

## ***In silico* Characterization and Homology Modeling of Cyanobacterial RuBisCo (LS) with Computational Tools and Bioinformatic Servers**

\*<sup>1</sup>S. Meenakshi, <sup>2</sup>S. Srisudha

<sup>1,2</sup>Department of Botany and Microbiology, Lady Doak College, Madurai  
Phone: 9047520826, Email ID: \*microkshi82@gmail.com

Received: Jun 22<sup>nd</sup> 2012, Accepted: Jul 6<sup>th</sup> 2012, Published: Sep 1<sup>st</sup> 2012

### **Abstract:**

Photosynthetic prokaryotes and eukaryotes, namely phytoplankton, chromophytes and cyanobacteria use the Calvin – Benson – Bassham (CBB) pathway to metabolize CO<sub>2</sub> [1]. The key enzyme of the CBB pathway is ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCo). RuBisCo-based CO<sub>2</sub> fixation occurs in all cyanobacteria through C3 intermediates. The RuBisCo plays a pivotal role in linking the inorganic (CO<sub>2</sub>) and the organic phases (biomass) of the global carbon cycle. It is a slow and confused catalyst, limiting productivity and resource (e.g., water, nutrients) use efficiency in many plants [2]. The aim of this study was to determine the physicochemical properties of cyanobacterial RuBisCo and to develop 3-dimensional models of the selected enzyme. The biocomputational analysis was performed *in silico* using web-based software and servers. The alignment of cyanobacterial enzymes with the other related sequences and secondary structure analysis revealed that there are conserved amino acid substitutions and polymorphisms in the RuBisCo sequence. The large subunit of the selected cyanobacterial RuBisCo has many phosphorylation sites; alpha helical in structure, high homology with other cyanobacterial sequences (99%) and has close relation in phylogenetic tree. Furthermore, some cyanobacterial strains seem to possess unique amino acid stretches that may modulate various aspects of catalysis and regulation. The physicochemical characteristics and the 3D models provide a framework for the purification and characterization of cyanobacterial RuBisCos.

### **Key words:**

Ribulose-1,5-bisphosphate carboxylase/oxygenase, cyanobacteria, (LS) Large subunit, computational analysis, proteomic tools.

### **Introduction:**

D-ribulose-1,5-bisphosphate carboxylase / oxygenase (RuBisCo, EC4.1.1.39), the most abundant enzyme on earth catalyses the atmospheric CO<sub>2</sub> uptake in plants. It catalyses the addition of CO<sub>2</sub> to ribulose-1,5-bisphosphate (RuBP) and gives two molecules of 3-phosphoglycerate (3-PGA). This reaction is essential for the synthesis of sugars through the Calvin cycle. Thereby, the carboxylase activity of RuBisCo is an important process in

photosynthesis [3]. However, RuBisCo also catalyses the addition of O<sub>2</sub> to RuBP to produce one molecule of 3-PGA and one molecule of 2-phosphoglycolate [4]. These pathways are relatively smaller contributors to global carbon fixation than that catalyzed by RuBisCo. Since RuBisCo is an important component in cellular metabolism, the focus has been emphasized more in learning about the molecular basis of RuBisCo's ability to distinguish between CO<sub>2</sub> and O<sub>2</sub>. Different forms of RuBisCo are widely distributed among the Photosynthetic organisms, Eukaryotic algae, Autotrophic Proteobacteria and some Archaeae.

The catalytic incorporation of CO<sub>2</sub> into RuBP by RuBisCo is the first step in the production of carbohydrates by plants, which are used to build biomass and produce energy during growth and development. Despite the pivotal role of RuBisCo in linking the inorganic (CO<sub>2</sub>) and the organic phases (biomass) of the global carbon cycle, it is a slow and confused catalyst, limiting the productivity and the use of resources (e.g., water, nutrients) in many plants [2]. Understandably, RuBisCo has been studied intensively and is a prime target for genetic engineering to improve photosynthetic efficiency [6]. Many amino acid residues have been manipulated in both subunits to understand the structure-function relationship of RuBisCo. The active site residues of LS (Large Subunit) are highly conserved among more than thousand LS sequences [7].

In Plant, Algae, Cyanobacteria and phototrophic and chemoautotrophic proteobacteria, RuBisCo comprises large subunit (LSU) chains and small subunit (SSU) chains. The substrate binding sites are located in the large Subunit. Many archaeal RuBisCo enzymes are found in obligate anaerobes that have evolved in the absence of molecular oxygen. Interestingly, those archaeal RuBisCo enzymes that have been studied were shown to be extremely oxygen-sensitive.

