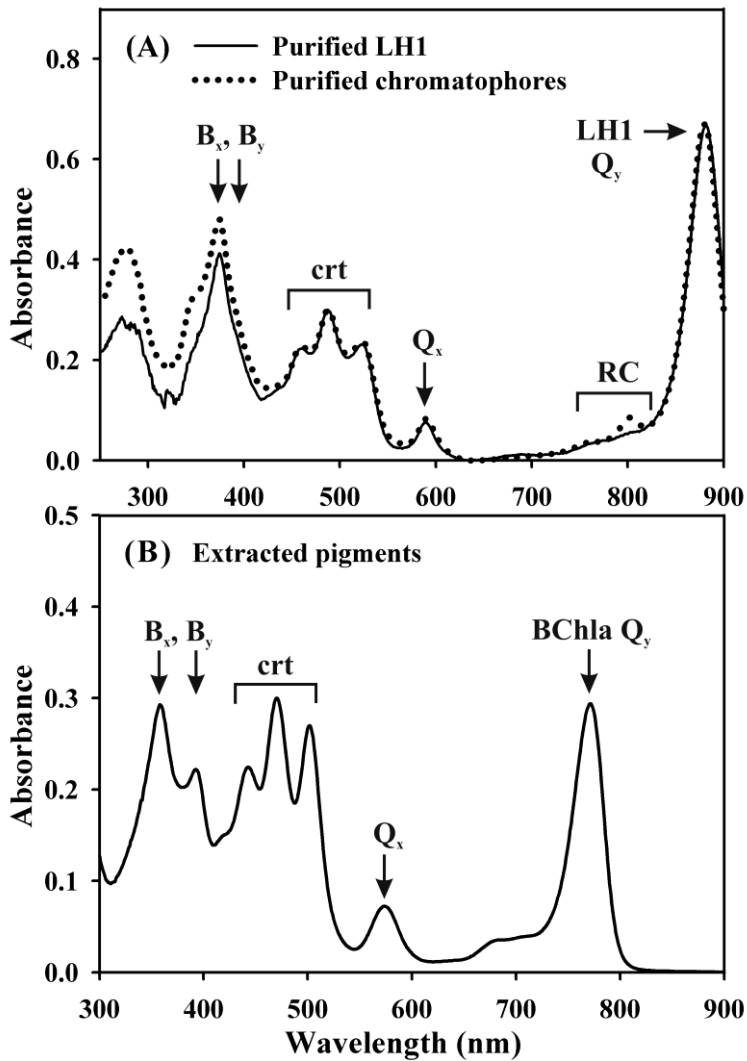
Fig. 5. (A) Comparison of the absorption spectrum of isolated SLYC18 chromatophores (the ICM fraction (dotted line)) with that of purified LH1 complex obtained from the ICM fraction. Note that the peak at 802 nm, which is observed for the ICM, and is due to the absorption of the accessory BChla of the RC, is not present in the LH1 preparation. (B) Absorption spectrum of the pigment fraction after methylene chloride/MeOH extraction (see Materials and Methods) of the purified LH1 complexes. The spectrum was obtained using diethyl ether as a solvent.
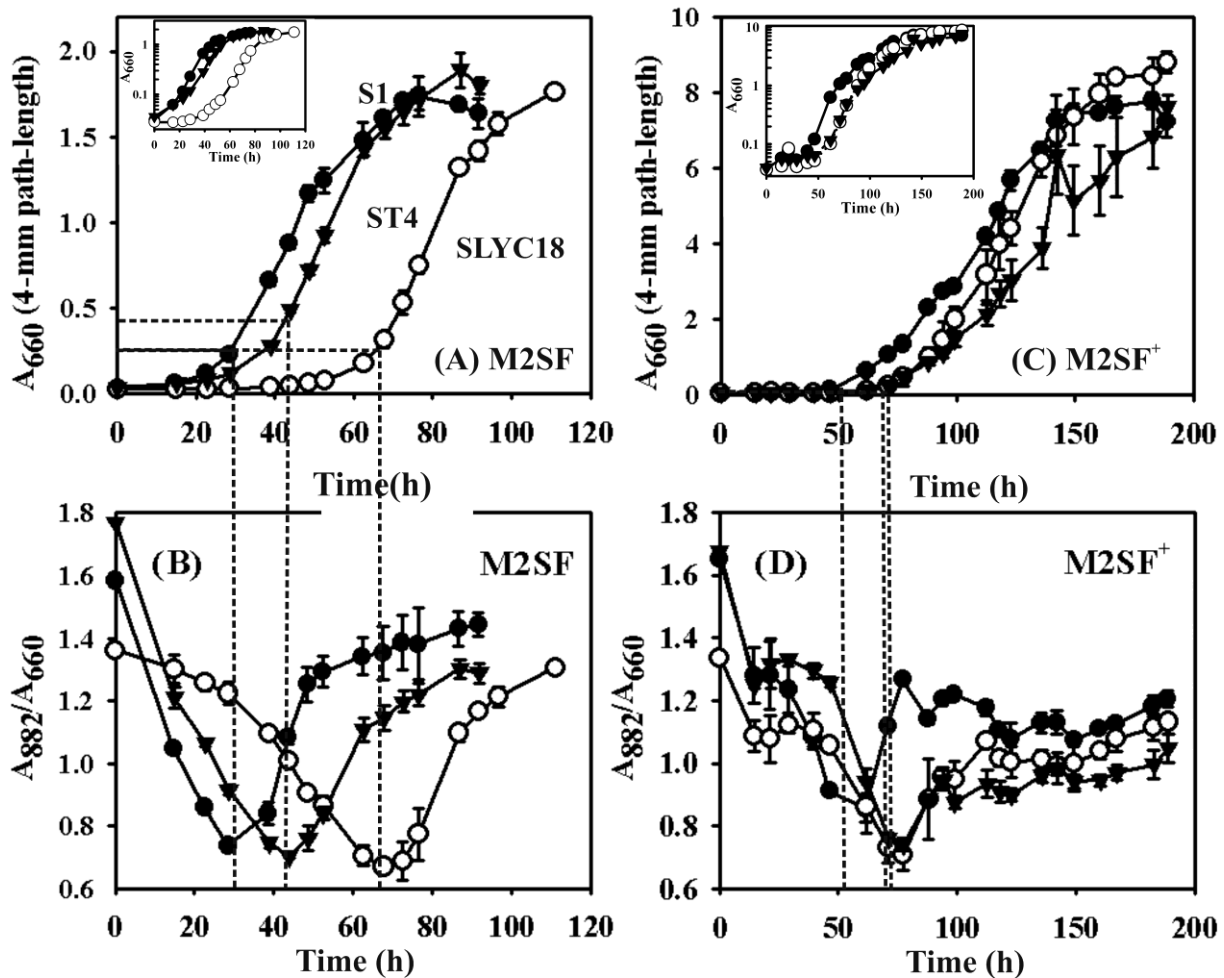
Fig. 6. Growth curves of wild-type (S1(●)) and crt mutants (SLYC18(○) and ST4(▼)) in (A) M2SF medium, (C) M2SF$^+$ medium. For each strain, in each growth medium, three independent growth curves were performed, but in the figure, only the average curve with error bars is shown. For each growth curve, the development of ICM (indicated by A_{882}/A_{660}) is also shown in (B), (D). The dashed lines have been included to provide visual aids showing the correlation between the onset of the microaerophilic state experienced by the cells, corresponding to the increase of ICM/cell (A_{882}/A_{660}), and the growth phase of the cells in the respective media. For comparative purposes, we have also included the commonly-used semi-logarithmic plot as an inset.
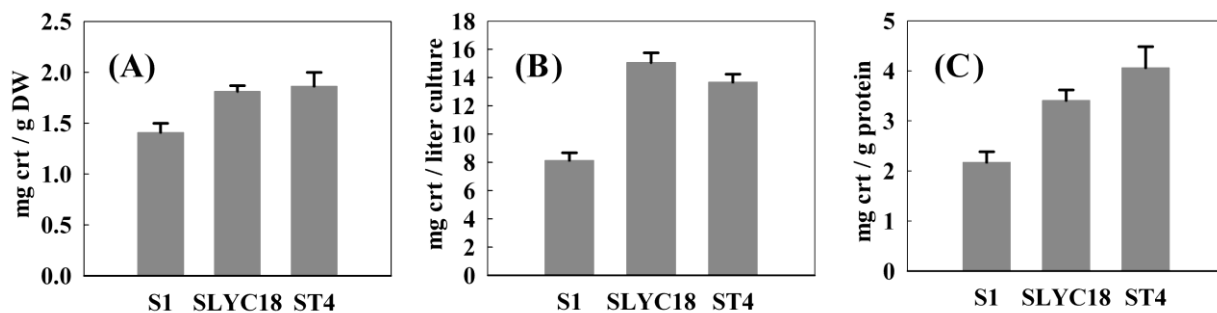
Fig. 7. Quantitative analysis of carotenoids from stationary phase cells grown in M2SF$^+$ medium. All determinations were performed in triplicate for each of the individual cultures shown in Fig. 6C. Thus, each of the bar chart values and their error bars have been obtained from a total of nine determinations. (A) mg crt/g DW cells; (B) mg crt/liter culture; (C) mg crt/g total protein, as determined by modified Lowry-Peterson procedure (40), see Methods.

Table 1. Strains and plasmids used in this study

Strain or plasmid	Relevant characteristics	Reference
***E. coli* strains**		
XL1MR	Δ(*mcrA*)183 Δ(*mcrCB-hsdSMR-mrn173 endA1 supE44 thi1 recA1 gyrA96 relA1 lac*)	Stratagene
***R. rubrum* strains**		
S1	Wild-type	7
ST4	S1-derived *crtD*⁻ mutant, Kan^r	27
SLYC18	S1-derived *crtC*⁻*D*⁻ site-directed mutant, Kan^r	This study
Plasmids		
pBsKSII+	High-copy cloning vector, ColE1, Amp^r	Stratagene
pBsLGKan	pBsKSII+ derivative, carrying the kanamycin resistance cassette, Kan^r	32
pRK404	Derivative of pRK290, *mob*$^+$	9
pRK2013	Mobilizing helper plasmid, *tra*$^+$, Kan^r	12
pBsSGE5	pBsKSII+ derivative, containing the 6.37 kb EcoRV fragment from the pSC4 cosmid	This study
pBsSGE5K1	pBsSGE5 derivative, containing the kanamycin resistance cassette in the opposite direction to the *crtD* gene	This study
pBsSGE5K2	pBsSGE5 derivative, containing the kanamycin resistance cassette in the same direction as the *crtD* gene	This study
pRKSGE5	pRK404 derivative, containing the 6.42 kb HindIII/XbaI insert from pBsSGE5	This study

pSUP202	Suicide vector, ColE1, *mob*$^+$, Amp^r Cm^r Tet^r	50
pSUPSGE5K1	pSUP202 derivative, containing the HindIII/SalI insert of pBsSGE5K1, Amp^r Kan^r Cm^r Tet^r	This study
pSUPSGE5K2	pSUP202 derivative, containing the HindIII/SalI insert of pBsSGE5K2, Amp^r Kan^r Cm^r Tet^r	This study

Supplementary material

Determination of the extinction coefficient for lycopene (lyc) in methanol. A small amount of extracted lycopene was dried to completion under nitrogen and then dissolved either in 0.5-ml hexane or 0.5-ml methanol. The spectra of both samples were then taken, and the extinction coefficient of lycopene in methanol (ε^{501nm} (1cm)) was calculated from the known extinction coefficient in hexane (ε^{502nm} (1cm)) as follows:

$$\varepsilon_{1cm}^{501nm}(lyc, MeOH) = \frac{A^{501nm}(MeOH)}{A^{502nm}(hexane)} \times \varepsilon_{1cm}^{502nm}(lyc, hexane) \qquad (1)$$

Similar calculations were performed for Spx. The results are shown below:

Supplementary Table 1

Carotenoids	Solvents	Extinction Coefficients	Standard Deviation
Lycopene[a]	hexane	$\varepsilon^{502nm}(1cm) = 172$ mM^{-1}cm^{-1}	4.98%
Lycopene[b]	MeOH	$\varepsilon^{501nm}(1cm) = 138$ mM^{-1}cm^{-1}	4.11%
Lycopene[b]	diethyl ether	$\varepsilon^{502nm}(1cm) = 184.2$ mM^{-1}cm^{-1}	4.73%
Spx[c]	petroleum ether	$\varepsilon^{488nm}(1cm) = 140$ mM^{-1}cm^{-1}	0.65%
Spx[b]	hexane	$\varepsilon^{484nm}(1cm) = 134.7$ mM^{-1}cm^{-1}	0.96%
Spx[b]	MeOH	$\varepsilon^{484nm}(1cm) = 132$ mM^{-1}cm^{-1}	1.54%

[a], taken from (S3);

[b], determined in this study;

[c], taken from (S1).

The standard deviation was derived from three separate measurements.

Absorption spectra of carotenoids. The absorption spectra of lycopene, thSpx, and Spx in hexane. The characteristic absorption maxima are also indicated.
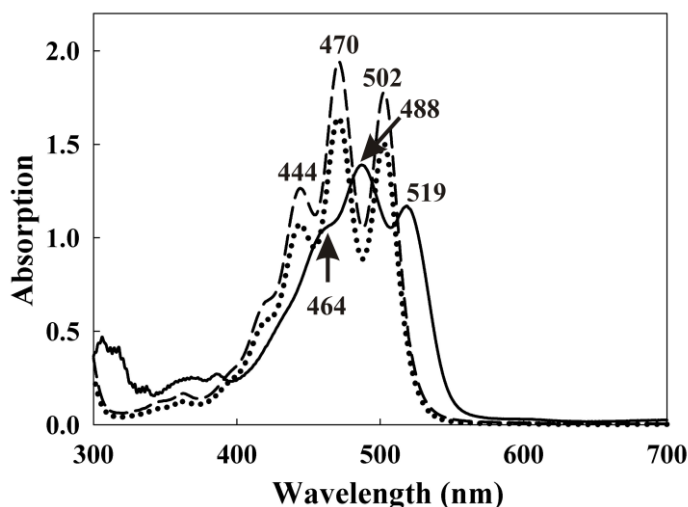
Supplementary Fig. 1

Typical absorption spectra for the carotenoids Spx (solid line), lycopene (dashed line), and thSpx (dotted line) taken in hexane. The relative intensities are arbitrary.

Determination of the lag phase. One example (SLYC18_1, M2SF$^+$ growth curve).

The growth behavior following the lag phase fits the well-known logistic equation (see (S2)):

$$n = \frac{n_0 K * e^{rt}}{K + n_0(e^{rt} - 1)} \qquad (2)$$

In equation (2), n is the number of cells at measurement time t, n_o is the initial cell number, K is the "carrying capacity" of the medium, and r is the growth rate. Since we have performed cell turbidity measurements within the linear response region, we can replace n with the turbidity (A_{660}). The difficulty in applying equation (2) is to objectively determine the end of the lag phase, in particular, because the inclusion of a small lag phase can also be fitted reasonably well. We reasoned that since the best fit is indicated by the variance:

$$S^2 = \frac{\sum_{1}^{N}(n_{exp\,tl} - n_{fit})^2}{N-1} \qquad (3)$$

we could define the end of the lag phase by choosing the lowest variance to fitting trials where the lag phase region has been successively truncated.

We illustrate this procedure by showing the detailed data processing of one of the SLYC18 growth curves (Supplementary Fig. 2, shown below). The first curve shows the initial data set without lag phase truncation. The following two curves show successive lag phase truncations at 21.33 h and

29.08 h, respectively, where the long lag phase caused the fitting program (a defined algorithm in the SIGMAPLOT program library) to abort. The last three curves show succesive lag phase truncations at times 39.47 h, 46.42 h, and 61.75 h (these times correspond to the individual (t, A_{660}) data points), and their corresponding residuals (residual = $(A_{660}(\text{exptl}) - A_{660}(\text{fitted}))$. The residual allows a visual assessment of the fit. The variance of each lag phase truncation is shown inside the residual box. Thus, despite the fact that (in this case) the residual plots appear to be visually rather similar, the lowest variance is obtained after the truncation at 39.47 h, which was taken to be the length of the lag phase.

It should be noted that the fitted growth rate (r) is rather insensitive to the exact positioning of the lag phase on the final three truncations. Thus, this value is always reliable. The growth rates for all three strains examined are listed in Supplementary Table 2.

Supplementary Fig. 2. Determination of the lag phase for one of the growth curves obtained for SLYC18 grown semi-aerobically in M2SF$^+$ medium. The lag phase truncation is indicated by the time shown in the growth curve boxes. The variances of the last three fitted curves are indicated inside their residual boxes.
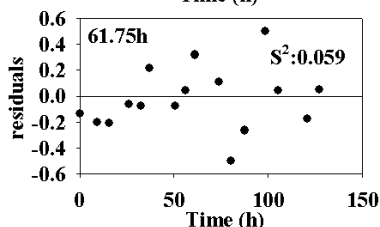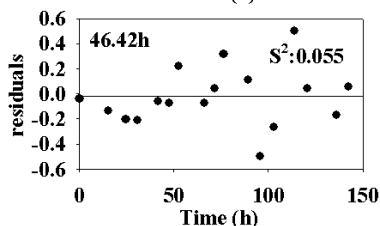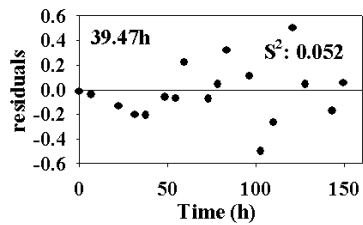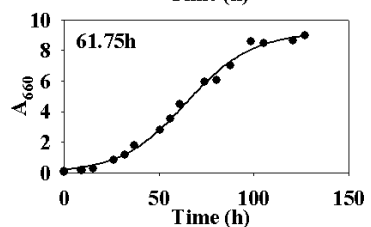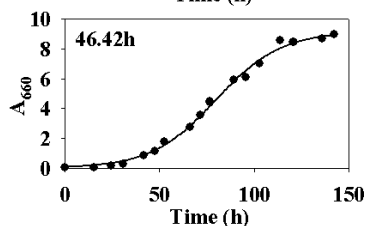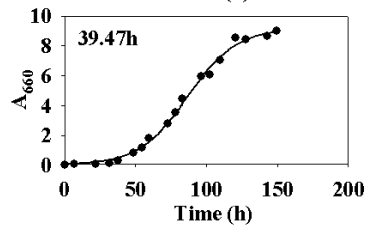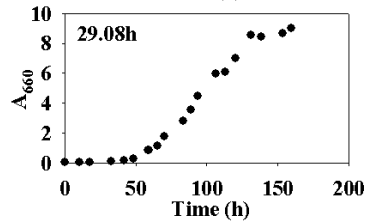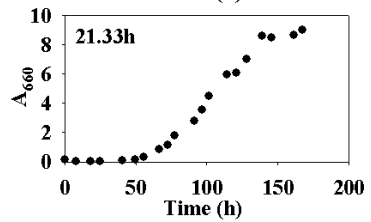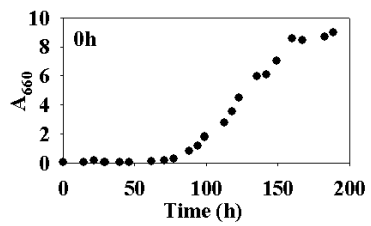
Supplementary Table 2

Strain	Medium					
	M2SF			M2SF$^+$		
	Growth rate (r) ($A_{660}.h^{-1}$)	Generation time (h)	Lag phase (h)	Growth rate (r) ($A_{660}.h^{-1}$)	Generation time (h)	Lag phase (h)
S1	0.12±0.01	5.99±0.53	14.7	0.06±0.00	12.42±0.38	21.8
SLYC18	0.10±0.00	7.07±0.14	52.5	0.06±0.00	11.42±0.52	39.5
ST4	0.11±0.00	6.44±0.08	22.8	0.05±0.01	13.81±2.02	46.6

Supplementary Fig. 3

Sequence homologies of functionally defined phytoene desaturases (CrtIs) from various sources

Homology set 1: Sequence homologies of all the CrtIs considered within the present study.

```
            1         10        20        30        40        50        60        70        80        90        100       110       120       130
rrCrtI      MKLHPAP---TVPTQDRRP H AVVIGSGFGGLAAAVRLGARGYRVTVLEKLGGPGGRGGVFKQDGFTFDAGPTIVT L PSLFEELWTLCGRQMSDDVDLRPLTPFYTIRFNDGETFRAQSDPEAMRAEVARL
rpCrtI      MLDPGLNPAPRPSRDRAP H AVVIGSGFGGLAAAVRLGAKGYRVTVLEKLDKAGGRAYVHKQDGFSFDAGPTIVT A PYLFEELWKLCGKRMSDDITLKPMSPFYRIRFDDGTHFDYSDDRDAVLDQIAKF
rgCrtI      MKSDAAAPRAAGKQDPQAP V ALVVIGSGFGGMAAAVRLAAKGYRVTVLEKLDAPGGRAYVHRREGHVFDAGPTIVT V PYLFDELWALAGRKFSDDIELKSLDPFYRIRFDDGDHFDYSGDPARMRAEVRRI
rcCrtI            MSKNTEGMG R AVVIGAGLGGLAAAMRLGAKGYKVTVDRLDRPGGRGSSITKGGHRFDLGPTIVT V PDRLRELWADCGRDFDKDVSLVPMEPFYTIDFPDGEKYTAYGDDAKVKAEVARL
rsCrtI         MPSISPASDAD R ALVIGSGLGGLAAAMRLGAKGWRVTVIDKLDVPGGRGSSITQEGHRFDLGPTIVT V PQSLRDLWKTCGRDFDADVELKPIDPFYEVRWPDGSHFTVRQSTEAMKAEVARL
paCrtI                   MK P TTVIGAGFGGIALAIRLQAAGIPVLLLEQRDKPGGRAYVYEDQGFTFDAGPVTIT D PSAIEELFALAGKQLKEYVELLPVTPFYRLCWESGKVFNYDNDQTRLEAQIQQF
psCrtI                   MK P TTVIGAGFGGIALAIRLQAAGIPVLLLEQRDKPGGRAYVYQEQGFTFDAGPVTIT D PSAIEELFALAGKQLKDYVELLPVTPFYRLCWESGKVFNYDNDQAQLEAQIQQF
rsCrtIATCC17025          Q(5)                                                                       D(5)
```

```
                      140       150       160       170       180       190       200       210       220       230       240       250
rrCrtI      SPGDVEGYERFIAESEDIYKVGFEQ M A DKPFSSITDMLKVVPQ M LKLGALRSVWGHVAKHVADPRLRIALSFHPLFIGGNPLSVTSIYSMI A FLERR W GVHFAMGGTHSLVKGLVGLIEGQGNHLRY
rpCrtI      CPDDVPAYDRFMAASHEIFKVGFEQ L G DQPFSHFTDMLKIAPA M IKLESYRSVYGLVAKHFKDPKLRQVFSFHPLLIGGNPFMSSSVYCLI T YLEKQ W GVHSAMGGTGALVTGLVNLIEGQGNTIRY
rgCrtI      SPSDAEGFEREMREADQCYELGFRT L G DKAFDTVGDLIKAAPL I VKLRGWRSLHQMVSSHLKHPKLRIAMSLQSLLIGGNPFSVTSMYALV N ALERQ W GVHWAMGGTGELIRGLVDVFEGMGGTMRL
rcCrtI      SPGDVEGFRHEMWDAKARYEFGYEN L G RKPMSKLWDLIKVLPT F GWLRADRSVYGHAKKMVKDDHLRFALSFHPLFIGGDPFHVTSMYILV S QLEKK F GVHYAIGGVQAIADAMAKVITDQGGEMRL
rsCrtI      SPGDVAGYEFLKDSEKRYWFGYED L G RRSMHKLWDLIKVLPT F GMMRADRTVYQHAALRVKDERLRMALSFHPLFIGGDPFNVTSMYILV S QLEKE F GVHYAIGGVAAIAAAMAKVIEGQGGSFRM
paCrtI      NPRDVEGYRQFLDYSRAVFKEGYLK L G TVPFLSFRDMLRAAPQ L AKLQAWRSVYSKVASYIEDEHLRQAFSFHSLLVGGNPFATSSIYTLI H ALERE W GVWFPRGGTGALVQGMIKLFQDLGGEVVL
psCrtI      NPRDVAGYRAFLDYSRAVFNEGYLK L G TVPFLSFKDMLRAAPQ L AKLQAWRSVYSKVAGYIEDEHLRQAFSFHSLLVGGNPFATSSIYTLI H ALERE W GVWFPRGGTGALVNGMIKLFQDLGGEVVL
rsCrtIATCC17025          H(5) S(4)                I(5)                                                   S(4)        S(5)
```

```
            260       270       280       290       300       310       320       330       340       350       360       370       380       390
rrCrtI      NAGVAEILVKNDRACGV-RLETGEEIAADIVVSNADSAATYKNLLPAAHRRWWSDRRLAWSRYSMSLFIWHFGTKG---TYPDVAHHTILMGPRYHTLIADIFQAKTLAKDFSL L Y LYRPTATDPSMAPEGCDS
rpCrtI      NQDVRQIVVENGTACGV-KLADGEVIKADIVVSNADSASTYRYLLPPETRKRWTDAKIEKSRYSMSLFVWYFGTKR---RYEDVKHHTILLGPRYKELISDIFSRKVVAEDFSL Y LHRPTATDPSLAPQGCDT
rgCrtI      KAEVKRIEVDNGVATGV-TLADGERIPADIVVCNGDTGYLYKNIVDARWRKHWTDARIERGHYSMGLFVWYFGTDR---RYEDVPHHMMVLGPRYRELLDDIFRKKKLASDFSIY LHRPTATDPSMAPAGCDT
rcCrtI      NTEVDEILVSRDGKATGIRLMDGTELPAQVVVSNADAGHTYKRLLRNRDRWRWTDEKLDKKRWSMGLFVWYFGTKGTAKMWKDVGHHTVVVGPRYKEHVQDIFIKGELAEDMSL Y VHRPSVTDPTAAPKGDDT
rsCrtI      NTEVDEILVEKGTATGV-RLASGEVLRAGLVVSNADAGHTYMRLLRNHPRRWTDAHVKSRRWSMGLFVWYFGTKGTKGMWPDVGHHTIVNAPRYKGLVEDIFLKGKLAKDMSL Y IHRPSITDPTVAPEGDDT
paCrtI      NARVSHMETTGNKIEAV-HLEDGRRFLTQAVASNADVVHTYRDLLSQHPAAVKQSNKLQTKRMSNSLFVLYFGLNH---HHDQLAHHTVCFGPRYRELIDEIFNHDGLAEDFSL Y LHAPCVTDSSLAPEGCGS
psCrtI      NARVSHMETVGDKIQAV-QLEDGRRFETCAVASNADVVHTYRDLLSQHPAAAKQAKKLQSKRMSNSLFVLYFGLNH---HHDQLAHHTVCFGPRYRELIHEIFNHDGLAEDFSL Y LHAPCVTDPSLAPEGCGS
rsCrtIATCC17025                                                                                                          C(4)
```

```
                      400       410       420       430       440       450       460       470       480       490       500       510       520
rrCrtI      FYVLSPVPHLGGGQ--DWALEAEPYRKAILAELERTVLPGLGAKIVTSKVVTPVDFRDGLSSLNGAAFGLEPAFTQSAWFRPHNKSEEIDRLYLVGAGTHPGAGLPGVLSSAKILDTVVPDASVL V
rpCrtI      FYVLSPVPNLLGDT--DWHTKAETYRASIAKMLGATVLPDLENQIATSKITTPIDFQDRLSSFRGAAFGLEPVLWQSAWFRPHNQSEDVKRLYLVGAGTHPGAGLPGVLSSARVLDALVPEADSL V TS
rgCrtI      FYALMPVPHLGSGT--DWTTQAEPYRQSVQEALERTVLPGLGQHLRVSFCTTPLDFQHRLLSYKGAGFGLEPLLLQSAYFRPHNRSEDVKNLFMVGASTHPGAGVPGIMSAKALESVLPDPATF R R
rcCrtI      FYVLSPVPNLGFDNGVDWSVEAEKYKAKVLKVIEERLLPGVAEKITEEVVFTPETFRDRYLSPLGAGFSLEPRILQSAWFRPHNASEEVDGLYLVGAGTHPGAGVPSVIGSGELVAQMIPDAPKP E TPAA
rsCrtI      FYALSPVPHLKQAQPVDWQAVAEPYRESVLEVLEQS-MPGIGERIGPSLVFTPETFRDRYLSPWGAGFSIEPRILQSAWFRPHNISEEVANLFLVGAGTHPGAGVPGVIGSAEVMAKLAPDAPRA R REAE
paCrtI      YYVLAPVPHLGTAN-LDWTVEGPKLRDRIFAYLEQHYMPGLRSQLVTHRMFTPFDFRDQLNAYHGSAFSVEPVLTQSAWFRPHNRDKTITNLYLVGAGTHPGAGIPGVIGSAKATAGLMLEDLI
psCrtI      YYVLAPVPHLGTAN-LDWAVEGPRLRDRIFDYLEQHYMPGLRSQLVTHRMFTPFDFRDELNAWQGSAFSVEPILTQSAWFRPHNRDKHIDNLYLVGAGTHPGAGIPGVIGSAKATAGLMLEDLI
rsCrtIATCC17025                                                                                                                        G(5)
```

```
            530
rrCrtI
rpCrtI
rgCrtI
rcCrtI      AAPKARTPRAKAAQ
```

```
rsCrtI      PAERLAAE
paCrtI
psCrtI
rsCrtIATCC17025
```

We have shown the comparison using a functional definition: (top) lycopene-producing CrtIs from photosynthetic sources (*R. rubrum* (rrCrtI) and *Rps. palustris* (rpCrtI)); the lycopene+neurosporene-producing CrtI from the photosynthetic *Rvi. gelatinosus* (rgCrtI); neurosporene-producing CrtIs from photosynthetic sources (*Rb. capsulatus* (rcCrtI) and *Rb. sphaeroides* (rsCrtI)), and lycopene-producing CrtIs from non-photosynthetic sources (*P. ananatis* (paCrtI) and *P. stewartii* (psCrtI)). The accession numbers for all the CrtIs are as follows: rrCrtI: EMBL ABC21298.1, rpCrtI: EMBL ACF00227.1, rgCrtI: EMBL BAA94063.1, rcCrtI: EMBL CAA36533.1, rsCrtI: EMBL AAB31138.1, paCrtI: EMBL BAA14127.1, psCrtI: EMBL AAN85599.1. The black and grey shadings indicate amino acid residues conserved throughout the family and only the photosynthetic CrtIs, respectively. The bottom line under the open boxes shows the amino acid changes found by (1) Wang and Liao (5, S5) and also by (2) Stickforth and Sandmann (4, S4) to increase the product yield of lycopene using the CrtIs from *Rb. sphaeroides* ATCC 17025 (note that this has a slightly different amino acid sequence to that from the more well-studied *Rb. sphaeroides* 2.4.1 shown in the comparison) and *Rvi. gelatinosus*, respectively. It can be seen that, in general, these amino acid changes do not correspond to the amino acids in either photosynthetic or non-photosynthetic, lycopene-producing CrtIs (open boxes). Exceptions are the V68D mutation of *Rb. sphaeroides* CrtI, which is also found in the *Pantoea* CrtIs, and the F166I mutation of *Rb. sphaeroides* CrtI, which corresponds to the *Rvi. gelatinosus* CrtI sequence.(The sequence number 68 and 166 have been taken from the original literature(S5), respectively corresponding to 76 and 174 in our numbering system.).

Homology set 2: Sequence homologies of only the lycopene-producing CrtIs considered within the present study.
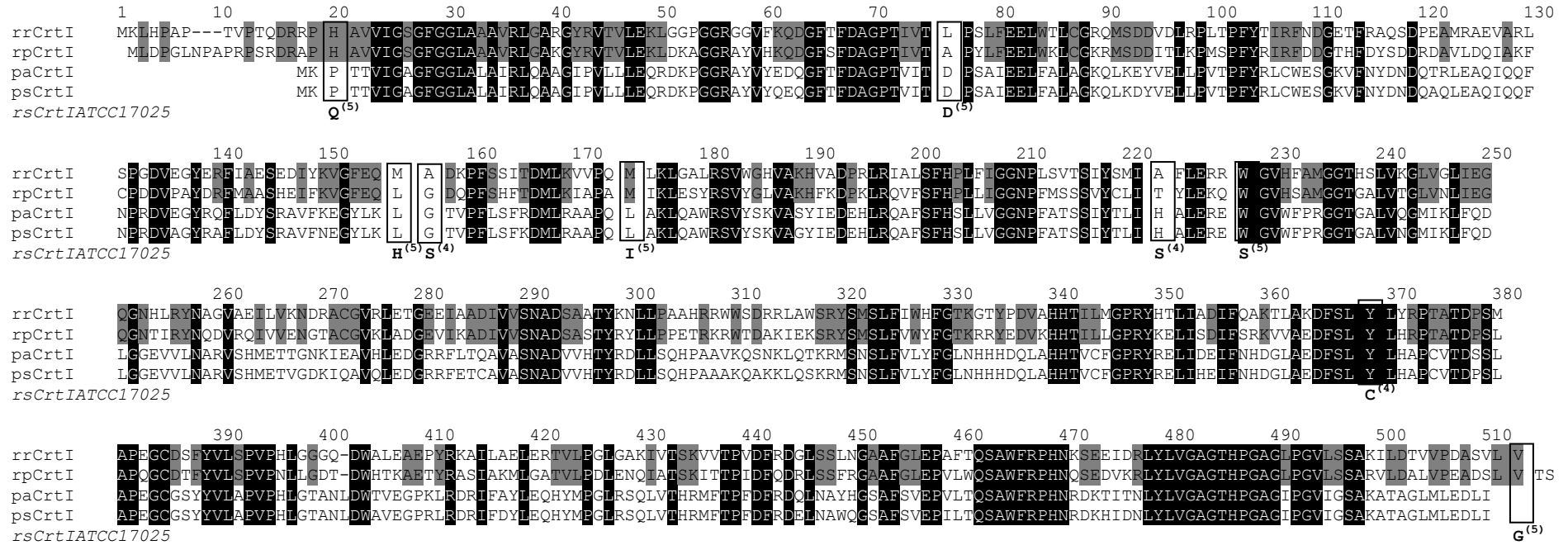
A comparison of the homology sets 1 and 2 shows that the lycopene-producing CrtIs show significantly more sequence homology (37.2%) than those observed for the whole family (22.5%). However, the sequences of the lycopene-producing CrtIs from the photosynthetic organisms have less identity to each other (64.1%) than to those from the nonphotosynthetic *Pantoea* sequences (93.5%).

Supplementary references

S1. **Jensen, S. L., G. Cohen-Bazire, T. O. M. Nakayama, and R. Y. Stanier.** 1958. The path of carotenoid synthesis in a photosynthetic bacterium. Biochim. Biophys. Acta. **29**:477-498.

S2. **Novak, M. A.** 2006. Evolutionary dynamics: exploring the equations of life. p. 13. Harvard University Press.

S3. **Ravelo-Pérez, L. M., J. Hernández-Borges, M. Á. Rodríguez-Delgado, and T. Borges-Miquel.** 2008. Spectrophotometric analysis of lycopene in tomatoes and watermelons: A practical class. Chem. Educator. **13**:10.1333/s00897082104a.

S4. **Stickforth, P., and G. Sandmann.** 2007. Kinetic variations determine the product pattern of phytoene desaturase from *Rubrivivax gelatinosus*. Arch. Bioch. Biophys. **461**:235-241.

S5. **Wang, C.-W., and J. C. Liao.** 2001. Alteration of the product specificity of *Rhodobacter sphaeroides* phytoene desaturase by directed evolution. J. Biol. Chem. **276**:41161-41164.

Chapter 4   High-level of β-carotene expression in *R. rubrum*

4.1. Introduction

The high medical and nutrient value of astaxanthin has drawn the interest of more and more groups to try to produce it at high level. In the biosynthetic pathway of astaxanthin, β-carotene is a very important intermediate which is also the precursor of vitamin A in mammals and is utilized industrially as food or feed supplements (Krinsky, 1989; Miura, *et al.*, 1998).

Non-carotenogenic microbes have been engineered to produce β-carotene by introduction of exogenous β-carotene synthesis genes from carotenogenic organisms, for example, *Erwinia* species (*Erwinia herbicola* (now called *Pantoea stewartii*), *Erwinia uredovora* (now called *Pantoea ananatis*)), marine bacteria (*Agrobacterium aurantiacum*, *Alcaligens sp.*) and plants.
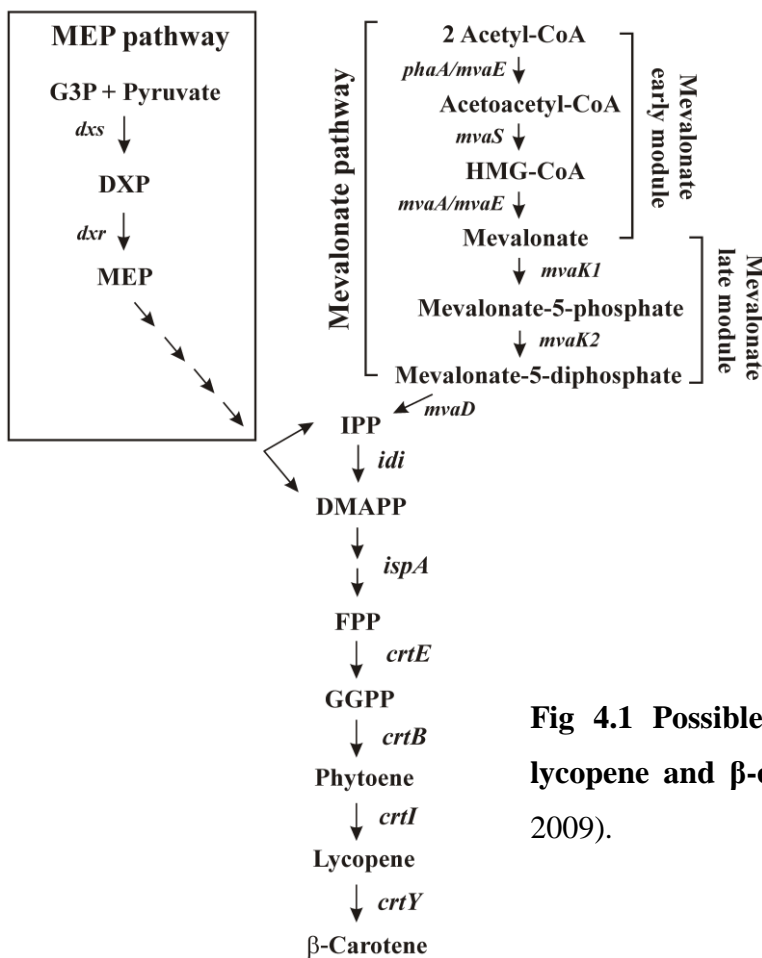
Fig 4.1 Possible pathways for the production of lycopene and β-carotene (adapted from Yoon *et al.*, 2009).

So far, the most commonly employed organism for the expression of β-carotene is *E. coli*, which is non-carotenogenic but can produce isoprenoid compounds using FPP as precursor. In general, isoprenoid precursors are produced from the glycolytic intermediates through 1-deoxy-D-xylulose-5-phosphate (DXP) and 2-C-methyl-D-erythritol-4-phosphate (MEP) in bacteria, cyanobacteria and plant plastids (**Fig 4.1**). In the plant cytoplasm, the isoprenoid precursors are obtained from the

mevalonate (MVA) pathway which is composed of two distinguishable modules, the MVA early module (from acetyl-CoA to mevalonate through HMG-CoA and the MVA late module (from mevalonate to IPP in three steps catalyzed by *mvaK1, mvaK2*, and *mvaD*). Strategies that have been used to produce β-carotene in *E. coli* have either engineered the MEP pathway to increase IPP and DMAPP, or have introduced a heterologous MVA pathway to increase the flux directing to IPP. In an early study, D-1-deoxyxylulose 5-phosphate synthase (DXS) and *crtEBI* from *E. uredovora* were overexpressed in *E. coli* (Matthews and Wurtzel, 2000) and lycopene levels of **1.3 mg/g DW** were obtained. Later, Yoon *et al.* (2006, 2008) introduced the late module of the MVA pathway and IDI from *Streptococcus pneumonia* into *E. coli* which was cultivated in 2YT medium containing 2% (w/v) glycerol and 6.6 mM MVA to achieve very high lycopene levels of **102 mg/l** and up to **503 mg/l** β-carotene. To solve the problem of the addition of exogenous MVA for the industrial scale, Yoon *et al.* (2009) constructed a complete MVA pathway in *E. coli* by introducing two expression vectors carrying the MVA early module from *Enterococcus faecalis* and the MVA late module from *S. pneumonia* respectively. The recombinant *E. coli* was cultivated in 2YT medium containing 2% glycerol and able to produce **465 mg/l** β-carotene.

The industrially important conventional yeast, *Saccharomyces cerevisiae*, is also non-carotenogenic but can accumulate ergosterol from FPP by the sterol biosynthesis pathway. Yamano *et al.* (1994) directed the carbon flux partially from ergosterol biosynthesis pathway into that of carotenoid by introducing expression vectors containing *crtEBI* and *crtEBIY* from *E. uredovora* respectively into *S. cerevisiae* and were able to produce lycopene and β-carotene at levels of **0.11 mg/g DW** and **0.10 mg/g DW**, respectively. Later, Miura *et al.* (1998) constructed plasmids to contain *crtEBIY* from *E. uredovora*, as well as *crtEBIY*/*crtWZ* from *E. uredovora /A. aurantiacum* which were modified according to *C. utilis* codon usage. The engineered *C. utilis* harboring those plasmids were able to produce lycopene, β-carotene and astaxanthin at levels of **1.1 mg/g DW, 0.4 mg/g DW** and **0.4 mg/g DW**, respectively. In a recent strategy Yan *et al.* (2012) overexpressed HMG-CoA reductase and *crEBIY* from *Xanthophyllomyces dendrorhous* to produce β-carotene of **6.29 mg/g DW** in the presence of an ergosterol synthesis inhibitor.

There has been only one study (Hunter *et al.*, 1994) documented so far which has attempted to produce β-carotene in purple photosynthetic bacteria. In this latter work, a 12.8 kb crt cluster containing *crtBIYZ* from *E. herbicola* was introduced into different *Rb. sphaeroides* carotenoid mutants on an expression vector, and β-carotene was obtained at very low level. The highest β-carotene level achieved was 88% β-carotene content of the total carotenoid with a *crtD⁻ Rb. sphaeroides* mutant with *crtBIYZ* on vector pRER1B. There are several points in the study of

Hunter *et al.* (1994) which will be discussed in details in the "Discussion": (1) Although it was reported that β-carotene functioned successfully in the LH complexes, the growth rate under photoheterotrophic conditions was unobtainable. (2) The LH1 complexes level achieved with β-carotene as major carotenoid was only 20% compared to that achieved with spheroidenone (Hunter *et al.*, 1994), and as indicated by the absorption spectrum shown, the total ICM level was very low. (3) It was also mentioned that the stress imposed by photosynthetic growth on the β-carotene-producing strain could be avoided by aerobic growth and that there might be some shift in the absorbance peaks of LH1 and LH2 due to the presence of different carotenoids.

Carotenoids with cyclic groups are very common in plant species. However, the corresponding carotenoid genes are rarely applied for engineering non-carotenogenic organisms to produce cyclic carotenoids. Plant lycopene cyclases, belonging to the CrtL family (Yu *et al.*, 2010), are categorized into lycopene β-cyclases and lycopene ε-cyclases, which are capable of producing β or ε-ionone end groups, respectively (**Fig 4.2**) (Cunningham and Gantt, 2001). The lycopene β-cyclase introduces β-rings on both ends of lycopene and forms β-carotene, whereas the lycopene ε-cyclase catalyzes δ-carotene by introduction only one ε-ring to lycopene. In the presence of two cyclases, α-carotene is produced by addition of β-ring and ε-ring on two ends of lycopene.
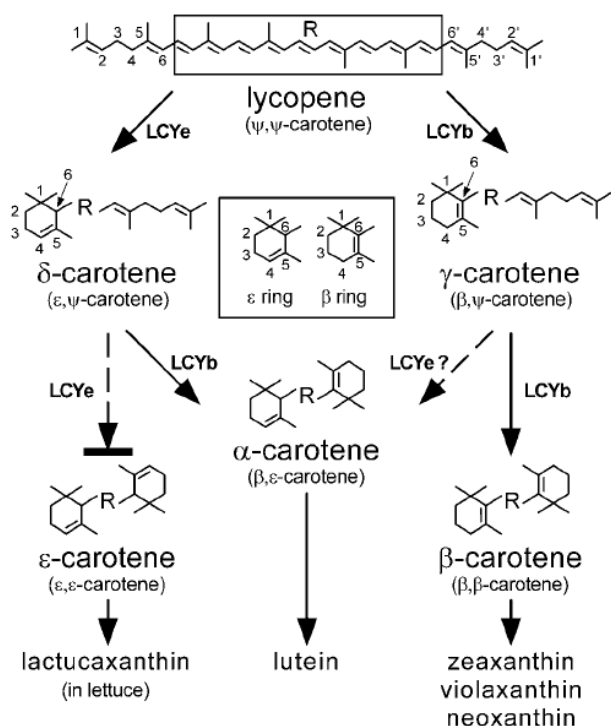
Fig 4.2 Biosynthesis pathway from lycopene to cyclic carotenoids in plants (taken from Cunningham and Gantt, 2001). LCYe and LCYb are the abbreviations for lycopene ε-cyclase and lycopene β-cyclase, respectively.

In this study, two different strategies were applied for β-carotene expression in *R. rubrum* by insertion of *A. thaliana* lycopene β-cyclase (In this study we maintain the designation CrtL, the same enzyme is referred to LCYb in plant and EBI databases). The reason of choosing the *A.*

thaliana CrtL is due to its known genome sequence data with easy access at the European Arabidopsis Stock centre (uNASC).

