Genome sequencing of *Leptolyngbya* Heron Island, 2Å crystal structure of phycoerythrin and spectroscopic investigation of chromatic acclimation

by

Robin Paul

A Dissertation presented in Partial Fulfillment

of the requirements for the Degree

Doctorate of Philosophy

Approved April 2014 by the
Graduate Supervisory Committee:

Dr. Petra Fromme, Chair
Dr. Alexandra Ros
Dr. Robert Roberson

ARIZONA STATE UNIVERSITY

April, 2014

ABSTRACT

Photosynthesis is the primary source of energy for most living organisms. Light harvesting complexes (LHC) play a vital role in harvesting sunlight and passing it on to the protein complexes of the electron transfer chain which create the electrochemical potential across the membrane which drives ATP synthesis. phycobilisomes (PBS) are the most important LHCs in cyanobacteria. PBS is a complex of three light harvesting proteins: phycoerythrin (PE), phycocyanin (PC) and allophycocyanin (APC). This work has been done on a newly discovered cyanobacterium called *Leptolyngbya Heron Island* (*L.HI*).

This study has three important goals:

1) Sequencing, assembly and annotation of the *L.HI* genome – Since this is a newly discovered cyanobacterium, its genome was not previously elucidated. Illumina sequencing, a type of next generation sequencing (NGS) technology was employed to sequence the genome. Unfortunately, the natural isolate contained other contaminating and potentially symbiotic bacterial populations. A novel bioinformatics strategy for separating DNA from contaminating bacterial populations from that of *L.HI* was devised which involves a combination of tetranucleotide frequency, %(G+C), BLAST analysis and gene annotation.

2) Structural elucidation of phycoerythrin - Phycoerythrin is the most important protein in the PBS assembly because it is one of the few light harvesting proteins which absorbs green light. The protein was crystallized and its structure solved to a resolution of 2Å.

This protein contains two chemically distinct types of chromophores: phycourobilin and phycoerythrobilin. Energy transfer calculations indicate that there is unidirectional flow of energy from phycourobilin to phycoerythrobilin. Energy transfer time constants using Forster energy transfer theory have been found to be consistent with experimental data available in literature.

3) Effect of chromatic acclimation on photosystems – Chromatic acclimation is a phenomenon in which an organism modulates the ratio of PE/PC with change in light conditions. Our investigation in case of *L.HI* has revealed that the PE is expressed more in green light than PC in red light. This leads to unequal harvesting of light in these two states. Therefore, photosystem II expression is increased in red-light acclimatized cells coupled with an increase in number of PBS.

ACKNOWLEDGEMENTS

This work has been possible as a result of the support of many people at the Arizona State University. I would like to thank Dr. Petra Fromme for giving me this wonderful opportunity to work in this project. I would also like to thank her for her support, suggestions and constructive criticisms on my research. She also helped me in critiquing and reviewing the manuscripts which have been either accepted or submitted in reputed peer-reviewed journals.

There have been some other people without whom carrying out this work would have been either very difficult or impossible:

1) Dr. Robert Jinkerson, Department of Chemistry and Geochemistry, Colorado School of mines – He proposed the idea for identifying the genomic scaffolds of *Leptolyngbya Heron Islands* from the raw Illumina sequencing data.

2) Dr. Raimund Fromme, Department of Chemistry and Biochemistry, Arizona State University – He helped in devising the crystal trays for crystallizing phycoerythrin. He also helped in analyzing the phycoerythrin diffraction patterns and the X-ray structure determination.

3) Dr. Joshua Labaer, Jason Steel and Kristina Buss, Biodesign Institute, Arizona State University – Dr. Joshua Labaer allowed us to use the Illumina sequencer for sequencing the *Leptolyngbya Heron Island* genome. Jason Steel and Kristina Buss were involved in preparing the *Leptolyngbya Heron Island* genomic DNA for sequencing.

I would like to thank all members of the Fromme lab for their help and support. This work was supported in part by the Galvin Chair grant awarded to Dr. Petra Fromme.

I would like to thank Dr. Oliver Beckstein, Department of Physics, Arizona State University for introducing me to theoretical sciences. His course on protein simulations taught me all the basics of how to manipulate PDB files. He also taught me python programming which I used widely for analyzing genomic and protein crystal structure data.

I would also like to thank my parents and my sister for their help and support throughout my academic career. My mother herself being a research scientist in the field of microbiology also gave me necessary guidance and also critiqued my research work.

I would also like to thank my wife Anasuya Pal who is also a graduate student in the department of Chemistry & Biochemistry for her support and constructive criticisms on my research. Her mental and monetary support has also enabled me to concentrate more in my research during the final semester of my Ph.D. I would also like to thank my mother-in-law for her support during the final days of my Ph.D.

TABLE OF CONTENTS

Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

ABBREVIATIONS

PE: Phycoerythrin

PC: Phycocyanin

APC: Allophycocyanin

PBS: Phycobilisomes

NGS: Next-generation sequencing

PS I: Photosystem I

PS II: Photosystem II

L.HI: *Leptolyngbya* Heron Island

CA: Chromatic acclimation

PUB: Phycourobilin

PEB: Phycoerythrobilin

PCB: Phycocyanobilin

LGT: Lateral gene transfer

PMSF: Phenylmethanesulfonyl fluoride

IEC: Ion-exchange chromatography

MWCO: Molecular weight cut-off

KDa: Kilo Daltons

MES: (2-[N-morpholino]ethanesulfonic acid)

MCM : MES, Calcium Chloride, Magnesium Chloride

bp: Base-pair

RMSD: Root mean square deviation

qPCR: Quantitative polymerase chain reaction

qRT-PCR: Quantitative reverse transcriptase - polymerase chain reaction

$NADP^+$: Nicotinamide adenine dinucleotide phosphate

NADPH: Nicotinamide adenine dinucleotide phosphate (reduced)

ATP: Adenosine triphosphate

ADP: Adenosine diphosphate

CHAPTER 1: INTRODUCTION

1.1 Photosynthesis

Photosynthesis is the most important source of energy for all living organism on earth. All living organisms either directly capture energy using photosynthesis or consume products of others which do so. Among organisms which can directly capture light energy using oxygenic photosynthesis for production of biomass are cyanobacteria, red algae and plants. The energy captured is converted in the light reactions to drive synthesis of the high energy products NADPH and ATP. These are used in the dark reactions to fix CO_2 from the air to form complex carbohydrates [1].

In cyanobacteria and plants the site of the light reactions of the photosynthetic activity is located in a membrane system called thylakoids [1]. The space outside the thylakoid membrane is called stroma and the space inside it is called the lumen [1, 2]. The function of the light reactions is to capture the light energy from the sun and use it to create a transmembrane electrochemical gradient that drives synthesis of ATP from ADP and P_i [3, 4] and reduction of $NADP^+$ to NADPH [5]. The dark reactions take place in the stroma and convert CO_2 into carbohydrates using the ATP and NADPH synthesized in the light reactions [6, 7].

The light reactions consist of the reactions which occur in the large protein complexes which form the photosynthetic electron transfer chain (Figure 1-1). Energy is captured

by the light harvesting complexes such as phycobilisomes (PBS) and transferred to photosystem II and photosystem I.

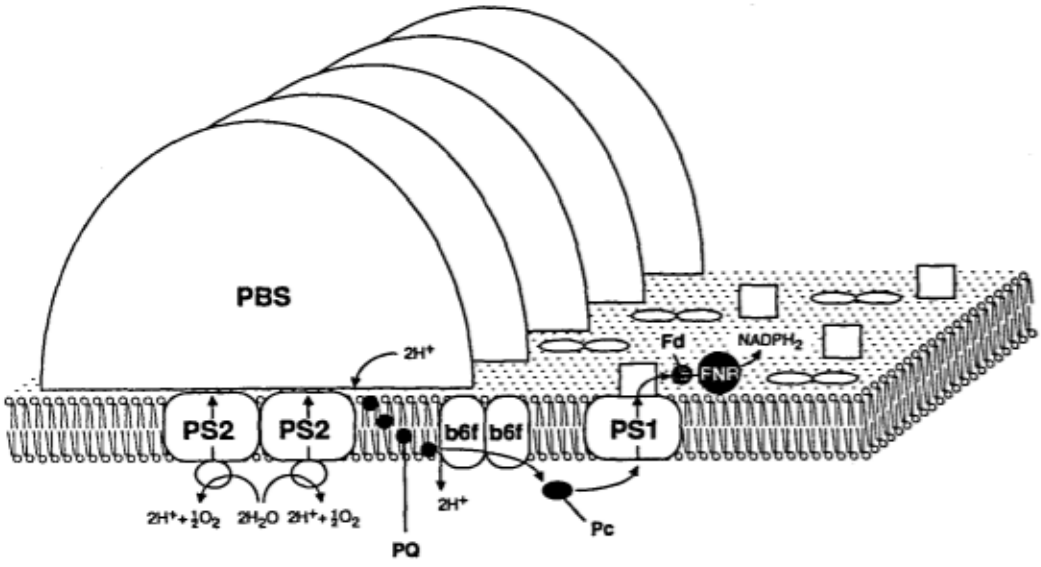


FIGURE 1-1

Diagram of the electron transfer chain in photosynthesis. The figure is taken from reference [8].

In photosystem II, light-driven transmembrane charge separation from water to platoquinone takes place. In this process, 2 H_2O molecules are oxidized to form O_2, $4e^-$ and $4H^+$ ions. The protons are released into the lumen thereby contributing to the formation of a proton gradient across the thylakoid membrane. The electrons are transferred via a chain of electron carriers to plastoquinone PQ [9] . After two charge separation events PQ takes up two protons and leave the binding site as plastoquinol PQH_2, which is subsequently replaced by a plastoquinone from the PQ-redox pool [9]. PQH_2 transfers the electron to the cytochrome b_6f complex [10] which in turn transfers the electrons to photosystem I using the soluble electron transfer protein Plastocyanin. In

this process the cytochrome b_6f complex also pumps protons across the membrane thereby contributing to the enhancement of the proton gradient. Photosystem I catalyzes the second light-driven reaction [5]. It provides the electrons for the reduction of $NADP^+$ to NADPH via ferredoxin and the ferredoxin-NADP-reductase FNR by accepting light energy from the light harvesting complexes and chlorophylls attached to core antenna part of the protein. The electrochemcial proton gradient created across the thylakoid membrane by the electron transfer chain is utilized by the ATP synthase for the synthesis of ATP.

The overall reaction in photosynthesis is the following:

$$6CO_2 + 12NADPH + 18ATP \longrightarrow C_6H_{12}O_6 + 12NADP + 18ADP + 6H_2O + 18P_i$$

1.2 Phycobilisomes

Phycobilisomes are the most important light harvesting complexes in cyanobacteria and red algae [11]. They consist of water soluble light harvesting proteins which protrude from the thylakoid membrane. Their function is to capture light energy and pass it on to photosystem I and II for charge separation. The phycobilisomes consist of two parts. The central core which consists of allophycocyanin (APC) which form the cylinders from which individual rods protrude out (Figure 1-2). The rods protruding out consist of two other light harvesting proteins: phycoerythrin (PE) and phycocyanin (PC). Phycoerythrin has the most blue-shifted absorption spectrum and absorbs in the range from 490-560 nm [12]. phycocyanin absorbs from 500-650 nm and Allophycocyanin from 640-680 nm

[13]. Therefore, the phycobiliproteins are arranged in the phycobilisomes in such a way such that the most blue-shifted protein is situated right at the top while the most red-shifted protein is situated right at the bottom core of the complex. This helps in capturing maximum amount of energy. The PBS complex is further stabilized by small linker proteins which do not contain chromophores.

1.3 Chromatic Acclimation

Chromatic acclimation (CA) is the phenomenon in which an organism changes the composition of the phycobiliproteins in response to changes in light conditions [14]. In chromatic acclimation, the PE/PC ratio changes with change in light conditions. There are three types of chromatic acclimation responses: Type I cyanobacteria do not show any change in the PBS composition (this is usually not considered a chromatic acclimation response), in the second type (type II) there is increase in PE composition in the PBS in green-light acclimatized cells but no corresponding increase in PC in red-light acclimatized cells. In the third type of response (type III) there is increase in composition of both PE and PC in green and red-light acclimatized cells respectively.

Type III chromatic acclimation (CA) in the cyanobacterium *Fremyella Diplosiphona* has been studied extensively [14]. CA response is dependent upon three regulator proteins (Regulation of chromatic adaptation) RcaE, RcaF and RcaC. The RcaE contains a phytochrome-class chromophore which differentiates green/red light [15]. The RcaE receptor also contains histidine kinase which phosphorylates the protein in red-light

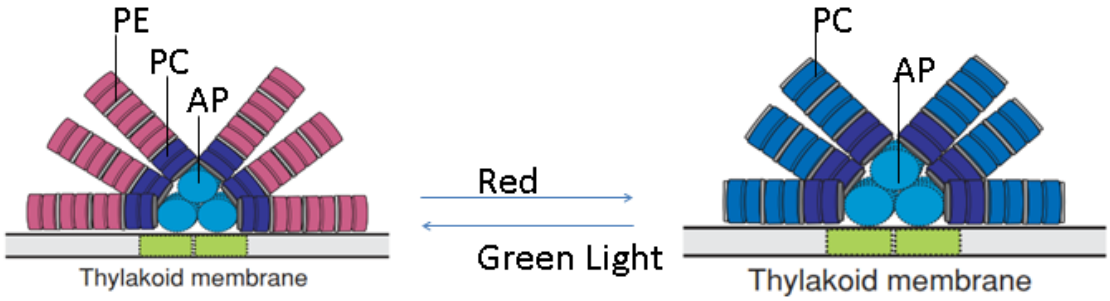


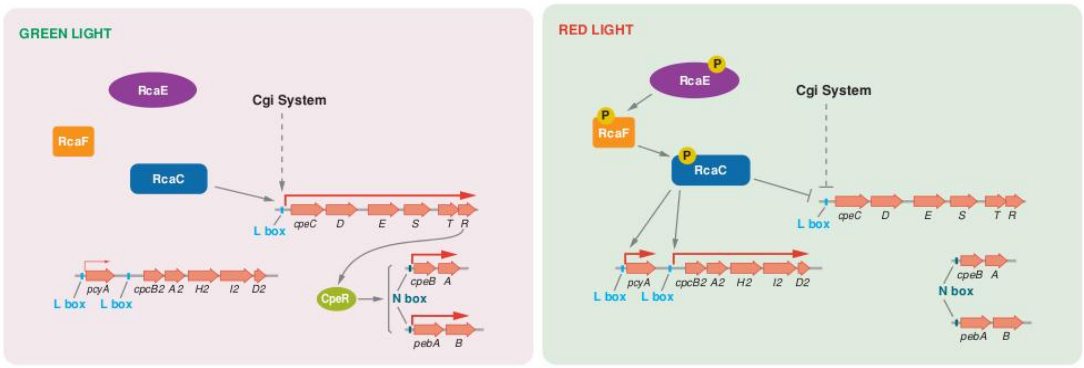
Figure 1-2

Diagram showing the structure of Phycobilisomes. The composition of the phycobilisomes change with change in light depicting the phenomenon of chromatic acclimation. The figure has been adapted from [14].

conditions which in turn phosphorylates RcaF and RcaC. The RcaC in turn induces the operon which helps in the expression of PC (Figure 1-3B). In the case of green-light conditions, the RcaE, RcaF and RcaC receptors are unphosphorylated and therefore induce the operon expressing PE.

1.4 The cyanobacterium - *Leptolyngbya* Heron Island

The rise of cyanobacteria is marked by the change of the atmosphere of the Earth from anoxygenic to oxygenic conditions about 2.5 Billion years ago [16]. Therefore they have been able to evolve over billions of years and are thus able to survive in different types of ecosystems. It is believed that larger eukaryotic organism engulfed prokaryotic cyanobacteria into their own eukaryotic cells [17, 18]. This is known as the endosymbiotic theory. They are therefore considered to be the ancestors of modern-day plants.



(A) (B)

FIGURE 1-3

Mechanism of chromatic acclimation in green and red-light acclimatized cells. (A) In green light the regulators are unphosphorylated and therefore induce expression of phycoerythrin whereas in (B) red light the regulators are phosphorylated and induce expression of phycocyanin. The figure has been adapted from [14].

To study the phycobilisomes and chromatic acclimation in greater detail, we have studied a newly discovered cyanobacterium *Leptolyngbya* Heron Island (*L.HI*). *L.HI* is a section III filamentous cyanobacterium which is filamentous and grow in a single plane (without branching) [19] (Figure 4-4 and Figure 4-9). This species was isolated from the marine ecosystem of Heron Island, Australia. There is large amount of lateral gene transfer (LGT) due to the presence of cyanophages [20, 21] (Phages that specifically infect cyanobacteria). This creates unique combination of genes coming from different organisms making the metabolic pathways and gene structure of such organisms really complex. For instance, it has been found that some cyanophages contain the D1 and D2 core subunits of photosystem II. They degrade the host cyanobacterium's D1 and

D2 genes making them dependent on the phage's D1 and D2 genes thereby facilitating the expression of phage genes [22]. Also many cyanobacteria living in oceans tend to live deep inside the water column where there is very little light available for photosynthesis [23]. These low light adapted cyanobacteria usually employ light harvesting complexes such as the phycobilisomes as well as membrane intrinsic antenna complexes called the Pcb proteins for harvesting of light [24]. As marine cyanobacteria exhibit a large variation of strategies to acclimate to different ecosystems and light conditions they were chosen for study of chromatic acclimation.

CHAPTER 2: GENOME SEQUENCING, ASSEMBLY AND ANNOTATION OF *LEPTOLYNGBYA* HERON ISLAND

2.1 Introduction

Leptolyngbya Heron Island is a newly isolated cyanobacterium from Heron island from the great barrier reef in Australia. Thus its genome has not been previously elucidated. Characterization of the genome is a prerequisite for the carrying out and interpretation of any kind of spectroscopic or crystallographic studies on proteins from this organism. However, since this cyanobacterium has been directly isolated from nature the cell cultures also contain other bacterial non-photosynthetic species, some of them might represent symbionts. Therefore, the natural isolate is not a pure culture but contains contaminants. For sequencing and elucidation of genomes, usually a pure sample of that organism is necessary. Physical methods such as antibiotic treatment for selectively killing the bacterial contaminants in the dark [25], where cyanobacteria do not undergo cell division were attempted but were unsuccessful. Also such methods tend to damage the DNA of the organism under study [25] so this method was not pursued further.

In-silico identification of *L.HI* gene sequences from unaxenic (cell cultures containing contaminating organisms) cultures is a more benign method for elucidating the genome of *L.HI* under these conditions. Several methods such as BLAST, tetranucleotide frequency and %(G+C) analysis have been employed for identifying *L.HI* gene sequences.

2.2 Methods

2.2.1 Isolation of genomic DNA from *Leptolyngbya* Heron Island cells

DNA was isolated from the unaxenic *Leptolyngbya sp.* Heron Island (*L.HI*) cell cultures using the Qiagen Plant mini kit. The 16S rDNA was sequenced using 27f and 1525r primers so as to ensure that a sufficiently strong signal of the 16S rDNA from *L.HI* could be observed. It was assumed that a sample containing a sufficiently strong 16S rDNA signal from *L.HI* would also contain sufficient copies of the entire *L.HI* genome. Once the presence of *L.HI* genome was confirmed, the sample was sonicated to obtain DNA fragments of approximately 100bp. Then appropriate adapters were added to the DNA fragments and were sequenced using the Illumina Hiseq 2000 using the short-insert paired-end sequencing protocol to a depth of coverage of 100X.

2.2.2 Analysis of 16S rDNA using Polymerase Chain Reaction

16S rDNA was amplified from the genomic DNA using Polymerase Chain Reaction (PCR). 27f forward and 1525r reverse primers were used [26] for amplification of 16S rDNA. The protocol for the PCR is as follows:

The PCR product was sequenced and compared using blastn [27] with the NCBI non-redundant gene database. The top 16S rDNA hit having the maximum sequence identity with the query sequence (the input gene sequence) was probed for its organismal origin. It was assumed that this organism was closely related to the organism containing the query gene.

S.No	Steps	Temperature	Time
1.	Initial Denaturation	95 ^{o}C	2 mins
2.	Denaturation Annealing Extension (25 cycles)	95 ^{o}C 47 ^{o}C 72 ^{o}C	30 seconds 30 seconds 90 seconds
3.	Final extension	72 ^{o}C	7 mins

Table 1 - Protocol for polymerase chain reaction.

A number of PCR reactions were performed, and the sample showing the clearest 16S rDNA signal which matched with a 16S rDNA gene of the species *Leptolyngbya* was chosen for sequencing the genome.

2.2.3 Sequencing of *Leptolyngbya* Heron Island genomic DNA using Illumina sequencer

The genomic DNA was broken into approximately 100bp using a sonicator and was subsequently sequenced using the Illumina Hiseq 2000 using the short-insert paired-end sequencing protocol to a depth of coverage of 100X.

2.2.4 Assembling genomic DNA using the Abyss assembler

Genomic reads were assembled using the assembler Abyss version 1.3.5 [28] using a k-value equal to 64. 1071 different scaffolds were obtained. Scaffolds having a length less

than 5000bp were eliminated. Only 401 scaffolds were found to have length greater than 5000bp and were used for the following steps.

2.2.5 Selection of *Leptolyngbya* Heron Island scaffolds using custom BLAST Algorithm

For selection of *L.HI* scaffolds, a BLAST algorithm was devised such that only those scaffolds were selected which had atleast one gene which matched with some gene in the reference genome. For the purpose of this analysis, genes from the genome of *Leptolyngbya* PCC 7375, *Leptolyngbya* PCC 7376 and *Nostoc* Punctiforme were used.

The algorithm described in (see appendix A Script 1) was used for identifying *L.HI* scaffolds using BLAST. The query file contains all the gene sequences from *Leptolyngbya* PCC 7375, *Leptolyngbya* PCC 7376 and *Nostoc* Punctiforme. The reference file contains all the gene sequences identified in the previous genome assembly step. The NCBI blastn [27] is used with an e-value of 0.001. This script uses numpy [29], matplotlib [30] and biopython [31] packages for execution. 123 blast positive scaffolds were selected out of 401.

2.2.6 Tetranucleotide analysis of BLAST positive scaffolds

Tetranucleotide frequencies of BLAST positive scaffolds was calculated using the software TETRA [32]. The tetranucleotide frequencies were calculated using the python script given in Appendix A Script 2.

This script uses principal component analysis [33] to analyze the tetranucleotide frequencies. It uses those axis which have the three highest eigenvalues. All scaffolds are represented by dots in the 3D graph (Figure 2-2).

For selecting scaffolds found to be clustered in the 3D graph, (Appendix A Script 3) was used. This script requires a point and a radius in the 3D space of the principal component axis such that the volume covers all the scaffolds that are clustered together in it. 120 scaffolds were selected out of 123 blast positive scaffolds.

2.2.7 G+C analysis of selected scaffolds from Tetranucleotide analysis

G+C analysis of Tetranucleotide positive scaffolds were further analyzed by measuring %(G+C) of each of the scaffolds. In spite of BLAST and tetranucleotide frequency analysis multiple peaks in the %(G+C) graph were still observed. For isolation of a single G+C peak, (Appendix A Script 4) was used. This script asks the user to enter a cut-off %(G+C) value. Based on this cut-off, the script generates two fasta files, one containing all fasta files below the given cut-off and the other above this cut-off.

2.3 Results and discussions

Leptolyngbya Heron Island (*L.HI*) [6] is a class III filamentous cyanobacterium isolated from the Great barrier reefs in Australia. Since this was isolated directly from a natural habitat the cell cultures also had other contaminating bacterial species. Therefore, before genomic DNA samples were submitted for sequencing, the 16S rDNAs were sequenced to ensure that the cyanobacterial 16S rDNA signal was visible. The presence of 16S

rDNA in a genomic sample gives an approximate estimate of the percentage of genomic DNA from the organism of interest in a given sample [34]. In case of the *L.HI* genomic sample, the 16S rDNA was amplified using PCR and subsequently cloned into *E.Coli*. Each of the 16S rDNA were then sequenced to determine the community profile of the sample. The 16S rDNA of *L.HI* was found along with the those of other heterotrophic bacteria. Unfortunately vast numbers of oceanic bacteria remain unknown, so by comparing sequence homology of 16S rDNA of the heterotrophic bacteria in the NCBI database did not reveal the identity of most of these heterotrophic bacteria. The only heterotroph that could be identified using this analysis was *Mucus* bacterium. The *Mucus* bacterium has been previously found to protect its host from other pathogenic microorganisms [35]. The results of the 16S rDNA analysis are displayed in APPENDIX B.

Identification of organismal origin of DNA reads from next-generation sequencing (new DNA sequencing methods which can sequence prokaryotic genomes in a very short period of time eg. Illumina) is a complicated problem. Therefore no single method such as BLAST analysis, tetranucleotide frequency analysis and %(G+C) analysis can completely identify all the scaffolds which have a common organismal origin. Therefore a correct choice of methods applied in the right combination can solve this complex problem. The order of steps have been described in Figure 2-1.

2.3.1 BLAST analysis

In the BLAST step, scaffolds were selected on the basis of sequence homology with respect to the reference genes. The selection of reference cyanobacterial genomes is

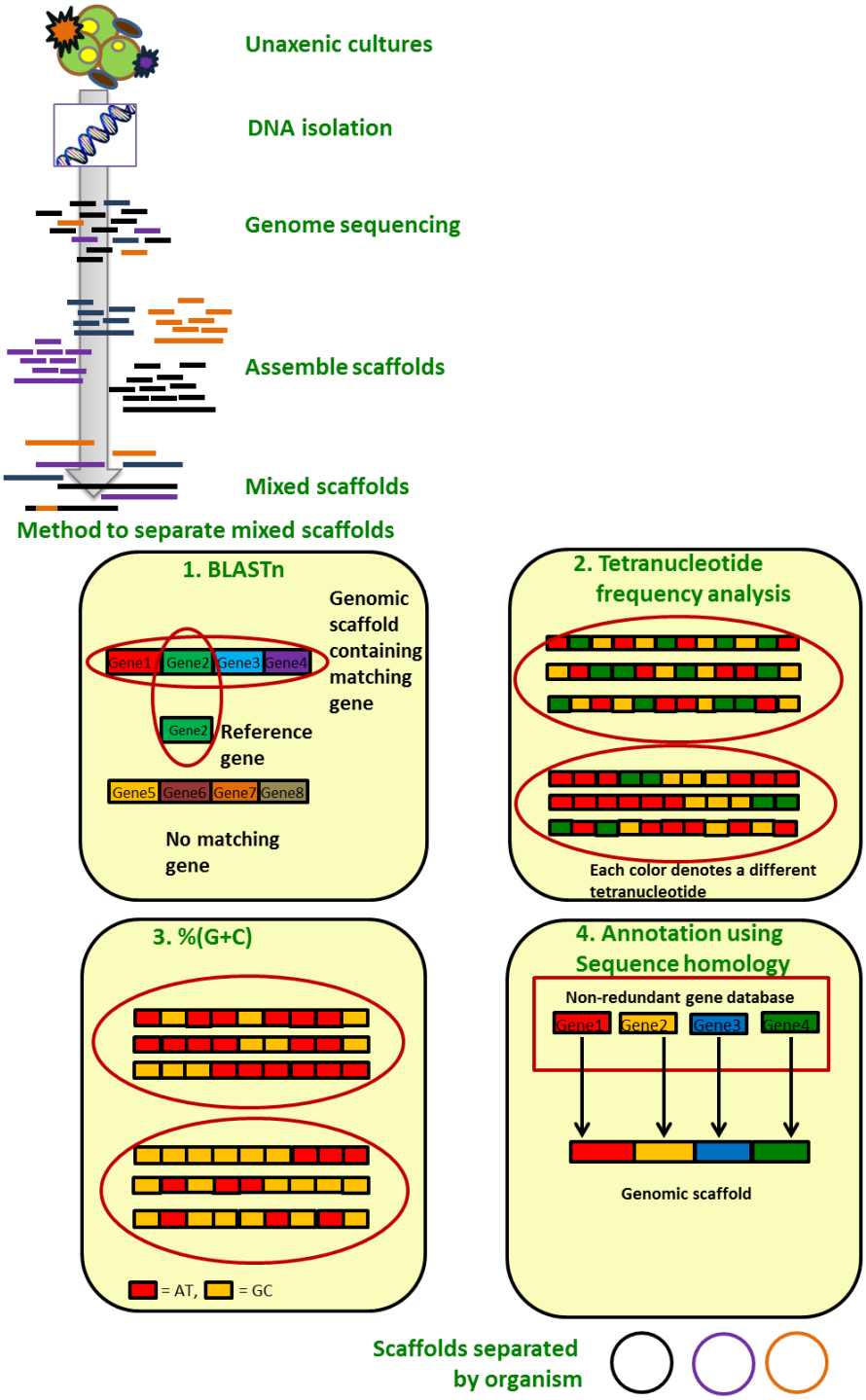


FIGURE 2-1

Flow chart of the protocol used for the identification of *L.HI* gene sequences involving sequencing of genomic reads, assembly of DNA reads, BLAST, tetranucleotide frequency and %(G+C) analysis.

critical to the analysis. For this purpose, the genes from the cyanobacteria *Leptolyngbya* PCC 7375 [36], *Leptolyngbya* PCC 7376 [37] and *Nostoc* Punctiforme [38] were used. The first two cyanobacteria are closely related to the *Leptolyngbya* Heron Island and were therefore selected as a reference genome. *Nostoc* Punctiforme is a somewhat distant relative of *Leptolyngbya* Heron Island. This was used so as to test the selection of scaffolds when using the previous two reference genomes. It was found that almost the same set of scaffolds were selected when using *Nostoc* Punctiforme as the reference genome as in the case of *Leptolyngbya* PCC 7375 and *Leptolyngbya* PCC 7376. This showed that the selection of scaffolds was complete and there were no false negatives.

However, this method is prone to selecting false positives i.e choosing scaffolds which are actually not part of the *L.HI* genome but have been incorrectly selected as *L.HI* scaffolds. This usually can happen because of the presence of common genes such as the 16S rDNA gene which are conserved among all bacterial species. To remove these BLAST false positives, methods which do not simply rely on the sequence homology need to be employed. For this reason, tetranucleotide frequencies and %(G+C) analysis were employed.

2.3.2 Tetranucleotide frequency analysis

Tetranucleotide frequencies have been found to be a good marker for identifying the organismal origin of genome sequences [39]. To interpret this data multivariate analysis such as principal component analysis is essential.

In Figure 2-2, the results indicate that most of the scaffolds are oriented vertically on top of each other. Since those scaffolds which have the same organismal origin have approximately the same tetranucleotide frequencies. Therefore such scaffolds are likely to be closely clustered together in the principal component analysis 3D graph (Figure 2-2). Thus the scaffolds marked in red were selected as those representing the *L.HI* genome. Only 3 scaffolds were found to be separated from the rest of the scaffolds. These were not considered to be part of the *L.HI* genome.

One drawback of tetranucleotide hypothesis method is that it is not accurate for scaffolds of small sizes with less than 5000bp. This is because, the tetranucleotide frequencies calculated are not statistically accurate due to the small size of the scaffold. For this reason, all scaffolds featuring sizes lower than 5000bps were removed prior to calculation of tetranucleotide frequencies.

2.3.3 %(G+C) analysis of tetranucleotide positive scaffolds

As the final step, %(G+C) analysis was used for finding contaminant scaffolds. All gene sequences having a single organismal origin should have similar %(G+C) values. Therefore by using the algorithm (see appendix A Script 4) we can select a threshold %(G+C) value such that all scaffolds above this value can be discarded.