The availability of internet based tools provides an excellent opportunity to characterize the physicochemical properties of Cyanobacterial RuBisCo as well as their primary, secondary and three dimensional structural properties. The purpose of this study was primarily to

report the *in silico* analysis and characterization of Cyanobacterial RuBisCo in comparison to a few other plant RuBisCo systems<sup>[8]</sup>. Since the LS of RuBisCo is believed to be regulated by phosphorylation which may have some impact on its enzyme activity, the study was focused on computational analysis in identifying the phosphorylation sites<sup>[9]</sup>. Detailed understanding of RuBisCo catalysis has taken advantage of the exponential rise in computer power and the increasing accuracy of theoretical protein models. Several groups have applied up-to-date computational tools to examine the energetics and atomistic details of RuBisCo's carboxylation and oxygenation reactions<sup>[10]</sup>. While these calculations reveal molecular details of the contribution of active-site residues to RuBisCos catalysis, the large size of RuBisCo prevents inclusion of all atoms in the calculation. As a result, the design of "better" RuBisCo using *in-silico* modelling tools will enable the understanding of its structure and function<sup>[11]</sup>.

#### Materials and methods:

##### RuBisCo protein sequences:

Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCo) protein sequences were retrieved from the National Center for Biotechnology Information (NCBI) [www.ncbi.nlm.nih.gov]. Ten LS of RuBisCo protein sequences from Cyanobacterial strains and three Plant RuBisCo protein sequences were selected. The selected Cyanobacterial and Plant sequences were converted to FASTA format<sup>[12]</sup>.

##### Sequence analysis:

##### Sequence alignments:

Multiple sequence alignment was performed with ClustalW<sup>[13]</sup>. The ClustalW alignment file was generated. Phylogenetic analysis of protein sequences was generated using the alignment obtained with ClustalW and visualized in tree top viewer.

##### Structural analysis:

The Amino acid composition of the cyanobacterial and plant RuBisCo sequences was computed using the PEPSTATS analysis tool<sup>[14]</sup>. The physico-chemical parameters such as the molecular weight, isoelectric point (pI), instability index (Ii), aliphatic index (Ai), and grand average of hydropathicity (GRAVY), -R: No. of negative residues and +R: No. of positive residues were determined using the ProtParam tool of the ExPASy proteomics server. Secondary structure prediction was conducted using double prediction method for cyanobacterial and plant RuBisCo sequences. This enables the prediction of secondary structure such as

helix, sheet and coil<sup>[15]</sup>. The presence of disulfide bridge was analyzed using CYS-REC tool which predicts the most probable bonding patterns between available cysteine residues. The amino acid sequences of the selected cyanobacteria and plants were analyzed for the putative phosphorylation sites using the NetPhos 2.0 Server available at the Center for Biological Sequence Analysis. (<http://www.cbs.dtu.dk/services/NetPhos/>). The 3-D model dimensional structure RuBisCo protein sequences were constructed using SWISS-MODEL<sup>[16]</sup>. The Rasmol and PyMOL were used to visualize and refine the models. The modeled 3D structures were evaluated and validated with the RAMPAGE programs<sup>[17]</sup>.

#### Result & Discussion:

##### RuBisCo protein sequences:

Ten LS of cyanobacterial RuBisCo sequences and three LS of plant RuBisCo sequences were selected from NCBI databases. The sequences were converted into FASTA format.

The primary structure analysis of LS of Cyanobacterial RuBisCo indicates that the percentage of tryptophan residues is less than 2% across species. The most abundant amino acid is alanine and glycine which makes up approximately 10 - 9% of the primary structures in both cyanobacterial and plant sequences. Leucine and glutamine constituted 5 - 10%. In plant RuBisCo, alanine comprised 7-8% of primary structure.

##### Computation of Physico-chemical parameters of RuBisCo using ProtParam tool:

The average molecular weight of RuBisCo was slightly larger in cyanobacterial LSs i.e., 53279.7 g mol<sup>-1</sup> and whereas in plant RuBisCos, the molecular weight was 52729.1 g mol<sup>-1</sup>. The computed isoelectric point (pI) of all cyanobacterial RuBisCos was below 7 indicating that the proteins will most likely precipitate in acidic buffers but plant RuBisCo have pI above 7 indicating that the proteins are soluble in basic buffers. According to the ProtParam server, a protein whose Instability Index (Ii) is larger than 40 may be unstable therefore, the program predicted that all plant RuBisCos would be unstable in solution but in LS of cyanobacterial RuBisCo of *Planktothrix rubescens*, *Cylindrospermopsis raciborskii* CS-505, *Raphidiopsis brookii* D9, *Fischerella* sp and *Thermosynechococcus vulcanus*, Ii index was smaller than 40 which predicts that the proteins are stable. The aliphatic index which gives a measure of the relative volume occupied by alanine, valine, isoleucine and leucine, ranged from 74.62 to 84.08 for LS of

cyanobacterial RuBisCos except *Planktothrix rubescens*. In plant RuBisCo Ai ranged from 77.34 to 78.53 in LS. The Grand Average Hydropathy (GRAVY) values for LS of cyanobacterial RuBisCos ranged from -0.252 to -0.388 indicating that the proteins will interact favorably with water, the range from -0.324 to -0.580 GRAVY values for LS of plant RuBisCo protein. A GRAVY score indicates the relative value for the hydrophobic residues of the protein. Although no positional or interaction effects for adjacent residues are taken into consideration by the GRAVY score, it still provides some indication of the physical state of the protein. <sup>[18]</sup>

