International Research Journal of Biological Sciences _____ ISSN 2278-3202
Vol. **3(4),** 11-17, April **(2014)** *Int. Res. J. Biological Sci.*

In silico analyses of Rubisco Enzymes from different classes of Algae

Bhavisha P. Sheth* and Vrinda S. Thaker
The Virtual Institute of Bioinformatics, Department of Biosciences, Saurashtra University, Rajkot 360005, Gujarat, INDIA

Available online at: www.isca.in, www.isca.me
Received 21st September 2013, revised 24th October 2013, accepted 3rd December 2013

Abstract

Rubisco (Ribulose 1, 5 Bisphosphate Carboxylase Oxygenase) is the most predominant enzyme of one of the few carbon assimilatory processes in nature i.e. Photosynthesis. The rbcL and rbcS genes code for the large and small subunits of ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) respectively. In this study the rbcL protein sequences selected from various classes of algae were phylogenetically analyzed. Expasy's Prot-param server and Cys_rec tool were used for physico-chemical and functional characterization of these proteins. For comparative structural analysis, experimental structures (X-ray and NMR) of rubisco proteins of representative species of Rhodophyta (Galderia sp. PDBID 1IWA) and Chlorophyta (Chlamydomonas sp. 1GK8) were used. Also, as no experimental structure of rubisco from any member of phaeophyta group is available, homology modeling approach was employed in order to derive structure of the same from Lessonia vadosa, a representative species of phaeophyta group. The validity of the modeled protein was further checked by RAMPAGE, Procheck, WHATIF, Errat, and Verify-3d servers. Studies of secondary structure of these proteins were carried out by the SSCP server. The in silico analysis, confirmed the close correlation between the rhodophyte and the phaeophyte rubico proteins at the functional level due to similarity in adaptability of the enzyme.

Keywords: *rbcL*, rubisco, bioinformatics, phylogenetics, protein.

Introduction

There is a simultaneous increase in the world hunger, with the increase in the global population. Photosynthesis, the process of conversion of light energy to consumable chemical energy, is by far the most predominant of the few carbon assimilatory processes, in contrast to most of the carbon dissimilation processes of nature. Rubisco, responsible for all photosynthetic carbon fixation is the most abundant enzyme on the earth.

The properties of rubisco like effectiveness and specificity determine the photosynthetic efficiency and ultimately the productivity of photosynthetic organisms[1] however, is often thought of as a highly conserved and sluggish enzyme. The *rbcL* gene encodes the large subunit of Rubisco enzyme. As Rubisco catalyzes the rate-limiting step of photosynthesis, the structural basis for catalysis can identify potential targets for useful engineering. However, rubisco enzymes from different species have different catalytic constants. If the structural basis for such differences were known, a rationale could be developed for genetically engineering an improved enzyme. So, the further beneficial genetic engineering of Rubisco may result in substantial increases in crop-plant productivity.

The bibliometric and bioinformatic analysis might offer great help in designing better alternatives of the enzyme *in silico*. Kinetic comparisons of the enzyme with other more potent natural Rubiscos of interest in the evolutionary study will help to authenticate the obtained results. The evolutionary analysis with the integration of bioinformatics tools and experimental validation will bring out the best results in the Rubisco research. The structural analysis of the photosynthetically most efficient, rubisco protein is reported from the red alga. However, no report on brown algae is listed. The present study highlights the correlation between the structural and phylogenetic aspects of the rubisco enzymes, between these major groups of algae i.e. Rhodophyta, Chlorophyta and Phaeophyta and comparison of the same with that of other groups of algae is carried out and the results are discussed.

Material and Methods

Sequence analyses: Protein sequence phylogenetic analysis: The amino acid sequences of secondary metabolite proteins of various algal *rbcL* proteins from three representative families viz. Rhodophyceae (Rh) (Rh1-7), Chlorophyceae (Ch) (Ch1-8) and Phaeophyceae (Ph) (Ph1-10) were retrieved from SWISSPROT, a public resource of curated protein sequences (table 1). The multiple sequence alignment of the sequences used in the study was performed in ClustalX. Further, the evolutionary history was inferred using the Neighbor-Joining method and the bootstrap consensus tree was inferred from 2700 replicates using the MEGA5 software[2] (figure 1).