Figure 2-3 shows the %(G+C) graph of the original set of scaffolds obtained after assembly (red). A large portion of the original set of scaffolds was removed by BLAST and tetranucleotide frequency analysis. Figure 2-3 represents a plot of the %(G+C) graph of tetranucleotide positive scaffolds (green).

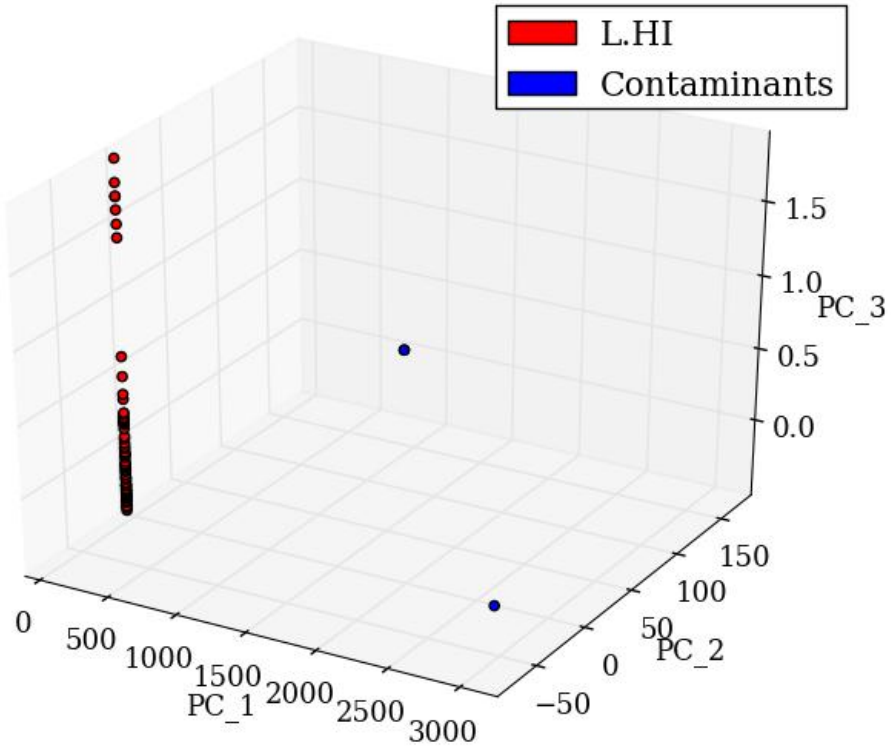


FIGURE 2-2

Principal component analysis of tetranucleotide frequencies obtained from BLAST positive scaffolds. The clustered scaffolds circled, have been chosen as *L.HI* scaffolds.

There are still some minor peaks beside the large *L.HI* peak which were not removed by the first two methods. It was observed that these small peaks contained %(G+C) values which were comparable to some of those scaffolds that were previously removed by BLAST/tetranucleotide frequency analysis. Thus it is quite likely that these scaffolds belonged to the same contaminant genome. All these small peaks had %(G+C) greater than 51. These were therefore removed by using a cut-off of 51%. This resulted in a pure set of scaffolds (Figure 2-3) (yellow).

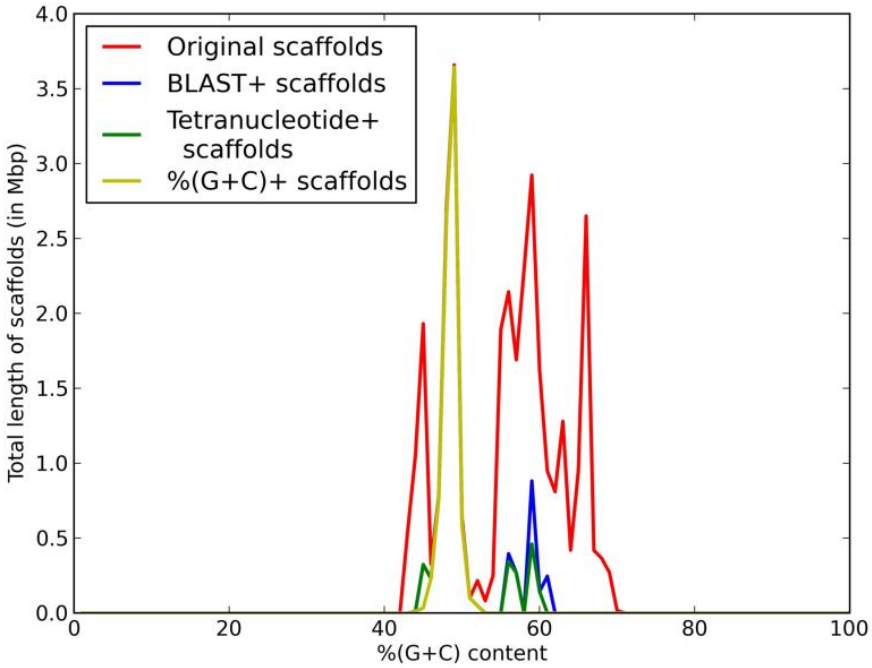


FIGURE 2-3

%(G+C) graph of the original set of scaffolds (red), %(G+C) graph of BLAST positive scaffolds (blue), %(G+C) graph of tetranucleotide positive scaffolds (green) and scaffolds retained after applying %(G+C) cut-off of 51% (yellow).

2.3.4 Annotation of *L.HI* genome

Annotating a genome is critical for deriving any useful information out of it. The prokaryotic genome annotation pipeline (PGAP) was used. Those genes that were not annotated by PGAP were annotated using BLAST. Also in our case annotation is the final proof for identification of correct scaffolds of *L.HI*. To verify the organismal identity of the scaffolds, the genes in each of the scaffold were compared with non-redundant BLAST database in NCBI. The organismal origin of the gene having the highest sequence similarity with the query gene were analyzed.

96% of the genes were found to be of cyanobacterial origin. Out of all the cyanobacterial genes, 78% of the genes matched with a gene found in the *Leptolyngbya* species, 80% of the genes belonged to the group of *Oscillatoria* (the family to which *Leptolyngbya* clade of cyanobacteria belong).

At first, 96% of the genes to be of cyanobacterial origin seem to be a bit low. However, this species of cyanobacteria has been isolated from the Great barrier reefs in Australia. Therefore, its natural habitat is marine water where lot of lateral gene transfer take place due to the presence of phages [20, 40]. Lateral gene transfer (LGT) can account for upto 24% of total number of genes in a prokaryotic genome [41]. As proof of phage infection, we have found upto 9 CRISPR arrays in the *L.HI* genome. CRISPR arrays are a defense mechanism against infecting phages [42]. Small portions of viral DNA are incorporated into the host genome which also contain nucleases downstream (itself an example of LGT), so if the same phage attacks the same host, the host can activate the CRISPR arrays so that nucleases are secreted to cleave the phage DNA. This shows that the remaining 4% of the genome could genuinely belong to the *L.HI* genome although it may not be of cyanobacterial origin.

To ensure that all genuine *L.HI* scaffolds were selected, the number of blast hits for each of the genomic scaffolds obtained after assembly were calculated (see Figure 2-4). The number of hits for each genomic scaffold are indicated in red and the final set of scaffolds is indicated in blue. It is observed that scaffolds having high blast hits were selected as *L.HI* genomic scaffolds and no scaffolds having greater than 20 blast hits remained unselected. This shows that the scaffolds that were left out had either no blast hits or very few blast hits. This suggest that in the initial blast step these

scaffolds got selected due a match with some conserved bacterial gene found in all species of bacteria for eg. 16S rDNA gene.

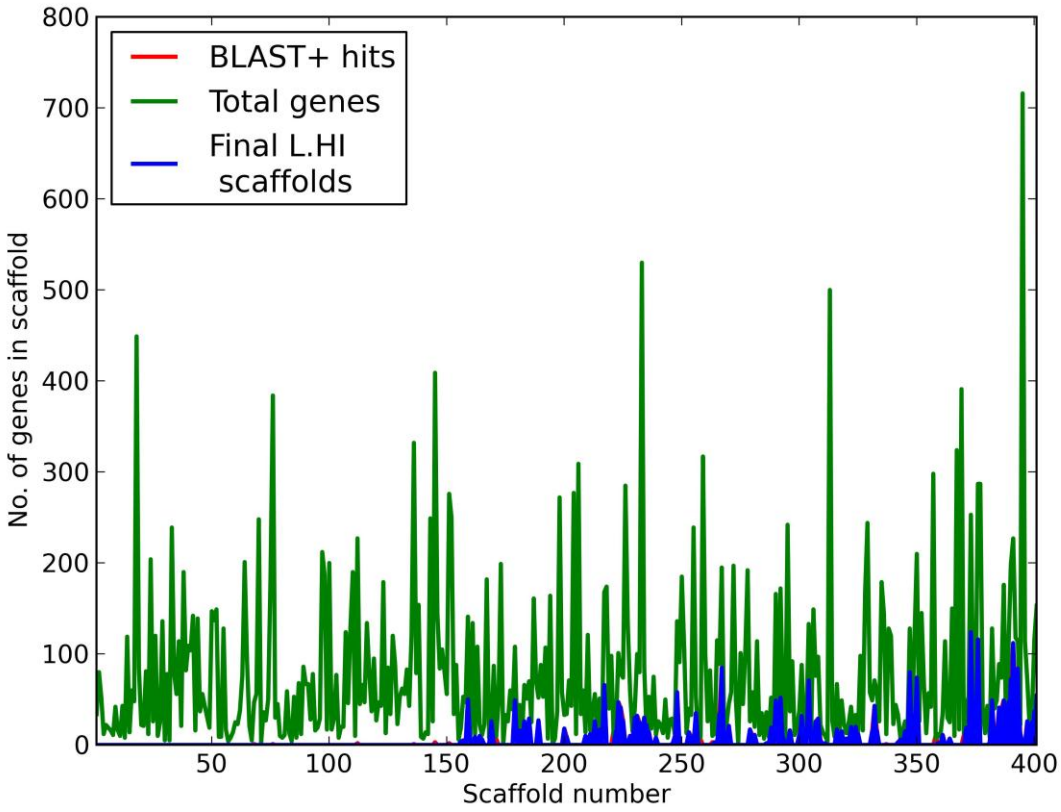


FIGURE 2-4

Histogram of number of BLAST hits for each scaffold obtained after genome assembly for *L.HI*. Red line indicate the BLAST+ scaffolds whereas blue line represents the final set of scaffolds obtained after tetranucleotide and %(G+C) analysis.

Determination of the species of bacterial contaminants in the culture was attempted. For this, the 16S rDNA gene was extracted from the assembled scaffolds and matched using the BLAST algorithm against the non-redundant gene database in NCBI. Most of these genes matched with 16S rDNA gene of uncultured bacteria. One was found to match with

16S rDNA gene of *Mucus bacterium*. This species of bacteria is known to be in symbiosis with other organisms living in the coral reefs [35]. Therefore it is quite likely that a similar species of bacteria is in symbiosis with *L.HI*.

2.4 Conclusion

The genome of *Leptolyngbya* Heron Island has been successfully elucidated. Using a combination of BLAST, tetranucleotide frequencies and %(G+C) analysis the gene sequences of *L.HI* were successfully evaluated. Using sequence homology it was found that 96% of the genes match with some gene found in some other cyanobacterial species. This shows that our identification of *L.HI* gene sequences from the symbiotic pool containing the cyanobacteria and other bacterial contaminants is correct.

CHAPTER 3: 2Å CRYSTAL STRUCTURE OF PHYCOERYTHRIN FROM *LEPTOLYNGBYA* HERON ISLAND

3.1 Introduction

Phycobilisomes in cyanobacteria and red algae consist of three light-harvesting proteins: Phycoerythrin, phycocyanin and allophycocyanin. Phycoerythrin is one of the most important light-harvesting proteins since it absorbs in the green region of the visible-spectrum and thereby fills the "green-gap" [43].

However, till-date there is no crystal structure of phycoerythrin from cyanobacteria published in a peer-reviewed journal. Therefore the crystal structure of phycoerythrin from the red alga *Polysiphonia Urcelolata* was used for comparison since energy transfer studies using femtosecond time-resolved spectroscopy have been carried out on this protein previously [44]. There is only a sequence identity of 65% between PE of *Polysiphonia Urcelolata* (eukaryote) and that of *Leptolyngbya* Heron Island. Also there are no studies of energy transfer done using the Forster resonance energy transfer theory on phycoerythrin. Therefore in this work, energy transfer between chromophores have been estimated using this theory to verify the results obtained using femtosecond time-resolved spectroscopy by Chen et al. [44].

3.2 Methods

3.2.1 Phycoerythrin amino acid sequence determination

DNA was isolated from *L.HI* cells using the Qiagen Plant mini kit. The DNA was sheared into small fragments of 100bp and sequenced using the Illumina Hi Seq 2000.To determine the amino acid sequence of the PE genes, we employed whole genome sequencing using the Illumina Hiseq 2000 platform to generate paired end reads [6], which is described in chapter 2. To acquire the sequence of PE before the whole genome assembly was finished we initially sought only the PE genes from the raw reads. A subset of the reads (30%) were used as a BLASTx [45] query against the PE genes found in *Leptolyngbya* sp. KC 45 (AEO80318.1 and AEO80317.1). Reads identified by BLASTx were filtered by the percent of identical matches (>70%) and alignment length (>10 AA) to the PE queries. Reads remaining were assembled by the program Geneious [46]. Two copies of the PE αβ are present in *L.HI,* which contain 22 nucleotide differences in the CDS (6 in the α, and 16 in the β), the majority of which are synonymous substitutions. However two nucleotide differences in the β subunit result in AA changes: β1-87I and β1-159I or β2-87V and β2-159T (APPENDIX C). This shows that the selection pressure is on conservation of the amino acid sequence of phycoerythrin than nucleotide sequence.

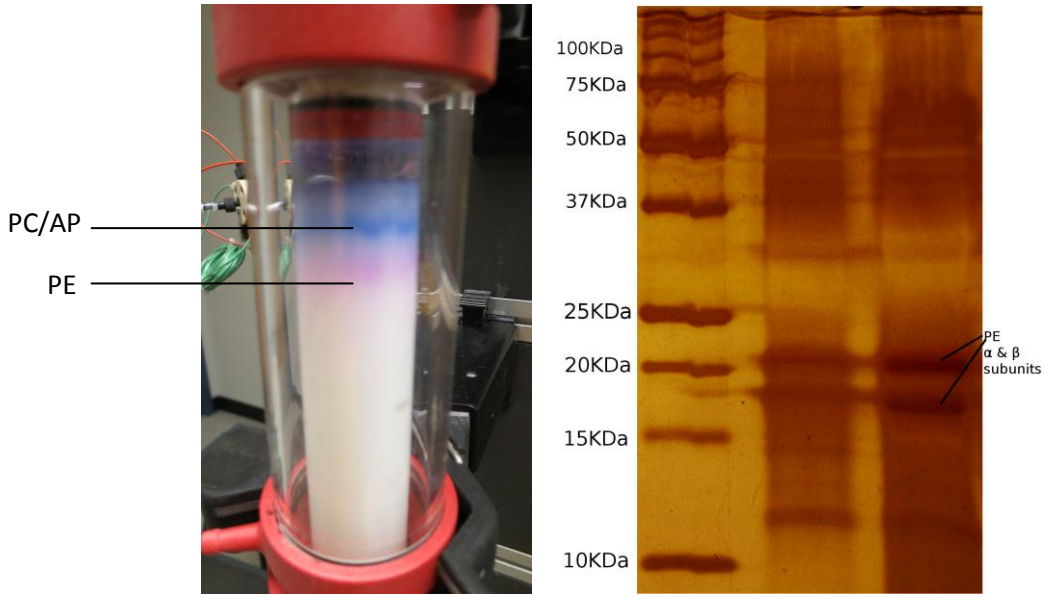
3.2.2 Cell culture, Purification and Crystallization of phycoerythrin

The *L.HI* cells were grown at 28^{o}C in 500ml of sterilized media containing 0.47M NaCl, 10mM $CaCl_2 \cdot 2H_2O$, 9.1mM KCl, 3.1mM $MgSO_4 \cdot 7\ H_2O$ and 1.7mM $NaNO_3$. To this solution, 0.5ml each of 2.68mM Na_2EDTA, 3.1mM citric acid, 0.37M Na_2CO_3 and

0.23M K_2HPO_4. 0.5ml of trace element solution containing 4.5mM H_3BO_3, 10mM $MnCl_2 \cdot 4H_2O$, 65µM $ZnSO_4 \cdot 7H_2O$, 0.82mM $Na_2MoO_4.2H_2O$, 0.24mM $CuSO_4 \cdot 5H_2O$ and 84µM $CoCl_2 \cdot 6H_2O$ were also added. *L.HI* is a filamentous cyanobacterium and prefers to grow on the surface, such as the colonies on rocks or corals in their natural habitat at Heron Island. Therefore, cotton strings were added to the solution to provide surface area for the growth of the cyanobacteria. For the purpose of extracting maximum amount of phycoerythrin from the cells, the cell cultures were exposed to green light. (Refer to chapter 4 for more details about chromatic acclimation)

L.HI cells corresponding to a wet weight of approximately 4-5g were harvested from the *L.HI* cell culture. They were suspended in a buffer containing 50mM MES pH 6.4, 10mM $MgCl_2$ and 10mM $CaCl_2$ (MCM buffer). Cells were lysed using a sonicator in 1 min cycles (at 20W) with gaps of 5 mins in between at a temperature of 4^oC. Since phycobiliproteins are soluble proteins, they can easily be separated from lipids and membrane proteins by centrifugation. The cell lysate was therefore centrifuged at a speed of 30000xg. The supernatant was kept in saturated $(NH_4)_2SO_4$ for two days at a temperature of 4^oC in the presence of PMSF (phenylmethanesulfonyl fluoride). The pellet of the protein was then dialyzed against 2L of buffer using dialysis membranes with a cut-off of 12KDa MWCO (molecular weight cut-off) for a time period of 48hrs. The buffer was replaced thrice at equal intervals during this time frame (see SDS-gel in Figure 3-1B). After dialysis was complete, the protein was further purified by ion-exchange chromatography (IEC). A Q-Sepharose column was equilibrated with 50mM Phosphate buffer. A salt gradient from 0-1M NaCl was used at a flow rate of 1ml/min. About 50mg of protein was loaded onto the column. 100ml of

50mM Phosphate buffer was passed through the column to elute the protein. The eluted protein was collected by means of a fraction collector in volumes of 1ml. The peak fractions of the ion-exchange purified phycoerythrin were concentrated to a concentration of 10mg/ml using 10KDa MWCO filters and rinsed with 50mM MES (2-[N-morpholino]ethanesulfonic acid) pH 6.4 buffer.



(A) (B)

FIGURE 3-1

(A) Picture of the Ion-exchange column used for the purification of phycoerythrin. The ion-exchange column separates the solution containing the impure phycoerythrin (PE) into two separate bands. (B) SDS-gel of phycoerythrin obtained after dialysis of phycobiliproteins in saturated ammonium sulphate.

3.2.3 Crystallization of phycoerythrin

To obtain ordered structures of phycoerythrin for X-ray diffraction, the proteins must be crystallized. For this purpose, hanging-drop method was used. This method is based on Figure 3-2.

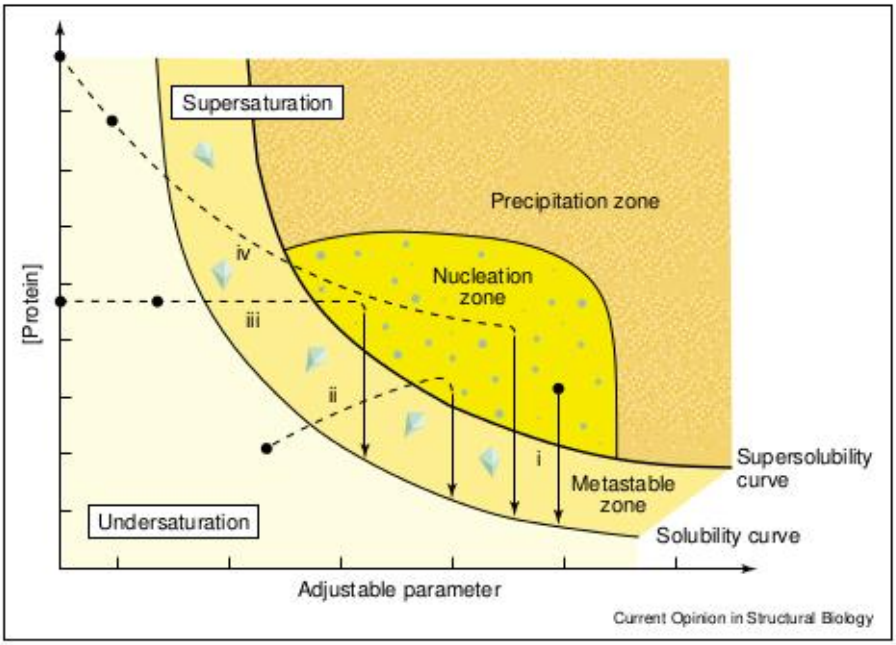


FIGURE 3-2

Phase diagram of protein crystallization. It shows the various phases such as undersaturation, the metastable zone and the precipitation zone. Figure adapted from [47].

In the hanging drop method, a well containing the reservoir solution (with a volume of approximately 500µl containing a precipitant) has a small drop of protein solution on the inside surface of the lid of that well. The protein solution consists of 1-2µl of the protein mixed with the same volume of reservoir solution in a 1:1 ratio. Since the protein solution only consists of 50% of the reservoir solution, due to vapor diffusion there is a net transfer of water from the drop to the reservoir at the bottom of the well. This leads to an increase in the concentration of the precipitant and the protein so that the drop passes onto the nucleation zone to form "seed crystals". As crystals form, the concentration of the protein in the solution decreases and it passes onto the metastable

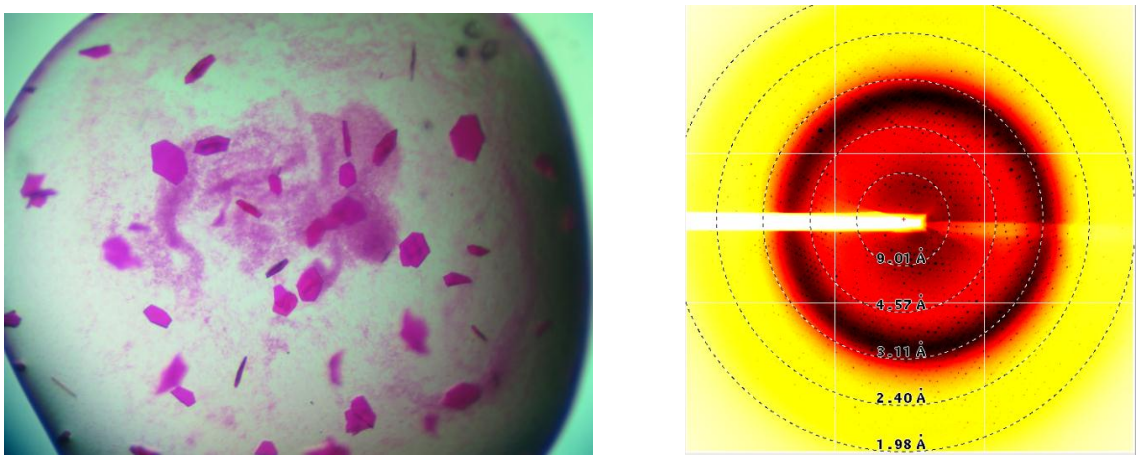
phase. In this phase no new crystals are formed but the liquid protein crystallizes on the "seed crystals". For attaining microcrystals, using suitable precipitant and buffer the protein must enter briefly into the nucleation phase and then must go into the metastable zone rapidly. If the drop enters deep into the nucleation zone or stays in it for a long period of time too many small crystals may form together which aggregate together to give an amorphous precipitate which do not diffract X-rays. On the other hand, if the drop does not enter the nucleation zone at all, no crystals or precipitates will be formed and the protein will remain in solution. Thus a careful choice of buffer and precipitants is required to obtain ordered protein microcrystals.

For obtaining crystals suitable for X-ray diffraction, 500µl crystal trays were used. Plates were set using Sodium dihydrogen phosphate as the reservoir buffer whose concentration was varied from 0.05-0.2M in steps of 0.05M. PEG (Polyethylene Glycol) of molecular weight 3350Da was used as the precipitant. The concentration of PEG 3350 was varied from 12-27% (w/v). Four conditions for freezing of crystals were tested: 35% PEG 3350 with 0.05, 0.1, 0.15 and 0.2M Sodium Dihydrogen Phosphate.

Hexagonal looking crystals were obtained (Figure 3-3a) at 0.2M Sodium Dihydrogen Phosphate in the presence of 15% PEG 3350. Crystals at other conditions had no regular shape or were too small. Crystals frozen with 35% PEG 3350 and 0.15M Sodium Dihydrogen Phosphate were found to yield the highest resolution structure.

3.2.4 Structure determination of phycoerythrin from diffraction pattern

The X-ray diffraction data for phycoerythrin were collected at the APS diffraction facility in Argonne, in Chicago using the AESC315R detector with exposure time of 10s/image



(a) (b)

FIGURE 3-3

(a) Crystals of phycoerythrin obtained at 0.2M Sodium dihydrogen phosphate in the presence of 15% PEG 3350. Crystals were frozen using 35% PEG 3350 and 0.15M Sodium Dihydrogen Phosphate. (b) 2Å diffraction of PE crystals using APS ID-19 light source, distance 300 mm, wavelength 0.979Å (space group: $P2_1$).

(λ=0.979Å) (Figure 3-3b). The X-ray diffraction data was processed using the program XDS [48]. For deriving the initial phases the phycoerythrin structure from *Griffithsia Monilis* [49] (Protein Data Bank filename, 1B8D) was used as a search-model. Phenix Autobuild [50] was used for rebuilding and 50 cycles of refinement were carried out using the Phenix Autobuild [50]. For validating stereochemistry, the program Phenix Validation [51] was used. For elucidating the amino acid sequence of phycoerythrin, the entire genome was sequenced, assembled and annotated [6] as described above. All figures were made either in pymol or VMD [52]. For aligning protein subunits the align function in pymol was used. For aligning chromophores, the pair-fit function in pymol was used.

3.2.5 Energy transfer

Phycoerythrin is usually the top-most component of the phycobilisomes (Figure 1-1) and responsible for capturing the high-energy (blue-shifted) radiation. Therefore, studying energy transfer in this protein is critical to understanding the functioning of the phycobilisomes. Energy transfer was calculated on the basis of Forster's resonance energy transfer theory. It depends on the spectroscopic characteristics of the two chromophores such as the extinction coefficient, overlap integral between the donor and the acceptor, orientation between the chromophores and the distance between them. The formula for calculating energy transfer is as follows [53, 54]:

$$k_{et} = C \times G \times S \times I$$

where constant, $C = 2.72 \times 10^{-28}$ mol. This constant takes into account various other constants such as Avogadro number, index of refraction in water etc [53, 54].

$G = (\kappa_{DA})^2/(R_{DA})^6$

where κ_{DA} = Orientation between donor and acceptor, R_{DA} = Distance between donor and acceptor

$S = (\phi_D/\tau_D)\epsilon_A$

where ϕ_D = Fluorescence quantum yield of chromophores, τ_D = Fluorescence lifetime of chromophore, ϵ_A = maximum visible molar absorptivity (in cm^2mol^{-1})

I = Overlap integral of the given donor and acceptor

The spectroscopic parameters ϕ_D, τ_D, ϵ_A contributing towards S and the overlap integral (I) have been calculated by Jose et al. [12] for PEB and PUB chromophores. These values were used for calculating energy transfer in phycoerythrin from *L.HI*.

3.3 Results

3.3.1 Structure of phycoerythrin

The phycoerythrin heterodimer consists of two α and β subunits which form a αβ heterodimer., which can be regarded as the smallest building block of the PE assembly. The α subunit contains two phycoerythrobilin (PEB) chromophores whereas the β subunit has two PEB chromophores and one phycourobilin chromophore (PUB). The αβ heterodimer trimerizes to from the $(\alpha\beta)_3$ assembly. The $(\alpha\beta)_3$ assembly in turn associates to form the hexamer $(\alpha\beta)_6$, which is shown in (Figure 3-4) [55].

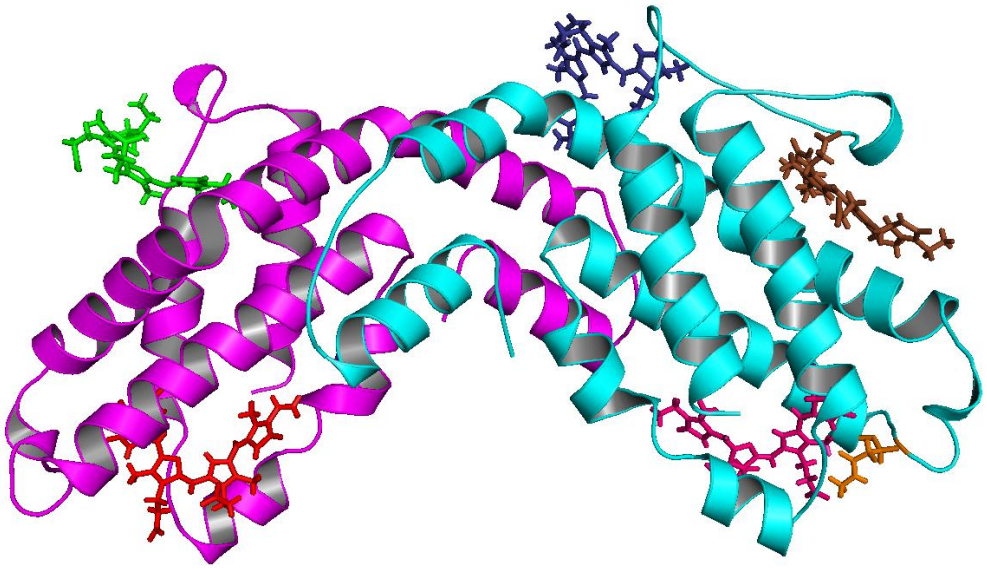
3.3.2 Structure validation

Ramachandran plot is a way to visualize backbone dihedral angles ϕ vs Ψ of amino acid residues in a protein [56]. Usually ϕ and Ψ have only certain combination of values that are allowed. If a certain residue has a disallowed combination of ϕ and Ψ values it suggests of there is some special kind of interaction with some other molecule resulting in a strained geometry. The Ramachandran plot of all the residues have been shown in Figure 3-5. Most residues were found to have allowed torsion angles. Those that did not have allowed torsion angles were found to be glycine residues which do not have a side-chain and can adopt any ϕ and Ψ angles in any of the four quadrants of the Ramachandran plot. Other disallowed conformations were found to be present in the

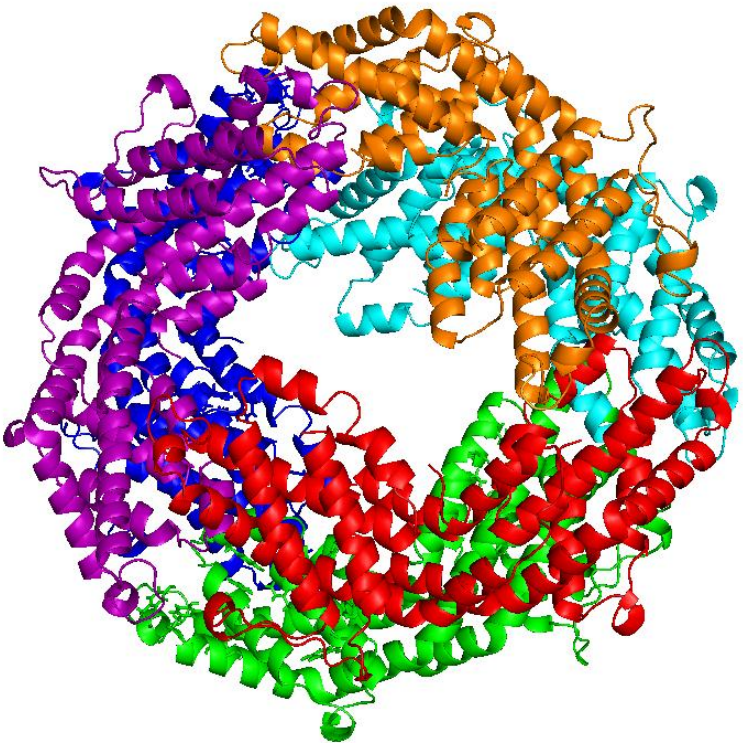
Wavelength (Å)	0.979
Resolution range (Å)	40.21 - 2.018 (2.09 - 2.018)
Space group	P 1 21 1
Unit cell	89.38 108.39 119.21 90 106.54 90
Total reflections	1551123 (130752)
Unique reflections	140484 (13737)
Multiplicity	11.0 (9.5)
Completeness (%)	98.23 (95.86)
Mean I/sigma(I)	40.73 (4.88)
Wilson B-factor	19.14
R-merge	0.7604 (1.522)
R-meas	0.799
CC1/2	0.814 (0.274)
CC*	0.947 (0.655)
R-work	0.1644 (0.2404)
R-free	0.2245 (0.3037)
Number of non-hydrogen atoms	18705
macromolecules	15306
ligands	1290
water	2109
Protein residues	2100
RMS(bonds)	0.013
RMS(angles)	1.59
Ramachandran favored (%)	98
Ramachandran outliers (%)	0.2
Clashscore	7.85
Average B-factor	23.00
macromolecules	21.90
ligands	24.10
solvent	30.20

TABLE 2

Data collection and refinement statistics of phycoerythrin crystal structure.