#### **Secondary Structure Prediction for Cyanobacterial and Plant RuBisCo Sequences:**

Despite the low cysteine content, the disulfide bridge prediction tool CYS-REC computed disulfide bonds in most cyanobacterial and plant RuBisCo. The methods used to generate the consensus were: The Double Prediction Method (DPM) <sup>[19]</sup>. The secondary structure of  $\alpha$ -helix predicted ranged from 35 to 38% for LS of cyanobacterial RuBisCos and plant has 36%. The secondary structure of  $\beta$ -sheet predicted ranged from 4 to 5% for LS of cyanobacterial RuBisCos. The results of the consensus in Table 4 indicate that cyanobacterial RuBisCos are largely alpha helical with less than 10% beta structures. Coil comprised 48 -50% in both cyanobacterial and plant RuBisCo sequences. In *Planktothrix rubescens*, no  $\alpha$ ,  $\beta$  helical structure were detected and 100% of the secondary structure was represented by coils.

#### **Putative Phosphorylation site prediction server for Cyanobacterial and plant RuBisCo:**

LS of RuBisCo are regulated by phosphorylation which has an impact on its activity. Computational analysis helps us to identify the phosphorylation site for enhancing the function of RuBisCo. The server predicts phosphorylation sites for all the three amino acids serine, threonine and tyrosine. Protein phosphorylation at serine, threonine or tyrosine residues affects a multitude of cellular signaling processes. Here an artificial neural network method has been employed for prediction of phosphorylation sites in independent sequences with sensitivity in the range from 69% to 96%. The LS of *Raphidiopsis brookii D9* has maximum 16 amino acid residues but in LS of plant *Oryza sativa Japonica Group* contain 17 amino acid residues.

#### **Homology Modeling:**

The homological comparisons of RuBisCo sequences were accomplished by online tool BLAST 2 sequences.

The results indicate high homologies of RuBisCo sequences between cyanobacterial *Geitlerinema sp. PCC 8501*, *Microcystis aeruginosa PCC 7806*, *Planktothrix rubescens*, *Trichodesmium erythraeum IMS101*, *Cylindrospermopsis raciborskii CS-505*, *Raphidiopsis brookii D9*, *Fischerella sp. JSC-11*, *Lyngbya majuscula 3L*, *Arthrospira platensis NIES-39*, *Thermosynechococcus vulcanus* and plant RuBisCo sequences of *Zea mays*, *Sorghum bicolor* and *Oryza sativa Japonica Group*. The identities are 85%, 85% and 87% respectively. The identities of further deduced cyanobacterial RuBisCo sequence to cyanobacterial sp. comparisons using the BLAST 2 sequences indicated 99% similarity between *Raphidiopsis brookii D9* and *Lyngbya majuscula 3L* to *Cylindrospermopsis raciborskii CS-505* and 96% similarity between *Fischerella sp. JSC-11* and *Cylindrospermopsis raciborskii CS-505*.

#### **Multiple Sequence Alignment using ClustalW:**

Leucine residues are found as dimers in cyanobacterial and plant RuBisCo. Most significantly, the sequences LLLHIH and GFVDLL are found in all cyanobacterial and plant RuBisCos. The sequences LTDLT, RAVYECLR, LNVTAPTC and TPDYTPK are present in all cyanobacterial strains and the sequences LTSLT, RACYECLR, LNATAGTC and TPEYETK in plant RuBisCo sequences indicating that these sequences are conserved. These sequences indicate alpha helical nature of RuBisCo. The sequences LLACF is found in RuBisCo of *Microcystis* but in all other cyanobacterial and plant RuBisCo, Glycine-rich stretch, GGLDF and SGGDH are found. (Fig 1)

A multiple sequence alignment of cyanobacterial and plant RuBisCo was performed with ClustalW <sup>[13]</sup>. The alignment of cyanobacterial proteins revealed that there are conserved amino acid substitutions between the marine and fresh water strains <sup>[19]</sup>. The multiple sequence alignment also revealed that there are stretches of amino acids that are exclusive to the marine and fresh water algal RuBisCo protein. The sequence DYTP, DNQF, ARWC and ACIQ is found in RuBisCo of marine and fresh water algae.

