

**Analyzing Structural and Functional Evolutionary links between
RuBisCO and Carbon Concentrating Mechanism (CCM) proteins**

THESIS

Submitted in partial fulfilment
of the requirements for the degree of
DOCTOR OF PHILOSOPHY

by

GURPREET KAUR SIDHU

Under the Supervision of
Dr. Sandhya Mehrotra



BITS Pilani
Pilani | Dubai | Goa | Hyderabad

**BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE
PILANI (RAJASTHAN) INDIA**

2016

**BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE,
PILANI**

CERTIFICATE

This thesis is submitted under Regulation 8.20 (a) of the Academic Regulations for Doctoral Programmes which allows a faculty member of the Institute Professional to do Ph.D. research without the benefit of a supervisor.

This is to certify that the thesis entitled “**Analyzing Structural and Functional Evolutionary links between RuBisCO and Carbon Concentrating Mechanism (CCM) proteins**” and submitted by **Gurpreet Kaur Sidhu** ID No **2011PHXF011P** for award of Ph.D. Degree of the Institute embodies my original work.

Signature

: *Gurpreet*

Name

: GURPREET KAUR SIDHU

ID No.

: 2011PHXF011P

Date

: **27.12.16**

**BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE,
PILANI**

CERTIFICATE

This is to certify that the thesis entitled “Analyzing Structural and Functional Evolutionary links between RuBisCO and Carbon Concentrating Mechanism (CCM) proteins” and submitted by Gurpreet Kaur Sidhu, ID No 2011PHXF011P for award of Ph.D. of the Institute embodies original work done by her under my supervision.

Signature of the Supervisor

: 

Name (Supervisor)

: SANDHYA MEHROTRA, Ph.D.

Designation

: Assistant Professor,

Department of Biological Sciences,

BITS Pilani, Pilani Campus.

Date

: 27/12/2016

CONTENTS

| | |
|--|------|
| Preface | i |
| Acknowledgement | iv |
| List of Tables | vii |
| List of Figures | x |
| Abbreviations | xv |
| Chapter 1 Introduction and Review of Literature | |
| 1.1 Introduction | 1-2 |
| 1.2 <i>Gloeobacter violaceus</i> PCC 7421 | 1-3 |
| 1.3 RuBisCO- A Relic of Bygone Age | 1-5 |
| 1.3.1 The distinct forms of RuBisCO | 1-5 |
| 1.3.2 The active site of RuBisCO | 1-6 |
| 1.3.3 The mechanism of action | 1-7 |
| 1.3.4 RuBisCO activase | 1-8 |
| 1.3.5 Assembly of RuBisCO | 1-8 |
| 1.3.6 Selection of RuBisCO towards a better enzyme if not a perfect one | 1-9 |
| 1.4 Carbon Concentrating Mechanisms (CCMs) | 1-10 |
| 1.4.1 CCMs in cyanobacteria | 1-10 |
| 1.4.1.1 Two types of carboxysomes with distinct composition and evolution | 1-11 |
| 1.4.1.2 Components of carboxysomes | 1-14 |
| 1.4.1.3 Biogenesis of carboxysomes | 1-17 |
| 1.4.1.4 Regulation of carboxysome genes | 1-18 |
| 1.4.1.5 Microcompartments | 1-18 |
| 1.4.2 CCMs in proteobacteria | 1-19 |
| 1.4.3 CCMs in algae | 1-20 |
| 1.4.3.1 CCM based on enhancement of CO ₂ concentration by acidification | 1-20 |
| 1.4.3.2 CCM in diatoms | 1-21 |
| 1.4.4 CCMs in higher plants | 1-22 |
| 1.4.5 C ₃ plants do not have any CCM | 1-23 |
| 1.5 Gaps in existing research | 1-23 |
| 1.6 Objectives of present research | 1-24 |
| Chapter 2 Materials and Methods | |
| 2.1 <i>In-silico</i> analysis of RuBisCO and CCM proteins of cyanobacteria | 2-2 |
| 2.1.1 The distribution of CCM proteins across Eubacteria and Archaea | 2-2 |
| 2.1.2 Identification of conserved domains and phylogenetic analysis | 2-2 |
| 2.1.3 The study of arrangement of RuBisCO encoding genes in | 2-2 |

| | | |
|--------|---|------|
| | cyanobacteria to determine if it is random or has functional significance | |
| 2.1.4 | The analysis for evolutionary links between RuBisCO and CCM proteins | 2-3 |
| 2.1.5 | <i>In-silico</i> modeling of CCM proteins and RuBisCO proteins of <i>Gloeobacter violaceus</i> PCC 7421 | 2-3 |
| 2.1.6 | Designing of primers for amplification of RuBisCO and CCM encoding genes of <i>Gloeobacter violaceus</i> PCC 7421 | 2-4 |
| 2.2 | <i>In-vitro In-vivo</i> analysis | 2-6 |
| 2.2.1 | Amplification of the genes under study from the genome of <i>Gloeobacter violaceus</i> PCC 7421 | 2-6 |
| 2.2.2 | Agarose gel electrophoresis | 2-8 |
| 2.2.3 | PCR amplicon purification by Gel Elution Protocol (using gel extraction kit Qiagen) | 2-9 |
| 2.2.4 | Restriction of purified PCR product and vector | 2-10 |
| 2.2.5 | Ligation of PCR product into a TA vector | 2-12 |
| 2.2.6 | Cloning into expression vector | 2-14 |
| 2.2.7 | Alkaline lysis method of isolation of plasmid | 2-15 |
| 2.2.8 | Transformation of plasmid/ligation product into host cells | 2-18 |
| | 2.2.8.1 Preparation of <i>E. coli</i> DH5 α and <i>E. coli</i> BL21(DE3)pLysS competent cells | 2-19 |
| | 2.2.8.2 Transformation of recombinant DNA into <i>E. coli</i> DH5 α competent cells | 2-20 |
| 2.2.9 | The assessment of expression of constructs generated | 2-22 |
| 2.2.10 | Solubility of the expressed proteins | 2-23 |
| 2.2.11 | Purification of the proteins by Immobilized Metal Affinity Chromatography (IMAC) | 2-25 |
| | 2.2.11.1 Protein purification by IMAC under native conditions | 2-26 |
| | 2.2.11.2 Desalting of the eluted fractions | 2-28 |
| | 2.2.11.3 Histidine tag cleavage | 2-29 |
| | 2.2.11.4 Protein purification under denaturing conditions | 2-30 |
| 2.2.12 | Mass Spectrometry analysis | 2-33 |
| 2.2.13 | Circular Dichroism (CD) | 2-34 |
| 2.2.14 | RuBP assay to study the enzyme kinetics of RuBisCO | 2-35 |
| | 2.2.14.1 Preparation of coupling enzyme mixture | 2-35 |
| | 2.2.14.2 Preparation of reaction components of assay | 2-36 |
| | 2.2.14.3 Estimation of Protein concentration | 2-36 |
| | 2.2.14.4 Activation of RuBisCO | 2-36 |
| | 2.2.14.5 Preparation of assay mixture | 2-36 |
| | 2.2.14.6 Determination of Michaelis Menten constant and turnover number of the enzyme | 2-37 |
| 2.2.15 | The effect of RuBisCO SSU and CcmM on activity of RuBisCO LSU | 2-38 |

Chapter 3 **Distribution and Diversity of Carbon Concentrating Mechanism Proteins of *β* Cyanobacteria**

| | | |
|-------|---|------|
| 3.1 | Introduction | 3-2 |
| 3.2 | Methodology | 3-3 |
| 3.2.1 | Identification of early diverging cyanobacteria | 3-3 |
| 3.2.2 | Identification of CCM homologs across all Eubacterial phyla | 3-4 |
| 3.2.3 | Phylogenetic analysis of CCM proteins | 3-5 |
| 3.3 | Results and Discussions | 3-7 |
| 3.3.1 | <i>Gloeobacter violaceus</i> PCC 7421 is an early diverging cyanobacterium | 3-7 |
| 3.3.2 | The complete set of β carboxysome proteins found in cyanobacteria only | 3-10 |
| 3.3.3 | The CCM operon was formed <i>in-situ</i> and passed on by vertical succession | 3-17 |
| 3.3.4 | Phylogenetic analysis of CCM proteins | 3-21 |
| 3.3.5 | Diversity of CCM proteins | 3-32 |
| 3.3.6 | <i>G. violaceus</i> PCC 7421- cyanobacterium with unique characteristics | 3-37 |
| 3.4 | Conclusions | 3-40 |

Chapter 4 **Cloning and Characterization of RuBisCO Encoding Genes of *Gloeobacter violaceus* PCC 7421**

| | | |
|---------|--|------|
| 4.1 | Introduction | 4-2 |
| 4.2 | Materials and Methods | 4-8 |
| 4.2.1 | <i>In-Silico</i> sequence analysis | 4-8 |
| 4.2.1.1 | RuBisCO diversity and co-evolution of constituting proteins | 4-8 |
| 4.2.1.2 | Analysis of locus of RuBisCO Subunit encoding genes | 4-8 |
| 4.2.1.3 | Scanning different Cyanobacterial species for presence of RbcX protein | 4-8 |
| 4.2.2 | Construction of expression plasmid constituting RuBisCO encoding genes of <i>G. violaceus</i> | 4-9 |
| 4.2.3 | Expression of recombinant plasmids in prokaryotic system | 4-9 |
| 4.2.4 | Test for the solubility of the protein expressed <i>in-vivo</i> in native and/or denaturing buffer | 4-10 |
| 4.2.5 | Purification of <i>G. violaceus</i> PCC 7421 RuBisCO subunits by IMAC | 4-10 |
| 4.2.6 | MS/MS and CD analysis of <i>G. violaceus</i> RuBisCO proteins purified by IMAC | 4-11 |
| 4.2.7 | RuBisCO spectrophotometric assay | 4-11 |
| 4.2.8 | <i>In-silico</i> homology modeling | 4-12 |

| | | |
|--------|--|------|
| 4.3 | Results and Discussions | 4-12 |
| 4.3.1 | The protein RbcL is most conserved while RbcS and RbcX more diverse | 4-12 |
| 4.3.2 | The RuBisCO operon seems to have co-evolved in most β cyanobacteria, but not in α cyanobacteria | 4-13 |
| 4.3.3 | Positional analysis of RbcX with respect to other RuBisCO subunit encoding genes | 4-17 |
| 4.3.4 | The expression of <i>G. violaceus</i> RuBisCO as soluble or insoluble form and purification of <i>G. violaceus</i> RuBisCO | 4-23 |
| 4.3.5 | Validation of protein identity and conformation (MS/MS and CD analysis) | 4-31 |
| 4.3.6 | Catalytic Activity of <i>G. violaceus</i> RuBisCO <i>in-vitro</i> | 4-34 |
| 4.3.7 | <i>In-silico</i> generated tertiary model of RuBisCO LSU and SSU of <i>G. violaceus</i> | 4-36 |
| 4.3.8 | Role of chaperones in RuBisCO assembly | 4-39 |
| 4.3.9 | Cyanobacterial RuBisCO Vs <i>G. violaceus</i> RuBisCO | 4-39 |
| 4.3.10 | Congruency in growth rate and RuBisCO activity | 4-40 |
| 4.3.11 | RuBisCO activity not dependent on the active site alone | 4-40 |
| 4.4 | Conclusions | 4-41 |

Chapter 5 Cloning and Characterization of Carbon Concentrating Mechanism (CCM) Proteins of *Gloeobacter violaceus* PCC 7421

| | | |
|-------|---|------|
| 5.1 | Introduction | 5-2 |
| 5.2 | Methodology | 5-5 |
| 5.3 | Results and Discussions | 5-5 |
| 5.3.1 | CcmK | 5-5 |
| 5.3.2 | CcmL | 5-10 |
| 5.3.3 | CcmM | 5-13 |
| | 5.3.3.1 Prediction of IRES | 5-13 |
| | 5.3.3.2 Analysis of the conserved residues for CA domain in CcmM | 5-15 |
| | 5.3.3.3 Tapping for evolutionary link between CcmM and RuBisCO | 5-17 |
| | 5.3.3.4 The <i>in-silico</i> generated protein model for CcmM IRES1 | 5-21 |
| | 5.3.3.5 Cloning and <i>in-vivo</i> expression of CcmM | 5-22 |
| 5.3.4 | CcmN | 5-25 |
| 5.3.5 | CcmO | 5-28 |
| 5.3.6 | Amplification and cloning of the <i>ccm</i> operon of <i>G. violaceus</i> | 5-32 |
| 5.4 | Conclusions | 5-33 |

Chapter 6 The Effect of CcmM and RbcS of *Gloeobacter violaceus* on Catalytic Activity of a Form I (*Gloeobacter violaceus*) and Form II (*Rhodospirillum rubrum*)

| | | |
|------------------------------|---|------|
| RuBisCO Large Subunit | | |
| 6.1 | Introduction | 6-2 |
| 6.2 | Methodology | 6-2 |
| 6.3 | Results and Discussions | 6-3 |
| 6.3.1 | <i>Rhodospirillum rubrum</i> RuBisCO | 6-3 |
| 6.3.2 | The Effect of <i>G. violaceus</i> RuBisCO SSU on <i>R. rubrum</i> RuBisCO | 6-7 |
| 6.3.3 | The effect of <i>G. violaceus</i> RuBisCO SSU on activity of LSU | 6-15 |
| 6.3.4 | Effect of <i>G. violaceus</i> CcmM on activity of <i>R. rubrum</i> RuBisCO | 6-20 |
| 6.3.5 | Effect of <i>G. violaceus</i> CcmM on the activity of <i>G. violaceus</i> RuBisCO LSU | 6-27 |
| 6.4 | Conclusions | 6-32 |

Chapter 7 Conclusions and Future Scope

| | | |
|-----|--------------|-----|
| 7.1 | Conclusions | 7-2 |
| 7.2 | Future Scope | 7-6 |

References

Appendices

| | |
|----|------------------------------------|
| i | List of Publications |
| ii | Conferences and Workshops attended |

Biography of Guide

Biography of Self

Reprint of Publications

Preface

Photosynthesis is among the most important biochemical pathways, being responsible for metabolizing the unusable inorganic forms of carbon to readily usable source of energy and building blocks of life for all living forms. The efficiency of photosynthesis directly effects the crop yield, which in turn depends upon the rate limiting step in the Calvin cycle catalyzed by the enzyme RuBisCO (Ribulose-1,5-Bisphosphate Carboxylase Oxygenase). Hence, the efficiency of the protein RuBisCO, evaluated by its catalytic rate and specificity, directly affects the rate of photosynthesis and ultimately the yield.

RuBisCO, despite its origin dating back to millions of years from now and its importance in nature suffers from low catalytic rate and fickle specificity for substrate. The catalytic site being conserved in all the autotrophic organisms ranging from cyanobacteria to higher plants shows the significance of the arrangement of the residues in the active site to be the most apt for the carboxylation of RuBP. Evolution works on the principle of random mutations in the genetic code and nature works on the phenomenon of survival of the fittest. Hence, evolution being inevitable did apply to RuBisCO as well, but only the protein having the active site residues conserved were able to survive or in this case were capable of performing their function.

The enzyme has been of interest to the scientific world for more than 70 years now, for basically two main reasons. Firstly, despite the investment of ample amount of time and resources in all these years, a number of aspects to the enzyme still remain elusive viz. the molecular basis of cause of variation in characteristics of the protein, the factors involved in the assembly of the protein, etc. Secondly, being the rate limiting factor in photosynthesis and having some major drawbacks, it is a subject with lots of scope of improvement in crop yield, keeping in mind the impending danger of food scarcity. Characterization of RuBisCO and the factors playing role in its assembly and catalysis will impart some knowledge to further understand this protein. Moreover, since every RuBisCO has its own set of characteristics by virtue of its complexity, and only about 100 RuBisCO proteins have been characterized till date, the study of catalysis of this protein from different sources is an open avenue to widening the horizons of knowledge about this protein.

The present study involves the characterization of RuBisCO from an early diverging cyanobacteria *Gloeobacter violaceus* PCC 7421. *G. violaceus* is in itself an interesting organism owing to its unique set of characteristics. Further, considered to be an organism of primitive

origin, it could also be among the initial set of organism to have acquired oxygenic photosynthesis and hence RuBisCO. Attempts have been made in this study to express and characterize the protein in the presence and absence of the chaperone and the small subunit so as to obtain a soluble form. In an attempt to identify the role of the small subunit of RuBisCO, the activity of form I and form II RuBisCO were also evaluated in the presence of the small subunit RuBisCO, *in-vitro*. Carbon Concentrating Mechanism (CCM) is a mechanism which came into existence as a consequence of loose specificity of the RuBisCO for carbon dioxide. This mechanism constituting carboxysomes in cyanobacteria is responsible for concentrating carbon dioxide around RuBisCO-encapsulated in protein microcompartments. As an objective of my research, the CCM proteins of *G. violaceus* were also cloned and characterized. The ultimate aim of the study of CCM proteins is successful transformation of CCM into C₃ plants, which are devoid of any such mechanism.

The work is presented in form of seven chapters:

1. Introduction and literature review: This Section summarizes information on RuBisCO and Carbon Concentrating Mechanisms and their importance.
2. Material and Methods: This Section describes details of *in-silico* and various experimental procedures and techniques, which were employed in the present study.
3. The chapter 3 is a descriptive analysis of the distribution of the CCM proteins across the Eubacterial and Archaeobacteria phyla, in an attempt to find ancestral proteins for the same.
4. The chapter 4 presents the amplification, expression, purification and characterization RuBisCO encoding genes of *G. violaceus* in *E. coli* system (*in-vitro* and *in-silico*).
5. The chapter 5 involves the cloning and characterization of the CCM proteins of *G. violaceus* by expression in *E. coli* and protein modeling for CCM proteins.
6. The chapter 6 is about the protein-protein interactions between RuBisCO subunits and with CcmM and their functional implications.
7. The chapter 7 provides the conclusive note on the findings from the present study and the future scope.

Acknowledgements

ਨਾਨਕ ਨਾਮ ਜਹਾਜ ਹੈ, ਚੜ੍ਹੇ ਸੇ ਉਤਰੇ ਪਾਰ,
Nanak naam jahaj hai, charhe jo utre paar

ਜੇ ਸ਼ਰਧਾ ਕਰ ਸੇਵਦੇ, ਗੁਰੂ ਪਾਰ ਉਤਾਰਨ ਹਾਰ।
Jo shradha kar sevde, guru paar utaran haar

First and foremost I extend my deepest thanks to the almighty for making me fortunate to enjoy endless fruits of humanity and become capable of having the intuitiveness to expand my limits for oneself and anyone destined to be succoured through me.

Success is the outcome of inquisitive mind, directed by adept guide, reinforced by perseverant efforts and the guts to treat defeat with a positive outlook. I am thankful to almighty for giving me the opportunity to learn in supervision of my guide, Dr. Sandhya Mehrotra, a perfect blend of an educator and benevolence. I would like to thank her for all her timely encouragements and for directing me onto the right path when I got distracted, which I believe was quiet often. Under her tutelage, I had cherished enormous freedom to pursue new ideas and perceptions. I also want to use this medium, to convey my utmost regard for her, for giving me the opportunity to work under her able guidance. Thank you again for the splendid voyage and I eagerly look forward to future expeditions down the lane.

Apart from my advisor, I would like to extend by sincere thanks to members of my Doctoral Advisory Committee members, Prof. Rajesh Mehrotra and Prof. Shibasish Chowdhury, for their valuable guidance and contribution to the furtherance of this research, insightful discussion, scientific advice, knowledge and also for their time and effort in checking the manuscript in short time period. I offer my gratitude, to Director Prof. A.K. Sarkar, Dean, Academic & Research Division Prof. S.K. Verma BITS-Pilani, Pilani Campus, for allowing me to use institute's infrastructure. I also thank my alma mater, BITS Pilani for providing me with this and various other opportune possibilities which resulted in the culmination of this research.

Special thanks to the present and former heads of the Department of Biological Sciences, Prof. Prabhat Nath Jha, Prof. Rajesh Mehrotra, Prof. Jitendra Panwar and Prof. Shibasish Chowdhury for their timely support and efforts in making the department resourceful and congenial to work. I would further like to be thankful to DRC members Prof. Jitendra Panwar, Prof. Ashis Kumar Das,

Prof. Shibashis Chowdhary, Prof. Lalita Gupta, Dr. Sandhya Mehrotra and Prof. Prabhat Nath Jha for their time to time co-operation and words of encouragement and also for charting policies in support of students where and when necessary. The experience of coming in contact with the faculty members of Department of Biological Sciences, in the duration of my degree, was thoroughly productive and I do hereby acknowledge all of them. The involvement in teaching practice under the expert guidance of teachers of high caliber and credentials in the world class setup of BITS Pilani was a knowledgeable experience.

A friend is someone who knows the song in your heart, and can sing it back to you when you have forgotten the words. I am blessed to have wonderful friends Panchsheela, Naveen and Dr. Purva, who were buddies in times of fun and pillars of strength during stressful situations. I owe them big thanks for being a moral support and enable me to see through difficult times. I rejoice their immaculate sense of humour and motivation that often rejuvenated me to come stronger and smarter after every failure. I am indebted to jovial company and homely atmosphere created by my friends Zarna, Isha didi, Shobha, Gagandeep sir, Dr. Kuldeep and Dr. Arpit, throughout the time I spent at BITS, Pilani.

Heartfelt thanks to my colleagues and labmates Vandana Tomar, Nisha Jangir, Asha Jhakar, Vidushi Asati, Chetna Sangwan and Zaiba Hasan Khan, thank you for offering me valued advice, untimely help and moral support whenever needed and also for maintaining an amiable environment in the lab. The short rejuvenating breaks, in company of my labmates breaking the monotony of busy days was always stimulating and joyous.

Learning is a never ending process and all you know is put to challenge when one has to extend it to someone else, in a manner that it can be executed. I thank Garima Singh and Rashmi for helping me perform experiments, your agog approach to grasp concepts was indeed empowering from new perspectives. I would further like to thank Jaspreet Singh, Ankit Bansal and Rohan for their expertise in running codes for some *in-silico* analysis. I am grateful for your time and effort that resulted into outputs that form an important part of this dissertation.

This acknowledgement would be incomplete without special mention of my seniors Dr. Deepak Pakalpati, Dr. Amit Subuhi, P.A. Boopathi sir, Gagandeep sir, Dr. Arpit and Dr. Purva who taught me lab culture and a sense of belongingness for lab. I as a neophyte looked upon to their footsteps and suggestions in performing every duty assigned or handling every minor/major situation in research. I am thankful to the staff of office of Department of Biological Sciences for


their help in every official matter. A special word of gratefulness for Mr. Naresh Kumar Saini, Mr. Raghuveer, Mr. Mahipal and Mr. Manoj Kumar who have helped me with several official works during my stay in BITS Pilani.

The very existence of 'I', its intellect and characteristics were bestowed upon me by my beloved parents. I bow to my adorable parents, Mohinderjit Singh and Davinderjit Kaur for their encouragement, support, sacrifices and unbound love showered upon me throughout my life. They have always kept me upbeat to meet new challenges and succeed in every walk of life. Emotions of my heart should find new boundaries to contain my love for my brother Gurkirat Singh, I extend my earnest gratitude to him for his care and affection. I am deeply obliged to my grandfather- Gurcharan Singh for his love and tireless support since the day I stepped outside my home with the urge to learn and become capable of extending and applying what I learn. I don't imagine a life without their love, concern and blessings. Your prayers and strong belief in me had helped me go on and sustain thus far. I dedicate this thesis to my beloved parents, whose role in my life is and will always remain immense. For, it was because of their faith and confidence in my abilities, that I kept working through the hurdles and failures that I had faced during this entire time.

The financial assistance provided by DST-fasttrack project 2011-2012 and UGC-BSR 2012-2016, has enabled me to pursue my research with ease, which is gratefully acknowledged. Last but not the least, I feel privileged for my stay at BITS, Pilani which is home to many knowledgeable people with great work culture. Until now I heard people say, but now I realize that there is some formless, invisible energy transcended into you being a BITSian that forms an integral part of your personality and acts as an unconscious guiding force in your life ahead. I feel empowered being a part of this esteemed institute. I take an oath to serve for the mankind with whatever knowledge and wisdom I possess and will acquire in future by virtue of furtherance of my research work and results thereof.

Date: 27-12-2016

Gurpreet Kaur Sidhu



LIST OF TABLES

| No. | Table Title | Page No. |
|------|--|----------|
| 2A | The Accession numbers and the conserved domain profile of the proteins of CCM from <i>G. violaceus</i> | 2-2 |
| 2B | List of organisms used for study of evolutionary link between CcmM and RuBisCO proteins | 2-3 |
| 2C | List of primers used in the amplification of the genes under study | 2-5 |
| 2D | The PCR conditions standardized for various genes under study | 2-7 |
| 2E | PCR conditions standardized for <i>ccmM</i> and <i>ccm</i> operon sequences | 2-8 |
| 2F | Characteristics of Agarose Low Electroendosmosis | 2-9 |
| 2G | The concentration of agarose gel suitable for different DNA sized fragments | 2-9 |
| 2H | The detail of the restriction enzyme used for cloning procedures | 2-11 |
| 2I | The composition of reaction mixture used for a standard restriction reaction | 2-12 |
| 2J | The composition of ligation reaction along with recommended controls for TA cloning | 2-13 |
| 2K | The important characteristics of the vectors used, in context to the expression and purification studies | 2-15 |
| 2L | The important characteristics of the BD Talon Co ²⁺ resin | 2-26 |
| 2M | Composition of the buffers used protein purification by IMAC under native conditions | 2-27 |
| 2N | Column characteristics of DS03 | 2-29 |
| 2O | Buffers used in protein purification under denaturing conditions | 2-31 |
| 2W | Buffers used for renaturing of the purified proteins | 2-32 |
| 3A | The list of phyla used for the analysis of distribution of CCM proteins | 3-6 |
| 3B | The distribution of the β carboxysome proteins across phyla | 3-14 |
| 3C | The Correlation between the phylogeny of β -cyanobacteria on the basis of CcmM protein C terminal domain (CA domain) and N terminal domain (SSU repeat domain) generated by MirrorTree | 3-25 |
| 3D | The degree of similarity of the SSU repeats present in CcmM proteins of β -cyanobacteria to the bonafide RuBisCO SSU domain | 3-25 |
| 3A-S | BLASTP hits for amino acid sequence of CcmK from <i>G. violaceus</i> PCC 7421 from various phyla of Eubacteria and Archaea | 3-42 |

| | | |
|-------------|---|------|
| 3B-S | BLASTP hits for amino acid sequence of CcmL from <i>G. violaceus</i> PCC 7421 from various phyla of Eubacteria and Archaea | 3-42 |
| 3C-S | BLASTP hits for amino acid sequence of CcmM from <i>G. violaceus</i> PCC 7421 from various phyla of Eubacteria and Archaea | 3-43 |
| 3D-S | BLASTP hits for amino acid sequence of CcmN from <i>G. violaceus</i> PCC 7421 from various phyla of Eubacteria and Archaea | 3-44 |
| 3E-S | BLASTP hits for amino acid sequence of CcmO from <i>G. violaceus</i> PCC 7421 from various phyla of Eubacteria and Archaea | 3-45 |
| 3F-S | List of <i>ccmK</i> gene complements in β -cyanobacteria arranged according to the phylogenetic topology. | 3-46 |
| 4A | Locus of L and S subunit of RuBisCO in the genome of the organisms with RbcX | 4-19 |
| 4B | Locus of L and S subunit of RuBisCO in the genome of the organisms lacking RbcX | 4-20 |
| 4C | Diverse chaperones present in Cyanobacteria without RuBisCO chaperone | 4-21 |
| 4D | The CD spectra data of IMAC purified and renatured <i>G. violaceus</i> LSU | 4-34 |
| 5A | The most conserved regions of the CcmL protein | 5-11 |
| 5B | Results of SD sequence, followed by start codon, search in the <i>ccmM</i> gene sequence | 5-14 |
| 5C | The number of SSU repeats in the CcmM from different sources and the degree of identity with the best found hits of RbcS in pBLAST search | 5-19 |
| 6A | The parameters of <i>R. rubrum</i> RuBisCO reported by different research groups | 6-7 |
| 6B | The ratios and corresponding concentrations of the <i>R. rubrum</i> RuBisCO and <i>G. violaceus</i> RuBisCO SSU used for complementation studies | 6-7 |
| 6C | The residues involved in interaction between LSU (A and I- Row 1 and 2) and SSU (B/H/J/P – rows 3-6) in <i>T. elongatus</i> RuBisCO (2YBV) retrieved from PDB | 6-10 |
| 6D | The attributes for the most possible interactions between <i>R. rubrum</i> RuBisCO and <i>G. violaceus</i> RbcS (SSU) | 6-14 |
| 6E | The ratios and corresponding concentrations of the <i>G. violaceus</i> RbcL (LSU) and <i>G. violaceus</i> RbcS (SSU) used for complementation studies. | 6-15 |
| 6F | The attributes for the most possible interactions between <i>G. violaceus</i> RbcL (LSU) dimer and <i>G. violaceus</i> RbcS (SSU) | 6-19 |
| 6G | The ratios and corresponding concentrations of the <i>R. rubrum</i> RbcL (LSU) and <i>G. violaceus</i> CcmM used for complementation studies. | 6-21 |

| | | |
|-----------|---|------|
| 6H | The RuBisCO SSU domains of <i>G. violaceus</i> CcmM as estimated by CDD search and the conserved residues in each of the domains with respect to the <i>G. violaceus</i> RuBisCO small subunit identified by sequence alignment | 6-22 |
| 6I | The domains of <i>G. violaceus</i> CcmM IRES1 identified on the basis of the tertiary structure of the protein estimated <i>in-silico</i> | 6-23 |
| 6J | The attributes for the most possible interactions between <i>R. rubrum</i> Rbel. (LSU) dimer and <i>G. violaceus</i> CcmM IRES1 | 6-25 |
| 6K | The ratios and corresponding concentrations of the <i>G. violaceus</i> Rbel. (LSU) and <i>G. violaceus</i> CcmM used for complementation studies. | 6-27 |
| 6L | The attributes for the most possible interactions between <i>G. violaceus</i> Rbel. (LSU) dimer and <i>G. violaceus</i> CcmM IRES1 | 6-29 |

LIST OF FIGURES

| No. | Title of Figure | Page No. |
|--------|--|----------|
| 1A | Active site residues of <i>Spinacia oleracea</i> RuBisCO | 1-7 |
| 1B | The comparison of K_{cat} (catalytic turnover) and S (selectivity) of enzyme RuBisCO from various organisms | 1-12 |
| 1C | The comparison of K_MCO_2 and K_MO_2 of enzyme RuBisCO from various organisms | 1-13 |
| 2A | Vector systems used for cloning and expression of proteins under study | 2-17 |
| 3A(i) | Maximum Likelihood phylogenetic representation of cyanobacteria based on 16S rRNA sequence | 3-8 |
| 3A(ii) | Maximum Likelihood phylogenetic representation of cyanobacteria based on 16S rRNA- <i>pyrH-rbcL</i> sequence | 3-9 |
| 3B | The distribution of β carboxysome proteins across various phyla of Eubacteria | 3-16 |
| 3C | Maximum Likelihood phylogenetic representation of cyanobacteria based on carboxysome operon sequence | 3-19 |
| 3D | Phylogenetic analysis of α and β -cyanobacteria separately on the basis of <i>ccm</i> operon | 3-20 |
| 3E(i) | The duplication events of CcmK in the course of evolution of β -cyanobacteria | 3-23 |
| 3E(ii) | The evolutionary history of CcmK amino acid sequences inferred by using the Maximum Likelihood method | 3-26 |
| 3F | The evolutionary history of CcmL was inferred by using the Maximum Likelihood method based on the Le_Gascuel_2008 model | 3-27 |
| 3G | The evolutionary history of CcmM was inferred by using the Maximum Likelihood method based on the Whelan And Goldman + Freq. model | 3-29 |
| 3H | The evolutionary history of CcmN was inferred by using the Maximum Likelihood method based on the Whelan And Goldman + Freq. model | 3-30 |
| 3I | The evolutionary history of CcmO was inferred by using the Maximum Likelihood method based on the Le_Gascuel_2008 model | 3-31 |
| 3J | The distance of the cyanobacteria from <i>G. violaceus</i> PCC 7421 on the basis of 16S rRNA- <i>pyrH-rbcL</i> sequences and carboxysome encoding operon derived by Jukes-Cantor model in MEGA 6.0 | 3-34 |
| 3K | The comparison of distance of various cyanobacteria from <i>G. violaceus</i> PCC 7421 derived by phylogenetic analysis of the 16S rRNA sequence, | 3-35 |

| | | |
|-----------|---|------|
| | <i>ccm</i> operon, CcmK, CcmL, CcmM, CcmN and CcmO sequences of β -cyanobacteria estimated in MEGA 6.0 | |
| 3L | Box and Whiskers plot for the distance of the 16S rRNA, <i>ccm</i> operon, CcmK, CcmL, CcmM, CcmN and CcmO sequences with respect to <i>G. violaceus</i> PCC 7421 sequences | 3-36 |
| 3M | The closest homologs of CCM proteins of <i>G. violaceus</i> PCC 7421, among the β -cyanobacteria available at Cyanobase found by phylogenetic analysis | 3-39 |
| 3N | The speculated time frame of events from origin of cyanobacteria to evolution of CCM in cyanobacteria and higher forms | 3-40 |
| 4A | Timeline of the important discoveries and observations about the protein RuBisCO from 1947 to 2016 | 4-4 |
| 4B(i-iii) | The sequence logo for the RbcL, RbcS and RbcX sequence from 37 cyanobacterial species available at Kazusa Genome Resource, respectively | 4-14 |
| 4C(i-ii) | Evolution pattern of RbcL, RbcX and RbcS in β -cyanobacteria and α -cyanobacteria, respectively | 4-16 |
| 4D | The alignment of RbcL amino acid sequence of five organisms lacking RbcX and five having RbcX, analysed for the presence or the absence of the RbcX interacting Domain | 4-23 |
| 4E | The agarose (1%) gel image of the PCR products of the genes under study (<i>rbcl</i> , <i>rbcX</i> , <i>rbcS</i> and <i>rubisco</i> operon) | 4-23 |
| 4F | The SDS-PAGE (12%) image of IPTG induced cultures of <i>pCOLDII-Gv-rbcl</i> , <i>pCOLDII-Gv-rbcS</i> , <i>pCOLDII-Gv-rbcX</i> and <i>pCOLDII-Gv-rbcLSX</i> in <i>E. coli</i> BL21(DE3)pLysS | 4-26 |
| 4G | SDS-PAGE (12%) showing solubility of the overexpressed protein in presence/absence of denaturants | 4-28 |
| 4H | The SDS-PAGE (12%) for fractions collected during metal affinity chromatography from IPTG induced cultures of <i>pCOLD-Gv-rbcl</i> , <i>pCOLD-Gv-rbcS</i> and <i>pCOLD-Gv-rbcX</i> in <i>E. coli</i> BL21(DE3)pLysS | 4-30 |
| 4I | The image of 6% native PAGE for <i>G. violaceus</i> RbcL, after purification in denaturing condition followed by subsequent renaturing by steady removal of denaturing agents | 4-30 |
| 4J(i-ii) | The m/z vs intensity plot of the <i>G. violaceus</i> RuBisCO LSU and SSU protein analyzed by MALDI-TOF after in-gel tryptic digestion, respectively | 4-32 |
| 4K(i-ii) | The m/z vs intensity plot of the <i>G. violaceus</i> RuBisCO LSU and SSU obtained by <i>in-silico</i> trypsin digested, respectively | 4-33 |

| | | |
|-----------------|--|------|
| 4L | The CD spectra of <i>G. violaceus</i> RbcL purified by IMAC | 4-34 |
| 4M | The activity (change in absorbance per minute) of <i>G. violaceus</i> RuBisCO RbcL and Holoenzyme at different RuBP concentrations | 4-35 |
| 4N | The effect of RuBP concentration on the specific activity of <i>G. violaceus</i> RuBisCO large subunit and RuBisCO holoenzyme | 4-36 |
| 4O | The protein model for <i>G. violaceus</i> RbcL generated by homology modeling | 4-37 |
| 4P | The protein model for <i>G. violaceus</i> RbcS generated by homology modeling | 4-38 |
| 5A | Timeline of major discoveries and observations in carboxysomes from 1961-2016 | 5-3 |
| 5B | The sequence logo for CcmK sequences from 20 β -cyanobacteria, aligned by T Coffee program. | 5-6 |
| 5C | The protein model for CcmK generated by homology modeling | 5-8 |
| 5D(i-ii) | The Agarose gel (1%) image for the polymerase chain reaction product for <i>ccm</i> genes and restriction analysis of <i>pET15b-ccmK</i> clone | 5-9 |
| 5E | SDS-PAGE (12%) to check the expression of 12.5kDa <i>G. violaceus</i> PCC 7421 <i>ccmK</i> (approx 300bp gene cloned in <i>pET15b</i>) in <i>E. coli</i> BL21(DE3)pLysS by 1mM IPTG induction. | 5-9 |
| 5F | The sequence logo for CcmL sequences from 20 β -cyanobacteria, aligned by T Coffee program | 5-11 |
| 5G | The CcmL protein showing the residues involved in interactions between the monomers to form the hexamer | 5-12 |
| 5H | Confirmation of <i>pCOLDII-ccmL</i> construct by restriction digestion with enzymes <i>NdeI</i> and <i>XhoI</i> | 5-12 |
| 5I | The SDS-PAGE (12%) image for the expression profile of ~12kDa protein of <i>G. violaceus</i> PCC 7421 <i>ccmL</i> (approx 350bp gene cloned in <i>pCOLDII</i>) in <i>E. coli</i> BL21(DE3)pLysS by 1mM IPTG induction | 5-12 |
| 5J | The sequence logo for the CcmM N-terminal domain of the CcmM protein from 20 β -cyanobacteria | 5-15 |
| 5K | The sequence logo for the CcmM N-terminal domain of the CcmM protein from <i>G. violaceus</i> and γ CA of <i>Methanosarcina thermophila</i> | 5-16 |
| 5L | The sequence logo for the CcmM C-terminal domain of the CcmM protein from 20 β -cyanobacteria | 5-17 |
| 5M | Sequence logo for the multiple sequence alignment of the RuBisCO small subunit from β -cyanobacteria | 5-18 |

| | | |
|-----------------|--|------|
| 5N | The tertiary structure of IRES1 CcmM by homology modeling | 5-21 |
| 5O | The Agarose gel (0.8%) image and PCR conditions for <i>ccmM</i> of <i>G. violaceus</i> PCC 7421 using Phusion polymerase | 5-23 |
| 5P | Agarose gel (0.8%) image for confirmation of <i>pET15b-ccmM</i> construct by restriction digestion with enzymes <i>NdeI</i> and <i>XhoI</i> | 5-24 |
| 5Q | SDS-PAGE (12%) image to check the expression of 71KDa <i>G. violaceus</i> PCC 7421 <i>ccmM</i> (approx 2000bp gene cloned in <i>pET15b</i>) in BL21(DE3)pLysS by 1mM IPTG induction | 5-24 |
| 5R | SDS-PAGE (12%) image for purification of CcmM of <i>G. violaceus</i> PCC 7421 by affinity chromatography | 5-25 |
| 5S | Sequence logo for the multiple sequence alignment of the CcmN from β -cyanobacteria, obtained by alignment in MEGA6.0 | 5-26 |
| 5T | The tertiary structure of CcmN by homology modeling | 5-27 |
| 5U | Agarose gel (0.8%) for the confirmation of <i>pET15b-ccmN</i> construct by restriction with enzymes <i>NdeI</i> and <i>BamHI</i> and SDS-PAGE for expression analysis from the construct | 5-28 |
| 5V | Sequence logo for the multiple sequence alignment of the CcmO from β -cyanobacteria | 5-29 |
| 5W | Agarose gel (0.8%) for the confirmation of <i>pET15b-ccmO</i> construct by restriction with enzymes <i>NdeI</i> and <i>BamHI</i> and SDS-PAGE for expression analysis from the construct | 5-29 |
| 5X | The tertiary structure of CcmO by homology modeling | 5-30 |
| 5Y | Agarose gel (0.8%) for the amplification of the carboxysome operon of <i>G. violaceus</i> PCC 7421 (~4.5Kbp) along with the conditions used for PCR by Phusion Polymerase and confirmation of <i>pET15b-ccm</i> operon clone by restriction analysis | 5-32 |
| 6A | Schematic of the objective of the study presented in chapter 6 | 6-2 |
| 6B | The image of SDS-PAGE (12%) for analyzing the expression profile of <i>pET15b-R rubrum</i> RuBisCO by IPTG induction | 6-4 |
| 6C | The SDS-PAGE (12%) for the IMAC purification of <i>R. rubrum</i> RuBisCO | 6-4 |
| 6D | The mass-spectrometry results for <i>R. rubrum</i> RuBisCO analyzed by MASCOT | 6-5 |
| 6E(i-ii) | The trend line for decrease in optical density at 340nm of the reaction mixtures containing different concentrations of the <i>R. rubrum</i> RuBisCO and RuBP, respectively | 6-5 |
| 6F(i-ii) | The graphical representation of the effect of <i>R. rubrum</i> RuBisCO concentration and RuBP (substrate) concentration on the activity of <i>R.</i> | 6-6 |

rubrum RuBisCO, respectively

| | | |
|----------|--|------|
| 6G | The Lineweaver-Burk plot for the RuBP dependent assay for <i>R. rubrum</i> RuBisCO | 6-6 |
| 6H | The graphical representation of the effect of substrate concentration on the specific activity of the <i>R. rubrum</i> RuBisCO (40µg) | 6-6 |
| 6I | The trend line for decrease in optical density at 340nm of the reaction mixtures containing <i>R. rubrum</i> RuBisCO and different concentrations of <i>G. violaceus</i> RuBisCO SSU after addition of the substrate RuBP | 6-8 |
| 6J | The effect of different concentrations of <i>G. violaceus</i> RuBisCO SSU on specific activity of <i>R. rubrum</i> RuBisCO | 6-9 |
| 6K | The arrangement of the RuBisCO subunits with respect to each other in <i>T. elongatus</i> RuBisCO | 6-9 |
| 6L | The arrangement of <i>G. violaceus</i> RuBisCO small subunit with respect to <i>R. rubrum</i> RuBisCO large subunit dimer, according to the 10 output files | 6-13 |
| 6M | The interacting residues (as generated by GRAMMX output files 1 and 3) between the <i>R. rubrum</i> RuBisCO LSU dimer and the <i>G. violaceus</i> RuBisCO SSU | 6-14 |
| 6N | The effect of different concentrations of <i>G. violaceus</i> RuBisCO SSU on specific activity of <i>G. violaceus</i> RbcL (LSU) | 6-16 |
| 6O | The substrate (RuBP) dependent assay of the <i>G. violaceus</i> RbcL and <i>G. violaceus</i> RbcL in presence of <i>G. violaceus</i> RbcS in ratio of 1:1 | 6-16 |
| 6P | The arrangement of <i>G. violaceus</i> RuBisCO small subunit with respect to <i>G. violaceus</i> RuBisCO large subunit dimer, according to the 10 output files | 6-18 |
| 6Q | The interacting residues (as generated by GRAMMX output file 8) between the <i>G. violaceus</i> RuBisCO LSU dimer and the <i>G. violaceus</i> RuBisCO SSU | 6-19 |
| 6R | The effect of different concentrations of <i>G. violaceus</i> CcmM on specific activity of <i>R. rubrum</i> RbcL (LSU) | 6-21 |
| 6S(i-ii) | The substrate (RuBP) dependent assay of the <i>R. rubrum</i> RbcL in presence of <i>G. violaceus</i> CcmM in ratio of 1:3, respectively; The Lineweaver-Burk plot for the RuBP dependent assay for <i>R. rubrum</i> RbcL + <i>G. violaceus</i> CcmM complex, <i>in-vitro</i> in ratio of 1:3, respectively | 6-21 |
| 6T | CPK model of CcmM IRES1 coloured according to SSU domains as mentioned in table 6I; The arrangement of <i>G. violaceus</i> CcmM with respect to <i>R. rubrum</i> RuBisCO large subunit dimer, according to the 10 output files | 6-24 |

| | | |
|-----------------|--|------|
| 6U | The interacting residues (as generated by GRAMMX output files 1, 3 and 6, respectively) between the <i>R. rubrum</i> RuBisCO LSU dimer and the <i>G. violaceus</i> CcmM IRES1 | 6-26 |
| 6V | The effect of different concentrations of <i>G. violaceus</i> CcmM on specific activity of <i>G. violaceus</i> RbcL (LSU) | 6-27 |
| 6W(i-ii) | The substrate (RuBP) dependent assay of the <i>G. violaceus</i> RbcL in presence of <i>G. violaceus</i> CcmM in ratio of 1:3, respectively; The Lineweaver-Burk plot for the RuBP dependent assay for <i>G. violaceus</i> RbcL · <i>G. violaceus</i> CcmM complex, <i>in-vitro</i> in ratio of 1:3, respectively | 6-28 |
| 6X | The arrangement of <i>G. violaceus</i> CcmM with respect to <i>G. violaceus</i> RuBisCO large subunit dimer, according to the 10 output files | 6-30 |
| 6Y | The interacting residues (as generated by GRAMMX output files 3 (3.1 and 3.2), 5 (5.5 and 5.2) and 9 (9.1 and 9.2) between the <i>G. violaceus</i> RuBisCO LSU dimer and the <i>G. violaceus</i> CcmM IRES1 | 6-31 |

Abbreviations

| | |
|---------------------|--|
| μ | Micro |
| 3-PGA | 3-Phosphoglycerate |
| ATP | Adenosine Triphosphate |
| BMC | Bacterial Microcompartment |
| CaCl_2 | Calcium chloride |
| CAM | Crassulacean Acid Metabolism |
| CCM | Carbon Concentrating Mechanism |
| Ci | Inorganic Carbon |
| CO_2 | Carbon Dioxide |
| DNA | Deoxyribonucleic Acid |
| DNase | Deoxyribonuclease |
| dNTPs | Deoxy nucleoside triphosphate |
| DTT | Dithiothreitol |
| <i>E. coli</i> | <i>Escherichia coli</i> |
| EDTA | Ethylene diamine tetra acetic acid |
| g | gram |
| <i>G. violaceus</i> | <i>Gloeobacter violaceus</i> PCC 7421 |
| Gya | Gillion years ago |
| h | Hour |
| IPTG | Isopropyl β -D thiogalactopyranoside |
| IRES | Internal Ribosomal Entry Site |
| K_{cat} | Catalytic turnover |
| kDa | kilo Dalton |
| K_M | Michaelis Menten constant |
| LB | Luria Bertani broth medium |
| LBA | Luria Bertani agar medium |
| LSU | Large subunit of RuBisCO |
| M | Molar |
| mg | Milligram |
| MgCl_2 | Magnesium chloride |
| min | Minute |
| mM | millimolar |
| NaCl | Sodium chloride |

| | |
|------------------|--|
| NADPH | Nicotinamide Adenine Dinucleotide Phosphate |
| NaOH | Sodium hydroxide |
| ng | nanogram |
| nm | nanometer |
| O.D. | Optical density |
| O ₂ | Oxygen |
| PAGE | Polyacrylamide Gel Electrophoresis |
| PCR | Polymerase chain reaction |
| <i>R. rubrum</i> | <i>Rhodospirillum rubrum</i> |
| RNA | Ribonucleic acid |
| RuBisCO | Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase |
| RuBP | Ribulose-1,5-bisphosphate |
| SDS | Sodium dodecyl sulphate |
| sec | Second |
| SSU | Small subunit of RuBisCO |
| SSU repeat | Small Subunit repeat |
| TAE | Tris- acetate- EDTA |
| TE | Tris-EDTA |
| TEMED | N, N, N', n' -Tetramethylethylenediamine |
| Tris | 2-Hydroxy methylamine |
| UV | Ultra-violet |
| X-gal | 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside |
| γ-CA | Gamma carbonic anhydrase |
| Ω | Specificity |

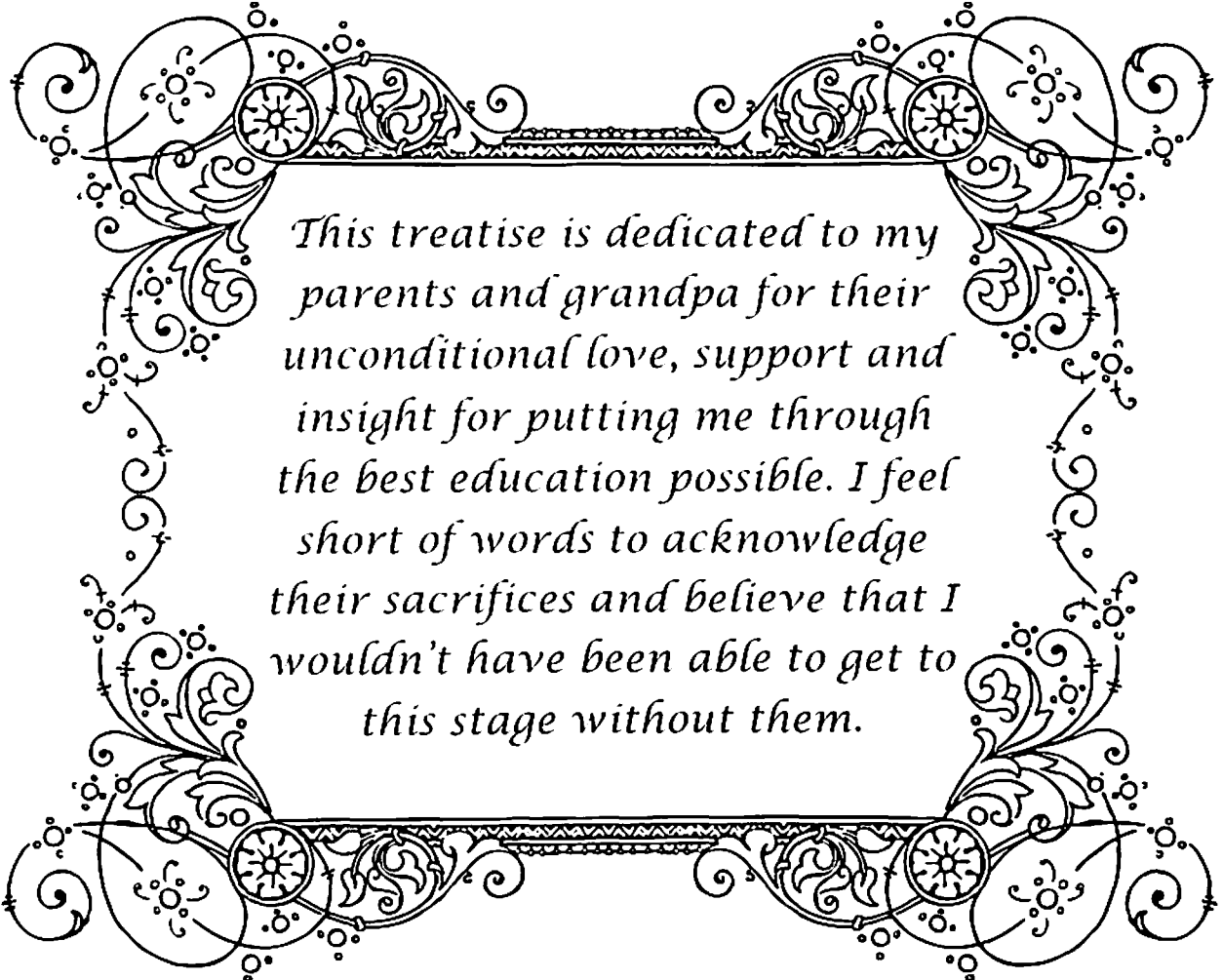
Amino Acids

| | | |
|---------------|-----|---|
| Alanine | ala | A |
| Arginine | arg | R |
| Asparagine | asn | N |
| Aspartic Acid | asp | D |
| Cysteine | cys | C |
| Glutamine | gln | Q |
| Glutamic Acid | glu | E |

| | | |
|---------------|-----|---|
| Glycine | gly | G |
| Histidine | his | H |
| Isoleucine | ile | I |
| Leucine | leu | L |
| Lysine | lys | K |
| Methionine | met | M |
| Phenylalanine | phe | F |
| Proline | pro | P |
| Serine | ser | S |
| Threonine | thr | T |
| Tryptophane | trp | W |
| Tyrosine | tyr | Y |
| Valine | val | V |

Nucleotide Bases

| | |
|----------|---|
| Adenine | A |
| Guanine | G |
| Cytosine | C |
| Thymine | T |
| Uracil | U |

A decorative border with intricate floral and scrollwork patterns, featuring circular motifs and swirling lines that frame the central text.

*This treatise is dedicated to my
parents and grandpa for their
unconditional love, support and
insight for putting me through
the best education possible. I feel
short of words to acknowledge
their sacrifices and believe that I
wouldn't have been able to get to
this stage without them.*

CHAPTER 1

Introduction and Review of Literature

1.1. Introduction

Ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCO, hereafter) is the primary enzyme involved in the process of photosynthesis. It catalyzes the fixation of atmospheric CO₂ (carbon dioxide) to Ribulose-1,5-bisphosphate (RuBP) to form two molecules of 3-phosphoglyceric acid (3-PGA) which is further cycled through subsequent reactions to form the basic organic molecules of life. But RuBisCO is not a perfect catalyst. It suffers from fickle specificity for CO₂ and opportunistic molecules like O₂. Oxygen directs RuBP to products a part of which eventually lead to energy inefficient processes and hence a setback for the process that directly or indirectly feeds all basic metabolisms of life (Gutteridge and Pierce 2006). Further, the catalytic turnover rate is painfully low, as it can fix only about three CO₂ molecules per second. RuBisCO catalyzes the most important step of photosynthesis and hence in order to compensate its inefficiency (Badger and Bek 2008), plants need to subject as much as 50% of their leaf nitrogen for the synthesis of RuBisCO, hence no wonder it turns out to be the most abundant protein on earth.

RuBisCO came into existence when the atmospheric composition was different from that of today. It has evolved to improve its performance under CO₂ limitation, O₂ inhibition, and scarcity of nitrogen. Nature, in order to enhance RuBisCO efficiency, has evoked changes that have led to RuBisCO with high specificity for CO₂, salvage pathway to recapture losses made because of accepting oxygen as a substrate, carbon concentrating mechanisms (CCM) and improved stomatal regulation. It could be attributed to these evolutionary changes that despite 500 fold higher O₂ levels as compared to CO₂ in the gaseous form, only 25% of the RuBisCO reactions are oxygenase. It is to be noted that the losses made at the hand of photorespiration are expected to rise in the coming future as an outcome of global warming which is predicted to favor oxygenation over carboxylation.

CCMs in their simplest form constitute a compartment for CO₂ accumulation, Ci (inorganic carbon) transporters and carbonic anhydrases that cause an increase in the CO₂/O₂ ratio at the site of RuBisCO activity and hence a decrease in oxygenase activity. Although, such a system adds to the resource expenditure of the organism for its synthesis, maintenance, and operation but the net outcome of the process is energy efficient (Sulstremeyer and Rinast 1996). The CCMs prevalent in various organisms include, carboxysomes - the most primitive known CCM in cyanobacteria and proteobacteria (Price et al., 2008), pyrenoids in algae (Giordano et al., 2005), C₄ and CAM metabolism based mechanisms in higher plants and certain algae (Edwards et al., 2004) and C₃-C₄ intermediates. The C₃ plants have not been found to have any CCM.

Cyanobacteria came into origin nearly 3500 million years ago (Badger et al., 2002), when the atmosphere had higher levels of CO₂ and low levels of O₂ favoring carboxylation activity of RuBisCO. About 400 million years ago, there was a large decline in CO₂ levels and an almost doubling of the oxygen concentration because of enhanced levels of oxygenic photosynthesis (Berner and Kothavala 2001). Raven (1997) states that this could have been the first major pressure faced by photosynthetic organisms to develop CCM. The increase in atmospheric oxygen occurred after the decrease in nickel content of the oceans, an outcome of the cooling mantle of the earth (Konhauser et al., 2009). Methanogens; one of the initial living forms and competitors of sulfate-reducing microorganisms for the resources need nickel as a cofactor for most of their biochemical processes and hence the scarcity of the same led the downfall of methanogens and eventually methane – a hurdle for an oxygenic environment.

The ultimate aim of the study of CCMs is to test the possibility of introducing CCM into C₃ plants. The approach of introducing a CCM into C₃ plants appears to be promising in terms of productivity but is hurdled by the fact that none of these mechanisms are completely deciphered till date and the extent of genetic engineering involved to establish such systems is also an issue. Among the various naturally occurring CCMs viz. carboxysomes, pyrenoids C₄ metabolism and CAM, the carboxysomes are comparatively best understood and introducing such a system in C₃ would involve limited genetic engineering. Further, the precise function of RuBisCO small subunit (RbcS) being elusive, the complete understanding of the enzyme RuBisCO is also lacking. The SSU repeats in protein CcmM are thought to be responsible for the initial assembly of the carboxysomes and such an arrangement of the proteins also brings forth the possibility of an evolutionary link between the two proteins. The function of another carboxysomal protein enclosed in the shell i.e. CcmN is also not clear. For establishing a cyanobacterial CCM in C₃ plants, knowing the functions and correlation between the proteins would not be sufficient as the regulation of synthesis of the various proteins under the effect of differential C_i conditions involved in the same is also important and still remains undeciphered. In an attempt to fulfill a part of the gaps in the existing knowledge of RuBisCO, carboxysome proteins, and their evolutionary links, the following objectives were designed.

1.2. *Gloeobacter violaceus* PCC 7421

Cyanobacteria having a global impact by virtue of their oxygenic photosynthetic on biogeochemical cycles since ancient Earth are among the most important group of prokaryotes. *Gloeobacter violaceus* (Rippka et al., 1974) is the most primitive among all living cyanobacteria. The *G. violaceus* PCC 7421 strain was isolated from the surface of a limestone rock in Kernwald (Switzerland). The phylogenetic comparison of the SSU 16S rRNA and other genes from the various cyanobacteria suggests that *G. violaceus* diverged at

an early stage of evolution of cyanobacterial lineage (Seo and Yokota 2003; Hoffmann et al., 2005; Gupta and Mathews 2010) and is hence an important organism with respect to evolution related studies. *G. violaceus* possess a simple cell organization without thylakoids and an unusual structure of photosynthetic apparatus (Rippka et al., 1974; Bryant et al., 1981; Guglielmi et al., 1981). *G. violaceus* possess several distinct characteristics which are mentioned below:

- The cells are devoid of any thylakoid structures and as reported from the microscopic images the photosynthetic machinery of the organism is situated in the cytoplasmic membrane while in other cyanobacteria it is found in the thylakoid membrane (Rippka et al., 1974). This observation suggests that the components facing the lumen in other cyanobacteria are facing the periplasm in the case of *G. violaceus*, which raises the possibility that the photosynthetic electron transfer system and the respiratory systems which also might occur in the cytoplasmic membrane co-exist and possibly also share some machinery (Nakamura et al., 2003).
- The phycobilisome morphology in *G. violaceus* is quite distinct from that found in other cyanobacteria. In *G. violaceus* the phycobiliproteins form 6 peripheral rod-shaped elements which in turn form bundle shaped aggregates with five horizontal rods of an allophycocyanin core, situated vertically, adjacent to the inner surface of the cytoplasmic membrane.
- Further, the fatty acid composition is also different. Sulfoquinovosyl diacylglycerol (SQDG) although considered to play an important role in photosystem stabilization is absent in *G. violaceus*.

G. violaceus has been used as a model organism for studies based on evolution of oxygenic photosynthesis (Bernat et al., 2012) and is also among the initial set of cyanobacteria whose complete genome was sequenced (Nakamura et al., 2003). Various aspects of genus *Gloeobacter* remain comparatively unexplored including habitat, distribution, life cycle and identification of variability in the genus. The *Gloeobacter* strains studied till date include PPC 7421, PCC 9601 and PCC 8105. Rippka et al. (1974) suggest a close relationship of *G. violaceus* to a not very well studied botanical species *Gloeotheca coerulea* Geitler 1927, because of the similar cell morphology, life cycle and the presence of polar granules similar to those found in *G. violaceus* (Geitler 1927). Further, *Aphanothece caldariorum* Richter 1880, another rock-inhabiting cyanobacterium (Hansgirg 1892) is similar to *Gloeotheca coerulea* in morphology and life cycle and is thus also a plausible close relative of *G. violaceus* (Komarek and Anagnostidis 1999).

1.3. RuBisCO – “A Relic of a bygone age” (Ellis, 2010)

Although, photosynthesis is one of the most important biological processes, its evolution is still elusive and is believed to have come into existence after the divergence of the Archaeal – bacterial and eukaryal lineages, as no chlorophyll has been detected in archaeal species. Although, recently Meng et al. (2009) have reported the presence of an *in-vivo* functional bacteriochlorophyll in an uncultivated crenarchaeota, the group implicates that possibly the bacteriochlorophyll encoding gene identified could have a function other than that expressed *in vivo*, but it definitely gives new perspective to the evolution of the photosynthetic machinery. Further, on the basis of various forms of RuBisCO sequences available in the database, Tabita et al. (2008) speculate that the form III RuBisCO, arising from *Methanomicrobia* is possibly the ancestor of all forms of RuBisCO and RuBisCO-like proteins.

1.3.1 The distinct forms of RuBisCO

All RuBisCO enzymes are multimeric. Two different types of subunits viz. the large catalytic subunit (L, 50-55 kDa) and small subunit (S, 12-18 kDa), in various combinations constitute the different forms of RuBisCO. Basically, there are four forms i.e. I, II, III, and IV. Forms I, II and III have RuBP dependent CO₂ fixing ability. Form I is present in most chemoautotrophic bacteria, cyanobacteria, eukaryotic algae and all higher plants (Andersson 2008). While form II is found in purple non-sulphur bacteria, certain chemoautotrophic bacteria (Shively et al., 1998) and in eukaryotic dinoflagellates (Morse et al., 1995; Whitney et al., 1995; Rowan et al., 1996). The archaea have been found to have form III RuBisCO, however, they lack phosphoribulokinase (PRK), which is necessary for the synthesis of RuBP (the substrate for RuBisCO). The occurrence of RuBisCO form III in archaea in absence of PRK could be justified by two possibilities, i.e. firstly, RuBP is not the substrate for RuBisCO in archaea or secondly, the archaea have an alternative pathway for RuBP synthesis (Tabita et al., 2007). As stated by Finn and Tabita (2004) even an extensive analysis, suggest absence of an alternative substrate for carboxylation by RuBisCO in any organism. Further, Tabita et al. (2007) deciphered the occurrence of potential PRK encoding genes in recently sequenced genomes of certain archaeal organisms. Form IV termed as RuBisCO-like proteins (RLPs) are putative RuBisCO sequences, found in organisms that do not use CO₂ as a major source of carbon and are involved in the methionine salvage pathway in many bacteria (Tabita et al., 2007).

Form I shows great diversification and is further differentiated into four types i.e. IA, IB, IC, and ID. Forms IA and IB belong to the green grouping and are found in proteobacteria, cyanobacteria, green algae and higher plants. Forms IC and ID belong to the red group and

are found in proteobacteria and non-green algae. The form IA enzymes are further of two types IAc and IAq depending on two distinct types of small subunits and gene arrangements. Form IAc is associated with carboxysome genes (Cannon et al., 2001, 2003) and form IAq is not. The major difference between the proteins encoded by the IA genes is in the small subunit of L_8S_8 enzymes. All IAq small subunit sequences are characterized by a six-amino-acid insertion at the N-terminal end of the protein. From the crystal structure of form IAc RuBisCO from *Halothiobacillus neapolitanus* (Kerfeld et al., 2005) and the crystal structure of cyanobacterial form Ibc enzyme (Newman et al., 1993; Newman and Gutteridge 1994) it can be deduced that the region of insertion of small subunit sequence in Ia_q occurs at a position on top exterior face of small subunit, with potential interactions between adjoining small subunits and external proteins (Long et al., 2005). An insertion at this point would disrupt these interactions (Badger and Bek 2008).

The large subunits of forms I-IV exhibit an overall 25-30 percent sequence identity and have differences in function also (in case of RLPs), but secondary structure of the catalytic subunit of the enzyme is well conserved and the differences, if any, are localized to a few loops (Andersson and Buckland 2008). The overall fold common to all large subunits constitutes a smaller amino-terminal domain consisting of four to five stranded β sheets with helices on one side of the sheet and a larger carboxy-terminal domain featuring an eight stranded α/β barrel structure (Andersson 2008). The active site is located at the intradimer interface between the carboxy-terminal domain of one large subunit and the amino terminal of the second large subunit in the L2 dimer and the loops 1, 2 and 5-8 at the carboxy terminal end of the β -strands of large subunit contribute residues involved in catalysis (Andersson 2008). The small subunit of RuBisCO is more diverse. The common core comprises a four stranded antiparallel β -sheet with two helices on one side (Knight et al., 1989, 1990). The most distinct variations occur in two main regions i.e. the βA - βB -loop and the carboxy terminus (reviewed in Spreitzer 2003). The βA - βB -loop of four small subunits line the opening of the solvent channel that traverses the holoenzyme and are involved in various interactions both to the large subunits as well as the small subunits (Andersson 2008).

1.3.2. The Active site of RuBisCO

The active site of the enzyme is harbored in the large subunit. The function of the small subunit of RuBisCO is still not clear. Considering the existence of organisms with form II of RuBisCO which consists of the large subunit alone, it is expected that RbcS is not directly involved in catalysis, but possibly has a role that aids efficient carboxylation of its substrate. Meyer et al. (2012) have reported the involvement of RbcS in pyrenoid formation in *Chlamydomonas reinhardtii*. The study revealed that the reported function could only be carried out by the *C. reinhardtii* RbcS and could not be augmented if a higher organism RbcS

was used instead. In cyanobacteria, RbcS has been reported to be involved in the carboxysome assembly. Despite of the variation in the subunit compositions and the amino acid sequences of the various forms of RuBisCO, the basic folding and arrangement of the active site is conserved in all RuBisCOs. Further, it should also be noted that despite the conservation of the active site of all RuBisCO proteins existing in nature, the variation in the specificity factor between forms I and II can vary by factor of 15 folds and the rate of carboxylation can vary by 6 folds (Techerkez et al., 2006; Parry et al., 2007).

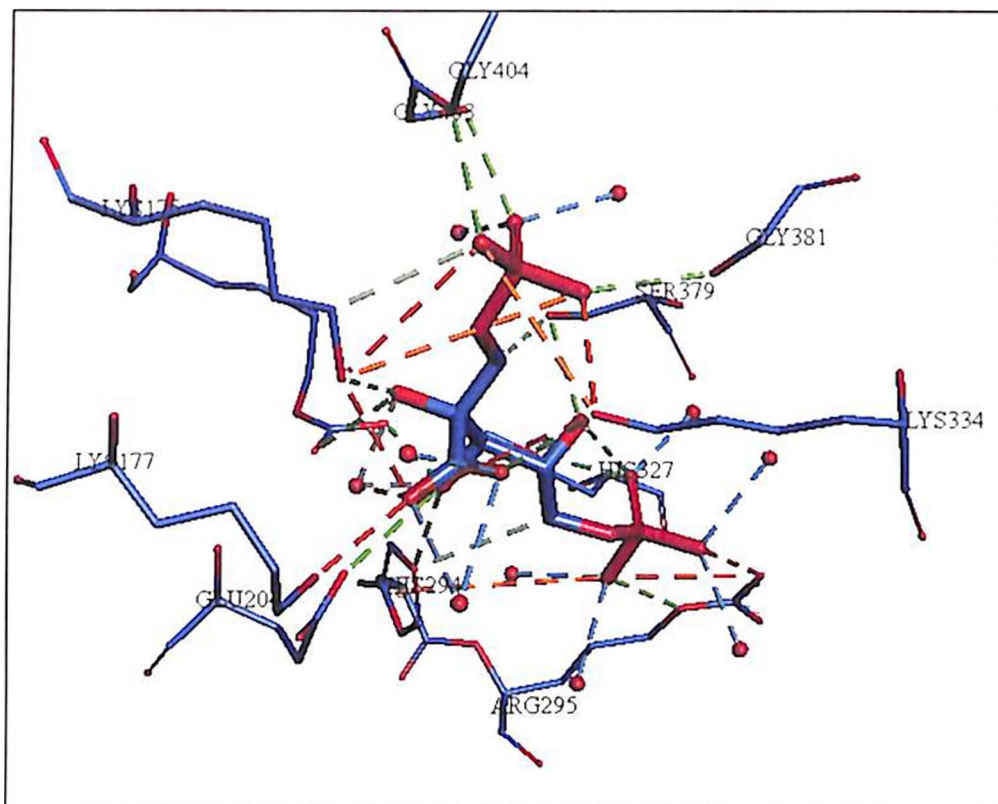


Figure 1A: Active site residues of *Spinacia oleracea* RuBisCO (Karkehabadi et al., 2003)

The Lysine residues suggested by Hartman et al. (1986) to be play an essential role in RuBisCO catalysis occupy positions 166 and 329 in *R. rubrum* RuBisCO and positions 175 and 334 in spinach RuBisCO. Further, the pKa values of the suggested sites and their enhanced nucleophilicities argue that these residues are involved in catalysis rather than in substrate-enzyme assembly (Hartman et al., 1986). The group also reports that the distance between the two active site lysine residues lies within the range of 12Å. The **Figure 1A** shows the arrangement of the catalytic site of Spinach RuBisCO large subunit. The residues involved are found to remain conserved in all forms of functional RuBisCO.

1.3.3. Mechanism of action

All bonafide RuBisCO proteins must first be activated by carbamylation at ϵ amino group of active site lysine residue Lys201 (Lorimer and Miziorko, 1980) by a CO_2 molecule which is distinct from substrate CO_2 . The carbamylated Lys201 is stabilized by divalent cations (Mg^{2+})

bound to adjacent acidic residues. The carbamate and divalent complex participation in the catalysis constrains the ways in which the active site can be modified to impact specificity or turnover (Hartman and Harpel 1994). Once activated the carboxylation reaction involves 5 steps and at least three transition states. The steps involved include enolization of RuBP, carboxylation of the 2,3-enediolate, hydration of the resulting ketone, carbon-carbon scission and stereospecific protonation of the resulting carboxylate of one of the product 3PGA (Andersson 2008). Carbamylation causes only minor changes in the conformation of RuBisCO large subunit affecting residues in the loop between β strands 2 and 3 of the N-terminal domain (Taylor and Andersson 1997). But, the binding of RuBP and other phosphorylated ligands induces several structural changes in and around the active site, the most distinct being a 12Å shift of the $\alpha\beta$ barrel loop 6 from a retracted (open) to an extended (closed) position (Schreuder et al., 1993; Taylor and Andersson 1997).

Magnesium has a very important role to play in the RuBisCO catalysis. Magnesium, along with carbamylated Lys201 (in Spinach RuBisCO), which provides monodentate ligand, is liganded by two monodentate ligands obtained from Asp203 and Glu204 (for Spinach RuBisCO) along with three water moieties (Andersson et al., 1989). The Asp203 and Glu204 interact via their oxygen atoms with Lys175 and Lys177, respectively so as to prevent the carboxylate groups of the later to interact with the metal ion, which if happens prevents binding of gaseous substrates. RuBP binds by replacing two water molecules and CO_2 gets bound by replacement of the third water molecule. Alonso et al. (2009) report ~90% and ~55%, reduction in RuBP catalysis by replacement of Mg^{2+} for activation, with Ca^{2+} , and Co^{2+} , respectively.

1.3.4. RuBisCO Activase

The multiple sequence alignment analysis suggests RuBisCO activase to be a member of extended AAA⁺ family of ATPases (Neuwald et al., 1999). It is reported to be involved in light-dependent regulation of RuBisCO by relieving the inhibition of phosphorylated compounds that bind to both carbamylated and non-carbamylated RuBisCO active sites (Portis, 2003; Portis et al., 2008). The tight binding of RuBP to uncarbamylated enzyme and 2CAP to the carbamylated enzyme renders conformational changes that cause the enzyme to be inactive and hence the activase is required to release the sugar phosphates and reopen the catalytic site.

1.3.5. Assembly of RuBisCO

As reported by Alonso et al. (2009) and Goloubinoff et al. (1989), *Methanococcooides burtonii* RuBisCO and *R. rubrum* RuBisCO (form II RuBisCOs), respectively, can assemble into dimers in *E. coli*, but further assembly into L₄, L₆, L₁₀ conformations requires the involvement

of chaperones. The studies of Alonso et al. (2009) have shown RuBP induced assembly of *M. burtonii* L₂ RuBisCO into decamers. It is to be noted that the group provides evidence for the fact that the catalytic properties of the dimers and the decamers are quite similar. While the more complex hexadecameric assembly of form I RuBisCO involving small subunit and the large subunit require RuBisCO-specific chaperones, which not only limit their assembly and hence impair their catalytic activity in prokaryotic host cells.

The assembly of RuBisCO is assisted by several chaperones viz. GroEL, GroES, and RbcX. Initially, the RbcL monomers undergo folding with the help of GroEL-GroES system, with C-terminal of the protein remaining bound to GroEL. The extended C-terminal region of RbcL is identified by RbcX₂ by virtue of conserved motif of the former. Simultaneously, RbcX₂ interacts with the N-terminal of another RbcL monomer leading to the formation of dimers, following which RbcX₂-bound RbcL dimers assemble to tetramers, forming the RbcL₄-RbcX₂₈ complex. Finally, RbcS binds to the complex leading to a conformational change in the RbcL C-terminal which in turn facilitates the release of RbcX₂ and formation of a stable holoenzyme.

The importance of the enzyme has resulted in a plethora of reports on various aspects of catalysis and assembly. Some important residues have been identified in the LSU and SSU for holoenzyme formation. Ser112F and Gly322S residues are involved in LSU-LSU interactions in tobacco RuBisCO (Avni et al., 1989; Shikanai et al., 1996). The Cys172 residue in *C. reinhardtii* RuBisCO large subunit when replaced with Ser by Moreno and Spreitzer (1999) caused an increase in the stability of the holoenzyme. The importance of the residues 346-348 of *Synechococcus* RuBisCO LSU in holoenzyme assembly came to the forefront when Ramage et al. (1998) and Romanova et al. (1997) substituted VDL to YNT or YHT, resulting in impaired assembly of the holoenzyme. The “long distance” interactions between residue 290 and residues 222 and 262 are particularly interesting because all three residues are in contact with the βA-βB loop of the small subunit.

1.3.6. Selection of RuBisCO towards a better enzyme if not a perfect one

The best way to compare the catalytic efficiencies of different enzymes or the turnover of different substrates by the same enzyme is to compare V_{\max}/K_M ratio. The preference of RuBisCO for CO₂ versus O₂ is represented by specificity factor (Ω), the ratio of catalytic efficiency of carboxylation (V_c/K_c) to that of oxygenation (V_o/K_o). The Ω values reveal that there has been a selection for an enzyme with higher specificity for CO₂ at the cost of carboxylation rate, with bacteria displaying low specificity values but high turnover and the higher plants displaying high specificity and low turnover rate (Andersson 2008). **Figures 1B** and **1C** depict the comparison of the kinetics of RuBisCO in terms of catalytic rate (K_{cat}) vs

selectivity $S(\Omega)$ and $K_M \text{ CO}_2$ vs $K_M \text{ O}_2$ (K_M - Michaelis-Menten constant) among various organisms, respectively. The general trend observed is that the RuBisCO from lower organisms viz. archaeobacteria (for which data is available) have lower affinity and selectivity for CO_2 and yet a higher catalytic turnover, which can be explained by the fact that these organisms inhabit the anaerobic environment. Further, it is to be noted that if we compare kinetics of RuBisCO from C_3 and C_4 plants, it can be elucidated that even though C_3 plants are devoid of a CCM, they have a higher affinity and selectivity for CO_2 and catalytic rate comparable to that of RuBisCO from C_4 plants, validating the fact that RuBisCO in each organism has undergone adaptations which make it suitable to work best in its native conditions. The non-green algae RuBisCO forming the other extreme exhibit highest selectivity, the lowest turnover number and a moderate affinity for O_2 and CO_2 .

Techerkez et al. (2006) have hypothesized the variations in the kinetic properties of RuBisCO are an outcome of the optimization between two parameters i.e., specificity to CO_2 and catalytic turnover rate of the enzyme. The enzymes that now have a higher specificity have evolved such that they attain a transition state very similar to the end product that binds the enzyme tightly so that the carboxylation reactions get favored, but such binding restricts maximum catalytic throughput. The hypothesis suggests that although RuBisCO has undergone modifications that have not enhanced its activity to the maximum but have provided it with certain optimization according to its environment.

1.4. Carbon Concentrating Mechanisms (CCMs)

CCMs, as mentioned before, are mechanisms adopted by various photosynthetic organisms to overcome the confused specificity of RuBisCO by concentrating CO_2 in its vicinity and hence reduce the effect of competing substrates. According to the pathways of CCM evolution derived by Badger et al. (2002) the carboxysomes being the most primitive developed first in cyanobacteria and then differentiated into α and β forms with the splitting of α and β -cyanobacteria. Sicheritz-Ponten and Andersson (2001) suggest that the appearance of carboxysomes in proteobacteria could possibly be an outcome of lateral gene transfer. According to Tomitani et al. (1999) the event of divergence of α and β -cyanobacteria occurred before the primary endosymbiosis. Under such conditions, as speculated by Badger and Price (2003) the carboxysomes would have evolved independently in both α and β -carboxysomes.

1.4.1. CCMs in cyanobacteria

CCM of cyanobacteria viz. carboxysomes is a proteinaceous microcompartment, enclosing RuBisCO and carbonic anhydrase (CA). The generation of CO_2 by CAs in the vicinity of RuBisCO coupled with restriction of its diffusive efflux from carboxysomes leads to the

accumulation. The substrate for carboxysomes i.e., HCO_3^- (bicarbonate) is assimilated by CO_2 and HCO_3^- active transporters both in the plasma membrane and the thylakoid membrane. Badger et al. (2002) state that the first steps towards developing a cyanobacterial CCM would have been the evolution of a carboxysome structure for RuBisCO accumulation. The inorganic carbon sources available to the photosynthetic organisms are in the form of both CO_2 and HCO_3^- , but only the former can be utilized by RuBisCO, hence the presence of CAs would have been essential as otherwise the conversion of HCO_3^- to CO_2 would be very slow. Further, Badger and Price (2003) state that the origin of the low-affinity transporters and subsequently high-affinity HCO_3^- and CO_2 -uptake systems would have been necessary as the CO_2 stress conditions became more severe.

1.4.1.1. *Two types of carboxysomes with distinct components and evolution*

As elucidated by Badger et al. (2002), on the basis of RuBisCO, the two primary groups of cyanobacteria i.e. α -cyanobacteria and β -cyanobacteria possess form IA and form IB of RuBisCO, respectively. The sequencing has shown that the carboxysomes possessed by α -cyanobacteria (encoded by *csa* operon) are quite different from the ones possessed by β -cyanobacteria (encoded by *ccm* operon) (Badger and Price 2003) and are termed as α -carboxysomes and β -carboxysomes, respectively. The α -carboxysomes are present in *Prochlorococcus* as well as some marine *Synechococcus* species, the genetic components of carboxysome are organized into an operon and it encapsulates β -CA and a yet to be characterized structural protein, CsoS2. The β -carboxysomes are found in all cyanobacteria except *Prochlorococcus* and *Synechococcus* and the genes are more dispersed as compared to that of α -carboxysomes.

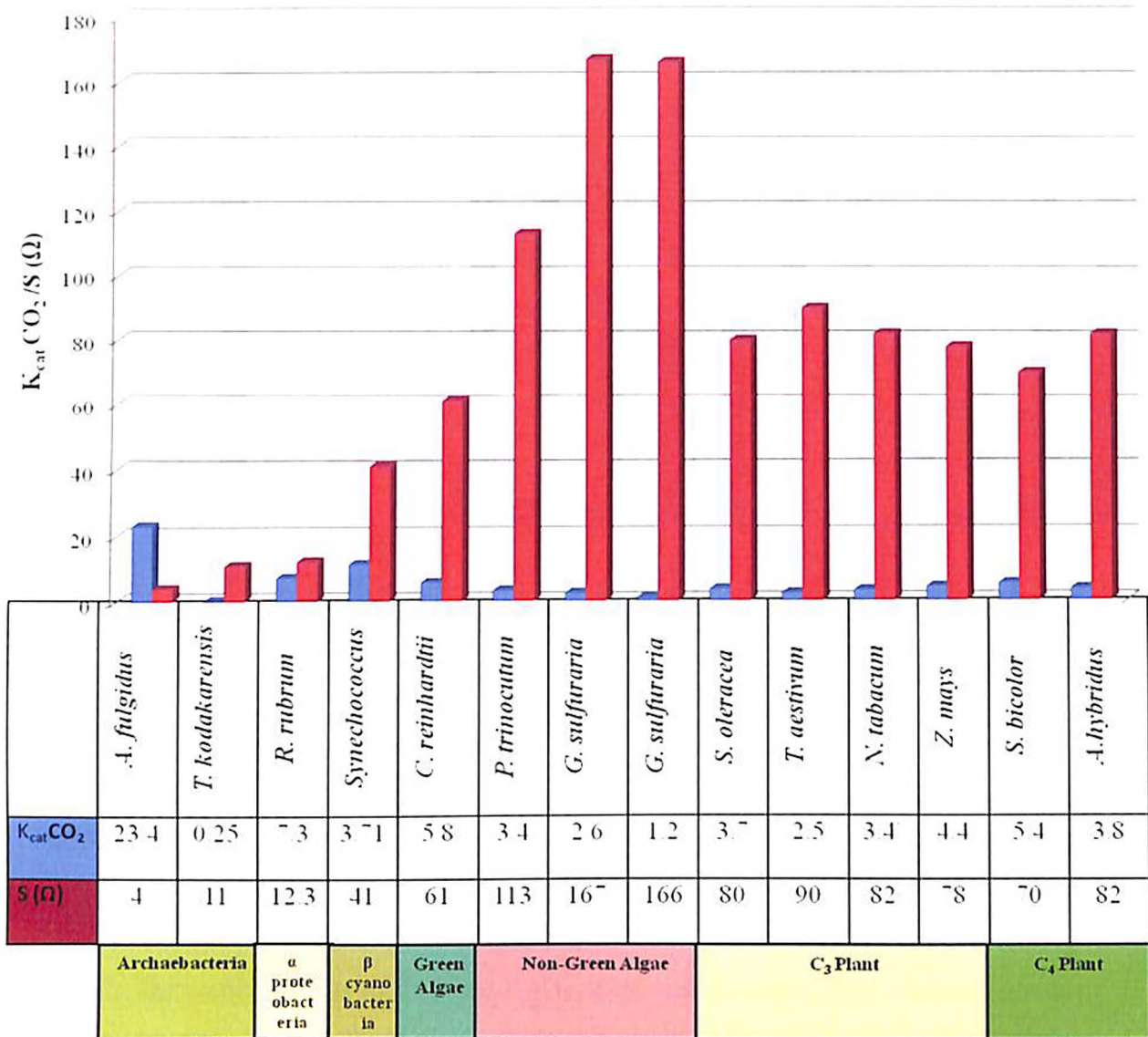


Figure 1B: The comparison of K_{cat} (catalytic turnover) and S (selectivity) of enzyme RuBisCO from various organisms. The organisms included in the comparison include Archaeobacteria: *A. fulgidus* (*Archaeoglobus fulgidus*; Kreel and Tabita 2007), *T. kodakarensis* (*Thermococcus kodakarensis*; Yoshida et al., 2007) Proteobacteria: *R. rubrum* (*Rhodospirillum rubrum*; Kane et al., 1994) Cyanobacteria: *Synechococcus* (*Synechococcus* sp. strain PCC 7002 Yoshida et al., 2007) Green Algae: *C. reinhardtii* (*Chlamydomonas reinhardtii*; Jordan and Ogren 1981) Non-Green Algae—*P. tricornutum* (*Phaeodactylum tricornutum*; Whitney et al., 2001), *G. monilis* (*Griffithsia monilis*; Whitney et al., 2001), *G. sulfuraria* (*Galdieria sulfuraria*; Whitney et al., 2001) C_3 plants—*S. oleracea* (*Spinacia oleracea*; Jordan and Ogren 1981; Sage and Seemann 1993), *T. aestivum* (*Triticum aestivum*; Jordan and Ogren 1981; Zhu et al., 1998), *N. tabacum* (*Nicotiana tabacum*; Whitney et al., 2001), C_4 Plants—*Z. mays* (*Zea mays*; Jordan and Ogren 1981; Sage and Seemann 1993), *S. bicolor* (*Sorghum bicolor*; Jordan and Ogren 1981), *A. hybridus* (*Amaranthus hybridus*; Jordan and Ogren 1981).

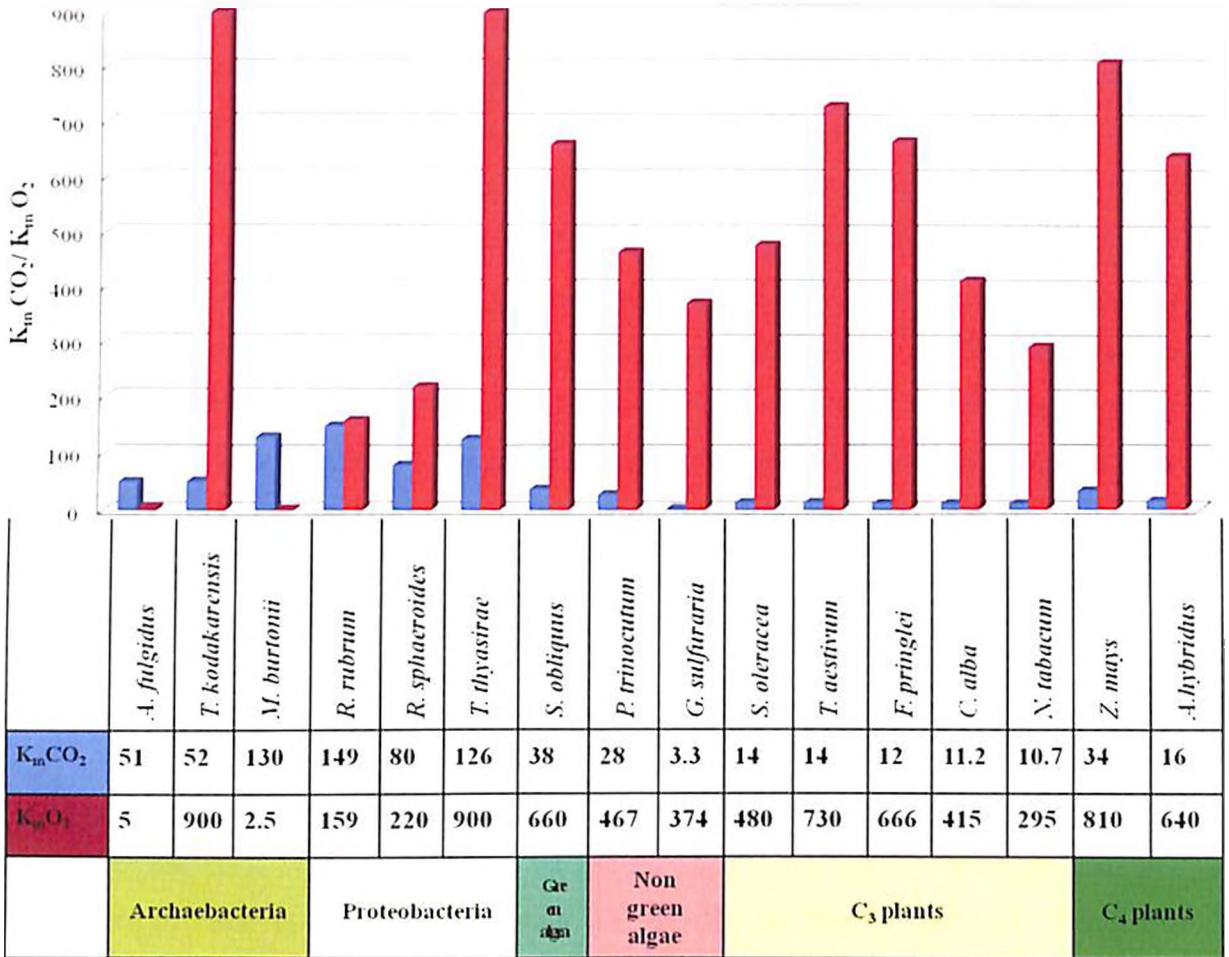


Figure 1C: The comparison of $K_m \text{CO}_2$ and $K_m \text{O}_2$ of enzyme RuBisCO from various organisms. The organisms included in the comparison are Archaeobacteria—*A. fulgidus* (*Archaeoglobus fulgidus*, Kreel and Tabita 2007), *T. kodakarensis* (*Thermococcus kodakarensis*, Yoshida et al., 2007), *M. burtonii* (*Methanococcoid burtonii* Alonso et al., 2009) Proteobacteria—*R. rubrum* (*Rhodospirillum rubrum*, Kane et al., 1994), *R. sphaeroides* (*Rhodopseudomonas sphaeroides*, Savira et al., 2010), *Thiomicrospira thyasirae* (*T. thyasirae*, Cook et al., 1991) Green Algae—*Scenedesmus obliquus* (*S. obliquus*, Savira et al., 2010) Non- Green Algae—*Phaeodactylum tricomutum* (*P. tricomutum*, Whitney et al., 2001), *Galdieria sulfuraria* (*G. sulfuraria*, Whitney et al., 2001) C₃ plants—*Spinacia oleracea* (*S. oleracea*, Jordan and Ogren 1981; Sage and Seemann 1993), *Flaveria pringlei* (*F. pringlei*, Savira et al., 2010), *Chenopodium alba* (*C. alba*, Savira et al., 2010), *Triticum aestivum* (*T. aestivum*, Jordan and Ogren 1981), *Nicotiana tabacum* (*N. tabacum*, Whitney et al., 2001) C₄ Plants—*Zea mays* (*Z. mays*, Jordan and Ogren 1981; Sage and Seemann 1993), *Amaranthus hybridus* (*A. hybridus*, Jordan and Ogren 1981).

Carboxysomes are found in all cyanobacteria except for 'Cyanobacterium UCYN-A' which as per recent reports lacks major photosynthetic pathway (Zehr et al., 2008, Thompson et al., 2012) and although being related to Cyanothecce genus of cyanobacteria has adopted symbiotic lifestyle (Criscuolo and Gribaldo 2011). As proposed by Rae et al. (2011) and Blank and Sanchez-Baracaldo (2010) α - β divergence of cyanobacteria occurred approximately 1.0 Gya and the α -cyanobacteria gained their *csa* operon from gammaproteobacteria such as *Nitrococcus* by horizontal gene transfer (HGT) (Marin et al., 2007). Further, on the basis of the facts that α - β cyanobacteria came into existence by divergent evolution and the carboxysomes constituted by both are encoded by a distinct set of genes, it is suggested by Rae et al. (2013) that the carboxysomes in each of the cyanobacteria evolved independently of each other. The above-mentioned suggestion is further supported by the fact that none of the α -cyanobacteria has been found to possess β -carboxysomes and vice versa. Another important observation made by Espie and Kimber (2011) i.e. there is a greater homology between the each of the carboxysome shell proteins and the shell proteins of other microcompartments than among each other, further supports convergent evolution of the carboxysomes. The two types of carboxysomes although have similar function and a similar outward appearance, they have quite distinct internal arrangement of the proteins.

It should be noted that all cyanobacteria contain carboxysomes or at least possess the gene repertoire for the same (except for Cyanobacterium UCYN-A) but the occurrence of the carboxysomes in bacteria is sporadic. Further, as reported by Badger and Bek (2008) the RuBisCO protein sequences of autotrophic bacteria contain signals for the occurrence of carboxysomes. The RuBisCO form IAc is associated with carboxysomes and the existence of form IAq suggests the absence of carboxysome operon.

1.4.1.2. Components of carboxysomes:

Carboxysomes have an icosahedral geometry. The proteins that form the basis of this geometry belong to two domain classes: BMC /Pfam00936 domain i.e. Bacterial microcompartment proteins constituting hexamers that assemble to lead to the formation of the flat facets of the shell and CcmL/EutN (Pfam 03319) domain (Kerfeld et al., 2005) proteins that form the pentamers that occupy the vertices and hence provide curvature to the carboxysome shell. The BMC domain proteins of α -carboxysomes characterized from *H. neapolitanus* are CsoS1A and CsoS1C (Tsai et al., 2009). In β -carboxysomes, CcmK2 and CcmK4 have been characterized from *Synechocystis* sp PCC 6803 to have BMC domain (Kerfeld et al., 2005). CcmK2 protein units polymerize to form an even layer of hexamers oriented in the same direction and CcmK4 forms a layer of hexamer faces oriented in alternating convex and concave faces. With the advent of genomic sequencing two more *ccmK* homologs were identified in β -cyanobacteria (Price et al., 1998, Cannon et al., 2001,

2002). Rae et al. (2012) reported that although these CcmK homologs are involved in the formation of a shell of β -carboxysomes, but the functionality of the microcompartment is not hampered in the complete absence of the two as compared to the other shell proteins, which if absent result into high C_i requiring phenotype.

The BMC domain proteins of α -carboxysomes (Cso-type homologs) and β -carboxysomes (Ccm-type homologs) share a conserved fold important for hexamer-hexamer interaction that is D-X-X-X-K motif and a comparatively less conserved domain R-P-H-X-N motif at the edges of the hexamer (Kinney et al., 2011). The α -carboxysomes have CsoS4A and CsoS4B proteins with Pfam03319 domain while the β -carboxysomes have CcmL as the Pfam03319 domain protein (Kinney et al., 2011).

The central pore (4.7 Å diameter) of the subunit facets of the shell proteins forms another important feature of the carboxysome shell (Kerfeld et al., 2005). These pores in the shell protein hexamers act as a channel for metabolites, HCO_3^- ions, RuBP and 3PGA. The conserved residues K-I-G-S and R-(A/V)-G-S constitute the wall of the pores in CcmK2 and CcmK4 protein hexamers, respectively (Kinney et al., 2011). The pores of till date structurally characterized carboxysomes are predominantly positively charged which is favorable for the passage of negatively charged metabolites such as HCO_3^- ions and at the same time, CO_2 leakage to the cytoplasm and O_2 entry into the carboxysomes is inhibited. The conserved residues of pores of pentamers are R-G-S-A-A for CcmL and G-S-S-A-A for CsoS4A (Kinney et al., 2011).

The proteins CsoSID in α -carboxysomes and CcmO in β -carboxysomes are the fused BMC domain proteins that form part of the carboxysome shell. CsoSID forms trimers resulting in pseudo hexamers dimensionally similar to hexamers with conserved domain D-X-X-X-K at the edges. CcmO, a tandem BMC domain protein in β -cyanobacteria is apparently a fusion of two CcmK-like proteins. Hence, the carboxysome shell is composed of CcmK2, CcmK4, CcmL and CcmO in β -carboxysomes and CsoS1A, CsoSID, CsoS4A, CsoS4B and CsoSID in α -carboxysomes. The other proteins identified in β -carboxysomes involved in the functioning of the carboxysomes include CcmM and CcmN. CcmM encodes a protein with an N-terminal homologous to γ -CA (Alber and Ferry 1996) and a C-terminal region having three to four RuBisCO small subunit repeats (SSU repeats) (Ludwig et al., 2000). CcmM has been confirmed to be an essential structural component of the carboxysomes (Long et al., 2005) as its disruption leads to non-functional carboxysomes and the inability of the organism to grow under limiting C_i conditions (Woodger et al., 2005). Studies on *Synechococcus* PCC 7942 have revealed the occurrence of a full-length CcmM form (58kDa) i.e., M58 and a shorter 35kDa CcmM i.e., M35 resulting from a putative Internal Ribosomal Entry Site (IRES) within the *ccmM* gene transcript (Long et al., 2007). Similar reports for the expression of

multiple CcmM forms from a single gene have been observed in case of *Synechocystis* PCC 6803 (Ludwig et al., 2000; Cot et al., 2008), *Synechococcus* PCC 7002 (Ludwig et al., 2000), and *Acaryochloris marina* MBIC11017 (Long et al., 2010). Each form constitutes three RbcS repeat motifs, each motif is capable of binding RbcL to form a CcmM-RuBisCO complex and possibly also crosslink with other RuBisCO molecules within the carboxysomes shell. Evidence reveals that such interactions have potential to be the basis for the initial assembly of the carboxysomes (Long et al., 2007). On the basis of studies conducted by Long et al. (2010, 2011), M58 is confined to the inner shell of the carboxysome where it is responsible for recruitment of carboxysomal carbonic anhydrase and outer shell as well as interlinking of RuBisCO molecules (Long et al., 2007, 2010; Cot et al., 2008). The M58 interacts with the outer shell directly via CcmK2 protein and indirectly via CcmN protein (which possibly is involved in the interaction between the inner shell and outer shell proteins) (Kinney et al., 2012). Further, the M35 form is restricted to the shell lumen involved in interactions with the RuBisCO molecules and as suggested by electron microscopic studies of carboxysomes causing arrangement of the later in two planes (Kaneko et al., 2006). Conceivably CsoS2 plays a similar role as CcmM in α -carboxysomes, but affirmation is so far lacking.

The N-terminal of CcmM has a γ -CA like domain. CcaA, which is likely to be associated with the shell, interacts with the N-terminal of M58. Long et al. (2007) speculate that the N-terminal region of M58 performs the role as a scaffold for CcaA and other carboxysomal proteins. Recently, the crystal structure of the CA-like domain of *Thermosynechococcus elongates* BP-1 CcmM has been found out and it has been shown to be an active carbonic anhydrase in itself and other species lacking CcaA (Pena et al., 2010). The carboxysomal protein CcmN contains bacterial transferase hexapeptide repeat found in the N-terminal of CcmM as well, but whether this indicates potential structural homology between CcmM and CcmN, is still undeciphered (Long et al., 2007). Another carboxysome protein recently identified i.e. CcmP is reported to be an orthologue of α -carboxysomal protein CsoSID (Cai et al., 2012). CsoSID, as speculated from its crystal structure, provides a potential pore for the passage of large metabolites in α -carboxysomes.