Physiochemical characterization: For physiochemical characterization, amino acid composition, theoretical instability index, aliphatic index and GRAVY (grand average hydropathy) were computed using the Expasy's ProtParam server for set of proteins[3]. The results are shown in table 2 and table 3.

Structural analyses: The protein 3D structures of the photosynthetically most efficient alga *Galderia partita* (PDBID:

1IWA)[4] and that of a green alga *Clamydomonas reinhardtii* (PDBID: 1GK8)[5] are retrieved from the RCSB PDB. As experimental structures of any organism from the Phaeophyte group are not available, homology modeling approach was used in order to derive the structure of *Lessonia vadosa,* a representative of the Phaeophyte group of algae.

Homology modelling of *Lessonia vadosa* Rubisco: The protein sequence of *Lessonia vadosa,* retrieved from the SwissProt Database was used for BlastP analysis from the NCBI BLAST Suite[6]. Based on high score, lower e-value and maximum sequence identity, the best template (1IWA) was selected which was then used as reference structure to build a 3D model. The homology model can be seen in figure 2.

Model building and evaluation: The three dimensional structures of proteins were modeled using EsyPred Server[7]. Quality of generated models was evaluated with PROCHECK[8] and RAMPAGE[9] server by Ramachandran plot analysis. Validation of generated models was further performed by VERIFY 3D[10], WHATIF[11] and ERRAT[12] programs. The results obtained are shown in figure 3A-D.

Submission of the modeled protein in protein model database (PMDB): The model generated for *Lessonia vadosa* Rubisco was submitted in Protein Model Database, PMDB[13]. The submitted model can be accessed via its PMID.

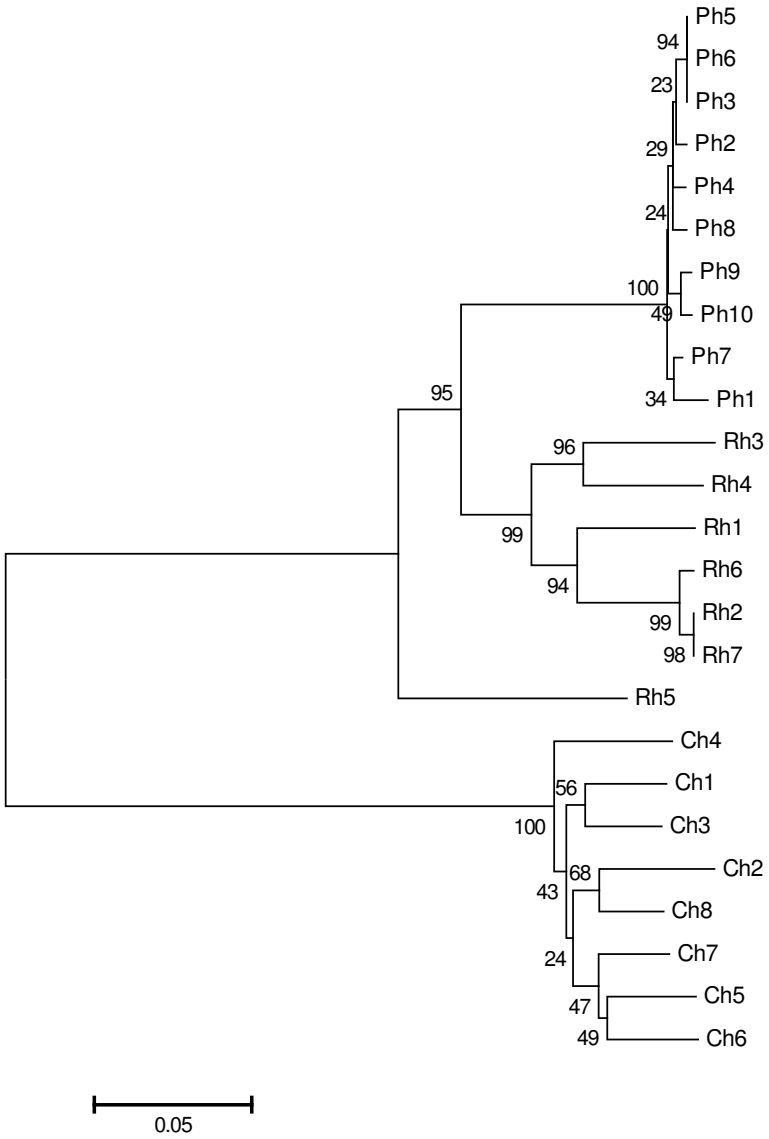


Figure-1
Phylogenetic analysis of *rbcL* protein sequences from different classes of algae using MEGA5 (Ph1-5: Phaeophyta; Rh1-5: Rhodophyta; Ch1-5: Chlorophyta)