Fig: 1 Multiple sequence alignment of RuBisCo (LS) from Cyanobacteria and plant; the black blocks contain sequences that are unique to a group of cyanobacteria and plant protein sequences. The alignment was generated by Clustal W

gi Thermosynechococcus	-MAYTQSKSQKVGYQAGVKDYRLTYYPDTPKDDILAAFRVTPQPGVP	49
gi Arthrospira	MSYSQTQTKSKAGYQAGVKDYKLTYYTPDTPKDDILAAFRVSPQPGVP	50
gi Zea	MSPQTEKASVGFKAGVKDYKLTYYTPEYETKDDILAAFRVTPQLGVP	49
gi Sorghum	-MSPQTEKASVGFKAGVKDYKLTYYTPEYETKDDILAAFRVTPQLGVP	49
gi Oryza	-MSPQTEKASVGFKAGVKDYKLTYYTPEYETKDDILAAFRVTPQPGVP	49
gi Raphidiopsis	MSYAQTQTQKSGYQAGVKDYRLTYYPDTPKDDLLAAFRMTPQPGVP	50
gi Fischerella	MSYAQTQTQKAGYQAGVKDYRLTYYPDTPKDDILAAFRVTPQPGVP	50
gi Microcystis	----MVQAKSKGFQAGVKDYRLTYYPDTPKDDLLACFRVTPQPGVP	45
gi Lyngbya	MSYSSRTQTKAGYDAGVKDYKLTYYTPDTPKDDILAAFRVTPQPGVP	50
:		
gi Thermosynechococcus	FEEAAAVAAESSTGTWTTVWTDLLDLDLRYKGRCYDIEPLPGEDNQFIA	99
gi Arthrospira	PEEAGAAVAAESSTGTWTTVWTDLLDLDLRYKGRCYHIEPVPGEDNQFFC	100
gi Zea	PEEAGAAVAAESSTGTWTTVWTDGLTSLDRYKGRCYHIEPVPDQDYIC	99
gi Sorghum	PEEAGAAVAAESSTGTWTTVWTDGLTSLDRYKGRCYHIEPVPDQDYIC	99
gi Oryza	PEEAGAAVAAESSTGTWTTVWTDGLTSLDRYKGRCYHIEPVPGEDNQYIA	99
gi Raphidiopsis	PEEAGAAVAAESSTGTWTTVWTDLLDLDLRYKGRCYDIEPVAGEDNQYIA	100
gi Fischerella	PEEAGAAVAAESSTGTWTTVWTDLLDLDLRYKGRCYDIEPVPGEDNQYIC	100
gi Microcystis	PEEAGAAVAAESSTGTWTTVWTDNLTLDLRYKGRCYDIEPVPNEDNQFFC	95
gi Lyngbya	PEEAGAAVAAESSTGTWTTVWTDLLDLDLRYKGRCYDIEPVRGEDNQFIC	100
:		
gi Thermosynechococcus	VIAYPLDLFEEGSVTNMLTSIVGNVFGFKALKALRLEDLRIPVAYLKTFQ	149
gi Arthrospira	FVAYPLDLFEEGSVTNMLTSIVGNVFGFKALRGLRLEDMRIPIAYLKTFQ	150
gi Zea	YVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLEDLRIPPAYSKTFQ	149
gi Sorghum	YVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLEDLRIPPAYVKTQ	149
gi Oryza	YVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLEDLRIPPTYSKTFQ	149
gi Raphidiopsis	YVAYPLDLFEEGSVTNMFTSIVGNVFGFKALRALRLEDLRIPVAYLKTFQ	150
gi Fischerella	YVAYPLDLFEEGSVTNMLTSIVGNVFGFKALRALRLEDLRIPVAYLKTFQ	150
gi Microcystis	FVAYPLDLFEEGSVTNMLTSIVGNVFGFKALRGLRLEDIRFPVALIKTFQ	145
gi Lyngbya	FIAYPLDLFEEGSVTNMLTSIVGNVFGFKALRALRLEDLRIPVAYLKTFQ	150
:		
gi Thermosynechococcus	GPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDF	199
gi Arthrospira	GPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDF	200
gi Zea	GPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRACVYECLRGGLDF	199
gi Sorghum	GPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRACVYECLRGGLDF	199
gi Oryza	GPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRACVYECLRGGLDF	199
gi Raphidiopsis	GPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDF	200
gi Fischerella	GPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDF	200
gi Microcystis	GPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDF	195
gi Lyngbya	GPPHGIQVERDKLNKYGRPLLGCTIKPKLGLSAKNYGRAVYECLRGGLDF	200
:		
gi Thermosynechococcus	TKDDENINSQPFQRWRDRFLFVADAIHKAQAETGEIKGHYLNVTAPTCEE	249
gi Arthrospira	TKDDENINSQPFMRWRDRFLFVQEAIEKAQAETNEIKGHYLNVTAPTCEE	250
gi Zea	TKDDENINSQPFMRWRDRFVFCAEAIYKAQAETGEIKGHYLNATAGTCEE	249
gi Sorghum	TKDDENINSQPFMRWRDRFVFCAEAIYKAQAETGEIKGHYLNATAGTCEE	249
gi Oryza	TKDDENINSQPFMRWRDRFVFCAEAIYKSAQAETGEIKGHYLNATAGTCEE	249
gi Raphidiopsis	TKDDENINSAPFQRWRDRFLFVAEAIKKSQAETGEIKGHYLNVTAPTCEE	250
gi Fischerella	TKDDENINSQPFQRWRDRFLFVADAIHKAQAETGEIKGHYLNVTAPTCEE	250
gi Microcystis	TKDDENINSQPFMRWRDRFLFVQEAIVKSAQETNEVKGHYLNVTAPTCEE	245
gi Lyngbya	TKDDENINSQPFMRWRDRFLFVADAIHKSQAETGEIKGHYLNVTAPTCEE	250