4.2. Construction of a β-carotene production strain of *R. rubrum*

4.2.1. Initial trial of the β-carotene production in *R. rubrum* (This early work was performed by R. Ghosh, H. Grammel, A. Heyer, unpublished data)

As we proved that the carotenoid lycopene is incorporated into *R. rubrum* LH1 complexes, we assume that β-carotene should also localize to the ICM. ICM expression is under control of the *reg* signal transduction pathway which regulates genes expression in the PGC including the *puf* operon. The *puf* operon of *R. rubrum* encodes the LH1 complex and the RC, whose expression is regulated by the O_2 and redox levels (Belanger and Gingras, 1988). Under conditions of low O_2 and high redox, the *R. rubrum* cell is filled with ICM (LH1+RC), which indicates that the capacity of *R. rubrum* cell for β-carotene should be high. As the insertion of recombinant genes into an expression vector is usually the standard route for heterologous protein expression, the initial strategy for the production of β-carotene was to express the *crtL* under control of the *puf* O_2 promoter *in trans* in *R. rubrum*. Heterologous gene expression can create translational problems in the host organism due to the presence of rare codons. As shown in **Table 4.1**, the *A. thaliana crtL* showed a rare codons (AGG/AGA) usage of up to 4.4 %, which is higher than the optimal for *E. coli* (Wada *et al.*, 1992). However, the *R. rubrum cycA* gene (encoding cytochrome c_2) has a rare codons (AGG/AGA) usage of 11.47% and still expresses the cytochrome c_2 at a very high level (5% of the total cellular protein). Therefore we considered the expression of an unusual CrtL in *R. rubrum* to be feasible.

```
  M   D   T   L   L   K   T   P   N   K   L   D   F   F   I   P   Q   F   H   G   F   E   R   L   C
ATG GAT ACT CTG TTG AAA ACA CCC AAC AAG CTC GAT TTT TTC ATC CCT CAG TTT CAT GGG TTT GAG AGA TTA TGC
  S   N   N   P   Y   H   S   R   V   R   L   G   V   K   K   R   A   I   K   I   V   S   S   V   V
AGT AAC AAT CCA TAC CAT TCA AGG GTT AGG CTT GGT GTG AAG AAA AGG GCT ATC AAA ATT GTC TCT AGT GTA GTG
  S   G   S   A   A   L   L   D   L   V   P   E   T   K   K   E   N   L   D   F   E   L   P   L   Y
AGT GGT AGC GCT GCT CTT TTG GAT CTT GTT CCT GAA ACT AAG AAG GAG AAT CTT GAC TTT GAG CTT CCT TTG TAC
  D   T   S   K   S   Q   V   V   D   L   A   I   V   G   G   G   P   A   G   L   A   V   A   Q   Q
GAC ACT TCC AAG AGT CAA GTT GTT GAT TTG GCT ATT GTT GGT GGT GGT CCT GCT GGT TTA GCC GTG GCT CAG CAG
  V   S   E   A   G   L   S   V   C   S   I   D   P   S   P   K   L   I   W   P   N   N   Y   G   V
GTT TCT GAA GCT GGA CTC TCT GTT TGT TCC ATT GAT CCT TCT CCT AAG CTC ATA TGG CCT AAC AAT TAT GGA GTT
  W   V   D   E   F   E   A   M   D   L   L   D   C   L   D   T   T   W   S   G   A   V   V   Y   V
TGG GTT GAT GAG TTT GAG GCT ATG GAT TTA CTA GAC TGC CTG GAT ACC ACA TGG TCT GGT GCT GTT GTC TAT GTC
  D   E   G   V   K   K   D   L   S   R   P   Y   G   R   V   N   R   K   Q   L   K   S   K   M   L
GAT GAA GGT GTC AAG AAG GAT TTG AGC CGG CCT TAT GGG AGA GTT AAC CGG AAA CAG CTC AAA TCC AAA ATG CTT
  Q   K   C   I   T   N   G   V   K   F   H   Q   S   K   V   T   N   V   V   H   E   E   A   N   S
CAG AAA TGT ATT ACC AAC GGT GTT AAA TTT CAT CAG TCT AAG GTC ACT AAT GTG GTT CAC GAG GAG GCA AAC TCC
  T   V   V   C   S   D   G   V   K   I   Q   A   S   V   V   L   D   A   T   G   F   S   R   C   L
ACT GTG GTC TGC AGT GAC GGT GTA AAG ATT CAG GCT TCC GTG GTT CTT GAT GCC ACT GGG TTT TCC CGA TGC TTG
  V   Q   Y   D   K   P   Y   N   P   G   Y   Q   V   A   Y   G   I   V   A   E   V   D   G   H   P
GTT CAG TAT GAC AAA CCT TAC AAC CCT GGG TAC CAA GTA GCT TAC GGG ATT GTA GCT GAA GTT GAT GGT CAC CCA
  F   D   V   D   K   M   V   F   M   D   W   R   D   K   H   L   D   S   Y   P   E   L   K   E   R
TTC GAT GTA GAC AAA ATG GTG TTC ATG GAT TGG AGA GAC AAA CAT CTG GAC TCA TAT CCT GAG CTG AAA GAA CGG
  N   S   K   I   P   T   F   L   Y   A   M   P   F   S   S   N   R   I   F   L   E   E   T   S   L
AAC AGC AAG ATC CCA ACG TTC TTG TAC GCT ATG CCA TTT TCT TCC AAC CGA ATA TTT CTT GAA GAA ACT TCT TTA
  V   A   R   P   G   L   R   M   E   D   I   Q   E   R   M   A   A   R   L   K   H   L   G   I   N
GTT GCT AGA CCT GGT CTG AGA ATG GAA GAT ATC CAA GAA AGA ATG GCT GCT AGA CTG AAA CAT CTG GGG ATC AAT
  V   K   R   I   E   E   D   E   R   C   V   I   P   M   G   G   P   L   P   V   L   P   Q   R   V
GTG AAG AGG ATT GAG GAA GAC GAG CGT TGT GTG ATC CCG ATG GGC GGT CCT TTA CCA GTC TTA CCT CAA CGG GTT
  V   G   I   G   G   T   A   G   M   V   H   P   S   T   G   Y   M   V   A   R   T   L   A   A   A
GTG GGG ATT GGT GGG ACA GCA GGA ATG GTT CAT CCT TCA ACT GGT TAC ATG GTT GCT AGG ACT CTT GCA GCT GCA
  P   I   V   A   N   A   I   V   R   Y   L   G   S   P   S   S   N   S   L   R   G   D   Q   L   S
CCA ATA GTT GCA AAT GCC ATT GTG AGA TAC CTC GGT TCA CCA AGT AGT AAT AGC CTG AGA GGA GAT CAA CTC TCT
  A   E   V   W   R   D   L   W   P   I   E   R   R   Q   R   E   F   F   C   F   G   M   D   I
GCT GAG GTT TGG AGA GAC TTG TGG CCT ATC GAA CGG CGT AGA CAG AGG GAG TTC TTC TGT TTT GGA ATG GAT ATT
  L   L   K   L   D   L   D   A   T   R   R   F   F   D   A   F   F   D   L   Q   P   H   Y   W   H
CTG CTG AAA CTC GAT TTA GAC GCT ACT AGA AGG TTC TTT GAT GCA TTC TTT GAT CTG CAA CCT CAT TAC TGG CAC
  G   F   L   S   S   R   L   F   L   P   E   L   L   V   F   G   L   S   L   F   S   H   A   S   N
GGA TTC TTG TCT TCC AGG CTG TTT CTC CCG GAA CTG TTG GTC TTC GGG TTG TCG CTC TTC TCA CAC GCT TCC AAT
  T   S   R   L   E   I   M   T   K   G   T   V   P   L   A   K   M   I   N   N   L   V   Q   D   R
ACC TCA AGA TTG GAG ATC ATG ACA AAG GGG ACT GTT CCT CTT GCT AAG ATG ATC AAC AAT TTG GTA CAA GAT AGA
  D
GAC TAA
```

Table 4.1 The DNA sequence of *crtL* from *A*. *thaliana* and its translated polypeptide. The rare codons are highlighted in red.

The *A*. *thaliana crtL* gene, present on vector pU10665 and localized between the EcoRI/SalI sites, (**Fig 4.3**) was obtained from uNASC.
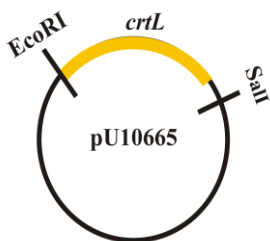
Fig 4.3 Gene map of vector pU10665. The lycopene β-cyclase (indicated in yellow) from *A*. *thaliana* presents on vector pU10665.

Several molecular biological studies have indicated the O_2-regulated *puf* O_2 promoter in *Rb. capsulatus* to be -733 to -751 bp upstream of *pufB*, and to be -679 to -729 bp upstream of *pufB* in *Rb. sphaeroides* (Adams *et al.*, 1989; Narro *et al.*, 1990). A similar sequence motif was identified in *R. rubrum* to be -521 to -571 bp upstream of *pufB* (Lee and Collins, 1993).
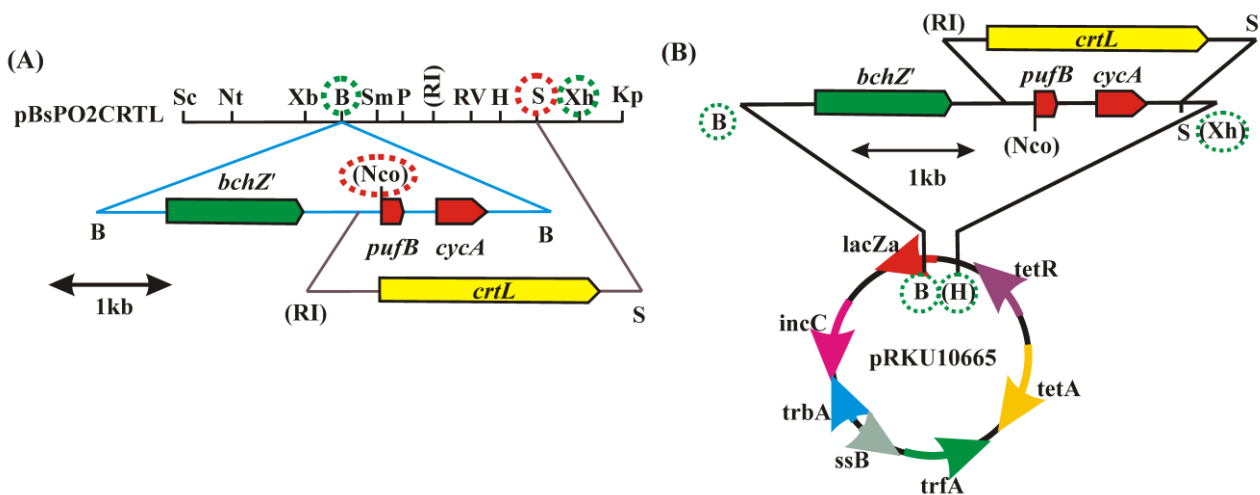
Fig 4.4 Gene maps of pBsPO2CRTL and pRKU10665. The vector pBsPO2CRTL contained the *puf* O_2 promoter and the *crtL* gene, which were present on the BamHI fragment (indicated in blue line) and the (NcoI)/SalI fragment (indicated in plum line), respectively (A). The BamHI/(XhoI) fragment (indicated in dotted green circles) of vector pBsPO2CRTL were subcloned into pRK404, yielding pRKU10665 (B).

A 2.7 kb BamHI fragment, containing *bchZ'* (*bchZ* upstream region, *bchZ*, *puf* O_2 promoter), *pufB* and *cycA*, was subcloned into the BamHI site of pBsKSII+ (**Fig 4.4 (A)**), yielding pBsPO2. The (NcoI)/SalI fragment (restriction site shown in parentheses has been blunt-ended) of pBsPO2 was subsequently replaced with the *crtL* gene (an (EcoRI)/Sal fragment), yielding pBsPO2CRTL. The BamHI/(XhoI) fragment, containing the *puf* O_2 promoter and *crtL*, was then subcloned into pRK404 which was BamHI/(HindIII) digested and dephosphorylated, yielding pRKU10665 (**Fig 4.4 (B)**). pRKU10665 was triparentally conjugated with SLYC18 and the HPLC-MS analysis of the recombinant *R. rubrum* (pRKU10665) showed that only about 5% of the total carotenoid produced corresponded to β-carotene, the remaining 95% due to lycopene (R. Ghosh, H. Grammel, A. Heyer, unpublished data). Therefore, the initial trial of β-carotene production, using a recombinant *R. rubrum* strain, was only partially successful.

4.2.2. An alternative strategy for the production of β-carotene in *R. rubrum*

4.2.2.1. Preliminary considerations

An alternative strategy for the production of β-carotene would be to insert the *A. thaliana crtL* gene into the *R. rubrum* chromosome and express the gene using an endogenous *cis* promoter. This strategy needs to consider several fundamental questions. (1) An appropriate chromosomal locus has to be chosen for foreign gene insertion. The SLYC18 mutant proved that a partial *crtC-crtD* deletion leads to lycopene (the substrate of the lycopene β-cyclase) production at wild-type levels

and that the *crtC* and *crtD* genes are dispensable. (2) The elements to be inserted into *R. rubrum* chromosome have to be decided as well. (a) **A strong promoter**. Since the *crtD* activity can saturate every LH1 complex with Spx under anaerobic and low light conditions for wild-type S1, the activity of *crtD* promoter is assumed to be strong. (b) **A synthetic ribosome binding site (rbs).** As the eukaryotic gene does not have a suitable rbs for prokaryotic organisms, a synthetic rbs needs to be inserted for maximal expression. (c) **Selection antibiotics**. The *npt* cassette encoding neomycin phosphotransferase (Lupo and Ghosh, 2004), derived from Tn*5*, is required as well. (3) Site-directed insertion of the rbs-*crtL-npt* fragment into the *R. rubrum* chromosome. The rbs-*crtL-npt* fragment, flanked by about 1217 bp on *R. rubrum* chromosomal DNA (the truncated *crtC* and *crtD* genes), is inserted into a conjugable suicide vector (pSUP202). After conjugation, the Kan^{R}Tet^S phenotype indicates a site-directed double recombination event.

4.2.2.2. Short overview of the construction of mutant SWGK46.

Considering the above requirements, the construction started using the 6.37-kb EcoRV fragment which contains *crtC*, *crtD* and *crtE*. The basic idea of the construction (**Fig 4.5**) was to replace the original BstEII fragment (partial *crtC-crtD*) with a synthetic BstEII linker (step1-4, corresponding to **Fig 4.6-4.8**) which contains several restriction enzyme sites for the insertion of foreign genes. Selection of restriction enzymes to be present on the BstEII linker should meet the criteria that those enzymes do not cut in the middle of EcoRV fragment, nor in the polylinker of construction vector, the *npt* cassette and the *crtL* gene. As we can see from **Table 4.2**, only MluI and SwaI meet these criteria and were chosen to be present on the linker. The BglII and PstI sites were chosen initially to facilitate the directional insertions of the *npt* and *crtL* cassettes in desired orientations, respectively (**Fig 4.5**). The *npt* cassette should be placed after the *crtL* gene (step 5-6, corresponding to **Fig 4.10-4.11**), ideally with both of them in the same orientation as *crtD*. *crtL* with the rbs, placed about 9-16 bp spaces in front of the start codon, should run from *crtD* promoter (step7-8, corresponding to **Fig 4.12-4.13**). Since the upstream region of the *crtD* is still present, a stop codon should be placed in front of the rbs (**Fig 4.10**) to prevent the creation of toxic fusion proteins. The insert was transferred from the construction vector to the suicide vector by two unique restriction enzymes (step 9, **Fig 4.14**), flanking the insert.
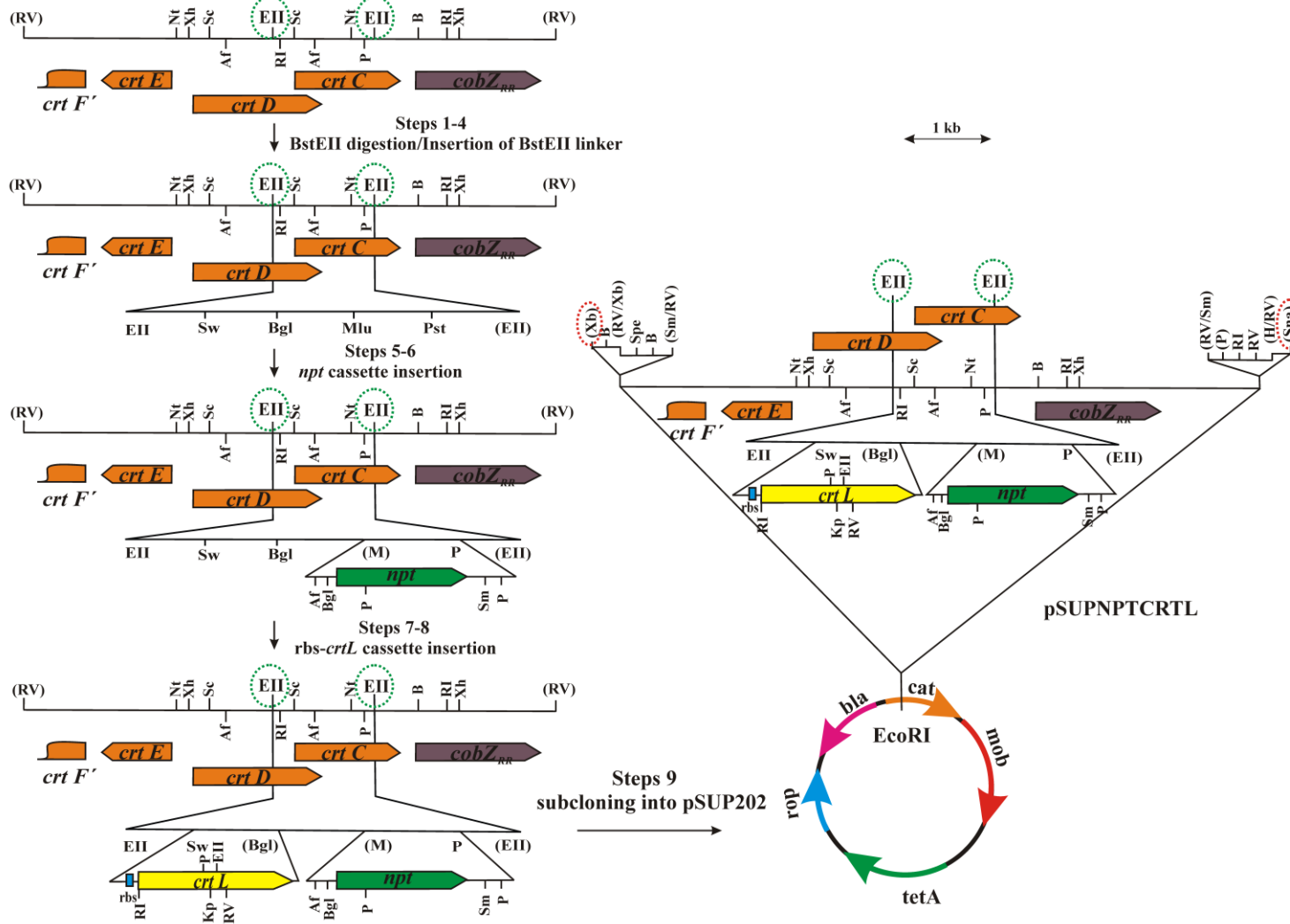
Fig 4.5 Overview of construction of suicide vector for site-directed insertion of *crtL* into the *crtD-crtC* locus. Four steps were taken to insert a synthetic BstEII linker to replace the partial *crtC-crtD* region. Both sides have flanking regions. Steps 5-6 are the processes of *npt* cassette insertion. Steps 7-8 placed the eukaryotic *crtL* in the same orientation as the *crtD* with a synthetic rbs in front of its starting codon. The last step transferred the construction from the construction vector to the suicide vector by two unique enzymes.

Table 4.2 The presence of restriction enzymes in pBsKSII+, the SGE5 insert, pBsWGE5, the *npt* cassette, the *crtL* gene and pSUP202.

Enzymes/ (Abb.)	pBsKSII+	SGE5	npt	crtL	pSUP202	pBsWGE5
SnaBI/Sna	0	0	0	0	0	1
MluI/M	0	0	0	0	0	0
SwaI/Sw	0	0	0	0	0	0
BglII/Bgl	0	0	1	0	0	0
BstEII/EII	0	2	0	1	1	2
SacI/Sc	1	2	0	0	0	2
NotI/Nt	1	2	0	0	0	2
XbaI/Xb	1	0	0	0	0	1
SpeI/Spe	1	1	0	1	0	1
BamHI/B	1	2	0	0	1	3
SmaI/Sm	1	13	1	0	0	13
PstI/P	1	2	2	1	1	1
EcoRI/RI	1	3	0	1	1	2
EcoRV/RV	1	1	0	1	1	1
HindIII/H	1	0	0	0	1	0
ClaI	1	2	0	0	1	2
SalI/S	1	5	0	0	1	5
XhoI/Xh	1	2	0	0	0	2
KpnI/Kp	1	0	0	1	0	1
HpaI	0	0	0	1	0	0
NdeI	0	0	0	1	1	0
PvuII	2	0	3	1	2	2
AfeI	0	2	0	1	6	1
NcoI/Nco	0	8	1	0	1	8
SphI	0	4	1	0	1	4
AflIII/Af	1	1	1	0	1	1
HindII	1	9	0	1	2	9

Note: All the abbrevations of the enzymes shown on the construction maps are listed in this table.

4.2.2.3. The nine steps for the construction of mutant SWGK46.

As mentioned in chapter 3, vector pBsSGE5 contains the 6.37-kb EcoRV fragment, however, it is not the ideal vector for the lycopene β-cyclase construction because of its complicated polylinker. Therefore, the first construction step was to simplify the polylinker of pBsKSII+.

Step 1

Some of the restriction enzyme sites in the polylinker of pBsKSII+ (indicated in red dotted line) was removed (**Fig 4.6**) by XhoI/BamHI digestion, dephosphorylated and ligated with a new synthetic XhoI-SnaBI-EcoRV-BamHI linker (**Table 4.3**), indicated by the green dotted line, yielding pBsWG1. The reasons of choosing these four restriction enzyme sites are as follows: First, the polylinker should be as simple as possible: XhoI/ BamHI digestion removed most of the restriction sites. Secondly, SnaBI and XbaI (see **Table 4.2**) are the candidates for the last step to transfer the construction as they are not present in the SGE5 insert, the *npt* cassette and the *crtL* gene. Thirdly, EcoRV site should be present on the linker for blunt-end ligation in the next step.
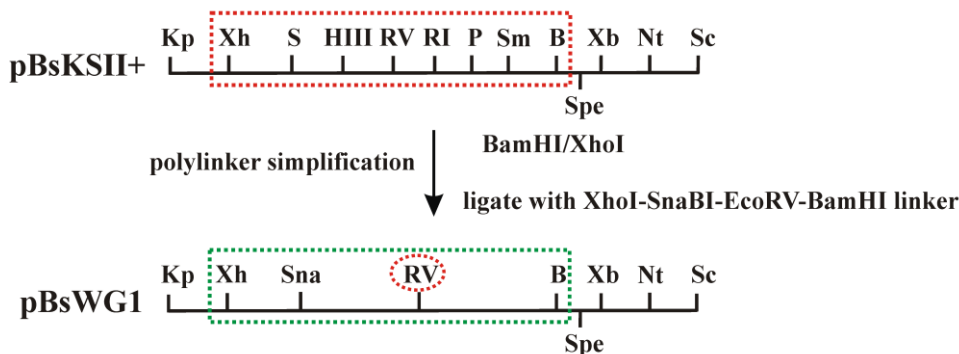
Fig 4.6 (step1) pBsKSII+ polylinker simplification and construction of pBsWG1/RV. Vector pBsWG1 was constructed by replacing the polylinker of pBsKSII+ with a short synthetic XhoI-SnaBI-EcoRV-BamHI linker.

Table 4.3 XhoI-SnaBI-EcoRV-BamHI linker.

Linker	Sequence
XhoI-SnaBI-EcoRV-BamHI linker	

Steps 2 and 3

The 6.42-kb SGE5 insert was obtained by (HindIII)/(XbaI) double digestion of pBsSGE5 (**Fig 4.7**, step 2), ligated with dephosphorylated pBsWG1/RV, yielding pBsWGE5 (**Fig 4.7**, step 3). pBsWGE5 is the *crtL* construction vector with SnaBI and XbaI on two ends which will be used to cut the construction out in the last step and transfer it to a suicide vector.
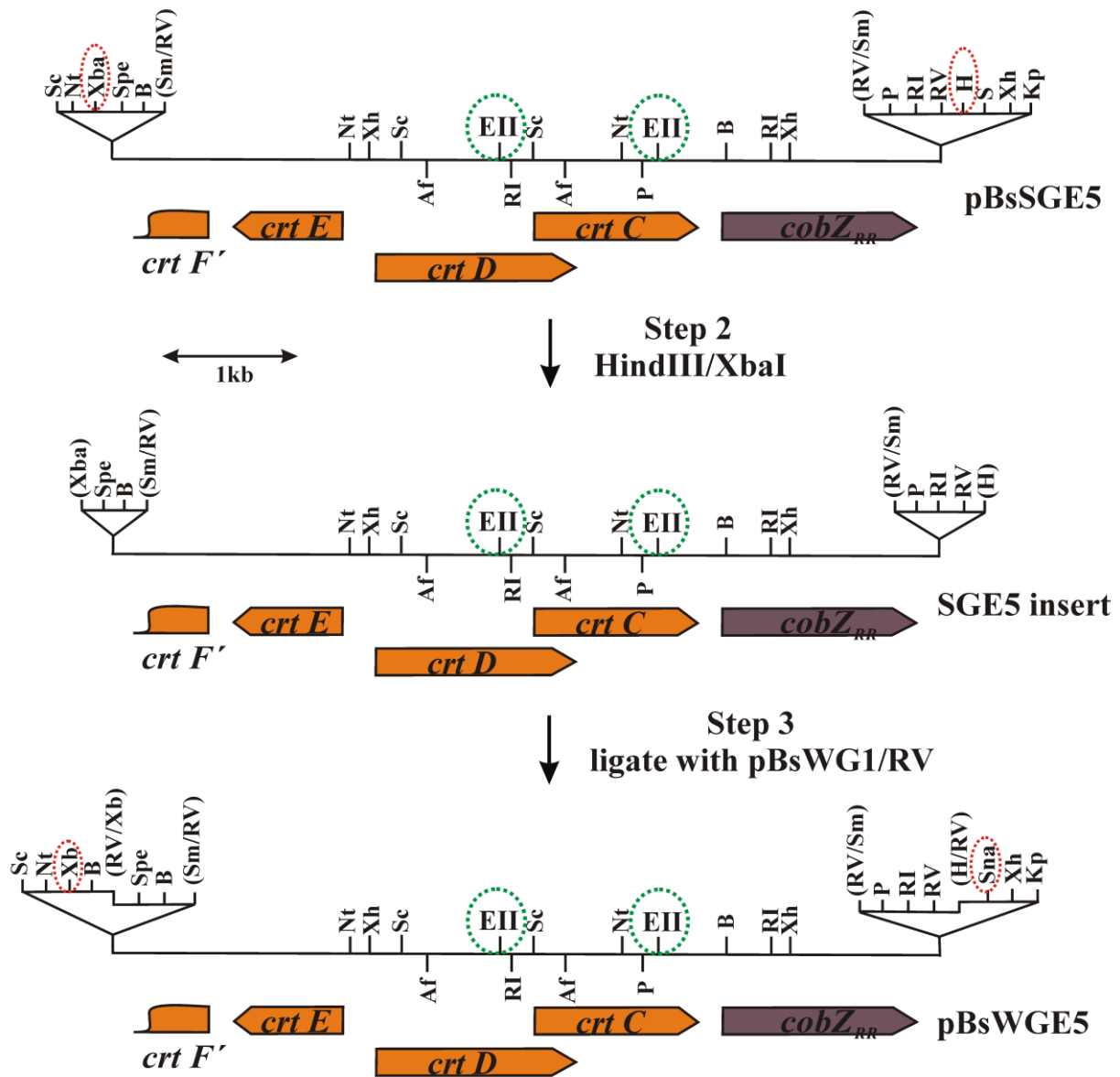
Fig 4.7 (steps 2 and 3) Gene maps of the SGE5 insert and the construction vector pBsWGE5. The SGE5 insert was a 6.42-kb (HindIII)/(XbaI) fragment. SnaBI and XbaI are indicated in red dotted circles, which were present on both ends of vector pBsWGE5.

Step 4

The PstI site present in the pBsWGE5 polylinker was removed by site-directed mutagenesis, yielding pBsWGE6ΔPst (**Fig 4.8**, step 4). A synthetic BstEII-SwaI-BglII-MluI-PstI-BstEII linker replaced the original BstEII fragment of pBsWGE6ΔPst, leading to partial *crtC-crtD* deletion and yielding pBsWGE8L.
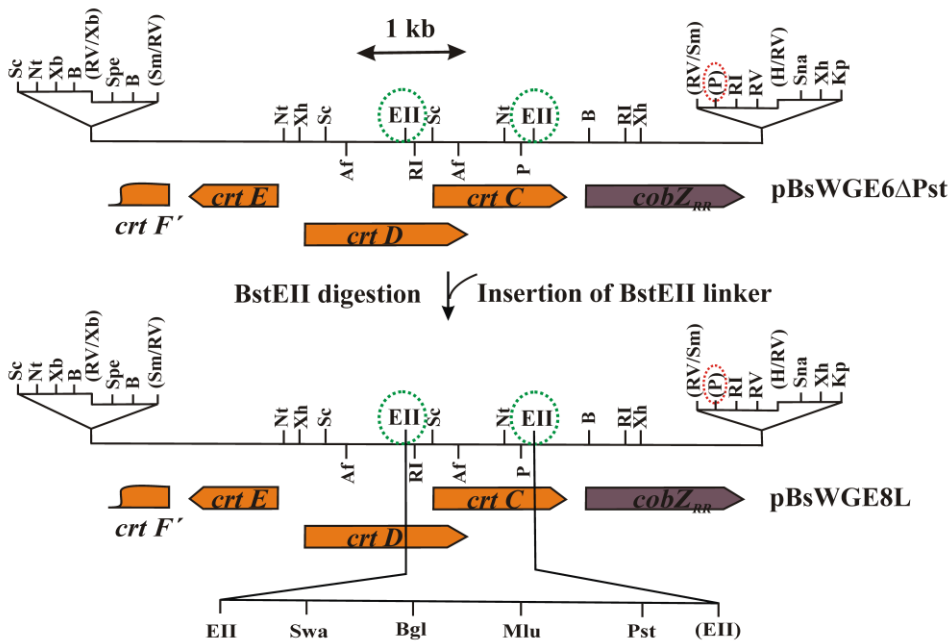
Fig 4.8 (step 4) Construction of pBsWGE8L. pBsWGE8L was constructed by BstEII digestion of pBsWGE6ΔPst, dephosphorylated and ligated with the BstEII-SwaI-BglII-MluI-PstI-BstEII linker.

SwaI and MluI are rare cutters and are not present in the SGE5 insert, the *npt* cassette and the *crtL* cassette. Two in-frame stop codons (indicated by green boxes, **Fig 4.9**) were incorporated into the BstEII linker to arrest the expression of the 5'-region of *crtD*, thus preventing the creation of toxic fusion proteins. SwaI was also chosen for the BstEII linker because there is an out-of-frame stop codon existing in the SwaI recognition site, which prevents an open reading frame arising due to ribosomal frame shifting.

Fig 4.9 BstEII synthetic linker. BstEII, SwaI, BglII, MluI, PstI sites were indicated in red boxes (A). Two in-frame and one out-of-frame stop codons (indicated in green boxes) were designed in the linker to ensure that the expression of the 5' residues of *crtD* stops in front of *crtL*. (B) BstEII linker fused in-frame with the rest of *crtC* and *crtD* on both ends.

Step 5

Initially, we planned to insert the *npt* cassette vectorially into the vector pBsWGE8L. The strategy was to construct a multipurpose vector which could also be used for a later cloning step with *crtL*. Vector pBsWG7 (**Fig 4.10**, step 5) was constructed by insertion of two synthetic linkers into pBsKSII+. The EcoRI-rbs-SwaI-BamHI linker was designed for placing the rbs in the front of the start codon of *crtL*. The XhoI-MluI-PstI-SalI linker (indicated in green dotted line) was designed with the goal of inserting the *npt* cassette in the same orientation as *crtD* into pBsWGE8L. The plan was to place *npt* cassette between the SalI and HindIII sites in vector pBsWG7 and cut it out with PstI/(HindIII), ligate with pBsWGE8L which is (MluI)/PstI digested and dephosphorylated. However, since there are two PstI sites (one of which was initially overlooked) in the *npt* cassette (**Table 4.2**), a simpler strategy was employed (random insertion) for *npt* cassette insertion (see step 6).
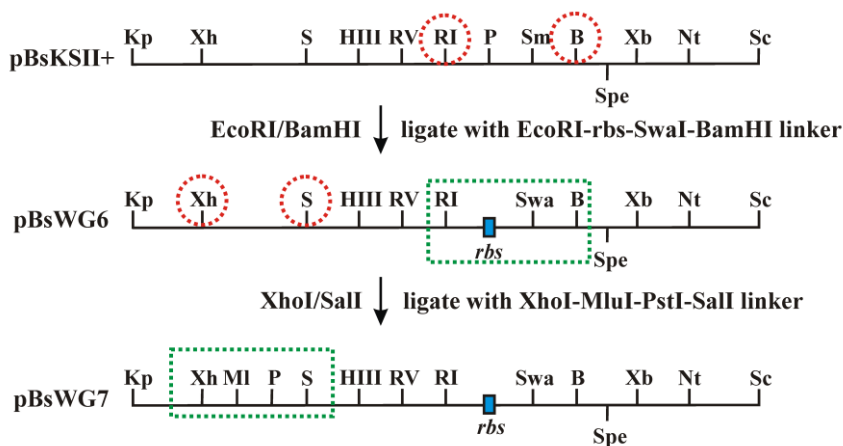
Fig 4.10 (step 5) Construction of the vector pBsWG7. pBsKS+ ligated with two synthetic linkers (indicated in green dotted line), yielding pBsWG6 and pBsWG7, respectively. pBsWG7 was digested with HindIII/SalI, ligated with the *npt* cassette, yielding pBsWG8NPT. The *npt* cassette was obtained by HindIII/SalI digestion of pBsWG8NPT followed by blunt-ending with the Klenow polymerase fragment.

Table 4.4 EcoRI-SD-SwaI-BamHI linker and XhoI-MluI-PstI-SalI linker.

Linkers	Sequences
EcoRI-SD-SwaI-BamHI linker	EcoRI SD SwaI BamHI 5′ AAT TCA TCC CTC CTC CAT TTA AAT GAG 3′ 3′ GT AGG GAG GAG GTA AAT TTA CTC CTAG 5′
XhoI-MluI-PstI-SalI linker	XhoI MluI PstI SalI 5′ TCG AGA CGC GTC CTG CAG 3′ 3′ CT GCG CAG GAC GTC AGCT 5′

Step 6

The *npt* cassette with (HindIII)/(SalI) blunt ends, was ligated with pBsWGE8L which was (MluI)/(PstI) digested and dephosphorylated. After selection, only pBsWGE9NPT (**Fig 4.11**, step 6) was obtained, with *npt* in the opposite orientation to *crtD*.
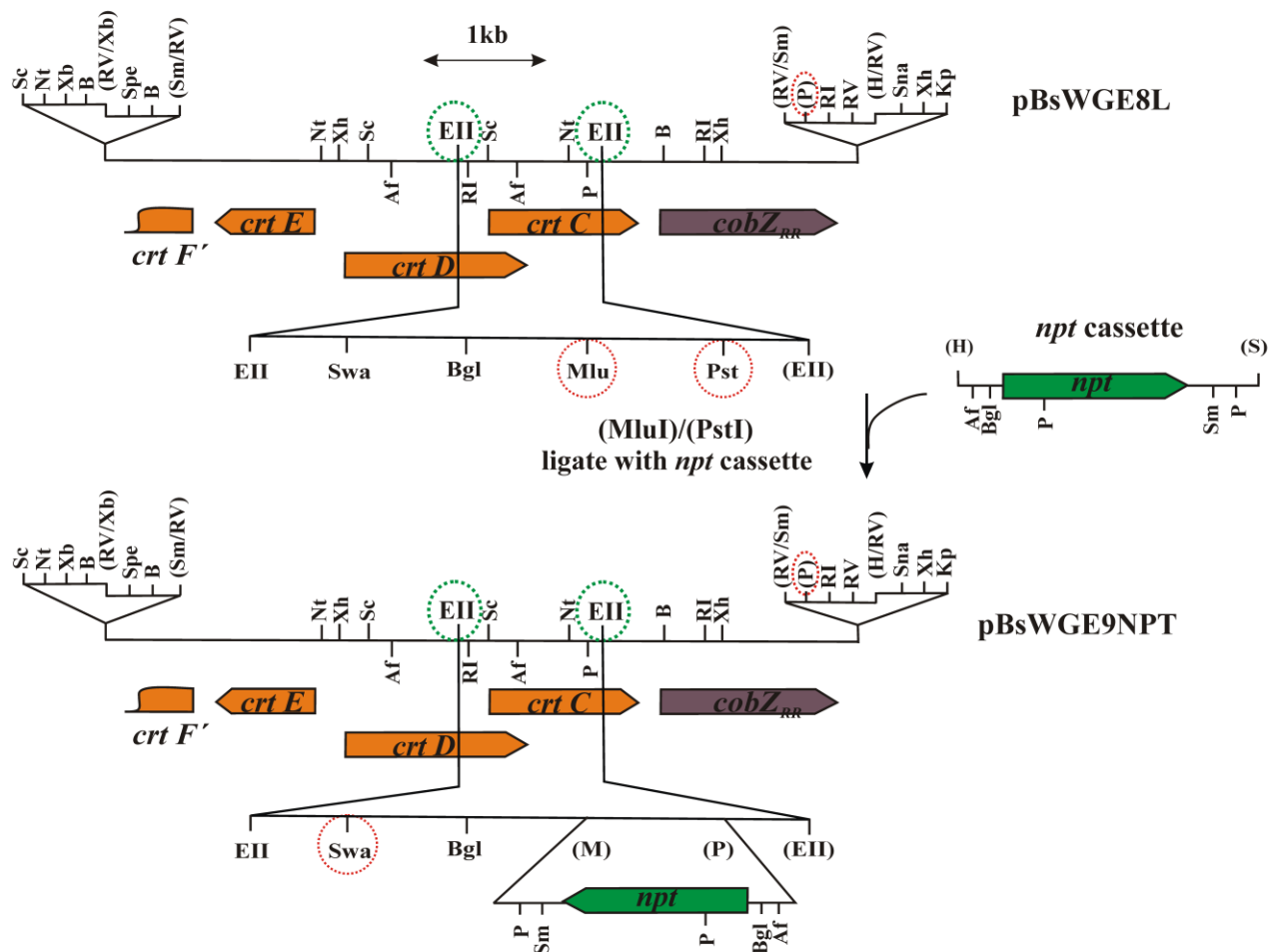
Fig 4.11 (step 6) Construction of the vector pBsWGE9NPT. pBsWGE8L was digested with MluI/PstI, blunted, dephosphorylated and ligated with *npt* cassette (also was blunted), yielding pBsWGE9NPT.

Step 7

pBsWG7 was EcoRI/SalI digested, dephosphorylated and ligated with an EcoRI/SalI digested *crtL* fragment to obtain the vector pBsWG9CRTL (**Fig 4.12**, step 7). The purpose of this step is to ensure that the synthetic rbs is placed in front of the ATG start codon of *crtL* for maximal expression.
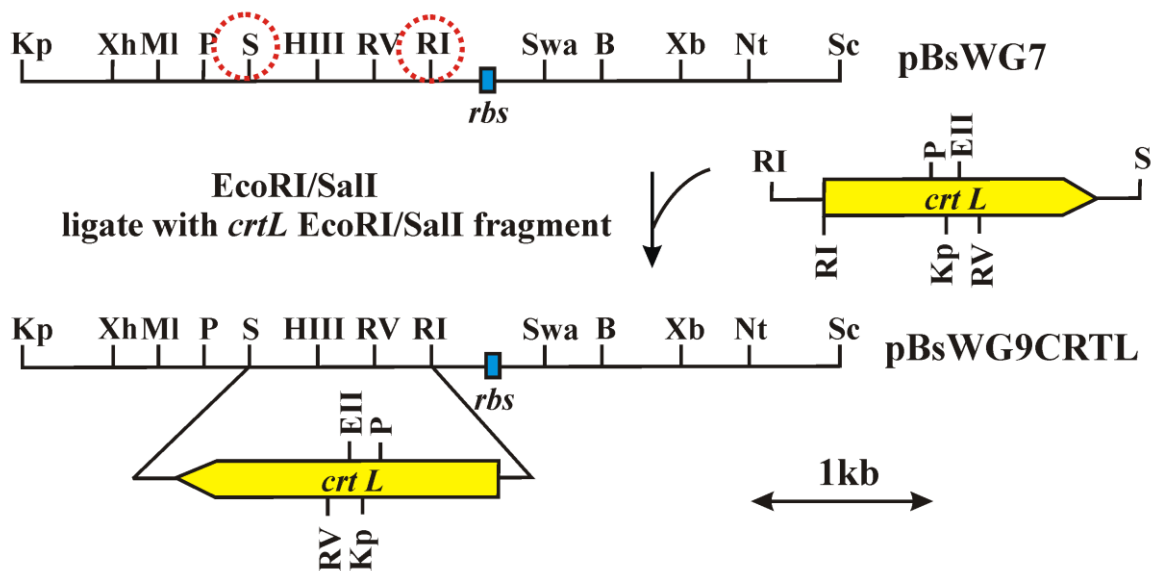
Fig 4.12 (step 7) Preparation of the *crtL* cassette for subsequent insertion into final vector. In vector pBsWG9CRTL, a synthetic rbs was in front of *crtL* cassette.

Step 8

The most important step of the construction is to ensure *crtL* will be inserted in the same orientation as *crtD*, and can be expressed from the *crtD* promoter. The *crtL* cassette was obtained by (Swa)/(SalI) digestion of pBsWG9CRTL, ligated with pBsWGE9NPT which had been digested by SwaI and dephosphorylated (**Fig 4.13 (A)**). After selection, pBsWGE9NPTCRTL was obtained, with *crtL* present in the same orientation as *crtD*. Before the insert was transferred to suicide vector, DNA sequencing of pBsWGE9NPTCRTL was performed. It was confirmed that the *crtL* is in the same orientation as the *crtD* fragment (**Fig 4.13 (B)**). The sequence data showed the rbs was -10 bp to -13 bp in front of a start codon of *crtL*, an in-frame stop codon which is necessary for arresting the truncated *crtD* ORF (*crtD'*).
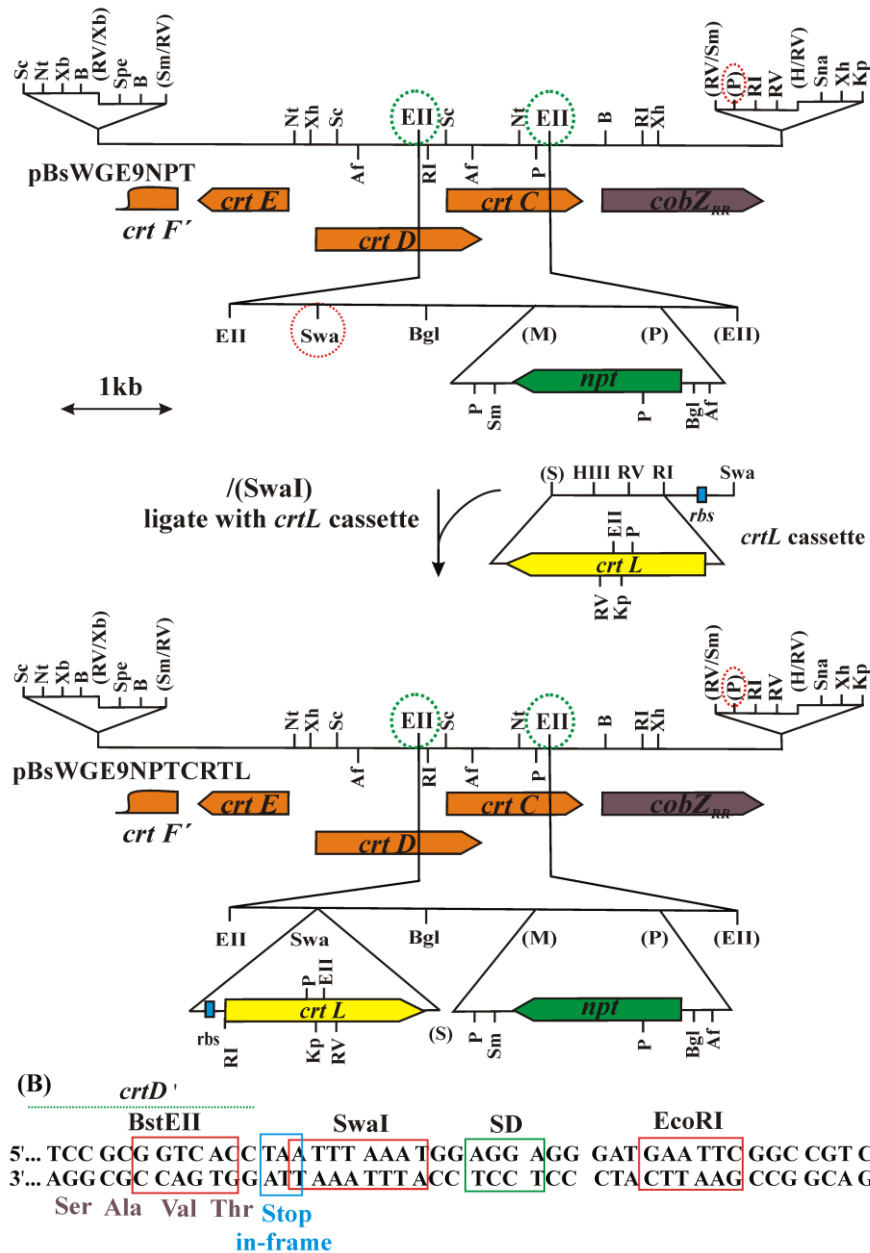
Fig 4.13 (step 8) (A) Construction of the vector pBsWGE9NPTCRTL. The *crtL cassette* was inserted into the SwaI site of pBsWGE9NPT, yielding pBsWGE9NPTCRTL. (B) Fusion region of *crtL* with the rest of *crtD* in pBsWGE9NPTCRTL.

Step 9

Finally, the insert from the pBsWGE9NPTCRTL was digested by XbaI and SnaBI, blunt-ended and then ligated with the suicide vector pSUP202, which had been digested by EcoRI, blunted and dephosphorylated. The final vector pSUPNPTCRTL is shown in **Fig 4.14**.
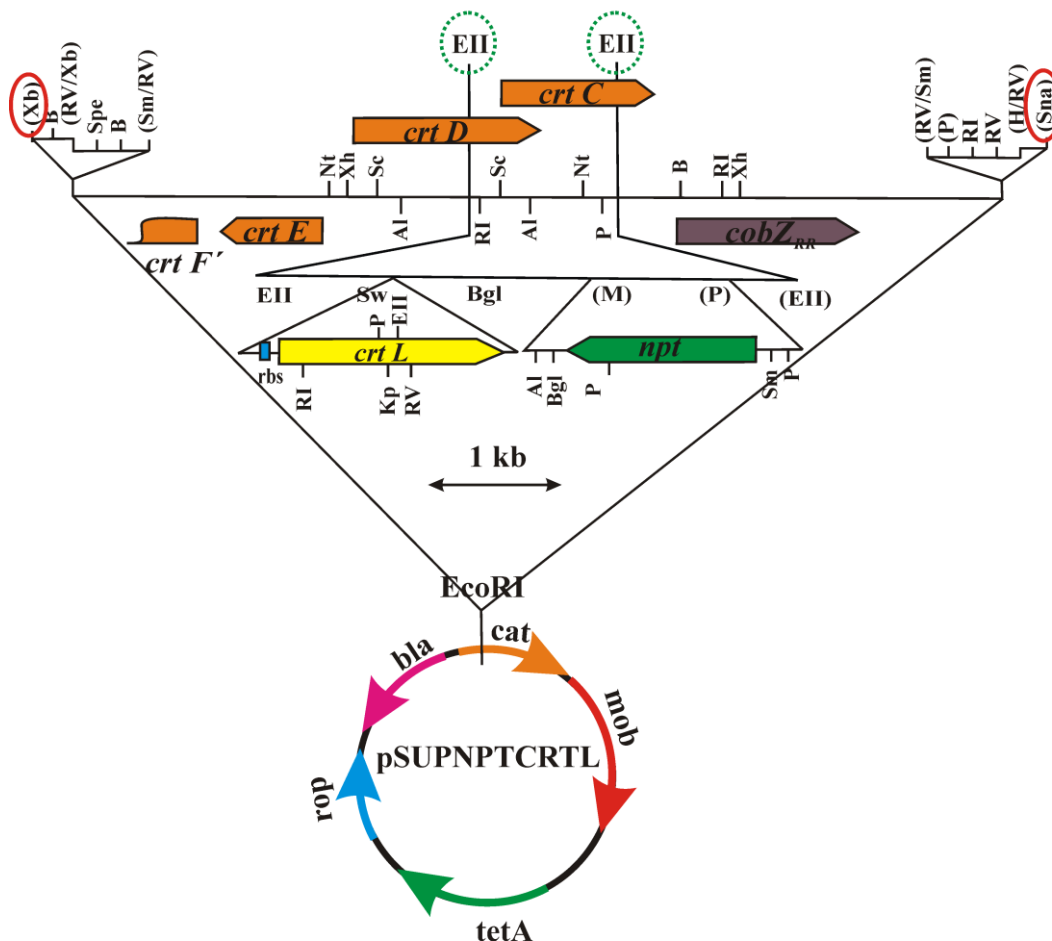
Fig 4.14 (step 9) Gene map of the suicide vector pSUPNPTCRTL. The 8.5 kb (SnaBI)/(XbaI) fragment, containing the construction of rbs-*crtL-npt* with homologous regions on both ends, was subcloned into pSUP202 EcoRI site, yielding pSUPNPTCRTL.

4.2.3. Introduction of pSUPNPTCRTL into *R. rubrum* by conjugation

pSUPNPTCRTL was subsequently triparentally conjugated with *R. rubrum* S1. Selection of double recombinants was determined by their Kan[r]Tet[s] phenotype. The chromosomal localization of the *npt-crtL* gene was confirmed using Southern hybridization.

4.2.4. Spectral analysis of whole cells.

Spectra of cells from stationary phase at the end of growth curve were measured after suspension in M-medium containing 80% glycerol as described before.

4.2.5. Spectral and TLC analysis of extracted carotenoid.

Carotenoid extraction from stationary phase at the end of growth curve was performed as described before, the hexane extract was used for spectral and TLC analysis.

4.2.6. Growth characterization.

The growth properties of SWGK46 mutant under either anaerobic photoheterotrophic conditions or dark, chemoheterotrophic conditions were examined in this study.

4.2.7. Carotenoid quantification.

Carotenoid quantification was performed from the stationary phase cells grown in M2SF medium at the scale of 100 ml and 500 ml.

4.2.8. *crtD* promoter activity determination

(1) The construction of promoter expression vector pPHUcrtD22.

A 2.1 kb fragment of pBsSGE5, which contains the *crtD* promoter region, was obtained by (XbaI)/(SacI) digestion (**Fig 4.15**) and subcloned into expression vector pPHU236 (Hübner *et al.*, 1991) which was (XbaI)/(ScaI) digested and dephosphorylated, yielding pPHUcrtD22. Then, the *lacZ* in the pPHUcrtD22 was designed to express from the *crtD* promoter.
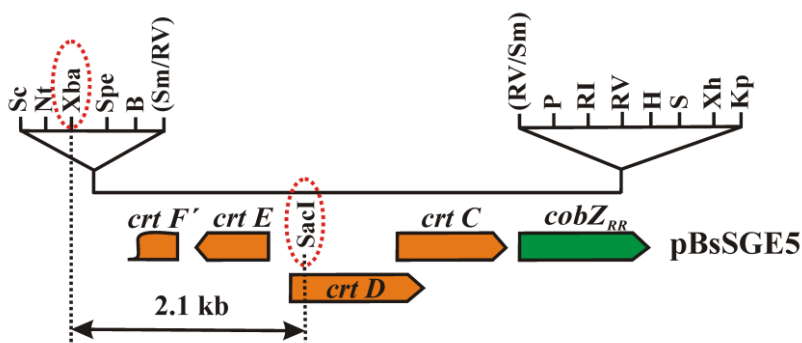
Fig 4.15 The *crtD* promoter containing fragment.

(2) Conjugation of pPHUcrtD22 and pPHU236 with wild-type S1

The conjugation was performed as described in the Chapter 2. The pPHUcrtD22/S1 and pPHU236/S1 were replated until *E. coli*-free and when anaerobic slope cultures were made from single colonies.