Till date as many as five Ci uptake systems have been identified in cyanobacteria, which include three HCO_3^- transporters and two CO_2 uptake systems, among which some are constitutive while others are inducible only when grown under Ci limiting conditions. The energy sources for these Ci uptake systems could be in the form of ATP (BCT1 HCO_3^- transporters), NADPH or reduced ferredoxin (CO_2 uptake) and coupling to electrochemical Na^+ gradient (as in SbtA or BicA HCO_3^- transporters) (Badger et al., 2002; Badger and Price 2003). HCO_3^- being ionic is approximately 1000 fold less permeable to lipid membranes than the uncharged CO_2 molecule and is hence the preferred form of Ci for accumulation in the

cytoplasm (Price et al., 2008). Upon entry into the carboxysomes, HCO_3^- is acted upon by CAs for conversion to CO_2 . In β -carboxysomes Cca lefA and in α -carboxysomes shell protein CsoSC carry out the CA activity (So et al., 2004).

Under conditions of Ci sufficiency as well, cyanobacteria exhibit a low transport affinity form of CCM, indicating that a number of genes involved in CCM are constitutive (Kaplan and Reinhold 1999). Under Ci limited conditions several β -cyanobacteria are reported to be able to upregulate the expression of high-affinity Ci transporters such as NDH-1_c, SbtA, BCT1 or BicA (Price et al., 2008) controlled by LysR type transcription regulators namely CcmR (NdhR) or CmpR (Wang et al., 2004). Although, the exact mechanism involved in this regulation is not known, but it was found that the expression of these CO_2 responsive genes is inversely proportional to the size of the internal Ci pool during illumination (Woodger et al., 2003). The carboxysome proteins are under study by various research groups in different parts of the world providing information about the knowledge gaps in carboxysome structure and function.

1.4.1.3. Biogenesis of Carboxysomes

The structural differences in the proteins constituting the two forms of carboxysomes along with putative variation in their internal organization suggest the existence of distinct pathways of biogenesis for both kinds of carboxysomes. The β -carboxysome constitutes outer shell (CcmK, CcmL, and CcmO), RuBisCO organizing layer/inner shell (M58, RuBisCO, CcmK, CcmL, CcmO, CcmN and CcaA) and the lumen (M35, RuBisCO SSU and RuBisCO LSU). As reported by Long et al. (2010, 2011), β -carboxysomes get expressed even with nonstandard stoichiometries. Further, Kinney et al. (2012) and Rae et al. (2012) revealed the occurrence of shell-free β -carboxysomes, although at a comparatively lesser number under Ci sufficiency conditions, suggesting β -carboxysomes to be lumen centric. Kinney et al. (2012) suggest CcmN as the key player in the interaction between the inner shell and the outer shell along with CcmM (part of the inner shell) which interacts with CcmK2 of the outer shell. Rae et al. (2013) along with Orus et al. (1995) hypothesize that β -carboxysome formation is initiated by assembly of lumen proteins organized by CcmM, followed by encapsulation by the inner shell proteins and the outer shell proteins with involvement of the CcmN and CcmM in organization of each of the layers. The α -carboxysomes, in contrast, are likely to be shell-centric (Menon et al., 2008; Baker et al., 1998; Iancu et al., 2010). Gonzales et al. (2005) showed the involvement of CsoS2 in carboxysome assembly by direct interaction with RuBisCO, CsoSCA, and the outer shell proteins. The overall mechanism of α -carboxysome still remains elusive.

1.4.1.4. Regulation of Carboxysome Genes

In general, within a range of C_i concentration, the β -carboxysome constituting genes are constitutively expressed. Reports show little or temporary change (2 to 5 fold rise) in the expression of *ccmKLMNO*, *ccaA* and *rbcLXS* at low CO_2 concentrations in comparison to the low CO_2 responsive genes (*bicA*, *sbtAB*, *cmpABCD*, *porB* and *ndhF3-ndhD3-chpY*) which undergo drastic changes with respect to transcript formation. Further, electron microscopic analysis reveals formation of a greater number of carboxysomes under low C_i conditions as compared to high C_i conditions (McKay et al., 1993). On the other hand, the regulation in the formation of α -carboxysomes is not much studied in α -cyanobacteria. In proteobacteria possessing α -carboxysomes not much change has been observed in expression of carboxysomal RuBisCO operons in response to varying C_i concentrations (Heinhorst et al., 2006) while in some cases the noncarboxysomal RuBisCO has been found to be preferred in C_i depleting conditions or even complete repression of CO_2 fixation (Kusian and Bowien 1997; Dubs and Tabita 2004).

1.4.1.5 Microcompartments

Cell compartmentalization was considered an asset of eukaryotes alone until the discovery of carboxysomes in cyanobacteria in 1961 (Jensen and Bowen 1961). The discovery of such compartments was delayed largely because of the icosahedral shape and nano-scale size as these were mistaken as viral capsids. BMCs are polyhedral protein complexes (40-200 nm in diameter), encasing metabolic enzymes encapsulated in a protein shell and are important for bacteria for optimization of metabolic processes such as carbon fixation, ethanolamine utilization, 1,2 propanediol utilization etc. (Kerfeld et al., 2010). The BMCs serve to exclude the cell from toxic metabolites formed by the enclosed enzymes, omit the loss of volatile intermediates and prevent the interference of competing substrates (Penrod and Roth 2006; Sampson and Bobik 2008). Confinement of metabolic pathways further is assumed to provide better efficiency and enhanced protein stability due to exclusion of oxidative damage (Huseby and Roth 2013). The major kinds of BMCs discovered till date include carboxysomes (carbon fixation) (Jensen and Bowen 1961), Pdu BMC (1,2 propanediol utilization) (Chen et al., 1994) and Eut BMC (Ethanolamine utilization) (Kofoid et al., 1999).

The most extensively studied BMC is carboxysomes. Carboxysomes are a part of a CCM of cyanobacteria and some proteobacteria. The carboxysomes evolved in order to overcome the inefficiency of enzyme RuBisCO, which it suffers at the hands of photorespiration because of its fickle specificity to carbon dioxide and oxygen (Badger et al., 1998).

1.4.2 CCM in Proteobacteria

The proteobacteria include most of the gram-negative chemoheterotrophic bacteria and are presumed to have arisen from a common photosynthetic ancestor. However, only a few are now photosynthetic as several different metabolic and nutritional capacities have arisen to replace this characteristic. The subgroups of proteobacteria are designated by Greek letters i.e., α , β , and γ proteobacteria.

The α proteobacteria are capable of growth at very low levels of nutrients, in a wide range of environments which differ in CO_2 and O_2 concentrations i.e., soil, sludge, water and plant roots and stems environments. They contain four main forms of RuBisCO i.e., form II, IC, IAq and IAe (Badger and Bek 2008) possibly as an adaptation to survive in diverse environments. The photo organotroph *Bradyrhizobium sp.* BTAl has form IAe RuBisCO, (Giraud et al., 2000) and two facultative chemolithomixotrophs – *Nitrobacter hamburgensis* and *Nitrobacter winogradski* have α -carboxysomes, form IAe RuBisCO and form IC RuBisCO, suggesting the possible occurrence of a CCM under low CO_2 conditions (Strakenberg et al., 2006).

The β -proteobacteria is generally less flexible in their metabolic activities as compared to α -proteobacteria. The forms of RuBisCO found in various members of the group are form II, IC, IAq, and IAe. *Thiobacillus denitrificans* an obligate chemolithotroph has form IAq RuBisCO and α -carboxysome genes placed in close proximity, suggesting it to be in a transition state between metabolic state adapted to low CO_2 conditions or to a future state (Badger and Bek 2008). *Thiomonas intermedia* contains both form IAe and form II RuBisCO which can be attributed to the variable environments of aerobic soils and low CO_2 conditions of aerobic aquatic environments (Badger and Bek 2008).

The γ proteobacteria include a range of obligate, photo and chemolithotrophs and also some organotrophs. The RuBisCO forms IAq and IAe are most dominant in this group but there are some occurrences of form IC and II also. *Thiomicrospira crunogena* is a chemolithoautotroph (sulfur oxidizing) inhabiting deep sea hydrothermal vents, oscillating between environments with a CO_2 concentration of 20 to 1000 μMoles (Goffredi et al., 1997). Dobrinski et al. (2005) has reported rapid accumulation of C_i (in both forms i.e., HCO_3^- and CO_2) in *T. crunogena* when grown at limiting CO_2 concentration. *Nitrococcus mobilis* is a nitrite-oxidizing bacterium and contains a single form IAe RuBisCO and α -carboxysome genes. *Allochromatium vinosum* has mostly form IAq RuBisCO, but it also contains genes for form IAe and hence has potential to form CCM (Badger and Bek 2008). The halotolerant sulfur oxidizing species *H. neapolitanus* has form IAe and form II RuBisCO indicating the potential to inhabit low CO_2 environment (Badger and Bek 2008).

1.4.3. CCM in algae

Algae being mostly autotrophic are responsible for a large proportion (50%) of the global primary productivity (Behrenfeld et al., 2001). Along with the shortcomings of RuBisCO, algae, in comparison to land plants face another challenge that is the availability of CO₂ in an aqueous environment is far less than that in air, due to 10,000 times slower diffusion of CO₂ in water. Further, algae are also exposed to flickering C_i-supply at hands of respiration of soil biota and the effect of pH on the relative distribution between dissolved CO₂ and HCO₃⁻ (Spalding 2008). Different algal species possess different forms of RuBisCO, green algae have form IB and S_c value ranging from 35-90; rhodophytes, cryptophytes, heterokonts and haptophytes have form II RuBisCO and high S_c (Ω) values; and the dinophytes exhibit form II RuBisCO with low S_c value of ~30 (Raven and Beardall 2003). The various kinds of CCM in algae are based on localization of CO₂ accumulation in compartments to have an acidic environment, or on biochemical mechanisms such as C₄ photosynthesis and crassulacean acid metabolism (CAM) (Giordano et al., 2005).

1.4.3.1 CCM based on enhancement of CO₂ concentration by acidification – Pyrenoids

The basis of the CCM that causes localized elevation of CO₂ concentration, as elucidated by Pronina, Ramazanov, and Semenenko (1981) relies on the set up of pH gradient across the chloroplast- thylakoid membrane, in the presence of light, i.e., the chloroplast stroma has a pH close to 8.0, and the thylakoid lumen has a pH between 4 and 5. The pK_a of interconversion of HCO₃⁻-to-CO₂ is approximately 6.3; hence chloroplast stroma has HCO₃⁻ as the predominant C_i species while thylakoid lumen is rich in CO₂. In presence of such a mechanism, the bicarbonate transported into lumen of the thylakoids is converted to carbon dioxide, hence creating a high CO₂ environment (Pronina, Ramazanov and Semenenko 1981). These thylakoid structures have been reported to be penetrating the pyrenoid structures where RuBisCO is present and hence the latter is facilitated with above ambient CO₂ concentration. Another important component for this mechanism to be possible is the HCO₃⁻ and CO₂ transporters in the chloroplast and thylakoid membranes. Moreover, the set up of the pH gradient across the two compartments is due to the light-driven electron transport chain and is hence not possible in the absence of light. This type of mechanism is evident in the eukaryotic green alga *C. reinhardtii* induced by both low C_i concentration and low light intensity (Yamano et al., 2008). Structurally, pyrenoids are present in the chloroplast as a dense matrix material, constituting tightly packed fibrils approximately 60 Å in diameter (Gibbs 1962).

The components of CCM in *Chlamydomonas* have been identified by studies based on mutant and transcriptome analysis (Yamano and Fukuzawa 2009). The nucleus-encoded and constitutively expressed Ccm1/Cia5 has been reported to be the epical controller of the CCM

of *Chlamydomonas* (Wang et al., 2005). Further *Ler1*, *Lei14*, *Lei16* and *Lei25* which are putative MYB-DNA-binding regulatory factor, putative GTPase, TB2-DP1 and U-box domain, respectively, induced by limiting CO₂ (reviewed by Yamano and Fukuzawa 2009). The LCR1 transcriptional regulator, relays and amplifies the low CO₂ signal from CCM1 to at least three CO₂-responsive genes, *cah1*, *lei1*, and *lei6* (Yoshioka et al., 2004).

The Ci transporters identified in *Chlamydomonas* include *Lei1*, *LeiA*, *LeiB*, *Cep1*, *Cep2*, *Hla3*, and *Rh1*. Among these *Lei1*, *LeiB*, *Cep1*, and *Cep2* are low CO₂ induced while *Hla3* is low CO₂ and high light induced and *Rh1* is high CO₂ induced (reviewed by Yamano and Fukuzawa 2009). The *Lei1* and *Rh1* transporters are found in the cytoplasmic membrane while the *LeiA* and *Cep1* transporters are in the chloroplast membrane and *LeiB* is in the vicinity of the pyrenoids (reviewed by Yamano and Fukuzawa 2009).

A number of CA have also been characterized in *Chlamydomonas*, some of which are found to be involved in the conversion of HCO₃⁻ to CO₂ and vice versa according to the pH of the environment of its respective location. The CA identified include α -CA namely CAH1, CAH2, CAH3, and β -CA namely CAH4, CAH5, CAH6, CAH7, CAH8 and CAH9. CAH1 is periplasmic and its role is to facilitate entry of CO₂ into the algal cell and convert any HCO₃⁻ species to CO₂. The low CO₂ condition leads to induction of *cah1* gene (Fukuzawa et al., 1990) and has been found to be regulated by a silencer region, which represses transcription under high-CO₂ conditions or in the dark, and an enhancer region, which activates the same under low-CO₂ conditions in the presence of light (Kueho et al., 1999).

1.4.3.2 CCM in diatoms

Diatoms contribute as much as 40% to the primary productivity of the ocean. Diatom plastids originated from a secondary endosymbiosis with a red alga (Roberts et al., 2007). Although diatom RuBisCO has relatively high CO₂ selectivity over O₂, the low diffusion of CO₂ in water reduces the rate of photosynthesis. There are very significant gaps in our understanding of CCM in diatoms.

The occurrence of both CO₂ and HCO₃⁻ transporters is known, but the location of these transporters is still not known. The confirmation of the exact location of such transporters i.e., in the thylakoid membrane or the plasmalemma or both could make the kind of CCM present in diatom evident. Hence, the studies of the transport of the Ci species across the plastid envelope need to be done to find out any such possibility. There is also the affirmation of presence of a C₄-like biochemistry in one marine diatom which could act as, or be part of a CCM. Another mechanism which could be feasible in diatoms is the accumulation of CO₂ from HCO₃⁻ – at low pH maintained by a H⁺ pump, in vicinity of RuBisCO. Although no

CaH₃ analogue (carbonic anhydrase for conversion of HCO₃⁻ to CO₂) is reported in diatoms, but the pervasive luminal photosystem II protein PsbO has CA activity (Enami et al., 2005).

The basic biochemistry of C₄ and CAM metabolism are discussed in the next section. Reinfelder and colleagues (Reinfelder et al., 2004) found profound evidence of C₄ photosynthesis in the marine diatom *Thalassiosira weissflogii* under low zinc nutrition and low CO₂ (Reinfelder et al., 2000), but whether it functions as a CCM has been questioned. There is also substantial affirmation for function of a CCM through a C₄ cycle in the marine macroalga *Ulota flabellum* (Bowes et al., 2002).

1.4.4. CCM in higher plants - C₄ metabolism and Crassulacean Acid Metabolism (CAM)

Terrestrial plants and a few aquatic macrophytes evolved a biochemically and anatomically complex organic carbon pump called the C₄ pathway. Such a mechanism appeared in plants particularly found in regions with high temperatures and developed characteristics of a specialized leaf anatomy that ultimately led to high photosynthetic rates, high growth rates, low photorespiration rates and low rates of water loss. Photosynthesis in the leaves of C₄ plants is carried out in two leaf types: mesophyll and bundle sheath cells. The CO₂ taken up by the mesophyll cells in the presence of enzyme phosphoenolpyruvate carboxylase is fixed to form a four carbon compound oxaloacetate. The oxaloacetate formed is either reduced to malate using NADPH or converted to aspartate by transamination. Either of the products formed is then transported to bundle sheath cells via plasmodesmata. In the bundle sheath cells malate undergoes oxidation and decarboxylation to yield pyruvate and CO₂ by the reduction of NADP⁺ in the presence of malic enzymes. Hence CO₂ is made available for use by RuBisCO. On the other hand when the product formed in mesophyll cells is aspartate, it is transaminated to form oxaloacetate and reduced to malate once in the bundle sheath cells and finally CO₂ is released by malic enzyme or PEP carboxykinase. The pyruvate formed in bundle sheath cells is transferred back to mesophyll cells where it is converted to PEP by pyruvate phosphate dikinase.

CAM is a photosynthetic adaptation to periodic drought as it allows the gas exchange to occur in the night time when due to lower temperatures the rate of water loss is reduced. As opposed to C₄ metabolism where the assimilation and fixing of CO₂ are separated by carrying out the two processes in different compartments, in CAM they are partitioned by a temporal separation. The CO₂ is assimilated in the dark in the form of malic acid and stored in the vacuoles. During the day the malate is released from the vacuole and utilized by RuBisCO for fixation into RuBP to form PGA.

1.4.5 C₃ plants do not have any CCM

The C₃ plants constitute many commercially important plants but they do not have any CCM and are hence at the mercy of RuBisCO design alone. The C₃ plants have RuBisCO with higher specificity but lower turnover as compared to the C₄ plants which in the presence of the CCM are capable of creating a CO₂ saturating condition in which the RuBisCO can afford to have lower specificity which in turn gives it higher turnover. Moreover, the increase in temperature reduces the amount of CO₂ available to the organism and hence the photosynthetic efficiency is further reduced.

Extensive studies are being carried out worldwide to explore the possibility of enhancing C₃ photosynthesis by various strategies, one among which is by introduction of components of CCM or the complete CCM into C₃ plants. The C₄ plants being most closely related to the C₃ plants should be the most plausible candidate for consideration. But the introduction of C₄ metabolism would involve biochemical as well as anatomical changes and is hence challenging (Zhu et al., 2010). In that case the single cell C₄ metabolism discovered in *Borszczowia aralocaspica* and *Bieneria cycloptera* could be considered (Smith et al., 2009). Moreover, the cyanobacterial bicarbonate transporters or the entire CCM of cyanobacteria could be introduced into plants systems leading to reduced photorespiration and enhanced turnover (Price et al., 2008). Similarly, the pyrenoids possessed by eukaryotic algae also act to concentrate CO₂ in vicinity of RuBisCO. Anthocerotae is the only terrestrial plant class with a recognized pyrenoid and the genes involved in its formation are (Smith and Griffiths 1996), possible candidates for being introduced into C₃ plants for enhanced photosynthetic yield in the latter.

Introducing each of these systems would involve a complex feat of gene engineering and given that several components of each of these systems are still uncharacterized, the introduction of a complete CCM could be successful only in the long run (Raines 2011). In short term, the introduction Ci transporters alone could give positive effect on the CO₂ assimilation by RuBisCO (Price et al., 2011).

1.5. Gaps in Existing Research

The inefficiency of RuBisCO as mentioned above has led the various photosynthetic organisms to develop mechanisms viz. CCMs to overcome the loss suffered at the hands of fickle specificity of the former. According to the current information, the CCM evolved about 400 million years ago when the atmospheric conditions were too stressful for the RuBisCO (with the design suitable for CO₂ rich atmosphere) to carry out its function. Further, it has been proposed that the CCM first appeared in cyanobacteria in the form of carboxysomes which underwent changes in various organisms to take more complex forms. Many aspects of

carboxysomes, as well as pyrenoids, are still unexplored and hence limit our understanding. The function of some component of carboxysomes and the regulation of the assembly of the carboxysome are still not known. Further, the scientific world is still struggling to elucidate the role of the small subunit of RuBisCO. Although several studies annotate RuBisCO SSU to be directly indirectly involved in catalysis, while others consider it to be part of stabilizing machinery of RuBisCO multimer, but the exact function is still elusive. The studies targeting RuBisCO from diverse sources is timely as it could possibly unlock the role of various factors involved in its assembly and catalysis and hence raise the possibility of enhancing its activity. *G. violaceus* PCC 7421 being an early diverging cyanobacterium is an important organism in context to studying evolution, and is hence targeted in the current study.

1.6. Objectives of the Proposed Research

In light of the available information and in-lieu of understanding the aspects of RuBisCO and CCM proteins from an organism expected to be of ancestral origin these proteins were analyzed *in-vitro* and *in-silico*. Further, attempts were made to assess the effect of the SSU of RuBisCO and CcmM on the activity of the LSU of RuBisCO. In addition, the evolutionary relation between these proteins was also analyzed. The present study is the first report on the characterization of RuBisCO and CCM encoding genes from this ancestral organism. The objectives designed to address these queries are mentioned as follows:

- To clone RuBisCO operon genes (*rbcL*, *rbcS*, and *rbcX*) and genes encoding carbon concentration mechanism proteins (*ccmK*, *L*, *M*, *N*, *O*) from the genome of the early diverging cyanobacterium *Gloeobacter violaceus* and characterize them at the molecular level.
- To check if the *ccmM* and/or *rbcS* of *Gloeobacter violaceus* could augment the catalytic activity of a RuBisCO (e.g., *Rhodospirillum rubrum*) lacking small subunit.
- To elucidate structural and functional evolutionary links between RuBisCO small subunit gene and CCM proteins (*ccmM* encoded) from *Gloeobacter violaceus* by checking if CcmM is capable of substituting RbcS activity *in vitro*.

1.4.5 C_3 plants do not have any CCM

The C_3 plants constitute many commercially important plants but they do not have any CCM and are hence at the mercy of RuBisCO design alone. The C_3 plants have RuBisCO with higher specificity but lower turnover as compared to the C_4 plants which in the presence of the CCM are capable of creating a CO_2 saturating condition in which the RuBisCO can afford to have lower specificity which in turn gives it higher turnover. Moreover, the increase in temperature reduces the amount of CO_2 available to the organism and hence the photosynthetic efficiency is further reduced.

Extensive studies are being carried out worldwide to explore the possibility of enhancing C_3 photosynthesis by various strategies, one among which is by introduction of components of CCM or the complete CCM into C_3 plants. The C_4 plants being most closely related to the C_3 plants should be the most plausible candidate for consideration. But the introduction of C_4 metabolism would involve biochemical as well as anatomical changes and is hence challenging (Zhu et al., 2010). In that case the single cell C_4 metabolism discovered in *Borszczowia aralocaspica* and *Biennertia cycloptera* could be a considered (Smith et al., 2009). Moreover, the cyanobacterial bicarbonate transporters or the entire CCM of cyanobacteria could be introduced into plants systems leading to reduced photorespiration and enhanced turnover (Price et al., 2008). Similarly, the pyrenoids possessed by eukaryotic algae also act to concentrate CO_2 in vicinity of RuBisCO. Anthocerotae is the only terrestrial plant class with a recognized pyrenoid and the genes involved in its formation are (Smith and Griffiths 1996), possible candidates for being introduced into C_3 plants for enhanced photosynthetic yield in the latter.

Introducing each of these systems would involve a complex feat of gene engineering and given that several components of each of these systems are still uncharacterized, the introduction of a complete CCM could be successful only in the long run (Raines 2011). In short term, the introduction C_i transporters alone could give positive effect on the CO_2 assimilation by RuBisCO (Price et al., 2011).

1.5. Gaps in Existing Research

The inefficiency of RuBisCO as mentioned above has led the various photosynthetic organisms to develop mechanisms viz. CCMs to overcome the loss suffered at the hands of fickle specificity of the former. According to the current information, the CCM evolved about 400 million years ago when the atmospheric conditions were too stressful for the RuBisCO (with the design suitable for CO_2 rich atmosphere) to carry out its function. Further, it has been proposed that the CCM first appeared in cyanobacteria in the form of carboxysomes which underwent changes in various organisms to take more complex forms. Many aspects of

CHAPTER 2

Materials and Methods

[†] This chapter details the general protocols used in the present research and any specific methods followed are mentioned in subsequent chapters.

The chapter has been divided into two major parts viz. the *in-silico* analysis and the *in-vivo/in-vitro* analysis (not necessarily in the exact order of its performance).

2.1 *In-silico* analysis of RuBisCO and CCM proteins of cyanobacteria

2.1.1. The distribution of CCM proteins across Eubacteria and Archaea

The amino acid sequences of carbon concentrating mechanism protein of *Gloeobacter violaceus* PCC 7421 retrieved from NCBI database of 2015 (www.ncbi.nlm.nih.gov) (Geer et al., 2010) were used as the query against 31 phyla in BLASTP, since according to 16S rRNA analysis, it is an early diverging cyanobacteria. The accession numbers for each carbon concentrating mechanism (CCM) proteins of *G. violaceus* used in the study are mentioned in Table 2A. The retrieved sequences were then saved as text files (.txt) and used for finding homologs in the NCBI database by pBLAST (Altschul et al., 1997). The sequences of the hits obtained with significant E-value and identity score were obtained and saved in .txt files.

Table 2A: The Accession numbers and the conserved domain profile of the proteins of CCM from *G. violaceus* as mentioned in NCBI database

| S.No | Protein Name | Accession Number | Amino acid length | Domain present |
|------|--------------|------------------|-------------------|-------------------------------------|
| 1 | CcmK | BAC90037.1 | 102aa | BMC ccmK domain |
| 2 | CcmI | WP_011142092.1 | 100aa | EutN ccmI domain |
| 3 | CcmM | WP_011142091.1 | 668aa | Lbh-gamma CA and RuBisCO SSU domain |
| 4 | CcmN | NP_925038.1 | 201aa | LbetaII superfamily domain |
| 5 | CcmO | NP_925037.1 | 245aa | BMC ccmK domain |

2.1.2. Identification of conserved domains and phylogenetic analysis

The sequences of the RuBisCO and carbon concentrating mechanism proteins from *G. violaceus* PCC 7421 retrieved from pBLAST were used for this analysis (as mentioned in section 2.1.1). The amino acid sequences were used to carry out multiple sequence alignment in MEGA 6.0 (Tamura et al., 2013) by ClustalW (Larkin et al., 2007) (using Gonnet protein weight matrix) for phylogenetic analysis and T-Coffee server (Notredame et al., 2000) (BLOSUM62 weights matrix) for sequence logo. The phylogenetic analysis was done in MEGA 6.0 using maximum likelihood method. The MSA was used to identify conserved domains and sequence logos were generated using Weblogo3 (Schneider and Stephens 1990; Crooks et al., 2004) with MSA fasta file as input. (The exact details of the algorithm used are mentioned where and when performed)

2.1.3. The study of arrangement of RuBisCO encoding genes in cyanobacteria to determine if it is random or has functional significance

In order to ascertain the role of *rbcX*, we analysed genome of various cyanobacteria retrieved from Kazusa genome resource (<http://genome.microbedb.jp/cyanobase/>) for the analysis of the location of *rbcS*, *rbcX*, and *rbcL* and whether these subunits are located adjacent to each

other or are dispersed in the genome. The analysis also aimed to implicate whether the distance of *rbcA* with respect to *rbcL* and *rbcS* is making any significant contribution towards its function.

2.1.4. The analysis for evolutionary links between RuBisCO and CCM proteins

The CD (conserved domain) (Marchler-Bauer et al., 2015) search of the CcmM protein mentioned Table 2A revealed the presence of RuBisCO SSU repeats in CcmM protein. This observation was further investigated. The CcmM protein sequence was retrieved from all β -cyanobacteria available on Kazusa Genome Resource. The list of organisms and their corresponding proteins are mentioned in Table 2B. Each of the sequences retrieved were used a query for pBLAST in NCBI, separately, and the best hits for RuBisCO proteins were identified. Based on e value and percent identity, the best hits obtained were listed. The best homologs obtained were then aligned with each of the query sequence to identify variation in the identity of the repeats found.

Table 2B: List of organisms used for study of evolutionary link between CcmM and RuBisCO proteins.

| S.No | Name of Organism | CcmM protein loci |
|------|---|-----------------------------------|
| 1 | <i>Microcystis aeruginosa</i> NIES843 | MAE47910; 4393189-4395147 |
| 2 | <i>Synechocystis</i> sp PCC 6803 | sll1031; 216264-218327 |
| 3 | <i>Anabaena</i> PCC 7120 | all0865; 994401-996068 |
| 4 | <i>Arthrospira platensis</i> NIES39 | NIES39_K04800; 4369863-4371503 |
| 5 | <i>Cyanothece</i> sp PCC 7424 | PCC7424_1369; 1506143-1508170 |
| 6 | <i>Synechococcus</i> sp JA-2-3B' (2-13) | CYB_1794; 1881751-1883712 |
| 7 | <i>Synechococcus</i> sp JA-3-3Ab | CYA_1614; 1617937-1619901 |
| 8 | <i>Synechococcus elongatus</i> PCC 6301 | syc0133_c; 143122-144741 |
| 9 | <i>Cyanothece</i> sp ATCC 51142 | ccc_4280; 4407231-4409231 |
| 10 | <i>Nostoc punctiforme</i> ATCC 29133 | Npun_F4294; 5380153-5382162 |
| 11 | <i>Gloeobacter violaceus</i> PCC 7421 | gll2093; 2246256-2248262 |
| 12 | <i>Thermosynechococcus elongatus</i> BP-1 | tlh0944; 964745-966703 |
| 13 | <i>Synechococcus elongatus</i> PCC 7942 | Synpcc7942_1423; 1476059-1477678 |
| 14 | <i>Anabaena variabilis</i> ATCC 29413 | Ava_4469; 5595463-5597133 |
| 15 | <i>Synechococcus</i> sp PCC 7002 | SYNPCC7002_A1800; 1885447-1887432 |
| 16 | <i>Acaryochloris marina</i> MBIC11017 | AM1_3052; 3074749-3075681 |
| 17 | <i>Trichodesmium erythraeum</i> IMS101 | Tery_3849; 5946968-5948986 |
| 18 | <i>Cyanothece</i> sp PCC 7425 | Cyan7425_1615; 1463687-1465639 |
| 19 | <i>Cyanothece</i> sp PCC 8801 | PCC8801_1599; 1673139-1675139 |

2.1.5. In-silico modeling of CCM proteins and RuBisCO proteins of *G. violaceus* PCC 7421

The *G. violaceus* PCC 7421 CCM and RuBisCO proteins' models are not available; hence the protein sequences were used for homology modeling by MODELER (Webb and Sali 2014) to

generate tertiary structures. The models generated were further refined using online tool Modrefiner (Xu and Zhang 2011). The models generated were further validated by Ramachandran plot (Ramachandran et al., 1963), Verify3D (Bowie et al., 1991), QMEAN score (Benkert et al., 2008), IM score (Zhang and Skolnick 2004) and PDBsum (Laskowski 2007; de Beer et al., 2014). The models generated were then used to identify possible interactions between the proteins speculated to be involved to form quaternary structures on the basis of literature available.

2.1.6. Designing of primers for amplification of RuBisCO and CCM encoding genes of *G. violaceus* PCC 7421

Primer designing lays foundation for cloning strategy hence needs to be done carefully. The two important properties of primers are specificity and efficiency. Specificity is frequency with which the mispriming of the primers is avoided and efficiency is defined as how close the primer pair is capable of amplifying the desired region of the template. Specificity is controlled by the length and the annealing temperature of the primer pair. Primers between 18-24bp is considered to be specific if the annealing temperature is set within a few degrees of the melting temperature of the primers. Perfect base pairing between the 3' end of the primer and the template is optimum for obtaining good amplification results. The restriction site is incorporated towards the 5' end of the oligonucleotide. The restriction enzyme requires 3-6 bases at the 5' end as a sitting position to carry out the digestion of the phosphodiester linkage. A reasonable GC content of the primers should be maintained. The preferred GC content is around 50% so as to obtain a sufficient thermal window for efficient annealing. Further the primer self complementarity, primer dimer formation and primer non specificity must be taken care of to obtain optimum primer set.

The nucleotide sequence data sets for the RuBisCO encoding genes and the CCM genes were retrieved from NCBI was used for primer designing. The sigma DNA calculator tool (<http://www.sigma-genosys.com/calc/DNAcalc.asp>) was used to obtain melting temperature, molecular weight, occurrence of secondary structure and formation of primer dimer (if any) of the primer sequence. The uniqueness of the primer sequence was determined by nBLAST (www.ncbi.nlm.nih.gov/BLAST/BASTn). The primers were synthesized by outsourcing to Eurofins Analytical Services India Pvt. Ltd. The sequences of the primers used are mentioned in **Table 2C**.

Table 2C: List of primers used in the amplification of the genes under study³

| SNo | Primer name | Primer sequence (5' to 3') |
|-----|-----------------------------------|---|
| 1 | GV-RbcS-For- <i>NdeI</i> | CGG AGI CCC ATC <u>ATA TGC</u> AAA CGC TGC CCA |
| 2 | GV-RbcS-Rev- <i>XhoI</i> | GCA TCG GGC <u>CTC GAG</u> CTA GGG CTT GTA GAC |
| 3 | GV-RbcX- <i>NdeI</i> -F | ACG CGI <u>CAT ATG</u> GAT CAT CAG |
| 4 | GV-RbcX- <i>XhoI</i> -R | TCC GTA <u>CTC GAG</u> TCA ATC GTT |
| 5 | GV-RbcL- <i>NdeI</i> -F | CAG GAG GTC <u>CAT ATG</u> TCC TAT ACG |
| 6 | GV-RbcL- <i>BamHI</i> -R | CGC AGG <u>GGA TCC</u> CTA GAG TTT |
| 7 | GV-CemO- <i>NdeI</i> -F | GG CAA AAC GGA GGA <u>CAT ATG</u> ATT GAG AGG ATT GCC |
| 8 | GV-CemO- <i>BamHI</i> -R | CGC CCA CCA <u>GGA TCC</u> CTA ACG CTC GGG |
| 9 | GV-CemN- <i>NdeI</i> -F | GC GGC CCT GAA <u>CAT ATG</u> GCT TCT CTG CC |
| 10 | GV-CemN- <i>BamHI</i> -R | C CAA GGA TGG <u>GGA TCC</u> CTA CGC TCC TCC |
| 11 | GV-CemM- <i>NdeI</i> -F | GG AGG AAT CGG <u>CAT ATG</u> GTT GCA CAA AGC |
| 12 | GV-CemM- <i>XhoI</i> -R | GA GAA GCC <u>CTC GAG</u> TCA GGG CCG CTG G |
| 13 | GV-CemL- <i>NdeI</i> -F | CAG ACC <u>CAT ATG</u> CAG ATT GGC |
| 14 | GV-CemL- <i>BamHI</i> -R | GCA TTC <u>GGA TCC</u> CTA GCG G |
| 15 | GV-CemK1- <i>NdeI</i> -F | GGG AGG TTG <u>CAT ATG</u> CCT ATT |
| 16 | GV-CemK1- <i>BamHI</i> -R | AAA CGC <u>GGA TCC</u> TCA AGT |
| 17 | GV-CemK2- <i>NdeI</i> -F | GGA ATC <u>CAT ATG</u> GCA GTT GC |
| 18 | GV-CemK2- <i>BamHI</i> -R | TGC CAA <u>GGA TCC</u> TCA GGG |
| 19 | GV-RuBisCO operon- <i>NdeI</i> -F | CAG GAG GTC <u>CAT ATG</u> TCC TAT ACG AAG |
| 20 | GV-RuBisCO operon- <i>XhoI</i> -R | TCG GGC <u>CTC GAG</u> CTA GGG |
| 21 | GV-Cem operon- <i>NdeI</i> -F | GGG AGG TTG <u>CAT ATG</u> CCT ATT GCT GTT GGA ATG ATC |
| 22 | GV-Cem operon- <i>BamHI</i> -R | GC CCA CCA <u>GGA TCC</u> CTA ACG CTC GGG AG |

³ The restriction site inserted in the primers is underlined. The nomenclature followed for naming the primers is GV-name of organism- name of protein encoded by the gene for which the primers are designed - name of restriction site inserted - forward reverse primer.

2.2 *In-vitro/In-vivo* analysis

2.2.1. Amplification of the genes under study from the genome of *G. violaceus* PCC 7421

Polymerase Chain Reaction (PCR) is a powerful amplification technique capable of generating ample supply of DNA segment of interest from a small amount of starting material. Although, simple and reproducible, PCR can get complicated at times. The problems faced in PCR could be non-specific amplifications or at times no amplifications. Further, depending upon the enzyme used there could be introduction of unintentional mutations in the amplicons generated. These problems can be avoided by meticulous planning and patient efforts towards troubleshooting.

The recommended concentrations of the various PCR components are mentioned below:

| Component | Working Concentration |
|------------------------------|-------------------------------------|
| MgCl ₂ (Buffer A) | 1.5 mM |
| Primers | 10-30 pMoles |
| dNTPs | 200 μMoles (50 μMoles of each dNTP) |
| Polymerase | 1-2 units |
| Template | 10 ng |

It is essential to completely denature the template DNA at the beginning of PCR to ensure efficient utilization of the template during the first amplification cycle. If the GC content of the template is 50% or less, an initial denaturation of 1-3 minutes at 95°C is sufficient. For GC rich templates, this step should be prolonged to 10 minutes. In case even longer denaturation is required, *Taq* DNA polymerase should be added after the initial denaturation. A DNA denaturation time of 30 seconds per cycle at 95°C is normally sufficient. For GC rich DNA templates, this step can be prolonged to 3-4 minutes. DNA denaturation can be enhanced by addition of additives. The melting temperature of the primer-template complex decreases significantly in the presence of these additives, hence annealing temperature is adjusted accordingly and amount of enzyme should be increased. The annealing temperature is kept 3-5°C lower than the average of the melting temperatures of both the primers. Annealing for a period of 30 seconds is generally sufficient. If non-specific PCR products appear, the annealing temperature should be optimized stepwise in 1-2°C increments.

The PCR conditions standardized for various genes using *Taq* DNA Polymerase (Bangalore Genei) are mentioned in **Table 2D** using Applied Biosystems Veriti Thermocycler. *Taq* DNA polymerase is a 94KDa thermostable enzyme, its source organism being *Thermus aquaticus*, and hence its application in PCR. It possesses 5' to 3' polymerase activity and 5' to 3' exonuclease activity. One unit (U) of DNA polymerase is defined as the amount of enzyme

required to convert 10nmoles of total deoxyribonucleoside-triphosphates into precipitable DNA in 30 minutes at 72 °C. *Taq* DNA polymerase lacks 3' to 5' exonuclease activity (proofreading), hence the amplicons generated have adenine (A) overhangs at 3' end. It has a half life of greater than 2 hours at 92.5 °C, 40 minutes at 95 °C, 9 minutes at 97.5 °C and can replicate a 1000bp strand in less than 10 seconds at 72 °C. The conditions optimized for genes under study are mentioned in **Table 2D**.

Table 2D: The PCR conditions standardized for various genes

| S No | Step in PCR | <i>Ccmk1</i> | <i>Ccmk2</i> | <i>ccml</i> | <i>ccmN</i> | <i>ccmO</i> | <i>rbcS</i> | <i>rbcX</i> | <i>rbcL</i> | <i>Rubisco operon</i> |
|------|----------------------|----------------|------------------|--------------|---------------|---------------|----------------|------------------|---------------|-----------------------|
| 1 | Initial denaturation | 95 °C; 3min | 95 °C; 3min | 95 °C; 3min | 95 °C; 3min | 95 °C; 3min | 95 °C; 3min | 95 °C; 3min | 95 °C; 3min | 95 °C; 3min |
| 2 | Denaturation | 95 °C; 1min | 95 °C; 1min | 95 °C; 1min | 95 °C; 1min | 95 °C; 1min | 95 °C; 1min | 95 °C; 1min | 95 °C; 1min | 95 °C; 1min |
| 3 | Annealing | 48 °C; 1:30min | 50 °C; 1:30min | 52 °C; 1 min | 63 °C; 1 min | 63 °C; 1 min | 61 °C; 1:30min | 50 °C; 1:30min | 51 °C; 1 min | 58 °C; 1:30min |
| 4 | Extension | 72 °C; 1:30min | 72 °C; 1:30 mins | 72 °C; 1 min | 72 °C; 2 mins | 72 °C; 2 mins | 72 °C; 2 mins | 72 °C; 1:30 mins | 72 °C; 2 mins | 72 °C; 2:30 mins |
| 5 | Final extension | 72 °C 3mins | 72 °C 3 mins | 72 °C 4 mins | 72 °C 4 mins | 72 °C 4mins | 72 °C 4 mins | 72 °C 3 mins | 72 °C 4 mins | 72 °C 4 mins |
| 6 | Cycles | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 | 30 |
| | Expected size (bp) | 309 | 342 | 303 | 606 | 738 | 321 | 378 | 1425 | 2124 |

The conditions were standardized for the polymerase chain reactions of the genes under study viz. *ccmK*, *ccmL*, *ccmN*, *ccmO*, *rbcX*, *rbcS*, *rbcL*, and *rubisco operon* of *G. violaceus* PCC 7421 using *Taq* DNA polymerase (Genei Bangalore). The amplification for *ccmM* and *ccm operon* was obtained using Phusion Polymerase (Thermo Scientific). Thermo Scientific Phusion High-Fidelity DNA Polymerase brings together a novel *Pyrococcus*-like enzyme with a processivity-enhancing domain. The enzyme is capable of generating long templates with high speed and accuracy. The error rate of Phusion DNA Polymerase in Phusion HF Buffer is determined to be 4.4×10^{-7} , which is approximately 50-fold lower than that of *Thermus aquaticus* DNA polymerase, and 6-fold lower than that of *Pyrococcus furiosus* DNA polymerase. Phusion DNA Polymerase possesses 5'→3' DNA polymerase activity and 3'→5' exonuclease activity. The enzyme generates amplicons with blunt ends. The polymerase can generate amplicons as long as 7.5 kb and 20 kb. The PCR conditions standardized for *ccm operon* and *ccmM* using Phusion polymerase are mentioned in **Table 2E**.

Table 2E: PCR conditions standardized for *ccmM* and *ccm operon* sequences.

| | |
|---|---|
| <p>PCR Reaction Mix: <i>ccm operon</i> Water – 12 μl. 5X Phusion GC buffer – 4 μl. 10mM dNTPs – 0.4 μl. Primer F – 1 μl. Primer R – 1 μl. Template – 0.5 μl. DMSO – 0.4 μl. Phusion polymerase – 0.2 μl.</p> <p>PCR Reaction Mix: <i>ccmM</i> Water – 12 μl. 5X Phusion GC buffer – 4 μl. 10mM dNTPs – 0.4 μl. Primer F – 1 μl. Primer R – 1 μl. Template – 0.5 μl. DMSO – 0.4 μl. Phusion polymerase – 0.2 μl.</p> | <p>PCR Conditions: <i>ccm operon</i> Initial Denaturation – 98 C – 1 min Denaturation – 98 C – 7 secs Annealing – 72 C – 1:20 mins Extension – 72 C – 1:20 mins Final Extension – 72 C – 5 mins 4 C until removed from cycler</p> <p>PCR Conditions: <i>ccmM</i> Initial Denaturation – 98 C – 1 min Denaturation – 98 C – 7 secs Annealing – 72 C – 35 secs Extension – 72 C – 35 secs Final Extension – 72 C – 5 mins 4 C until removed from cycler</p> |
|---|---|

The PCR products obtained were checked on agarose gel by electrophoresis. The percentage of gel used was decided on the bases of the expected size of the amplicon with reference to Table 2G.

2.2.2. Agarose gel electrophoresis

The amplified products were analyzed on agarose gel. The agarose gel electrophoresis was carried out in submarine horizontal agarose slab gel apparatus procured from Tarson and Bio-Rad (Sambrook et al., 1989). The reagents used are mentioned below.

Gel loading buffer (Tracking dye 6X): EDTA (pH 8.0) = 1mM; Glycerol = 50% (v/v); Bromophenol Blue = 0.4% (w/v); Xylene Cyanol FF = 0.4% (w/v)

The 6X loading buffer is used as a DNA tracking dye in agarose gel electrophoresis. The buffer consists two dyes – Bromophenol Blue and Xylene cyanol FF. Bromophenol Blue migrates through agarose gels at approximately the same rate as linear ds DNA 300bp in length whereas Xylene cyanol FF migrates at approximately the same rate as linear ds DNA 4kb in length.

TAE buffer 50X 1L, pH 8.0 (autoclaved): 242 gm Tris Base; 57.1 mL Glacial acetic acid; 100 mL EDTA pH 8.0

TE Buffer 10 mL, pH 8.0: 100mM TrisHCl (pH 8.0); 10mM EDTA (pH 8.0)

Table 2F: Characteristics of Agarose Low Electroendosmosis (Low EEO) (DNase, RNase Free)

| Feature | Characteristics |
|-------------------------|-----------------------------|
| Gel strength (1.5% gel) | 1500-3000 g/cm ² |
| Gelling Temperature | 36 °C |
| Melting Temperature | 86-88 °C |

The agarose (characteristic of agarose mentioned in **Table 2F**) gels were made and electrophoresis was done in 1X TAE buffer. Agarose powder was suspended in 1X TAE buffer (concentration of the agarose gel was decided on basis of size of fragment to be examined, mentioned in **Table 2G**), and was dissolved by heating in a microwave oven (for uniform heating). The melted agarose gel was cooled down to approximately 50 °C, and just before pouring in the casting tray, ethidium bromide (EtBr) (0.5 µg/ml in TE buffer, pH 8.0) was added in the gel solution. After adding EtBr, the gel was poured in the electrophoresis boat. Wells were made by fixing the comb over the boat, prior to pouring of the molten agarose. After solidification the comb was removed and the gel was transferred to electrophoresis tank containing 1X TAE buffer. The DNA samples and molecular markers were mixed with tracking dye 1X (1:6 volume of DNA sample). Electrophoresis was carried out at 80-100V. The gel was then examined under transmitted UV light (302-312 nm) using a gel documentation system (Biorad). Gels of different strengths were prepared depending on the size of DNA to be analyzed. A table which illustrates appropriate gel strength for different sizes of DNA is given below (**Table 2G**):

Table 2G: The concentration of agarose gel suitable for different DNA sized fragments to obtain good resolution

| Gel strength (w/v) | Size of DNA (bp) |
|--------------------|------------------|
| 4% | 50-120 |
| 3% | 120-120 |
| 2% | 300-800 |
| 1.5% | 800-1000 |
| 1.2% | 1000-1200 |
| 1% | 1200-2000 |
| 0.8% | 2000 longer |

The amplicons obtained were purified by Gel Extraction kit (Qiagen). The protocol followed for the purification is mentioned in the following text.

2.2.3. PCR amplicon purification by Gel Elution Protocol (using gel extraction kit Qiagen)

- 1) The PCR product was run on agarose gel in fresh 1X TAE buffer, and after clear resolution of bands the DNA fragment of interest was excised with a clean scalpel.

Table 2F: Characteristics of Agarose Low Electroendosmosis (Low EEO) (DNase, RNase Free)

| Feature | Characteristics |
|-------------------------|------------------------------|
| Gel strength (1.5% gel) | 1500-3000 g cm ⁻¹ |
| Gelling Temperature | 36 °C |
| Melting Temperature | 86-88 °C |

The agarose (characteristic of agarose mentioned in **Table 2F**) gels were made and electrophoresis was done in 1X TAE buffer. Agarose powder was suspended in 1X TAE buffer (concentration of the agarose gel was decided on basis of size of fragment to be examined, mentioned in **Table 2G**), and was dissolved by heating in a microwave oven (for uniform heating). The melted agarose gel was cooled down to approximately 50 °C, and just before pouring in the casting tray, ethidium bromide (EtBr) (0.5 µg ml⁻¹ in TE buffer, pH 8.0) was added in the gel solution. After adding EtBr, the gel was poured in the electrophoresis boat. Wells were made by fixing the comb over the boat, prior to pouring of the molten agarose. After solidification the comb was removed and the gel was transferred to electrophoresis tank containing 1XTAE buffer. The DNA samples and molecular markers were mixed with tracking dye 1X (1/6 volume of DNA sample). Electrophoresis was carried out at 80-100V. The gel was then examined under transmitted UV light (302-312 nm) using a gel documentation system (Biorad). Gels of different strengths were prepared depending on the size of DNA to be analyzed. A table which illustrates appropriate gel strength for different sizes of DNA is given below (**Table 2G**):

Table 2G: The concentration of agarose gel suitable for different DNA sized fragments to obtain good resolution

| Gel strength (w/v) | Size of DNA (bp) |
|--------------------|------------------|
| 4% | 50-120 |
| 3% | 120-120 |
| 2% | 300-800 |
| 1.5% | 800-1000 |
| 1.2% | 1000-1200 |
| 1% | 1200-2000 |
| 0.8% | 2000 longer |

The amplicons obtained were purified by Gel Extraction kit (Qiagen). The protocol followed for the purification is mentioned in the following text.

2.2.3. PCR amplicon purification by Gel Elution Protocol (using gel extraction kit Qiagen)

- 1) The PCR product was run on agarose gel in fresh 1X TAE buffer, and after clear resolution of bands the DNA fragment of interest was excised with a clean scalpel.

- 2) The gel was weighed in pre-weighed eppendorf tube and 3 volumes of buffer QG was added to 1 volume of gel (100 mg = 100 μ L x 3).
- 3) The samples were incubated at 50 °C (water bath) until a clear solution was obtained. One volume of isopropanol was added to the sample and mixed properly to cause DNA precipitation.
- 4) The samples were kept at room temperature for 5 min.
- 5) A volume of 750 μ L of sample was applied to the QIA quick spin column and placed in a 2 mL collection tube and centrifuged at 13000 rpm for 1min at room temperature.
- 6) The flow through was discarded and the column was placed back in the same collection tube. The previous step was repeated until the complete sample volume was consumed.
- 7) A volume of 500 μ L of QG buffer was added to QIA quick spin columns and centrifuged at 13000 rpm for 1min at room temperature.
- 8) Flow through was discarded and 750 μ L of PE buffer was added to the column.
- 9) The QIA quick spin column was incubated at room temperature for 5 min and centrifuged at 13000 rpm for 1min.
- 10) The flow through was discarded and an empty spin was given at 13000 rpm for 1min to remove any remaining liquid in the column.
- 11) The spin columns were transferred to fresh 1.5 mL micro-centrifuge tubes and 20 μ L (depending upon the initial DNA concentration) of nucleases free Milli-Q water was added to centre of the spin column. Tubes were then incubated at room temperature for 5-10 min.
- 12) After incubation, tubes were centrifuged at 13,000 rpm for 1.5 min at room temperature. The samples were checked on agarose gel, quantified and stored at -20°C.

The amplified and purified products hence obtained were restriction digested and cloned into *pGEM*[®]-T vector of 3 kb pair in the multiple cloning site region of the plasmid.

2.2.4. Restriction of purified PCR product and vector

Restriction enzymes bind specifically to and cleave double-stranded DNA at specific sites within, or adjacent to, a unique sequence called the recognition sequence. The restriction enzymes can be classified into Type I, Type II and Type III. The Type I and Type III carry out both modification and ATP-dependent restriction activities. Type I and Type III, both bind to specific site, but Type III cleaves at specific site and then gets released from substrate and Type I causes cleavage at random sites on the bound DNA. Hence, Type I and Type III are not used in molecular cloning. Type II restriction/modification enzymes are binary systems consisting of a restriction endonuclease that cleaves at specific site and

a separate methylase that carries out the modification of the same site. A large number of Type II restriction endonucleases have been isolated, characterized and are being widely used in procedures of molecular cloning. The restriction sites for Type II enzymes are mostly, 4, 5 or 6 nucleotides in length and are palindromic sequences. The location of the cleavage site within the recognition site differs from enzyme to enzyme. Some enzymes cause cleavage at the same axis of symmetry in both strands generating blunt end DNA, while some enzymes cleave the DNA at the same location of the recognition sequence on opposite sides of axis of symmetry of both the strands giving cohesive ends of DNA.

The restriction sites used for present molecular cloning were selected such that there is occurrence of single enzyme sites in the vector i.e. in the multiple cloning site and there is no restriction site for these enzymes in the DNA insert. The *in-silico* restriction analysis of the DNA of interest was done using NEB cutter (freely available tool online; Vincze et al., 2003), and the restriction enzymes were chosen accordingly to design primers for PCR. The amplicons and the vector of interest were then digested with chosen enzymes to generate compatible ends. The restriction enzymes were procured from New England Biolabs (NEB). The enzymes used were *NdeI*, *BamHI* and *XhoI*. The detail of each of the enzymes is mentioned in **Table 2H**.

Table 2H: The detail of the restriction enzyme used for cloning procedures of various DNA fragments

| S.No | Name of Enzyme | Source (<i>Escherichia coli</i> strain that contains cloned gene) | Restriction Site* |
|------|-------------------|---|-------------------|
| 1 | <i>NdeI</i> | <i>NdeI</i> gene from <i>Neisseria denitrificans</i> | CATATG GTATAC |
| 2 | <i>BamHI-HF</i> * | Modified <i>BamHI</i> gene (E163A/E167T) from <i>Bacillus amyloliquifaciens H</i> | GGATCC CCTAGG |
| 3 | <i>XhoI</i> | <i>XhoI</i> gene from <i>Xanthomonas holcicola</i> | CTCGAG GAGCTC |

*The restriction enzyme cuts between the highlighted nucleotides.

One unit of the restriction enzyme is defined as the amount of enzyme required to digest 1 µg of λ DNA in 1 hour at 37°C in a total reaction volume of 50 µL. The enzymes were available in concentration of 20U/µL. A 20-fold over digestion with these enzymes is claimed to provide > 95% DNA fragments with required termini and hence an efficient raw material for ligation. Hence, for restriction digestion, 1 µg of DNA was used in presence of 1X buffer (compatible buffer according to the enzyme used) and 1 µL (20U) of enzyme in a 50 µL reaction. Fortunately, the enzymes used for digestion were 100% efficient in buffer compatible for all three enzymes; hence double digestions could be set up. The digestion reaction set up in general is as mentioned in **Table 2I**.

Table 21: The composition of reaction mixture used for a standard restriction reaction

| Component | Amount |
|--------------------------------|-------------------------------|
| DNA | 1 μg |
| Buffer | 5 μl . (10X to 1X) |
| Enzyme 1 (<i>NdeI</i>) | 1 μl . |
| Enzyme 2 (<i>XhoI/BamHI</i>) | 1 μl . |
| Water | Upto 50 μl . |
| Incubate at 37 °C for 2-3 hrs | |

The digested PCR products and vector were purified by gel extraction procedure, before proceeding for ligation reactions. The PCR product and vector after digestion and purification were ligated. Alternatively, TA cloning can be followed as restriction digestion can at times be time consuming and cumbersome.

2.2.5. Ligation of PCR product into a TA vector

The pGEM⁺-T vector (Promega) is a linearized vector with a single 3' – terminal thymidine at both ends. The T overhangs of pGEM-T readily ligate with A overhangs of insert generated by polymerases lacking 3' to 5' exonuclease activity, in the absence of recircularization of the vector. Further, pGEM-T is a high copy number plasmid containing T7 and SP6 RNA polymerase promoters on either side of the multiple cloning sites, within the α -peptide coding region of β -galactosidase. This arrangement allows identification of positive recombinants by blue/white screening. A few important considerations must be kept in mind for successful T-vector cloning. Firstly, the introduction of nucleases (common source being water, hence nuclease free water should be used for the reactions) into the ligation mixture can cause degradation of T-overhangs hence it should be strictly avoided. Secondly, the ligation between fragments with single base overhangs can be inefficient; hence competent cells with transformation efficiency $\geq 1 \times 10^8$ cfu/ μg DNA should be used. Thirdly, as a general consideration, the exposure of PCR product to short wave UV light should be minimized to prevent formation of pyrimidine dimers. Fourthly, the clean-up of the PCR product is recommended before ligation so as to remove any primer-dimers or other undesired reaction products. The advantage of using TA vector before cloning into expression vector is that there is no need of restriction enzymes making the procedure much simpler and faster. At the same time the major drawback of TA cloning is that directional cloning is not possible so the gene has 50% chance of getting cloned in the reverse direction. Since, it is only a sub-cloning and the insert will finally be digested and ligated into expression vector, the orientation of the gene is not an issue. This additional cloning into TA vector makes restriction digestion 100% sure as the double digested (with enzymes for which restriction sites have been introduced in the primers) pGEMT⁺ insert clone gives the insert, which is eluted and used for further cloning.

The ligation reaction involves *pGEM-T* vector (50-100ng), 2X ligation buffer (working concentration is 1X), PCR product with A overhangs (concentration depending upon the molar ratio between vector and insert) and T4 DNA ligase (3 Weiss U/ μ L). The formula for calculation of required concentration of insert depending upon the molar ratio is as follows:

$$\frac{\text{ng of vector} \times \text{Kbp size of insert} \times \text{insert:vector molar ratio}}{\text{Kbp size of vector}^*} = \text{ng of insert}$$

* Vector size is 3000bp

Ratios from 3:1 to 1:3 between insert and vector provide good results in most reactions, to start with but ratios like 1:1, 6:1 or 1:6 were used. The biomath calculator (www.promega.com/biomath) can also be used for calculation of amount of insert required to set up a ligation reaction. The procedure followed for T-vector cloning is as follows (Table 2J):

- 1) The buffer, vector, control DNA and DNA of interest were kept on ice to thaw, before setting up the reaction. Once thawed, the contents were spinned shortly to bring them to the bottom of the tube. Vortexed the 2X buffer, vigorously, before use.
- 2) Added all the contents sequentially as mentioned below into a 0.5 mL tube with low DNA binding capacity. Mixed the contents by pipetting.
- 3) A positive control with T-vector and control DNA (with A overhangs) is set up to check the proficiency of the enzyme and vector. Further, a background control is set up to check self ligation of the vector. After mixing, incubated the reaction for 1 hour at room temperature, or alternatively, at 4°C for overnight to obtain maximum number of transformants.
- 4) The complete ligation mixture is then transformed into high-efficiency competent cells. After transformation the screening is done initially by the colour of the colony on an IPTG+X-gal plate, followed by plasmid isolation and restriction digestion of the plasmid from white colonies.

Table 2J: The composition of ligation reaction for TA cloning

| Reaction component | Test Reaction | Positive control | Background control |
|---|---------------|------------------|--------------------|
| 2X rapid Ligation Buffer, T4 DNA Ligase | 5 μ L | 5 μ L | 5 μ L |
| pGEM-T Vector (50ng) | 1 μ L | 1 μ L | 1 μ L |
| PCR product | X μ L* | - | - |
| Control Insert DNA | - | 2 μ L | - |
| T4 DNA Ligase (3 Weiss U/ μ L) | 1 μ L | 1 μ L | 1 μ L |
| Nuclease free water to final volume | 10 μ L | 10 μ L | 10 μ L |
| *Molar Ratio of insert:vector may need optimization | | | |

- 7) Once the clone is confirmed by restriction analysis, insert is digested out with enzymes for which the sites are present in the primers, so as to generate cohesive ends. Different enzyme sites are used in forward and reverse primers so as to cause in-frame directional insertion of the insert in the vector of interest (*pET15b/pCOLIDH*).
- 8) The double digested insert is then eluted using Qiagen gel extraction kit and used for further cloning.

Note: *Taq* DNA polymerase gives 3' A overhangs as it lacks 3' to 5' exonuclease activity. *Taq* DNA polymerase was used for amplification of all amplicons except *ccmM* and *ccm operon* for which Phusion polymerase was used. Phusion polymerase does not give 3' A overhangs, therefore the amplicons generated need to be processed by a procedure called extension so as to obtain 3' A overhangs.

Procedure for A tailing of blunt PCR product

- 1) Took 1-7 μL of purified PCR fragment generated by a proofreading polymerase (Phusion Polymerase).
- 2) Added 1 μL *Taq* DNA polymerase 10X reaction buffer containing MgCl_2 , followed by dATP to a final concentration of 0.2mM. To the reaction mix added 5U of *Taq* DNA polymerase and then nuclease free water to make the final reaction volume upto 10 μL .
- 3) Mixed the contents by pipetting and then spin down all the contents to the bottom of the tube. Incubated the reaction mix at 72°C for 30 minutes to 1 hour 30 minutes. 1-2 μL of the reaction mix was then used in ligation with *pGEM⁺-T* vector.

2.2.6. Cloning into expression vector

The pET system is among the most powerful systems developed till date for cloning and expression of recombinant proteins in *E. coli*. Target genes are cloned in pET plasmids under control of strong bacteriophage T7 transcription signals. T7 RNA polymerase is so selective and active that, when fully induced, almost all of the cell's resources are converted to target gene expression. Further, this system can be attenuated by lowering the concentration of the inducer, which can also result in increase in solubility of some proteins. Target genes are first cloned using hosts that do not contain T7 RNA polymerase gene, thus preventing formation of potentially toxic proteins in the cell and hence maintaining plasmid stability. Once established in a non-expression host, target protein expression may be initiated either by infecting the host with λCE6 , a phage that carries T7 RNA polymerase gene under control of λ p_i and p_t promoters, or by transforming the plasmid into expression host containing the chromosomal copy of the T7 RNA polymerase

gene under control of *lacUV5* which can be induced by IPTG (*E. coli* BL21(DE3)pLysS was used for the present study). The cold shock vector *pCOLDI* is designed to perform efficient protein expression utilizing the promoter derived from *cspA* gene, a cold shock gene. At the downstream of the *cspA* promoter is inserted a *lac* operator for strictly controlled expression. As the promoter used in the vector is derived from *E. coli*, most *E. coli* hosts can be used for expression and purification of the protein of interest. The

The important characteristics of the vectors used in the study are mentioned in **Table 2K** and shown in **Figure 2A**.

Table 2K: The important characteristics of the vectors used, in context to the expression and purification studies

| S.No | Feature | <i>pET15b</i> | <i>pCOLDIH</i> |
|------|-----------------------|---------------|----------------------|
| 1 | Size | 5708bp | 4392bp |
| 2 | Origin of replication | pBR322 | ColE1 |
| 3 | Antibiotic resistance | Ampicillin | Ampicillin |
| 4 | Promoter | l7 promoter | <i>cspA</i> promoter |

2.2.7. Alkaline lysis method of isolation of plasmid

Reagents used for plasmid isolation

Solution I GET Buffer (50mM Glucose; 25mM TrisHCl (pH 8.0); 10mM EDTA (pH 8.0); Solution II Lysis Buffer (NaOH 0.2N; 10% SDS (Sodium dodecyl sulphate)); Solution III (Sodium acetate 3M (pH 5.2); Tris saturated phenol (Merck); Chloroform and Isoamyl alcohol; 70% Ethanol (Merck); Isopropanol; Nuclease free Milli-Q water and/ or TE buffer pH 8.0.

Protocol for mini preparation of plasmid isolation described by Birnboim and Doly (1979) in Sambrook et al. (1989)

- 1) Single bacterial colony (*E. coli* DH5a cells transformed with plasmid of interest) was inoculated in 5 ml. Luria Bertani broth (LB) medium antibiotic (ampicillin for *pGEMT*, *pET15b* and *pCOLDIH*) and incubated at 37°C for 12-16 hours with continuous shaking at 250 rpm in a shaker incubator. A volume of 1.5 ml. of the overnight grown culture was transferred to a micro-centrifuge tube and the cells were centrifuged at 6000 rpm for 5 min at room temperature (Eppendorf cooling centrifuge). The supernatant was discarded and 1.5 ml. culture was added to the same tube, it was centrifuged at 6000 rpm for 5 minutes.
- 2) The pellet was suspended in 100 µl. of ice cold solution I (GET buffer - 0.05M glucose, 0.025 M Tris HCl pH 8.0 and 0.01 M EDTA, pH 8.0) by vortexing.

- 3) To the resuspended cells, 200 μ l. of freshly prepared solution II (alkaline SDS - 1% SDS and 0.2N NaOH) was added and the cells were lysed by gently inverting the tubes 2-5 times.
- 4) The sample was incubated at room temperature for 5 min. This is a critical step, incubation at this step should not exceed 5 minutes.
- 5) After 5 min incubation, 150 μ l. of chilled solution III (3M sodium acetate pH 5.2) was added and samples were gently mixed by inverting the tubes 2-5 times. The solution III causes renaturation of the plasmid DNA and the genomic DNA being larger in size is incapable of denaturation and hence settles with the debris upon centrifugation. The sample was again incubated on ice (since it involves an exothermic reaction) for 15-30 min. The lysate was centrifuged at 10,000 rpm for 20 min at 4°C. The clear transparent supernatant was transferred into fresh micro-centrifuge tube. DNase free RNaseA (20 μ g/ml) was added and the sample was incubated at 37°C for 1 hour. RNase solution is heated at 100°C before use, to inactivate any DNase (if present). RNase has ability to renature after undergoing boiling while DNase denatures.
- 6) Plasmid Purification:-Equal volume of phenol: chloroform: isoamyl alcohol (25:24:1 v/v) was added to the supernatant obtained at previous step. The solutions were mixed by inverting the tubes to get a homogenous mixture. The sample was centrifuged at 10,000 rpm for 10 min.
- 7) After centrifugation, the upper aqueous phase was removed carefully and this layer was transferred to fresh 1.5 ml. micro-centrifuge tube. Equal volume of chloroform:isoamyl alcohol (24:1) was added and the tubes were mixed by inverting. The sample was centrifuged at 10,000 rpm for 10 min.
- 8) The upper aqueous phase was removed carefully and this layer was transferred to fresh 1.5 ml. micro-centrifuge tube.
- 9) DNA precipitation was done after plasmid purification.
 - i) The upper aqueous phase obtained after chloroform: isoamyl alcohol treatment was taken and equal volume of absolute isopropanol was added.
 - ii) The samples were kept at -20°C for 15-30 min.
 - iii) The samples were centrifuged at 10,000 rpm for 20 min at 4°C.
 - iv) The pellet was washed with 70% ethanol by centrifuging at 10,000 rpm for 10 min at 4°C.

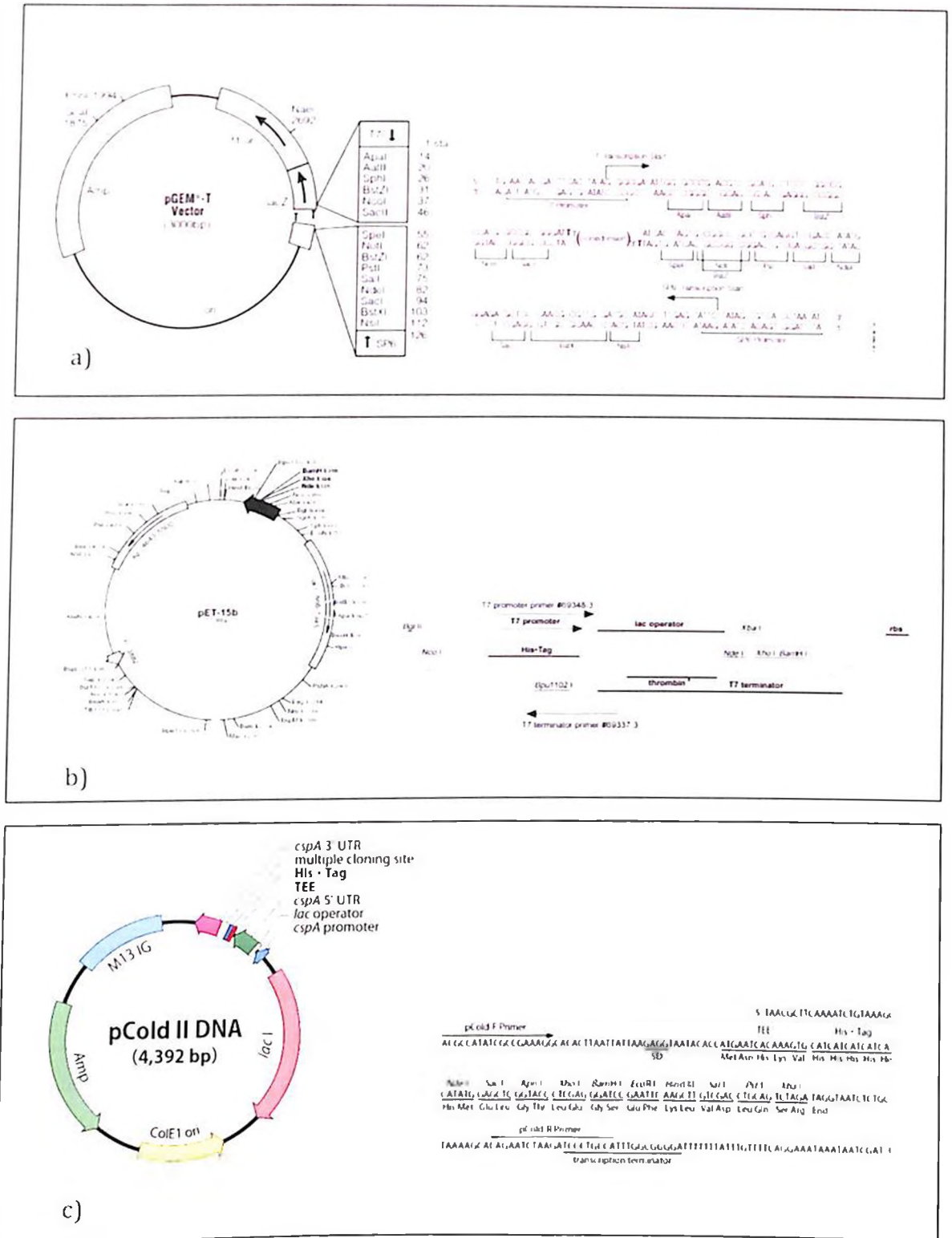


Figure 2A: Vector systems used for cloning and expression of proteins under study. a) *pGEM⁺-T* vector (3.0Kbp); b) *pET15b* vector (5.7Kbp); c) *pColdII* vector (4.3Kbp)

v) The precipitated nucleic acid was dried and dissolved in 20 μ l. of TE buffer (10mM pH 8.0) or nuclease free water. The samples were stored at -20 °C. The plasmid DNA was checked on 0.8% agarose gel.

Restriction analysis of amplified PCR products and plasmid DNA

Endonuclease digestion of amplified PCR products and the plasmid was done by *NdeI* and *BamHI*/*XhoI* restriction enzymes with their respective buffers. A 50 μ l. reaction volume was set up. The digestion mixture was incubated at 37°C for 3 hr and checked on 1% agarose gel. To obtain pure vector and the amplified PCR products for cloning purpose, gel elution of the vector and amplified PCR products were done using QIAGEN Gel Extraction kit.

Ligation of digested purified PCR products in the cloning vector

Double digested and purified PCR amplicons and vector was quantified, and the inserts were cloned in MCS of *pET15b/pCOLDII* vector between *NdeI* and *BamHI*/*XhoI* enzyme site. Ligation was set on the basis of the molar ratio of insert to vector according to principles described before and transformed into host cells.

2.2.8. Transformation of plasmid/ligation product into host cells

The introduction of plasmid or recombinant plasmid into a compatible and competent host cell is called transformation. The recombinant plasmids generated by cloning techniques were introduced into *E. coli* cells for protein expression and purification. The procedures used for the same are mentioned in forthcoming text.

Artificial transformation of *E. coli* with plasmid DNA in the presence of monovalent ions is a commonly used method in genetic engineering. The measurements of both steady state and time-resolved anisotropies of fluorescent dye trimethyl ammonium diphenyl hexatriene (TMA-DPH), bound to cellular outer membrane revealed heat shock of 0°C to 42°C causes decrease in outer membrane fluidity of cells considerably, because of release of lipids to external medium. A subsequent cold shock of 42°C to 0°C causes increase in fluidity to its original state. Further, the heat-cold pulse causes depolarization of the inner cell membrane. In a nutshell, the heat-cold shock facilitates transformation of DNA into *E. coli* cells by two ways, firstly, the heat shock creates transient pores on the outer cell membrane due to release of lipids hence breaking the first barrier and secondly, lowering of membrane potential helps DNA cross the inner cell membrane as well. To achieve high transformation efficiencies the host cells need to be processed in order to get competent cells. There are various methods for preparation of competent cells, which include, Hanahan method, Inoue method, calcium chloride method and electroporation. Hanahan method gives competent cells with transformation efficiencies as high as 10^8 cfu/ μ g of

supercoiled DNA transformed, if the protocol is followed strictly, using high quality fresh reagents, ultra clean glassware plasticware and cultures grown from original frozen stock. Inoue method yields equally efficient competent cells but is less finicky and more reproducible, using a fairly standard procedure, except that the cells are grown at 18°C rather than 37°C, which gives the cell membrane certain characteristics making it favorable to take up foreign DNA more efficiently. Further, the calcium chloride method published by Cohen et al (1972) is simple and rapid variation of the technique, yielding efficiencies of 10^6 to 10^7 cfu μ g of supercoiled DNA transformed. Competent cells prepared using this method may be preserved at -70°C, although prolonged storage may result into deterioration in efficiency. In the present study calcium chloride and Inoue methods were used.

2.2.8.1. Preparation of *E. coli* DH5 α and *E. coli* BL21(DE3)pLysS competent cells

Requirements

100mM CaCl₂ (50 mL) - autoclaved and pre-cooled to 4°C; 60% glycerol solution- autoclaved pre-cooled to 4°C; 5 mL LB broth for 1^o culture - autoclaved (no antibiotic for *E. coli* DH5 α ; chloramphenicol for *E. coli* BL21(DE3)pLysS); 50 mL LB broth for 2^o culture - autoclaved (no antibiotic for *E. coli* DH5 α ; chloramphenicol for *E. coli* BL21(DE3)pLysS); Oakridge tubes- autoclaved and stored at 4°C before use; Eppendorf tubes 1.5 & 2 mL – autoclaved and stored at 4°C; Autoclaved 1 mL and 200 μ L tips; Autoclaved 1 mL and 200 μ L cut tips; LB agar + antibiotic (depending upon the resistance gene in the plasmid to be transformed)

Procedure

- 1) Preparation of primary (1^o) culture: 5 mL of LB was inoculated with a single colony of *E. coli* DH5 α culture and incubated at 37°C for 14-16 h. Preparation of secondary culture:
- 2) After 14-16 h of 1^o culture, 1% inoculum from 1^o culture was taken, i.e. 250 μ L for 25 mL medium, and secondary culture was inoculated and incubated at 37°C for 2-3 h, until the OD at 600 nm reached 0.3-0.4.
- 3) It was centrifuged at 6000 rpm at 4°C for 8 min. The supernatant was discarded without disturbing the pellet (allow the tubes to stand in inverted position on paper towels to drain away any traces of liquid media) and ice cold 100mM CaCl₂ (30 mL for 50 mL culture) was added to each tube.
- 4) The pellet was dislodged completely and carefully using cut tips with repeated pipetting. During the process of resuspension, do not allow the cells to stay out of ice for more than 2-3 minutes. The purpose of gentle resuspension is to allow each of the cell's surfaces to

come in contact with the monovalent ions in solution (which in turn will be responsible for interaction with the DNA to be transformed), without causing any damage to the cells due to shearing.

- 5) Once the resuspension was done properly, it was centrifuged at 8000 rpm, 4°C for 8 min.
- 6) The supernatant was discarded very gently and the pellet was resuspended in 100mM CaCl₂ (2 ml. CaCl₂ for 50 ml. culture). To the cell suspension 0.9 ml. of 60% glycerol was added so as to make the final concentration 15%. It was mixed well by tapping. Mini cooler was used to keep the eppendorfs cool.
- 7) Finally, 0.3 ml. aliquots were transferred in sterile eppendorf tubes, labelled and frozen in liquid nitrogen (for snap freezing) and stored at -80°C.

2.2.8.2 Transformation of recombinant DNA into *E. coli* DH5α competent cells

- 1) The competent cells *E. coli* (DH5α) were thawed on ice (for approximately 20-30 minutes). The ligation mixture plasmid DNA was added to 100 μl. of competent cells, mixed by gentle tapping and incubated on ice for 30 min.
- 2) A heat shock was given at 42°C (water bath) for 90s and the tubes were incubated on ice for 2-3 min. (this critical step in transformation, hence should be done carefully and accurately)
- 3) A volume of 800 μl. of LB medium was added to each of the tubes, mixed well and incubated at 37°C incubator for 45 min with shaking at 150 rpm.
- 4) The cells were pelleted by spinning at 3000 rpm for 5 min. From each of the tubes, 750 μl. of LB media was removed from the supernatant without disturbing the pellet.
- 5) The bacterial pellet was resuspended in 150 μl. remaining supernatant and was spread on LB agar plate with antibiotic (depending on resistance in plasmid transformed). The plates were incubated overnight (12-14 hrs) at 37°C. The bacterial colonies were checked for recombinant plasmid by isolation of plasmid and restriction analysis.
- 6) In every transformation experiment, setting up of controls is very important, in order to measure transformation efficiency and to eliminate possibility of contamination. The controls set up in transformation experiment include:
 - An aliquot of competent cells to which no DNA is added should be carried through the transformation protocol. This aliquot should be plated on agar plate containing antibiotic used to select the transformants. There should not be any growth on this plate. In case there is growth, the possible reasons could be contamination of competent cells with the

bacteria containing the antibiotic resistance, the antibiotic used is degraded or the plates used were contaminated with antibiotic resistant bacteria.

- An aliquot of cells to which no DNA is added and is carried through the transformation procedure, should be plated on agar plate without antibiotic in order to check the survival rate of the competent cells. There should be a lawn of cells on this plate if the competent cells are prepared properly and are stored under ideal conditions.
- An aliquot of competent cells should be transformed with a known concentration of plasmid DNA, so as to check the efficiency of the competent cells and the transformation procedure. When a ligation reaction is used as source of transforming DNA, the transformation efficiency is reduced by at least two orders of magnitude compared with supercoiled plasmid DNA (linear DNA does not result in successful transformants). The actual number of transformants obtained per μg of ligated DNA will depend on the amount of recombinant plasmids generated during the ligation reaction, which in turn will depend upon the concentration and purity of the vector and insert used.

Colonies were observed and master plate of these colonies was prepared and plasmid isolation was done by alkaline lysis method. After plasmid isolation, positive clones were analysed by digestion and finally, recombinant plasmids were confirmed by sequencing. After confirmation of the recombinant, the recombinant vector was transformed into expression host.

Bacterial Hosts

E. coli DH5 α has mutations which include *delta lacZ Delta M15 DeltalacZYA-argF' U169 recA1 endA1 hsdR17(rK-mK⁺) supE44 thi-1 gyrA96 relA*. The *lacZ Delta M15* mutation allows for blue-white screening for recombinant cells. The *endA1* mutation permits for lower endonuclease degradation and hence a higher transformation efficiency. The *recA1* mutation reduces homologous recombination and hence provides a more stable insert. It is a recombination deficient amber suppressing strain used for plating and growth of plasmids. The *lacZ delta M15* mutation is responsible for α complementation with the amino terminus of β -galactosidase encoded in blue-white screening vectors.

E. coli BL21pLysS is a strain used for high level of expression of genes cloned into vectors containing bacteriophage T7 promoter. The T7 RNA polymerase gene of λ DE3 bacteriophage, integrated in the bacterial genome under control of *lacUV5* promoter. The strain has T7 lysozyme gene in the pLysS plasmid with chloramphenicol resistance. T7 lysozyme suppresses the activity of T7 RNA polymerase promoter, reducing the basal level expression until induced by isopropyl-1-thio- β -D-galactopyranoside (IPTG). The

advantage of T7 lysozyme is that it increases the tolerance of the cell against toxicity. The absence of lon protease and ompT membrane protease remove the ability of the cell to degrade expressed proteins. The genotype of the strain is: $F^+ ompT hsdS(r_B^- m_B^-) gal dcm \lambda(DE3) pLysS (Cam^r) (\lambda(DE3): lacI, lacUV5-T7 \text{ gene } 1, ind1, sam7, nin5)$.

E. coli DH5 α is used for plating and growth of plasmids and *E. coli* BL21(DE3)pLysS is used for expression and purification of expressed proteins.

The positive clones obtained were validated by restriction digestion and sequencing analysis. The samples were outsourced to sequencing facility of Delhi University south Campus.

2.2.9. The assessment of expression of constructs generated

The choice of system for expression of protein in *E. coli* is affected by various factors. These include, the size of the protein to be expressed (proteins with <100 residues are easy to express in soluble form), the amount of protein needed and form of protein required (whether the native state of the protein is required or not). The proteins to be purified in the present study are required in native form and in large quantities as the downstream application of the purified proteins is evaluation of the activity of the proteins. Therefore, *E. coli* BL21(DE3)pLysS strain and vectors with IPTG inducible T7 promoter were used for expression. The vectors used possess His-tag and hence the proteins expressed can be purified by affinity chromatography. The protocol followed for assessment of expression of various constructs is mentioned as follows:

- 1) Primary inoculation: Inoculated 5 mL LB broth (plus suitable antibiotics) with the *E. coli* BL21(DE3)pLysS strain carrying the construct with single colony from freshly streaked LB agar plate (with appropriate antibiotics) from glycerol stock and incubate at 37°C, 150rpm for overnight (12-14hrs).
- 2) Secondary inoculation: Inoculated 20 mL LB broth (with appropriate antibiotics) with the overnight grown primary culture at rate of 1%, and incubate at 37°C and 150rpm, till the optical density (at 600nm) of the culture reaches 0.4-0.5. Once the desired OD₆₀₀ is reached, 1 mL of the culture is drawn into a fresh 1.5 mL tube. To the rest of the culture, IPTG is added. The amount of IPTG to be added varies from protein to protein. The maximum recommended concentration is 1mM, while one needs to standardize from a range for optimum IPTG concentration required, which varies from 0.01-1mM. In case of some proteins higher IPTG concentration can be lethal for the cell and some get enclosed into insoluble inclusion bodies. Hence, in such cases, lower inducer concentrations should be tried. Further, in some cases the T7 promoter induced proteins get expressed but get

enclosed into inclusion bodies due to very high expression, in such cases, alternative vector systems like cold shock inducible vectors can be used.

- 3) As soon as IPTG (suitable concentration depending upon standardization) is added, 1 mL of culture is drawn into fresh 1.5 mL tube, and the culture is incubated at 37°C, 150 rpm. Some proteins do not get expressed or due to overexpression turn out to be toxic to the cell, hence to make conditions suitable for the cell, so that it is capable of expressing the protein to an optimal level and also maintain normal cell metabolism, the culture after induction is grown at lower temperatures. Hence, the temperature optimal for protein under study needs to be standardized.
- 4) After every one hour, 1 mL sample is drawn from the culture and transferred to a fresh tube, upto 8-9 hrs of incubation. The cultures grown at lower temperatures should be assessed for expression for longer period of time (about 18 hrs) to identify the time required for maximum expression.
- 5) All the samples collected must be stored at 4°C, till used for further processing. Once all samples are collected, the tubes are centrifuged at 13,000 rpm, 1 minute at room temperature. The supernatant is discarded and any residual liquid is removed by tapping on paper towel. To the pellet obtained 100 µL 1X SDS-gel loading buffer is added and the suspension is kept at 100°C for 10 minutes. This treatment causes the cells lysis and the proteins released are denatured to their primary structure. The samples are then centrifuged at 13,000 rpm for 1 minute at room temperature, in order to cause debris to settle to the bottom.
- 6) Loaded 40µg of each suspension on 10% SDS-PAGE. All the RuBisCO protein of *G. violaceus* (Chapter 4) and *Rhodospirillum rubrum* (Chapter 6) and CCM proteins of *G. violaceus* (Chapter 5) were analyzed for expression by above mentioned method.

2.2.10. Solubility of the expressed proteins

- 1) Primary inoculation: Inoculated 5 mL LB broth (plus suitable antibiotics) with the *E. coli* BL21(DE3)pLysS strain carrying the construct with single colony from freshly streaked LB agar plate (with appropriate antibiotics) from glycerol stock and incubate at 37°C, 150rpm for overnight (12-14hrs).
- 2) Secondary inoculation: Inoculated 20 mL LB broth (with appropriate antibiotics) with the overnight grown primary culture at rate of 1%, and incubate at 37°C and 150rpm, till the optical density (at 600nm) of the culture reaches 0.4-0.5. Once the OD reaches 0.4-0.5, IPTG is added to the culture. The amount of IPTG to be added varies from protein to protein. The maximum recommended concentration is 1mM, while one needs to

- standardize from a range for optimum IPTG concentration required, which varies from 0.01-1mM.
- 3) As soon as IPTG (suitable concentration depending upon standardization) is added, the culture is incubated at 37°C, 150 rpm (or temperature standardized for the protein), for suitable time (as standardized for maximal expression).
 - 4) The cells are pelleted at 8000 rpm, 20 minutes at 4°C. The supernatant is discarded and any remaining liquid is drained by inverting the tubes on paper towels. The each gram of pellet, 4 µl. of 100mM PMSF and protease inhibitor (working concentration 1X) is added. The pellet is then resuspended in buffer 1 (5mM imidazole; 300mM sodium chloride; 250mM sodium phosphate, pH 8.0), by adding 3 ml. for each gram of pellet.
 - 5) Added 80 µl. of 10mg/ml. Lysozyme to the cell suspension and mixed by gentle vortexing and avoiding frothing.
 - 6) Incubated at 30°C for 1hr with continuous shaking. After Lysozyme treatment the sample was put on ice for 30 minutes to allow cooling down.
 - 7) The sample was then sonicated at 10KHz by giving 10 short bursts of 10 seconds with 30 seconds rest time. (While sonicating keep the probe properly submerged, increase sample volume if needed but not more than 10 mL; Give short pulses; While sonicating keep the sample resting on ice to avoid overheating)
 - 8) Optional: If lysate still viscous add DNase at rate of 1mg/ml. per gram of pellet and incubate at 4°C for 30 minutes in presence of 10mM MgCl₂.
 - 9) After sonication the sample was centrifuged at 7850rpm, 20 minutes at 4°C. The supernatant was collected in fresh tube, labeled as fraction A and stored at 4°C.
 - 10) The pellet obtained in step 9 was washed in buffer 2 (5mM imidazole; 300mM sodium chloride; 250mM sodium phosphate, pH 8.0; triton-X100 0.5%) and centrifuged at 7850rpm, 20 minutes at 4°C. The supernatant was collected in fresh tube, labeled as fraction B and stored at 4°C.
 - 11) Resuspended the pellet obtained in step 10 in buffer 3 (5mM imidazole; 300mM sodium chloride; 250mM sodium phosphate, pH 8.0; Urea 8M; PMSF), sonicated as mentioned in step 7 and centrifuged at 7850rpm, 20 minutes at 4°C. The supernatant was collected in fresh tube, labeled as fraction C and stored at 4°C.
 - 12) Analyzed the fractions A, B and C on SDS-PAGE (12% resolving gel and 5% stacking gel). *R. rubrum* RuBisCO (Chapter 6), *G. violaceus* RbcL, *G. violaceus* RbcS, *G.*

violaceus RbcX and *G. violaceus* RuBisCO operon (Chapter 4) containing recombinant plasmids when expressed in *E. coli* BL21(DE3)pLysS were analyzed for solubility.

In many cases the expressed protein is insoluble and accumulates in inclusion bodies, especially under conditions of high level expression. The solubility of such proteins can be increased by several strategies. Firstly, reducing the rate of protein synthesis, which in turn can be done by lowering growth temperature, using weaker promoter, using low copy number plasmid and by lowering inducer concentration, can cause increased solubility of the expressed protein. Further, changes in the growth medium like, addition of prosthetic groups or cofactors essential for protein stability and folding, addition of buffer to the medium, addition of 1% glucose to suppress induction of lac promoter by lactose, addition of polyols (like sorbitol), sucrose cause an increase in osmotic pressure and hence an accumulation of osmoprotectants (which stabilize protein structure), can be beneficial for obtaining expressed protein in native form. Moreover, coexpression with chaperones and foldases also enhance protein solubility. These methods alone or in combination need to be standardized for desirable results. In cases, where none of the above mentioned result into a soluble form of the expressed protein, then the denaturing protocol of protein purification is used followed by renaturing of the protein by dialysis.

2.2.11. Purification of the proteins by Immobilized Metal Affinity Chromatography (IMAC)

IMAC was introduced in 1975 (Porath et al 1975), as technique for protein purification by interaction of the protein residues with immobilized metal ions.

Most commonly, the amino acids involved in interactions with the metal ions include histidine, cysteine and tryptophan. In the present study BD TALON IMAC resins were used for protein purification. These are cobalt-based resins for purification of polyhistidine-tagged proteins. The advantage of using these resins is their durability, compatibility with a wide range of reagents, functionality under both native and denaturing conditions and can be used in a variety of formats viz. small scale, large scale, gravity flow and spin columns. As an advantage over other metal affinity resins, BD TALON utilizes a tetradentate metal chelator, which tightly holds the electropositive metal in an electronegative pocket, creating ideal conditions for binding of metals like cobalt. The binding pocket is an octahedral structure in which four out of six coordination sites are occupied by the resin ligand, thus enhancing the binding capacity of the resin. BD TALON cobalt IMAC resin being more specific to the polyhistidine tag exhibits reduced affinity for unwanted proteins.

2.2.11.1 Protein Purification under native conditions

The characteristics of the BD TALON resin used in present study are mentioned in **Table 2L**. The protocol followed for protein purification is as follows:

- 1) Primary inoculation of *E. coli* BL21(DE3)pLysS (*pET15b-R.rubrum* RuBisCO) or other recombinant DNA prepared in present study) in 30 mL LB + Amp (100mg/mL) + Chloramp (34mg/mL) and incubation at 37°C and 150rpm for 12-13 hrs.
- 2) Secondary inoculation of *E. coli* BL21(DE3)pLysS (*pET15b-R.rubrum* RuBisCO) in 3000 mL LB + Amp (100mg/mL) + Chloramp (3mg/mL) and incubation at 37°C and 150rpm till OD_{600nm} reaches approximately 0.5 (approx. 2-3Hrs).

Table 2L: The important characteristics of the BD Talon Co²⁺ resin

| Feature | BD TALON Resin |
|--------------------------------|-------------------------|
| Capacity (mg protein/mL resin) | 5-10 |
| Matrix | Sepharose CL-6B |
| Bead Size (µm) | 45-165 |
| Max Linear flow rate (cm/hr) | 75-150 |
| Max Vol flow rate (mL/min) | 0.5-1.0 |
| pH stability | 2-14(2hrs); 3-14(24hrs) |
| Protein exclusion limit (Da) | 4 X 10 ⁶ |

- 3) Once the optimal OD₆₀₀ is attained induction for protein expression using IPTG (@ 1mM (working concentration) was done followed by incubation at 18°C and 150rpm for 18hrs.
- 4) After 18hrs of incubation the cells were pelleted down by centrifuging the culture at 7850rpm, 4°C for 20 minutes. The supernatant was discarded and the pellet dried by inverting on tissue paper.
- 5) The pellet was then stored at -80°C till used. For purification the pellet was removed from -80°C and put on ice for 15-20 minutes. The buffers used for purification are mentioned in **Table 2M**.
- 6) Removed the pellet from -80°C and kept in ice for 20-30 minutes. For each gram of pellet added 4 µL of 100mM PMSF, 80 µL of 10mg/mL Lysozyme*, 1X working concentration of ProteoGuard EDTA-free Protease inhibitor cocktail** (Clontech) and three times the weight of the pellet of buffer 1. Resuspended the pellet in the buffer by gentle vortexing and avoiding frothing.

* Lysozyme is an enzyme used to breakdown bacterial cell walls and hence enhances protein recovery. The enzyme acts by hydrolysis of 1,4 beta linkages between N-acetylmuramic acid and N-acetyl D-glucosamine residues in peptidoglycans and between the N-acetyl-D-glucosamine residues in chitodextrins.

**Clontech's ProteoGuard EDTA-free protease inhibitor cocktail is a mixture of protease inhibitors to protect target protein from proteolysis by endogenous proteases released during cell lysis. The cocktail is a 100X solution in DMSO, 1X of which consists of Aprotinin (0.8 μ M), Bestatin (50 μ M), Leupeptin (20 μ M), Pepstatin A (10 μ M) and PMSF (1mM).

Table 2M: Composition of the buffers used protein purification by IMAC under native conditions⁴

| Components | Buffer 1 (50 ml.) | Buffer 2 (50 ml.) | Buffer 3 (10 ml.) |
|---------------------------|-------------------|-------------------|-------------------|
| Imidazole | 20mM | 40mM | 1M |
| Sodium Chloride | 300mM | 300mM | 300mM |
| Sodium Phosphate (pH 8.0) | 50mM | 50mM | 50mM |

- 7) Incubated at 30 °C for 1hr with continuous shaking. After Lysozyme treatment the sample was put on ice for 30 minutes to allow cooling down.
- 8) The sample was then sonicated at 10KHz by giving 10 short bursts of 10 seconds with 30 seconds rest time. (While sonicating keep the probe properly submerged, increase sample volume if needed but not more than 10 ml.; Give short pulses; While sonicating keep the sample resting on ice to avoid overheating)
- 9) Optional: If lysate still viscous add DNase at rate of 1mg/ml. per gram of pellet and incubate at 4°C for 30 minutes in presence of 10mM MgCl₂.
- 10) After sonication the sample was centrifuged at 7850rpm, 20 minutes at 4°C. The supernatant was collected in fresh tube, labeled as fraction A and stored at 4°C.
- 11) Simultaneously, the column was prepared. The BD Talon Cobalt resin (Clontech) was thoroughly resuspended. (Use 2 ml. of resin suspension per ~3mg of anticipated polyhistidine tagged protein. 2 mL of homogenously resuspended resin will provide 1 mL bed volume of the BD Talon resin).
- 12) Immediately transferred the required amount of resin suspension to a sterile tube that can accommodate 10-20 times the resin bed volume (For 1 mL bed volume 20 mL tube).
- 13) Centrifuged at 700 x g for 2 minutes to pellet the resin (Centrifugation at higher speed can damage the resin). The supernatant was removed and discarded.
- 14) Added 10 bed volumes of buffer 1 (equilibration buffer/Sonication buffer) and mixed briefly to pre-equilibrate the resin. Re-centrifuged at 700 x g for 2 minutes to pellet the

⁴Prepared buffers freshly; Checked pH accurately; Filtered the buffers before use; Stored the buffers properly (if needed)

PMSF (Sigma) 100mM stock in 100% Isopropanol
Lysozyme (Merck) 10mg/ml in Autoclaved Milli-Q water

resin. The supernatant was removed and discarded without disturbing the resin. Step 15 was repeated.

- 15) The clarified sample from step 11 was added to the equilibrated resin. Gently agitated at 4°C for 1-2hrs on platform shaker to allow the polyhistidine-tagged protein to bind the resin. The resin was then centrifuged at 700 x g for 5 minutes at 4°C. Carefully removed the supernatant without disturbing the pellet (resin) and collected it in a fresh tube and labeled as flow through.
- 16) The resin was then washed by adding 10-20 bed volumes of buffer 2 (Wash buffer). Gently agitated the suspension at 4°C for 10 minutes on a platform shaker to promote thorough washing. The pellet was then centrifuged at 700 x g for 5 minutes at 4°C. The supernatant was collected without disturbing the pellet and collected in fresh tube labeled as wash 1. Step 17 was repeated and supernatant collected in fresh tube labeled as wash 2.
- 17) Added one bed volume of the wash buffer to the resin bed from step 17 and resuspended. The resin was then transferred to 2 ml. gravity column with end cap in place and the resin was allowed to settle out of suspension. The end-cap was then removed and the buffer was allowed to drain until it reached the top of the bed resin, making sure no air bubbles are trapped in the resin bed.
- 18) The column was then again washed with 10 ml. bed volume of wash buffer and the fractions collected in 1 ml. aliquots and labeled as wash 3-12. (washing is done till the Bradford reagent test shows no change in color).
- 19) Finally eluted the polyhistidine-tagged protein by adding 5 ml. bed volume of buffer 3 (elution buffer) to the column and the fractions were collected (0.5 ml. each) and labeled as E1-E10. (Under most conditions, the majority of the polyhistidine-tagged protein will be recovered in the first two bed volumes). Spectrophotometric method, Bradford assay and SDS-PAGE analysis was done to determine which fractions contained the majority of tagged protein.

2.2.11.2. Desalting of the eluted fractions

The desalting of the purified protein was done using size exclusion chromatography. The matrix of the column used is made of cross-linked dextran with epichlorohydrin. The fractionation range for globular proteins is between 1,000 - 5,000 Da, with an exclusion limit of approximately 5,000Da. The matrix is compatible with a wide range of buffers and solutions and is stable within the pH range of 2-12 (column details in **Table 2N**).

- 1) The elute fractions were pooled and desalted using Desalting column (DS03 Bangalore Genei).

Table 2N: Column characteristics of DS03 used for protein desalting

| | |
|---|-------|
| Column Volume (ml.) | 10 |
| Void Volume (ml.) | 3.5 |
| Recommended sample volume (ml.) | 1-2.5 |
| Sample dilution, fold of applied volume (ml.) | 1.5-3 |

- 2) The sample concentration does not influence the separation as long as the viscosity does not differ more than a factor of 1.5 from that of the buffer used. This corresponds to a maximum concentration of 50mg/ml. for proteins or 5mg/ml. for high molecular weight polymers such as dextran, when normal aqueous phase buffers are used. The sample should be completely dissolved. Centrifuged or filtered through 0.45µm filter to remove particulate material, if necessary.
- 3) The column and the buffers were brought to ambient temperature. Removed first the top cap and then the bottom cap of the column. Care should be taken to avoid air bubbles in the column. The column is equilibrated with 50 ml. thrombin buffer i.e. 5-bed volumes (Thrombin Buffer - Tris-Cl pH 8.0 = 20mM; Sodium Chloride = 150mM; Calcium Chloride = 2.5mM). The buffer was allowed to flow through under gravitational force.
- 4) Added the recommended amount of sample (2.5 ml. for DS03) gently and allowed it to drain completely. The flow through was collected in fresh tube and labeled as flow through desalting. Then add recommended volume of buffer (0.5 ml.) and collect as discard.
- 5) Eluted the protein with the recommended volumes according to the user manual and collected the fractions (0.5 ml. each) in fresh tubes and labeled as E1-E8 (4.3 ml. for DS03). The protein rich fractions determined by Spectrophotometric and SDS-PAGE analysis were pooled and used for further steps.