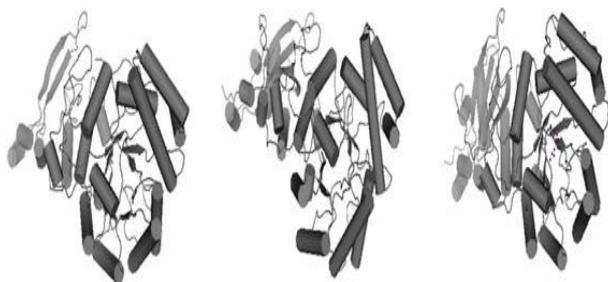
gi Thermosynechococcus	MLKRAEFAKEL <b>EM</b> PIIMHDF <b>LA</b> GFTANTT <b>LSKW</b> CRDNG <b>MLLHIHRAMHA</b> 299
gi Arthrospira	<b>MM</b> KRAEFAKE <b>IGT</b> PIIMHDF <b>FA</b> GFTANTT <b>LARW</b> CRDNG <b>LLLHIHRAMHA</b> 300
gi Zea	<b>MIKRAVFA</b> RELGVPIVMHDYLTGGFTANT <b>LSHY</b> CRDNG <b>LLLHIHRAMHA</b> 299
gi Sorghum	<b>MIKRAVFA</b> RELGVPIVMHDYLTGGFTANT <b>LSHY</b> CRDNG <b>LLLHIHRAMHA</b> 299
gi Oryza	<b>MIKRAVFA</b> RELGVPIVMHDYLTGGFTANT <b>SLAHY</b> CRDNG <b>LLLHIHRAMHA</b> 299
gi Raphidiopsis	MLKRAEFAKEL <b>EM</b> PIIMHDY <b>LA</b> GFTANTT <b>LARW</b> CRDNG <b>LLLHIHRAMHA</b> 300
gi Fischerella	MLKRAEFAKEL <b>KM</b> PIIMHDY <b>LA</b> GFTANTT <b>LAKW</b> CRDNG <b>LLLHIHRAMHA</b> 300
gi Microcystis	<b>MMQRAEFAAEIKT</b> PIIMHDYLTGGFTANT <b>LAKF</b> CRD <b>K</b> GLLHIHRAMHA 295
gi Lynngbya	<b>MMKRAEFAKSLN</b> MPIIMHDYLTGGFTANT <b>LARW</b> CRDNG <b>VLLHIHRAMHA</b> 300
gi Thermosynechococcus	<b>VMDRQKNHG</b> IHFRVLAKCLRM <b>SGGDHIHT</b> GT <b>VVGKLE</b> GD <b>KAVTLGFVDLL</b> 349
gi Arthrospira	<b>VVDRQR</b> NHG <b>IHFRVLAKCLRM</b> SGGD <b>HL</b> HS <b>GT</b> V <b>VGKLE</b> GE <b>KGIT</b> MG <b>FVDLM</b> 350
gi Zea	<b>VIDRQKNHGM</b> HFRVLAK <b>ALRM</b> SGGD <b>HIHSGT</b> V <b>VGKLE</b> GER <b>EITLGFVDLL</b> 349
gi Sorghum	<b>VIDRQKNHGM</b> HFRVLAK <b>ALRM</b> SGGD <b>HIHSGT</b> V <b>VGKLE</b> GER <b>EITLGFVDLL</b> 349
gi Oryza	<b>VIDRQKNHGM</b> HFRVLAK <b>ALRM</b> SGGD <b>HIHA</b> GT <b>VVGKLE</b> GER <b>EMITLGFVDLL</b> 349
gi Raphidiopsis	<b>VIDRQKNHG</b> IHFRVLAK <b>ALRM</b> SGGD <b>HIHT</b> GT <b>VVGKLE</b> GD <b>RAITMGFVDLL</b> 350
gi Fischerella	<b>VIDRQKNHG</b> IHFRVLAKCLRM <b>SGGDHIHT</b> GT <b>VVGKLE</b> GER <b>GITMGFVDLL</b> 350
gi Microcystis	<b>VIDRQKNHG</b> IHFRVLAK <b>CLRL</b> SGGD <b>HL</b> HS <b>GT</b> V <b>VGKLE</b> GER <b>GITMGFVDLM</b> 345
gi Lynngbya	<b>VIDRQKHGM</b> HFRVLAKCLRM <b>SGGDHL</b> HS <b>GT</b> V <b>VGKLE</b> GE <b>KGITMGFVDLM</b> 350
gi Thermosynechococcus	<b>REN</b> Y <b>IEQDR</b> SG <b>Y</b> FTQDWASMPG <b>VM</b> AVASGG <b>IHVWHMPALVDIFGDDAV</b> 399
gi Arthrospira	<b>REDHIE</b> EDRSR <b>G</b> IFFTQDWASMPG <b>VMP</b> VASGG <b>IHVWHMPALVEIFGDDSC</b> 400