Table-1
List of selected *rbcL* protein sequences

Sr.No	Family	Code	Accession No.	Name of organism
1.	Rhodophyta	Rh1	Q09119	*Porphyridium aerugineum*
2.		Rh2	Q760S7	*Porphyra dentata*
3.		Rh3	P24624	*Antithamnion sp.*
4.		Rh4	P48691	*Chondrus crispus*
5.		Rh5	P23755	*Galdieria sulphuraria*
6.		Rh6	Q760R5	*Porphyra haitanensis*
7.		Rh7	Q760T5	*Porphyra yezoensis*
8.	Chlorophyta	Ch1	Q20EX7	*Oltmannsiellopsis viridis*
9.		Ch2	Q0P3J3	*Ostreococcus tauri*
10.		Ch3	P12466	*Chlorella vulgaris*
11.		Ch4	P26958	*Bryopsis maxima*
12.		Ch5	P08211	*Chlamydomonas moewusii*
13.		Ch6	Q2TGZ2	*Dunaliella tertiolecta*
14.		Ch7	B2X1Y2	*Oedogonium cardiacum*
15.		Ch8	Q9T4F2	*Nephroselmis olivacea*
16.	Phaeophyceae	Ph1	A1BNI3	*Pseudolessonia laminarioides*
17.		Ph2	A1BNI4	*Lessonia nigrescens*
18.		Ph3	A1BNI5	*Lessonia tholiformis*
19.		Ph4	A1BNI7	*Lessonia trabeculata*
20.		Ph5	A1BNI9	*Lessonia vadosa*
21.		Ph6	A1BNJ1	*Lessonia sp. CNUK PL216*
22.		Ph7	A1BNJ2	*Costaria costata*
23.		Ph8	A1BNJ3	*Ecklonia radiata*
24.		Ph9	A1BNJ4	*Saccharina sessilis*
25.		Ph10	A1BNJ6	*Macrocystis pyrifera*

Table-2
Physicochemical characterization of selected proteins

Sr. No.	Accession No. (Uniprot)	Instability Index	Aliphatic Index	Grand Average of Hydropathicity (GRAVY)
1	Q09119	30.25	84.36	-0.171
2	Q760S7	32.25	85.39	-0.142
3	P24624	29.92	88.77	-0.116
4	P48691	25.50	83.98	-0.084
5	P23755	33.14	87.30	-0.195
6	Q760R5	30.95	85.39	-0.136
7	Q760T5	31.54	85.39	-0.142
8	Q20EX7	40.38	76.20	-0.330
9	Q0P3J3	35.54	79.12	-0.276
10	P12466	40.32	78.72	-0.301
11	P26958	39.52	79.54	-0.252
12	P08211	39.45	78.67	-0.247
13	Q2TGZ2	40.69	78.86	-0.254
14	B2X1Y2	40.82	79.94	-0.262
15	Q9T4F2	35.14	79.75	-0.243
16	A1BNI3	25.60	86.56	-0.081
17	A1BNI4	28.38	86.97	-0.089
18	A1BNI5	28.90	86.76	-0.094
19	A1BNI7	28.60	87.38	-0.089
20	A1BNI9	28.96	86.76	-0.095
21	A1BNJ1	28.96	86.76	-0.095
22	A1BNJ2	26.48	86.17	-0.094
23	A1BNJ3	27.11	86.97	-0.087
24	A1BNJ4	28.54	86.97	-0.092
25	A1BNJ6	28.37	87.38	-0.077

Table-3
Amino acid composition of *rbcL* proteins from various classes of algae

Sr. No.	Accession No.	Amino Acid (%)																					
		A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	O	U
1	P23755	10.1	5.7	4.5	5.5	0.6	3.2	6.3	8.1	2.0	5.7	9.1	5.5	3.4	3.9	3.7	3.9	6.5	1.6	4.1	6.7	-	-