(A)



(B)

FIGURE 3-4

(A) Structure of the αβ heterodimer of PE. The chain colored in magenta is the α-subunit, the chain colored in cyan is the β-subunit. Color coding of chromophores: α84, PEB (red); α140, PEB (green); β50/61, PUB (brown); β84, PEB (pink); β155, PEB (deep blue) (B) Hexameric structure of phycoerythrin. The αβ heterodimers labeled cyan, blue and green associate to form the upper trimer. Red, orange and purple indicate the αβ heterodimers which associate to form the lower trimer. Both the trimers interact to form the hexamer.

other non-helical parts of the protein which do not have a defined secondary structure. However, one exception is α-144 Methionine (red dot in Figure 3-5). This residue was found to be at close proximity to one of the chromophores. Therefore, this distorted geometry of the chromophore may be due to an interaction with the chromophore. This kind of distorted geometry has been found in the case of phycocyanin (PC) from the cyanobacteria *Fremyella Diplosiphon* and *Cyanidium Caldarium* where the β-Threonine has a distorted geometry which is located close to the β-84 chromophore [57, 58].

Similar to PC, PE also has many post-translational modifications. For example, one of the asparagine residues in PE is methylated (FIgure 3-6). It has been proposed that the methyl groups prevent the access of water to the chromophore thereby increasing efficiency of the chromophore [58].

So far there is no crystal structure available of phycoerythrin from a cyanobacterium. Therefore the sequence of the α and β-subunit was compared with that of eukaryotic phycoerythrin from *Polysiphonia* Urceolata. It was found that the sequence identity between

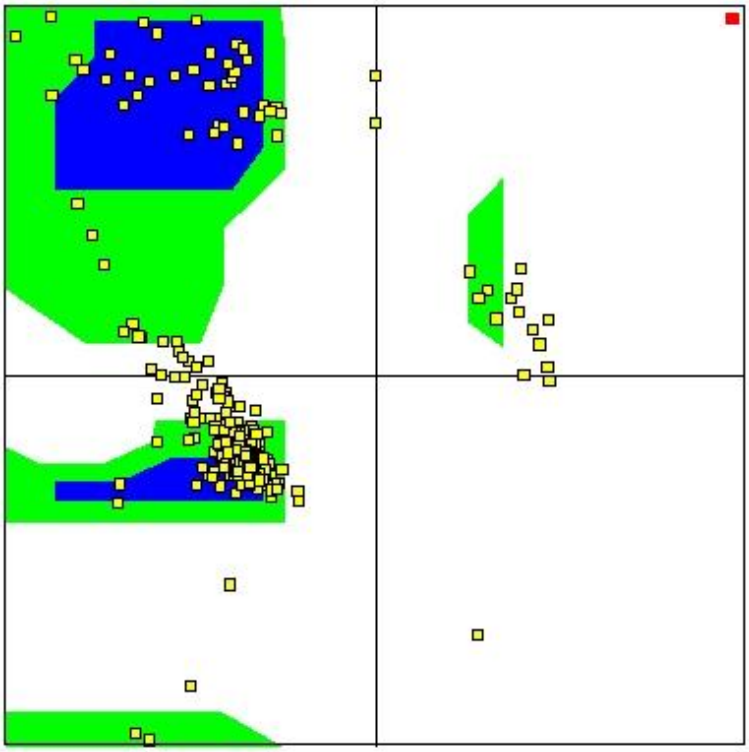


FIGURE 3-5

Ramachandran plot of the αβ heterodimer. The x and y-axis represents ϕ and Ψ angles from -180 to 180 degrees respectively. The blue region represents fully allowed conformations. The green and white regions represent partially and disallowed regions. The red point represents the residue α-144 Met.

PE from *L.HI* and *Polysiphonia* Urceolata was found to be only 67.6% and 64.2% homologous for the α and β-subunit respectively (APPENDIX D). When the structures of the αβ heterodimers of PE from both organisms were aligned from both *L.HI* as well as *Polysiphonia* Urceolata it was found that the root mean square deviation (RMSD) of all the atoms was 2.245Å.

The root mean square deviation (RMSD) of only the Cα atoms was 1.338Å. It has been proposed by Baker et al. [59] that two different proteins having a sequence identity of 50% usually have a RMSD of 1Å. In the case of PE from *L.HI* which has a sequence

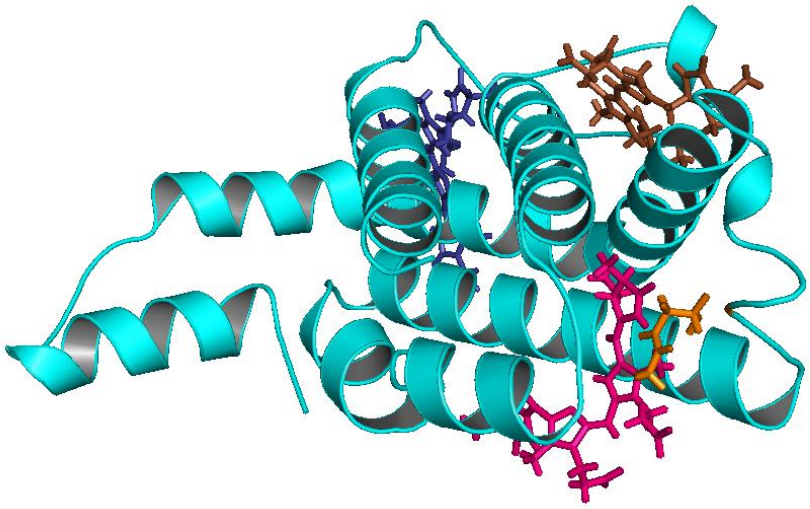


FIGURE 3-6

Diagram showing the methylated asparagine (orange) in the β-subunit. Color coding of chromophores: β50/61, PUB (brown); β84, PEB (pink); β155, PEB (deep blue).

homology of 65.9% (average of both subunits) with *Polysiphonia* Urceolata there is a RMSD of 1.338Å. This suggest that although phycobiliproteins have a common ancestry, they have diverged significantly among different organisms particularly between prokaryotes and eukaryotes. In Figure 3-7, it is observed that the differences in the 3D structure are in those regions which do not have a definite secondary structure (neither α-helix or β-sheet). These regions are located in proximity to the chromophores and may affect their orientation.

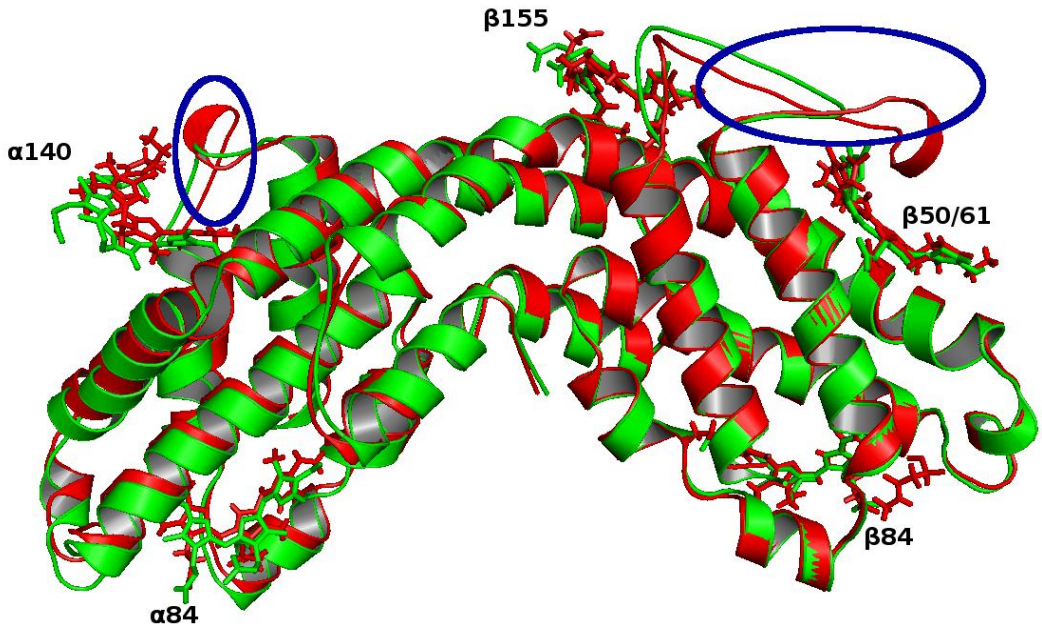


FIGURE 3-7

Diagram showing the overlay of the αβ heterodimer of phycoerythrin from *L.HI* (red) and *Polysiphonia* Urceolata (green). The irregular secondary structures which have maximum structural divergence are highlighted with a blue oval.

3.3.3 Chromophores

There are two different chromophores present in PE: phycoerythrobilin (PEB) and phycourobilin (PUB). The α-subunit contains two PEB's whereas the β-subunit contains two PEB's and one PUB.

As shown in Figure 3-8, the PEB chromophore is chair-shaped in which one of the tetrapyrrole rings points in the direction of the reader, two tetrapyrrole rings are in a plane and the fourth tetrapyrrole ring points in the direction away from the reader. The tetrapyrrole atoms in both the chromophores almost completely align with each other. However, only one of the ethyl side-chains in one of the terminal tetrapyrrole point in opposite directions.

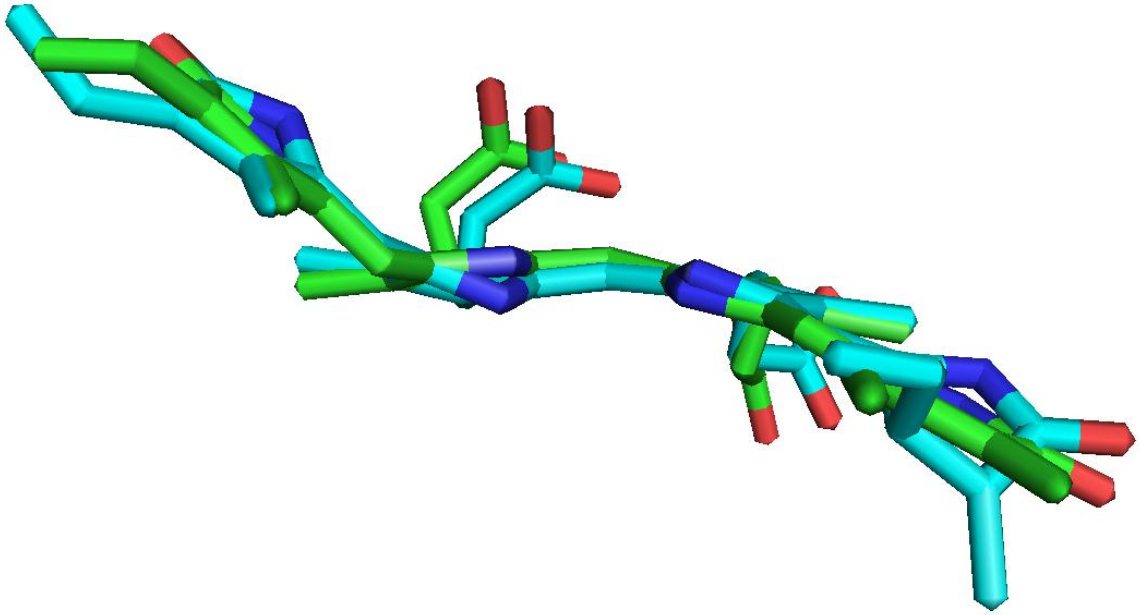


FIGURE 3-8

Diagram showing the overlay of chromophore PEB from PE of *L.HI* (green) and *Polysiphonia* Urceolata (blue) [60]. C-atoms are colored blue/green, N-atoms are colored deep blue, O-atoms are colored red. RMSD between the two PCB chromophores is 0.375Å.

The PUB chromophore of PE from *L.HI* was also aligned with that of PE from *Polysiphonia* Urceolata (Figure 3-9) [60]. In contrast to PEB, the PUB chromophore has a boat shaped structure in which the terminal tetrapyrrole rings were pointing towards the reader. Although the tetrapyrrole rings do not seem to be aligned in PUB as compared to PEB, it must be kept in mind that the resolution of the PE structure from *Polysiphonia* Urceolata is only 2.8Å (compared to a resolution of PE from *L. HI*, making such deviations in the structure of the chromophores much more likely.

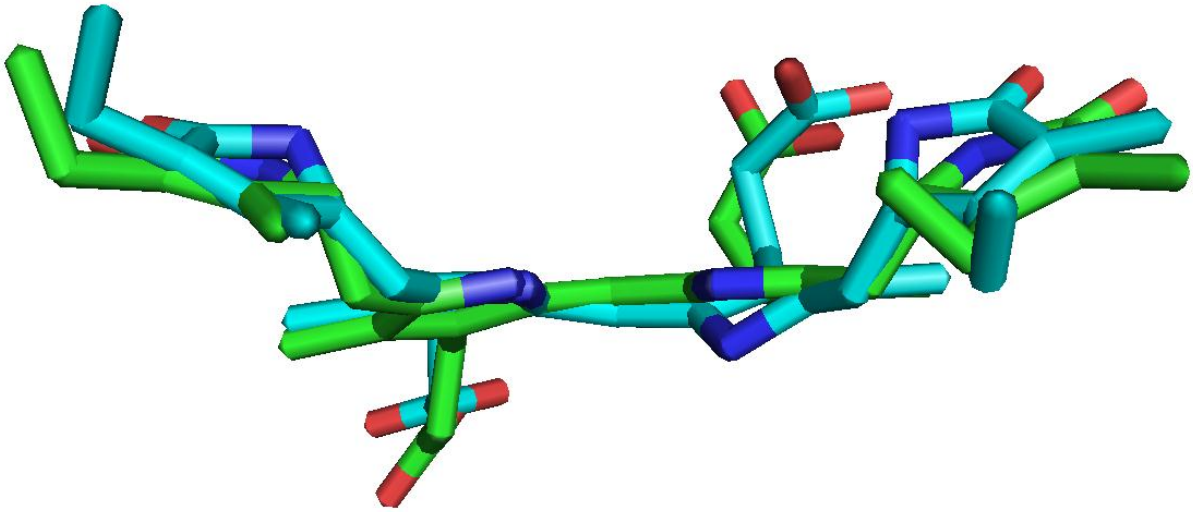


FIGURE 3-9

Diagram showing the overlay of the chromophore phycourobilin from PE of *L.HI* (green) and *Polysiphonia* Urceolata (blue) [60]. C-atoms are colored blue/green, N-atoms are colored deep blue, O-atoms are colored red. RMSD between the two PUB chromophores is 0.553Å.

Thus, in case of both PEB and PUB we find that the main chain tetrapyrrole atoms have the same conformation in three dimensional space as previously reported in eukaryotic PE. There are only some minor changes in the side-chains of the chromophores. This is unlikely to affect the spectroscopic properties of the chromophores in PE.

3.3.4 Energy transfer

Phycoerythrin is a light-harvesting protein which captures sunlight and transfers this energy to other phycobiliproteins (PC and APC) which in-turn pass it on to the core

electron-transfer chain. Therefore understanding of the energy-transfer mechanism in this protein is critical.

The energy transfer in PE is dependent upon the spectroscopic characteristics of the donor and acceptor chromophores, as well as the distance and orientation between them. The 2Å crystal structure provides the 3D atomic coordinates of each of the atoms present in the protein (except the hydrogen atoms) which enables the calculation of distance and orientation between any two given chromophores. However, the crystal structure does not provide any spectroscopic information about the chromophores. This spectroscopic characterization of the chromophores was carried out by Jose et al. [12].

The energy transfer between chromophores within the hexamer was calculated using the custom python script (see appendix A script 5).

The script first of all identifies all the chromophores in the protein. In the PDB file, all the chromophores are labeled either with “PUB”, “CYC” or “BLA”. (The names have not been assigned according to the identity of the chromophores but rather with the stereochemistry of molecules found commonly in the protein data bank. The true identity of the chromophores was confirmed by measuring the absorption spectrum of phycoerythrin (Figure 3-11b)). As shown in the previous section, the PUB and PEB chromophores contain open tetrapyrroles. The coordinates of each of the atoms in the open tetrapyrroles were used for calculating the center of mass of each of the chromophores. Linear curve fits for all the tetrapyrrole atoms were used for used for determining the plane of the chromophore. This line in turn was used for determining the

orientation between any two given chromophores. The formula used for calculating the energy transfer rate has been described previously in the methods section.

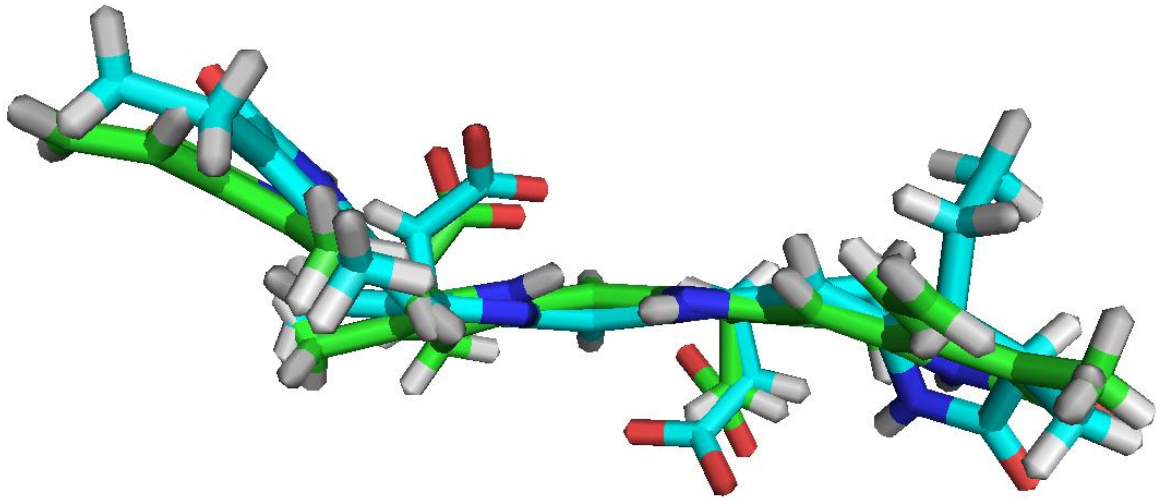
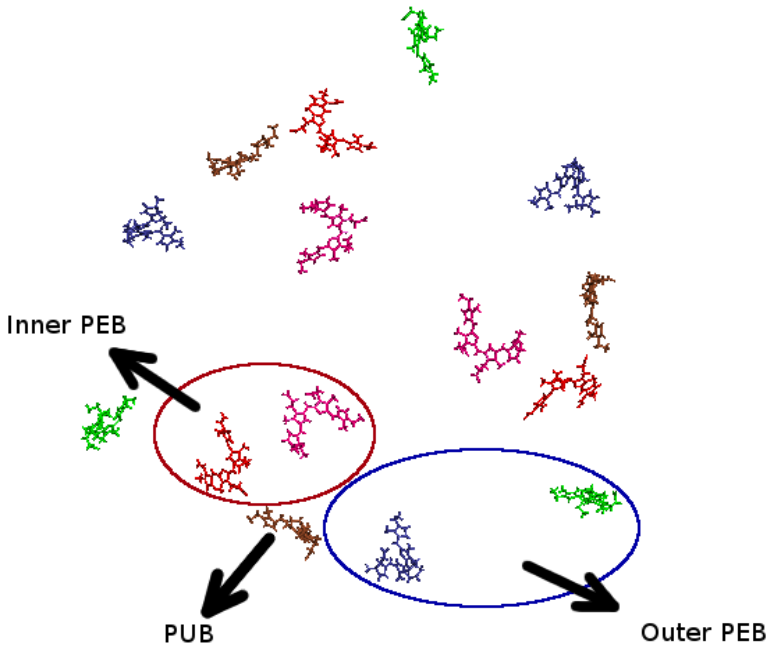


FIGURE 3-10

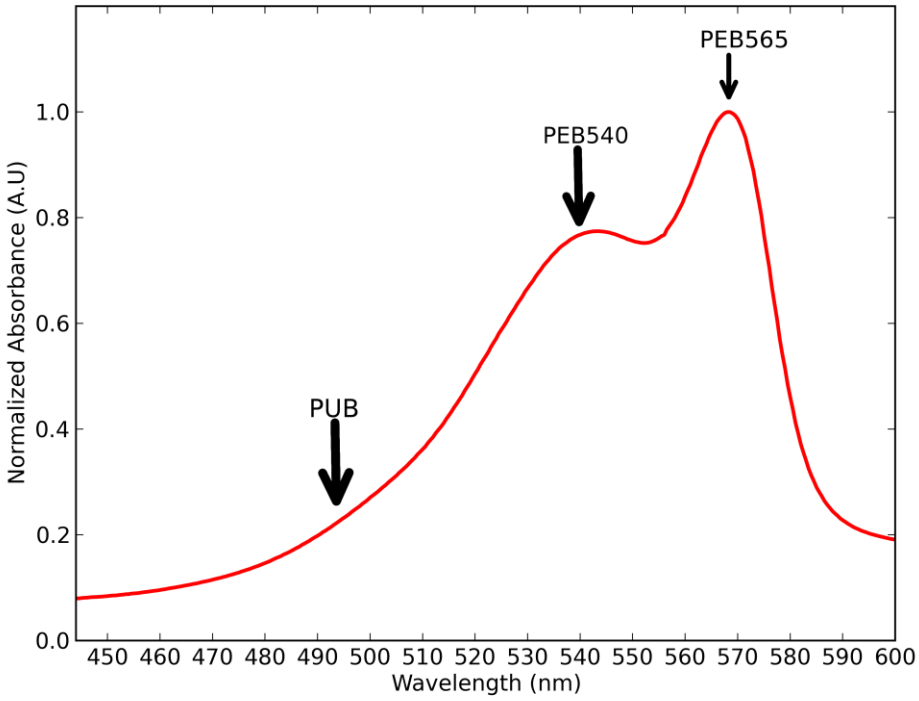
Diagram showing the overlay of the inner (green) and outer (light blue) PEB chromophores present in the hexameric phycoerythrin in *L.HI*.

The energy transfer mechanism is unique in phycoerythrin as compared to other phycobiliproteins such as phycocyanin or allophycocyanin. Phycoerythrin has two different chromophores: phycoerythrobilin (PEB) and phycourobilin (PUB). The PEB has two different λ_{max} depending upon the position of the chromophore in the protein. If it is located in the inner part of the hexameric phycoerythrin then it has a λ_{max} of 565 nm (Figure 3-11) [44]. PEB located at the outer portion of the chromophore have a λ_{max} of 540 nm [44]. In case of the inner chromophore, the outer pyrrole's are closer to the inner

pyrrole's and thus have a greater electron delocalization resulting in a more red-shifted λ_{max} (Figure 3-10). On the other hand in case of the outer chromophore, the outer pyrrole's are further away from the inner pyrrole's resulting in comparatively less electron delocalization resulting in a more blue-shifted λ_{max}. In hexameric phycoerythrin, the α84 and β84 are considered to be the inner PEB chromophores, and α140 and β155 are considered to be the outer PEB chromophores [44].



(a)



(b)

FIGURE 3-11

(a) Diagram depicting the position of chromophores PEB and PUB in trimeric phycoerythrin from *L.HI*. (Pink: β84, PEB , λ_{max} = 565 nm; Red: α84, PEB, λ_{max} = 565 nm; Green: α140, PEB, λ_{max} = 540 nm; Blue: β155, PEB , λ_{max} = 540 nm; Brown: β50/61, PUB, λ_{max} = 495 nm). Red oval contains the inner PEB chromophores and the blue oval contains the outer PEB chromophores. (b) Absorption spectra of phycoerythrin showing the peaks for PUB (β50/61), λ_{max} = 495 nm; PEB565 (α84 and β84), λ_{max} = 565 nm; PEB540 (α140 and β155), λ_{max} = 540 nm.

PUB has a λ_{max} of 498 nm [44] which is also significantly different from the inner and outer PEB λ_{max}. Such great differences in absorption spectra lead to small spectral overlaps between these chromophores. So in comparison to phycocyanin which has only one type of chromophore, phycoerythrobilin has three types of chromophores (two PEBs

absorbing at different wavelengths and PUB). As a result, the energy transfer time constant from PEB to PUB is by order of magnitudes greater than the one from PUB to PEB. This is clearly evident from the overlap integrals of PEB and PUB [12]. The average energy transfer time constants of all chromophore interactions have been listed in

Table 3.

In

Table 3, for any two given chromophores A and B, the energy transfer time constant from (A to B) as well as from (B to A) are given. In such a case, the distance and orientation between the chromophores is the same, the only parameter that varies is the overlap integral. For example between β84 and β50/61, the overlap integral between β50/61 as the donor and β84 as the acceptor is 2.015×10^{-18} cm^4 whereas when β84 is the donor and β50/61 is the acceptor the overlap integral is 0.005×10^{-18} cm^4. Thus there is a 400-fold decrease in the overlap integral which drastically affects the energy transfer time constant. All similar cases where the forward and backward rates are significantly different by orders of magnitude are labeled in bold. Thus in such cases there is only unidirectional flow of energy transfer. This is in stark contrast to phycocyanin in which the overlap integrals between various chromophores are comparable to each other [53].

Donor	Acceptor	Overlap Integral, I ($I \times 10^{18} cm^4$)	Distance (R) (in Å)	Orientation factor, κ^2	Energy transfer (in ps)($1/k_{et}$)
α140	α84	3.822	25.57	1.37	20.3125
α84	α140	2.568	25.57	1.37	28.6175
β84	α84	2.395	20.31	1.46	6.404
α84	β84	2.395	20.31	1.46	6.404
β50/61	β84	2.015	21.41	1.14	**13.184**
β84	β50/61	0.005	21.41	1.14	22347.82
β50/61	α84	2.015	31.90	1.80	**47.373**
α84	β50/61	0.005	31.90	1.80	205760.721
β50/61	α84	2.015	18.96	1.47	**3.866**
α84	β50/61	0.005	18.96	1.47	6560.251
β50/61	α140	6.080	28.13	1.89	**7.47**
α140	β50/61	0.056	28.13	1.89	3602.926
α140	β155	1.962	22.50	2.65	4.28
β155	α140	1.962	22.50	2.65	4.28
β50/61	β155	6.080	23.67	0.78	**4.813**
β155	β50/61	0.056	23.67	0.78	2321.647
β84	β84	2.395	35.37	1.51	160.291
β155	β155	1.962	28.13	0.42	124.758

Table 3

Energy transfer time constants between various chromophores. Energy transfer transitions whose magnitude is significantly greater than that of its reverse transitions have been labeled in bold.

3.3.5 Comparison of energy transfer results with data available in literature

For calculating energy transfer time constants, Chen et al. used femtosecond time-resolved spectroscopic technique to elucidate energy transfer between various chromophores in phycoerythrin from *Polysiphonia* Urceolata [44]. The energy transfer time constants described in Table 3 and Figure 3-12 are approximately of the same magnitude as previously reported by Chen et al. For example, Chen et al. [44] have estimated the energy transfer time constant between the α84 and β84 chromophores to be 1-2ps. Our calculations predict this energy transfer time constant to be 6.4ps. This minor

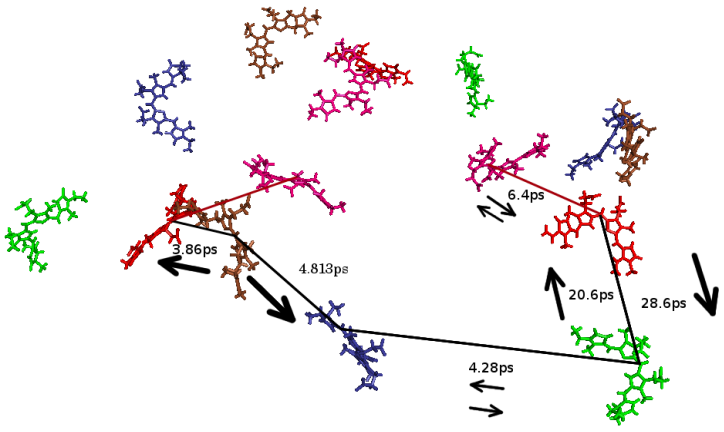


FIGURE 3-12

Figure of phycoerythrin trimer showing the possible energy transfer pathways between different chromophores.

difference can be attributed to small differences in the amino acids between phycoerythrin from *L.HI* and *Polysiphonia* Urceolata. Chen et al. [44] have also proposed an energy transfer time constant from PUB to PEB to be 2.5-3ps. Our calculations have predicted that energy transfer from β50/61 (PUB) to β155 (PEB540) to be 4.8ps whereas that of β50/61 (PUB) to α84 (PEB565) to be 3.866ps. Thus there is very close agreement with the predicted and literature values. Chen et al. [44] have also reported that they did not observe any energy transfer from PEB to PUB. Our calculations indicate that the energy transfer time constant for β155 (PEB540) to β50/61 (PUB) to be 2321ps and that of α84 (PEB565) to β50/61 (PUB) 6560ps. Thus the time required to carry out this transition is too long as compared to their following reverse transition.

3.4 Conclusion

The structure of phycoerythrin from the cyanobacterium *Leptolyngbya* Heron Island has been elucidated to a resolution of 2Å. This light harvesting protein contains of two different types of chromophores: phycoerythrobilin (PEB) and phycourobilin (PUB). Phycoerythrobilin (PEB) has a chair-shaped structure in which the inner two pyrrole rings are in a single plane and one of the outer pyrrole rings are above this plane and the other below it (Figure 3-8). In contrast, phycourobilin (PUB) has a boat-shaped structure in which the terminal pyrrole rings are both on the same side of the plane containing the inner two pyrrole rings.