gi Zea	<b>RDDFIEKDR</b> RG <b>IF</b> FTQDW <b>V</b> SMPG <b>VIP</b> VASGG <b>IHVWHMPALTEIFGDDSV</b> 399
gi Sorghum	<b>RDDFIEKDR</b> RG <b>IF</b> FTQDW <b>V</b> SMPG <b>VIP</b> VASGG <b>IHVWHMPALTEIFGDDSV</b> 399
gi Oryza	<b>R</b> ----- 350
gi Raphidiopsis	<b>REN</b> Y <b>VEQDKSR</b> G <b>Y</b> FTQDWASMPG <b>VM</b> AVASGG <b>IHVWHMPALVEIFGDDSV</b> 400
gi Fischerella	<b>REN</b> Y <b>VEQDKSR</b> G <b>Y</b> FTQDWASMPG <b>VM</b> AVASGG <b>IHVWHMPALVEIFGDDSV</b> 400
gi Microcystis	<b>REDYVE</b> ED <b>RA</b> RG <b>IF</b> FTQD <b>Y</b> AS <b>LP</b> G <b>VMP</b> VASGG <b>IHVWHMPALVEIFGDDSC</b> 395
gi Lynngbya	<b>REDHIEQDR</b> ERG <b>Y</b> FTQDWASMPG <b>VMP</b> VASGG <b>IHVWHMPALVEIFGDDSC</b> 400
gi Thermosynechococcus	<b>LQFGGGTLGHPWGNAPGATANRVALEACI</b> QAR <b>NEGRD</b> <b>LMREGGDI</b> REAA 449
gi Arthrospira	<b>LQFGGGTLGHPWGNAPGATANRVALEACI</b> QAR <b>NEGRN</b> <b>IFREGGD</b> VIREA <b>C</b> 450
gi Zea	<b>LQFGGGTLGHPWGNAPGA</b> AAN <b>RVALEACV</b> QAR <b>NEGRD</b> LAREG <b>NEIKAA</b> C 449
gi Sorghum	<b>LQFGGGTLGHPWGNAPGA</b> AAN <b>RVALEACV</b> QAR <b>NEGRD</b> LAREG <b>NEIKAA</b> C 449
gi Oryza	-----
gi Raphidiopsis	<b>LQFGGGTLGHPWGNAPGATANRVALEACV</b> QAR <b>NEGRN</b> LAREG <b>NDI</b> REAA 450
gi Fischerella	<b>LQFGGGTLGHPWGNAPGATANRVALEACI</b> QAR <b>NEGR</b> NLAREG <b>NDV</b> IREAA 450
gi Microcystis	<b>LQFGGGTLGHPWGNAPGATANRVALEACI</b> QAR <b>NEGR</b> SLAREG <b>NDV</b> IREA <b>C</b> 445
gi Lynngbya	<b>LQFGGGTLGHPWGNAPGATANRVALEACI</b> QAR <b>NEGRN</b> <b>LMREGGD</b> VIREAA 450
gi Thermosynechococcus	<b>RWSPELAAACELWKEIKFE</b> FEA <b>QDTI</b> ----- 475
gi Arthrospira	<b>KWSPELAV</b> ACELWKEIK <b>FE</b> FEAM <b>DTL</b> ----- 476
gi Zea	<b>KWSA</b> ELAA <b>ACEI</b> WKEIK <b>FDG</b> FKAM <b>DTI</b> ----- 476
gi Sorghum	<b>KWSA</b> ELAA <b>ACEI</b> WKEIK <b>FDG</b> FKAM <b>DTL</b> ----- 476
gi Oryza	-----
gi Raphidiopsis	<b>KWSPELA</b> T <b>ACELWKEIKFE</b> FE <b>TVD</b> T <b>V</b> ----- 476
gi Fischerella	<b>KWSPELAAACELWKEIKFE</b> FE <b>AVD</b> T <b>V</b> ----- 476
gi Microcystis	<b>RWSPELAAACELWKEIKFE</b> FEAM <b>DTL</b> ----- 471
189 gi Lynngbya	<b>KWSPELAVACELWKEIKFE</b> FEAM <b>DTV</b> ----- 476