2.2.11.3. Histidine tag cleavage

Thrombin is a site specific endoprotease, tested for cleavage of the sequence LeuValProArg↓GlySer, in recombinant fusion proteins. A 1:2000 weight:weight ratio of thrombin (one unit per mg of protein) to the target protein is sufficient for cleavage in 1X thrombin buffer at 16°C for 16hrs. However, since each protein represents the cleavage site differently, it is advisable to set trials at different thrombin concentrations, incubation temperatures and incubation time to obtain optimal enzyme efficiency. Excess enzyme concentration can cause non-specific proteolysis.

- 1) The protein once exchanged into thrombin buffer was then subjected to thrombin treatment (1U of thrombin for 10µg of protein) at 16°C for 16 hrs.
- 2) After the thrombin treatment the protein was subjected to desalting using the desalting column into 20mM Tris-Cl (pH 8.0) buffer according to steps 24-26. The protein fractions

collected were analyzed by SDS-PAGE for purity and UV spectrophotometric analysis for concentration and pooled. The protein was then stored at -80 °C after adding 20% glycerol.

Methods mentioned in section 2.2.12.1 – 2.2.12.3 were used for purification of proteins *R. rubrum* RuBisCO (Chapter 6) and *G. violaceus* CcmM (Chapter 5).

2.2.11.4 Protein purification under denaturing conditions

- 1) Inoculated 10 mL Luria Bertani broth (100mg/mL ampicillin and 34mg/mL chloramphenicol) with *E. coli* BL21(DE3)pLysS-*pET15b-GvRO* (*G. violaceus* PCC 7421 *rubisco* operon cloned in *pET15b* vector in host *E. coli* BL(DE3)21pLysS; or recombinant DNA transformed into host) from glycerol stock and incubated at 37°C, 150rpm for 12 hrs.
- 2) Used the 12 hr grown primary culture (mentioned in point 1 above) for inoculation of 100 mL Luria Bertani broth (in 250 mL flask) in presence of 100mg/mL ampicillin and 34mg/mL chloramphenicol at rate of 1% and incubated at 37°C, 150rpm till OD₆₀₀ reaches 0.4-0.5.
- 3) Once desired OD₆₀₀ is reached, collected 1 mL sample (uninduced culture in a fresh tube, pelleted at 5000rpm, 10 mins, room temperature). The pellet obtained was resuspended in SDS-lysis buffer (100mM Tris-Cl (pH 6.8); 4% SDS; 0.2% Bromophenol Blue; 20% Glycerol; 200mM β-Mercaptoethanol), boiled at 100°C for 10 minutes and then centrifuged at 13000 rpm for 1 minute at room temperature.
- 4) Added IPTG to the rest of the culture at rate of 0.5% and incubated at 18°C for 8 hrs at 150rpm.
- 5) After 8hrs of incubation, the cells were pelleted by centrifuging at 7800rpm, 20 minutes, 4°C. The supernatant was discarded and the pellet was stored at -80°C until used.
- 6) Prepared the buffers as mentioned in **Table 20**.
- 7) Removed the pellet from -80°C and kept in ice for 20-30 minutes. For each gram of pellet added 4 μL of 100mM PMSF, 80 μL of 10mg/mL Lysozyme and three times the weight of the pellet of buffer 1. Resuspended the pellet in the buffer by gentle vortexing and avoiding frothing.
- 8) Incubated at 30°C for 1hr with continuous shaking. After Lysozyme treatment the sample was put on ice for 30 minutes to allow cooling down.
- 9) The sample was then sonicated at 10KHz by giving 10 short bursts of 10 seconds with 30 seconds rest time. (While sonicating keep the probe properly submerged, increase sample volume if needed but not more than 10 mL; Give short pulses; While sonicating keep the sample resting on ice to avoid overheating)

Table 20: Buffers used in protein purification under denaturing conditions⁵

| Components | Buffer 1 (10 mL.) | Buffer 2 (10 mL.) | Buffer 3 (50 mL.) | Buffer 4 (50 mL.) | Buffer 5 (10 mL.) |
|------------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
| Imidazole | 5mM | 5mM | 20mM | 40mM | 1M |
| Sodium Chloride | 300mM | 300mM | 300mM | 300mM | 300mM |
| Sodium Phosphate (pH 8.0) | 50mM | 50mM | 50mM | 50mM | 50mM |
| Triton-X 100 | - | 0.5% | - | - | - |
| Urea | - | - | 8M | 8M | 8M |
| Glycerol | 10% | - | 10% | - | - |

- 10) Optional: If lysate still viscous add DNase at rate of 1mg/ml. per gram of pellet and incubate at 4 °C for 30 minutes in presence of 10mM MgCl₂.
- 11) After sonication the sample was centrifuged at 7850rpm, 20 minutes at 4°C. The supernatant was collected in fresh tube, labeled as fraction A and stored at 4°C.
- 12) The pellet obtained in step 11 was washed in buffer 2 and centrifuged at 7850rpm, 20 minutes at 4 °C. The supernatant was collected in fresh tube, labeled as fraction B and stored at 4 °C.
- 13) Resuspended the pellet obtained in step 12 in buffer 3, sonicated as mentioned in step 9 and centrifuged at 7850rpm, 20 minutes at 4°C. The supernatant was collected in fresh tube, labeled as fraction C and stored at 4°C.
- 14) Analyzed the fractions A, B and C on SDS-PAGE (12% resolving gel and 5% stacking gel).
- 15) To ensure proper resolution the samples A, B and C were boiled with SDS-Lysis Buffer in three ratios viz. 1:1, 1:3 and 3:1 (SDS lysis buffer : sample) and labeled as 1, 2 and 3, respectively.
- 16) The results of the above mentioned exercise show the desired protein i.e. *G. violaceus* RuBisCO (52.8 kDa according to *in-silico* tool generated molecular weight from the amino acid sequence) in the fraction C i.e., the one sonicated in the presence of 8M Urea.
- 17) Hence, the purification by affinity chromatography was carried out under denaturing conditions i.e., in presence of 8M Urea.
- 18) The BD-Talon Co²⁺ resin (1 mL bed volume) was equilibrated with 20 mL equilibration buffer viz. buffer 3.

⁵ Prepared buffers freshly; Checked pH accurately; Filtered the buffers before use; Stored the buffers properly (if needed). PMSF 100mM stock in 100% Isopropanol; Lysozyme 10mg/ml in Autoclaved Milli-Q water

- 19) The sample i.e. fraction C was applied to the equilibrated resin and incubated at 4°C for 1-2 hrs.
- 20) The flow through was then collected and stored at 4°C. The column was then washed with 20 ml. Wash buffer i.e., buffer 4. The wash fractions were collected of 6 ml. each.
- 21) Finally the elution was done using buffer 5 and elute fractions of 2 ml. each were collected. The wash fractions and the elute fractions obtained were analyzed by SDS-PAGE (12% separating gel and 5% stacking gel).
- Protein refolding of the solubilised proteins is initiated by the removal of the denaturant. The efficiency of the refolding procedure depends on competition between correct folding and protein aggregation. In order to slow down the process of aggregation, the refolding is carried out at low protein concentrations (10-100µg/ml). The variables important for protein refolding include buffer composition, temperature and additives to the buffer. The types of additives used include denaturant (eg urea; concentration is decreased in subsequent dialysis setups), inhibitors of aggregation (eg sugars), stabilizers (glycerol/PEG) and redox system to facilitate disulfide bond formation (reduced glutathione and oxidized glutathione). With these considerations the renaturing buffers were designed and optimized for appropriately folded protein.
- 22) The elute fractions E4-E12 were pooled exchanged sequentially into renaturing buffers as mentioned in **Table 2W**.
- 23) The proteins bound to the resin were washed with renaturing buffers sequentially using 25 volumes of column. The protein was then desalted and exchanged into Thrombin buffer (Tris-Cl (pH 8.0), Sodium Chloride 150mM and Calcium Chloride 2.5mM) using Desalting column.
- 24) The protein was then treated with thrombin protein (1U for 10µg of protein) and incubated at 16°C for 16Hrs. The thrombin digested protein was then checked on SDS-PAGE.

Table 2W: Buffers used for renaturing of the purified proteins

| Components | Buffer 1 | Buffer 2 | Buffer 3 | Buffer 4 |
|---------------------------|----------|----------|----------|----------|
| Sodium Phosphate (pH 8.0) | 50mM | 50mM | 25mM | 25mM |
| Sodium Chloride | 300mM | 300mM | 150mM | 100mM |
| Imidazole | 500mM | 250mM | 100mM | 50mM |
| Urea | 6M | 4M | 2M | - |
| Glycine | 100mM | 50mM | 10mM | - |
| Glycerol | 20% | 10% | 10% | 10% |

25) The Thrombin digested protein was finally desalted into 20mM Tris-Cl (pH 8.0) and concentrated using 10KDa cut-off filter. The protein was then checked on SDS-PAGE and native PAGE. The protein was stored in -80 °C in 20% glycerol.

The method mentioned in section 2.2.12.4 was used for purification of *G. violaceus* RbcL, RbcS and RbcX, (Chapter 4) as these proteins were found to be in the insoluble fraction according to the protocol followed in section 2.2.11.

2.2.12. Mass Spectrometry analysis

In gel digestion coupled with mass spectrometric analysis is a powerful tool for identification and characterization of proteins. The purified protein was run on SDS-PAGE, desired band excised and in gel digestion was carried out (ThermoScientific In-gel tryptic digestion kit). Trypsin is a serine protease that specifically cleaves peptide bonds at the carboxyl side of lysine and arginine residues. However, the cleavage can be blocked by proximal acid, aromatic or proline residues. The modified Trypsin provided in the kit has limited autocatalytic activity that should not interfere with mass spectral analysis. A trypsin fragment of mass 842.51 (m/z, M + H) will be the most common and should be used as an internal standard.

- 1) The excised band of *R. rubrum* RuBisCO (Chapter 6), *G. violaceus* RbcL, RbcS (Chapter 4) and CemM (Chapter 5) was kept in presence of destaining solution for 30 minutes, followed by reduction and alkylation of cystine using TCEP (Tris[2-carboxyethyl]phosphine) and IAA (Iodoacetamide), respectively, which improves recovery of cystine-containing peptides from in-gel digests and reduces the appearance of unknown masses in MS analysis.
- 2) Further, the gel was washed in 50% acetonitrile, dried and then rehydrated with digestion buffer and activated trypsin.
- 3) The overnight digested sample was then used for MS analysis. Target protein identification was done by Matrix Assisted Laser Desorption Ionization Time of Flight-Mass Spectrometry (MALDI-TOF-MS) (4800 plus MALDI TOF/TOF Analyzer, AB SCIEX, USA) at Delhi University South Campus CIF.
- 4) Samples for MALDI-TOF-MS was prepared by mixing 6 μ L of saturated α -cyano-4-hydroxycinnamic acid matrix (10 mg/ml of 50 % acetonitrile and 0.05 % trifluoroacetic acid in 0.2 micron filtered water) with 6 μ L of eluted peptide solution for each spot.

- 5) The samples were prepared immediately and vortexed to make a homogenous solution and 1 μ l of the mixture was dropped onto the metal target plate. The samples were air-dried completely.
- 6) The plate was properly aligned and calibrated using calibration mixture (4700 proteomics analyzer calibration mix, AB SCIEX, USA). Samples for calibration were made according to the manufacturer's instructions. Samples were analyzed using MS mode to obtain the peptide mass fingerprint (pmf) of each spot.

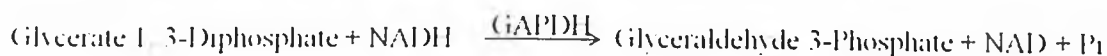
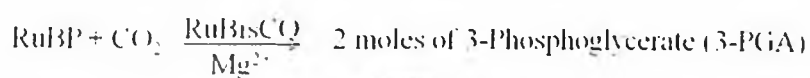
The raw data obtained for *R. rubrum* RuBisCO was analyzed by Mascot (Perkins et al., 1999; Koeing et al., 2008). Mascot uses probability (probability of occurrence of a search hit by chance) based scoring to judge whether a result is significant or not. Probability (p) is represented as $-10\log_{10}(p)$. Low probability means high score. Area in green represents – 95% of the probability density function. Scores to the right of green region have <5% chance to occur by chance. The MS data for *G. violaceus* RbcL, RbcS and CcmM were obtained in the form of charge mass ratio peak spectra.

2.2.13. Circular Dichroism (CD)

In optics, a dichroic material is either one which causes visible light to be split up into distinct beams of different wavelengths, or one in which light rays having polarization are absorbed by different amounts. Circular dichroism is dichroism involving circularly polarized light, i.e., the differential absorption of left and right handed light. CD is exhibited by any optically active molecule and hence biological molecules because of their dextrorotary and levorotary components are no exception. Further, the secondary structure of the biological molecules also imparts a CD signature to the molecule under study. Although, CD spectra can be used to estimate fraction of molecule in alpha helix conformation, beta sheet conformation or any other structure, but the location of these structural conformations cannot be identified. CD is an important tool to study the effect of external factors like temperature and denaturants on the secondary structure of the molecule, and hence can be used to verify that the protein under study is in native form or denatured form.

CD spectra of *G. violaceus* RuBisCO (Chapter 4) were recorded between 190 and 240nm with a JASCO J-815 spectropolarimeter, in a 0.1cm pathlength suprasil quartz cuvette at 20°C, scan rate of 50nm min⁻¹, 1nm bandwidth, calibrated with 0.06% (w/v) aqueous solution of (+)-10-camphorsulfonic acid at Delhi University South Campus CIF. The spectra were corrected for baseline and the secondary structure prediction analysis was done using DICHIROWEB (Whitmore and Wallace 2004).

2.2.14. RuBP assay to study the enzyme kinetics of RuBisCO⁶



RuBisCO assay was performed at 25 °C in a medium containing 100mM HEPES-KOH pH 8.0, 20mM MgCl₂, 50mM NaHCO₃, 0.2mM NADH, 5mM ATP, 20mM Phosphocreatine, 12.5U per ml. PCK, 11.25U per ml. PGK, 10U per ml. GAPD and 0.05mg CA. The initial activity was determined by adding 1µg of the protein to 1 mL reaction mixture. The reaction without RuBP was taken as the baseline for taking absorbance till the OD₃₄₀ stabilized, following which, 2mM RuBP (final concentration) was added and change in absorbance observed. The protein was activated in presence of activation buffer (HEPES-KOH pH 8.0 – 50mM, MgCl₂ – 25mM, EDTA – 1mM, NaHCO₃ – 25mM (freshly added) at 25 °C for 60 minutes. The absorbance at 340nm was measured using Shimadzu spectrophotometer (BITS Pharmacy Department). All the materials used for the assay were procured from Sigma.

2.2.14.1. Preparation of coupling enzyme mixture

The lyophilized proteins tend to be very hygroscopic, so they should not be opened in humid areas. The vials were taken out from refrigeration about half an hour before use and kept in a desiccator. The proteins were then dissolved in appropriate buffers. In the present study the all the enzymes were dissolved in 1 mL assay buffer (HEPES – 100mM; MgCl₂ – 20mM pH 8.0), to obtain final concentrations of 5KU/mL, 10KU/mL, 17.5KU/mL and 100mg/mL for 3-Phosphoglyceric Phosphokinase, Glyceraldehyde-3-phosphate dehydrogenase, Creatine Phosphokinase and carbonic anhydrase, respectively. The enzymes were then mixed to obtain final concentrations of 2.25U/µL, 2U/µL, 2.5U/µL and 0.01mg/µL for 3-Phosphoglyceric Phosphokinase, Glyceraldehyde-3-phosphate dehydrogenase, Creatine Phosphokinase and carbonic anhydrase, respectively. The mixture obtained was then dialyzed against assay buffer for 6hrs at 4°C, in order to remove any ammonium salts. The working concentration required for each of the enzymes is 11.25U, 10U, 12.5U and 0.05mg for 3-Phosphoglyceric Phosphokinase, Glyceraldehyde-3-phosphate dehydrogenase, Creatine Phosphokinase and carbonic anhydrase, respectively, for 1 mL RuBP assay according to Pearce and Andrews (2003). To achieve this working concentration, 5 µL of coupling enzyme mixture can be used. The coupling enzyme mixture hence prepared was stored at -80°C in small aliquots.

⁶ Coupling enzymes: PGK – Phosphoglycerate Kinase; GAPDH – Glyceraldehyde 3 phosphate Dehydrogenase

2.2.14.2. Preparation of reaction components of assay

NADH, β -Nicotinamide Adenine Dinucleotide (reduced form), is stable at 0-5°C, if stored in dark and under desiccated conditions. The solution of 10mM was prepared in sterile buffer (Assay buffer), as unbuffered water may be acidic and can decompose β -NADH, and stored at -20 °C in small aliquots. The solution was prepared fresh as potent enzyme inhibitors have been reported to form in frozen solutions. These inhibitors if formed cannot be detected as they have the same absorbance at 340nm as β -NADH. ATP (Adenosine Triphosphate) 0.5M solution, Phosphocreatine 250mM solution and RuBP 330mM solution were prepared in sterile water in small aliquots and stored at -20°C. The NaHCO₃, 0.5M solution was prepared freshly before every use. The RuBisCO activation Buffer consists of HEPES (pH 8.3) – 50mM, MgCl₂ – 25mM, EDTA – 1mM and NaHCO₃ – 25mM (freshly added).

Protocol:

- 1) The RuBisCO purified by affinity after excision of His-tag was exchanged into RuBisCO activation buffer by using desalting column.
- 2) The protein was concentrated using Amicon Ultra-2 centrifugal filter devices.
- 3) Then the concentration of the protein was estimated.

2.2.14.3. Estimation of Protein concentration

$$A_{280} (1\text{mg/ml}) = \frac{5690 n(W) + 1280 n(Y) + 120 n(C)}{\text{Molecular Weight of Protein}}$$

Where W – Tryptophan; Y – Tyrosine; C - Cysteine

2.2.14.4. Activation of RuBisCO

The RuBisCO was first activated by incubating 1 μ g of purified protein in 20 μ l. activation buffer (50mM HEPES pH 8.3; 25mM MgCl₂; 1mM EDTA; 25mM Sodium Bicarbonate) at room temperature for 20 minutes.

2.2.14.5. Preparation of assay mixture

All the assay components were removed from -20°C and the coupling enzyme from -80°C and allowed to thaw on ice. A clean set of quartz cuvettes with a path length of 1cm was used for the assay. All the components were mixed directly in the cuvette, in the order and volume mentioned below. Both NAD⁺ and NADH strongly absorb UV light because of adenine but at slightly different wavelengths. The absorbance maxima for NAD⁺ is at 259nm with extinction coefficient of 16,900M⁻¹cm⁻¹ and for NADH it is at 339nm, with extinction coefficient of 6220M⁻¹cm⁻¹. This difference in absorption properties makes it simple to measure conversion of one to another in enzymatic assays by taking optical density measurements at 340nm.

| Component | Volume added from stock (μL) | Final Working concentration |
|---|---|-------------------------------------|
| Assay Buffer | 778.94 μL . | HEPES 200mM; MgCl_2 - 20mM |
| NADH | 20 μL . | 0.2mM |
| Sodium Bicarbonate | 100 μL . | 50mM |
| Phosphocreatine | 80 μL . | 20mM |
| Coupling Enzymes | 5 μL . | PCK-12.5U, PGK-11.25U, GAPD-10U |
| Add 10 μL ATP (working conc - 5mM) | | |
| Take OD_{340} and let run out | | |
| Then add activated RuBisCO | | |
| Take OD_{340} and let run out | | |
| Add 0.2mM RuBP and take OD_{340} at interval of 1 sec upto 400-500secs | | |

The mass extinction coefficient and the molar extinction coefficient are the parameters defining how strongly a substance absorbs light at a given wavelength, per mass density or per molar concentration, respectively. The expected observation in the present assay is sudden drop in OD_{340} upon addition of substrate. From the slope of the linear part of the progress curve, the enzyme velocity is obtained as the amount of substrate converted during a unit time. A catalyzed reaction must initially follow a linear relationship, from which its velocity can be calculated. Further, upon depletion of the substrate, the reaction velocity decrease and then ultimately ceases. Therefore, it is essential that in order to obtain velocity of the reaction, only linear part of the progress curve should be evaluated.

Considering the linear part of the curve, the change in absorbance per minute was found out and used for calculation of enzyme activity. The formulas used are as follows:

$$\text{Units/mL enzyme} = \frac{(\Delta A_{340\text{nm}}/\text{min Test} - \Delta A_{340\text{nm}}/\text{min Blank}) (1) (\text{dilution factor})}{(4) (6.22) (\text{volume of enzyme used})}$$

1 = total volume (in mL) of assay

4 = 4 μmoles of $\beta\text{-NADH}$ are oxidized for each μmole of RuBP utilized

6.22 = millimolar extinction coefficient for $\beta\text{-NADH}$ at 340nm

$$\text{Units/mg solid} = \frac{\text{Units/mL enzyme}}{\text{mg/mL of enzyme}}$$

One unit of enzyme will convert 1 μmole of RuBP and CO_2 to 2 μmoles of 3-phosphoglycerate per minute at pH 7.8 at 25°C.

2.2.14.6. Determination of Michaelis Menten constant and turnover number of the enzyme

The K_M RuBP is measured by adding activated RuBisCO to assay mixtures containing RuBP concentrations 0.5, 1.0, 1.5, 2.0 and 2.5mM (working concentraion). K_M is the concentration

of the substrate where the initial velocity is half the maximum rate possible under the conditions of the experiment and V_{max} is the maximum initial velocity. The value of K_M reflects the stability of the enzyme-substrate interaction. It should be noted that more stability is achieved by using the linear form of the Michaelis – Menten equation, viz. Lineweaver-Burk or Hanes to calculate K_M .

The carboxylation turnover rate was obtained by dividing the Michaelis Menten extrapolated measure of V_{max} by the concentration of RuBisCO.

$$K_{cat} = \frac{V_{max}}{E_t \text{ (initial enzyme concentration)}}$$

The *R. rubrum* RuBisCO (Chapter 6) and *G. violaceus* RuBisCO (Chapter 4) were subjected to RuBisCO assay according to the protocol mentioned above and the V_{max} , K_M RuBP and K_{cat} for both proteins was analyzed.

2.2.15. The effect of RuBisCO SSU and CemM on activity of RuBisCO LSU

In order to evaluate the effect of RuBisCO small subunit and CemM proteins on the activity of the RuBisCO protein, the active site of which is stationed in the large subunit of the protein, the RuBisCO assay was conducted for the following combinations of purified proteins:

- *R. rubrum* RuBisCO
- *G. violaceus* PCC 7421 RuBisCO Large subunit
- *G. violaceus* PCC 7421 RuBisCO Large subunit + RuBisCO small subunit
- *R. rubrum* RuBisCO + *G. violaceus* PCC 7421 RuBisCO small subunit
- *R. rubrum* RuBisCO + *G. violaceus* PCC 7421 CemM

The proteins were incubated in 1:1/1:3 stoichiometric ratios at room temperature for duration of 60 minutes to allow *in-vitro* protein-protein interactions to take place. The K_M and K_{cat} was calculated for each of the enzyme combinations studied (Chapter 6).

of the substrate where the initial velocity is half the maximum rate possible under the conditions of the experiment and V_{max} is the maximum initial velocity. The value of K_M reflects the stability of the enzyme-substrate interaction. It should be noted that more stability is achieved by using the linear form of the Michaelis – Menten equation, viz. Lineweaver-Burk or Hanes to calculate K_M .

The carboxylation turnover rate was obtained by dividing the Michaelis Menten extrapolated measure of V_{max} by the concentration of RuBisCO.

$$K_{cat} = \frac{V_{max}}{E_i \text{ (initial enzyme concentration)}}$$

The *R. rubrum* RuBisCO (Chapter 6) and *G. violaceus* RuBisCO (Chapter 4) were subjected to RuBisCO assay according to the protocol mentioned above and the V_{max} , K_M RuBP and K_{cat} for both proteins was analyzed.

2.2.15. The effect of RuBisCO SSU and CcmM on activity of RuBisCO LSU

In order to evaluate the effect of RuBisCO small subunit and CcmM proteins on the activity of the RuBisCO protein, the active site of which is stationed in the large subunit of the protein, the RuBisCO assay was conducted for the following combinations of purified proteins:

- *R. rubrum* RuBisCO
- *G. violaceus* PCC 7421 RuBisCO Large subunit
- *G. violaceus* PCC 7421 RuBisCO Large subunit + RuBisCO small subunit
- *R. rubrum* RuBisCO + *G. violaceus* PCC 7421 RuBisCO small subunit
- *R. rubrum* RuBisCO + *G. violaceus* PCC 7421 CcmM

The proteins were incubated in 1:1/1:3 stoichiometric ratios at room temperature for duration of 60 minutes to allow *in-vitro* protein-protein interactions to take place. The K_M and K_{cat} was calculated for each of the enzyme combinations studied (Chapter 6).

CHAPTER 3

Distribution and Diversity of Carbon Concentrating Mechanism Proteins of β - cyanobacteria

3.1. Introduction

The belief of occurrence of cell compartmentalization in eukaryotes alone was challenged by the discovery of carboxysomes in cyanobacteria in 1961 (Jensen and Bowen 1961), followed by subsequent observation of more such microcompartments in proteobacteria. Bacterial microcompartments (BMCs) are polyhedral protein complexes (40-200 nm in diameter), encasing metabolic enzymes encapsulated in a protein shell and are utilized by several prokaryotes to create optimal conditions for certain metabolic processes such as carbon fixation, ethanolamine utilization, 1,2 propanediol utilization etc. (Kerfeld et al., 2010). The BMCs serve to conserve the metabolic pathways enclosed in it from the rest of the cellular environment for several possible reasons, which could be prevention of release of toxic metabolites, enhancing the metabolic capability of certain molecules, etc. (Penrod and Roth 2006; Sampson and Bobik 2008). The major kinds of BMCs discovered till date include carboxysomes (carbon fixation) (Jensen and Bowen 1961), Pdu BMC (1,2 propanediol utilization) (Chen et al., 1994) and Eut BMC (Ethanolamine utilization) (Kofoid et al., 1999).

The most extensively studied BMC is carboxysomes. Carboxysomes are a part of a carbon concentrating mechanism of cyanobacteria and some proteobacteria. The carboxysomes evolved in order to overcome the inefficiency of enzyme Ribulose-1, 5-Bisphosphate Carboxylase/Oxygenase (RuBisCO), which it suffers at the hands of photorespiration because of its fickle specificity to carbon dioxide and oxygen (Badger et al., 1998). The β -carboxysomes constitute structural proteins CcmK, CcmL and CcmO which form the carboxysome shell and CcmM, CcmN which are enclosed by the carboxysome shell along with RuBisCO (Kaplan and Reinhold 1999; Badger et al., 2002; Cannon et al., 2002). The CcmK and CcmO possess the BMC domain while the CcmL protein constitutes the Pfam03319 domain. The BMC domain proteins form the 20 flat facets of the shell while the Pfam03319 domain proteins form the pentamers, that provide curvature to the carboxysome shell by occupying the 12 vertices (Tanaka et al., 2008).

The Pdu microcompartments sequester an intermediate of 1,2 propanediol degradation (propionaldehyde) so as to avert toxicity. The Pdu BMCs are formed of 14 different polypeptides viz. PduABB'CDEGHJKOPTU (Havemann and Bobik 2003). The shell is composed of BMC domain proteins PduABB'JKTU and the enzymes of the metabolic pathway constitute Pdu CDE (B₁₂ dependent dehydratase), PduGH (diol dehydratase), PduO (adenosyl transferase) and PduP (Propionaldehyde dehydrogenase) (Havemann and Bobik 2003). Eut BMC is found in bacteria utilizing ethanolamine as carbon, nitrogen and energy source, formed as a result of phosphatidyl ethanolamine degradation in mammalian gastrointestinal tract (Kerfeld et al., 2010). The exact protein composition of Eut BMCs is unknown till date because the purification of the Eut BMCs is still not reported. The

versatility in the role of BMCs suggests that they contribute to metabolic innovation in bacteria in a broad range of environments (Kerfeld et al., 2010).

The evolutionary relationships between different types of microcompartments are not very clear but evidently they all contain a conserved outer shell structure, enclosing certain enzymatic reactions. As discussed by Rae et al., (2013), the formation of two types of carboxysomes (α and β) is by convergent evolution. The β -carboxysomes are lumen centric and possess higher intrinsic order with the lumen proteins capable of self-assembly, while the α carboxysomes are shell centric being capable of assembling the shell even in the absence of the core proteins (Menon et al., 2008). Such observations suggest that possibly the α carboxysomes came into existence by recruitment of carboxysome core proteins to pre-existing BMC shell in primeval cell (Rae et al., 2013). Moreover, as reported by Fan et al., (2012) α carboxysomes and Pdu microcompartments are capable of incorporating enzyme complexes through connection with the BMC shell. Alternatively, as conceived by Rae et al., (2013) RuBisCO IA could have incorporated into a pre-existing microcompartment shell by the targeting method, provided there was co-existence of another type of BMC in the same cell. There are reports of organisms possessing more than one type of BMC viz. *Salmonella enterica serovar typhimurium* contains both Pdu and Eut microcompartments (McClelland et al., 2001).

The mechanism by which the bacterial microcompartments control the passage of various respective metabolic molecules is not yet known, but analyzing these pathways may bring forth novel properties of biological protein sheets such as carboxysome shells. The protein databases reveal numerous microcompartment forming proteins involved in diverse metabolic pathways. Furthermore, in many cases several copies of the shell protein are prevalent in the same organism. The study of differences in these homologs shall highlight certain characteristic features for each one and hence aid in analyzing the role of these proteins. The present study involves the screening of various phyla of Eubacteria in order to identify homologs of carboxysome proteins and to evaluate the diversity of each of the CCM proteins.

3.2. Methodology

3.2.1. Identification of early diverging cyanobacteria

The 16S rRNA sequences were retrieved from 37 cyanobacterial strains available at Kazusa Genome Resource (<http://genome.microbedb.jp/>). The sequences retrieved were aligned using ClustalW (Thompson et al., 1994) in GUIDANCE2 server (Sela et al., 2015, Landan and Graur 2008). The GUIDANCE2 server accounts for errors in process of indel formation, formation of guide tree and provides optimal solutions for pairwise alignments which in turn form the building blocks of multiple sequence alignment. The server provides MSA of the

sequences as well as MSA with unreliable columns/sequences removed. The trimmed alignment file was then used for evaluation of best fit substitution model in MEGA 6.0 (Tamura et al., 2013). The best model on basis of BIC score was then used to carry out phylogenetic analysis with 1000 replicates of bootstrapping. The phylogenetic tree based on 16S rRNA sequence was then compared with phylogenetic trees generated by similar methodology for DNA sequences for *pyrH* and *rbcL* from 37 cyanobacterial strains. Sequences of *Escherichia coli* str K-12, *Staphylococcus aureus*, *Rhodospseudomonas palustris* CGA009 and *Chlorobium tepidum* TLS (where appropriate) were used for outgrouping the phylogenetic trees.

The MSA files of 16S rRNA, *rbcL* and *pyrH* were then concatenated by sequence matrix software (Gaurav et al., 2010) and the output file hence generated was analyzed phylogenetically after best fit substitution model testing in MEGA6.0. This practice will give more robust inference of early diverging cyanobacteria. The best fit models for various datasets evaluated are Kimura 2-parameter model (Kimura 1980) for 16S rRNA, General Time Reversible model (Nei and Kumar 2000) for concatenated sequences and operon sequences. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. A discrete Gamma distribution was used to model evolutionary rate differences among sites. The rate variation model allowed for some sites to be evolutionarily invariable.

3.2.2. Identification of CCM homologs across all Eubacterial phyla

The amino acid sequences of carbon concentrating mechanism proteins of *Gloeobacter violaceus* PCC 7421 were retrieved from Kazusa Genome Resource. The *G. violaceus* PCC 7421 proteins were used as the query since according to 16S rRNA analysis, it is an early diverging cyanobacteria. The accession numbers for each CCM proteins of *G. violaceus* used in the study are mentioned in **Table 2A**.

The protein sequences were used as query in BLASTP against phyla of Eubacteria (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) (Altschul et al., 1990, 1997). The exercise brought forth several proteins as hits which either perform the same function as the query sequence or perform a different function but still possess appreciable sequence homology. The BLASTP hits obtained were surveyed such that the matching sequences with an expect value equal to or lower than 10^{-5} to atleast one species of each of the phyla of Eubacteria were selected. The various phyla included in the study are mentioned in **Table 3A**.

To further validate the BLASTP search results, the query sequences were used to carry out TBLASTN search to make conclusive analysis on absence or wrong annotation of a gene.

3.2.3. Phylogenetic analysis of CCM proteins

In order to carry out the phylogenetic analysis, the amino acid sequences of the homologs of the proteins under study were retrieved from NCBI and a dataset created after removing redundant sequences. These individual datasets were then uploaded into GUIDANCE2 server to carry out MSA by ClustalW and remove the poorly aligned regions of the sequence. The output file from GUIDANCE2 server after removal of unreliable columns was then used for testing phylogenetic hypothesis by performing model comparison for robust inferences in MEGA6.0. The model with lowest BIC score is considered best for describing substitution pattern and is hence used for phylogenetic analysis with 1000 bootstrapping replicates in MEGA6.0. The best models evaluated for various datasets were Le_Gascuel_2008 model (Le and Gascuel 2008) for CcmK, CcmL, CcmN and CcmO sequences, and Whelan and Goldman · Freq. model (Whelan and Goldman 2001) for CcmM sequences. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model. A discrete Gamma distribution was used to model evolutionary rate differences among sites. The rate variation model allowed for some sites to be evolutionarily invariable. Further, the complete carboxysome operon (*csa* and *ccm* operon) sequences of both α and β -cyanobacteria were retrieved and analyzed for genes present and their arrangement. The sequences of the operons and the individual CCM proteins obtained were aligned by MSA and analyzed phylogenetically (maximum likelihood method based on best fit substitution model with 1000 replicate bootstrapping) using MEGA 6.0 (Tamura et al., 2013). The distance of each of the proteins from the corresponding protein of *G. violaceus* PCC 7421 was used to identify closest relatives of the early diverging cyanobacteria. The distances are basically number of amino acid/nucleotide substitutions per site between sequences, estimated by equal input model (Tajima and Nei 1984) for amino acid sequences and Jukes Cantor model (Jukes and Cantor 1969) for nucleotide sequences in MEGA6.0. The rate variation among sites was estimated using gamma distribution.

Table 3A: The list of phyla used for the analysis of distribution of CCM proteins

| S. No | Name of phylum | Taxonomy ID | Habitat and Description |
|-------|---------------------------|-------------|--|
| 1 | Archaea | 2157 | Found in broad range of habitats and are usually mutualists or commensals |
| 2 | Actinobacteria | 201174 | Can be terrestrial or aquatic; pathogenic or commensals |
| 3 | Aquificae | 187857 | Autotrophs |
| 4 | Armatimonadetes | 67819 | Aerobic Chemoheterotroph |
| 5 | Bacteroidetes Chlorobi | 68336 | Opportunistic pathogens/Anaerobic photoautotroph |
| 6 | Caldisei | 67814 | Thiosulphate reducing bacteria |
| 7 | Chlamydiae | 51290 | Obligate intracellular pathogens |
| 8 | Chloroflexi | 200795 | Anoxygenic phototrophs/Anaerobic halorespirers |
| 9 | Chrysiogenetes | 118001 | Respire using oxidized form of arsenic |
| 10 | Cyanobacteria | 1117 | Photoautotrophs |
| 11 | Deferribacteres | 68337 | Chemoorganotrophic heterotrophs |
| 12 | Deinococcus-thermus | 1297 | Obligate aerobic chemoorganoheterotroph |
| 13 | Dictyoglomi | 68297 | Chemoorganotrophic |
| 14 | Elusimicrobia | 74152 | Symbionts in termite |
| 15 | Fibrobacteres/Acidobacter | 131550 | Includes mostly rumen bacteria; capable of plant based cellulose degradation in ruminant animals |
| 16 | Firmicutes | 1239 | Pathogenic/Photoautotrophs |
| 17 | Fusobacteria | 32066 | Pathogens |
| 18 | Gemmatimonadetes | 142182 | Slow growing heterotrophic bacteria |
| 19 | Nitrospinae | 1293497 | Aerobic chemolithotrophic bacteria |
| 20 | Nitrospirae | 40117 | Chemolithotrophic nitrite oxidizing bacteria |
| 21 | Planctomycetes | 203682 | Anammox oxidation of ammonia to nitrogen gas |
| 22 | Alpha Proteobacteria | 28211 | Phototrophic, Symbionts, Intracellular pathogens |
| 23 | Beta Proteobacteria | 28216 | Ammonia oxidizing, phototrophs |
| 24 | Gamma Proteobacteria | 1236 | Pathogenic, Chromatium-photosynthetic and oxidize Hydrogen sulfide |
| 25 | Delta Proteobacteria | 28221 | Sulfate and sulfur reducing bacteria |
| 26 | Epsilon Proteobacteria | 29547 | Pathogenic, Symbionts, Chemolithotrophic, phototrophic |
| 27 | Spirochaetes | 136 | Chemoheterotrophic |
| 28 | Synergistetes | 508458 | Opportunistic pathogens |
| 29 | Tenericutes | 544448 | Pathogenic |
| 30 | Thermodesulfobacteria | 200940 | Sulfate reducing bacteria |
| 31 | Thermotogae | 200918 | Use complex sugars to produce H ₂ |

3.3. Results and Discussions

3.3.1. *Gloeobacter violaceus* PCC 7421 is an early diverging cyanobacterium

The 16S rRNA sequence based cluster analysis shows two distinct groups of cyanobacteria i.e., the α -cyanobacteria and the β -cyanobacteria (**Figure 3A(i)**). The clustering observed in the phylogenetic tree (**Figure 3A(i)**) is congruent to one observed in other similar analysis reported (Memon et al., 2013) for α -cyanobacteria while there are some differences in the arrangement of the β -cyanobacteria with respect to each other. The observed disparity could possibly be because of slight variation in the list of β -cyanobacteria used in the analysis. Memon et al., (2013) used *Cyanothece* sp PCC 7822, *Cyanothece* sp PCC 8802, *Cyanobacterium* UCYN-A, *Nostoc* sp PCC 7120 and *Anabaena azollae* 0708 in addition to the ones included in present analysis.

Two major clades are formed, Clade A constituting *Synechococcus* sp JA-2-3B'a(2-13), *Synechococcus* sp JA-3-3Ab and *G. violaceus* PCC 7421 and the other the rest of the cyanobacteria. The second clade further branches into two sub clades for α -cyanobacteria (clade C) and β -cyanobacteria (clade B). It is to be noted that *Synechococcus elongatus* PCC 6301 and *Synechococcus elongatus* PCC 7942 although are β -cyanobacteria, are found to be clustered with α -cyanobacteria. Similar results were also obtained by 16S rRNA analysis of cyanobacteria done by Memon et al., (2013). The fact that the clade constituting *G. violaceus* PCC 7421, branches into the rest of the cyanobacteria, implies that it is equally related to the rest of the cyanobacteria and is hence early diverging. *G. violaceus* PCC 7421, a unicellular cyanobacterium dwells on the calcareous rocks in mountainous regions of Switzerland (Rippka et al., 1974). *G. violaceus* possesses several unique characteristics, it lacks thylakoid membrane and the photosynthetic machinery is situated in the cytoplasmic membrane (Rippka et al., 1974), presence of morphologically distinct phycobilisomes (Guglielmi et al., 1981) and absence of SQDG (Sulfoquinovosyl diacylglycerol), which has an important role in photosystem stabilization (Salstam and Campbell 1996). These characteristics suggest that this organism has retained ancestral cyanobacterial properties. Further, 16S rRNA phylogenetic analysis revealed *G. violaceus* PCC 7421 to be a primitive cyanobacterial organism (Nelissen et al., 1995).

To further validate the early divergence of *G. violaceus* the phylogeny constituting concatenated sequences for 16S rRNA, *pyrH* and *rbcL* was generated (**Figure 3A(ii)**). The topology of the tree depicts two major clades viz. I and II. Clade I constitutes *Synechococcus* sp JA-2-3B'a(2-13), *Synechococcus* sp JA-3-3Ab and *G. violaceus* PCC 7421. The clade II constitutes two sub clades, IIA for all α -cyanobacteria and IIB for all β -cyanobacteria. These observations direct towards early divergence of *G. violaceus* PCC 7421.

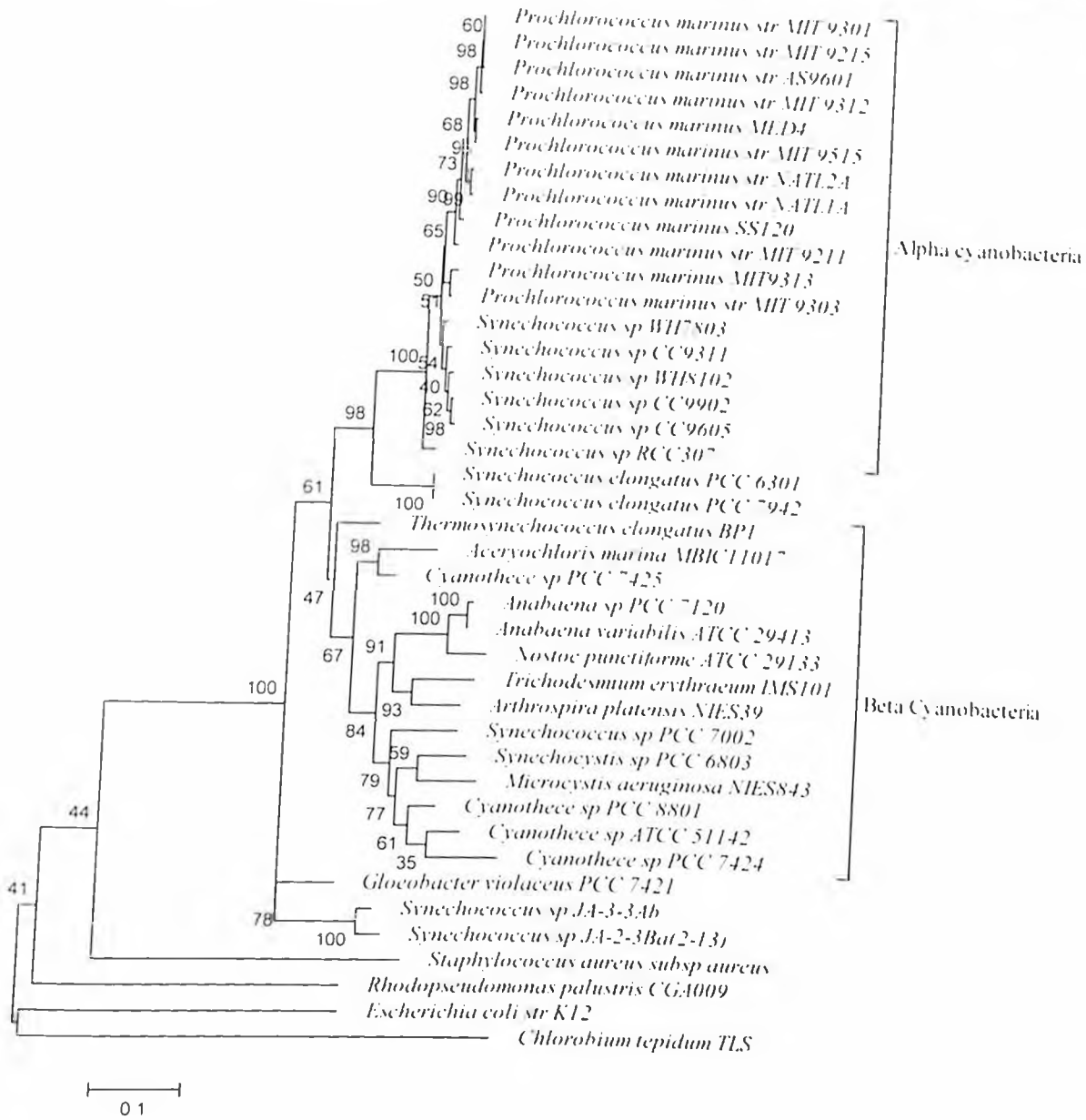


Figure 3A(i): Maximum Likelihood phylogenetic representation of cyanobacteria based on 16S rRNA sequence. The evolutionary history was estimated on basis of Kimura 2-parameter model with 1000 bootstrapping replicates. The tree with the highest log likelihood (-10787.2203) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (-G, parameter = 0.3414)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 28.2124% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 41 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 1274 positions in the final dataset. Evolutionary analyses were conducted in MEGA6. The outgrouping of the tree was done using the 16S rRNA sequences of *Rhodospseudomonas palustris* CGA009, *Chlorobium tepidum* TLS, *Staphylococcus aureus* and *Escherichia coli*.

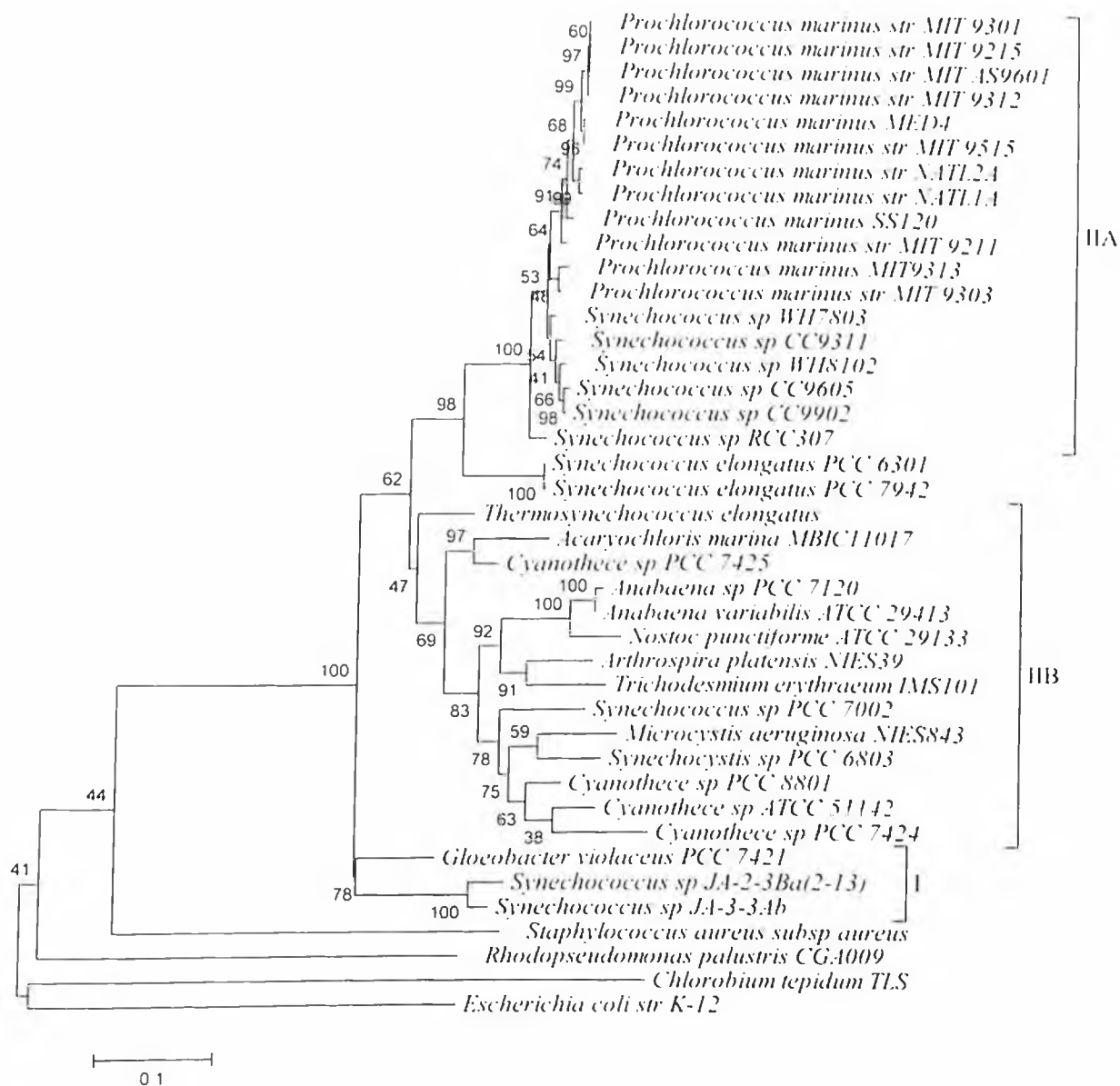


Figure 3A(ii): Maximum Likelihood phylogenetic representation of cyanobacteria based on 16S rRNA-*pyrH-rbcL* sequence. The sequences were concatenated using sequence matrix software. The evolutionary history was estimated on basis of General Time Reversible model with 1000 bootstrapping replicates. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.3531)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 0.0000% sites). The outgrouping of the tree was done using the sequences of *Staphylococcus aureus*, *Escherichia coli*, *Rhodospseudomonas palustris* CGA009 and *Chlorobium tepidum* TLS.

3.3.2. The complete set of β carboxysome proteins found in cyanobacteria only

The BLASTP results revealed hits from several of the phyla included in the study. The hits with $E \leq 10^{-5}$ were considered significant and used for carrying out the analysis. The protein encoded by *cemK* possess a BMC domain (IPR000249). The CemK protein has been reported to form hexameric units that assemble to form the microcompartments involved in biochemical pathways such as the primary dark reaction of photosynthesis viz. RuBisCO catalyzed carboxylation. The BLASTP results for CemK have been listed in **Table 3A-S**. The analysis revealed no significant homologs for this protein in Archaea, Aquificae, Caldisei, Chlamydiae, Chrysiogenetes, Deferribacteres, Deinococcus-Thermus, Dictyoglomi, Elusimicrobia, Nitrospinae, Nitrospirae, Epsilon proteobacteria, Thermodesulfobacteria and Thermotogae. Among the cyanobacteria almost all were found to possess the CemK protein as the hits obtained were of very low expect value and appreciably high identity. Further, in the Actinobacteria and Proteobacteria phyla, several significant hits were noted, which include both CemK protein as well as other proteins (proteins with function other than formation of carboxysomes but with similar domain). Other phyla in which other proteins formed the major portion of the hits obtained include Fibrobacteres/Acidobacter, Firmicutes, Fusobacteria, Planctomycetes and Synergistetes. Further, some phyla, namely, Bacteroidetes Chlorobi, Chloroflexi, Gemmatimonadetes, Spirochaetes and Tenericutes gave hits of proteins other than CemK only.

The homologs of CemK obtained in the analysis, apart from CemK itself include, Ethanolamine utilization protein EutM, Ethanolamine utilization protein EutK, Propanediol utilization protein PduA, Propanediol utilization protein PduJ, Propanediol utilization protein PduT and Biotin-(acetyl CoA carboxylase) ligase.

Seven out of 23 reported genes for Pdu microcompartment are homologous to the carboxysome shell proteins (Chen et al., 1994). The Pf00936 domain is found to have a ~80 residue sequence common in carboxysome, Eut and Pdu microcompartment proteins (Kerfeld et al., 2010). Electron microscopic studies on *Salmonella enterica* Serovar *Typhimurium* LT2, *Klebsiella oxytoca* (Shively et al., 1998), and *Lactobacillus reuteri* DSM 20016 (Sriramulu et al., 2008) grown on 1,2 propanediol, show formation of several polyhedral bodies bearing resemblance to carboxysomes in shape, size and electron density. The proteins found to have structural and functional analogy to the CemK and CemO proteins of carboxysomes are CsoS1A-D and CsoS2 in α carboxysomes, PduB, PduA, PduJ, PduK, PduU and PduT in Pdu microcompartments and EutL, EutM and EutS in Eut microcompartments (Frank et al., 2013).

The CcmL protein possesses the BMC domain and forms pentameric units constituting the vertices of the microcompartments in various Eubacteria. The BLASTP results for CcmL are

mentioned in Table 3B-S. The analysis did not bring forth any significant CcmL homologs from Archaea, Aquificae, Armatimonadetes, Caldiseptica, Chlamydiae, Chrysiogenetes, Deferribacteres, Dictyoglomi, Elusimicrobia, Nitrospinae, Nitrospirae, Epsilon proteobacteria, Thermodesulfobacteria and Thermotogae. The search from the cyanobacteria protein database revealed several significant hits which include both CcmL protein as well as proteins such as Ethanolamine utilization protein EutN. Similarly, α proteobacteria and β proteobacteria gave equal number of significant hits for CcmL and other proteins i.e., EutN. The proteins in Pdu and Eut microcompartments responsible for forming the vertices are PduN and EutN, respectively and hence share appreciable sequence homology.

The hits from phyla Actinobacteria, Firmicutes, Fusobacteria, Planctomycetes and Synergistetes gave majorly proteins other than CcmL as results of BLASTP. Furthermore, Bacteroidetes Chlorobi, Chloroflexi, Fibrobacteres/Acidobacter, Gemmatimonadetes, Gamma proteobacteria, Spirochaetes and Tenericutes do not possess CcmL protein, instead they have a hypothetical protein with significant homology to CcmL. The hypothetical protein could either be CcmL protein or a protein involved in formation of microcompartment for some pathway other than Calvin cycle of photosynthesis.

The domains present in CcmM are Lbh gamma CA and RuBisCO small subunit (SSU) domain. The BLASTP results for CcmM are mentioned in Table 3C-S. All the BLAST hits obtained for CcmM in Archaea were of very low expect value and were of proteins other than CcmM like carbonic anhydrases, acetyltransferases or ferripyochelin binding protein. The BLAST hits from cyanobacteria phylum were CcmM proteins from other cyanobacteria as well as several proteins with significant homology i.e. carbonic anhydrase and cytochrome C biogenesis protein. No complete homologs of CcmM were found in proteobacteria. All the hits obtained were proteins other than CcmM i.e., Hexapeptide repeat containing transferase, carbonate dehydratase, acetyltransferase, sulfate permease, protein YrdA, Phenylacetic acid degradation protein PaaY, UDP glucosamine N acyltransferase, etc. Further, Firmicutes were found to have appreciable homologs other than CcmM. The hits obtained were carbonic anhydrase, carbonate dehydratase, transferase or hypothetical protein with partial query coverage and appreciable percent identity. This implies that the homolog identified as hypothetical protein is not complete CcmM protein, but the proteins have the domains in common.

The several of the hits obtained are carbonic anhydrase as the N-terminal of CcmM has γ carbonic anhydrase domain. The *Synechococcus elongatus* PCC 7942 CcmM N-terminal bears 60% amino acid similarity to γ carbonic anhydrase of *Methanosarcina thermophila* and possibly forms a similar structural arrangement. Apart from this, N-terminal region of the CcmM (M58) has glutamine and three histidine residues conserved which are required for

Zn²⁺ coordination and carbonic anhydrase (Ludwig et al., 2000; Alber and Ferry 1996; Kisker et al., 1996).

Apart from these, all phyla under study were observed to possess homologs other than CcmM protein. The various homologs obtained include carbonic anhydrase, acetyltransferase, ferrityochelin binding protein, siderophore binding protein, phenylacetic acid degradation protein PaaY, 2,3,4,5-tetrahydropyridine-2,6-dicarboxylate-N-acetyltransferase, hexapeptide repeat containing protein, cytochrome C biogenesis protein CcmM, NUDIX protein, UDP-3-O-[3 hydroxymyristoyl] glucosamine N-acyltransferase, protein YrdA and hypothetical protein.

The CcmN proteins possess LbetaII domain i.e. left handed parallel β helix domain. These proteins contain three imperfect tandem repeats of a hexapeptide repeat motif. Proteins containing hexapeptide repeats are often enzymes showing acyltransferase activity. Very few CcmN homologs were found in the various phyla under investigation (Table 3D-S). Archaeal hits with significant homology were few and the ones obtained were hexapeptide repeat containing protein, carbonic anhydrase or acetyltransferase. Several phyla were found to have no significant homologs, which include, Aquificae, Armatimonadetes, Chrysiogenetes, Deferribacteres, Dictyoglomi, Elusimicrobia, Fibrobacteres/Acidobacter, Nitrospinae, Nitrospirae, Plantomycetes, ϵ proteobacteria, Spirochaetes, Synergistetes, Tenericutes, Thermodesulfobacteria and Thermotogae. Only cyanobacteria and firmicutes gave hits of CcmN protein protein and other proteins which include hypothetical protein, hexapeptide repeat containing transferase, transferase, 2,3,4,5-tetrahydropyridine-2,6-dicarboxylate-N-acetyltransferase and carbonic anhydrase. The analysis of CcmM and CcmN gave very few direct hits. Most of the hits obtained were due to homology with the LbetaII-gamma CA and LbetaII domains for CcmM and CcmN, respectively.

Duplication is one of the major driving forces for creation of new genes (Lynch and Cronery 2000). Once duplicated, the gene can evolve a novel function by sequence divergence or by combination with other domains, under a strict selection pressure, to form multidomain proteins (Vogel et al., 2004). The most common molecular mechanisms responsible for formation of multidomain proteins include non-homologous recombination (also called domain shuffling) (Vogel et al., 2004), fusion of genes (Enright et al., 1999; Marcotte et al., 1999) and fission of genes (Snel et al., 2000; Yanai et al., 2002). At the time of emergence of CCM in primitive cyanobacteria, the pre-existing domains possibly underwent certain evolutionary mechanism viz. domain shuffling, gene fusion or gene fission in order to develop proteins such as CcmM and CcmN, which are crucial for a functional low environmental Ci phenotype. This theory could be supported by the fact that CcmM and

CcmN do not have any complete homologs in any of the reported phyla available at NCBI, apart from the domain centric homologs.

The CcmO protein contains two repeats of the BMC domain. No significant homologs were found in the Archaeal phyla. In cyanobacteria phyla, hits were obtained for both CcmO and other proteins which include hypothetical protein and EutM (Table 3E-S). The analysis revealed no significant hits from several phyla of Eubacteria, namely, Aquificae, Armatimonadetes, Chrysiogenetes, Deferribacteres, Deinococcus-Thermus, Dictyoglomi, Elusimicrobia, Nitrospinae, Nitrospirae, Thermodesulfobacteria and Thermotogae. The phyla which gave hits other than CcmO, include Bacteroidetes:Chlorobi, Chlamydiae, Chloroflexi, Gemmatimonadetes, Spirochaetes and Tenericutes. The hits obtained in these phyla include EutM, carboxysomes shell protein, hypothetical protein, PduI, microcompartment protein and propanediol utilization protein. Very few direct homologs were obtained in Actinobacteria, Fibrobacteres/Acidobacter, Firmicutes, Fusobacteria, Planctomycetes and γ proteobacteria.

The phyla found to possess homologs for CcmK, CcmI, and CcmO include actinobacteria, fibrobacteres, fusobacteria, planctomycetes, α proteobacteria, β proteobacteria, gamma proteobacteria, cyanobacteria and firmicutes. It should be noted that none of the phyla except cyanobacteria were found to possess any homologs for proteins CcmM and CcmN.

The organisms other than cyanobacteria found to possess CcmK or proteins with similar function in α carboxysomes include *Acidimicrobium ferrooxidans* (Actinobacteria), *Leptotrichia trevisanii* sp oral taxon 879 (Fusobacteria), *Nitrobacter* sp Nb-311A, *Nitrobacter winogradskyi*, *Nitrobacter hamburgensis*, *Bradyrhizobium* sp BTAi1, *Bradyrhizobium* sp ORS278, *Bradyrhizobium* sp STM 3809 (Alpha proteobacteria), *Comamonadaceae bacterium* B1, *Comamonadaceae bacterium* H1, *Comamonadaceae bacterium* B1CA1-1, *Ferrovum myxofaciens*, *Thiobacillus thioparus*, *Thiobacillus denitrificans*, *Nitromonas eutropha*, *Tepidimonas taiwanensis*, *Burkholderiales bacterium* GJ-E10 and *Thiomonas arsenitoxydans*.

The organisms apart from cyanobacteria found to possess CcmI, include *Propionibacterium acedifaciens* (Actinobacteria), *Candidates Solibacter usitatus* (Fibrobacteres), *Leptotrichia* oral taxon 879 str F0557 (Fusobacteria), *Rhodopirellula maiorica*, *Rhodopirellula europaea*, *Rhodopirellula baltica*, *Rhodopirellula sallentina*, *Blastopirellula marina* DSM 3645 (Planctomycetes) and *Schlesneria paludicola* (Plantomycetes).

The genome of the organisms possessing CcmK/I, homologs were screened for the presence of other CCM proteins. *Acidimicrobium ferrooxidans* (Actinobacteria), *Bradyrhizobium* sp ORS 278 and *Bradyrhizobium* sp BTAi1 (α proteobacteria) were found to possess RuBisCO subunit encoding genes and CCM encoding genes adjacent to each other. Although these

organisms have shell proteins similar to those present in β -cyanobacteria, but they are devoid of CcmM and CcmN homologs which are important for formation of functional carboxysomes. These bacteria have been reported to have α carboxysomes which have a different set of proteins apart from the shell proteins for carboxysome structure formation (α carboxysomes).

Acidimicrobium ferrooxidans is a sulfur oxidizing Actinobacteria which is moderately thermophilic and acidophilic. It has been reported to be autotrophic as the RuBisCO present is found to be active in carbon dioxide fixation. Further, it has been proposed that the RuBisCO encoding genes present in *Acidimicrobium ferrooxidans* have been acquired from *Acidithiobacillus ferrooxidans* or its ancestors. *Bradyrhizobium* sp ORS 278 and *Bradyrhizobium* BTA11 are tropical, photosynthetic bacteria which fix nitrogen and are symbionts of legumes.

The α carboxysomes are found in α -cyanobacteria (Badger et al., 2002), sulfur oxidizing bacteria of α proteobacteria (Kelly and Harrison 1989), β proteobacteria, γ proteobacteria (Kelly and Wood 2000), Actinobacteria (Rae et al., 2013), nitrifying bacteria (Rae et al., 2013) and cyanobacterial chromatophore of *Paulinella chromatophora* (Marin et al., 2007). While as also observed in the present analysis there are no reports for presence of the β -carboxysomes or its gene repertoire in any other group of organisms except the β -cyanobacteria (Table 3B and Figure 3B).

Table 3B: The distribution of the β carboxysome proteins across phyla

| Phyla with no CCM protein homologs | Phyla with homologs for CcmK, CcmL and CcmO | Phyla with domain homologs for CcmM and CcmN domains | Phyla with homologs for all CCM (β carboxysome) proteins |
|---|---|---|---|
| Armatimonadetes ϵ Proteobacteria Thermodesulfobacteria | Actinobacteria Fibrobacteres Fusobacteria Planctomycetes α Proteobacteria β Proteobacteria γ Proteobacteria Firmicutes | Archeae Aquificae Caldiserica Chlamydiae Chrysiogenetes Deferribacteres Deinococcus Dictyoglomi Elusimicrobia Gemmatimonadetes Nitrospinae Nitrospirae Spirochaetes Tenericutes Thermotogae | Cyanobacteria |

The analysis reveals that the carbon concentrating mechanism microcompartment proteins, the ethanolamine utilization pathway shell proteins and the propanediol utilization mechanism shell proteins are evolutionarily linked to each other. The phylogenetic analysis of BMC domain proteins from carboxysomes, Pdu BMC, Eut BMC and grp-type BMC shows segregation of the carboxysomal shell proteins from the rest and moreover, the occurrence of BMCs other than carboxysomes is not uniform among organisms of a particular phyla or even genera (Abdul-Rahman et al., 2013). These observations suggest that these BMC types are of recent evolutionary origin (Abdul-Rehman et al., 2013). Further, Cai et al., (2015), revealed that CsoS1 protein of *Prochlorococcus marinus* str. MIT9313 (Pro9313) shares only 44% similarity to CcmK2 of *Synechococcus elongatus* PCC 7942 and 54% and 60% similarity to EutM (Tanaka et al., 2010) and PduA (Crowley et al., 2010), respectively. Cai et al., (2015) further report that on the basis of phylogenetic analysis, shell proteins of α and β -cyanobacteria are much different from each other and comparatively, are more closely related to major Pdu and Eut shell proteins. These observations suggest independent emergence of CCM in α and β -cyanobacteria after their divergence. Moreover, the α carboxysome shell proteins are closely related to the BMC shell proteins of other microcompartment enclosed pathways prevalent in proteobacteria than those of β -carboxysomes.

As mentioned before the CcmK, CcmL and CcmO proteins contain the BMC domain and are responsible for forming the shell of the carboxysomes. The analogous proteins in α -cyanobacteria that is the BMC domain proteins include CsoS1A, CsoS2A, CsoS2B, CsoS1B, CsoS1C, CsoS4A and CsoS1D, which are responsible for the carboxysome shell formation. Further, the annotation of many of the BMC domain proteins is incomplete and they are named as microcompartment proteins or carboxysome shell proteins. In order to identify a BMC protein to be CcmK or CsoS1A, the genome of the particular organism needs to be tapped for presence of other carboxysome proteins on the basis of which the kind of carboxysome present in that particular organism can be identified. The similar exercise when performed for the organisms found to have only the shell proteins and not CcmN and CcmM, it was found that they possess the α carboxysome proteins. Further, the phyla with hits for CcmN and CcmM (other than cyanobacteria) possess proteins with similar domains and not the complete proteins. The carboxysome and BMC (Eut, Pdu and Grp) shell proteins share pfam03319 (for CcmL in carboxysomes) and pfam00936 (for CcmK in carboxysomes) protein domains and hence the phyla found to possess only the shell proteins of CCM and not the complete carboxysome proteome, contain other BMCs. The above observations reveal that β -carboxysomes are found in β -cyanobacteria only while the α carboxysomes are found in α -cyanobacteria and few other phyla of Eubacteria, as discussed by Rae et al., (2013).

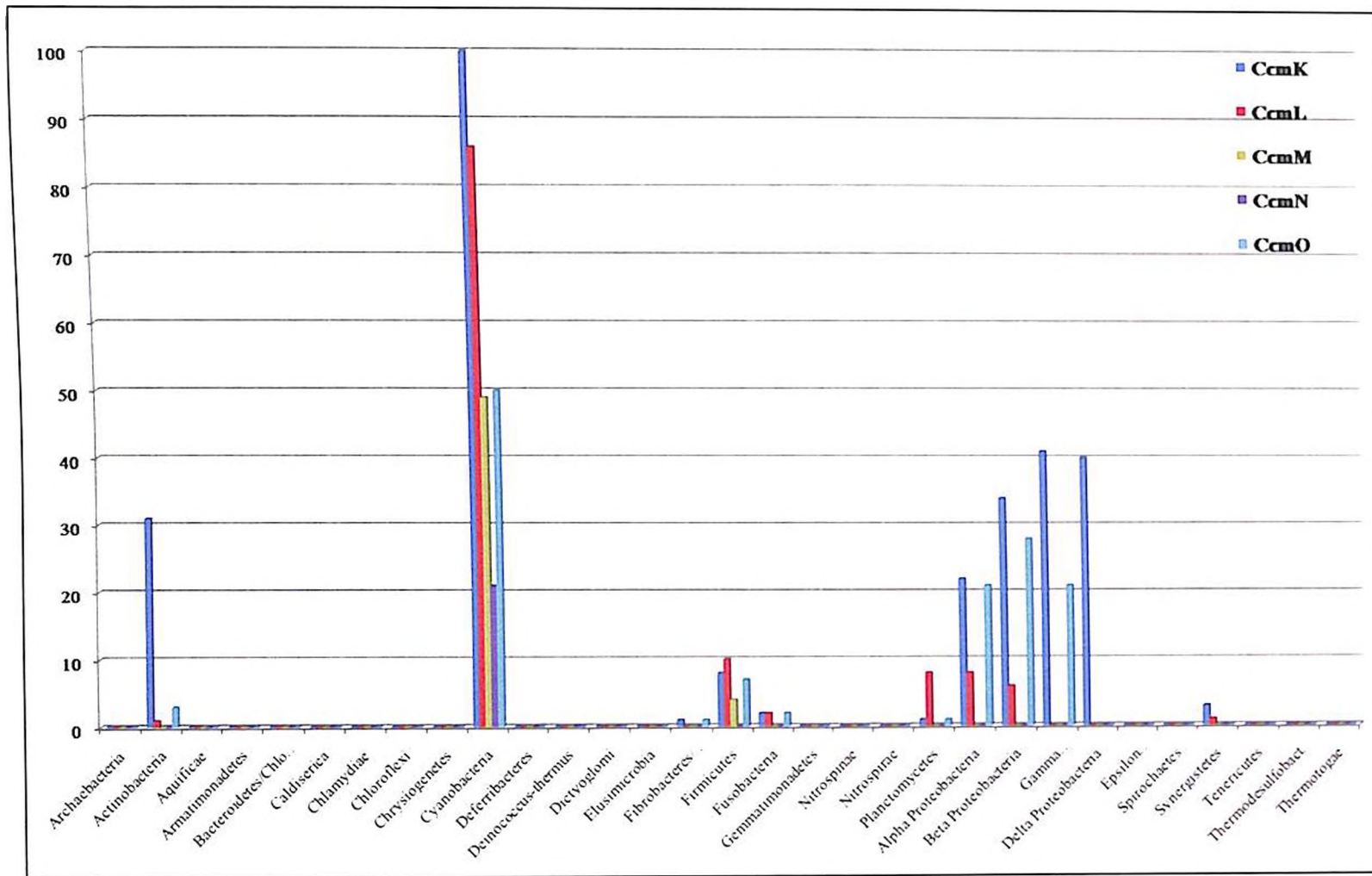


Figure 3B: The distribution of β carboxysome proteins across various phyla of Eubacteria. The X-axis shows the 31 eubacterial phyla tapped for CCM proteins using the BLASTP tool available online and the Y-axis shows the number of hits obtained from each of the phyla per first 100 searches or the total number of hits (either of which is more) for each of the CCM proteins using *G. violaceus* PCC 7421 CCM proteins as query sequence

3.3.3. The CCM operon was formed *in-situ* and passed on by vertical succession

The carboxysome operon phylogenetic tree (Figure 3C) shows two major clades constituting α -cyanobacteria and β -cyanobacteria, respectively. This result is in accordance with the expectation, since both α carboxysomes and β -carboxysomes are formed of different set of proteins with the exception of BMC proteins, which are found in both. The subclades formed for α -cyanobacteria are in accordance with the 16S rRNA phylogeny, validating the fact that the α carboxysome operon was formed *in-situ* and passed in vertical succession. The topological arrangement of β -cyanobacteria is at some variance from the 16S rRNA based phylogeny. *Thermosynechococcus elongatus* BP-1 clustering with *Acaryochloris marina* MBIC 11017 and *Cyanothece* sp PCC 7425 in 16S rRNA phylogram is found to be closer to early diverging group of cyanobacteria viz. *G. violaceus* PCC 7421, *Synechococcus* sp JA-2-3Ba*(2-13) and *Synechococcus* sp JA-3-3Ab. Further, *Trichodesmium erythraeum* IMS101 and *Microcystis aeruginosa* NIES39, forming a sub-clade in 16S rRNA analysis are clustered separately in the carboxysome operon phylogeny. An important observation that can be made from the topology of carboxysome operon based tree is that although the arrangement of the various organisms is at some variance from that in case of 16S rRNA tree but the clustering of the organisms in the sub-clades formed is conserved. Further, although the early diverging forms cluster together, but it would not be correct to consider them to be the first forms to have acquired CCM.

The evolution of the operon is based upon two conflicting theories of *in-situ* formation for co-regulation and horizontal gene transfer of functionally related genes. Roughly 50% of prokaryotic genome is reported to be operonic (Price et al., 2006) evolved under selection pressure. Memon et al., (2013) hypothesize that the most conserved and moderately conserved operons are formed *in-situ* and inherited vertically, on the basis of comparative phylogenetic analysis between the operon under study and the 16S rRNA sequence. An appreciable similarity between the phylogenetics based on 16S rRNA sequence and any other genetic sequence, if observed is possible only if the later is passed on by vertical inheritance (Memon et al., 2013).

During the study it was noted that the *csa* operon constitutes the carboxysome encoding genes as well as the RuBisCO encoding genes (Figure 3D), while most of the β -cyanobacterial carboxysomes constitute only the CCM genes. Only five of the β -cyanobacteria, available in Kazusa Genome Resource namely, *Synechococcus elongatus* PCC 6301, *Synechococcus elongatus* PCC 7942, *Synechococcus* sp PCC 7002, *Microcystis aeruginosa* NIES 843 and *Cyanothece* sp PCC 7424 were found to have RuBisCO genes in tandem with the carboxysome encoding genes. The carboxysome operon phylogenetic analysis was done using the complete operon (i.e. with RuBisCO genes if present in tandem), while the individual β -

cyanobacteria phylogenetic analysis was done using both kinds of sequences i.e. with RuBisCO genes and without RuBisCO genes (**Figure 3D**).

Further, while most of the β -cyanobacteria were found to have *ccmO* as a part of the operon, several had *CcmO* encoding gene distant from the other CCM encoding genes. Moreover, while *ccmK* *ccmK1* *ccmK2* were found to be located adjacent to the rest of CCM genes, *ccmK3* *ccmK4* if present were always found to be located at different loci of the genome (**Table 3F-S**).

It must be noted that the CCM genes, present on direct or complementary strand, were found to have a conserved arrangement viz. the order of the genes is *ccmO*, *ccmX*, *ccmM*, *ccmL* and *ccmK*. Among the organisms considered for the study, only in *Trichodesmium erythraeum* IMS101 there is an insertion of a hypothetical protein encoding gene in-between *ccmM* and *ccmX*, which hence forms a separate branch in the phylogenetic tree of CCM operon of β -cyanobacteria (**Figure 3D**). Among the three major clades after branching of *Synechocystis* sp PCC 6803 and *Trichodesmium erythraeum* IMS101, the RuBisCO genes containing operons cluster together with *Cyanothece* sp ATCC 51142 and *Cyanothece* sp PCC 7425 (which do not have RuBisCO genes in tandem with *ccm* genes). The other two sub-clades are formed on the basis of the differences/homology between the sequences alone. In order to validate this, the phylogenetic analysis for CCM operon was repeated with a dataset without the RuBisCO genes included in the operon sequence (**Figure 3D**). *Trichodesmium erythraeum* IMS101 was again found to form a separate branch due to insertion of a hypothetical protein in the operon. Among the three major sub-clades formed, some were observed to have conserved their topological position. The change in the topological position of *Acartyochloris marina* MBIC11017 and *Cyanothece* sp PCC 7425 in the *ccm* based-RuBisCO independent phylogeny shows them to be more closely related to early diverging cyanobacteria.

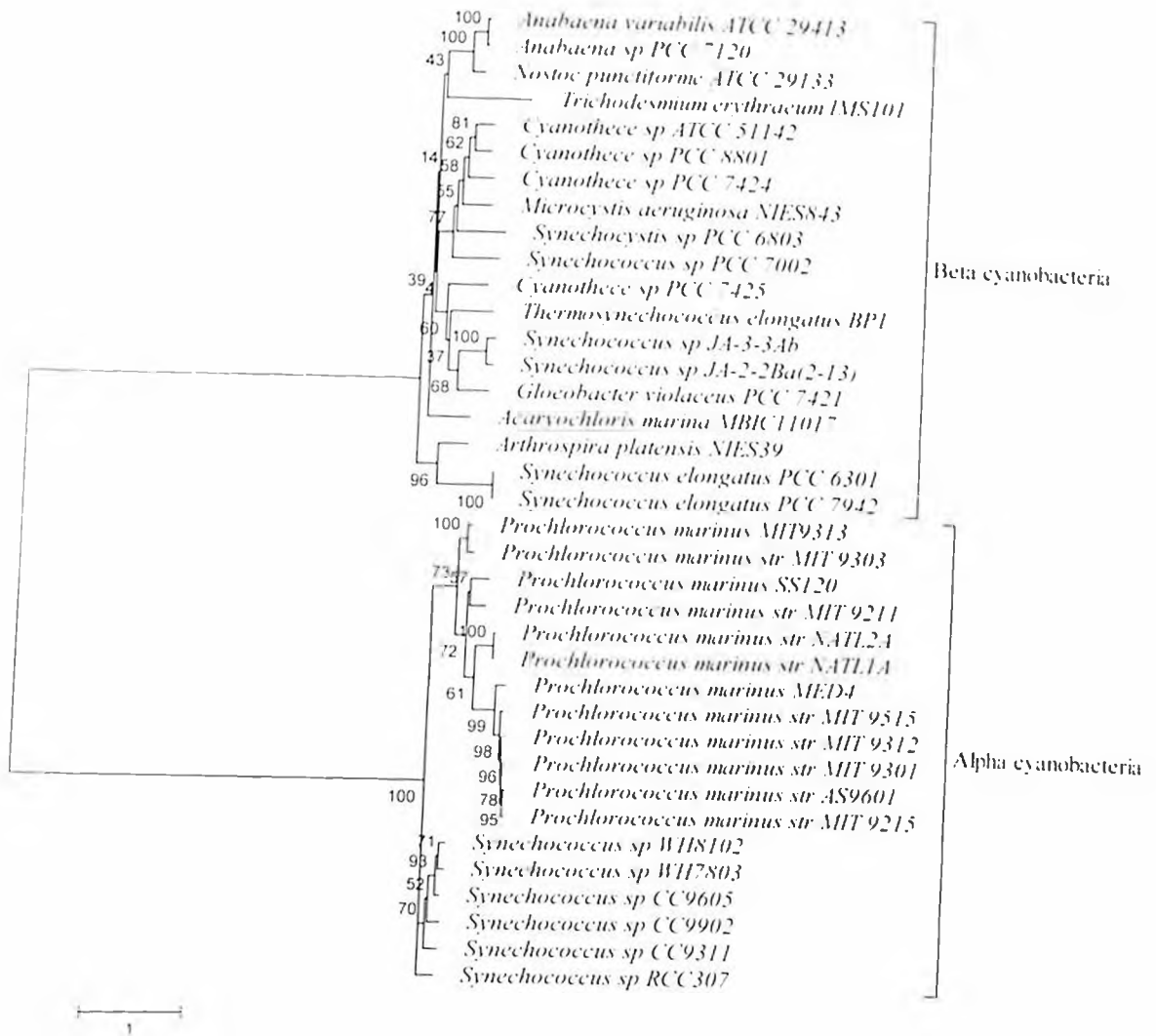


Figure 3C: Maximum Likelihood phylogenetic representation of cyanobacteria based on carboxysome operon sequence. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. The evolutionary distances were computed using the General Time Reversible model with 1000 bootstrapping replicates. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.8049)). The rate variation model allowed for some sites to be evolutionarily invariable ([-I], 0.0000% sites).

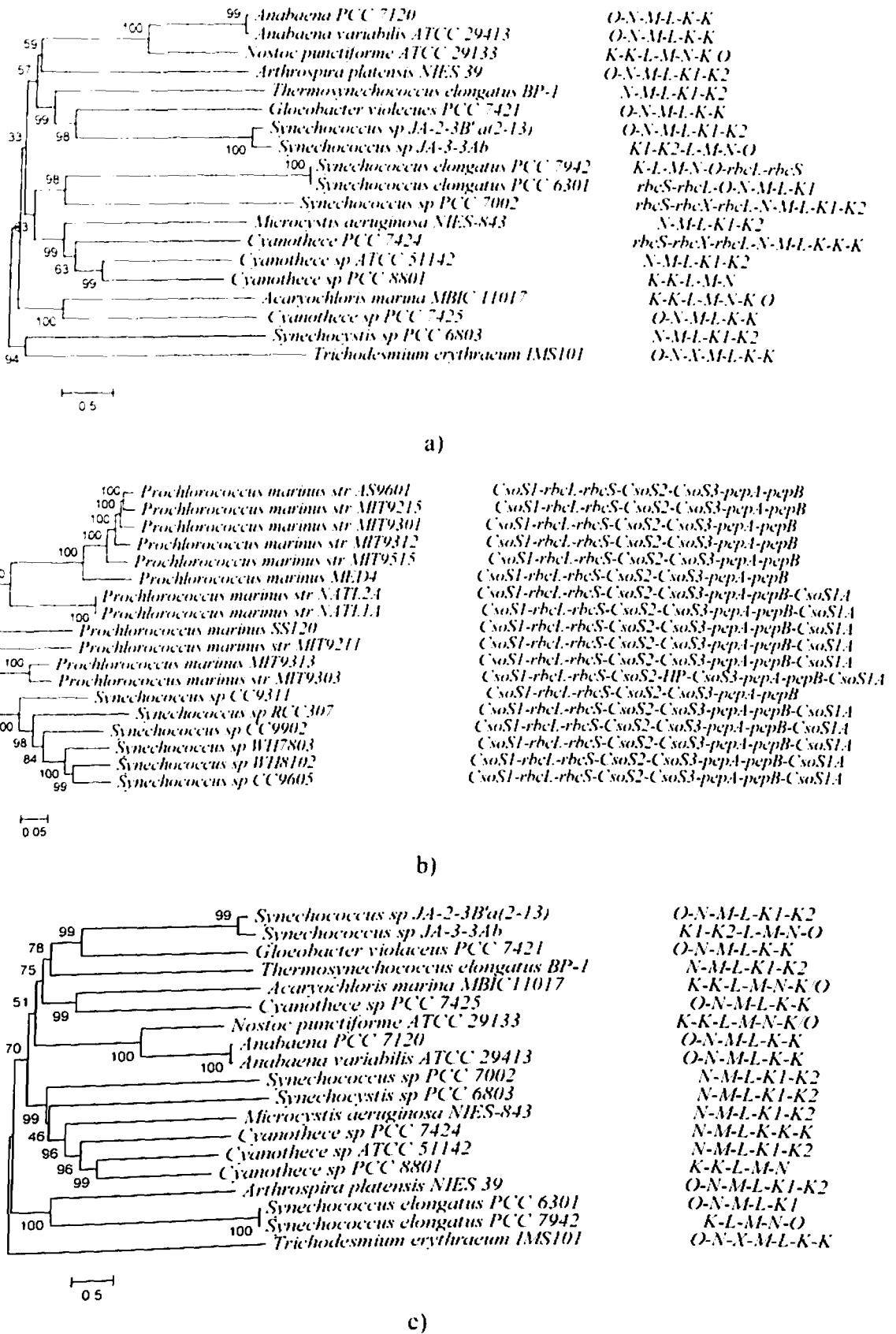


Figure 3D (a) Phylogenetic analysis of α -cyanobacteria on basis of *cso* operon; (b) Phylogenetic analysis of β -cyanobacteria based on *ccm* operon with (b) and without (c) RuBisCO genes.

3.3.4. Phylogenetic analysis of CCM proteins

The CcmK homologs are found in various phyla of Eubacteria, namely proteobacteria (72.8%), firmicutes (20.3%), cyanobacteria (4.5%), actinobacteria (0.7%), fusobacteria (0.6%), planctomycetes (0.2%), synergistales (0.1%), verrucomicrobia (0.1%) and eukaryotes (0.01%). *Paulinella chromatophora*, a eukaryote (Rhizaria) possess BMC domain proteins. Further, *Plutella xylostela*, a metazoan is also having a 1194 amino acid uncharacterized protein with PRP38 assoc superfamily domain and CcmK domain.

It is evident from the genomic data available that CcmK exists in more than one copy in all β carboxysome forming organisms (Table 3F-S). The minimum number of *ccmK* sequences observed till date is two and the maximum number is six. As mentioned earlier, atleast one *ccmK* gene exists with the remaining *ccm* operon while the rest are found elsewhere in the genome. In a study conducted by Rae et al., (2012) and Cameron et al., (2013), it was found that the *Synechococcus elongatus* PCC 7942 has three *ccmK* genes viz. *ccmK2*, *ccmK3*, *ccmK4*, out of which *ccmK2* deficient mutants showed carboxysome less phenotype, signifying the importance of *ccmK2*. The knockout of *ccmK3* and/or *ccmK4* on the other hand did not prevent the formation of the functional carboxysomes. This shows that possibly, the *ccmK* present as a part of the *ccm* operon are minimal requirement for a functional carboxysome assembly while the rest of the *ccmK* gene complements encode proteins with secondary roles, still unelucidated. The clustering of various β -cyanobacteria CcmK amino acid sequences shows that they are orthologs i.e. encoded by gene originating from common ancestor by speciation and not a result of individual duplication events in most cases (Figure 3E(i)). Duplication did take place, but in a primitive species and the gene sets hence formed were passed on by vertical succession. This is clear from the phylogenetic analysis as the CcmK protein encoded by genes adjacent to *ccmL* is clustered with CcmK protein from other cyanobacteria than to its own CcmK. CcmK encoded by gene adjacent to *ccmL* cluster separately from CcmK encoded by gene second adjacent to *ccmL* (Figure 3(ii) and Table 3F-S). Further, the relation between the various cyanobacteria on basis of CcmK sequence is analogous to that observed in case of the rest of the *ccm* proteins as well as 16S rRNA sequence, suggesting evidence for vertical succession.

It should be noted that CcmK proteins encoded by genes existing separately in the genome away from *ccm* operon, in case of *Thermosynechococcus elongatus* BP-1 and *Cyanothece* sp PCC 7425 form a distinctly separate clade, showing separation from the other sequences. Further, CcmK3 and CcmK4 sequences from these organisms show more similarity among each other than with any other sequences used in analysis. This result suggests an event of duplication of *ccmK* gene after divergence from the primitive cyanobacteria as *T. elongatus* BP-1 and *Cyanothece* sp PCC 7425 have been observed to be close relatives of early

diverging cyanobacteria. Further, these duplication events were not passed on in more recent cyanobacteria, as the more recent cyanobacteria have been observed to possess *CcmK3* and *CcmK4* encoding genes adjacent to each other. Hence, the duplication events in *T. elongatus* BP-1 and *Cyanothece* sp PCC 7425 leading to form *CcmK3* and *CcmK4*, are independent of each other. Hence, at a particular point of evolution of cyanobacteria, after the divergence of the primitive cyanobacteria, *ccmK* duplications took place leading to greater number of *ccmK* homologs.

If we follow the trend of emergence of cyanobacteria according to 16S rRNA analysis considering *G. violaceus* PCC 7421 to be the ancestor and correlate it with *ccmK* gene occurrence in various cyanobacteria, we observe an increase in number of *ccmK* genes with time (**Figure 3E(ii)** and **Table 3F-S**). The primitive cyanobacteria viz. *G. violaceus* PCC 7421, *Synechococcus* sp JA-2-3B'a(2-13) and *Synechococcus* sp JA-3-3Ab have two *ccmK* genes each, and *T. elongatus* BP-1 and *Cyanothece* sp PCC 7425 as mentioned earlier have two *ccmK* genes closely related to *G. violaceus* PCC 7421, while the outliers forming a separate clade. This shows that the event of duplication occurred after speciation of the recent cyanobacteria from the early ones. *Acarlyochloris marina* MBIC 11017, although being comparatively primitive has six *ccmK* genes. The *AMI_0656 ccmK* sequence is more closely related to *ccmK3* of *S. elongatus* PCC 6301 and *S. elongatus* PCC 7942. Further, *AMI_5778* is clustered close to *AMI_0655* and *AMI_3280*, showing a recent duplication event, i.e. initially it had four *ccmK* genes but recently it acquired another two by duplication events. *Cyanothece* sp PCC 7425, is the most recent according to 16S rRNA analysis and it has five *ccmK* genes. The phylogenetic analysis shows that the two *ccmK* genes apart from the ones present in operon have been obtained by vertical succession while the fifth one lying close to the *ccmK* genes present in operon is a result of recent duplication (**Figure 3E(i) and (ii)**; **Table 3F-S**).

The *CcmL* homologs are found in several eubacterial phyla. The maximum *CcmL* homologs are found in γ proteobacteria (77%), followed by Firmicutes (15.5%), Cyanobacteria (3.6%), Actinobacteria (0.9%), Planctomycetes (0.6%), Fusobacteria (0.6%), Verrucomicrobia (0.2%), Synergistales (0.1%), Archaea (0.01%) and eukaryotes (0.1%). The Archaea with protein containing the EutN/*ccmL* domain is *Haloquadratum walsbyi*. Further, there are several *CcmL* homologs in kingdom Animalia. *Toxocara canis*, a nematode (metazoan) has a hypothetical protein with EutN/*ccmL* domain. The function of the protein is not yet elucidated. *Galeopterus variegates*, a chordate (Mammalia) is also found to possess a predicted-proline rich protein 27 with EutN/*ccmL* domain. Moreover, certain fungi, viz. *Paulinella chromatophora*, *Neurospora tetrasperma*, *Pisolithus inctorius* Marx 270, *Rhizoctonia solani* AG-a, *Theileria orientalis* strain shintoku and *Enterocytozoon bieneusi* H348 have protein with domain present in the protein under study.

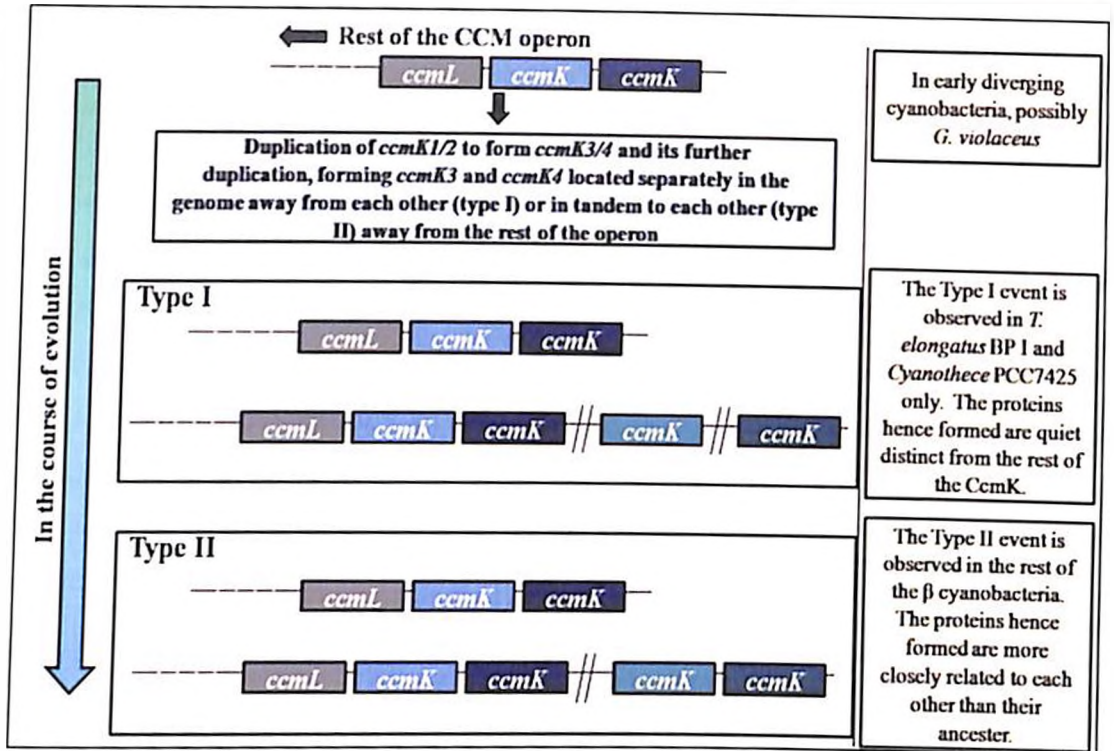


Figure 3E(i): The duplication events of *ccmK* in the course of evolution beta cyanobacteria, inferred from the degree of similarity between the *ccmK* homologues.

The various CcmL homolog sequences were retrieved from NCBI and analyzed phylogenetically (**Figure 3D**). The phylogenetic analysis of the CcmL protein shows formation of two major clusters viz. I and II (**Figure 3D**). Clade I has two sub clades constituting proteobacterial sequences in IA and β -cyanobacteria sequences in IB. The clade II branches into IIA constituting α -cyanobacteria and IIB constituting proteobacteria.

Further, the α cyanobacteria were found to have two proteins with CcmL domain i.e., peptide A and peptide B. It should be noted that such clustering can be explained by the fact that although cyanobacteria and other Eubacteria possess BMC and hence CcmL, the role of the later could be partly variable in both set of organisms. Hence, although the domain responsible for forming vertices of microcompartments is common in the two set of organisms, there are several evident differences in the sequences that have role specific to the type of microcompartment they constitute.

The CcmM protein was distinctly found in cyanobacteria apart from a few sequences from Firmicutes (**Figure 3G**). The sequences from Firmicutes and other unclassified bacteria were from organisms namely *Caldisaliniibacter kiritimatiensis*, *Bacillaceae bacterium* MTCC 10057, *Bacillaceae bacterium* MTCC 8252, *Bacillus thermotolerans* and *Haloplasma contractile* SSD-17B. It should be noted that although these sequences have been annotated as CcmM proteins and have the Lbh_CA domain, but they are devoid of the RuBisCO SSU domains, and hence were not included in phylogenetic analysis. Further, many cyanobacteria

CcmM proteins having CA and RuBisCO SSU domains were wrongly annotated as cytochrome C biogenesis proteins.

The CDD search of the CcmM proteins reveals the SSU repeats at the C-terminal and annotates a Bitscore to define the degree of identity of the region found in the query with the bonafide domain. **Table 3D** shows the bitscore for each of the SSU repeats (Red-maximum, followed by blue, green and yellow for minimum). The SSU1 and SSU4 domains (numbered in order of their appearance from N to C terminal) were found to be closest to the bonafide domain as compared to SSU2 and SSU3. Although, the present information does not signal any trend of origin duplication of the repeats, but the greater degree of relatedness of SSU1 and SSU2 directs to their early origin in CcmM protein, followed by origin of SSU2 and SSU3 by duplication events.

The increase of SSU repeats post speciation is further supported by the fact that there exists a weak correlation between phylogeny based on N terminal of CcmM and C terminal of CcmM (**Table 3C**). The N terminal of protein (~250 amino acids) is largely conserved, while the C terminal has mosaic regions of conservation (mentioned in chapter 5). The higher correlation between phylogeny based on CcmM and SSU domain as compared to that with CA domain, can be explained by the fact that the SSU repeat region occupies ~60% of the protein, determining the phylogentic topology of the CcmM protein.

| SSU1 | SSU2 | SSU3 | SSU4 | SSU5 |
|------|------|------|------|------|
| 134 | 119 | 133 | 142 | |
| 151 | 173 | 139 | 150 | |
| 135 | 135 | 144 | 145 | |
| 135 | 129 | 147 | 145 | |
| 131 | 137 | 151 | 153 | |
| 137 | 133 | 152 | 158 | |
| 143 | 141 | 136 | | |
| 143 | 141 | 137 | | |
| 146 | 144 | 148 | 144 | 146 |
| 156 | 132 | 143 | 135 | |
| 143 | 152 | 124 | 138 | |
| 130 | 149 | 134 | 115 | |
| 143 | 133 | 131 | 150 | |
| 141 | 130 | 129 | 155 | |
| 143 | 146 | 148 | 142 | |
| 142 | 136 | 142 | | |
| 130 | 146 | 142 | 137 | |
| 139 | 147 | 145 | | |

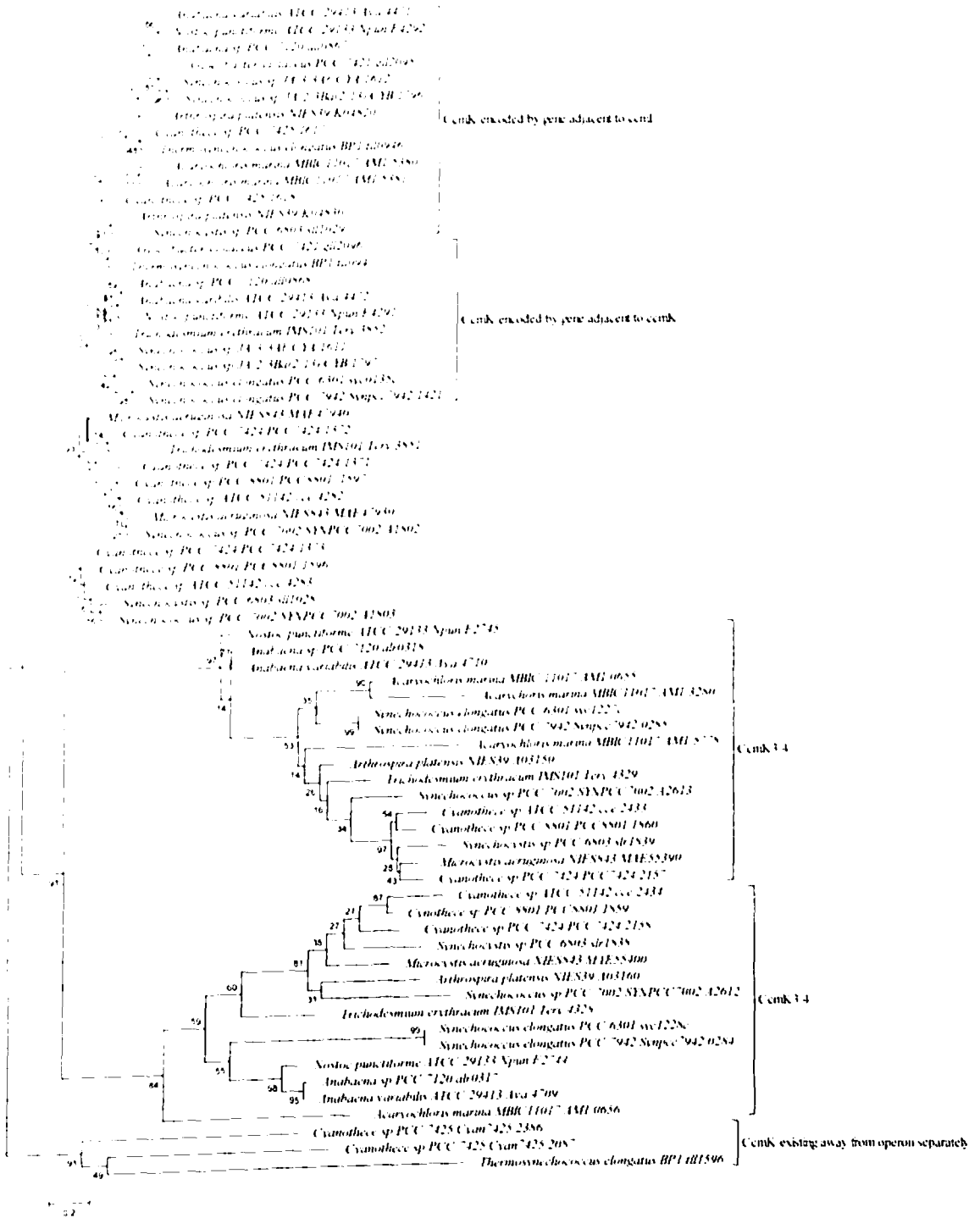


Figure 3E(ii): The evolutionary history of CcmK amino acid sequences inferred by using the Maximum Likelihood method based on the Le_Gascuel_2008 model. The tree with the highest log likelihood (-4553.9584) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.6365)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 70 amino acid sequences. There were a total of 135 positions in the final dataset. Evolutionary analyses were conducted in MEGA6.