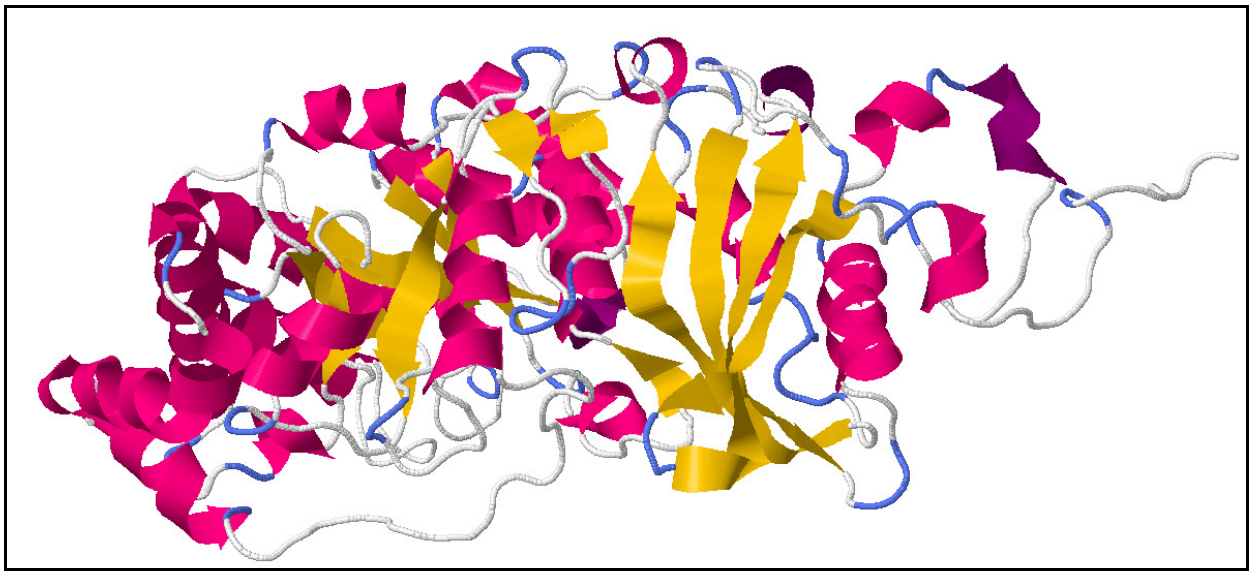
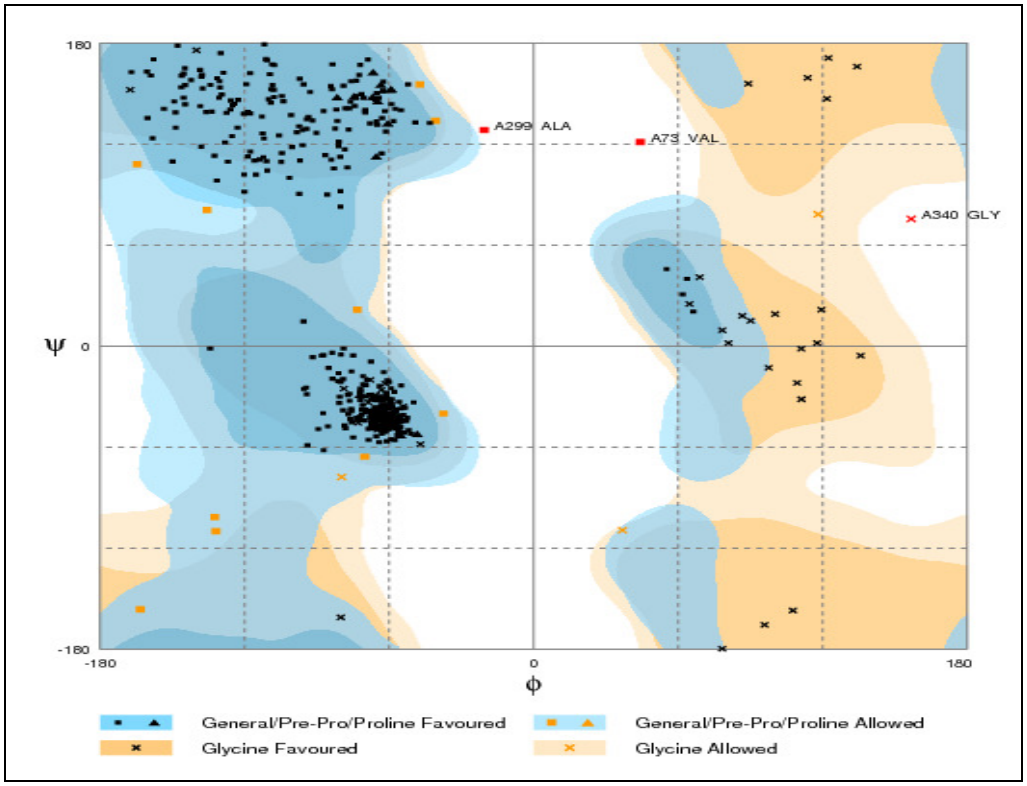