(3) β-galactosidase activity assay

Exponential phase cells of pPHUcrtD22/S1 and pPHU236/S1 were harvested by centrifugation of the cultures grown in M, M2S and M2SF medium, respectively. β-galactosidase activity was determined on permeabilized cells by using 0.5% N-lauroylsarcosine in the enzyme assay. β-galactosidase assay, using a method modified based upon that applied by Young *et al.* (1989), was performed in ice-cold 100 mM Na-phosphate buffer pH 7.0 containing appropriate amounts of cells, 10 mM KCl, 1 mM MgSO$_4$, 0.5% N-lauroylsarcosine and 2.19 mM synthetic substrate *o*-

nitrophenyl-β-D-galactopyranoside in a volume of 2.43 ml, after briefly vortexing, the mixture was incubated at 35°C for 30 min which was followed by the addition of 0.5 ml 1 M Na_2CO_3 to bring pH to 10.0 at which the molar extinction coefficient of nitrophenol is 4860 (Doub and Vandenbelt, 1949). The enzyme activity was calculated in units of μmol nitrophenol/min/standard OD.

4.2.9. Scaling up.

As $M2SF^+$ medium was developed for the high cell density and high levels expression of photosynthetic genes when cultures grown under dark, semi-aerobic conditions at large scale. The yield of β-carotene produced by SWGK46 was determined from 3 L culture in $M2SF^+$ medium. 500 ml of SWGK46 stationary phase cells grown in M2SF medium was used as an inoculum.

4.3 Results

4.3.1. Isolation of the β-carotene-producing strain SWGK46. Transconjugants were only obtained using pSUPNPTCRTL, where the *crtL* in the same orientation as *crtD* and *npt* in the opposite to *crtD*. Approximately 200 transconjugants were isolated using an initial kanamycin selection on Sistrom (M) agar under light, anaerobic conditions. 48 transconjugates were selected for $Kan^R Tet^S$ testing (**Fig 4.16**).

Fig 4.16 The first selection of the transconjugates. 48 transconjugates were plated on MK_{20} plate.

Four streaks showed $Kan^R Tet^S$ phenotypes: yellow, purple, pink, dark orange or brown. An initial spectral analysis showed the yellow colonies to contain β-carotene. One of them, designated SWGK46, was chosen for further analysis because it grew homogeneously on the plate both aerobically and anaerobically.

4.3.2. Detailed spectral analysis of the SWGK46 mutant. The absorption spectrum of SWGK46 whole cells (**Fig 4.17**) showed that the near-infrared (near-IR) Q_y absorption maximum of the LH1 complexes is present at 877 nm, which is different from the usual crt-containing LH1 complexes

(Q_y at 882 nm) documented so far (see Chapter 3) and also different to the LH1 complexes lacking carotenoids (Q_y at 874 nm). In whole cells, β-carotene absorption maxima were observed at 457, 483 and 518 nm. Superposition of the absorption spectra obtained from whole cells of SLYC18 and SWGK46 (same cell density) indicated increased levels of β-carotene and decreased levels of LH1 complexes in SWGK46. This observation has never been observed in any photosynthetic bacterium so far, we have designated this "imbalance" between carotenoid production and LH1 saturation with carotenoids as the "β-carotene decoupling effect".
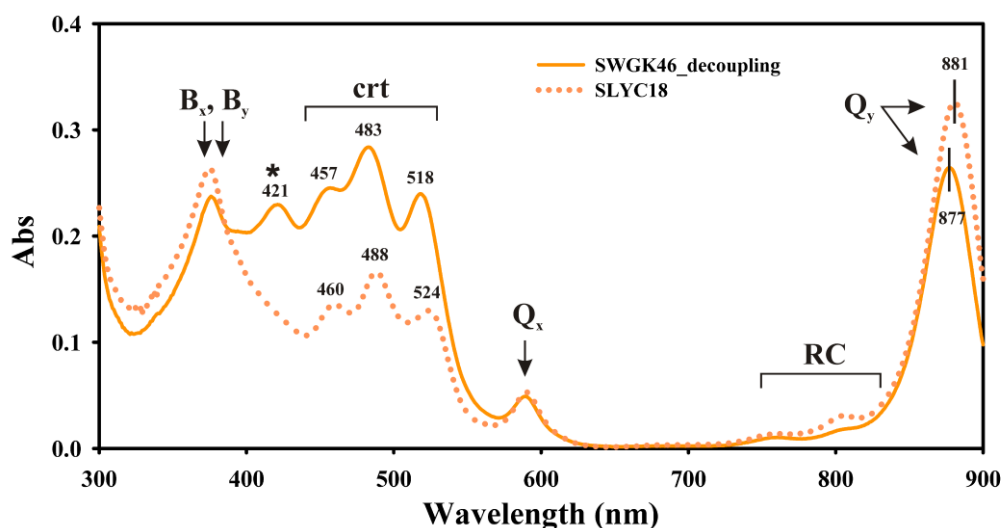
Fig 4.17 Spectral analysis of whole cells of SWGK46 and SLYC18. Indicated are the absorption peaks corresponding to β-carotene and lycopene, as well as the BChla near-IR maxima (Q_x and Q_y) and the Soret bands (B_x and B_y) of the LH1 and RC. The peak at 421 nm (marked with an asterisk) indicated a species probably corresponding to a BChla precursor (see main text for details).

4.3.3. Organic solvent extraction and TLC analysis. Organic solvent extraction (MeOH, hexane) from SWGK46 late exponential phase or stationary phase cells was performed and the absorption spectrum was determined from the hexane extract. The spectrum of extracted β-carotene from SWGK46 in hexane was superposed upon that of pure β-carotene in hexane for comparison. The two spectra fit to each other very well and show the same absorption maxima at 426, 450, 478 nm (**Fig 4.18 (A)**). This was the first primary proof that SWGK46 produces principally β-carotene. TLC analysis was performed and showed a large yellow β-carotene band running close to the front, corresponding to that of a pure β-carotene sample. In addition, a small orange band corresponded to lycopene (**Fig 4.18 (B)**). This was consistent with the slight difference at 478 nm peak between the spectra of hexane extract from SWGK46 and pure β-carotene in hexane. MS analysis showed that the molecular weight of SWGK46 hexane extract was 536 g. As β-carotene and lycopene have the

same molecular weight (536 g), in combination with the spectral and TLC analysis, we were confirmed that SWGK46 produces largely β-carotene as well as a small amount of lycopene.
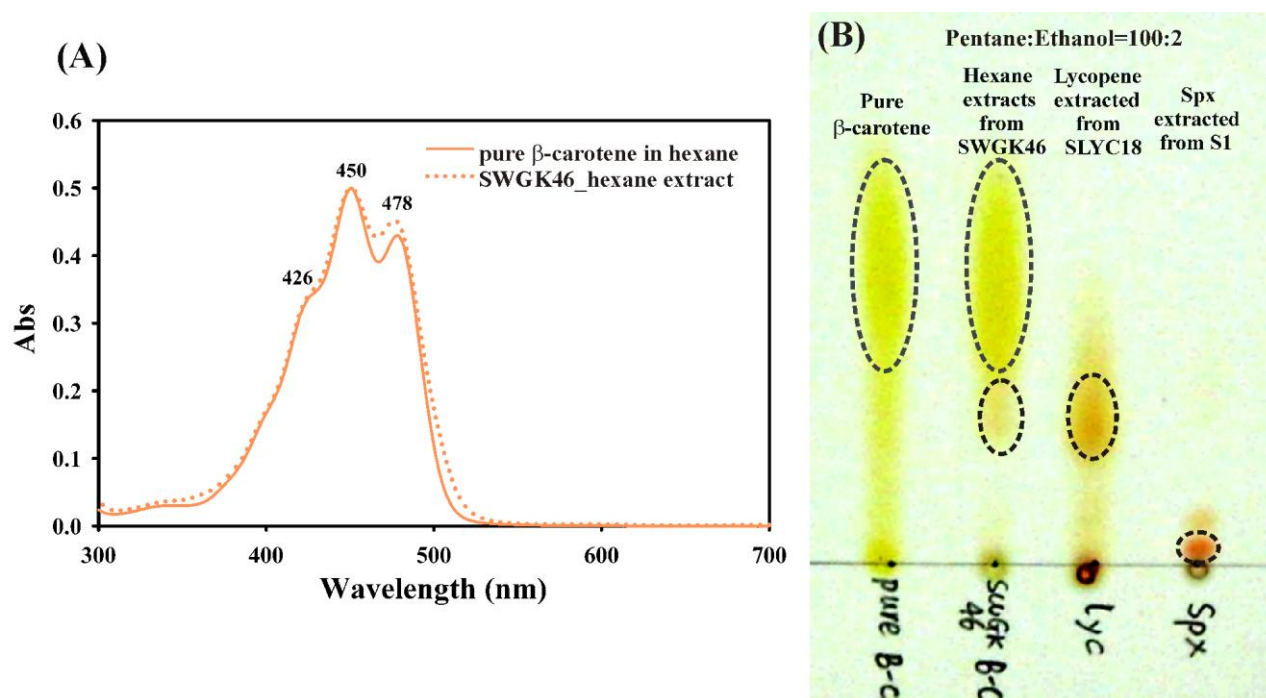
Fig 4.18 (A) Spectral analysis of extracted β-carotene from SWGK46 mutant. The spectrum of extracted β-carotene from SWGK46 was superposed upon of that of pure β-carotene for comparison. **(B) TLC analysis.** From left to right, the samples are commercial pure β-carotene, SWGK46 hexane extract, lycopene extracted from SLYC18 and Spx extracted from S1. The TLC plate was developed using pentane/ethanol (100:2 (v/v)) as the running solvent.

4.3.4. Southern hybridization of SWGK46 chromosomal DNA to an AflIII fragment of the pBsSGE5 insert: the proof of the correct insertion of the *crtL* and *npt* cassettes in the chromosome.

SWGK46 chromosomal DNA was digested by EcoRV, XhoI and AflIII respectively, yielding five overlapping fragments which were in the same sizes as predicted from the gene map (**Fig 4.19 (A), (B)**). Southern hybridization of the EcoRV-digested chromosomal DNA with the 4.3 kb AflIII fragment from pBsSGE5, which containing both the *crtL* and *npt* genes, gives a strong signal due to the chromosomal 3.9 kb fragment and a somewhat weaker signal at 4.5 kb, as expected. The AflIII-digested chromosomal DNA showed predominately the expected 4.3 kb band and very weak smaller bands, which were possibly due to "star activity". The XhoI-digested chromosomal DNA was only partially digested, but nevertheless the expected 7.7 kb and 5.0 kb bands were still observable. Southern hybridization was also performed using pSUP202 digested by EcoRI as a probe (**Fig 4.19 (C)**) which showed no hybridization signal and confirmed that no carrier plasmid

had been inserted into *R. rubrum* chromosome. We conclude that the *crtL-npt* construct was correctly inserted into the *R. rubrum* chromosome via a double recombinational event.
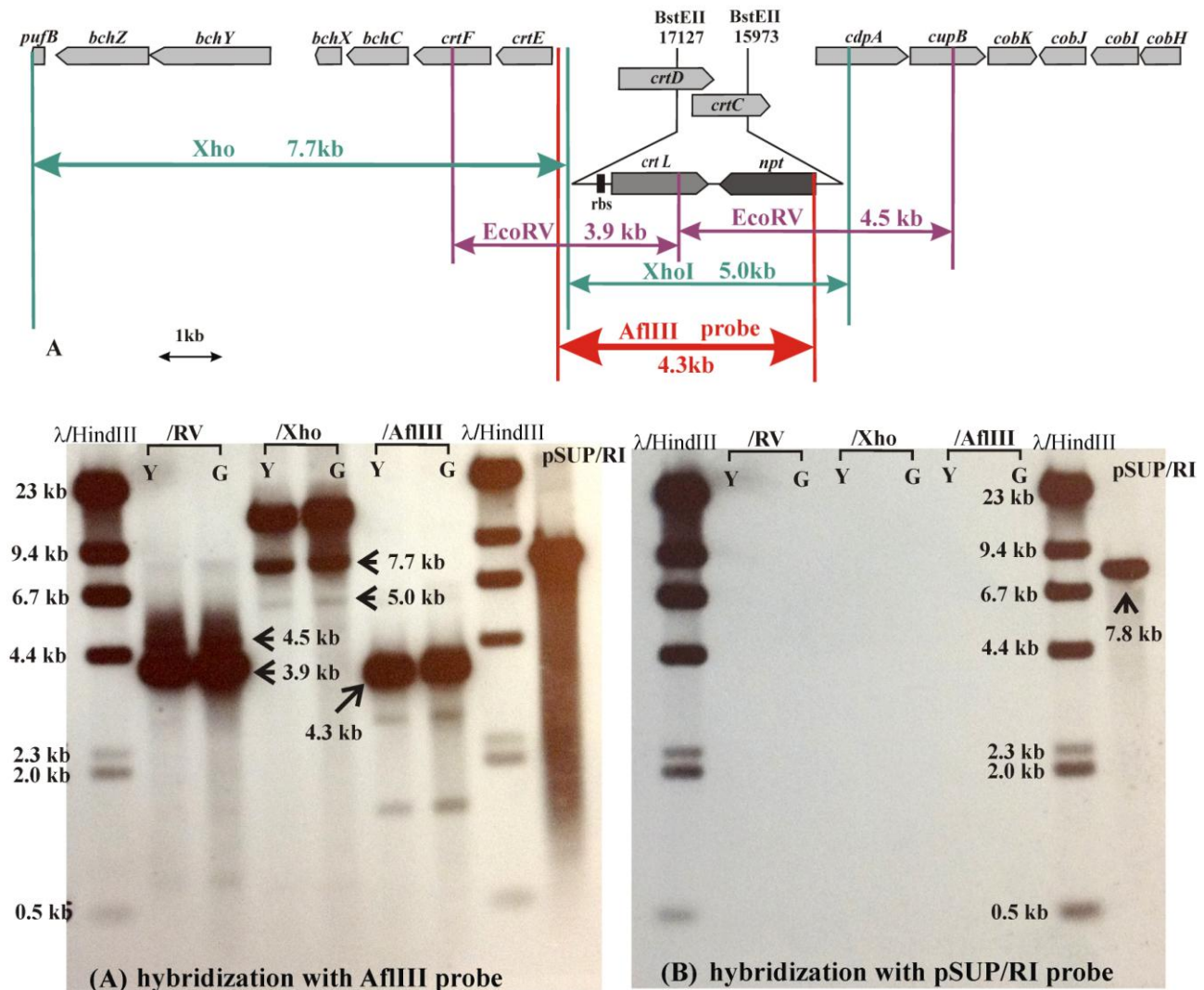
Fig 4.19 Southern hybridizaiton of SWGK46 chromosomal DNA to a 4.3 kb AflIII probe and a 7.8 kb pSUP202 (digested by EcoRI) probe. Chromosomal DNA was digested by EcoRV, XhoI, and AflIII, respectively. The sizes of the expected chromosomal fragments are indicated in (A). SWGK46 DNA hybridized to the 4.3 kb AflIII probe (B). SWGK46 DNA hybridized to 7.8 kb pSUP202 (digested by EcoRI) probe (C).

4.3.5. Growth physiology of the SWGK46 mutant

The growth properties of SWGK46 were studied under dark, semi-aerobic conditions in M2SF medium at the 100 ml scale (**Fig 4.20**). Compared to the parental strain SLYC18, SWGK46 showed a longer lag phase in the initial growth phase (**Fig 4.20 (A)**). However, following the commencement of growth, SWGK46 showed the same growth rate as that of SLYC18 and the final cell density obtained at the stationary phase of the growth curve was almost the same ($A_{660} \approx 1.8$). The appearance of ICM in SWGK46 under semi-aerobic conditions showed three unique

characteristics compared to that of SLYC18 (**Fig 4.20 (B)**). First, in the initial phase, when O_2 level is high, the ICM level showed a slight *increase* before it dropped to the trough position where the O_2 level went down below 0.3% and the ICM started to increase. Secondly, the trough position was around 15 h later than SLYC18 which was consistent with the lag phase observed from the A_{660} profile. Thirdly, in the stationary phase, the final ICM level/cell of SWGK46 was considerably lower than those of SLYC18, as well as S1.
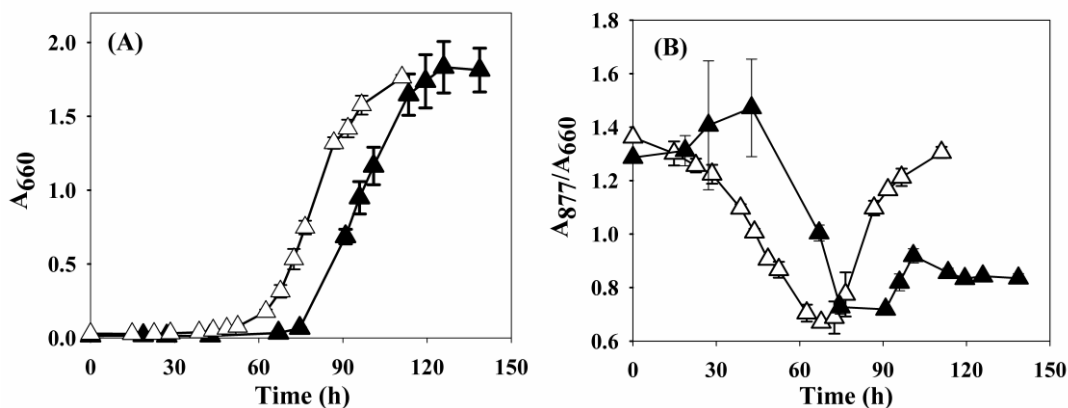
Fig 4.20 Dark, semi-aerobic growth of SWGK46 in 100 ml M2SF medium. The error bars were derived from three independent growth curves. (A) Cell density (indicated by A_{660}). ICM development (indicated by A_{877}/A_{660}) is shown in (B).

The SWGK46 growth properties were also examined under photoheterotrophic conditions (**Fig 4.21**). An interesting observation was that the cell density achieved here was not as high as that of the semi-aerobic culture (**Fig 4.21 (A)**). The photoheterotrophic growth curve lasted 400 h, which was more than twice as long as that of the dark, semi-aerobic culture. The ICM profile was as expected (**Fig 4.21 (B)**). As it grew anaerobically, in the initial phase, the ICM did not increase so much until the dissolved O_2 in the medium was consumed and O_2 level dropped below 0.3%.
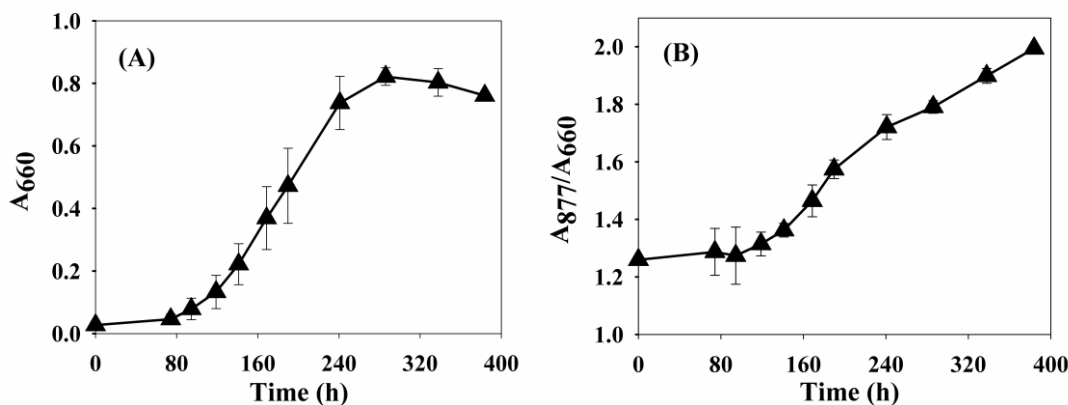
Fig 4.21 Anaerobic growth curve of SWGK46 in 100 ml M medium. Three independent growth curves were performed to show the reproducibility. (A) Cell density (indicated by A_{660}). ICM development (indicated by A_{877}/A_{660}) is shown in (B).

4.3.6. Quantitation of β-carotene level

The levels of β-carotene achieved in SWGK46 stationary phase cells grown in M2SF medium at the 100 ml and 500 ml scales, respectively, were determined by MeOH/hexane extraction followed by spectral analysis (**Fig 4.22**). When grown under semi-aerobic conditions in 100 ml M2SF medium, SWGK46 yielded values of **1.6 mg β-carotene/g DW**, **4.2 mg β-carotene/litre** and **5.1 mg β-carotene/g protein** respectively (**Fig 4.22 (A)**). The yield from the 500 ml scale was even higher, **4.4 mg β-carotene/g DW**, **5.1 mg β-carotene/litre** and **7.9 mg β-carotene/g protein** respectively were obtained (**Fig 4.22 (B)**).
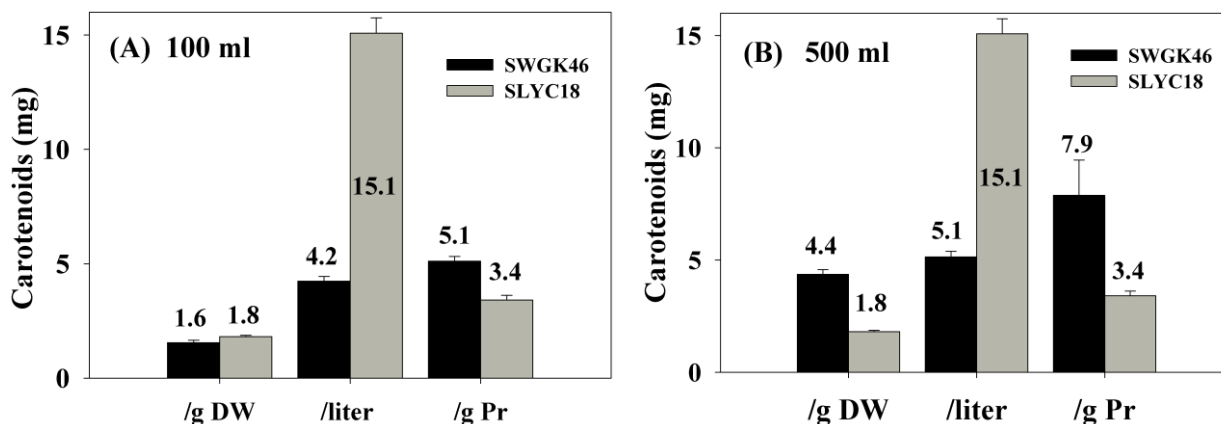
Fig 4.22 Quantitative analysis of β-carotene from stationary phase cells grown in M2SF medium at scale of 100 ml (A) and 500 ml (B) in comparison with lycopene from stationary phase cells grown in M2SF⁺ medium at a 3 litre scale. All determinations were performed in triplicate for each of the individual cultures. Thus, each of the bar chart values and their error bars have been obtained from a total of nine determinations.

4.3.7. Large scale growth experiment (3 L) in M2SF⁺ medium

The SWGK46 3 L M2SF⁺ culture used an M2SF-grown inoculum which had shown a characteristic β-carotene cell absorption spectrum prior to inoculation (**Fig 4.23 (A)**). The cell spectrum of the stationary phase cells of SWGK46 3 L culture was also performed (**Fig 4.23 (B)**) which has two major characteristics: first, the carotenoid level was suppressed comparing to that of the 500 ml M2SF culture. Secondly, the carotenoid absorption maxima seemed no longer corresponding to those of β-carotene. Carotenoid extraction and spectral analysis showed that the 3 L M2SF⁺ culture produced lycopene (**Fig 4.24**).
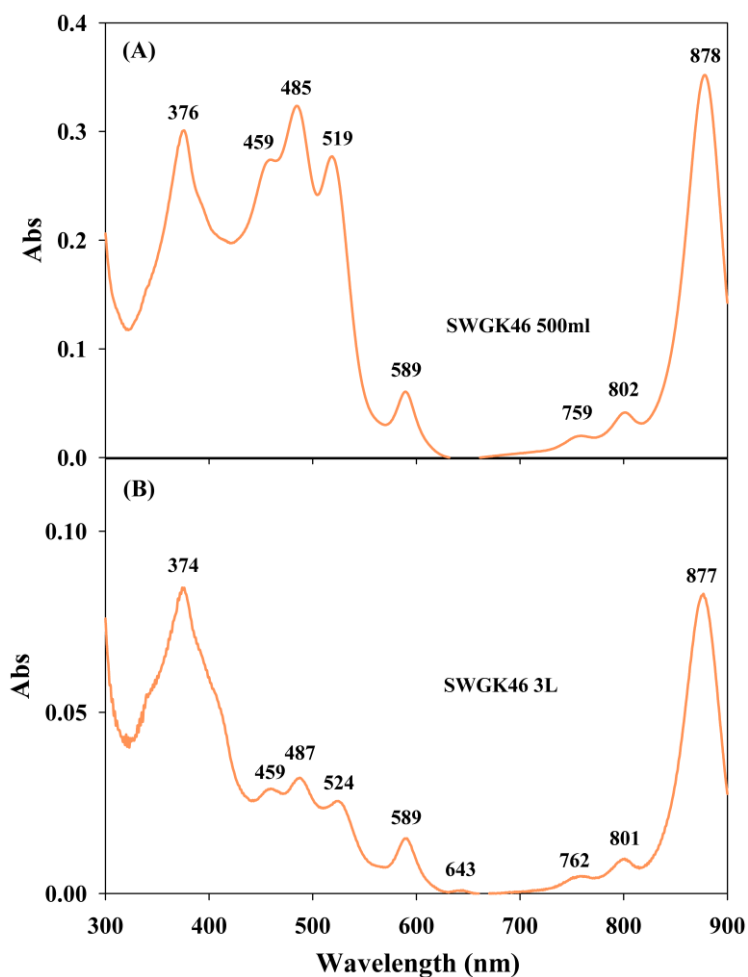
Fig 4.23 Whole-cell spectral analysis of SWGK46. Cell spectra of SWGK46 whole cells at the stationary phase from 500 ml M2SF culture (A) and 3 L M2SF$^+$ culture (B).
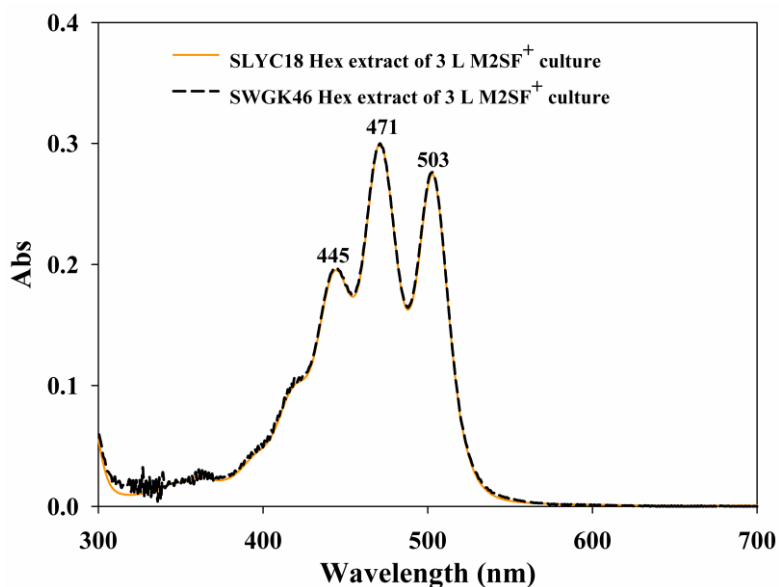
Fig 4.24 Spectral analysis of hexane extracts from whole cells of SWGK46 (black short dashed line) and SLYC18 (orange solid line), respectively. Cells were grown in 3 L M2SF$^+$ and harvested in the stationary phase.

4.3.8. *crtD* promoter activity

pPHUcrtD22 was selected on an X-gal plate and showed a light blue color which became much intense after cooling to 4°C, confirming that the *lacZ* was active on vector pPHUcrtD22 (**Fig 4.25**

(A)). *lacZ* utilizes the *tet*[R] promoter on vector pPHU236 and develops intense color compared to the weak blue color developed by *lacZ* using the *crtD* promoter on vector pPHUcrtD22.
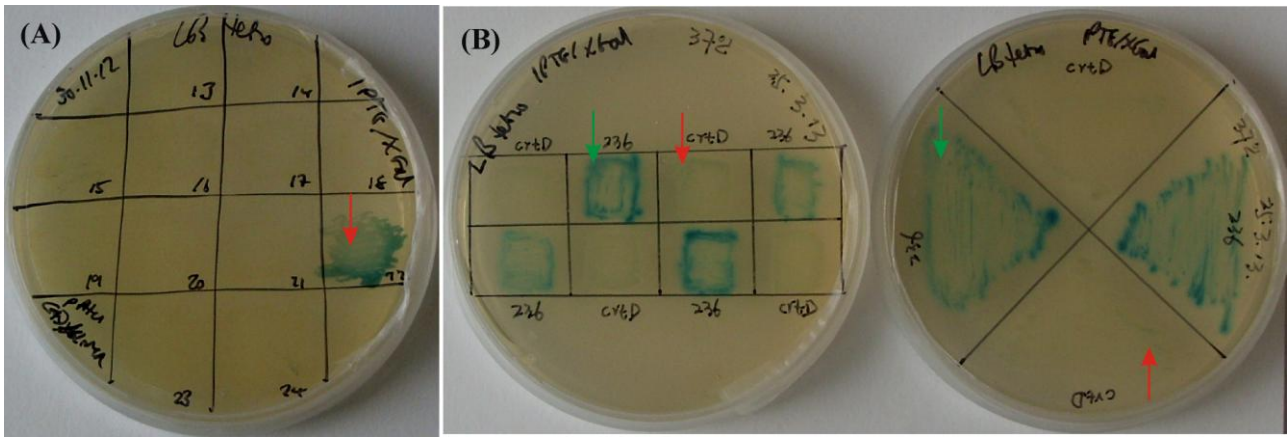
Fig 4.25 pPHUcrtD22 selection and comparison to pPHU236. pPHUcrtD22 was selected on LBtet$_{10}$ plate containing X-gal, indicated with the red arrow (A). pPHU236 showed a much more intense blue color compared to pPHUcrtD22 (B), which are indicated with red and green arrows, respectively.

```
                                                                                                              crtD
     crtD´´        ATG GAC ATT TGG TAG GCT CGA GTC ATG GCG ATC GAC AAA GGA TAA AGG CTT GTG AAG ACC CAG CAC GTG GTC GTG GTA GGC GCC GGC│ATG
 pPHUCRTD_seq      ATG GAC ATT TGG TAG GCT CGA GTC ATG GCG ATC GAC AAA GGA TAA AGG CTT GCN AAG ACC CAG CAC GTG GTC GTG GTA GGC GCC GGC ATG
                   Met Asp Ile Trp *** Ala Arg Val Met Ala Ile Asp Lys Gly *** Arg Leu Ala Lys Thr Gln His Val Val Val Val Gly Ala Gly Met

     crtD´´        GGC GGA CTG GCG GCG GCC ATC GAT CTC GCA TCG CGT GGC CTG CGC GTC ACC GTC CTT GAG CGA TCC CCG TCG CCC GGC GGC AAG ATG CGC
 pPHUCRTD_seq      GGC GGA CTG GCG GCG GCC ATC GAT CTC GCA TCG CGT GGC CTG CGC GTC ACC GTC CTT GAG CGA TCC CCG TCG CCC GGC GGC AAG ATG CGC
                   Gly Gly Leu Ala Ala Ala Ile Asp Leu Ala Ser Arg Gly Leu Arg Val Thr Val Leu Glu Arg Ser Pro Ser Pro Gly Gly Lys Met Arg

     crtD´´        GAG ATC GCC GTC GGC GGC GCC CGG CTC GAC GCC GGG CCG ACG GTT TTC ACC ATG CGC TGG GTG TTC GAG G                    PstI       SphI
 pPHUCRTD_seq      GAG ATC GCC GTC GGC GGC GCC CGG CTC GAC GCC GGG CCG ACG GTT TTC ACC ATG CGC TGG GTG TTC GAG GAC TCG ACC TGC AGG CAT GCA
                   Glu Ile Ala Val Gly Gly Ala Arg Leu Asp Ala Gly Pro Thr Val Phe Thr Met Arg Trp Val Phe Glu Asp Ser Thr Cys Arg His Ala
       lacZ                                                                                                 ATG ACC ATG ATT ACG
                                                                                                                lacZ

                   HindIII
 pPHUCRTD_seq      AGC TTC GAT CCC GTC GTT TTA CAA CGT CGT GAC TGG GAA AAC CCT GGC GTT ACC CAA CTT AAT CGC CTT GCA GCA CAT CCC CNC TTT CGC
                   Ser Phe Asp Pro Val Val Leu Gln Arg Arg Asp Trp Glu Asn Pro Gly Val Thr Gln Leu Asn Arg Leu Ala Ala His Pro Xxx Phe Arg
       lacZ        GAT TCA CTG GCC GTC GTT TTA CAA CGT CGT GAC TGG GAA AAC CCT GGC GTT ACC CAA CTT AAT CGC CTT GCA GCA CAT CCC CC- TTT CGC

 pPHUCRTD_seq      CAG CTG GCN GTA ATA GCN GAA GAG GCC CGC ACC NGA TCG CCC TTC CCA ACA GTT GCG CAG CCT GAA TGG CGA ATG GCG CTT TGC CTG GTT
                   Gln Leu Ala Val Ile Ala Glu Glu Ala Arg Thr Xxx Ser Pro Phe Pro Thr Val Ala Gln Pro Glu Trp Arg Met Ala Leu Cys Leu Val
       lacZ        CAG CTG GCG -TA ATA GC- GAA GAG GCC CGC ACC -GA TCG CCC TTC CCA ACA GTT GCG CAG CCT GAA TGG CGA ATG GCG CTT TGC CTG GTT

 pPHUCRTD_seq      TCC GGC ACC AGA AGC GGT GCC GGA AAG CTG GCT GGA GTG CGA TCT TCC TGA GGC CGA TAC TGT CGT CGT CCC CTC AGA CTG GCA GAT GCA
                   Ser Gly Thr Arg Ser Gly Ala Gly Lys Leu Ala Gly Val Arg Ser Ser *** Gly Arg Tyr Cys Arg Arg Pro Leu Arg Leu Ala Asp Ala
       lacZ        TCC GGC ACC AGA AGC GGT GCC GGA AAG CTG GCT GGA GTG CGA TCT TCC TGA GGC CGA TAC TGT CGT CGT CCC CTC AAA CTG GCA GAT GCA

 pPHUCRTD_seq      CGG TTA CGA TGC GCC CAT CTA CAC CAA CGT GAC CTA TCC CAT TAC GGT CAA TCC GCC GTT TGT TCC CAC GGA GAA TCC GAC GNG NTG TTA
                   Arg Leu Arg Cys Ala His Leu His Gln Arg Asp Leu Ser His Tyr Gly Gln Ser Ala Val Cys Ser His Gly Glu Ser Asp Xxx Xxx Leu
       lacZ        CGG TTA CGA TGC GCC CAT CTA CAC CAA CGT GAC CTA TCC CAT TAC GGT CAA TCC GCC GTT TGT TCC CAC GGA GAA TCC GAC GGG TTG TTA

 pPHUCRTD_seq      CTC GCT CAC ATT TAA TGT TGA TGA AAG NTG GCT ACA GGA AGG CCA G
                   Leu Ala His Ile *** Cys *** *** Lys Xxx Ala Thr Gly Arg Pro
       lacZ        CTC GCT CAC ATT TAA TGT TGA TGA AAG CTG GCT ACA GGA AGG CCA GAC GCG AAT TAT TTT TGA TGG CGT TAA CTC GGC GTT TCA TCT GTC....
```

Fig 4.26 pPHUcrtD22 sequence data. The alignments between pPHUcrtD22 sequence data with *crtD* (indicated in green) and with *lacZ* (indicated in magenta) confirmed that the fusion region of *crtD* with pPHU236 was in-frame and that the *crtD* promoter was in the same direction as *lacZ*.

The DNA of pPHUcrtD22 was sequenced and analysed by sequence alignment with *R. rubrum* *crtD* and *lacZ* (AJ308295) (**Fig 4.26**), which confirmed that the 2.1 kb (XbaI)/(SacI) fragment containing *crtD* promoter and *crtD′* truncated ORF was fused in frame with *lacZ*. The activity of the *crtD* promoter in wild-type *R. rubrum* S1 under different growth conditions was measured using a β-galactosidase assay. pPHUcrtD22/S1 cells harvested from semi-aerobic culture grown in M2S and M2SF media, as well as from anaerobic culture grown in M medium were used. Control experiments using pPHU236/S1 were also performed. The results are summarized in **Table 4.5**.

Strain	*crtD* promoter activity µmol/min/OD=1	Differential activity (%)	Deviation (%)	Relative activity	Deviation
pPHUcrtD22/S1_M2S	0.097	9.28	1.28	13.86	0.93
pPHUcrtD22/S1_M2SF	0.105	1.90	0.7	15.00	0.35
pPHUcrtD22/S1_M	0.009	22.22	0.3	1.29	0.37
pPHU236/S1_M2S	0.088			12.57	0.17
pPHU236/S1_M2SF	0.103			14.71	0.66
pPHU236/S1_M	0.007			1.00	0.20

Table 4.5 *crtD* promoter activities obtained using pPHUcrtD22/S1 cells under semi-aerobic (M2S and M2SF media) and anaerobic (M medium) growth conditions.

The β-galactosidase activities measured with pPHUcrtD22/S1 cells grown in M2S, M2SF and M media are only slightly higher than those measured with pPHU236/S1 cells grown in the same conditions. For better comparison, the *crtD* promoter activity was presented in another two ways: (1) Differential activity of *crtD* promoter under each growth condition, calculated as follows: (pPHUcrtD22/S1- pPHU236/S1)/ pPHUcrtD22/S1 (**Fig 4.27 (A)**). Under anaerobic growth conditions, the *crtD* promoter differential activity appears to be the highest (22.22%) compared to semi-aerobic conditions. The *crtD* promoter differential activity in M2S medium (9.8%) is slightly higher than in M2SF medium (1.90%). (2) Relative activity of the *crtD* promoter. The *crtD* promoter activity of pPHU236/S1 in M-medium was defined as 1, then the other *crtD* promoter activities measured with pPHUcrtD22/S1 and pPHU236/S1 cells grown with each conditions were divided by the *crtD* promoter activity of pPHU236/S1 in M-medium (**Fig 4.27 (B)**). The relative activity shows that the *crtD* promoter activities in M2S medium are larger than those in M2SF medium which in turn are larger than those in M medium.
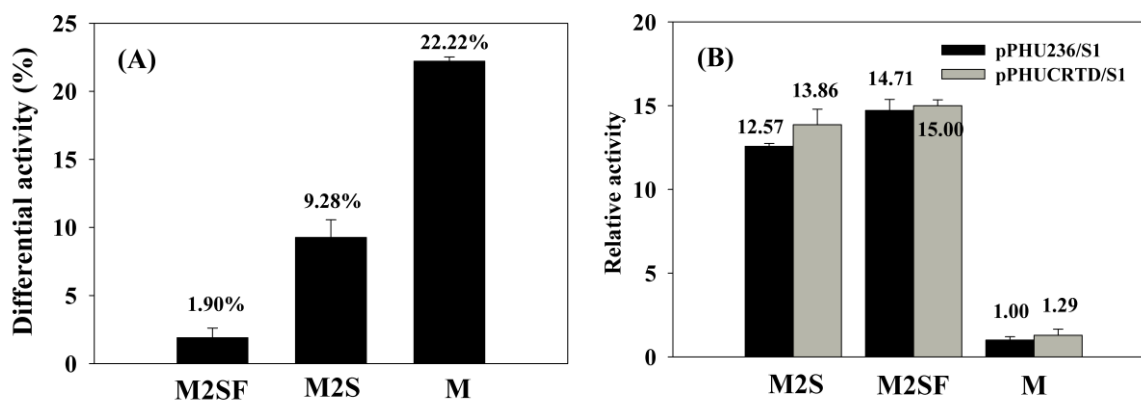
Fig 4.27 The _crtD_ promoter activity was presented in differential activity and relative activity two ways.

4.4. Discussion

Using the same strategy as for the production of lycopene in _R. rubrum_, SWGK46 was constructed by replacing part of the chromosomal _crtCD_ region with the _A. thaliana_ CrtL gene, together with a synthetic rbs and an _npt_ cassette. The strategy worked very well and high levels of β-carotene of 4.4 mg/g DW were achieved with SWGK46 grown in M2SF medium at 500 ml scale. However, SWGK46 started to produce lycopene instead of β-carotene when grown at 3 L scale in M2SF$^+$ medium. The reasons for this will be examined in Chapter 5.

New observations with SWGK46 which have never been documented so far are listed here:

(1) A unique form of the LH1 complexes with Q_y at 877 nm, was generated probably due to the conformational changes due the incorporation of β-carotene into the LH1 complexes.

(2) SWGK46 suffering from oxidative stress was diagnosed by the peak at 420 nm which was consistent with another O_2-sensitive _R. rubrum_ mutant (C. Autenrieth and R. Ghosh, unpublished).

(3) The "β-carotene decoupling effect".

Another important diagnostic tool employed is the ratio between the Q_y absorption maximum and the carotenoid absorption maximum. The localization of carotenoids in _R. rubrum_ is the ICM (16 carotenoids/LH1 complex, 1 carotenoid/RC). Since we know that _R. rubrum_ contains only one LH complex which has a characteristic Q_y absorption at 882 nm in the presence of carotenoids (except SWGK46Y (877 nm)). The level of LH1 complexes is proportional to that of ICM and this proportionality hardly changes regardless of different growth conditions. Therefore, A_{882}, being used to indicate the LH1 complexes level, can also indicate the ICM level, and the ratio A_{882}/A_{488} (SLYC18) or A_{877}/A_{483} (SWGK46) can be used to diagnose the coordination between carotenoids

capacity and carotenoids level in the cell. Under semi-aerobic, dark conditions, the SLYC18 cells are full of ICM and the carotenoid capacity in the cell is utilized to its maximum (A_{882}/A_{488} around 1.96). The carotenoids level appears to be coordinated with (or coupled to) that of BChla. This is indicated by the inability of the BChla⁻ phenotype (ST3) to produce carotenoids even though the lesion is far from and does not affect the carotenoid biosynthesis genes (R. Ghosh, unpublished). The coupling between BChla and carotenoid level is also indicated by the fixed ratio between BChla and carotenoid absorption maxima, which is respectively about 2.31 for S1 and 1.96 for SLYC18. We hypothesize that the carotenoid expression is activated by BChla, and that the carotenoid production level is coupled to that of BChla, both of which are localized in the ICM.

SWGK46Y showed a A_{877}/A_{483} of 0.93, which might indicate that SWGK46Y increased the carotenoid capacity via releasing β-carotene into the membrane or somewhere else. This observation also suggests that β-carotene production can be decoupled from the BChla level. This conclusion may be crucial for increasing the carotenoid production level significantly. Therefore, this study presents a convincing starting point for β-carotene production in purple bacteria. Higher carotenoid levels might be achieved when the β-carotene decoupling mechanism is well understood so that it may be abolished using molecular biological methods.

The mechanisms involved in β-carotene decoupling effects might be due to the mutations in the LH1 complex, in combination with the *A. thaliana* lycopene β-cyclase functioning as a blue light sensor which could interact with the carotenoid biosynthesis regulating enzymes. However, the response of lycopene β-cyclase to light has never even been reported or examined previously (to our knowledge). More details will be presented in the Chapter 7.

The result of the *crtD* promoter activity was suprising, which showed the *crtD* promoter activity on the vector pPHUcrtD22 in *R. rubrum in trans* is only slightly higher than that of the control vector pPHU236. There are several possible reasons for this result: (a) The *crtD* promoter is weak. This indicates that saturating the LH1 complexes with carotenoid does not need a strong promoter. (b) The vector pPHUcrtD22 contains the *crtD* promoter, in tandem to the tet^R promoter. For a weak target promoter this may be a problem, as it is known that the tet^R promoter shows a weak readthrough activity for the downstream regions. For weak promoters, this may also be "an interference effect" which attenuates the target promoter activity. This may explain why the total promoter activity of the two promoters in tandem did not give twice as much as the control.

We are also interested to know whether synthetic *crtWZ* from *A. aurantiacum* could be expressed after being introduced into SWGK46 *in trans* and which level of astaxanthin that can be achieved. In the course of this work, the *A. aurantiacum crtWZ* cistron was redesigned according to *R. rubrum* codon usage preference and the rules mentioned in Appendix 4. Gene synthesis was performed commercially (GenScript HK Inc). Expression studies using the synthetic genes are in progress in the Ghosh lab at this time.

Table . Strains and plasmids used in SWGK46 study

Strain or plasmid	Relevant characteristics	Reference
E. coli strains		
XL1MR	Δ(*mcrA*)183 Δ(*mcrCB-hsdSMR-mrn173 endA1 supE44 thi1 recA1 gyrA96 relA1 lac*)	Stratagene
R. rubrum strains		
S1	Wild-type	Cohen-Bazire *et al*., 1956
SWGK46	SLYC18-derived *crtC⁻D⁻* site-directed mutant, Kan^r	This study
Plasmids		
pBsKSII+	High-copy cloning vector, ColE1, Amp^r	Stratagene
pBsWG1	pBsKSII+ derivative, containing a BamHI_EcoRI_SnaBI_XhoI linker	This study
pBsWG6	pBsKSII+ derivative, containing a BamHI_SwaI_SD_EcoRI linker	This study
pBsWG7	pBsWG6 derivative, containing a XhoI_MluI_PstI_SalI linker	This study
pBsWG8NPT	pBsWG7 derivative, containing an *npt* cassette derived from the transposon Tn*5*	This study
pU10655	pUNI 51 derivative, containing a EcoRI/Sal *crtL* cassette from *A. thaliana*	Heyer, unpublished
pBsWG9CRTL	pBsWG7 derivative, containing *crtL* gene from *A. thaliana*	This study
pBsSGE5	pBsKSII+ derivative, containing the 6.37 kb EcoRV fragment from the pSC4 cosmid	Wang *et al.*, 2012
pBsWGE5	pBsWG1 derivative, containing the 6.42 kb HindIII/XbaI insert from pBsSGE5	This study
pBsWGE6ΔPstI	pBsWGE5 derivative, after the PstI site in the linker being removed	This study

pBsWGE7ΔPB	pBsWGE6ΔPstI derivative, after removal of the a BglII site created after PstI site-directed mutagenesis	This study
pBsWGE8L	pBsWGE7ΔPB derivative, containing a BstEII_SwaI_PstI_BstEII linker	This study
pBsWGE9NPT	pBsWGE8L derivative, containing the Tn*5* derived *npt* cassette	This study
pBsWGE9NPTCRTL	pBsWGE9NPT derivative, containing the A. thaliana *crtL* gene	This study
pBsLGKan	pBsKSII+ derivative, containing the Tn*5* derived *npt* cassette	Lupo and Ghosh, 2004
pRK2013	Mobilizing helper plasmid, *tra*[+], Kan[r]	Figurski and Helinski, 1979
pSUP202	Suicide vector, ColE1, *mob*[+], Amp[r] Cm[r] Tet[r]	Simon *et al.*, 1983
pSUPNPTCRTL	pSUP202 derivative, containing the 8.49 kb SnaBI/XbaI fragment from pBsWGE9NPTCRTL	This study
pPHU236	Broad-host-range *lacZ* fusion vector	Hübner *et al*., 1991
pPHUcrtD22	pPHU236 derivative, containing the *crtD* promoter in a 2.1 kb (XbaI)/(SacI) fragment from pBsSEG5	This study

Chapter 5　　The observations of the instability of the mutant SWGK46

5.1 Introduction

As mentioned in Chapter 4, SWGK46 can produce β-carotene at high levels when grown under semi-aerobic, dark conditions in M2SF medium at the 100 ml and the 500 ml scale, yielding **1.6 mg β-carotene/g DW** and **4.4 mg β-carotene/g DW** respectively. This phenotype was designated as **SWGK46Y**, which showed a characteristic unique form of the LH1 complexes (Q_y at 877 nm), a diagnostic peak of a BChl a precursor at 420 nm which arises due to oxidative stress, and the unusual decoupling between β-carotene and BChla.

We repeatedly observed that SWGK46 displayed instability during the process of up-scaling under semi-aerobic, dark conditions, and showed light-induced secondary mutations under photoheterotrophic conditions. The phenotypes observed were:

(1) A SWGK46 **brown** phenotype (designated as **SWGK46B**) was observed when SWGK46Y was grown in M2SF$^+$ medium at the 3-litre scale. The absorption spectrum of whole cells of SWGK46B showed depressed levels of ICM and carotenoid compared to those of SWGK46Y (**Fig 5.1 (A)**).

(2) A SWGK46 **green** phenotype (designated as **SWGK46G$_R$**) was selected under photoheterotrophic conditions which could resume the SWGK46Y phenotype under dark, aerobic conditions. Absorption spectra from the whole cells of SWGK46G$_R$ showed depressed levels of carotenoid, although the ICM level was maintained the same as that of SWGK46Y (**Fig 5.1 (B)**).