Figure 3F: The evolutionary history of CcmI was inferred by using the Maximum Likelihood method based on the Le Gascuel 2008 model. The tree with the highest log likelihood (-2457.0556) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.7523)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 82 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 50 positions in the final dataset. Evolutionary analyses were conducted in MEGA6. (Legend for figure on previous page)

The CcmN homologs were found exclusively in cyanobacteria and in no other phyla of Eubacteria. The CcmN protein sequences were retrieved from NCBI and used for phylogenetic analysis by maximum likelihood method using MEGA 6.1. The phylogenetic tree (**Figure 3H**) shows two major clades. The cyanobacteria falling in Nostocales lineage cluster together while the organisms belonging to Chroococales and Oscillatoriales were found to be dispersed in the phylogenetic tree.

The CcmO protein of CCM is found majorly in the cyanobacteria (75%) phyla and only a few protein homologs were found in Firmicutes (25%). In Firmicutes, only *Listeria monocytogenes* is found to have the protein homolog. The sequences for the protein homologs for CcmO were downloaded from NCBI and analyzed by maximum likelihood method of phylogenetics (**Figure 3I**). The analysis gave a tree with two distinct clades, one constituting the *Listeria monocytogenes* (Firmicutes) protein sequences while the second clade constitutes the protein sequences from cyanobacteria. The cyanobacteria clade has two subclades after branching out of *G. violaceus* PCC 7421 and *G. kilaueensis* JS1, one consisting the CcmO from *Synechococcus elongatus* PCC 6301 and *Synechococcus elongatus* PCC 7942 (both belong to Chroococales) while the second subclade contains the rest of the sequences.

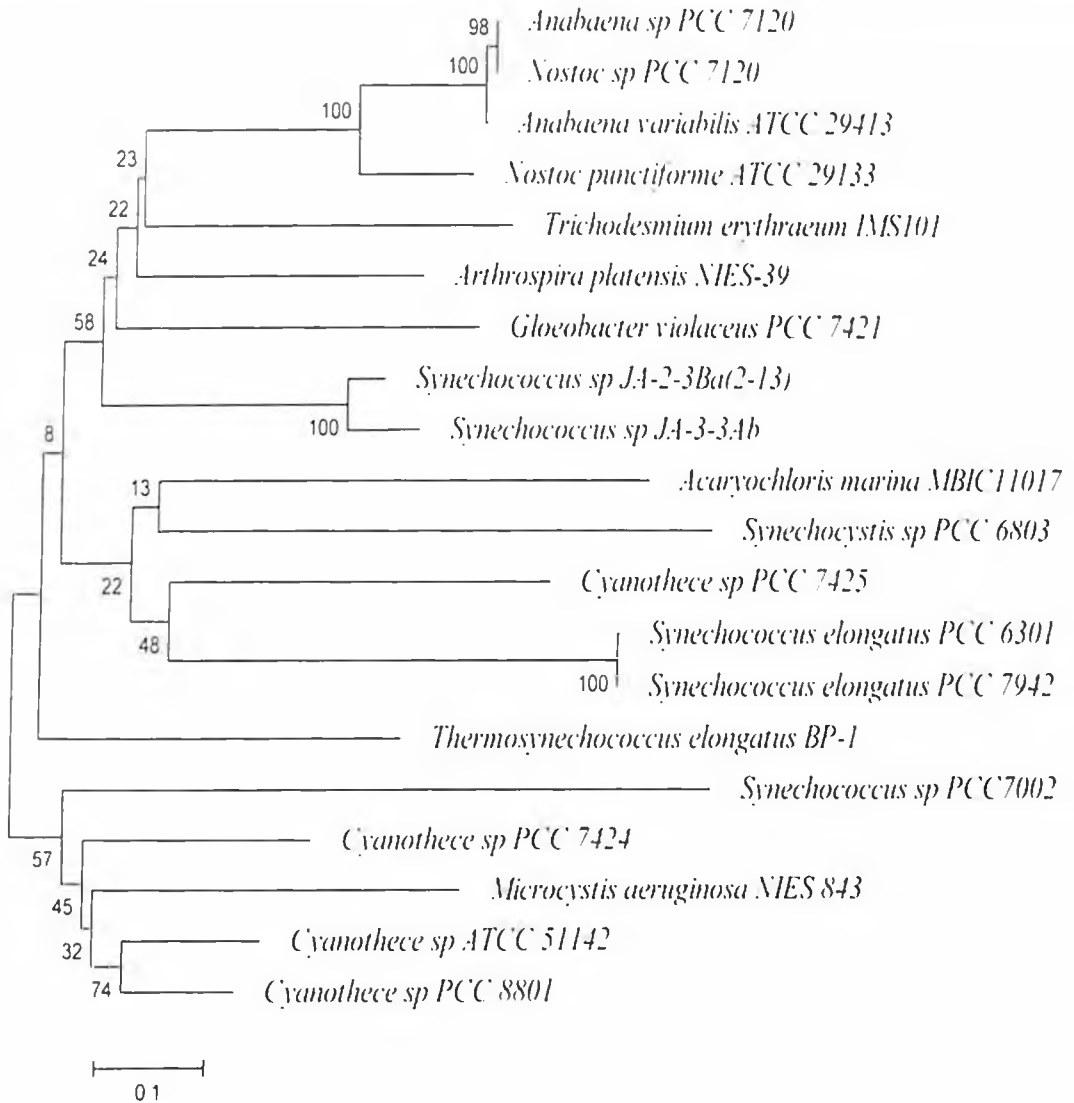


Figure 3G: The evolutionary history of CcmM was inferred by using the Maximum Likelihood method based on the Whelan and Goldman + Freq. model. The tree with the highest log likelihood (-5629.7750) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (-G, parameter = 1.0923)). The rate variation model allowed for some sites to be evolutionarily invariable ([-I], 11.3713% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 58 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 164 positions in the final dataset. Evolutionary analyses were conducted in MEGA6.

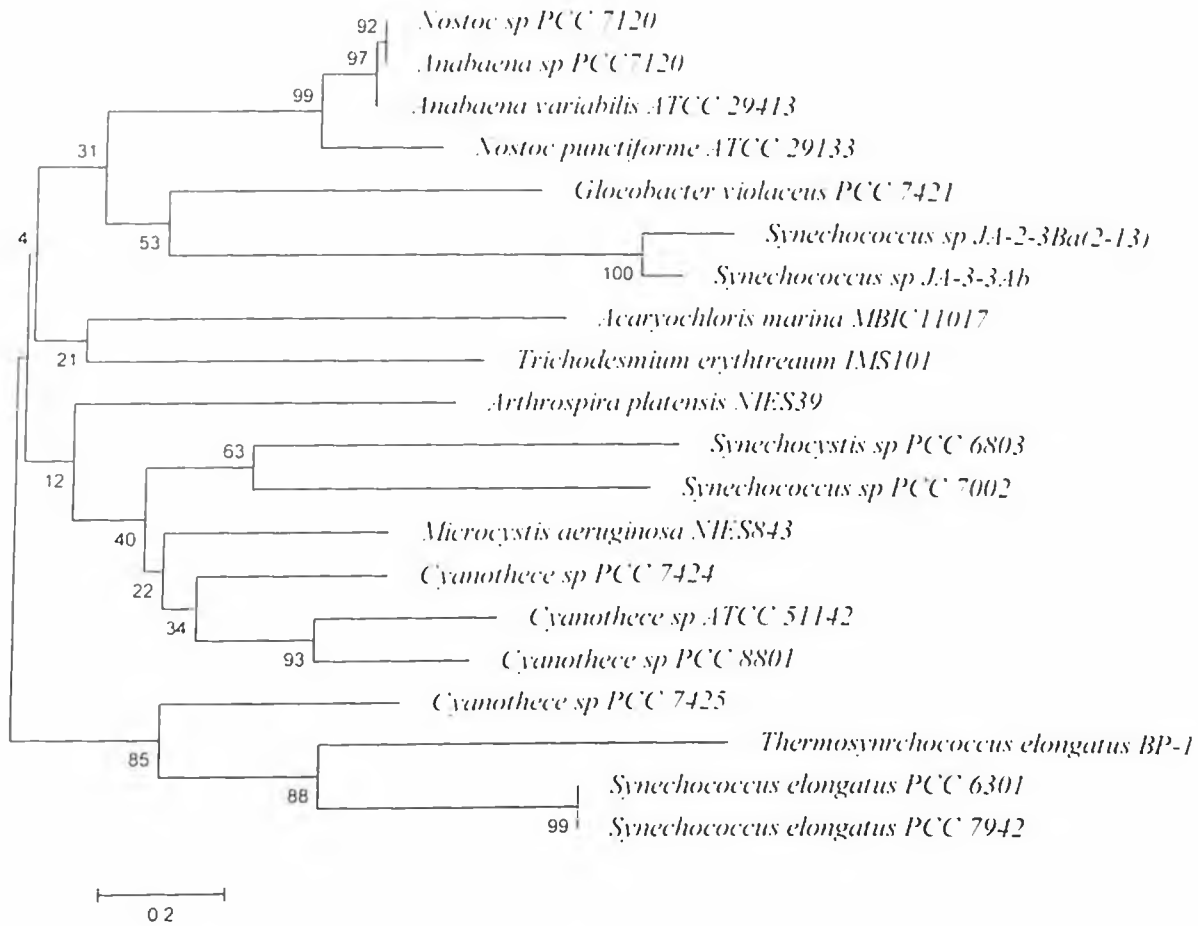


Figure 3H: The evolutionary history of CcmN was inferred by using the Maximum Likelihood method based on the Whelan And Goldman + Freq. model. The tree with the highest log likelihood (-3343.9371) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.7630)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 6.6871% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 29 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 109 positions in the final dataset. Evolutionary analyses were conducted in MEGA6.

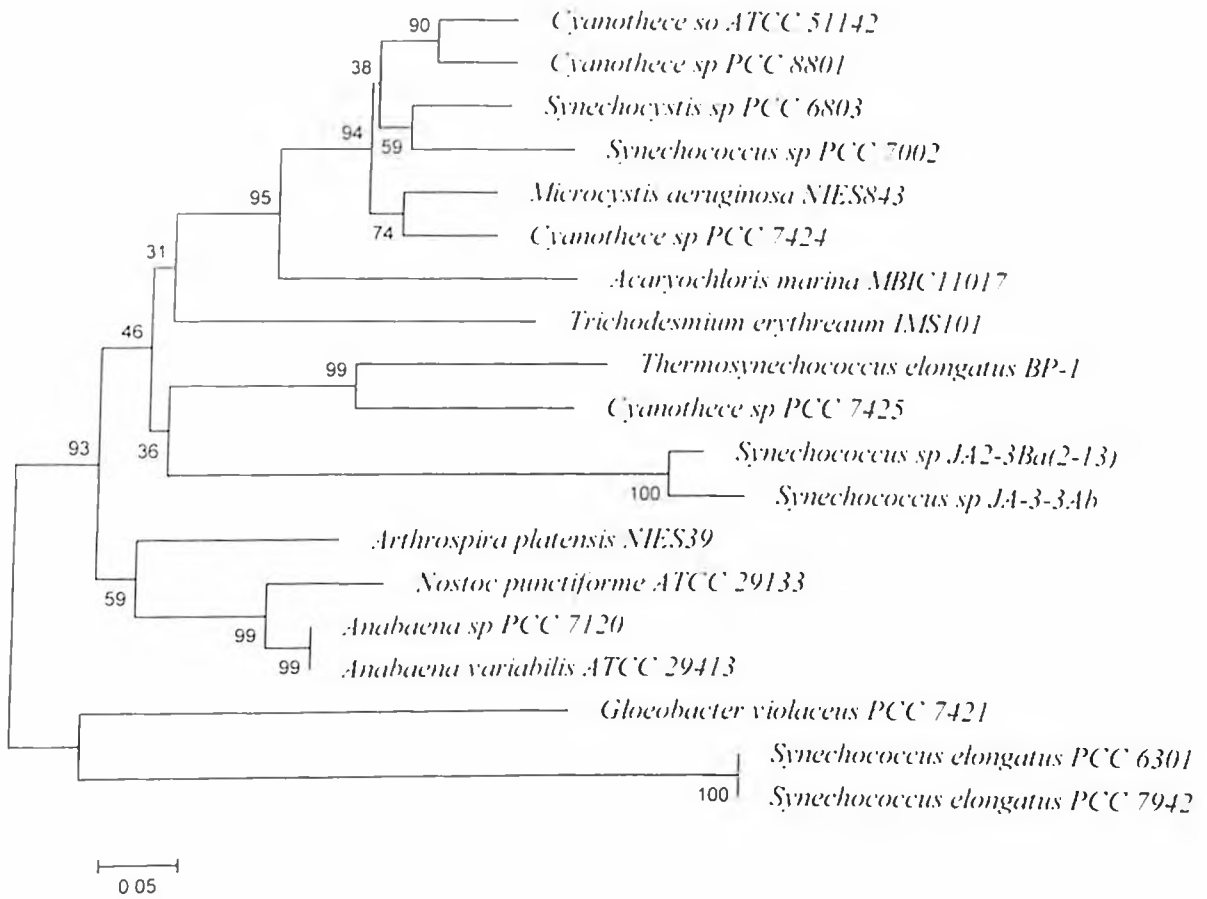


Figure 3I: The evolutionary history of CcmO was inferred by using the Maximum Likelihood method based on the Le Gascuel 2008 model. The tree with the highest log likelihood (-2646.1393) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.1462)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 28 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 148 positions in the final dataset. Evolutionary analyses were conducted in MEGA6.

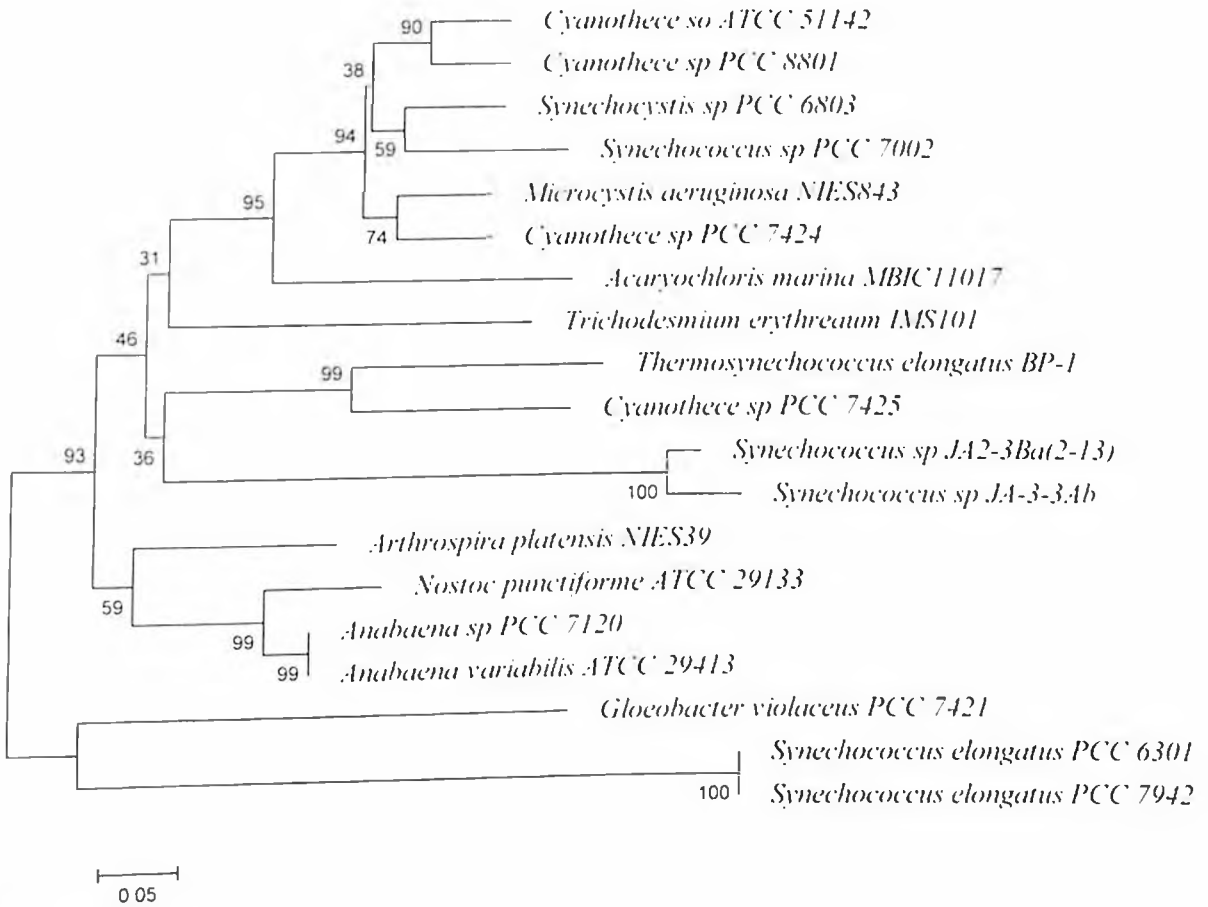


Figure 3I: The evolutionary history of CcmO was inferred by using the Maximum Likelihood method based on the Le Gascuel 2008 model. The tree with the highest log likelihood (-2646.1393) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.1462)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 28 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 148 positions in the final dataset. Evolutionary analyses were conducted in MEGA6.

3.3.5. Diversity of CCM proteins

All CCM proteins were analyzed for pairwise distances to find out the relation of one protein with respect to the others on the basis of sequence homology. Further, the distance of various cyanobacteria was analyzed with respect to *G. violaceus* PCC 7421, the most primitive cyanobacteria, using the number of substitutions per site of the sequences under study. The analysis brought forth the closest relatives of *G. violaceus* PCC 7421 on the basis of 16S rRNA and CCM protein sequence homology.

The distance obtained for individual sequence depicts the distance of each cyanobacteria from *G. violaceus* PCC 7421, for each protein respectively and hence the probable time of emergence of CCM in that particular cyanobacteria. The variation in distance of various cyanobacteria (both α and β) from *G. violaceus* PCC 7421, on basis of 16S rRNA-*pyrH-rbcL* (concatenated) and *cem operon* is plotted in **Figure 3J**. The data shows *T. elongatus* BP-1, *Synechococcus sp* JA-3-3Ab, *Cyanothece sp* PCC 7425, *Synechococcus sp* JA-2-3Ba(2-13) and *Acaryochloris marina* MBIC11017 as closest relatives of *G. violaceus* PCC 7421 on basis of 16S rRNA-*pyrH-rbcL* sequence. Further according to *cem operon* sequence the closest relatives observed were the same set of organisms viz. *Synechococcus sp* JA-2-3Ba(2-13), *Synechococcus sp* JA-3-3Ab, *Cyanothece sp* PCC 7425, *T. elongatus* BP-1 and *Acaryochloris marina* MBIC11017. Among β -cyanobacteria, *Trichodesmium erythraeum* IMS101 is the most distant organisms, which otherwise does not appear to be so distant (according to 16S rRNA sequence). The variation is probably because of the variation developed in the operon with course of time (insertion of a hypothetical protein encoding gene in the *cem* operon). Similar plot when made from the CCM protein sequence distances (**Figure 3K**), showed various peaks and dips of relatedness to *G. violaceus* PCC 7421. If observed closely, it can be observed that there is very little variation among various cyanobacteria with respect to CcmK, CcmL and CcmO proteins as compared to CcmM and CcmN protein sequences. There is a general pattern of variation in the distances among all the organisms under study for all the CCM proteins. A few exceptions to this general pattern are *Synechococcus sp* PCC 7002 (CcmM), *Nostoc punctiforme* (CcmM), *T. elongatus* BP-1 (CcmN), *Trichodesmium erythraeum* IMS101 (CcmN), *Synechocystis sp* PCC 6803 (CcmN) and *T. elongatus* BP-1 (CcmO). These abrupt variations for *Synechococcus sp* PCC 7002 (CcmM) and *Nostoc punctiforme* (CcmM) which show greater relatedness as compared to the other proteins from their respective operons, show that these proteins have somehow not undergone an appreciable variation with time and hence portray greater relatedness. It should be noted that the CcmM protein sequences of all 19 β -cyanobacteria available at Kazusa genome resource when used as query for BLASTP in NCBI, give hits for various CcmM proteins and RuBisCO SSU protein of *Nostoc punctiforme*. Possibly, this signals the

involvement of *N. punctiforme* in evolution of the protein. Further, *T. elongatus* BP-1 (CemN), *Trichodesmium erythraeum* IMS101 (CemN), *Synechocystis* sp PCC 6803 (CemN) and *T. elongatus* BP-1 (CemO) show abrupt peaks signaling greater degree of variation developed by the protein. These representations (**Figure 3J** and **3K**) fail to give the exact closest relatives of *G. violaceus* PCC 7421 due to difference in scale of distance for each protein, hence a box and whiskers plot was used to analyze the data.

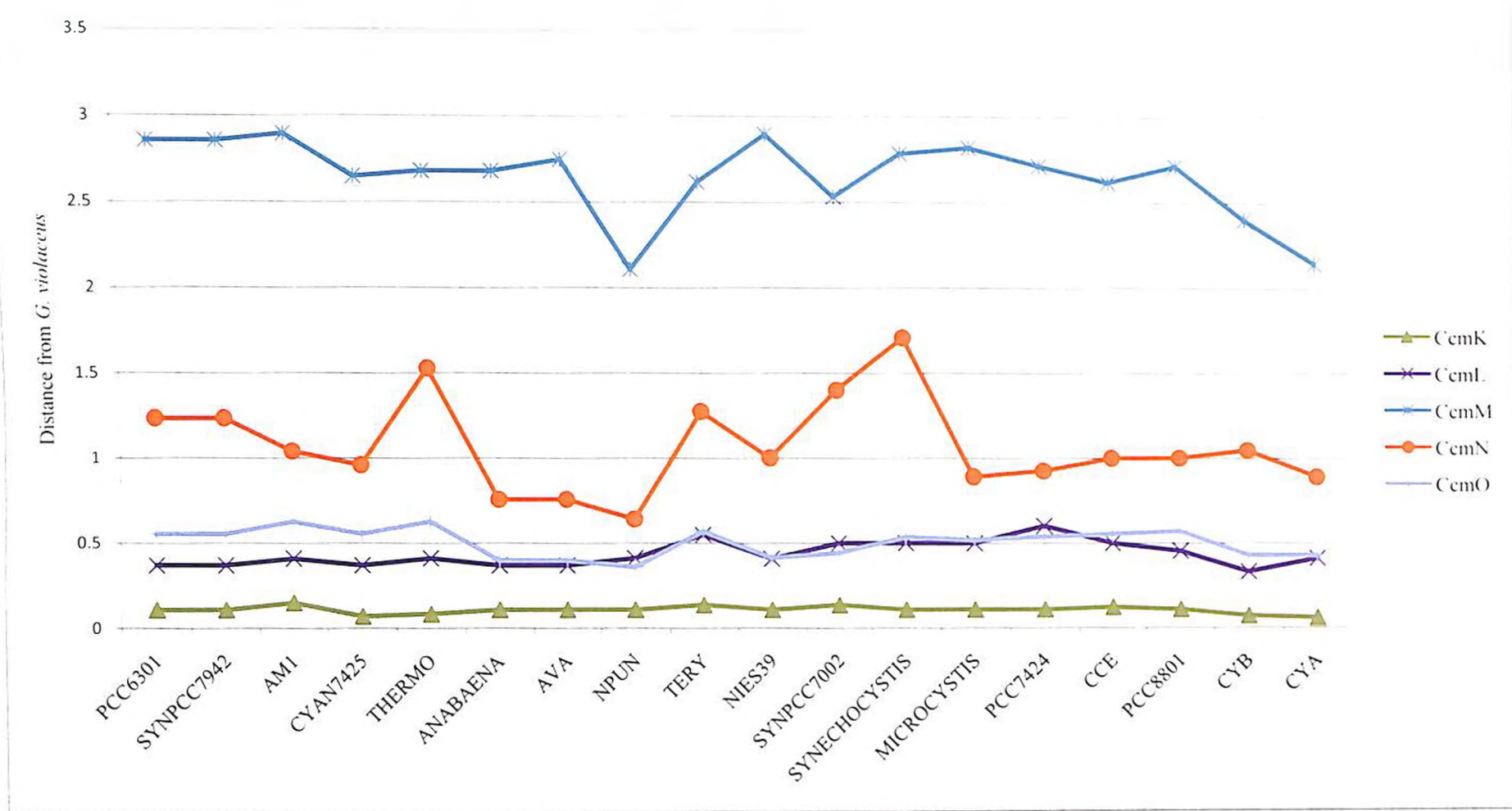


Figure 3K: The comparison of distance of various cyanobacteria from *G. violaceus* PCC 7421 derived by phylogenetic analysis of the 16S rRNA sequence, *ccm* operon. CcmK, CcmL, CcmM, CcmN and CcmO sequences of β -cyanobacteria estimated in MEGA 6.0.

The thickness of the bars shows degree of variation in the distance of the various β -cyanobacteria from *G. violaceus* PCC 7421 (Figure 3L). The distance of various cyanobacteria from *G. violaceus* PCC 7421 on basis of 16S rRNA shows very little variation, as expected, since it is the most conserved region of the genome. The proteins CemN and CemM show greater degree of variation. CemK, although expected to be very diverse, due to several duplicates of the protein found in each genome, the bar on the graph shows it to be comparatively conserved. This result was obtained, because only CemK/CemK1/CemK2 were used for this particular analysis (i.e. the CemK protein closest to *G. violaceus* PCC 7421 CemK from ML tree).

Further, information that can be extracted from the plot is the identification of close relatives of *G. violaceus* PCC 7421 with respect to the CCM proteins. The cyanobacteria corresponding to distance values lying between the minimum value and the first quartile of the set of values for a particular protein were considered as close relatives (Figure 3M).

According to 16S rRNA sequence, *T. elongatus* BP-1, *Synechococcus* sp JA-3-3Ab, *Cyanothece* sp PCC 7425, *Synechococcus* sp JA-2-3B'at(2-13) and *Acaryochloris marina* MBIC11017 are closest to *G. violaceus* PCC 7421. Further according to *ccm* operon sequence *Synechococcus* sp JA-2-3Ba(2-13), *Synechococcus* sp JA-3-3Ab, *Cyanothece* sp PCC 7425, *T. elongatus* BP-1 and *Acaryochloris marina* MBIC11017 are closest relatives according to distance. Similarly, for CemK and CemL, the closest homologs are *Synechococcus* sp JA-2-3B'at(2-13), *Synechococcus* sp JA-3-3Ab, *Cyanothece* sp PCC 7425, *T. elongatus* BP-1 and *S. elongatus* PCC 6301 and *Synechococcus* sp JA-2-3B'at(2-13), *Cyanothece* sp PCC 7425, *Anabaena* sp PCC 7120, *A. variabilis* ATCC 29413 and *S. elongatus* PCC 6301, respectively. For CemM, CemN and CemO are *Nostoc punctiforme* ATCC29133, *Synechococcus* sp JA-2-3B'at(2-13), *Synechococcus* sp JA-3-3Ab, *Synechococcus* sp PCC 7002, *T. erythraeum* IMS101; *Nostoc punctiforme* ATCC29133, *Anabaena* sp PCC 7120, *A. variabilis* ATCC 29413, *M. aeruginosa* NIES843, *Synechococcus* sp JA-3-3Ab; *Nostoc punctiforme* ATCC29133, *Anabaena* sp PCC 7120, *A. variabilis* ATCC 29413, *A. platensis* NIES39, *Synechococcus* sp JA-2-3B'at(2-13), respectively.

The phylogentic analysis of β -cyanobacteria on basis of 16S rRNA and *ccm* proteins shows some congruencies, suggesting *in-situ* formation of the operon in ancestral organisms and its subsequent vertical succession.

3.3.6. *G. violaceus* PCC 7421 – cyanobacterium with unique characteristics

It is a well accepted fact that *G. violaceus* PCC 7421 is an early diverging cyanobacteria. The unique characteristics of this organism and its phylogenetic position qualify it to be among the most primitive forms of life prevalent till date. Des Marais (2000) suggests on the basis of

accumulated evidence that photosynthetic organisms originated 3.2 to 3.5 billion years ago in forms related to cyanobacteria. Life did exist before that but with a different set of metabolisms from oxygenic photosynthesis. Jan Mares et al., (2013) provide detailed morphological, ultrastructural, pigment, and phylogenetic signals in support of primitive nature of *G. violaceus*. Although, *G. violaceus* is among the most initial cyanobacteria to come to life the present analysis does not suggest it have first acquired the CCM. The phylogenetic signals do not clearly identify an ancestor. It should be noted that no organism has been reported till date to have characteristics similar to that of an ancient organism like *G. violaceus* PCC 7421, and hence possibly is the most primitive cyanobacterium. The further lineage of cyanobacteria possibly came into existence after divergence of α and β forms, and hence the subsequent CCMs. This theory can be supported by the facts that both forms of cyanobacteria have completely different set of proteins for CCM and *G. violaceus* although being primitive does not seem to be ancestral for CCM proteins. Some intermediate organism evolved to develop CCM, passed it on to *G. violaceus* and further evolved to give rise to subsequent cyanobacterial lineage (**Figure 3N**).

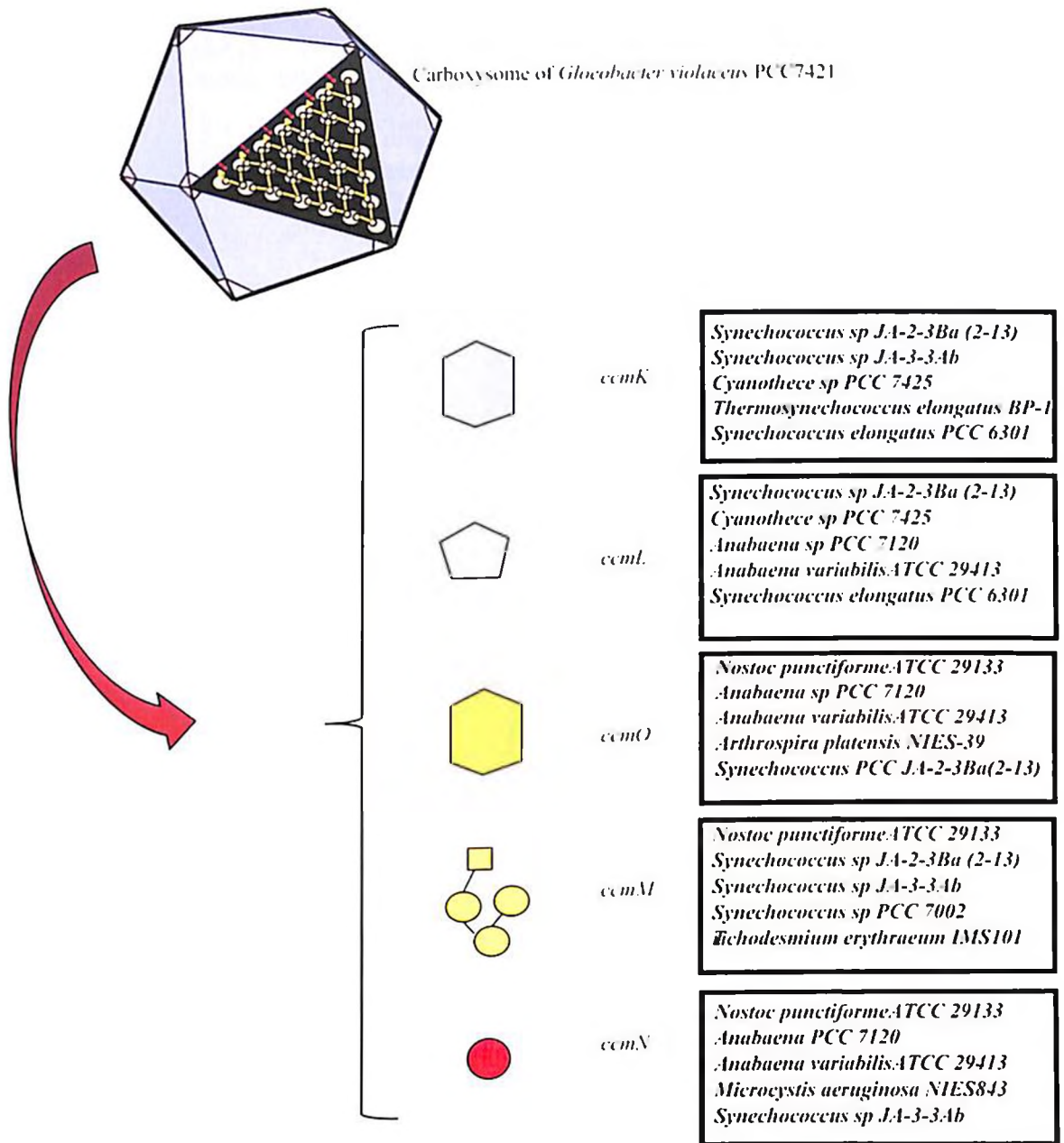


Figure 3M: The closest homologs of CCM proteins of *G. violaceus* PCC 7421, among the β -cyanobacteria available at Cyanobase found by phylogenetic analysis

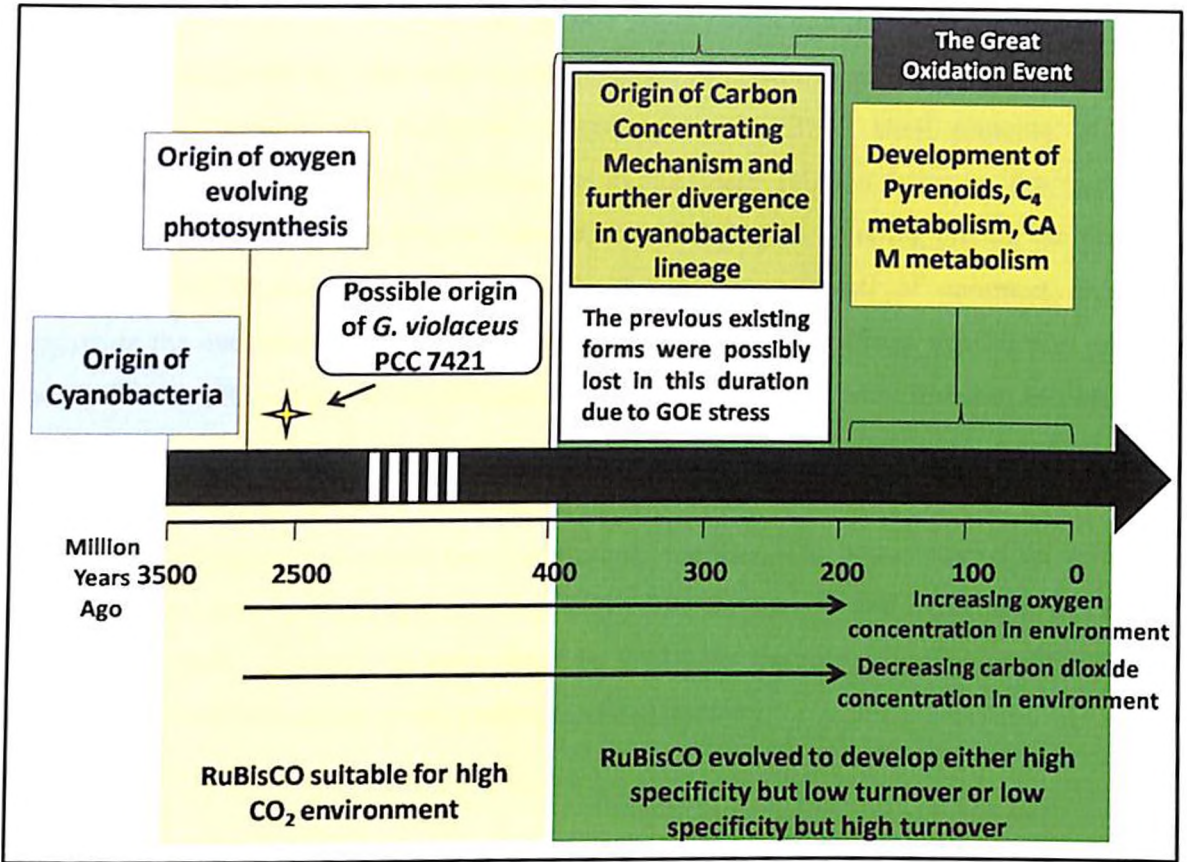


Figure 3N: The speculated time frame of events from origin of cyanobacteria to evolution of CCM in cyanobacteria and higher forms

3.4. Conclusions

The present analysis reveals that α and β -carboxysomes are encoded by distinct set of proteins apart from the shell proteins that share the BMC domain, and hence form distinct phylogenetic clades. Further, the comparison of the carboxysome genes based phylogeny with that of the 16S rRNA show appreciable congruency between the two and hence suggests in-situ formation of the carboxysome and its subsequent vertical succession. As pointed out by several researchers, the carboxysomes are believed to have come into emergence after the divergence of the α and β -cyanobacteria, considering the difference in the set of genes involved in the carboxysomes of both. The origin of *G. violaceus* PCC 7421 could possibly be dated back to the time of origin of oxygenic photosynthesis during the initial times of cyanobacterial origin. Although being ancestral the CCM seems to have originated in some other intermediate/ancestral organism, passed on to *G. violaceus* by horizontal gene transfer and also to further cyanobacterial lineage formed in vertical succession. As per the selfish operon theory, the weakly selected functions are more likely to exist in operon form as this way the chances of being lost in genetic drift are reduced. The strongly selected *ccm* operon of β -cyanobacteria, is dispersed in the genome, as in accordance with the selfish operon theory. But the α carboxysome genes although being strongly selected exist together in the

genome. The phylogenetic analysis, the genetic arrangement and the degree of conservation for the β carboxysome proteins adds to the evidence of in-situ formation of the *ccm* operon. Further, the structural and functional homology to the BMC shell proteins of other microcompartments (Pdu /Eut) signifies the evolutionary relation between the two. The CcmM and CcmN proteins were not found to have no protein homolog among the proteins available at NCBI database apart from the homologs as result of common domains suggesting the evolution of these by mechanisms like domain shuffling, gene fusion or gene fission. The CcmK proteins of the β -cyanobacteria when analyzed were found to be present in several copies, emphasizing the importance of the protein in shell formation. Further, it was found that the number of CcmK encoding genes has increased along the lane of evolution, as the early diverging cyanobacteria had the minimal number, which was passed on vertically and then subsequent duplication events took place. The further role and importance of each of the CcmK proteins can only be appreciated by analyzing the role of each and the effect of each on the overall efficiency of the carboxysome machinery.

Supplementary Tables

Table 3A-S: BLASTP hits for amino acid sequence of CcmK from *G. violaceus* PCC 7421 from various phyla of Eubacteria and Archaea

| S.No | Phylum | Total Hits | E: 10 ⁻⁵ | CcmK and E: 10 ⁻⁵ | Other proteins and E: 10 ⁻⁵ | Names of other proteins |
|------|---------------------------|------------|---------------------|------------------------------|--|---|
| 1 | Archaea | 9 | 0 | 0 | 0 | - |
| 2 | Actinobacteria | 100 | 100 | 31 | 69 | PduA, EutM, BMC domain protein |
| 3 | Aquificae | 11 | 0 | 0 | 0 | - |
| 4 | Armatimonadetes | | | | | |
| 5 | Bacteroidetes Chlorobi | 100 | 11 | - | 11 | Carboxysome shell protein, EutM, pduI, hypothetical protein |
| 6 | Caldiserica | 42 | 0 | 0 | 0 | - |
| 7 | Chlamydiae | 100 | 0 | 0 | 0 | - |
| 8 | Chloroflexi | 14 | 3 | 0 | 3 | EutM, carboxysome shell protein |
| 9 | Chrysiogenetes | 13 | 0 | 0 | 0 | - |
| 10 | Cyanobacteria | 100 | 100 | 100 | 0 | - |
| 11 | Deferribacteres | 2 | 0 | 0 | 0 | - |
| 12 | Deinococcus-thermus | 2 | 0 | 0 | 0 | - |
| 13 | Dictyoglomi | 4 | 0 | 0 | 0 | - |
| 14 | Elusimicrobia | 6 | 0 | 0 | 0 | - |
| 15 | Fibrobacteres Acidobacter | 21 | 12 | 1 | 11 | EutM, Carboxysome shell protein, microcompartment protein |
| 16 | Firmicutes | 100 | 100 | 8 | 92 | EutM, pduA, carboxysome shell protein |
| 17 | Fusobacteria | 100 | 79 | 2 | 77 | Carboxysome shell protein, EutM, pduI, hypothetical protein |
| 18 | Gemmatimonadetes | 11 | 5 | 0 | 5 | EutM |
| 19 | Nitrospirae | 5 | 0 | 0 | 0 | - |
| 20 | Nitrospirae | 101 | 0 | 0 | 0 | - |
| 21 | Planctomycetes | 46 | 36 | 1 | 35 | Carboxysome shell protein, EutM, microcompartment protein |
| 22 | Alpha Proteobacteria | 52 | 33 | 22 | 10 | EutM, pduA |
| 23 | Beta Proteobacteria | 54 | 46 | 34 | 12 | EutM, pduA, EutK, pduI |
| 24 | Gamma Proteobacteria | 100 | 100 | 41 | 59 | EutM, pduI, pduA |
| 25 | Delta Proteobacteria | 101 | 70 | 40 | 30 | EutM, pduA, EutK, Biotin ligase |
| 26 | Epsilon Proteobacteria | 2 | 0 | 0 | 0 | - |
| 27 | Spirochaetes | 24 | 20 | 0 | 20 | Carboxysome shell protein, EutM, hypothetical protein |
| 28 | Synergistetes | 71 | 36 | 3 | 33 | Carboxysome shell protein, EutM, pduI |
| 29 | Tenericutes | 12 | 4 | 0 | 4 | Propanediol utilization protein, EutM |
| 30 | Thermodesulfobacteria | 12 | 0 | 0 | 0 | - |
| 31 | Thermotogae | 13 | 0 | 0 | 0 | - |

Table 3B-S: BLASTP hits for amino acid sequence of CcmI from *G. violaceus* PCC 7421 from various phyla of Eubacteria and Archaea

| S. No | Phylum | Total Hits | E: 10 ⁻⁵ | CcmI. and E: 10 ⁻⁵ | Other proteins and E: 10 ⁻⁵ | Names of other proteins |
|-------|---------------------------|------------|---------------------|-------------------------------|--|---|
| 1 | Archaea | 2 | 0 | 0 | 0 | - |
| 2 | Actinobacteria | 78 | 22 | 1 | 21 | EutN, PduN, hypothetical protein |
| 3 | Aquificae | 1 | 0 | 0 | 0 | - |
| 4 | Armatimonadetes | 9 | 0 | 0 | 0 | - |
| 5 | Bacteroidetes/Chlorobi | 100 | 13 | 0 | 13 | EutN, hypothetical protein |
| 6 | Caldiserica | 3 | 0 | 0 | 0 | - |
| 7 | Chlamydiae | 3 | 0 | 0 | 0 | - |
| 8 | Chloroflexi | 10 | 1 | 0 | 1 | EutN |
| 9 | Chrysiogenetes | 8 | 0 | 0 | 0 | - |
| 10 | Cyanobacteria | 100 | 100 | 86 | 14 | EutN |
| 11 | Deferribacteres | 2 | 0 | 0 | 0 | - |
| 12 | Deinococcus-thermus | 0 | 0 | 0 | 0 | - |
| 13 | Dictyoglomi | 8 | 0 | 0 | 0 | - |
| 14 | Elusimicrobia | 50 | 0 | 0 | 0 | - |
| 15 | Fibrobacteres/Acidobacter | 18 | 11 | 0 | 11 | EutN, hypothetical protein |
| 16 | Firmicutes | 100 | 100 | 10 | 90 | EutN, propanediol utilization protein, hypothetical protein |
| 17 | Fusobacteria | 38 | 33 | 2 | 31 | EutN, hypothetical protein |
| 18 | Gemmatimonadetes | 13 | 3 | 0 | 3 | EutN, hypothetical protein |
| 19 | Nitrospirae | 9 | 0 | 0 | 0 | - |
| 20 | Nitrospirae | 9 | 0 | 0 | 0 | - |
| 21 | Planctomycetes | 52 | 31 | 8 | 23 | EutN |
| 22 | Alpha Proteobacteria | 38 | 17 | 8 | 9 | EutN |

| | | | | | | |
|----|------------------------|-----|-----|---|-----|----------------------------|
| 23 | Beta Proteobacteria | 41 | 10 | 6 | 4 | FutN |
| 24 | Gamma Proteobacteria | 100 | 100 | 0 | 100 | FutN |
| 25 | Delta Proteobacteria | | | | | |
| 26 | Epsilon Proteobacteria | 3 | 0 | 0 | 0 | - |
| 27 | Spirochaetes | 11 | 8 | 0 | 8 | Hypothetical protein, FutN |
| 28 | Synergistetes | 15 | 11 | 1 | 10 | FutN |
| 29 | Tenericutes | 15 | 3 | 0 | 3 | Hypothetical protein, FutN |
| 30 | Thermodesulfobacteria | 0 | 0 | 0 | 0 | - |
| 31 | Thermotogae | 8 | 0 | 0 | 0 | - |

Table 3C-S: BLASTP hits for amino acid sequence of CemM from *G. violaceus* PCC 7421 from various phyla of Eubacteria and archaea

| S No | Phylum | Total Hits | 1-10 | CemM and 1-10 | Other proteins and 1-10 | Names of other proteins |
|------|---------------------------|------------|------|---------------|-------------------------|---|
| 1 | Archaea | 100 | 100 | 0 | 100 | Carbonic anhydrase, acetyltransferase, ferripyochelin binding protein |
| 2 | Actinobacteria | 100 | 100 | 0 | 100 | Carbonate dehydratase, anhydrase, hypothetical protein, siderophore binding protein, phenylacetic acid degradation protein PaaY, 2,3,4,5-tetrahydropyridine-2,6-dicarboxylate-N-acetyltransferase |
| 3 | Aquificae | 60 | 20 | 0 | 20 | Acetyltransferase, transferase, hypothetical protein, putative carbonic anhydrase |
| 4 | Armatimonadetes | 20 | 0 | 0 | 0 | - |
| 5 | Bacteroidetes Chlorobi | 100 | 100 | 0 | 100 | Acetyltransferase, transferase, hypothetical protein, carbonic anhydrase, hexapeptide repeat containing protein |
| 6 | Caldisevica | 12 | 1 | 0 | 1 | Hypothetical protein |
| 7 | Chlamydiae | 30 | 0 | 0 | 0 | - |
| 8 | Chloroflexi | 73 | 13 | 0 | 13 | Anhydrase, anhydratase, Transferase hexapeptide repeat containing protein, hypothetical protein |
| 9 | Chrysiogenetes | 16 | 2 | 0 | 2 | Transferase, anhydratase |
| 10 | Cyanobacteria | 100 | 100 | 49 | 51 | Carbonic anhydrase, cytochrome C biogenesis protein cemM |
| 11 | Deferribacteres | 33 | 8 | 0 | 8 | Hypothetical protein, acetyltransferase |
| 12 | Deinococcus-thermus | 76 | 22 | 0 | 22 | Carbonate dehydratase, carbonic anhydrase, ferripyochelin binding protein, hypothetical protein, NUDIX protein |
| 13 | Dietvoglomi | 15 | 2 | 0 | 2 | Hypothetical protein |
| 14 | Flusimicrobia | 4 | 1 | 0 | 1 | Hypothetical protein |
| 15 | Fibrobacteres Acidobacter | 43 | 20 | 0 | 20 | Transferase, Hypothetical protein, carbonic anhydrase |
| 16 | Firmicutes | 100 | 100 | 4 | 96 | Carbonic anhydrase, carbonate dehydratase, hypothetical protein, transferase |
| 17 | Fusobacteria | 77 | 14 | 0 | 14 | Bacterial transferase hexapeptide repeat containing protein, acetyltransferase, Hypothetical protein |
| 18 | Gemmatimonadetes | 38 | 9 | 0 | 9 | Transferase hexapeptide repeat containing protein, phenylacetic acid degradation protein PaaY, Hypothetical protein, transferase |
| 19 | Nitrospinae | 22 | 3 | 0 | 3 | Hypothetical protein, putative transferase, hexapeptide repeat protein |
| 20 | Nitrospirae | 42 | 10 | 0 | 10 | Carbonic anhydrase, putative transferase, acetyltransferase, Hypothetical protein |
| 21 | Planctomycetes | 63 | 24 | 0 | 24 | Ferripyochelin binding protein, Hypothetical protein, anhydrase, phenylacetic acid degradation protein PaaY |

| | | | | | | |
|----|------------------------|-----|-----|---|-----|---|
| 22 | Alpha Proteobacteria | 100 | 100 | 0 | 100 | Hexapeptide repeat containing transferase, carbonate dehydratase, acetyltransferase |
| 23 | Beta Proteobacteria | 100 | 100 | 0 | 100 | Carbonate dehydratase, acetyltransferase |
| 24 | Gamma Proteobacteria | 100 | 100 | 0 | 100 | Carbonate dehydratase, UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acetyltransferase |
| 25 | Delta Proteobacteria | 101 | 101 | 0 | 101 | Sulfate permease, carbonic anhydrase, transferase, protein YrdA, phenylacetic acid degradation protein PaaY, UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acetyltransferase |
| 26 | Epsilon Proteobacteria | | | | | - |
| 27 | Spirochaetes | 100 | 100 | 0 | 100 | Transferase hexapeptide repeat protein, acetyltransferase, carbonic anhydrase |
| 28 | Synergistetes | 33 | 7 | 0 | 7 | Hypothetical protein, anhydrase, transferase |
| 29 | Ferriplasmatales | 11 | 1 | 0 | 1 | Hypothetical protein |
| 30 | Thermodesulfobacteria | 9 | 0 | 0 | 0 | - |
| 31 | Thermotogae | 74 | 2 | 0 | 2 | Hypothetical protein, acetyltransferase |

Table 3D-S: BLASTP hits for amino acid sequence of CcmN from *G. violaceus* PCC 7421 from various phyla of Eubacteria and Archaea

| S. No | Phylum | Total Hits | $1 \cdot 10^4$ | CcmN and $1 \cdot 10^4$ | Other proteins and $1 \cdot 10^4$ | Names of other proteins |
|-------|---------------------------|------------|----------------|-------------------------|-----------------------------------|--|
| 1 | Archaea | 107 | 15 | 0 | 15 | Hexapeptide repeat containing transferase, carbonic anhydrase, acetyltransferase |
| 2 | Actinobacteria | 103 | 51 | 0 | 51 | Carbonic anhydrase, isoleucine patch superfamily enzyme, putative siderophore binding protein |
| 3 | Aquificae | 11 | 0 | 0 | 0 | - |
| 4 | Armatimonadetes | 20 | 0 | 0 | 0 | - |
| 5 | Bacteroidetes Chlorobi | 101 | 17 | 0 | 17 | Carbonic anhydrase, bacterial transferase hexapeptide repeat protein, acetyltransferase |
| 6 | Caldiseica | | | | | |
| 7 | Chlamydiae | 34 | 1 | 0 | 1 | Carbonic anhydrase |
| 8 | Chloroflexi | 26 | 7 | 0 | 7 | Transferase hexapeptide repeat containing protein, anhydratase, acetyltransferase |
| 9 | Chrysiogenetes | 15 | 0 | 0 | 0 | - |
| 10 | Cyanobacteria | 100 | 100 | 21 | 79 | Hypothetical protein, hexapeptide repeat containing transferase, transferase, 2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-acetyltransferase |
| 11 | Deferribacteres | 18 | 0 | 0 | 0 | - |
| 12 | Deinococcus-thermus | 66 | 11 | 0 | 11 | Carbonic anhydrase, hypothetical protein, ferripyochelin binding protein, isoleucine patch superfamily enzyme |
| 13 | Dietyogloni | 9 | 0 | 0 | 0 | - |
| 14 | Flusimicrobia | 13 | 0 | 0 | 0 | - |
| 15 | Fibrobacteres/Acidobacter | 36 | 0 | 0 | 0 | - |
| 16 | Firmicutes | 102 | 10 | 0 | 10 | Hypothetical protein, transferase, carbonic anhydrase |
| 17 | Fusobacteria | 43 | 1 | 0 | 1 | Bacterial transferase hexapeptide repeat protein |
| 18 | Gemmatimonadetes | 27 | 1 | 0 | 1 | Transferase hexapeptide repeat containing protein |
| 19 | Nitrospinae | 16 | 0 | 0 | 0 | - |
| 20 | Nitrospirae | 28 | 0 | 0 | 0 | - |
| 21 | Planctomycetes | 37 | 0 | 0 | 0 | - |
| 22 | Alpha Proteobacteria | 102 | 33 | 0 | 33 | Anhydrase, bacterial transferase hexapeptide repeat protein, acetyltransferase |

| | | | | | | |
|----|------------------------|-----|----|---|----|---|
| 23 | Beta Proteobacteria | 105 | 11 | 0 | 11 | Anhydrase, Transferase hexapeptide repeat containing protein |
| 24 | Gamma Proteobacteria | 114 | 7 | 0 | 7 | Carbonic anhydrase, hypothetical protein, acetyltransferase, bacterial transferase hexapeptide repeat protein |
| 25 | Delta Proteobacteria | 0 | 0 | 0 | 0 | - |
| 26 | Epsilon Proteobacteria | 29 | 0 | 0 | 0 | - |
| 27 | Spirochaetes | 100 | 0 | 0 | 0 | - |
| 28 | Synergistetes | 35 | 0 | 0 | 0 | - |
| 29 | Fenericutes | 9 | 0 | 0 | 0 | - |
| 30 | Thermodesulfobacteria | 8 | 0 | 0 | 0 | - |
| 31 | Thermotogae | 39 | 0 | 0 | 0 | - |

Table 3E-8: BLASTP hits for amino acid sequence of CcmO from *G. violaceus* PCC 7421 from various phyla of Eubacteria and Archaea

| S No | Phylum | Total Hits | $E = 10^{-7}$ | CcmO and $E = 10^{-10}$ | Other proteins and $E = 10^{-7}$ | Names of other proteins |
|------|---------------------------|------------|---------------|-------------------------|----------------------------------|---|
| 1 | Archaea | 10 | 0 | 0 | 0 | - |
| 2 | Actinobacteria | 100 | 98 | 3 | 95 | Carboxysome shell protein, FutM, PduA, PduK |
| 3 | Aquificae | 4 | 0 | 0 | 0 | - |
| 4 | Armatimonadetes | 8 | 0 | 0 | 0 | - |
| 5 | Bacteroidetes Chlorobi | 26 | 9 | 0 | 9 | FutM, carboxysome shell protein, hypothetical protein PduI |
| 6 | Caldisevica | | | | | |
| 7 | Chlamydiae | 50 | 20 | 0 | 20 | FutM, carboxysome shell protein, hypothetical protein, microcompartment protein |
| 8 | Chloroflexi | 12 | 3 | 0 | 3 | FutM, carboxysome shell protein |
| 9 | Chrysiogenetes | 9 | 0 | 0 | 0 | - |
| 10 | Cyanobacteria | 100 | 100 | 50 | 50 | Hypothetical protein, FutM |
| 11 | Deferribacteres | 7 | 0 | 0 | 0 | - |
| 12 | Demococcus-thermus | 10 | 0 | 0 | 0 | - |
| 13 | Dictyoglom | 9 | 0 | 0 | 0 | - |
| 14 | Elusimicrobia | 8 | 0 | 0 | 0 | - |
| 15 | Fibrobacteres Acidobacter | 29 | 13 | 1 | 12 | FutM carboxysome shell protein, microcompartment protein PduI |
| 16 | Firmicutes | 100 | 100 | 7 | 93 | FutM, PduA, BMC domain protein, carboxysome shell protein |
| 17 | Fusobacteria | 100 | 67 | 2 | 65 | FutM, carboxysome shell protein |
| 18 | Gemmatimonadetes | 14 | 5 | 0 | 5 | FutM |
| 19 | Nitrospinae | 11 | 0 | 0 | 0 | - |
| 20 | Nitrospirae | 5 | 0 | 0 | 0 | - |
| 21 | Planctomycetes | 85 | 35 | 1 | 34 | FutM, carboxysome shell protein, microcompartment protein, hypothetical protein |
| 22 | Alpha Proteobacteria | 97 | 31 | 21 | 10 | FutM, PduA |
| 23 | Beta Proteobacteria | 49 | 43 | 28 | 11 | FutM, PduA, Hypothetical protein, PduI |
| 24 | Gamma Proteobacteria | 100 | 100 | 21 | 79 | FutM, PduJ, detox protein, PduA, FutN, hypothetical protein |
| 25 | Delta Proteobacteria | 0 | 0 | 0 | 0 | - |
| 26 | Epsilon Proteobacteria | 2 | 0 | 0 | 0 | - |
| 27 | Spirochaetes | 57 | 14 | 0 | 14 | FutM, carboxysome shell protein, hypothetical protein, microcompartment protein |
| 28 | Synergistetes | 96 | 27 | | | FutM, PduI, carboxysome shell protein |
| 29 | Fenericutes | 17 | 4 | 0 | 4 | FutM, Propanediol utilization protein |
| 30 | Thermodesulfobacteria | 11 | 0 | 0 | 0 | - |
| 31 | Thermotogae | 3 | 0 | 0 | 0 | - |

Table 3E-8: List of *ccmK* gene complements in β -cyanobacteria arranged according to the phylogenetic topology in Figure 3I.

| Clade | Name of organism | CcmK gene complements |
|----------------------------------|--|---|
| M1 | <i>S. elongatus</i> PCC 6301 | <i>Syc0134 (ccm1)-syc0135 (ccm1) -syc1227 (ccm4)-syc1228 (ccm3)</i> |
| | <i>S. elongatus</i> PCC 7942 | <i>Synpcc7942_1422 (ccm1)-Synpcc 7942_1421(ccm1) -Synpcc7942_0284 (ccm3)-Synpcc 7942_0285 (ccm4)</i> |
| | <i>Synechococcus</i> sp.JA-2-3B at2-13) | <i>CYB_1795 (ccm1)-CYB_1796 (ccm1)-CYB_1797 (ccm2)</i> |
| | <i>Synechococcus</i> sp.JA-3-3A4b | <i>CYA_1613 (ccm1)-CYA_1612 (ccm2)-CYA_1611 (ccm1)</i> |
| | <i>T. elongatus</i> BP-1 | <i>tl10945 (ccm1)-tl10946 (ccm1)-tl10947 (ccm2) -tlr0954 (ccm3) -tl11596 (ccm4)</i> |
| | <i>G. violaceus</i> PCC 7421 | <i>gll2094 (ccm1)-gll2095 (ccm1)-gll2096 (ccm1)</i> |
| | <i>T. erythraeum</i> IMS 101 | <i>Tery_3850 (ccm1)-Tery_3851 (MCP)-Tery_3852 (MCP) -Tery_4328 (MCP)-Tery_4329 (MCP)</i> |
| | <i>N. punctiforme</i> STCC 29133 | <i>Npun_F4293 (ccm1)-Npun_F4292 (MCP)-Npun_F4291 (MCP) -Npun_F2745 (MCP)-Npun_2744 (MCP)</i> |
| | <i>Anabaena</i> PCC 7120 | <i>all0866 (ccm1)-all0867 (ccm1)-all0868 (ccm1) -alr0317 (ccm1)-alr0318 (ccm1)</i> |
| | <i>A. variabilis</i> ATCC 29413 | <i>Ava_4470 (ccm1)-Ava_4471 (MCP)-Ava_4472 (MCP) -Ava_4709 (MCP)-Ava_4710 (MCP)</i> |
| | <i>Synechococcus</i> sp PCC 7002 | <i>SYNPCC7002_A1801 (ccm1)-SYNPCC7002_A1802 (ccm1)-SYNPCC7002_A1803 (ccm2) -SYNPCC7002_A2612 (ccm1)-SYNPCC7002_A2613 (ccm1)</i> |
| | <i>Synechocystis</i> sp PCC 6803 | <i>sl11030 (ccm1)-sl11029 (ccm1)-sl11028 (ccm2) -slr1838 (ccm3)-slr1839 (ccm4)</i> |
| | <i>Cyanotheca</i> sp PCC 7424 | <i>PCC7424_1370 (ccm1)-PCC7424_1371 (MCP)-PCC7424_1372 (MCP)-PCC7424_1373 (MCP) -PCC7424_2157 (MCP)-PCC7424_2158 (MCP)</i> |
| | <i>Cyanotheca</i> sp ATCC 51142 | <i>ccc_4281 (ccm1)-ccc_4282 (ccm1)-ccc_4283 (ccm2) -ccc_2433 (ccm4)-ccc_2434 (ccm3)</i> |
| | <i>Cyanotheca</i> sp PCC 8801 | <i>PCC8801_1598 (ccm1)-PCC8801_1597 (MCP)-PCC8801_1596 (MCP) -PCC8801_1859 (MCP)-PCC8801_1860 (MCP)</i> |
| | <i>A. marina</i> MBIC 11017 | <i>AMI_5382 (ccm1)-AMI_5381 (ccm1)-AMI_5380 (ccm1) -AMI_0655 (ccm1)-AMI_0656 (ccm1) -AMI_3280 (ccm1) -AMI_5778 (ccm1)</i> |
| | <i>Cyanotheca</i> sp PCC 7425 | <i>Cyan7425_1616 (ccm1)-Cyan7425_1617 (ccm1)-Cyan7425_1618 (MCP) -Cyan7425_2087 (MCP) -Cyan7425_2386 (MCP)</i> |
| | <i>Synechocystis</i> sp PCC 6803 | <i>sl11030 (ccm1)-sl11029 (ccm1)-sl11028 (ccm2) -slr1838 (ccm3)-slr1839 (ccm4)</i> |
| | <i>A. platensis</i> NIES 39 | <i>NIES39_K04810 (ccm1)-NIES39_K04820 (ccm1)-NIES39_K04830 (ccm2) -NIES39_A03150 (ccm4)-NIES39_A03160 (ccm3)</i> |
| | <i>A. platensis</i> NIES 39 | <i>NIES39_K04810 (ccm1)-NIES39_K04820 (ccm1)-NIES39_K04830 (ccm2) -NIES39_A03150 (ccm4)-NIES39_A03160 (ccm3)</i> |
| | <i>Cyanotheca</i> sp PCC 7425 | <i>Cyan7425_1616 (ccm1)-Cyan7425_1617 (ccm1)-Cyan7425_1618 (MCP) -Cyan7425_2087 (MCP) -Cyan7425_2386 (MCP)</i> |
| | <i>T. elongatus</i> BP-1 | <i>tl10945 (ccm1)-tl10946 (ccm1)-tl10947 (ccm2) -tlr0954 (ccm3) -tl11596 (ccm4)</i> |
| | <i>G. violaceus</i> PCC 7421 | <i>gll2094 (ccm1)-gll2095 (ccm1)-gll2096 (ccm1)</i> |
| | <i>Synechococcus</i> sp JA-2-3B at2-13) | <i>CYB_1795 (ccm1)-CYB_1796 (ccm1)-CYB_1797 (ccm2)</i> |
| | <i>Synechococcus</i> sp JA-3-3A4b | <i>CYA_1613 (ccm1)-CYA_1612 (ccm2)-CYA_1611 (ccm1)</i> |
| | <i>M. aeruginosa</i> NIES 843 | <i>MAE47920 (ccm1)-MAE47930 (ccm1)-MAE47940 (ccm2) -MAE55390 (ccm3)-mae55400 (ccm4)</i> |
| | <i>Cyanotheca</i> sp ATCC 51142 | <i>ccc_4281 (ccm1)-ccc_4282 (ccm1)-ccc_4283 (ccm2) -ccc_2433 (ccm4)-ccc_2434 (ccm3)</i> |
| | <i>Synechococcus</i> sp PCC 7002 | <i>SYNPCC7002_A1801 (ccm1)-SYNPCC7002_A1802 (ccm1)-SYNPCC7002_A1803 (ccm2) -SYNPCC7002_A2612 (ccm1)-SYNPCC7002_A2613 (ccm1)</i> |
| <i>Cyanotheca</i> sp PCC 8801 | <i>PCC8801_1598 (ccm1)-PCC8801_1597 (MCP)-PCC8801_1596 (MCP) -PCC8801_1859 (MCP)-PCC8801_1860 (MCP)</i> | |
| <i>M. aeruginosa</i> NIES 843 | <i>MAE47920 (ccm1)-MAE47930 (ccm1)-MAE47940 (ccm2) -MAE55390 (ccm3)-mae55400 (ccm4)</i> | |
| <i>Cyanotheca</i> sp PCC 7424 | <i>PCC7424_1370 (ccm1)-PCC7424_1371 (MCP)-PCC7424_1372 (MCP)-PCC7424_1373 (MCP) -PCC7424_2157 (MCP)-PCC7424_2158 (MCP)</i> | |
| <i>Cyanotheca</i> sp PCC 7424 | <i>PCC7424_1370 (ccm1)-PCC7424_1371 (MCP)-PCC7424_1372 (MCP)-PCC7424_1373 (MCP) -PCC7424_2157 (MCP)-PCC7424_2158 (MCP)</i> | |
| <i>T. erythraeum</i> IMS 101 | <i>Tery_3850 (ccm1)-Tery_3851 (MCP)-Tery_3852 (MCP) -Tery_4328 (MCP)-Tery_4329 (MCP)</i> | |
| <i>A. marina</i> MBIC 11017 | <i>AMI_5382 (ccm1)-AMI_5381 (ccm1)-AMI_5380 (ccm1) -AMI_0655 (ccm1)-AMI_0656 (ccm1) -AMI_3280 (ccm1) -AMI_5778 (ccm1)</i> | |
| <i>Anabaena</i> PCC 7120 | <i>all0866 (ccm1)-all0867 (ccm1)-all0868 (ccm1) -alr0317 (ccm1)-alr0318 (ccm1)</i> | |
| <i>A. variabilis</i> ATCC 29413 | <i>Ava_4470 (ccm1)-Ava_4471 (MCP)-Ava_4472 (MCP) -Ava_4709 (MCP)-Ava_4710 (MCP)</i> | |
| <i>N. punctiforme</i> STCC 29133 | <i>Npun_F4293 (ccm1)-Npun_F4292 (MCP)-Npun_F4291 (MCP) -Npun_F2745 (MCP)-Npun_2744 (MCP)</i> | |

| | | |
|------------------------------|---|---|
| A2 | <i>T. platensis</i> NIES 39 | NIES39_K04810 (ccml)-NIES39_K04820 (ccmk1)-NIES39_K04830 (ccmk2)-NIES39_A03150 (ccmk4)-NIES39_A03160 (ccmk3) |
| | <i>M. aeruginosa</i> NIES 843 | MAE47920 (ccml)-MAE47930 (ccmk1)-MAE47940 (ccmk2)-MAE55390 (ccmk3)-MAE55400 (ccmk4) |
| | <i>Synechocystis</i> sp PCC 6803 | sl1030 (ccml)-sl1029 (ccmk1)-sl1028 (ccmk2)-slr1838 (ccmk3)-slr1839 (ccmk4) |
| | <i>Cyanothecce</i> sp PCC 7424 | PCC7424_1370 (ccml)-PCC7424_1371 (MCP)-PCC7424_1372 (MCP)-PCC7424_1373 (MCP)-PCC7424_2157 (MCP)-PCC7424_2158 (MCP) |
| | <i>Cyanothecce</i> sp ATCC 51142 | ccc_4281 (ccml)-ccc_4282 (ccmk1)-ccc_4283 (ccmk2)-ccc_2433 (ccmk4)-ccc_2434 (ccmk3) |
| | <i>Cyanothecce</i> sp PCC 8801 | PCC8801_1598 (ccml)-PCC8801_1597 (MCP)-PCC8801_1596 (MCP)-PCC8801_1859 (MCP)-PCC8801_1860 (MCP) |
| | <i>Synechococcus</i> sp PCC 7002 | SYNPCC7002_A1801 (ccml)-SYNPCC7002_A1802 (ccmk1)-SYNPCC7002_A1803 (ccmk2)-SYNPCC7002_A2612 (ccmk)-SYNPCC7002_A2613 (ccmk) |
| | <i>T. erythraeum</i> IMS 101 | Tery_3850 (ccml)-Tery_3851 (MCP)-Tery_3852 (MCP)-Tery_4328 (MCP)-Tery_4329 (MCP) |
| | <i>Anabaena</i> PCC 7120 | all0866 (ccml)-all0867 (ccmk1)-all0868 (ccmk)-alr0317 (ccmk)-alr0318 (ccmk) |
| | <i>A. variabilis</i> ATCC 29413 | Ava_4470 (ccml)-Ava_4471 (MCP)-Ava_4472 (MCP)-Ava_4709 (MCP)-Ava_4710 (MCP) |
| | <i>S. elongatus</i> PCC 6301 | Syc0134 (ccml)-syc0135 (ccmk1)-syc1227 (ccmk4)-syc1228 (ccmk3) |
| | <i>S. elongatus</i> PCC 7942 | Synpcc7942_1422 (ccml)-Synpcc7942_1421 (ccmk)-Synpcc7942_0284 (ccmk3)-Synpcc7942_0285 (ccmk4) |
| | <i>A. marina</i> MBIC 11017 | AMI_5382 (ccml)-AMI_5381 (ccmk1)-AMI_5380 (ccmk)-AMI_0655 (ccmk)-AMI_0656 (ccmk)/AMI_3280 (ccmk)-AMI_5778 (ccmk) |
| | <i>Anabaena</i> PCC 7120 | all0866 (ccml)-all0867 (ccmk)-all0868 (ccmk)-alr0317 (ccmk)-alr0318 (ccmk) |
| | <i>A. variabilis</i> ATCC 29413 | Ava_4470 (ccml)-Ava_4471 (MCP)-Ava_4472 (MCP)-Ava_4709 (MCP)-Ava_4710 (MCP) |
| | <i>Cyanothecce</i> sp ATCC 51142 | ccc_4281 (ccml)-ccc_4282 (ccmk1)-ccc_4283 (ccmk2)-ccc_2433 (ccmk4)-ccc_2434 (ccmk3) |
| | <i>Cyanothecce</i> sp PCC 7424 | PCC7424_1370 (ccml)-PCC7424_1371 (MCP)-PCC7424_1372 (MCP)-PCC7424_1373 (MCP)-PCC7424_2157 (MCP)-PCC7424_2158 (MCP) |
| | <i>M. aeruginosa</i> NIES 843 | MAE47920 (ccml)-MAE47930 (ccmk1)-MAE47940 (ccmk2)-MAE55390 (ccmk3)-MAE55400 (ccmk4) |
| | <i>Cyanothecce</i> sp PCC 8801 | PCC8801_1598 (ccml)-PCC8801_1597 (MCP)-PCC8801_1596 (MCP)-PCC8801_1859 (MCP)-PCC8801_1860 (MCP) |
| | <i>Synechocystis</i> sp PCC 6803 | sl1030 (ccml)-sl1029 (ccmk1)-sl1028 (ccmk2)-slr1838 (ccmk3)-slr1839 (ccmk4) |
| | <i>Synechococcus</i> sp PCC 7002 | SYNPCC7002_A1801 (ccml)-SYNPCC7002_A1802 (ccmk1)-SYNPCC7002_A1803 (ccmk2)-SYNPCC7002_A2612 (ccmk)-SYNPCC7002_A2613 (ccmk) |
| | <i>A. marina</i> MBIC 11017 | AMI_5382 (ccml)-AMI_5381 (ccmk1)-AMI_5380 (ccmk)-AMI_0655 (ccmk)-AMI_0656 (ccmk)-AMI_3280 (ccmk)-AMI_5778 (ccmk) |
| | <i>T. erythraeum</i> IMS 101 | Tery_3850 (ccml)-Tery_3851 (MCP)-Tery_3852 (MCP)-Tery_4328 (MCP)-Tery_4329 (MCP) |
| | <i>A. marina</i> MBIC 11017 | AMI_5382 (ccml)-AMI_5381 (ccmk1)-AMI_5380 (ccmk)-AMI_0655 (ccmk)-AMI_0656 (ccmk)-AMI_3280 (ccmk)-AMI_5778 (ccmk) |
| | <i>A. marina</i> MBIC 11017 | AMI_5382 (ccml)-AMI_5381 (ccmk1)-AMI_5380 (ccmk)-AMI_0655 (ccmk)-AMI_0656 (ccmk)-AMI_3280 (ccmk)-AMI_5778 (ccmk) |
| | <i>S. elongatus</i> PCC 6301 | Syc0134 (ccml)-syc0135 (ccmk1)-syc1227 (ccmk4)-syc1228 (ccmk3) |
| <i>S. elongatus</i> PCC 7942 | Synpcc7942_1422 (ccml)-Synpcc7942_1421 (ccmk)-Synpcc7942_0284 (ccmk3)-Synpcc7942_0285 (ccmk4) | |
| B | <i>T. elongatus</i> BP-1 | tl10945 (ccml)-tl10946 (ccmk1)-tl10947 (ccmk2)-tlr0954 (ccmk3)-tl11596 (ccmk4) |
| | <i>Cyanothecce</i> sp PCC 7425 | Cyan7425_1616 (ccml)-Cyan7425_1617 (ccmk)-Cyan7425_1618 (MCP)-Cyan7425_2087 (MCP)-Cyan7425_2386 (MCP) |
| | <i>Cyanothecce</i> sp PCC 7425 | Cyan7425_1616 (ccml)-Cyan7425_1617 (ccmk)-Cyan7425_1618 (MCP)-Cyan7425_2087 (MCP)-Cyan7425_2386 (MCP) |
| | <i>T. elongatus</i> BP-1 | tl10945 (ccml)-tl10946 (ccmk1)-tl10947 (ccmk2)-tlr0954 (ccmk3)-tl11596 (ccmk4) |

CHAPTER 4

Cloning and Characterization of RuBisCO Encoding Genes of *Gloeobacter violaceus* PCC 7421

4.1. Introduction

Cyanobacteria are among the most important group of prokaryotes, by virtue of their ability to carry out oxygenic photosynthesis. The phylogenetic analysis of the 16S rRNA from the various cyanobacteria, in the present study and previous reports suggests that *Gloeobacter violaceus* is an early diverging cyanobacteria (Seo and Yokota 2003; Hoffmann et al., 2005; Gupta and Mathews 2010) and is hence an important organism with respect to evolution related studies. The *G. violaceus* PCC 7421 (Rippka et al., 1974) was first isolated from limestone rock in Kernwald (Switzerland). *G. violaceus* possess a simple cell organization without thylakoids and an unusual structure of photosynthetic apparatus (Rippka et al., 1974; Guglielmi et al., 1981; Bryant et al., 1981).

G. violaceus has been used as a model organism for studies based on the evolution of oxygenic photosynthesis (Bernat et al., 2012) and is also among the initial set of cyanobacteria whose complete genome was sequenced (Nakamura et al., 2003). The insight into the RuBisCO of *G. violaceus* could facilitate widening of horizons of knowledge about the protein.

RuBisCO (4.1.1.39) is a key enzyme in photosynthesis, catalyzing the carboxylation/oxygenation of Ribulose-1,5-bisphosphate depending upon the environmental conditions and the characteristics of the protein. RuBisCO exists in four basic forms viz. I, II, III and IV, all constituted by the large subunit (L) and the small subunit (S). Form I RuBisCO, found in proteobacteria, cyanobacteria and higher plants consist of the large subunit and the small subunit in an L_8S_8 arrangement. Form II RuBisCO found predominantly in proteobacteria and dinoflagellates, exists in $(L_2)_n$ conformation, i.e. in the complete absence of the small subunit. The form III RuBisCO exists in $(L_2)_n$, is found only in Archaea and is involved in the removal of RuBP, produced as a result of isomerization of ribose bisphosphate in purine/pyrimidine metabolism (Finn and Tabita 2004; Sato et al., 2007). Form IV is called the RuBisCO-like protein (RLP) because of its incapability to catalyze carboxylation reaction. It exists as a dimer of the large subunit in proteobacteria, cyanobacteria, archaea and algae. The enzyme despite its prevalent importance in metabolizing the inorganic carbon into organic forms, since billions of years, suffers from loose specificity and low catalytic rate. In pursuit of overcoming its shortcomings, the enzyme is subject to a tradeoff between specificity and catalytic turnover (Tcherkez et al., 2006). The enzyme has been incapable of removing its drawbacks but has evolved according to its environmental conditions to give the best possible catalysis. The organisms capable of carbon concentrating mechanism have a RuBisCO with low specificity but high turnover and the RuBisCO devoid of carbon concentrating mechanism have high specificity and low turnover rate (Andersson 2008).

RuBisCO being the rate limiting step in the photosynthesis is the prime target in studies involving productivity enhancement of crops. RuBisCO came into origin millions of years ago but has been incapable of overcoming this limitation because, in order to attain a functional carboxylation environment in RuBisCO, the protein demands a balance between a number of factors, viz. the catalytic site, interaction between the LSU and SSU, optimal biogenesis by chaperones and regulation by ancillary protein RuBisCO activase. Hence, in the presence of the acquired evolutionary changes by the subunits, there exists a balanced co-evolution. As a consequence, the kinetic measurements from around 100 species reveal that there exists a considerable variation in the catalytic properties of the protein. Further variation in the properties of the protein definitely exists in the nature, waiting to be discovered by structural and functional analysis. Moreover, in pursuit of developing a RuBisCO with best possible set of characteristics requires a detailed understanding of every aspect of the protein. These studies will help chalk-out the most feasible strategy for bioengineering the protein itself or its immediate environment to attain the goal of an efficient enzyme. Having said about the importance of RuBisCO and expectations of solving the problem of impending food scarcity, no wonder, there exist a history of more than 70 years (**Figure 4A**) for the protein and yet newer concepts about it are discovered each year.

The present study is an attempt to express RuBisCO from early diverging cyanobacteria, *G. violaceus* PCC 7421 in *E. coli*. In accordance with previous reports on the importance of expressing RuBisCO form I as an operon rather than separate subunits in order to obtain soluble protein, *G. violaceus* RbcL was expressed both in the presence and absence of the chaperone protein RbcX and small subunit RbcS. Further, the effect of the presence of RbcS on the catalysis of RbcL was also analyzed *in-vitro*.

Figure 4A: Timeline of the important discoveries and observations about the protein RuBisCO from 1947 to 2016 (Page 4-7). While efforts have been made to include all the major events, this analysis will be inevitably incomplete and would reflect the research of specific interest to the author. (Figure on next page)

Wildman and Bonner (1947) isolated a protein from leaves by ammonium sulphate precipitation, the fraction of which was elucidated a decade later

1947

Benson (1951) reported that RuBP is an early product of photosynthesis and it undergoes oxidation to form PGA and phosphoglycolate

1951

Wilson and Calvin (1955) provided data to support that RuBP undergoes carboxylation

1955

Based on similarity in physical properties Wildman speculated that the fraction I protein purified in 1947 was same as the carboxylation enzyme reported by **Weissbach et al (1956)**.

1957

Melvin Calvin and Andrew Benson identified early products of photosynthesis (**Calvin and Benson 1948, 1949; Benson et al 1950**) **Tamiya and Huzisige (1949)** reported that CO₂ and O₂ compete in photosynthesis

1948-50

Basylum et al (1954) successfully elucidated the complete cyclic pathway of photosynthesis **Quayle et al (1954)** and **Weissbach et al (1954)** reported carboxylation activity in extracts from Chlorella and Spinach respectively

1954

Purification of phosphoribulokinase (**Huwitz et al 1956**) synthesis of RuBP using phosphoribulokinase (**Horecker et al 1956**) and purification of enzyme involved in carboxylation of RuBP to form PGA (**Weissbach et al 1956**) **Ja-Koby et al (1956)** also reported synthesis of RuBP and PGA in presence of partially purified enzyme from spinach leaves

1956

Moses and Calvin (1958) attempted to purify the 6 carbon intermediate 2-carboxy-3-ketopentitol-1,5-bisphosphate, which was successfully trapped in 1982 by **Schloss and Lorimer (1982)**

1958

Electron micrographs of spinach chloroplast were reported to show localization of fraction I protein in chloroplast stroma (**Park and Pon 1961**).

1961

Reports appeared for the activation of carboxylating enzyme in the presence of magnesium and bicarbonate (**Pon et al 1963; Akoyunoglou and Calvin 1963**)

1963

Haselkorn et al (1965) reported the first high resolution electron micrograph of fraction I protein showing a depression today identified as the solvent channel hole **Trown (1965)** and **Thorner et al (1965)** demonstrated fraction I to be a crude enzyme preparation

1965

Rutner and Lane (1967), Sugiyama and Akazawa (1967) reported that the spinach ribulose is formed of two non-identical subunits

1967

On the basis of observation of increase in glycolate formation in presence of oxygen in Chlorella, **Basylum and Kirk (1962)** suggested the possible reason for the same to be oxidation of an early intermediate of photosynthesis

1962

Mullhofer and Rose (1964) demonstrated that the formation of PGA by carboxylation of RuBP involves cleavage of RuBP between 2nd and 3rd carbon of the chain

1964

Palsen and Lane (1966) successfully purified ribulose from spinach (free from contamination from other photosynthetic proteins) and determined the molecular weight to be 557,000 daltons

1966

First report on limitation of RuBisCO in carboxylation reaction appeared (**Waring et al 1968; Bjorkman 1968a, b**)

1968

Anderson et al (1969) reported ribulose from some microorganisms to have molecular weight different from fraction I protein

16

2

4

6

8

9

11

13

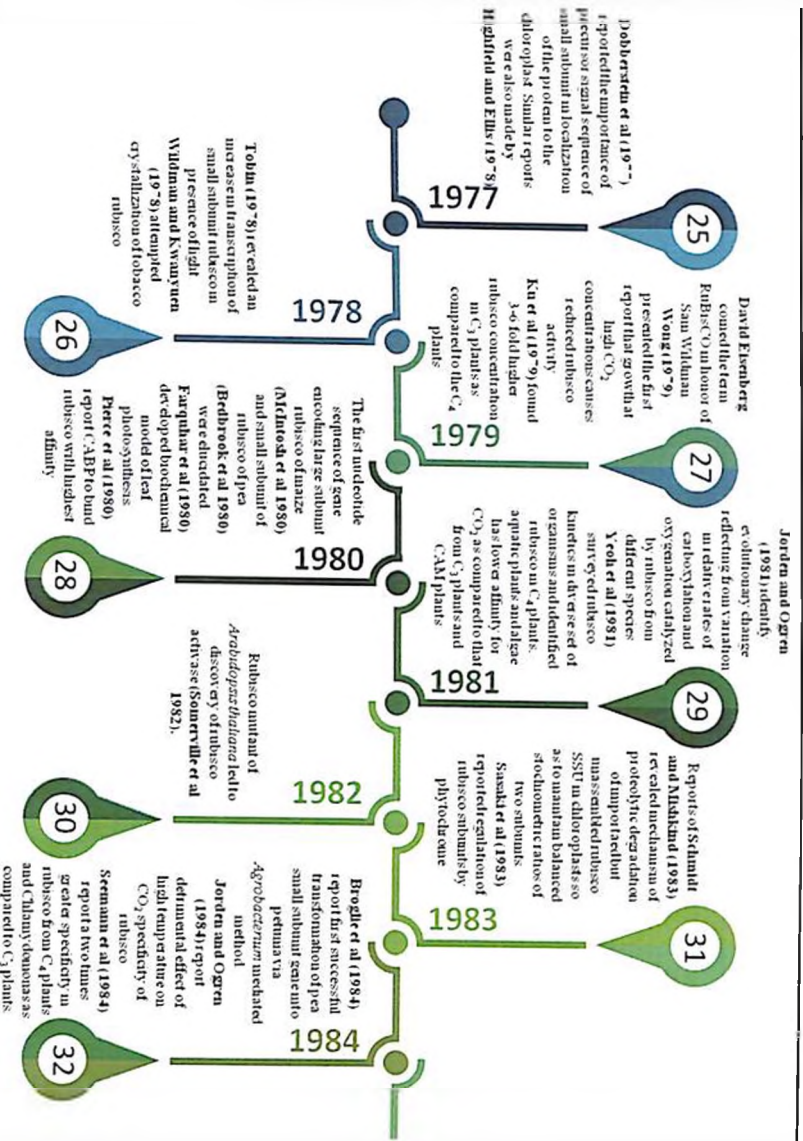
15

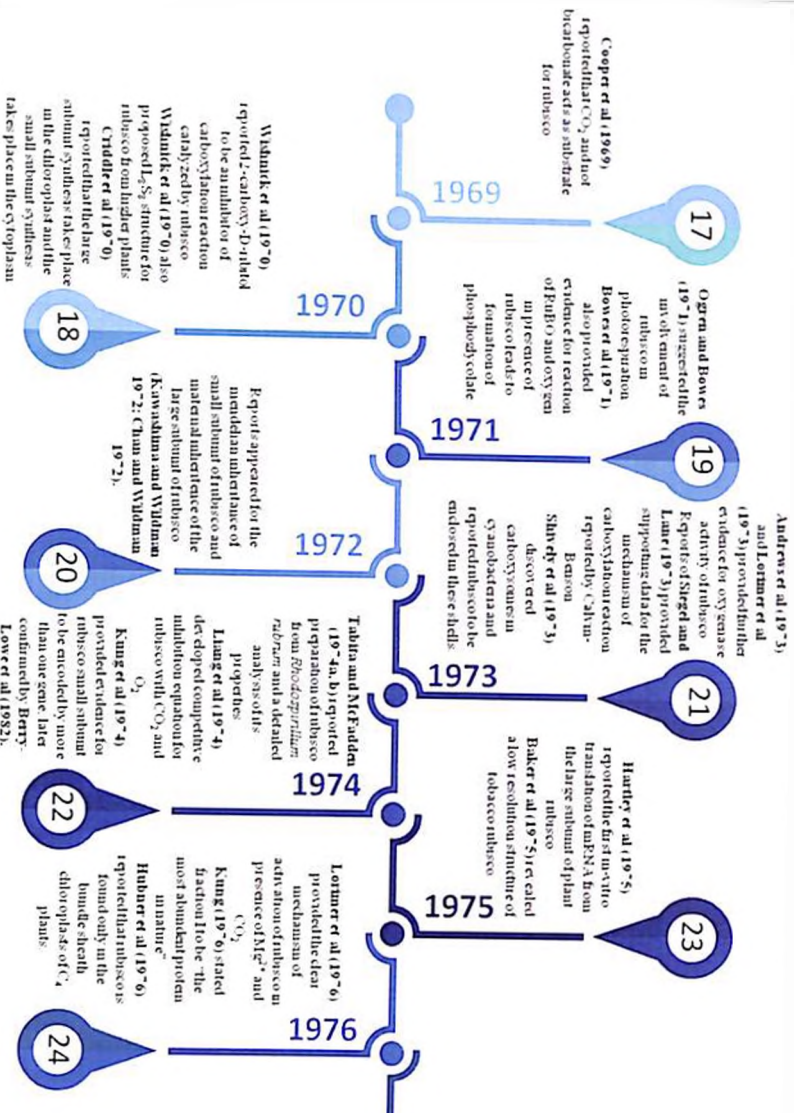
10

12

14

16





17

Cooper et al (1969)
reported that $^{14}\text{C}_2$ and not
bicarbonate acts as substrate
for rubisco

19

Ogren and Bowers
(1971) suggested the
involvement of
rubisco in
photorespiration
Bowers et al (1971)
also provided
evidence for reaction
of PEPCase and oxygen
in presence of
rubisco leads to
formation of
phlophosphoglycolate

21

Andrews et al (1973)
and Lottner et al
(1973) provided further
evidence for oxygenase
activity of rubisco
Reports of Siegel and
Lamer (1973) provided
supporting data for the
mechanism of
carboxylation reaction
reported by Calvin-
Benson
Shrivastya et al (1973)
discovered
carboxyoxenium
cyanobacteria and
reported rubisco to be
endosymbiotic in these shells

23

Hartley et al (1975)
reported the first in-vitro
transamination of mPNA from
the large subunit of plant
rubisco
Baker et al (1975) related
a low resolution structure of
tobacco rubisco

1969

1970

1971

1972

1973

1974

1975

1976

Widmann et al (1970)
reported 2-carboxy-D-ribidol
to be an inhibitor of
carboxylase reaction
catalyzed by rubisco

Widmann et al (1970) also
proposed C_2 structure for
rubisco from higher plants

Criddle et al (1970)
reported that the large
subunit synthesis takes place
in the chloroplast and the
small subunit synthesis
takes place in the cytoplasm

18

Reports appeared for the
meridian whereance of
small subunit of rubisco and
material evidence of the
large subunit of rubisco
(Kawashima and Widmann
1972; Chan and Widmann
1972).

20

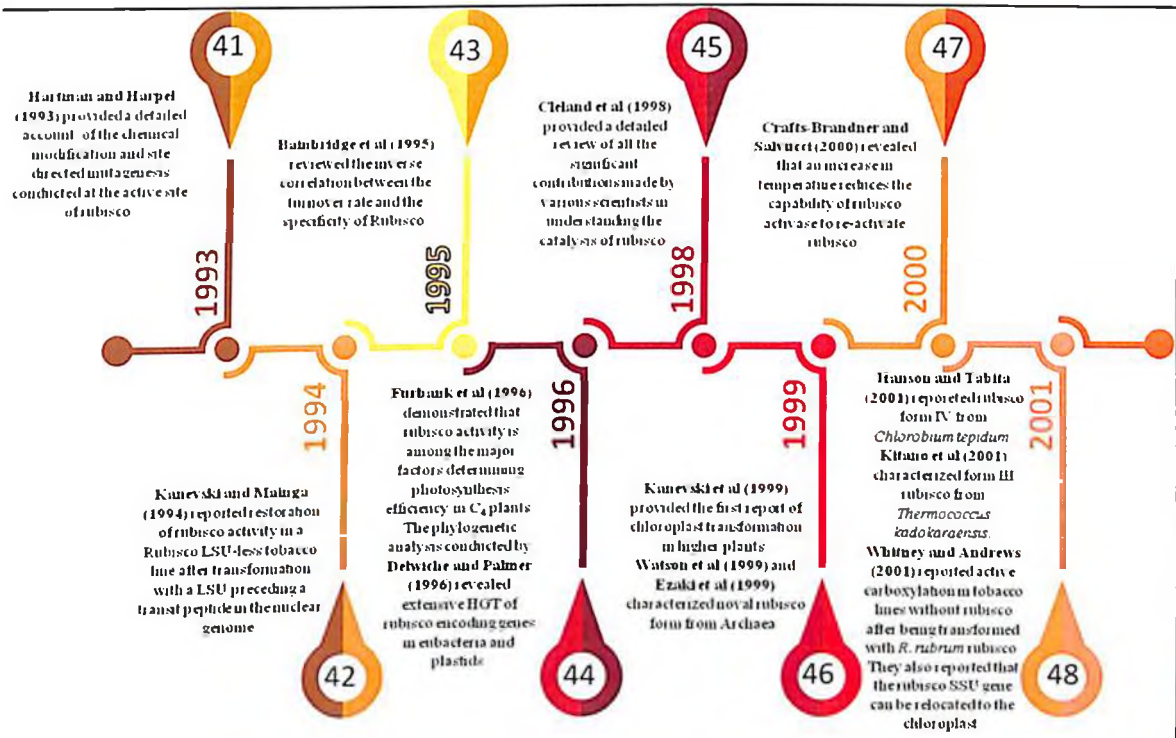
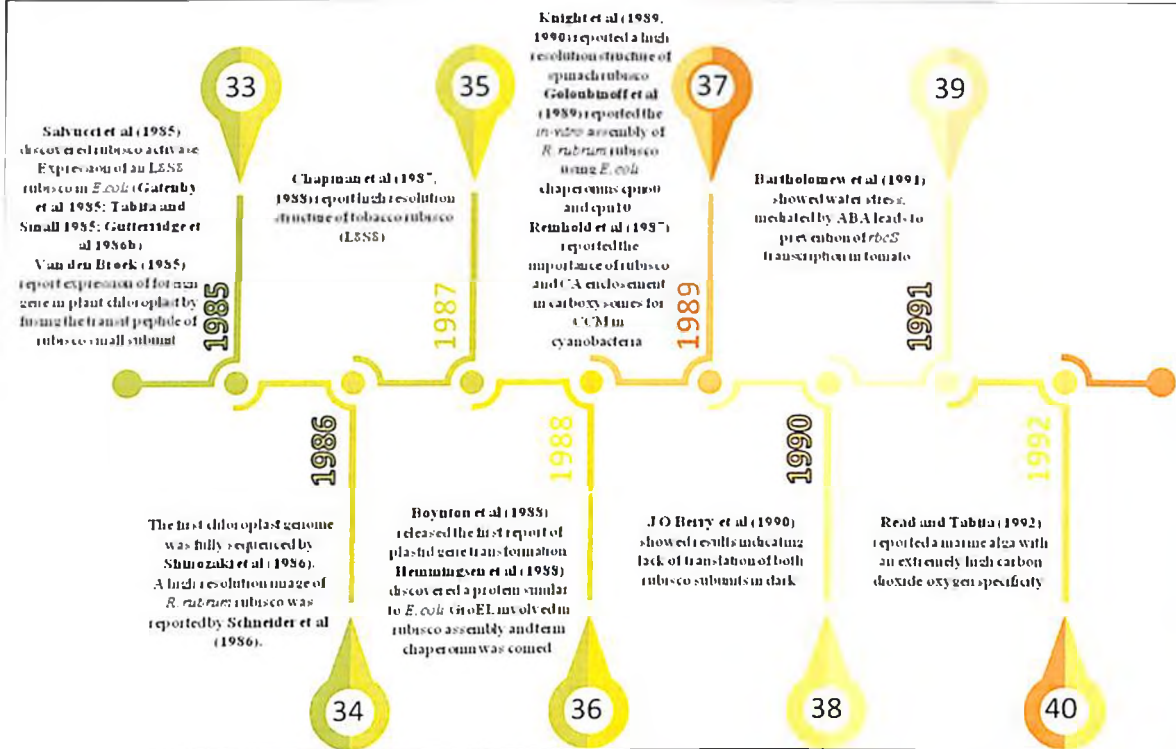
Tabita and McFadden
(1974a, b) reported
preparation of rubisco
from *Rhodospirillum
rubrum* and a detailed
analysis of its
properties
Liang et al (1974)
developed competitive
inhibition equation for
rubisco with CO_2 and
 O_2

22

Lottner et al (1976)
provided evidence for
rubisco small subunit
to be encoded by more
than one gene later
confirmed by Berry-
Lower et al (1982).

24

Kung (1976) stated
fraction II to be the
most abundant protein
in nature
Hubner et al (1976)
reported that rubiscos
found only in the
bundle sheath
chloroplasts of C_4
plants



Lin et al (2010)
demonstrated the role of diapaerases in assembly of hexadecameric ribisco. **Wittre et al (2010)** reported isolation of functional and novel ribisco obtained from genes cloned and characterized from oceanic metagenomic library.

57

2010

Kapranov et al (2012)
provided evidence to support the hypothesis that parallel amino acid replacements are associated with adaptive changes in ribisco. **Suzuki et al (2012)** identified that RBCs positively regulates the expression of hcd at the transposon level. **Meyer et al (2012)** reported the role of SST helices in pyruvate formation.

59

2012

Lin et al (2014)
successfully generated tobacco lines with cyanobacterial ribisco showing higher rates of carboxylation as compared to the wild types. **Joshi et al (2014)** reported the crucial role of SST in assembly of red-type ribisco, independently of any diapaerases.

61

2014

Srinagappan and Tabita (2016)
conducted ribisco selection using vesicle bacterium *Ralstonia eutropha*. **Markkander et al (2016)** discovered essential pyruvate component I (EPYCI), a protein centrally involved in assembly of ribisco into pyruvates in *Chlamydomonas*.

63

2016

Murderer-Cajnar et al (2011)
and **Storz et al (2011)** characterized ribisco activase for red-type and green-type ribisco, respectively.

2011

Loth et al (2013)
reported that ribisco is not the most abundant protein in phytoplankton. **Wächter et al (2013)** suggested that ribisco SST does not effect the species-specific interactions between ribisco and ribisco activase.

2013

Whitney et al (2015)
suggested the importance of ancillary proteins involved in recombinant ribisco biogenesis. **Tchirikov (2015)** discussed and analyzed the available data to explain the mechanism of oxygenation catalyzed by ribisco.

2015

Brancher et al (2015)
suggested co-evolution of diapaerone RBCX along with its cognate ribisco for optimal functioning.

2015

Coutton et al (2011)
demonstrated the importance of bacterial importase of bacterial ribisco in rhizobial-legume interaction.

2011

Morita et al (2013)
provided the first report on organ specific expression of RBCs leading to enhancement of ribisco activity.

2013

Brancher et al (2015)
suggested co-evolution of diapaerone RBCX along with its cognate ribisco for optimal functioning.

2015

Brancher et al (2015)
suggested co-evolution of diapaerone RBCX along with its cognate ribisco for optimal functioning.

2015

Blayney et al (2011)
re-evaluated interactions of ribisco with nucleotides and sugar phosphates and their implications.

2011

Morita et al (2013)
provided the first report on organ specific expression of RBCs leading to enhancement of ribisco activity.

2013

Brancher et al (2015)
suggested co-evolution of diapaerone RBCX along with its cognate ribisco for optimal functioning.

2015

Brancher et al (2015)
suggested co-evolution of diapaerone RBCX along with its cognate ribisco for optimal functioning.

2015

58

60

62

Sage (2002) reviewed the effect of temperature and CO_2 on C_3 and C_4 plant rubisco. Further, the evolution of rubisco properties towards improvement in performance was also reported.