Energy transfer studies of phycoerythrin reveal that the energy transfer in this protein is significantly different from that other phycobiliproteins such as phycocyanin. Phycocyanin has only about three PCB (phycocyanobilin) chromophores in the αβ heterodimer as compared to five in phycoerythrin which contains three chromophores PEB540 (outer PEB), PEB565 (inner PEB) and PUB. So far there are no energy transfer studies in literature making use of the forster energy transfer theory for calculating energy transfer theory between chromophores in phycoerythrin. In this work the forster energy transfer theory has been used for calculating energy transfer between chromophores. Difference in the magnitude of overlap integrals between forward and reverse energy transfer makes a dramatic difference in the time constant of energy transfer rate between the forward and reverse transitions between PUB and PEB.

There are suttle differences between phycoerythrin from *Polysiphonia* Urceolata and *L.HI*. Sequence homology (APPENDIX D) and structural

overlay (Figure 3-7) suggest that two loops having irregular secondary structure (neither α-helix nor β-sheet) are different in the two αβ heterodimers. But otherwise, the results obtained using the forster energy transfer method on phycoerythrin from *L.HI* are in good agreement with the results obtained using femtosecond time-resolved spectroscopic technique on phycoerythrin from *Polysiphonia* Urceolata.

CHAPTER 4: EFFECT OF CHROMATIC ACCLIMATION ON PHOTOSYSTEM II/PHOTOSYSTEM I RATIO

4.1 Introduction

Marine ecosystems contain a wide diversity of photoautotrophic organisms which have remained largely unexplored so far. These organisms include cyanobacteria, red and green algae, diatoms, plants as well as several other algal groups, which compete with each other for sunlight. *Leptolyngbya* Heron Island *(L.HI)* is a filamentous cyanobacterium newly isolated from Heron Island (Great Barrier Reefs, Australia). While photosynthesis in green algae and plants and some cyanobacteria such as *Prochlorococcus* depends on membrane intrinsic light harvesting antenna complexes (LHCI and LHCII) which contain chlorophyll *a* and *b* as major light absorption pigments, other cyanobacterial species have evolved membrane extrinsic antenna complexes, the phycobilisomes (PBS) which absorb light in the "green-window" of chlorophyll. The PBS consist of phycoerythrin (PE), phycocyanin (PC) and allophycocyanin (APC). PE has a pink color and absorbs light in the green region of the spectrum from 520-590 nm. APC and PC absorb at 620-670 nm and give the organism their characteristic blue-green color. *L.HI* contains PBS which possess APC, PC and PE and can absorb light of the whole visible spectrum and thus confer a selective advantage over plants and green algae, which possess only chlorophyll and thus absorb only blue and red light. A sophisticated process of regulation of light absorption, called the "chromatic acclimation" (CA) (earlier known as "Complementary Chromatic Adaptation", CCA)[14, 61-63] is observed in which

the organism selectively expresses phycobiliprotein genes according to the available light conditions; i.e., it enhances the expression ratio of PE/PC [64] in green light and decreases it in red light [62, 64, 65] (with exception of type IV in which the ratio of the chromophores of PE change and have nothing to do with the PE/PC ratio). This process seems to affect the chromophorylation of the α-subunit of PE. The cyanobacteria can be grouped into four types in this regard. In group I CA response, there is no change in PE/PC ratio under changing light conditions [66]. This is usually not considered a CA response. In group II CA reponse, there is increased expression of PE in green light but no corresponding change for PC in red light (type II CA response)[14, 64]. Cyanobacteria of group III regulate the expression levels of both PE and PC in green and red light respectively (type III CA response)[66]. In type IV CA response, shifts from blue to green light result in change in the ratio of phycoerythrobilin/phycourobilin changes in PE so as to adapt to the given light condition[67-69].

State transitions are another type of light acclimation which involves functional decoupling of the PBS from photosystem II(PS II) and its re-association with photosystem I(PS I) [70]. However state transitions are short-term changes that occur in the order of seconds and do not change the ratio of PS II/PS I [71]. In contrast, the change in stoichiometry between PS II and PS I is a long-term acclimation carried out by oxygenic photoautotrophs in response to the given light conditions [72].

While the response of the PBS composition to CA is studied extensively, the effect of CA on photosystems is still not very well understood and is a field of extensive study [73-76]. One of the most important techniques used for measuring the PS II/PS I

ratio is the steady-state fluorescence at 77K [74]. A typical fluorescence emission spectra constitutes bands at 685, 695 and 710-735 nm which correspond to PS II, the PS II-antenna complex and PS I respectively using excitation wavelength of 440 nm (Soret band of Chl a)[77]. The emission spectral range of λ_{max} for PS I (710 – 735 nm) is due to the varying amounts of "red-shifted" chlorophylls [78] in different species.

The change in stoichiometry between PS II and PS I is a long-term acclimation carried out by oxygenic photoautotrophs in response to the given light conditions. This phenomenon has been observed in prokaryotic cyanobacteria, algae and vascular plants [72], where the spectral composition of light plays an important role in determining the ratio of PS II/PS I [79-84]. In higher plants longer-wavelength light in the visible spectral regimes induces the increased ratio of PS II/PS I whereas this ratio is decreased in shorter-wavelength visible light [72]. In this work I have investigated the CA acclimation responses and their influence on PS I and PS II in *L.HI*.

4.2 Methods

4.2.1 Cell Growth

The *L.HI* cells were grown at 28^{o}C in 500ml of sterilized media containing 0.47M NaCl, 10mM $CaCl_2 \cdot 2H_2O$, 9.1mM KCl, 3.1mM $MgSO_4 \cdot 7\ H_2O$ and 1.7mM $NaNO_3$. To this solution, 0.5ml each of 2.68mM Na_2EDTA, 3.1mM citric acid, 0.37M Na_2CO_3 and 0.23M K_2HPO_4. 0.5ml of trace element solution containing 4.5mM H_3BO_3, 10mM $MnCl_2 \cdot 4H_2O$, 65µM $ZnSO_4 \cdot 7\ H_2O$, 0.82mM $Na_2MoO_4.2H_2O$, 0.24mM $CuSO_4 \cdot 5H_2O$ and 84µM $CoCl_2 \cdot 6H_2O$ were also added. *L.HI* is a filamentous cyanobacterium and prefers to grow on the surface, such as the colonies on rocks or corals in their natural habitat

at Heron Island. Therefore, cotton strings were added to the solution to provide surface area for the growth of the cyanobacteria. The white light acclimatized cells show a brown color, which turns green when the cells were grown under either a red LED λ = 620-630 nm(L620) (purchased from www.electron.com, model no. STRF4-3528-R60-12V) or red filters with broad spectral range of λ = 620-700 nm (F640)(self-constructed). (Pictures of cells grown using red LEDs are shown in Fig. 2; a picture of a cell culture of cells grown with broad spectral range red filters are shown in supplementary information Fig. S2). *L.HI* cells changed their color to red when they were grown under either green LED λ = 585-595 nm (L585) (model no. STRF4-3528-G60-12V) light or a broad green filter with a λ = 420-580 nm (F560). (The absorption spectra of the filters are shown in supplementary information Fig. S1). Cells using either filters or LED lights were grown at an irradiance of 2-3 W/m^2. Natural sunlight was used as the source of white light with its irradiance reduced to 2-3 W/m^2. Cells were harvested after 2 months of growth when the cotton strings were completely covered with cells.

4.2.2 Thylakoid membrane preparation

Cells corresponding to a wet weight of 1.0g were resuspended in 50mM MES pH 6.4, 10mM $MgCl_2$ and 10mM $CaCl_2$ (MCM buffer). Whole cells were lysed by sonication in 1-min cycles (at 20 W) interspersed with 5 min on ice, and membranes were pelleted by centrifugation at 30,000xg.

4.2.3 16S rDNA sequencing

Cells corresponding to a wet weight of 0.5g were broken with a 1ml cell homogenizer purchased from Kontes glass corporation. DNA was isolated using the Qiagen DNeasy plant mini kit. The 16S rDNA was amplified by PCR using 27f and 1525r primers as described in [85]. Annealing reaction was carried out at 50^{o}C. The resulting 16S rDNA fragments were ligated into a pCR 2.1 cloning vector obtained from Invitrogen containing a kanamycin resistance gene for selection. The vector was transformed into competent DH5α *E. coli* cells. The cells were incubated in a plate containing LB media and X-gal in the presence of kanamycin overnight. Several white cell colonies were picked randomly and incubated overnight in LB media. The plasmid from *E. coli* cells was isolated and the insert sequenced using the primers 27f and T7 promoter primer. The size of the partial gene cloned was about 1.5kb. Clustal X was used for aligning the various 16S rDNA sequences. MEGA 5.0 [86] software was used for constructing the phylogenetic tree (Fig. 1).

4.2.4 Transmission electron microscopy

Electron micrographs of other strains of *Leptolyngbya* have been previously published by Belleza et al. [87]. Belleza's staining protocol was slightly modified to remove Tannic acid, Ruthenium red and Alcian blue. The cells were prepared for TEM as follows: 0.5ml of F560 and F640 acclimatized cells containing about 5-10µM chlorophyll were taken and centrifuged at 5000xg. The cell pellet was treated with 2% glutaraldehyde in 0.1M pH 7.2 cacodylate buffer and kept in room temperature for 15mins. The cells were washed with the 0.1M cacodylate buffer three times by centrifuging at 5000xg. Then the

cells were treated with 0.5ml of 2% OsO_4 in 0.1M cacodylate buffer pH 7.2 for 1 hr. The sample was then washed twice by centrifugation at 5000xg with 0.1M cacodylate buffer pH 7.2. The cells were then incubated in 0.5% uranyl acetate for about 12hrs followed by washing with deionized water three times. The cells were dehydrated in acetone by using increasing concentrations of acetone that slowly infiltrated into the epoxy resin. The resin was polymerized by incubating it at 60^oC for 24 hrs. Using an ultramicrotome (Leica Ultracut from Leica Microsystems, Wetzlar, Germany), 60 nm thin sections were cut. Sections were collected onto copper mesh grids and post-stained with 2% uranyl acetate dissolved in 50% Ethanol and Sato's lead citrate[88]. They were then viewed using a JEM- 1200EX transmission electron microscope (JEOL Ltd., Tokyo, Japan). The microscope was operated at 80kV while viewing the cells.

4.2.5 Steady-state fluorescence & absorption difference measurements.

For steady-state fluorescence, the cells were resuspended in cell media containing 60% glycerol to obtain an optical density (O.D) less than 0.5 at 670 nm. Fluorescence spectra were then taken at 77 K with a Fluoromax-3 spectrofluorometer at 440 nm and 573 nm excitation wavelength.

The flash induced absorbance difference spectra of P700 were recorded at room temperature with a JTS-10 LED kinetic spectrometer (Bio-Logic, Grenoble, France). The sample was excited by actinic light (650 nm; intensity 3000uE/m^2; duration 10ms)

A Beckman Coulter DU 640 spectrophotometer was used for measuring the absorbance of cells. The extinction coefficients for calculating concentration of phycoerythrin (PE) and phycocyanin (PC) ($\varepsilon_{PE\ ,560}$ = 500,000 $M^{-1}cm^{-1}$ and $\varepsilon_{PC,620}$ =

270,000 $M^{-1}cm^{-1}$) were obtained from Bryant et al.[89]. The photosynthetic membrane contains large number of pigments which complicates quantification of phycobiliproteins in whole cells especially for allophycocyanin (APC) for which absorption overlaps strongly with the Q_X band absorption of the chlorophylls. Pigments in the PBS (from mesophiles such as *L.HI*) are unstable at high temperature while chlorophylls and carotenoids maintain their color at moderate temperatures of 60^{o}C. A method for de-convolution of the chlorophyll and phycobilipigment absorptions using heat denaturation is described by Zhao et al. [90, 91]. Cells corresponding to Chl concentrations of 3-5μmol/ml were homogenized using a glass homogenizer. One fraction of the whole cell suspension was kept at room temperature while the second half was incubated at 65^{o}C for 8 minutes in a water bath . Spectra were collected from both samples. The spectra of the PBS were obtained by subtracting the spectra of the heat incubated sample from that of the room temperature cells. The concentration of APC was determined using the method using three simultaneous equations derived using Beer-Lambert's law at wavelengths of 560, 620 and 650 nm (see Supplementary information, section 4.2) as described by Tandeau De Marsac et al.[92].

4.2.6 O_2 activity measurements

An oxylab32 electrode from Hansatech Instruments Limited was used for measuring the O_2 activity. A cell suspension in 1ml MCM buffer that contained a wet-weight of 0.1g cells were placed inside the chamber of the electrode. The cells were dark acclimatized and the cell was flushed with N_2 to decrease the O_2 background. They were exposed at

an irradiance of 60.90W/m^2 with red LEDs (650 nm) to measure the rate of oxygen production. The rate of increase in O_2 production was recorded during light intervals for calculating the rate of photosynthesis. The rate of decrease of O_2 during dark intervals was recorded to calculate the rate of respiration. At the end of the experiment, 100µl of water saturated with O_2 was added as an internal O_2 standard. The maximum solubility of 268 nmol/ml O_2 in water at 21^oC was calculated using the data published by Truesdale et al. [93]. This result was used for determining the rate of photosynthesis and respiration. The total O_2 activity was obtained by adding the rate of photosynthesis and respiration since during light intervals both pathways run simultaneously.

4.3 Results

4.3.1 Phylogenetic analysis of *Leptolyngbya* Heron Island

For the phylogenetic classification of *Leptolyngbya* Heron Island *(L.HI)*, the 16S rDNA of this organism was sequenced and aligned with the sequences of known strains of cyanobacteria. Since *L.HI* was directly isolated from natural habitat, other bacterial strains were found along with the cyanobacteria. Therefore, the 16S rDNAs sequences were amplified using PCR from the genomic DNA obtained from the cells. They were subsequently transformed into competent DH5α *E. coli* cells as described for environmental 16S rDNA sampling [16]. Individual colonies were selected and their 16S rDNA inserts sequenced. Several individual colonies were identified that contained the pure 16S rDNA of *L.HI*. Reference 16S rDNAs were identified using Genbank for the cyanobacterial species classified by Schirrmeister [16] and Turner [94]. From comparing the 16S rDNA sequence of newly isolated *L.HI* with those of the known

species it was established that this strain of cyanobacteria belongs to section III cyanobacteria (Figure 4-1, also see Figure 4-4 & Figure 4-9). This class constitutes filamentous cyanobacteria which grow in one plane and are unbranched. Section I and II consist of single-celled cyanobacteria [16]. In comparison, section IV and V consists of filamentous cyanobacteria that have terminal differentiation [16]. Section IV and V cyanobacteria form heterocysts [95] where they fix N_2 in the absence of PS II. Our *L.HI* cells do not show terminal differentiation nor do they show any indication in the electron micrographs of the presence of heterocysts. Therefore, we conclude *L.HI* belongs to the class III cyanobacteria which do not form heterocysts and do not fix N_2. The phylogenetic analysis based on the 16S rDNA (Figure 4-1) fits very well with the light and electron microscopic images shown in Figure 4-9. Hundreds of cells were imaged, but neither branches nor larger differentiated cells were observed.

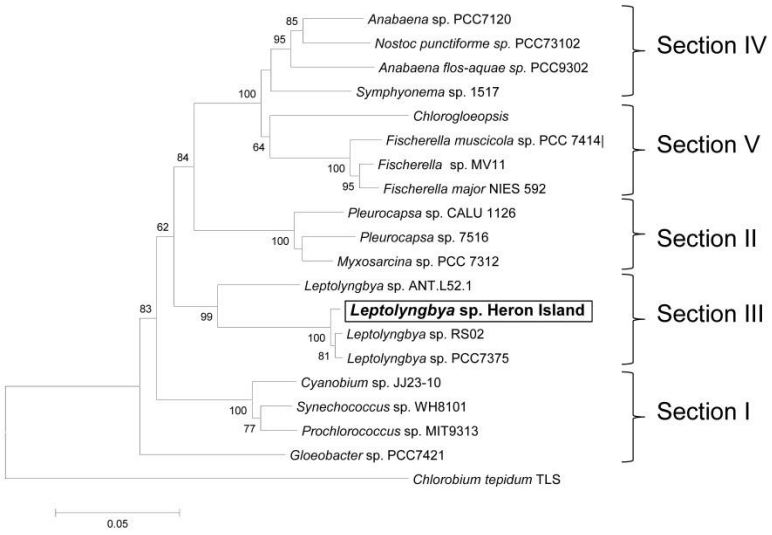


FIGURE 4-1

Phylogenetic position of *L.HI* (boxed) within the cyanobacterial phylum - The phylogenetic tree is based on the 16S rDNA sequences. The five different sections of cyanobacteria according to Rippka et al. [19] are indicated. The numbers at nodes refer to bootstrap support values (10,000 repetitions) if >60%. The sequence of *Chlorobium tepidum* TLS served as outgroup. The phylogenetic tree was generated using the Minimum Evolution method within MEGA5 [86]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances are given in the units of the number of base substitutions per site. All positions with less than 70% site coverage were eliminated. That is, fewer than 30% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 1433 positions in the final dataset. Reference 16S genes taken from [2, 16, 94, 96-101].

4.3.2 Chromatic acclimation in *Leptolyngbya* Heron Island cells

To study the CA response, *L.HI* cells grown under white light were used for inoculating the new cultures that would grow under different light conditions, i.e. under either red or green light illumination. Two different types of light sources were used: LEDs and broad wavelength filters. LEDs emit comparatively a much narrow range of wavelength.

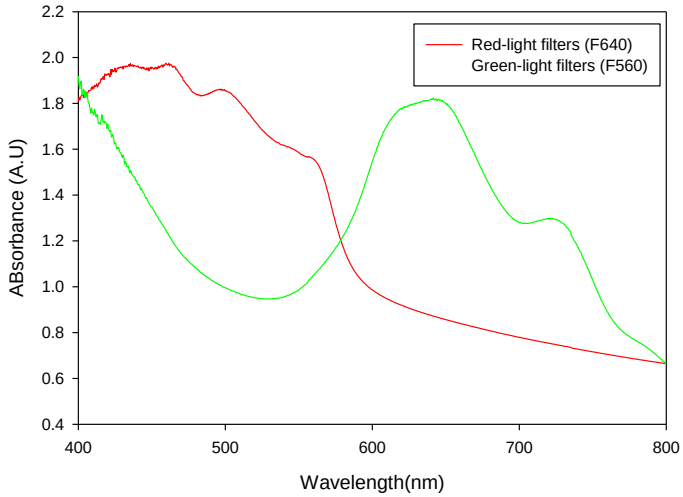
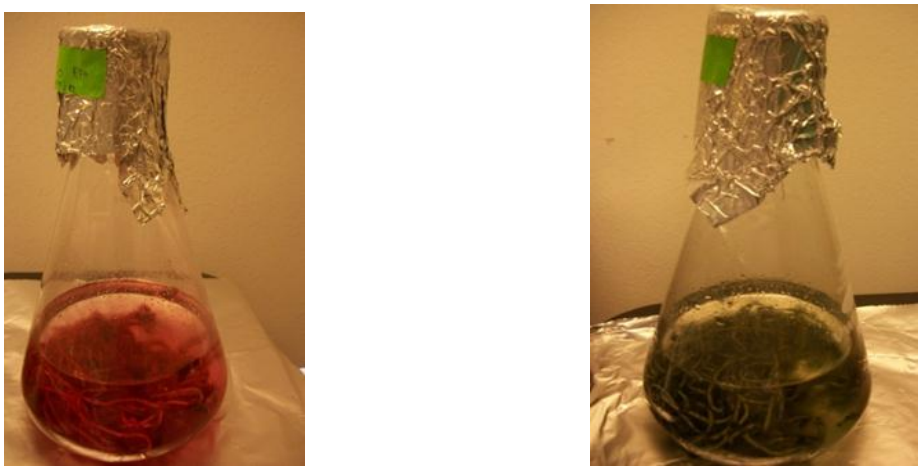


FIGURE 4-2

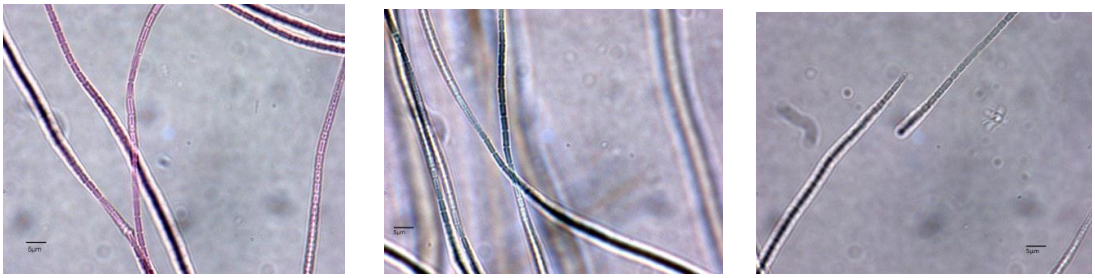
Absorption spectra of F560 and F640 filters used for inducing CA in *L.HI* cells.



(a) (b)

FIGURE 4-3

Pictures of *L.HI* cells acclimatized in (a) green light (F560) (b) red light (F640).



(a) (b) (c)

FIGURE 4-4

Light micrographs of cells acclimatized to varying light conditions
(a) Cells acclimatized to green monochromatic LED light L585 (b) cells acclimatized to red monochromatic LED light L620 (c) cells acclimatized to white light.

LEDs (λ = 585-595 nm (L585) and λ = 620-630 nm (L620)) were used as green and red light source. While LEDs provide monochromatic light of specific wavelengths, which might be used to excite the phycobilins selectively, the LED illumination does not mimic light conditions found in nature. In contrast, the filters primarily allow a broader spectral range of light covering the green or red light regions to pass through and provide much wider wavelength light (λ = 420-580 nm(F560) and λ = 620-700 nm(F640)), (see Figure 4-2) which thereby better mimicking environmental conditions.

Figure 4-3a and b show cells cultures acclimatized in green and red light respectively. Figure 4-4a and b show the cells grown under L585 and L620 light condition, respectively. Visible reddish color is observed in the cells grown under L585 (Figure 4-4a) light because CA increases the content of phycoerythrin under green light. The cells in Figure 4-4b are green because red light decreases PE content. Cells in Figure 4-4c are white-light cells which have a brownish-green color.

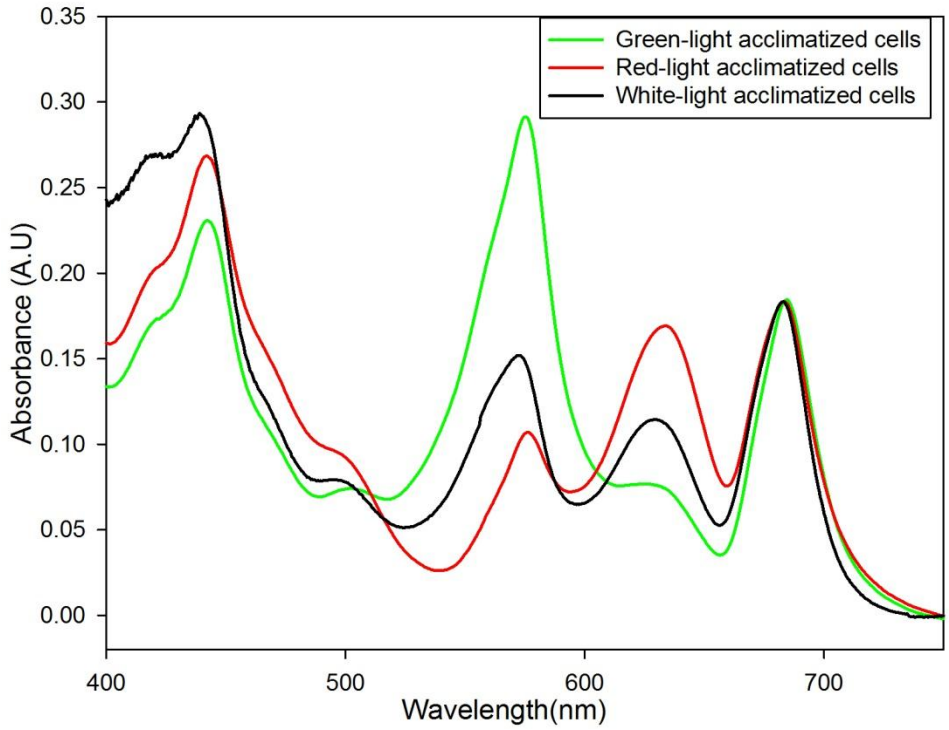


FIGURE 4-5

Absorption spectra of *L.HI* cells grown under different wavelength light conditions
red line: cells grown using λ_{max} = 420-580 nm filter, F560; and green line: cells grown using λ_{max} = 620-700 nm, F640 acclimatized cells. The black line represents the cells grown under white light condition.

The visible absorption spectral composition in Figure 4-5 indicate the changes of pigment composition in the process of CA. It is observed that the PE absorption band (λ_{max} = 572 nm) is higher in green-light acclimatized F560 cells than in red light acclimatized F640 cells. Cells acclimatized to white light show a peak intensity of PE which is less than that of F560 acclimatized cells but equal to F640 acclimatized cells. The color of cells grown under red and white light in Figure 4-4b and Figure 4-4c is mostly green and brownish-green respectively whereas color of cells grown under green light in Figure 4-4a is red also suggesting that PE is excited primarily under green light. This acclimation mechanism may be useful for the organism living in deeper ocean waters

where most of the red light may be is diminished by the water column and may also have been taken up by chlorophyll containing photoautotrophs at the surface of the ocean. This CA acclimation would enable *L.HI* to dilute PC by synthesizing more PE.

In case of PC (λ_{max} = 628 nm), the peak intensity of F640 cells is the highest followed by white-light acclimatized cells (Figure 4-5). The F560 acclimatized cells have lowest peak intensity of PC.

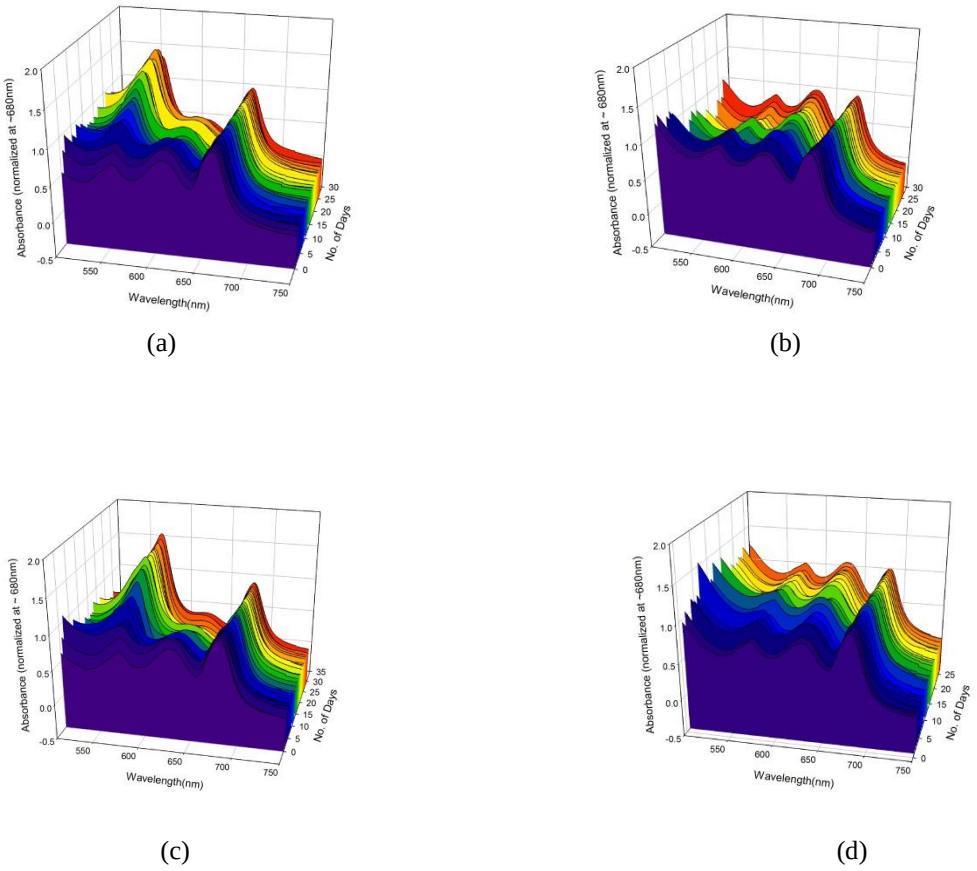


FIGURE 4-6

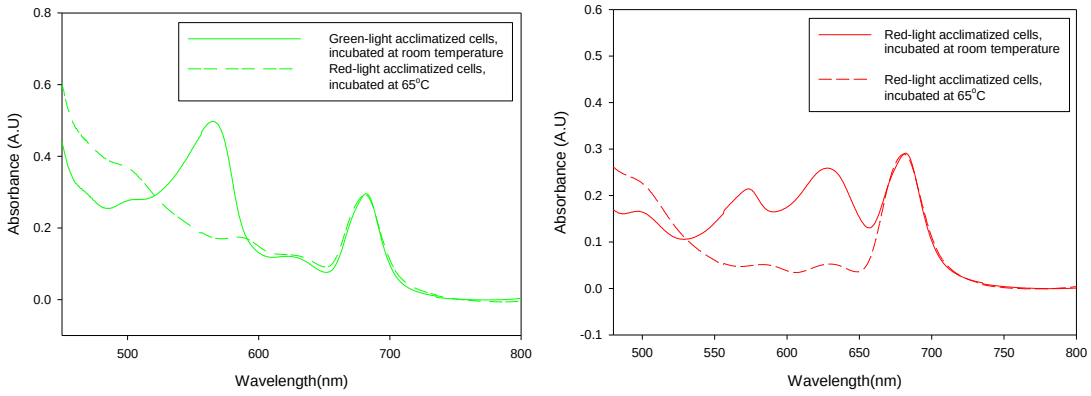
Absorbance spectra of *L.HI* as a function of time. Cells adapted to LED light (a) L585 (b) L620 or passing through broad band filters (c) F560 (d) F640.

To study the CA of *L.HI* in greater detail, the acclimation changes from white to red or to green light over a time period of a month (30 days) at time intervals of 24 hrs using a visible spectrophotometer were recorded (Figure 4-6). White light acclimatized cells were used as the starting cultures for studying the cell acclimation to light selected by green or red filters (F560 and F640) or single wavelength green or red LED illumination (L585 and L620). Interestingly, the cells showed a very similar CA response when they are acclimatized to the single wavelength LED light or the broad wavelength filters.

It was observed that there was an increase in the PE absorption band in cells acclimatized to the green filter light F560 and single wavelength LED light L585 acclimatized cells. The cells in both cases increase their light harvesting capacity by a dramatic increase in PE, which absorbs in the green-light region (see results of acclimation in Figure 4-6a & c). In contrast to the strong effect of green light on the PE content of the cells, red light exposure (both LED and filtered light) showed only a very small increase in content of PC in F640 and L620 acclimatized cells as shown in Figure 4-6b & d. On calculating the respective concentrations of PE and PC it was found that for the cells grown with the green LEDs (L585) the PE/PC ratio increased more than five times over a time span of 30 days in comparison to the starting culture of white light acclimatized cells. In contrast, in red LED light (L620) had only marginal influence on the PE/PC content as the ratio only slightly decreased during the same time (see Table 4). In case of cells grown using filters, a similar trend was observed.

4.3.3 Length of phycobilisome rods in green and red-light acclimatized cells

Determination of number of PBS/cell is important for correlating effect of CA on other components of the photosynthetic electron transfer chain. APC is located at the base of the PBS [102]. Therefore, its concentration is directly proportional to the number of PBS cores present in the cell [90]. When PBS are incubated at 65^{o}C, the phycobiliproteins with covalently bound chromophores denature while chlorophylls and carotenoids stay intact which results in disappearance of the absorbance from the phycobiliproteins [91].