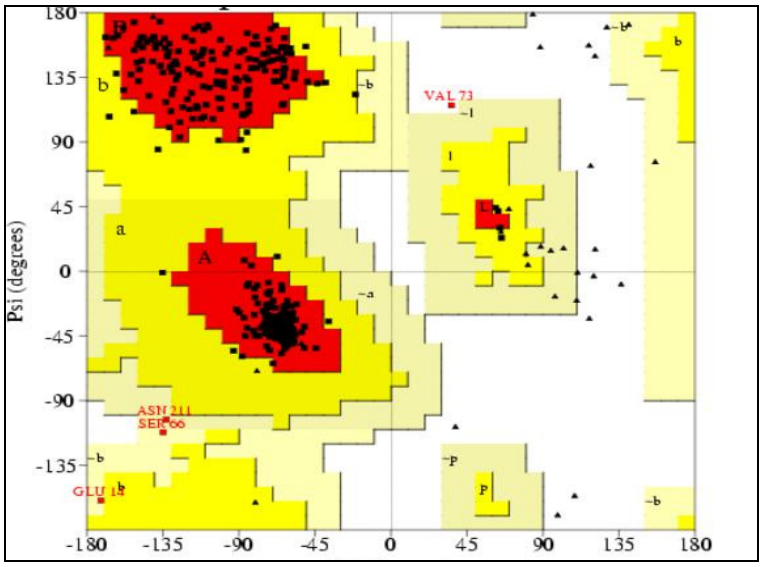