49

2002

Pain and Tabita (2004) identified an alternative route of formation of PEP in *E. coli* rubisco, but form III rubisco, but devoid of PEP. Onizuka et al. (2004) demonstrated the role of PEP in enhancement of production and assembly of rubisco in *E. coli*.

51

2004

Parth et al. (2006) reported improved catalytic activity in recombinant *Synechococcus rubisco* by its mutation and directed evolution in *E. coli*. Carré-Alfonso et al. (2006) discovered *Microcystis* to contain both form I and form IV rubisco.

53

2006

Parry et al. (2005) reviewed the regulation of rubisco from the perspective of role of inhibitors. In an evolutionary analysis of forms of rubisco Tabita et al. (2005) suggested that both rubisco and PLP appear to have originated from an ancestral protein of methanogenic archaea.

55

2008

2009

Eichmann et al. (2009) showed correlation between rubisco and photosynthetic electron transport proteins.

Ahida et al. (2003) reported PLP from *Bacillus subtilis* and demonstrated its role in methionine salvage pathway.

50

2003

Spreitzer et al. (2005) reported the enzymes that could potentially provide plant like properties to rubisco from *Chlamydomonas* Li et al. (2005) identified residues in rubisco activase with role in substrate recognition in rubisco.

52

2005

Saschenbrecker et al. (2007) provided an insightful detail of the role of PEP as a chaperone in rubisco assembly. Tamaka et al. (2007) reported the structure of rubisco chaperone PEP from *Synechococcus* Baranowski and Sire (2007) reported crystallization of rubisco from red alga.

54

2007

Whitney et al. (2009) expressed tobacco rubisco by linking the LST and SST encoding sequences in the chloroplast, leading to an enzyme with an enhanced specificity. Genkov and Spreitzer (2009) identified residues of SST of *Chlamydomonas* which improve its action with LST affect catalytic

56

4.2. Materials and Methods

4.2.1. *In-Silico* sequence analysis

4.2.1.1. *RuBisCO diversity and co-evolution of constituting proteins*

The amino acid sequences for RbcL, RbcS and RbcX from 37 cyanobacteria available in Kazusa Genome Resource were retrieved and analyzed for conserved residues (alignment by T-Coffee server; sequence logo by WebLogo3) and phylogenetic analysis (using MEGA 6.0). Further, the RbcL, RbcS and RbcX sequences from β cyanobacteria were used to compute evolutionary distance of a protein from the corresponding protein in *G. violaceus*, considering the fact that the later is an early diverging cyanobacteria. The estimates of evolutionary divergence between sequences were calculated using MEGA 6.0 by the equal input model. These values were then plotted simultaneously for the three proteins to understand the trend of evolution among the proteins.

4.2.1.2. *Analysis of locus of RuBisCO Subunit encoding genes (as also mentioned in section 2.1.3)*

In order to ascertain the role of *rbcX*, we analyzed genome of various Cyanobacteria retrieved from KAZUSA genome resource for the analysis of the location of *rbcS*, *rbcX*, and *rbcL* and whether these subunits are located adjacent to each other or are dispersed in the genome. The analysis also aimed to implicate whether the distance of RbcX with respect to RbcL and RbcS is making any significant contribution towards its function.

4.2.1.3. *Scanning different Cyanobacterial species for RbcX*

Several of the cyanobacterial species were found to lack Rbcx encoding genes. The absence of RbcX gene was confirmed by browsing the individual genes as well as by genome blast. RuBisCO being a multimeric protein is expected to require a chaperone for folding of an active protein. The absence of RbcX in several Cyanobacteria directs the thought process towards the idea, that there exist certain chaperones which are yet to be identified. To investigate this, the organism devoid of RbcX were analyzed for the chaperone present in them. The chaperone encoding gene sequences or proteins sequences were retrieved from Kazusa genome resource and aligned in MEGA 5.1 software with RbcX to identify any similarity in sequences.

4.2.2. Construction of expression plasmid constituting RuBisCO encoding genes of *G. violaceus* (polymerase chain reaction and cloning according to protocols mentioned in chapter 2)

The RuBisCO large subunit (*rbcL*), RuBisCO small subunit (*rbcS*) and RuBisCO chaperone (*rbcX*) encoding genes of *G. violaceus* PCC 7421 were amplified using the genome of the same by polymerase chain reaction. The *RuBisCO operon* was also amplified by polymerase chain reaction from the genome. The primers used for the amplifications carried out for the present study are mentioned in chapter 2 (Table 2C). The sequences for primer designing were retrieved from Kazusa genome resource (<http://genome.microbedb.jp/CyanoBase>) database. The PCR conditions were standardized for *rbcL*, *rbcS*, *rbcX* and *RuBisCO operon* using Taq polymerase (Merck). The conditions used are mentioned in chapter 2 (Table 2D). The expected size of the PCR products from these amplifications are 321bp, 1425bp, 378bp and 2124bp for *rbcS*, *rbcL*, *rbcX* and *RuBisCO operon*.

The PCR products hence obtained (Figure 4E) were purified (Qiagen Gel Extraction kit) and digested with respective restriction enzymes viz. *NdeI* and *BamHI* or *NdeI* and *XhoI* (New England Biolabs) depending on the restriction site introduced into the primers. The digested products obtained were purified and ligated into appropriate sites of *pCOLDII* and *pET15b*, downstream the *espA* and T7 promoters, respectively, using T4 DNA ligase (Promega) by standard ligation procedure. The gene sequence for His-tag present between the promoter and the start codon of the recombinant gene shall be translated on the N-terminal of the protein synthesized. The sequences of the plasmids obtained after cloning were confirmed by sequencing analysis. The constructs generated viz. *pCOLD-Gv-rbcL*, *pCOLD-Gv-rbcS*, *pCOLDII-Gv-rbcX*, *pCOLDII-Gv-rbcLSX*, *pET15b-Gv-rbcL*, *pET15b-Gv-rbcS*, *pET15b-Gv-rbcX* and *pET15b-Gv-rbcLSX* were transformed into *E. coli* BL21(DE3)pLysS cells and the protein expression was analyzed under IPTG induction.

4.2.3. Expression of recombinant plasmids in prokaryotic system

The *E. coli* BL21(DE3)pLysS cells transformed with *pET15b* and *pCOLDII* constructs were inoculated into Luria-Bertani broth containing ampicillin (100µg/mL) and chloramphenicol (34µg/mL) and incubated at 37°C for overnight, followed by secondary inoculation into Luria-Bertani broth containing ampicillin (100µg/mL) and chloramphenicol (34µg/mL) @ 1% and incubated at 37°C till the OD₆₀₀ reaches 0.3-0.4. Then the culture was induced for recombinant protein expression by cold shock and/or addition of different concentrations IPTG. The IPTG concentrations used to evaluate best expression conditions were 0.1, 0.5 and 1mM, followed by incubation at 15°C, 25°C and 37°C. The *pCOLDII* constructs showed

better expression when grown at sub-optimal temperatures and were hence used for further procedures.

4.2.4. Test for the solubility of the protein expressed *in-vivo* in native and/or denaturing buffer (with reference to section 2.2.11 in chapter 2)

The 12-14 hrs culture of the *E. coli* BL21(DE3)pLysS strain carrying the expression plasmids under study was inoculated into Luria-Bertani broth @ 1% containing 100mg/mL ampicillin and 34mg/mL chloramphenicol. The culture was grown at 37° C, with vigorous shaking at 150rpm until the OD₆₀₀ of 0.5-0.6 was obtained. The gene expression was induced by addition of IPTG at final concentration of 0.5mM. The cells were harvested after 18hrs incubation at 18° C, 150rpm by centrifugation at 7800rpm, 4° C for 20 minutes. The cell pellet obtained was stored at -80° C. The cells were resuspended in buffer 1 (Imidazole – 5mM, Sodium chloride – 300mM, Sodium Phosphate pH 8.0 – 50mM, Glycerol 10%) in presence of 1mM PMSF (Sigma-Aldrich) and 1X protease inhibitor cocktail (Clontech). The cell lysis was facilitated by lysozyme (Sigma-Aldrich) treatment at 30° C with continuous shaking, followed by incubation on ice for 30 minutes. The cells were then lysed by sonication (30 X 10 seconds, 10KHz, Ultrasonic Homogenizer Biochem Lifesciences) on ice and centrifuged at 7850rpm, 4° C for 20 minutes. The supernatant obtained was collected, labeled as Supernatant A and stored at 4° C. The pellet obtained was washed in buffer 2 (Imidazole – 5mM, Sodium chloride – 300mM, Sodium Phosphate pH 8.0 – 50mM, Triton-X 100 – 0.5%) and centrifuged at 7850rpm, 4° C for 20 minutes. The supernatant obtained was collected in fresh tube, labeled as supernatant B and stored at 4° C. Finally the pellet obtained was resuspended in buffer 3 (Imidazole – 20mM, Sodium chloride – 300mM, Sodium Phosphate pH 8.0 – 50mM, Urea – 8M, Glycerol 10%), sonicated (30 X 10 seconds, 10KHz, Ultrasonic Homogenizer Biochem Lifesciences) and centrifuged at 7850rpm, 4° C for 20 minutes. The supernatant obtained was collected, labeled as supernatant C and stored at 4° C. The supernatants collected i.e. A, B and C were then analyzed on 12% SDS-PAGE.

4.2.5. Purification of *G. violaceus* PCC 7421 RuBisCO subunits by IMAC (with reference to section 2.2.12.4 in chapter 2)

The supernatant C was used for protein purification as the desired protein was found to be soluble in buffer 3. The supernatant was incubated with pre-equilibrated (with buffer 3) BD Talon Co² resin (Clontech) for 2 hrs with continuous shaking at 25° C, so as to allow the binding of the histidine tagged protein to the resin. The resin was then washed with wash buffers 4a-4f (with reducing concentrations of urea), 5 volumes, each. The wash buffer 4a constitutes Imidazole – 20mM, Sodium chloride – 300mM, Sodium Phosphate pH 8.0 – 50mM, Urea – 6M; while buffers 1-5 (Table 2O and W) have same composition except for

urea concentrations i.e. 4M, 2M, 1M, 0.5M and 0M, respectively for 1-5. The metal bound recombinant protein when washed with reducing denaturant concentrations gets a renaturing environment without the possibility of abrupt aggregation of the proteins. The protein is finally eluted in buffer 5 (Imidazole – 1M, Sodium chloride – 300mM, Sodium Phosphate pH 8.0 – 50mM), 5 volumes in 500 μ l aliquots. The protein concentration in mg/mL was determined by taking absorbance at 280nm and dividing it by the absorption coefficient of the protein. The purity of the eluted fractions was determined by 12% SDS-PAGE. The fractions with purified protein of interest were pooled and used for further procedures.

The eluted protein was then centrifuged at 10000 rpm, 4°C for 20 minutes and the supernatant was exchanged into thrombin buffer (Tris-Cl pH 8.0 – 20mM, Sodium Chloride – 150mM, Calcium Chloride – 2.5mM) using PD10 column (Banglore Genei). The desalted protein was then treated with thrombin (1U/ μ g of protein; Calbiochem) at 16°C for 16 hrs. After thrombin treatment, the protein was exchanged into storage buffer (Tris-Cl pH 8.0 – 20mM) and concentrated 50 times using Amicon centrifugal filter devices (10K cutoff). The concentrated protein was stored in aliquots at -80°C in presence of 20% glycerol.

4.2.6. The MS/MS and CD analysis of *G. violaceus* RuBisCO proteins purified by IMAC (with reference to section 2.2.13 in chapter 2)

The purified protein was run on SDS-PAGE, the desired band excised and in-gel digestion was carried out (ThermoScientific In-gel tryptic digestion kit). The overnight digested sample was then used for MS analysis. Target protein identification is done by Matrix Assisted Laser Desorption Ionization-Time of Flight-Mass Spectrometry (MALDI-TOF-MS) (4800 plus MALDI TOF/TOF Analyzer, AB SCIEX, USA).

Further, the CD spectra of *G. violaceus* RuBisCO (RbcL) were recorded between 190 and 240nm with a JASCO J-815 spectropolarimeter, in a 0.1cm pathlength suprasil quartz cuvette at 20°C, scan rate of 50nm min⁻¹, 1nm bandwidth, calibrated with 0.06% (w/v) aqueous solution of (\pm)-10-camphorsulfonic acid. The spectra were corrected for baseline and the secondary structure prediction analysis was done using DICHROWEB.

4.2.7. RuBisCO spectrophotometric assay (with reference to section 2.2.15 in chapter 2)

RuBisCO assay was performed at 25°C in a medium containing 100mM HEPES-KOH pH 8.0, 20mM MgCl₂, 50mM NaHCO₃, 0.2mM NADH, 5mM ATP, 20mM Phosphocreatine, 12.5U per mL PCK, 11.25U per mL PGK, 10U per mL GAPD and 0.05mg CA. The initial activity was determined by adding 6 μ g of the protein to 1mL reaction mixture. The reaction without RuBP was taken as the baseline for taking absorbance till the OD₃₄₀ stabilized, following which, 2mM RuBP (final concentration) was added and change in absorbance observed. The protein was activated in presence of activation buffer (HEPES-KOH pH 8.0 –

50mM, $MgCl_2$ – 25mM, EDTA – 1mM, $NaHCO_3$ – 25mM freshly added) at 25° C for 60 minutes. The absorbance at 340nm was measured using Shimadzu spectrophotometer. The Michaelis-Menten constant for RuBP (K_{RuBP}) was determined by addition of different amounts of RuBP to bring final concentration from 0.01-3mM.

4.2.8. *In-silico* modeling (with reference to section 2.1.5 in chapter 2)

The *in-silico* model was generated for RbcL and RbcS of *G. violaceus* using Modeller tool. The models generated were subjected to energy minimization (Modrefiner) and then validated by online server (PDBsum). The inter subunit interactions were estimated using online server Grammx are mentioned in chapter 6.

4.3. Results and Discussions

4.3.1. The protein RbcL is most conserved while RbcS and RbcX more diverse

If we compare the α cyanobacteria RbcL, the MSA shows very little variation with as much as 90% of the sequence conserved. Further, the β cyanobacteria RbcL when aligned, show ~70% residues conserved. The most variable region is the initial 10-12 residues. The regions of the protein sequence harboring the active site residues i.e. Lys174, Kex200, Asp202, Glu203, His293, Lys333 (in *G. violaceus* RbcL identified by aligning with *T. elongatus* RbcL and *S. oleracea* RbcL, whose active sites are characterized) are conserved in all β cyanobacteria RuBisCO. Further, the substrate binding sites T172, K176, R294 and H326 in *G. violaceus* RbcL (identified by aligning with RbcL of *T. elongatus*) and metal binding sites D202 and E203 in *G. violaceus* (identified by aligning with *T. elongatus* RbcL) were also completely conserved. The Form IA and Form IB RuBisCO when aligned show ~60% sequence conserved, with functionally important residues completely conserved.

The alignment of RbcS sequence from α cyanobacteria shows ~80% residues conserved while the β cyanobacteria RbcS is very variable showing only 33.3% conservation of sequence upon multiple sequence alignment. Further, when α and β cyanobacterial RbcS was aligned together, the MSA showed major difference in the initial 10-15 residues of the sequence. A stretch of 6 residues, 'OSTVGD' is found in α cyanobacterial RbcS only. The amino acid sequence of all cyanobacterial RbcS when aligned shows a mere 17% conservation.

The chaperone protein viz. RbcX in β cyanobacteria and CbbX in α cyanobacteria have different story from their counterparts in the *RuBisCO operon*. RbcX is found in all β cyanobacteria while CbbX is found only in 5 out of 17 α cyanobacteria used in the present analysis. Apart from these 5, proteins annotated as ATPase in *Synechococcus* sp PCC9605 and *Synechococcus* sp PCC9902, show 79.5% identity to the CbbX protein and hence are possibly RuBisCO chaperonins. The other 10 α cyanobacteria do not have any homolog for

the CbbX protein. The closest homologs for CbbX in these cyanobacteria with an identity of 24-31% are holliday junction DNA helicase B, ATPase AAA family and cell division protein FtsH.

The CbbX sequence from 7 α cyanobacteria is 62% conserved, while the RbcX sequence from 19 β cyanobacteria is merely 12% conserved. Furthermore, the CbbX and RbcX sequences when aligned showed an insignificant degree of conservation i.e. 1.3%. This analysis clarifies a few things about the RuBisCO chaperonin which are mentioned below. Firstly, it is not an ultimately important factor involved in RuBisCO biogenesis as a number of α cyanobacteria are completely devoid of it. Secondly, the chaperone as the RbcL and RbcS of α cyanobacteria is much conserved while RbcX is not. Thirdly, although the RbcL of α and β cyanobacteria is 60% conserved, the RbcS and RbcX/CbbX in these two group of cyanobacteria when compared are quite different from each other. The RbcX of β cyanobacteria shows 27-42% identity to proteins like ABC transporter, cell division protein FtsH, DNA topoisomerase I, etc in α cyanobacteria. Further the CbbX of α cyanobacteria shows identity to proteins like putative ATPase, cell division protein FtsH, etc in β cyanobacteria. This shows the α and β cyanobacteria have different set of chaperone proteins which possibly have a common ancestor. It should be noted that in this search, the only two proteins that showed similarity to both the set of chaperones were ATPase proteins from *Synechococcus* sp PCC 9605 and *Synechococcus* sp PCC 9902 showing ~80% identity to CbbX and 36% identity to RbcX. **Figures 4B(i-iii)** show the conserved regions of RbcL, RbcS and RbcX, respectively. The sequence logo shows major portion of RbcL is conserved while that of RbcS and RbcX is less conserved.

4.3.2. The RuBisCO operon seems to have co-evolved in most β cyanobacteria, but not in α cyanobacteria

These observations show RbcL RuBisCO to be highly conserved, while the RbcS and RbcX/CbbX have undergone a great degree of variation. Since, the functional RuBisCO constitutes both the LSU and the SSU, in the presence of the chaperone, the three proteins should have co-evolved. Considering the fact that *G. violaceus* is an early diverging cyanobacteria, the distances of all cyanobacterial LSU, RbcX and SSU RuBisCO (separately) were computed from the former's sequence in order to identify the pattern of evolution of the two proteins (**Figure 4C(i and ii)**). The analysis showed co-evolution of the proteins in all β cyanobacteria except in case of *Arthrospira platensis* NIES39, *Synechococcus* sp JA-3-3B and *Synechococcus* sp JA-2-3Ab(2-13). While the pattern of evolution of α cyanobacterial LSU RuBisCO and SSU RuBisCO does not match with each other in most of the organisms under study. But it should also be noted that the identity in α cyanobacterial LSU and SSU

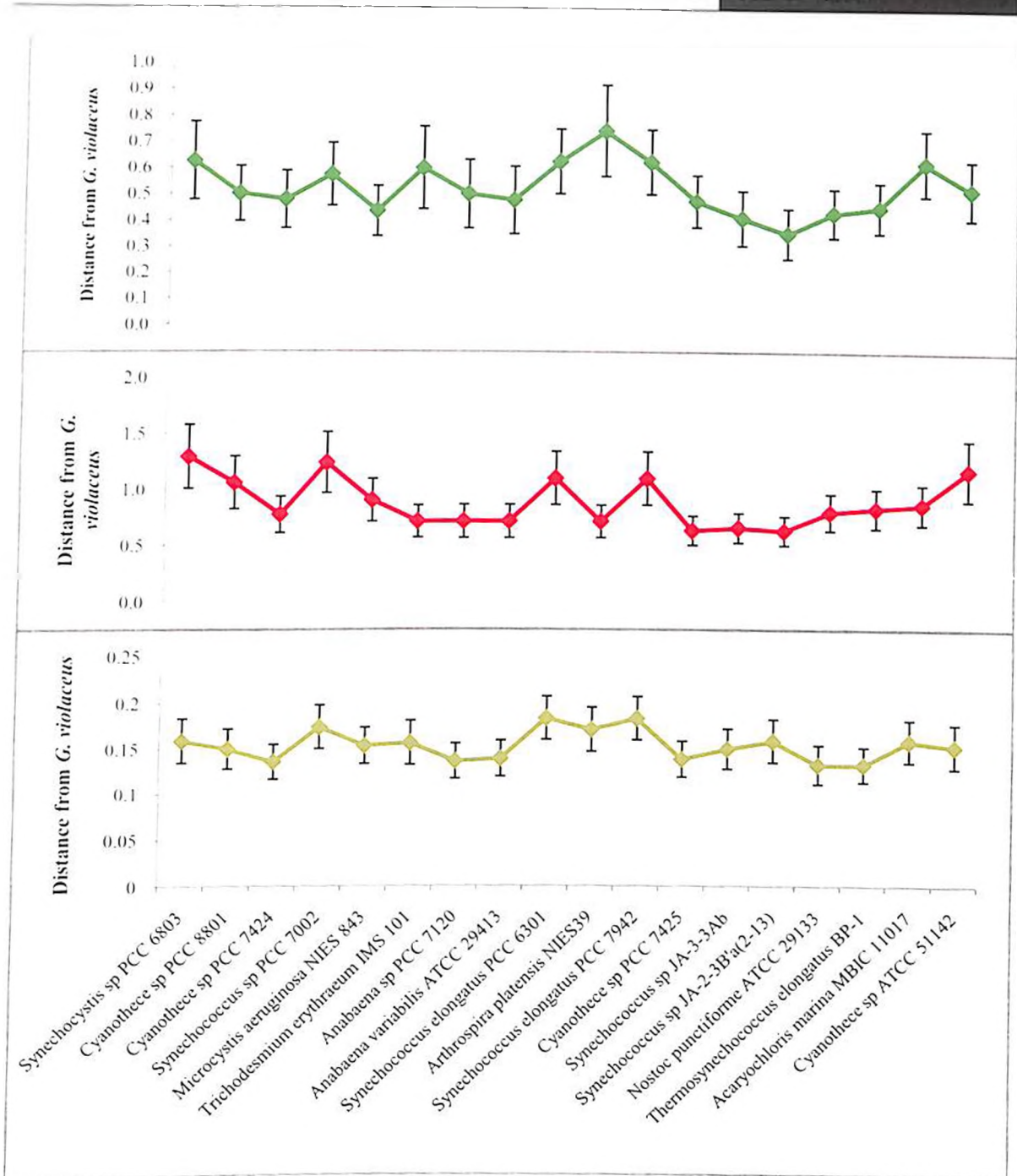


Figure 4C(i): Evolution pattern of RbcL (bottom panel-light green), RbcX (middle panel-red) and RbcS (top panel-dark green) in β cyanobacteria. The evolutionary distance of each of the proteins from corresponding protein in *G. violaceus* in the 19 cyanobacteria was plotted simultaneously in order to observe pattern of evolution. The pair wise distances were computed in MEGA6.0 using the equal input model. The bars on each plotted point show the standard error.

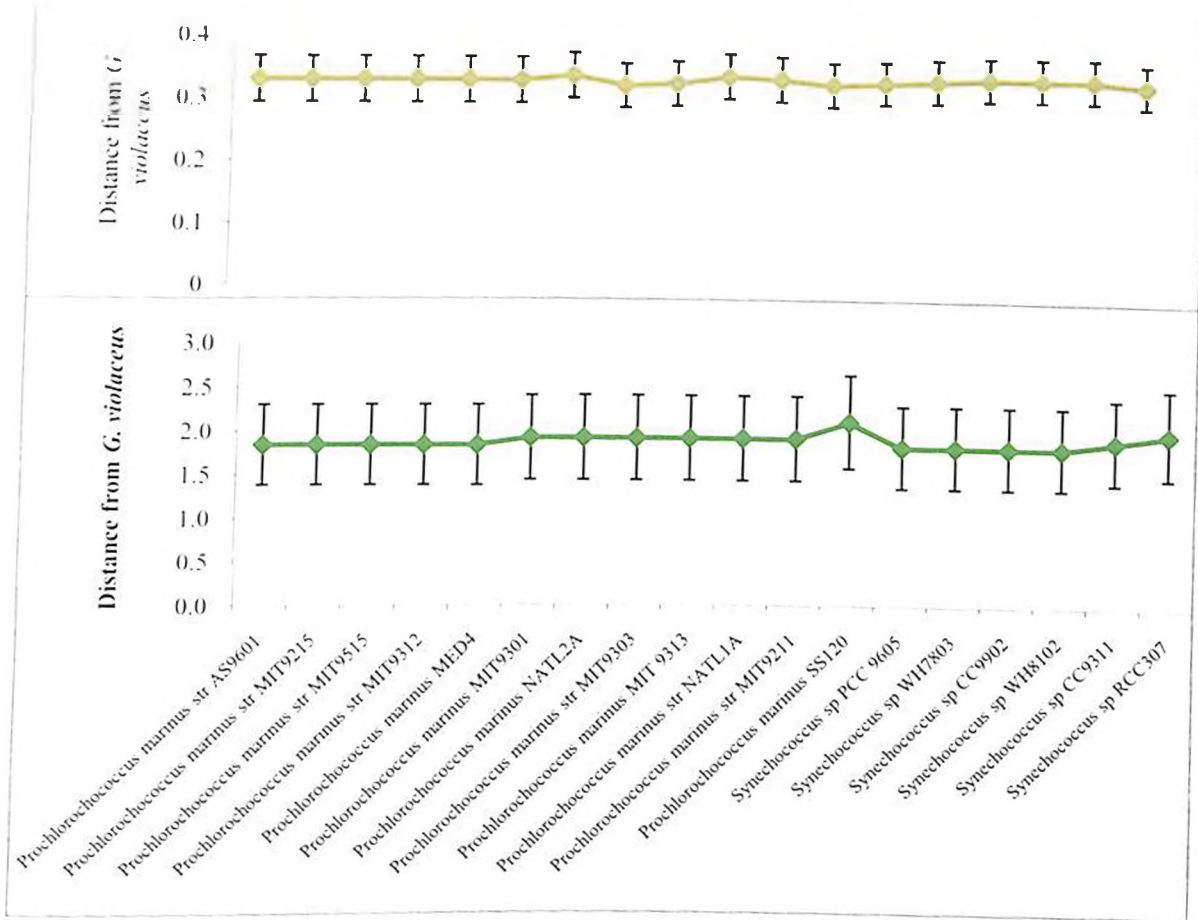


Figure 4C(ii): Evolution pattern of RbcL (top panel-light green) and RbcS (bottom panel-dark green) of α cyanobacteria. The evolutionary distance of each of the proteins from corresponding protein in *G. violaceus* in the 17 cyanobacteria was plotted simultaneously in order to observe pattern of evolution. The pair wise distances were computed in MEGA6.0 using the equal input model. The bars on each plotted point show the standard error. The CbbX sequences could not be included in the graph as it is found only in 5 out of 17 cyanobacteria used in present analysis.

4.3.3. Positional analysis of RbcX with respect to other RuBisCO subunit encoding genes

During the analysis of 37 cyanobacterial species we observed that X subunit was found either in between L and S subunit, or succeeding L and S subunit encoding genes. For example the arrangements in *Synechococcus elongatus* PCC 7002 and *Synechococcus elongatus* PCC 7942. In some cyanobacterial sp. there are multiple S-subunits viz. *Anabaena virabilis* ATCC 29413, *Nostoc punctiforme* ATCC29133, *Trichodesmium erythraeum* IMS101, *Cyanothece* sp. PCC 7425. Some cyanobacterial sp. also have more than one large subunit, for instance, *Cyanothece* sp. PCC8801, *Cyanothece* sp. PCC7424. In other cases like *Synechococcus elongatus* PCC7942, X subunit is located very far from its L and S subunit. The location of RbcL, RbcS, RbcX of some cyanobacterial species is given in **Table 4A** and **Table 4B**.

We also found certain cyanobacterial sp. in which RbcX/CbbX was absent or not yet recognized (**Table 4A** and **4B**), eg. *Synechococcus* sp., *Prochlorococcus marinus* sp. For

confirming the absence of RbcX, genome BLAST of all the organisms from **Table 4B** was done with RbcX of *G. violaceus* (since it is reported to be the most primitive organism according to 16S rRNA phylogenetic analysis) to find the similar or somewhat similar proteins. However, the blast result did not show any significant hits. In order to find out how these species fold their L subunit without rbcX, we started looking for other chaperonins which might aid in the process. We found out that *Prochlorococcus marinus* SS120 has 17 genes for different chaperones like DNA J, DNA K etc., which showed homology with RbcX. The list of chaperonins found in organisms not having documented RbcX are given in the **Table 4C**.

Table 4A: Locus of L and S subunit of RuBisCO in the genome of the organisms with RbcX

| Organism | <i>rbcL</i> (locus in bp) | <i>rbcS</i> (locus in bp) | <i>rbcX</i> (locus in bp) |
|---|----------------------------------|---|---------------------------|
| <i>Gloeobacter violaceus</i> PCC 7421 | 2307046-2308470 | 2308900-2309200 | 2308500-2308877 |
| <i>Synechocystis</i> sp. PC C 6803 | 2478414-2479826 | 2480477-2480818 | 2480034-2480444 |
| <i>Anabaena</i> sp. PCC 7120 | 1785970-1787400 | 1787950-1788279 | 1787495-1787893 |
| <i>Thermosynechococcus elongatus</i> BP-1 | 1574633-1576060 | 1573812-1574168 | 1574199-1574579 |
| <i>Microcystis aeruginosa</i> NIES-843 | 4390428-4391843 | 4389548-4389883 | 4389899-4390297 |
| <i>Synechococcus elongatus</i> PCC 6301 | 139920-141338 | 139494-139829 | 2692684-2693169 |
| <i>synechococcus</i> sp. PCC7002 | 1882749-1884164 | 1881910-1882245 | 1882276-1882680 |
| <i>Acaryochloris marina</i> MBIC 11017 | 1775408-1776838 | 1774474-1774821 | 1774874-1775275 |
| <i>Anabaena variabilis</i> ATCC 29413 | 4857469-4858899 | 4853884-4855128, 4856590-4856919 | 4856976-4857374 |
| <i>Synechococcus elongatus</i> PCC 7942 | 1479461-1480879 | 1480970-1481305 | 1595486-1595944 |
| <i>Synechococcus</i> sp. JA-2-3B'a(2-13) | 2682338-2683762 | 2681523-2681855 | 2681864-2682271 |
| <i>Synechococcus</i> sp. JA-3-3Ab | 1207204-1208628 | 1209115-1209447 | 1208699-1209106 |
| <i>Cyanothece</i> sp. ATCC 51142 | 3281510-3282925 | 3280671-3281006 | 3281030-3281428 |
| <i>Nostoc punctiforme</i> ATCC 29133 | 5263600-5265030 | 5265663-5265992, 5267505-5268779, 5380153-5382162 | 5265208-5265615 |
| <i>Trichodesmium erythraeum</i> IMS101 | 6791736-6793166 | 2407801-2409072, 6790587-6790922 | 6791096-6791482 |
| <i>Cyanothece</i> sp. PCC 7424 | 1061844-1063007, 1503225-1504643 | 1502367-1502702 | 1502734-1503141 |
| <i>Cyanothece</i> sp. PCC 7425 | 3372918-3374348 | 3372003-3372347, 4356729-4358000 | 3372423-3372818 |
| <i>Cyanothece</i> sp. PCC 8801 | 1677472-1678890, 2144196-2145281 | 1679403-1679738 | 1678968-1679372 |
| <i>Arthrospira platensis</i> NIES-39 | 4152769-4154199 | 4151888-4152223 | 4152272-4152637 |

Table 4B: Locus of L and S subunit of RuBisCO in the genome of the organisms lacking RbcX

| Organisms | <i>rbcL</i> (locus in bp) | <i>rbcS</i> (locus in bp) |
|--|---------------------------|---------------------------|
| <i>Prochlorococcus marinus</i> SS120 | 524454-525866 | 525970-526311 |
| <i>Prochlorococcus marinus</i> MED4 | 519087-520502 | 520592-520933 |
| <i>Prochlorococcus marinus</i> MIT9313 | 1293386-1294798 | 1292937-1293278 |
| <i>Synechococcus</i> sp. WH8102 | 1651727-1653142 | 1651326-1651667 |
| <i>Synechococcus</i> sp. CC9902 | 1563311-1564723 | 1562850-1563191 |
| <i>Synechococcus</i> sp. CC9605 | 716826-718241 | 718301-718642 |
| <i>Prochlorococcus marinus</i> str. MIT 9312 | 513733-515148 | 515243-515584 |
| <i>Prochlorococcus marinus</i> str. AS9601 | 529175-530590 | 530681-531022 |
| <i>Prochlorococcus marinus</i> str. MIT 9515 | 549602-551017 | 551111-551452 |
| <i>Prochlorococcus marinus</i> str. MIT 9303 | 737613-739025 | 739134-739475 |
| <i>Prochlorococcus marinus</i> str. NATL1A | 549278-550690 | 550791-551132 |
| <i>Prochlorococcus marinus</i> str. MIT 9301 | 503604-505019 | 505109-505450 |
| <i>Synechococcus</i> sp. RCC307 | 744350-745765 | 745825-746166 |
| <i>Synechococcus</i> sp. WH 7803 | 674560-675975 | 676035-676376 |
| <i>Prochlorococcus marinus</i> str. MIT 9215 | 555111-556526 | 556620-556961 |
| <i>Prochlorococcus marinus</i> str. MIT 9211 | 517060-518472 | 518572-518913 |
| <i>Chlorobium tepidum</i> TLS | 1677980-1679287 | |
| <i>Rhodopseudomonas palustris</i> CGA009 | 1731700-1733157 | 1733171-1733593 |
| <i>Synechococcus</i> sp. PCC 9311 | 1750450-1751862 | 1750005-1750346 |
| <i>Prochlorococcus marinus</i> NATL2A | 536918-538330 | 538431-538772 |

Table 4C: Diverse chaperones present in Cyanobacteria without RuBisCO chaperone

| Organism | HSP40 | HSP70 | HSP70-2 | ATPas es | HSP60 | GroEL | GroES | Other chaperones |
|-------------------------------|-------|-------|-------------------|----------|-------|-------|-------|---------------------|
| <i>P. marinus</i> SS120 | - | DNAK | - | + | - | + | + | ----- |
| <i>P. marinus</i> MED4 | - | DNAK | DNAK ₂ | - | - | - | + | ----- |
| <i>P. marinus</i> MIT9313 | - | DNAK | DNAK ₂ | - | - | + | + | ----- |
| <i>Synechococcus</i> WH102 | - | DNAK | DNAK ₂ | - | - | + | + | HSP33 |
| <i>Synechococcus</i> CC9902 | - | DNAK | - | - | - | - | + | HSP33, Cpn60, Tsp-1 |
| <i>Synechococcus</i> CC9605 | - | DNAK | - | - | - | - | + | HSP33, Cpn60, Tsp-1 |
| <i>P. marinus</i> SS120 | - | - | - | - | + | + | + | ----- |
| <i>P. marinus</i> MED4 | - | DNAK | DNAK ₂ | - | - | + | + | ----- |
| <i>P. marinus</i> MIT9515 | - | DNAK | DNAK ₂ | - | - | + | + | ----- |
| <i>P. marinus</i> MIT9303 | - | DNAK | DNAK ₂ | - | - | + | - | ----- |
| <i>P. marinus</i> NATL1A | - | DNAK | DNAK ₂ | - | - | + | + | ----- |
| <i>P. marinus</i> MIT9301 | - | DNAK | DNAK ₂ | - | - | + | + | ----- |
| <i>Synechococcus</i> RCC307 | DNAJ | DNAK | - | + | - | + | + | Clpb, Cpn10 |
| <i>Synechococcus</i> wh7803 | DNAJ | DNAK | - | - | - | + | + | Grpe, clpb |
| <i>P. marinus</i> 9215 | - | DNAK | DNAK ₂ | - | - | + | + | Cpn60 |
| <i>P. marinus</i> str MIT9211 | DNAJ | DNAK | DNAK ₂ | - | - | + | + | ----- |
| <i>Rhodo</i> CGA1009 | - | - | - | - | - | - | - | ----- |
| <i>Synechococcus</i> 9311 | DNAJ | DNAK | DNAK ₂ | + | - | + | - | Grpe, HSP 33 |
| <i>P. marinus</i> NATL2A | - | DNAK | - | - | - | + | + | Grpe |

Table 4C: '+' and '-' indicate the presence or the absence of the chaperone in the organism respectively.

Gene arrangement of RbcL, RbcS and RbcX is different in different cyanobacterial species which suggests that there is no set rule for RbcL to appear first, it may come after RbcS and RbcX. This observation indicates that these subunits were not transferred as a cluster during the course of evolution. Further, the organism which do not possess RbcX indicate that these organisms got these subunits independently and at different point of time, and eventually, RbcX subunit lost its relevance (at least in some species) which further indicates parallel evolution of these species. Moreover, the lack of rbcX also implicates development of

alternative strategies for protein folding. There are reports that *Synechococcus* sp. RbcX is not an ultimate requirement for RuBisCO large subunit folding implicating that even if it is present in organism like *Synechococcus*, it is not the most essential chaperone in RuBisCO large subunits folding. Presence of more than one small or large subunit might indicate gene duplication. The repeating subunit is found away from the core RuBisCO operon. Some organisms (e.g., *Prochlorococcus marinus*) do not have any reported RbcX. In order to know about the overall impact of the presence or the absence of RbcX, small and large subunit encoding genes were analyzed. There was no significant difference found in the other two subunits. An interesting observation was in the variation in the number of chaperones present in various organisms under study. The organisms which donot have RbcX, were found to have a large number of other folding genes (eg. *Prochlorococcus marinus*), which might give them an advantage of not wasting their energy in synthesizing RbcX and still having properly folded RbcL subunits.

As mentioned earlier, some organismare able to perform photosynthesis without RbcX eg. *Synechococcus* species. Possible reasons for it could be, (1) These organism possess some other proteins that overtake the functioning of RbcX or (2) structure of some of the RbcL and RbcS have evolved such that they don't require RbcX anymore.

As we know that rbcX interacts with EIKFEFD motif in RbcL and some marine cyanobacteria which lack RbcX still have this motif in their large subunits eg. *Prochlorococcus marinus* SS120, it may be due to several reasons. Several cyanobacteria have evolved certain folding proteins with similar domain which aids RuBisCO folding, hence removing the need of having RbcX. Further, as we know that RbcX acts downstream of GroEL and GroES in folding of RbcL. RbcX increases the folding of RbcL subunits ,so in organism which do not possess RbcX, their GroEL and GroES might be more efficient than other species due to unknown reasons. Surprisingly, Upon further investigating the occurrence of the RbcX binding domain in RbcL of organisms with RbcX and without RbcX, we came across another interesting observation, the organisms that donot have any reported RbcX have the exact RbcX interacting domain while the organisms that donot have any reported RbcX have the exact RbcX interacting domain while the organisms that have RbcX have a little change in the domain (**Figure 4D**).

| | W | K | E | I | K | F | E | F | D |
|---|---|---|---|---|---|---|---|---|---|
| <i>Prochlorococcus marinus</i> SS120 | . | . | . | . | . | . | . | . | . |
| <i>Synechococcus</i> sp. W1718 | . | . | . | . | . | . | . | . | . |
| <i>Prochlorococcus marinus</i> str MIT 9312 | . | . | . | . | . | . | . | . | . |
| <i>Synechococcus</i> sp. CC9605 | . | . | . | . | . | . | . | . | . |
| <i>Prochlorococcus marinus</i> NATL1 06041 | . | . | . | . | . | . | . | . | . |
| <i>Gloeobacter violaceus</i> PCC 7421 | . | . | . | . | . | . | . | Y | E |
| <i>Synechocystis</i> sp. PCC 6803 | . | . | . | . | . | . | . | . | E |
| <i>Anabanea</i> sp. PCC 7120 | . | . | . | . | . | . | . | . | E |
| <i>Thermosynechococcus elongatus</i> BP-1 | . | . | . | . | . | . | . | . | E |
| <i>Microcystis aeruginosa</i> NIES-843 | . | . | . | . | . | . | . | . | E |

Figure 4D: The alignment of RbcL amino acid sequence of five organisms lacking RbcX and five having RbcX, analyzed for the presence or the absence of the RbcX interacting Domain.

4.3.4. The expression of *G. violaceus* RuBisCO as soluble or insoluble form and purification of *G. violaceus* RuBisCO

The constructs generated viz. *pCOLDII-Gv-rbcL*, *pCOLDII-Gv-rbcS*, *pCOLDII-Gv-rbcX* and *pCOLDII-Gv-rbcLSX* were used for expression in *E. coli* BL21(DE3)pLysS cells. The cultures were induced by cold shock at 15°C for 30 minutes, followed by addition of IPTG at 0.5mM (working concentration, standardized so as to obtain optimal expression) and then incubated at 15°C for 18hrs (temperature condition standardized for optimal recombinant protein expression). The expression profile for *pCOLDII-Gv-rbc*, *pCOLDII-Gv-rbcS*, *pCOLDII-Gv-rbcX* and *pCOLDII-Gv-rbcLSX* are shown in **Figures 4F**, respectively.

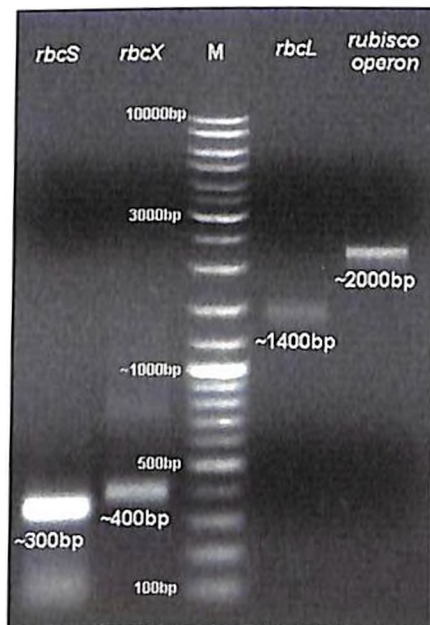


Figure 4E: The agarose (1%) gel image of the PCR products of the genes under study. lane 1: *rbcS*; lane 2: *rbcX*. lane 3: marker SM0331 (Thermoscientific); lane 4: *rbcL*; lane 5: *rubisco operon*. The estimated sizes of the PCR products obtained are labeled below the respective bands.

The expected protein molecular weight is 52.8kDa for *G. violaceus* RuBisCO LSU, 13.6kDa for *G. violaceus* RuBisCO SSU and 14kDa for RbcX (molecular weight estimated from the protein sequence available). The IPTG-induced (0.5mM) culture of *pCOLDII-Gv-rbcLSX* showed similar profile to that of *pCOLDII-Gv-rbcL*, showing the absence of RbcS and RbcX proteins in the former, which are either not expressed at all or are present in very low amounts that cannot be detected on SDS-PAGE. When RbcS and RbcX encoding genes were placed under separate promoters as in *pCOLD-Gv-rbcS* and *pCOLD-Gv-rbcX* constructs, the expected proteins were found to be synthesized in the *E. coli* system. Similar observations were also made in *Anabaena* RuBisCO expression in *E. coli* (Gurevitz et al., 1985), where when the complete *RuBisCO operon (rbcLXS)* containing construct lead to the expression of *rbcL* but not the subsequent proteins. Possibly, the intergenic region between *rbcL* and *rbcX* of *G. violaceus* and *Anabaena* is causing transcription termination in *E. coli* which as such does not occur in the actual organisms. If true, the problem can be overcome by using vectors with separate promoters for each of the genes of the operon and co-expressed.

The *G. violaceus* RuBisCO subunits RbcL and RbcS were found to be insoluble in buffer 1 (Imidazole – 5mM, Sodium chloride – 300mM, Sodium Phosphate pH 8.0 – 50mM, Glycerol 10%) and buffer 2 (Imidazole – 5mM, Sodium chloride – 300mM, Sodium Phosphate pH 8.0 – 50mM, Triton-X 100 – 0.5%), but soluble in buffer 3 (Imidazole – 20mM, Sodium chloride – 300mM, Sodium Phosphate pH 8.0 – 50mM, Urea – 8M, Glycerol 10%) (**Figure 4G**). The RbcX protein was found to be soluble in buffer 1. The proteins expressed by the constructs generated were found to get accumulated as inclusion bodies. In expression of recombinant proteins in *E. coli*, inclusion bodies are formed as a consequence of expression of misfolded proteins which in turn occurs in the case of proteins constituting more than one subunits and/or having disulfide linkages or prosthetic groups.

The type II RuBisCO of *Rhodospirillum rubrum (pET15b-R. rubrum* RuBisCO transformed in *E. coli* BL21(DE3)pLysS) when expressed in *E. coli* by IPTG induction (0.5mM working concentration) yielded soluble and functionally active protein (results shown in chapter 6), in the present study as well as in previous reports. Further, *Thermosynechococcus elongatus* BP1 type I RuBisCO (RbcL and RbcS of *T. elongatus* BP1 show 87% and 70% sequence identity to the RbcL and RbcS of *G. violaceus*, respectively) when expressed using *pUC18rbcLXS* in *E. coli* DH5 α , led to translation into a soluble protein with biological activity (Gubernator et al., 2008). Similarly, for *Synechococcus sp* PCC 6301, the *RuBisCO operon* expressed in *E. coli* yielded soluble protein in native conformation. The type III RuBisCO from *Archaeoglobus fulgidus* cloned into *pET11a* and expressed in *E. coli* BL21, by IPTG induction was purified in soluble-native form as determined by gel filtration chromatography and enzyme assay (Kreel and Tabita 2007).

Figure 4F (i-iv): The SDS-PAGE (12%) image of IPTG induced cultures of *pCOLDH-Gv-rbcL*, *pCOLDH-Gv-rbcS*, *pCOLDH-Gv-rbcX* and *pCOLDH-Gv-rbcLSX* in *E. coli* BL21(DE3)pLysS, respectively. a) SDS-PAGE (12%) for *E. coli* BL21(DE3)pLysS *pCOLDH-Gv-rbcL* culture after cold shock and 0.5mM IPTG induction. Lane1- marker, Lane 2- Crude extract from uninduced culture, Lane 3-10- Crude extract from cultures after 1-18 hrs of induction, respectively. b) SDS-PAGE (12%) for *E. coli* BL21(DE3)pLysS *pCOLDH-Gv-rbcS* culture after cold shock and 0.5mM IPTG induction. Lane1- marker, Lane 2- Crude extract from uninduced culture, Lane 3-10- Crude extract from cultures after 1-18 hrs of induction, respectively. c) SDS-PAGE (12%) for *E. coli* BL21(DE3)pLysS *pCOLDH-Gv-rbcX* culture after cold shock and 0.5mM IPTG induction. Lane1- marker, Lane 2- Crude extract from uninduced culture, Lane 3-10- Crude extract from cultures after 1-18 hrs of induction, respectively. d) SDS-PAGE (12%) for *E. coli* BL21(DE3)pLysS *pCOLDH-Gv-rbcLSX* culture after cold shock and 0.5mM IPTG induction. Lane1- marker, Lane 2- Crude extract from uninduced culture, Lane 3-10- Crude extract from cultures after 1-18 hrs of induction, respectively.

1-0hr

3hr

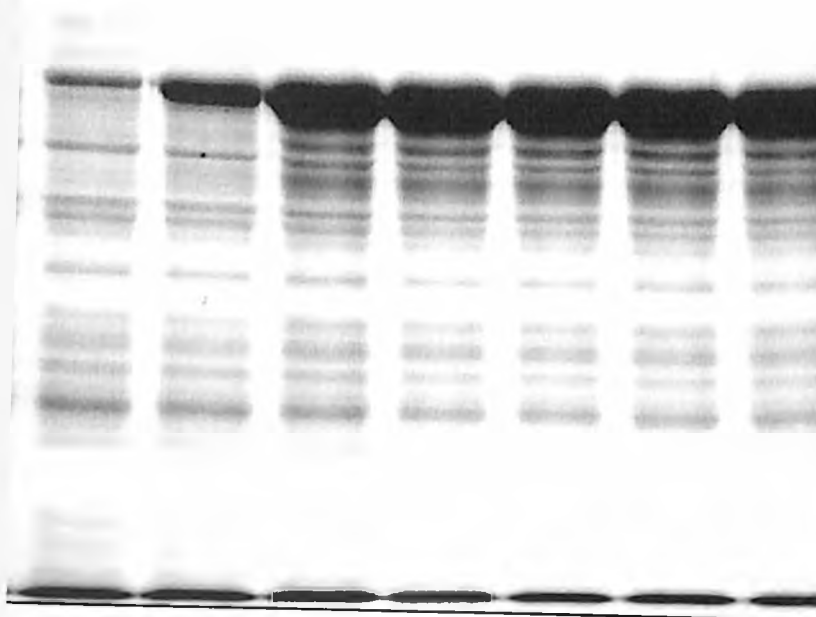
6hr

9hr

12hr

15hr

18hr



0hr

3hr

6hr

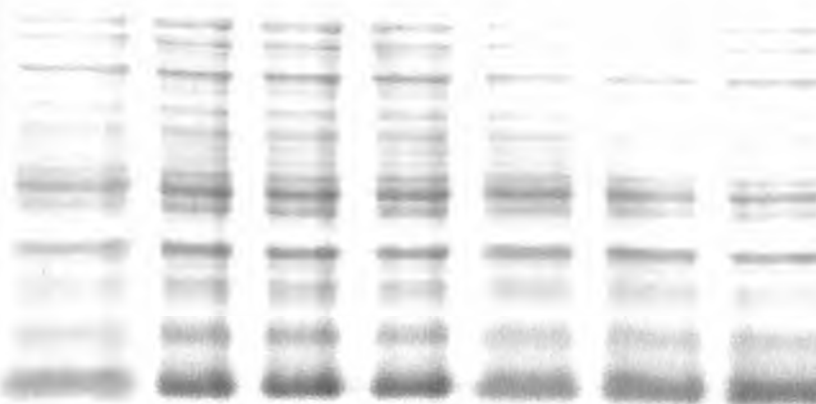
9hr

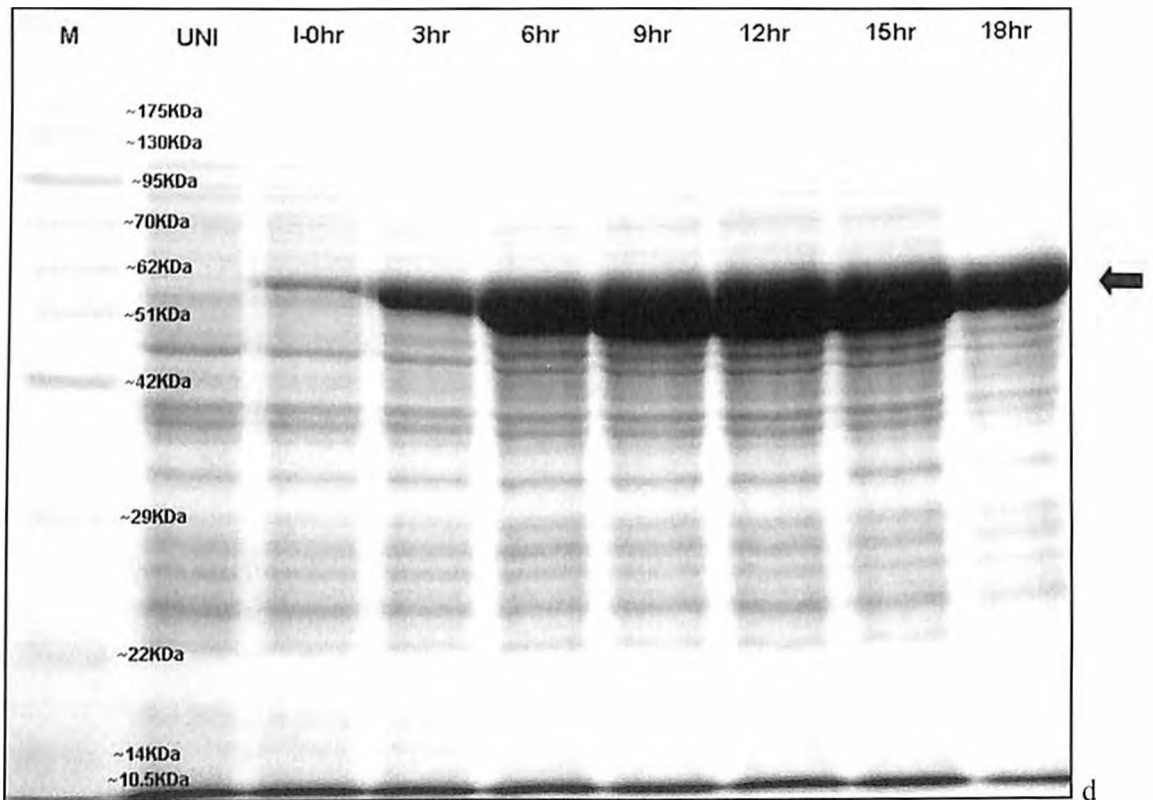
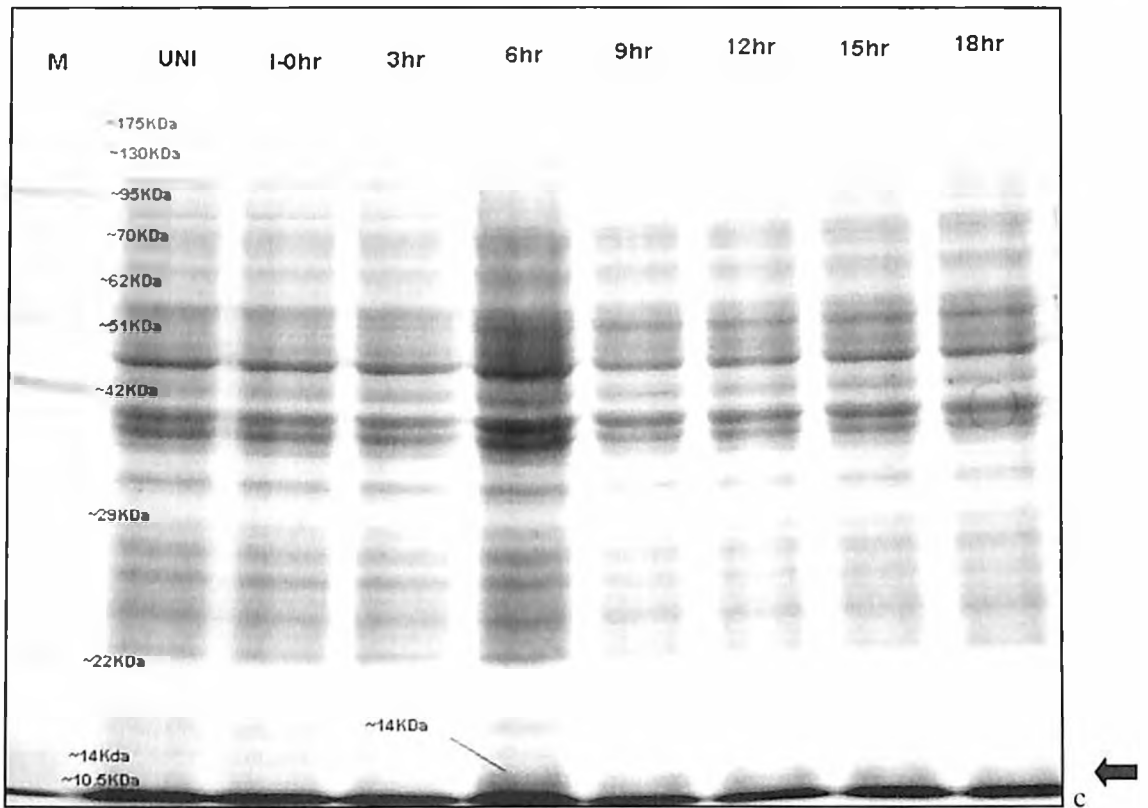
12hr

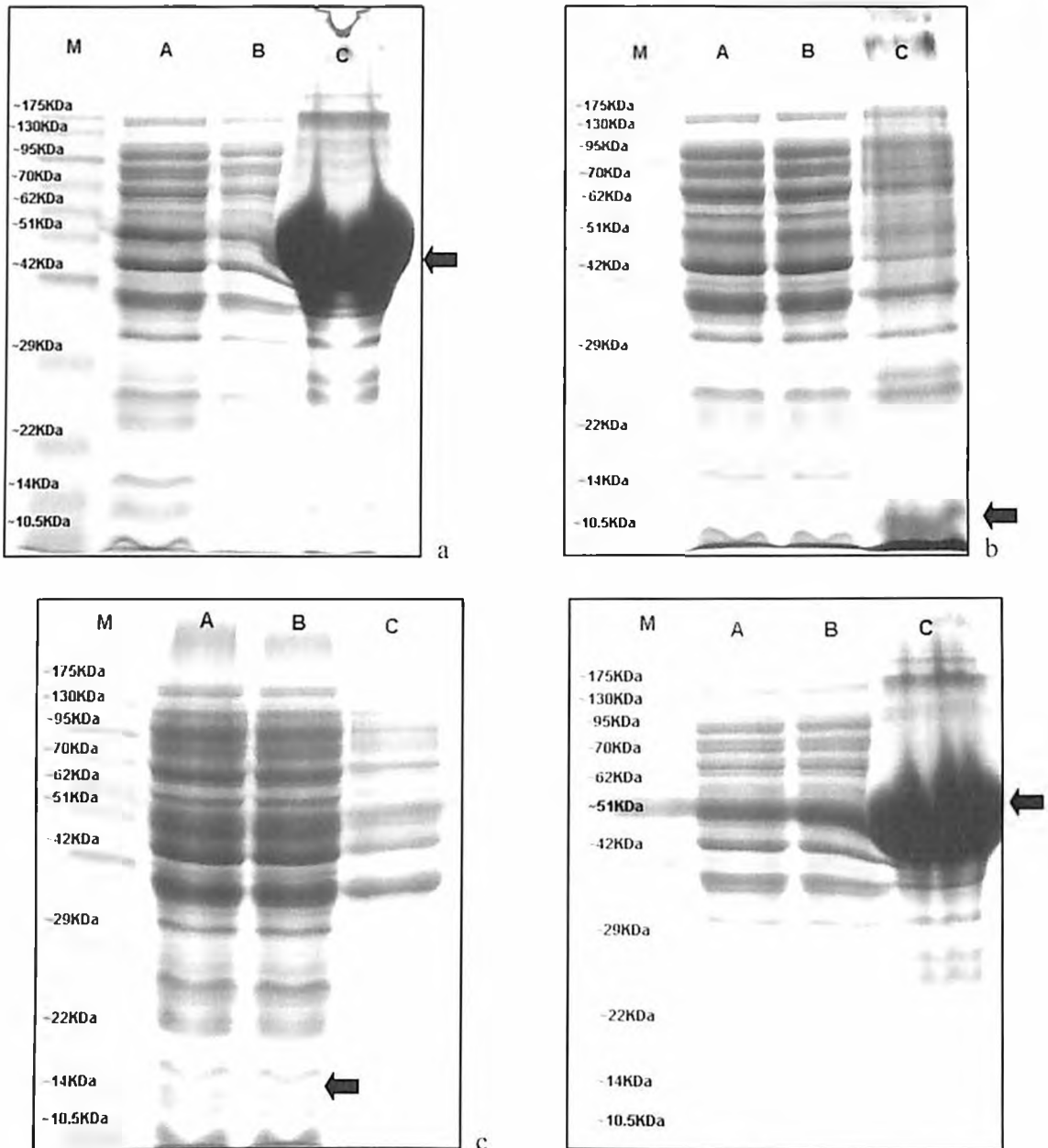
15hr

18hr

~13.6KDa







d

Figure 4G(i-iv): SDS-PAGE (12%) showing solubility of the overexpressed protein in presence/absence of denaturants. a) Solubility of RbcL expressed from *E. coli* BL21(DE3)pLysS *pCOLD-Gv-rbcL* in buffer 1 (Imidazole – 5mM, Sodium chloride – 300mM, Sodium Phosphate pH 8.0 – 50mM, Glycerol 10%) and buffer 2 (Imidazole – 5mM, Sodium chloride – 300mM, Sodium Phosphate pH 8.0 – 50mM, Triton-X 100 – 0.5%) and buffer 3 (Imidazole – 20mM, Sodium chloride – 300mM, Sodium Phosphate pH 8.0 – 50mM, Urea – 8M, Glycerol 10%), shown in A, B and C, respectively. b) Solubility of RbcS expressed from *E. coli* BL21(DE3)pLysS *pCOLD-Gv-rbcS* in buffer 1, buffer 2 and buffer 3, shown in A, B and C, respectively. c) Solubility of RbcX expressed from *E. coli* BL21(DE3)pLysS *pCOLD-Gv-rbcX* in buffer 1, buffer 2 and buffer 3, shown in A, B and C, respectively. d) Solubility of RbcL expressed from *E. coli* BL21(DE3)pLysS *pCOLD-Gv-rbcLXS* in buffer 1, buffer 2 and buffer 3, shown in A, B and C, respectively.

Although a plethora of evidence exists reporting the expression of form I, II and III RuBisCO in *E. coli* in native soluble form, there are also an equally convincing set of reports showing misfolded or aggregated RuBisCO expression in *E. coli*. The initial attempts to purify natively folded *Anabaena* RuBisCO (form I) from recombinant *E. coli* failed, possibly due to unavailability of adequate folding machinery (Gurevitz et al., 1985). Subsequent research by Li and Tabita (1997) evidenced resolving this problem by co-expressing the *Anabaena* RuBisCO encoding genes with chaperone proteins. Further, the higher plant RuBisCO (form I) from *Nicotiana tabacum* (Whitney and Sharwood 2007), *Zea mays* and *Triticum* (Cloney et al., 1993) could not be expressed in *E. coli* in a soluble form. The form I RuBisCO from non-green algae *Galdieria sulphuraria* and *Phaedactylum tricorutum* have also been reported to require specific folding proteins for proper assembly of subunits (Whitney et al., 2001).

The *G. violaceus* RbcL and RbcS recombinant chimeric proteins with N-terminal His tag was recovered as a homogenous preparation from the metal chelate column in presence of urea (affinity chromatography using BD Talon Co²⁺ resin). The eluted protein migrated on the SDS-PAGE as a protein with molecular weight ~52KDa and ~14KDa, respectively (**Figure 4H i and ii**) which is close to the theoretical molecular weight of the protein. The **Figure 4H(iii)** shows the purification profile of *G. violaceus* RbcX under native conditions. The N-terminal His tag of the protein was removed by thrombin treatment. The non-denaturing 6% polyacrylamide gel (**Figure 4I**) of RbcL (renatured on the column after binding in denaturing conditions) showed the protein at a position suggesting a homodimer with molecular weight ~150KDa. The occurrence of the trimeric and tetrameric form of the protein on the native PAGE shows the success of the refolding strategy followed. Elian Clark (1998) reviewed a wealth of literature on methods of refolding denatured recombinant protein by using *in-vitro* chaperone activity assisted by additives or slow removal of the denaturants. Colangeli et al., (1998) have reported refolding of denatured antigens of *Mycobacterium tuberculosis* by on-column removal of contaminating proteins and gradual exchange into a non-denaturing buffer before elution.

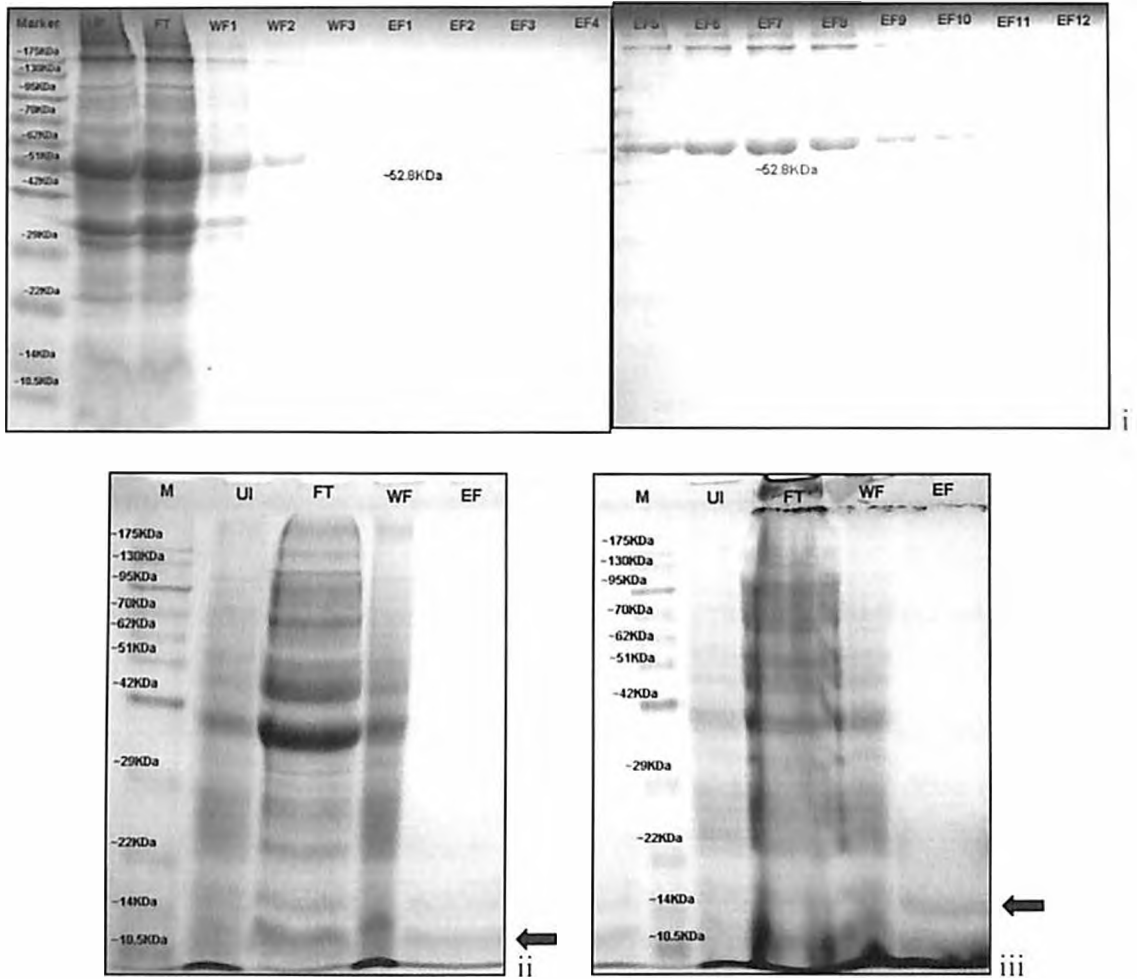


Figure 4H(i-iii): The SDS-PAGE (12%) for fractions collected during metal affinity chromatography from IPTG induced cultures of *pCOLD-Gv-rbcL*, *pCOLD-Gv-rbcS* and *pCOLD-Gv-rbcX* in *E. coli* BL21(DE3)pLysS.

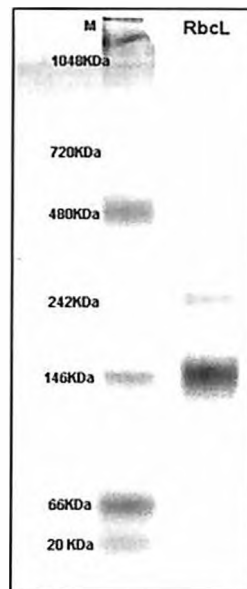


Figure 4I: The image of 6% native PAGE for *G. violaceus* RbcL, after purification in denaturing condition followed by subsequent renaturing by steady removal of denaturing agents.

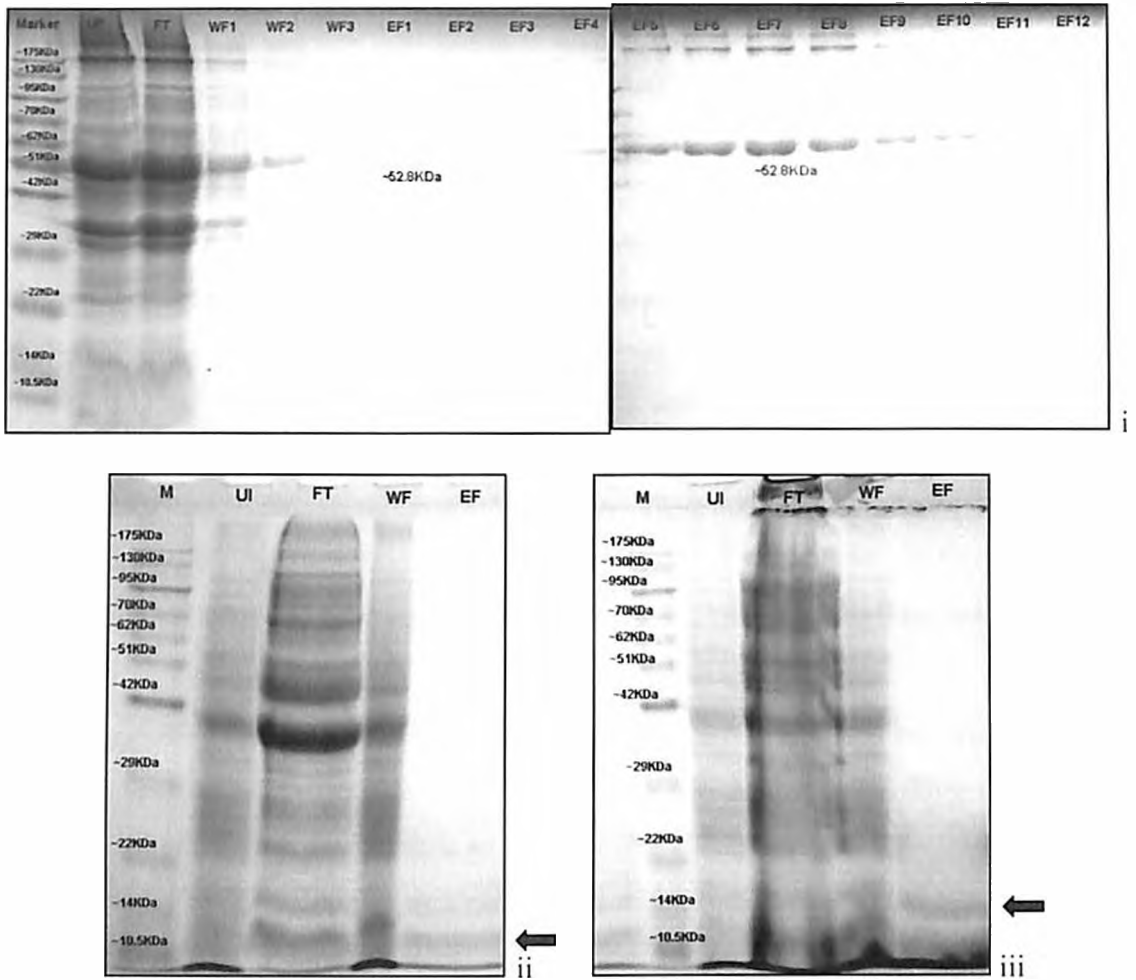


Figure 4H(i-iii): The SDS-PAGE (12%) for fractions collected during metal affinity chromatography from IPTG induced cultures of *pCOLD-Gv-rbcL*, *pCOLD-Gv-rbcS* and *pCOLD-Gv-rbcX* in *E. coli* BL21(DE3)pLysS.

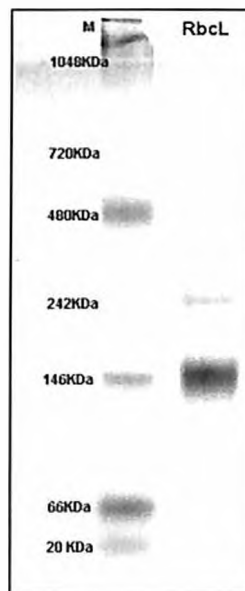


Figure 4I: The image of 6% native PAGE for *G. violaceus* RbcL, after purification in denaturing condition followed by subsequent renaturing by steady removal of denaturing agents.

4.3.5. Validation of protein identity and conformation (MS/MS and CD analysis)

The purified protein observed on SDS-PAGE was confirmed by MALDI-TOF analysis. The **Figure 4J i and ii** shows plot for mass/charge ratio against intensity for the trypsin digested *G. violaceus* RuBisCO LSU (RbcL) and SSU (RbcS). The peaks with greater than equal to 10 signal/noise ratio were considered and compared with the expected mass/charge ratios of trypsin digested protein obtained from the *in-silico* analysis. The fragments with highest intensities viz. fragments with m/z of 431.910, 447.890, 455.775, 461.897, 562.947, 650.940, 1175.815 and 1448.980 were found to be common between the expected and observed trypsin digested *G. violaceus* RuBisCO LSU protein sequences, validating the expression of the expected protein (**Figure 4K(i)**). Similarly, for RuBisCO SSU, the fragments with m/z value of 837.6, 940.736, 1062.01 and 1444.02 were found both in the expected and the observed trypsin digested *G. violaceus* RuBisCO SSU (**Figure 4K(ii)**).

The conformational stability against denaturation of the native RuBisCO enzyme was studied using CD spectroscopy. The CD spectroscopy data was analyzed by three methods viz. CONTIN, SELCON3 and CDSSTR for reliable results to bring out the correspondence between the structural variations and CD spectra. The CD spectrum of the native RuBisCO enzyme reflects the backbone conformation of the protein molecule (**Figure 4L**). It reveals positive band at 195 nm and negative band at 212 nm, characteristic for the α -helical structure. Around 212 nm, the β -sheet structure also contributes to the ellipticity of the α -helix bands. The low NRMSD, a goodness of fit parameter, as obtained for CDSSTR (**Table 4D**) indicates optimal secondary structure prediction solution and that the database used is suitable

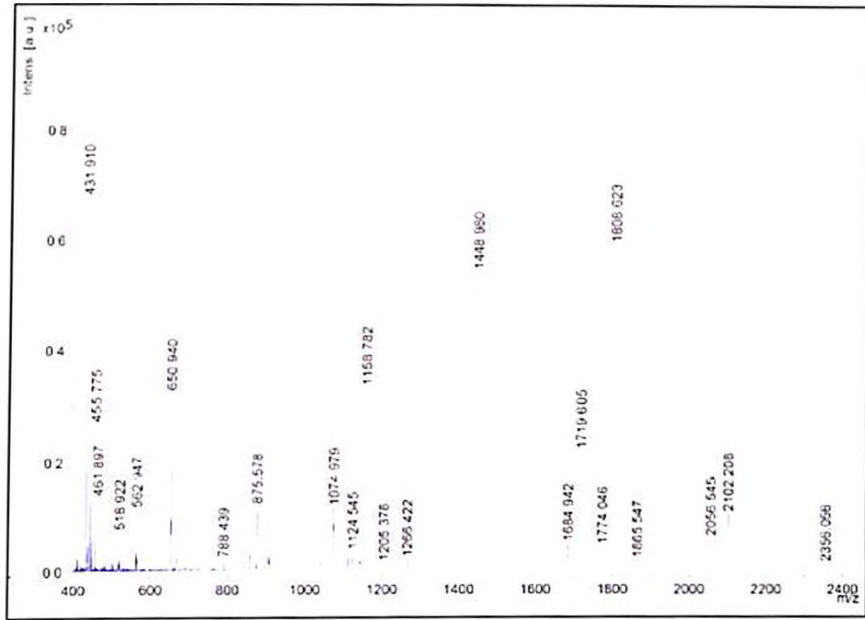


Figure 4J(i): The m/z vs intensity plot of the RuBisCO LSU protein analyzed by MALDI-TOF after in-gel tryptic digestion

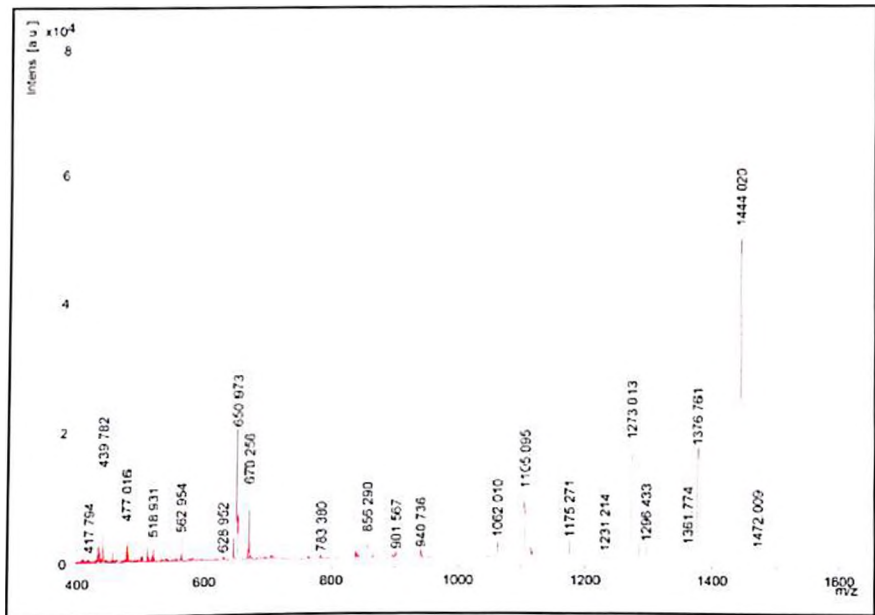


Figure 4J(ii): The m/z vs intensity plot of the RuBisCO SSU protein analyzed by MALDI-TOF after in-gel tryptic digestion

Table 4D: The CD spectra data of IMAC purified and renatured *G. violaceus* LSU (RbcL) analyzed by CONTIN, SELCON3 and CDSSTR by using DICHROWEB.

| Method | NRMSD | Helix 1 | Helix 2 | Strand 1 | Strand 2 | Turns | Unordered | Total |
|-----------|--------|---------|---------|----------|----------|-------|-----------|-------|
| Contin-LL | 24.430 | 0.000 | 0.047 | 0.234 | 0.119 | 0.189 | 0.411 | 1 |
| | | 0.002 | 0.023 | 0.272 | 0.138 | 0.195 | 0.370 | 1 |
| CDSSTR | 0 | 0.36 | 0.06 | 0.18 | 0.12 | 0.07 | 0.21 | 1 |
| Selcon3 | 0.732 | 0.000 | 0.028 | 0.292 | 0.195 | 0.228 | 0.250 | 0.993 |

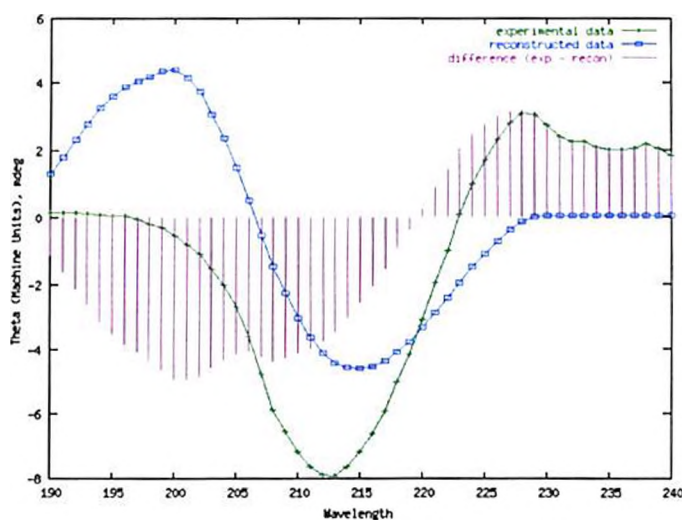


Figure 4L: The CD spectra of *G. violaceus* RbcL purified by IMAC. The experimental data obtained was normalized with respect to the data obtained from standard proteins in the database to obtain the reconstructed data.

4.3.6. Catalytic Activity of *G. violaceus* RuBisCO *in-vitro*

The specific activity of *G. violaceus* PCC 421 RuBisCO LSU at 2mM RuBP concentration was found to be 0.005 μ moles/min/mg. The activity of the protein increased with an increase in RuBP concentration with the K_M RuBP value of 0.1917mM and V_{max} value of 0.0002 per minute. The turnover number of the protein was found to be 0.0029/sec (**Figure 4M**).

Renatured *G. violaceus* RbcL, when reconstituted with RbcS in presence of RbcX, lead to four-fold enhancement in the activity of the protein (**Figure 4M** and **4N**). The *G. violaceus* RuBisCO LSU (0.5 μ M) was first incubated with RbcX (0.1 μ M) and incubated at 25°C for 30 minutes followed by addition of *G. violaceus* RuBisCO SSU (2 μ M) and ATP (4mM) for 30 minutes in presence of RuBisCO activation buffer and 25mM NaHCO₃ created an enhanced carboxylation environment for RuBisCO LSU. The specific activity observed after complementation was 0.0196 μ moles/min/mg. The K_M RuBP of the protein for the LSU+SSU complex also improved to 0.102mM and the V_{max} of the protein complex was found to

increase up to 0.011/minute. The turnover number of the protein complex was found to be 0.016 second.

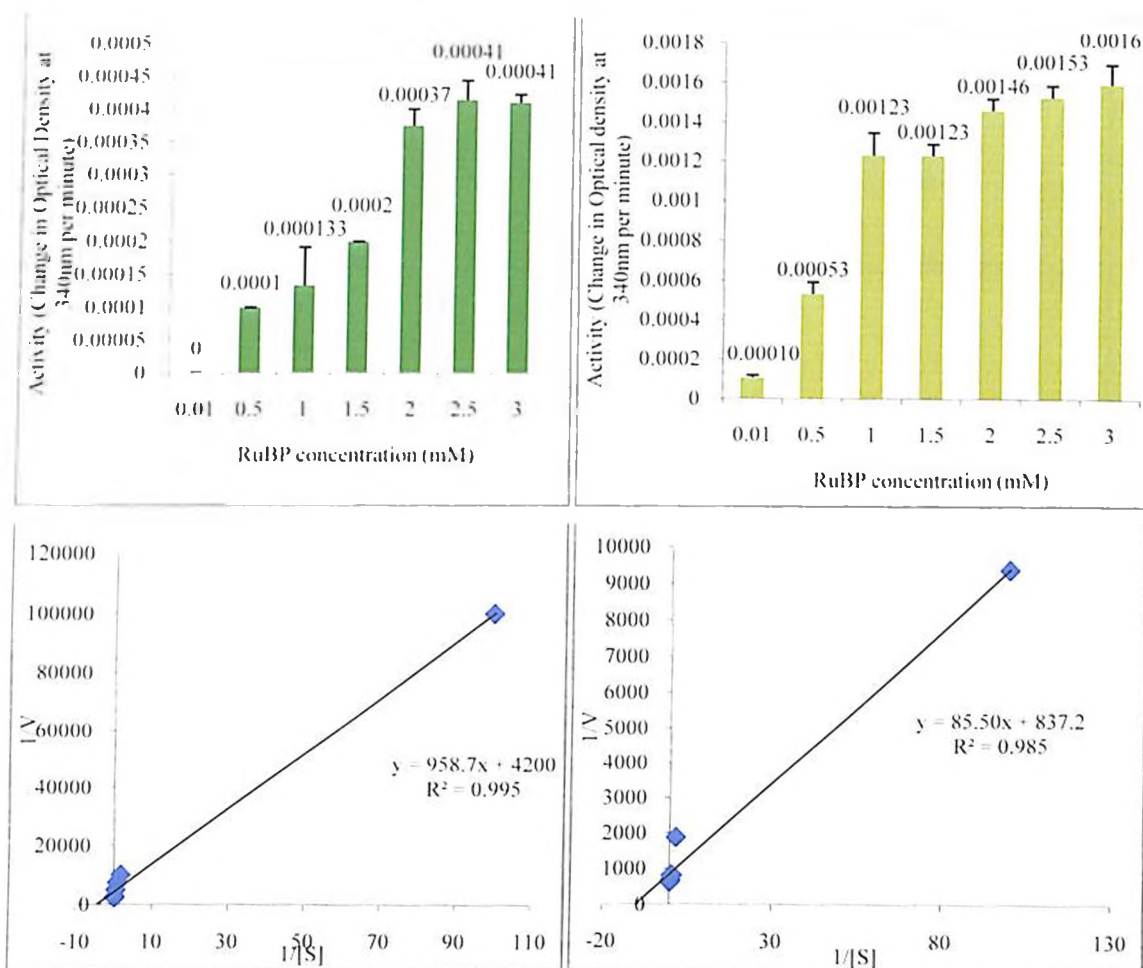


Figure 4M: The activity (change in absorbance per minute) of *G. violaceus* RuBisCO RbcL (● top left panel) and Holoenzyme (● top right panel) at different RuBP concentrations. The bottom left and the bottom right graphs show the Lineweaver-Burk plot for *G. violaceus* RbcL and holoenzyme, respectively for the RuBisCO assay performed.

The catalytic activity observed for *G. violaceus* is quite lower than what is usually observed in RuBisCO from cyanobacteria. Possibly, it might reflect our current inability to isolate the protein in its near native state or measure its activity under optimal conditions.

The specific activity, K_M RuBP and K_{cat} of *T. elongatus* BP1 are 1.83 μ moles of PGA/min/mg, 0.0204mM, and 2.65/sec, respectively (Gubernator et al., 2008). The K_M RuBP and K_{cat} reported for *Synechococcus* sp PCC 6301 is 0.0639mM and 11.4/sec, respectively (Greene et al., 2007). The *Anabaena* RuBisCO expressed using the operon *rbcLXS* in *E. coli*, in presence of chaperones GroEl and GroES, yielded a natively folded protein with the specific activity of 13.8nmoles/min/mg. It should be noted that in absence of these chaperones, the protein obtained showed ~5 fold decrease in activity, i.e. 2.8nmoles/min/mg.

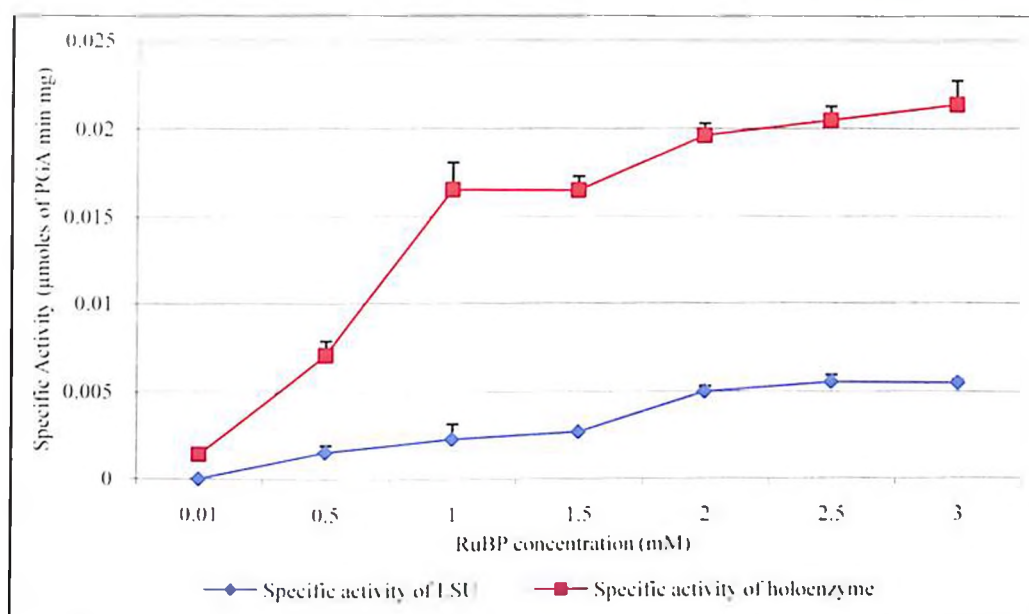


Figure 4N: The effect of RuBP concentration on the specific activity of *G. violaceus* RuBisCO large subunit (RbeL- blue trendline) and RuBisCO holoenzyme (RbeL + RbeS- red trendline).

4.3.7. *In-silico* generated tertiary model of RuBisCO LSU and SSU of *G. violaceus*

The **Figure 4O** and **4P** represent the tertiary structure of the *G. violaceus* RuBisCO LSU and SSU obtained by homology modeling using MODELLER. The Figure shows the secondary structures of *G. violaceus* LSU and SSU obtained by PDBsum. The models generated when validated by PDBsum server showed 95.6% residues in the most favored region on a Ramachandran plot (**Figure 4O**) for LSU RuBisCO and 91.4% residues in the most favored region for SSU RuBisCO (**Figure 4P**).

The tertiary model generated for LSU RuBisCO of *G. violaceus* shows that the major portion of the protein is in α helical form. Further, approximately the middle portion of the protein forms α -helix- β sheet barrel-like structure, with alternating α -helix and β sheet structures. It should also be noted that the residues involved substrate binding, active site and metal ion binding are stationed in this α -helix- β sheet barrel-like structure. The SSU RuBisCO model shows that almost equal portions of the protein take up α helical and β sheet structures.

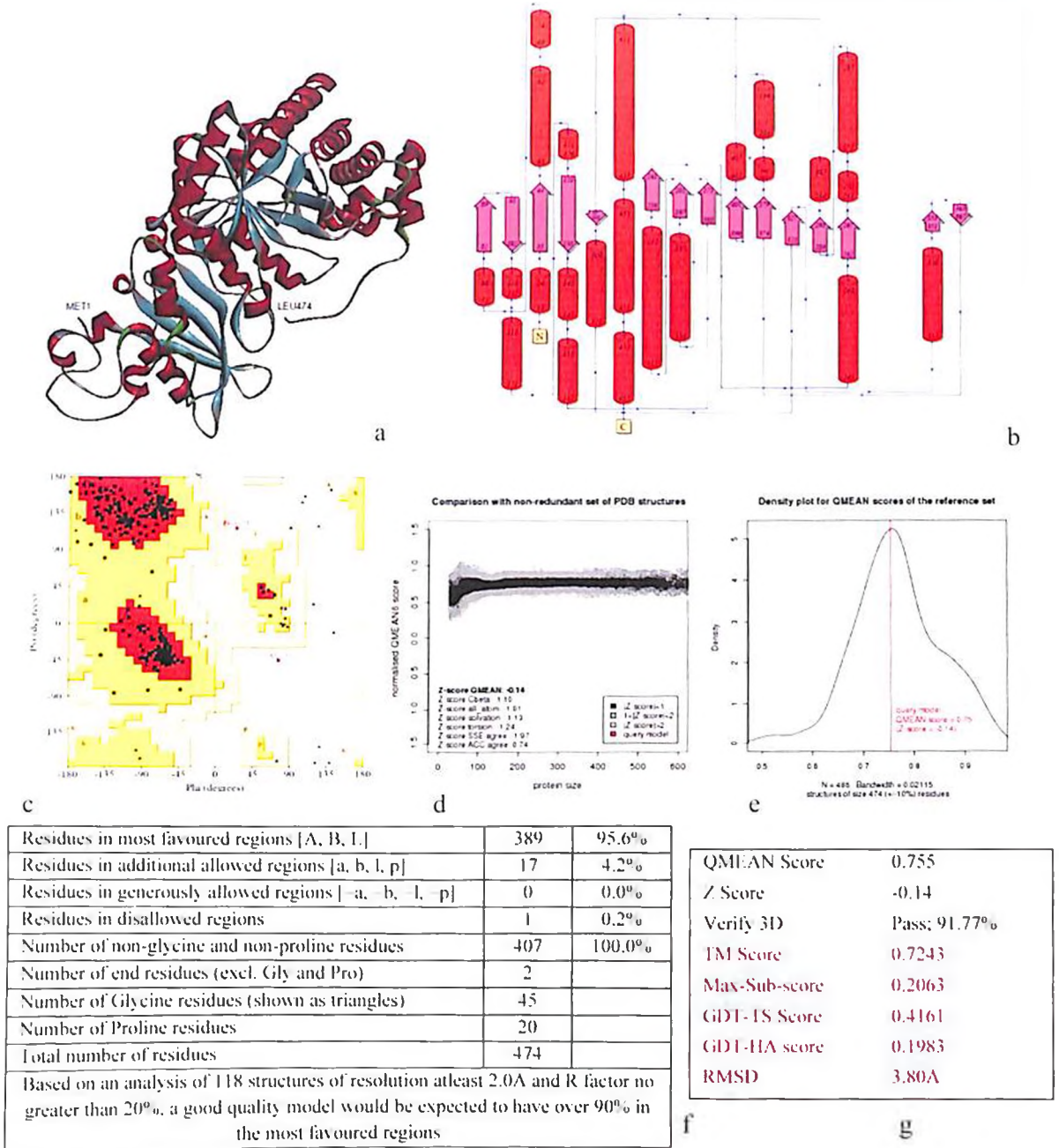


Figure 40: The protein model for *G. violaceus* RbcL generated by homology modeling and its validation. a –The tertiary structure of RbcL showing the N terminal and the C terminal of the protein. b - The topology of the protein showing the secondary structure. c – Ramachandran plot for the protein model. d and e – QMEAN score of the RbcL model in comparison to the standard protein models. f – Ramachandran plot statistics showing >90% residues to be in the favorable region of the plot, hence validating it to be a good quality model. g – Scoring of the RbcL model according to QMEAN, Verify3d and TMscore (in red font, obtained by superposition of the RbcL model with template i.e. 2YBV) servers.

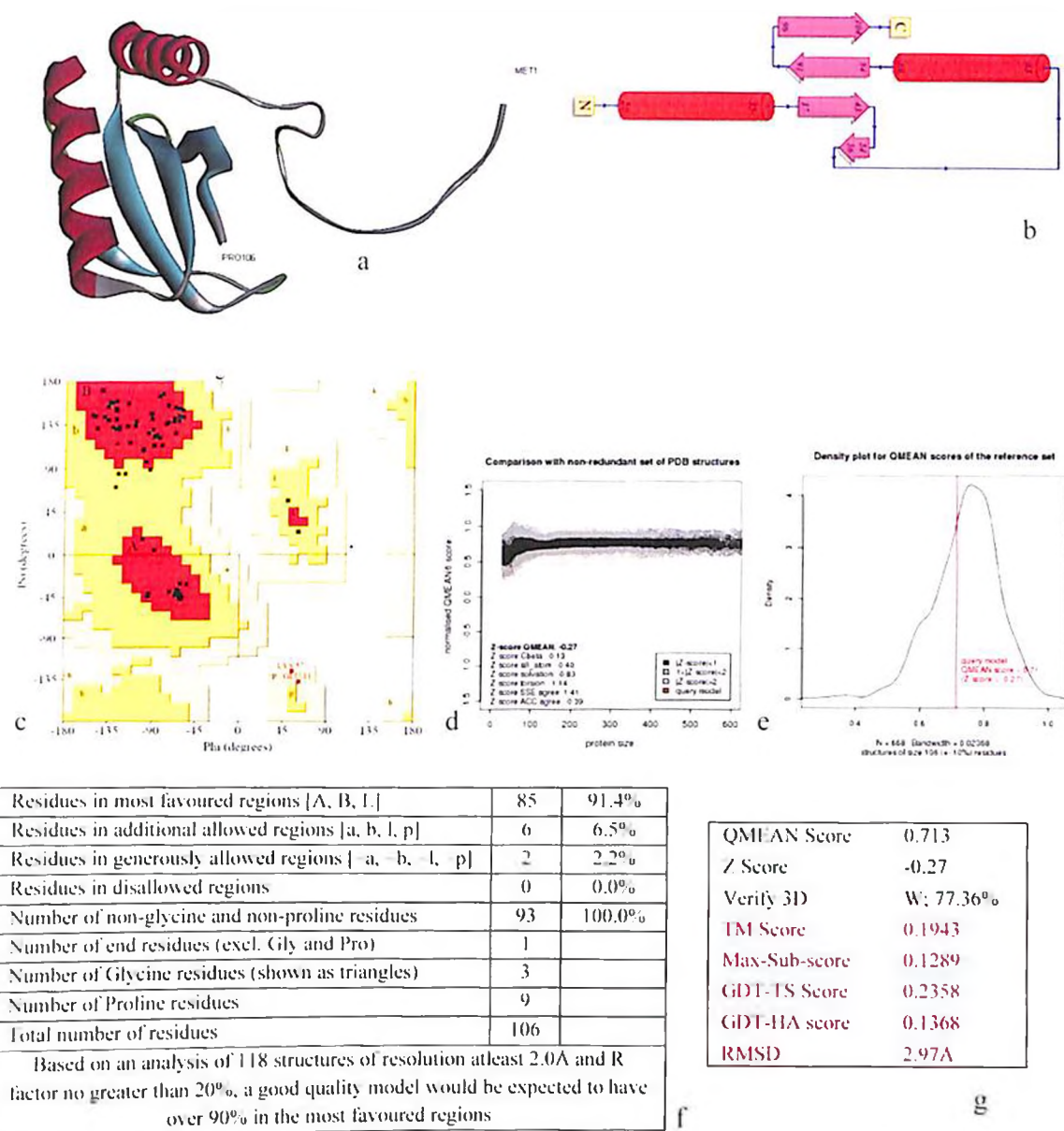


Figure 4P: The protein model for *G. violaceus* RbcS generated by homology modeling and its validation. a –The tertiary structure of RbcS showing the N terminal and the C terminal of the protein. b - The topology of the protein showing the secondary structure. c – Ramachandran plot for the protein model. d and e – QMEAN score of the RbcL model in comparison to the standard protein models. f – Ramachandran plot statistics showing >90% residues to be in the favorable region of the plot, hence validating it to be a good quality model. g – Scoring of the RbcS model according to QMEAN, Verify3d and TMscore (in red font, obtained by superposition of the RbcS model with template i.e. IRBL) servers.

4.3.8. Role of chaperones in RuBisCO assembly

As observed in the present study and in previous reports, the form I RuBisCO from certain cyanobacteria and all reported higher plants is difficult to express in *E. coli* in a correctly folded form. These difficulties have been attributed to the absence of a companion protein in *E. coli* system which can keep the large subunit in solution until it comes in contact with the small subunit of RuBisCO and forms a holoenzyme (Gurevitz et al., 1985). Li and Tabita (1997), reported that the *Anabaena* RuBisCO subunits when expressed in absence of each other lead to the formation of insoluble aggregates. They further reveal the importance of chaperone protein RbcX, which when present with RuBisCO LSU and SSU, in-vivo, caused a 14 fold enhancement in the RuBisCO activity. Their work also demonstrated that the involvement of other chaperone proteins like GroEL and GroES can further increase the solubility of the holoenzyme and hence its activity. On the contrary, these chaperones are not necessary for the formation of a biologically active recombinant RuBisCO in *E. coli* from cyanobacteria like *Synechococcus* sp PCC 6301 and *T. elongatus* BP1. Koay et al. (2016) created chimeric RuBisCO LSU by replacing segments of readily expressed and soluble *Synechococcus* LSU with corresponding segments of a difficult to assemble (*in-vivo* in *E. coli*) *Chlamydomonas reinhardtii* LSU of RuBisCO and expressed them in *E. coli*. The study leads to the identification of regions of *Synechococcus* LSU RuBisCO, possibly responsible for its easy folding and assembly in *E. coli* which includes residues 98-197 and 248-447. The rest of the regions even if replaced by corresponding regions from *C. reinhardtii* did not affect the solubility of the protein.