(3) A second SWGK46 **green** phenotype (designated as **SWGK46G$_{IRR}$**) was also selected under photoheterotrophic conditions which remained green in color when grown under dark, aerobic conditions.
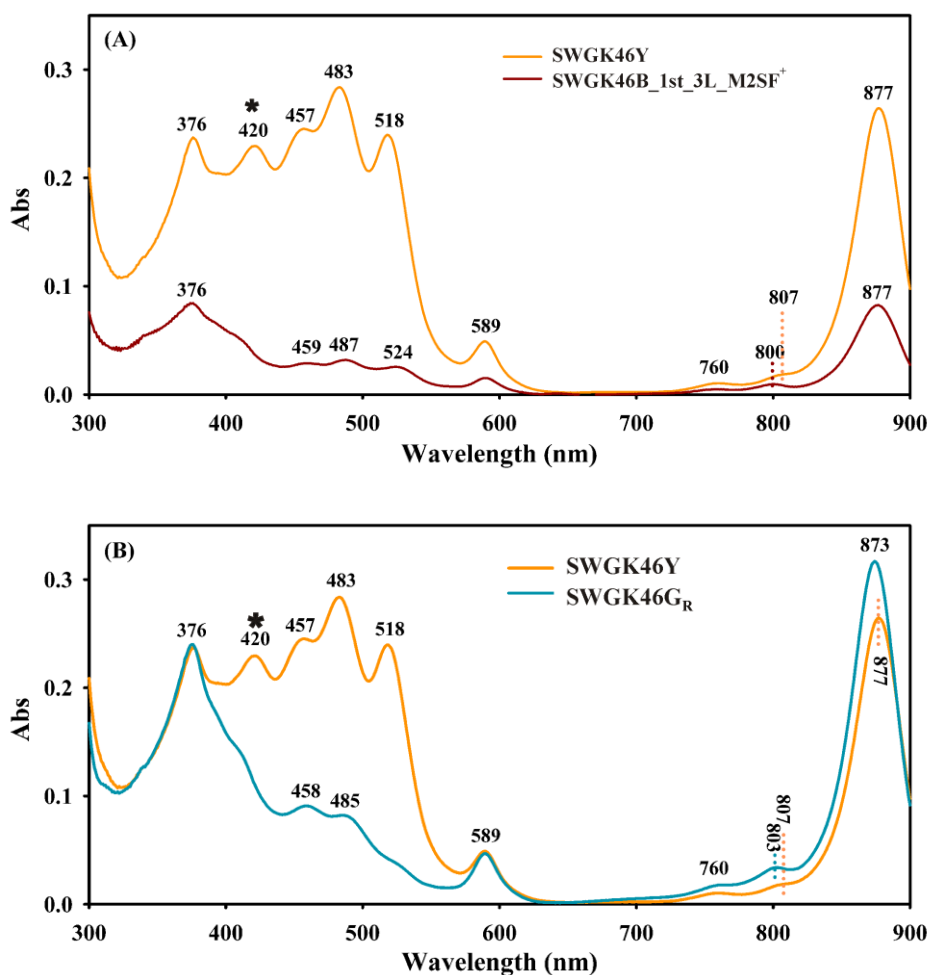
Fig 5.1 Absorption spectra of the unstable SWGK46 phenotypes. The whole cell spectra of SWGK46B (A) and SWGK46G$_R$ (B) are shown in comparison to that of SWGK46Y, respectively. The peak at 420 nm, which is diagnostic for oxidative stress, is indicated as an asterisk.

This chapter focuses on the SWGK46 secondary mutant phenotypes (SWGK46B and SWGK46G$_R$), which appear to arise due to susceptibility of SWGK46Y to oxidative and light stress, respectively. A series of experiments were designed to help to exclude possibilities and focus on the limiting factors.

5.2 The properties of the SWGK46B strain

5.2.1 Whole cell spectral and carotenoid analysis of the SWGK46B strain

In comparison to those of SLYC18 and SWGK46Y, the cell spectrum of SWGK46B displayed depressed carotenoid and ICM levels (**Fig 5.2 (A)**). The LH1 Q$_y$ absorption maximum of SWGK46B was observed at 877 nm which is the unique for LH1 complexes of SWGK46Y, probably due to the incorporation of β-carotene. However, the carotenoid absorption maxima of SWGK46B were observed at 459, 487, and 524 nm, very similar to those of SLYC18 (460, 488 and 524 nm). Carotenoid extraction was performed from the stationary phase cells of SWGK46B grown in 3-litre M2SF$^+$ medium. The absorption spectrum was taken from the hexane extract of the SWGK46B which

surprisingly fitted exactly with that of lycopene extracted from a 3-litre M2SF$^+$ culture of SLYC18 (**Fig 5.2 (B)**). The product of the SWGK46B culture grown in 3-litre M2SF$^+$ medium was lycopene.

The ratio between the absorption maxima of the LH1 complexes Q$_y$ and the carotenoid is used to indicate the BChla to carotenoid level. The A$_{877}$/A$_{487}$ of SWGK46B was 2.59, which was different from that of SWGK46Y (A$_{877}$/A$_{483}$, 0.93) which is characteristic of decoupled carotenoid from BChla, as well as that of SLYC18 (A$_{882}$/A$_{488,}$ 1.96) which shows typical carotenoid level coupled to that of BChla.
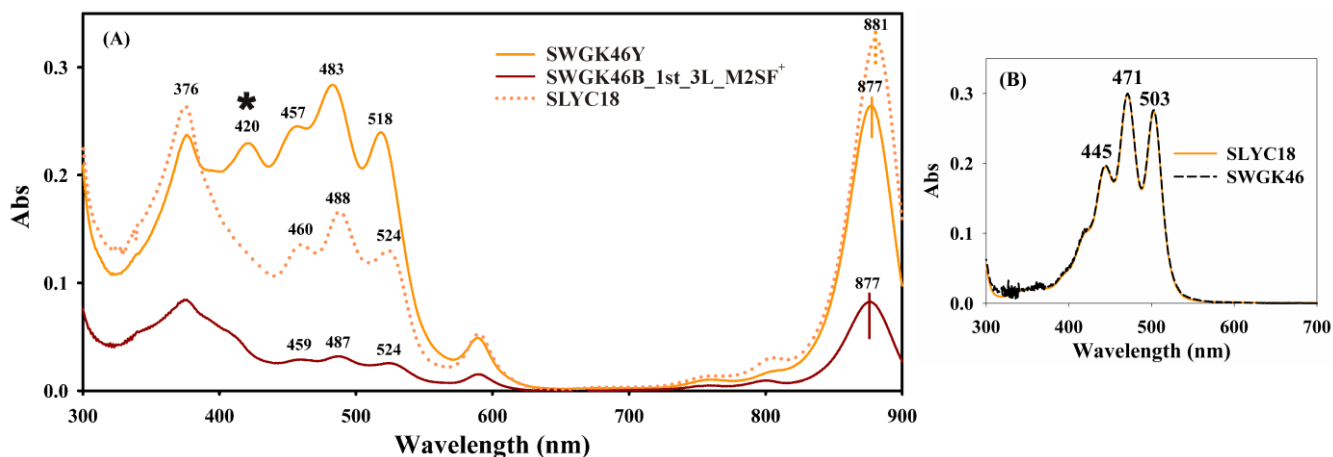
Fig 5.2 Spectral analysis of SWGK46B cells. The absorption spectrum of whole cells of SWGK46B was compared to those of SWGK46Y and SLYC18 (A). Spectral analysis of the hexane extract from stationary phase cells of the SWGK46B grown in 3-litre M2SF$^+$ medium is shown in comparison with that taken from the hexane extract from SLYC18 culture grown in 3-litre M2SF$^+$ medium (B).

The second derivative spectra are often a useful aid for analysing overlapping spectral features. An interesting observation was that the peak at 420 nm which is diagnostic for oxidative stress (**C. Autenrieth and R. Ghosh, unpublished**), was observed from both the second derivatives of the cell spectra of SWGK46Y and SLYC18 **Fig 5.3 (A) and (B)**), but was absent from that of SWGK46B (**Fig 5.3 (C)**).
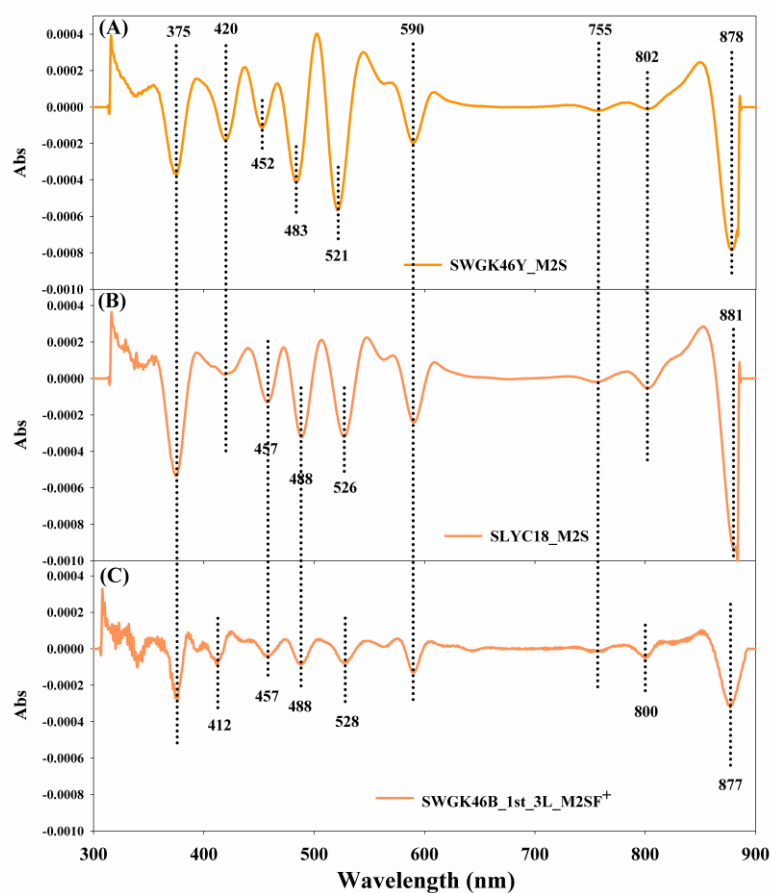
Fig 5.3 Second derivative analysis of the cell spectrum of SWGK46B. The second derivatives of the cell spectra of SWGK46Y (A), SLYC18 (B), and SWGK46B (C) were shown for comparison. Equal amounts of cells have been used for all experiments shown here.

5.2.2 The possibilities that lead to the creation of the SWGK46B

SWGK46B produced mainly lycopene when it grew under semi-aerobic, dark conditions in M2SF$^+$ medium at 3-litre scale. There are several possibilities which might lead to this case:

(1) **One reason could be that the SWGK46 3-litre M2SF$^+$ culture was contaminated with a non-β-carotene producing strain.** As a control, 200 μl aliquots of the stationary phase culture were plated on LBKan$_{50}$ plate and grown at 37°C overnight. No growth on the LBKan$_{50}$ plate was observed. Therefore, the possibility of contamination with non-β-carotene producing *E. coli* during the process of scaling up was excluded.

(2) **M2SF$^+$ medium might cause physiological stress which results in the CrtL inactivation and leads to the down-regulation of the expression of ICM, as well as carotenoid.** The second derivative of cell spectrum of SWGK46B was analysed in comparison with those of SWGK46Y and SLYC18 (**Fig 5.3**). The three characteristic β-carotene absorption maxima were observed at 452, 483 and 521 nm from the second derivative of the cell spectrum of SWGK46Y (**Fig 5.3 (A)**). Though SWGK46Y suffered oxidative stress (indicated by the peak at 420 nm) due to the incorporation of β-carotene into the LH1 complexes, the β-carotene level obtained is still 1.7 times of that of lycopene

132

produced by SLYC18, whereas the ICM level in SWGK46Y is only 81% of that of SLYC18 (**Fig 5.2 (A)**). The LH1 Q_y absorption maximum of SWGK46B was observed at 877 nm without the presence of the 420 nm absorption peak (**Fig 5.3 (C)**). However, when the SWGK46B was regrown in a M2S medium, the brown phenotype was retained. This might indicate that no special M2SF[+] stress is involved.

(3) **pH dependent compositional changes in the M2SF[+] medium**

New M2SF medium at 500 ml scale and M2SF[+] medium at 3-litre were made. SWGK46 was grown in a 500 ml M2SF medium under semi-aerobic conditions, which was subsequently used to inoculate the SWGK46 3-litre M2SF[+] culture.
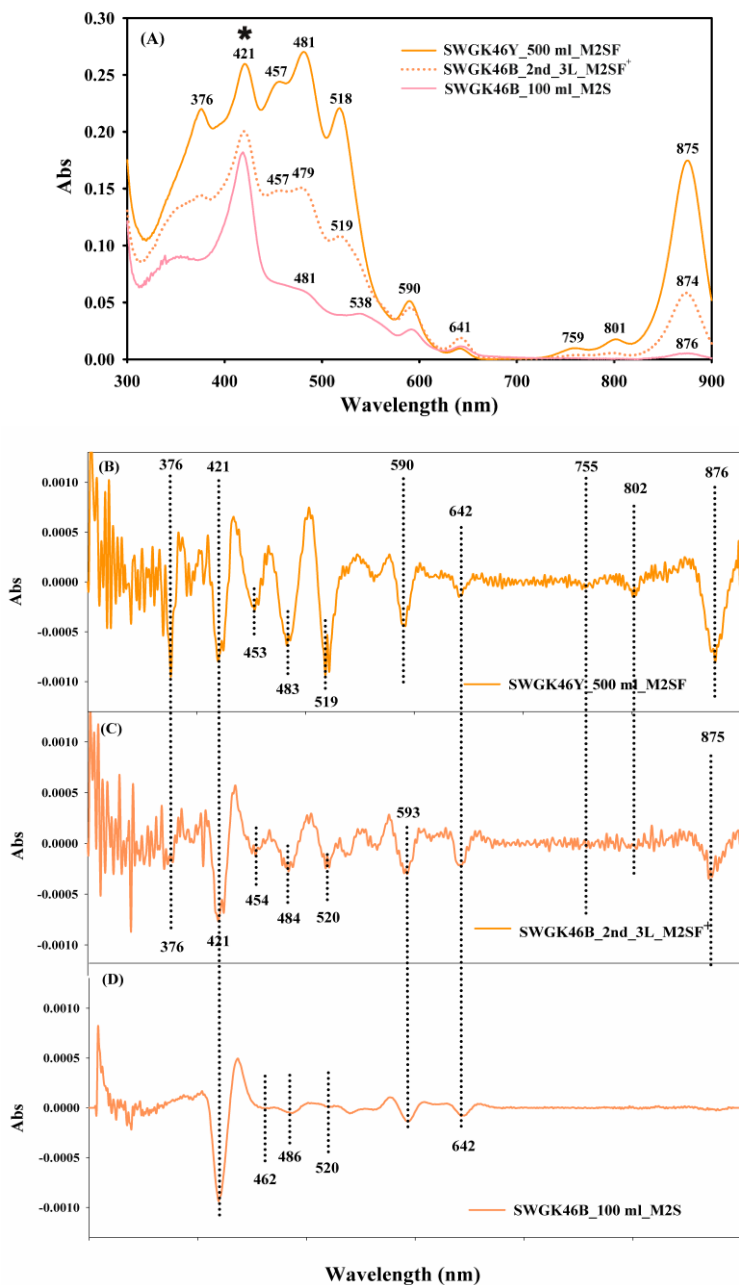
Fig 5.4 Whole cell spectra analysis of the SWGK46B. The cell spectrum of SWGK46B grown in 3-litre M2SF[+] medium was compared to those of SWGK46Y grown in 500 ml M2SF medium, and the second culture of SWGK46B grown in 100 ml M2S medium (A). The second derivative of the cell spectrum of SWGK46Y grown in 500 ml M2SF medium (B) was shown in comparison with those of SWGK46B (C) and the second culture of SWGK46B in 100 ml M2S medium (D). Equal amounts of cells were employed for these spectra.

The whole cell spectrum of SWGK46Y culture grown in 500 ml M2SF medium was analysed (**Fig 5.4 (A),** solid yellow line). Characteristic β-carotene absorption maxima at 457, 481, 518 nm and the absorption maximum of Q_y of the LH1 complexes at 875 nm were observed. The β-carotene level was decoupled from that of BChla, which was indicated by the $A_{875}/A_{481}=0.63$. This is the highest β-carotene level we can achieve so far (5.1 mg/ litre, see **Fig 4.22 (B)**). The amplitude of the peak at 421 nm increased (**Fig 5.4 (B)**) and even exceeded that of the β-carotene absorption maximum at 483 nm. It seemed the stress which is related to the effect of β-carotene upon the LH1 complexes increased. The blue shift of the Q_y absorption maximum (875 nm) was again observed.

The whole cell spectrum of the SWGK46B culture, grown in 3-litre M2SF$^+$ medium, exhibited depressed levels of ICM and carotenoids (**Fig 5.4 (A)**, dotted line). It was also observed that the Q_y of the LH1 complexes absorbed maximally at 874 nm. The carotenoid absorption maxima were observed at 457, 479, 519 nm which were partially similar to those of β-carotene (457, 483, 518 nm). The peak at 421 nm dominated over the carotenoid absorption maximum at 483 nm (**Fig 5.4 (A, C)**). And the irreversibility of this stress was checked by regrowing SWGK46B in a 100 ml M2S medium and from which the absorption spectrum of whole cells was performed (**Fig 5.4 (A), (D)**, solid pink line) and the expression of membrane and carotenoids was even more depressed. The 420 nm absorption peak, together with absorption peaks at 534 nm and 643 nm, indicated the availability of BChla precursor which was also observed in mutant OSPUHB (C. Autenrieth and R. Ghosh, unpublished data) due to O_2 stress (**Fig 5.5**). The SWGK46B phenotype and the presence of BChla precursor are characteristic of oxidative stress. Presumably, with successive growth passages, a small brown viable population (has less oxidative stress) in the culture starts to dominate the SWGK46Y ongoing population.
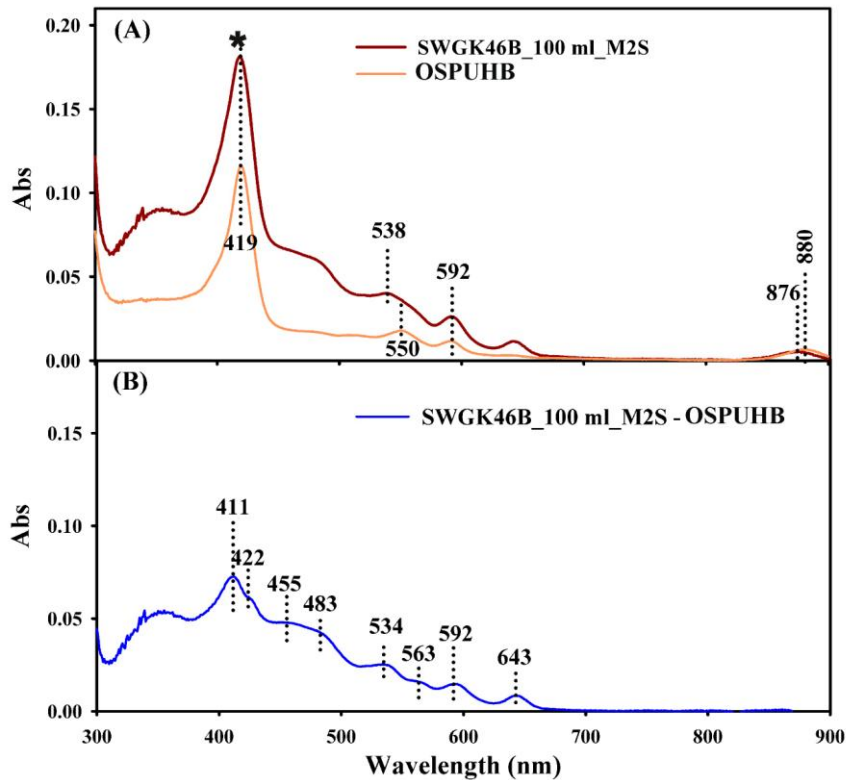
Fig 5.5 Whole cell spectra analysis of the second culture of the SWGK46B. The cell spectrum of the second culture of the SWGK46B was compared to that of the OSPUHB (A). The difference between these two spectra was also shown (B).

I propose that the possible reasons of the SWGK46B phenotype are as follow: (1) Upon the incorporation of β-carotene into the LH1 complexes, the interaction between carotenoid and the residues in the carotenoid binding site changes which could cause conformational changes of the LH1 complexes and make the BChla more accessible to O_2. (2) When lycopene is replaced with β-carotene, the quenching of the $BChl^T$ shows low efficiency because the energy level of β-carotene (9 conjugated double bonds) is higher than lycopene (11 conjugated double bonds). As soon as $BChl^T$ accumulates in the ICM, harmful free radicals due to the reaction with O_2 ($BChl^T + O_2 \rightarrow BChl + 2O^*$) will be produced, which create extreme stress for the organism (Cogdell *et al.*, 2000). Presumably, the survival strategy of SWGK46B is to down-regulate the expression of carotenoid and ICM. The changes of the carotenoid absorption maxima could be due to the changes of the interaction between carotenoid and residues in its binding site in the membrane (**see more discussion in 5.5.3 OCP**).

5.3. The observation of the SWGK46 grey/green phenotype

200 µl aliquots taken from the SWGK46 **original** slope culture, obtained from the initial conjugation mix and screened under photoheterotrophic conditions, was plated on MK_{20} plate grown under photoheterotrophic conditions. For single colonies, a three-fold streak was performed on MK_{20} plate and grew under the same conditions. A grey/green phenotype of SWGK46 was observed (**Fig 5.6**) and

was designated as **SWGK46G**. The homogenous grey/green lawn could lead to the interpretation that a further secondary mutation of SWGK46 was produced by some environmental or genetic factors.

Fig 5.6 The SWGK46G phenotype. A three-fold streak was performed on an MK_{20} plate from the SWGK46 original photoheterotrophic slope culture grown under photoheterotrophic conditions.

5.3.1. The cell spectral analysis of the SWGK46G culture

The cell spectral characteristics of the SWGK46G culture which grown in 100 ml M2S medium were as follows: (1) The photosynthetic membrane level was nearly the same as that of SLYC18, however, the LH1 Q_y absorption maximum was observed at 873 nm, which is characteristic for LH1 complexes formed without bound carotenoid, as observed in carotenoid-less strains G9 and ST2 (Wiggli *et al*., 1996) (**Fig 5.7 (A)**). (2) The carotenoid level was depressed but the ICM and LH1 complex levels were maintained the same as those of SLYC18 (A_{873}/A_{485}=3.84). (3) The carotenoid species in the SWGK46G absorbed maximally at 459, 491 and 527 nm (**Fig 5.7 (D)**), which are different from those of SWGK46Y (452, 483 and 521 nm) (**Fig 5.7 (B)**), and similar to those of SLYC18 (457, 488 and 526 nm) (**Fig 5.7 (C)**). (4) The diagnostic peak for stress at 420 nm was not present in the second derivative of the cell spectra of the SWGK46G.
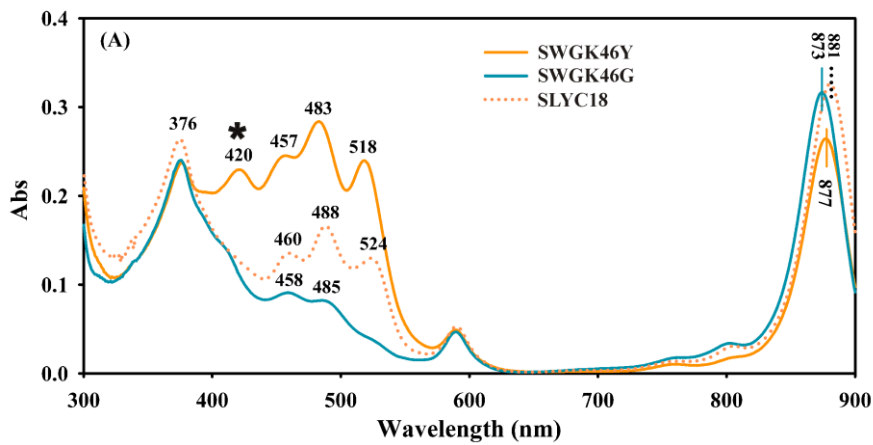
Fig 5.7 Whole cell spectra analysis of the instable SWGK46G. The cell spectrum of SWGK46G was compared to those of SWGK46Y and SLYC18 (A). The second derivatives of the cell spectra of SWGK46Y (B), SLYC18 (C) and SWGK46G (D) are shown in comparison, and the region near Q_y (850 nm-890 nm) has been enlarged for better examination of the difference among the three strains (E).
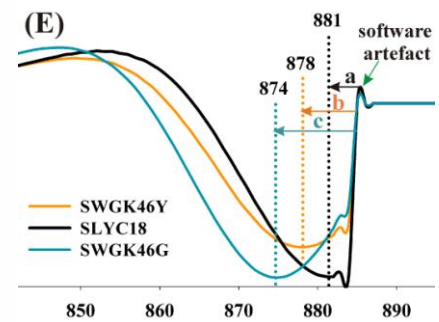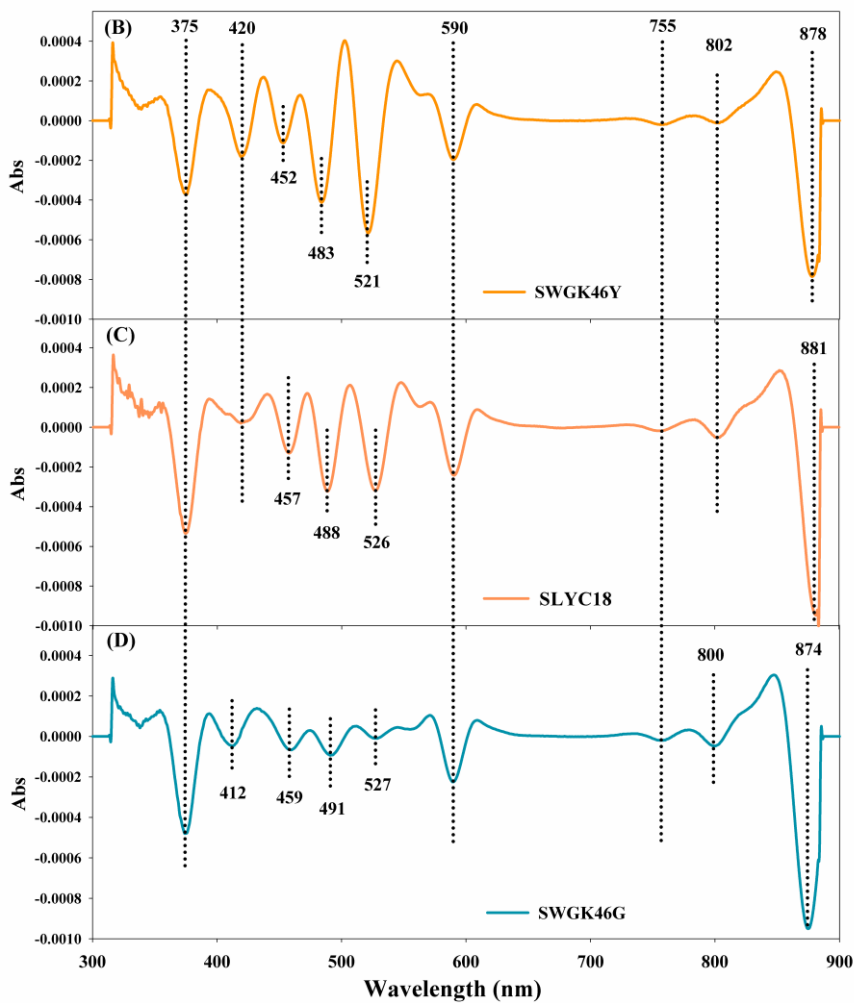
5.3.2. The recovery of SWGK46G to SWGK46Y phenotype

Since SWGK46G was selected under photoheterotrophic conditions, I tested whether SWGK46G could be recovered to form the SWGK46Y phenotype under aerobic, dark conditions? A three-fold streak

plating was performed on MK_{20} plate from single colony of the SWGK46G grown under aerobic, dark conditions at 30° and RT, respectively.

The results of the parallel experiments at 30° and RT were consistent. The SWGK46G phenotype separated into two groups as follows: (1) A SWGK46G phenotype, selected under photoheterotrophic conditions which could regain the SWGK46Y phenotype under aerobic and dark conditions, was designated as **SWGK46G$_R$** (the yellow streaks in **Fig 5.8**); (2) Another SWGK46G phenotype, also selected under photoheterotrophic conditions, but did not turn yellow SWGK46Y under aerobic and dark conditions, was designated as **SWGK46G$_{IRR}$** (indicated in **Fig 5.8**, IRR stands for *irr*eversible).
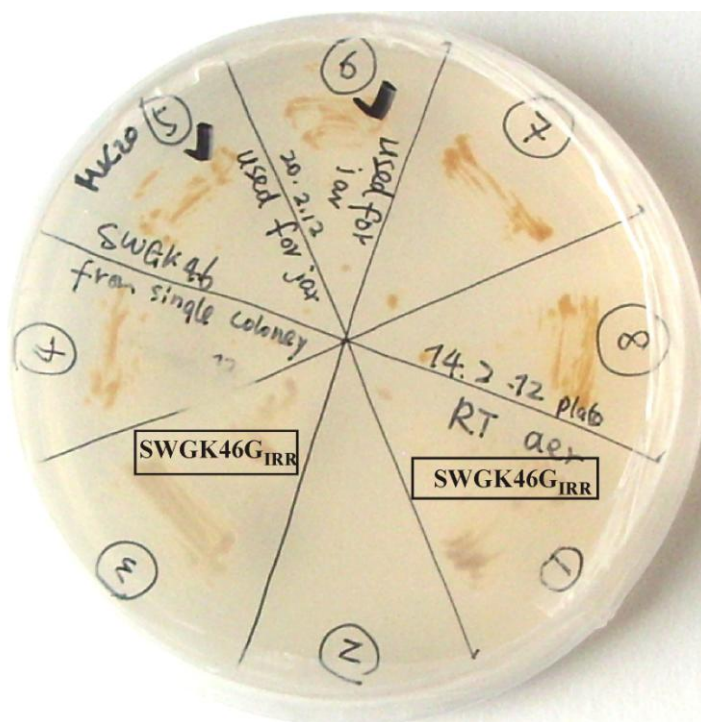
Fig 5.8 The transition from SWGK46G phenotype to SWGK46Y phenotype under dark, aerobic conditions. SWGK46G$_R$ showed a yellow phenotype (sections 4-8), and the SWGK46G$_{IRR}$ showed a grey/green phenotype (sections 1 and 3).

5.3.3. The reproducibility of the SWGK46Y to SWGK46G transition

Whether the aerobic SWGK46Y phenotype will stay yellow when it grows under photoheterotrophic conditions? A three-fold streak plate of SWGK46Y was grown under photoheterotrophic conditions. Again, two phenotypes were observed, the SWGK46Y and SWGK46G (**Fig 5.9**). The observation indicated that under photoheterotrophic conditions, there could be some regulating factors (for example, light intensity, light quality) which resulted in different phenotypes even from the same SWGK46Y phenotype.
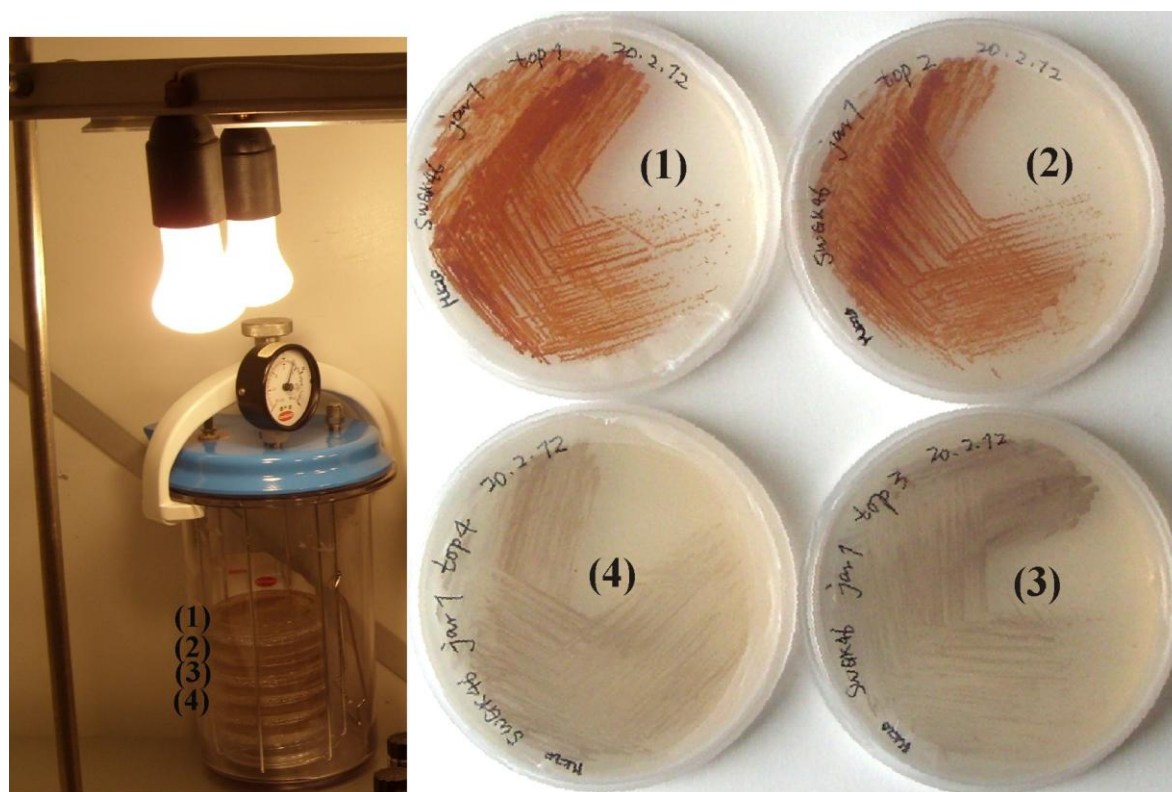
Fig 5.9 SWGK46Y↔SWGK46G transition. Some of the aerobic, dark SWGK46Y phenotypes changed back to SWGK46G phenotypes under photoheterotrophic conditions (plates (3) and (4) as indicated) and some did not (plates (1) and (2)). The numbering of the plates also corresponds to the numbering of the plate level in the jar shown on the left image.

5.3.4 The growth experiments of SWGK46Y and SWGK46G phenotypes under photoheterotrophic conditions

SWGK46Y and SWGK46G cultures were plated out again on MK_{20} plate grown under photoheterotrophic conditions. Three experiments were performed respectively which basically proved that SWGK46Y phenotype stayed yellow under photoheterotrophic conditions. SWGK46G showed diversity, some stayed grey (**Fig 5.10 (A)**), some changed to yellow (**Fig 5.10 (C)**) and some exhibited mixed yellow/grey colors (**Fig 5.10 (B)**). A growth and color gradient was observed (**Fig 5.10 (C)**), which displayed that both the phenotypes of SWGK46Y and SWGK46G exhibited intense color and more growth near the edges of the plate, and weak color and less growth in the middle of the plate, which observation was called "**edge effects**". This is the first time that "edge effects" have been observed in photosynthetic bacteria so far. The cell mass distribution is independent of the direction of the illuminating light, which is also not SWGK46Y or SWGK46G phenotype-dependent. The edge effect is probably due to the diffusion of a low molecular weight substance, which possibly inhibits

139

growth. At the edges of the colony, this component (not yet identified) can diffuse away and normal growth can continue.
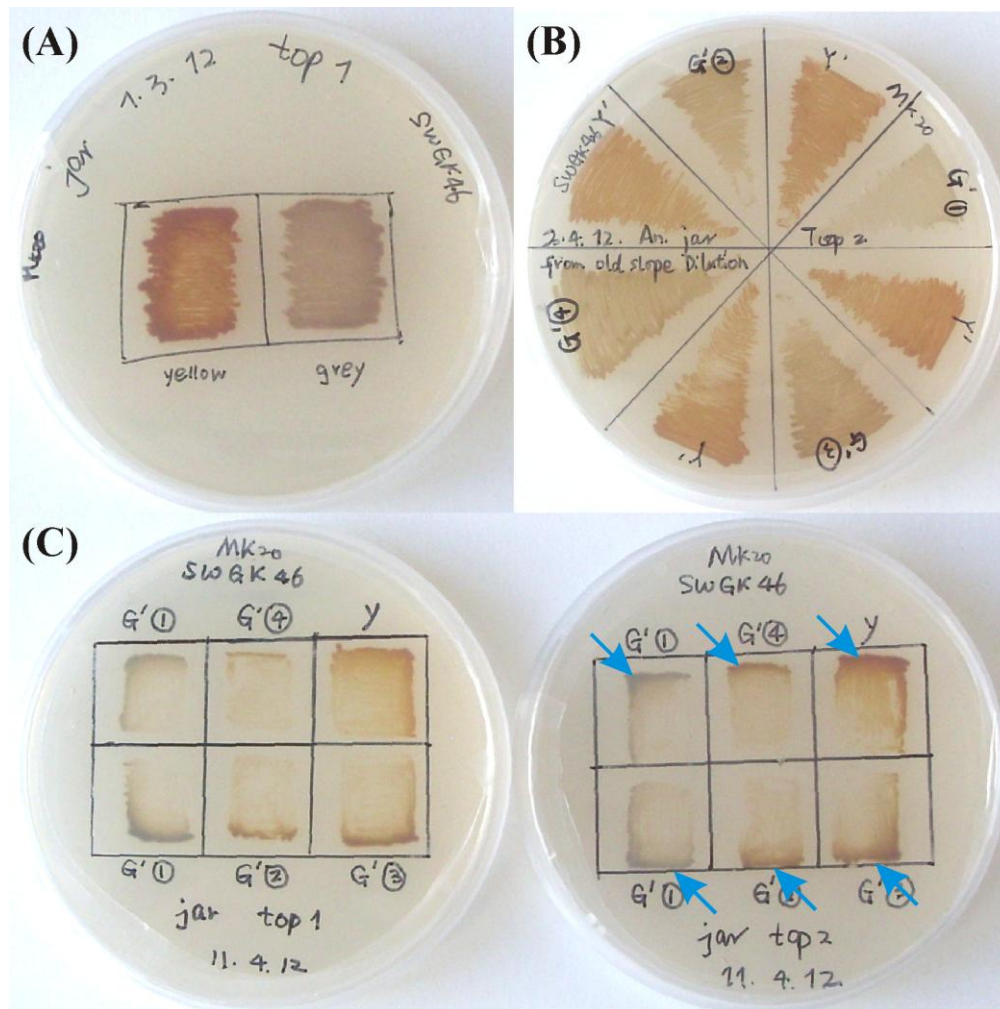
Fig 5.10 The diversity of SWGK46Y and SWGK46G phenotypes under photoheterotrophic conditions. (A) The SWGK46Y and SWGK46G phenotypes stayed yellow and grey, respectively. (B) The SWGK46Y phenotype stayed yellow and the SWGK46G phenotype exhibited a mixture of yellow and grey colors. (C) Both SWGK46Y and SWGK46G phenotypes exhibited characteristic "edge effects" (blue arrow).

5.3.5. Investigation of the populations in the initial anaerobic slope stock culture of SWGK46Y

The initial anaerobic slope culture of SWGK46Y, derived from the primary screening after conjugation, was maintained at 4°C for several months. We investigated the populations in the initial anaerobic slope culture by serial dilution of the anaerobic slope supernatant followed by plating (200 µl of each sample) onto MK_{20} and $LBKan_{50}$ agar, grown under aerobic, dark conditions at 30°C.

On the MK_{20} plate, a white lawn was formed with a weak and obscure color. There were a few colonies on top of the lawn but the color was also weak compared to the parallel experiments on $LBKan_{50}$ plates. The initial anaerobic slope stock culture of SWGK46Y, four phenotypes showed that (the result for the $LBKan_{50}$ plate):

(1) One phenotype of SWGK46 exhibited in large, yellow colonies (designated as **SWGK46YR**).

(2) Another phenotype of SWGK46 exhibited in small, yellow colonies (designated as **SWGK46SY**).

(3) The third phenotype of SWGK46 exhibited in large, green colonies (designated as **SWGK46G**).

(4) The last phenotype of SWGK46 exhibited in small, green colonies (designated as **SWGK46SG**).
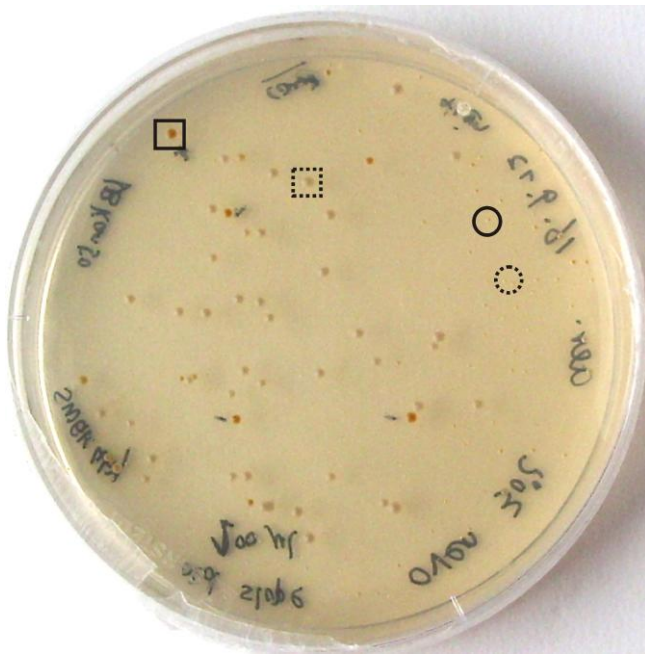
Fig 5.11 Populations in the initial anaerobic slope stock culture of SWGK46Y. Four phenotypes were indicated as follows: SWGK46YR (solid line square), SWGK46G (dotted line square), SWGK46SY (solid line circle) and SWGK46SG (dotted line circle).

The major population (72%) of the initial SWGK46Y anaerobic slope culture was the SWGK46SG strain with a colony diameter of about 0.2-0.3 mm. The SWGK46SY strain has the same colony size as the SWGK46SG and comprises 14% of the population. The SWGK46YR and SWGK46G strains have a colony diameter of about 1.5-2 mm, and occupy 2% and 12%, respectively. The SWGK46G strain, with growth and survival advantages, grows more quickly than other phenotypes and could thus take over and dominate the population under photoheterotrophic or other stress conditions. The subpopulations except SWGK46YR selected by initial anaerobic selection probably only become apparent after multiple growth passages.

Single colonies of the four phenotypes purified from the initial anaerobic slope stock culture of SWGK46Y were plated on MK_{20} and $LBKan_{50}$ plates and grown under dark, aerobic conditions at RT (23°C) and photoheterotrophic conditions at 30°C, respectively.
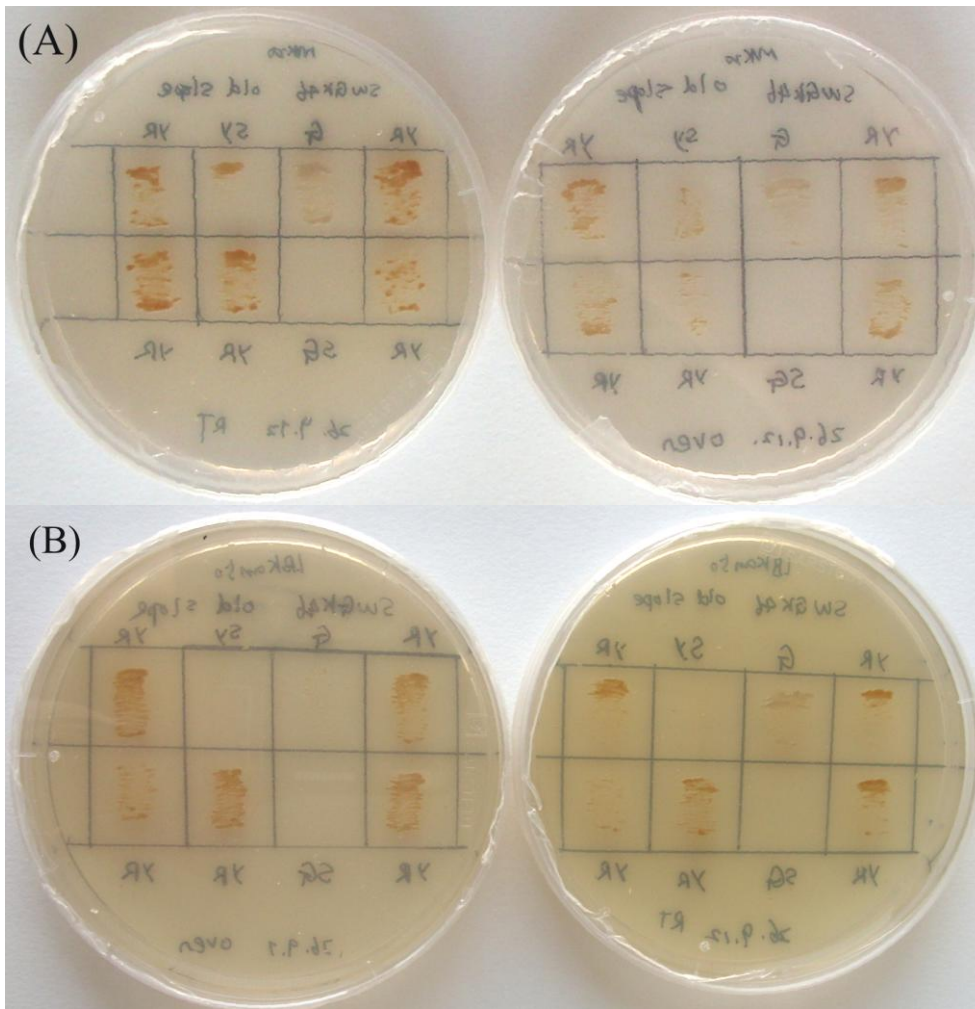
Fig 5.12 Reversibility of the four phenotypes in the initial anaerobic slope culture of SWGK46Y under dark aerobic conditions. SWGK46 colonies of the four phenotypes (YR, G, SY, SG) were replated on MK$_{20}$ plate (A) and LBKan$_{50}$ (B).

Results: (1) SWGK46YR and SWGK46G colonies retained their initial phenotypes under aerobic, dark conditions, both at 30°C and RT (**Fig 5.12**). (2) SWGK46SY grew slower, and SWGK46SG did not grow at all. The same results were obtained from the parallel experiment under photoheterotrophic conditions at 30°C.

5.4 The possible reasons of the SWGK46G$_R$ phenotype

The SWGK46G$_R$ phenotype could be due to several possible reasons as follows:

(1) CrtL may be mutated or deleted by chromosomal DNA arrangement.

(2) There may be an effect due to **light intensity** and **light quality**.

(3) The **light wavelength** may affect the expression of CrtL.

(4) The folic acid content in the medium may affect the expression of CrtL.

These possibilities mentioned above were investigated and are mentioned in details as follows.

5.4.1. Is the CrtL mutated in the SWGK46G$_R$ phenotype?

Two pairs of PCR primers were designed to obtain the complete gene sequences of *crtL* from both the SWGK46Y and SWGK46G$_R$ phenotypes, using their chromosomal DNA as templates. One pair of the primers, crtLup and NPT1390, were designed according to the gene sequences in the upstream and downstream of *crtL*. The other pair of the primers, FcrtL828 and RcrtL933 were designed according to the gene sequences in the middle of the *crtL* gene with 100 bp overlapping in two directions (**Fig 5.13**).
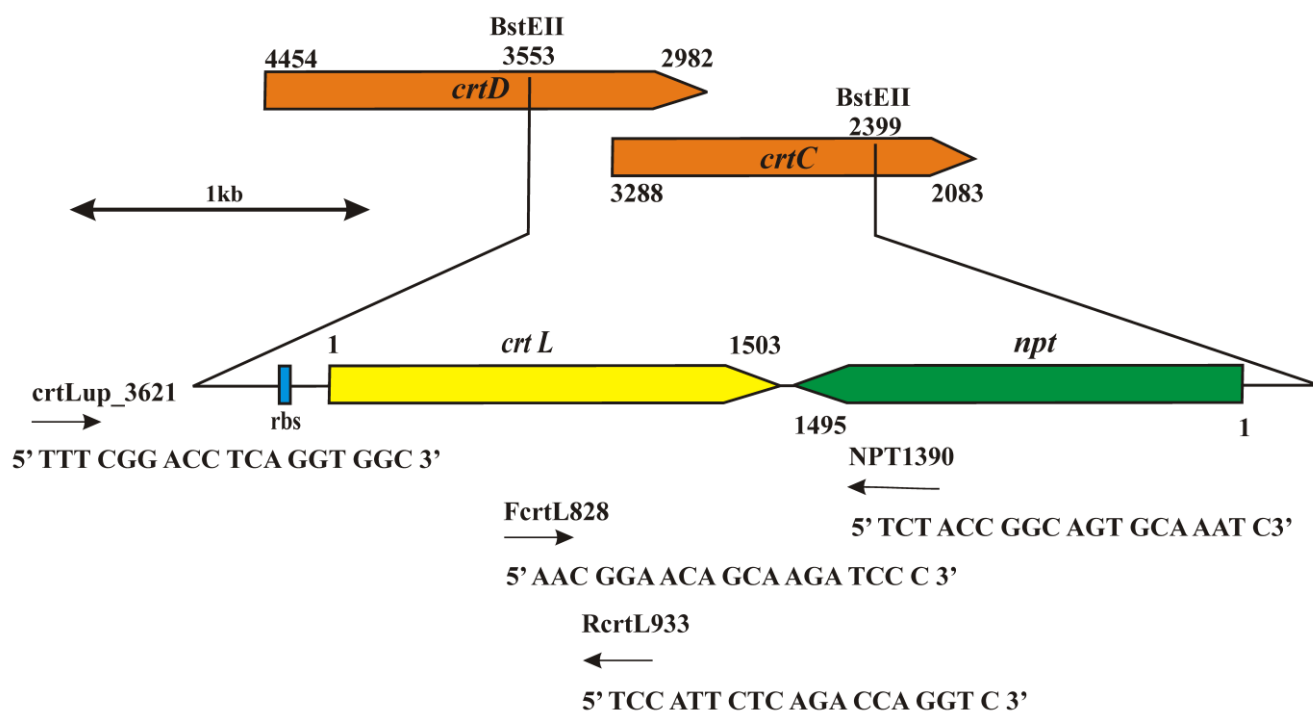
Fig 5.13 Scheme of the obtaining the *crtL* gene from SWGK46Y and SWGK46G$_R$ phenotypes by PCR. The sequences of the two pairs of primers (crtLup/NPT1390 and FcrtL828/RcrtL933) are indicated.