**Phylogenetic Tree Prediction using Treetop Software:**

Phylogenetic and bioinformatic analysis of RuBisCo provide a useful framework to understand the relationship of the different forms and how they have evolved from a common ancestor. Inclusions of short sequences present a major difficulty in reconstruction of phylogenetic relationships as they limit the number of informative positions in an alignment. A phylogenetic tree based on RuBisCo LS, showed three different clades. The first clade was occupied by the members of Nostocales (*Cylindrospermopsis*) as well as Oscillatoriales (*Geitlerinema* sp. PCC 8501 and *Trichodesmium erythraeum* IMS101). This was supported by bootstrap value of 94%. *Trichodesmium erythraeum* IMS101 formed a slightly distinct sub clade when compared to other strains. Clade two consisted of three sub clades. *Microcystis* of Chroococcales family formed the first sub clade (IIA). The sub clade IIB was found to be intermixed with *Raphidiopsis brookii* D9, *Fischerella* and *Thermosynechococcus vulcanus*. The third sub clade (IIc) was occupied by the members of Oscillatoriales (*Lyngbya* and *Arthrospira*). The above mention topology was supported by high bootstrap values.

The third clade was completely occupied by plant RuBisCo LS *Zea mays* and *Sorghum bicolor* exhibited 100% sequence similarity followed by *Oryza sativa*. In the phylogenetic tree based on LS *Raphidiopsis brookii* D9 and *Fischerella* were found to exhibit high similarity followed by *Lyngbya* and *Arthrospira*. From the predicted phylogenetic trees, *Zea mays* and *Sorghum bicolor* were found to have highly similar sequences when compared to *Oryza sativa*. The homology models of RuBisCo from cyanobacteria and plant species were predicted using SWISS-MODEL (Fig. 3).

**Fig: 3 Modeled structure of RuBisCo from (a) *Geitlerinema* sp. PCC 8501 (b) *Raphidiopsis brookii* D9 and (c) *Arthrospira platensis* NIES-39 viewed from the N-terminal end of the central  $\beta$ -barrel. The models were generated with SWISS MODEL and viewed with PyMOL**



The stereochemical and energetic properties of the models were evaluated with the RAMPAGE servers. According to the Ramachandran plot, 95 to 96.6% residues were in the most favoured regions in LS of cyanobacterial RuBisCo. LS of plant RuBisCo have 96.5% residues in the favoured region. These results indicate that the models were geometrically viable.

During evolution, different organisms have developed diverse arrangements of the RuBisCo L2 building blocks. Form I RuBisCos are the most abundant form found in plants, algae and many photosynthetic bacteria. The L2 subunits in Form I RuBisCos are arranged in an (L2)<sub>4</sub> core with two groups of four small (S) subunits (~13 to 17 kDa) capping the L8 core to form an L8S8 molecule. Although not strictly required for CO<sub>2</sub>-fixation, the S-subunits are essential for maximal activity and provide structural stability [21]. Recent evidence suggests that S-subunits in the unicellular green algae *Chlamydomonas reinhardtii* may also play a structural role in pyrenoid development. RuBisCos classed as Form II and Form III lack S-subunits, containing only L-subunits arranged into L2 to (L2)<sub>5</sub> complexes. These RuBisCos are primarily found in phototrophic proteobacteria, chemoautotrophs, dinoflagellates and Archaea.

**Conclusion:**

Cyanobacterial RuBisCo LSs were selected from NCBI database and characterized for their physico-chemical properties. Molecular weight, isoelectric point, aliphatic index, instability, number of negative residues both positive and negative, Grand Average Hydropathy (GRAVY) and disulfide bond were computed. Physicochemical parameters provided useful data for the purification of cyanobacterial RuBisCo. Secondary structure analysis revealed that the proteins are largely alpha helical; this was supported by the homology modeling of RuBisCo from *Geitlerinema* sp. PCC 8501, *Trichodesmium erythraeum* IMS101, *Arthrospira platensis* NIES-39 and *Thermosynechococcus vulcanus*. Even though LS shows moderate active site and other properties the subunit enhances the functional activity of LS. Putative phosphorylation site can act as an inducer to LS to enhance the activity. The percentage of homology is 85% of *Cylindrospermopsis* to *Zea mays* and *Sorghum bicolor*, 87% to *Oryza sativa japonica* group. The percent of cyanobacterial homology shows 99% of *Cylindrospermopsis* to *Raphidiopsis brookii* D9 and *Lyngbya majuscula* 3L. The predicted phylogenetic trees based on LS *Raphidiopsis brookii* D9 and *Fischerella* were found to exhibit high similarity than other strains. The 3D

model of *Geitlerinema* sp. PCC 8501 RuBisCo suggests that the molecule contain more beta-sheet and alpha helix.