4.3.9. Cyanobacterial RuBisCO Vs *G. violaceus* RuBisCO

If we compare the α cyanobacteria LSU, the MSA shows very little variation with as much as 90% of the sequence conserved. Further, the β cyanobacteria LSUs when aligned show ~70% residues conserved. The most variable region is the initial 10-12 residues. The regions of the protein sequence harboring the active site residues i.e. Lys174, Kcx200, Asp202, Glu203, His293, Lys333 (in *G. violaceus* LSU identified by aligning with *T. elongatus* LSU and *S. oleracea* LSU, whose active sites are characterized) are conserved in all β cyanobacteria RuBisCO. Further, the substrate binding sites T172, K176, R294 and H326 in *G. violaceus* LSU (identified by aligning with LSU of *T. elongatus*) and metal binding sites D202 and E203 in *G. violaceus* (identified by aligning with *T. elongatus* LSU) were also completely conserved. The Form IA and Form IB RuBisCO, when aligned, show ~60% sequence conserved, with functionally important residues completely conserved.

The alignment of SSU sequence from α cyanobacteria shows ~80% residues conserved while the β cyanobacteria SSU is very variable showing only 33.3% conservation of sequence upon

multiple sequence alignment. Further, when α and β cyanobacterial SSU RuBisCO was aligned together, the MSA showed a major difference in the initial 10-15 residues of the sequence. A stretch of 6 residues, 'OSTVGD' is found in α cyanobacterial SSU RuBisCO only. The amino acid sequence of all cyanobacterial SSU RuBisCO when aligned shows mere 17% conservation.

4.3.10. Congruency in growth rate and RuBisCO activity

The *T. elongatus* BP-1 RuBisCO is the most closely related to that of *G. violaceus* RuBisCO among the fully characterized RuBisCO proteins of cyanobacteria. The growth parameters, protein activity and the crystal structure of the RuBisCO protein from this organism are well reported. The comparison of the protein from the two organism was done in order to elucidate information from the data available and to justify the differences observed.

As reported by Yamaoka et al. (1978), the growth rate is in congruence with the photosynthetic activity of thermophilic cyanobacteria, i.e. the factors affecting the growth rate have an effect on the photosynthetic efficiency of the organism as well. Hence, the slow RuBisCO activity of *G. violaceus* could be one of the reasons for the slow doubling time (72hrs) of the organism. Other factors possibly responsible for the slow growth could be the absence of several genes for proteins for PSI and PSH viz. PsaI, PsaJ, PsaK, PsaX, PsbY, PsbZ and Psb27 (Nakamura et al., 2003). Further, as observed by Yu et al., (2014), *Synechococcus elongatus* UTEX 2973 (doubling time 2.3hrs) has multi folds faster growth and photosynthetic efficiency as compared to *Synechococcus elongatus* PCC 7942 (8.5hrs), the possible reasons for which is the occurrence of a thick layer of thylakoids in former than in the later. And *G. violaceus* being devoid of any thylakoids is incapable of attaining a greater growth rate.

4.3.11. RuBisCO activity not dependent on the active site alone

The active site residues of RuBisCO are located in the large subunit. The conservation of these active site residues as well as the other functionally important residues viz. the substrate binding site and the Mg-binding site, across all the cyanobacterial RuBisCO (both form IA and IB), shows the importance of these residues. Further, the active site residues of *G. violaceus* also overlap with those of *Spinacea oleraceae* RuBisCO. These observations signify two important facts i.e. firstly, the importance of the residues involved in carboxylation reaction and secondly, the carboxylation is determined by the active site but the rate of carboxylation and its specificity are determined by several other domains of the protein apart from the active site.

The specific activity of *T. elongatus* RuBisCO almost 100 times that of *G. violaceus* RuBisCO activity. Despite the 87% and 70% sequence identity in LSU and SSU RuBisCO,

CHAPTER 5

Cloning and Characterization of Carbon Concentrating Mechanism (CCM) Proteins of *Gloeobacter violaceus* PCC 7421

5.1. Introduction

CCMs as mentioned before are mechanisms adopted by various photosynthetic organisms to overcome the confused specificity of RuBisCO by concentrating CO₂ in its vicinity and hence reduce the effect of competing substrates. CCM of cyanobacteria *viz.* carboxysomes is a proteinaceous microcompartment, enclosing RuBisCO and carbonic anhydrase. The generation of CO₂ by carbonic anhydrases (CA) in vicinity of RuBisCO coupled with restriction of its diffusive efflux from carboxysomes leads to the accumulation. The substrate for carboxysomes *i.e.*, HCO₃⁻ is assimilated by CO₂ and HCO₃⁻ active transporters both in the plasma membrane and the thylakoid membrane.

Carboxysomes have an icosahedral geometry. The proteins that form the proteinaceous shell belong to two domain classes: Bacterial microcompartment (BMC /Pfam00936) domain proteins existing as hexamers that form the flat facets of the shell and CcmL/EutN (Pfam 03319) domain proteins existing as pentamers that constitute the vertices of the shell. The residues of the BMC proteins identified to be important for hexamer-hexamer interactions are D-X-X-X-K motif and a less conserved domain R-P-H-X-N motif at the hexamer edges (Kinney et al., 2011). The pores acting as important channels for metabolites have conserved residues K-I-G-S, R-(A/V)-G-S and R-G-S-A-A in CcmK2, CcmK4 and CcmL, respectively (Kinney et al., 2011). Hence, the carboxysome shell is composed of CcmK2, CcmK4, CcmL and CcmO in β carboxysomes.

The other proteins identified in β carboxysomes involved in the functioning of the carboxysomes include CcmM and CcmN. CcmM encodes a protein with an N-terminal homologous to γ -CA and a C-terminal region having three to four RuBisCO small subunit repeats (SSU repeats) (Long et al., 2007). CcmM has been confirmed to be an essential structural component of the carboxysomes as its disruption leads to non-functional carboxysomes and inability of the organism to grow under limiting C_i conditions (detailed in chapter 1). The carboxysomal protein CcmN contains bacterial transferase hexapeptide repeat found in the N-terminal of CcmM as well, but whether this indicates potential structural homology between CcmM and CcmN, is still undeciphered.

The present chapter details the cloning and characterization of the CCM proteins mentioned above from an early diverging cyanobacteria *viz.* *Gloeobacter violaceus*.

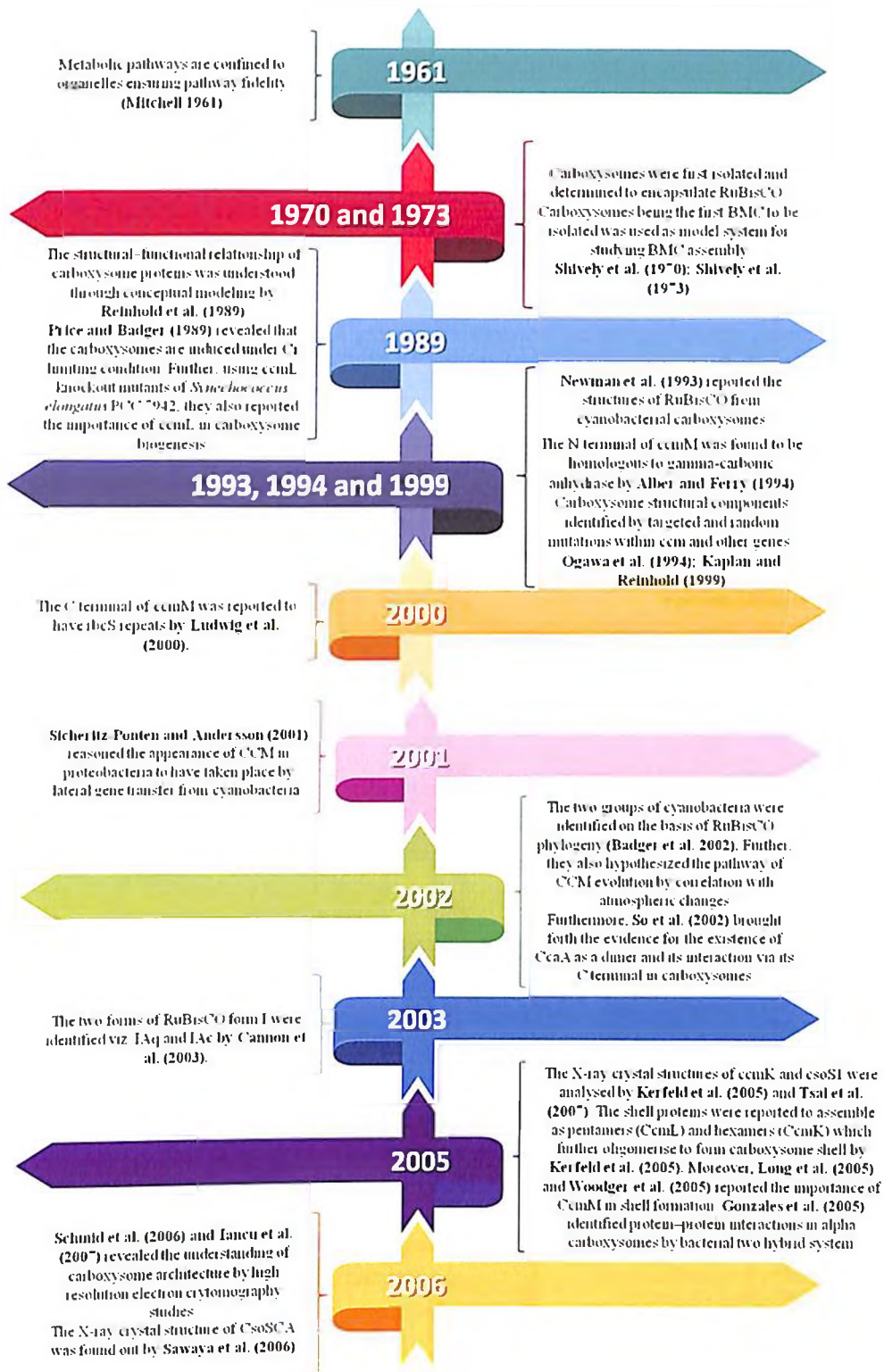


Figure 5A: Timeline of major discoveries and observations in carboxysomes from 1961-2016. While efforts have been made to include all the major events, this analysis will be inevitably incomplete and would reflect the research of specific interest to the author. (Pages 2-3)

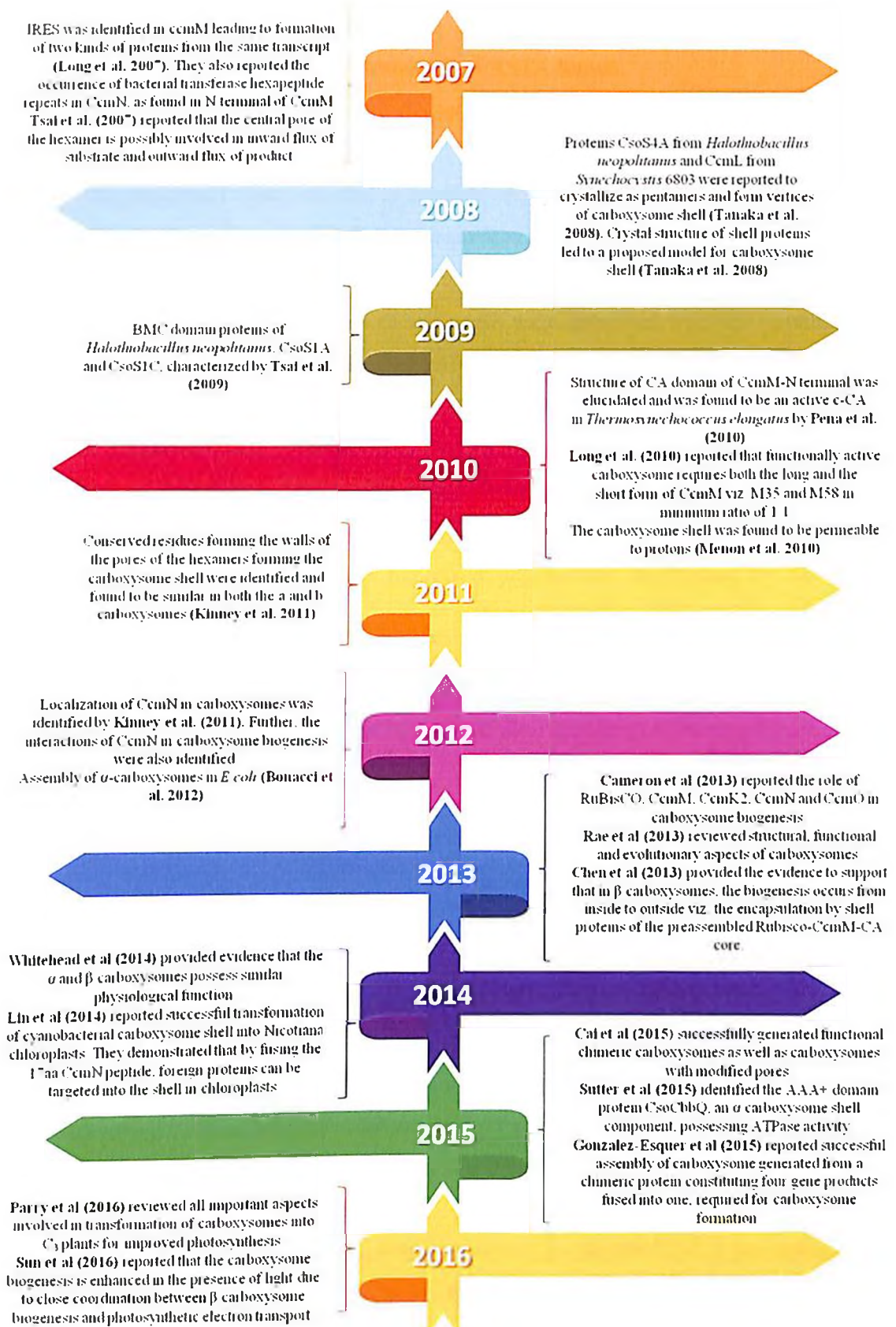
5.1. Introduction

CCMs as mentioned before are mechanisms adopted by various photosynthetic organisms to overcome the confused specificity of RuBisCO by concentrating CO₂ in its vicinity and hence reduce the effect of competing substrates. CCM of cyanobacteria viz. carboxysomes is a proteinaceous microcompartment, enclosing RuBisCO and carbonic anhydrase. The generation of CO₂ by carbonic anhydrases (CA) in vicinity of RuBisCO coupled with restriction of its diffusive efflux from carboxysomes leads to the accumulation. The substrate for carboxysomes *i.e.*, HCO₃⁻ is assimilated by CO₂ and HCO₃⁻ active transporters both in the plasma membrane and the thylakoid membrane.

Carboxysomes have an icosahedral geometry. The proteins that form the proteinaceous shell belong to two domain classes: Bacterial microcompartment (BMC /Pfam00936) domain proteins existing as hexamers that form the flat facets of the shell and CcmL/EutN (Pfam 03319) domain proteins existing as pentamers that constitute the vertices of the shell. The residues of the BMC proteins identified to be important for hexamer-hexamer interactions are D-X-X-X-K motif and a less conserved domain R-P-H-X-N motif at the hexamer edges (Kinney et al., 2011). The pores acting as important channels for metabolites have conserved residues K-I-G-S, R-(A/V)-G-S and R-G-S-A-A in CcmK2, CcmK4 and CcmL, respectively (Kinney et al., 2011). Hence, the carboxysome shell is composed of CcmK2, CcmK4, CcmL and CcmO in β carboxysomes.

The other proteins identified in β carboxysomes involved in the functioning of the carboxysomes include CcmM and CcmN. CcmM encodes a protein with an N-terminal homologous to γ-CA and a C-terminal region having three to four RuBisCO small subunit repeats (SSU repeats) (Long et al., 2007). CcmM has been confirmed to be an essential structural component of the carboxysomes as its disruption leads to non-functional carboxysomes and inability of the organism to grow under limiting Ci conditions (detailed in chapter 1). The carboxysomal protein CcmN contains bacterial transferase hexapeptide repeat found in the N-terminal of CcmM as well, but whether this indicates potential structural homology between CcmM and CcmN, is still undeciphered.

The present chapter details the cloning and characterization of the CCM proteins mentioned above from an early diverging cyanobacteria viz. *Gloeobacter violaceus*.



5.2. Methodology (detailed in chapter 2)

- Retrieved amino acid sequences for the CCM proteins from 20 β -cyanobacteria available at Kazusa Genome Resource in FASTA format.
- The amino acid sequences were aligned to identify conserved regions of the protein sequences using T-Coffee server.
- The alignments were then used to generate sequence logo using WebLogo3 server.
- The phylogenetic analysis of these sequences is mentioned in chapter 3, done using MEGA6.0.
- The CCM encoding genes of *G. violaceus* were amplified from the genome using PCR (conditions mentioned in chapter 2).
- The amplified fragments were then cloned into cloning vector and subsequently expression vector using standard cloning techniques.
- The recombinant plasmids generated were then used for transformation into expression host viz. *E. coli* BL21(DE3)pLysS and induced in the presence of IPTG.
- The molecular weight of the expressed protein was determined by SDS-PAGE analysis.
- The *in-silico* modeling of the proteins was done using homology modeling and the models hence generated were validated by Ramachandran plot, QMEAN score, Verify3D and TM scores.

5.3. Results and Discussions

5.3.1. CcmK

The CcmK protein is the shell protein of the carboxysome existing in the form of hexamers, which assemble to form the flat facets of the icosahedral structure. As shown in chapter 3, there exist at least two CcmK homologs in a β cyanobacterium.

The CcmK sequences from 20 β cyanobacteria when aligned by T Coffee server and alignment then used for generating the sequence logo (**Figure 5B**), showed 6 highly conserved regions. These regions were marked as site I, II, III, IV, V and VI. Site I is GXXE/Q, site II is GFPXX, site III is DXKK, site IV is R/VXTXV, site V is RGXXXEV and site VI is R/NPH/PXN. According to the available literature, site III is involved in hexamer-hexamer interaction and site IV although less conserved is also involved in interactions at the edges of the hexamers (Kinney et al., 2011). The pores of the hexamers is formed of KIGS residues in CcmK2, R(A/V)GS in CcmK4 and not characterized in other CcmK forms. Since the present sequence logo is generated by alignment of all available forms of CcmK (CcmL, CcmK1, CcmK2, CcmK3 and CcmK4), the pore residues reported do not appear to be

conserved. Such variation in the residues flanking the pore possibly has an effect on the movement of metabolites across the pore as adaptation to survive diverse metabolic environments. It should be noted that the KIGS and RA/VGS motifs (marked in **Figure 5B**) are less conserved than sites I, II, IV and V, which although more conserved have not been associated with any function as yet.

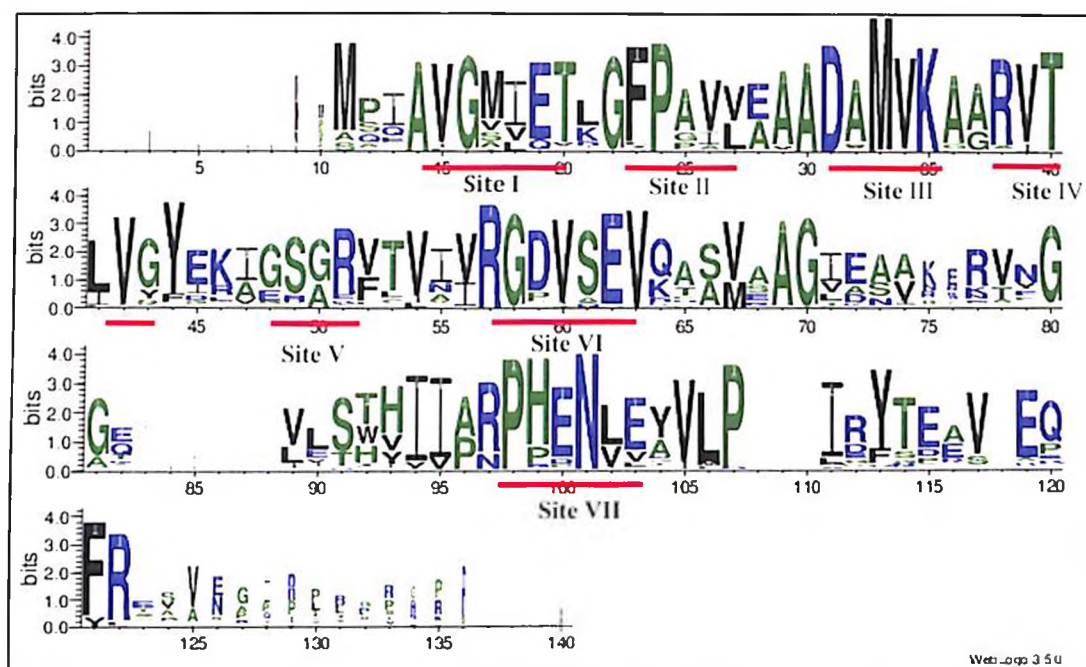


Figure 5B: The sequence logo for CcmK sequences from 20 β -cyanobacteria, aligned by T Coffee program. The underlined areas labeled as Site I-VII show the conserved regions of the sequence.

The carboxysome shell is constituted mainly by the CcmK protein, which form the flat facets enclosing the RuBisCO-CcmM-CA assembly. The enclosing of the cargo is mediated by proteins like CcmN which interacts with the shell and the enclosed protein complexes. Possibly, these uncharacterized conserved regions (mentioned above) could be responsible for the interaction with CcmN and hence mediate carboxysome assembly.

Figure 5C shows the tertiary structure of CcmK protein of *G. violaceus* derived by homology modeling. Out of the 5000 models generated by the Modeler program, the protein model shown was chosen owing to its lowest normalized DOPE score value i.e. -0.75352. The pdb. File of the model generated was then subjected to energy minimization by uploading it to ModRefiner server. The energy minimized model was then validated by PROCHECK analysis by Ramachandran plot (**Figure 5C**) which shows more than 90% residues lay in the most favorable region of the plot, hence the model generated is of good quality. The protein model was then analyzed by v.3.0 of Gail Hutchinson's PROMOTIF program to identify the motifs. The analysis revealed the protein structure to have 1 beta sheet constituting 4 strands

(type – antiparallel with topology -2X 1 2X); 1 beta hairpin (constituting strand 1 and strand 2) of class 2:2IP; 3 beta bulges; 4 strands (Val3-Leu11; Val29-Ser38; Arg41-Asp49; Gln71-Ile78); 4 helices (Phe13-Ala26; Val50-Arg66; Glu83-Val88; Glu95-Gln99); 1 helix-helix interaction (between helix 1 and 2); 5 beta turns (Ser38-Arg41(SSGR); Val67-Gly71(VYGG); Ile78-Pro81(IARP); Leu89-Arg92(LPIR); Ile107-Tyr110(IRPY)) and 1 gamma turn (Ala27-Val29(ARV)- of inverse type).

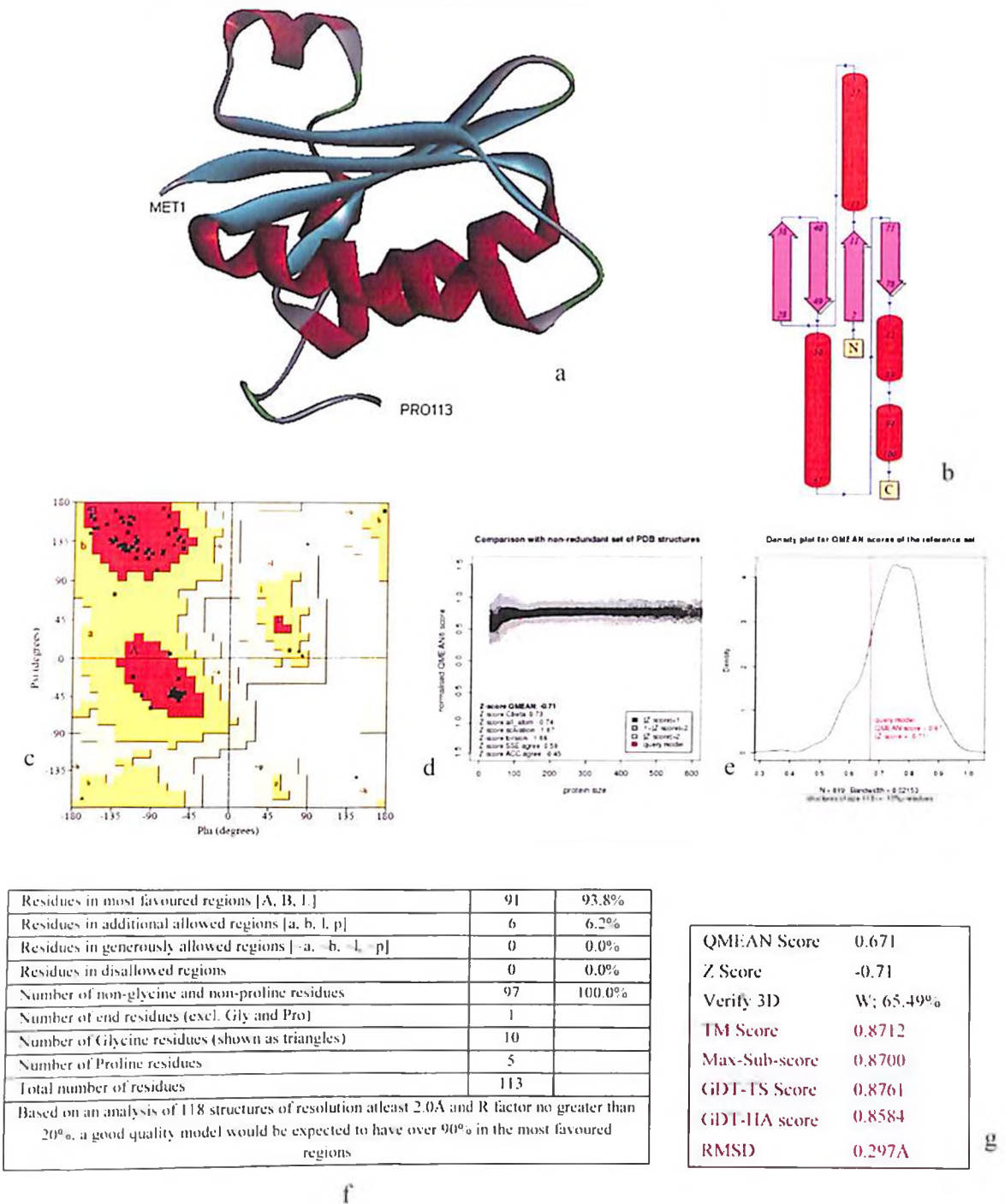


Figure 5C: The protein model for *G. violaceus* CcmK generated by homology modeling and its validation. a – The tertiary structure of CcmK showing the N terminal and the C terminal of the protein. b - The topology of the protein showing the secondary structure. c – Ramachandran plot for the protein model. d and e – QMEAN score of the CcmK model in comparison to the standard protein models. f – Ramachandran plot statistics showing >90% residues to be in the favorable region of the plot, hence validating it to be a good quality model. g – Scoring of the CcmK model according to QMEAN, Verify3d and TMScore (in red font, obtained by superposition of the CcmK model with template i.e. 3SSR) servers.

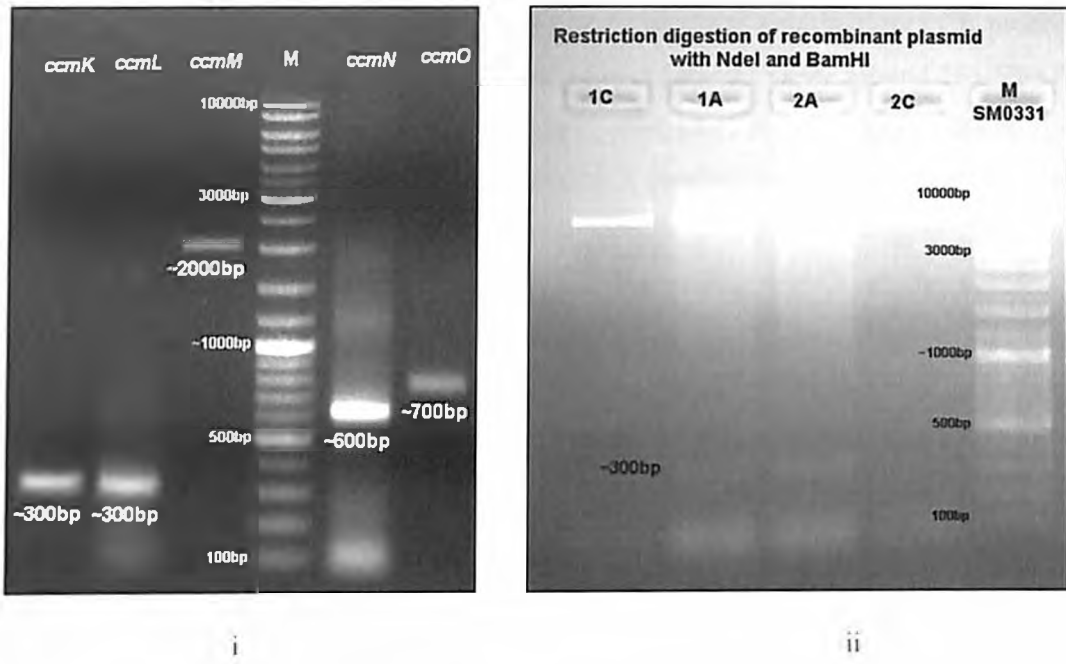


Figure 5D: i – The Agarose gel (1%) image for the polymerase chain reaction product for *ccmK* (first lane), *ccmL* (second lane), *ccmM* (third lane), Marker SM0331 (fourth lane); *ccmN* (fifth lane), *ccmO* (sixth lane). The respective sizes of the products obtained are mentioned below each band. ii - Confirmation of *pET15b-ccmK* construct by restriction digestion with enzymes *NdeI* and *BamHI*. The single digestion of the *pET15b-ccmK* gave single band of size ~6Kbp (1C and 2C lane 1 and 4, respectively). The double digestion of the construct (Lane 1A and 2A – lane 2 and 3) gave two bands of sizes ~5.7Kbp (size of empty vector *pET15b*) and ~300bp (size of *ccmK* gene)

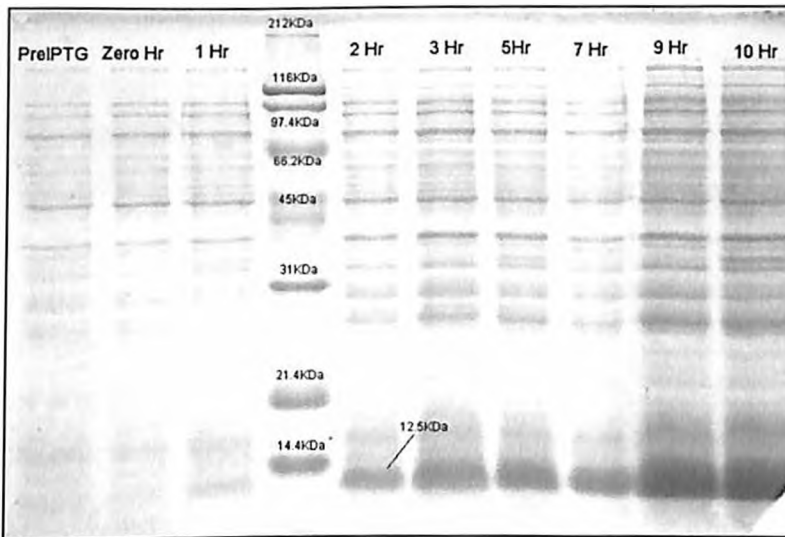


Figure 5E: SDS-PAGE (12%) to check the expression of 12.5kDa *G. violaceus* PCC 7421 *ccmK* (approx 300bp gene cloned in *pET15b*) in *E. coli* BL21(DE3)pLysS by 1mM IPTG induction. Lane 1- Uninduced crude extract, lane 2 – 0 hrs after IPTG induction, lane 3 – 1 hr after induction, Lane 4 – standard protein ladder, lane 5-10 – crude extract after 2, 3, 5, 7, 9 and 10 hrs of induction, respectively.

The *ccmK* region of the genome of *G. violaceus* PCC 7421 was amplified by PCR procedure using conditions and reaction mixture as mentioned in chapter 2 (Table 2E and 2F). The PCR product was analyzed on agarose gel (1%) (Figure 5D(i)) and found to be of ~300bp (exact size is 342bp). The PCR product hence obtained was purified by gel extraction and cloned into *pET15b* vector using standard cloning procedures (mentioned in chapter 2) between restriction sites *NdeI-BamHI*, downstream of his-tag coding region of the vector. The *pET15b-ccmK* clone was confirmed by restriction digestion (Figure 5D(ii)). The double digestion of the vector *pET15b-ccmK* yielded 2 bands of sizes ~5.7Kbp (size of *pET15b*) and ~300bp (approximate size of *ccmK*). The sequence of the construct was also confirmed by DNA sequencing facility.

Once confirmed, the *pET15b-ccmK* construct was transformed into *E. coli* BL21(DE3)pLysS and analyzed for expression of protein by IPTG induction (procedure mentioned in chapter 2). The theoretical pI and molecular weight as determined by ExPASy tool pI/MW (available online - http://web.expasy.org/compute_pi/) was found to be 6.84 and 12234.13 daltons, respectively. The expressed protein *in-vivo* also has the his-tag, hence the expected molecular weight of the protein is ~13234.13 daltons (molecular weight of his-tag is approximately 1000 daltons). The protein of expected size viz. ~13kDa was found to express after 2 hrs induction by IPTG @ 1mM (working concentration). The expression was found to increase upto 2 hrs after IPTG induction (Figure 5E).

5.3.2. CcmL

CcmL protein assembles in the form of pentamers which form the vertices of the icosahedrons (carboxysome). The knockout studies of the Δ *ccmL* mutants have shown formation of leaky carboxysomes, incapable of surviving low Ci conditions. The alignment of CcmL (amino acid sequences) from 20 β cyanobacteria shows eight most conserved regions of the sequence, labeled as site I-VIII (underlined red) in Figure 5F. The important residues reported till date, to be involved in pentamer interaction and the pore residues have been marked with open arrows and closed arrows, respectively in Figure 5F. The location of these conserved residues with respect to the secondary structure of CcmL is shown in Table 5A.

The residues involved in pentamer interaction constitute sites I, II, IV, VII and VIII. The conserved residues in site V have been reported to be a part of the pore of the pentameric structure. The literature provides evidence for only some of the residues in the conserved regions, while the rest viz. site III, site VI and some residues in other sites have not been associated with any particular role. Since, CcmL is reported to form only the vertices of the carboxysome, only 12 pentamers are possibly required to form a functional microcompartment, i.e. 60 protein molecules. Further, as reported by Sutter et al. (2013) the

pentamer has a pore of diameter 4Å, (comparatively small as the pore diameter of CemK1 and CemK2 is 4.8Å and 5.5Å, respectively), which makes it unsuitable to act as entry/exit point for the metabolites. The uncharacterized sites opens quest for identification of their role in carboxysome structure and function.

Table 5A: The most conserved regions of the CemL protein sequence alignment generated using CemL amino acid sequence from 20 β-cyanobacteria, labeled as site I-VIII in **Figure 5F**. The table shows the location of the conserved regions with respect to the secondary structure of the protein. In the conserved sequences mentioned in the table below, the variation of only two residues at one loci is represented by mentioning both the residues separated by an oblique and the variation of more than two residues at a loci in represented as X.

| Site Name | Conserved sequence | Location in secondary structure |
|-----------|--------------------|---------------------------------|
| Site I | GTVV/TST/SXK | Strand β1 |
| Site II | GXKF/LLL/VXQXXD | Strand β2 |
| Site III | YXVAXD | Strand β3 |
| Site IV | VGAG | Turn 3 |
| Site V | EWVLXXXGS/GXXR | Strand β4 and helix α1 |
| Site VI | R/KPXDAXV | Turn 5 |
| Site VII | IIDTV/I | Between strand β5 and β6 |
| Site VIII | YXK/QK/RXXXR | End of β7 |

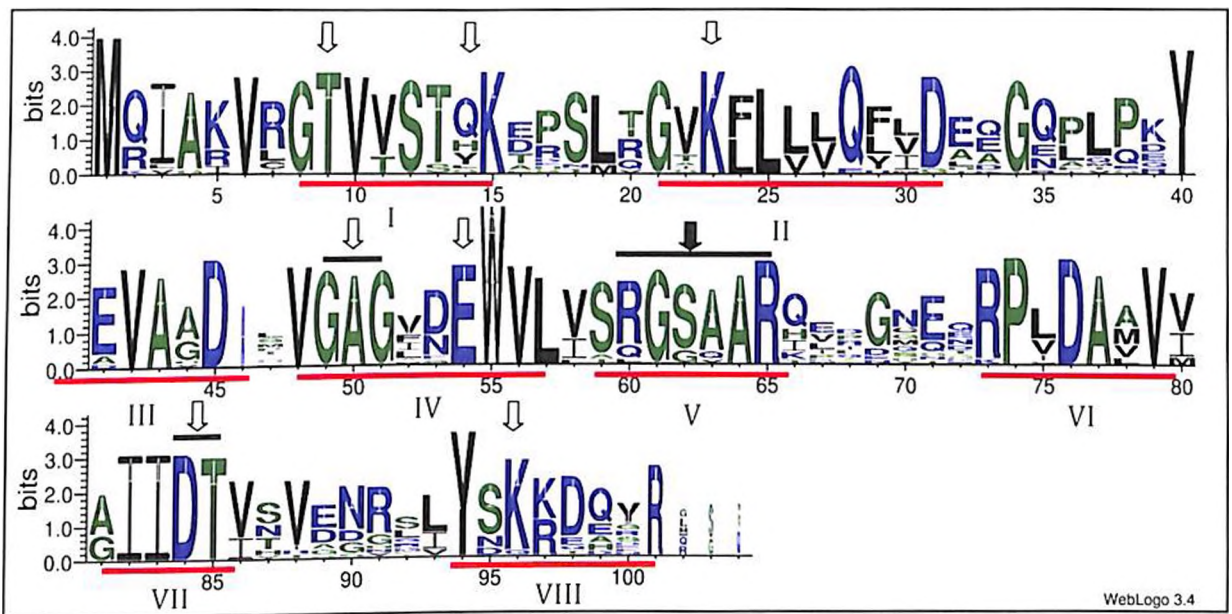


Figure 5F: The sequence logo for CemL sequences from 20 β-cyanobacteria, aligned by T Coffee program. The underlined areas labeled as Site I-VIII show the conserved regions of the sequence. The residues marked in open arrows represent the residues involved in interaction to form pentamer and the ones marked in closed arrows represent the pore residues of the pentamer.

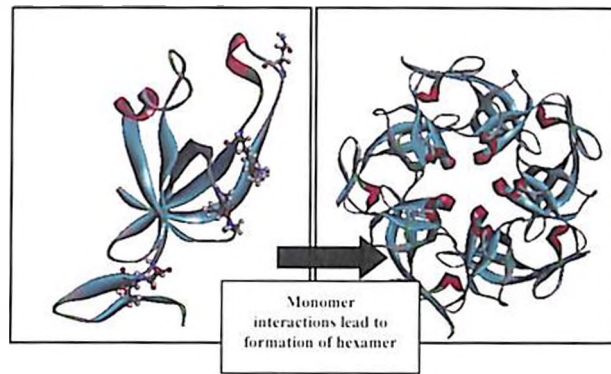


Figure 5G: The CcmL protein showing the residues involved in interactions between the monomers to form the hexamer.

Figure 5H: Confirmation of *pCOLDDII-ccmL* construct by restriction digestion with enzymes *NdeI* and *XhoI*. The digestion of the construct gave two bands of sizes ~4.4 Kbp (size of empty vector *pCOLDDII*) and 350 bp (size of *ccmL* gene). Lane 1- Standard DNA marker, Lane 2- *pCOLDDII-ccmL* digested with *NdeI* and *XhoI*, Lane 3- *pCOLDDII-ccmL* digested with *NdeI* and Lane 4- undigested *pCOLDDII-ccmL*.

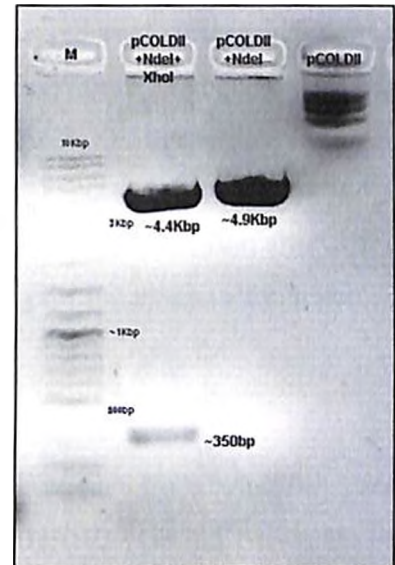
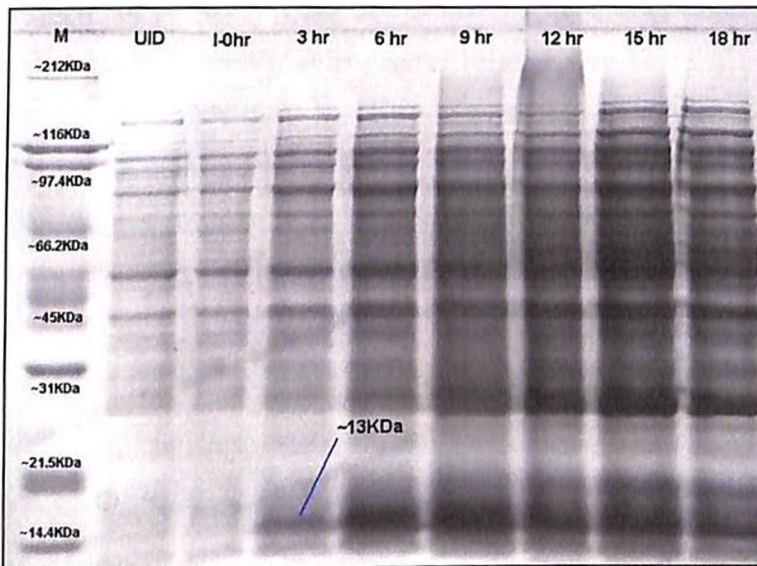


Image 5I: The SDS-PAGE (12%) image for the expression profile of ~12kDa protein of *G. violaceus* PCC 7421. CcmL (approx 350bp gene cloned in *pCOLDDII*) in *E. coli* BL21(DE3)pLysS by 1mM IPTG induction. Lane 1- Standard protein marker, Lane 2- crude extract from uninduced culture, Lane 3-10- Crude extract from cultures after 1, 2, 3, 5, 7, 9 and 10 hrs of induction.



The crystal structure of *G. violaceus* PCC 7421 CcmL is already reported by Sutter et al. (2013). The **Figure 5G** shows the residues involved in monomer interactions to form the hexamer. The *ccmL* region of the genome of *G. violaceus* PCC 7421 was amplified using polymerase chain reaction using primers, reaction mixture and reaction conditions as

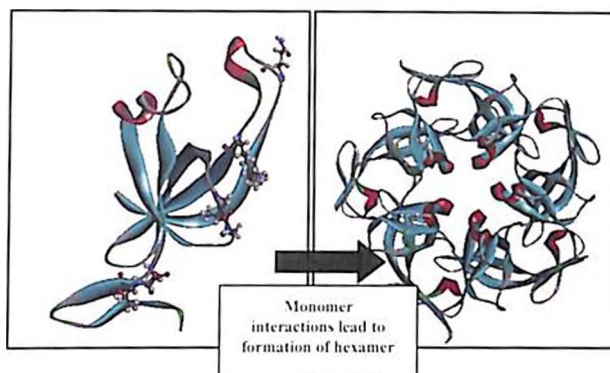


Figure 5G: The CcmL protein showing the residues involved in interactions between the monomers to form the hexamer.

Figure 5H: Confirmation of *pCOLDII-ccmL* construct by restriction digestion with enzymes *NdeI* and *XhoI*. The digestion of the construct gave two bands of sizes ~4.4 Kbp (size of empty vector *pCOLDII*) and 350 bp (size of *ccmL* gene). Lane 1- Standard DNA marker, Lane 2- *pCOLDII-ccmL* digested with *NdeI* and *XhoI*, Lane 3- *pCOLDII-ccmL* digested with *NdeI* and Lane 4- undigested *pCOLDII-ccmL*.

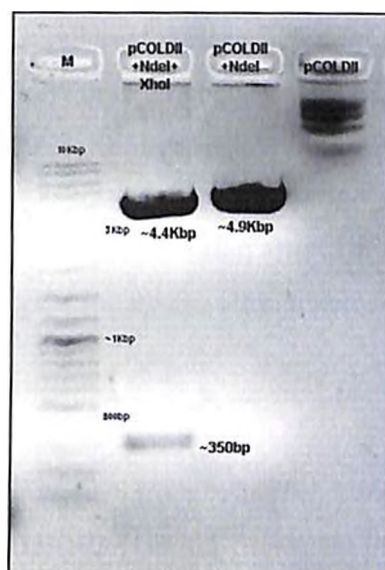
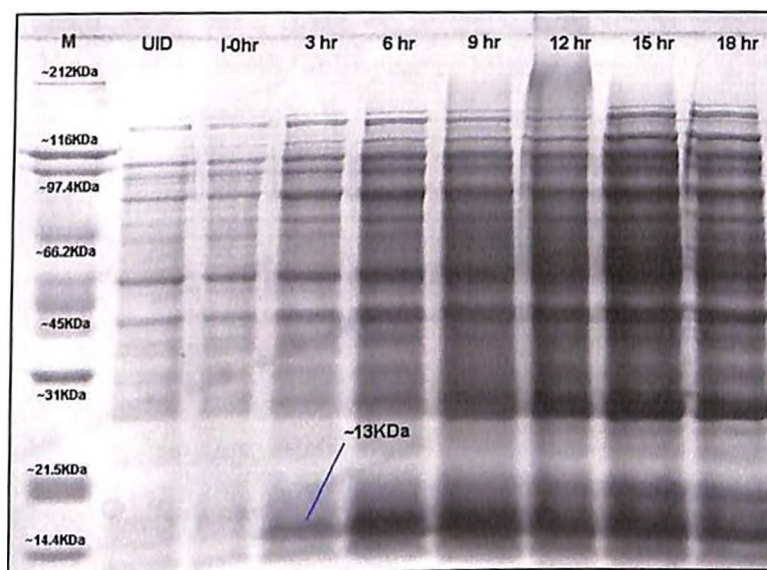


Image 5I: The SDS-PAGE (12%) image for the expression profile of ~12kDa protein of *G. violaceus* PCC 7421, CcmL (approx 350bp gene cloned in *pCOLDII*) in *E. coli* BL21(DE3)pLysS by 1mM IPTG induction. Lane 1- Standard protein marker, Lane 2- crude extract from uninduced culture, Lane 3-10- Crude extract from cultures after 1, 2, 3, 5, 7, 9 and 10 hrs of induction.



The crystal structure of *G. violaceus* PCC 7421 CcmL is already reported by Sutter et al. (2013). The **Figure 5G** shows the residues involved in monomer interactions to form the hexamer. The *ccmL* region of the genome of *G. violaceus* PCC 7421 was amplified using polymerase chain reaction using primers, reaction mixture and reaction conditions as

mentioned in chapter 2 in Tables 2D, 2E and 2F, respectively. The Figure 5D(i) shows the PCR amplified *cemL* region, of approximately 300bp (exact size of the gene is 303bp), of the *G. violaceus* genome. The product hence obtained was cloned into *pCOL.DII* vector between restriction sites *NdeI* and *XhoI* using standard cloning protocol (mentioned in chapter 2). The recombinant plasmid generated was confirmed by restriction digestion with the *NdeI* and *NdeI-XhoI*, to observe the shift in the size of the DNA single digested and double digested, respectively (Figure 5H). The sequence of the recombinant plasmid was also confirmed by DNA sequencing. The clone once confirmed was transformed into *E. coli* BL21(DE3)pLysS cells and the *cemL* gene was induced in the presence of IPTG @ 1mM (working concentration), after a cold shock at 15°C for 30 minutes. The expected molecular weight of the CcmL protein estimated by using online tool by ExPASy (http://web.expasy.org/cgi-bin/compute_pi/pi_tool) is 10829.39 daltons. Since, the protein is chimeric with a fused his-tag (approx molecular weight of his-tag is 1000daltons), the expected molecular weight is about 13kDa (Figure 5I). The cold shock and IPTG induction lead to expression after 1 hr of induction. The expression was found to increase upto 7 hrs post induction, after which the protein expression decreased.

5.3.3. CcmM

The CcmM protein is a multidomain protein. There is an N terminal γ carbonic anhydrase (CA) domain and C terminal RuBisCO small subunit (SSU) repeats. The γ CA domain has been reported to be functional carbonic anhydrase in *Thermosynechococcus elongatus* BPI CcmM protein. Further, as reported by Long et al. (2007) the CcmM protein, a product of ~2000bp gene, leads to formation of two proteins from the same genetic code by virtue of the internal ribosomal entry site (IRES) present in its sequence. The M58 and M35 are the two forms of CcmM encoded from genome of *Synechococcus elongatus* PCC 7942. Long et al. (2007) have reported M35 to be involved in assembly of RuBisCO and M58 to have role in encapsulation of the RuBisCO-M35 complex by the shell proteins. The *G. violaceus* PCC 7421, CcmM protein was analyzed to identify the putative IRES. Further, the N terminal of the protein was searched to identify the conserved, functionally important residues of γ CA in order to predict its role.

5.3.3.1. Prediction of IRES

The gene sequence of CcmM of *G. violaceus* PCC 7421 was retrieved from Kazusa Genome Resource. The IRES is characterized by the occurrence of the Shine-Dalgarno (SD) sequence i.e. a ribosomal entry site in mRNA generally located ~8bp upstream of the start codon. The Shine-Dalgarno sequence occurs in bacterial as well as Archaea sequences and also the chloroplast and the mitochondrial transcripts. It is a six-base consensus sequence AGGAGG,

which may vary to some extent, for eg. it is found to be AGGAGGU in *E. coli* and GAGG in T4 virus genes. The SD sequence identified in transcripts of CcmM from various other cyanobacteria as reported by Long et al. (2007) is GAGG, AGGAG, AGAA, GAA and AGA in *S. elongatus* PCC 7942, *Anabaena* sp PCC 7120, *Synechococcus* sp WH8105, *Synechococcus* sp PCC 7002 and *Synechocystis* sp PCC 6803, respectively.

Table 5B: Results of SD sequence, followed by start codon, search in the *cmmM* gene sequence. The most plausible SD sequences from the search were named as IRES 1-3 (in order of their favorability).

| Possible SD seq | Locs on gene | GTG(+/-);distance from SD | ATG(+/-);distance from SD | Plausible IRES |
|-----------------|--------------|---------------------------|---------------------------|----------------|
| GGGAA | 147 | +; 9bp | - | |
| GGA | 210 | +; 4bp | +; 10bp | IRES 3 |
| GGA | 237 | - | - | |
| GGA | 429 | -; 1bp | - | |
| GGA | 435 | - | - | |
| GGA | 516 | - | - | |
| GAGAG | 622 | - | - | |
| GGAGG | 834 | +; 0bp | - | |
| GGAG | 852 | +; 3bp | - | IRES 1 |
| GAGG | 967 | - | - | |
| GGAGG | 1059 | - | - | |
| GGAAG | 1188 | -; 1bp | - | |
| GGAG | 1221 | - | - | |
| GGGGGAG | 1324 | - | - | |
| GGAGG | 1527 | +; 0bp | - | |
| GGAG | 1545 | +; 3bp | - | IRES 2 |
| GAGG | 1759 | - | - | |
| GGAGG | 1887 | - | - | |
| GGAG | 1905 | +; 3bp | - | |
| GGGGGAG | 1930 | +; 3bp | - | |

The *G. violaceus* PCC 7421 *cmmM* sequence was searched for all the reported SD sequences followed by a start codon (ATG/GTG in range of ~8bp). **Table 5B** shows the results of the search conducted. Three loci in the *cmmM* gene were found to be possible SD sequences viz IRES 1, IRES 2 and IRES 3. IRES 1 is located at 852bp loci which is the region where the CA domain ends and the SSU repeats have not started. IRES 2 is located at 1545 i.e. between the first and the second SSU repeat. While IRES 3 located at 210bp, is lying in between of the CA domain. The expected protein products from these IRES are of 382aa, 151aa and 596aa, for IRES 1, 2 and 3, respectively. The IRES leads to the formation of protein with RuBisCO SSU repeats. The expected IRES 2 protein product is too short and IRES 3 product is too long to fit in the criterion of a functional CcmM short form protein. Theoretically, the long form of the protein has both the CA domain and the SSU repeats domain, while the short form of the protein has only the RuBisCO SSU repeats. Hence, the IRES must lie just before the beginning of the SSU repeats and after the CA domain regions. The MSA, represented in the form of sequence logo (**Figure 5L**) for CcmM shows residues M and V (for codons ATG and

GTTG, respectively) to be significantly conserved showing the putative location of the IRES to overlap in all the reported CcmM sequences. Hence, IRES 1 is supposedly the expected entry site for the formation short form of CcmM of *G. violaceus*.

5.3.3.2. Analysis of the conserved residues for CA domain in CcmM

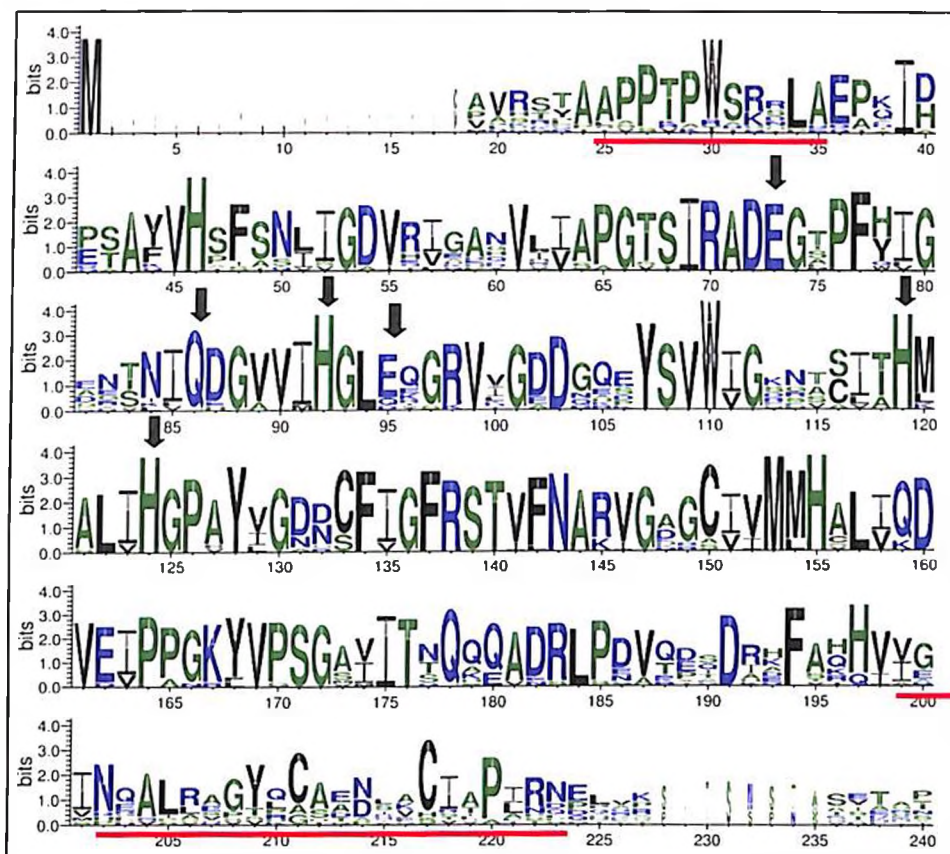


Figure 5J: The sequence logo for the CcmM N terminal domain of the CcmM protein from 20 β cyanobacteria. The residues marked in closed arrows are the residues important for the γ CA activity and the residues underlined in red are regions important for formation for a stable CA protein (APPTPWSRDLA and VVGINEALLSGYQCAENIACIAPIR) as reported by Pena et al. (2010).

Long et al. (2007) have reported the N-terminal of CcmM to possess γ CA like domain. Pena et al. (2010) have reported *T. elongatus* BP1 CcmM to have carbonic anhydrase activity. The CcmM protein sequence of *G. violaceus* and from other β cyanobacteria was analyzed for the presence of residues important for γ CA activity (Figure 5J and 5K). Iverson et al. (2000) characterized the active site residues of the γ CA from *Methanosarcina thermophila*. As per their reports, the active site of the enzyme is constituted by a trimeric arrangement of the protein. This trimer contains two active sites. In one active site three histidine residues are involved, 2 contributed from one monomer and the third one from the adjacent monomer. The histidine residues involved are His81 and His122 from one monomer and His 117 from adjacent monomer in *Methanosarcina thermophila*. Further, Glu62 and Glu84 from one monomer and Gln75 from adjacent monomer are also important for active site.

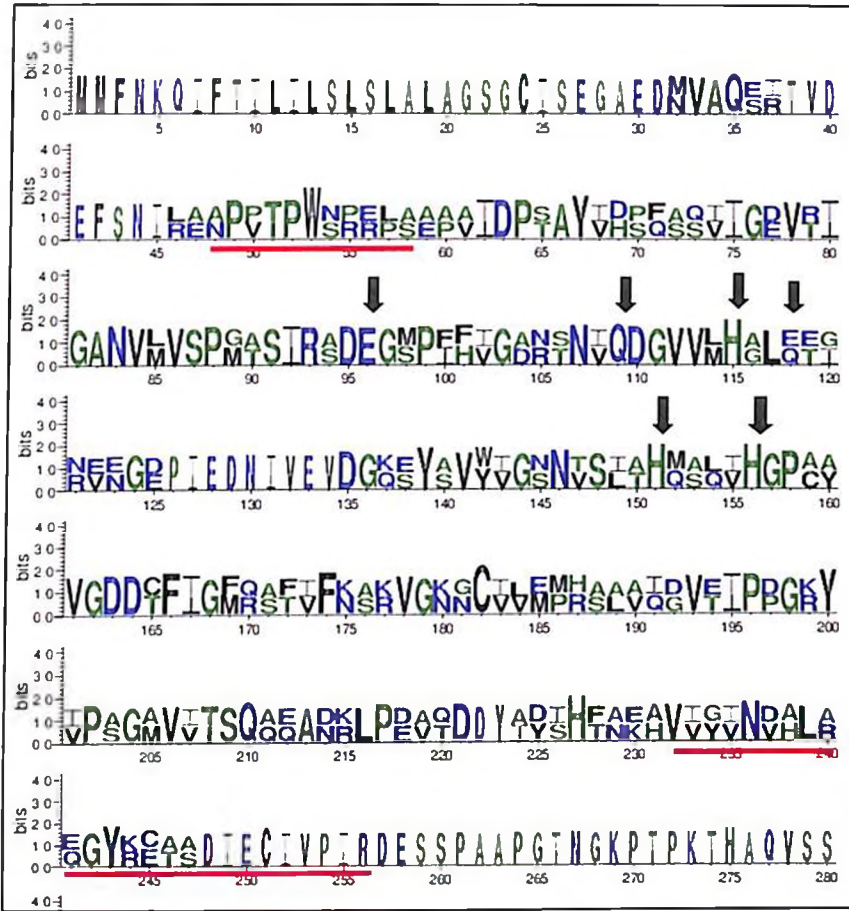


Figure 5K: The sequence logo for the CcmM N terminal domain of the CcmM protein from *G. violaceus* and γ CA of *Methanosarcina thermophila*. The residues marked in closed arrows are the residues important for the γ CA activity and the residues underlined in red are regions important for formation for a stable CA protein (APPTPWSRDLA and VVGINEALLSGYQCAENIACIPIR) as reported by Pena et al (2010).

These residues being conserved in the 20 CcmM sequences aligned in present study shows the possibility of existence of active CA domain. Further, as Pena et al. (2010) APPTPWSRDLA and VVGINEALLSGYQCAENIACIPIR sequences present in CA domain of *T. elongatus* are necessary for stabilizing the protein in an active conformation. These residues are conserved in *G. violaceus* (Figure 5K) but not in all the cyanobacterial CcmM sequences (Figure 5J), hence, *G. violaceus* CcmM has a plausible chance to possess γ CA activity as in case of *T. elongatus* CcmM.

small subunit (possibly from the same organism, as the one who's CcmM is used as query). The exercise when performed, focusing on the RbcS hits in the search, RuBisCO SSU from other organisms was found to be more closely related as compared to the SSU from the same organism. Further, surprisingly the closest hit obtained in all 20 searches resulted in the same set of top hits, suggesting not only an evolutionary link between the CcmM and RbcS but also the speculated set of organisms from which this protein might have originated. The top five RbcS hits obtained in all searches (the order of identity was variable) included the protein from *Leptolyngbya sp O-77*, *Chamaesiphon minutes*, *Oscillatoria nigro-viridis*, *Nodularia spumigena*, *Chrysochloris ovalisporum*, *Crinalium epipsammum*, *Microcoleus vaginatus* and *Fischerella*. The degree of identity observed was variable for each of the SSU repeats in the CcmM. Furthermore, the number of SSU repeats was also variable in the CcmM proteins under study. **Table 5C** shows the number of SSU repeats in the CcmM from different sources and the degree of identity with the best found hits of RbcS in BLASTP search.

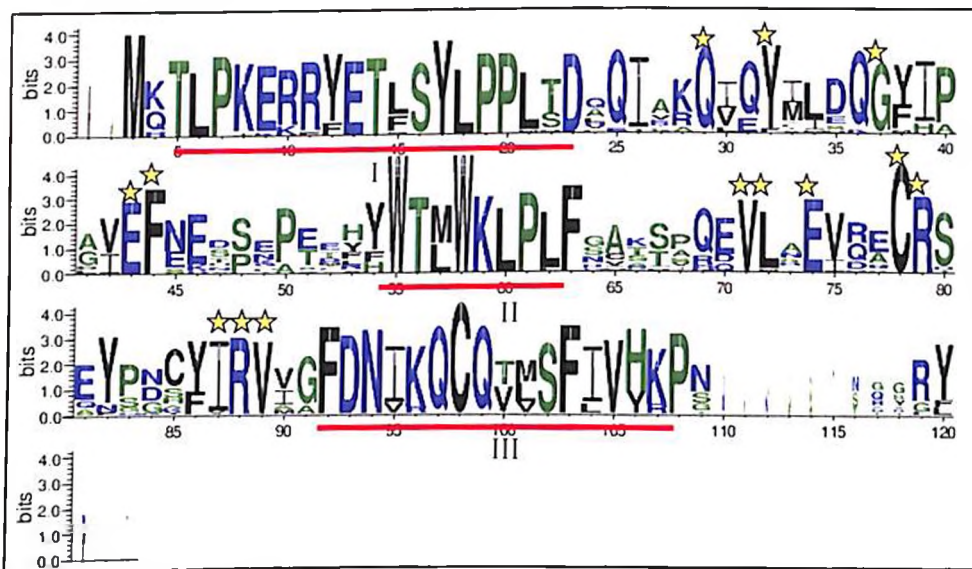


Figure 5M: Sequence logo for the multiple sequence alignment of the RuBisCO small subunit from β cyanobacteria, obtained by alignment in MEGA6.0 using Clustal W followed by input of alignment in FASTA format to the WebLogo3 tool. The red underlined regions and the starred residues are the most conserved regions of the protein. The sites marked as I, II and III are the most conserved regions apart from the residues marked with yellow stars.

Table 5C: The number of SSU repeats in the CcmM from different sources and the degree of identity with the best found hits of RbcS in BLASTP search

| Name of Organisms | RbcS best hits from | Number of SSU repeats | Query cover (%) | % Identity |
|--|---------------------------------------|-----------------------|-----------------|------------|
| <i>Gloeobacter violaceus</i> PCC 7421 | <i>Leptolyngbya sp O-77</i> | 4 | 65 | 47 |
| | <i>Chamaesiphon minutus</i> | | 65 | 45 |
| | <i>Nodularia spumigena</i> | | 61 | 54 |
| | <i>Chrysochloris ovalisporum</i> | | 62 | 50 |
| | <i>Crinalium epipsammum</i> | | 58 | 58 |
| <i>Synechocystis sp PCC</i> 6803 | <i>Leptolyngbya sp O-77</i> | 4 | 65 | 49 |
| | <i>Chamaesiphon minutus</i> | | 64 | 42 |
| | <i>Oscillatoria nigro-viridis</i> | | 60 | 52 |
| | <i>Crinalium epipsammum</i> | | 62 | 47 |
| | <i>Microcoleus vaginatus</i> | | 61 | 50 |
| <i>Anabaena sp PCC</i> 7120 | <i>Leptolyngbya sp O-77</i> | 3 | 60 | 46 |
| | <i>Chamaesiphon minutus</i> | | 59 | 44 |
| | <i>Chrysochloris ovalisporum</i> | | 55 | 56 |
| | <i>Nostoc sp PCC 7524</i> | | 57 | 53 |
| | <i>Cylindrospermum stagnale</i> | | 55 | 68 |
| <i>Anabaena variabilis</i> ATCC 29413 | <i>Leptolyngbya sp O-77</i> | 3 | 60 | 46 |
| | <i>Chamaesiphon minutus</i> | | 63 | 44 |
| | <i>Chrysochloris ovalisporum</i> | | 56 | 56 |
| | <i>Cylindrospermum stagnale</i> | | 55 | 68 |
| | <i>Nostoc sp PCC 7524</i> | | 57 | 66 |
| <i>Nostoc punctiforme</i> ATCC 29133 | <i>Leptolyngbya sp O-77</i> | 4 | 66 | 48 |
| | <i>Chamaesiphon minutus</i> | | 65 | 42 |
| | <i>Nostoc piscinale</i> | | 66 | 38 |
| | <i>Nodularia spumigena</i> | | 62 | 55 |
| | <i>Chrysochloris ovalisporum</i> | | 61 | 53 |
| <i>Arthrospira platensis</i> NIES 39 | <i>Leptolyngbya sp O-77</i> | 3 | 61 | 46 |
| | <i>Chamaesiphon minutus</i> | | 60 | 47 |
| | <i>Oscillatoriales cyanobacterium</i> | | 48 | 67 |
| | <i>Cylindrospermum stagnale</i> | | 60 | 54 |
| | <i>Microcoleus vaginatus</i> | | 61 | 39 |
| <i>Cyanothece sp PCC</i> 8801 | <i>Leptolyngbya sp O-77</i> | 4 | 66 | 50 |
| | <i>Chamaesiphon minutus</i> | | 65 | 46 |
| | <i>Fischerella muscicola</i> | | 64 | 53 |
| | <i>Nostoc sp PCC 7524</i> | | 61 | 55 |
| | <i>Oscillatoria acuminata</i> | | 60 | 53 |
| <i>Cyanothece sp ATCC</i> 51142 | <i>Leptolyngbya sp O-77</i> | 4 | 67 | 48 |
| | <i>Chamaesiphon minutus</i> | | 65 | 48 |
| | <i>Microcoleus vaginatus</i> | | 59 | 51 |
| | <i>Fischerella sp NIES-3754</i> | | 74 | 56 |
| <i>Cyanothece sp PCC</i> | <i>Leptolyngbya sp O-77</i> | 4 | 67 | 47 |

| | | | | |
|--|---------------------------------------|---|----|----|
| 7424 | <i>Chamaesiphon minutus</i> | | 66 | 41 |
| | <i>Oscillatoriales cyanobacterium</i> | | 57 | 52 |
| <i>Trichodesmium erythraeum</i> IMS101 | <i>Leptolyngbya</i> sp O-77 | 4 | 67 | 48 |
| | <i>Chamaesiphon minutus</i> | | 65 | 43 |
| <i>Synechococcus</i> sp JA-3-3.Ah | <i>Chamaesiphon minutus</i> | 4 | 64 | 47 |
| | <i>Leptolyngbya</i> sp O-77 | | 63 | 46 |
| <i>Thermosynechococcus elongatus</i> BP1 | <i>Leptolyngbya</i> sp O-77 | 4 | 64 | 44 |
| | <i>Chamaesiphon minutus</i> | | 65 | 42 |
| | <i>Leptolyngbya</i> sp PCC 6406 | | 60 | 43 |
| <i>Synechococcus</i> sp JA-2-3B'at(2-13) | <i>Chamaesiphon minutus</i> | 4 | 64 | 47 |
| | <i>Leptolyngbya</i> sp O-77 | | 66 | 47 |
| <i>Microcystis aeruginosa</i> NIES 843 | <i>Leptolyngbya</i> sp O-77 | 4 | 69 | 50 |
| | <i>Chamaesiphon minutus</i> | | 66 | 45 |
| | <i>Fischerella muscicola</i> | | 75 | 50 |
| | <i>Fischerella</i> sp PCC 9605 | | 66 | 52 |
| <i>Cyanothece</i> sp PCC 7425 | <i>Leptolyngbya</i> sp O-77 | 4 | 66 | 47 |
| | <i>Chamaesiphon minutus</i> | | 68 | 41 |
| | <i>Oscillatoris nigro-viridis</i> | | 64 | 46 |
| | <i>Microcoleus vaginatus</i> | | 64 | 45 |
| | <i>Oscillatoriales cyanobacterium</i> | | 53 | 65 |
| <i>Synechococcus elongatus</i> PCC 6301 | <i>Leptolyngbya</i> sp O-77 | 3 | 61 | 48 |
| | <i>Chamaesiphon minutus</i> | | 59 | 45 |
| | <i>Oscillatoris nigro-viridis</i> | | 73 | 39 |
| | <i>Cyanobacterium</i> PCC 7702 | | 58 | 47 |
| | <i>Microcoleus vaginatus</i> | | 73 | 37 |
| <i>Synechococcus elongatus</i> PCC 7942 | <i>Leptolyngbya</i> sp O-77 | 3 | 61 | 47 |
| | <i>Chamaesiphon minutus</i> | | 59 | 45 |
| | <i>Oscillatoris nigro-viridis</i> | | 73 | 39 |
| | <i>Cyanobacterium</i> PCC 7702 | | 58 | 47 |
| | <i>Microcoleus vaginatus</i> | | 73 | 37 |
| <i>Acaryochloris marina</i> MBIC11017 | <i>Leptolyngbya</i> sp O-77 | 5 | 71 | 49 |
| | <i>Chamaesiphon minutus</i> | | 71 | 46 |
| | <i>Microcoleus vaginatus</i> | | 70 | 41 |
| | <i>Oscillatoris nigro-viridis</i> | | 70 | 38 |
| | <i>Nostoc</i> sp PCC 7524 | | 65 | 49 |
| <i>Synechococcus</i> sp PCC 7002 | <i>Chamaesiphon minutus</i> | 4 | 65 | 39 |
| | <i>Leptolyngbya</i> sp O-77 | | 65 | 38 |

5.3.3.4. The in-silico generated protein model for CcmM IRES1

The **Figure 5N** shows the tertiary structure for IRES1 generated by homology modeling. The model with the lowest DOPE score value from among the 5000 models generated by Modeller was selected and subjected to further energy minimization by ModRefiner (online server). The energy minimized pdb. file was then submitted to PDBsum for validation. The Ramachandran plot statistics show ~89% residues in the most favored region of the plot, which is less than the criterion required for a good quality model. The crystal structure for CcmM has not been reported till date, hence the homology modeling was done in segments of the protein. The model (**Figure 5N**) shows the SSU repeat regions arranged at approximately the three vertices of a triangle. *G. violaceus* CcmM has four SSU repeats, three constituting the complete domain and one is partially conserved. The arrangement of the SSU domain is in accordance with the proposed interaction between M35 and RuBisCO protein in process of carboxysome assembly (by Long et al., 2010). The model for the complete CcmM protein requires modeling using different templates followed by threading due to absence of any CcmM structure which could be used as a template.

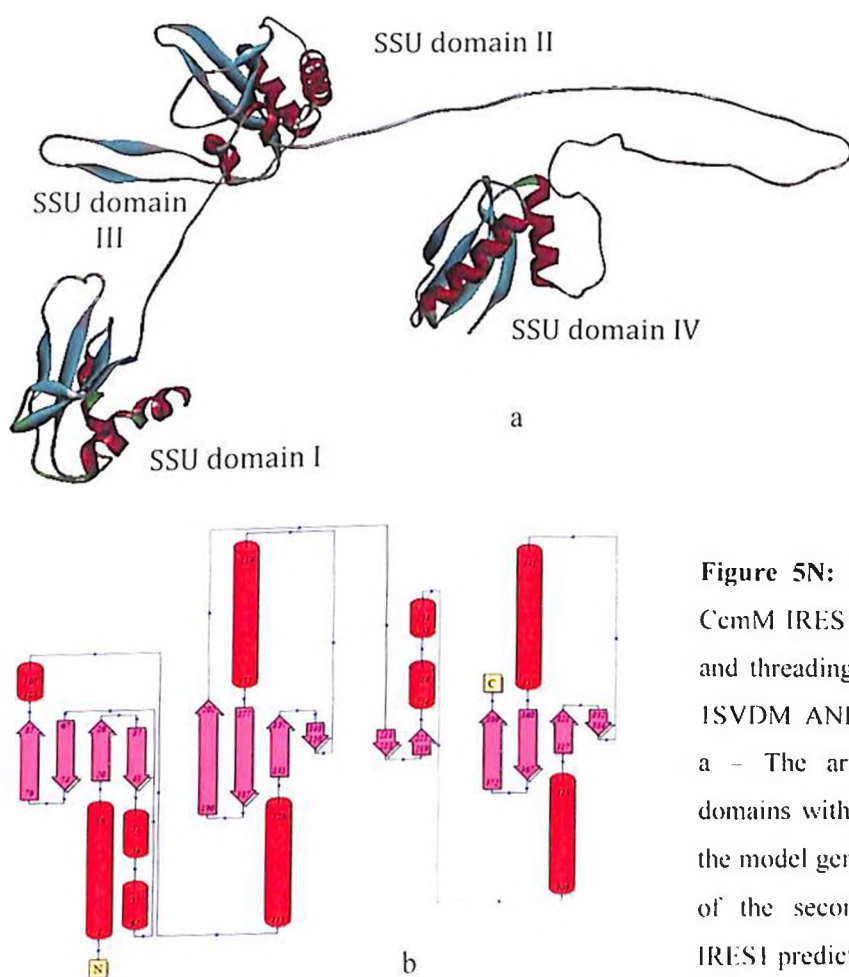
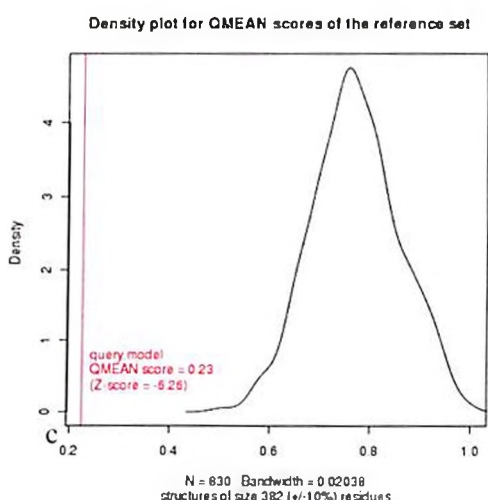
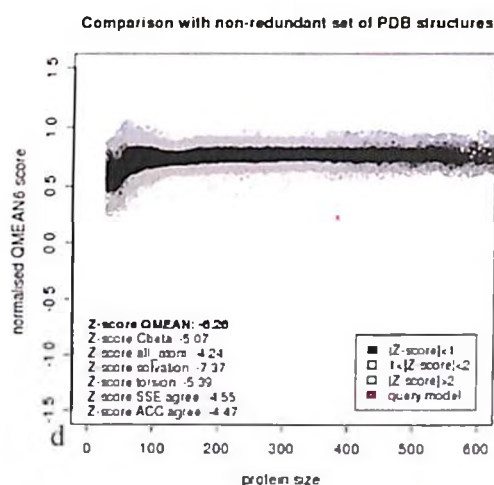
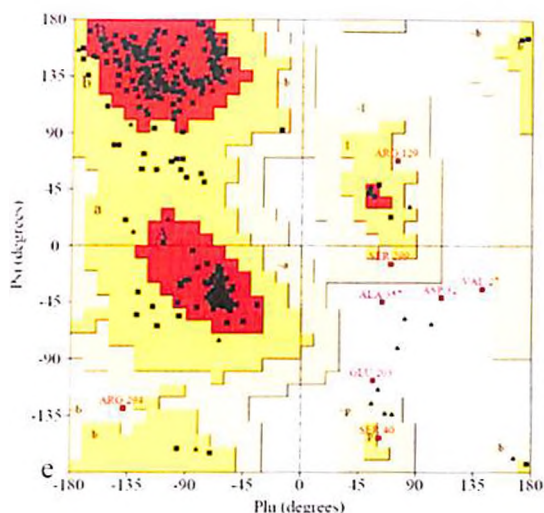


Figure 5N: The tertiary structure of CcmM IRES1 by homology modeling and threading using 1BWVS, 1GK81, 1SVDM AND 3ZXWB as templates. a – The arrangement of the SSU domains with respect to each other in the model generated. b – The topology of the secondary structures of the IRES1 predicted by PDBsum



| | | |
|--|-----|--------|
| Residues in most favoured regions [A, B, I,] | 280 | 88.9% |
| Residues in additional allowed regions [a, b, l, p] | 27 | 8.6% |
| Residues in generously allowed regions [-a, -b, -l, -p] | 4 | 1.3% |
| Residues in disallowed regions | 4 | 1.3% |
| Number of non-glycine and non-proline residues | 315 | 100.0% |
| Number of end residues (excl. Gly and Pro) | 1 | |
| Number of Glycine residues (shown as triangles) | 38 | |
| Number of Proline residues | 28 | |
| Total number of residues | 382 | |
| Based on an analysis of 118 structures of resolution atleast 2.0Å and R factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions | | |

| | |
|---------------|-----------|
| QMEAN Score | 0.227 |
| Z Score | -6.26 |
| Verify 3D | I; 64.83% |
| TM Score | 0.1972 |
| Max-Sub-score | 0.1572 |
| GDT-TS Score | 0.2382 |
| GDT-HA score | 0.1706 |
| RMSD | 14.384Å |

f

g

Figure 5N: c – The Ramachandran plot for IRES1. d and e - QMEAN score of the CcmM IRES1 model in comparison to the standard protein models. f – The Ramachandran plot statistics for CcmM IRES1 model generated. g - Scoring of the CcmM IRES1 model according to QMEAN, Verify3d and TMscore (in red font, obtained by superposition of the CcmM IRES1 model with template i.e. 1BWVS, 1GK8I, 1SVDM AND 3ZXWB) servers.

5.3.3.5. Cloning and *in-vivo* expression of CcmM

The *ccmM* is a 2008bp gene in *G. violaceus*. Even after repeated attempts Taq Polymerase did not yield the amplification of the desired product and hence Phusion Polymerase, a high

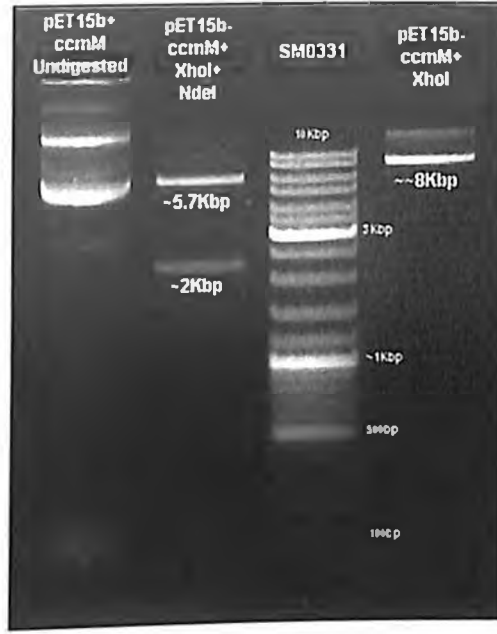


Figure 5P: Agarose gel (0.8%) image for confirmation of *pET15b-ccmM* construct by restriction digestion with enzymes *NdeI* and *XhoI*. The digestion of the construct gave two bands of sizes ~5.7Kbp (size of empty vector *pET15b*) and ~2Kbp (size of *ccmM* gene). Lane 1- Undigested *pET15b-ccmM*; Lane 2- Double digested *pET15b-ccmM* with *NdeI* and *XhoI*; Lane 3- Standard DNA marker SM0331; Lane 4- Single digested *pET15b-ccmM* with *XhoI*.

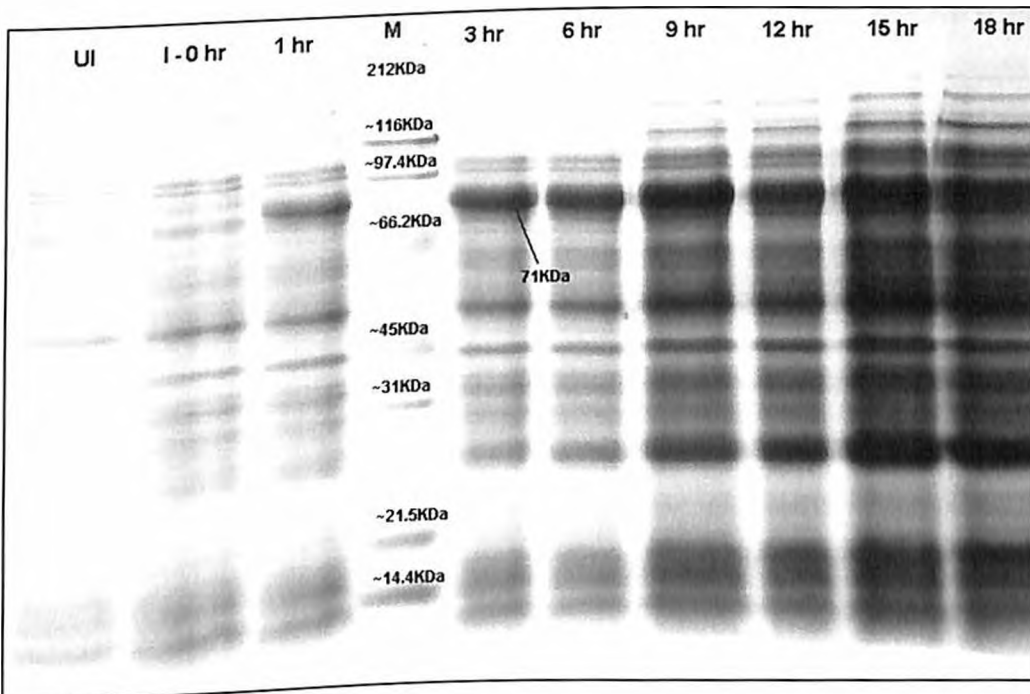


Figure 5Q: SDS-PAGE (12%) image to check the expression of 71kDa *G. violaceus* PCC 7421 *ccmM* (approx 2000bp gene cloned in *pET15b*) in BL21(DE3)pLysS by 1mM IPTG induction. Lane 1- Uninduced culture; Lane 2- Induced culture at 0hrs; Lane 3- Induced culture after 1 hr; Lane 4- protein Marker; Lane 5-10- The crude extract from induced culture after 2, 3, 4, 5, 6, and 7 hrs of induction, respectively.

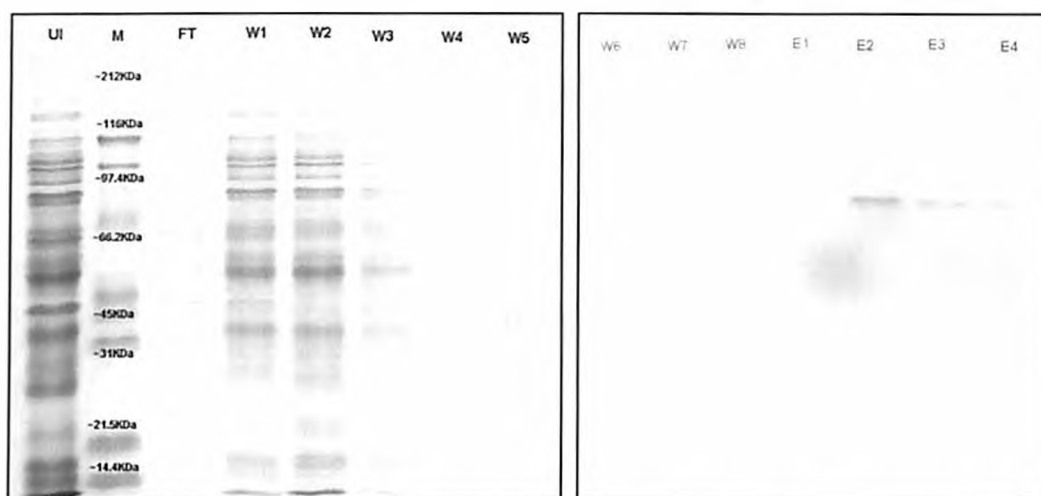


Figure 5R: SDS-PAGE (12%) image for purification of CcmM of *G. violaceus* PCC 7421 by affinity chromatography. Lane 1- Crude extract from induced culture; Lane 2- standard protein marker; lane 3- flow through; Lane 4-8- wash fractions; Lane 1-3 of second gel – wash fractions; Lane 4-7- Elute fractions. (FT: Flow Through, W1-W8: Wash Fractions, E1-E4: Elute Fractions).

5.3.4. CcmN

CcmN is not a very well understood protein. It possess a trimeric LpxA-like domain i.e. a single stranded left handed beta helix fold, constituting tandem repeats of hexapeptide. Similar arrangement is also observed in the N terminal of CcmM. Long et al. (2007) and Cot et al. (2008) have reported interaction of CcmN with RuBisCO and CcmM, respectively. Kinney et al. (2012) reported that CcmN localizes in carboxysomes with N terminal involved in interaction with CcmM-RuBisCO complex and short peptide at C terminal involves in interaction with the shell protein CcmK. In the present analysis, CcmN sequence conserved regions were identified, the CcmN structure from *G. violaceus* was elucidated *in-silico* by homology modeling and the CcmN encoding gene was cloned and analyzed for expression in *E. coli*.

The **Figure 5S** shows the conserved residues of CcmN of β cyanobacteria. The sequence logo shows the comparatively conserved N terminal (bracketed) and a small C terminal region (underlined) connected by not so conserved linker. These are the regions involved in interaction with CcmM-RuBisCO and CcmK, respectively, as already reported. **Figure 5T** highlights these regions in the tertiary structure of the protein from *G. violaceus* generated by homology modeling. The Ramachandran plot for the model shows ~85% residues lie in the most favored region, suggesting a sub-optimal model generated. There is no crystal structure reported till date, hence the modeling was done using other homologous proteins as templates.

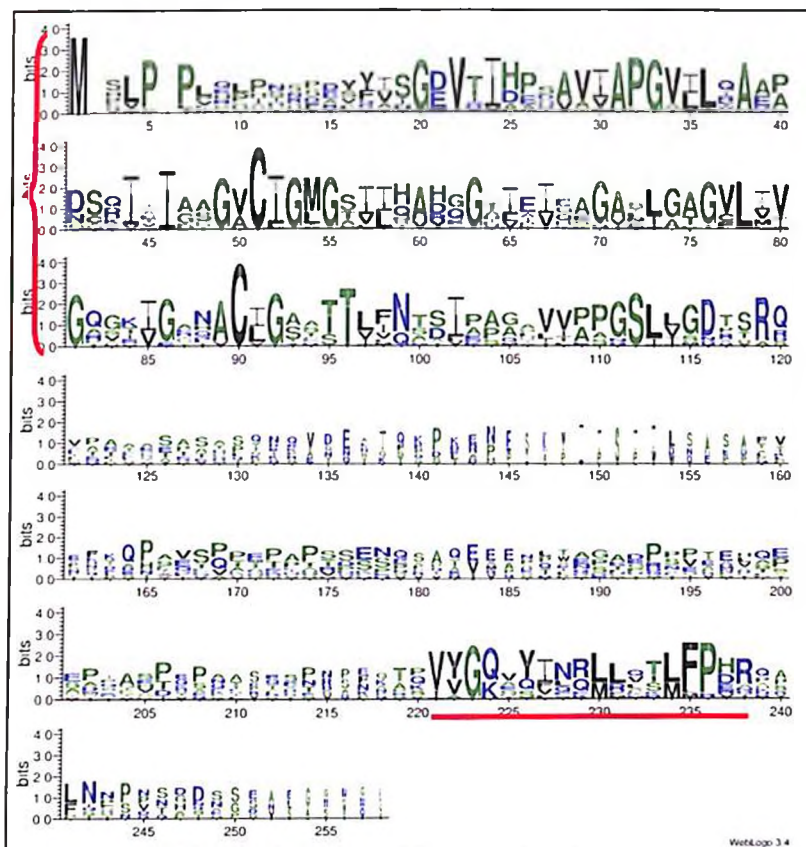


Figure 5S: Sequence logo for the multiple sequence alignment of the CcmN from β cyanobacteria, obtained by alignment in MEGA6.0 using Clustal W followed by input of alignment in FASTA format to the WebLogo3 tool. The red bracketed/ underlined regions are the most conserved regions of the protein.

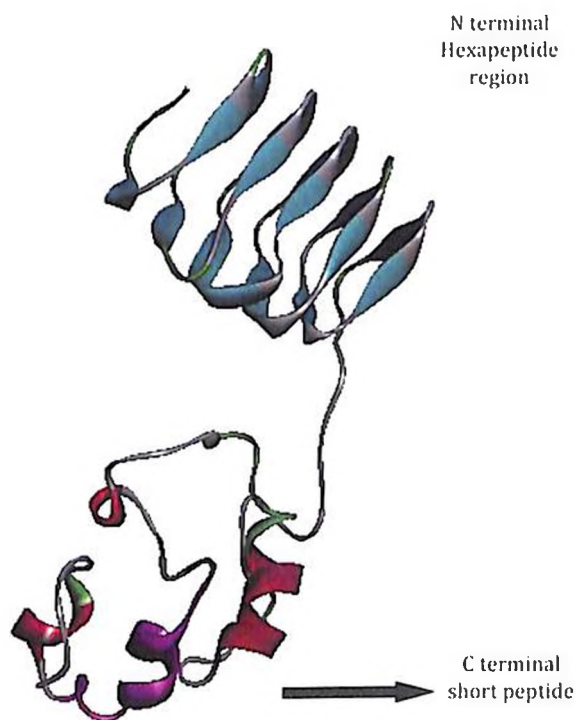
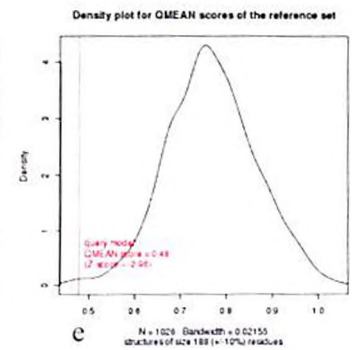
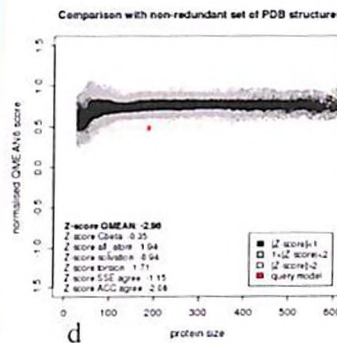
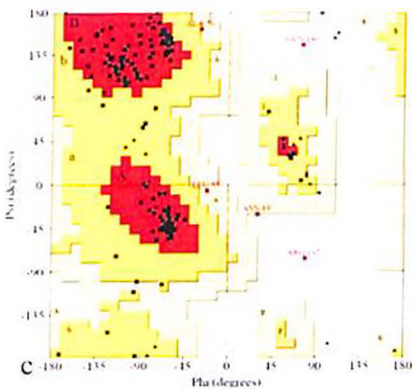
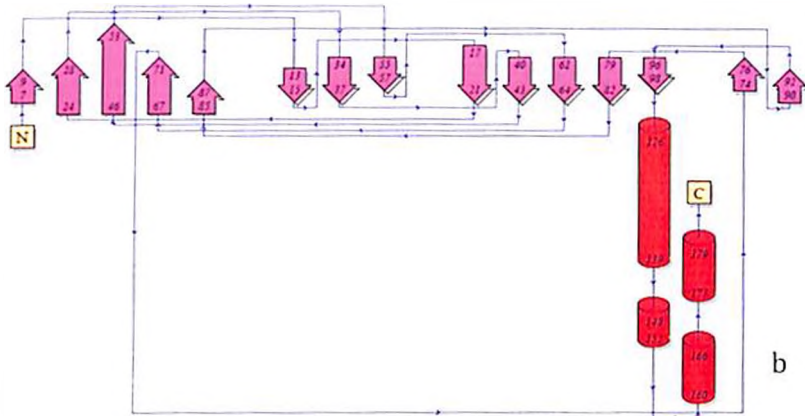


Figure 5T: The tertiary structure of CcmN by homology modeling. a – The arrangement of the conserved domains with respect to each other.



| | | |
|--|-----|--------|
| Residues in most favoured regions [A, B, I,] | 118 | 84.9% |
| Residues in additional allowed regions [a, b, l, p] | 16 | 11.5% |
| Residues in generously allowed regions [-a, -b, -l, -p] | 2 | 1.2% |
| Residues in disallowed regions | 3 | 2.2% |
| Number of non-glycine and non-proline residues | 139 | 100.0% |
| Number of end residues (excl. Gly and Pro) | 1 | |
| Number of Glycine residues (shown as triangles) | 26 | |
| Number of Proline residues | 22 | |
| Total number of residues | 188 | |
| Based on an analysis of 118 structures of resolution atleast 2.0Å and R factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions | | |

| | |
|---------------|-----------|
| QMEAN Score | 0.48 |
| Z Score | -2.96 |
| Verify 3D | W: 77.66% |
| TM Score | 0.2192 |
| Max-Sub-score | 0.0810 |
| GDT-TS Score | 0.1436 |
| GDT-HA score | 0.0718 |
| RMSD | 16.451Å |

f

Figure 5T: The tertiary structure of CcmN by homology modeling. b – The topology of the secondary structures of the CcmN predicted by PDBsum. c – The Ramachandran plot for CcmN model. d and e – QMEAN score of the CcmN model in comparison to the standard protein models. f – The Ramachandran plot statistics for CcmN model generated. g - Scoring of the CcmN model according to QMEAN, Verify3d and TMscore (in red font, obtained by superposition of the CcmN model with template i.e. 3KWCA and 2J00A) servers.

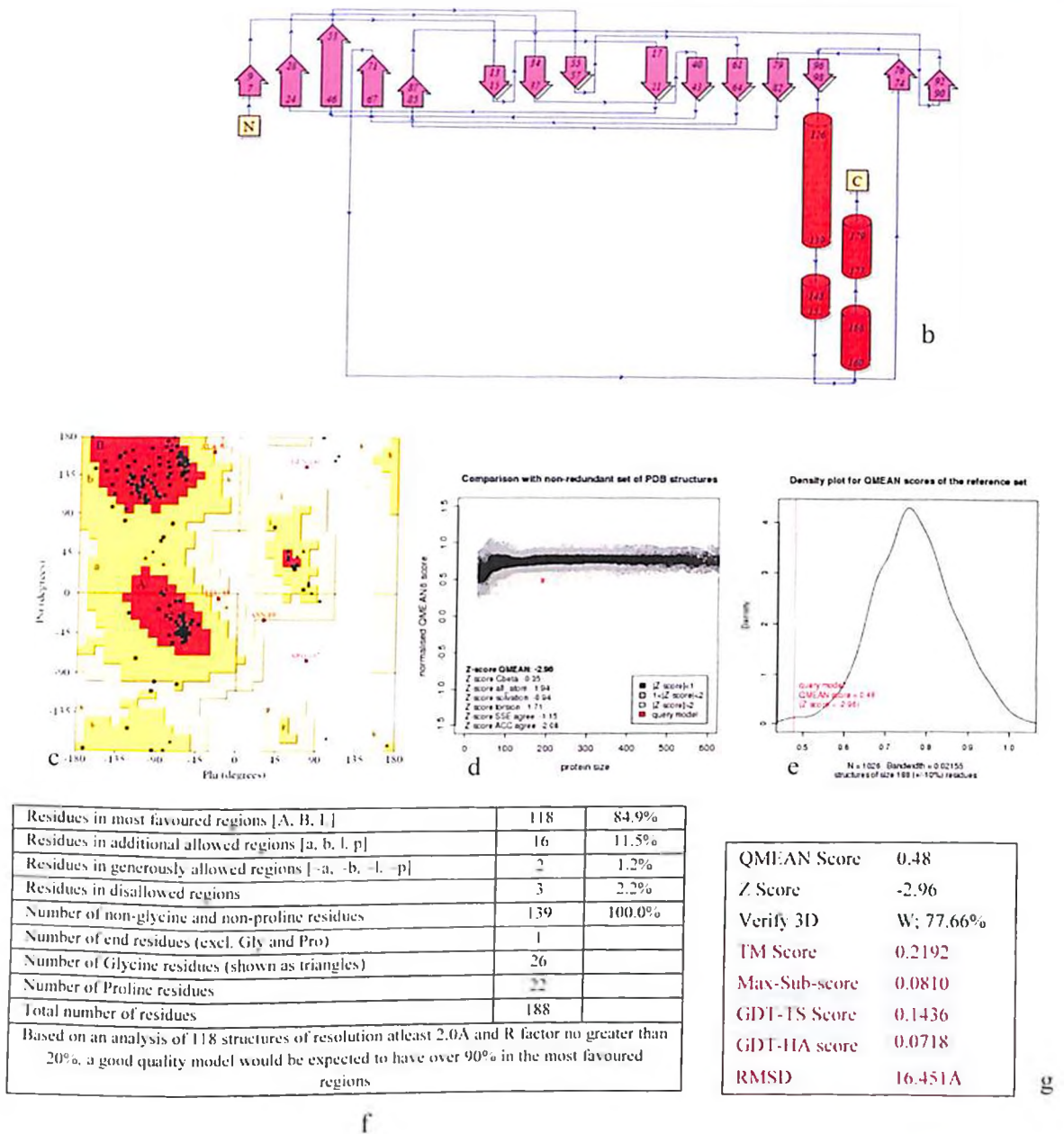


Figure 5F: The tertiary structure of CcmN by homology modeling. **b** – The topology of the secondary structures of the CcmN predicted by PDBsum. **c** – The Ramachandran plot for CcmN model. **d** and **e** – QMEAN score of the CcmN model in comparison to the standard protein models. **f** – The Ramachandran plot statistics for CcmN model generated. **g** - Scoring of the CcmN model according to QMEAN, Verify3d and TMscore (in red font, obtained by superposition of the CcmN model with template i.e. 3KWCA and 2J00A) servers.