(a) (b)

FIGURE 4-7

Absorption spectra of (a) Green-light and (b) Red-light acclimatized cells in both room temperature and at 65^{o}C. All spectra have been normalized to their respective chlorophyll concentrations.

This effect can be used to "subtract" the chlorophyll and carotenoid spectrum from the spectrum of all pigments in the cells as described in Figure 4-7. The dotted lines in the spectra have virtually no peaks corresponding to any of the phycobiliproteins which

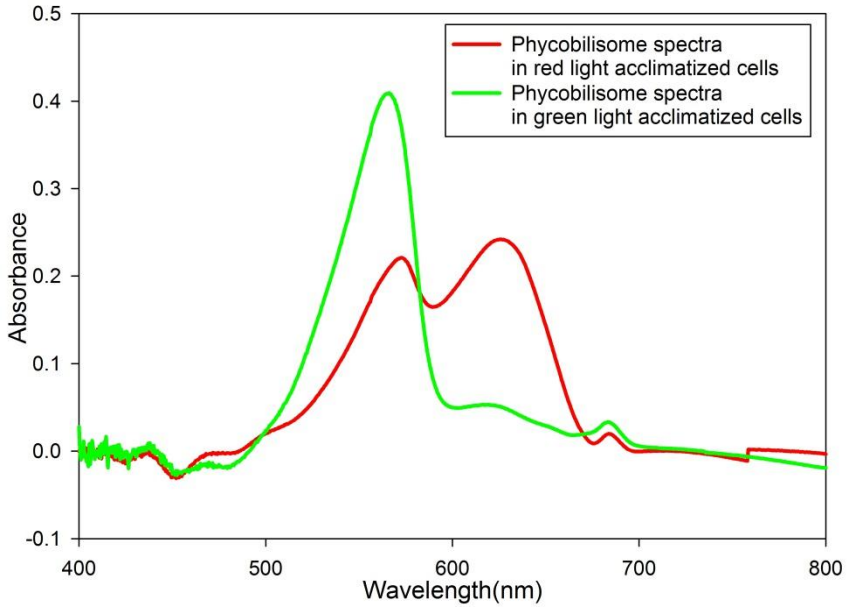


FIGURE 4-8

Absorption spectra of PBS from red and green *L.HI* cells.
The spectra were obtained after subtracting the absorbance spectra of heat-incubated cells at 65^{o}C from room-temperature incubated cells. The spectra of both red and red-light acclimatized cells have been normalized to their respective chlorophyll concentrations.

suggests PBS decompose at this temperature. The absorption spectra of the PBS were obtained by subtracting the spectra of the heat-incubated cells from the spectra of the room-temperature incubated cells (Figure 4-8). Since the absorption peak of APC is also eclipsed by the PC peak, an indirect method was used for the calculation of APC concentration. The simultaneous three-variable equations used as proposed by Bennett and Bogorad [13]. They were devised by considering the individual absorbance at three different wavelengths by all the three phycobiliproteins. The equations are as follows (The extinction coefficients used in the simultaneous equations were obtained from Bryant et al. [89]):

$$A_{560} = \varepsilon_{560,PE}[PE] + \varepsilon_{560,PC}[PC] + \varepsilon_{560,APC}[APC]$$

$$A_{620} = \varepsilon_{620,PE}[PE] + \varepsilon_{620,PC}[PC] + \varepsilon_{620,APC}[APC]$$

$$A_{650} = \varepsilon_{650,PE}[PE] + \varepsilon_{650,PC}[PC] + \varepsilon_{650,APC}[APC]$$

Table 4 compares the calculated PE, PC and APC contents of green and red-light acclimatized cells. Ideally, the PE, PC and APC concentration should be measured with the number of cells in the sample, however the PE, PC and APC were compared with respect to the Chl concentration since number of cells counted in case of filamentous cyanobacteria is impossible. Other indirect measures such as mass of cell was not found to be reproducible since thickness of the sheath covering the filaments was not constant among all filaments of *L.HI*. Thus the cell growth and spectroscopic data were related to the Chl *a* concentration. There is a very small difference in the ratio of PC/APC ratio between green-light acclimatized cells and red-light acclimatized cells. However, there is approximately seventeen fold increase (from 0.48 to 8.44) in the PE/APC ratio from red to green-light acclimatized cells. The (PE+PC)/APC is proportional to the length of the PBS rods. On comparing this ratio from the two light-acclimatized cells, a nine-fold difference (Table 4) in the length of the PBS between the red and green-light acclimatized cells is calculated. The APC content in red-light acclimatized cells is five times higher than that of green-light acclimatized cells. This is further confirmed by steady-state fluorescence experiments as shown in Figure 4-10a & b. In this figure, all spectra have been normalized with respect to the PC fluorescence peak (~640 nm). It is

observed that the APC peak at 660 nm is higher in red-acclimatized cells than in green-acclimatized cells. Peak assignments and fluorescence spectra have been discussed in greater detail in section 4.3.5. Since APC forms the base of the PBS, the red-light acclimatized cells must have dramatically increased the number of PBS/Chl compared to the green-light acclimatized cells. This is further confirmed by observing the peak for the PBS-PSII complex in Figure 4-10a & b. The peak for the PBS-PSII complex is increasing in red-light acclimatized cells which can only be possible if there is an increase in number of PBS complexes.

	µmol PE /mol Chl	µmol PC /mol Chl	µmol APC/ mol Chl	PE/APC	PC/APC	(PE+PC) /APC
Red-light acclimatized cells	87.4	216.6	182.4	0.48	1.18	1.66
Green-light acclimatized cells	288.8	49.4	34.2	8.44	1.44	9.88

TABLE 4

Relative concentrations of phycoerythrin (PE), phycocyanin (PC) and allophycocyanin (APC) in green (F560) and red-light acclimatized cells (F640) obtained from Figure 4-8.

To calculate concentration of PE, PC and APC, absorption difference spectra of cells measured at 65^{o}C from that of room temperature were plotted (Figure 4-7 and Figure 4-8). The concentrations of PE, PC and APC were calculated using simultaneous equations as described above. Concentration of PE and PC were calculated using extinction coefficients of 500,000 and 270,000$M^{-1}cm^{-1}$ respectively[89].

4.3.4 Electron micrographs of green and red-light acclimatized *Leptolyngbya* Heron Island cells

Ultrastructural analysis of *L.HI* by transmission electron microscopy was undertaken with the goal to study changes in cell morphology with change in light conditions. For this purpose, cells acclimatized in both λ_{max} = 420-580 nm(F560) and λ_{max} = 620-700 nm (F640) were harvested after full acclimation was completed and prepared for the TEM (Transmission electron microscopy) analysis. The TEM data show that the cells have a clearly delineated cell membrane and are covered by an extracellular sheath made of carbohydrates that might help in the uptake of nutrients [87].

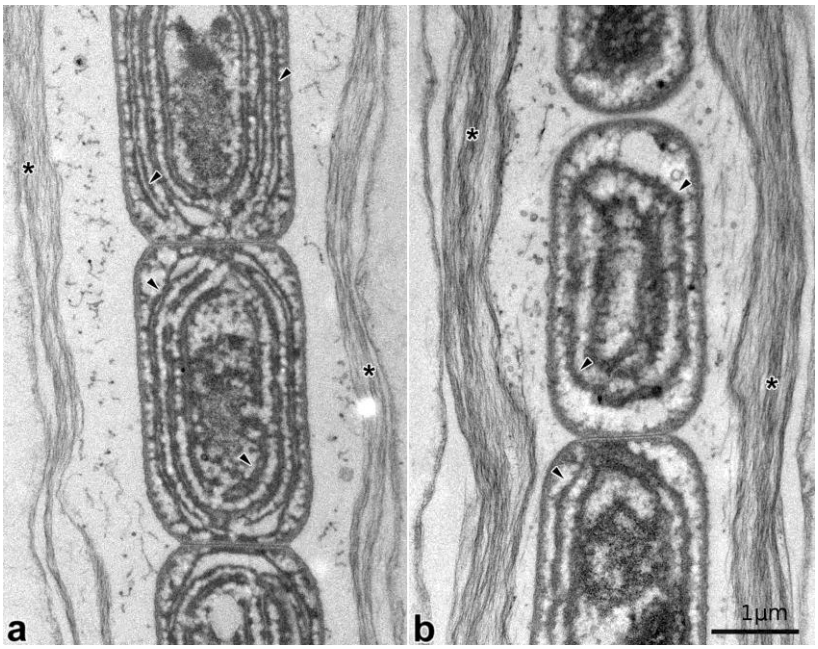


FIGURE 4-9

Transmission electron micrographs of *L.HI* cells grown under various light conditions

(a) Green-light acclimatized cells (grown under the filter of λ_{max} = 420-580 nm (F560) and the (b) red-light acclimatized cells grown under the filter of λ_{max} = 620-700 nm (F640). Scale bar is 1 µm. The outer polysaccharide layer of the cell are marked with a (*) whereas the thylakoid layers are marked with an arrow.

The width of the cells including the extracellular layer is approximately 2.5-3µm. The length and breadth of each of the cells (without the extracellular later) is about 3 and 1.5 µm respectively.

It is observed that there are higher number of thylakoid layers in λ_{max} = 420 -580 nm (F560) green light acclimatized cells as compared to λ_{max} = 620-700 nm (F640) red light acclimatized cells, as shown in Figure 4-9. In case of green-light acclimatized cells (F560 acclimatized cells, Figure 4-9a) an average of 3-4 thylakoid layers per cell were observed as compared to red-light acclimatized cells (F640 acclimatized cells, Figure 4-9b) which only had an average of 1-2 thylakoid rings. In another separate study, Khanna et al. [103] also observed in *Anacystis nidulans* mutants lacking in PC that the thylakoid area was reduced. Although *Anacystis nidulans* does not exhibit CA, in both cases the trend is the same that on reducing the length of the PBS rods there is a decrease in number of thylakoids. This leads to the hypothesis that to arrange each of the PBS on the thylakoid membrane more space is required if the rods are bigger. If this hypothesis is true, more thylakoids would be required to fit all the PBS in green-light acclimatized cells. Although PBS increase in red-light acclimatized cells is five-fold, the length of the rods decrease nine-fold thus increasing the density of these "shorter" PBS in the thylakoid membrane of red-acclimatized cells. This may lead to a higher number of thylakoids in green-light acclimatized cells. It may be possible that *L.HI* is optimized to harvest green light since its found at a lower level in the water column of the coral reefs where red light is not abundant which may lead to longer PBS rods and greater number of thylakoids in green light.

4.3.5 Steady-state fluorescence of *Leptolyngbya* Heron Island cells

Steady-state fluorescence spectra were measured (Figure 4-10a & b) at cryogenic temperatures after 20 days from the start of inoculation of the culture after switch from white to red or green light (After 20 days there was no more appreciable changes in PE and PC amounts, Figure 4-6). At this time, cells were fully acclimatized to the new light conditions and have reached a constant PE/PC to Chl as shown in Figure 4-6. The fluorescence spectra were normalized at 712 nm.

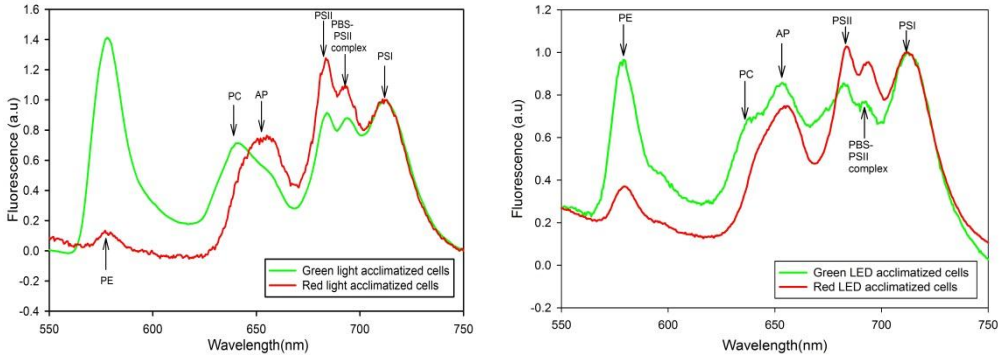


FIGURE 4-10

Steady-state fluorescence spectra of *L.HI* cells measured after cells acclimatized completely to their respective light conditions.
(a) λ_{max} = 620-700 nm (F640) & F560 (b) λ_{max} = 620 nm (L620) & L585. The fluorescence spectra of each was measured at a temperature of 77K at λ_{exc} = 440 nm. All spectra have been normalized at 712 nm.

The band at 580 nm corresponds to PE fluorescence [104], which is higher in green-light acclimatized cells (grown under green light F560 & L585) than in red-light acclimatized cells (grown under red light F640 & L620). There are also peaks at approximately 640 and 660 nm which correspond to PC and APC fluorescence respectively [105]. The peak assignment of PE and PC were verified by isolating each of these proteins separately and measuring their steady-state fluorescence. The peak at 682 nm can be assigned to

fluorescence emitted by PS II since its more blue-shifted and the peak at 712 nm is emitted by PS I since its more red-shifted. To further prove the peak assignments of the various proteins, thylakoid membranes devoid of PBS (as described in methods section) were isolated from the cells and their steady-state fluorescence spectra was measured (Figure 4-11).

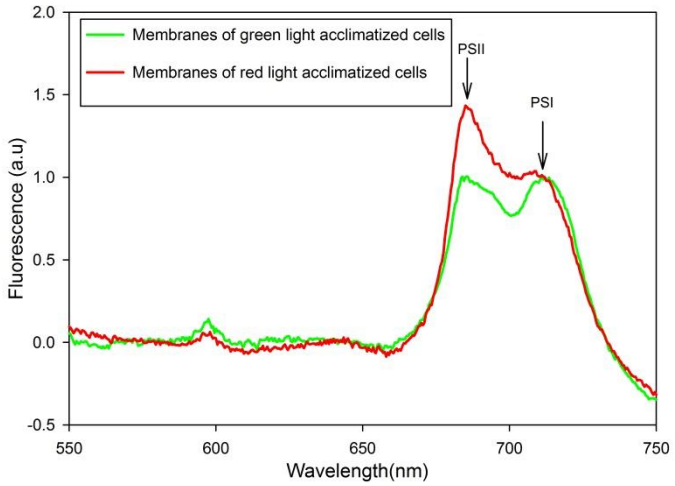


FIGURE 4-11

Steady-state fluorescence spectra of *L.HI* membranes isolated from intact cells. Cells grown in λ_{max} = 420-580 nm (F560) and F640. The fluorescence spectra of each cell culture was measured at cryogenic temperature of 77K with excitation at λ_{exc} = 440 nm. All spectra have been normalized to the fluorescence at 712 nm.

In Figure 4-11, the bands for the phycobiliproteins i.e. bands at 580, 640 and 660 nm are either very small or absent whereas the bands at 682 and 712 nm corresponding to the photosystems are clearly visible. This is expected as the PBS separate from the thylakoid membrane during the isolation of the membranes by differential centrifugation. Since the band at 660 nm is absent in the membrane spectra we can infer that it must have come from the PBS. Since APC is the only component of PBS whose peak remains

unassigned, we may conclude that the 660 nm peak could be emitted by APC. Also the band at 691 nm peak is missing from the fluorescence spectra of the membranes. This also suggests that this peak comes from the PBS. A small peak at approximately 691 nm has previously been assigned to the terminal emitter from a complex formed between the PBS and the PS II core complex [105]. Thus we can assign this peak to the PBS-PS II core complex. The results deduced from Figure 4-11 match with our earlier assignment of the peaks from literature [104, 105].

It is interesting to note that in green-light acclimatized cells Figure 4-10a the PC peak is higher than that of APC whereas in Figure 4-10b it is the opposite. This is because the filters allow some light of λ = 620 nm to pass through thereby the cells may have a higher PC content. In contrast, in Figure 4-10b cells illuminated with LEDs are exposed to a very narrow range of light λ = 585-595 nm which is outside the absorption spectrum of PC leading to a decrease in the PC content of the cells.

It is also noted that in Figure 4-10a & b the PS II/PS I ratio in red-light acclimatized cells was higher than in green-light acclimatized cells. This interesting finding was observed both under LED light illumination (L585 & L620) and during illumination with broad red and green filter (F560 & F640) light. Changes in the ratio of PS II/PS I are often a regulatory mechanism to maintain turnover balance between the two photosystems [74]. PBS are the major external antenna complexes in cyanobacteria which funnel the excitation energy mainly to PS II. Under green light, the PBS increase their efficiency by elongating the rods with PE which absorbs in the green, so light excitation in the green will funnel most of the excitation energy to PS II. Red light

is mainly absorbed by PC as well as the chlorophylls of the core antenna in PS I and PSII. PS I has a large core antenna system of 96 Chl/P700 [106] which is tightly coupled (quantum yield of excitation energy transfer is 99.9% at RT [107, 108]) , and it contains chlorophylls that are "red shifted" and absorb at wavelengths > 700 nm. In contrast, PS II has a much smaller core antenna system of only 38 Chl/P680 [109, 110] and does not absorb light in the red (> 700 nm). Furthermore the quantum yield of excitation energy transfer in PS II is only 80% , i.e. 20% of the excitation energy at RT is lost in form of fluorescence [107, 108]). To compensate for the lower efficiency of the core antenna system, PS II receives light from the PBS. Although PBS is known to couple with PS I also by a mechanism called state transitions, this is unlikely since state transitions are temporary acclimations which last in the order of seconds or minutes [111, 112] whereas CA occurs in about a week. Campbell et al. [113] have proposed that state transitions may operate in green-red light acclimation by long-term residency in state 1 or 2. However, state transitions are even less likely for *L.HI* because of the higher salt concentration its adapted to living in. High salt conditions lead to high osmotic strength. *L.HI* was isolated from open ocean sea water containing higher amount of salts than cyanobacteria living in freshwater. Joshua et al. [114] have shown that by using a phosphate buffer of 0.2M they increase the osmotic strength of the buffer thereby reducing PBS mobility and thus eliminating state transitions. The cell media for growing *L.HI* contains about 0.5-0.75M NaCl. In this case, it is unlikely that state transitions can take place. Another indication that the PBS may bind more to PS II in red light, the peak at 691 nm (the PBS-PS II complex) actually increases in red-light acclimatized cells. This shouldn't have been in the case if

a transition to state I would have taken place. This observation also further supports our hypothesis that the amount of both PS II and PBS is increasing in red light. However, state-transitions have not been exclusively studied in marine cyanobacteria so far.

Some cyanobacterial strains feature type III CA where PC, which absorbs at 620 nm, is dramatically overexpressed under red light illumination. However, the *L.HI* cells adapt to red light in a novel way. The PC content of the PBS in under red light illumination is less than that of PE in green-light. This is coupled with a dramatic increase of the ratio of total PBS complexes per Chl. At the same time the ratio of PS II/PS I is increased. This leads to the conclusion that the cells balance their light capturing ability under red light by increasing the PS II/PS I & PBS/chl ratio. This conclusion led us to the next set of experiments where we measured directly the PS I activity of the cells by measuring light induced absorption changes at 700 nm and the PS II activity by determination of the oxygen evolving activity.

4.3.6 Estimation of photosystem I in *Leptolyngbya* Heron Island cells

Since light conditions affect the PS II/PS I ratio, it is important to determine which one of the two photosystems is up or down regulated. The photosystem contents were measured with respect to Chl concentrations because counting number of cells was not possible due to the filamentous nature of *L.HI*. Light induced absorption changes of P700 were measured in the millisecond time range as described in the methods section.

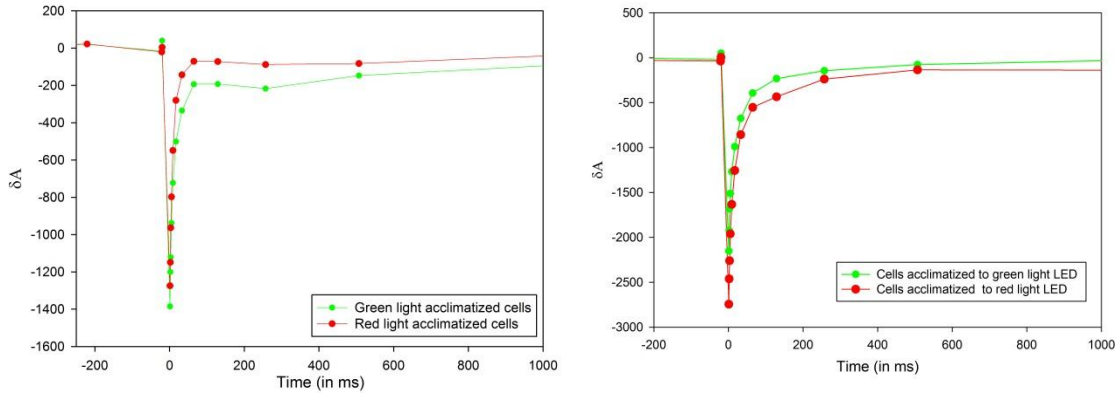


FIGURE 4-12

Light-induced P700 absorption difference spectra of *L.HI* cells.
Cells are grown in green/red light with broad band filters (a) λ_{max} = F560 & F640 (b) λ_{max} = L585 & L620 acclimatized cells. The experimental error of the measurements is 10%, i.e. there is no significant difference in the photosystem I content of red and green-light acclimatized cells.

The results are depicted in Figure 4-12. It was found that the P700/Chl ratios were very similar in all samples. Within the experimental error of 10% the P700/Chl ratio there is no significant difference in PS I content between red and green-light acclimatized cells for both, the illumination with broad band filters and for illumination with LED light. This suggests that it might be primarily PS II, which undergoes changes in expression with change in light conditions.

This result is expected considering that PBS primarily dock to PS II [8, 115, 116] (see Figure 4-11). Although they can also bind to PS I, docking of PBS to PS I mainly happens as a result of state transitions. State transitions depend on the redox-pool of the electron transfer chain where PBS temporarily decouple from PS II to bind PS I, when the PQ pool becomes reduces Indicating too much influx of electrons from PSII and to less efflux from PSI. However state transitions are short-term acclimations which occur in the span of a few minutes or seconds[72]. Therefore, these acclimations are not likely

to affect the changes in PS II/PS I ratio. CA was induced in the time frame of several weeks and such long-term acclimations caused by CA directly involve changes in expression of PS I and PS II and do not involve short term movement of PBS [72, 117]. Also *L.HI* is grown under high salt which may further prevent PBS mobility, thereby further reducing the possibility of PBS migrating from PS II to PS I. Thus we hypothesize that the PBS primarily attaches to PS II in our experiments.

4.3.7 Effect of chromatic acclimation on photosystem II content

The rate of O_2 evolution/chl was monitored in whole cells to determine if red and green light acclimatized cells show differences in PS II activity. The O_2 liberated by a cell is proportional to the amount of PS II present in the cell. All cells were grown using the same amount of nutrients and light except that different wavelengths of lights were used, so differences in O_2 activity would be due to change in light conditions.

S. No	Light Source	$[O_2]$/[chl] (in nmol O_2/min µmol chl)
1	F560	6.9
2	F640	12.2
3	L585	3.8
4	L620	5.7

TABLE 5

O_2 activity per chlorophyll for *L.HI* cells grown in F560, F640, L585, L620. The data represent the average of three different experiments. O_2 activity was measured after 30days from start of inoculation.

Therefore O_2 evolution per chlorophyll was measured for all the four light

acclimatized cells (Table 5). The O_2 evolution was determined with respect to chl concentration and not with respect to the mass of cell since thickness of the sheath covering the filaments was not constant among all filaments of *L.HI* and the cells can not be counted due to the filamentous nature of *L.HI*.

It is observed that green-light acclimatized cells (F560 & L585) show lower O_2 evolution/Chl compared to the red-light acclimatized cells illuminated with the corresponding LED or filter setup (F640& L620). These results fit very well with the increase in PS II fluorescence shown in Fig. 6. It has been shown by Zhao et al.[90] that on removal of the phycocyanin gene from *Synechococcus* sp. PCC 7002, the expression of PS II increases so as to compensate for the loss of the light harvesting capacity of PC. It is very interesting that in the case of red-light acclimated *L.HI* cells (which have a green apppearance), we observe a similar trend where the reduction of the cross-sectional area of the PBS in the red-light acclimatized cells (in comparison with green-light acclimatized cells) is compensated by a dramatic increase in the expression of PS II. Although *Synechococcus* sp. PCC 7002 does not exhibit CA a similar trend is observed. The PS II is increased to compensate for the decrease in harvesting ability of the PBS due to reduction of the length of the PBS rods. The increase in PS II also results in increase in the number of docking sites for the PBS which also increases in red light.

4.4 Discussion

Cyanobacteria representing one of the oldest and most diverse prokaryotic photoautotrophs[16] can adapt to various environmental light conditions. Figure 4-1 shows that *Leptolyngbya* Heron Island *(L.HI)* belongs to the class of section III

cyanobacteria, consistent with their morphology that they form unbranched filaments (Figure 4-4a, b and c). In Figure 4-5 and Figure 4-6a & c we find that the amount of phycoerythrin (PE) increases in green-light acclimatized cells with time (in both cases independent cell-cultures were illuminated with broad band filters F560 and by narrow band LED light L585). In contrast phycocyanin (PC) shows only slight changes under CA conditions under both filter F640 and LED L620 illumination as presented in Figure 4-6b & d. The observation of CA in LED acclimatized cells suggests that the regulation of CA acclimatized cells is triggered by single wavelength and broad wavelength illumination in the same way. The relative concentrations of PE, PC and allophycocyanin (APC) were measured in both red and green light acclimatized cells (Table 4). It was found that the PE/APC ratio was nine times higher in green-light acclimatized cells as compared to red-light acclimatized cells. In contrast, the PC/APC ratio increased only slightly from green to red-light acclimatized cells. This suggests that the phycobilisome (PBS) rods in green-light acclimatized cells are longer than those in red-light acclimatized cells (Table 4). The concentration of APC in red-light acclimatized cells was five times higher than in green-light acclimatized cells suggesting that there are much more PBS bound to the thylakoid membrane in red-light acclimatized cells as compared to green-light acclimatized cells. This may be an acclimation to compensate for the decreased ability to harvest light due to small size of the rods in the PBS. and is accompaied by an increase in the PSII content.

The investigation of the cell morphology by transmission electron microscopy (Figure 4-9) shows that *L.HI* are filamentous having clearly delineated cell membranes enclosed by a sheath of polysaccharide layer. However, the most

interesting observation in the EM micrographs of Figure 4-9 is that the red-light acclimatized cells have decreased number of thylakoid layers than green-light acclimatized cells. In case of *A. nidulans* it was also observed that the mutant lacking PC also showed a decrease in the number of thylakoids [104]. Both these results suggest that the decrease in thylakoids might be a result of decrease in the PBS size in both studies. In green-light the large size of the PBS, by dramatic increase of PS/PBS, more than compensates for the lower number of PBS (as compared to red-light acclimatized cells). The larger PBS may requires more membrane area for it to be bound. This possibly increases the number of thylakoids in green-light acclimatized cells.

Figure 4-10a & b show that red-light acclimatized cells have a higher PS II/PS I ratio than green-light acclimatized cells. Modulation of the stoichiometry between the two photosystems is seen as a mechanism to balance turnover rates [73]. Our results (both fluorescence and O_2 activity measurements (Table 5) show that the PSII/PSI ratio in red-light acclimatized cells is increased compared to green-light acclimatized cells to compensate for the lower excitation energy transfer rate. Interestingly, a similar regulation mechanism was reported earlier by Zhao et al [90] for a mutant strain that lacks PC which support our results. They mutated the PC gene using interposon mutagenesis in *Synechococcous* sp. PCC 7002 and found that the mutant had a higher PS II/PS I ratio than the wild-type. They proposed that due to the mutation, the cross-section of the PBS was much smaller as compared to the wild-type. In order to compensate for it the PS II concentration was increased in order to harvest more light. In another separate study, Khanna et al. [103] also mutated the PC in *A.*

nidulans. They also reported that PS II content increased along with increase in the number of PBS. Although none of these strains exhibit CA, the CA response in *L.HI* features the same trend. This leads us to propose that *L.HI* increases PSII and PBS content under red light illumination to compensate for the smaller PBS light harvesting capacity.

To further prove if this kind of acclimation actually happens as a result of CA, LED lights were used since only PE and PC can absorb at those wavelengths. In these conditions an increase in PBS and PS II in red light occurred confirming that this effect actually occurred as a result of CA. The same effect was observed with filters which allow a much wider range of wavelength to pass through thereby mimicking more closely environmental conditions. This suggests that this acclimation might be actually taking place in nature. CA may be advantage for *L.HI* as this strain of marine cyanobacteria is more adapted to living deeper in the water column of the coral reef , where red light is scarce and blue-green light is more abundant. Thus its able to harvest light more easily in green-light conditions.

4.5 Conclusions

To conclude, *L.HI* is a section III filamentous cyanobacterium which exhibits reduction in the length of PBS when cells are acclimatized to red light compared to cells acclimatized to green light. In addition, the content of thylakoids/cell is higher in green-light acclimatized cells so as to accommodate the large PBS. To compensate for the decreased ability to harvest light in red light, the PS II content of the thylakoids is

increased coupled with an increase in the number of PBS, so as to maintain turnover rate of the photosynthetic electron transfer chain. This strain was found deeper inside the water column of the coral reef is and therefore possibly more acclimatized to harvesting blue-green light resulting in such a unique CA response.

CHAPTER 5: CONCLUSIONS

In this dissertation, new knowledge has been gained about the mechanism of light harvesting by the phycobilisomes. Also the genome of the cyanobacterium *Leptolyngbya* Heron Island was elucidated in this work.

In chapter 1, the genome of the cyanobacterium *Leptolyngbya* Heron Island was elucidated. Since this cyanobacterium was isolated from natural habitats it had other bacterial contaminants present in its cell culture (Figure 2-3). The genomic DNA was isolated and the content of *L.HI* DNA in the sample was tested using 16S rDNA studies. For this each of the 16S rDNA PCR products were transformed into competent *E.Coli* cells using a plasmid. Approximately half of the PCR products were found to have 16S rDNA's that matched with those of other *Leptolyngbya* species. This step was essential to verify that all parts of the cyanobacterial genome received good coverage during sequencing and are represented in the final genomic dataset.

A four-step strategy was used for identifying the genomic scaffolds of *Leptolyngbya* Heron Island (Figure 2-1). The first step involved BLAST analysis in which scaffolds which had any of its genes that matched with another in the reference genome were selected (section 2.2.5). This is a highly sensitive technique for detecting *L.HI* gene sequences but also ended up in wrongly selecting those scaffolds which were not part of the *L.HI* genome but got selected because they had some common conserved bacterial gene which matched with the reference cyanobacterial gene dataset. To remove these false positives, tetranucleotide frequencies coupled with principal component analysis was used (Figure 2-2). Tetranucleotide frequencies are usually conserved

throughout the length of a particular genome [32, 39]. This method thus helped in removing some of the false positive BLAST scaffolds in the previous step. %(G+C) analysis was also used for removing other false positive results in the first step (Figure 2-3). Finally this gene model was verified by comparing the gene sequences with non-redundant prokaryotic gene database in NCBI. It was found that 96% of the genes in the gene model matched with other cyanobacterial genes in NCBI. 78% of the genes that matched with cyanobacterial genes in NCBI database were found to be from other *Leptolyngbya* species and 80% of them were from the *Oscillatoria* group (which contains the cyanobacterial species *Leptolyngbya*). The remaining 4% of the non-cyanobacterial genes could be in as a result of horizontal gene transfer [40] by cyanophages. As part of the gene annotation, 11 CRISPR arrays [42] were detected which act as an immunity against cyanophages. This suggests that these cyanophages are highly active in the habitat of *L.HI*. By annotating the genome of *L.HI*, the phycoerythrin α and β-subunit sequences were found and used for elucidation of the phycoerythrin crystal structure.