Figure-2
Homology modelling of *Lessonia vadosa* rubisco (viewed using Jmol) (PMDBID: PM0078256)



Number of residues in favoured region (~98.0% expected): 456 (96.8%), Number of residues in allowed region (~2.0% expected): 12 (2.5%), Number of residues in outlier region: 3 (0.6%)

Figure-3A
Ramchandran Plot analysis using RAMPAGE



94.1% residues in favourable regions; 4.9% residues in additional residue regions; 0.7% residues in generously regions; 0.2% residues in disallowed regions; Over all G-factor: -0.01

Figure-3B
Ramchandran Plot analysis using PROCHECK

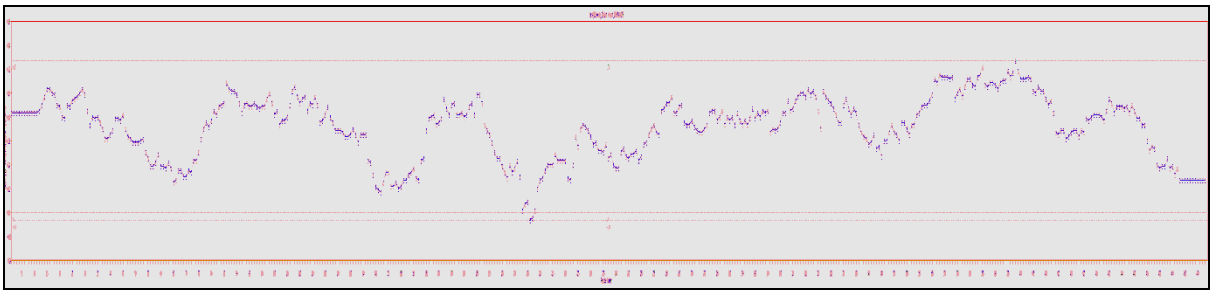


Figure-3C
Verify-3D plot for validity check of the modeled structure

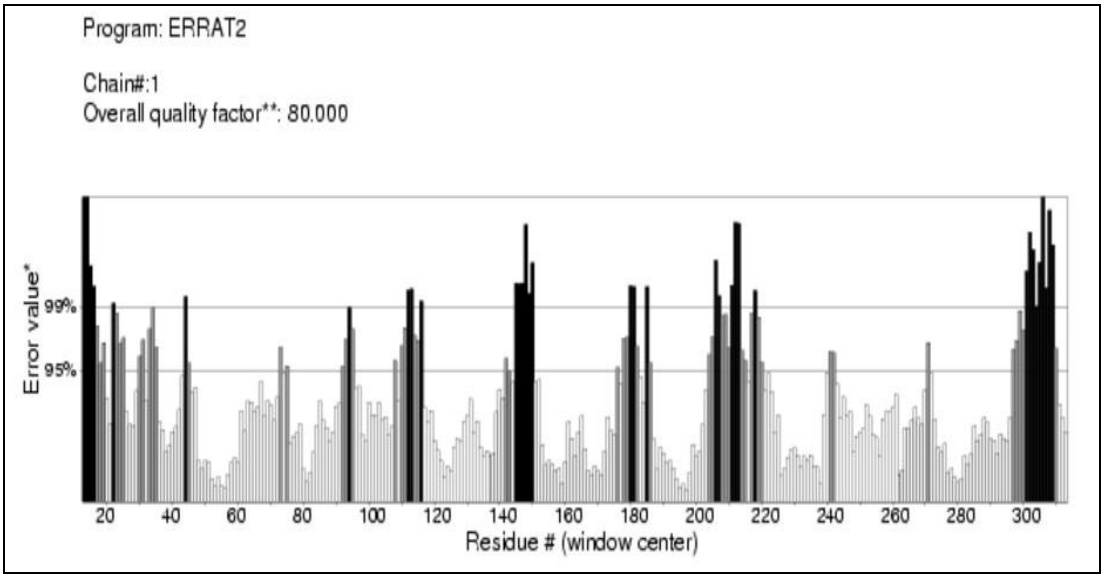


Figure-3D
Errat Plot for validity check of the modeled structure

Functional characterization: CYS_REC (http://sunl.softberry.com/berry.phtml/topic) was used to locate "SS bond" between the pair of cystein residues, if present. The tool yields position of cysteins, total number of cysteins present and pattern, if present, of pairs in the protein sequence as output. The sequences for the chosen structures were only chosen for the determination of this parameter. The results are presented in table-4.

Secondary structure content prediction: Secondary structure elements prediction was performed using the Secondary Structural Content Prediction (SSCP) server[14]. The results are shown in table-4.

Table-4
Functional characterization and secondary structure content prediction

Accession Number	Cys_Rec	Score	Alpha helix	Beta turns	Coils
P23755	No S-S bonds	Negative values	43.4	17.0	39.5
P08211	172-192	57.7	19.8	23.3	56.9
	449-459	85.1			
A1BNI9	No S-S bonds	Negative values	51.7	8.5	39.9

Superimposition and pairwise structural alignment: The selected 3D structures of the protein of the Chlorophyte (1GK8) and Phaeophyte (PM0078256) groups are superimposed and aligned pairwise with help of SuperPose version 1.0 [15], which is a protein superposition server using the rhodophyte protein (1IWA) as the template and the chlorophyte (1GK8) and phaeophyte (PM0078256) proteins as the targets . Root Mean Square Deviation (RMSD) values were calculated between the set of targets and template protein to see how much modeled protein deviates from the template protein structure. The results are shown in table-5.