The amplified PCR fragment containing the *crtL* genes from both phenotypes of SWGK46Y and SWGK46G$_R$ were sequenced directly. From the sequence data, we are confirmed that the *crtL* genes from both phenotypes show 100% homology to the gene sequence of the wild-type lycopene β-cyclase (**Fig 5.14**). Therefore, the CrtL in the SWGK46G$_R$ strain is intact and unchanged in sequence. Thus the inactivation of CrtL in SWGK46G is not due to a rearrangement of *crtL* gene. However, there is still the possibility that CrtL has been "switched off" by an unknown regulation mechanism which will be considered later.
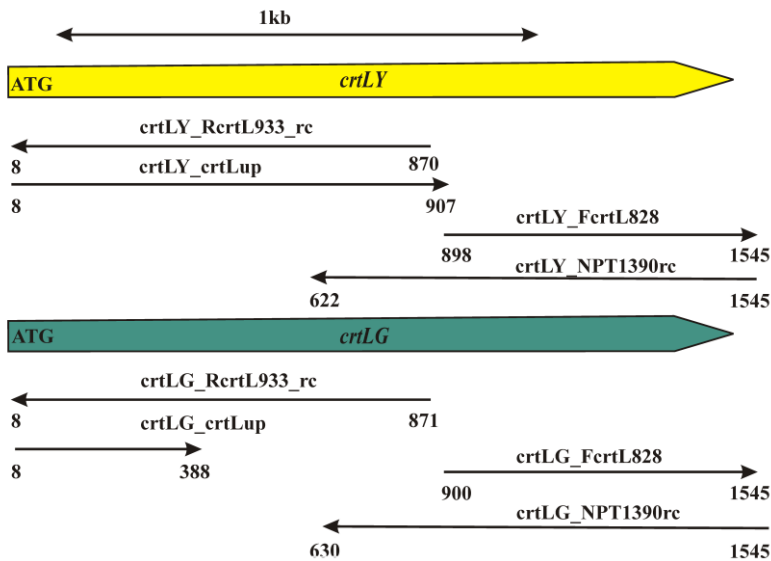
Fig 5.14 The scheme of amplification of *crtL*s from SWGK46Y and SWGK46G phenotypes by PCR.

5.4.2. Light intensity and quality effect

We discovered that the light intensity and quality does affect the expression of the lycopene β-cyclase during the reversible SWGK46Y to SWGK46G transition under photoheterotrophic conditions. At the beginning of my PhD, the culture room was originally set up with tungsten bulbs (more intensity in the green (495 nm) to infrared region (900 nm)), which were gradually replaced with energy saving bulbs (Krypton, Osram) (more energy in the blue region (450-495 nm), similar to LEDs) (**Fig 5.15**).

We know that blue light induces photochemistry in many photoreceptors and regulates many activities in photosynthetic organisms. We observed that SWGK46G responded differently to tungsten light (**Fig 5.12**) and the light of the energy saving bulbs (**Fig 5.9** and **Fig 5.10**). In particular, the grey/green SWGK46G$_R$ phenotype arose more rapidly when the culture of SWGK46Y were illuminated with the bluish light of the energy saving bulbs. Thus, blue light may be one factor causing the down-regulation of carotenoid production and the inactivation of CrtL in SWGK46G phenotype.
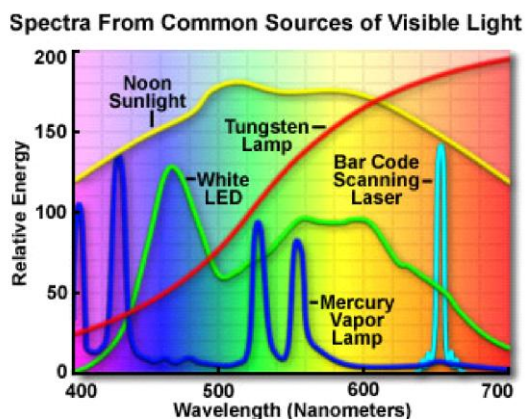
Fig 5.15 Scheme of relative energy of common sources of visible light. The light spectra of tungsten lamp and white LED are indicated in red and green, respectively (taken from freedomlightbulb.blogspot.com).

144

To test the effect of light quality upon CrtL expression, SWGK46Y was cultivated under photoheterotrophic conditions, using light filtered by a green plastic cover and a Kodak 92 filter (gelatin filter, Kodak), respectively. Light passing through the green plastic cover is limited to the 475-570 nm and 830-900 nm regions, and has a transmission of around 50%-73%. The Kodak 92 filter can filter the light from violet to orange, with 90% filtered light corresponding to the 620-900 nm region (**Fig 5.16 (A)**).

From the spectral analysis of the whole cells, the carotenoid level of the SWGK46 photoheterotrophic culture cultivated with light filtered by the green plastic cover was only one third of that cultivated with Kodak 92 filter. However, with the Kodak 92 filter, the decoupling between carotenoid and BChla was observed. Even in the second derivative spectra, the diagnostic peak for stress at 420 nm was absent for both of the cultures. However, the carotenoids expressed with the green plastic cover (459, 489, 528 nm) and the Kodak 92 filter (455, 486, 525 nm) are different from β-carotene (452, 483, 521 nm). It seems that the light wavelength indeed affects the carotenoid production level. These light-dependent down-/up-regulations of gene expression are not observed in normal physiological *R. rubrum* strains or other physiological carotenoid-containing strains. Thus, the light-dependent down-/up-regulation of carotenoid production is probably due to the interaction of the lycopene β-cyclase with the early carotenoid biosynthesis enzymes.
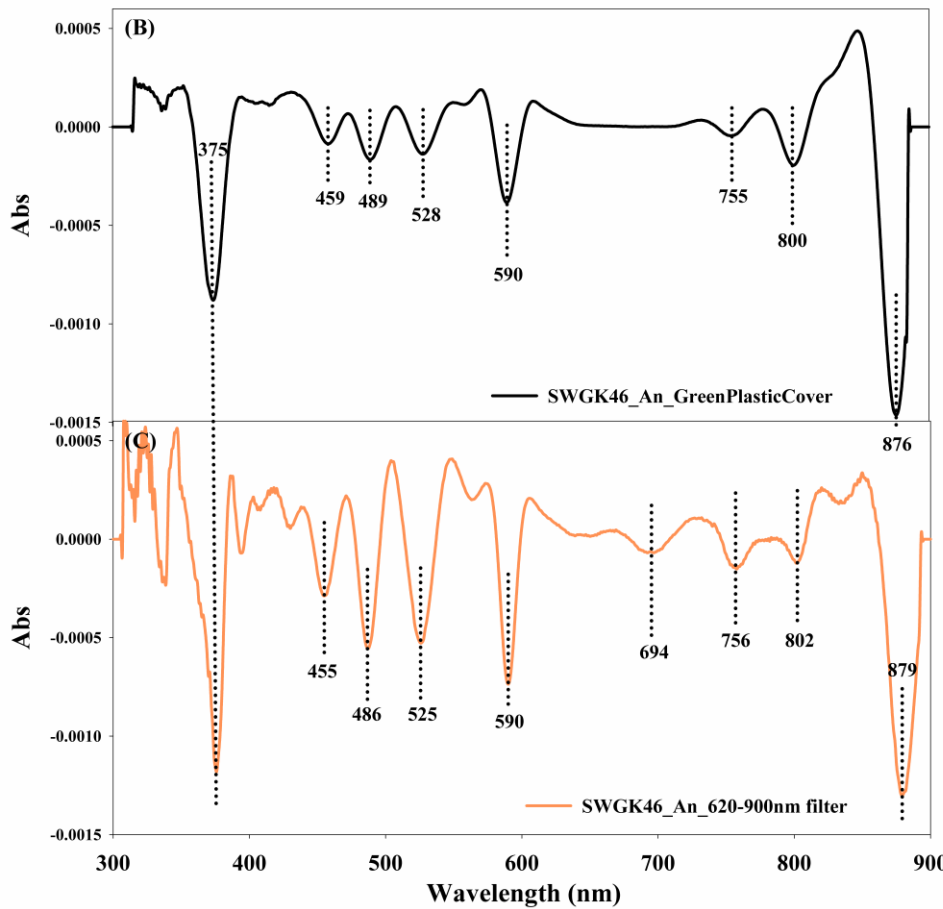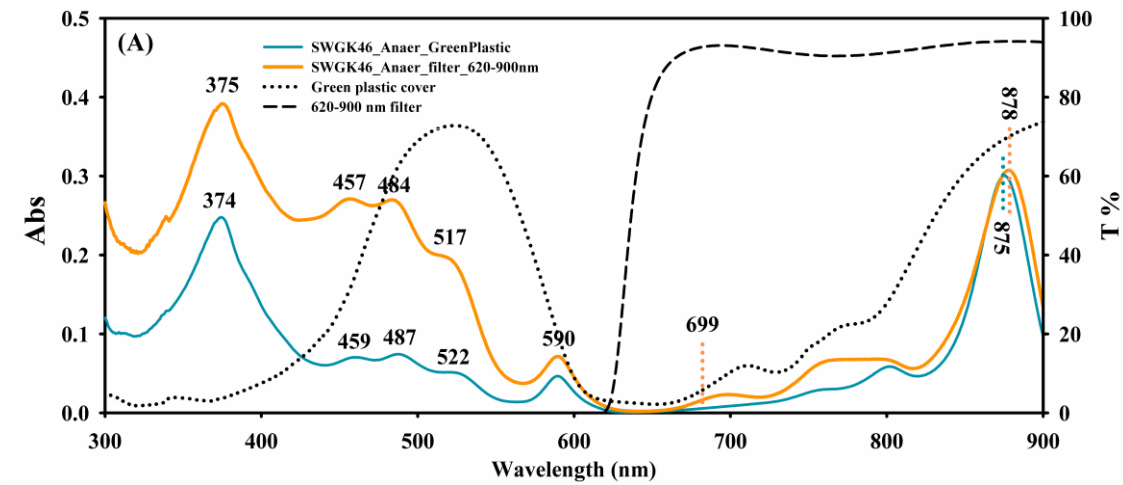
Fig 5.16 The effect of light wavelength upon SWGK46Y carotenoid expression. The light spectra of the green plastic cover and Kodak 92 filter are shown with the cell spectra from the two photoheterotrophic cultures cultivated with light filtered by them (A). Analysis of the second derivatives of the two spectra is also shown in comparison (B) and (C).

5.4.3 Light wavelength effect

To confirm the effect of blue-green light upon the lycopene β-cyclase expression in combination with the light wavelength effect, red (625 nm) and infrared (850 nm) LEDs were used to cultivate SWGK46Y under photoheterotrophic conditions. Two SWGK46 phenotypes were observed, the

146

intense orange streaks and the light orange streaks. Carotenoid extraction from both of the phenotypes was performed, followed by Tlc analysis. It was shown that the intense orange SWGK46 phenotype was mainly due to β-carotene, but a small amount of lycopene and some BChla precursor were also observed. In contrast, the light orange SWGK46 phenotype produced only low levels of lycopene.
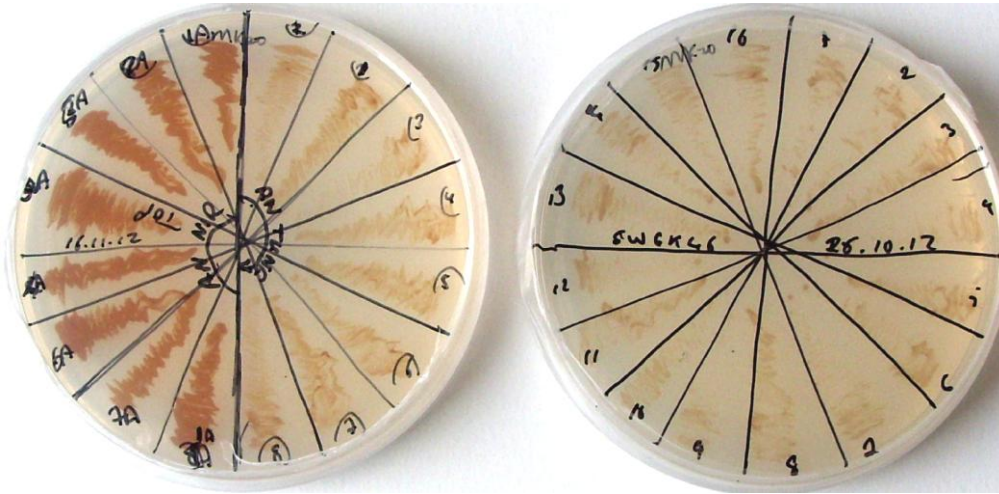
Fig 5.17 SWGK46Y replated from streaks grown photoheterotrophically either with an LED light source (wavelength at 620 nm and 835 nm, dark orange streaks) or with a tungsten light source (light orange streaks). The secondary plates shown in the figure have both been grown under photoheterotrophic conditions with tungsten light.

5.4.4 Folic acid effect upon CrtL expression

The "edge effects" were continuously observed even with the dark, aerobic SWGK46Y culture (**Fig 5.18**) which confirms that the "edge effects" are not light-dependent.
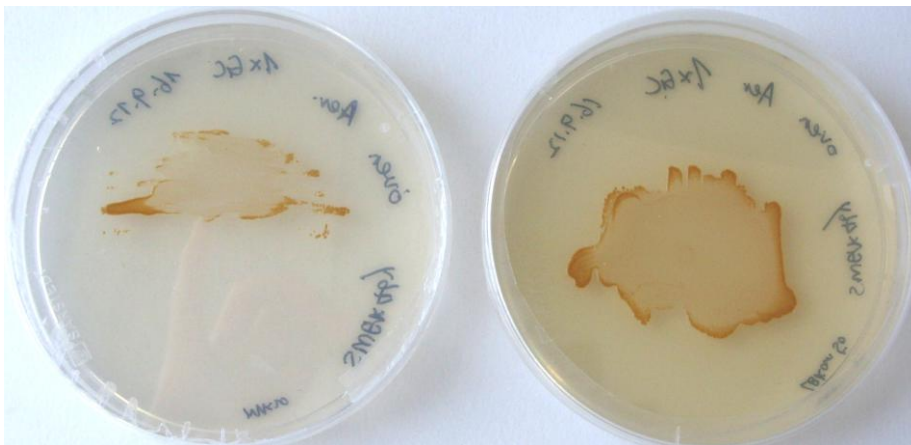
Fig 5.18 SWGK46Y grown under aerobic, dark conditions. SWGK46Y obtained from the glycerol cultures were grown on the MK_{20} and $LBKan_{50}$ plates under aerobic, dark conditions at 30°C.

To test whether the differences in the expression levels of carotenoids were due to the position in the plate, the cells from the whitish region of the lawn and the intense yellow edge region were replated on MK_{20} and $LBKan_{50}$ as eight sections, respectively. The results of parallel experiments on MK_{20} or $LBKan_{50}$ plates under dark, aerobic conditions were consistent: there was no detectable color difference between the cultures replated from the whitish region and that from the intense yellow edge. All strains exhibited yellow color, however, the "edge effects" were retained (**Fig 5.19**).
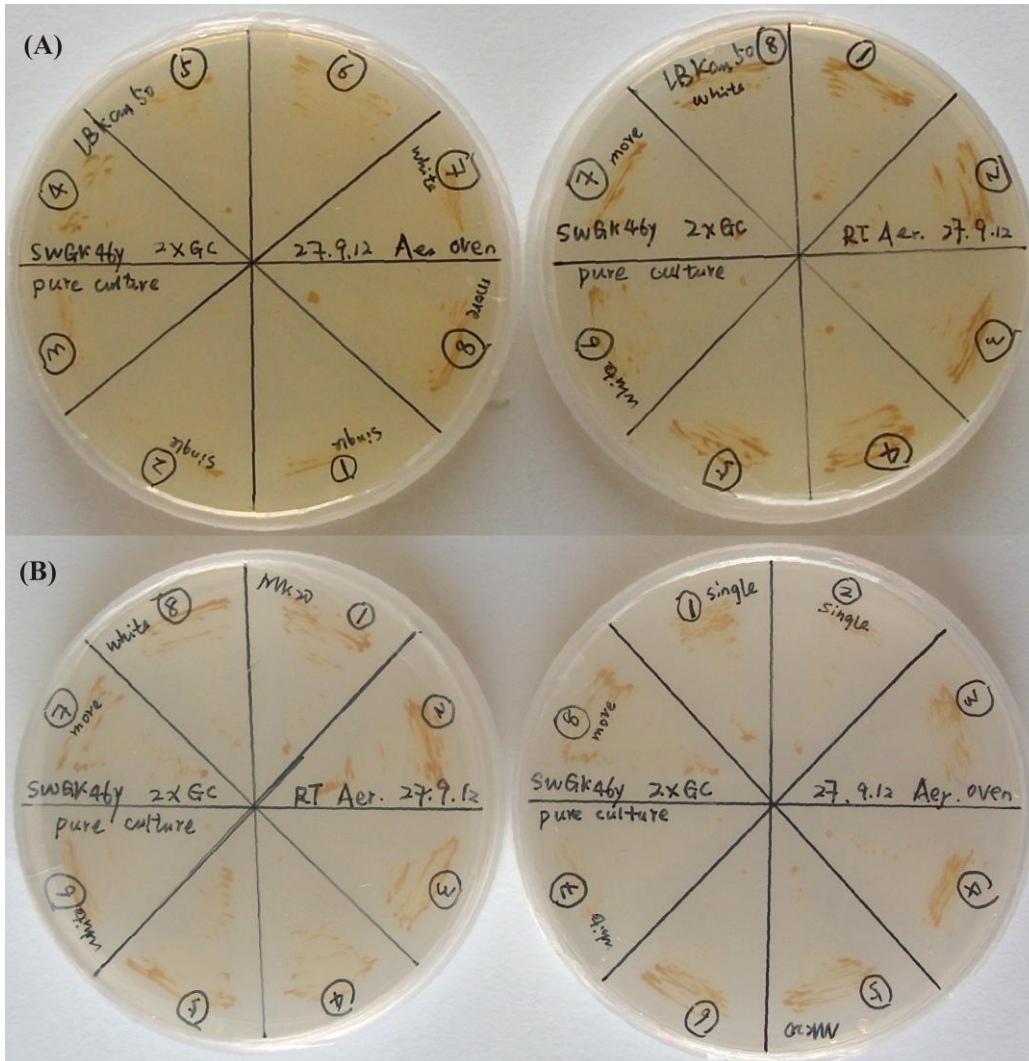
Fig 5.19 The "edge effects" observed in SWGK46Y grown under dark, aerobic conditions. Parallel experiments were performed on $LBKan_{50}$ plates (A) and MK_{20} plates (B) at 30°C and RT, respectively.

Which factors are involved leading to the "edge effects"? More access to the substances or nutrients in the medium could be one possibility. When cells grow on the edges, it has more access to the

substances or nutrients in the plate. We observed that higher expression levels of β-carotene were achieved with LB medium compared to that with M-medium. As we had begun to suspect that CrtL may also contain a blue light receptor domain, and we searched for cofactor dependence of well-known blue-light sensors. Of these, we realized that the Cry-DASH (see later) contains 5,10-methenyltetrahydrofolate (MTHF) as cofactor (Klar *et al.*, 2007). We therefore considered the possibility that folate may have an effect upon carotenoid production. LB medium is rich in yeast extract which is the major source for folic acid. As lycopene β-cyclase is a flavoprotein using FAD as cofactor, folic acid might also be one of the factors affecting β-carotene production, and which may lead to the conversion from SWGK46Y to SWGK46G phenotype when the medium is folic acid-limited. To examine this possibility, 200 µl aliquots from the SWGK46G M2S culture in the stationary phase were replated on MK_{20} and LBK_{50} plates containing 16.7, 33.3, 50, and 66.7 µM folic acid, and grown under dark, aerobic conditions at 30°C, also under photoheterotrophic conditions, respectively. Under dark, aerobic conditions, folic acid inhibited the growth of SWGK46Y on $LBKan_{50}$ but the parallel experiment on MK_{20} plates did not show inhibition effect. The same result was obtained under photoheterotrophic conditions. At present, we cannot deduce the significane of this effect.

5.5. Discussion

In this chapter, the secondary mutant SWGK46B has been described, which arises due to the oxidative stress caused by the presence of *A. thaliana* lycopene β-cyclase. SWGK46B showed the unique form of the LH1 complexes with Q_y at 877 nm whereas the carotenoid is lycopene at low level instead of β-carotene. This might indicate that the conformation of the LH1 complexes has adapted to the unique form to bind β-carotene via some mutations. This still should be examined by DNA sequencing of the LH genes. Unfortunately, the unique conformation of the LH1 complexes which enables them to bind β-carotene might also make the BChla more accessible to O_2. This would explain why the SWGK46B showed more sensitivity to O_2. The depressed levels of the ICM and carotenoids of SWGK46B are likely to be caused by a global down-regulation of the photosynthesis genes due to the oxidative stress created by the production of O_2 free radicals followed reaction with the $BChl^T$ species.

The Q_y absorption maximum, employed as a very important diagnostic tool for spectral analysis in this chapter, indicates the conformation of the LH1 complexes. It is known that the incorporation of carotenoid into the LH1 complexes caused conformational changes in the N-terminal regions of the

apo-proteins of the LH1 complexes (Brunisholz *et al.*, 1986). Lycopene has been proved to be incorporated into the LH1 complexes which showed Q_y absorption maximum at 882 nm, even though SLYC18 suffered some stress which was indicated by the small absorption peak at 420 nm from the second derivative of the cell spectrum (**Fig 5.3 (B)**). However, this stress was small and did not affect the expression levels of the ICM and lycopene in SLYC18. The Q_y absorption maximum of SWGK46Y shifts to 877 nm, which indicates that some conformational change happens in the LH1 complexes, probably due to the incorporation of β-carotene. It seemed that the incorporation of β-carotene into the LH1 complexes caused more stress for SWGK46Y, because the absorption peak at 420 nm was larger (**Fig 5.3 (A)**). However, the stress, SWGK46Y suffered under semi-aerobic, dark conditions in 100 ml M2S medium, did not effect the production of β-carotene.

Another secondary mutant, SWGK46G$_R$, showed a complex effects due to the O_2, light intensity and light quality in the presence of *A. thaliana* lycopene β-cyclase. The SWGK46G$_R$ carotenoid down-regulation in response to blue light, and the reversible transition from SWGK46G$_R$ to SWGK46Y in the dark revealed that *A. thaliana* lycopene β-cyclase might also function as a blue light sensor. The reversibility of SWGK46G$_R$ to SWGK46Y, as well as the confirmation of an unchanged DNA sequence in the SWGK46G$_R$ strain indicates that lycopene β-cyclase in SWGK46G$_R$ is intact, and that the carotenoid down-regulation mechanism might be mediated via interaction with carotenoid biosynthesis enzymes which are at regulating point, instead of lycopene β-cyclase mutation (which was also experimentally confirmed, see **5.4.1**).

What is the nature of the light-induced anaerobic stress that leads to the green phenotype? It seems likely that the altered (877 nm) conformation of β-carotene-containing LH1 coupled to the mismatch between the S_2 level of β-carotene and the O_x excited state of BChl, causing the carotenoid-mediated quenching of the light-induced BChl^T state to be very inefficient. In the absence of O_2, a long-lived BChl^T species would probably react with ground state BChla. This photodegradation process would cause the structure of the LH1 ring, which is the basic building block of the chromatophore membrane, to degrade, thereby causes severe physiological stress. It is reasonable to suppose that the cell responds to the stress by mutating a key early carotenoid biosynthesis enzyme(s) (probably CrtIB, which is indicated by preliminary DNA sequencing), thus allow the more stable carotenoid-less (blue-green) LH1 complex to be formed, and thus allow the formation of normal ICM level.

Since the *A. thaliana* lycopene β-cyclase is hypothesized to function as a new blue light sensor, it is necessary to examine the homologies of lycopene cyclase from different sources, e.g. plant, cyanobacteria and *Pantoea* species. It is also relevant to perform a bioinformatic analysis of the *A. thaliana* lycopene β-cyclase in comparison to those known blue light sensors, e.g. LOV protein, BLUF protein, cryptochromes and orange carotenoid protein (OCP). This will be described in Chapter 6.

Chapter 6   Bioinformatics analysis of lycopene cyclase

6.1. Introduction

Plant lycopene cyclases belong to the flavoprotein family and their activities require non-covalently bound FAD (Mialoundama *et al.*, 2010). The proteobacterial lycopene cyclase from *Pantoea ananatis* also catalyzes an FAD-dependent non-redox reaction (Yu *et al.*, 2010). The cofactor might act as a dipole stabilizing a cationic intermediate or transition state of the reaction in the process of cyclization. Since flavin absorbs blue-green light and the effect of blue-green light absorption upon crt expression was observed in Chapter 5, we consider that *A. thaliana* CrtL may also function as a light sensor? This chapter focuses on the bioinformatics analysis of *A. thaliana* CrtL by protein sequence alignment, comparisons of the hydropathy plot and secondary structure with those known photoreceptors which have been mentioned in the Introduction.

6.2. Methods

Protein sequence alignment was performed by using Vector NTI Align X program. **Hydropathy plots** were performed from data generated by using a Qbasic program (written by Dr. Caroline Autenrieth) to scan protein sequences with the Kyte and Doolittle scale and a moving window width of 5. **Secondary structure prediction** was performed online (Jpred3 secondary structure prediction server, http://www.compbio.dundee.ac.uk/www-jpred/).

6.3. Investigation of the relationships among the lycopene cyclases from different sources

To know more about the characteristics of *A. thaliana* lycopene β-cyclase (At_CrtL) (EMBL AAB53337), the protein sequence homology to its counterparts from *Daucus carota* (Dc_LCYB) (EMBL ABB52071), *Synechococcus elongatus* PCC7942 (Sy_CrtL) (EMBL ABB58092), *Pantoea stewartii* (Ps_CrtY) (EMBL AAA64980) and *Pantoea ananatis* (Pa_CrtY) (EMBL BAA14126) was analyzed (**Fig 6.1**).

The homology (indicated in yellow) among these five proteins is 6.9%, which is distributed along the overall sequence. The sequence homology between the two plant lycopene cyclases is high up to 76.6%, whereas that between the two *Pantoea* lycopene cyclases is 56%. The homology tree reveals that the plant lycopene cyclases and *Pantoea* lycopene cyclases belong to two different types, whereas that from cyanobacteria belongs to a third type (**Fig 6.1**) which is also consistent with the multiple

sequence alignment. *S. elongatus* CrtL shares 27.1% sequence homology with the plant counterparts and 10.8% with *Pantoea* counterparts.

```
Sy_CrtL  ----------------------------------------------------------------------------------M
Dc_CrtL  MKVMDTLLKTHNKLEFFNPIHGFPDKVGTLSCLKFRNQELRFGSRRSNVNWGKNGSVKASSSALLELVQETKKENLEFDLPLYDPSNGLV
At_CrtL  ---MDTLLKTPNKLDFFIPQFHGFERLCSNNPYHSR---VRLGVKKRAIKIVS--SVVSGSAALLDLVPETKKENLDFELPLYDTSKSQV
Ps_CrtL  ----------------------------------------------------------------------------------M
Pa_CrtL  --------------------------------------------------------------------------------MQPH
             10        20        30        40        50        60        70        80        90

Sy_CrtL  FDALVIGSGPAGLAIAAELAQRG--LKVQGLSPVDPFHPWENTYGIWGPELDSIGLEHLFGHRWSNCVSYFGEAPVQHQYN-YGLFDR--
Dc_CrtL  VDLAVVGGGPAGLAVAQQVSEAG--LAVVSIDPSPKLI-WPNNYGVWVDEFEAMDLLDCLDTTWSSAIVYIDDQTTKELGRPYGRVNR--
At_CrtL  VDLAIVGGGPAGLAVAQQVSEAG--LSVCSIDPSPKLI-WPNNYGVWVDEFEAMDLLDCLDTTWSGAVVYVDEGVKKDLSRPYGRVNR--
Ps_CrtL  RDLILVGGGLANGLIAWRLRQRYPQLNILLLIEAG------EQPGGNHTWSFHEDDLTPGQHAWLAPLVAHAWPGYEVQFPDLRRRLARGY
Pa_CrtL  YDLILVGAGLANGLIALRLQQQQPDMRILLIDAA------PQAGGNHTWSFHHDDLTESQHRWIAPLVVHHWPDYQVRFPTRRRKLNSGY
             100       110       120       130       140       150       160       170       180

Sy_CrtL  AQLQQHWLRQCEQGGLQWQLGKAAAIAHDSHHSCVTTAAGQELQARLVVDTTGHQAAFIQRPHSDAIAYQAAYGIIGQFSQPPIEPHQFV
Dc_CrtL  KQLKSKMMQKCISNGVKFHQAKVVKVVHEEAKSLLICDDGVTIQAAVVLDATGFSRCLVQYDKPYNPGYQVAYGIVAEVEEHPFDVNKMI
At_CrtL  KQLKSKMLQKCITNGVKFHQSKVTNVVHEEANSTVVCSDGVKIQASVVLDATGFSRCLVQYDKPYNPGYQVAYGIVAEVDGHPFDVDKMV
Ps_CrtL  YSITSERFAEALHQALGENIWLNCSVSEVLPNSVRLA-NGEALLAGAVIDGRGVT-----ASSAMQTGYQLFLGQQWRLT-QPHGLTVPI
Pa_CrtL  FCITSQRFAEVLQRQFGPHLWMDTAVAEVNAESVRLK-KGQVIGARAVIDGRGYA-----ANSALSVGFQAFIGQEWRLS-HPHGLSSPI
             190       200       210       220       230       240       250       260       270

Sy_CrtL  LMDYRSDHLSP----EE-RQLPPTFLYAMDLGNDVYFVEETSLAACPAIPYDRLKQRLYQRLATRGVTVQVIQHEEY--CLFPMNLPLPD
Dc_CrtL  FMDWRDSHLNGNTELKERNSKIPTFLYAMPFSSDRIFLEETSLVARPGLAMGDIQERMVARLRHLGIKVKSIEEDER--CVIPMGGPLPV
At_CrtL  FMDWRDKHLDSYPELKERNSKIPTFLYAMPFSSNRIFLEETSLVARPGLRMEDIQERMAARLKHLGINVKRIEEDER--CVIPMGGPLPV
Ps_CrtL  LMDATVAQQQG----------YR-FVYTLPLSADTLLIIEDTRYANVPQRDDNALRQTVTDYAHSKGWQLAQLEREETGCLPITLAGDIQA
Pa_CrtL  IMDATVDQQNG----------YR-FVYSLPLSPTRLLIEDTHYIDNATLDPECARQNICDYAAQQGWQLQTLLREEQGALPITLSGNADA
             280       290       300       310       320       330       340       350       360

Sy_CrtL  LTQSVVGFGGA---ASMVHPASGYMVGALLRRAPDLANAIAAGLNASSS--LTTAELATQAWRGLWPTEKIRKHYIYQFGLEKLMRFSEA
Dc_CrtL  LPQRVVGIGGT---AGMVHPSTGYMVARTLAAAPVVANAIVQYLGG-SKKGALGNELSAEVWKDLWPIERRRQREFFCFGMDILLKLDLP
At_CrtL  LPQRVVGIGGT---AGMVHPSTGYMVARTLAAAPIVANAIVRYLGSPSSNSLRGDQLSAEVWRDLWPIERRRQREFFCFGMDILLKLDID
Ps_CrtL  LWADAPGVPRSGMRAGLFHPTTGYSLPIAVALADAIADSPRLGSVP-------LYQLTRQFAERHWRR----QGFFRLLNRMLFLAGREE
Pa_CrtL  FWQQRP-LACSGLRAGLFHPTTGYSLPLAVAVADRLSALDVFTSAS------IHHAITHFARERWQQ----QGFFRMLNRMLFLAGPAD
             370       380       390       400       410       420       430       440       450

Sy_CrtL  QLNHHFQTFFGLPKEQWYGFLTNTLSLPELIQAMLRLFAQAPNDVRWGLMEQQGRELQLFWQAIAAR--
Dc_CrtL  GTRRFFSAFFDLEPRYWHGFLSSRLFLPELFFFGLSLFSNASNTSRIEIMAKGTVPLVNMVNNLIKDRE
At_CrtL  ATRRFFDAFFDLQPHYWHGFLSSRLFLPEILVFGLSLFSHASNTSRLEIMTKGTVPLAKMINNLVQDRD
Ps_CrtL  NRWRVMQRFYGLPEPTVERFYAGRLSLFDKARILTGKPPVPLGEAWRAAINHFPDRRDKG---------
Pa_CrtL  SRWRVMQRFYGLPEDLIARFYAGKLTLTDRLRILSGKPPVPVLAALQAIMTTHR-------------
             460       470       480       490       500       510
```

```
 ┌─ A. thaliana_CrtL_AAB53337
 │
 ┤  D. carota_LCYB_ABB52071
 │
 │  ┌─ P. stewartii_CrtY_AAA64980
 ┤  │
 │  └─ P. ananatis_CrtY_BAA14126
 │
 └──── S. elongatus_CrtL_ABB58092
```

Fig 6.1 The multiple sequence alignment and the homology tree among the lycopene cyclases from *S. elongatus, D. carota, A. thaliana, P. stewartii* and *P. ananatis*. The identical regions among the overall sequence alignment are indicated in yellow, whereas the identical regions between *S. elongatus* and the two plant lycopene cyclases, *S. elongatus* and the two *Pantoea* source lycopene cyclases are indicated in grey and blue, respectively.

The sequence similarity among the five lycopene cyclases, included in the multiple sequence alignment, are also reflected from the hydropathy plots which are indicated with colored boxes (**Fig 6.2**). The hydropathy similarities between the two plant cyclases and those between the two *Pantoea* lycopene cyclases are significant, which are indicated with blue and pink boxes, respectively. *S. elongatus* CrtL shows many hydropathy similarities to the two plant lycopene cyclases (indicated in green boxes), whereas the two plant lycopene cyclases exhibits less similarities to the *Pantoea* counterparts (indicated in red boxes).
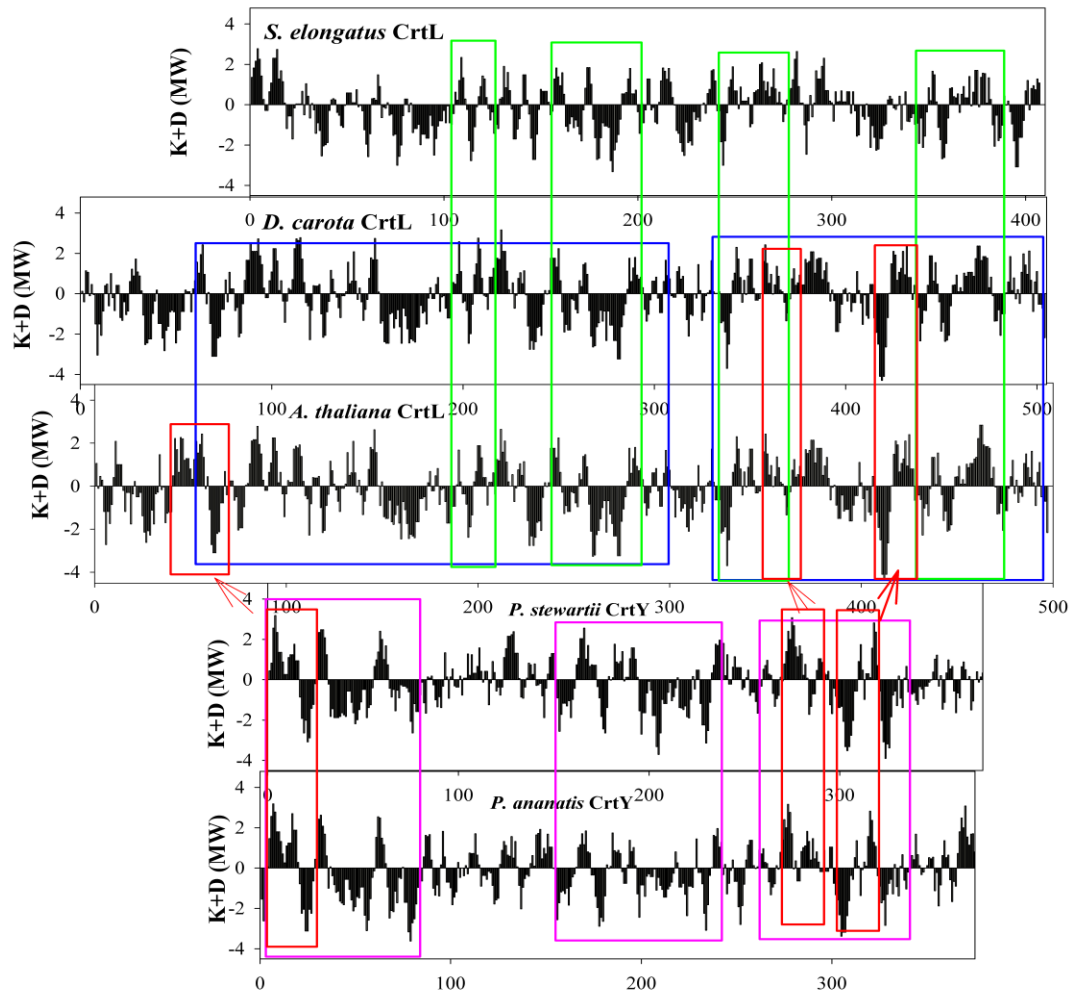
Fig 6.2 The hydropathy plot of lycopene cyclase from *S. elongatus*, *D. carota*, *A. thaliana*, *P. stewartii*, and *P. ananatis*. The similarity between the hydropathy plots of CrtLs from *D. carota* and *A. thaliana* is indicated by blue boxes whereas the pink boxes indicate the similarity between the CrtYs from *P. stewartii* and *P. ananatis*. The similar domains between *A. thaliana* CrtL and the two *Pantoea* CrtYs are indicated in red boxes whereas the similar domains among *A. thaliana* CrtL, *D. carota* CrtL and *S. elongatus* CrtL are indicated in green boxes.

6.4. The predicted secondary structure of *A. thaliana* CrtL

The secondary structure of *A. thaliana CrtL* was predicted by the Jpred software and is shown alongside the protein sequence data (**Fig 6.3**).

```
At_CrtL   MDTLLKTPNKLDFFIPQFHGFERLCSNNPYHSRVRLGVKKRAIKIVSSVVSGSAALLDLVPETKKENLDFELPLYDTSKSQVVDLAIVGG
                                                                                        ➤

              10        20        30        40        50        60        70        80        90

At_CrtL   GPAGLAVAQQVSEAGLSVCSIDPSPKLIWPNNYGVWVDEFEAMDLLDCLDTTWSGAVVYVDEGVKKDLSRPYGRVNRKQLKSKMLQKCIT
              ▭            ➤                                        ➤                   ▭

             100       110       120       130       140       150       160       170       180

At_CrtL   NGVKFHQSKVTNVVHEEANSTVVCSDGVKIQASVVLDATGFSRCLVQYDKPYNPGYQVAYGIVAEVDGHPFDVDKMVFMDWRDKHLDSYP
             ➤          ➤          ➤                      ➤                ➤

             190       200       210       220       230       240       250       260       270

At_CrtL   ELKERNSKIPTFLYAMPFSSNRIFLEETSLVARPGLRMEDIQERMAARLKHLGINVKRIEEDERCVIPMGGPLPVLPQRVVGIGGTAGMV
                 ➤          ➤           ▭                ➤              ➤

             280       290       300       310       320       330       340       350       360

At_CrtL   HPSTGYMVARTLAAAPIVANAIVRYLGSPSSNSLRGDQLSAEVWRDLWPIERRRQREFFCFGMDILLKLDLDATRRFFDAFFDLQPHYWH
             ▭                         ▭                   ▭       ▭

             370       380       390       400       410       420       430       440       450

At_CrtL   GFLSSRLFLPELLVFGLSLFSHASNTSRLEIMTKGTVPLAKMINNLVQDRD
             ▭     ▭       ▭       ▭

             460       470       480       490       500
```

Fig 6.3 The predicted secondary structure of the *A. thaliana* CrtL along its sequence. The predicted α-helices and β-sheets are indicated with cylinders and arrows, respectively.

In the following, we are going to explore the possible homologies of *A. thaliana* CrtL with the known blue light sensors, including LOV, BLUF, Crytochromes and OCP protein family.

6.5. Comparion of *A. thaliana* CrtL to the LOV protein family

Multiple sequence alignment was performed among the known LOV proteins and *A. thaliana* CrtL (EMBL AAB53337). The LOV proteins included in the alignment are as follows: phototropin phy3 from *Adiantum capillus-veneris* (EMBL BAA36192.2), LOV histidine kinase from *Caulobacter crescentus* (EMBL AAK22272.2), LOV STAS from *Listeria monocytogenes* (EMBL CAC98877.1), LOV phosphodiesterases from *Synechocystis* PCC 6803 (EMBL BAA10080.1), LOV circadian rhythm regulating protein FKF1 from *A. thaliana* (EMBL AF216523.2) (**Fig 6.4**).

The multiple sequence alignment shows that the residues 146-261 of *A. thaliana* CrtL share some sequence similarities with the LOV proteins. Interestingly, the conserved residues in LOV proteins

which are important for interacting with FMN (F968, Q970, V978, N998, N1008) can be found in *A. thaliana* CrtL, corresponding to F185, Q187, V208, Q227, Q237, respectively. The photoactive C966, which rotates and forms an adduct with the C4a of FMN upon blue light induction, is five amino acids upstream shifted in *A. thaliana* CrtL (indicated in red circle). The conserved salt bridge between E960 and K1001, which is important for the interaction between the LOV domain and its partner, is not conserved *A. thaliana* CrtL, because only K1001 can be found (corresponding to K230) and the E960 does not have a counterpart in *A. thaliana* CrtL (**Fig 6.4**).

Fig 6.4 Multiple sequence alignment and the homology tree among LOV domains and *A. thaliana* CrtL. The LOV sequences in this alignment include phototropin phy3 (residues 929-1032), LOV histidine kinase (residues 33-136), hypothetical LOV STAS (residues 19-121), LOV phosphodiesterases (residues 307-413), *A. thaliana* circadian rhythm regulating protein LOV FKF1 (residues 54-166), and *A. thaliana* CrtL (residues 146-261). The conserved key residues, which interact with the chromophore, are indicated with an arrow whereas the residues involved in forming the connecting structural pathway and the salt bridge, are indicated with one and two asterisks, respectively.

From the multiple sequence alignment we find that the small region (residues 146-247) of *A. thaliana* CrtL has some sequence similarities to the known LOV proteins (**Fig 6.4**), which is further investigated by the hydropathy comparison between *A. thaliana* CrtL and the LOV proteins. This region (residues 146-247) of *A. thaliana* CrtL, indicated with a grey bar (**Fig 6.5**), contains two small regions showing hydropathy similarities to the N- and C- terminus of the LOV domains, which are indicated in red boxes (**Fig 6.5**). In between, there are some variations.
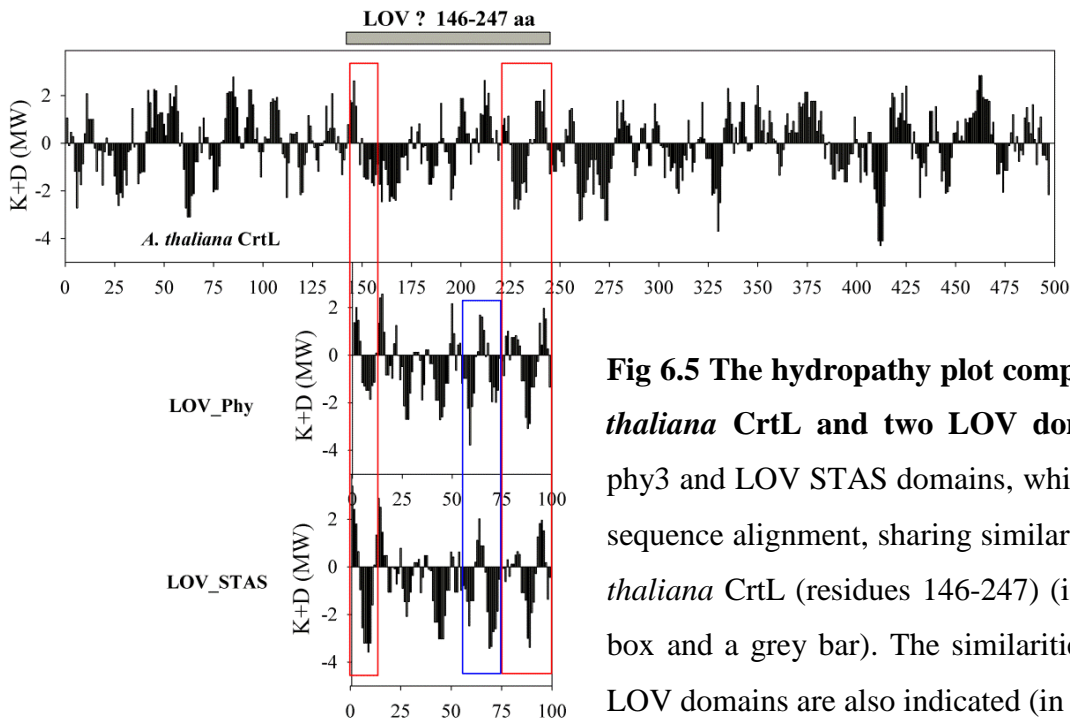
Fig 6.5 The hydropathy plot comparison between _A. thaliana_ CrtL and two LOV domains. Phototropin phy3 and LOV STAS domains, which were used in the sequence alignment, sharing similarity to a region of _A. thaliana_ CrtL (residues 146-247) (indicated with a red box and a grey bar). The similarities between the two LOV domains are also indicated (in blue boxes).

The secondary structure of _A. thaliana_ CrtL was predicted and shown in **Fig 6.6**. The residues 146-247 of _A. thaliana_ CrtL was predicted to be a domain composed of β4-α3-β5-β6-β7-β8 which corresponds to the β-β-α-α-α-α-β-β-β of the phy3 LOV domain (PDB: 1JNU). The secondary structure comparison is not exactly the same, but it might be that the large α3 helix of _A. thaliana_ CrtL functions similarly to the four small helices in the LOV domain.
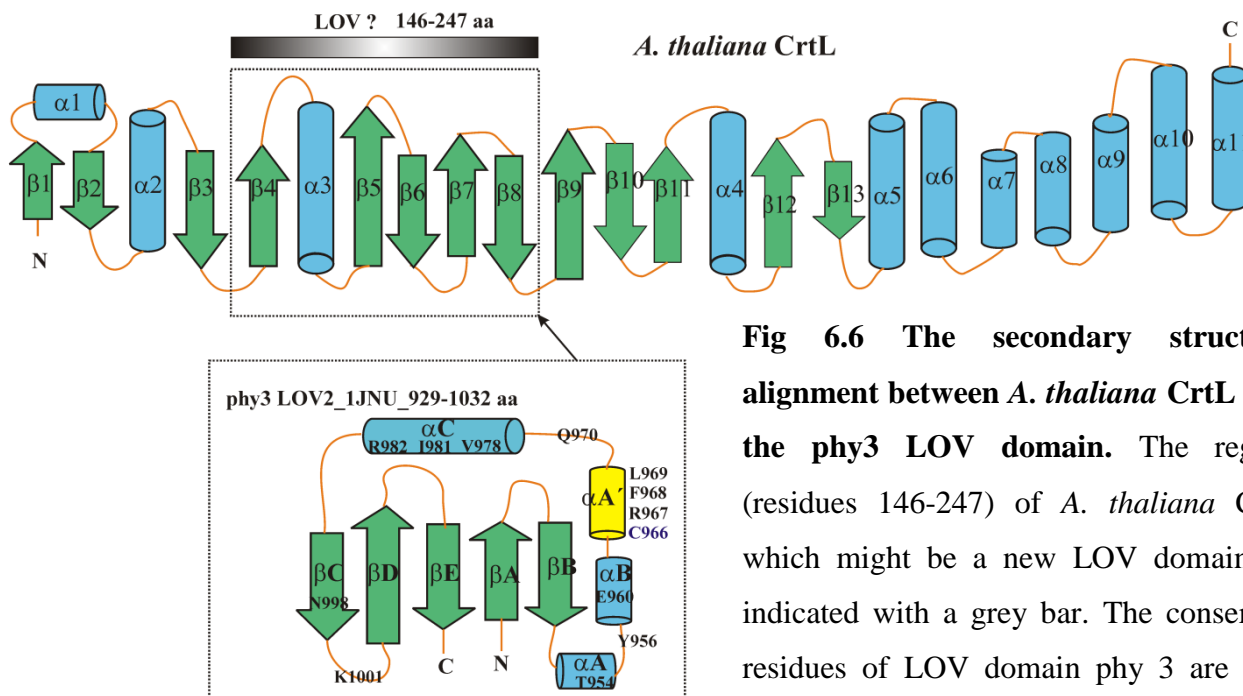
Fig 6.6 The secondary structure alignment between _A. thaliana_ CrtL and the phy3 LOV domain. The region (residues 146-247) of _A. thaliana_ CrtL which might be a new LOV domain, is indicated with a grey bar. The conserved residues of LOV domain phy 3 are also indicated.

157

In combination of the multiple sequence alignment, hydropathy comparison and secondary structure alignment, there would seem to be the possibility that the residues 146-247 of *A. thaliana* CrtL might be a new member of the LOV protein family.