In chapter 3, the phycoerythrin crystal structure has been described. Phycoerythrin was extracted from *Leptolyngbya* Heron Island and purified using ion-exchange chromatography (Figure 3-1). The protein was crystallized using vapor-diffusion technique (Figure 3-3). The protein crystals diffracted to a resolution of 2Å. Phycoerythrin has two α-helical subunits called α and β (Figure 3-4). This αβ heterodimer trimerizes to form a cyclic trimer $(\alpha\beta)_3$ which in turn dimerizes to give the hexamer $(\alpha\beta)_6$ (Figure 3-4). The α-subunit consists of two phycoerythrobilin chromophores , the β-subunit consists of two phycoerythrobilin chromophores and one phycourobilin chromophore (Figure 3-4a). The phycoerythrobilin has two different conformations in

which the two outer pyrrole's are further out in the outer PEB (Figure 3-10) resulting in lower electron delocalization and a more blue-shifted λ_{max} (Figure 3-11a & b). This dramatically decreases the overlap integrals between these three chromophores (PEB540, PEB565 and PUB495) as compared to transitions within these same type of chromophores [12]. The Forster resonance energy transfer theory was used to calculate the energy transfer time constant between two given chromophores in phycoerythrin. The energy transfer time constant values were compared to experimental data obtained using time-resolved fluorescence [44]. It was found that the theoretically calculated time constant were close to experimentally determined ones (Table 3).

In Chapter 4, the effect of chromatic acclimation on photosystems have been described. Chromatic acclimation is the change in PE/PC ratio with change in green/red light conditions. It was found that in the case of *Leptolyngbya* Heron Island, phycoerythrin expression is increased more in green light acclimatized cells than phycocyanin in red-light acclimatized cells (Figure 4-5 and Figure 4-6). Quantitative estimation of phycoerythrin and phycocyanin in both the light acclimatized states suggest that length of the phycobilisomes is nine times higher in green-light acclimatized state than in red-light conditions (Table 4). This leads to unequal harvesting of light by the phycobilisomes in the two light acclimatized cells. It was found that the PS II/PS I ratio was higher in red-light acclimatized cells (Figure 4-10 and Figure 4-11). Quantitative estimation of PS I using P700 absorption difference spectra revealed that PS I is constant in both light acclimations (Figure 4-12). O_2 activity studies revealed that red-light acclimatized cells liberated more O_2/Chlorophyll which suggested that PS II was higher in this light acclimatized state (Table 5).

CHAPTER 6: FUTURE WORK

6.1 Verification of chromatic acclimation results using molecular biology techniques

6.1.1 Introduction - Quantitative polymerase chain reaction (qPCR)

Quantitative polymerase chain reaction is a very accurate method for detecting copies of a particular gene [118]. This method involves use of sequence-specific DNA probes which are complementary to the DNA probe of interest. This DNA sequence contains a fluorescent probe which emits fluorescence only on hybridization with the DNA sequence of interest.

Expression of genes is usually measured by monitoring the corresponding mRNA of a particular gene. Since qPCR usually involves monitoring of DNA, the mRNA must be converted back into DNA. For this purpose, quantitative reverse transcriptase - polymerase chain reaction is used (qRT-PCR) [119, 120]. Thus the corresponding mRNA of a gene of interest is converted back into DNA.

6.1.2 Verification of spectroscopic results on chromatic acclimation

The results reported in chapter 3 are based on spectroscopic measurements. This work was done prior to the sequencing of *L.HI* genome. Therefore verification using molecular biology techniques such as quantitative polymerase chain reaction were not feasible at that time. Subsequently, the entire genome of *Leptolyngbya* Heron Island was sequenced. Now it has become feasible to carry out this verification.

In chapter 3, it was found that expression of phycoerythrin, phycocyanin, photosystem II changed with change in light conditions. The mRNA's of the α and β-subunit of phycoerythrin and phycocyanin must be probed. The psbA gene (of photosystem II) transcript also needs to be probed. Although spectroscopic results suggest that photosystem I is constant in both the light acclimatized cells this need to be verified using quantitative reverse transcriptase polymerase chain reaction.

6.2 Possible use of *Leptolyngbya* Heron Island for bioenergy production

Cyanobacteria have been widely used for the production of biofuels [121, 122]. Photosynthetic organisms in aquatic environments have been found to be more efficient as compared to their terrestrial counterparts [123]. This may be because aquatic photoautotrophs are exposed to less amount of light. Therefore they usually contain light harvesting complexes such as the phycobilisomes so that they can harvest light more efficiently.

However, energy harvested using other technologies such as photovoltaic cell has better efficiencies as compared to photosynthetic organisms currently being used [124]. This is partly due to most photosynthetic organisms used for biofuel production rely on chlorophyll for harvesting light which is only able to absorb a very small part of the visible light spectrum [124]. Thus cyanobacterial species such as *Leptolyngbya* Heron Island are ideal candidates for this purpose. The phycobilisomes in *L.HI* contain phycoerythrin, phycocyanin and allophycocyanin which along with chlorophyll cover a large part part of the visible spectrum. This dramatically increases the light harvesting capability of this cyanobacterium. Also since it is able to change the

composition of the phycobilisome with change in light conditions this further increases the photosynthetic efficiency of the cyanobacterium.

APPENDIX A: PYTHON SCRIPTS

Script 1 - Blast Algorithm for identifying *L.HI* scaffolds

```
# This script requires biopython, numpy and matplotlib for
execution.

#                                   NOTES
# 1) It is important that both the query and the fasta file have the
fasta id as a number eg. 1, 2, 3 etc.
# For this please use the script createfasta.py
# 2) Before using this script, it is necessary to make a blast
database of the reference fasta file (filename2).
# The syntax is as follows:
# formatdb -i filename2 -p F -o T -n filename2
#This script is licensed under GNU GPL v3
from Bio.Blast.Applications import NcbiblastnCommandline
from Bio.Blast import NCBIXML
from Bio.SeqUtils import GC
from Bio import SeqIO
import numpy as np
import matplotlib.pyplot as plt
import os
import collections
# Query scaffold
filename1 = input('Enter name of query fasta file (enclose name in
""):')
# Reference scaffold
filename2 = input('Enter name of reference fasta file (enclose name
in ""):')
blastx_cline = NcbiblastnCommandline(query=filename1, db=filename2,
evalue=0.001,outfmt=6, out="lepblast.xml")
print blastx_cline
stdout, stderr = blastx_cline()

print "Blast over, now doing analysis...."

with open("lepblast2.xml","w") as abc:
    for line in open("lepblast.xml"):
                line = line.replace("Query","")
                line = line.replace("_","")
                abc.write("%s" % (line))
abc.close()

x = np.loadtxt("lepblast2.xml", unpack = True, dtype = int)
redun = x[1]
uni = np.unique(redun) # Selects only unique sequences
uni = uni.astype(int)
print "Number of scaffolds in the resultant fasta file is:",
len(uni)
seq_records2 = SeqIO.parse(filename2, "fasta")
seq_records2 = list(seq_records2)
```

```
print "Number of scaffolds in the original fasta file
is:",len(seq_records2)

with open("Purecontigs.fasta","w") as defg:
  for v in uni:
          if len(str(seq_records2[v-1].seq)) >= 10000: # Remove
sequences with length less than 10000 since they dont match with the
L.HI sequence
                  defg.write("%s %s\n %s\n" % (">",seq_records2[v-
1].id, seq_records2[v-1].seq)) # Making the fasta file

defg.close()
# Remove temporary file "lepblast.xml"
os.remove("lepblast.xml")

#Plotting histogram of GC of purecontigs

with open("temp2.txt","w") as ijk:

   for seq_record2 in SeqIO.parse(filename2, "fasta"):
       ijk.write("%d %f\n" % (len(seq_record2.seq),
GC(seq_record2.seq)))#Calculating GC content and saving it to
Purecontigs.txt


ijk.close()


seq_records3 = SeqIO.parse("Purecontigs.fasta", "fasta")
seq_records3 = list(seq_records3)

with open("temp3.txt","w") as mno:

   for seq_record3 in SeqIO.parse("Purecontigs.fasta", "fasta"):
       mno.write("%d %f\n" % (len(seq_record3.seq),
GC(seq_record3.seq)))#Calculating GC content and saving it to
Purecontigs.txt


mno.close()

efgh = np.loadtxt("temp2.txt",dtype = float)
abcd = np.loadtxt("temp3.txt",dtype = float)

with open("Purecontigs.txt","w") as pqrs:
  for r1 in range(100):
    r2 = r1+1
    tot = 0
    for j in range(len(abcd)):
       if abcd[j][1]>r1 and abcd[j][1]<=r2:
            tot += abcd[j][0]
    pqrs.write("%i\t%i\n" %(r2,tot))
pqrs.close()
```

```
with open("original.txt","w") as uvwx:
  for r1 in range(100):
    r2 = r1+1
    tot = 0
    for j in range(len(efgh)):
       if efgh[j][1]>r1 and efgh[j][1]<=r2:
            tot += efgh[j][0]
    uvwx.write("%i\t%i\n" %(r2,tot))
uvwx.close()

xyz = np.loadtxt("Purecontigs.txt", unpack=True)
lmn = np.loadtxt("original.txt", unpack=True)

plt.plot(lmn[0],lmn[1]/1000000,'g-', lw=2, label = "Original \n
scaffolds")
plt.plot(xyz[0],xyz[1]/1000000,'r-', lw=2, label = "BLAST+ \n
scaffolds")
plt.ylabel("Total length of scaffolds (in Mbp)")
plt.xlabel("%(G+C) content")
plt.legend(loc="upper left")
plt.savefig("Purecontigs.png", dpi=500)
plt.close()

# Calculating number of hits for each scaffolds in query fasta file
bins = np.arange(1,len(seq_records2)+2,1)
a=np.arange(1,len(seq_records2)+1,1)
N,B = np.histogram(redun,bins)
plt.plot(a,N,'r-', lw=2)
plt.ylabel("No. of BLAST hits")
plt.xlabel("Scaffold number")

# Save figure as high resolution PNG
plt.savefig("Blasthits.jpg", dpi=500)
plt.close()

with open("Blasthits.txt",'w') as pqr:
  for i in range(len(N)):
      pqr.write("%i\t%i\n" %(a[i],N[i]))

print "Analysis done"


#Removing temporary files
os.remove("temp2.txt")
os.remove("temp3.txt")
os.remove("original.txt")
```

Script 2 - Script for carrying out Principal Component Analysis after calculation of tetranucleotide frequencies using TETRA [32]

```
#This script requires the installation of numpy, biopython and
matplotlib
#This script requires the csv file from TETRA (Teeling H, Waldmann
J, Lombardot T, Bauer M, Glockner FO: TETRA: a web-service and a
stand-alone program for the analysis and comparison of
tetranucleotide usage patterns in DNA sequences. BMC bioinformatics
2004, 5(1):163.)
#This script is licensed under GNU GPL v3
import os
import numpy as np
from mpl_toolkits.mplot3d import Axes3D
import matplotlib.pyplot as plt
from Bio.Cluster import pca  # For PCA analysis


filename = input('Enter name of csv file from TETRA (enclose name in
""):')
with open("temp.txt",'w') as ijk:
            for line in open(filename):  # Getting rid of unwanted
characters and spaces
                line = line.replace(";","\t")
                line = line.replace("#","")
                line = line.replace("_","")
                ijk.write("%s" % (line))
ijk.close()
data = np.loadtxt("temp.txt", unpack = "True")
columnmean, coordinates, components, eigenvalues = pca(data) # PCA
analysis
eigenlist = eigenvalues
length = len(eigenlist)
coord = np.zeros(3)
b=0
for i in range(3):   # Calculating the 3 greatest eigenvalues to be
chosen as the axes for the PCA graph
    a = 0
    for j in xrange(length):
              if eigenlist[j] >= a:
                 a = eigenlist[j]

    for k in xrange(length):
              if eigenlist[k] == a:
                 coord[b] = k
                 eigenlist[k] = 0
                 b+=1
```

```
# Plotting results of PCA analysis
fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')
plt.rc('font', family='serif', size=13)
with open("coordinates.txt",'w') as xyz:
   for i in range(length):
       xs = coordinates[i,coord[0]]
       ys = coordinates[i,coord[1]]
       zs = coordinates[i,coord[2]]
       ax.scatter(xs, ys, zs, c='c', marker='o')
       xyz.write("%g\t%f\t %f\t%f\n" %(i+1,xs,ys,zs)) # Writing
coordinates
   ax.set_xlabel('PC_1')
   ax.set_ylabel('PC_2')
   ax.set_zlabel('PC_3')
   plt.savefig("PCA.png")

xyz.close()
os.remove("temp.txt")
```

Script 3 - Script for selecting closely clustered scaffolds in PCA graph (Figure 2-2).

```
#This script requires installation of numpy, biopython and
matplotlib libraries
#This script is licensed under the GNU GPLv3
#This script requires the file "coordinates.txt" which is an output
of PCA.py
from Bio import SeqIO
from Bio.SeqUtils import GC
from mpl_toolkits.mplot3d import Axes3D
import matplotlib.pyplot as plt

import numpy as np
import math
import os


def distance(x,y,z,x0,y0,z0): # Calculating distance between two
scaffolds
    x1 = x - x0
    y1 = y - y0
    z1 = z - z0
    return math.sqrt(x1*x1 + y1*y1 + z1*z1)


x = np.zeros(3)
#Input fasta file
filename1 = input('Enter name of fasta file (enclose name in ""):')
x = input('Enter the coordinates (e.g [-36.048591,-2.548105,0.6]:')
threshold = input('What is the threshold radius?:')
print "Coordinates chosen are:",x[0],x[1],x[2]
print "Threshold radius is:", threshold

seq_records1 = SeqIO.parse(filename1, "fasta")
seq_records1 = list(seq_records1)
length = len(seq_records1)

coordinates = np.loadtxt("coordinates.txt", unpack=True)
s = coordinates[0]
xs = coordinates[1]
ys = coordinates[2]
zs = coordinates[3]

plt.rc('font', family='serif', size=13)
# Plotting results of selected genes via tetranucleotide PCA
analysis
fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')
```

```
with open("PCAcontigs.fasta","w") as mno:
  with open("temp.txt","w") as xyz:
       for i in range(length):
           if distance(xs[i],ys[i],zs[i],x[0],x[1],x[2]) <=
threshold:  # See if the gene is within the prescribed coordinates
                  mno.write("%s %s\n %s\n" %
(">",seq_records1[i].id, seq_records1[i].seq)) # Making the fasta
file in PCA contigs
                  xyz.write("%s %f\n" % (len(seq_records1[i].seq),
GC(seq_records1[i].seq)))#Calculating G+C content and saving it to
temp.txt
                  ax.scatter(xs[i], ys[i], zs[i], c='r', marker='o')
           else:
                  ax.scatter(xs[i], ys[i], zs[i], c='b', marker='o')
mno.close()
xyz.close()
blue_proxy = plt.Rectangle((0, 0), 1, 1, fc="b")
red_proxy = plt.Rectangle((0, 0), 1, 1, fc="r")
ax.legend([red_proxy,blue_proxy],['L.HI','Contaminants'])
ax.set_xlabel('PC_1')
ax.set_ylabel('PC_2')
ax.set_zlabel('PC_3')
plt.savefig("PCA2.png")
plt.clf()
abcd = np.loadtxt("temp.txt",dtype = float)


with open("PCAcontigs.txt","w") as pqrs:
  for r1 in range(100):
    r2 = r1+1
    tot = 0
    for j in range(len(abcd)):
       if abcd[j][1]>r1 and abcd[j][1]<=r2:
            tot += abcd[j][0]
    pqrs.write("%i\t%i\n" %(r2,tot))
pqrs.close()

xyz = np.loadtxt("PCAcontigs.txt", unpack=True)

plt.plot(xyz[0],xyz[1]/1000000,'r-', lw=2)
plt.ylabel("Total length of scaffolds (in Mbp)")
plt.xlabel("%(G+C) content")


# Save figure as high resolution PNG
plt.savefig("PCAcontigsGC.png", dpi=500)
plt.close()
```

Script 4

This script separates gene scaffolds on the basis of a user-defined %(G+C) cut-off. It creates two fasta files one containing gene sequences below this cut-off and the other above this cut-off.

```
#This script requires installation of numpy, matplotlib and
biopython libraries
#This script is licensed under the GNU GPLv3
from Bio import SeqIO
from Bio.SeqUtils import GC
import matplotlib.pyplot as plt
import numpy as np
import math
import os

# Enter name of the fasta file
filename1 = input('Enter name of fasta file (enclose name in ""):')
seq_records1 = SeqIO.parse(filename1, "fasta")
seq_records1 = list(seq_records1)

x = input('Enter cutoff %(G+C):') # Asks for the G+C cut-off value

with open("PCAcontigs1.fasta","w") as abc:
  with open("temp1.txt","w") as xyz:
    with open("PCAcontigs2.fasta","w") as ghi:
      with open("temp2.txt","w") as jkl:
        for i in range(len(seq_records1)): # Seperating contigs on
the basis of the given cutoff
                        if GC(seq_records1[i].seq) < x: # Enforcing
the cut-off criteria
                           abc.write("%s %s\n %s\n" %
(">",seq_records1[i].id, seq_records1[i].seq)) # Making the fasta
file
                           xyz.write("%i %f\n" %
(len(seq_records1[i].seq), GC(seq_records1[i].seq)))#Calculating G+C
content
                        else :
                           ghi.write("%s %s\n %s\n" %
(">",seq_records1[i].id, seq_records1[i].seq)) # Making the fasta
file
                           jkl.write("%i %f\n" %
(len(seq_records1[i].seq), GC(seq_records1[i].seq)))#Calculating G+C
content

abc.close()
xyz.close()
ghi.close()
```

```
jkl.close()


temp1 = np.loadtxt("temp1.txt",dtype = float)

with open("PCAcontigs1.txt","w") as pqrs:
  for r1 in range(100):
    r2 = r1+1
    tot = 0
    for j in range(len(temp1)):
       if temp1[j][1]>r1 and temp1[j][1]<=r2:
            tot += temp1[j][0]
    pqrs.write("%i\t%i\n" %(r2,tot))
pqrs.close()

efgh = np.loadtxt("PCAcontigs1.txt", unpack=True)

os.remove("temp1.txt")

temp2 = np.loadtxt("temp2.txt",dtype = float)

with open("PCAcontigs2.txt","w") as pqrs:
  for r1 in range(100):
    r2 = r1+1
    tot = 0
    for j in range(len(temp2)):
       if temp2[j][1]>r1 and temp2[j][1]<=r2:
            tot += temp2[j][0]
    pqrs.write("%i\t%i\n" %(r2,tot))
pqrs.close()

uvwx = np.loadtxt("PCAcontigs2.txt", unpack=True)

plt.plot(efgh[0],efgh[1]/1000000,'r-', lw=2)
plt.plot(uvwx[0],uvwx[1]/1000000,'g-', lw=2)
plt.ylabel("Total length of scaffolds (in Mbp)")
plt.xlabel("%(G+C) content")


# Save figure as high resolution PNG
plt.savefig("GCseparation.png", dpi=500)
plt.clf()

os.remove("temp2.txt")
```