#### Acknowledgement:

This work was financially supported, by Women Scientist Scheme (Basic Research Area) from Department of Science and technology (DST), New Delhi.

#### References:

- [1]Caldwell, PE., MacLean MR and Norris PR, 2007. "Ribulose biphosphate carboxylase activity and a Calvin cycle gene cluster in *Sulfobacillus* species". Microbiology, 153: 2231-2240.
- [2]Long SP, Zhu XG, Naidu SL, Ort DR, 2006. "Can improvement in photosynthesis increase crop yields?" Plant Cell Environ 29: 315–330.
- [3]Portis AR, 1992. "Regulation of ribulose 1,5-bisphosphate carboxylase/oxygenase activity," Annu. Rev. Plant Physiol. Plant Mol.Biol., 43:415-437.
- [4]Spreitzer RJ, 1999. "Questions about the complexity of chloroplast ribulose-1,5-bisphosphate carboxylase/oxygenase," Photosynthesis Research, 60:29-42.
- [5]Leegood, R.C., 2002. "C<sub>4</sub> photosynthesis: Principles of CO<sub>2</sub> concentration and Prospects for its introduction into C<sub>3</sub> plants". J. Exp. Bot., 53: 581-590.
- [6]Raines CA, 2006. "Transgenic approaches to manipulate the environmental responses of the C<sub>3</sub> carbon fixation cycle". Plant Cell Environ, 29: 331–339.
- [7]Spreitzer R.J., 2002. "RuBisCo: Structures, regulatory interactions, and a better enzyme," Annu. Rev. Plant Biol., 53:441-475.
- [8]Smith A.A and Plazas M.C, 2011. "In silico Characterization and Homology Modeling of Cyanobacterial Phosphoenolpyruvate Carboxylase Enzymes with Computational Tools and Bioinformatics Servers". American Journal of Biochemistry and Molecular Biology, 1: 319-336.
- [9]<http://www.cc.ntut.edu.tw/~wwwwoaa/oaa-nwww/oaa-bt/bt-data/lowan/14.pdf>.
- [10]Kannappan B and Gready JE., 2008. "Redefinition of RuBisCo carboxylase reaction reveals origin of water for hydration and new roles for active-site residues". J Am Chem Soc 130:15063-15080.
- [11]Whitney S.M, 2010. "Running head: Understanding and manipulating Rubisco biogenesis", Plant Physiology Preview, 1104: 110.164814.
- [12]Gilbert, D., 2003. "Sequence file format conversion with command-line readseq". Curr. Protoc Bioinformatics, 10.1002/0471250953
- [13]Larkin, M.A., Blackshields G, Brown N.P, Chenna R and McGettigan P.A, 2007. "Clustal W and clustal X version 2.0". Bioinformatics, 23: 2947-2948.
- [14]Rice, P., Longden I and Bleasby A, 2000. "Emboss: The european molecular biology open software suite". Trends Genet., 16: 276-277.
- [15]Deleage G. and Roux B, 1987. "An algorithm for protein secondary structure prediction based on class prediction". Protein Eng., 1: 289-294.
- [16]Kyce J., and Doolittle R.F., 1982, "A simple method for displaying the hydropathic character of a protein", J. Mol. Biol., 157(6): 105-132
- [17]Arnold, K., L. Bordoli, J. Kopp and T. Schwede, 2006. "The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. Bioinformatics", 22: 195-201.
- [18]Vriend, G., 1990. "What If: A molecular modeling and drug design program". J. Mol. Graphics, 8: 52-56.
- [19]Smith, A.A., Coomes M.W and Smith T.E, 2008. "Isolation and sequence of the phosphoenolpyruvate carboxylase gene of the marine cyanobacterium synechococcus PCC 7002". J. Biol. Sci., 8: 1261-1270.
- [20]Kai, Y., Matsumura H, Inoue T, Terada K and Nagara Y., 1999. "Three-dimensional structure of phosphoenolpyruvate carboxylase: A proposed mechanism for allosteric inhibition". Proc. Natl. Acad Sci. USA., 96: 823-828.
- [21]Andersson I, Taylor TC. 2003. "Structural framework for catalysis and regulation in ribulose-1,5-bisphosphate carboxylase/oxygenase". Archives of Biochemistry and Biophysics, 414:130-140.