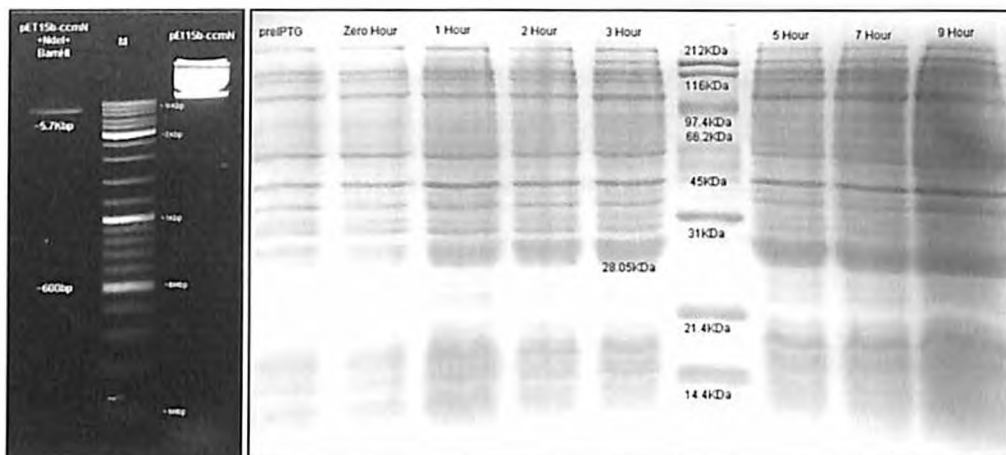


Figure 5U: Left - Agarose gel (0.8%) for the confirmation of *pET15b-ccmN* construct by restriction with enzymes *NdeI* and *BamHI*. The digestion of the construct gave two bands of sizes ~5.7Kbp (size of empty vector *pET15b*) and ~600bp (size of *ccmN* gene). Right - Gel image for SDS-PAGE (12%) to check the expression of ~28kDa *G. violaceus* PCC 7421 CcmN (approx 800bp gene cloned in *pET15b*) in BL21(DE3)pLysS by 1mM IPTG induction.

The *ccmN* from *G. violaceus* genome was amplified using mixture and conditions mentioned in chapter 2 (Table 2E and 2F). The gel image of the product hence obtained is shown in figure 51VA. The amplified product was cloned into *pET15b* vector between restriction sites *NdeI* and *BamHI*. The recombinant plasmid was confirmed to contain *ccmN* by restriction analysis (Figure 5U-left) and DNA sequencing. The confirmed clone was then transformed into *E. coli* BL21(DE3)pLysS and expression of the protein was analyzed in presence of IPTG @ 1mM. The expected band of ~29kDa was observed on the SDS-PAGE upon IPTG induction (Figure 5U right).

5.3.5. CcmO

The CcmO protein constitutes two BMC domains and forms a part of the flat facets of the carboxysomes. The residues inherent to BMC domain were found to be conserved in the β cyanobacterial CcmO protein sequences analyzed by multiple sequence alignment performed using T Coffee server. The Figure 5V shows the sequence logo generated by WebLogo3 from the multiple sequence alignment file provided to the server.

The gene encoding *G. violaceus* CcmO, viz. *ccmO* was amplified by PCR from the genome of *G. violaceus* using conditions and reaction mixture mentioned in chapter 2 (Table 2D, 2E and 2F). The Figure 5D(i) shows the image of the agarose gel for *ccmO* PCR product (~700bp). The PCR product hence obtained was purified by gel elution and used for ligation into *pET15b* vector after digestion with *NdeI* and *BamHI* restriction enzymes. The Figure 5W (left) shows the agarose gel for the recombinant vector *pET15b-ccmO*, undigested, single digested with *NdeI/bamHI* (yielding product of size ~6.4Kbp) and double digested with *NdeI*

and *Bam*HI (yielding two bands of sizes ~5.7Kbp and ~700bp, for *pET15b* and *ccmO*, respectively). The sequence of the recombinant vector was confirmed by sequencing. The confirmed *pET15b-ccmO* was then transformed into *E. coli* BL21(DE3)pLys host to analyze the expression of the protein. The transformed culture upon induction in the presence of IPTG at the concentration of 1mM, expressed the expected protein of ~27kDa after 2 hrs of induction (Figure 5W right).



Figure 5V: Sequence logo for the multiple sequence alignment of the CcmO from β cyanobacteria, obtained by alignment in MEGA6.0 using Clustal W followed by input of alignment in FASTA format to the WebLogo3 tool. The red bracketed/underlined regions are the most conserved regions of the protein.

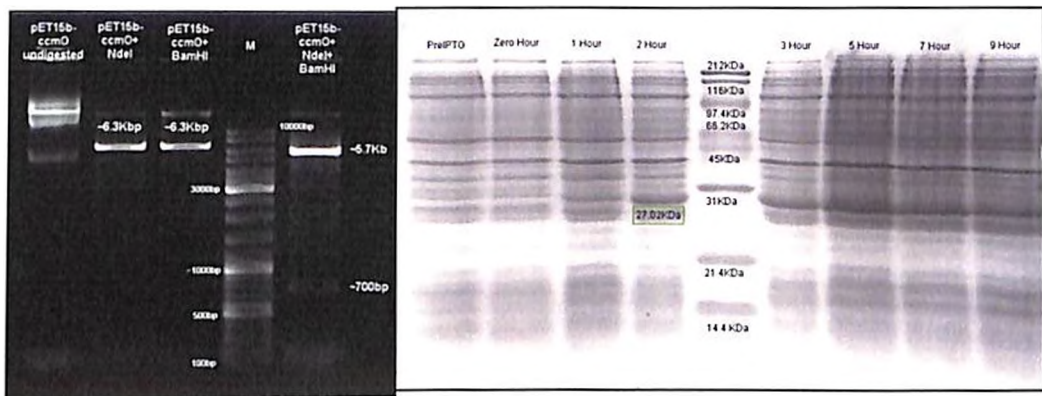
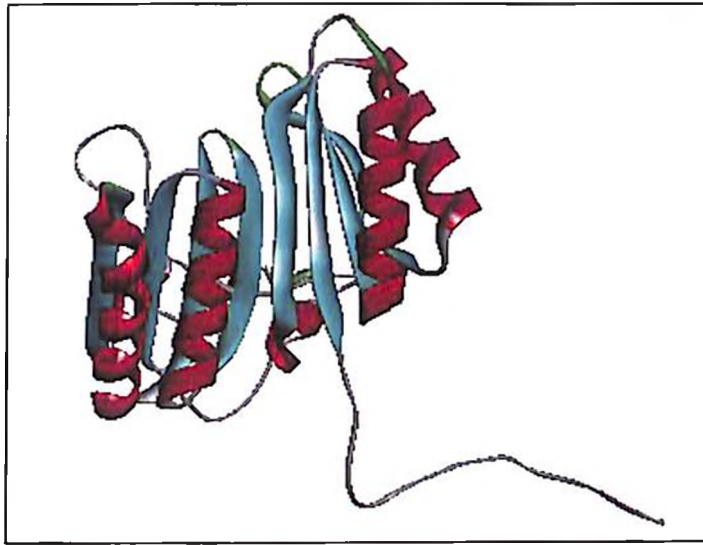


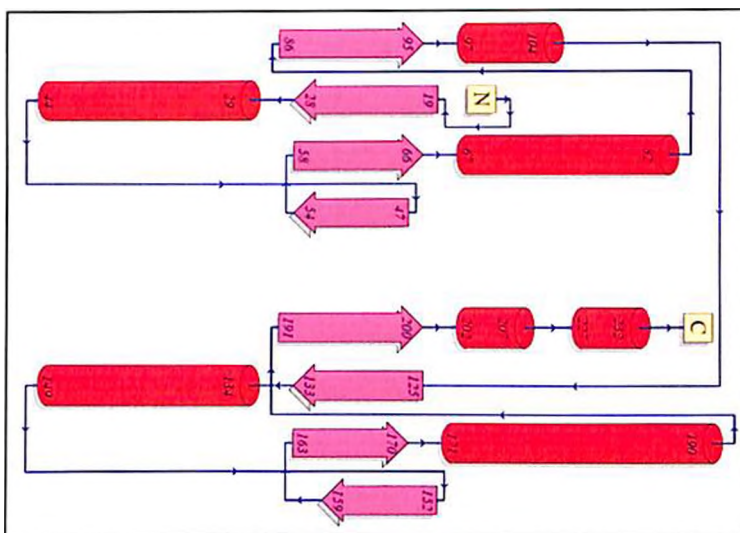
Figure 5W: Left - Agarose gel (0.8%) for the confirmation of *pET15b-ccmO* construct by restriction with enzymes *Nde*I and *Bam*HI. The digestion of the construct gave two bands of sizes ~5.7kbp (size of empty vector *pET15b*) and ~700bp (size of *ccmO* gene). Right - Gel image for SDS-PAGE (12%) to check the expression of ~28kDa *G. violaceus* PCC 7421 CcmO (approx 800bp gene cloned in *pET15b*) in BL21(DE3)pLysS by 1mM IPTG induction.

The CcmO protein of *G. violaceus* was modeled *in-silico* to elucidate the tertiary structure of the protein, by homology modeling. The Figure 5X shows the tertiary structure for CcmO, topology of the protein (as predicted by PDBsum) and the Ramachandran plot for the model

generated (for validation of the model) showing ~90% residues occupying the most favored regions of the plot.

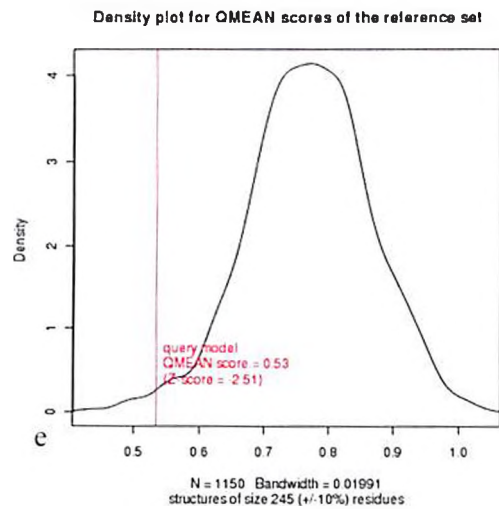
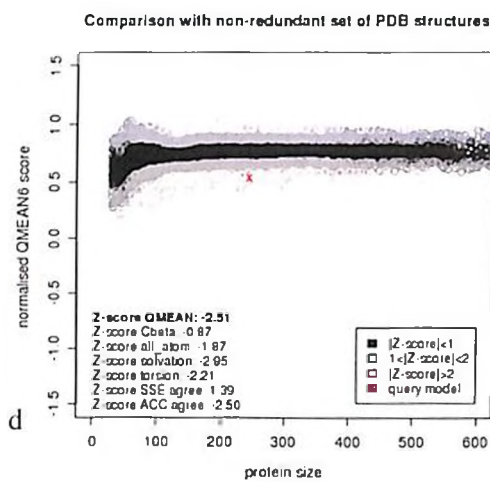
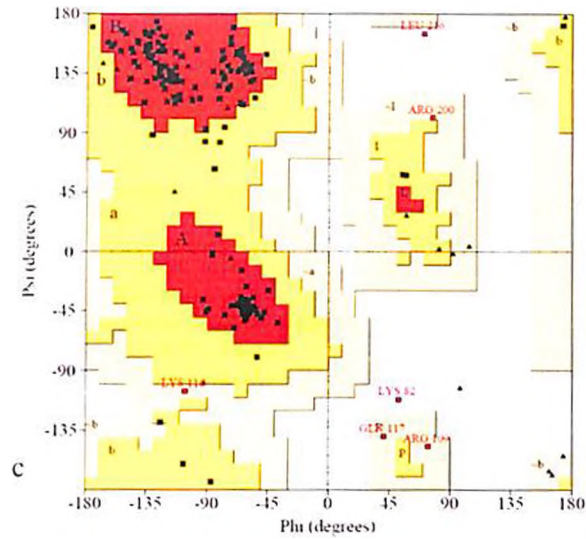


a



b

Figure 5X: The tertiary structure of CcmO by homology modeling. a – The two views showing the repeated domains of CcmO. b – The topology of the secondary structures of the CcmO predicted by PDBsum.



| | | |
|--|-----|--------|
| Residues in most favoured regions [A, B, I] | 185 | 90.7% |
| Residues in additional allowed regions [a, b, l, p] | 13 | 6.4% |
| Residues in generously allowed regions [-a, -b, -l, -p] | 4 | 2.0% |
| Residues in disallowed regions | 2 | 1.0% |
| Number of non-glycine and non-proline residues | 204 | 100.0% |
| Number of end residues (excl. Gly and Pro) | 2 | |
| Number of Glycine residues (shown as triangles) | 22 | |
| Number of Proline residues | 17 | |
| Total number of residues | 245 | |
| Based on an analysis of 118 structures of resolution atleast 2.0Å and R factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions | | |

| | |
|---------------|-----------|
| QMEAN Score | 0.534 |
| Z Score | -2.51 |
| Verify 3D | E: 64.49% |
| TM Score | 0.1538 |
| Max-Sub-score | 0.0420 |
| GDT-TS Score | 0.0776 |
| GDT-HA score | 0.0418 |
| RMSD | 17.529Å |

f

Figure 5X: c – The Ramachandran plot for CcmO model. d and e – QMEAN score of the CcmN model in comparison to the standard protein models. f – The Ramachandran plot statistics for CcmO model generated. g - Scoring of the CcmO model according to QMEAN, Verify3d and TMscore (in red font, obtained by superposition of the CcmN model with template i.e.3PAC) servers.

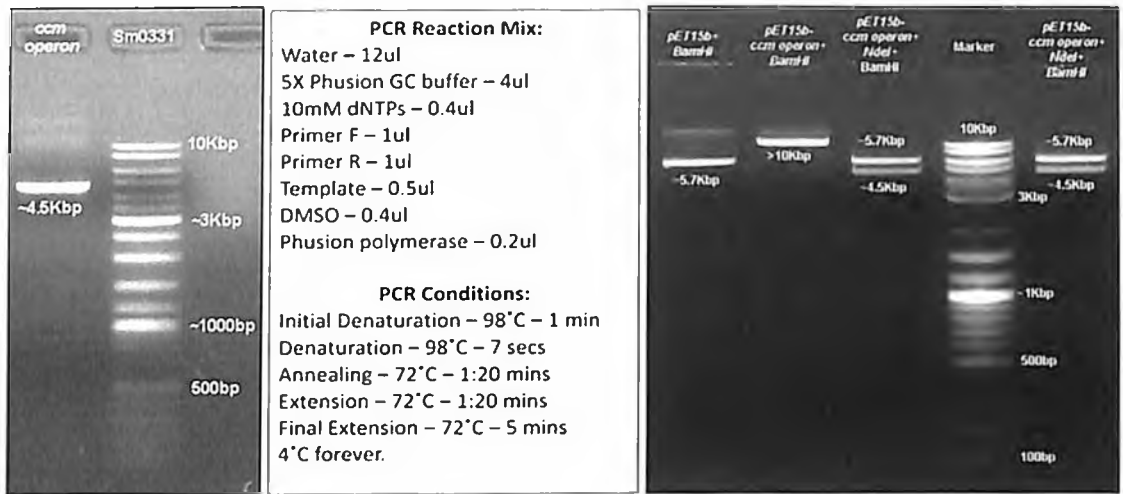


Figure 5Y: Left - Agarose gel (0.8%) for the amplification of the carboxysome operon of *G. violaceus* PCC 7421 (~4.5Kbp) along with the conditions used for PCR by Phusion Polymerase. Right- The image for Agarose gel (0.8%) for the confirmation of *pET15b-ccm operon* construct by restriction with enzymes *NdeI* and *BamHI*. The digestion of the construct gave two bands of sizes ~5.7Kbp (size of empty vector *pET15b*) and ~4.5Kbp (size of *ccm operon*).

5.3.6. Amplification and cloning of the *ccm operon* of *G. violaceus*

The *ccm operon* of *G. violaceus* PCC 7421 was also amplified and cloned into *pET15b*. The *ccm operon* is a ~4.5Kbp region, which was difficult to amplify using Taq polymerase enzyme. The Phusion polymerase designed specifically for long amplifications and providing an added advantage of high fidelity and speed was used for the amplification of this region. The primers designed for this amplification are mentioned in chapter 2 (table 2D). The **Figure 5Y** (left) shows the amplification product and the conditions used for the same. The PCR product obtained was purified by gel elution and used for ligation into *pET15b* vector between the restriction sites *NdeI* and *BamHI* using standard cloning procedures. The **Figure 5Y** (right) shows the image for restriction analysis of *pET15b-ccm operon* recombinant vector. Since the standard DNA marker used is 100bp-10Kbp and the recombinant vector's expected size is ~10.2Kbp (greater than the largest band of the marker), empty vector (*pET15b* digested with *BamHI*) was used as control for single digestion of the recombinant vector with *BamHI*. The lag in the migration of the recombinant vector shows the presence of insert, the size of which was confirmed by double digestion of the recombinant vector with *NdeI* and *BamHI*, yielding two bands of 5.7Kbp (*pET15b*, which migrated to the length same as that for the single digested empty vector) and 4.5Kbp (size of the *ccm operon*). Only initial 700bp and the 3700bp region of the clone prepared was confirmed by sequencing as the complete sequence confirmation requires primer walking.

5.4. Conclusions

The characterization of the carboxysome proteins forms an integral part of the study to achieve the farfetched goal of augmenting C_3 plant productivity. Although many aspects of these proteins have been unraveled but still several characters remain unknown.

The present study shows the presence of several conserved regions of the proteins CcmK and CcmL, which have not been associated with any function. These conserved regions could be investigated for their role in carboxysome biogenesis. Further, the importance of pore residues could be analyzed for feasibility of creating entry points for modified set of metabolites. The carboxysome proteins viz. CcmK, CcmL, CcmM, CcmN and CcmO, when expressed under control of IPTG inducible promoter individually, were found to express proteins of expected sizes. While the *ccm operon* when expressed as a whole, did not yield all the proteins of encoded by the operon.

The CcmM IRES of *G. violaceus* was predicted and the tertiary structure of the protein hence expected was estimated by modeling. The structure obtained is in accordance with proposed set of interactions between the RuBisCO and the CcmM (short form) proteins by Long et al (2010). Further, the conservation of the residues crucial for stable carbonic anhydrase activity in the CcmM protein, signals the possibility of the protein to be capable of causing interconversion of HCO_3^- and CO_2 . The CcmN protein on the other hand, although having the carbonic anhydrase domain was not found to have these conserved residues.

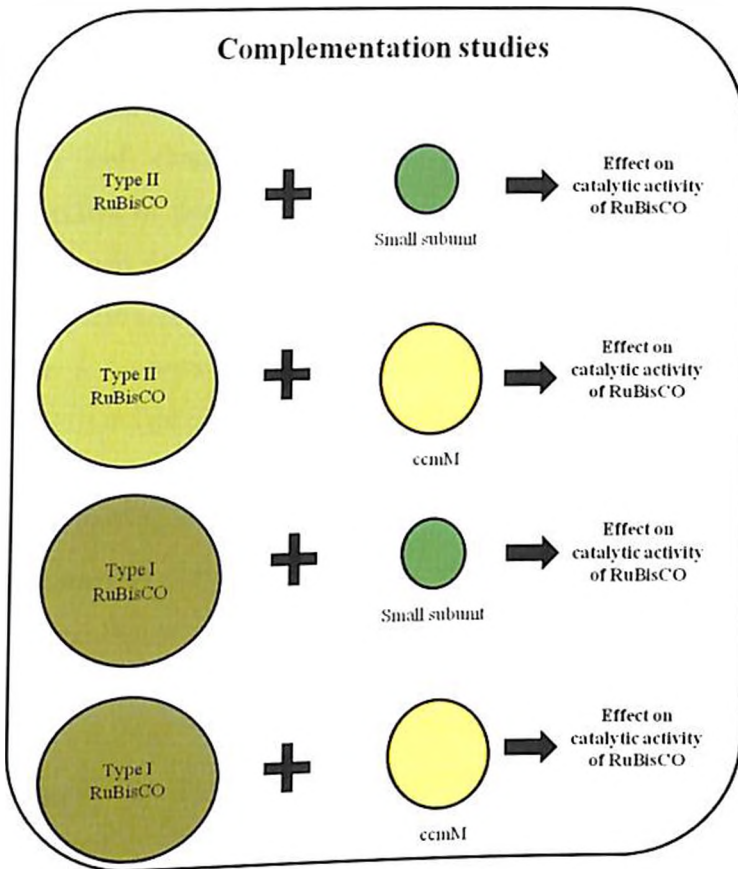
The CcmM protein C terminal SSU repeats were found to be more closely related to RuBisCO SSU of *Leptolyngbya* than to the RuBisCO SSU of their own, suggesting a common point of origin of the CcmM protein. The lack in concordance in evolution of C and N terminals of the protein direct towards the appearance of the duplicated repeats at the C terminal post the event of speciation.

The present study opens quests for further analysis of the importance of the CcmK and CcmL proteins in carboxysome functioning and biogenesis. Further, the expression of the complete operon of the carboxysome could yield valuable information for the assembly of the microcompartment.

CHAPTER 6

**The Effect of CcmM and RbcS of
Gloeobacter violaceus on Catalytic Activity
of a Form I (*Gloeobacter violaceus*) and
Form II (*Rhodospirillum rubrum*) RuBisCO
Large Subunit**

6.1. Introduction



The analysis was done using IMAC purified proteins and *in-silico* generated models and interactions.

Type II RuBisCO -
Rhodospirillum rubrum
RuBisCO

RuBisCO SSU -
Gloeobacter violaceus PCC
7421 RuBisCO SSU

CcmM -
Gloeobacter violaceus PCC
7421 CcmM

Type I RuBisCO -
Gloeobacter violaceus PCC
7421 RuBisCO

6.2. Methodology

The proteins purified viz. *Rhodospirillum rubrum* RuBisCO, *Gloeobacter violaceus* RuBisCO large subunit, *G. violaceus* RuBisCO small subunit, *G. violaceus* RbcX and *G. violaceus* CcmM by IMAC, subsequently desalted and stored in RuBisCO activation buffer were quantified (protocols mentioned in chapter 2). The *R. rubrum* RuBisCO was purified in native form and hence as a holoenzyme constituted by large subunit only, was used directly for the assay. *G. violaceus* RuBisCO (form I) LSU was reconstituted in the presence of RbcX. The RbcL, in RuBisCO activation buffer was incubated at 25°C for 60 minutes in presence of RbcX and/or RbcS/CcmM. The reaction was initiated by addition of ATP (4mM) and NaHCO₃ (25mM). The optimal concentrations of the proteins required for reconstitution were determined by using different ratios and then selecting the one giving the best activity. The catalytic activity of the resulting protein after reconstitution was determined by the standard RuBisCO assay. The speculated interactions between the proteins under study were also analyzed *in-silico* by performing protein-protein docking using GRAMMX (Tovchigrechko and Vakser 2006) tool. The interactions estimated were visualized by the Discovery studio software.

6.3. Results and Discussions

6.3.1. *Rhodospirillum rubrum* RuBisCO

R. rubrum is a proteobacteria and possess form II RuBisCO. The RuBisCO of *R. rubrum* is very well characterized and the crystal structure for the same is also determined. The objective of purification of *R. rubrum* RuBisCO and analysis of its kinetics in the present study was firstly, to test the accuracy of the purification procedures as well as the RuBisCO assay and secondly, to determine the effect foreign RuBisCO SSU on the activity of the same. The *R. rubrum* RuBisCO (*pET15b-R. rubrum* RuBisCO), readily expressed in *E. coli* BL21(DE3)pLysS by IPTG induction @ 0.5mM after 1 hr, upto 10hrs (**Figure 6B**). The 2nd culture of *E. coli* BL21(DE3)pLysS transformed with *pET15b-R. rubrum* RuBisCO, after induction @ 0.5mM and incubation at 18°C, 150rpm for 18 hrs was pelleted down. The protein was purified under native conditions according to protocol mentioned in chapter 2. The protein was purified to homogeneity by IMAC (**Figure 6C**), desalted by PD10 column into thrombin buffer. The protein purified was gel eluted by tryptic digestion and analyzed by MALDI-TOF. **Figure 6D** shows the MASCOT results of the protein, with score of 75 which suggests that the probability of detecting the sequence of *R. rubrum* RuBisCO in the sample analyzed by chance is less than 5%. After cleavage of the his-tag, protein was exchanged into RuBisCO activation buffer and stored at -80°C until further used.

RuBisCO activity was analyzed at different enzyme concentrations of the protein as well as in the presence of different concentrations of the substrate RuBP. The **Figure 6E i** and **ii** show the trend of decrease in OD₃₄₀ with an increase in concentration of the protein and substrate, respectively. The decrease in optical density after the initial 10-30 seconds i.e. after the addition of substrate RuBP is found to be maximum upto 40µg (working concentration) of the protein in the reaction mixture. The maximum of 1.18 change in optical density at 340nm per minute was observed. **Figure 6F(i)** shows the activity of *R. rubrum* RuBisCO at different working concentrations, which is found to increase upto 40µg, beyond which all the active sites of the protein appear to have saturated in the presence of the provided reaction conditions or the reaction intermediates are limited to cause any further increase in activity. Further, **Figure 6F(ii)** shows the effect of RuBP concentration on the activity of the protein at 40µg working concentration.

3. Results and Discussions

3.1. *Rhodospirillum rubrum* RuBisCO

R. rubrum is a proteobacteria and possess form II RuBisCO. The RuBisCO of *R. rubrum* is very well characterized and the crystal structure for the same is also determined. The objective of purification of *R. rubrum* RuBisCO and analysis of its kinetics in the present study was firstly, to test the accuracy of the purification procedures as well as the RuBisCO assay and secondly, to determine the effect foreign RuBisCO SSU on the activity of the same. The *R. rubrum* RuBisCO (*pET15b-R. rubrum* RuBisCO), readily expressed in *E. coli* BL21(DE3)pLysS by IPTG induction @ 0.5mM after 1 hr, upto 10hrs (**Figure 6B**). The 2L culture of *E. coli* BL21(DE3)pLysS transformed with *pET15b-R. rubrum* RuBisCO, after induction @ 0.5mM and incubation at 18°C, 150rpm for 18 hrs was pelleted down. The protein was purified under native conditions according to protocol mentioned in chapter 2. The protein was purified to homogeneity by IMAC (**Figure 6C**), desalted by PD10 column into thrombin buffer. The protein purified was gel eluted by tryptic digestion and analyzed by MALDI-TOF. **Figure 6D** shows the MASCOT results of the protein, with score of 75 which suggests that the probability of detecting the sequence of *R. rubrum* RuBisCO in the sample analyzed by chance is less than 5%. After cleavage of the his-tag, protein was exchanged into RuBisCO activation buffer and stored at -80°C until further used.

RuBisCO activity was analyzed at different enzyme concentrations of the protein as well as in the presence of different concentrations of the substrate RuBP. The **Figure 6E i** and **ii** show the trend of decrease in OD₃₄₀ with an increase in concentration of the protein and substrate, respectively. The decrease in optical density after the initial 10-30 seconds i.e. after the addition of substrate RuBP is found to be maximum upto 40µg (working concentration) of the protein in the reaction mixture. The maximum of 1.18 change in optical density at 340nm per minute was observed. **Figure 6F(i)** shows the activity of *R. rubrum* RuBisCO at different working concentrations, which is found to increase upto 40µg, beyond which all the active sites of the protein appear to have saturated in the presence of the provided reaction conditions or the reaction intermediates are limited to cause any further increase in activity. Further, **Figure 6F(ii)** shows the effect of RuBP concentration on the activity of the protein at 40µg working concentration.

6.3. Results and Discussions

6.3.1. *Rhodospirillum rubrum* RuBisCO

R. rubrum is a proteobacteria and possess form II RuBisCO. The RuBisCO of *R. rubrum* is very well characterized and the crystal structure for the same is also determined. The objective of purification of *R. rubrum* RuBisCO and analysis of its kinetics in the present study was firstly, to test the accuracy of the purification procedures as well as the RuBisCO assay and secondly, to determine the effect foreign RuBisCO SSU on the activity of the same. The *R. rubrum* RuBisCO (*pET15b-R. rubrum* RuBisCO), readily expressed in *E. coli* BL21(DE3)pLysS by IPTG induction @ 0.5mM after 1 hr, upto 10hrs (**Figure 6B**). The 2L culture of *E. coli* BL21(DE3)pLysS transformed with *pET15b-R. rubrum* RuBisCO, after induction @ 0.5mM and incubation at 18°C, 150rpm for 18 hrs was pelleted down. The protein was purified under native conditions according to protocol mentioned in chapter 2. The protein was purified to homogeneity by IMAC (**Figure 6C**), desalted by PD10 column into thrombin buffer. The protein purified was gel eluted by tryptic digestion and analyzed by MALDI-TOF. **Figure 6D** shows the MASCOT results of the protein, with score of 75 which suggests that the probability of detecting the sequence of *R. rubrum* RuBisCO in the sample analyzed by chance is less than 5%. After cleavage of the his-tag, protein was exchanged into RuBisCO activation buffer and stored at -80°C until further used.

RuBisCO activity was analyzed at different enzyme concentrations of the protein as well as in the presence of different concentrations of the substrate RuBP. The **Figure 6E i and ii** show the trend of decrease in OD₃₄₀ with an increase in concentration of the protein and substrate, respectively. The decrease in optical density after the initial 10-30 seconds i.e. after the addition of substrate RuBP is found to be maximum upto 40µg (working concentration) of the protein in the reaction mixture. The maximum of 1.18 change in optical density at 340nm per minute was observed. **Figure 6F(i)** shows the activity of *R. rubrum* RuBisCO at different working concentrations, which is found to increase upto 40µg, beyond which all the active sites of the protein appear to have saturated in the presence of the provided reaction conditions or the reaction intermediates are limited to cause any further increase in activity. Further, **Figure 6F(ii)** shows the effect of RuBP concentration on the activity of the protein at 40µg working concentration.

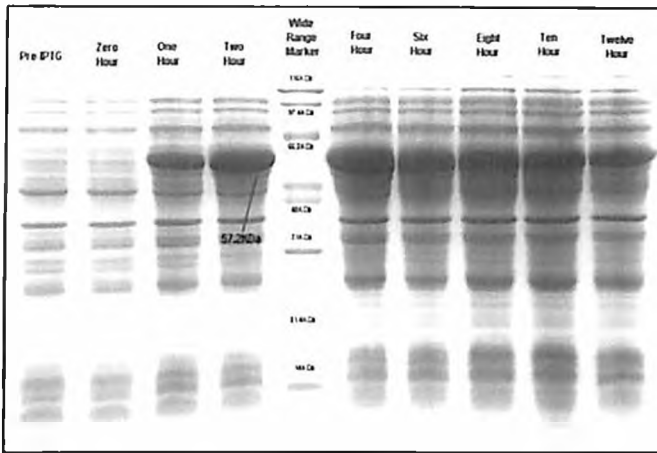


Figure 6B: The image of SDS-PAGE (12%) for analyzing the expression profile of *pET15b-R. rubrum* RuBisCO by IPTG induction at rate of 0.5mM working concentration and incubation at 18°C for 18hrs. Lane 1- Uninduced culture; Lane 2-4- crude extract from culture after 0, 1 and 2 hrs of induction, respectively; Lane 5- standard protein marker; Lane 6-10- crude extract from culture after 4, 6, 8, 10 and 12hrs of induction, respectively.

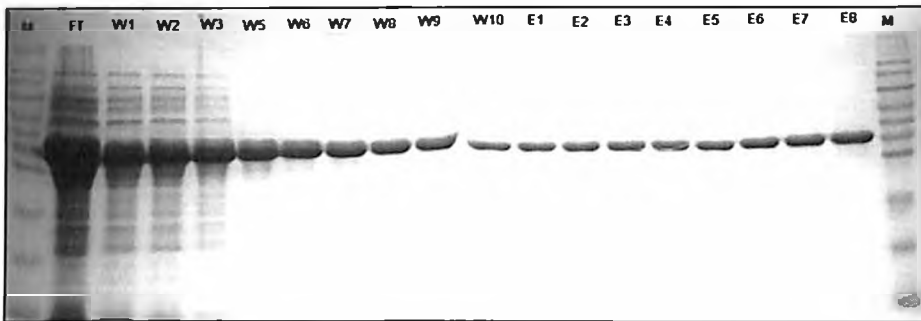


Figure 6C: The SDS-PAGE (12%) for the IMAC purification of *R. rubrum* RuBisCO. Lane 1- Standard protein marker; Lane 2- Flow through; Lane 3-10 and lane 1 of second gel – wash fractions; Lane 2-9 of second gel – elute fractions; lane 10 of second gel – standard protein marker.

The protein was found to show significant activity even in the presence of substrate concentrations as low as 0.005mM. The maximum activity was observed at 2mM working concentration of the substrate. The **Figures 6G** and **6H** show the Lineweaver-Burk plot and specific activity of *R. rubrum* RuBisCO at different substrate concentrations, respectively. The specific activity, K_m RuBP, V_{max} and catalytic turnover rate for *R. rubrum* RuBisCO was found to be 270.003 μ Moles of PGA/ min/mg of protein, 0.0490mM, 0.5737/minute and 10/second, respectively. **Table 6A** shows the kinetic parameters determined for *R. rubrum* RuBisCO by various research groups. The comparison shows that the results obtained in present study are in accordance with those reported earlier. Hence the

purification procedures and the RuBisCO activity estimated can be considered dependable for determining the same for unknown proteins.

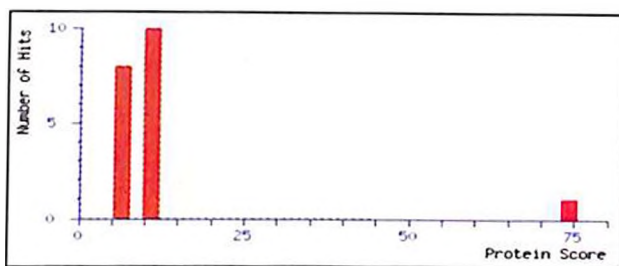


Figure 6D: The mass-spectrometry results analyzed by MASCOT. Mascot uses probability (probability of occurrence of a search hit by chance) based scoring to judge whether a result is significant or not. Area in green represents – 95% of the probability density function. Scores to the right of green region have <5% chance to occur by chance.

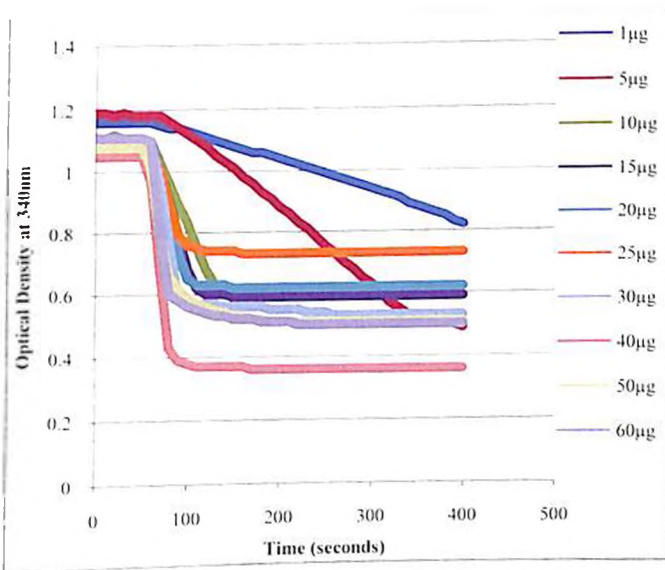
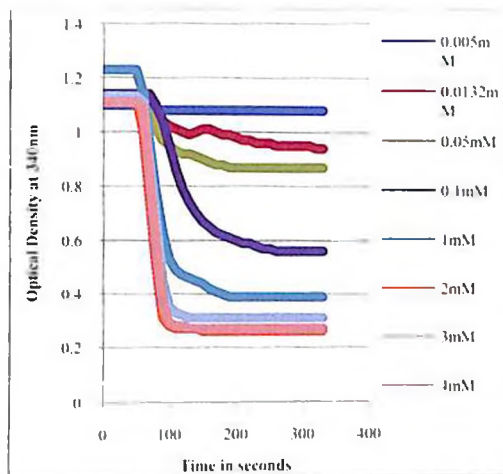
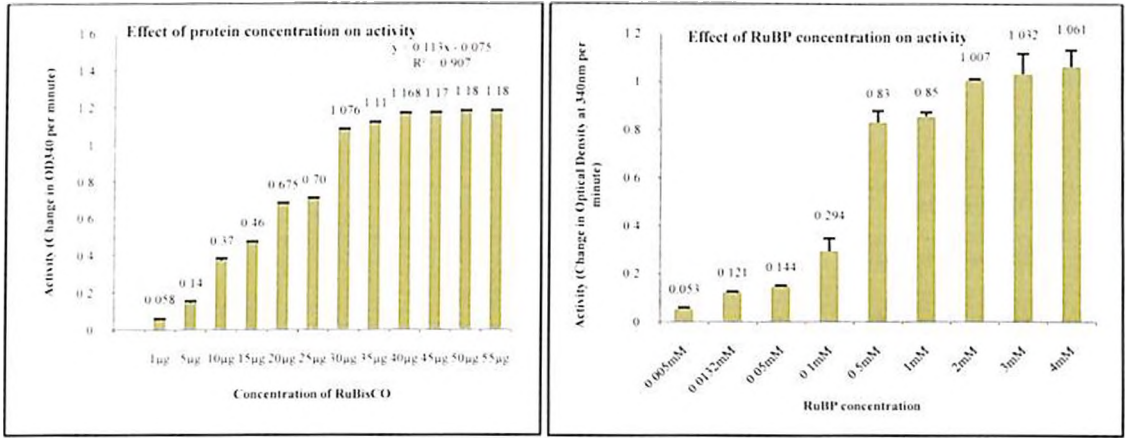


Figure 6E(i): The trend line for decrease in optical density at 340nm of the reaction mixtures containing different concentrations of the *R. rubrum* RuBisCO after addition of the substrate RuBP. The initial 50-60 seconds show no change in OD₃₄₀ , followed by an abrupt decrease as soon as substrate is added. The trendlines of different colours are for different protein concentrations, mentioned in the legend in the graph.

Figure 6E(ii): The trend line for decrease in optical density at 340nm of the reaction mixtures containing different concentrations of RuBP added to the *R. rubrum* RuBisCO reaction mixture. The initial 50-60 seconds show no change in OD₃₄₀ , followed by an abrupt decrease as soon as substrate is added. The trendlines of different colours are for different RuBP concentrations, mentioned in the legend in the graph.





i

ii

Figure 6F i and ii: The graphical representation of the effect of *R. rubrum* RuBisCO concentration and RuBP (substrate) concentration on the activity of *R. rubrum* RuBisCO, respectively. i – The *R. rubrum* RuBisCO concentrations used for the assay are shown on the X-axis of the graph and the change in optical density per minute (activity) are shown on the Y-axis. The RuBP concentration used for concentration dependent assay was 2mM. The assay was performed in triplicates and the error bars show the standard deviation. ii - The concentration of *R. rubrum* RuBisCO used for substrate dependent assay was 40µg. The X-axis show the substrate concentrations used and the Y axis show the activity of the protein. The assay was performed in triplicates and the error bars represent the standard deviation.

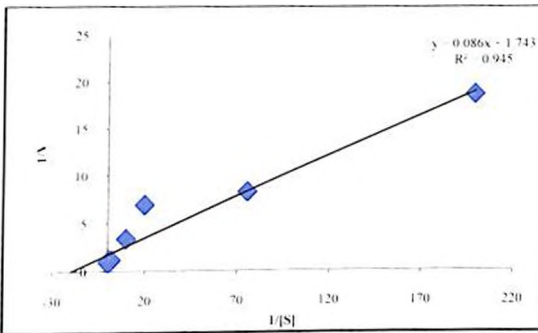


Figure 6G: The Lineweaver-Burk plot for the RuBP dependent assay for *R. rubrum* RuBisCO. The V_{max} and K_m RuBP calculated from the plot are 0.5737 and 0.049mM, respectively.

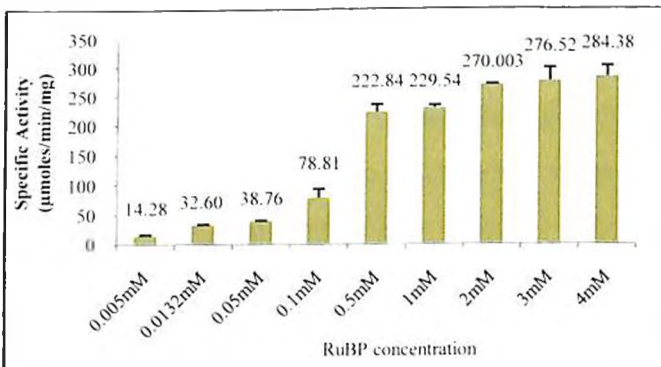


Figure 6H: The graphical representation of the effect of substrate concentration on the specific activity of the *R. rubrum* RuBisCO (40µg). The X-axis show the substrate concentrations used and the Y-axis show the specific activity (µMoles of PGA formed/min/mg of the protein used).

Table 6A: The parameters of *R. rubrum* RuBisCO reported by different research groups using recombinant protein and the spectrophotometric assay similar to the one used in present analysis.

| Parameter | Values | Reference |
|------------------------------|--------|---------------------|
| K_m RuBP (mM) | 0.0128 | Liggin et al 2009 |
| | 0.03 | Nakano et al 2010 |
| | 0.05 | Terzaghi et al 1986 |
| Turnover Number (per second) | 5 | Nakano et al 2010 |
| | 6.28 | Liggin et al 2009 |

6.3.2. The Effect of *G. violaceus* RuBisCO SSU on *R. rubrum* RuBisCO

The *R. rubrum* RuBisCO was allowed to interact with *G. violaceus* RuBisCO SSU, *in-vitro* according to the protocol mentioned earlier. **Table 6B** shows the various concentrations of the protein used in the assay.

Table 6B: The ratios and corresponding concentrations of the *R. rubrum* RuBisCO and *G. violaceus* RuBisCO SSU used for complementation studies.

| <i>R. rubrum</i> RuBisCO (μ g) | <i>G. violaceus</i> RuBisCO SSU (μ g) | Ratio |
|--|---|--------|
| 40 | - | 1:0 |
| 40 | 20 | 1:0.5 |
| 40 | 30 | 1:0.75 |
| 40 | 40 | 1:1 |
| 40 | 60 | 1:1.5 |
| 40 | 80 | 1:2 |
| 40 | 100 | 1:3 |
| - | 40 | 0:1 |

The **Figures 6I** and **6J** show the trendline of the change in optical density after addition of substrate to the different combinations of *R. rubrum* RuBisCO and *G. violaceus* SSU and the effect of same on the specific activity, respectively. The maximum activity of the protein was observed in complete absence of the small subunit. An increase in the concentration of the SSU only caused a decrease in the activity of the protein. The K_M RuBP of the protein was found to increase and the V_{max} decreased for *R. rubrum* RuBisCO in presence of *G. violaceus* RuBisCO SSU.

The in-silico analysis of interactions between R. rubrum RuBisCO and G. violaceus RuBisCO SSU

The pdb. file of *R. rubrum* RuBisCO was retrieved from PDB viz. 9RUB. Further, in order to identify the residues involved in interaction between the large subunit and the small subunit, the protein model for RuBisCO *Thermosynechococcus elongatus* (2YBV) was downloaded from PDB and visualized in Discovery Studio 4.5 software. In the receptor ligand module of the visualize, the intermolecular interactions between the large subunit and the small subunit of the protein were noted for *T. elongatus* (using amino acid chains A, I, B, H, J and P). a total of sixteen residues of large subunit were found to be involved in interaction with their corresponding small subunit in *T. elongatus*. **Table 6C** shows the residues involved in interaction and **Figure 6K** shows the arrangement of the subunits with respect to each other and a snapshot of the interacting residues.

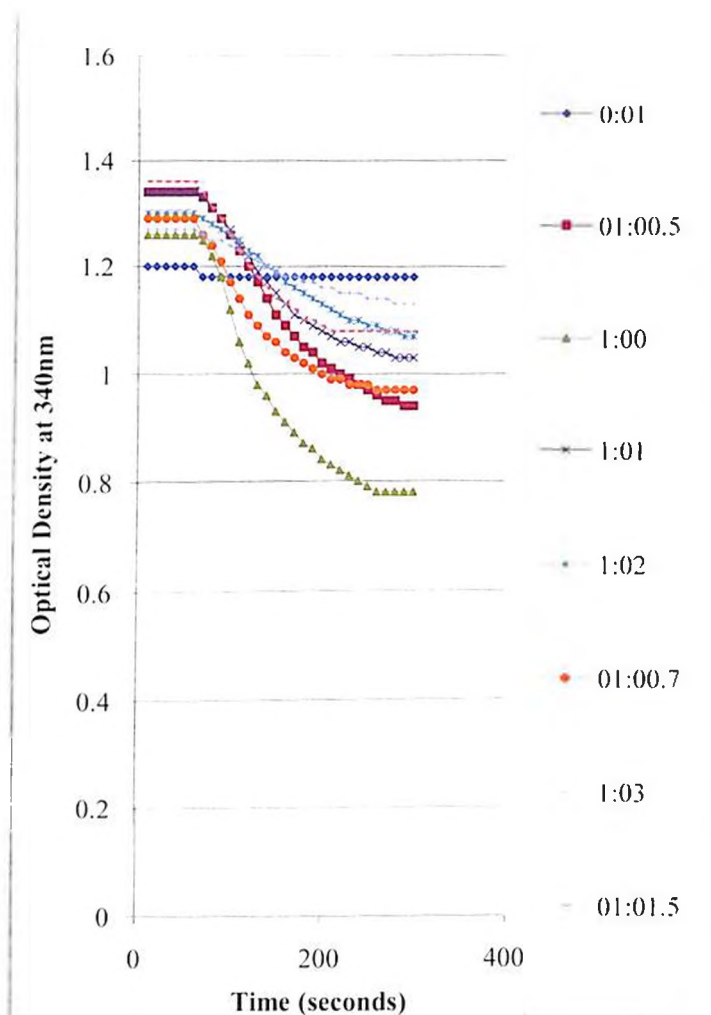


Figure 6I: The trend line for decrease in optical density at 340nm of the reaction mixtures containing *R. rubrum* RuBisCO and different concentrations of *G. violaceus* RuBisCO SSU after addition of the substrate RuBP. The initial 50-60 seconds show no change in OD₃₄₀, followed by an abrupt decrease as soon as substrate is added. The trendlines of different colours are for different ratios of the two proteins (*R. rubrum* RuBisCO: *G. violaceus* SSU), mentioned in the legend in the graph.

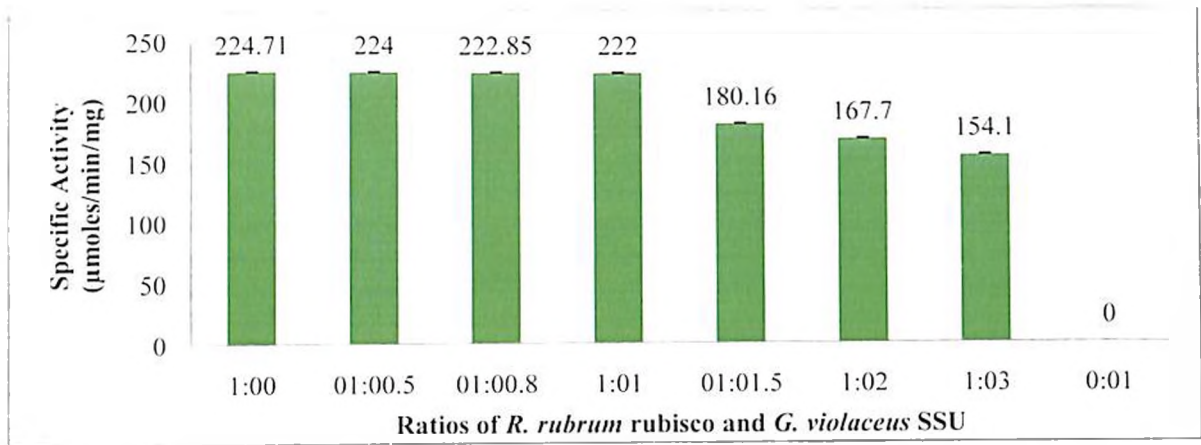


Figure 6J: The effect of different concentrations of *G. violaceus* RuBisCO SSU on specific activity of *R. rubrum* RuBisCO. The concentration of *R. rubrum* RuBisCO used was 40μg and for *G. violaceus* RuBisCO SSU 0, 20, 30, 40, 60, 80, 100 and 40μg from left to right on the X axis of the graph.

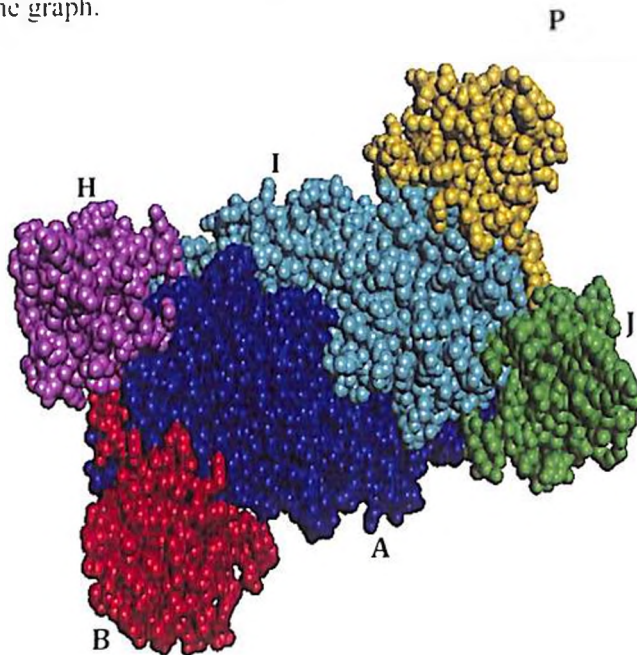


Figure 6K: The arrangement of the RuBisCO subunits with respect to each other in *T. elongatus* RuBisCO. The large subunits-A and I; Small subunits – P, B, H and J. the interactions between these subunits are mentioned in Table 6C.

Table 6C: The residues involved in interaction between LSU (A and I- Row 1 and 2) and SSU (B/H/J/P – rows 3-6) in *T. elongatus* RuBisCO (2YBV) retrieved from PDB

| A | I | B | H | J | P |
|--------|--------|-------|-------|-------|-------|
| Lys164 | | Glu11 | | | |
| Arg167 | | Glu11 | | | |
| Arg421 | | Glu11 | | | |
| Glu231 | | Lys6 | | | |
| Glu433 | | Arg26 | | | |
| | Lys164 | | | | Glu11 |
| | Arg167 | | | | Glu11 |
| | Arg421 | | | | Glu11 |
| | Glu231 | | | | Lys6 |
| Asp76 | | | | Asn92 | |
| Arg79 | | | | Ile93 | |
| Ser181 | | | Gln95 | | |
| Asn184 | | | Gln95 | | |
| Tyr190 | | | Thr54 | | |
| Arg194 | | Leu4 | | | |
| Lys227 | | | Asn43 | | |
| Lys227 | | | Tyr52 | | |
| Gln229 | | Glu49 | | | |
| Asn432 | | Gln27 | | | |
| Ala230 | | Arg8 | | | |
| Thr232 | | Arg9 | | | |
| Tyr165 | | Thr12 | | | |
| Glu425 | | Tyr15 | | | |
| Asp160 | | Arg51 | | | |
| Lys161 | | Arg51 | | | |
| Lys183 | | | Tyr52 | | |
| Glu191 | | | Met55 | | |
| | Leu74 | | Gln95 | | |
| | Arg79 | | Ile93 | | |
| | Ser81 | | | Gln95 | |
| | Asn184 | | | Gln95 | |
| | Arg187 | | | Gln97 | |
| | Tyr190 | | | Thr54 | |
| | Arg194 | | | | Leu4 |
| | Lys227 | | | Asn43 | |
| | Gln229 | | | | Glu49 |
| | Asn432 | | | | Gln27 |
| | Lys183 | | | Tyr52 | |
| | Glu191 | | | Met55 | |
| Leu74 | | | | Gln95 | |
| | Thr232 | | | | Arg9 |
| | Tyr165 | | | | Thr12 |
| | Glu425 | | | | Tyr15 |
| | Gln429 | | | | Gln27 |
| | Asp160 | | | | Arg51 |
| | Lys161 | | | | Arg51 |
| Gly166 | | Val98 | | | |
| Gly196 | | Tyr15 | | | |
| | Gly166 | | | | Val98 |

There are two types of interactions between LSU and SSU as shown in **Figure 6K** viz. involving one LSU and one SSU only and other involving 2 LSUs and one SSU. In the first type the large subunit residues identified to be involved in heteromer interaction include Lys164, Arg167, Arg421, Arg194, Glu229, Asn432, Gly166, Gly196, Glu231, Glu433, Ala230, Thr232, Tyr165, Glu425, Asp160 and Lys161. It should be noted that the large subunit residues involved in interaction with the small subunit lie in the 9th α helix, loop between 9th α helix and 5th β strand, 10th α helix, loop between 10th α helix and 6th β strand, 11th α helix, 22nd α helix and loop between 22nd α helix and 23rd α helix. The SSU residues participating in this interaction with large subunit were also noted and are mentioned in **Table 6C**. The SSU residues involved in this interaction include Leu4, Lys6, Arg8, Arg9, Glu11, Thr12, Tyr15, Arg26, Gln27, Glu49, Arg51 and Val98. These residues occupy loop 1 (Leu4, Lys6, Arg8, Arg9, Glu11, Thr12 and Tyr15), α helix 1 (Arg26 and Gln27), loop 3 i.e. between β strand 1 and β strand 2 (Gln49 and Arg51) and β strand 4 (val98). The LSU residues involved in second type of interaction with SSU include Asp76, Arg79, Ser181, Asn184, Tyr190, Lys227, Lys183, Glu191, Leu74 and Arg187. The SSU residues in this type of interaction include Asn92, Ile93, Gln95, Thr54, Asn43, Tyr52, Met55, Gln95 and Gln97.

The next step was to identify if these residues are conserved in the LSU of *R. rubrum* and SSU of *G. violaceus*. Out of the 16 LSU residues involved in first type of interaction with SSU mentioned above for *T. elongatus*, only 6 were found to be conserved in *R. rubrum* LSU, viz. Gly157, Gly187, Gln219, Glu221, Thr222 and Arg411. Further, among the SSU residues only 6 out of 12 residues involved in first type of interaction mentioned above were conserved which include Leu4, Lys6, Glu11, Thr12, Tyr15 and Gln27.

The interactions estimated by GRAMMX were reported in form of 10 output files. Each of the outputs were analyzed for the residues involved in interaction and compared with the already reported residues involved in similar interactions as mentioned above. Among the 10 outputs generated, only 2 were found significant on the basis of similarity in residues involved. **Figure 6L** shows the position of SSU with respect to LSU dimer for both the outputs and the respective close up showing the residues involved in interaction for this arrangement. **Table 6D** shows the attributes of the interactions found significant for LSU and SSU.

In form I RuBisCO, the arrangement of LSU-SSU can be of two types as shown in **Figure 6K**. The figure shows A and I (the LSUs) forming a dimer and B, H, J and P (the SSUs) interacting with LSU in different arrangements. The arrangement of A-B/I-P involves LSU-16 and SSU-11 residues and arrangement A-H/J and I-H/J involves LSU-10 and SSU-8 residues. From the interaction estimated in present study the LSU residues involved are 4 and SSU

residues involved are 3-4. The results can be justified by the fact that *R. rubrum* RuBisCO LSU (Form II) is very different from *T. elongatus* RuBisCO LSU (Form I), and hence many interacting residues are not conserved. Although, 6 out of 16 residues are conserved but they were not found to be involved in interaction with SSU, possibly because of difference in tertiary structure arrangement of the residues with respect to SSU, as compared to that in case of form I RuBisCO LSU.

The results obtained suggest that the interaction between Form II LSU and Form I SSU does take place but is different from that observed in Form I RuBisCO. As reported by Lun et al (2014), RuBisCO small subunit is speculated to work as a CO₂ reservoir by virtue of its hydrophobic amino acids that have greater tendency to bind CO₂. The enzymatic assay performed in the present study showed a decrease in activity of *R. rubrum* RuBisCO in presence of increasing concentration of a foreign SSU. This could possibly be firstly, because of overcrowding by the SSU around the LSU making the active site inaccessible to the reaction metabolites or secondly because of binding of CO₂ to SSU, which did not interact with LSU as much as expected/ interacted in a manner different from expected and hence created a Ci deficient environment.

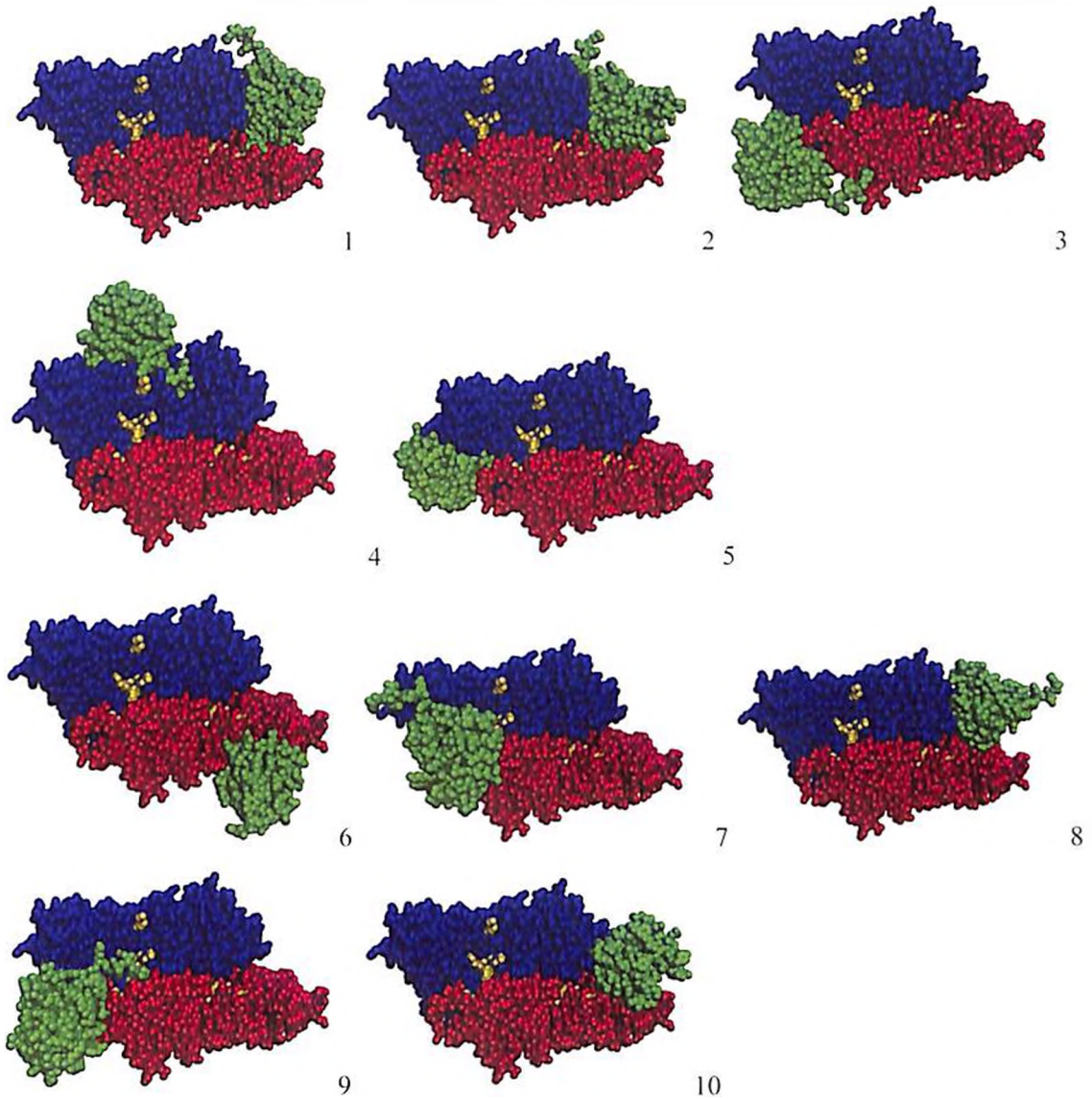
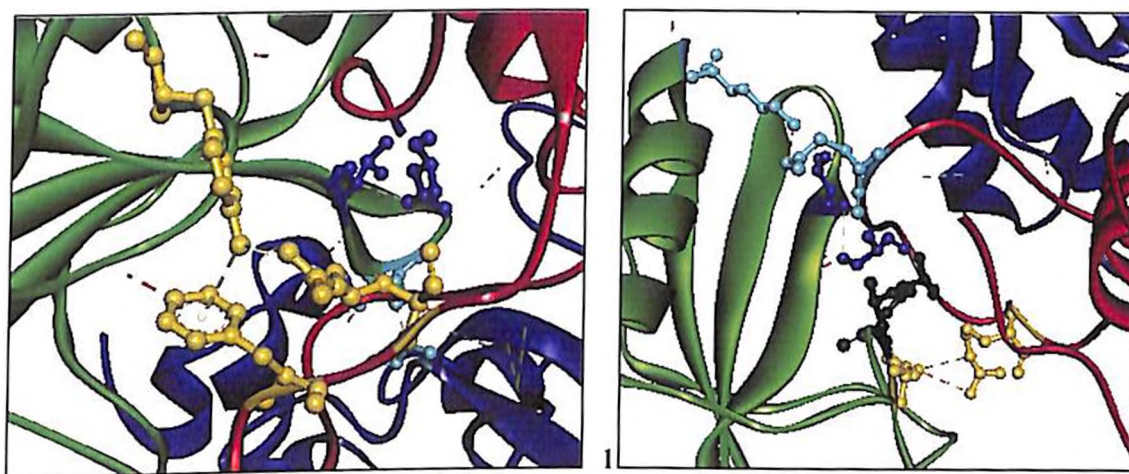


Figure 6L: The arrangement of *G. violaceus* RuBisCO small subunit with respect to *R. rubrum* RuBisCO large subunit dimer, according to the 10 output files (numbered 1-10 in the same order as generated by the server) generated by GRAMMX for interaction between the two proteins *in-silico*. The blue and the red CPK models represent the subunits (LSU) of the large subunit dimers, with the active site residues in *R. rubrum* colored as yellow (Asn111, Lys166, Lys168, Glu194, His287, Arg288, His321, Lys329, Ser368). The green CPK model represents the *G. violaceus* small subunit of RuBisCO arranged according to its interactions with the LSU dimer.

Table 6D: The attributes for the most possible interactions between *R. rubrum* RuBisCO and *G. violaceus* RbcS (SSU)

| Output | Attributes | Distance | Category | H-Donor | H-Acceptor |
|--------|-------------------------|----------|-------------------|---------|------------|
| 1 | A:Arg65:NH1-C:Tyr15:OH | 2.92377 | H-bond | Arg65 | Tyr15 |
| | B:Gly394:CA-C:Gln95:O | 2.52966 | H-bond | Gly394 | Gln95 |
| | C:Thr98:CA-A:Asp2:OD2 | 3.09876 | H-bond | Thr98 | Asp2 |
| | C:Tyr15:OH-A:Phe63 | 3.4705 | H-bond (Pi donor) | Tyr15 | Phe63 |
| 3 | B:Arg65:NH2-C:Glu11:OE1 | 3.56872 | H-bond | Arg65 | Glu11 |
| | C:Lys34:NZ-B:Glu56:OE1 | 3.57374 | H-bond | Lys34 | Glu56 |
| | B:Cys58:SG-C:Thr98:O | 3.21733 | H-bond | Cys58 | Thr98 |
| | B:Arg65:NE-C:Glu11:OE1 | 3.39419 | H-bond | Arg65 | Glu11 |
| | C:Thr12:OG1-B:Asp61:OD1 | 2.84633 | H-bond | Thr12 | Asp61 |



3

Figure 6M: The interacting residues (as generated by GRAMMX output files 1 and 3) between the *R. rubrum* RuBisCO LSU dimer and the *G. violaceus* RuBisCO SSU. The blue and the red chains represent the LSU of *R. rubrum* and the green chain represents the *G. violaceus* SSU. The interacting residues from the two subunits are shown in the ball and stick style. 1. The interacting residues according to output file 1 from GRAMMX are marked yellow (LSU-Arg65; SSU-Tyr15), light blue (LSU-Gly394; SSU-Gln95), purple (LSU-Asp2; SSU-Thr98) and yellow (LSU-Phe63; SSU-Tyr15). 2. The interacting residues according to output file 3 from GRAMMX are marked yellow (LSU-Arg65; SSU-Glu11), light blue (LSU-Glu56; SSU-Lys34), purple (LSU-Cys58; SSU-Thr98) and black (LSU-Asp61; SSU-Thr12).

6.3.3. The effect of *G. violaceus* RuBisCO SSU on activity of LSU

The *G. violaceus* RbcL (LSU), as mentioned in chapter 4 shows specific activity, V_{MAX} , K_M RuBP and K_{cat} of $0.005\mu\text{Moles}/\text{min}/\text{mg}$ (at 2mM RuBP concentration), $0.0002/\text{minute}$, 0.1917mM and $0.0029/\text{second}$, respectively. In an attempt to study the effect of RbcS (SSU) on activity of RbcL, *G. violaceus* RbcS purified by IMAC after overexpression in *E. coli* BL21(DE3)pLysS (chapter 4) was incubated with *G. violaceus* RbcL, in presence of RbcX (chaperone protein), ATP and NaHCO_3 (as mentioned earlier in this chapter), *in-vitro*. The ratios of RbcL and RbcS concentrations of *G. violaceus* used for the analysis are mentioned in Table 6E. The Figure 6N shows the effect of concentration of RbcS on the specific activity of *G. violaceus* RbcL.

Table 6E: The ratios and corresponding concentrations of the *G. violaceus* RbcL (LSU) and *G. violaceus* RbcS (SSU) used for complementation studies.

| Ratio (RbcL:RbcS) | <i>G. violaceus</i> RbcL (μg) | <i>G. violaceus</i> RbcS (μg) |
|-------------------|--|--|
| 1:0 | 60 | - |
| 1:0.5 | 60 | 30 |
| 1:0.75 | 60 | 45 |
| 1:1 | 60 | 60 |
| 1:1.5 | 60 | 90 |
| 1:2 | 60 | 120 |
| 1:3 | 60 | 150 |
| 0:1 | - | 60 |

The in-silico estimation of interaction between G. violaceus RbcL and G. violaceus RbcS

The *in-silico* generated models for *G. violaceus* RbcL and RbcS were subjected to energy minimization and validation as mentioned in chapter 4. The RuBisCO protein assembly as reported by Liu et al (2010) is initiated by formation of the RbcL dimer with the help of RbcX which eventually replaced by the RbcS subunits. Hence, the interaction analysis was performed according to similar pattern, viz. the interaction between the RbcL dimer and the RbcS subunits was estimated.

The *G. violaceus* RbcL protein dimer interactions were estimated by protein-protein docking using GRAMMX online server. The outputs generated were analyzed to identify the most biologically significant set of interactions by comparing with the already reported residues involved in such interactions (mentioned in **Table 6C**). The outputs were validated by binding energy involved and the surface area in contact during such interaction. The interaction file after validation by PDBePISA (Krissinel and Henrick 2007) showed $T\Delta S^{\text{dis}}$ of 15.1 Kcal/mol and buried surface area of 2383.6 \AA^2 . As reported by Day et al (2012), the most efficient protein-protein complexes involve $\sim 20 \text{ Kcal/mol/ \AA}^2$ of binding energy and require $\sim 500 \text{ \AA}^2$ area of contact to form a stable complex.

In the next step, the *G. violaceus* RbcL dimer and the *G. violaceus* RbcS docking was conducted. The GRAMMX outputs generated were analyzed. **Figure 6P** shows the different arrangements of the subunits as estimated by the outputs generated by the docking server. The output files 1 and 2 showed interactions of RbcS with RbcLA (monomer 1) and RbcLB (monomer 2), respectively via the same set of residues. Among the residues participating from the RbcL viz. Arg338, Arg359, Asp395, Arg430, Asp356, Arg388, Gly394, Arg438 and Gly433, none were found to match with the residues of *T. elongatus* reported to be involved in RbcL-RbcS interactions. While the RbcS residues viz. Glu11, Thr12, Gln95 and Met55 are also involved in similar interaction in the RuBisCO protein from *T. elongatus*.

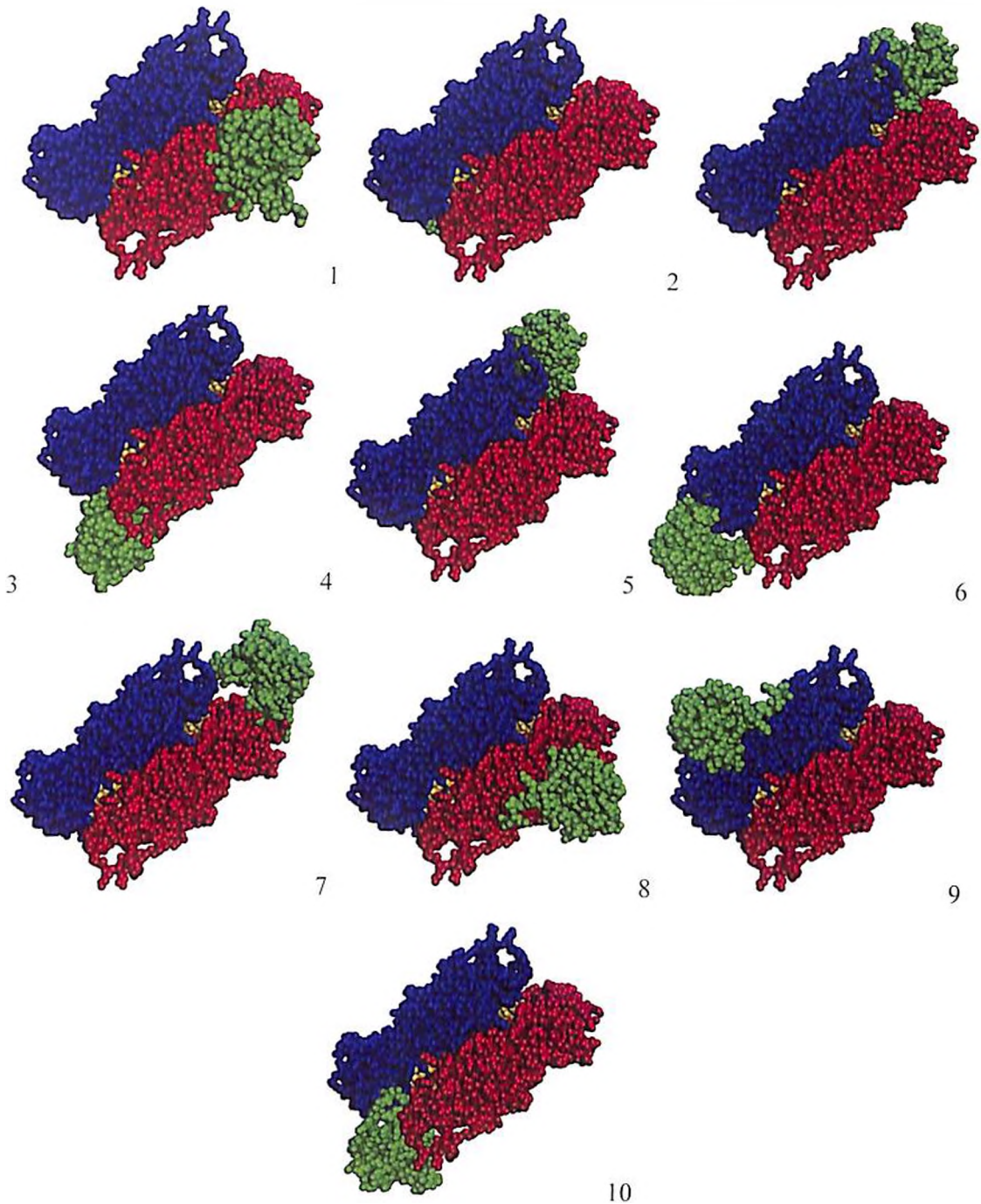


Figure 6P: The arrangement of *G. violaceus* RuBisCO small subunit with respect to *G. violaceus* RuBisCO large subunit dimer, according to the 10 output files (numbered 1-10 in the same order as generated by the server) generated by GRAMMX for interaction between the two proteins *in-silico*. The blue and the red CPK models represent the subunits (LSU) of the large subunit dimers, with the active site residues in *G. violaceus* colored as yellow (Lys174, Lys200, Asp202, Glu203, His293, Lys333, Thr172, Lys176, Arg294 and His326). The green CPK model represents the *G. violaceus* small subunit of RuBisCO arranged according to its interactions with the LSU dimer.

Table 6F: The attributes for the most possible interactions between *G. violaceus* RbcL (LSU) dimer and *G. violaceus* RbcS (SSU)

| Attributes | Distance | Category | H donor | H acceptor |
|--------------------------|----------|-------------------|---------|------------|
| A:Asn162:ND2-C:Phe9:O | 2.90491 | H-Bond | Asn162 | Phe9 |
| A:Gln228:NE2-C:Glu11:OE2 | 3.31896 | H-Bond | Gln228 | Glu11 |
| A:Gln228:NE2-C:Thr12:OG1 | 3.02441 | H-Bond | Gln228 | Thr12 |
| A:Lys226:CA-C:Gln97:O | 2.71977 | H-Bond | Lys226 | Gln97 |
| A:Lys226:CE-C:Thr98:OG1 | 3.4817 | H-Bond | Lys226 | Thr98 |
| A:Gly232:CA-C:Met50:O | 3.15809 | H-Bond | Gly232 | Met50 |
| C:Thr49:CA-A:Glu233:OE1 | 3.41894 | H-Bond | Thr49 | Glu233 |
| C:Val51:CA-A:Thr231:O | 3.08036 | H-Bond | Val51 | Thr231 |
| C:Cys96:SG-A:Tyr189 | 4.03024 | H-Bond (pi donor) | Cys96 | Tyr189 |



Figure 6Q: The interacting residues (as generated by GRAMMX output file 8) between the *G. violaceus* RuBisCO LSU dimer and the *G. violaceus* RuBisCO SSU. The blue and the red chains represent the LSU of *G. violaceus* and the green chain represents the *G. violaceus* SSU. The interacting residues from the two subunits are shown in the ball and stick style. 8. The interacting residues according to output file 8 from GRAMMX are marked yellow (LSU-Asn162; SSU-Phe9), light blue (LSU-Gln228; SSU-Glu11 and Thr12), purple (LSU-Lys226; SSU-Gln97 and Thr98), Pink (LSU-Gly232; SSU-Met50), Orange (LSU-Glu233; SSU-Thr49), magents (LSU-Thr231; SSU-Val51) and light green (Tyr189; SSU-Cys96).

The output file 8 (**Table 6F** and **Figure 6Q**) on the other hand showed conserved residues from both the subunits to be participating in the interaction. The RbcL residues involved are Asn162, Gln228 (corresponding residue in *T. elongatus* is Gln229), Lys226 (corresponding residue in *T. elongatus* Lys227) Gly232, Glu233, Thr231 (corresponding residue in *T.*

was also conducted. The assay revealed a slight enhancement in the K_M RuBP of the resulting protein complex (Figure 6S i and ii). The K_M RuBP of *R. rubrum* RbcL + *G. violaceus* CcmM was found to be 0.0316mM against the K_M RuBP of *R. rubrum* RuBisCO to be 0.049mM. The V_{MAX} of the RbcL + CcmM complex was found to decrease by as much as ~50% from 0.5737/minute to 0.2876/minute.

Table 6G: The ratios and corresponding concentrations of the *R. rubrum* RbcL (LSU) and *G. violaceus* CcmM used for complementation studies.

| Ratio (RbcL:CcmM) | <i>R. rubrum</i> RbcL (μg) | <i>G. violaceus</i> CcmM (μg) |
|-------------------|---|--|
| 1:0 | 40 | 0 |
| 1:1 | 40 | 40 |
| 1:2 | 40 | 80 |
| 1:3 | 40 | 120 |
| 1:6 | 40 | 240 |
| 0:1 | 0 | 40 |

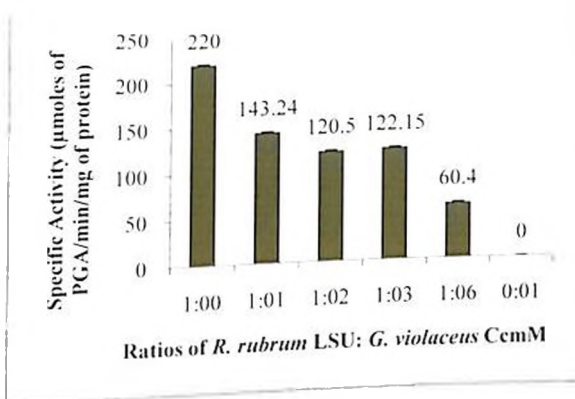


Figure 6R: The effect of different concentrations of *G. violaceus* CcmM on specific activity of *R. rubrum* RbcL (LSU). The concentration of *R. rubrum* RbcL used was 40 μg and for *G. violaceus* CcmM 0, 40, 80, 120, 240 and 40 μg from left to right on the X axis of the graph.

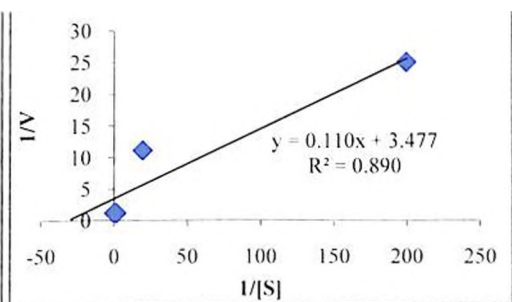
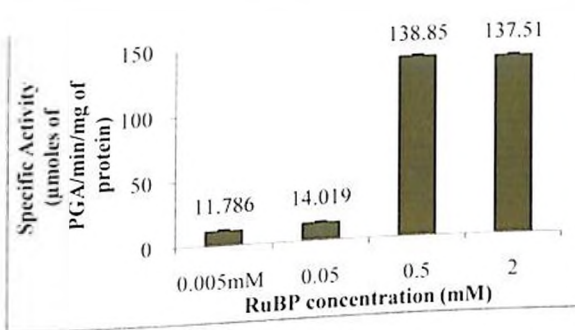


Figure 6S i and ii: i) The substrate (RuBP) dependent assay of the *R. rubrum* RbcL in presence of *G. violaceus* CcmM in ratio of 1:3, respectively. The X axis show the RuBP concentrations used for the assay and Y axis show the specific activity. The error bars show the standard deviation observed for the specific activity calculated. ii) The Lineweaver-Burk plot for the RuBP dependent assay for *R. rubrum* RbcL + *G. violaceus* CcmM complex, in-vitro in ratio of 1:3, respectively.

The in-silico analysis of the interactions between R. rubrum RuBisCO and G. violaceus CcmM

The *in-silico* generated model for CcmM was energy minimized and validated by ModelRefiner and PDBsum, respectively. The Ramachandran plot for the CcmM full form of the protein, suggests it to be of poor quality, which is because of the absence of any suitable template for the protein. The CcmM IRES1 although also a product of *in-silico* homology modeling and threading, was found to be of a comparatively better quality and was hence used for present analysis. The CcmM IRES1 protein has four SSU domains as predicted by the CDD search of the protein. **Table 6H** and **Figure 6T** show the arrangement of these domains on the protein. According to the CDD search, it was expected that the protein will have four folded entities (one each for the SSU domains) connected by linkers. As shown in the **Figure 6T**, the CcmM IRES1 has three folded entities connected by two linkers. **Table 6I** shows the residues involved in formation of the folded entities viz. I', II' and III' and the linkers viz. I'-II' and II'-III'.

For the interaction studies between the RbcL and CcmM, the *R. rubrum* RbcL dimer and *G. violaceus* CcmM IRES1 models were submitted to GRAMMX. The outputs generated were analyzed for the residues involved and the arrangement of the two proteins with respect to each other.

Table 6H: The RuBisCO SSU domains of *G. violaceus* CcmM as estimated by CDD search and the conserved residues in each of the domains with respect to the *G. violaceus* RuBisCO small subunit identified by sequence alignment

| Range of residues | SSU Domain | % Identity with RbcS | Conserved residues w.r.t RbcS (<i>G. violaceus</i>) |
|-------------------|------------|----------------------|---|
| 1-4 | - | - | |
| 5-31 | I | 22 | G7, R11, I13, G14, D16, P31 |
| 32-63 | - | - | |
| 64-149 | II | 27 | K20, E25, T26, P52, P53, I64, E68, A70, Q72, Q74, L77, G80, Y81, E86, W98, P103, K108, P110, E111, V112, L113, C119, G125, R129, I131, G132, D134, I145, P149 |
| 150-164 | - | - | |
| 165-262 | III | 21 | Y165, T167, S169, D179, A180, L190, G193, V198, E199, F200, W211, P215, E224, V225, S227, G238, R242, I244, G245, D247, P262 |
| 263-299 | - | - | |
| 300-382 | IV | 20 | E301, Q305, L310, G313, V318, E319, F320, W331, P335, E344, G358, R362, I364, G365, D367, I378, P382 |

Table 6I: The domains of *G. violaceus* CcmM IRES1 identified on the basis of the tertiary structure of the protein estimated *in-silico*

| Range of residues | Domain |
|-------------------|----------|
| 1-86 | I' |
| 87-102 | I'-II' |
| 103-236 | II' |
| 237-299 | II'-III' |
| 300-382 | III' |

The **Figure 6T** shows the arrangement of the proteins with respect to each other according to the 10 outputs generated by GRAMMX. In order to identify the expected residues, the conserved residues between *G. violaceus* CcmM SSU domain and *G. violaceus* RbcS were noted (**Table 6H**). Only a few residues identified in these outputs were found to be common with the residues conserved. According to the residues involved and the expected arrangement of RuBisCO with respect to CcmM at the three folded entities, the outputs 1, 3 and 6 were found to be most appropriate. The residues conserved in the SSU domain which were also found to be involved in LSU-SSU interaction of *R. rubrum* and *G. violaceus*, respectively, were Glu301 (Glu23 of *G. violaceus* RbcS), Pro382 (Pro106 of *G. violaceus* RbcS) and Pro103 (Pro59 of *G. violaceus* RbcS), although their interacting partners were different. Some of the residues involved in interaction between *R. rubrum* RbcL and *G. violaceus* RbcS and between *R. rubrum* RbcL and *G. violaceus* CcmM were found to be common i.e. Arg65, Asp2 and Glu56 for RbcL of *R. rubrum* and Tyr15, Glu11, Thr12 and Thr98 for *G. violaceus* RbcS.

The arrangement of RbcL with respect to CcmM IRES1 as shown in **Figure 6T** (1, 3 and 6), is in accordance with as speculated by Long et al (2007), and if one CcmM IRES1 interacts with three RuBisCO holoenzymes, the residues involved in these interactions viz. SSU Domain I', SSU Domain II' and SSU domain III', would be as mentioned in **Table 6I** and shown in **Figure 6U**.

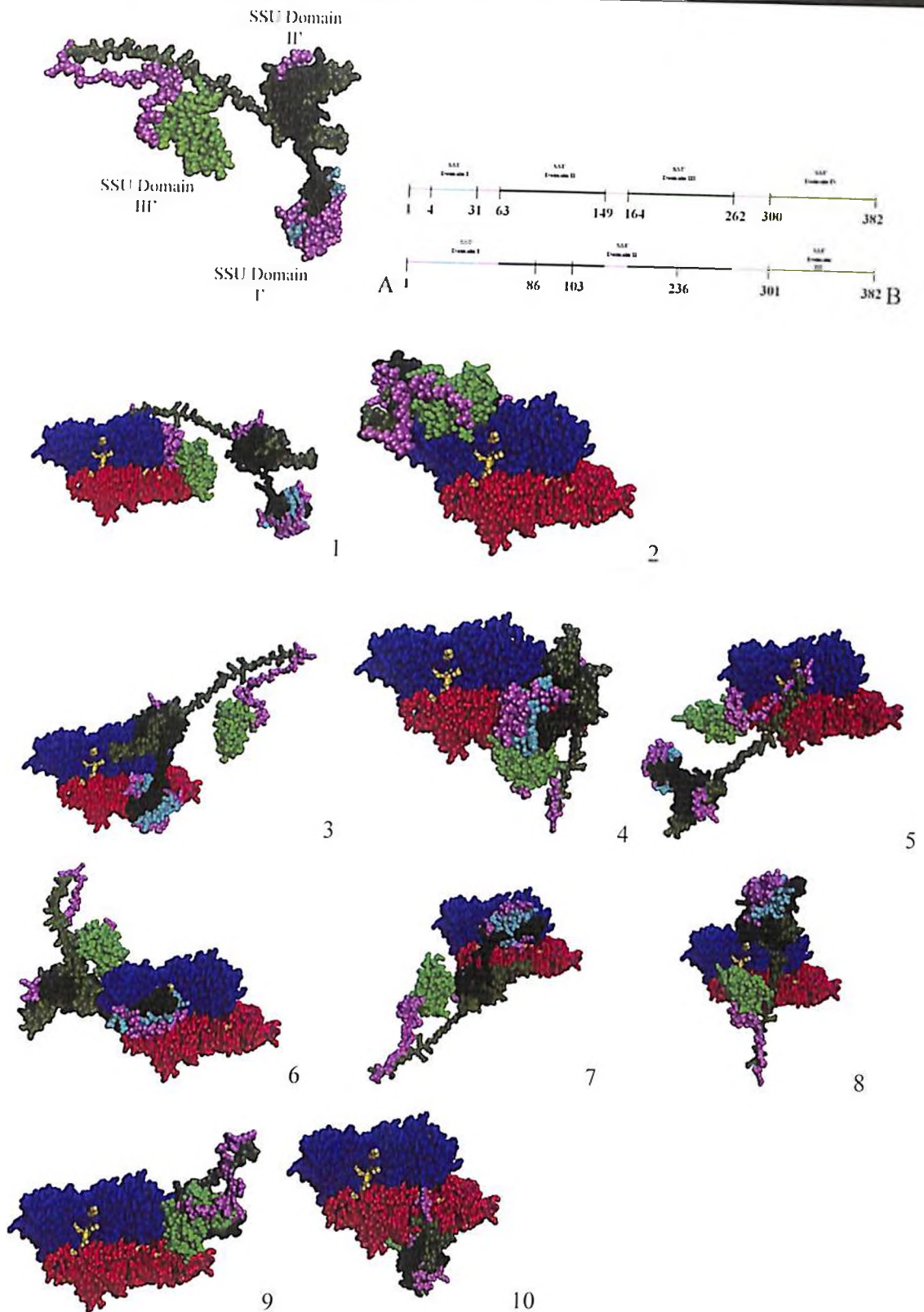


Figure 6T: A- CPK model of CcmM IRES1 coloured according to SSU domains as mentioned in table 6IX and B. 1-10: The arrangement of *G. violaceus* CcmM with respect to *R. rubrum* RuBisCO large subunit dimer, according to the 10 output files (numbered 1-10 in the same order as generated by the server) generated by GRAMMX for interaction between the two proteins *in-silico*. The blue and the red CPK models represent the subunits (LSU) of the large subunit dimers, with the active site residues in *R. rubrum* colored as yellow. The green-pink CPK model represents the *G. violaceus* CcmM arranged according to its interactions with the LSU dimer.

Table 6J: The attributes for the most possible interactions between *R. rubrum* RbcL (LSU) dimer and *G. violaceus* CcmM IRES1

| Out put No | Attributes | Distance | Category | H Donor | H acceptor | SSU Domain |
|------------------------|--------------------------|----------|----------|---------|------------|------------|
| 1 | A:Lys13:NZ-C:Glu256:OE2 | 3.73503 | H Bond | Lys13 | Glu256 | II'-III' |
| | B:Gly442:N-C:Glu301:OE1 | 2.36388 | H Bond | Gly442 | Glu301 | III' |
| | B:Asp443:N-C:Glu301:OE1 | 3.27812 | H Bond | Asp443 | Glu301 | III' |
| | C:Thr276:OG1-A:Val8:O | 2.603 | H Bond | Thr276 | Val8 | II'-III' |
| | B:Pro441:CD-C:Ser299:O | 3.33195 | H Bond | Pro441 | Ser299 | II'-III' |
| | B:Gly450:CA-C:Pro382:O | 3.7792 | H Bond | Gly450 | Pro382 | III' |
| | C:Pro262:CD-A:Gly20:O | 3.48717 | H Bond | Pro262 | Gly20 | II'-III' |
| C:Thr273:CB-A:Arg265:O | 2.63743 | H Bond | Thr273 | Arg265 | II'-III' | |
| 3 | A:Arg78:NH1-C:Glu143:OE2 | 2.69351 | H Bond | Arg78 | Glu143 | II' |
| | C:Arg207:NH1-A:Glu76:OE1 | 2.90353 | H Bond | Arg207 | Glu76 | II' |
| | A:Glu22:N-C:Leu189:O | 3.20501 | H Bond | Glu22 | Leu189 | II' |
| | A:Glu322:N-C:Gly62:O | 2.33132 | H Bond | Glu322 | Gly62 | II' |
| | C:Arg91:NH2-A:Ala34:O | 2.57114 | H Bond | Arg91 | Ala34 | I'-II' |
| | A:Ser5:CB-C:Pro103:O | 2.77304 | H Bond | Ser5 | Pro103 | II' |
| | A:Gly20:CA-C:Ser188:O | 2.5888 | H Bond | Gly20 | Ser188 | II' |
| | A:Asp75:CA-C:Glu222:OE1 | 3.35822 | H Bond | Asp75 | Glu222 | II' |
| C:Leu189:CA-A:Gly20:O | 2.65871 | H Bond | Leu189 | Gly20 | II' | |
| 6 | C:Arg92:NH1-B:Glu56:OE1 | 2.88673 | H Bond | Arg92 | Glu56 | I'-II' |
| | A:Lys431:NZ-C:Ser97:OG | 3.38587 | H Bond | Lys431 | Ser97 | I'-II' |
| | A:Arg435:NH2-C:Ser97:OG | 2.94697 | H Bond | Arg435 | Ser97 | I'-II' |
| | B:Tyr36:OH-C:Ala50:O | 2.4379 | H Bond | Tyr36 | Ala50 | I' |
| | C:Phe9:N-A:Ile400:O | 3.18951 | H Bond | Phe9 | Ile400 | I'-II' |
| | A:Ile400:CA-C:Arg91:O | 3.49018 | H Bond | Ile400 | Arg91 | I'-II' |
| | C:Arg58:CA-B:Asp2:OD1 | 3.49453 | H Bond | Arg58 | Asp2 | I' |
| C:Thr85:CB-A:Gly442:O | 3.65641 | H Bond | Thr85 | Gly442 | I' | |

The interaction file after validation by PDBEPIA showed $T\Delta S^{\text{diss}}$ of 14.6Kcal/mol and buried surface area of 10260.1\AA^2 suggesting a stable complex hence formed. As per the in-vitro and in-silico analysis, the CcmM protein of *G. violaceus* does form stable complex with *R. rubrum* RuBisCO (form II). The complexes formed are via residues different from the most conserved residues in the SSU domains. As from the in-vitro studies the protein upon its interaction with the large subunit is found to cause a decrease in activity, possibly due to blockage of the active site of the RbcL in many of the PPI as shown in **Figure 6T**. Further, CcmM possibly has affinity for RuBP, causing it to concentrate around the active site of the large subunit and hence causing a decrease in the effective K_M RuBP of the large subunit.

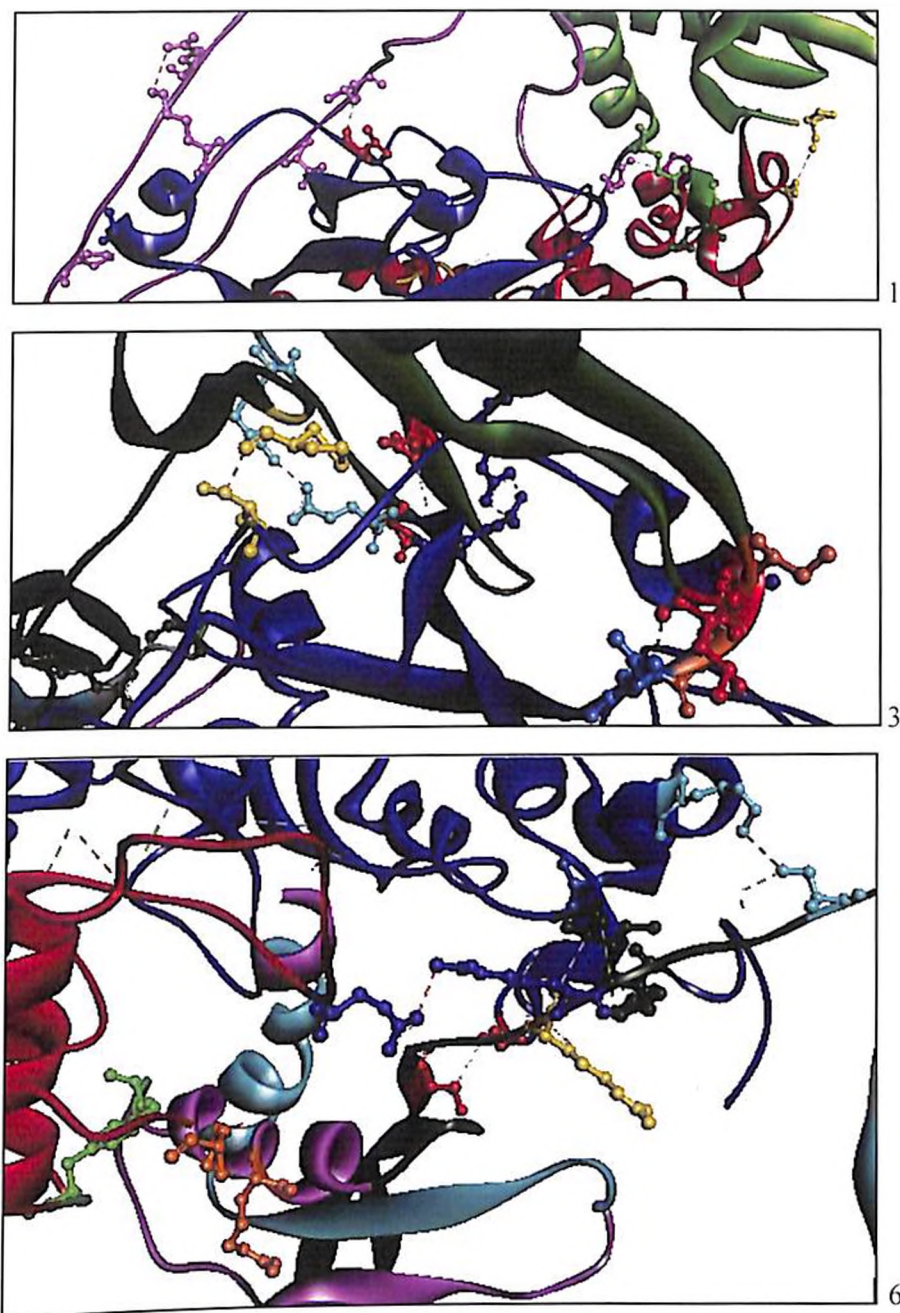


Figure 6U: The interacting residues (as generated by GRAMMX output files 1, 3 and 6, respectively) between the *R. rubrum* RuBisCO LSU dimer and the *G. violaceus* CcmM IRES1. The blue and the red chains represent the LSU of *R. rubrum* and the others represent the *G. violaceus* CcmM IRES1 coloured according to the location of residues on the chain. The interacting residues from the two subunits are shown in the ball and stick style. 1, 3 and 6. The interacting residues according to output file 1, 3 and 6 from GRAMMX are marked violet, Light Blue, Blue, Green, Dark Green, Yellow, Light Orange, dark Orange and Red in the order as mentioned in Table 6J.

6.3.5. Effect of *G. violaceus* CcmM on the activity of *G. violaceus* RuBisCO LSU

The *G. violaceus* RbcL and CcmM purified by IMAC were reconstituted according to methodology as mentioned earlier. The ratios of the concentrations of the two proteins used for in-vitro activity assay are mentioned in **Table 6K**. The concentration of *G. violaceus* RbcL used was 60 μ g. The **Figure 6V** shows the decrease in carboxylation activity of RbcL (*G. violaceus*) upon addition of CcmM (*G. violaceus*).