Script 5

Script for calculating energy transfer in phycoerythrin

```
# run /home/robin/SkyDrive/Phycoerythrin/energy
transfer/trimers/pymoloverlap5.py
# You will need numpy and biopython to run this script. This code
will only work from the pymol commandline.
import numpy as np
from pymol import cmd, stored
from Bio.PDB.PDBParser import PDBParser
import os
selection = "PE"

#Input path of PDB file
path = "/home/robin/SkyDrive/Phycoerythrin/energy transfer/trimers/"
filename = path + "PERP_refine_38.pdb" #Input name of pdb file
#filename = path + "HK.pdb" #Input name of pdb file
cmd.load(filename,selection)

p = PDBParser(PERMISSIVE=1)
structure_id = "PE"

# parsing PDB file
structure = p.get_structure(structure_id, filename)
model=structure[0]

def getcosine(xi,yi,zi,xj,yj,zj):
    R1 = np.sqrt(xi*xi+yi*yi+zi*zi)
    R2 = np.sqrt(xj*xj+yj*yj+zj*zj)
    return ((xi*xj+yi*yj+zi*zj)/(R1*R2))


# This function assigns line color of energy transfer according to
the ket value
def drawline(atomno1,atomno2,ket):
    cmd.distance("d%i%i"%(atomno1,atomno2),"id %i" %(atomno1),"id
%i" %(atomno2))
    cmd.set("dash_width",50)
    cmd.hide("labels","d%i%i"%(atomno1,atomno2))
    if ket>=400:
       cmd.color("blue","d%i%i"%(atomno1,atomno2))
       return 0

    elif ket>=100:
       cmd.color("red","d%i%i"%(atomno1,atomno2))
       return 0
    elif ket>=10:
       cmd.color("purple","d%i%i"%(atomno1,atomno2))
       return 0
```

```
    else:
       cmd.color("green","d%i%i"%(atomno1,atomno2))
       return 0

def rb_cartoon(selection):
    cmd.hide("everything",selection)
    cmd.set("dot_width",0.00001)
    cmd.set("bg_rgb",[1,1,1]) #white background
    cmd.select("res","resname PUB or resname CYC or resname BLA")
    cmd.show("sticks","res")
    cmd.set("fog_start",0.0) #remove fog
    cmd.set("fog",0.0) #remove fog
    #cmd.do("color red,%s"%(selection))
    cmd.color("red","resname BLA and resi 166")#a84
    cmd.color("green","resname CYC and resi 168")#a140
    cmd.color("brown","resname PUB and resi 200") #b50/61
    cmd.color("hotpink","resname CYC and resi 201")#b84
    cmd.color("deepblue","resname CYC and resi 202")#b155

# Calculate number of chromophores
n=0
for chain in model:
 for residue in chain:
   if residue.get_resname() == "CYC" or "BLA" or "PUB":
     for atoms in residue:
        if atoms.get_name() == "NA":
           n+=1


h=0
b = ".txt"
c = str(h+1) + b

resnam = []
ch = []
resno = []
atomno = []
m=1
# Loop for calculating the coordinates of the atoms in the pyrrole
of the chromophore

for chain in model:
 for residue in chain:
  if residue.get_resname() == "CYC":
     m+=1
     resnam.append("CYC")
     ch.append(chain)
     resno.append(residue.get_id())
     c = path + str(h+1) + b
     with open(c,'w') as xyz:
       i=0
       j=0
       k=0
       l=0
```

```
        for atoms in residue:
            if atoms.get_name() == "NA":
               coordinatesaN = atoms.get_coord()
               i+=1
            elif atoms.get_name() == "C1A":
               coordinatesa1 = atoms.get_coord()
               i+=1
            elif atoms.get_name() == "C2A":
               coordinatesa2 = atoms.get_coord()
               i+=1
            elif atoms.get_name() == "C3A":
               coordinatesa3 = atoms.get_coord()
               i+=1
            elif atoms.get_name() == "C4A":
               coordinatesa4 = atoms.get_coord()
               i+=1
            elif atoms.get_name() == "CHA":
               coordinatescha = atoms.get_coord()
               atomno.append(atoms.get_serial_number())
               i+=1
            elif atoms.get_name() == "NB":
               coordinatesbN = atoms.get_coord()
               j+=1
            elif atoms.get_name() == "C1B":
               coordinatesb1 = atoms.get_coord()
               j+=1
            elif atoms.get_name() == "C2B":
               coordinatesb2 = atoms.get_coord()
               j+=1
            elif atoms.get_name() == "C3B":
               coordinatesb3 = atoms.get_coord()
               j+=1
            elif atoms.get_name() == "C4B":
               coordinatesb4 = atoms.get_coord()
               j+=1
            elif atoms.get_name() == "CHB":
               coordinateschb = atoms.get_coord()
               j+=1
            elif atoms.get_name() == "NC":
               coordinatescN = atoms.get_coord()
               k+=1
            elif atoms.get_name() == "C1C":
               coordinatesc1 = atoms.get_coord()
               k+=1
            elif atoms.get_name() == "C2C":
               coordinatesc2 = atoms.get_coord()
               k+=1
            elif atoms.get_name() == "C3C":
               coordinatesc3 = atoms.get_coord()
               k+=1
            elif atoms.get_name() == "C4C":
               coordinatesc4 = atoms.get_coord()
               k+=1
            elif atoms.get_name() == "ND":
```

```
            coordinatesdN = atoms.get_coord()
            l+=1
         elif atoms.get_name() == "C1D":
            coordinatesd1 = atoms.get_coord()
            l+=1
         elif atoms.get_name() == "C2D":
            coordinatesd2 = atoms.get_coord()
            l+=1
         elif atoms.get_name() == "C3D":
            coordinatesd3 = atoms.get_coord()
            l+=1
         elif atoms.get_name() == "C4D":
            coordinatesd4 = atoms.get_coord()
            l+=1
         elif atoms.get_name() == "CHD":
            coordinateschd = atoms.get_coord()
            l+=1

         if i==6:

           xyz.write("%f\t%f\t%f\n" %
(coordinatesaN[0],coordinatesaN[1],coordinatesaN[2]))
           xyz.write("%f\t%f\t%f\n" %
(coordinatesa1[0],coordinatesa1[1],coordinatesa1[2]))
           xyz.write("%f\t%f\t%f\n" %
(coordinatesa2[0],coordinatesa2[1],coordinatesa2[2]))
           xyz.write("%f\t%f\t%f\n" %
(coordinatesa3[0],coordinatesa3[1],coordinatesa3[2]))
           xyz.write("%f\t%f\t%f\n" %
(coordinatesa4[0],coordinatesa4[1],coordinatesa4[2]))
           xyz.write("%f\t%f\t%f\n" %
(coordinatescha[0],coordinatescha[1],coordinatescha[2]))
           i=0
         if j==6:

           xyz.write("%f\t%f\t%f\n" %
(coordinatesbN[0],coordinatesbN[1],coordinatesbN[2]))
           xyz.write("%f\t%f\t%f\n" %
(coordinatesb1[0],coordinatesb1[1],coordinatesb1[2]))
           xyz.write("%f\t%f\t%f\n" %
(coordinatesb2[0],coordinatesb2[1],coordinatesb2[2]))
           xyz.write("%f\t%f\t%f\n" %
(coordinatesb3[0],coordinatesb3[1],coordinatesb3[2]))
           xyz.write("%f\t%f\t%f\n" %
(coordinatesb4[0],coordinatesb4[1],coordinatesb4[2]))
           xyz.write("%f\t%f\t%f\n" %
(coordinateschb[0],coordinateschb[1],coordinateschb[2]))
           j=0
         if k==5:

           xyz.write("%f\t%f\t%f\n" %
(coordinatescN[0],coordinatescN[1],coordinatescN[2]))
           xyz.write("%f\t%f\t%f\n" %
(coordinatesc1[0],coordinatesc1[1],coordinatesc1[2]))
```

```
            xyz.write("%f\t%f\t%f\n" %
(coordinatesc2[0],coordinatesc2[1],coordinatesc2[2]))
            xyz.write("%f\t%f\t%f\n" %
(coordinatesc3[0],coordinatesc3[1],coordinatesc3[2]))
            xyz.write("%f\t%f\t%f\n" %
(coordinatesc4[0],coordinatesc4[1],coordinatesc4[2]))
            k=0
          if l==6:

            xyz.write("%f\t%f\t%f\n" %
(coordinatesdN[0],coordinatesdN[1],coordinatesdN[2]))
            xyz.write("%f\t%f\t%f\n" %
(coordinatesd1[0],coordinatesd1[1],coordinatesd1[2]))
            xyz.write("%f\t%f\t%f\n" %
(coordinatesd2[0],coordinatesd2[1],coordinatesd2[2]))
            xyz.write("%f\t%f\t%f\n" %
(coordinatesd3[0],coordinatesd3[1],coordinatesd3[2]))
            xyz.write("%f\t%f\t%f\n" %
(coordinatesd4[0],coordinatesd4[1],coordinatesd4[2]))
            xyz.write("%f\t%f\t%f\n" %
(coordinateschd[0],coordinateschd[1],coordinateschd[2]))
            l=0
    h+=1
    xyz.close()
 elif residue.get_resname() == "BLA":
    m+=1
    resnam.append("BLA")
    ch.append(chain)
    resno.append(residue.get_id())
    c = path + str(h+1) + b
    with open(c,'w') as pqr:
      i=0
      j=0
      k=0
      l=0
      for atoms in residue:
          if atoms.get_name() == "NA":
             coordinatesaN = atoms.get_coord()
             i+=1
          elif atoms.get_name() == "C1A":
             coordinatesa1 = atoms.get_coord()
             i+=1
          elif atoms.get_name() == "C2A":
             coordinatesa2 = atoms.get_coord()
             i+=1
          elif atoms.get_name() == "C3A":
             coordinatesa3 = atoms.get_coord()
             i+=1
          elif atoms.get_name() == "C4A":
             coordinatesa4 = atoms.get_coord()
             i+=1
          elif atoms.get_name() == "CHA":
             coordinatescha = atoms.get_coord()
             atomno.append(atoms.get_serial_number())
```

```
            i+=1
        elif atoms.get_name() == "NB":
            coordinatesbN = atoms.get_coord()
            j+=1
        elif atoms.get_name() == "C1B":
            coordinatesb1 = atoms.get_coord()
            j+=1
        elif atoms.get_name() == "C2B":
            coordinatesb2 = atoms.get_coord()
            j+=1
        elif atoms.get_name() == "C3B":
            coordinatesb3 = atoms.get_coord()
            j+=1
        elif atoms.get_name() == "C4B":
            coordinatesb4 = atoms.get_coord()
            j+=1
        elif atoms.get_name() == "CHB":
            coordinateschb = atoms.get_coord()
            j+=1
        elif atoms.get_name() == "NC":
            coordinatescN = atoms.get_coord()
            k+=1
        elif atoms.get_name() == "C1C":
            coordinatesc1 = atoms.get_coord()
            k+=1
        elif atoms.get_name() == "C2C":
            coordinatesc2 = atoms.get_coord()
            k+=1
        elif atoms.get_name() == "C3C":
            coordinatesc3 = atoms.get_coord()
            k+=1
        elif atoms.get_name() == "C4C":
            coordinatesc4 = atoms.get_coord()
            k+=1
        elif atoms.get_name() == "ND":
            coordinatesdN = atoms.get_coord()
            l+=1
        elif atoms.get_name() == "C1D":
            coordinatesd1 = atoms.get_coord()
            l+=1
        elif atoms.get_name() == "C2D":
            coordinatesd2 = atoms.get_coord()
            l+=1
        elif atoms.get_name() == "C3D":
            coordinatesd3 = atoms.get_coord()
            l+=1
        elif atoms.get_name() == "C4D":
            coordinatesd4 = atoms.get_coord()
            l+=1
        elif atoms.get_name() == "CHD":
            coordinateschd = atoms.get_coord()
            l+=1
```

```
          if i==6:

            pqr.write("%f\t%f\t%f\n" %
(coordinatesaN[0],coordinatesaN[1],coordinatesaN[2]))
            pqr.write("%f\t%f\t%f\n" %
(coordinatesa1[0],coordinatesa1[1],coordinatesa1[2]))
            pqr.write("%f\t%f\t%f\n" %
(coordinatesa2[0],coordinatesa2[1],coordinatesa2[2]))
            pqr.write("%f\t%f\t%f\n" %
(coordinatesa3[0],coordinatesa3[1],coordinatesa3[2]))
            pqr.write("%f\t%f\t%f\n" %
(coordinatesa4[0],coordinatesa4[1],coordinatesa4[2]))
            pqr.write("%f\t%f\t%f\n" %
(coordinatescha[0],coordinatescha[1],coordinatescha[2]))
            i=0
          if j==6:

            pqr.write("%f\t%f\t%f\n" %
(coordinatesbN[0],coordinatesbN[1],coordinatesbN[2]))
            pqr.write("%f\t%f\t%f\n" %
(coordinatesb1[0],coordinatesb1[1],coordinatesb1[2]))
            pqr.write("%f\t%f\t%f\n" %
(coordinatesb2[0],coordinatesb2[1],coordinatesb2[2]))
            pqr.write("%f\t%f\t%f\n" %
(coordinatesb3[0],coordinatesb3[1],coordinatesb3[2]))
            pqr.write("%f\t%f\t%f\n" %
(coordinatesb4[0],coordinatesb4[1],coordinatesb4[2]))
            pqr.write("%f\t%f\t%f\n" %
(coordinateschb[0],coordinateschb[1],coordinateschb[2]))
            j=0
          if k==5:

            pqr.write("%f\t%f\t%f\n" %
(coordinatescN[0],coordinatescN[1],coordinatescN[2]))
            pqr.write("%f\t%f\t%f\n" %
(coordinatesc1[0],coordinatesc1[1],coordinatesc1[2]))
            pqr.write("%f\t%f\t%f\n" %
(coordinatesc2[0],coordinatesc2[1],coordinatesc2[2]))
            pqr.write("%f\t%f\t%f\n" %
(coordinatesc3[0],coordinatesc3[1],coordinatesc3[2]))
            pqr.write("%f\t%f\t%f\n" %
(coordinatesc4[0],coordinatesc4[1],coordinatesc4[2]))
            k=0
          if l==6:

            pqr.write("%f\t%f\t%f\n" %
(coordinatesdN[0],coordinatesdN[1],coordinatesdN[2]))
            pqr.write("%f\t%f\t%f\n" %
(coordinatesd1[0],coordinatesd1[1],coordinatesd1[2]))
            pqr.write("%f\t%f\t%f\n" %
(coordinatesd2[0],coordinatesd2[1],coordinatesd2[2]))
            pqr.write("%f\t%f\t%f\n" %
(coordinatesd3[0],coordinatesd3[1],coordinatesd3[2]))
```

```
            pqr.write("%f\t%f\t%f\n" %
(coordinatesd4[0],coordinatesd4[1],coordinatesd4[2]))
            pqr.write("%f\t%f\t%f\n" %
(coordinateschd[0],coordinateschd[1],coordinateschd[2]))
            l=0
    h+=1
    pqr.close()

 elif residue.get_resname() == "PUB":
    m+=1
    resnam.append("PUB")
    ch.append(chain)
    resno.append(residue.get_id())
    c = path + str(h+1) + b
    with open(c,'w') as mno:
      i=0
      j=0
      k=0
      l=0
      for atoms in residue:
          if atoms.get_name() == "NA":
             coordinatesaN = atoms.get_coord()
             i+=1
          elif atoms.get_name() == "C1A":
             coordinatesa1 = atoms.get_coord()
             i+=1
          elif atoms.get_name() == "C2A":
             coordinatesa2 = atoms.get_coord()
             i+=1
          elif atoms.get_name() == "C3A":
             coordinatesa3 = atoms.get_coord()
             i+=1
          elif atoms.get_name() == "C4A":
             coordinatesa4 = atoms.get_coord()
             i+=1
          elif atoms.get_name() == "CHA":
             coordinatescha = atoms.get_coord()
             i+=1
          elif atoms.get_name() == "NB":
             coordinatesbN = atoms.get_coord()
             j+=1
          elif atoms.get_name() == "C1B":
             coordinatesb1 = atoms.get_coord()
             j+=1
          elif atoms.get_name() == "C2B":
             coordinatesb2 = atoms.get_coord()
             j+=1
          elif atoms.get_name() == "C3B":
             coordinatesb3 = atoms.get_coord()
             j+=1
          elif atoms.get_name() == "C4B":
             coordinatesb4 = atoms.get_coord()
             j+=1
          elif atoms.get_name() == "CHB":
```

```
               coordinateschb = atoms.get_coord()
               atomno.append(atoms.get_serial_number())
               j+=1
            elif atoms.get_name() == "NC":
               coordinatescN = atoms.get_coord()
               k+=1
            elif atoms.get_name() == "C1C":
               coordinatesc1 = atoms.get_coord()
               k+=1
            elif atoms.get_name() == "C2C":
               coordinatesc2 = atoms.get_coord()
               k+=1
            elif atoms.get_name() == "C3C":
               coordinatesc3 = atoms.get_coord()
               k+=1
            elif atoms.get_name() == "C4C":
               coordinatesc4 = atoms.get_coord()
               k+=1
            elif atoms.get_name() == "CHC":
               coordinateschc = atoms.get_coord()
               k+=1
            elif atoms.get_name() == "ND":
               coordinatesdN = atoms.get_coord()
               l+=1
            elif atoms.get_name() == "C1D":
               coordinatesd1 = atoms.get_coord()
               l+=1
            elif atoms.get_name() == "C2D":
               coordinatesd2 = atoms.get_coord()
               l+=1
            elif atoms.get_name() == "C3D":
               coordinatesd3 = atoms.get_coord()
               l+=1
            elif atoms.get_name() == "C4D":
               coordinatesd4 = atoms.get_coord()
               l+=1

            if i==6:

              mno.write("%f\t%f\t%f\n" %
(coordinatesaN[0],coordinatesaN[1],coordinatesaN[2]))
              mno.write("%f\t%f\t%f\n" %
(coordinatesa1[0],coordinatesa1[1],coordinatesa1[2]))
              mno.write("%f\t%f\t%f\n" %
(coordinatesa2[0],coordinatesa2[1],coordinatesa2[2]))
              mno.write("%f\t%f\t%f\n" %
(coordinatesa3[0],coordinatesa3[1],coordinatesa3[2]))
              mno.write("%f\t%f\t%f\n" %
(coordinatesa4[0],coordinatesa4[1],coordinatesa4[2]))
              mno.write("%f\t%f\t%f\n" %
(coordinatescha[0],coordinatescha[1],coordinatescha[2]))
              i=0
            if j==6:
```

```
            mno.write("%f\t%f\t%f\n" %
(coordinatesbN[0],coordinatesbN[1],coordinatesbN[2]))
            mno.write("%f\t%f\t%f\n" %
(coordinatesb1[0],coordinatesb1[1],coordinatesb1[2]))
            mno.write("%f\t%f\t%f\n" %
(coordinatesb2[0],coordinatesb2[1],coordinatesb2[2]))
            mno.write("%f\t%f\t%f\n" %
(coordinatesb3[0],coordinatesb3[1],coordinatesb3[2]))
            mno.write("%f\t%f\t%f\n" %
(coordinatesb4[0],coordinatesb4[1],coordinatesb4[2]))
            mno.write("%f\t%f\t%f\n" %
(coordinateschb[0],coordinateschb[1],coordinateschb[2]))
            j=0
          if k==6:

            mno.write("%f\t%f\t%f\n" %
(coordinatescN[0],coordinatescN[1],coordinatescN[2]))
            mno.write("%f\t%f\t%f\n" %
(coordinatesc1[0],coordinatesc1[1],coordinatesc1[2]))
            mno.write("%f\t%f\t%f\n" %
(coordinatesc2[0],coordinatesc2[1],coordinatesc2[2]))
            mno.write("%f\t%f\t%f\n" %
(coordinatesc3[0],coordinatesc3[1],coordinatesc3[2]))
            mno.write("%f\t%f\t%f\n" %
(coordinatesc4[0],coordinatesc4[1],coordinatesc4[2]))
            mno.write("%f\t%f\t%f\n" %
(coordinateschc[0],coordinateschc[1],coordinateschc[2]))
            k=0
          if l==5:

            mno.write("%f\t%f\t%f\n" %
(coordinatesdN[0],coordinatesdN[1],coordinatesdN[2]))
            mno.write("%f\t%f\t%f\n" %
(coordinatesd1[0],coordinatesd1[1],coordinatesd1[2]))
            mno.write("%f\t%f\t%f\n" %
(coordinatesd2[0],coordinatesd2[1],coordinatesd2[2]))
            mno.write("%f\t%f\t%f\n" %
(coordinatesd3[0],coordinatesd3[1],coordinatesd3[2]))
            mno.write("%f\t%f\t%f\n" %
(coordinatesd4[0],coordinatesd4[1],coordinatesd4[2]))
            l=0
    h+=1
    mno.close()
  else:
    m+=1
    continue

X = np.zeros(n)
Y = np.zeros(n)
Z = np.zeros(n)
avg = np.zeros((n,3))

# Loop for calculating the best fit line for all the pyrrole's in
the chromophore
```

```
for I in range(n):
  a = str(I+1)
  abc = np.loadtxt(path+a+".txt",unpack = True)
  xcf = abc[0]
  ycf = abc[1]
  zcf = abc[2]

  data = np.concatenate((xcf[:, np.newaxis],
                         ycf[:, np.newaxis],
                         zcf[:, np.newaxis]),
                        axis=1)
  # Calculate the mean of the points, i.e. the 'center' of the cloud
  datamean = data.mean(axis=0)

  # Do an SVD on the mean-centered data.
  uu, dd, vv = np.linalg.svd(data - datamean)

  # Now vv[0] contains the first principal component, i.e. the
direction
  # vector of the 'best fit' line in the least squares sense.

  # Now generate some points along this best fit line, for plotting.

  #linepts = vv[0] * np.mgrid[-7:7:2j][:, np.newaxis]
  linepts = vv[0] * zcf[:, np.newaxis]

  # shift by the mean to get the line in the right place
  linepts += datamean

  X[I] = linepts[1][0] - linepts[0][0]
  Y[I] = linepts[1][1] - linepts[0][1]
  Z[I] = linepts[1][2] - linepts[0][2]

  # Calculating mid-point of chromophore
  for j in range(len(abc[0])):
     avg[I][0] += abc[0][j]
     avg[I][1] += abc[1][j]
     avg[I][2] += abc[2][j]
  avg[I][0] = avg[I][0]/len(abc[0])
  avg[I][1] = avg[I][1]/len(abc[1])
  avg[I][2] = avg[I][2]/len(abc[2])

# Loop for calculating the energy transfer between two chromophores

with open(path + "dodecamer.txt",'w') as pqr:
 #pqr.write("%s %i %s" %("Number of chromophores is:",n,"\n"))
 #pqr.write("%s\t%s\t%s\t%s\t%s\t%s\t%s\t%s\t%s\t%s\n" %("Chain
A","Residue name A","Pos A","|","Chain B","Residue name B","Pos
B","R (in Angstroms)","k2","Energy transfer (in ns-1)"))
 for I in range(n):
   #J=I+1
   for J in range(n):
     # x-component of line joining X and Y
     # y-component of line joining X and Y
```

```
     # z-component of line joining X and Y
     ri = X[I] - X[J]
     rj = Y[I] - Y[J]
     rk = Z[I] - Z[J]
     cos0T = getcosine(X[I],Y[I],Z[I],X[J],Y[J],Z[J])
     cos0D = getcosine(X[I],Y[I],Z[I],ri,rj,rk)
     cos0A = getcosine(X[J],Y[J],Z[J],ri,rj,rk)
     #print cos0T,cos0D,cos0A
     k2 = (cos0T - 3*cos0D*cos0A)*(cos0T - 3*cos0D*cos0A)
     R = np.sqrt((avg[I][0]-avg[J][0])*(avg[I][0]-
avg[J][0])+(avg[I][1]-avg[J][1])*(avg[I][1]-avg[J][1])+(avg[I][2]-
avg[J][2])*(avg[I][2]-avg[J][2]))


     str1 = str(resno[I])
     str1 = str1[6:9]
     str2 = str(resno[J])
     str2 = str2[6:9]

     if str1 == "201":
        c1 = "b84"
        q1 = 0.25          # Inner b84 chromophore
        td1 = 2.12
        e1 = 0.97
        residuename1 = "PEB565"
     elif str1 == "200":
        c1 = "b50/61"
        q1 = 0.2           # PUB chromophore
        td1 = 3.78
        e1 = 1.83
        residuename1 = "PUB495"
     elif str1 == "202":
        c1 = "b155"
        q1 = 0.28          # Outer b155 chromophore
        td1 = 2.02
        e1 = 1.08
        residuename1 = "PEB540"
     elif str1 == "166":
        c1 = "a84"
        q1 = 0.25          # Outer a84 chromophore
        td1 = 2.12
        e1 = 0.97
        residuename1 = "PEB565"
     elif str1 == "168":
        c1 = "a140"
        q1 = 0.28          # Outer a140 chromophore
        td1 = 2.02
        e1 = 1.08
        residuename1 = "PEB540"

     if str2 == "201":
        c2 = "b84"
        q2 = 0.25          # Inner b84 chromophore
        td2 = 2.12
```

```
        e2 = 0.97
        residuename2 = "PEB565"
     elif str2 == "200":
        c2 = "b50/61"
        q2 = 0.2           # PUB chromophore
        td2 = 3.78
        e2 = 1.83
        residuename2 = "PUB495"
     elif str2 == "202":
        c2 = "b155"
        q2 = 0.28          # Outer b155 chromophore
        td2 = 2.02
        e2 = 1.08
        residuename2 = "PEB540"
     elif str2 == "166":
        c2 = "a84"
        q2 = 0.25          # Outer a84 chromophore
        td2 = 2.12
        e2 = 0.97
        residuename2 = "PEB565"
     elif str2 == "168":
        c2 = "a140"
        q2 = 0.28          # Outer a140 chromophore
        td2 = 2.02
        e2 = 1.08
        residuename2 = "PEB540"

# Deciding overlap integral

     if q1 == 0.2 and q2 == 0.2:
        Ov = 0.388
     elif q1 == 0.2 and q2 == 0.28:
        Ov = 6.080
     elif q1 == 0.2 and q2 == 0.25:
        Ov = 2.015
     elif q1 == 0.28 and q2 == 0.2:
        Ov = 0.056
     elif q1 == 0.28 and q2 == 0.28:
        Ov = 1.962
     elif q1 == 0.28 and q2 == 0.25:
        Ov = 3.822
     elif q1 == 0.25 and q2 == 0.2:
        Ov = 0.005
     elif q1 == 0.25 and q2 == 0.28:
        Ov = 2.568
     elif q1 == 0.25 and q2 == 0.25:
        Ov = 2.395

     ket = 2.72*(k2/(R*R*R*R*R*R*td1))*q1*e2*Ov*(10**10) # Formula
for calculating energy transfer (in ns)


     if ket>=0:
```

```
       #pqr.write("%s\t%s\t%s\t%s\t%s\t%s\t%s\t%s\t%s\t%s\n"
%(ch[I],residuename1,c1,"|",ch[J],residuename2,c2,str(R),str(k2),str
(ket)))
       pqr.write("%s\t%s\t%s\t%s\t%s\t%s\n"
%(c1,c2,Ov,str(R),str(k2),str(1000/ket)))
       drawline(atomno[I],atomno[J],ket)
     J+=1
pqr.close()

# Removing temporary files of the chromophores
for i in range(n):
  c = str(i+1) + b
  os.remove(path+c)

#---------------------------------------------------------------------




# Set the background, color of chromophores in pymol
rb_cartoon("PE")



#cmd.quit()

# Spectroscopic data is from the article "Spectroscopic Parameters
of phycoerythrobilin and phycourobilin on phycoerythrin from
Gracilaria chilensis"
```

APPENDIX B: BLAST RESULTS OF PCR PRODUCTS OBTAINED BY USING 16S DNA PRIMERS ON GENOMIC DNA ISOLATED FROM UNAXENIC *L.HI* CELL CULTURES

(a) *Leptolyngbya* BLAST result

```
L.HI    57    CCTTCCGGTACGGCTACCTTGTTACGACTTCACCCCAGTCATCAGCCCTGCCTTCGGCGC  116
              ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||| |
L.7375  1438  CCTTCCGGTACG-CTACCTTGTTACGACTTCACCCCAGTCATCAGCCCTGCCTTCGGCAC  1380

L.HI    117   CCCCCTCCGAAAACGGTTAGGGTAACGACTTCGGGCATGGCCAACTTCCATGGTGTGACG  176
              |||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||
L.7375  1379  CCCCCTCCGCGAACGGTTAGGGTAACGACTTCGGGCATGGCCAACTTCCATGGTGTGACG  1320

L.HI    177   GGCGGCGTGTACAAGGCCCGGGAACGTATTCACCGCAGTATGCTGACCTGCGATTACTAG  236
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
L.7375  1319  GGCGGTGTGTACAAGGCCCGGGAACGTATTCACCGCAGTATGCTGACCTGCGATTACTAG  1260

L.HI    237   CGATTCCTCCTTCATGCAGGCGAGTTGCAGCCTGCAATCTGAACTGAGCAACGGTTTATG  296
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
L.7375  1259  CGATTCCTCCTTCATGCAGGCGAGTTGCAGCCTGCAATCTGAACTGAGCAACGGTTTATG  1200

L.HI    297   GGATTAGCTCACTATCGCTAGTTGGCTGCCCTTTGTCCGTTGCATTGTAGTACGTGTGTA  356
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
L.7375  1199  GGATTAGCTCACTATCGCTAGCTGGCTGCCCTTTGTCCGTTGCATTGTAGTACGTGTGTA  1140

L.HI    357   GCCCAGAACGTAAGGGGCATGATGACTTGACGTCGTCCACACCTTCCTCCGAGTTGTCCC  416
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
L.7375  1139  GCCCAGAACGTAAGGGGCATGATGACTTGACGTCGTCCACACCTTCCTCCGAGTTGTCCC  1080

L.HI    417   CGGCAGTCTCTCCAGAGTGCCCAACTGAATGATGGCAACTAAAGACGTGGGTTGCGCTCG  476
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
L.7375  1079  CGGCAGTCTCTCCAGAGTGCCCAACTGAATGATGGCAACTAAAGACGTGGGTTGCGCTCG  1020

L.HI    477   TTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAGCCATGCACCACCTGTC  536
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
L.7375  1019  TTGCGGGACTTAACCCAACATCTCACGACACGAGCTGACGACAGCCATGCACCACCTGTC  960

L.HI    537   TCCTCGTTCCCGAAGGCACTATCTAATTTCTTAGATATTCGAGGGATGTCAAGTCCTGGT  596
              |||||| |||||||||||||| ||||||||||||||||||||||||||||||||||||||
L.7375  959   TCCTCGCTCCCGAAGGCACTACCTAATTTCTTAGATATTCGAGGGATGTCAAGTCCTGGT  900

L.HI    597   AAGGTTCTTCGCGTTGCATCGAATTAAACCACATACTCCACCGCTTGTGCGGGCCCCCGT  656
              |||||||| ||||||||||||||||||||||||||||||||||||| |||||||||
L.7375  899   AAGGTTCTCCGCGTTGCATCGAATTAAACCACATACTCCACCGCTTGTGC-GGCCCCCGT  841

L.HI    657   CAATTCCTTTGAGTTTCACACTTGCGTGCGTACTCCCCAGGCGGGATACTTAACGCGTTT  716
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
L.7375  840   CAATTCCTTTGAGTTTCACACTTGCGTGCGTACTCCCCAGGCGGGATACTTAACGCGTTT  781

L.HI    717   GCTTCGGTACTGCACGGGTCGATACGCACAACACCTAGTATCCATCGTTTACAGCTAGGA  776
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
L.7375  780   GCTTCGGTACTGCACGGGTCGATACGCACAACACCTAGTATCCATCGTTTACAGCTAGGA  721

L.HI    777   CTACNNGGGTATCTAATCCCTTTCGCTNCCCTAGCTTTCGTCCATCAGCGTCAGTCTTGG  836
              ||||  |||||||||||||||||||| ||||||||||||||||||||||||||||||||
L.7375  720   CTACAGGGGTATCTAATCCCTTTCGCTCCCCTAGCTTTCGTCCATCAGCGTCAGTCTTGG  661
```

(b) Contaminant

```
Con 59  CGAACGCTGGCGGCAGGCCTAACACATGCAAGTCGAACGCCCTCTTCGGAGGGAGTGGCG  118
```

BLAST results of PCR products obtained using 27f and 1525r primers used for sequencing 16S rDNA. Obtained mainly the two PCR products: One that matched with (a) 16S rDNA gene of Leptolyngbya PCC 7375 and (b) 16S rDNA gene of mucus bacterium 96 (M.b). The 16S rDNA gene of the contaminant has been named as “Con” in table.

APPENDIX C: NUCLEOTIDE AND AMINO ACID SEQUENCE COMPARISON OF PHYCOERYTHRIN GENES BETWEEN DIFFERENT STRAINS OF *LEPTOLYNGBYA*

Sequence alignment of nucleotide and amino acid sequences of α and β-subunits of phycoerythrin from two different sets of phycoerythrin genes in *L.HI* and one gene from *L.KC45*. (*) denotes all three sets of genes have the same nucleotide at that position. (:) denotes that the nucleotide of *L.KC45* is different from that of the two sets of genes obtained from *L.HI*. (#) denotes that the nucleotide position is not conserved between the two sets of genes obtained from *L.HI*.

(a) PE α-subunit nucleotide sequence

```
L.HI1
ATGAAATCTGTTGTTACTACTGTGATCGCTGCTGCAGATGCAGCTGGTCGTTTTCCTTCT 60
L.HI2
ATGAAATCTGTTGTTACTACTGTGATCGCTGCTGCAGATGCAGCTGGTCGTTTTCCTTCC 60
L.KC45
ATGAAATCTGTTGTTACTACCGTGATTGCAGCGGCTGATGCAGCCGGACGGTTTCCGAGC 60

********************:*****:**:**:**:********:**:**:*****:::#

L.HI1
ACTTCCGACCTAGAGTCCGTACAAGGTAGCCTACAGCGTGCTGCTGCCCGTTTGGAAGCC 120
L.HI2
ACTTCTGACCTAGAGTCCGTACAAGGTAGCCTACAGCGTGCTGCTGCCCGTTTGGAAGCC 120
L.KC45
ACCTCGGATCTCGAATCGGTTCAGGGTTCGATCCAACGGGCCGCTGCTCGCCTAGAAGCT 120

**:**#**:**:**:**:**:**:***::::*:**:**:**:*****:**::*:*****:

L.HI1
GCTGAGAAGCTAGCTGGCAACTTGGACAACGTGGCTAAAGAAGCTTACGATGCTTGCATC 180
L.HI2
GCTGAGAAGCTAGCTGGCAACTTGGACAATGTGGCTAAAGAAGCTTACGATGCTTGCATC 180
L.KC45
GCAGAAAAGCTGGCCAACAACCTCGATAACGTAGCTCGGGAAGCTTATGATGCGTGCATC 180

**:**:*****:**:::****:*:**:**#**:***:::********:*****:******

L.HI1
TCCAAGTACCCCTACTTGAACAATGCGGGTGAAGCTAACTCCACTGACACGTTCAAGGCA 240
L.HI2
TCCAAGTACCCCTACTTGAACAATGCGGGTGAAGCTAACTCCACTGACACGTTCAAGGCA 240
L.KC45
AAGAAATATCCCTACCTGAACAATGCGGGTGAAGCAAACTCTACCGACACCTTCAAAGCC 240

:::**:**:******:********************:*****:**:*****:*****:**:

L.HI1
AAGTGCCTACGTGACGTGAAGCACTACATGCGTCTGATCAGCTACAGCCTAGTTGTGGGT 300
L.HI2
AAGTGCCTACGTGACGTGAAGCACTACATGCGTCTGATCAGCTACAGCCTAGTTGTGGGT 300
L.KC45
AAGTGTCTGCGTGACATCAAGCACTACATGCGCCTGATTCAATACTGCTTGGTGGTTGGC 300

*****:**:******:*:**************:*****::::***:**:*:**:**:**:

L.HI1
GGTACTGGTCCTCTGGATGAGTGGGGAATTGCCGGTCAGCGTGAAGTATACCGTGCTCTG 360
L.HI2
GGTACTGGTCCTCTGGATGAGTGGGGAATTGCCGGTCAGCGTGAAGTATACCGCGCTCTG 360
L.KC45
```

```
GGTACCGGTCCGCTCGATGAGTGGGGCATTGCTGGCCAGAAGGAAGTGTATCGTGCCCTC 360

*****:*****:**:**********:*****:**:***:::*****:**:**#**:**:

L.HI1
AACCTGCCTACTGCTCCTTACGTAGCGGCTCTAAGCTTTGCTCGTAATCGTGGTTGCGCT 420
L.HI2
AATCTACCTACTGCTCCTTACGTAGCGGCTCTAAGCTTTGCTCGTAATCGTGGTTGCGCT 420
L.KC45
GGTCTGCCCACCGCACCCTATGTCGAAGCCCTCAGCTTCGCTCGCAACCGTGGCTGTGCT 420

::#**#**:**:**:**:**:**:*::**:**:*****:*****:**:*****:**:***

L.HI1
CCTCGTGACATGTCTGCTCAGGCTCTGACTGAGTACAACGCACTGGTTGACTATGTAATC 480
L.HI2
CCTCGTGACATGTCTGCTCAGGCTCTGACTGAGTACAACGCACTGGTTGACTATGTAATC 480
L.KC45
CCGCGCGATATGTCGGCTCAAGCCCTGACCGAGTACAACGCCCTGCTCGACTACGCCATC 480

**:**:**:*****:*****:**:*****:**********:***:*:*****:*::***

L.HI1                AACTCTCTTTCCTAG 495
L.HI2                AACTCTCTTTCCTAG 495
L.KC45               AACTCGCTGTCCTAG 495
                     *****:**:******
```

(b) PE α-subunit protein sequence

```
L.HI1
MKSVVTTVIAAADAAGRFPSTSDLESVQGSLQRAAARLEAAEKLAGNLDNVAKEAYDACI 60
L.HI2
MKSVVTTVIAAADAAGRFPSTSDLESVQGSLQRAAARLEAAEKLAGNLDNVAKEAYDACI 60
L.KC45
MKSVVTTVIAAADAAGRFPSTSDLESVQGSIQRAAARLEAAEKLANNLDNVAREAYDACI 60

*****************************:**************:******:*******

L.HI1
SKYPYLNNAGEANSTDTFKAKCLRDVKHYMRLISYSLVVGGTGPLDEWGIAGQREVYRAL 120
L.HI2
SKYPYLNNAGEANSTDTFKAKCLRDVKHYMRLISYSLVVGGTGPLDEWGIAGQREVYRAL 120
L.KC45
KKYPYLNNAGEANSTDTFKAKCLRDIKHYMRLIQYCLVVGGTGPLDEWGIAGQKEVYRAL 120

:************************:*******:*:*****************:******

L.HI1            NLPTAPYVAALSFARNRGCAPRDMSAQALTEYNALVDYVINSL 163
L.HI2            NLPTAPYVAALSFARNRGCAPRDMSAQALTEYNALVDYVINSL 163
L.KC45           GLPTAPYVEALSFARNRGCAPRDMSAQALTEYNALLDYAINSL 163
                 :*******:****************************:**:****
```

(c) PE β-subunit nucleotide sequence

```
L.HI1
ATGCTTGACGCTTTTTCTAGAGCGGTAGTATCCGCTGATGCCAGTACTGCACCCGTAGGT 60
L.HI2
ATGCTTGACGCTTTTTCTAGAGCGGTAGTGTCCGCTGATGCCAGTACTGCACCCGTAGGT 60
L.KC45
ATGCTTGACGCTTTTTCTAGAGCTGTTGTCGCAGCCGATGCCAGCACGTCTGTTGTATCA 60
```

```
**********************:**:**#:*:**:********:**::*::::***:::

L.HI1
GACCTAGCTGCTCTTAAGGCATTTGTTGCTAGCGGCAACCGTCGTCTAGATGCTGTAAAC 120
L.HI2
GACTTAGCTGCTCTTAAGGCATTTGTTGCTAGCGGCAACCGTCGTCTAGATGCTGTAAAC 120
L.KC45
GATATCAGTGCATTAAGAAAGTTTGTGGCCGAAGGAAACCGTCGTCTGGATGCAGTGAAT 120

**:#*:::***::*:*::::::*****:**::::**:**********:*****:**:**:

L.HI1
GCAATTGCATCCAACGCTAGCTGCATGGTGTCCGATGCGATCGCTGGCATGATCTGTGAG 180
L.HI2
GCAATTGCATCCAACGCTAGCTGCATGGTATCCGATGCTATCGCTGGCATGATCTGTGAG 180
L.KC45
GCGATTGCGTCGAACGCGAGCTGCATGGTCTCGGACGCGGTAGCAGGCATGATCTGTGAA 180

**:*****:**:*****:**********#**:**:**#:*:**:**************:

L.HI1
AACCAGGGTTTGATCCAAGCTGGTGGTAACTGCTACCCCAACCGTCGTATGGCGGCTTGC 240
L.HI2
AACCAGGGTTTGATCCAAGCTGGTGGTAACTGCTACCCCAACCGTCGTATGGCGGCTTGC 240
L.KC45
AACCAAGGCTTGATCCAAGCGGGTGGCAACTGCTATCCCAACCGTCGGATGGCCGCTTGC 240

*****:**:**********:*****:********:**********:*****:******

L.HI1
CTCCGTGACGGCGAGATCGTTCTACGTTATGTTACCTACGCTCTACTAGCTGGTGATGCG 300
L.HI2
CTACGTGACGGCGAAATCATCTTGCGCTATGTTACCTACGCTCTGCTAGCTGGTGATGCG 300
L.KC45
CTGCGCGATGCCGAAATCATTTTGCGCTATGTCACCTATGCGCTGTTGGCTGGTGATGCC 300

**#**:**:*:***#***#*##*#**#*****:*****:**:**#:*:***********:

L.HI1
TCTGTTCTCGATGATCGTTGCTTGAACGGTCTCAAAGAGACTTACGCGGCTCTAGGCGTA 360
L.HI2
TCTGTTCTCGATGATCGTTGCTTGAACGGTCTCAAAGAGACTTACGCGGCTCTAGGCGTA 360
L.KC45
TCGGTGCTTGACGATCGCTGCCTCAACGGGTTGAAAGAAACCTATGCAGCGCTTGGCGTG 360

**:**:**:**:*****:***:*:*****::*:*****:**:**:**:**:**:*****:

L.HI1
CCTGCAACTTCCACTGTACGTGCGGTACAGATCATGAAGGCTCAAGCAGCTGCTCACATT 420
L.HI2
CCTGCAACTTCCACTGTACGCGCGGTACAGATCATGAAGGCTCAAGCAGCTGCTCACATT 420
L.KC45
CCCACCACCTCCACCGTGCGGGCGGTGCAAATCATGAAAGCGCAAGCAGCCGCTCATATT 420

**::*:**:*****:**:**#*****:**:********:**:********:*****:***

L.HI1
AAGGATGAGCCTAGCGAAGCTCGTGCTGGTGCTCGTCTACGTAAGATGGGCTCCACTGTA 480
L.HI2
AAGGATGAGCCTAGCGAAGCTCGTGCTGGTGCTCGTCTACGTAAGATGGGCTCCATTGTA 480
L.KC45
CAAGACAACCCCAGCGAGCAATTTGCCGGTGCCAAACTGCGCAAGATGGGAACTCCCGTC 480
```

```
:*:**::*:**:*****:::::::***:*****::::**:**:********::*::#:**:

L.HI1
ACTGAAGATCGTTGTGCTTCCTTGGTTGCTGAAGCTTCTAGCTACTTTGATCGCGTTATC 540
L.HI2
ACTGAAGATCGTTGTGCATCTTTGGTTGCTGAAGCTTCTAGCTACTTTGATCGCGTTATC 540
L.KC45
GTTGAGGATCGCTGCGCCAGCCTGGTGGCCGAAGCATCGAGCTACTTTGATCGCGTGATT 540

::***:*****:**:**#::#:****:**:*****:**:******************:**:

L.HI1                                              TCTGCTCTAGGTTAG 555
L.HI2                                              TCTGCTCTAGGTTAG 555
L.KC45                                             TCGGCCCTCAGCTAG 555
                                                   **:**:**::*:***
```