6.6. Comparison of *A. thaliana* CrtL to the BLUF protein family

Multiple sequence alignment was performed among the known BLUF proteins and *A. thaliana* CrtL (EMBL AAB53337). The BLUF proteins included in the alignment are as follows: Tll0078 from *Thermosynechococcus elongatus* BP-1 (EMBL BAC07631.1), Slr1694 from *Synechocystis sp.* PCC6803 (EMBL BAA18389.1), AppA from *R. sphaeroides* (EMBL ABA77707.1). The multiple sequence alignment shows that there is not so much sequence similarity between *A. thaliana* CrtL and the BLUF proteins. Important residues in the BLUF domains, Y8/Q50 (responsible for hydrogen switching between residues and the keto group of FAD) and N32/H72/W91 (interacting with FAD upon blue light illumination) are not conserved in the *A. thaliana* CrtL (**Fig 6.7**).
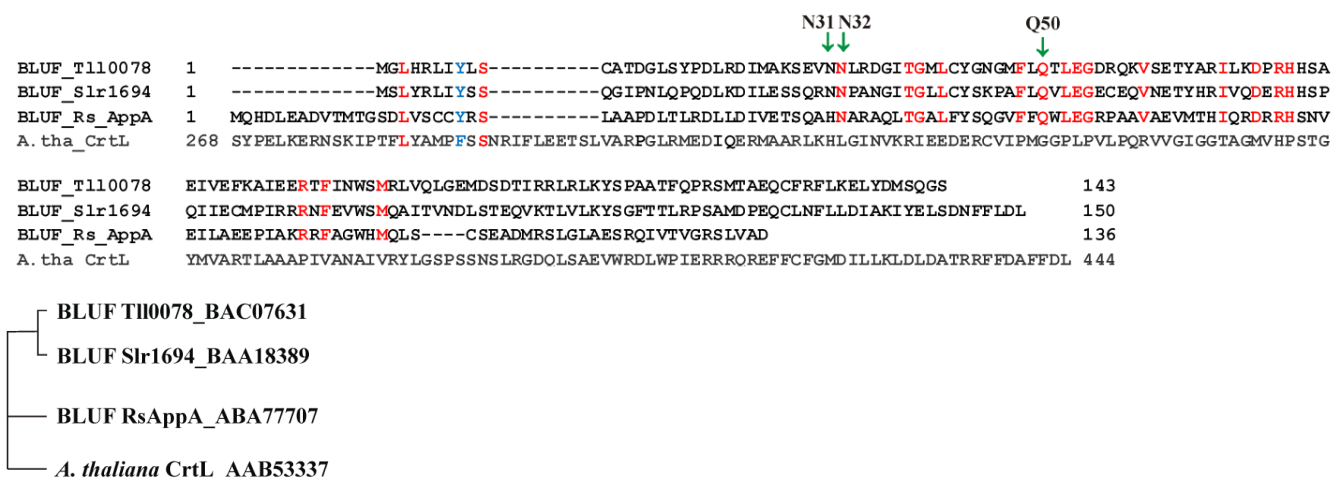
Fig 6.7 Multiple sequence alignment and the homology tree among BLUF proteins and *A. thaliana* CrtL. The BLUF proteins in this alignment include Tll0078 (residues 1-143), Slr1694 (residues 1-150), AppA (residues 1-136) and *A. thaliana* CrtL (residues 268-444). Important residues (N31, N32 and Q50) are indicated with a green arrow.

Alhough the multiple sequence alignment did not show many similarities between *A. thaliana* CrtL and BLUF proteins, the hydropathy comparison in contrast shows that *A. thaliana* CrtL (residues 225-308) shares similarity to Slr1694 (residues 1-150), indicated with a blue box, whereas the similarities between Slr1694 (residues 1-150) and AppA (residues 1-136) are indicated with red boxes (**Fig 6.8**).
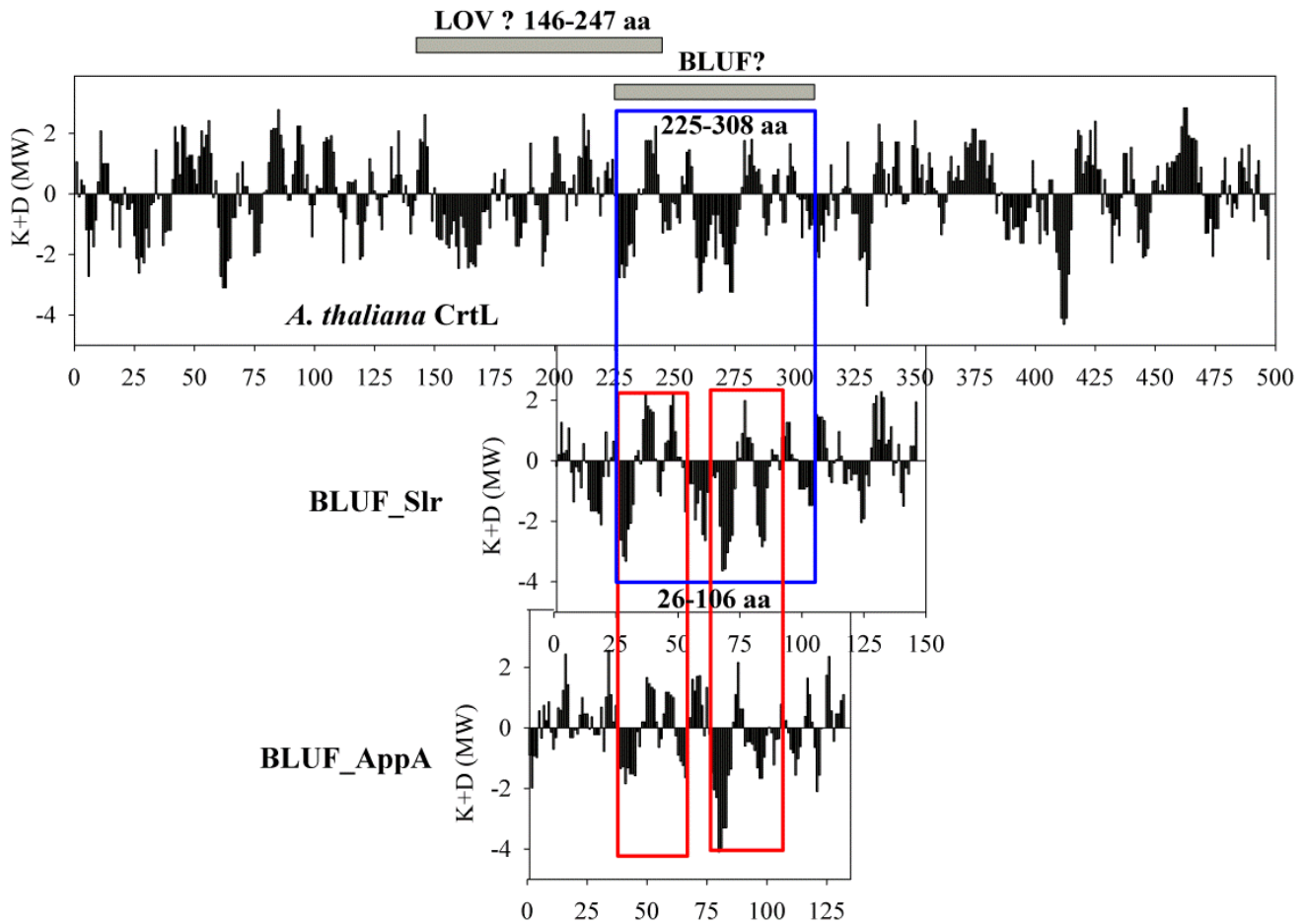
Fig 6.8 The hydropathy plot comparison between *A. thaliana* CrtL and two BLUF domains. Slr1694 (residues 1-150) and AppA (residues 1-136) domains, which were used in the sequence alignment, show hydropathy similarities to *A. thaliana* CrtL (residues 225-308). Residues 146-247 and residues 225-308 of *A. thaliana* CrtL are indicated with grey bars.

The residues 225-308 of *A. thaliana* CrtL is predicted to be composed of β8-β9-β10-β11-α4, which corresponds to the Slr1694 BLUF domain composed of β1α1β2β3α2β4β5 (**Fig 6.9**). *A. thaliana* CrtL might contain one new LOV domain (residues 146-247) and one new BLUF domain (residues 225-308), and these two regions are overlapping. However, until now the BLUF domains have been found only in bacteria.
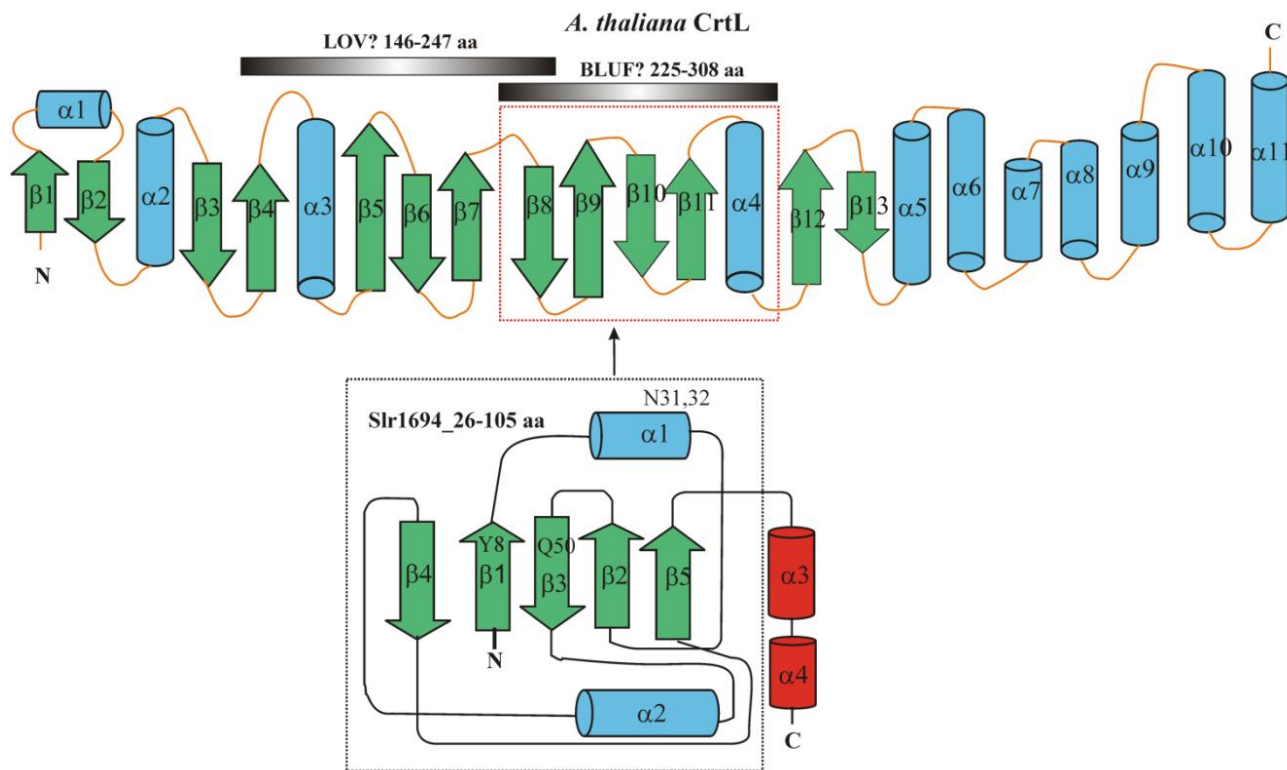
Fig 6.9 The secondary structure alignment between *A. thaliana* **CrtL and the Slr1694 domain (residues 26-105).** The regions, (residues 146-247) and (residues 225-308) of *A. thaliana* CrtL, are indicated with grey bars which might function as a new LOV domain and a new BLUF domain, respectively.

6.7. Comparison of *A. thaliana* CrtL to the cryptochrome protein family

Sequences used in the alignment include the Cry1 C-terminus from *A. thaliana* (EMBL AAB28724.1), the Cry2 C-terminus from *A. thaliana* (EMBL AAB04996.1), and the Cry3 (Cry-DASH) from *A. thaliana* (EMBL AF069716). The multiple sequence alignment between *A. thaliana* CrtL and cryptochromes did not lead to a good alignment. For better alignment, the DAS motif, which exists in the N-terminus (Cry3) or C-terminus (Cry1 and Cry2) of cryptochromes, was separated from the rest of the sequence. The multiple sequence alignment between *A. thaliana* CrtL and cryptochromes without DAS domains was performed, which shows a convincing sequence homology along the overall alignment, highlighted in yellow (**Fig 6.10**). It is important to mention that the D422 (conserved residue interacting with FAD) and W432 (conserved residue in electron transfer upon blue light induction) in cryptochromes are conserved in *A. thaliana* CrtL which correspond to D252 and W261, respectively (**Fig 6.10**). Partial sequence homology among cryptochromes, and that between cryptochromes and *A. thaliana* CrtL, are indicated in pink and blue, respectively.

160

```
Cry1noDAS   MSGSVSGCGSGGCSIVWFRRDLRVEDNPALAAAVRAG-PVIALFVWAP--EEEGHYH--- 54
Cry2noDAS   MKMDKK-------TIVWFRRDLRIEDNPALAAAAHEG-SVFPVFIWCP--EEEGQFY--- 47
Cry3noDAS   -----------VTTIWFRNDLRVLDNDALYKAWSSSDTILPVYCLDPRLFHTTHFFNFP 48
AtCrtL_1-294 ----------------------MDTLLKTPNKLDFFIPQFHGFERLCSNNPYH--- 31

Cry1noDAS   -PGRVSRWWLKNSLAQLDSSLRSLGTCLITKRSTDSVASLLDVVKSTGASQIFFNHLYDP 113
Cry2noDAS   -PGRASRWWMKQSLAHLSQSLKALGSDLTLIKTHNTISAILDCIRVTGATKVVFNHLYDP 106
Cry3noDAS   KTGALRGGFLMECLVDLRKNIMKRGLNLLIR-SGKPEEILPSIAKDFGARTVFAHKETCS 107
AtCrtL_1-294 --------------SRVRLGVKKRAIKIVSSVVSGSAALLDLVPETKKENLDFELPLYDT 77

Cry1noDAS   LSIVRDHRAKDVLTAQG--IAVRSFNADLLYEPWEVTDELGRPFSMFAAFWERCLSMPYD 171
Cry2noDAS   VSLVRDHTVKEKLVERG--ISVQSYNGDLLYEPWEIYCEKGKPFTSFNSYWKKCLDMSIE 164
Cry3noDAS   EEVDVERLVNQGLKRVGNSTKLELIWGSTMYHKDDLPFDVFDLPDVYTQFRKSVEAKCSI 167
AtCrtL_1-294 SKSQVVDLAIVGGGPAG------------------------LAVAQQVSEAGLSVCSI 111

Cry1noDAS   ----PESPLLPPKKIISGDVSKCVADPLVFEDDSEKGSNALLARAWSPGWSNGDKALTTF 227
Cry2noDAS   SVMLPPPWRLMPITAAAEAIWACSIEELGLENEAEKPSNALLTRAWSPGWSNADKLLNEF 224
Cry3noDAS   RSSTRIPLSLGPTPSVDDWGDVPTLEKLGVEPQEVTRGMRFVGGESAG----VGRVFEYF 223
AtCrtL_1-294 D---PSPKLIWPN----------NYGVWVDEFEAMDLLDCLDTTWSG----AVVYVDEG 153

Cry1noDAS   INGPLLEYSKNRRK--ADSATTSFLSPHLHFGEVSVRKVFHLVRIKQVAWANEGNEAGEE 285
Cry2noDAS   IEKQLIDYAKNSKK--VVGNSTSLLSPYLHFGEISVRHVFQCARMKQIIWARDKNSEGEE 282
Cry3noDAS   WKKDLLKVYKETRNGMLGPDYSTKFSPWLAFGCISPRFIYEEVQR------YEKERVANN 277
AtCrtL_1-294 VKKDLSRPYGRVNR-------KQLKSKMLQKCITNGVKFHQSKVT--------NVVHEE 197

Cry1noDAS   SVNLFLKSIGLREYS------------------------------------------- 300
Cry2noDAS   SADLFLRGIGLREYSRYICFNFPFTHEQSLLSHLRFFPWDADVDKFKAWRQGRTGYPLVD 342
Cry3noDAS   STYWVLFELIWRDYFRFLSIKCGNSLFHLGGPRNVQGKWSQDQKLFESWRDAKTGYPLID 337
AtCrtL_1-294 AN------------------------------------------------------- 199

Cry1noDAS   ---------------IRVVVSSFFVKVLQLPWRWGMKYFWDTLLDADLESDALGWQYITG 345
Cry2noDAS   AGMRELWATGWMHNRIRVIVSSFAVKFLLLPWKWGMKYFWDTLLDADLECDILGWQYISG 402
Cry3noDAS   ANMKELSTTGFMSNRGRQIVCSFLVRDMGLDWRMGAEWFETCLLDYDPCSNYGNWTYGAG 397
AtCrtL_1-294 ---------------STVVCSDGVKIQAS-VVLDATGFSRCLVQYDKPYNPG-YQVAYG 241

Cry1noDAS   TLPDSREFDRIDNPQFEGYKFDPNGEYVRRWLPELSRLPTDWIHHPWNAPESVLQAAGIE 405
Cry2noDAS   SIPDGHELDRLDNPALQGAKYDPEGEYIRQWLPELARLPTEWIHHPWDAPLTVLKASGVE 462
Cry3noDAS   VGNDPREDRYFSIP-KQAQNYDPEGEYVAFWLQQLRRLPKEKRHWPG----RLMYMDTVV 452
AtCrtL_1-294 IVAEVD----------GHPFDVDKMVFMDWRDKHLDSYP----------ELKERNSKI 279

Cry1noDAS   LGSNYPLPIVGIDEAKARLHEALSQMWQLEAAS 438
Cry2noDAS   LGTNYAKPIVDIDTARELLAKAISRTREAQIMI 495
Cry3noDAS   PLKHGNGPMAGGSKSGGGFRGSHSGRRSRHNGP 485
AtCrtL_1-294 PTFLYAMPFSSNRIF----------------- 294
```

```
 ┌─ Cry1_AAB28724
 │
 │  └─ Cry2_AAB04996
 │
 ├──── Cry3_AF069716
 │
 └──── A. thaliana CrtL_AAB53337
```
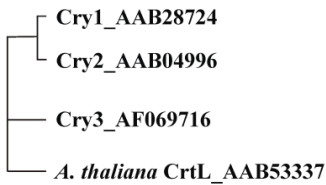
Fig 6.10 The multiple sequence alignment and the homology tree among the known cryptochromes (without DAS domain) and *A. thaliana* CrtL. Sequences used in this alignment include *A. thaliana* Cry1 (residues 1-438), *A. thaliana* Cry2 (residues 1-495), *A. thaliana* Cry3 (Cry DASH) (residues 1-485) and *A. thaliana* CrtL (residues 1-294).

A multiple sequence alignment between *A. thaliana* CrtL and the DAS domains of cryptochromes was also performed, shown in (**Fig 6.11**). The DAS domain, which is composed of three motifs (D-motif, A-motif and S-motif) and important for signaling transduction in all three types of cryptochromes, is conserved nearly perfectly in *A. thaliana* CrtL. This indicates that *A. thaliana* CrtL might employ similar signal transduction as cryptochromes.

```
                                                           D-motif                A-motif
At-cry1 499 RAAIENGSEEGLGDSAEVEEAPIEFPRDITMEETEPTRLNPNRRYEDQM--VESIT----SSLIRPEEDEES
At-cry2 496 GAAPD----------EIVADSFEALGANTIKEPG---LCPSVSSNDQQ--VESAVRYNGSKRVKPEEEEE-
At-cry3 45  -------------------------------------------NDHIHRVE---------ALTEEEIDS
At_CrtL 295 ---------------------------LEET-SLVARPGLRMEDIQERMAARLKHLGINVKRIEEDERC
                                                                          S
                                                                        motif
At-cry1 SLNLRNSVGDSRAEVPRNMVNTNQAQQRRAEPASNQVTAMIPEFNIRIVAESTEDSTAE--SSSSGRRERSG 634
At-cry2 --------------RDMKKS----------------------RGFDERELFSTAESSSSSS----- 575
At-cry3 -----------------------------------------VAIKTFERYALP-SSSSVKRKGKG 84
At_CrtL -V------PMGGPLPVLPQRVVGIGGTAGMVHPSTGYMVARTLAAAPIVANAIVRYLGSP-SSNSLRG---- 396
```

Cry3_AF069716

Cry1_AAB28724

Cry2_AAB04996

A. thaliana CrtL_AAB53337

Fig 6.11 The multiple sequence alignment and the homology tree among *A. thaliana* CrtL and the DAS motifs conserved in cryptochromes. Sequences used in this alignment include Cry1 C-terminus (residues 499-634), Cry2 C-terminus (residues 496-575), Cry3 (Cry DASH) N-terminus (residues 45-84), and *A. thaliana* CrtL (residues 295-396). The D-, A- and S-motif of the DAS domain are indicated with green, blue and red dotted boxes, respectively.

The hydropathy comparison shows that *A. thaliana* CrtL contains three regions sharing convincing hydropathy similarities to the three types of cryptochromes, the residues 90-125 (in light blue box), 262-300 (in pink box) and 295-396 (DAS motif, composed of three elements indicated in green, dark blue and red boxes) (**Fig 6.12**).
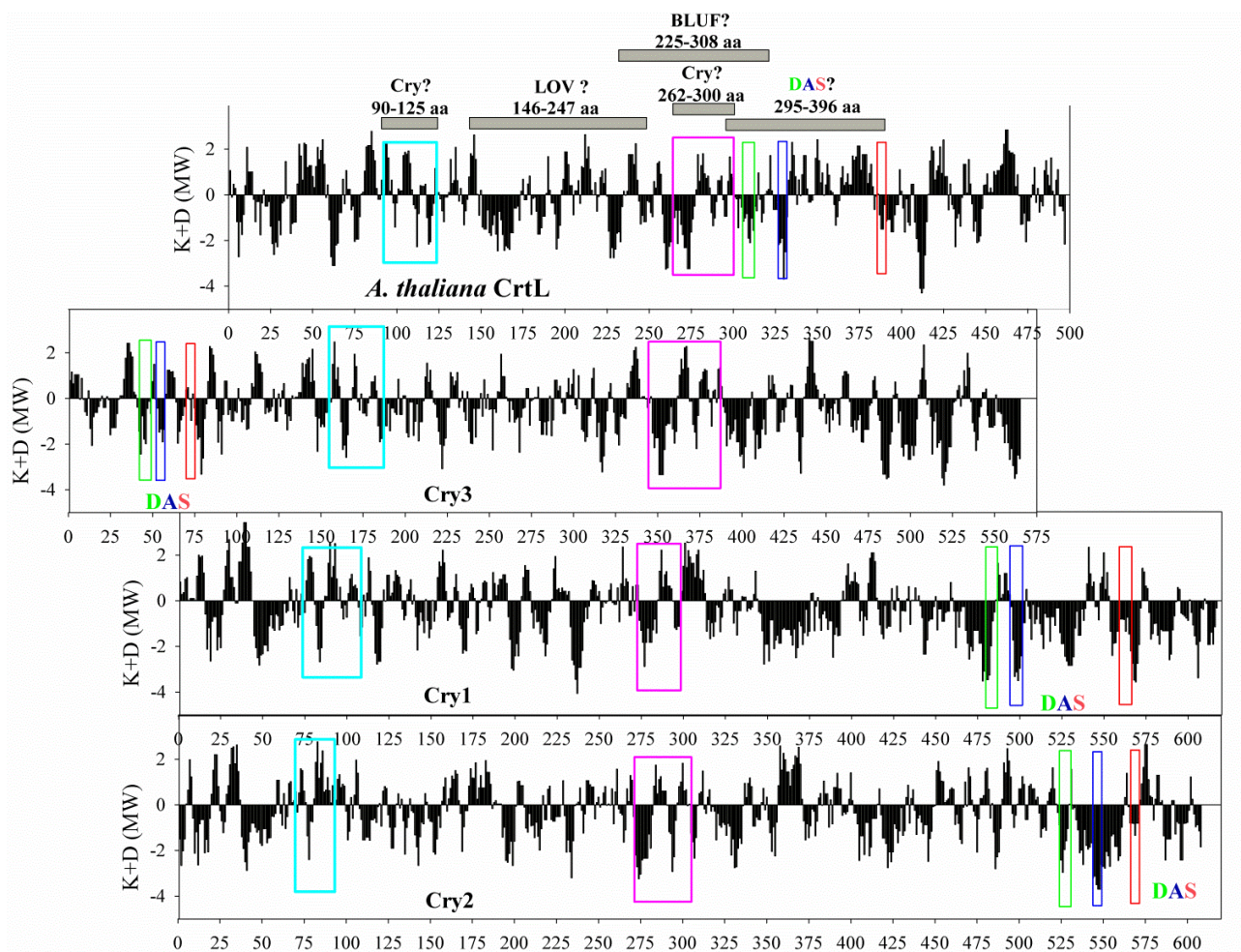
Fig 6.12 The hydropathy plot comparison between *A. thaliana* CrtL and three cryptochromes. Two similar regions among the overall sequences are indicated in cyan and pink boxes. The DAS motif is indicated in all the sequences of cryptochromes and *A. thaliana* CrtL (residues 295-396) with three colored boxes (green for D-motif, blue for A-motif and red for S-motif). The residues 90-125 and 262-300, 146-247, 225-308 of *A. thaliana* CrtL, which might function as a new Cry domain, LOV domain and BLUF domain, respectively, are indicated with grey bars as well.

The predicted secondary structure of *A. thaliana* CrtL is shown in comparison with that of cryptochrome 3 (PDB: 2J4D) (**Fig 6.13**). The three regions, revealed by hydropathy plot comparison, are indicated with grey bars, including the DAS motif (residues 295-396), the residues 90-125 (α2β3) and the residues 262-300 (β9β10β11). Conserved residues (D252, W261) are also indicated. The residues 90-125 (α2β3) of *A. thaliana* CrtL corresponds to the residues 157-189 of Cry3 (α6β5α7α8), whereas the residues 262-300 (β9β10β11) of *A. thaliana* CrtL corresponds to the residues 340-390 of

163

the Cry3 (α16α17α18). Interestingly, the residues 262-300 and the putative DAS domain (residues 295-396) overlap with the putative "new BLUF element" (residues 225-308).
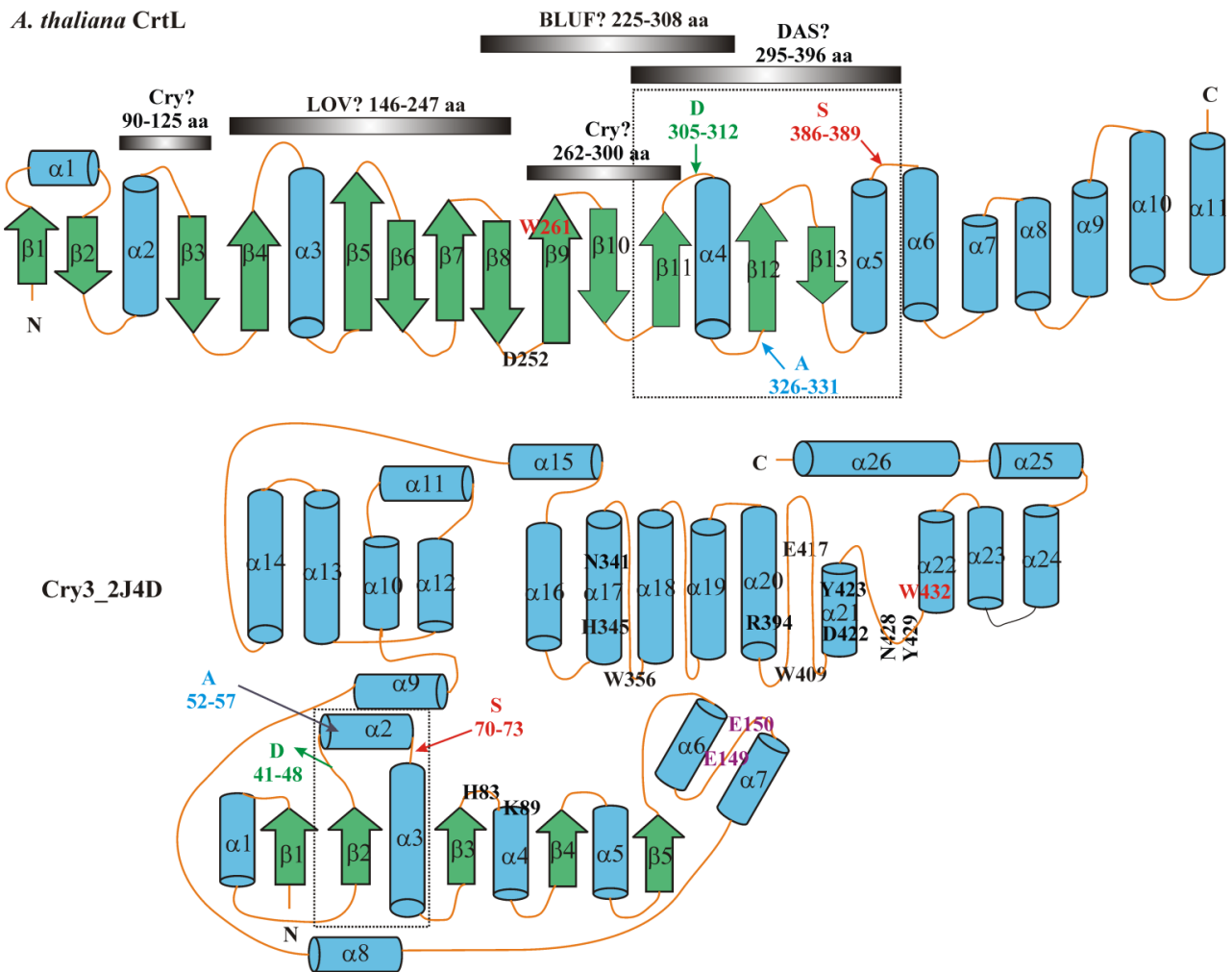
Fig 6.13 The secondary structure alignment between *A. thaliana* CrtL and the Cry3 protein. The regions of *A. thaliana* CrtL similar to cryptochromes (residues 90-125, 262-300 and 295-396), and those similar to BLUF (residues 225-308) and LOV (residues 146-247) are all indicated with grey bars. The key residues of Cry3 are also indicated.

6.8. Comparison of *A. thaliana* CrtL to OCP

Sequence alignment was performed between the OCP protein from *Synechocystis sp.* PCC 6803 (EMBL BAA18188.1) and *A. thaliana* CrtL (EMBL AAB53337), which shows convincing sequence homology along the overall alignment (**Fig 6.14**).

```
At_CrtL  MDTLLKTPNKLDFFIPQFHGFERLCSNNPYHSRVRLGVKKRAIKIVSSVVSGSAALLDLVPETKKENLDFELPLYDTSKS
OCP      ------------------------------------------------------------MPFTIDSARGIFPNTLA
              10        20        30        40        50        60        70        80

At_CrtL  QVVDLAIVGGGPAGLAVAQQVSEAGLSVCSIDPSPKLIWPNNYGVWVDEFEAMDLLDCLDTTWSGAVVYVDEGVKKDLSR
OCP      ADVVPATIAR-FSQLNAEDQLALIWFAYLEMGKTLTIAAPGAASMQLAENALKEIQAMGPLQQTQAMCDLANRADTPLCR
              90       100       110       120       130       140       150       160

At_CrtL  PYGRVNRKQLKSKMLQKCITNGVKFHQSKVTNVVHEEANSTVVCSDGVKIQASVVLDATGFSRCLVQYDKPYNPGYQVAY
OCP      TYASWSPN-IKLGFWYRLGEIMEQGFVAPIPAGYQLSANANAVLATIQGLES-------------------GQQITV
              170       180       190       200       210       220       230       240

At_CrtL  GIVAEVDGHPFDVDKMVFMDWRDKHLDSYPELKERNSKIPTFLYAMPFSSNRIFLEETSLVARPGLRMEDIQERMAARLK
OCP      LRNAVVD-------MGFTAGKDGKRIAEPVVPPQDTASRTKVSIEGVTNATVLNYMDNLNANDFDTLIELFTSDGALQP
              250       260       270       280       290       300       310       320

At_CrtL  HLGINVKRIEEDERCVIPMGGPLPVLPQRVVGIGGTAGMVHPSTGYMVARTLAAAPIVANAIVRYLGSPSSNSLRGDQLS
OCP      PFQRPIVGKENVLRFFREECQNLKLIPERGVTEPAEDGFTQIKVTGKVQTPWFGGNVGMNIAWRFLLNPE----------
              330       340       350       360       370       380       390       400

At_CrtL  AEVWRDLWPIERRRQREFFCFGMDILLKLDLDATRRFFDAFFDLQPHYWHGFLSSRLFLPELLVFGLSLFSHASNTSRLE
OCP      --------------------GKIFFVAIDLLASPKELLNFAR
              410       420       430       440       450       460       470       480

At_CrtL  IMTKGTVPLAKMINNLVQDRD
OCP
              490       500
```

Fig 6.14 Sequence alignment between OCP and *A. thaliana* CrtL. The important residues interacting with the carotenoid in the OCP are indicated with red asterisks, including Y44, W110, R155, E244, Y201, and W288.

The hydropathy comparison between *A. thaliana* CrtL and OCP indicates there are two domains in the *A. thaliana* CrtL sharing similarities to OCP, which are the residues 215-337 and 425-483, corresponding to the residues 146-263 and 275-317 of OCP, indicated with a red box and blue circle, respectively (**Fig 6.15**).
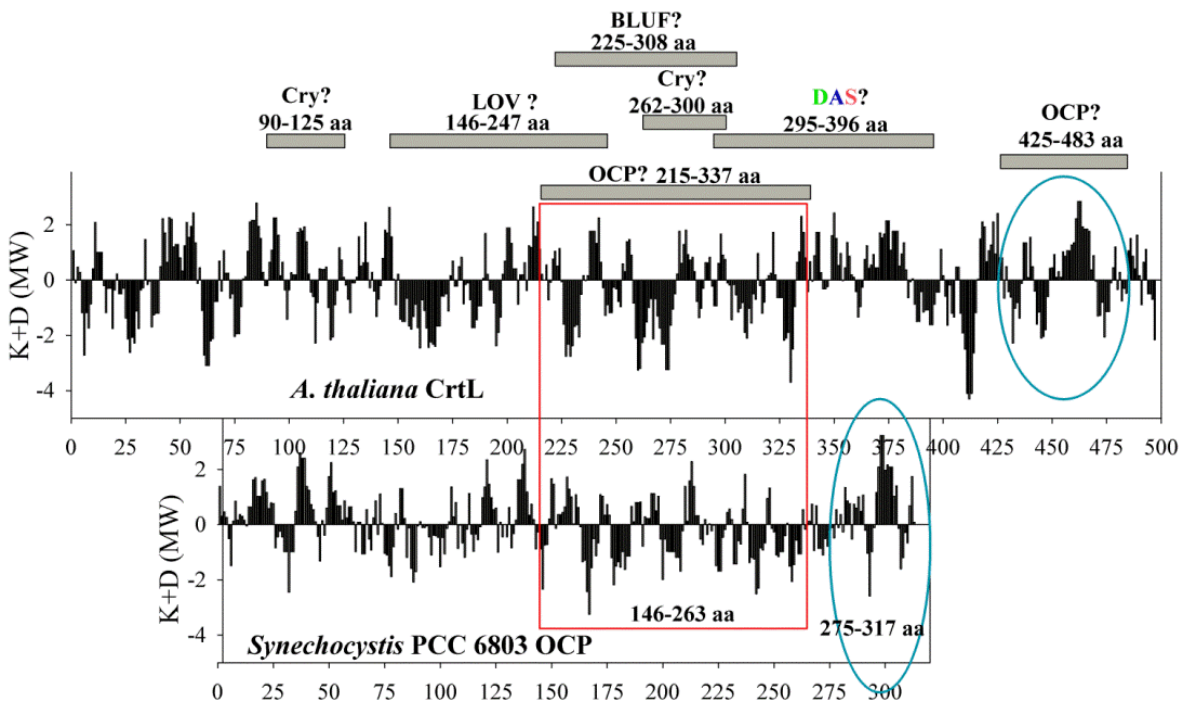
Fig 6.15 The hydropathy comparison between OCP and the *A. thaliana* CrtL. The residues 215-337 and 425-483 which show similarity to OCP, together with those similar to cryptochromes (residues 90-125, 262-300 and 295-396), and those similar to BLUF (residues 225-308) and LOV (residues 146-247) are indicated with grey bars.

The predicted secondary structure of *A. thaliana* CrtL is shown in comparison with that of OCP (PDB: 3MG1). The two domains in the *A. thaliana* CrtL (residues 215-337 and 425-483), which share similarity to those of OCP according to the hydropathy comparison, correspond to the residues 146-263 and 275-317 in OCP which are involved in the interaction with the carotenoid and catalytic function, respectively (Wilson *et al.*, 2011) (**Fig 6.15**).
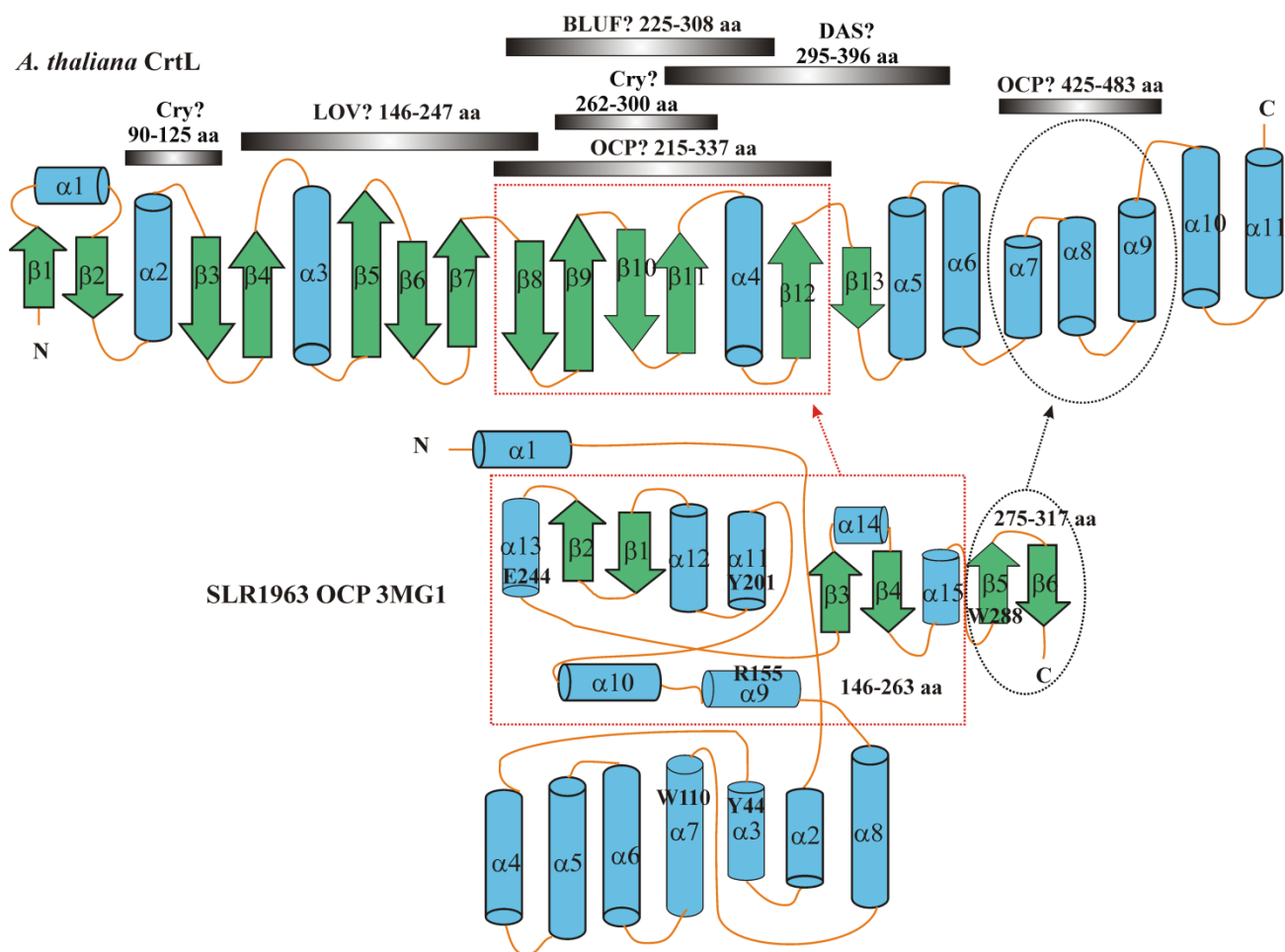
Fig 6.16 The comparison between the predicted secondary structure of *A. thaliana* CrtL and OCP Slr1963. The residues 215-337 and 425-483 which show similarity to OCP, together with those similar to cryptochromes (residues 90-125, 262-300, and 295-396), and those similar to BLUF (residues 225-308) and LOV (residues 146-247) are indicated with grey bars.

6.9. Discussion

Based upon the sequence alignment, hydropathy comparison and secondary structure alignment, it seems that *A. thaliana* CrtL might contain two major domains, one domain (residues 146-396) might contain several overlapping elements and function as the blue light sensor domain sharing similarities to LOV (residues 146-247 aa), BLUF (residues 225-308), Cry (residues 262-300 and 295-396) and OCP (residues 215-337). The other domain (residues 425-483) shares similarity to the catalytic domain of OCP (residues 275-317) and might function like the OCP in a similar mechanism. The residues 295-396 of *A. thaliana* CrtL might function as a connector between the two domains and might be responsible for the signal transduction, since *A. thaliana* CrtL contains a conserved DAS motif which is responsible for the signal transduction in all three types of cryptochromes. This will help to explain the SWGK46Y to SWGK46G_R reversible transition.

We hypothesize that the lycopene β-cyclase may contain a blue light sensing domain in addition to the catalytic domain for the enzyme. If the blue light sensing domain is similar to an OCP-type protein, then carotenoid (either lycopene or β-carotene) may undergo blue light-induced configurational changes. This in turn would cause the conformation of the sensing domain to be changed, which may interact with early crt biosynthesis enzymes (probably CrtIB) via the putative DAS motif, resulting in the down-regulation of carotenoid production in the SWGK46G_R. When the blue light is removed, the configuration of the carotenoid in the *A. thaliana* CrtL changes back and lead to the reversible transition from SWGK46G_R to SWGK46Y.

I emphasize here that those assignments in this chapter are highly tentative. In particular, it is very unlikely that a collection of blue-light sensing domains are distributed through all the protein sequence. However, the homologies between the cryptochrome DAS sequences, as well as that of OCP seem to be convincing. Further examination of the structures of the domains will require their separation (initial design will be based upon the above bioinformatics analysis) by molecular biological methods and the insertion of the domains into bacterial or plant hosts.

Chapter 7 Discussion and perspectives for further research

The major theme of my PhD work has been to establish *R. rubrum* as an organism for biotechnological application, in particular, the production of carotenoids at high levels and low cost. Carotenoids of high nutrient and medical values, for example, lycopene, β-carotene and astaxanthin are not the wild-type carotenoid in *R. rubrum*. Therefore engineering of the carotenoid biosynthetic pathway is required for the purpose of unusual carotenoid production using molecular biology.

In contrast to the *in trans* expression strategies employed in many other studies reported so far, the strategy of lycopene expression in *R. rubrum* was to insert a kanamycin resistance cassette into the *crtCD* locus in the chromosomal DNA to arrest the carotenoid biosynthesis pathway before lycopene is processed by CrtC and CrtD, giving lycopene as the major product. The lycopene-producing mutant, SLYC18, has more O_2 sensitivity and a longer lag phase compared to the wild-type S1. However, beyond the lag phase SLYC18 still shows nearly the same exponential growth rate as S1. In the stationary phase, SLYC18 can achieve the same cell density as S1, and express lycopene at high levels of **2 mg/g DW** growing chemoheterotrophically in the dark with M2SF$^+$ medium. This is the first time to report that *R. rubrum* can produce lycopene exclusively by biosynthetic pathway engineering. The strategy worked well and is the starting point for developing *R. rubrum* as an organism for unusual carotenoid expression because it proved that the *crtCD* gene locus is dispensible and can be replaced with other heterologous genes. A major result was the demonstration that the carotenoid lycopene, which is normally produced only in very small quantities during the carotenoid biosynthesis pathway, is still incorporated into the LH1 complexes at levels equivalent to those of Spx. This is particularly interesting, since lycopene, as a unusual carotenoid which is normally produced only at very low levels under physiological conditions, is almost impossible to incorporate into the LH1 complexes by *in vitro* reconstitution (Fiedor *et al.*, 2004; Kakitani *et al*., 2007).

From lycopene to β-carotene, only one step catalyzed by lycopene β cyclase is needed for β-carotene expression in *R. rubrum*. Therefore, using the same strategy as employed for the construction of SLYC18, a module containing the *A. thaliana* CrtL, as well as a synthetic rbs and a *npt* selection cassette, were inserted into the chromosomal DNA of *R. rubrum* at the same locus (*crtCD*). This strategy proved to be successful. SWGK46 indeed produced almost exclusively β-carotene. From the observation of the uniquely shifted LH1 Q_y maximum at 877 nm, it would also appear that the β-carotene can be bound by the LH1 complexes. However, SWGK46Y showed more sensitivity to O_2,

indicated by the longer lag phase compared to SLYC18 and a diagnostic peak at 420 nm due to the presence of a BChla precursor (PPIX) or a derivative thereof (e.g. PPIX-monomethylester). This observation is consistent with the observations in other O_2-sensitive mutants (C. Autenrieth and R. Ghosh, unpublished). However, SWGK46Y still showed the same exponential growth rate and achieved the same cell density at the stationary phase compared to SLYC18. β-carotene production at high levels of **4.4 mg/g DW** was obtained with M2SF medium at 500 ml scale. It was proved that the strategy worked well and it was successful compared to the only one documented study (Hunter *et al.*, 1994) on β-carotene expression in a photosynthetic bacterium, *Rb. sphaeroides*, which was not able to measure the growth rate of the photosynthetic culture and the level of the LH1 complexes was very low. SWGK46 also showed a **decoupling** of carotenoid from BChla biosynthesis, so that the amounts of β-carotene produced exceeded the amount necessary for the incorporation into the LH1. This is the first time that decoupling has been observed.

With the β-carotene producer, SWGK46, several unusual interesting phenomena were observed.

(1) **The O_2 effect**. The SWGK46B phenotype was selected after several passages, probably due to the accumulation of a viable subpopulation of secondary mutants. The repressed levels of carotenoid and ICM as well as the diagnostic peak at 420 nm are characteristic of SWGK46B, which probably indicates that a global down-regulation of photosynthesis genes was induced in SWGK46B by oxidative stress.

(2) **Unique forms of the LH1 complexes**. We also observed that the LH1 complexes of SWGK46Y and SWGK46B were in different forms indicated by the Q_y absorption at 877 nm and 875 nm, respectively. As mentioned before in Chapter 3 and Chapter 5, the LH1 complexes with and without bound carotenoid documented in the literature so far, have Q_y absorption maxima at 882 nm and 873 nm, respectively. *In vivo*, an unusual carotenoid, lycopene, which normally only accumulates in very small amount during carotenoid biosynthesis, also shows a Q_y absorption maximum at 882 nm (Wang *et al.*, 2012). However, the LH1 complexes of SWGK46Y show a unique Q_y absorption maximum at 877 nm. Therefore, we assume that β-carotene is incorporated into the LH1 complexes in SWGK46Y, and that the Q_y absorption at 877 nm could be due to different pigment-protein interaction and some conformational changes upon the incorporation of β-carotene into the LH1 complexes. This needs to be proved further experimentally by isolation of the pure LH1 complexes. This was consistent with the study of Hunter *et al.* (1994), using *Rb. sphaeroides* as a host organism, where it was shown that β-carotene incorporated successfully in the LH1 complexes and suggested that some conformational

changes might be caused due to β-carotene. In this latter study, however, the amount of carotenoid produced was extremely low and not quantified, and the LH1 Q_y peak could not be distinguished clearly from that of the LH2 complexes (also present in very low amounts). The Q_y absorption maximum of SWGK46B showed at 875 nm, which might be a mixture of LH1 complexes with Q_y at 877 nm and LH1 complexes with Q_y at 873 nm. This might indicate that the LH1 complexes in SWGK46B are mutated in such a way to exclude the carotenoid from the LH1 complexes for survival under higher oxidative stress.

(3) **A grey/green phenotype SWGK46G$_R$ arose after SWGK46 was grown under photoheterotrophic conditions for many passages**. The Q_y absorption maximum of SWGK46G$_R$ showed at 873 nm, which is characteristic of LH1 complexes without bound carotenoid. The total ICM levels were equivalent to those of the wild-type whereas the carotenoid level was down-regulated to a very low level. This appearance of the SWGK46G$_R$ phenotype was more rapid in the presence of blue-green light. We hypothesize that the SWGK46G$_R$ has a regulatory mechanism only to down-regulate the expression level of carotenoid in response to blue light with higher intensity, which could be recovered in the dark. This was confirmed by SWGK46Y photoheterotrophic growth experiments, which were performed by growing SWGK46Y under red as well as infrared light (625 nm and 850 nm, respectively) without blue light. Under these conditions, SWGK46Y remained yellow in color (see Chapter 5). It is tempting here to hypothesize that the light effector of the SWGK46G$_R$ strain is the *A. thaliana* CrtL, since this effect has never been observed in any other *R. rubrum* mutants so far either in the Ghosh lab or in the literature. The possibility that CrtL had been genetically modified (mutations or rearrangement) was excluded since the CrtL gene sequence of the SWGK46G$_R$ is identical to that of the wild-type CrtL sequence.