Table 6K: The ratios and corresponding concentrations of the *G. violaceus* RbcL (LSU) and *G. violaceus* CcmM used for complementation studies.

| Ratio (RbcL:CcmM) | <i>G. violaceus</i> RbcL (μ g) | <i>G. violaceus</i> CcmM (μ g) |
|-------------------|-------------------------------------|-------------------------------------|
| 1:0 | 60 | 0 |
| 1:1 | 60 | 60 |
| 1:2 | 60 | 120 |
| 1:3 | 60 | 180 |
| 1:6 | 60 | 360 |
| 0:1 | 0 | 60 |

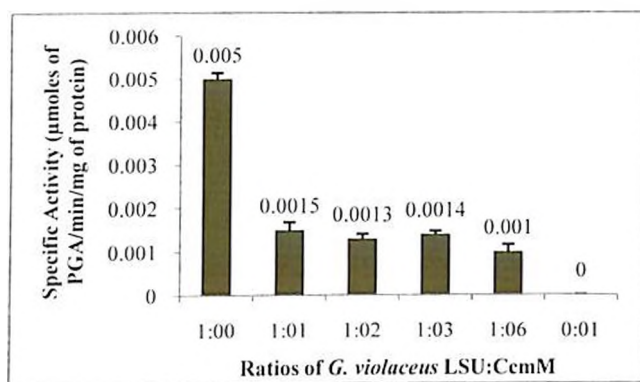


Figure 6V: The effect of different concentrations of *G. violaceus* CcmM on specific activity of *G. violaceus* RbcL (LSU). The concentration of *G. violaceus* RbcL used was 60 μ g and for *G. violaceus* CcmM 0, 60, 120, 180, 360 and 60 μ g from left to right on the X axis of the graph.

The specific activity of RbcL + CcmM complex at 1:1 ratio was as low as 0.0015 μ Moles of PGA/min/mg of protein which is $\sim 1/3^{\text{rd}}$ of the specific activity of RbcL alone (2mM RuBP concentration). Further **Figures 6W i** and **ii**, show the effect of RuBP concentration on the activity of *G. violaceus* RbcL + CcmM complex in ratio of 1:3, respectively. As observed for *R. rubrum* RbcL + *G. violaceus* CcmM complex, the K_M RuBP of *G. violaceus* RbcL + CcmM was also found to be slightly enhanced to 0.1150mM.

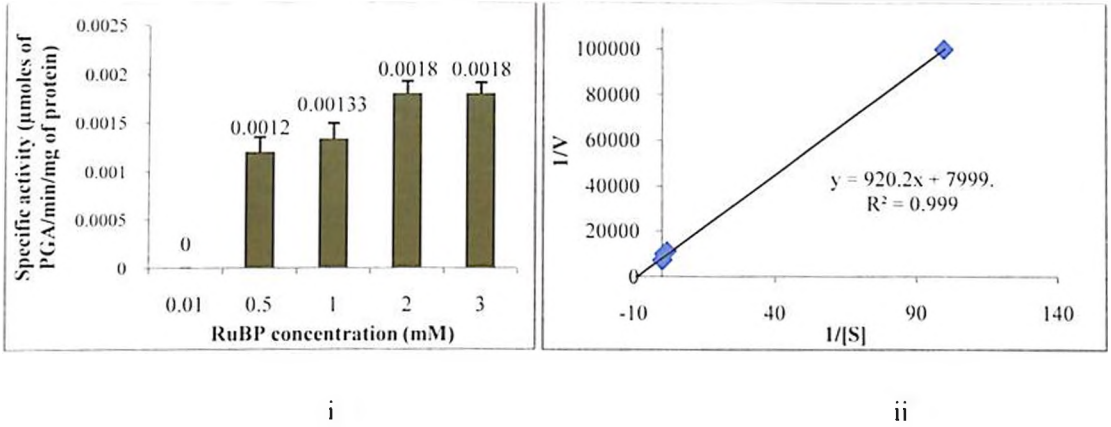


Figure 6W i and ii: i) The substrate (RuBP) dependent assay of the *G. violaceus* RbcL in presence of *G. violaceus* CcmM in ratio of 1:3, respectively. The X axis show the RuBP concentrations used for the assay and Y axis show the specific activity. The error bars show the standard deviation observed for the specific activity calculated. ii) The Lineweaver-Burk plot for the RuBP dependent assay for *G. violaceus* RbcL + *G. violaceus* CcmM complex, in-vitro in ratio of 1:3, respectively.

The in-silico interaction analysis between *G. violaceus* RbcL and *G. violaceus* CcmM

The in-silico studies for estimating interaction between *G. violaceus* RbcL and *G. violaceus* CcmM was done using energy minimized models of *G. violaceus* RbcL dimer and CcmM IRES1, respectively. The GRAMMX generated 10 output files which were visualized in Discovery Studio 4.5 and analyzed for arrangement of the proteins with respect to each other (Figure 6X 1-10) and the residues involved. According to the residues involved and the expected arrangement of the proteins in ratio 1:3 for RbcL and CcmM, respectively, output files 3, 6 and 9 were found to be most appropriate. Only two residues viz. Tyr81 (corresponding to Tyr36 in *G. violaceus* RbcS) and Trp211 (corresponding to Trp53 in *G. violaceus* RbcS) were the conserved residues found in both RbcS and CcmM of *G. violaceus* but found in interaction with RbcL via CcmM alone. It should be noted that the residues involved in interaction between *G. violaceus* RbcL and CcmM are completely different from the *G. violaceus* RbcS residues involved in interaction with RbcL. The analysis reveals some residues of CcmM IRES1 involved in interaction with RbcL. from both *R. rubrum* and *G. violaceus* viz. Arg92 (for binding to I'), Leu189 and Arg91 (for binding to II') and Glu256 and Ser299 (for binding to III'). The attributes of the residues involved RbcL and CcmM interaction are mentioned in Table 6L and shown in Figure 6Y.

Table 6L: The attributes for the most possible interactions between *G. violaceus* RbcL (LSU) dimer and *G. violaceus* CcmM IRES1

| Output No | Attributes | Distance | Category | H Donor | H acceptor | SSU Domain |
|-----------|-------------------------------|----------|----------|---------|------------|------------|
| 3 | C:Arg221:NH2- A:Glu467:OE2 | 2.87936 | H bond | Arg221 | Glu467 | II' |
| | A:Tyr84:OH-C:Tyr9:O | 3.34966 | H bond | Tyr84 | Tyr9 | I' |
| | A:Arg138:NH1-C:Tyr81:OH | 3.11542 | H bond | Arg138 | Tyr81 | I' |
| | A:Thr364:OG1-C:Tyr81:O | 2.76912 | H bond | Thr364 | Tyr81 | I' |
| | B:Val16:N-C:Glu230:OE1 | 3.03778 | H bond | Val16 | Glu230 | II' |
| | B:Arg20:NE-C:Tyr240:OH | 2.72534 | H bond | Arg20 | Tyr240 | II'-III' |
| | C:Trp211:NE1- A:Glu467:OE1 | 2.74998 | H bond | Trp211 | Glu467 | II' |
| | B:Gly46:CA-C:Glu203:OE2 | 3.77414 | H bond | Gly46 | Glu203 | II' |
| | C:Ala6:CA-A:Asp85:OD1 | 3.23005 | H bond | Ala6 | Asp85 | I' |
| | C:Tyr9:CA-A:Tyr84:OH | 3.1814 | H bond | Tyr9 | Tyr84 | I' |
| 5 | A:Asn94:ND2-C:Ser299:OG | 2.61982 | H bond | Asn94 | Ser299 | II'-III'* |
| | C:Val258:N-B:Glu459:O | 3.05394 | H bond | Val258 | Glu459 | II'-III' |
| | A:Pro43:CD-C:Leu298:O | 3.75897 | H bond | Pro43 | Leu298 | II'-III'* |
| | C:Glu256:CA- B:Glu463:OE1 | 3.29222 | H bond | Glu256 | Glu463 | II'-III' |
| | C:Ser275:CB-B:Tyr468:OH | 3.64679 | H bond | Ser275 | Tyr468 | II'-III' |
| 9 | C:Arg91:NH1-B:Glu87:OE2 | 3.1966 | H bond | Arg91 | Glu87 | I'-II' |
| | C:Arg30:NH1-B:Gly91:O | 3.23444 | H bond | Arg30 | Gly91 | I'-II' |
| | C:Cys101:SG-B:His352:O | 3.59249 | H bond | Cys101 | His352 | I'-II' |
| | B:His152:CA-C:Leu189:O | 3.72722 | H bond | His152 | Leu189 | II' |
| | B:Arg338:CD- C:Gln218:OE1 | 3.79153 | H bond | Arg338 | Gln218 | II' |
| | C:Ser191:CB- B:Asp159:OD1 | 3.60443 | H bond | Ser191 | Asp159 | II' |

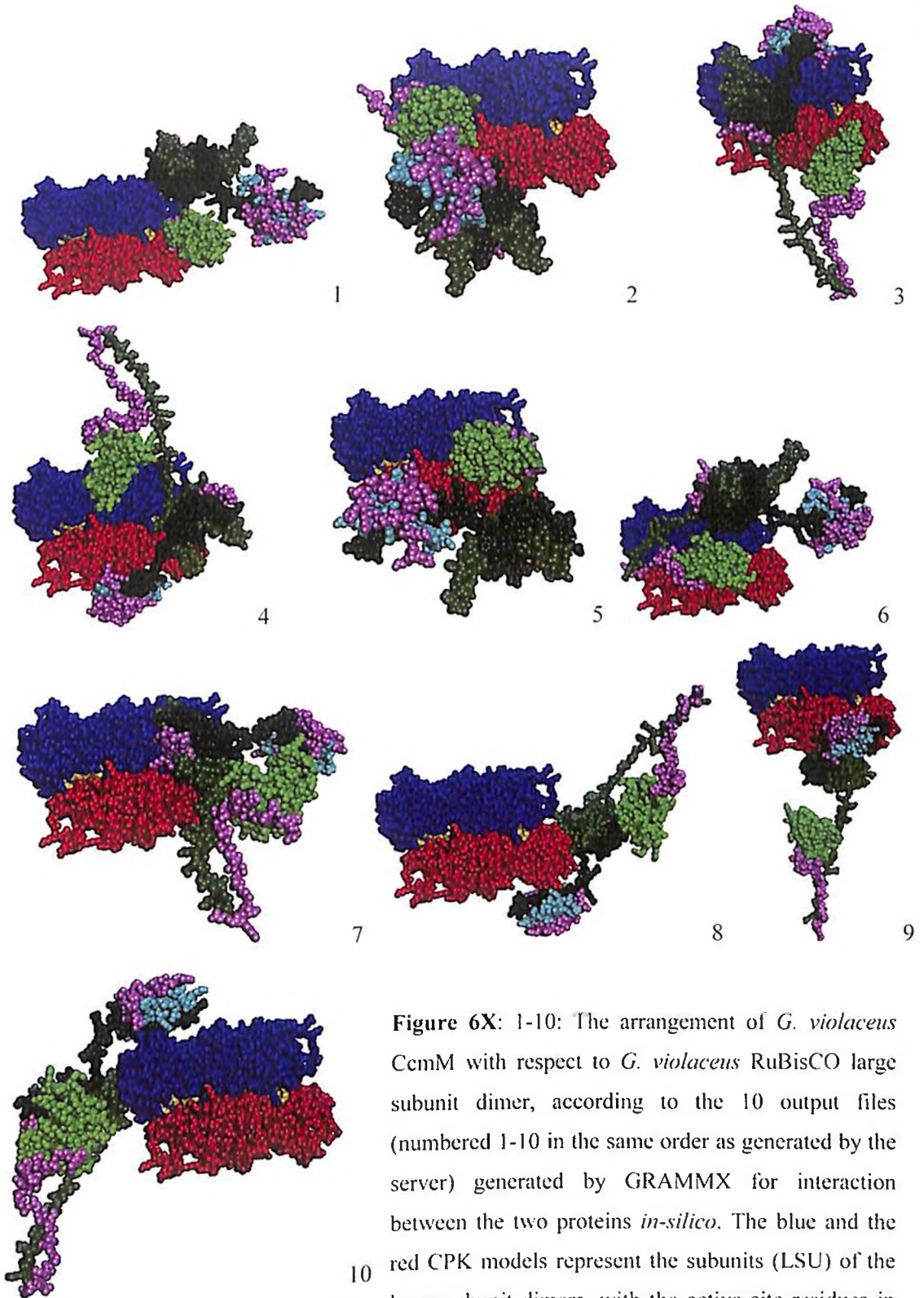


Figure 6X: 1-10: The arrangement of *G. violaceus* CcmM with respect to *G. violaceus* RuBisCO large subunit dimer, according to the 10 output files (numbered 1-10 in the same order as generated by the server) generated by GRAMMX for interaction between the two proteins *in-silico*. The blue and the red CPK models represent the subunits (LSU) of the large subunit dimers, with the active site residues in *G. violaceus* colored as yellow (Lys174, Lys200, Asp202, Glu203, His293, Lys333, Thr172, Lys176, Arg294 and His326). The green-pink CPK model represents the *G. violaceus* CcmM arranged according to its interactions with the LSU dimer.

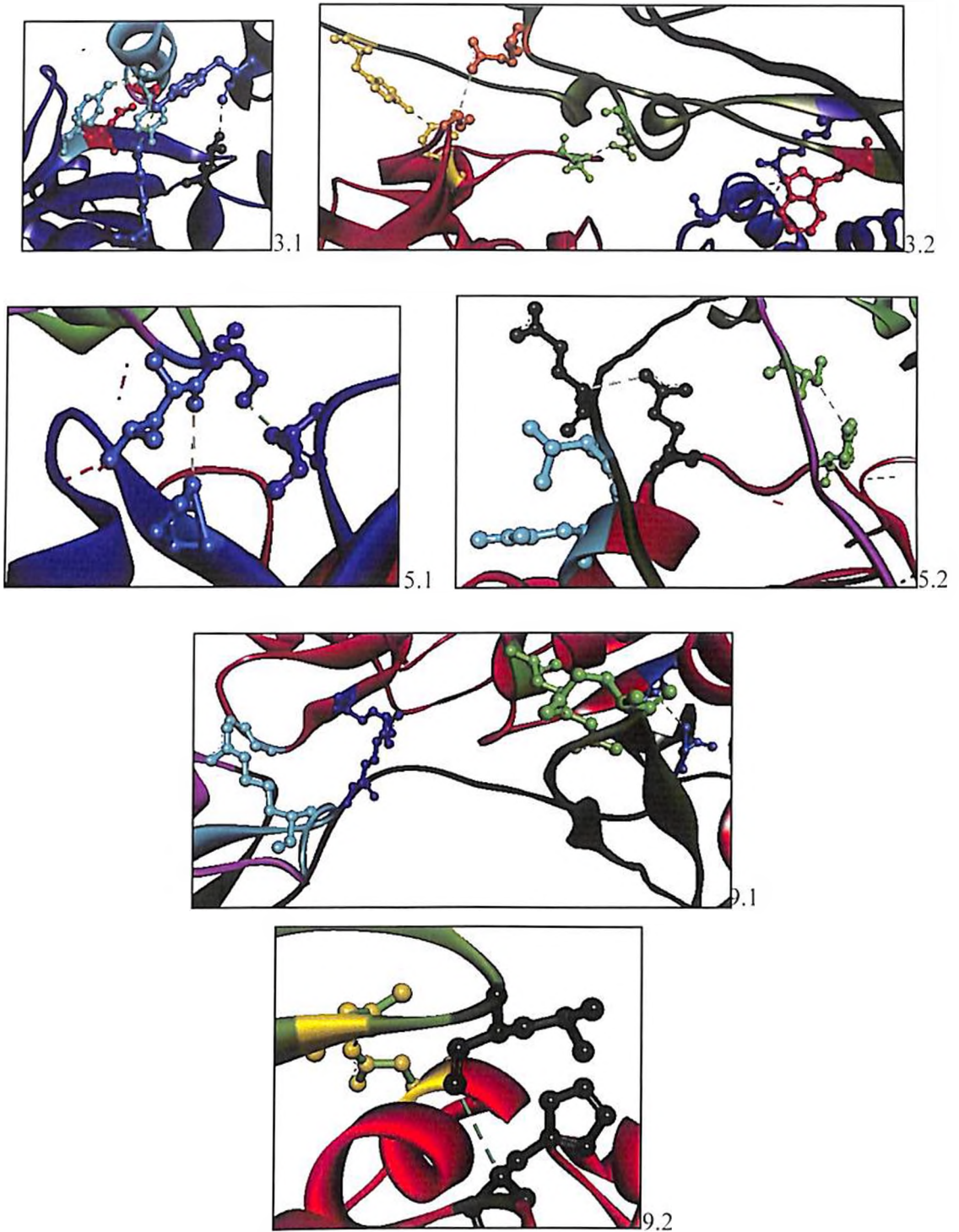


Figure 6Y: The interacting residues (as generated by GRAMMX output files 3 (3.1 and 3.2), 5 (5.5 and 5.2) and 9 (9.1 and 9.2)) between the *G. violaceus* RuBisCO LSU dimer and the *G. violaceus* CcmM IRES1. The blue and the red chains represent the LSU of *R. rubrum* and the others represent the *G. violaceus* CcmM IRES1 coloured according to the location of residues on the chain. The interacting residues from the two subunits are shown in the ball and stick style. The interacting residues according to output file 3, 5 and 9 from GRAMMX are marked violet, Light Blue, Blue, Dark Green, Light Green, Yellow, Light Orange, Dark Orange, Red and Pink in the order as mentioned in Table 6L.

The analysis shows possible interaction between *G. violaceus* RbcL dimer and CcmM IRES bearing TAS^{diss} of 14.6Kcal/mol and involving buried area of 7668.7\AA^2 . The in-vitro studies suggest, no effective augmentation in the specific activity of the large subunit but a slight enhancement in the affinity of the protein for the substrate RuBP. Further, some residues of CcmM IRES1 viz. Arg92, Leu189, Arg91, Glu256 and Ser299 have been identified to be involved in interaction with both the form I and form II RuBisCO large subunits.

6.4. Conclusions

The *in-vitro* analysis conducted in the present study viz. the complementation of large subunit of form I and form II RuBisCO with the small subunit and the CcmM protein from *G. violaceus* showed different kinds of effect on the activity of the former. The small subunit where found to augment the activity of the large subunit of form I RuBisCO, lead to deterioration of the activity of the form II RuBisCO. The same when analyzed *in-silico*, showed possible interaction between both set of proteins but via different set of residues and hence leading to different kinds of interactions between the proteins under study, one leading to augmentation in activity and the other leading to decrease in activity. Further, the CcmM protein of *G. violaceus* when analyzed for its effect on the two forms of RuBisCO used in the study, showed a similar effect. The catalytic rate of both the forms was found to decrease with increasing concentration of CcmM but the affinity for the substrate was found to be enhanced, indicating the role of certain residues of CcmM possibly having some role in binding RuBP around the catalytic site and hence making it more readily available for carboxylation.

CHAPTER 7

Conclusions and Future Scope

7.1 Conclusions

The conclusions drawn from the present work can be summarized as mentioned below in the order of their appearance in the thesis.

7.1.1. Distribution and Diversity of Carbon Concentrating Mechanism Proteins of β -Cyanobacteria

The analysis was conducted to find out the presence/absence of the carboxysome forming proteins across various phyla of Eubacteria in order to trace their evolutionary path. The analysis was conducted using the CCM proteins of *G. violaceus* PCC 7421, early diverging cyanobacteria. The CCM protein homologs were found in various eubacterial phyla.

- The shell proteins viz. CcmK, CcmL and CcmO homologs were found in Actinobacteria, Fibrobacteres, Fusobacteria, Planctomycetes, α Proteobacteria, β Proteobacteria and γ Proteobacteria. It is to be noted that these phyla are devoid of homologs for CcmM and CcmN proteins as they have α -carboxysomes. Only Cyanobacteria were found to have complete set of β -carboxysome constituting proteins.
- The close homology of the shell proteins by virtue of the BMC domains supports the fact the shell proteins of carboxysomes are evolutionarily linked to shell proteins of microcompartments involved in Ethanolamine utilization and Propanediol utilization pathways.
- Further, the absence of any homolog for the CcmM and CcmN proteins across Eubacteria suggests that the two proteins have possibly originated by domain shuffling (needs to be proven) between the carbonic anhydrase protein and the RuBisCO small subunit protein domains (both carbonic anhydrase and RuBisCO exist in carboxysomes).
- The congruency in the phylogenetic arrangement according to 16S rRNA and 16S rRNA-rbcL-pyrH with that of ccm/cso operon implies *in-situ* formation of the later and its vertical succession. Contrary to as expected, *G. violaceus* was not found to be ancestral with respect to CCM proteins.
- The origin of *G. violaceus* PCC 7421 could possibly be dated back to the time of origin of oxygenic photosynthesis. Although being ancestral the CCM seems to have originated in some other intermediate/ancestral organism, passed on to *G. violaceus* by horizontal gene transfer and also to further cyanobacterial lineage formed in vertical succession.

- The CcmK proteins of the β -cyanobacteria when analyzed were found to be present in several copies, emphasizing the importance of the protein in shell formation. Further, it was found that the number of CcmK encoding genes has increased along the line of evolution, as the early diverging cyanobacteria had the minimal number, which was passed on vertically and then subsequent duplication events took place.

7.1.2. Cloning and Characterization of RuBisCO Encoding Genes of *G. violaceus* PCC 7421

The RuBisCO sequences from both α and β -cyanobacteria were analyzed for conservation of sequence, arrangement of genes in the operon and degree of co-evolution. Further, the RuBisCO from early diverging cyanobacterium *G. violaceus* was characterized.

- The RbcL sequence was found to be most conserved, ~90% across all cyanobacteria. The proteins active site residues remain conserved even across phyla and different forms of the RuBisCO, suggesting the importance of the arrangement of residues involved in catalysis.
- The RbcS and RbcX sequences in the contrary were found to be less conserved, ~17% and 12%, respectively. All the β -cyanobacteria were found to possess RuBisCO chaperone but several α cyanobacteria were found to lack the same.
- The RuBisCO encoding sequences of β -cyanobacteria appears to have co-evolved in most organisms analyzed in present study, but the same could not be observed for α -cyanobacterial RuBisCO operon, although the degree of variation in the later was very low.
- RbcX helps in proper folding of RuBisCO by interacting with RbcL. However, it is not an absolute requirement for RuBisCO to attain proper folding only with the aid of RbcX. The present analysis led to the finding that cyanobacterial species lacking RbcX contain multitude of protein showing homology to chaperone like proteins. These proteins might be playing the same role as RbcX in these cyanobacterial species to help RuBisCO acquiring proper folding. Analyses also indicated that in general the *rbcx* gene to be present between *rbcl* and *rbcs*.
- The comparison of form I RuBisCO (with L₈S₈ arrangement) from diverse organisms shows conservation of the residues responsible for catalysis and yet a variation in characteristics i.e. specificity and turnover exists. The intriguing, unidentified properties of the protein make it unique even among the closely related organisms. *G. violaceus* PCC 7421, an early diverging cyanobacterium with unique characteristics and an exceptionally slow growth rate is an important organism in studies involving the evolution of photosynthesis. The present study is the first report on RuBisCO

from *G. violaceus* being expressed in *E. coli*. The protein, unlike in other cyanobacteria failed to express in a soluble form and hence assemble as a holoenzyme *in-vivo*. The protein purified yielded very low activity (0.005 μ Moles of PGA/min/mg of protein) which possibly is the actual activity of the protein supported by the fact that the organism has a slow growth rate or the renaturation procedure was incapable of assisting native folding of the protein. Further, the study reports existence of some signal in the intergenic region between *rbcL* and *rbcX* that caused inhibition of transcription of the downstream genes in *E. coli* which otherwise is not inhibitory in *G. violaceus*. The study opens quests for identification of factors responsible for assembly of RuBisCO apart from RbcX, which is solely responsible for folding RuBisCO in some; ineffective by itself in others and not required at all in the rest of form I RuBisCO studied till date.

- The specific activity of *G. violaceus* PCC 421 RuBisCO LSU at 2mM RuBP concentration was found to be 0.005 μ Moles/min/mg. The activity of the protein increased with an increase in RuBP concentration with the K_M RuBP value of 0.1917mM and V_{max} value of 0.0002 per minute. The turnover number of the protein was found to be 0.0029/sec. Further, RbcS of *G. violaceus* was found to augment the activity of the large subunit of RuBisCO of *G. violaceus*. The specific activity observed after complementation was 0.0196 μ Moles/min/mg. The K_M RuBP of the protein for the LSU+SSU complex also improved to 0.102mM and the V_{max} of the protein complex was found to increase up to 0.011/minute. The turnover number of the protein complex was found to be 0.016/second.

7.1.3. Cloning and Characterization of CCM Encoding Genes of *G. violaceus* PCC 7421

The CCM proteins from β cyanobacteria were analyzed for conserved regions and the genes for the same from *G. violaceus* were cloned and expressed in *E. coli*.

- Regions of CcmK, CcmL and CcmO apart from already reported were identified to be conserved in all homologs of the protein indicating their involvement in some important function, possibly the interactions with the lumen proteins of carboxysomes.
- Further, the internal ribosomal entry site for CcmM of *G. violaceus* was estimated by *in-silico* analysis. The IRES identified was used to predict the short for of the CcmM protein from *G. violaceus* and hence its tertiary structure by homology modeling and threading. The conservation of residues important for carbonic anhydrase activity were found to be conserved in CcmM of *G. violaceus*. The BLASTP of CcmM shows

greater homology of the protein with the RbcS of other organisms like *Leptolyngbya* than its own RbcS. The possible reason could be because of a common origin of the protein. Such homology represents an evolutionary link between the two proteins.

- The alignment of the CcmN proteins from β -cyanobacteria revealed conservation of regions trimeric LpxA-like domain and the C terminal short peptide, suggested to be involved in interaction with lumen proteins of carboxysomes. CcmO has conserved residues similar to those of CcmK, hence bringing forth the functional or evolutionary or both kind of relationship between the two proteins.
- All the CCM proteins were successfully cloned into expression vector and upon induction in the presence of IPTG lead to the expression of expected size of proteins viz. ~13kDa, ~13kDa, ~71kDa, ~29kDa and ~27kDa for CcmK, CcmL, CcmM, CcmN and CcmO, respectively. The tertiary structures for CcmK, CcmM IRES1, CcmN and CcmO were predicted by homology modeling.
- The CcmM protein of *G. violaceus* was successfully purified by affinity chromatography under native conditions. The purified protein was used in complementation studies mentioned in the next section. The protein sequence was validated by mass spectrometry.

7.1.4. The effect of CcmM and RbcS of *G. violaceus* on catalytic activity of a form I (*G. violaceus*) and form II (*Rhodospirillum rubrum*) RuBisCO large subunit

- The *R. rubrum* RuBisCO was purified and analyzed for its kinetics in the present study for firstly, to test the accuracy of the purification procedures as well as the RuBisCO assay and secondly, to determine the effect foreign RuBisCO SSU on the activity of the same. The specific activity, K_m RuBP, V_{max} and catalytic turnover rate for *R. rubrum* RuBisCO was found to be 270.003 μ Moles of PGA/ min/mg of protein, 0.0490mM, 0.5737/minute and 10 per second, respectively. The comparison with already available data shows that the results obtained in present study are in accordance with those reported earlier. Hence the purification procedures and the RuBisCO activity estimated can be considered dependable for determining the same for unknown proteins.
- The effect of RuBisCO small subunit of *G. violaceus* on large subunit of *R. rubrum* RuBisCO when analyzed suggested that the interaction between Form II LSU and Form I SSU does take place but is different from that observed in Form I RuBisCO. The enzymatic assay performed in the present study showed a decrease in activity of *R. rubrum* RuBisCO in presence of increasing concentration of a foreign SSU. This could possibly be firstly, because of overcrowding by the SSU around the LSU

making the active site inaccessible to the reaction metabolites or secondly because of binding of CO₂ to SSU, which did not interact with LSU as much as expected/interacted in a manner different from expected and hence created a Ci deficient environment.

- Further, the complementation of *G. violaceus* large subunit with the *G. violaceus* small subunit in the presence of the chaperone protein RbcX, showed augmentation of the activity of the former. The results show a stable interaction between RbcL and RbcS of *G. violaceus* RuBisCO form I. While as mentioned earlier the activity of the Form II RuBisCO large subunit was found to diminish in the presence of increasing concentration of the foreign small subunit of RuBisCO. The results suggest that small subunit of RuBisCO does cause an augmentation in the activity of the large subunit, but its activity is specific. Since ~50% residues reported to involve in LSU-SSU interaction were not found to be conserved in the large subunit of *R. rubrum*, hence possibly the loss in activity could be attributed to the formation of a complex not very suitable for the carboxylation catalysis.
- As per the *in-vitro* and *in-silico* analysis, the CcmM protein of *G. violaceus* does form stable complex with *R. rubrum* RuBisCO (form II) as well as with RbcL of *G. violaceus*. The complexes formed are via residues different from the most conserved residues in the SSU domains. As from the *in-vitro* studies the protein upon its interaction with the large subunit is found to cause a decrease in activity, possibly due to blockage of the active site of the RbcL. Further, CcmM possibly has affinity for RuBP, causing it to concentrate around the active site of the large subunit and hence causing a decrease in the effective K_M RuBP of the large subunit. Further, some residues of CcmM IRES1 viz. Arg92, Leu189, Arg91, Glu256 and Ser299 have been identified to be involved in interaction with both the form I and form II RuBisCO large subunits.

7.2 Future Scope

The present study opens quest for carrying out further investigations on several fronts. The possible future ventures enrooting from the conclusions drawn from aforementioned analysis have been listed below.

- The CCM constructs generated could be used for assembly of carboxysome in *E. coli*. The residues identified to be conserved in various CCM proteins could be analyzed for their role in carboxysome formation or function by generation of deletion constructs or chimeric constructs.
- The CcmM IRES1 identified in the present study can be analyzed for its protein-protein interaction with RuBisCO by *in vitro/in vivo* analysis.

- The carboxysome if found to assemble successfully in *E. coli* can be tested for its assembly in plant system by cloning into suitable plant transformation vector and the effect of the same on growth and development of plant hence analyzed.
- Further, studies could be carried out to identify other important proteins or regulatory factors that need to be introduced along with the CCM so as to achieve an enhancement in the plant productivity.
- In order to establish the effect of SSU repeats of CcmM protein on the activity of type I and II RuBisCO, deletion variants could be constructed containing only the SSU repeat terminal of the protein in combination with the complete protein, *in-vitro*. The analysis in the present study showed that the SSU repeats in CcmM of one cyanobacteria were more closely related to SSU RuBisCO of other cyanobacteria, rather than its own. This finding could be further investigated in order to add to the knowledge of the evolution of CCM proteins.
- The *G. violaceus* RuBisCO found to be insoluble *in-vivo* upon overexpression, despite co-expressing chaperone (RbcX) and using sub-optimal growth conditions, made renaturation of the denatured protein as the only resort to carrying out the study. Possibly, the co-expression of GroEL and GroES could augment the solubility of the protein *in-vivo*. Further, the absence of the RbcS and RbcX protein from the overexpressed operon construct could be overcome by using a duet vector. These alterations could improve the suitability of the conditions for expression of the complete operon and the folding of the protein into an active holoenzyme. These experimentations will give an insight into the role of various proteins in the folding of the RuBisCO holoenzyme and hence an opportunity to improve the catalysis.
- The *in-silico* generated CCM protein models could be analyzed for protein protein interactions and the same can be validated by *in-vitro/in-vivo* analysis.

References

- Abdul-Rahman F, Petit E, Blanchard JL (2013) The distribution of polyhedral bacterial microcompartments suggests frequent horizontal transfer and operon reassembly. *J Phylogen Evolution Biol* 1(4):1-7.
- Alber BE and Ferry JG (1996) Characterization of heterologously produced carbonic anhydrase from *Methanosarcina thermophila*. *J Bacteriol* 178:3270-3274.
- Alonso H, Blayney MJ, Beck JL, Whitney SM (2009) Substrate-induced assembly of *Methanococcoides burtonii* d-Ribulose-1,5-bisphosphate carboxylase/oxygenase dimers into decamers. *J Biol Chem* 284:33876-33882.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403-410.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389-3402.
- Andersson I (2008) Catalysis and regulation in RuBisCO. *J Exp Bot* 59:1555-1568.
- Andersson I and Backlund A (2008) Structure and Function of RuBisCO. *Plant Physio Biochem* 46(3):275-291.
- Andersson I, Knight S, Schneider G, Lindqvist Y, Lundqvist T, Brändén CI, Lorimer GH (1989) Crystal structure of the active site of ribulose-bisphosphate carboxylase. *Nature* 337:229-234.
- Avni A, Edelman M, Rachailovich I, Aviv D, Fluhr R (1989) A point mutation in the gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/ oxygenase affects holoenzyme assembly in *Nicotiana tabacum*. *EMBO J* 8:1915-1918.
- Badger M, Andrews TJ, Whitney SM, Ludwig M, Yellowlees DC, Leggat W, Price GD (1998) The diversity and coevolution of RuBisCO, plastids, pyrenoids, and chloroplast based CO₂ concentrating mechanisms in algae. *Canadian J Bot* 76(6):1052-1071.
- Badger MR and Bek EJ (2008) Multiple RuBisCO forms in proteobacteria: their functional significance in relation to CO₂ acquisition by the CBB cycle. *J Exp Bot* 59:1525-1541.

References

- Abdul-Rahman F, Petit E, Blanchard JL (2013) The distribution of polyhedral bacterial microcompartments suggests frequent horizontal transfer and operon reassembly. *J Phylogen Evolution Biol* 1(4):1-7.
- Alber BE and Ferry JG (1996) Characterization of heterologously produced carbonic anhydrase from *Methanosarcina thermophila*. *J Bacteriol* 178:3270-3274.
- Alonso H, Blayney MJ, Beck JL, Whitney SM (2009) Substrate-induced assembly of *Methanococcoides burtonii* d-Ribulose-1,5-bisphosphate carboxylase/oxygenase dimers into decamers. *J Biol Chem* 284:33876-33882.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403-410.
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389-3402.
- Andersson I (2008) Catalysis and regulation in RuBisCO. *J Exp Bot* 59:1555-1568.
- Andersson I and Backlund A (2008) Structure and Function of RuBisCO. *Plant Physio Biochem* 46(3):275-291.
- Andersson I, Knight S, Schneider G, Lindqvist Y, Lundqvist T, Brändén CI, Lorimer GH (1989) Crystal structure of the active site of ribulose-bisphosphate carboxylase. *Nature* 337:229-234.
- Avni A, Edelman M, Rachailovich I, Aviv D, Fluhr R (1989) A point mutation in the gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/ oxygenase affects holoenzyme assembly in *Nicotiana tabacum*. *EMBO J* 8:1915-1918.
- Badger M, Andrews TJ, Whitney SM, Ludwig M, Yellowlees DC, Leggat W, Price GD (1998) The diversity and coevolution of RuBisCO, plastids, pyrenoids, and chloroplast based CO₂ concentrating mechanisms in algae. *Canadian J Bot* 76(6):1052-1071.
- Badger MR and Bek EJ (2008) Multiple RuBisCO forms in proteobacteria: their functional significance in relation to CO₂ acquisition by the CBB cycle. *J Exp Bot* 59:1525-1541.

- Badger MR and Price GD (2003) CO₂ concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution. *J Exp Bot* 54:609-622.
- Badger MR, Hanson DT, Price GD (2002) Evolution and diversity of CO₂ concentrating mechanisms in cyanobacteria. *Funct Plant Biol* 29:161-173.
- Baker SH, Jin SM, Aldrich HC, Howard GT, Shively JM (1998) Insertion mutation of the form I *cbhL* gene encoding ribulose biphosphate carboxylase/oxygenase (RuBisCO) in *Thiobacillus neapolitanus* results in expression of form II RuBisCO, loss of carboxysomes, and an increased CO₂ requirement for growth. *J Bacteriol* 180:4133-4139.
- Behrenfeld MJ, Randerson JT, McClain CR, Feldman GC, Los SO, Tucker CJ et al. (2001) Biospheric primary production during an ENSO transition. *Science* 291:2594-2597.
- Benkert P, Tosatto SCE and Schomburg D (2008) QMEAN: A comprehensive scoring function for model quality assessment. *Proteins: Structure, Function and Bioinformatics*. 71(1):261-277.
- Bernát G, Schreiber U, Sendtko E, Stadnichuk IN, Rexroth S, Rögner M, Koenig F (2012) Unique properties vs. Common themes: The atypical cyanobacterium *Gloeobacter violaceus* PCC 7421 is capable of state transitions and blue-light-induced fluorescence quenching. *Plant Cell Physiol* 53:528–542.
- Berner RA, Kothavala Z (2001) GEOCARB III: a revised model of atmospheric CO₂ over phanerozoic time. *Am J Sci* 301:182-204.
- Birnboim HC and Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* 24: 7(6):1513-1523.
- Blank CE and Sanchez-Baracaldo P (2010) Timing of morphological and ecological innovations in the cyanobacteria—a key to understanding the rise in atmospheric oxygen. *Geobiology* 8:1-23.
- Bowes G, Rao SK, Estavillo GM, Reiskind JB (2002) C₄ mechanisms in aquatic angiosperms: comparisons with terrestrial C₄ systems. *Funct Plant Biol* 29:379-392.
- Bowie JU, Luthy R and Eisenberg D (1991) A method to identify protein sequences that fold into a known three-dimensional structure. *Science* 253(5016):164-170.
- Bryant DA, Cohen-Bazire G, Glazer AN (1981) Characterization of the biliproteins of *Gloeobacter violaceus* chromophore content of a cyanobacterial phycoerythrin carrying

phycourobilin chromophore. *Arch Microbiol* 129:190–198.

Cai F, Kerfeld CA, Sandh G (2012) Bioinformatic identification and structural characterization of a new carboxysome shell protein, p 345–356. *In* Burnap R, Vermaas W (ed), *Functional genomics and evolution of photosynthetic systems*, vol 33. Springer, Dordrecht, Netherlands.

Cai F, Sutter M, Bernstein SL, Kinney JN, Kerfeld CA (2015) Engineering Bacterial Microcompartment Shells: Chimeric Shell Proteins and Chimeric Carboxysome Shells. *ACS Synth Biol* 4(4):444–453.

Cameron, JC, Wilson, SC, Bernstein, SL, Kerfeld CA (2013) Biogenesis of a bacterial organelle: The carboxysome assembly pathway. *Cell* 155:1131–1140.

Cannon GC, Baker SH, Soyer F, Johnson DR, Bradburne CE, Mehlman JL et al. (2003) Organization of carboxysome genes in the thiobacilli. *Curr Microbiol* 46:115–119.

Cannon GC, Bradburne CE, Aldrich HC, Baker SH, Heinhorst S, Shively JM (2001) Microcompartments in prokaryotes: carboxysomes and related polyhedra. *Appl. Environ. Microbiol* 67:5351–5361.

Cannon GC, Heinhorst S, Bradburne CE, Shively JM (2002) Carboxysome genomics: a status report. *Funct Plant Biol* 29:175–182.

Chen P, Andersson D, Roth JR (1994) The control region of the pdu/cob regulon in *Salmonella typhimurium*. *J Bacteriol* 176:5474–5482.

Clark, E (1998) Refolding of recombinant proteins TL. *Curr Opin Biotechnol* 9 VN-re:157–163.

Cloney LP, Bekkaoui DR, Hemmingsen SM (1993) Co-expression of plastid chaperonin genes and a synthetic plant RuBisCO operon in *Escherichia coli*. *Plant Mol Biol* 23:1285–1290.

Cohen SN, Chang ACY, Hsu L (1972) Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA. *Proc Nat Acad Sci USA* 69:2110.

Colangeli R, Heijbel A, Williams AM, Manca C, Chan J, Lyashchenko K, Gennaro ML (1998) Three-step purification of lipopolysaccharide-free, polyhistidine-tagged recombinant

- Cook CM, Lanaras T, Wood AP, Codd GA, Kelly DP (1991) Kinetic properties of ribulose biphosphate carboxylase/oxygenase from *Thiobacillus thyasiris*, the putative symbiont to *Thyasira flexuosa* (Montagu), a bivalve mussel. *J Gen Microbiol* 137:1491-1496.
- Cot SSW, So AKC, Espie GS (2008) A multiprotein bicarbonate dehydration complex essential to carboxysome function in cyanobacteria. *J Bacteriol* 190:936-945.
- Criscuolo A and Gribaldo S (2011) Large-scale phylogenomic analyses indicate a deep origin of primary plastids within cyanobacteria. *Mol. Biol. Evol* 28:13019-13032.
- Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: A sequence logo generator. *Genome Research* 14:1188-1190.
- Crowley CS, Cascio D, Sawaya MR, Kopstein JS, Bobik TA, Yeates TO (2010) Structural insight into the mechanisms of transport across the *Salmonella enterica* Pdu microcompartment shell. *J Biol Chem* 285:37838-37846.
- Day ES, Cote SM and Whitty A (2012) Binding Efficiency of Protein-Protein Complexes. *Biochemistry* 51(45): 9124–9136.
- de Beer TAP, Berka K, Thornton JM, Laskowski RA (2014) PDBsum additions. *Nucleic Acids Res* 42:D292-D296.
- Des Marais DJ (2000) Evolution: When did photosynthesis emerge on Earth? *Science* 289:1703-1705.
- Dobranski KP, Longo DL, Scott KM (2005) A hydrothermal vent chemolithoautotroph with a carbon concentrating mechanism. *J Bacteriol* 187:5761-5766.
- Dubbs JM and Tabita FR (2004) Regulators of nonsulfur purple phototrophic bacteria and the interactive control of CO₂ assimilation, nitrogen fixation, hydrogen metabolism and energy generation. *FEMS Microbiol Rev* 28:353-376.
- Edwards GE, Franceschi VR, Voznesenskaya EV (2004) Single-cell C₄ photosynthesis versus the dual – cell (Kranz) paradigm. *Annu Rev Plant Biol* 55:173-196.
- Ellis J (2010) Biochemistry: Tackling unintelligent design. *Nature* 463:164-165.
- Enami I, Suzuki T, Tada O, Nakada Y, Nakamura K, Tohri A, Ohta H, Inoue I, Shen JR (2005) Distribution of the extrinsic proteins as a potential marker for the evolution of

Enright AJ, Iliopoulos I, Kyrpides NC, Ouzounis CA (1999) Protein interaction maps for complete genomes based on gene fusion events. *Nature* 402:86-90.

Espie GS and Kimber MS (2011) Carboxysomes: cyanobacterial RuBisCO comes in small packages. *Photosynth Res* 109:7-20.

Fan CG, Cheng SQ, Sinha S, Bobik TA (2012) Interactions between the termini of lumen enzymes and shell proteins mediate enzyme encapsulation into bacterial microcompartments. *Proc Natl Acad Sci USA* 109:14995-15000.

Finn MW and Tabita FR (2004) Modified pathway to synthesize ribulose 1,5-bisphosphate in methanogenic archaea. *J Bacteriol* 186:6360-6366.

Finn MW and Tabita FR (2004) Modified Pathway To Synthesize Ribulose 1,5-Bisphosphate in Methanogenic Archaea. *J Bacteriol* 186:6360-6366.

Frank S, Lawrence AD, Prentice MB, Warren MJ (2013) Bacterial microcompartments moving into a synthetic biological world. *J Biotechnol* 163: 273-279.

Fukuzawa H, Fujiwara S, Yamamoto Y, Dionisio-Sese ML, Miyachi S (1990) cDNA cloning, sequence, and expression of carbonic anhydrase in *Chlamydomonas reinhardtii*: regulation by environmental CO₂ concentration. *Proc Natl Acad Sci USA* 87:4383-4387.

Gaurav V, David JL, Rudolf M (2010) SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics* 27(2):171-180.

Geer LY, Marchler-Bauer A, Geer RC, Han L, He J, He S, Liu C, Shi W, Bryant SH (2010) The NCBI BioSystems database. *Nucleic Acids Res* D492-6.

Geitler L (1927) Neue Blaualgen aus Lunz. *Archiv für Protistenkunde* 60:440-448.

Gibbs SP (1962) The ultrastructure of the pyrenoids of greenalgae. *J Ultrastructure Res* 7:262-272.

Giordano M, Beardall, Raven JA (2005) CO₂ concentrating mechanisms in algae: mechanisms, environmental modulation, and evolution. *Annu Rev Plant Biol* 56:99-131.

Giraud E, Hannibal L, Fardoux J, Vermeir AB, Dreyfus (2000) Effect of *Bradyrhizobium* photosynthesis on stem nodulation of *Aeschynomene* sensitive. *Proc Natl Acad Sci U S A* 97:14795-14800

- Goffredi SK, Childress JJ, Desaulniers NT, Lee RW, Lallier FH, Hammond D (1997) Sulfide acquisition by the vent worm *Riftia pachyptila* appears to be via uptake of HS⁻, rather than H₂S. *J Exp Biol* 200:2609-2616.
- Goloubinoff P, Christeller JT, Gatenby AA, Lorimer GH (1989) Reconstitution of active dimeric ribulose biphosphate carboxylase from an unfolded state depends on two chaperonin proteins and Mg-ATP. *Nature* 342, 884-889.
- Gonzales AD, Light YK, Zhang Z, Iqbal T, Lane TW, Martino A (2005) Proteomic analysis of the CO₂-concentrating mechanism in the open-ocean cyanobacterium *Synechococcus* WH8102. *Can J Bot* 83:735-745.
- Greene DN, Whitney SM, Matsumura I (2007) Artificially evolved *Synechococcus* PCC6301 RuBisCO variants exhibit improvements in folding and catalytic efficiency. *Biochem J* 404:517-524.
- Gubernator B, Bartoszewski R, Kroliczewski J, Wildner G, Szczepaniak A (2008) Ribulose-1,5-bisphosphate carboxylase/oxygenase from thermophilic cyanobacterium *Thermosynechococcus elongatus*. *Photosynth Res* 95:101-109.
- Guglielmi G, Cohen-Bazire G, Bryant DA (1981) The structure of *Gloeobacter violaceus* and its phycobilisomes. *Arch Microbiol* 129:181-189.
- Gupta RS and Mathews DW (2010) Signature proteins for the major clades of Cyanobacteria. *BMC Evol Biol* 10:24.
- Gurevitz M, Somerville CR, McIntosh L (1985) Pathway of assembly of ribulosebisphosphate carboxylase/oxygenase from *Anabaena* 7120 expressed in *Escherichia coli*. *Proc Natl Acad Sci U. S. A.* 82:6546-6550.
- Gutteridge S and Pierce J (2006) A unified theory for the basis of the limitation of the primary reaction of photosynthetic CO₂ fixation: Was Dr. Pangloss right? *Proc Natl Acad Sci U S A* 103: 7203-7204.
- Hansgirg A (1892) Prodröm der Algenflora von Böhmen. *Arch Naturwiss Landesdurchforsch Böhmen* 8:1-268.
- Hartman FC and Harpel MR (1994) Structure, function, regulation, and assembly of D-ribulose-1,5-bisphosphate carboxylase oxygenase. *Ann Rev Biochem* 63:197-234.

- Hartman FC, Stringer CD, Milanez S, Lee EH (1986) The Active Site of RuBisCO. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 313:1162; 379-395.
- Havemann GD and Bobik TA (2003) Protein content of polyhedral organelles involved in coenzyme B12-dependent degradation of 1,2-propanediol in *Salmonella enterica serovar typhimurium* LT2. *J Bacteriol* 185:5086-5095.
- Heinhorst S, Cannon GC, Shively JM (2006) Carboxysomes and carboxysome-like inclusions, p 141–165. *In* Shively J (ed), *Complex intracellular structures in prokaryotes*, vol 2. Springer, Berlin, Germany.
- Hoffmann L, Kas̆tovský J, Komařek J (2005) System of cyanoprokaryotes (cyanobacteria) – state in 2004. *Arch Hydrobiol Suppl Algal Stud* 117:121.
- Huseby DL and Roth JR (2013) Evidence that a metabolic microcompartment contains and recycles private cofactor pools. *J Bacteriol* 195:2864-2879.
- Iancu CV, Morris DM, Dou ZC, Heinhorst S, Cannon GC, Jensen GJ (2010) Organization, structure, and assembly of alpha-carboxysomes determined by electron cryotomography of intact cells. *J Mol Biol* 396:105-117.
- Iverson TM, Alber BE, Kisker C, Ferry JG, Rees DC (2000) A closer look at the active site of gamma-class carbonic anhydrases: high-resolution crystallographic studies of the carbonic anhydrase from *Methanosarcina thermophila*. *Biochemistry* 39(31):9222-31.
- Jan Mares, Hrouzek P, Kana R, Ventura S, Strunecky O, Komarek J (2013) The primitive thylakoid-less cyanobacterium *Gloeobacter* is a common rock-dwelling organism. *Plos one* doi:10.1371/journal.pone.0066323.
- Jensen T and Bowen C (1961) Organization of the centropiasm in *Nostoc punctiforme*. *Proc Iowa Acad Sci* 68:89-96.
- Jordan DB and Ogren WL (1981) Species variation in the specificity of ribulose biphosphate carboxylase/oxygenase. *Nature* 291:513-515.
- Jukes TH and Cantor CR (1969) Evolution of protein molecules. *In* Munro HN, editor, *Mammalian Protein Metabolism*, pp. 21-132, Academic Press, New York.

- Kane HJ, Viil J, Entsch B, Paul K, Morell MK, Andrews TJ (1994) An improved method for measuring the CO₂/O₂ specificity of ribulose bisphosphate carboxylase/oxygenases. *Aust J Plant Physiol* 21:449-461.
- Kaneko Y, Danev R, Nagayama K, Nakamoto H (2006) Intact carboxysomes in a cyanobacterial cell visualized by Hilbert differential contrast transmission electron microscopy. *J Bacteriol* 188:805-808.
- Kaplan A and Reinhold L (1999) CO₂ concentrating mechanisms in photosynthetic organisms. *Annu Rev Plant Physiol Plant Mol Biol* 50:539-570.
- Karkehabadi S, Taylor TC, Andersson I (2003) Calcium supports loop closure but not catalysis in Rubisco. *J Mol Biol* 334:65-73.
- Kelly D, Harrison A. Genus *Thiobacillus*, p 1842-1858. In Staley JT, Bryant MP (ed), *Bergey's manual of systematic bacteriology*, vol 3. Williams and Wilkins, Baltimore, MD. 1989.
- Kelly DP and Wood AP (2000) Reclassification of some species of *Thiobacillus* to the newly designated genera *Acidithiobacillus* gen nov., *Halothiobacillus* gen. nov and *Thermothiobacillus* gen. nov. *Int J Syst Evol Microbiol.* 50: 511-516.
- Kerfeld CA, Heinhorst S, Cannon GC (2010) Bacterial microcompartments. *Annu Rev Microbiol* 64:391-408.
- Kerfeld CA, Sawaya MR, Tanaka S, Nguyen CV, Phillips M, Beeby M, Yeates TO (2005) Protein structures forming the shell of primitive bacterial organelles. *Science* 309:936-938.
- Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16:111-120.
- Kinney JN, Axen SD, Kerfeld CA (2011) Comparative analysis of carboxysome shell proteins. *Photosynth Res* 109:21-32.
- Kinney JN, Salmeen A, Cai F, Kerfeld CA (2012) Elucidating essential role of conserved carboxysomal protein CcmN reveals common feature of bacterial microcompartment assembly. *J Biol Chem* 287:17729-17736.
- Kisker C, Schindelin H, Alber BE, Ferry JG, Rees DC (1996) A left-hand beta-helix revealed by the crystal structure of a carbonic anhydrase from the archaeon *Methanosarcina thermophila*. *EMBO J* 15: 2323-2330.

- Kane HJ, Viil J, Entsch B, Paul K, Morell MK, Andrews TJ (1994) An improved method for measuring the CO₂/O₂ specificity of ribulose biphosphate carboxylase/oxygenases. *Aust J Plant Physiol* 21:449-461.
- Kaneko Y, Danev R, Nagayama K, Nakamoto H (2006) Intact carboxysomes in a cyanobacterial cell visualized by Hilbert differential contrast transmission electron microscopy. *J Bacteriol* 188:805-808.
- Kaplan A and Reinhold L (1999) CO₂ concentrating mechanisms in photosynthetic organisms. *Annu Rev Plant Physiol Plant Mol Biol* 50:539-570.
- Karkehabadi S, Taylor TC, Andersson I (2003) Calcium supports loop closure but not catalysis in Rubisco. *J Mol Biol* 334:65-73.
- Kelly D, Harrison A. Genus *Thiobacillus*, p 1842-1858. In Staley JT, Bryant MP (ed), *Bergey's manual of systematic bacteriology*, vol 3. Williams and Wilkins, Baltimore, MD. 1989.
- Kelly DP and Wood AP (2000) Reclassification of some species of *Thiobacillus* to the newly designated genera *Acidithiobacillus* gen nov., *Halothiobacillus* gen. nov and *Thermothiobacillus* gen. nov. *Int J Syst Evol Microbiol.* 50: 511-516.
- Kerfeld CA, Heinhorst S, Cannon GC (2010) Bacterial microcompartments. *Annu Rev Microbiol* 64:391-408.
- Kerfeld CA, Sawaya MR, Tanaka S, Nguyen CV, Phillips M, Beeby M, Yeates TO (2005) Protein structures forming the shell of primitive bacterial organelles. *Science* 309:936-938.
- Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16:111-120.
- Kinney JN, Axen SD, Kerfeld CA (2011) Comparative analysis of carboxysome shell proteins. *Photosynth Res* 109:21-32.
- Kinney JN, Salmeen A, Cai F, Kerfeld CA (2012) Elucidating essential role of conserved carboxysomal protein CcmN reveals common feature of bacterial microcompartment assembly. *J Biol Chem* 287:17729-17736.
- Kisker C, Schindelin H, Alber BE, Ferry JG, Rees DC (1996) A left-hand beta-helix revealed by the crystal structure of a carbonic anhydrase from the archaeon *Methanosarcina thermophila*. *EMBO J* 15: 2323-2330.

- Knight S, Andersson I, Brändén CI (1989) Re-examination of the three-dimensional structure of the small subunit of RuBisCO from higher plants. *Science* 244:702-705.
- Knight S, Andersson I, Brändén CI (1990) Crystallographic analysis of ribulose 1,5-bisphosphate carboxylase from spinach at 2.4 Å resolution: subunit interactions and the active site. *J Mol Biol* 215:113-160.
- Koay T, Wong HL, Lim BH (2016) Engineering of Chimeric Eukaryotic/Bacterial Rubisco Large Subunits in *Escherichia coli*. *Genes Genet Syst* 91(3):139-150.
- Koenig T, Menze BH, Kirchner M, Monigatti F, Parker KC, Patterson T, Steen JJ, Hamprecht FA, Steen H (2008) Robust prediction of the MASCOT score for an improved quality assessment in mass spectrometric proteomics. *J Proteome Res* 7(9):3708-3717.
- Kofoed E, Rappleye C, Stojiljkovic I, Roth J (1999) The 17-gene ethanolamine (eut) operon of *Salmonella typhimurium* encodes five homologues of carboxysome shell proteins. *J Bacteriol* 181:5317-5329.
- Komařek J and Anagnostidis K (1999) Cyanoprokaryota. 1. Teil: Chroococcales. Süßwasserflora von Mitteleuropa. Heidelberg, Berlin: Spektrum Akademischer Verlag GmbH. 548.
- Konhäuser KO, Pecoits E, Lalonde SV, Papineau D, Nisbet EG, Barley ME, Arndt NT, Zahnle K, Kamber BS (2009) Oceanic nickel depletion and a methanogen famine before the Great Oxidation Event. *Nature* 458:750-753.
- Kreel NE and Tabita FR (2007) Substitutions at methionine 295 of *Archaeoglobus fulgidus* ribulose-1,5-bisphosphate carboxylase/oxygenase affect oxygen binding and CO₂/O₂ specificity. *J Biol Chem* 282:1341-1351.
- Krissinel E and Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372:774-797.
- Kucho K, Ohyama K, Fukuzawa H (1999) CO₂-responsive transcriptional regulation of CAH1 encoding carbonic anhydrase is mediated by enhancer and silencer regions in *Chlamydomonas reinhardtii*. *Plant Physiol* 121:1329-1337.
- Kusian B and Bowien B (1997) Organization and regulation of *cbh* CO₂ assimilation genes in autotrophic bacteria. *FEMS Microbiol Rev* 21:135-155.

- Landan G and Graur D (2008) Local reliability measures from sets of co-optimal multiple sequence alignments. *Pac Symp Biocomput* 13:15-24.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ and Higgins DG (2007) ClustalW and ClustalX version 2 *Bioinformatics*. 23(21):2947-2948.
- Laskowski R A (2007). Enhancing the functional annotation of PDB structures in PDBsum using key figures extracted from the literature. *Bioinformatics* 23:1824-1827.
- Le SQ and Gascuel O (1993) An Improved General Amino Acid Replacement Matrix. *Mol Biol Evol* 25(7):1307-1320.
- Li LA and Tabita FR (1997) Maximum activity of recombinant ribulose 1,5-bisphosphate carboxylase/oxygenase of *Anabaena* sp. strain CA requires the product of the *rbcX* gene. *J Bacteriol* 179:3793-3796.
- Liu C, Young AL, Starling-Windhof A, Bracher A, Saschenbrecker S, Rao BV, Rao KV, Berninghausen O, Mielke T, Hartl FU, Beckmann R, Hayer-Hartl M (2010) Coupled chaperone action in folding and assembly of hexadecameric Rubisco. *Nature* 463: 197–202.
- Lloyd J and Farquhar GD (1994) $\delta^{13}\text{C}$ discrimination during CO_2 assimilation by the terrestrial biosphere. *Oecologia* 99:201-215.
- Long BM, Badger MR, Whitney SM, Price GD (2007) Analysis of carboxysomes from *Synechococcus* PCC 7942 reveals multiple RuBisCO complexes with carboxysomal proteins CcmM and CcaA. *J Biol Chem* 282:29323-29335.
- Long BM, Price GD, Badger MR (2005) Proteomic assessment of an established technique for carboxysome enrichment from *Synechococcus* PCC 7942. *Can J Bot* 83:746-757.
- Long BM, Rae BD, Badger MR, Price GD (2011) Over-expression of the carboxysomal CcmM protein in *Synechococcus* PCC7942 reveals a tight co-regulation of carboxysomal carbonic anhydrase (CcaA) and M58 content. *Photosynth Res* 109:33-45.
- Long BM, Tucker L, Badger MR, Price GD (2010) Functional cyanobacterial carboxysomes have an absolute requirement for both long and short forms of the CcmM protein. *Plant Physiol.* 153:285-293.

- Meng J, Wang F, Wang F, Zheng Y, Peng X, Zhou H, Xiao X (2009) An uncultivated crenarchaeota contains functional bacteriochlorophyll a synthase. *ISME J* 3:106-116.
- Menon BB, Dou Z, Heinhorst S, Shively JM, Cannon GC (2008) *Halothiobacillus neapolitanus* carboxysomes sequester heterologous and chimeric RuBisCO species. *PLoS One* 3:e3570. doi:10.1371/journal.pone.0003570.
- Meyer MT, Genkov T, Skepper JN, Jouhet J, Mitchell MC, Spreitzer RJ, Griffiths H (2012) RuBisCO small-subunit α -helices control pyrenoid formation in *Chlamydomonas*. *Proc Nat Acad Sci USA* 109:19474-19479.
- Moreno J and Spreitzer RJ (1999) Cys-172 to Ser substitution in the chloroplast encoded large subunit affects stability and stress-induced turnover of ribulose biphosphate carboxylase/oxygenase. *J Biol Chem* 274:26789-26793.
- Morse D, Salois P, Markovic P and Hastings JW (1995) A nuclear-encoded form II RuBisCO in dinoflagellates. *Science* 268(5217):1622-1624.
- Nakamura Y, Kaneko T, Sato S, Mimuro M, Miyashita H, Tsuchiya T, Sasamoto S, Watanabe A, Kawashima K, Kishida Y, Kiyokawa C, Kohara M, Matsumoto M, Matsuno A, Nakazaki N, Shimpo S, Takeuchi C, Yamada M, Tabata S (2003) Complete Genome Structure of *Gloeobacter violaceus* PCC 7421, a Cyanobacterium that Lacks Thylakoids. *DNA Res* 10:137-145.
- Nei M and Kumar S (2000) *Molecular Evolution and Phylogenetics*. Oxford University Press, New York.
- Neilssen B, Van de Peer Y, Wilmotte A, De Wachter R (1995) An early origin of plastids within the cyanobacterial divergence is suggested by evolutionary trees based on complete 16S rRNA sequences. *Mol Biol Evol* 12:1166-1173.
- Neuwald AF, Aravind L, Spouge JL, Koonin EV (1999) AAA(+): a class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Gen Res* 9:27-43.
- Newman J and Gutteridge S (1994) Structure of an effector-induced inactivated state of ribulose 1,5-bisphosphate carboxylase/oxygenase: the binary complex between enzyme and xylulose 1,5-bisphosphate. *Structure* 2:495-502.

- Newman J, Branden CI, Jones TA (1993) Structure determination and refinement of RuBisCO. *Acta Crystallogr D Biol Crystallogr* 49:548-560.
- Notredame, Higgins, Heringa (2000) T-Coffee: A novel method for multiple sequence alignments. *JMB* 302:205-217.
- Orus MI, Rodriguez ML, Martinez F, Marco E (1995) Biogenesis and ultrastructure of carboxysomes from wild type and mutants of *Synechococcus* sp strain PCC 7942. *Plant Physiol* 107:1159-1166.
- Parry MAJ, Madgwick PJ, Carvalho JFC, and Andraloje PJ (2007) Prospects for increasing photosynthesis by overcoming the limitations of RuBisCO. *J Agric Sci* 145:31-43.
- Pearce FG and Andrews TJ (2003) The relationship between side reactions and slow inhibition of ribulose-biphosphate carboxylase revealed by a loop 6 mutant of the tobacco enzyme. *J Biol Chem* 278:32526-32536.
- Pena KL, Castel SE, Araujo C, Espie GS, Kimber MS (2010) Structural basis of the oxidative activation of the carboxysomal γ -carbonic anhydrase, CcmM. *Proc Natl Acad Sci USA* 107:2455-2460.
- Penrod JT and Roth JR (2006) Conserving a volatile metabolite: a role for carboxysome – like organelles in *Salmonella enteric*. *J Bacteriol* 188:2865-2874.
- Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20(18):3551-3567.
- Porath J, Carlsson J, Olsson I, Belfrage G (1975) Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* 258:598-599.
- Portis A (2003) RuBisCO activase: RuBisCO's catalytic chaperone. *Photosyn Res* 75:11-27.
- Portis AR, Li C, Wang D, Salvucci ME (2008) Regulation of RuBisCO activase and its interaction with RuBisCO. *J Exp Bot* 59.
- Price GD, Badger MR, von Caemmerer S (2011) The prospect of using cyanobacterial bicarbonate transporters to improve leaf photosynthesis in C₃ crop plants. *Plant Physiol* 155:20-26.

Price GD, Badger MR, Woodger FJ, Long BM (2008) Advances in understanding the cyanobacterial CO₂-concentrating mechanism (CCM): functional components, Ci transporters, diversity, genetic regulation and prospects for engineering into plants. *J Exp Bot* 59:1441-1461.

Price GD, Howitt SM, Harrison K, Badger MR (1993) Analysis of a genomic DNA region from the cyanobacterium *Synechococcus sp* str PCC 7942. *J Bacteriol* 175(10):2871-2879.

Price GD, Sültemeyer D, Klughammer B, Ludwig M, Badger MR (1998) The functioning of the CO₂ concentrating mechanism in several cyanobacterial strains—a review of general physiological characteristics, genes, proteins, and recent advances. *Can J Bot* 76:973-1002.

Price MN, Arkin AP, Alm EJ (2006) The life-cycle of operons. *PLoS Genet* 2(6):e96.

Pronina NA, Ramazanov ZM, Semenenko VE (1981) Carbonic anhydrase activity of *Chlorella* cells as a function of CO₂ concentration. *Sov Plant Physiol* 28:345-351.

Rae BD, Forster B, Badger MR, Price GD (2011) The CO₂-concentrating mechanism of *Synechococcus* WH5701 is composed of native and horizontally- acquired components. *Photosynth Res* 109:59-72.

Rae BD, Long BM, Badger MR, Price GD (2012) Structural determinants of the outer shell of β-carboxysomes in *Synechococcus elongatus* PCC 7942: roles for CcmK2, K3-K4, CcmO, and CcmL. *PLoS One* 7:e43871. doi:10.1371/journal.pone.0043871.

Rae BD, Long BM, Badger MR, Price GD (2013) Functions, compositions, and evolution of the two types of carboxysomes: Polyhedral microcompartments that facilitate CO₂ fixation in cyanobacteria and some proteobacteria. *Microbiol Mol Biol Rev* 77:357-379.

Raines CA (2011) Increasing photosynthetic carbon assimilation in C₃ plants to improve crop yield: Current and future strategies. *Plant Physiol* 155:36-42.

Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. *J Mol Biol* 7:95-99.

Ramage RT, Read BA, Tabita FR (1998) Alteration of the helix region of cyanobacterial ribulose 1,5- bisphosphate carboxylase/oxygenase to reflect sequences found in high substrate specificity enzymes. *Arch Biochem Biophys* 349:81-88.

Raven JA (1997) CO₂ -concentrating mechanisms: a direct role for thylakoid lumen acidification? *Plant Cell Environ* 20:147-154.

Raven JA, Beardall J. Carbon acquisition mechanisms in algae: carbon dioxide diffusion and carbon dioxide concentrating mechanisms. In: A.W.D. Larkum, S.E. Douglas, J.A. Raven Editors. *Photosynthesis in Algae*. Dordrecht: Kluwer Academic Publishers, 2003. p 225-244.

Reinfelder JR, Kraepiel AM, Morel FM (2000) Unicellular C₄ photosynthesis in a marine diatom. *Nature* 407:996-999.

Reinfelder JR, Milligan AJ, Morel FMM (2004) The role of the C₄ pathway in carbon accumulation and fixation in a marine diatom. *Plant Physiol* 135:2106-2111.

Reinhold L, Zviman M, Kaplan A (1989) A quantitative model for inorganic carbon fluxes and photosynthesis in cyanobacteria. *Plant Physiol Biochem* 27:945-954.

Rippka R, Waterbury J, Cohen-Bazire G (1974) Cyanobacterium which lacks thylakoids. *Arch Microbiol* 100:419-436.

Roberts K, Granum E, Leegood RC, Raven JA (2007) Carbon acquisition by diatoms. *Photosynth Res* 93:79-88.

Romanova AK, Zhen-Qi-Cheng Z, Mc-Fadden BA (1997) Activity and carboxylation specificity factor of mutant ribulose 1,5-bisphosphate carboxylase/oxygenase from *Anacystis nidulans*. *Biochem Mol Biol Int* 42:299-307.

Rowen R, Whitney SM, Fowler A and Yellowlees D (1996) RuBisCO in Marine Symbiotic Dinoflagellates: Form II Enzymes in Eukaryotic Oxygenic Phototrophs Encoded by a Nuclear Multigene Family. *Plant Cell* 8:539-553.

Sage RF and Seemann JR (1993) Regulation of ribulose-1,5-bisphosphate carboxylase/oxygenase activity in response to reduced light intensity in C₄ plants. *Plant Physiol* 102:21-28.

Salstam E and Campbell D (1996) Membrane lipid composition of the unusual cyanobacterium *Gloeobacter violaceus* PCC 7421, which lacks sulfoquinovosyl diacylglycerol. *Arch Microbiol*. 166:132-35.

Sambrook J, Fritsch EF and Maniatis T (1989) *Molecular Cloning: A laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Sampson EM and Bobik TA (2008) Microcompartments for B-12 dependent 1,2-propanediol degradation provide protection from DNA and cellular damage by a reactive metabolic intermediate. *J Bacteriol* 190:2966-71.

- Sato T, Atomi H and Imanaka T (2007) Archaeal type III RuBisCOs function in a pathway for AMP metabolism. *Science* 315:1003-1006.
- Savira Y, Noorb E, Milob R, Tlustya T (2010) Cross-species analysis traces adaptation of RuBisCO toward optimality in a low-dimensional landscape. *Proc Natl Acad Sci USA* 107:3475-3480.
- Schneider TD and Stephens RM (1990) Sequence Logos: A New Way to Display Consensus Sequences. *Nucleic Acids Res* 18:6097-6100.
- Schreuder H, Knight S, Curmi PMG, Andersson I, Cascio D, Brändén C-I, Eisenberg D (1993) Formation of the active site of ribulose-1,5-bisphosphate carboxylase/oxygenase by a disorder-order transition from the unactivated to the activated form. *Proc Natl Acad Sci USA*. 90:9968-9972.
- Sela, I, Ashkenazy H, Katoh K, Pupko T (2015) GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Res.* (Web Server issue): W7-W14.; doi: 10.1093/nar/gkq443.
- Seo PS and Yokota A (2003) The phylogenetic relationships of cyanobacteria inferred from *16S rRNA*, *gyrB*, *rpoC1* and *rpoD1* gene sequences. *J Gen Appl Microbiol* 49:191-203.
- Shikanai T, Foyer CH, Dulieu H, Parry MAJ, Yokota A (1996) A point mutation in the gene encoding the RuBisCO large subunit interferes with holoenzyme assembly. *Plant Mol Biol* 31:399-403.
- Shively JM, Bradburne CE, Aldrich HC, Bobik TA, Mehlman JL, Jin S, Baker SH (1998) Sequence homologs of the carboxysomal polypeptide CsoS1 of the thiobacilli are present in cyanobacteria and enteric bacteria that form carboxysomes-polyhedral bodies. *Can J Bot Revue Can Bot* 76:906-916.
- Sicheritz-Ponten T and Andersson SGE (2001) A phylogenomic approach to microbial evolution. *Nucleic Acids Res* 29:545-552.
- Smith EC and Griffiths H (1996) A pyrenoid based carbon concentrating mechanism is present in terrestrial bryophytes of the class Anthocerotae. *Planta* 200:203-212.
- Smith ME, Koteyeva NK, Voznesenskaya EV, Okita TW, Edwards GE (2009) Photosynthetic features of non-Kranz type C-4 versus Kranz type C-4 and C-3 species in subfamily Suaedoideae (Chenopodiaceae). *Funct Plant Biol* 36:770-782.

Snel B, Bork P, Huynen M (2000) Genome evolution: Gene fusion versus gene fission. *Trends Genet* 16:9-11.

So AKC, Espie GS, Williams EB, Shively JM, Heinhorst S, Cannon GC (2004) A novel evolutionary lineage of carbonic anhydrase (ϵ class) is a component of carboxysome shell. *J Bacteriol* 186:623-630.

So AKC, John-McKay M, Espie GS (2002) Characterization of a mutant lacking carboxysomal carbonic anhydrase from the cyanobacterium *Synechocystis* PCC6803. *Planta* 214:456-467.

Spalding MH (2008) Microalgal carbon dioxide concentrating mechanisms: *Chlamydomonas* inorganic carbon transporters. *J Exp Bot* 59:1463–1473.

Spreitzer RJ (2003) Role of the small subunit in ribulose-1,5-bisphosphate carboxylase/oxygenase. *Arch Biochem Biophys* 414:141-149.

Sriramulu DD, Liang M, Hernandez-Romero D, Raux-Deery E, Lunsdorf H, Parsons JB, Warren MJ, Prentice MB (2008) *Lactobacillus reuteri* DSM 20016 produces cobamin dependent diol dehydratase in metabolosomes and metabolizes 1,2-propanediol by disproportionation. *J Bacteriol* 190:4554-4567.

Starkenburg SR, Chain PS, Sayavedra-Soto LA, Hauser L, Land ML, Larimer FW, Malfatti SA, Klotz MG, Bottomley PJ, Arp DJ, Hickey WJ (2006) Genome sequence of the chemolithoautotrophic nitrite-oxidizing bacterium *Nitrobacter winogradskyi* Nb-255. *App Environ Microbiol* 72:2050-2063.

Sültemeyer DF and Rinast KA (1996) The CO₂ permeability of the plasma membrane of *Chlamydomonas reinhardtii*: mass-spectrometric ¹⁸O exchange measurements from ¹³C¹⁸O₂ in suspension of carbonic anhydrase-loaded plasma-membrane vesicles. *Planta* 200:358-368.

Sutter M, Wilson SC, Deutsch S, Kerfeld CA (2013) Two new high-resolution crystal structures of carboxysome pentamer proteins reveal high structural conservation of CcmL orthologs among distantly related cyanobacterial species. *Photosynth Res* 118: 9.

Tabita FR, Hanson TE, Li H, Satagopan S, Singh J, Chan S (2007) Function, Structure, and Evolution of the RuBisCO-Like Proteins and their RuBisCO Homologs. *Microbiol Mol Biol Rev* 71:576-599.

- Tabita FR, Satagopan S, Hanson TE, Kreeel NE, Scott SS (2008) Distinct form I, II, III, and IV RuBisCO proteins from the three kingdoms of life provide clues about RuBisCO evolution and structure/function relationships. *J Exp Bot* 59:1515-1524.
- Tajima F and Nei M (1984) Estimation of evolutionary distance between nucleotide sequences. *Mol Biol Evol* 1:269-285.
- Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30:2725-2729.
- Tanaka S, Kerfeld CA, Sawaya MR, Cai F, Heinhorst S, Cannon GC, Yeates TO (2008) Atomic level models of the bacterial carboxysome shell. *Science* 319:1083-1086.
- Tanaka S, Sawaya MR, Yeates TO (2010) Structure and mechanisms of a protein based organelle in *Escherichia coli*. *Science* 327:81-84.
- Taylor TC and Andersson I (1997) Structure of a product complex of spinach ribulose-1,5-bisphosphate carboxylase/oxygenase. *Biochem* 36: 4041-4046.
- Tcherkez GGB, Farquhar GD and Andrews TJ (2006) Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proc Natl Acad Sci U. S. A.* 103:7246-7251.
- Thompson AW, Foster RA, Krupke A, Carter BJ, Musat N, Vaultot D, Kuypers MMM, Zehr JP (2012) Unicellular cyanobacterium symbiotic with a single-celled eukaryotic alga. *Science* 337:1546-1550.
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22(22):4673-4680.
- Tomitani A, Okada K, Miyashita H, Matthijs HCP, Ohno T, Tanaka A (1999) Chlorophyll b and phycobilins in the common ancestor of cyanobacteria and chloroplasts. *Nature* 400:159-162.
- Tovchigrechko A and Vakser IA (2006) GRAMM-X public web server for protein-protein docking. *Nucleic Acids Res* 34(Web Server issue): W310-W314.
- Tsai Y, Sawaya MR, Yeates TO (2009) Analysis of lattice-translocation disorder in the layered hexagonal structure of carboxysome shell protein CsoS1C. *Acta Crystallogr D: Biol Crystallogr* 65:980-988.

Vincze T, Posfai J, Roberts RJ (2003) NEBcutter: a program to cleave DNA with restriction enzymes *Nucleic Acids Res* 31:3688-3691.

Vogel C, Bashton M, Kerrison ND, Chothia C, Teichmann SA (2004) Structure, function and evolution of multidomain proteins. *Curr Opin Struct Biol* 14:208–216.

Wang HL, Postier BL, Burnap RL (2004) Alterations in global pattern of gene expression in *Synechocystis sp.* PCC 6803 in response to inorganic carbon limitation and the inactivation of *ndhR*, a LysR family regulator. *J Biol Chem* 279: 5739-5751.

Wang Y, Sun Z, Horken KM, Im C, Xiang Y, Grossman AR, Weeks DP (2005) Analysis of CIA5, the master regulator of the carbon- concentrating mechanisms in *Chlamydomonas reinhardtii*, and its control of gene expression. *Can J Bot* 83:765-779.

Webb B and Sali A (2014) Comparative Protein Structure Modeling Using Modeller. *Curr Protoc Bioinfo* 5.6.1-5.6.32.

Whelan S and Goldman N (2001) A general empirical model for protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular Biology and Evolution* 18(5):691-699.

Whitmore L and Wallace BA (2004) DICHROWEB: an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Research* 32:W668-673

Whitney SM, Baldet P, Hudson GS, John Andrews T (2001) Form I RuBisCOs from non-green algae are expressed abundantly but not assembled in tobacco chloroplasts. *Plant J* 26:535-547.

Whitney SM, Sharwood RE (2007) Linked RuBisCO subunits can assemble into functional oligomers without impeding catalytic performance. *J Biol Chem* 282:3809-3818.

Whitney SM, Shaw DC, Yellowlees D (1995) Evidence that some dinoflagellates contain a ribulose-1, 5-bisphosphate carboxylase/ oxygenase related to that of the α -proteobacteria, *Proc Royal Society London. B* 259:271-275.

Whitney SM, Baldet P, Hudson GS, Andrews TJ (2001) Form I RuBisCOs from non-green algae are expressed abundantly but not assembled in tobacco chloroplasts. *Plant J* 26:535-547.

- Woodger FJ, Badger MR, Price GD (2003) Inorganic carbon limitation induces transcripts encoding components of the CO₂ concentrating mechanism in *Synechococcus* sp. PCC 7942 through a redox independent pathway. *Plant Physiol* 133:2069-2080.
- Woodger FJ, Badger MR, Price GD (2005) Sensing of inorganic carbon limitation in *Synechococcus* PCC 7942 is correlated with the size of the internal inorganic carbon pool and involves oxygen. *Plant Physiol* 139:1959-1969.
- Xu D and Zhang Y (2011) Improving the Physical Realism and Structural Accuracy of Protein Models by a Two-step Atomic-level Energy Minimization. *Biophys J* 101:2525-2534.
- Yamano T and Fukuzawa H (2009) Carbon-concentrating mechanism in a green alga, *Chlamydomonas reinhardtii*, revealed by transcriptome analyses. *J Basic Microbiol* 49:42-51.
- Yamano T, Miura K, Fukuzawa H (2008) Expression analysis of genes associated with the induction of the carbon-concentrating mechanism in *Chlamydomonas reinhardtii*. *Plant Physiol* 147:340-354.
- Yamaoka T, Satoh K, Katoh S (1978) Photosynthetic activities of a thermophilic blue-green alga. *Plant Cell Physiol* 19:943-954.
- Yanai I, Wolf YI, Koonin EV (2002) Evolution of gene fusions: horizontal transfer versus independent events. *Genome Biol* 3:research0024.
- Yoshida S, Atomi H, Imanaka T (2007) Engineering of a type III RuBisCO from a hyperthermophilic archaeon in order to enhance catalytic performance in mesophilic host cells. *Appl Environ Microbiol* 73:6254-6261.
- Yoshioka S, Taniguchi F, Miura K, Inoue T, Yamano T, Fukuzawa H (2004) The novel Myb transcription factor LCRI regulates the CO₂ -responsive gene *Cah1*, encoding a periplasmic carbonic anhydrase in *Chlamydomonas reinhardtii*. *Plant Cell* 16:1466-1477.
- Yu J, Liberton M, Cliften PF, Head RD, Jacobs JM, Smith RD, Koppenaar DW, Brand JJ, Pakrasi HB (2015) *Synechococcus elongatus* UTEX 2973, a fast growing cyanobacterial chassis for biosynthesis using light and CO₂. *Sci Rep* 5:8132.
- Zehr JP, Bench SR, Carter BJ, Hewson I, Niazi F, Shi T, Tripp HJ, Affourtit JP (2008) Globally distributed uncultivated oceanic N₂-fixing cyanobacteria lack oxygenic photosystem II. *Science* 322:1110-1112.

Zhang Y and Skolnick J (2004) Scoring function for automated assessment of protein structure template quality. *Proteins* 57:702-710.

Zhu G, Jensen RG, Bohnert HJ, Wildner GF, Schlitter J (1998) Dependence of catalysis and CO_2/O_2 specificity of RuBisCO on the carboxy-terminus of the large subunit at different temperatures. *Photosynth Res* 57:71-79.

Zhu XG, Long SP, Ort DR (2010) C_3 rice- an ideal arena for systems biology research. *J Integr Plant Biol* 52:762-770.

Appendix (i) – List of Publications

1. Carbon concentrating mechanisms: In rescue of RubisCO inefficiency: **Gurpreet Kaur Sidhu**, Rajesh Mehrotra and Sandhya Mehrotra (2014) *Acta Physiol. plant.* DOI: 10.1007/s11738-014-1652-3
2. Analysis of 39 cyanobacterial species reveals rbcX subunit to be present between L and S Subunits. **Gurpreet Kaur Sidhu**, Rashmi Sharma, Rajesh Mehrotra, Sandhya Mehrotra (2015), *Journal of Advances in Biology* Vol. 7, No. 3, 1420-1426.
3. Capturing atmospheric carbon: biological and non biological methods. *International journal of low-carbon technologies.* Panchsheela Nogia, **Gurpreet Kaur Sidhu**, Rajesh Mehrotra and Sandhya Mehrotra*(2013). doi:10.1093/ijlct/ctt071-9.
4. Role of habitat and great oxidation event on the occurrence of three multi-subunit inorganic carbon uptake systems in cyanobacteria. Vandana Tomar, **Gurpreet Kaur Sidhu**, Panchsheela Nogia, Rajesh Mehrotra, Sandhya Mehrotra 2015, *Journal of genetics.*
5. Distribution of Homologs of Carboxysome Shell Proteins across Various Phyla of Eubacteria. **Gurpreet Kaur Sidhu**, Panchsheela Nogia, Vandana Tomar, Rajesh Mehrotra, Sandhya Mehrotra. 2016 (under review).
6. Ribulose-1,5-bisphosphate carboxylase/oxygenase from a unique cyanobacterium *Gloeobacter violaceus* PCC 7421. **Gurpreet Kaur Sidhu**, Panchsheela Nogia, Vandana Tomar, Rajesh Mehrotra, Sandhya Mehrotra. 2016 (in communication).
7. Heterologous Expression of carboxysome proteins of an early diverging cyanobacterium in *E. coli*. **Gurpreet Kaur Sidhu**, Panchsheela Nogia, Vandana Tomar, Rajesh Mehrotra, Sandhya Mehrotra. 2016 (in communication).
8. Cross Talks Among Regulatory Components Of Carbon Concentrating Mechanisms In Lower Organisms. Vandana Tomar, **Gurpreet Kaur Sidhu**, Panchsheela Nogia, Rajesh Mehrotra and Sandhya Mehrotra. 2016 (in communication).

Appendix (ii) – List of conferences and workshops attended in the course of the degree

1. Seminar and Workshop on Liquid Chromatography and Tandem Mass spectrometry, organized by department of Pharmacy, BITS Pilani on 23-24 August 2016 in BITS Pilani, Rajasthan.
2. Extensive duplication and subfunctionalization of a nitrate transporter gene in *Arabidopsis thaliana*: Multiple factors involved in paralog retention P. Nogia, V. Tomar, G.K. Sidhu, R. Mehrotra, S. Mehrotra, Plant genome evolution: A current opinion conference, 6-8th September 2015, Amsterdam, The Netherlands.
3. Gurpreet Kaur Sidhu, Sandhya Mehrotra and Rajesh Mehrotra. Characterization of RubisCO and Carbon Concentration (CCM) proteins of *Gloeobacter violaceus* PCC7421 FAF-115, International Conference on Emerging Trends in Biotechnology, ICETB 2014. November 6-9, 2014, School of environmental sciences. Jawaharlal Nehru University and Biotech Research Society of India, JNU, New Delhi, India.
4. National Workshop cum training programme on STRUCTURAL BIOINFORMATICS, organized by Bioinformatics Centre, Birla Institute of Scientific Research, Statue Circle, Jaipur, India, held on 10-12 January, 2014.
5. Characterizing operons encoding RubisCO and Carbon concentrating mechanism proteins in a primitive cyanobacterium, *Gloeobacter violaceus*. Gurpreet Kaur Sidhu, Rajesh Mehrotra and Sandhya Mehrotra. The 38th Lorne conference on protein structure and function, 10-14th February, 2013, Lorne, Victoria, Australia.
6. Analyzing structural and functional evolutionary links between RubisCO and Carbon concentration mechanism proteins. Gurpreet Kaur Sidhu and Sandhya Mehrotra. International Conference on Industrial Biotechnology, Nov 21-23, 2012, Punjabi University, Patiala, Punjab, India.

Appendix (iii) – Biography of Guide – Dr. Sandhya Mehrotra

Dr. Sandhya Mehrotra obtained her Ph.D degree from centre for plant molecular biology, National Botanical Research Institute, Lucknow. The dissertation involved studying expression and regulation of genes encoded by chloroplast genome and involved in photosynthesis. Though she analyzed several genes, her main interest was the primary carbon fixing enzyme RuBisCO and gene encoding its large subunit. Thereafter she did a short post doctoral work in the area of developing anti-osteoporosis agents from Indian medicinal plants from Central Drug Research, Lucknow. In 2005, she moved for a post doctoral fellowship to Nara Institute of Science and Technology, Nara, Japan, where she studied patterns of evolution of Calvin cycle genes. Her major research interests include carbon fixation and sequestration, evolution of Calvin cycle and its enzymes, stress responsive designer promoters in plants and lipase catalyzed biofuel production from microorganisms. Since November 2008, she has been serving as a faculty member in the Department of Biological Sciences of BITS, Pilani. While at BITS, Pilani, she has been principal investigator in projects funded by DST/SERB and UGC as well as co-investigator in Aditya Birla Group of companies funded project. She has recently been granted a SERB/DST major project envisaging installment of cyanobacterial bicarbonate transporters in C3 plants to increase crop productivity. At BITS, Pilani, she is involved in teaching several courses at undergraduate and postgraduate levels. She has also guided several project and dissertation students. At the department level she has served/been serving as a member of Departmental committee on Academics, departmental research committee, library committee, placement faculty, member of student faculty council and the institute level standing committee on student discipline. Her research endeavors at BITS, Pilani and before have resulted into publication of as many as 30 research articles in international and national journals of repute and more than 34 presentations in both national and international conferences. She has been awarded the Indo- Australia science and technology early career visiting fellowship and has also been selected for Indian National Science Academy's bilateral exchange program. She serves as editor and reviewer for several journals from time to time.

Appendix (iv) – Biography of Self

Ms. Gurpreet Kaur Sidhu, completed her M.Sc Microbiology from Punjab Agricultural University, Ludhiana, Punjab. In partial fulfillment of the requirement for the degree she was involved in dissertation on “Isolation and Characterization of Entomopathogenic Fungi Against *Helicoverpa armigera* (Hubner)”. In August 2011, she enrolled for the Ph.D programme in Birla Institute of Technology and Science, Pilani. During the course of her degree in BITS, Pilani, she received financial assistance from funding agencies namely, Department of Science and Technology (DST) and University Grants Commission – Basic Scientific Research (UGC-BSR). In partial fulfillment of the requirement for the degree she has been involved in teaching courses namely Biology Laboratory, Biology Project Laboratory, Genetic Engineering and Instrumental Methods of Analysis. She has authored research articles published in international journals and has presented papers in various national and international conferences.

Carbon concentrating mechanisms: in rescue of Rubisco inefficiency

Gurpreet Kaur Sidhu · Rajesh Mehrotra · Sandhya Mehrotra

Received: 11 February 2014 / Revised: 29 June 2014 / Accepted: 23 August 2014 / Published online: 4 September 2014
© Franciszek Górski Institute of Plant Physiology, Polish Academy of Sciences, Kraków 2014

Abstract Agricultural yields have kept pace with the rising demands in the recent past as a result of the breeding and improved farming practices, but these practices alone will not be able to meet the demands of the future. The focus is now on the enhancement of the photosynthetic machinery. In photosynthesis, the rate limiting step is the one catalyzed by RuBisCO- Ribulose-1,5-bisphosphate carboxylase/oxygenase (4.1.1.39), which, because of its loose specificity and low turnover rate, is the primary target of most research programs directed towards improved photosynthesis. The other avenues of photosynthetic machinery that are under investigation to enhance it include—improved stomatal regulation and membrane permeability, RuBisCO with high specificity for CO₂ and higher catalytic turnover; bypass of photorespiration and introduction of carbon concentrating mechanism (CCM) into the C₃ plants. Carbon concentrating mechanisms cause accumulation of carbon dioxide in vicinity of RuBisCO producing a high CO₂/O₂ ratio and hence an environment more suitable for carboxylation reactions than oxygenation reactions. This article includes the basic details of the major naturally occurring CCMs in various photosynthetic organisms to identify the knowledge gaps in each which could help study the prospects of its possible introduction into a non-native system as C₃ plants which are devoid of any CCM.

Keywords C₄ metabolism · Carbon concentrating mechanism · Carboxysomes · Pyrenoids · RuBisCO

Abbreviations

| | |
|-------------------------------|--|
| 3-PGA | 3-Phosphoglycerate |
| AMP | Adenosine 5' monophosphate |
| ATP | Adenosine triphosphate |
| BMC | Bacterial microcompartment |
| C ₃ plants | Plants with a three carbon compound first product after CO ₂ fixation |
| C ₄ plants | Plants with a four carbon compound as first product after CO ₂ fixation |
| CA | Carbonic anhydrase |
| CAM | Crassulacean acid metabolism |
| CCM | Carbon concentrating mechanism |
| Ci | Inorganic carbon |
| HCO ₃ ⁻ | Bicarbonate |
| IRES | Internal ribosomal entry site |
| K _{cat} | Catalytic rate |
| K _m | Michaelis constant |
| NADPH | Nicotinamide adenine dinucleotide phosphate |
| RLP | RuBisCO like protein |
| RuBisCO | Ribulose-1,5-bisphosphate carboxylase |
| RuBP | Ribulose-1,5-bisphosphate |

Introduction

RuBisCO i.e., Ribulose-1,5-bisphosphate carboxylase is the primary enzyme involved in the process of photosynthesis. It catalyses the fixation of atmospheric CO₂ (carbon dioxide) to Ribulose-1,5-bisphosphate (RuBP) to form two molecules of 3-phosphoglycerate

Communicated by A.K.Kononowicz.

G. K. Sidhu · R. Mehrotra · S. Mehrotra (✉)
Department of Biological Sciences, Birla Institute of Technology
and Sciences, Pilani 333031, Rajasthan, India
e-mail: sandhyamehrotrabits@gmail.com



Analysis of 39 cyanobacterial species reveals rbcX subunit to be present between L and S Subunits

Gurpreet Kaur Sidhu¹, Rashmi Sharma¹, Rajesh Mehrotra¹, Sandhya Mehrotra^{1*}

¹Department of Biological Sciences, Birla Institute of Technology and Science, Pilani, Rajasthan-33303, India
Corresponding Author: Sandhya Mehrotra,

Abstract

The impact of great oxidation event on RuBisCO has been tremendous. It has led to the competition between carbon dioxide and oxygen at the active site of the enzyme. Cyanobacteria developed strategies to combat this change by concentrating carbon dioxide in organelles called carboxysomes. RbcX helps in proper folding of RuBisCO by interacting with RbcL. However, it is not an absolute requirement for RuBisCO to attain proper folding only with the aid of RbcX. RbcX has a chaperone like activity. The present analysis led to the finding that cyanobacterial species lacking RbcX contain multitude of proteins showing homology to chaperone like proteins. These proteins might be playing the same role as RbcX in these cyanobacterial species to help RuBisCO acquire proper folding. Analyses also indicated that in general the rbcX motif is present between rbcL and rbcS.



Council for Innovative Research

Peer Review Research Publishing System

Journal of Advances in Biology

Vol. 7, No. 3

www.cirjab.com

editorsjab@gmail.com , editor@cirjab.com

Capturing atmospheric carbon: biological and nonbiological methods

Panchsheela Nogia, Gurpreet Kaur Sidhu, Rajesh Mehrotra and Sandhya Mehrotra*
*Department of Biological Science, Birla Institute of Technology and Science, Pilani,
Rajasthan 333031, India*

Abstract

Atmospheric carbon dioxide is one of the primary greenhouse gases on earth and its continuous emission by manmade activities is leading to a rise in atmospheric temperature. On the other hand, various natural phenomena exist that contribute to the sequestration of atmospheric carbon dioxide, i.e. its capture and long-term storage. These phenomena include oceanic, geological and chemical processes happening on earth. In addition to the above-mentioned nonbiological methods, various biological methods *viz.* soil carbon sequestration and phytosequestration have also been contributing to fixation of atmospheric carbon. Phytosequestration is mainly performed by several photosynthetic mechanisms such as C₃, C₄ and crassulacean acid metabolism (CAM) pathways of plants, carboxysomes of cyanobacteria and pyrenoids of microalgae. For an effective mitigation of global climate change, it is required to stabilize the CO₂ concentration to viable levels. It requires various permutations and combinations of naturally existing and engineering strategies. Although numerous strategies are in commodious use in the present times, the issues of sustainability and long-term stability still exist. We present an overview of the natural and manmade biological and nonbiological processes used today to reduce atmospheric CO₂ levels and discuss the scope and limitations of each of them.

Keywords: carbon; carbon sequestration; phytosequestration; carboxysomes; CO₂ fixation; global climate change; pyrenoids

Received 25 May 2013; revised 15 August 2013; accepted 30 September 2013

*Corresponding author:
sandhyamehrotrabits@
gmail.com

1 INTRODUCTION

Rise in global surface temperature is the cumulative effect of successively increasing concentration of greenhouse gases over several decades [1]. Green house gas emission results from the continuous use of fossil fuels which are fulfilling ~85% of world's energy requirement [2]. Most of the greenhouse gases like methane (CH₄), nitrous oxide (N₂O), chlorofluorocarbons (CFC) and carbon dioxide (CO₂) are contributing to overall climate change, wherein carbon dioxide is playing a major role. For a while, environmentalists who have been intensely involved in studying the adverse effects caused by greenhouse gases have been advocating the removal of excess CO₂ as the only way by which one can undo the harm that has already disturbed the natural balance [3].

In context of crop productivity, theoretically it seems that increasing atmospheric CO₂ level would enhance plant photosynthesis and ultimately crop productivity but the long-term effects are uncertain and might involve several side issues such as


negative effects on plant food web, decreased plant nutritional values, reduced N content of plant etc. Hence, there has to be a balance in atmospheric carbon level [4, 5].

Other part of the story says that carbon fixation by photosynthesis can also be considered as a natural method of capturing atmospheric carbon [6]. In order to overcome negative effects of global warming and climate change that is being caused by increased carbon level, people are trying to make improvements in the naturally occurring photosynthetic reactions by transferring few mechanisms or several genes from an efficient system like cyanobacteria, microalgae and C₄ plants to inefficient photosynthetic systems such as higher C₃ plants [7–10]. Success rate is, however doubtful and, therefore, we need to develop several strategies to minimize the ever increasing carbon levels so that side issues associated with it could also be avoided.

'Carbon sequestration' is the term given to capturing atmospheric carbon and converting it into forms unable to contribute to global warming. It is anticipated that in the coming centuries CO₂ emissions might double, so there is a serious need to

RESEARCH ARTICLE

Role of habitat and great oxidation event on the occurrence of three multisubunit inorganic carbon-uptake systems in cyanobacteria

VANDANA TOMAR, GURPREET KAUR SIDHU, PANCHSHEELA NOGIA, RAJESH MEHROTRA
and SANDHYA MEHROTRA* 

Department of Biological Sciences, Birla Institute of Technology and Science, Pilani 333 031, India

Abstract

The oxygenase reaction catalyzed by RuBisCO became an issue only after the evolution of the oxygenic photosynthesis in cyanobacteria. Several strategies were developed by autotrophic organisms as an evolutionary response to increase oxygen levels to help RuBisCO maximize its net carboxylation rate. One of the crucial advancements in this context was the development of more efficient inorganic carbon transporters which could help in increasing the influx of inorganic carbon (Ci) at the site of CO₂ fixation. We conducted a survey to find out the genes coding for cyanobacterial Ci transporters in 40 cyanobacterial phyla with respect to transporters present in *Gloeobacter violaceus* PCC 7421, an early-diverging cyanobacterium. An attempt was also made to correlate the prevalence of the kind of transporter present in the species with its habitat. Basically, two types of cyanobacterial inorganic carbon transporters exist, i.e. bicarbonate transporters and CO₂-uptake systems. The transporters also show variation in context to their structure as some exist as single subunit proteins (BicA and SbtA), while others exist as multisubunit proteins (namely BCT1, Ndh₃ and Ndh₄). The phylogeny and distribution of the former have been extensively studied and the present analysis provides an insight into the latter ones. The *in silico* analysis of the genes under study revealed that their distribution was greatly influenced by the habitat and major environmental changes such as the great oxidation event (GOE) in the course of their evolution.

[Tomar V., Sidhu G. K., Nogia P., Mehrotra R. and Mehrotra S. 2016 Role of habitat and great oxidation event on the occurrence of three multisubunit inorganic carbon-uptake systems in cyanobacteria. *J. Genet.* 95, xx–xx]

Introduction

Cyanobacteria or blue green algae are one of the most diverse groups of organisms which colonize various niches including freshwater (rivers, ponds and lakes), polar caps, hot springs, alkaline, estuarine, open as well as saline oceans. During the ancient environmental adversities, these organisms along with certain algal species developed active systems known as carbon-concentrating mechanisms (CCM) to concentrate CO₂ at the site of RuBisCO activity which in turn led to a positive impact on photosynthetic ability of these organisms (Caemmerer and Evans 2010). This strategy was also adopted by the land plants and they developed certain anatomical features (Kranz anatomy) to increase the carboxylation activity of RuBisCO, thus reducing the resultant photorespiratory losses (Raven *et al.* 2008).

The CCM present in cyanobacteria are highly efficient and can accumulate CO₂ around RuBisCO by a factor of

1000-fold above ambient levels (Price *et al.* 2011). CCMs in cyanobacteria and proteobacteria as a whole include (i) RuBisCO and carbonic anhydrases (CA) enclosed in micro-compartments known as carboxysomes and (ii) Ci transporters, which regulate the CO₂/HCO₃⁻ influx and efflux at the site of RuBisCO activity and hence lead to a marked elevation in the CO₂ concentration in the vicinity of RuBisCO.

Cyanobacteria are dependent on active accumulation of Ci to achieve a satisfactory rate of CO₂ fixation and growth (Badger *et al.* 2002). The efficacy of any CCM relies on the ability to minimize the loss of CO₂ from the CO₂ elevation zone or the dissolved Ci accumulation zone (Price *et al.* 2007). It can be accomplished in four ways. First, the accumulation of bicarbonate instead of CO₂ reduces the chances of Ci leakage, since the former is less permeable through the plasma membrane as compared to the latter. Secondly, the absence of CA activity in the cytosol minimizes leakage due to wasteful conversion to CO₂ and subsequent diffusion back to the external medium. Thirdly, the carboxysome protein shell is proposed to have specifically

*For correspondence. E-mail: sandhyamehrotrabits@gmail.com.

Keywords. inorganic carbon transporters; *Gloeobacter violaceus* PCC 7421; great oxidation event; cyanobacteria; evolution; carbon-concentrating mechanism.