Table-5
Pairwise structural alignment among the members of different classes of algae

Template	Target (PDB/ PMDB ID)	RMSD
1IWA (*Galdieria sp.)*	1GK8 (*Chlamydomonas sp.)*	1.369
	PM0078256 (*Lessonia vadosa)*	0.712

Results and Discussion

Sequence analyses: The Phylogenetic analyses of *rbcL* proteins gives important clues as each clade of the Phylogenetic tree comprised of the members of the same group viz. Rhodophyta, Chlorophyta or Phaeophyta (figure-1). Similar results were also observed elsewhere[16]. The physicochemical analyses of the *rbcL* was carried out using the ProtParam server (table-2), according to which a protein whose instability index is larger than 40 may be unstable[17]. Hence, the *rbcL* proteins from members of Rhodophyta and Phaeophyta groups were highly stable whereas those from Chlorophyta group were unstable or near to unstable. The aliphatic index of the proteins was found to be between 76.2-88.77; the higher indices suggesting they may be stable at high temperatures[18]. The GRAVY values[19] fall in range of -0.077 to -0.330; indicating that the proteins will interact favourably with water.

Structural analyses: The 3D structures of the two organisms, one each from Rhodophyta (*Galderia sp.)*[4] and Chlorophyta groups (*Chlamydomonas sp.)*[5] were further used for structural analyses. As there was no availability of experimental structures for any member of the Phaeophyta group, the 3D structure of the rubisco from *Lessonia vadosa*, a representative of the Phaeophyta group was modeled using the EsyPred Server (figure-2). The validity of the model was checked using RAMPAGE, Procheck, WHATIF, Errat and Verify-3d servers. Ramachandran plot for the same has been illustrated in figure-3A (RAMPAGE) and figure-3B (PROCHECK). Altogether more than 90% of the residues were found to be in favoured and allowed regions, which validate the quality of homology model suggesting the acceptability of the modeled structure. Ramachandran Z-score as inferred from the WHATIF Server is 0.457. The score expressing how well the backbone conformations of all residues correspond to the known allowed areas in the Ramachandran plot, is within expected ranges for well-refined structures. The modeled structures were also validated by other structure verification servers such as Verify 3D and Errat (figure-3C, figure-3D). Verify 3D assigned a 3D-1D score of >0.2 for 90.51% of the residues, representing the acceptability of the model**.** This implies that the model was compatible with its sequence. ERRAT showed overall quality factor of 80.00 for the model. The validated homology model was thus submitted to the Protein Model Database with the PMDBID: PM0078256.

The pairwise structure alignment was carried out using the red algal Rubisco (*Galderia sp.)*, known to be the most potent rubisco enzyme, as template with that of the 3D structures of Chlorophyte (*Chlamydomonas sp.*) and Phaeophyte (*Lessonia vadosa*) proteins. RMSD (Root Mean Square Deviation) values obtained were 1.369 and 0.712 respectively, suggesting a greater similarity of the Phaeophyte protein with the red algal protein (table-5). Also the evolutionary closeness of the red and brown algal proteins is evident from the phylogenetic sequence analyses (figure-1).

The present study shows close correlation between the Rhodophyte and the Phaeophyte rubisco at the functional level due to similarity in adaptability of the enzyme, which is comparable with ecology of both the type of organisms. The members of the Rhodophyte and Phaeophyte groups thrive in the marine habitat in contrast to most of the members of the Chlorophyta surviving in the freshwater habitat. Rubisco is the predominant enzymatic mechanism in the biosphere by which algae and other photosynthetic organisms fix CO_2 into organic biomass via the Calvin–Benson–Basham reductive pentose phosphate pathway[1]. Rubisco being the chief enzyme, in

photosynthesis, is acted upon by all forces of evolution[20]. Hence, adaptation of the organisms at their respective ecological niche plays an important role in the stereochemistry and structural similarities between their rubisco enzymes as evident from the present study between the members of different classes of algae.

Conclusion

The relationship of rubisco among the major three classes of algae was undertaken at the functional level by selection of the protein sequences (*rbcL*) and protein structures. The *in silico* studies indicated a close correlation between that of Rhodophyta and Phaeophyta at the sequence as well as the structural levels. This is very well indicated in the phylogenetic analysis as well as the structural alignment analysis. The members of Rhodophyta and Phaeophyta belong to the same clade, indicating phylogenetic relatedness. Further, relatedness of brown and red algal protein structure is also evident by pair wise structural alignment. These results suggest that Phaeophyta may have more efficient Rubisco than that of Chlorophyta.