The carotenoid down-regulation mechanisms in SWGK46B and SWGK46G$_R$ seemed to be different. In SWGK46B, the depressed levels of ICM and carotenoid observed were probably due to oxidative stress. In contrast, only the carotenoid expression level was affected in SWGK46G$_R$, which seemed to be due to a blue light effect upon the *A. thaliana* CrtL. The carotenoid expression down-regulation could be regulated at DNA level, or at the protein level. If the regulation happens at the DNA level, several possibilities are listed as follows: (1) The possibility, that *A. thaliana crtL* in SWGK46G$_R$ was mutated, has been experimentally excluded, but the *crtD* promoter region might still be mutated and needs to be checked. (2) It is also possible that either the *crtIB* promoter or the *crtIB* structural genes may be mutated. Preliminary sequencing of the *crtIB* promoter and the structural genes has indicated

that most of the DNA sequence is identical in both the SWGK46Y and SWGK46G$_R$ mutants, at least three small regions may show changes. This work is ongoing (Dr. C. Autenrieth, unpublished). (3) The possibility of *crtE* as the regulatory point is low, because if *crtE* is affected, it would lead to a change in the UQ levels and create growth limitation, which were not observed. More particularly, we did not observe a genuine *crtE*⁻ phenotype (R. Ghosh, unpublished) in SWGK46. We therefore excluded *crtE* as a candidate for the crt regulatory point. (4) The LH1 complexes might be mutated to be adjusted for stress for survival.

It might also be that the carotenoid down-regulation in SWGK46G$_R$ is at the protein level. Based upon the experimental results and the bioinformatics analysis (see Chapter 6), our hypothesis is as follows: (1) *A. thaliana* CrtL might possess a domain which functions as a blue light sensor, and a domain functions as a catalytic domain, similar to that documented for OCP. (2) The light sensor domain of the *A. thaliana* CrtL might bind carotenoid and the catalytic domain might interact with carotenoid biosynthesis regulating enzymes, e.g. CrtIB. (3) Upon blue light illumination, β-carotene in the *A. thaliana* CrtL undergoes conformational changes, which might lead to the conformational changes in the *A. thaliana* CrtL, then via the signal transduction region (high similarity to the conserved DAS motif in cryptochromes), the signal is transferred to the catalytic domain to interact with CrtIB differently. (4) Finally the carotenoid biosynthesis in the SWGK46G$_R$ is down-regulated, which however could be recovered in the dark.

Within the context of my PhD work, we have explicitly examined the question of crt coupling to BChla biosynthesis for the first time. This hypothesis was supported by the BChla⁻ mutant (ST3) which contains intact carotenoid biosynthesis genes but cannot produce carotenoids. We hypothesize that the reason is due to the absence of BChla (Ghosh, unpublished) which seems to have an "activation effect" upon the carotenoid biosynthesis. Another proof of the BChla "activation effect" upon crt biosynthesis is an another BChla biosynthesis mutant, which cannot produce BChla but is still capable of accumulating bacteriochlorophyllide, does produce carotenoid. The activation effect of BChla upon carotenoid has never been considered in any purple bacteria so far. However, all experimental observations of the expression of naturally occurring physiological carotenoids (e.g. lycopene, thSpx) suggest that the carotenoid expression is coupled and coordinated with the BChla level. The ratio between the LH1 Q$_y$ absorption maximum and carotenoid absorption maximum, (e.g. lycopene, (A$_{882}$/A$_{488}$)) was used to indicate the BChla to lycopene level in SLYC18, which is always fixed

(around 1.96). However, the A_{877}/A_{483} of SWGK46Y is 0.93 and it seemed that the coordination between BChla and β-carotene is not satisfied in this strain since the β-carotene level is higher than that is required for saturation of the LH1 complexes. This is the first time to report this unusual decoupling between BChla and carotenoid in all purple bacteria so far. We hypothesize that (1) *A. thaliana* CrtL can function as a decoupling agent to release the crt expression from that of BChla. Possibly, in SWGK46Y, *A. thaliana* CrtL interacts with CrtIB, since enzyme activity prior to this step would lead to recognizable phenotypic effect which are not observed. (2) Probably, the conformation of the CrtIB could be changed by interacting with CrtL which could lead to the down-/up-regulation of crt biosynthesis. Another explanation for the crt down-regulation is as follows: The presence of β-carotene in the LH1 complexes may make the LH1 BChla more susceptible to light-induced triplet formation. To avoid this, CrtIB possibly experiences a mutation which causes it to reduce carotenoid levels, thereby reducing the negative effect of higher amounts of BChla triplets. (3) The necessity of BChla activation of CrtIB is possibly relaxed by the binding to CrtL, which thus leads to high levels of β-carotene.

According to all the experimental results and bioinformatics analysis of the *A. thaliana* CrtL, my perspectives for the further research are as follows:

(1) The chromatophores and the LH1 complexes from SWGK46Y, SWGK46G$_R$ and SWGK46B should be purified and examined systematically. If β-carotene is mainly incorporated into the LH1 complexes of SWGK46Y, we should be able to obtain nearly the same absorption spectra of the chromatophores and the LH1 complexes of SWGK46Y. It would also be interesting to know whether the LH1 complexes of SWGK46B are composed of two types, one with bound carotenoid and the other one without bound carotenoid. Then the SWGK46G$_R$ will be the excellent model to investigate the localization of carotenoid if the carotenoid is not incorporated into the LH1 complexes. Is the carotenoid in the membrane or bound to protein (lycopene β-cyclase)?

(2) We hypothesize that *A. thaliana* CrtL might activate carotenoid expression, thus leading to the decoupling between BChla and carotenoid. This hypothesis might be proved by conjugating ST3 (BChla⁻ mutant) with pSUPNPTCRTL to insert *crtL* into the *crtCD* locus. If our hypothesis is correct, this insertion would allow carotenoid expression in ST3.

(3) To understand the regulatory mechanism of *A. thaliana* CrtL in *R. rubrum* and its interaction with carotenoid biosynthesis enzymes, it would be important to isolate the *A. thaliana* CrtL to investigate if it is in a complex with other carotenoid biosynthesis enzymes (e.g. CrtIB). It is also interesting to

investigate whether the interaction between *A. thaliana* CrtL and *R. rubrum* CrtIB needs lycopene and β-carotene as cofactor.

(4) According to the bioinformatics analysis, the putative light sensing and catalytic domains of *A. thaliana* CrtL might be separated and subcloned into expression vectors. SLYC18, containing only the light sensing or catalytic domain in an expression vector *in trans*, would be expected to obtain the same phenotype without any new phenotypic observations. Only when both of the light sensing and catalytic domains are presented in SLYC18 *in trans*, should the SWGK46G$_R$ phenotype be observed when grown under photoheterotrophic conditions in the presence of blue light.

(5) A future strategy for the expression of carotenoids would be to create a mutant according to the criteria as follows: (a) LH1 will be mutated so that carotenoid does not bind. (2) The hypothetical necessity for the BChla activation of CrtIB must be abolished. In this *R. rubrum* mutant, we would expect a normal level of ICM (thus an enlarged membrane compartment within the cell). Since the LH1 complexes are in the carotenoid-less blue-green form (Q$_y$ at 873 nm) that carotenoids will be released into the membrane.

Appendix 1: Solutions for Southern hybridization

20×SSC: 3 M NaCl

0.3 M Na_3-citrate

pH 7.0

Prehybridization solution: 5×SSC

1% blocking reagent

0.1% N-lauroylsarcosine

0.02% SDS (w/v)

Buffer 1: 0.1 M Maleic acid

0.15 M NaCl

Washing buffer: 0.3% Tween 20 in buffer 1

Buffer 2: 1% blocking reagent in buffer 1

Fab conjugate buffer: Anti-Dig fab fragment diluted with buffer 2 in the ratio of 1:20000

Buffer 3: 0.1 M Tris-HCl

0.1 M NaCl

50 mM $MgCl_2$

pH 9.5

CDP-Star solution: CDP-Star diluted with buffer 3 in the ratio of 1:100

Appendix 2: Additional figures of SLYC18 work

This part of work concerning SLYC18 was not included in the published manuscript.

A2.1

The MS analysis confirms that SLYC18 produces lycopene (molecular weight of 536 g) almost exclusively. The profile of the molecular weights of lycopene in different isotopic forms is shown as an insert in the **FigA2.1**.
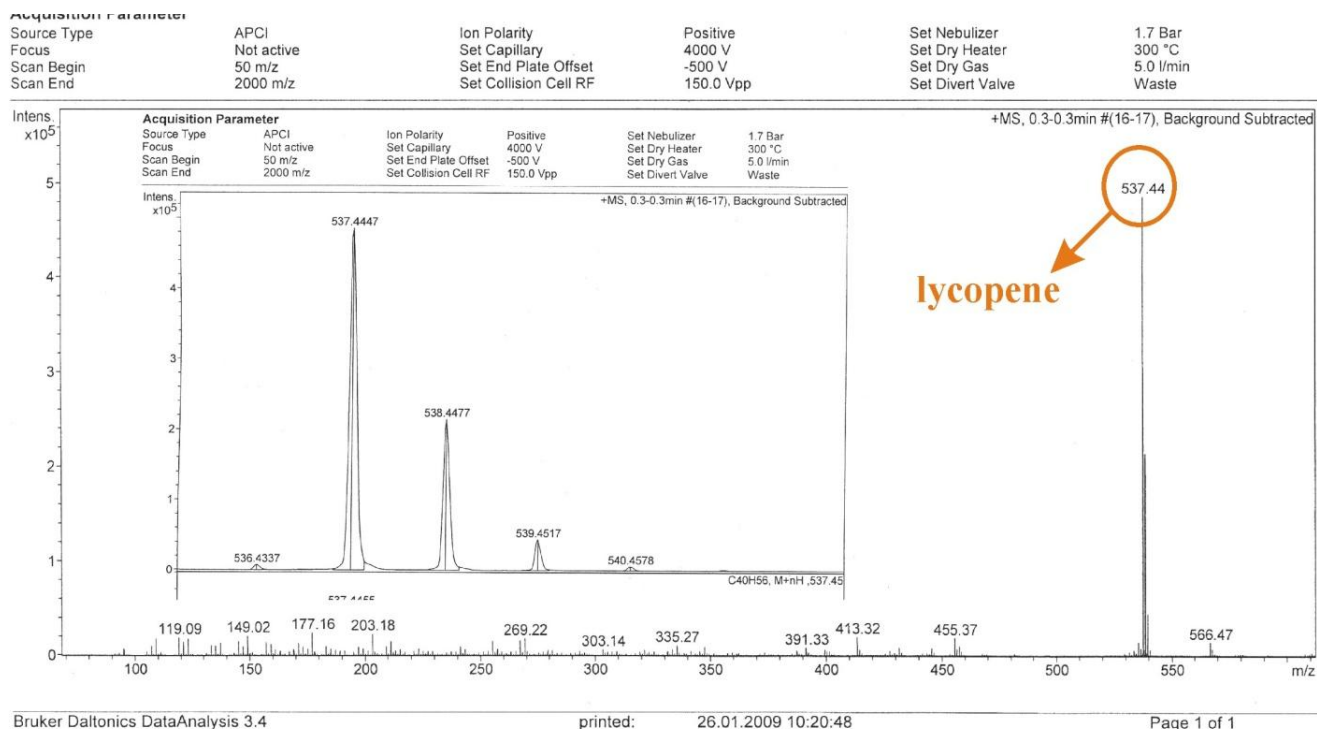
Fig A2.1 SLYC18 MS analysis. The major peak of the SLYC18 products, lycopene, is indicated. The profile of the lycopene labeled with different isotopes is shown as an insert.

A2.2

Absorption spectral analysis of the chromatophores of S1 and SLYC18.
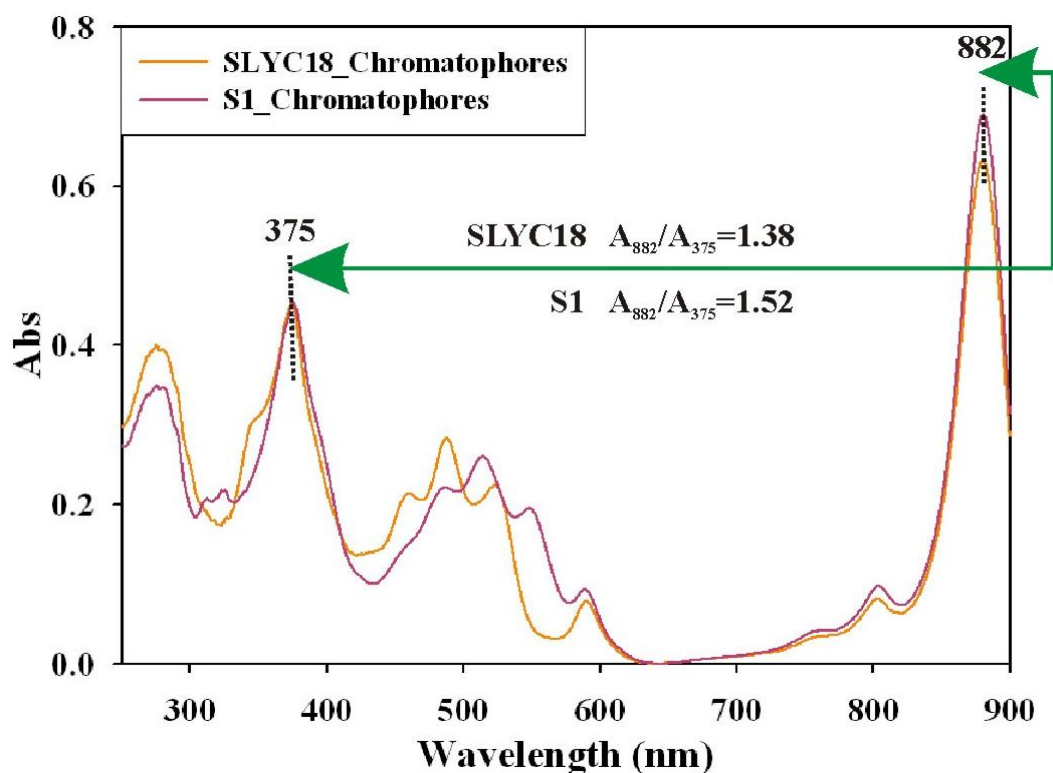
Fig A2.2 Absorption spectral analysis of SLYC18 chromatophores in comparison to that of S1. The ratio (A_{882}/A_{375}) was used to indicate whether the LH1 complexes of SLYC18 are in a native state compared to that of S1. The A_{882}/A_{375} of SLYC18 is 1.38, very close to that of S1 (1.52).

A2.3

For precise and sensitive examination of the details of the pigment-pigment interactions in LH1 and RC, CD spectral analysis of the chromatophores of SLYC18 and S1 was performed. The diagnostic S-shaped signal (Philipson and Sauer, 1973) due to the active RC contains a maximal and minimal absorption at 807 nm and 820 nm, respectively, with a crossover point at 813 nm (Philipson and Sauer, 1973). Here, we present that the S-shaped signals of the RCs from S1 and SLYC18 have nearly identical crossover points at 804 nm, maximal and minimal absorption peaks at 797 nm and 816 nm. Native LH1 complexes in chromatophores also show a diagnostic S-shaped signal (Cogdell and Scheer, 1985; Ghosh *et al.,* 1988) with a maximal and minimal absorption at 875 nm and 899 nm, respectively, as well as a crossover point at 886 nm. The NIR-CD S-shaped signal of the LH1 complexes of

SLYC18 exhibited a maximum, a minimum and a crossover point at 867 nm, 876 nm and 891 nm, respectively, corresponding well to those of S1.

Probably due to insufficient sonication of the membrane sample, the CD signal shown has somewhat lower intensity at 876/870 nm than shown by Cogdell and Scheer (1985), and Ghosh *et al.* (1988). This is typical for CD-signal distortion due to turbidity.
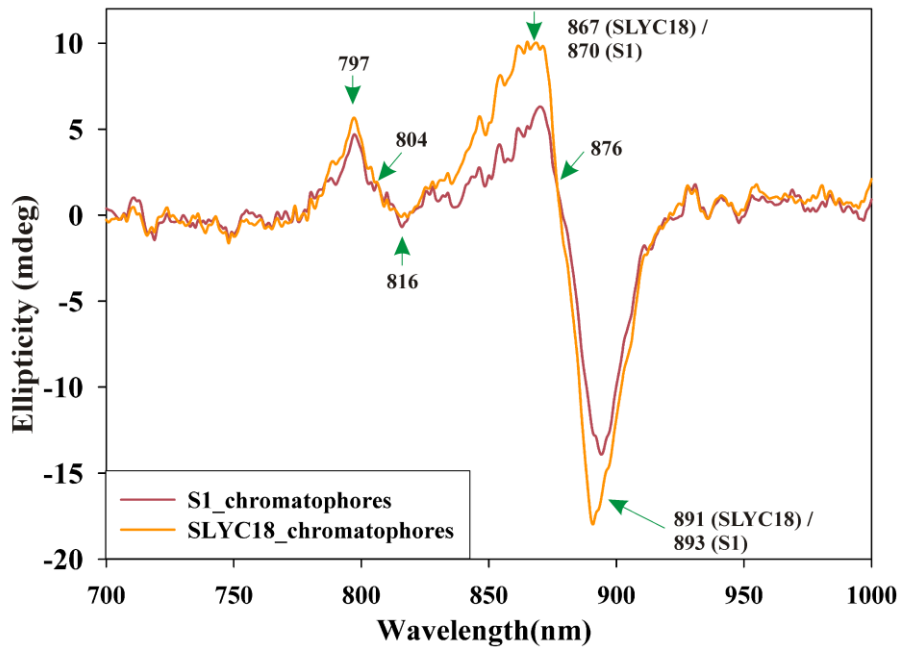
Fig A2.3 CD spectral analysis of SLYC18 chromatophores in comparison with that of S1. The characteristic S shape signal due to the native RC and LH1 are indicated with maxima, minima and crossover points.

Appendix 3: Construction vectors for SWGK46

pBsWG1
2.94 kb
Kp Xh Sna RV B Xb Nt Sc
21 bp
Spe

pBsWG6
2.97 kb
Kp Xh S HIII RV RI Swa B Xb Nt Sc
rbs
Spe
27 bp

pBsWG7
2.98 kb
Kp Xh Ml P S HIII RV RI Swa B Xb Nt Sc
rbs
18 bp Spe

pBsWG8NPT
4.36 kb
Kp Xh Ml P S HIII RV RI Swa B Xb Nt Sc
rbs
Spe
1.50 kb
npt
Al Bgl P Sm P
1kb

Kp Xh Ml P S HIII RV RI Swa B Xb Nt Sc **pBsWG9CRTL**
rbs **4.73 kb**
EII P
crt L
RV Kp 1.78 kb
1kb

pBsWGE5
9.36 kb

pBsWGE6ΔPst
9.36 kb

pBsWGE8L
8.24 kb

pBsWGE9NPT
9.73 kb

pBsWGE9NPTCRTL   11.51 kb

EII   EII

crt C

crt D

(Xb)
B
(RV/Xb)
Spe
B
(Sm/RV)

(RV/Sm)
(P)
RI
RV
(H/RV)
(Sna)

Nt
Xh
Sc

Sc

Nt

B

RI
Xh

crt F′

crt E

AI

RI

AI

P

cobZ$_{RR}$

EII   Sw   Bgl   (M)   (P)   (EII)

P
EII

crt L

npt

rbs

RI

Kp
RV

AI
Bgl
P

Sm
P

1 kb

pSUPNPTCRTL
16.36 kb

Insert: 8.53 kb

EcoRI

bla   cat

mob

roP

pSUP202
7.83 kb

tetA

181

Appendix 4 Re-design of the synthetic *crtWZ* cistron

For the production of astaxanthin *in trans* in *R. rubrum*, a two gene cistron *crtWZ* must be introduced into SWGK46Y. Here present the re-design of the *A. aurantiacum crtWZ* prior to synthesis.

A4.1 Investigation of the codon usage of genes in *R. rubrum* which are highly/weakly expressed

The codon usage of genes (listed in **Table A4.1**) in *R. rubrum* was investigated by using a Qbasic program written by C. Autenrieth and R. Ghosh, and the result is summarized in **Fig A4.1**.

R. rubrum genes	Proteins	Expression level*
petB	Cytochrome b subunit of the cytochrome bc_1 complexes	++
cycA	Cytochrome c_2	+++
cbbM	Ribulose bisphosphate carboxylase (RuBisCO)	+++
pufA	LH1- α subunit	+++
pufB	LH1- β subunit	+++
pufL	RC- L subunit	++
pufM	RC- M subunit	++
crtI	Carotenoid biosynthesis enzymes	+
crtB		+
crtC		+
crtD		+
crtE		+
crtF		+

Table A4.1 The expected expression levels of typical *R. rubrum* genes.

*Summary from literature and internal group discussions.
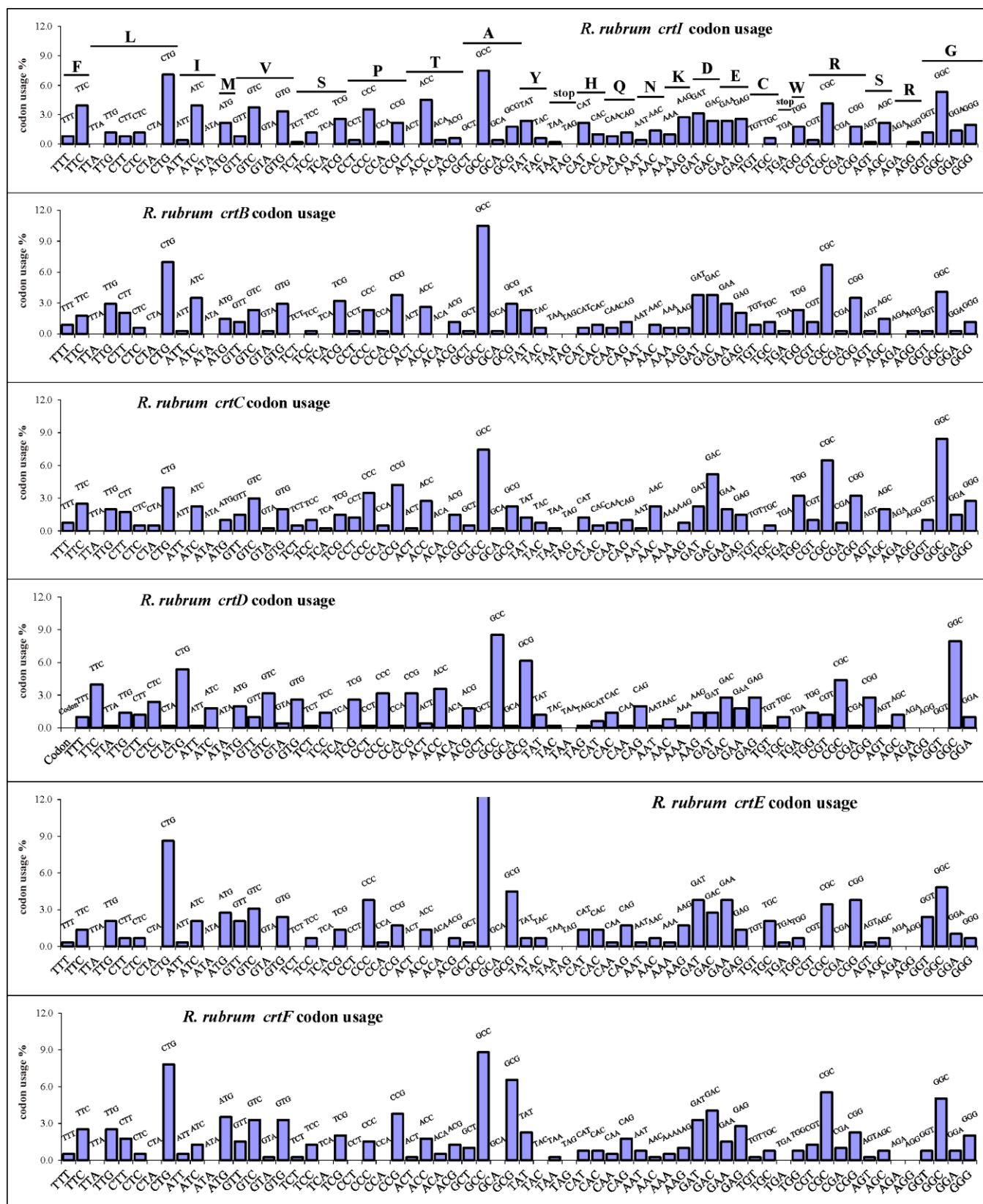
Fig A4.1 (Part I) Summary of the codon usage of the weakly expressed genes in *R. rubrum*, including *crtI, crtB, crtC, crtD, crtE, crtF* from *R. rubrum*.

183
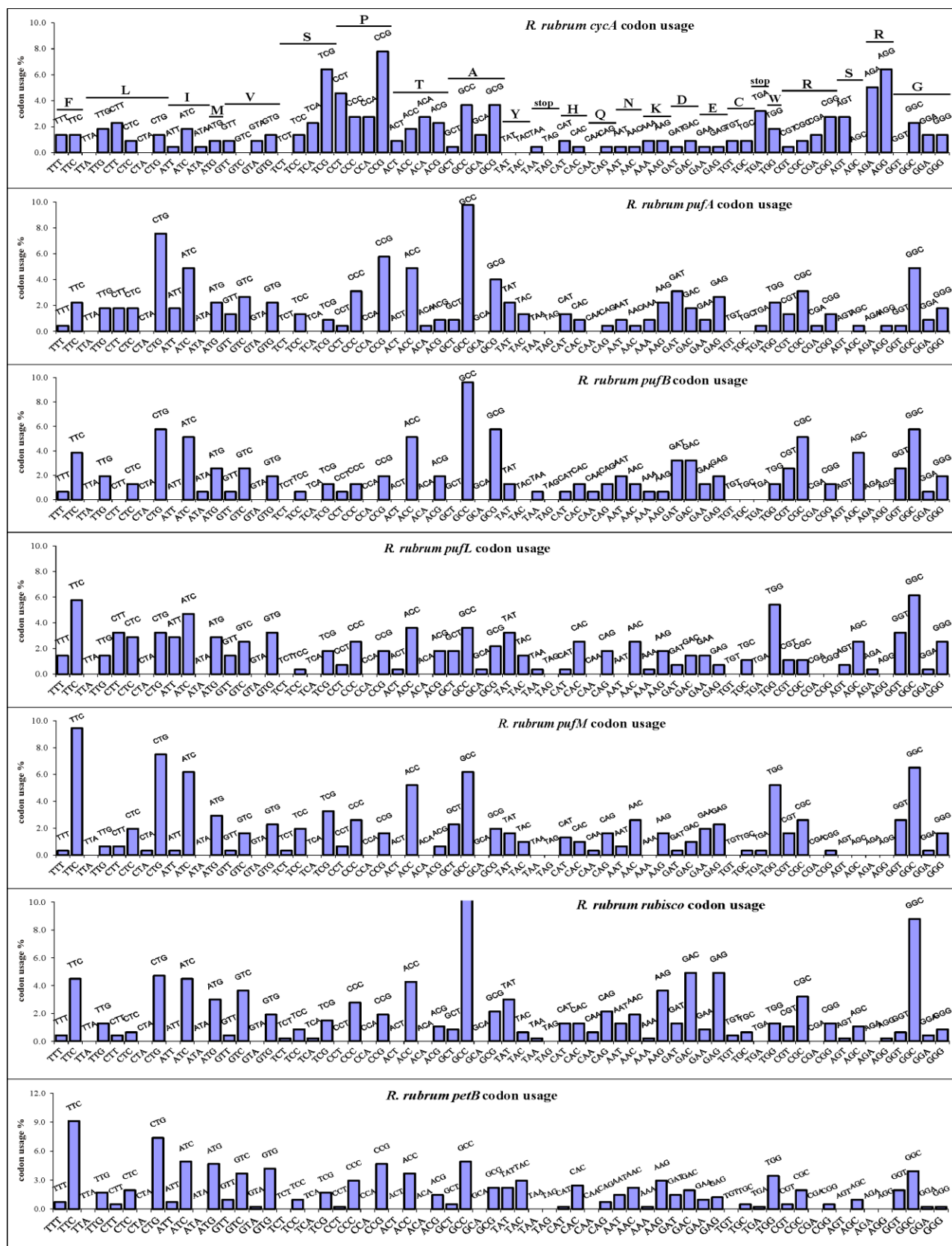
Fig A4.1 (Part II) Summary of the codon usage of the highly expressed genes in *R. rubrum*, including *cycA, pufA, pufB, pufL, pufM, rubisco* and *petB* from *R. rubrum*.

184

Table A4.2 The preferred codon usage (the highest frequency in **Fig A4.1**) **of genes in _R. rubrum_ which are highly expressed**

Phe	F	TTC
Leu	L	CTG
Ile	I	ATC
Met	M	ATG
Val	V	GTG
Ser	S	TCG
Pro	P	CCG
Thr	T	ACC
Ala	A	GCC
Tyr	Y	TAT
STOP		TAA
His	H	CAT
Gln	Q	CAG
Asn	N	AAC
Lys	K	AAG
Asp	D	GAC
Glu	E	GAG
Cys	C	TGC
Trp	W	TGG
Arg	R	CGC
Gly	G	GGC

A4.3 The _A. aurantiacum crtWZ_ original sequence data

The original sequences of _A. aurantiacum crtWZ_ were checked according to several rules as follows: (1) The codons for the first eight amino acids, especially the first two, are very important for protein expression. According to the N-end rule, Met-Ala- and Met-Gly- usually show high levels of protein expression whereas Met-Phe- usually shows low levels of protein expression (Bachmair _et al._, 1986). (2) For cytoplasmic membrane protein expression, the amino acids flanking the transmembrane helices should follow the rule that negative charges are localized in the periplasm and positive charges are present in the cytoplasm. Proteins containing three positive charges are usually not able to be exported to the periplasm in bacteria. (3) If AGA/AGG (R) are present in the _A. aurantiacum crtWZ_, which are rare codons in _E. coli_, those should be replaced with the preferred codon of the overexpressed genes of _R. rubrum_. (4) In prokaryotic organisms, efficient protein expression requires a rbs at a distance of 9-16 bp in the upstream of the starting codon. (5) Hairpins often present in the eukaryotic genes which can be removed by eukaryotic RNAase but will create problems for prokaryotic organisms. Therefore, it is

better to check whether there are hairpins in the *A. aurantiacum crtWZ.* After checking according to the rules mentioned above, only one rare codon AGA was found (Fig A4.2).

Agrobacterium aurantiacum crtW **242 aa**																								
M	S	A	H	A	L	P	K	A	D	L	T	A	T	S	L	I	V	S	G	G	I	I	A	A	W	L
atg	agc	gca	cat	gcc	ctg	ccc	aag	gca	gat	ctg	acc	gcc	acc	agc	ctg	atc	gtc	tcg	ggc	ggc	atc	atc	gcc	gct	tgg	ctg

A	L	H	V	H	A	L	W	F	L	D	A	A	A	H	P	I	L	A	I	A	N	F	L	G	L	T
gcc	ctg	cat	gtg	cat	gcg	ctg	tgg	ttt	ctg	gac	gca	gcg	gcg	cat	ccc	atc	ctg	gcg	atc	gca	aat	ttc	ctg	ggg	ctg	acc

W	L	S	V	G	L	F	I	I	A	H	D	A	M	H	G	S	V	V	P	G	R	P	R	A	N	A
tgg	ctg	tcg	gtc	gga	ttg	ttc	atc	atc	gcg	cat	gac	gcg	atg	cac	ggg	tcg	gtg	gtg	ccg	ggg	cgt	ccg	cgc	gcc	aat	gcg

A	M	G	Q	L	V	L	W	L	Y	A	G	F	S	W	R	K	M	I	V	K	H	M	A	H	H	R
gcg	atg	ggc	cag	ctt	gtc	ctg	tgg	ctg	tat	gcc	gga	ttt	tcg	tgg	cgc	aag	atg	atc	gtc	aag	cac	atg	gcc	cat	cac	cgc

H	A	G	T	D	D	D	P	D	F	D	H	G	G	P	V	R	W	Y	A	R	F	I	G	T	Y	F
cat	gcc	gga	acc	gac	gac	gac	ccc	gat	ttc	gac	cat	ggc	ggc	ccg	gtc	cgc	tgg	tac	gcc	cgc	ttc	atc	ggc	acc	tat	ttc

G	W	R	E	G	L	L	L	P	V	I	V	T	V	Y	A	L	I	L	G	D	R	W	M	Y	V	V
ggc	tgg	cgc	gag	ggg	ctg	ctg	ctg	ccc	gtc	atc	gtg	acg	gtc	tat	gcg	ctg	atc	ctt	ggg	gat	cgc	tgg	atg	tac	gtg	gtc

F	W	P	L	P	S	I	L	A	S	I	Q	L	F	V	F	G	T	W	L	P	H	R	P	G	H	D
ttc	tgg	ccg	ctg	ccg	tcg	atc	ctg	gcg	tcg	atc	cag	ctg	ttc	gtg	ttc	ggc	acc	tgg	ctg	ccg	cac	cgc	ccc	ggc	cac	gac

A	F	P	D	R	H	N	A	R	S	S	R	I	S	D	P	V	S	L	L	T	C	F	H	F	G	G
gcg	ttc	ccg	gac	cgc	cac	aat	gcg	cgg	tcg	tcg	cgg	atc	agc	gac	ccc	gtg	tcg	ctg	ctg	acc	tgc	ttt	cac	ttt	ggc	ggt

Y	H	H	E	H	H	L	H	P	T	V	P	W	W	R	L	P	S	T	R	T	K	G	D	T	A	Stop
tat	cat	cac	gaa	cac	cac	ctg	cac	ccg	acg	gtg	ccg	tgg	tgg	cgc	ctg	ccc	agc	acc	cgc	acc	aag	ggg	gac	acc	gca	tga

Agrobacterium aurantiacum crtZ **162 aa**																								
M	T	N	F	L	I	V	V	A	T	V	L	V	M	E	L	T	A	Y	S	V	H	R	W	I	M	H
atg	acc	aat	ttc	ctg	atc	gtc	gtc	gcc	acc	gtg	ctg	gtg	atg	gag	ttg	acg	gcc	tat	tcc	gtc	cac	cgc	tgg	atc	atg	cac

G	P	L	G	W	G	W	H	K	S	H	H	E	E	H	D	H	A	L	E	K	N	D	L	Y	G	L
ggc	ccc	ctg	ggc	tgg	ggc	tgg	cac	aag	tcc	cac	cac	gag	gaa	cac	gac	cac	gcg	ctg	gaa	aag	aac	gac	ctg	tac	ggc	ctg

V	F	A	V	I	A	T	V	L	F	T	V	G	W	I	W	A	P	V	L	W	W	I	A	L	G	M
gtc	ttt	gcg	gtg	atc	gcc	acg	gtg	ctg	ttc	acg	gtg	ggc	tgg	atc	tgg	gcg	ccg	gtc	ctg	tgg	tgg	atc	gcc	ttg	ggc	atg

T	V	Y	G	L	I	Y	F	V	L	H	D	G	L	V	H	Q	R	W	P	F	R	Y	I	P	R	K
act	gtc	tat	ggg	ctg	atc	tat	ttc	gtc	ctg	cat	gac	ggg	ctg	gtg	cat	cag	cgc	tgg	ccg	ttc	cgt	tat	atc	ccg	cgc	aag

G	Y	A	R	R	L	Y	Q	A	H	R	L	H	H	A	V	E	G	R	D	H	C	V	S	F	G	F
ggc	tat	gcc	aga	cgc	ctg	tat	cag	gcc	cac	cgc	ctg	cac	cat	gcg	gtc	gag	ggg	cgc	gac	cat	tgc	gtc	agc	ttc	ggc	ttc

I	Y	A	P	P	V	D	K	L	K	Q	D	L	K	M	S	G	V	L	R	A	E	A	Q	E	R	T
atc	tat	gcg	ccc	ccg	gtc	gac	aag	ctg	aag	cag	gac	ctg	aag	atg	tcg	ggc	gtg	ctg	cgg	gcc	gag	gcg	cag	gag	cgc	acg

| Stop |
| tga |

Fig A4.2 The original sequences of *A. aurantiacum crtWZ* genes. The charges and the rare codon have been indicated in colour.

A4.4 The *A. aurantiacum crtWZ* genes modified according to the preferred codon usage of *R. rubrum* (Aa*crtWZ*mod1)

All the codons were checked and replaced with the *R. rubrum* preferred codon usage (**Table A4.2**), and the rare codon was removed.

```
        Agrobacterium aurantiacum   crtW  242aa modified according to cycA_pufA_pufB usage

      M   S   A   H   A   L   P   K   A   D   L   T   A   T   S   L   I   V   S   G   G   I   I   A   A   W   L
ori   atg agc gca cat gcc ctg ccc aag gca gat ctg acc gcc acc agc ctg atc gtc tcg ggc ggc atc atc gcc gct tgg ctg
mod1  atg tcg gcc cat gcc ctg ccg aag gcc gac ctg acc gcc acc tcg ctg atc gtg tcg ggc ggc atc atc gcc gcc tgg ctg

      A   L   H   V   H   A   L   W   F   L   D   A   A   A   H   P   I   L   A   I   A   N   F   L   G   L   T
ori   gcc ctg cat gtg cat gcg ctg tgg ttt ctg gac gca gcg gcg cat ccc atc ctg gcg atc gca aat ttc ctg ggg ctg acc
mod1  gcc ctg cat gtg cat gcc ctg tgg ttc ctg gac gcc gcc gcc cat ccg atc ctg gcc atc gcc aac ttc ctg ggc ctg acc

      W   L   S   V   G   L   F   I   I   A   H   D   A   M   H   G   S   V   V   P   G   R   P   R   A   N   A
ori   tgg ctg tcg gtc gga ttg ttc atc atc gcg cat gac gcg atg cac ggg tcg gtg gtg ccg ggg cgt ccg cgc gcc aat gcg
mod1  tgg ctg tcg gtg ggc ctg ttc atc atc gcc cat gac gcc atg cat ggc tcg gtg gtg ccg ggc cgc ccg cgc gcc aac gcc

      A   M   G   Q   L   V   L   W   L   Y   A   G   F   S   W   R   K   M   I   V   K   H   M   A   H   H   R
ori   gcg atg ggc cag ctt gtc ctg tgg ctg tat gcc gga ttt tcg tgg cgc aag atg atc gtc aag cac atg gcc cat cac cgc
mod1  gcc atg ggc cag ctg gtg ctg tgg ctg tat gcc ggc ttc tcg tgg cgc aag atg atc gtg aag cat atg gcc cat cat cgc

      H   A   G   T   D   D   D   P   D   F   D   H   G   G   P   V   R   W   Y   A   R   F   I   G   T   Y   F
ori   cat gcc gga acc gac gac gac ccc gat ttc gac cat ggc ggc ccg gtc cgc tgg tac gcc cgc ttc atc ggc acc tat ttc
mod1  cat gcc ggc acc gac gac gac ccg gac ttc gac cat ggc ggc ccg gtg cgc tgg tat gcc cgc ttc atc ggc acc tat ttc

      G   W   R   E   G   L   L   L   P   V   I   V   T   V   Y   A   L   I   L   G   D   R   W   M   Y   V   V
ori   ggc tgg cgc gag ggg ctg ctg ctg ccc gtc atc gtg acg gtc tat gcg ctg atc ctt ggg gat cgc tgg atg tac gtg gtc
mod1  ggc tgg cgc gag ggc ctg ctg ctg ccg gtg atc gtg acc gtg tat gcc ctg atc ctg ggc gac cgc tgg atg tat gtg gtg

      F   W   P   L   P   S   I   L   A   S   I   Q   L   F   V   F   G   T   W   L   P   H   R   P   G   H   D
ori   ttc tgg ccg ctg ccg tcg atc ctg gcg tcg atc cag ctg ttc gtg ttc ggc acc tgg ctg ccg cac cgc ccc ggc cac gac
mod1  ttc tgg ccg ctg ccg tcg atc ctg gcc tcg atc cag ctg ttc gtg ttc ggc acc tgg ctg ccg cat cgc ccg ggc cat gac

      A   F   P   D   R   H   N   A   R   S   S   R   I   S   D   P   V   S   L   L   T   C   F   H   F   G   G
ori   gcg ttc ccg gac cgc cac aat gcg cgg tcg tcg cgg atc agc gac ccc gtg tcg ctg ctg acc tgc ttt cac ttt ggc ggt
mod1  gcc ttc ccg gac cgc cat aac gcc cgc tcg tcg cgc atc tcg gac ccg gtg tcg ctg ctg acc tgc ttc cat ttc ggc ggc

      Y   H   H   E   H   H   L   H   P   T   V   P   W   W   R   L   P   S   T   R   T   K   G   D   T   A   stop
ori   tat cat cac gaa cac cac ctg cac ccg acg gtg ccg tgg tgg cgc ctg ccc agc acc cgc acc aag ggg gac acc gca tga
mod1  tat cat cat gag cat cat ctg cat ccg acc gtg ccg tgg tgg cgc ctg ccg tcg acc cgc acc aag ggc gac acc gcc tga


                       Agrobacterium aurantiacum   crtZ    162aa

      M   T   N   F   L   I   V   V   A   T   V   L   V   M   E   L   T   A   Y   S   V   H   R   W   I   M   H
ori   atg acc aat ttc ctg atc gtc gtc gcc acc gtg ctg gtg atg gag ttg acg gcc tat tcc gtc cac cgc tgg atc atg cac
mod1  atg acc aac ttc ctg atc gtg gtg gcc acc gtg ctg gtg atg gag ctg acc gcc tat tcg gtg cat cgc tgg atc atg cat

      G   P   L   G   W   G   W   H   K   S   H   H   E   E   H   D   H   A   L   E   K   N   D   L   Y   G   L
ori   ggc ccc ctg ggc tgg ggc tgg cac aag tcc cac cac gag gaa cac gac cac gcg ctg gaa aag aac gac ctg tac ggc ctg
mod1  ggc ccg ctg ggc tgg ggc tgg cat aag tcg cat cat gag gag cat gac cat gcc ctg gag aag aac gac ctg tat ggc ctg

      V   F   A   V   I   A   T   V   L   F   T   V   G   W   I   W   A   P   V   L   W   W   I   A   L   G   M
ori   gtc ttt gcg gtg atc gcc acg gtg ctg ttc acg gtg ggc tgg atc tgg gcg ccg gtc ctg tgg tgg atc gcc ttg ggc atg
mod1  gtg ttc gcc gtg atc gcc acc gtg ctg ttc acc gtg ggc tgg atc tgg gcc ccg gtg ctg tgg tgg atc gcc ctg ggc atg

      T   V   Y   G   L   I   Y   F   V   L   H   D   G   L   V   H   Q   R   W   P   F   R   Y   I   P   R   K
ori   act gtc tat ggg ctg atc tat ttc gtc ctg cat gac ggg ctg gtg cat cag cgc tgg ccg ttc cgt tat atc ccg cgc aag
mod1  acc gtg tat ggc ctg atc tat ttc gtg ctg cat gac ggc ctg gtg cat cag cgc tgg ccg ttc cgc tat atc ccg cgc aag

      G   Y   A   R   R   L   Y   Q   A   H   R   L   H   H   A   V   E   G   R   D   H   C   V   S   F   G   F
ori   ggc tat gcc aga cgc ctg tat cag gcc cac cgc ctg cac cat gcg gtc gag ggg cgc gac cat tgc gtc agc ttc ggc ttc
mod1  ggc tat gcc cgc cgc ctg tat cag gcc cat cgc ctg cat cat gcc gtg gag ggc cgc gac cat tgc gtg tcg ttc ggc ttc

      I   Y   A   P   P   V   D   K   L   K   Q   D   L   K   M   S   G   V   L   R   A   E   A   Q   E   R   T
ori   atc tat gcg ccc ccg gtc gac aag ctg aag cag gac ctg aag atg tcg ggc gtg ctg cgg gcc gag gcg cag gag cgc acg
mod1  atc tat gcc ccg ccg gtg gac aag ctg aag cag gac ctg aag atg tcg ggc gtg ctg cgc gcc gag gcc cag gag cgc acc

      stop
      tga
```

FigA4.3 The modified *A. aurantiacum crtWZ* genes with *R. rubrum* preferred codons (in blue) (Aa*crtWZ*mod1). The rare codon, AGA indicated in red, was replaced with CGC.

A4.5 The *A. aurantiacum crtWZ* genes were modified again on the basis of Aa*crtWZ*mod1 to remove the second possible open reading frame as well as the SphI and NdeI sites (Aa*crtWZ*mod2) which probably will be used to modify the genes later on.

	Agrobacterium aurantiacum crtW 242aa_remove 2nd reading frame and SphI/NdeI sites	
	M S A H A L P **K** A **D** L T A T S L I V S G G I I A A W L	
ori	atg agc gca cat gcc ctg ccc aag gca gat ctg acc gcc acc agc ctg atc gtc tcg ggc ggc atc atc gcc gct tgg ctg	
hiexp	atg tcg gcc cat gcc ctg ccg aag gcc gac ctg acc gcc acc tcg ctg atc gtg tcg ggc ggc atc atc gcc gcc tgg ctg	
	A L H V H A L W F L **D** A A A H P I L A I A N F L G L T	
ori	gcc ctg cat gtg cat gcg ctg tgg ttt ctg gac gca gcg gcg cat ccc atc ctg gcg atc gca aat ttc ctg ggg ctg acc	
hiexp	gcc ctg cat gtg cac ggc ctg tgg ttc ctg gac gcc gcc gcc cat ccg atc ctg gcc atc gcc aac ttc ctg ggc ctg acc	
	W L S V G L F I I A H **D** A M H G S V V P G **R** P **R** A N A	
ori	tgg ctg tcg gtc gga ttg ttc atc atc gcg cat gac gcg atg cac ggg tcg gtg gtg ccg ggg cgt ccg cgc gcc aat gcg	
hiexp	tgg ctg tcg gtg ggc ctg ttc atc atc gcc cat gac gcc atg cat ggc tcg gtg gtg ccg ggc cgc ccg cgc gcc aac gcc	
	A M G Q L V L W L Y A G F S W **R** **K** M I V **K** H M A H H **R**	
ori	gcg atg ggc cag ctt gtc ctg tgg ctg tat gcc gga ttt tcg tgg cgc aag atg atc gtc aag cac atg gcc cat cac cgc	
hiexp	gcc atg ggc cag ctg gtg ctc tgg ctg tat gcc ggc ttc tcg tgg cgc aag atg atc gtg aag cac atg gcc cat cat cgc	
	H A G T **D** **D** **D** **D** P **D** F **D** H G G P V **R** W Y A **R** F I G T Y F	
ori	cat gcc gga acc gac gac gac ccc gat ttc gac cat ggc ggc ccg gtc cgc tgg tac gcc cgc ttc atc ggc acc tat ttc	
hiexp	cat gcc ggc acc gac gac gac ccg gac ttc gac cat ggc ggc ccg gtg cgc tgg tat gcc cgc ttc atc ggc acc tat ttc	
	G W **R** **E** G L L L P V I V T V Y A L I L G **D** **R** W M Y V V	
ori	ggc tgg cgc gag ggg ctg ctg ctg ccc gtc atc gtg acg gtc tat gcg ctg atc ctt ggg gat cgc tgg atg tac gtg gtc	
hiexp	ggc tgg cgc gag ggc ctg ctg ctg ccg gtg atc gtg acc gtg tat gcc ctg atc ctg ggc gac cgc tgg atg tat gtg gtg	
	F W P L P S I L A S I Q L F V F G T W L P H **R** P G H **D**	
ori	ttc tgg ccg ctg ccg tcg atc ctg gcg tcg atc cag ctg ttc gtg ttc ggc acc tgg ctg ccg cac cgc ccc ggc cac gac	
hiexp	ttc tgg ccg ctg ccg tcg atc ctg gcc tcg atc cag ctg ttc gtg ttc ggc acc tgg ctg ccg cat cgc ccg ggc cat gac	
	A F P **D** **R** H N A **R** S S **R** I S **D** P V S L L T C F H F G G	
ori	gcg ttc ccg gac cgc cac aat gcg cgg tcg tcg cgg atc agc gac ccc gtg tcg ctg ctg acc tgc ttt cac ttt ggc ggt	
hiexp	gcc ttc ccg gac cgc cat aac gcc cgc tcg tcg cgc atc tcg gac ccg gtg tcg ctg ctg acc tgc ttc cat ttc ggc ggc	
	Y H H **E** H H L H P T V P W W **R** L P S T **R** T **K** G **D** T A	
ori	tat cat cac gaa cac cac ctg cac ccg acg gtg ccg tgg tgg cgc ctg ccc agc acc cgc acc aag ggg gac acc gca tga	
hiexp	tat cat cat gag cat cat ctg cat ccg acc gtg ccg tgg tgg cgc ctg ccg tcg acc cgc acc aag ggc gac acc gcc tga	
	Agrobacterium aurantiacum crtZ 162aa	
	M T N F L I V V A T V L V M **E** L T A Y S V H **R** W I M H	
ori	atg acc aat ttc ctg atc gtc gtc gcc acc gtg ctg gtg atg gag ttg acg gcc tat tcc gtc cac cgc tgg atc atg cac	
hiexp	atg acc aac ttc ctg atc gtg gtg gcc acc gtg ctg gtg atg gag ctg acc gcc tat tcg gtg cat cgc tgg atc atg cat	
	G P L G W G W H **K** S H H **E** **E** H **D** H A L **E** **K** N **D** L Y G L	
ori	ggc ccc ctg ggc tgg ggc tgg cac aag tcc cac cac gag gaa cac gac cac gcg ctg gaa aag aac gac ctg tac ggc ctg	
hiexp	ggc ccg ctg ggc tgg ggc tgg cat aag tcg cat cat gag gag cat gac cat gcc ctg gag aag aac gac ctg tat ggc ctg	
	V F A V I A T V L F T V G W I W A P V L W W I A L G M	
ori	gtc ttt gcg gtg atc gcc acg gtg ctg ttc acg gtg ggc tgg atc tgg gcg ccg gtc ctg tgg tgg atc gcc ttg ggc atg	
hiexp	gtg ttc gcc gtg atc gcc acc gtg ctg ttc acc gtg ggc tgg atc tgg gcc ccg gtg ctg tgg tgg atc gcc ctg ggc atg	
	T V Y G L I Y F V L H **D** G L V H Q **R** W P F **R** Y I P **R** **K**	
ori	act gtc tat ggg ctg atc tat ttc gtc ctg cat gac ggg ctg gtg cat cag cgc tgg ccg ttc cgt tat atc ccg cgc aag	
hiexp	acc gtg tat ggc ctg atc tat ttc gtg ctg cat gac ggc ctg gtg cat cag cgc tgg ccg ttc cgc tat atc ccg cgc aag	
	G Y A **R** **R** L Y Q A H **R** L H H A V **E** G R **D** H C V S F G F	
ori	ggc tat gcc aga cgc ctg tat cag gcc cac cgc ctg cac cat gcg gtc gag ggg cgc gac cat tgc gtc agc ttc ggc ttc	
hiexp	ggc tat gcc cgc cgc ctg tat cag gcc cat cgc ctg cat cat gcc gtg gag ggc cgc gac cat tgc gtg tcg ttc ggc ttc	
	I Y A P P V **D** **K** L **K** Q **D** L **K** M S G V L **R** A **E** A Q **E** **R** T	
ori	atc tat gcg ccc ccg gtc gac aag ctg aag cag gac ctg aag atg tcg ggc gtg ctg cgg gcc gag gcg cag gag cgc acg	
hiexp	atc tat gcc ccg ccg gtg gac aag ctg aag cag gac ctg aag atg tcg ggc gtg ctg cgc gcc gag gcc cag gag cgc acc	
	tga	

Fig A4.4 The modified *A. auranticum crtWZ* genes (Aa*crtWZ*mod2). The second possible open reading frame in the middle of *crtW* was removed by replacing CTG with CTC (indicated with a green circle) and the SphI site (GCATGC) was removed by replacing CAT with CAC (indicated with a pink box), and the NdeI site (CATATA) was avoided by keeping the codon CAC instead of replacing it with CAT.