(d) PE β-subunit protein sequence

```
L.HI1
MLDAFSRAVVSADASTAPVGDLAALKAFVASGNRRLDAVNAIASNASCMVSDAIAGMICE 60
L.HI2
MLDAFSRAVVSADASTAPVGDLAALKAFVASGNRRLDAVNAIASNASCMVSDAIAGMICE 60
L.KC45
MLDAFSRAVVAADASTSVVSDISALRKFVAEGNRRLDAVNAIASNASCMVSDAVAGMICE 60

**********:*****::*:*::**::***:**********************:******

L.HI1
NQGLIQAGGNCYPNRRMAACLRDGEIVLRYVTYALLAGDASVLDDRCLNGLKETYAALGV 120
L.HI2
NQGLIQAGGNCYPNRRMAACLRDGEIILRYVTYALLAGDASVLDDRCLNGLKETYAALGV 120
L.KC45
NQGLIQAGGNCYPNRRMAACLRDAEIILRYVTYALLAGDASVLDDRCLNGLKETYAALGV 120

**********************:**#**********************************

L.HI1
PATSTVRAVQIMKAQAAAHIKDEPSEARAGARLRKMGSTVTEDRCASLVAEASSYFDRVI 180
L.HI2
PATSTVRAVQIMKAQAAAHIKDEPSEARAGARLRKMGSIVTEDRCASLVAEASSYFDRVI 180
L.KC45
PTTSTVRAVQIMKAQAAAHIQDNPSEQFAGAKLRKMGTPVVEDRCASLVAEASSYFDRVI 180

*:******************:*:***::***:*****:#*:*******************

L.HI1                                                        SAL 183
L.HI2                                                        SAL 183
L.KC45                                                       SAL 183
                                                             ***
```

APPENDIX D: SEQUENCE HOMOLOGY OF α AND β-SUBUNIT BETWEEN PHYCOERYTHRIN FROM *L.HI* AND *POLYSIPHONIA* URCEOLATA

α-subunit

L.HI

X Y

KSVVTTVIAAADAAGRFPSTSDLESVQGSLQRAAARLEAAEKLAGNLDNVAKEAYDACISK

*P.*urceolata

MKSVITTTISAADAAGRYPSTSDLQSVQGNIQRAAARLEAAEKLGSNHEAVVKEAGDACFSK
*** ** * ******* ****** **** ************* * * *** *** **

L.HI

A B C

YPYLNNAGEANSTDTFKAKCLRDVKHYMRLISYSLVVGGTGPLDEWGIAGQREVYRALNLP

PEB

*P.*urceolata

YGYNKNPGEAGENQEKINKCYRDIDHYMRLINYTLVVGGTGPLDEWGIAGAREVYRTLNLP
* * * *** ** ** ****** * ********** ***** ****

L.HI

D E

TAPYVAALSFARNRGCAPRDMSAQALTEYNALVDYVINSL-

PEB

*P.*urceolata

SAAYIAAFVFTRDRLCIPRDMSAQAGVEFCTALDYLINSLS
* * ** * * * * ******** * ** ****

β-subunit

L.HI

X Y

MLDAFSRAVVSADASTAPVG--DLAALKAFVASGNRRLDAVNAIASNASCMVSDAIAGMICE

PUB

*P.*urceolata

MLDAFSRVVVNSDSKAAYVSGSDLQALKTFINDGNKRLDAVNYIVSNSSCIVSDAISGMICE
******* ** * * * ** *** * ** ****** * ** ** ***** *****

L.HI

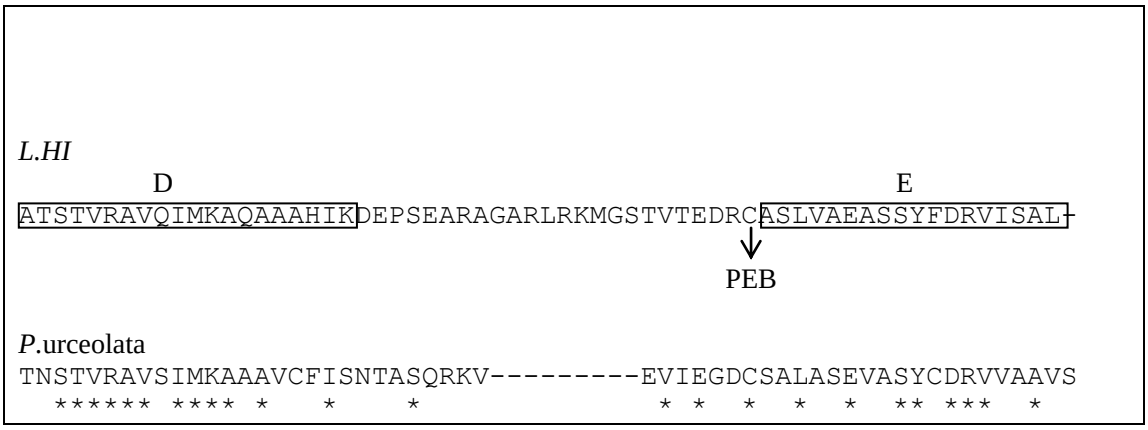
A B C

NQGLIQAGG-CYPNRRMAACLRDGEIVLRYVTYALLAGDASVLDDRCLNGLKETYAALGVP

PEB

*P.*urceolata

NPGLITPGGNCYTNRRMAACLRDGEIILRYVSYALLAGDASVLEDRCLNGLKETYIALGVP
* *** ** ** ********** **** ********** ********** *****


L.HI
D
E
ATSTVRAVQIMKAQAAAHIKDEPSEARAGARLRKMGSTVTEDRCASLVAEASSYFDRVISAL
PEB
P.urceolata
TNSTVRAVSIMKAAAVCFISNTASQRKV---------EVIEGDCSALASEVASYCDRVVAAVS

.

REFERENCES

1. Fromme P: **Photosynthetic protein complexes: a structural approach**: John Wiley & Sons; 2008.

2. Blankenship RE: **Molecular mechanisms of photosynthesis**: Wiley. com; 2008.

3. Boyer PD: **The ATP synthase-a splendid molecular machine**. *Annual review of biochemistry* 1997, **66**(1):717-749.

4. Yoshida M, Muneyuki E, Hisabori T: **ATP synthase—a marvellous rotary engine of the cell**. *Nature Reviews Molecular Cell Biology* 2001, **2**(9):669-677.

5. Jordan P, Fromme P, Witt HT, Klukas O, Saenger W, Krauß N: **Three-dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution**. *Nature* 2001, **411**(6840):909-917.

6. Robin Paul, Robert E. Jinkerson, Kristina Buss, Jason Steel, Remus Mohr, Wolfgang R. Hess, Min Chen, Fromme P: **Draft Genome Sequence of the Filamentous Cyanobacterium *Leptolyngbya* sp. Strain Heron Island J, Exhibiting Chromatic Acclimation**. *Genome Announcements* 2014, **2**(1):e01166-01113.

7. Farquhar G, von Caemmerer Sv, Berry J: **A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species**. *Planta* 1980, **149**(1):78-90.

8. Bald D, Kruip J, Rögner M: **Supramolecular architecture of cyanobacterial thylakoid membranes: How is the phycobilisome connected with the photosystems?** *Photosynthesis research* 1996, **49**(2):103-118.

9. Raven JA, Evans MC, Korb RE: **The role of trace metals in photosynthetic electron transport in O2-evolving organisms**. *Photosynthesis Research* 1999, **60**(2-3):111-150.

10. Kurisu G, Zhang H, Smith JL, Cramer WA: **Structure of the cytochrome b6f complex of oxygenic photosynthesis: tuning the cavity**. *Science* 2003, **302**(5647):1009-1014.

11. Gantt E: **Phycobilisomes**. *Annual Review of Plant Physiology* 1981, **32**(1):327-347.

12. Sepúlveda-Ugarte J, Brunet JE, Matamala AR, Martínez-Oyanedel J, Bunster M: **Spectroscopic parameters of phycoerythrobilin and phycourobilin on phycoerythrin from *Gracilaria chilensis***. *Journal of Photochemistry and Photobiology A: Chemistry* 2011, **219**(2):211-216.

13. Bennett A, Bogorad L: **Complementary chromatic adaptation in a filamentous blue-green alga**. *The journal of cell biology* 1973, **58**(2):419-435.

14. Kehoe DM, Gutu A: **Responding to color: the regulation of complementary chromatic adaptation**. *Annu Rev Plant Biol* 2006, **57**:127-150.

15. Kehoe DM, Grossman AR: **Similarity of a chromatic adaptation sensor to phytochrome and ethylene receptors**. *Science* 1996, **273**(5280):1409-1412.

16. Schirrmeister B, Antonelli A, Bagheri H: **The origin of multicellularity in cyanobacteria**. *BMC evolutionary biology* 2011, **11**(1):45.

17. Margulis L: **Serial endosymbiotic theory (SET) and composite individuality**. *Microbiology Today* 2004, **31**(4):172-175.

18. Giovannoni S, Turner S, Olsen G, Barns S, Lane D, Pace N: **Evolutionary relationships among cyanobacteria and green chloroplasts**. *Journal of Bacteriology* 1988, **170**(8):3584-3592.

19. Rippka R, Deruelles J, Waterbury JB, Herdman M, Stanier RY: **Generic assignments, strain histories and properties of pure cultures of cyanobacteria**. *Journal of General microbiology* 1979, **111**(1):1-61.

20. Ignacio-Espinoza JC, Sullivan MB: **Phylogenomics of T4 cyanophages: lateral gene transfer in the 'core'and origins of host genes**. *Environmental microbiology* 2012, **14**(8):2113-2126.

21. Sullivan MB, Waterbury JB, Chisholm SW: **Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus***. *Nature* 2003, **424**(6952):1047-1051.

22. Mann NH, Cook A, Millard A, Bailey S, Clokie M: **Marine ecosystems: bacterial photosynthesis genes in a virus**. *Nature* 2003, **424**(6950):741-741.

23. Lochte K, Turley C: **Bacteria and cyanobacteria associated with phytodetritus in the deep sea**. 1988.

24. Garczarek L, van der Staay GW, Hess WR, Le Gall F, Partensky F: **Expression and phylogeny of the multiple antenna genes of the low-light-adapted strain *Prochlorococcus* marinus SS120 (Oxyphotobacteria)**. *Plant molecular biology* 2001, **46**(6):683-693.

25. Droop M: **A procedure for routine purification of algal cultures with antibiotics**. *British Phycological Bulletin* 1967, **3**(2):295-297.

26. Rainey FA, Ward-Rainey NL, Janssen PH, Hippe H, Stackebrandt E: **Clostridium paradoxum DSM 7308T contains multiple 16S rRNA genes with heterogeneous intervening sequences**. *Microbiology* 1996, **142**(8):2087-2095.

27. Buhler JD, Lancaster JM, Jacob AC, Chamberlain RD: **Mercury BLASTN: Faster DNA sequence comparison using a streaming hardware architecture**. *Proc of Reconfigurable Systems Summer Institute* 2007.

28. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol İ: **ABySS: a parallel assembler for short read sequence data**. *Genome research* 2009, **19**(6):1117-1123.

29. Van Der Walt S, Colbert SC, Varoquaux G: **The NumPy array: a structure for efficient numerical computation**. *Computing in Science & Engineering* 2011, **13**(2):22-30.

30. Hunter JD: **Matplotlib: A 2D graphics environment**. *Computing in Science & Engineering* 2007, **9**(3):0090-0095.

31. Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B: **Biopython: freely available Python tools for computational molecular biology and bioinformatics**. *Bioinformatics* 2009, **25**(11):1422-1423.

32. Teeling H, Waldmann J, Lombardot T, Bauer M, Glöckner FO: **TETRA: a web-service and a stand-alone program for the analysis and comparison of tetranucleotide usage patterns in DNA sequences**. *BMC bioinformatics* 2004, **5**(1):163.
33. Ringnér M: **What is principal component analysis?** *Nature biotechnology* 2008, **26**(3):303-304.

34. Kunin V, Copeland A, Lapidus A, Mavromatis K, Hugenholtz P: **A bioinformatician's guide to metagenomics**. *Microbiology and Molecular Biology Reviews* 2008, **72**(4):557-578.

35. Shnit-Orland M, Kushmaro A: **Coral mucus-associated bacteria: a possible first line of defense**. *FEMS microbiology ecology* 2009, **67**(3):371-380.

36. Ishida T, Watanabe MM, Sugiyama J, Yokota A: **Evidence for polyphyletic origin of the members of the orders of *Oscillatoriales* and *Pleurocapsales* as determined by 16S rDNA analysis**. *FEMS Microbiology Letters* 2001, **201**(1):79-82.

37. Baran R, Ivanova NN, Jose N, Garcia-Pichel F, Kyrpides NC, Gugger M, Northen TR: **Functional Genomics of Novel Secondary Metabolites from Diverse Cyanobacteria Using Untargeted Metabolomics**. *Marine drugs* 2013, **11**(10):3617-3631.

38. Meeks JC, Elhai J, Thiel T, Potts M, Larimer F, Lamerdin J, Predki P, Atlas R: **An overview of the genome of *Nostoc* punctiforme, a multicellular, symbiotic cyanobacterium**. *Photosynthesis research* 2001, **70**(1):85-106.

39. Teeling H, Meyerdierks A, Bauer M, Amann R, Glöckner FO: **Application of tetranucleotide frequencies for the assignment of genomic fragments**. *Environmental microbiology* 2004, **6**(9):938-947.

40. Zhaxybayeva O, Gogarten JP, Charlebois RL, Doolittle WF, Papke RT: **Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events**. *Genome Research* 2006, **16**(9):1099-1108.

41. Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Nelson WC, Ketchum KA: **Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga* maritima**. *Nature* 1999, **399**(6734):323-329.

42. Sorek R, Kunin V, Hugenholtz P: **CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea**. *Nature Reviews Microbiology* 2008, **6**(3):181-186.

43. Gundlach K, Werwie M, Wiegand S, Paulsen H: **Filling the "green gap" of the major light-harvesting chlorophyll *a/b* complex by covalent attachment of Rhodamine Red**. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 2009, **1787**(12):1499-1504.

44. Chen H, Dang W, Xie J, Zhao J, Weng Y: **Ultrafast energy transfer pathways in R-phycoerythrin from *Polysiphonia* urceolata**. *Photosynthesis research* 2012, **111**(1-2):81-86.

45. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs**. *Nucleic acids research* 1997, **25**(17):3389-3402.

46. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C: **Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data**. *Bioinformatics* 2012, **28**(12):1647-1649.

47. Chayen NE: **Turning protein crystallisation from an art into a science**. *Current Opinion in Structural Biology* 2004, **14**(5):577-583.

48. Kabsch W: **Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants**. *Journal of applied crystallography* 1993, **26**(6):795-800.

49. Ritter S, Hiller RG, Wrench PM, Welte W, Diederichs K: **Crystal structure of a phycourobilin-containing phycoerythrin at 1.90-Å resolution**. *Journal of structural biology* 1999, **126**(2):86-97.

50. Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Moriarty NW, Zwart PH, Hung L-W, Read RJ, Adams PD: **Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard**. *Acta Crystallographica Section D: Biological Crystallography* 2007, **64**(1):61-69.

51. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW: **PHENIX: a comprehensive Python-based system for**

macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography* 2010, **66**(2):213-221.

52. Humphrey W, Dalke A, Schulten K: **VMD: visual molecular dynamics**. *Journal of molecular graphics* 1996, **14**(1):33-38.

53. Debreczeny MP, Sauer K, Zhou J, Bryant DA: **Comparison of Calculated and Experimentally Resolved Rate Constants for Excitation Energy Transfer in C-Phycocyanin. 1. Monomers**. *The Journal of Physical Chemistry* 1995, **99**(20):8412-8419.

54. Debreczeny MP, Sauer K, Zhou J, Bryant DA: **Comparison of calculated and experimentally resolved rate constants for excitation energy transfer in C-phycocyanin. 2. Trimers**. *The Journal of Physical Chemistry* 1995, **99**(20):8420-8431.

55. MacColl R: **Cyanobacterial phycobilisomes**. *Journal of structural biology* 1998, **124**(2):311-334.

56. Ramachandran G, Ramakrishnan Ct, Sasisekharan V: **Stereochemistry of polypeptide chain configurations**. *Journal of molecular biology* 1963, **7**(1):95-99.

57. Duerring M, Schmidt GB, Huber R: **Isolation, crystallization, crystal structure analysis and refinement of constitutive C-phycocyanin from the chromatically adapting cyanobacterium *Fremyella* diplosiphon at 1.66 Å resolution**. *Journal of molecular biology* 1991, **217**(3):577-592.

58. Stec B, Troxler RF, Teeter MM: **Crystal structure of C-phycocyanin from *Cyanidium* caldarium provides a new perspective on phycobilisome assembly**. *Biophysical journal* 1999, **76**(6):2912-2921.

59. Baker D, Sali A: **Protein structure prediction and structural genomics**. *Science* 2001, **294**(5540):93-96.

60. Chang W-r, Jiang T, Wan Z-l, Zhang J-p, Yang Z-x, Liang D-c: **Crystal Structure of R-phycoerythrin from *Polysiphonia* urceolata at 2.8 Å Resolution**. *Journal of molecular biology* 1996, **262**(5):721-722.

61. de Marsac NT: **Phycobiliproteins and phycobilisomes: the early observations**. *Photosynthesis research* 2003, **76**(1-3):193-205.

62. Grossman AR: **A molecular understanding of complementary chromatic adaptation**. *Photosynthesis research* 2003, **76**(1-3):207-215.
63. Gutu A, Kehoe DM: **Emerging perspectives on the mechanisms, regulation, and distribution of light color acclimation in cyanobacteria**. *Molecular plant* 2012, **5**(1):1-13.

64. Hirose Y, Narikawa R, Katayama M, Ikeuchi M: **Cyanobacteriochrome CcaS regulates phycoerythrin accumulation in *Nostoc* punctiforme, a group II chromatic adapter**. *Proceedings of the National Academy of Sciences* 2010, **107**(19):8854-8859.

65. Terauchi K, Montgomery BL, Grossman AR, Lagarias JC, Kehoe DM: **RcaE is a complementary chromatic adaptation photoreceptor required for green and red light responsiveness**. *Molecular microbiology* 2004, **51**(2):567-577.

66. De Marsac NT: **Occurrence and nature of chromatic adaptation in cyanobacteria**. *Journal of Bacteriology* 1977, **130**(1):82-91.

67. Everroad C, Six C, Partensky F, Thomas J-C, Holtzendorff J, Wood AM: **Biochemical bases of type IV chromatic adaptation in marine *Synechococcus* spp**. *Journal of bacteriology* 2006, **188**(9):3345-3356.

68. Shukla A, Biswas A, Blot N, Partensky F, Karty JA, Hammad LA, Garczarek L, Gutu A, Schluchter WM, Kehoe DM: **Phycoerythrin-specific bilin lyase–isomerase controls blue-green chromatic acclimation in marine *Synechococcus***. *Proceedings of the National Academy of Sciences* 2012, **109**(49):20136-20141.

69. Palenik B: **Chromatic Adaptation in Marine *Synechococcus* Strains**. *Applied and environmental microbiology* 2001, **67**(2):991-994.

70. Mullineaux CW: **Excitation energy transfer from phycobilisomes to photosystem I in a cyanobacterium**. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1992, **1100**(3):285-292.

71. Williams WP, Allen JF: **State 1/state 2 changes in higher plants and algae**. *Photosynthesis Research* 1987, **13**(1):19-45.

72. Fujita Y, Murakami A, Aizawa K, Ohki K: **Short-term and long-term adaptation of the photosynthetic apparatus: homeostatic properties of thylakoids**. In: *The molecular biology of cyanobacteria.* Springer; 1994: 677-692.

73. Murakami A, Fujita Y: **Steady state of photosynthetic electron transport in cells of the cyanophyti Synechocystis PCC 6714 having different stoichiometry between PS I and PS II: analysis of flash-induced oxidation-reduction of cytochrome f and P700 under steady state of photosynthesis**. *Plant and cell physiology* 1991, **32**(2):213-222.

74. Murakami A: **Quantitative analysis of 77K fluorescence emission spectra in Synechocystis sp. PCC 6714 and *Chlamydomonas* reinhardtii with variable PS I/PS II stoichiometries**. *Photosynthesis research* 1997, **53**(2-3):141-148.

75. Campbell D, Houmard J, De Marsac NT: **Electron transport regulates cellular differentiation in the filamentous cyanobacterium *Calothrix***. *The Plant Cell Online* 1993, **5**(4):451-463.

76. Damerval T, Guglielmi G, Houmard J, De Marsac NT: **Hormogonium differentiation in the cyanobacterium *Calothrix*: a photoregulated developmental process**. *The Plant Cell Online* 1991, **3**(2):191-201.

77. Murata N, Nishimura M, Takamiya A: **Fluorescence of chlorophyll in photosynthetic systems III. Emission and action spectra of fluorescence—Three emission bands of chlorophyll *a* and the energy transfer between two pigment systems**. *Biochimica et Biophysica Acta (BBA)-Biophysics including Photosynthesis* 1966, **126**(2):234-243.

78. Brecht M, Radics V, Nieder JB, Studier H, Bittl R: **Red Antenna States of Photosystem I from *Synechocystis* PCC 6803†**. *Biochemistry* 2008, **47**(20):5536-5543.

79. Myers J, Graham J-R, Wang RT: **Light harvesting in *Anacystis* nidulans studied in pigment mutants**. *Plant physiology* 1980, **66**(6):1144-1149.

80. Fujita Y, Ohki K, Murakami A: **Chromatic regulation of photosystem composition in the photosynthetic system of red and blue-green algae**. *Plant and cell physiology* 1985, **26**(8):1541-1548.

81. Cunningham FX, Dennenberg RJ, Jursinic PA, Gantt E: **Growth under red light enhances photosystem II relative to photosystem I and phycobilisomes in the red alga *Porphyridium* cruentum**. *Plant physiology* 1990, **93**(3):888-895.

82. Melis A, Murakami A, Nemson JA, Aizawa K, Ohki K, Fujita Y: **Chromatic regulation in *Chlamydomonas* reinhardtii alters photosystem stoichiometry and improves the quantum efficiency of photosynthesis**. *Photosynthesis research* 1996, **47**(3):253-265.

83. Glick RE, McCauley SW, Gruissem W, Melis A: **Light quality regulates expression of chloroplast genes and assembly of photosynthetic membrane complexes**. *Proceedings of the National Academy of Sciences* 1986, **83**(12):4287-4291.

84. Chow W, Anderson JM, Hope A: **Variable stoichiometries of photosystem II to photosystem I reaction centres**. *Photosynthesis research* 1988, **17**(3):277-281.

85. Stackebrandt E, Goodfellow M: **Nucleic acid techniques in bacterial systematics**, vol. 4: John Wiley & Son Ltd; 1991.

86. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods**. *Molecular biology and evolution* 2011, **28**(10):2731-2739.

87. Bellezza S, Paradossi G, De Philippis R, Albertano P: ***Leptolyngbya* strains from Roman hypogea: cytochemical and physico-chemical characterisation of exopolysaccharides**. *Journal of applied phycology* 2003, **15**(2-3):193-200.

88. Takagi I, Yamada K, Sato T, Hanaichi T, Iwamoto T, Jin L: **Penetration and stainability of modified Sato's lead staining solution**. *Journal of electron microscopy* 1990, **39**(1):67-68.

89. Bryant DA, Guglielmi G, de Marsac NT, Castets A-M, Cohen-Bazire G: **The structure of cyanobacterial phycobilisomes: a model**. *Archives of Microbiology* 1979, **123**(2):113-127.

90. Zhao J, Shen G, Bryant DA: **Photosystem stoichiometry and state transitions in a mutant of the cyanobacterium *Synechococcus* sp. PCC 7002 lacking phycocyanin**. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 2001, **1505**(2):248-257.

91. Zhao J, Brand JJ: **Specific bleaching of phycobiliproteins from cyanobacteria and red algae at high temperature in vivo**. *Archives of Microbiology* 1989, **152**(5):447-452.

92. De Marsac NT, Houmard J: **[34] Complementary chromatic adaptation: Physiological conditions and action spectra**. *Methods in enzymology* 1988, **167**:318-328.

93. Truesdale G, Downing A: **Solubility of oxygen in water**. 1954.

94. Turner S, Pryer KM, Miao VP, Palmer JD: **Investigating Deep Phylogenetic Relationships among Cyanobacteria and Plastids by Small Subunit rRNA Sequence Analysis1**. *Journal of Eukaryotic Microbiology* 1999, **46**(4):327-338.

95. Tomitani A, Knoll AH, Cavanaugh CM, Ohno T: **The evolutionary diversification of cyanobacteria: molecular–phylogenetic and paleontological perspectives**. *Proceedings of the National Academy of Sciences* 2006, **103**(14):5442-5447.

96. Ligon PJ, Meyer KG, Martin JA, Curtis SE: **Nucleotide sequence of a 16S rRNA gene from *Anabaena* sp. strain PCC 7120**. *Nucleic acids research* 1991, **19**(16):4553-4553.

97. Nelissen B, Van de Peer Y, Wilmotte A, De Wachter R: **An early origin of plastids within the cyanobacterial divergence is suggested by evolutionary trees based on complete 16S rRNA sequences**. *Molecular Biology and Evolution* 1995, **12**(6):1166-1173.

98. Urbach E, Scanlan DJ, Distel DL, Waterbury JB, Chisholm SW: **Rapid diversification of marine picophytoplankton with dissimilar light-harvesting structures inferred from sequences of *Prochlorococcus* and *Synechococcus* (Cyanobacteria)**. *Journal of molecular evolution* 1998, **46**(2):188-201.

99. Nelissen B, De Baere R, Wilmotte A, De Wachter R: **Phylogenetic relationships of nonaxenic filamentous cyanobacterial strains based on 16S rRNA sequence analysis**. *Journal of molecular evolution* 1996, **42**(2):194-200.

100. Iteman I, Rippka R, de Marsac NT, Herdman M: **rDNA analyses of planktonic heterocystous cyanobacteria, including members of the genera *Anabaenopsis* and *Cyanospira***. *Microbiology* 2002, **148**(2):481-496.

101. Hosoya R, Hamana K, Isobe M, Yokota A: **Polyamine distribution profiles within cyanobacteria**. *Microbiol Cult Coll* 2005, **21**:3-8.

102. Ducret A, Sidler W, Wehrli E, Frank G, Zuber H: **Isolation, characterization and electron microscopy analysis of a hemidiscoidal phycobilisome type from the cyanobacterium *Anabaena* sp. PCC 7120**. *European Journal of Biochemistry* 1996, **236**(3):1010-1024.

103. Khanna R, Graham J-R, Myers J, Gantt E: **Phycobilisome composition and possible relationship to reaction centers**. *Archives of Biochemistry and Biophysics* 1983, **224**(2):534-542.

104. French C, Smith JH, Virgin HI, Airth RL: **Fluorescence-Spectrum Curves of Chlorophylls, Pheophytins, Phycoerythrins, Phycocyanins and Hypericin**. *Plant Physiology* 1956, **31**(5):369.

105. Mullineaux CW: **Excitation energy transfer from phycobilisomes to photosystem I in a cyanobacterial mutant lacking photosystem II**. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1994, **1184**(1):71-77.

106. Jordan P, Fromme P, Witt HT, Klukas O, Saenger W, Krauß N: **Three-dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution**. *Nature* 2001, **411**(6840):909-917.

107. Trissl H-W, Wilhelm C: **Why do thylakoid membranes from higher plants form grana stacks?** *Trends in biochemical sciences* 1993, **18**(11):415-419.

108. Gobets B, Kennis JT, Ihalainen JA, Brazzoli M, Croce R, van Stokkum IH, Bassi R, Dekker JP, van Amerongen H, Fleming GR: **Excitation energy transfer in dimeric light harvesting complex I: a combined streak-camera/fluorescence upconversion study**. *The Journal of Physical Chemistry B* 2001, **105**(41):10132-10139.

109. Vrettos JS, Stewart DH, de Paula JC, Brudvig GW: **Low-temperature optical and resonance Raman spectra of a carotenoid cation radical in photosystem II**. *The Journal of Physical Chemistry B* 1999, **103**(31):6403-6406.

110. Un S, Tang X-S, Diner BA: **245 GHz high-field EPR study of tyrosine-D and tyrosine-Z in mutants of photosystem II**. *Biochemistry* 1996, **35**(3):679-684.

111. Wollman F-A: **State transitions reveal the dynamics and flexibility of the photosynthetic apparatus**. *The EMBO journal* 2001, **20**(14):3623-3630.

112. Allen JF, Mullineaux CW, Sanders CE, Melis A: **State transitions, photosystem stoichiometry adjustment and non-photochemical quenching in cyanobacterial cells acclimated to light absorbed by photosystem I or photosystem II**. *Photosynthesis research* 1989, **22**(2):157-166.
113. Campbell D, Oquist G: **Predicting light acclimation in cyanobacteria from nonphotochemical quenching of photosystem II fluorescence, which reflects state transitions in these organisms**. *Plant physiology* 1996, **111**(4):1293-1298.

114. Joshua S, Mullineaux CW: **Phycobilisome diffusion is required for light-state transitions in cyanobacteria**. *Plant physiology* 2004, **135**(4):2112-2119.

115. Mullineaux CW, Holzwarth AR: **Kinetics of excitation energy transfer in the cyanobacterial phycobilisome-Photosystem II complex**. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1991, **1098**(1):68-78.

116. Zilinskas BA, Greenwald LS: **Phycobilisome structure and function**. *Photosynthesis research* 1986, **10**(1-2):7-35.

117. Pfannschmidt T, Nilsson A, Allen JF: **Photosynthetic control of chloroplast gene expression**. *Nature* 1999, **397**(6720):625-628.

118. Heid CA, Stevens J, Livak KJ, Williams PM: **Real time quantitative PCR**. *Genome research* 1996, **6**(10):986-994.

119. Bustin SA, Nolan T: **Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction**. *Journal of biomolecular techniques: JBT* 2004, **15**(3):155.

120. Bustin S: **Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems**. *Journal of molecular endocrinology* 2002, **29**(1):23-39.

121. Parmar A, Singh NK, Pandey A, Gnansounou E, Madamwar D: **Cyanobacteria and microalgae: a positive prospect for biofuels**. *Bioresource technology* 2011, **102**(22):10163-10172.

122. Machado IM, Atsumi S: **Cyanobacterial biofuel production**. *Journal of biotechnology* 2012, **162**(1):50-56.

123. Dismukes GC, Carrieri D, Bennette N, Ananyev GM, Posewitz MC: **Aquatic phototrophs: efficient alternatives to land-based crops for biofuels**. *Current Opinion in Biotechnology* 2008, **19**(3):235-240.

124. Larkum A: **Limitations and prospects of natural photosynthesis for bioenergy production**. *Current opinion in biotechnology* 2010, **21**(3):271-276.