Acknowledgement

The instrumentation and computation facility under CAS (Gujarat govt.) and BIT Virtual, GSBTM, DST (Gujarat govt.) scheme are gratefully acknowledged. Financial support to the first author in form of UGC Research Fellowship in Sciences for Meritorious Students is also highly acknowledged.

References

1. Spreitzer R.J. and Salvucci M.E., Rubisco: structure, regulatory interactions, and possibilities for a better enzyme, *Annu. Rev. Plant Biol.* **53,** 449–475 **(2002)**

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

3. Wilkins M. R., Gasteiger E., Bairoch A., Sanchez J. C., Williams K. L., Appel R.D. and Hochstrasser D. F., *Methods Mol. Biol.* **112**, 531–552 **(1999)**

4. Sugawara H., Yamamoto H., Shibata N., Inoue T., Okada S., Miyake C., Yokota A. and Kai Y., Crystal structure of carboxylase reaction-oriented ribulose 1, 5-bisphosphate carboxylase/oxygenase from a thermophilic red alga, *Galdieria partita*. *J Biol Chem,* **274**,15655-61 **(1999)**

5. Taylor T. C., Backlund A., Bjorhall K., Spreitzer R. J. and Andersson, I., First crystal structure of Rubisco from a green alga, *Chlamydomonas reinhardtii*. *J Biol Chem* **276(51)**, 48159-48164 **(2001)**

6. Altschul S. F., Gish W., Miller W., Myers E. W. and Lipman, D. J., Basic local alignment search tool. *J. Mol. Biol*., **215(3)**, 403-410 **(1990)**

7. Lambert C., Léonard N., De Bolle X. and Depiereux E., ESyPred3D: Prediction of proteins 3D structures, *Bioinformatics* **18(9)**, 1250-1256 **(2002)**

8. Laskowski R. A., MacArthur, M. W., Moss, D. S. and Thornton J. M., PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr.,* **26(2)**, 283-291 **(1993)**

9. Lovell S.C., Davis I.W., Arendall W.B., de Bakker P.I., Word J.M., Prisant M.G., Richardson J.S. and Richardson D.C., Structure validation by Calpha geometry: phi, psi and Cbeta deviation, *Proteins*, **50**, 437–450 **(2003)**

10. Luthy R., Bowie J. U. and Eisenberg D. Assessment of protein models with three-dimensional profiles, *Nature*, **356(6364)**, 83-85 **(1992)**

11. Vriend G. WHAT IF: a molecular modeling and drug design program, *J. Mol. Graphics*, **8(1)**, 52-56 **(1990)**

12. Colovos C. and Yeates T. O., Verification of protein structures: patterns of nonbonded atomic interactions, *Protein Sci.,* **2(9)**, 1511-1519 **(1993)**

13. Castrignanò T., De Meo P. D. O., Cozzetto D., Talamo I. G. and Tramontano A. The PMDB protein model database. *Nucleic Acids Res.,* **34(1)**, D306-D309 **(2006)**

14. Eisenhaber F., Imperiale F., Argos P. and Frömmel C., Prediction of secondary structural content of proteins from their amino acid composition alone. I. New analytic vector decomposition methods. *Proteins: Struct. Func. Bioinformatics*, **25(2)**, 157-168 **(1996)**

15. Maiti R., Van Domselaar G. H., Zhang H. and Wishart D. S., SuperPose: a simple server for sophisticated structural superposition, *Nucleic Acids Res.,* **32(2)**, W590-W594 **(2004)**

16. Sheth B.P. and Thaker V.S. *rbcL*: a key to C value paradox in plants, *Plant Arch.*, **12(2)**, 915-919 **(2012)**

17. Guruprasad K., Reddy B. B. and Pandit M. W., Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence, *Protein Eng.,* **4(2)**, 155-161 **(1990)**

18. Atsushi Ikai., Thermostability and aliphatic index of globular proteins, *J. Biochem.* **88(6)**, 1895-1898 **(1980)**

19. Kyte J. and Doolittle R. F., A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.,* **157(1)**, 105-132 **(1982)**

20. Wang M., Kapralov M. and Anisimova M. Coevolution of amino acid residues in the key photosynthetic enzyme Rubisco. *BMC Evol. Biol.*, **11(1)**, 266 **(2011)**

21. (http://us.expasy.org/tools/protparam.html)