188

A4.6 The *A. auranticum crtWZ* genes were finally modified prior to synthesis as follows:

(1) A HindIII site, a rbs, and the upstream region of *puf* operon were placed in the front of *crtW*.

(2) The *crtW* stop codon TGA was replaced with TAA, and a synthetic *puf* linker, derived from the intergenic region of the *pufBA* cistron was added. The start codon of the second gene of the cistron (*crtZ*) was incorporated into a unique NdeI site.

(3) The end of *crtZ* was modified by replacing the TGA with a TAA (in-frame) stop codon. A NotI site, a *trp* terminator, an out-of-frame stop codon and a BamHI site were added in the end of *crtZ*.

crtW →

		M S A H A L P **K** A **D** L T A T S L I V S G G I I A A W L
ori	**HindIII** **SD** **puf operon upstream**	atg agc gca cat gcc ctg ccc aag gca gat ctg acc gcc acc agc ctg atc gtc tcg ggc ggc atc atc gcc gct tgg ctg
hiexp	TCG GAA GCT TCA GGA GGT AGC ACCC	atg tcg gcc cat gcc ctg ccg aag gcc gac ctg acc gcc acc tcg ctg atc gtg tcg ggc ggc atc atc gcc gcc tgg ctg

A L H V H A L W F L **D** A A A H P I L A I A N F L G L T W L S V G L F I

| **ori** | gcc ctg cat gtg cat gcg ctg tgg ttt ctg gac gca gcg gcg cat ccc atc ctg gcg atc gca aat ttc ctg ggg ctg acc tgg ctg tcg gtc gga ttg ttc atc |
| **hiexp** | gcc ctg cat gtg **cac** gcc ctg tgg ttc ctg gac gcc gcc gcc cat ccg atc ctg gcc atc gcc aac ttc ctg ggc ctg acc tgg ctg tcg gtg ggc ctg ttc atc |

I A H **D** A M H G S V V P G **R** P **R** A N A A M G Q L V L W L Y A G F S W R

| **ori** | atc gcg cat gac gcg atg cac ggg tcg gtg gtg ccg ggg cgt ccg cgc gcc aat gcg gcg atg ggc cag ctt gtc ctg tgg ctg tat gcc gga ttt tcg tgg cgc |
| **hiexp** | atc gcc cat gac gcc atg cat ggc tcg gtg gtg ccg ggc cgc ccg cgc gcc aac gcc gcc atg ggc cag ctg gtg ctc tgg ctg tat gcc ggc ttc tcg tgg cgc |

K M I **V** **K** H M A H H **R** **H** A **G** T **D** **D** **D** **P** **D** **F** **D** H G **G** **P** V **R** W Y A **R** F I **G**

| **ori** | aag atg atc gtc aag cac atg gcc cat cac cgc cat gcc gga acc gac gac gac ccc gat ttc gac cat ggc ggc ccg gtc cgc tgg tac gcc cgc ttc atc ggc |
| **hiexp** | aag atg atc gtg aag **cac** atg gcc cat cat cgc cat gcc ggc acc gac gac gac ccg gac ttc gac cat ggc ggc ccg gtg cgc tgg tat gcc cgc ttc atc ggc |

T Y F G W R E G L L L P V I V T V Y A L I L G D R W M Y V V F W P L P

| **ori** | acc tat ttc ggc tgg cgc gag ggg ctg ctg ctg ccc gtc atc gtg acg gtc tat gcg ctg atc ctt ggg gat cgc tgg atg tac gtg gtc ttc tgg ccg ctg ccg |
| **hiexp** | acc tat ttc ggc tgg cgc gag ggc ctg ctg ctg ccg gtg atc gtg acc gtg tat gcc ctg atc ctg ggc gac cgc tgg atg tat gtg gtg ttc tgg ccg ctg ccg |

S I L A S I Q L F V F G T W L P H **R** P G H **D** **A** F **P** **D** **R** H N **A** **R** S S **R** I

| **ori** | tcg atc ctg gcg tcg atc cag ctg ttc gtg ttc ggc acc tgg ctg ccg cac cgc ccc ggc cac gac gcg ttc ccg gac cgc cac aat gcg cgg tcg tcg cgg atc |
| **hiexp** | tcg atc ctg gcc tcg atc cag ctg ttc gtg ttc ggc acc tgg ctg ccg cat cgc ccg ggc cat gac gcc ttc ccg gac cgc cat aac gcc cgc tcg tcg cgc atc |

S **D** P V S L L T C F H F G G Y H H E H H L H P T V P W W R L P S T R T

| **ori** | agc gac ccc gtg tcg ctg ctg acc tgc ttt cac ttt ggc ggt tat cat cac gaa cac cac ctg cac ccg acg gtg ccg tgg tgg cgc ctg ccc agc acc cgc acc |
| **hiexp** | tcg gac ccg gtg tcg ctg ctg acc tgc ttc cat ttc ggc ggc tat cat cat gag cat cat ctg cat ccg acc gtg ccg tgg tgg cgc ctg ccg tcg acc cgc acc |

K G **D** T A **crtZ →** M T N F L I V V A T V L V M E L T A Y S V H R W I

| **ori** | aag ggg gac acc gca tga _puf_ **linker** ATG acc aat ttc ctg atc gtc gtc gcc acc gtg ctg gtg atg gag ttg acg gcc tat tcc gtc cac cgc tgg atc |
| **hiexp** | aag ggc gac acc gcc TAA GGA GAC GGC CAT ATG acc aac ttc ctg atc gtg gtg gcc acc gtg ctg gtg atg gag ctg acc gcc tat tcg gtg cat cgc tgg atc |

NdeI

M H G P L G W G W H **K** S H H E E H D H A L E K N D L Y G L V F A V I A

| **ori** | atg cac ggc ccc ctg ggc tgg ggc tgg cat tcc cac cac gag gaa cac gac cac gcg ctg gaa aag aac gac ctg tac ggc ctg gtc ttt gcg gtg atc gcc |
| **hiexp** | atg cat ggc ccg ctg ggc tgg ggc tgg cat aag tcg cat cat gag gag cat gac cat gcc ctg gag aag aac gac ctg tat ggc ctg gtg ttc gcc gtg atc gcc |

T V L F T V G W I W A P V L W W I A L G M T V Y G L I Y F V L H D G L

| **ori** | acg gtg ctg ttc acg gtg ggc tgg atc tgg gcg ccg gtc ctg tgg tgg atc gcc ttg ggc atg act gtc tat ggg ctg atc tat ttc gtc ctg cat gac ggg ctg |
| **hiexp** | acc gtg ctg ttc acc gtg ggc tgg atc tgg gcc ccg gtg ctg tgg tgg atc gcc ctg ggc atg acc gtg tat ggc ctg atc tat ttc gtg ctg cat gac ggc ctg |

V H Q **R** W P F **R** Y I P **R** K G Y A **R** **R** L Y Q A H **R** L H H A V E G **R** D H C

| **ori** | gtg cat cag cgc tgg ccg ttc cgt tat atc ccg cgc aag ggc tat gcc aga cgc ctg tat cag gcc cac cgc ctg cac cat gcg gtg gag ggg cgc gac cat tgc |
| **hiexp** | gtg cat cag cgc tgg ccg ttc cgc tat atc ccg cgc aag ggc tat gcc ⟨cgc⟩ cgc cgc ctg tat cag gcc cat cgc ctg cat cat gcc gtg gag ggc cgc gac cat tgc |

V S F G F I Y A P P V D K L K Q **D** L K M S G V L R A E A Q E R T

| **ori** | gtc agc ttc ggc ttc atc tat gcg ccc ccg gtc gac aag ctg aag cag gac ctg aag atg tcg ggc gtg ctg cgg gcc gag gcg cag gag cgc acg tga |
| **hiexp** | gtg tcg ttc ggc ttc atc tat gcc ccg ccg gtg gac aag ctg aag cag gac ctg aag atg tcg ggc gtg ctg ctg cgc gcc gag gcc cag gag cgc acc tga GTC TCA |

out-frame   **in-frame**

| **ori** | |
| **hiexp** | GCG CGG CCG GCC CAG CCC GCC TAA TGA GCG GGC TTT TTT TTT AGG GAT CC |

NotI   **Terminator**   **BamHI**

Fig A4.5 The final modified _A. auranticum crtWZ_ synthetic cistron. All the modifications mentioned above, including the codon usage changes (in blue), the removal of the rare codon (in red circle), the removal of the second reading frame and SphI/NdeI sites (in pink boxes), the modifications in the beginning of _crtW_ and the ends of _crtW_ and _crtZ_ are shown.

190

References

The literature relevant to the SLYC18 work has been listed separately at the end of Chapter 3.

Adams, C. W., Forrest, M. E., Cohen, S. N., and Beatty, J. T. 1989. Structural and functional analysis of transcriptional control of the *Rhodobacter capsulatus puf* operon. J. Bacteriol. **171**:473-482.

Addlesee, H. A., and Hunter, C. N. 1999. Physical mapping and functional assignment of the geranylgeranyl-bacteriochlorophyll reductase gene, *bchP*, of *Rhodobacter sphaeroides*. J. Bacteriol. **181**:7248-7255.

Addlesee, H. A., and Hunter, C. N. 2002. *Rhodospirillum rubrum* possesses a variant of the *bchP* gene, encoding geranylgeranyl-bacteriopheophytin reductase. J. Bacteriol. **184**:1578-1586.

Akbar, S., Gaidenko, T. A., Kang, C. M., O'Reilly, M., Devine, K. M., and Price, C. W. 2001. New family of regulators in the environmental signalling pathway which activates the general stress transcription factor sigma (B) of *Bacillus subtilis*. J. Bacteriol. **183**:1329-1338.

Albracht, S. P., and Hedderich, R. 2000. Learning from hydrogenase: location of a proton pump and of a second FMN in bovine NADH-ubiquinone oxidoreductase (Complex I). FEBS Lett. **485**:1-6.

Alper, H., Miyaoku, K., and Stephanopoulos, G. 2005. Construction of lycopene-overproducing *E. coli* strains by combining systematic and combinatorial gene knockout targets. Nat. Biotechnol. **23**:612-616.

Ausich, R. L., Brinkhaus, F. L., Mukharji, I., Proffitt, J. H., Yarger, J. G., Yen, H.-C. B. 1991. Biosynthesis of carotenoids in genetically engineered hosts. Patent Application PCT/US91/01458.

Bachmair, A., Finley, D., and Varshavsky, A.1986. In vivo half-life of a protein is a function of its amino-terminal residue. Science. **234**:179-186.

Ballottari, M., Girardon, J., Dall'osto, L., and Bassi, R. 2012. Evolution and functional properties of photosystem II light harvesting complexes in eukaryotes. Biochim. Biophys. Acta. **1817**:143-157.

Banerjee, R., and Batschauer, A. 2005. Plant blue-light receptors. Planta. **220**:498-502.

Bauer, C. E. 1995. Regulation of photosynthesis gene expression. pp. 1221-1234. *In* R. E. Blankenship, M. T. Madigan, and C. E. Bauer (eds), Anoxygenic Photosynthetic Bacteria. Kluwer Academic Publishers, Netherlands.

Bélanger, G., and Gingras, G. 1988. Structure and expression of the *puf* operon messenger RNA in *Rhodospirillum rubrum.* J. Biol. Chem. **263**:7639-7645.

Boatman, E. S., and Douglas, H. C. 1961. Fine structures of the photosynthetic bacterium *Rhodospirillum vannieli*. J. Cell. Biol. **11**:469-483.

Boulay, C., Wilson, A., D'Haene, S., Kirilovsky, D. 2010. Identification of a protein required for recovery of full antenna capacity in OCP-related photoprotective mechanism in cyanobacteria. Proc. Natl. Acad. Sci. USA. **107**:11620-11625.

Brock, T. D., Smith, D. W., and Madigan, M. T. 1984. "Biology of Microorganisms", 4th edition. Prentice-Hall, Inc., Englewood Cliffs, NJ.

Brudler, R., Hitomi, K., Daiyasu, H., Toh, H., Kucho, K., Ishiura, M., Kanehisa, M., Roberts, V. A., Todo, T., Tainer, J. A., and Getzoff, E. D. 2003. Identification of a new cryptochrome class. Structure, function, and evolution. Mol. Cell. **11**:59-67.

Brunisholz, R. A., Suter, F., and Zuber, H. 1984. The light-harvesting polypeptides of *Rhodospirillum rubrum.* I. The amino-acid sequence of the second light-harvesting polypeptide B880-β (B870-β) of *Rhodospirillum rubrum* S1 and the carotenoidless mutant G-9*. Aspects of the molecular structure of the two light-harvesting polypeptides B880-α (B870-α) and B880-β (B870-β) and of the antenna complex B880 (B870) from *Rhodospirillum rubrum.* Hoppe Seyler's Z. Physiol. Chem. **365**:675-688.

Bryant, D. A., Guglielmi, G., Tandequ de Marsac, N., Castets, A.-M., and Cohen-Bazire, G. 1979. The structure of cyanobacterial phycobilisomes: a model. Arch. Microbiol. **123**:113-127.

Clayton, R. K., and Haselkorn, R. 1972. Protein components of bacterial photosynthetic membranes. J. Mol. Biol. **68**:97-100.

Clayton, R. K., and Sistrom, W. R. 1978. "The Photosynthetic Bacteria". Plenum Press. New York. USA.

Cogdell, R. J., Lindsay, J. G., Valentine, J., and Durant, I. 1982. A further characterization of the B890 light-harvesting pigment-protein complex from *Rhodospirillum rubrum* strain S1. FEBS Lett. **150**:151-154.

Cogdell, R., and Scheer, H. 1985. Circular dichroism of light-harvesting complexes from purple photosynthetic bacteria. Photochem. Photobiol. **42**:669-678.

Cogdell, R. J. 1994. Introduction of new carotenoids into the bacterial photosynthetic apparatus by combining the carotenoid biosynthetic pathways of *Erwinia herbicola* and *Rhodobacter sphaeroides*. J. Bacteriol. **176**:3692-3697.

Cogdell, R. J., Isaacs, N. W., Howard, T. D., McLuskey, K., Fraser, N. J., and Prince, S. M. 1999. How photosynthetic bacteria harvest solar energy. J. Bacteriol. **181**:3869-3879.

Cogdell, R. J., Howard, T. D., Bittl, R., Schlodder, E., Geisenheimer, I., and Lubitz, W. 2000. How carotenoids protect bacterial photosynthesis. Phil. Trans. R. Soc. Lond. B. **355**:1345-1349.

Cohen-Bazire, G., and Kunisawa, R. 1960. Some observations on the synthesis and function of the photosynthetic apparatus in *Rhodospirillum rubrum*. Proc. Nat. Acad. Sc. USA. **46:**1543-1553.

Cohen-Bazire, G. 1963. Some observations on the organization of the photosynthetic apparatus in purple and green bacteria. pp. 89-114. *In* H. Gest, A. San Pietro, and L. P. Vernon, "Bacterial Photosynthesis" (eds), Antioch Press, Yellow Springs, Ohio.

Cohen-Bazire, G., and Sistrom, W. R. 1966. The prokaryotic photosynthetic apparatus. pp. 313-341. *In* L. P. Vernon and G. R. Seely, "The Chlorophylls" (eds), Academic Press, New York.

Collins, M. L. P., and Niederman, R. A. 1976. Membranes of *Rhodospirillum rubrum*: Physicochemical properties of chromatophore fractions isolated from osmotically and mechanically disrupted cells. J. Bacteriol. **126**:1326-1338.

Coomber, S. A., Chaudhri, M., Connor, A., Britton, G., and Hunter, C. N. 1990. Localized transposon Tn*5* mutagenesis of the photosynthetic gene cluster of *Rhodobacter sphaeroides*. Mol. Microbiol. **4**:977-989.

Cramer, W. A., and Crofts, A. R. 1982. Electron and proton transport. **7**:387-467. *In* Govindjee, Photosynthesis: Energy conversion by plants and bacteria, Academic Press, New York, USA.

Crosson S., and Moffat K. 2002. Photoexcited structure of a plant photoreceptor domain reveals a light-driven molecular switch. The Plant Cell. **14**:1067-1075.

Crosson, S., Rajagopal, S., and Moffat, K. 2003. The LOV domain family: photoresponsive signaling modules coupled to diverse output domains. Biochemistry. **42**:2-10.

Cuendet, P. A., and Zuber, H. 1977. Isolation and characterization of a bacteriochlorophyll-associated chromatophore protein from *Rhodospirillum rubrum* G-9. FEBS let. **79**:96-100.

Cunningham, F. X., Jr, and Gantt, E. 2001. One ring or two? Determination of ring number in carotenoids by lycopene ε-cyclases. Proc. Natl. Acad. Sci. USA. **98**:2905-2910.

Daiyasu, H., Ishikawa, T., Kuma, K., Iwai, S., Todo, T., Toh, H. 2004. Identification of cryptochrome DASH from vertebrates. Genes Cells. **9**:479-495.

Doub, L., and Vandenvelt, J. 1949. The ultraviolet absorption spectra of simple unsaturated compounds. II. m- and *o*-disubstituted benzene derivatives. J. Am. Chem. Soc. **71**:2412-2420.

Dus, K., Sletten, K., and Kamen, M. D. 1968. Cytochrome c_2 of *Rhodospirillum rubrum*. J. Biol. Chem. **243**:5507-5518.

Dutton, P. L., and Prince, R. C. 1978. Reaction center-driven cytochrome interactions in electron and proton translocation and energy coupling. pp. 525-570. *In* R. K. Clayton and W. R. Sistrom (eds), The photosynthetic bacteria. Plenum Press. New York. USA.

Elsen, S., Jaubert, M., Pignol, D., and Giraud, E. 2005. PpsR: a multifaceted regulator of photosynthesis gene expression in purple bacteria. Mol. Microbiol. **57**:17-26.

Feher, G., Allen, J. P., Okamura, M. Y., and Rees, D. C. 1989. Structure and function of bacterial photosynthetic reaction centres. Nature. **339**:111-116.

Ferguson, S. J., Jackson, J. B., and McEwan, A. G. 1987. Anaerobic respiration in the Rhodospirillaceae: characterization of pathways and evolution of roles in redox balancing during photosynthesis. FEMS Microbiology Reviews. **46**:117-143.

Figurski, D. H., and Helinski, D. R.1979. Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided *in trans*. Proc. Natl. Acad. Sci. USA. **76**:1648-1652.

Fraker, P. J., and Kaplan, S. 1972. Isolation and characterization of a bachteriochlorophyll-containing protein from *Rhodospirillum sphaeroides*. J. Biol. Chem. **247**:2732-2737.

Frenkel, A. W., and Hickman, D. D. 1959. Structure and photochemical activity of chlorophyll-containing particles from *Rhodospirillum rubrum.* J. Biophys. Biochem. Cytol. **6**:285-290.

Galperin, M. Y., Nikolskaya, A. N., and Koonin, E. V. 2001. Novel domains of the prokaryotic two-component signal transduction systems. FEMS. Microbiol. Lett. **203**:11-21.

Gauden, M., van Stokkum, I. H. M., Key, J. M., Lührs, D. C., van Grondelle, R., Hegemann, P., and Kennis, J. T. M. 2006. Hydrogen-bond switching through a radical pair mechanism in a flavin-binding photoreceptor. Proc. Natl. Acad. Sci. USA. **103**:10895-10900.

Gennis, R. B., Barquera, B., Hacker, B., Doren, S. R. von, Arnaud, S., Crofts, A. R., Davidson, E., Gray, K. A., and Daldal, F. 1993. The bc_1 complexes of *Rhodobacter sphaeroides* and *Rhodobacter capsulatus*. J. Bioenerg. Biomemb. **25**:195-209.

Ghosh, R., Bachofen, R., and Hauser, H. 1984. Structural changes accompanying the irreversible oxidation of the chromatophore membrane from *Rhodospirillum rubrum* G9. Biochim. Biophys. Acta. **765**:97-105.

Ghosh, R., Hauser, H., and Bachofen, R. 1988. Reversible dissociation of B873 light-harvesting complex from *Rhodospirillum rubrum* G9+. Biochemistry. **27**:1004-1014.

Ghosh, R., Hardmeyer, A., Thoenen, I., and Bachofen, R. 1994. Optimization of the Sistrom culture medium for large-scale batch cultivation of *Rhodospirillum rubrum* under semi-aerobic conditions with maximal yield of photosynthetic membranes. Appl. Environ. Microbiol. **60**:1698-1700.

Giuliano, G., Pollock, D., and Scolnik, P. A. 1986. The gene *crtI* mediates the conversion of phytoene into colored carotenoids in *Rhodopseudomonas capsulatus*. J. Biol. Chem. **261**:12925-12929.

Gomelsky, M., and Kaplan, S. 1997. Molecular genetic analysis suggesting interactions between AppA and PpsR in regulation of photosynthesis gene expression in *Rhodobacter sphaeroides* 2.4.1. J. Bacteriol. **179**:128-134.

Gomelsky, M., and Klug, G. 2002. BLUF: a novel FAD-binding domain involved in sensory transduction in microorganisms. Trends Biochem. Sci. **27**:497-500.

Grammel, H., Gilles, E.-D., and Ghosh, R. 2003. Microaerophilic cooperation of reductive and oxidative pathways allows maximal photosynthetic membrane biosynthesis in *Rhodospirillum rubrum*. Appl. Environ. Microbiol. **69**:6577-6586.

Hasunuma, T., Miyazawa, S.-I., Yoshimura, S., Shinzaki, Y., Tomizawa, K.-I., Shindo, K., Choi, S.-K., Misawa, N., and Miyake, C. 2008. Biosynthesis of astaxanthin in tobacco leaves by transplastomic engineering. Plant. J. **55**:857-868.

Holt, T. K., and Krogmann, D. W. 1981. A carotenoid-protein from cyanobacteria. Biochim. Biophys. Acta. **637**:408-414.

Huala, E., Oeller, P. W., Liscum, E., Han, I. S., Larsen, E., and Briggs, W. R. 1997. *Arabidopsis* NPH1: a protein kinase with a putative redox-sensing domain. Science. **278**:2120-2123.

Huang, Y., Baxter, R., Smith, B. S., Partch, C. L., Colbert, C. L., and Deisenhofer, J. 2006. Crystal structure of cryptochrome 3 from *Arabidopsis thaliana* and its implications for photolyase activity. Proc. Natl. Acad. Sci. USA. **103**:17701-17706.

Hübner, P., Willison, J. C., Vignais, P. M., and Bickle, T. A. 1991. Expression of regulatory *nif* genes in *Rhodobacter capsulatus*. J. Bacteriol. **173**:2993-2999.

Hunter, C. N., Hundle, B. S., Hearst, J. E., Lang, H. P., Gardiner, A. T., Takaichi, S., and Lee, I.Y., and Collins, M. L. 1993. Identification and partial sequence of the BchA gene of *Rhodospirillum rubrum*. Curr. Microbiol. **27**:85-90.

Hunter, C. N., Hundle, B. S., Hearst, J. E., Lang, H. P., Gardiner, A. T., Takaichi, S., and Cogdell, R. J. 1994. Introduction of new carotenoids into the bacterial photosynthetic apparatus by combining the carotenoid biosynthetic pathways of *Erwinia herbicola* and *Rhodobacter sphaeroides*. J. Bacteriol. **176**:3692-3697.

Irschik, H., and Oelze, J. 1973. Membrane differentiation in phototrophically growing *Rhodospirillum rubrum* during transition from low to high light intensity. Biochim. Biophys. Acta. **330**:80-89.

Iseki, M., Matsunaga, S., Murakami, A., Ohno, K., Shiga, K., Yoshida, K., Sugai, M., Takahashi, T., Hori, T., and Watanable, M. 2002. A blue-light-activated adenylyl cyclase mediates photoavoidance in *Euglena gracilis*. Nature. **415**:1047-1051.

Iverson, T. M., Luna, C., Cecchini, G., and Rees, D. C. 1999. Structure of the *Escherichia coli* fumarate reductase respiratory complex. Science. **284**:1961-1966.

Jackson, J. B. 1988. Bacterial photosynthesis. pp. 317-376. *In* C. Anthony (Ed), "Bacterial energy transduction", Academic Press, London.

Jamieson, S., Wang, P., Qian, P., Kirkland, J. Y., Conroy, M. J., Hunter, C. N., and Bullough, P. A. 2002. Projection structure of the photosynthetic reaction centre-antenna complex of *Rhodospirillum rubrum* at 8.5 Å resolution. EMBO J. **15**:3927-3935.

Jarillo, J. A., Gabrys, H., Capel, J., Alonso, J. M., Ecker, J. R., and Cashmore, A. R. 2001. Phototropin-related NPL1 controls chloroplast relocation induced by blue light. Nature. **410**:952-954.

Jenney, F. E., Jr, and Daldal, F. 1993. A novel membrane-associated c-type cytochrome, cyt c_y, can mediate the photosynthetic growth of *Rhodobacter sphaeroides* and *Rhodobacter capsulatus*. EMBO. J. **12**:1283-1292.

Jin, Y.-S., and Stephanopoulos, G. 2007. Multi-dimensional gene target search for improving lycopene biosynthesis in *Escherichia coli*. Metab. Eng. **9**:337-347.

Johnson, E. A., and Schroeder, W. A. 1996. Microbial carotenoids. Adv. Biochem. Eng. Biotechnol. **53**:119-178.

Jordan, P., Fromme, P., Witt, H. T., Klukas, O., Saenger, W., and Krauβ, N. 2001.Three-dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution. Nature. **411**:909-917.

Jung, A., Domratcheva, T., Tarutina, M., Wu, Q., Ko, W. H., Shoeman, R. L., Gomelsky, M., Gardner, K. H., and Schlichting, I. 2005. Structure of a bacterial BLUF photoreceptor: Insights into blue light-mediated signal transduction. Proc. Natl. Acad. Sci. USA. **102**:12350-12355.

Kajiwara, S., Fraser, P. D., Kondo, K., and Misawa, N. 1997. Expression of an exogeneous isopentenyl diphosphate isomerase gene enhances isoprenoid biosynthesis in *Escherichia coli*. Biochem. J. **324**:421-426.

Kang, M. J., Lee, Y. M., Yoon, S. H., Kim, J. H., Ock, S. W., Jung, K. H., Shin, Y. C., Keasling, J. D., and Kim, S. W. 2005. Identification of genes affecting lycopene accumulation in *Escherichia coli* using a shot-gun method. Biotechnol. Bioeng. **91**:636-642.

Karrasch, S., Bullough, P. A., and Ghosh, R. 1995. The 8.5 Å projection map of the light-harvesting complex I from *Rhodospirillum rubrum* reveals a ring composed of 16 subunits. EMBO J. **14**:631-638.

Kataoka, M., Inai, K., Ueki, T., and Yamashita, J. 1984. X-ray diffraction studies on chromatophore membrane from photosynthetic bacteria. II. Comparison of diffraction patterns of photosynthetic units from various purple bacteria. J. Biochem. **95**:567-573.

Kato, M., Ikoma, Y., Matsumoto, H., Sugiura, M., Hyodo, H., and Yano, M. 2004. Accumulation of carotenoids and expression of carotenoid biosynthetic genes during maturation in citrus fruit. Plant Physiol. **134**:824-837.

Kerfeld, C. A. 2004. Water-soluble carotenoid proteins of cyanobacteria. Arch. Bioch. Biophys. **430**:2-9.

Kirilovsky, D., and Kerfeld, C. A. 2012. The orange carotenoid protein in photoprotection of photosystem II in cyanobacteria. Biochim. Biophys. Acta. **1817**:158-166.

Kita, A., Okajima, K., Morimoto, Y., Ikeuchi, M., and Mili, K. 2005. Structure of a cyanobacterial BLUF protein, TII0078, containing a novel FAD-binding blue light sensor domain. J. Mol. Biol. **349**:1-9.

Klamt, S., Grammel, H., Straube, R., Ghosh, R., and Gilles, E. D. 2008. Modeling the electron transport chain of purple non-sulfur bacteria. Mol. Syst. Biol. **4**:1-18.

Klar, T., Pokorny, R., Moldt, J., Batschauer, A., and Essen, L. O. 2006. Cryptochrome 3 from *Arabidopsis thaliana*: Structural and functional analysis of its complex with a folate light antenna. J. Mol. Biol. **366**:954-964.

Krinsky, N. I. 1989. Carotenoids in medicine. pp. 279-291. *In* N. I. Krinsky, M. M. Mathews-Roth, and R. F. Tayor (ed.), "Carotenoids Chemistry and Biochemistry". Plenum Press, New York, N.Y.

Lancaster, C. R., Kroger, A., Auer, M., and Michel, H. 1999. Structure of fumarate reductase from *Wolinella succinogenes* at 2.2 Å resolution. Nature. **402**:377-385.

Lee, P. C., and Schmidt-Dannert, C. 2002. Metabolic engineering towards biotechnological production of carotenoids in microorganisms. Appl. Microbiol. Biotechnol. **60**:1-11.

Lin, C., Todo, T. 2005. The cryptochromes. Genome Biol. **6**:220-228.

Lupo, D., and Ghosh, R. 2004. The reaction center H subunit is not required for high levels of light-harvesting complex 1 in *Rhodospirillum rubrum* mutants. J. Bacteriol. **186**:5585-5595.

Masuda, S., and Bauer, C. E. 2002. AppA is a blue light photoreceptor that antirepresses photosynthesis gene expression in *Rhodobacter sphaeroides*. Cell. **110**:613-623.

Matthews, P. D., and Wurtzel, E. T. 2000. Metabolic engineering of carotenoid accumulation in *Escherichia coli* by modulation of the isoprenoid precursor pool with expression of deoxyxylulose phosphate synthase. Appl. Environ. Microbiol. **53**:396-400.

Mialoundama, A. S., Heintz, D., Jadid, N., Nkeng, P., Rahier, A., Deli, J., Camara, B., and Bouvier, F. 2010. Characterization of plant carotenoid cyclases as members of the flavoprotein family fuctioning with no net redox change. Plan. Physiol. **153**:970-979.

Miller, K. R., and Jacob, J. S. 1983. Two-dimensional crystals formed from photosynthetic reaction centers. J. Cell Biol. **97**:1266-1270.

Misawa, N., Yamano, S., and Ikenaga, H. 1991. Production of β-carotene in *Zymomonas mobilis* and *Agrobacterium tumefaciens* by introduction of the biosynthesis genes from *Erwinia uredovora*. Appl. Environ. Microbiol. **57**:1847-1849.

Misawa, N., Kajiwara, S., Kondo, K., Yokoyama, A., Satomi, Y., Saito, T., Miki, W., and Ohtani, T. 1995. Canthaxanthin biosynthesis by the conversion of methylene to keto groups in a hydrocarbon β-carotene by a single gene. Biochem. Biophys. Res. Commun. **209**:867-876.

Miura, Y., Kondo, K., Saito, T., Shimada, H., Fraser, P. D., and Misawa, N. 1998. Production of the carotenoids lycopene, β-carotene, and astaxanthin in the food yeast *Candida utilis*. Appl. Environ. Microbiol. **64**:1226-1229.

Möglich, A., Yang, X. J., Ayers, R. A., and Moffat K. 2010. Structure and function of plant photoreceptors. Annu. Rev. Plant Biol. **61**:21-47.

Molisch, H. 1907. Die Purpurbakterien nach neuen Untersuchungen. Fischer Verlag. Jena. pp. 95.

Stanier, R. Y., Doudoroff, M., Kunisawa, R., and Contopoulou, R. 1959. The role of organic substrates in bacterial photosynthesis. Proc. Nat. Acad. Sc. USA. **45**:1246-1260.

Mosley, C. S., Suzuki, J. Y., and Bauer, C. E. 1994. Identification and molecular genetic characterization of a sensor kinase responsible for coordinately regulating light harvesting and reaction centre gene expression in response to anaerobiosis. J. Bacteriol. **176**:7566-7573.

Narro, M. L., Adams, C. W., and Cohen, S. N. 1990. Isolation and characterization of *Rhodobacter capsulatus* mutants defective in oxygen regulation of the *puf* operon. J. Bacteriol. **172:**4549-4554.

Nelis, H. J., and De Leenheer, A. P. 1991. Microbial sources of carotenoid pigments used in foods and feeds. J. Appl. Bacteriol. **70**:181-191.

Nicholls, D. G., and Ferguson, S. J. 2001. Bioenergetics 3. Academic Press. New York.

Niederman, R. A. 1974. Membranes of *Rhodopseudomonas sphaeroides*: Interactions of chromatophores with the cell envelope. J. Bacteriol. **117**:19-28.

Oh, J. I., Eraso, J. M., and Kaplan, S. 2000. Interacting regulatory circuits involved in the orderly control of photosynthesis gene expression in *Rhodobacter sphaeroides* 2.4.1. J. Bacteriol. **182**:3081-3087.

Oh, J. I., and Kaplan, S. 2000. Redox signalling: globalization of gene expression. EMBO J. **19**:4237-4247.

Oh, J. I., and Kaplan, S. 2001. Generalized approach to the regulation and integration of gene expression. Mol. Microbiol. **39**:1116-1123.

Ohnishi, T., Moser, C. C., Page, C. C., Dutton, P. L., and Yano, T. 2000. Simple redox-linked proton-transfer design: new insights from structures of quinol-fumarate reductase. Struct. Fold. Des. **8**:R23-R32.

Pei, J. M., and Grishin, N. V. 2001. GGDEF domain is homologous to adenylyl cyclase. Proteins. **42**:210-216.

Philipson, K. D., and Sauer, K. 1973. Comparative study of the circular dichroism spectra of reaction centers from several photosynthetic bacteria. Biochemistry. **12**:535-539.

Picorel, R., Bélanger, G., and Gingras, G. 1983. Antenna holochrome B880 of *Rhodospirillum rubrum* S1. Pigment, phospholipid, and polypeptide composition. Biochemistry. **22**:2491-2497.

Polívka, T., Kerfeld, C. A., Pascher, T., and Sundström, V. 2005. Spectroscopic properties of the carotenoid 3'-hydroxyechinenone in the orange carotenoid protein from the cyanobacterium *Arthrospira maxima*. Biochemistry. **44**:3994-4003.

Ralley, L., Enfissi, E. M. A., Misawa, N., Schuch, W., Bramley, P. M., and Fraser, **P. D.** 2004. Metabolic engineering of ketocarotenoid formation in higher plants. Plant. J. **39**:477-486.

Ravanello, M. P., Ke, D., Alvarez, J., Huang, B., and Shewmaker, C. K. 2003. Coordinate expression of multiple bacterial carotenoid genes in canola leading to altered carotenoid production. Metab. Engin. **5**:255-263.

Römer, S., Fraser, P. D., Kiano, J. W., Shipton, C. A., Misawa, N., Schuch, W., and Bramley, P. M. 2000. Elevation of the provitamin A content of transgenic tomato plants. Nat. Biotechnol. **18**:666-669.

Sägesser, R. 1992. Identifikation und Charakterisierung des Photosynthese-Genclusters von *Rhodospirillum rubrum*. PhD thesis. Universität Zürich.

Sambrook, J., and Russell, D. W. 2001. Molecular cloning: a laboratory manual, 3rd ed. Cold Spring Harbor Press, Cold Spring Harbor, N.Y.

Sandmann, G., Albrecht, M., Schnurr, G., Knörzer, O., and Böger, P. 1999. The biotechnological potential and design of novel carotenoids by gene combination in *Escherichia coli*. TIBTECH. **19**:233-237.

Schachman, H. K., Pardee, A. B., and Stanier, R. Y. 1952. Studies on the macromolecular organization of microbial cells. Arch. Biochem. Biophys. **38**:245-260.

Scolnik, P. A., and Marrs, B. L. 1987. Genetic research with photosynthetic bacteria. Annu. Rev. Microbiol. **41**:703-726.

Self, S. J., Hunter, C. N., and Leatherbarrow, R. J. 1990. Molecular cloning, sequencing and expression of cytochrome c_2 from *Rhodospirillum rubrum*. Biochem. J. **265**:599-604.

Shalitin, D., Yang, H., Mockler, T. C., Maymon, M., Guo, H., Whitelam, G. C., and Lin, C. 2002. Regulation of *Arabidopsis* cryptochrome 2 by blue-light-dependent phosphorylation. Nature. **417**:763-767.

Shalitin, D., Yu, X., Maymon, M., Mockler, T. C., and Lin, C. 2003. Blue light-dependent *in vivo* and *in vitro* phosphorylation of *Arabidopsis* cryptochrome 1. Plant Cell. **15:**2421-2429.

Shimada, H., Kondo, K., Fraser, P. D., Miura, Y., Saito, T., and Misawa, N. 1998. Increased carotenoid production by the food yeast *Candida utilis* through metabolic engineering of the isoprenoid pathway. Appl. Environ. Microbiol. **64**:2676-2680.

Siebert, C. A., Qian, P., Fotiadis, D., Engel, A., Hunter, C. N., and Bullough, P. A. 2004. Molecular architecture of photosynthetic membranes in *Rhodobacter sphaeroides*: The role of PufX. EMBO J. **23**:690-700.

Simon, R., Priefer, U., and Pühler, A. 1983. A broad host range mobilization system for *in vivo* genetic engineering: transposon mutagenesis in gram negative bacteria. Nat. Biotechnol. **1**:784-791.

Sistrom, W. R. 1960. A requirement for sodium in the growth of *Rhodopseudomonas sphaeroides*. J. Gen. Microbiol. **22**:778-785.

Somers, D. E., Devlin, P. F., Kay, S. A. 1998. Phytochromes and cryptochromes in the entrainment of the *Arabidopsis* circadian clock. Science. **282**:1488-1490.

Stark, W., Kühlbrandt, W., Wildhaber, I., Wehrli, E., and Mühlethaler, K. 1984. The structure of the photoreceptor unit of *Rhodopseudomonas viridis*. EMBO. J. **3**:777-783.

Suzuki, S., Nishihara, M., Nakatsuka, T., Misawa, N., Ogiwara, I., and Yamamura, S. 2007. Flower color alteration in *Lotus japonicus* by modification of the carotenoid biosynthetic pathway. Plant. Cell. Rep. **26**:951-959.

Swem, D. L., and Bauer, C. E. 2002. Coordination of ubiquinol oxidase and cytochrome cbb_3 oxidase expression by multiple regulators in *Rhodobacter capsulatus.* J. Bacteriol. **184**:2815-2820.

Swem, L. R., Gong, X., Yu, C.-A., and Bauer, C. E. 2006. Identification of a ubiquinone-binding site that affects autophosphorylation of the sensor kinase RegB. J. Biol. Chem. **281**:6768-6775.

Swingley, W. D., Sadekar, S., Mastrian, S. D., Matthies, H. J., Hao, J., Ramos, H., Acharya, C. R., Conrad, A. L., Taylor, H. L., Dejesa, L. C., Shah, M. K., O'Huallachain, M. E., Lince, M. T., Blankenship, R. E., Beatty, J. T., and Touchman, J. W. 2007. The complete genome sequence of *Rosebacter denitrificans* reveals a mixotrophic rather than photosynthetic metabolism. J. Bacteriol. **189**:683-690.

Takemoto, J., and Lascelles, J. 1973. Coupling between bacteriochlorophyll and membrane protein synthesis in *Rhodopseudomonas sphaeroides*. Proc. Natl. Acad. Sci. USA. **70**:799-803.

Tauschel, H.-D., and Drews, G. 1967. Thylakoidmorphogenese bei *Rhodopseudomonas palustris*. Arch. Mikrobiol. **59**:381-404.

Ueki, T., Kataoka, M., and Mitsui, T. 1976. Structural order in chromatophore membranes of *Rhodospirillum rubrum.* Nature. **262**:809-810.

Vatter, A. E., and Wolfe, R. S. 1958. The structure of photosynthetic bacteria. J. Bacteriol. **75**:480-488.

van der Horst, M. A., and Hellingwerf, K. J. 2004. Photoreceptor proteins, "Star actors of modern times": A review of the functional dynamics in the structure of representative members of six different photoreceptor families. Acc. Chem. Res. **37**:13-20.

Wada, K., Wada, Y., Ishibashi, F., Gojobori, T., Ikemura, T. 1992. Codon usage tabulated from the GenBank genetic sequence data. Nucleic Acid Res. **20**:2111-2118.

Walz, T., and Ghosh, R. 1997. Two-dimensional crystallization of the light-harvesting I-reaction centre photounit from *Rhodospirillum rubrum*. J. Mol. Biol. **265**:107-111.

Wang, G.-S., Grammel, H., Abou-Aisha, K., Sagesser, R., and Ghosh, R. 2012. High-level production of the industrial product, lycopene, using the photosynthetic bacterium, *Rhodospirillum rubrum*. Appl. Environ. Microbiol. **78:**7205-7215.

West, A. H., and Stock, A. M. 2001. Histidine kinases and response regulator proteins in two-component signalling systems. Trends Biochem. Sci. **26**:369-376.

Whittenbury, P. B., and Sistrom, W. R. 1964. The preparation and properties of bacterial chromatophore fractions. J. Cell Biol. **23**:135-150.

Whittenbury, R., and McLee, A. G. 1967. *Rhodopseudomonas palustris* and *Rh. Viridis*-photosynthetic budding bacteria. Archiv für Mikrobiologie. **59**:324-334.

Wilson, A., Punglnell, C., Gall, A., Bonetti, C., Alexandre, M., Routaboul, J.-M., Kerfeld, C. A., van Grondelle, R., Robert, B., Kennls, J. T. M., and Kirilovsky, D. 2008. A photoactive carotenoid protein acting as light intensity sensor. Proc. Natl. Acad. Sci. USA. **105**:12075-12080.

Wilson, A., Punginelli, C., Couturier, M., Perreau, F., and Kirilovsky D. 2011. Essential role of two tyrosines and two tryptophans on the photoprotection activity of the orange carotenoid protein. Biochim. Biophys. Acta. **1807**:293-301.

Wu, Y. P., and Krogmann, D. W. 1997. The orange carotenoid protein of *Synechocystis* PCC 6803. Biochim. Biophys. Acta.**1322**:1-7.

Yagi, T., Yano, T., Di, B., and Matsuno, Y. 1998. Procaryotic complex I (NDH-1), an overview. Biochim. Biophys. Acta. **1364**:125-133.

Yamono, S., Ishii, T., Nakagawa, M., Ikenaga, H., and Misawa, N. 1994. Metabolic engineering for production of β-carotene and lycopene in *Saccharomyces cerevisiae*. Biosci. Biotechnol. Biochem. **58**:1112-1114.

Yan, G.-L., Wen, K.-R., Duan, C.-Q. 2012. Enhancement of β-carotene production by over-expression of HMG-CoA reductase coupled with addition of ergosterol biosynthesis inhibitors in recombinant *Saccharomyces cerevisiae*. Curr. Microbiol. **64**:159-163.

Yang, H. Q., Tang, R. H., Cashmore, A. R. 2001. The signalling mechanism of *Arabidopsis* CRY1 involves direct interaction with COP1. Plant Cell. **13**:2573-2587.

Yoon, S.-H., Lee, Y.-M., Kim, J.-E., Lee, S.-H., Lee, J.-H., Kim, J.-Y., Jung, K.-H., Shin, Y.-C., Keasling, J. D., and Kim, S.-W. 2006. Enhanced lycopene production in *Escherichia coli* engineered to synthesize isopentenyl diphosphate and dimethylallyl diphosphate from mevalonate. Biotechnol. Bioeng. **94**:1025-1032.

Yoon, S.-H., Park, H.-M., Kim, J.-E., Lee, S.-H., Choi, M.-S., Kim, J.-Y., Oh, D.-K., Keasling, J. D., and Kim, S.-W. 2008. Increased β-carotene production in recombinant *Escherichia coli* harbouring an engineered isoprenoid precursor pathway with mevalonate addition. Biotechnol. Prog. **23**:599-605.

Yoon, S.-H., Lee, S.-H., Das, A., Ryu, H.-K., Jang, H.-J., Kim, J.-Y., Oh, D.-K., Keasling, J. D., and Kim, S.-W. 2009. Combinatorial expression of bacterial whole mevalonate pathway for the production of β-carotene in *E. coli*. J. Biotechnol. **140**:218-226.

Young, D. A., Bauer, C. E., Williams, J. C., and Marrs, B. L. 1989. Genetic evidence for superoperonal organization of genes for photosynthetic pigments and pigment-binding proteins in *Rhodobacter capsulatus*. Mol. Gen. Genet. **218**:1-12.

Yu, B., Lydiate, D. J., Schäfer, U. A., and Hannoufa, A. 2010. Characterization of a β-carotene hydroxylase of *Adonis aestivalis* and its expression in *Arabidopsis thaliana*. Planta. **226**:181-192.

Zannoni, D. 1995. Aerobic and anaerobic electron transport chains in anoxygenic phototrophic bacteria. pp.949. *In* R. E. Blankenship, M. T. Madigan, and C. E. Bauer (ed.), "Anoxygenic photosynthetic bacteria". Kluwer Academic Publishers, Dordrecht, The Netherlands.

Zeilstra-Ryalls, J., and Kaplan, S. 1998. Role of the *fnrL* gene in photosystem gene expression and photosynthetic growth of *Rhodobacter sphaeroides* 2.4.1. J. Bacteriol. **180**:1496-1503.

Zouni, A., Witt, H.-T., Kern, J., Fromme, P., Krauβ, N., Saenger, W., and Orth, P. 2001. Crystal structure of photosystem II from *Synechococcus elongatus* at 3.8 Å resolution. Nature. **409**:739-743.

Zsebo, K. M., and Hearst, J. E. 1984. Genetic-physical mapping of a photosynthetic gene cluster from *Rb. capsulatus*. Cell. **37**:937-947.

Acknowledgements

Curriculum Vitae

Name: Guoshu Wang

Date of birth: 12.12.1981

Place of birth: ZiBo, Shandong Province, China

Nationality: China

Work address: Department of Bioenergetics,

Institute of Biology, University of Stuttgart

Stuttgart, Germany

Home address: Pfaffenwaldring 42B 104, Stuttgart

E-mail: guoshu.wang@bio.uni-stuttgart.de

Education

Sep. 2009-present: PhD student, major in Bioenergetics at the University of Stuttgart, Germany.

PhD thesis: Metabolic engineering of the photosynthetic bacterium *Rhodospirillum rubrum* to produce industrially interesting plant carotenoids at high level and low cost.

2008-2009: PhD course study, major in Biochemical Engineering at the Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian, China

2005-2008: M.Sc. in Hydrobiology at the Dalian Ocean University, Dalian, China.

Thesis: Studies on seedling incubation of filaments from crusts of *Grateloupia filicima* and *Chondrus ocellatus.*

Prize: Outstanding postgraduate in 2008.

2001-2005: B.Sc. in Biotechnology at the College of Life Science, Qufu Teachers University, China.

Thesis: Comparison of four methods to extract DNA from wheat.

Awards: Outstanding college graduate in 2005.

1998-2001: Senior high school in ZiBo, Shandong, China

1994-1998: Middle School in Zhoucun, ZiBo, Shandong, China

1990-1994: Elementary School in Zhoucun, ZiBo

1988-1990: Preschool in Zhoucun, ZiBo