JSMARTech
Journal of Smart Bioprospecting and Technology
April 2025, VOL 06 NO 01
p-ISSN: 2686-0805 **e-ISSN**: 2714-7894
DOI: https://doi.org/10.21776/ub.jsmartech.2025.006.01.04

Comparative *In-silico* Physicochemical and Structural Analysis of *RuBisCo* Proteins in Green, Red, and Brown Seaweeds Using Bioinformatics Tools

Muthusamy Thangaraj

Centre of Advanced Study in Marine Biology, Faculty of Marine Sciences,
Annamalai University, Parangipettai, Tamil Nadu, India-608 502

Submission: 19 October 2024; **Revised**: 3 February 2025; **Accepted**: 25 March 2025

e-mail: coralholders@gmail.com

ABSTRACT. *In bioinformatics, the comparison of protein structures is crucial for understanding their function, evolution, and relationships. This study focuses on Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) protein structures compared in three seaweeds: Ulva lactuca (green algae), Gracilaria edulis (red algae), and Sargassum fusiforme (brown algae). The physicochemical characteristics, secondary structures, and 3D structural properties of these proteins were analyzed using bioinformatics tools. The results revealed that the lowest percentage of amino acid composition was Cys (0.8%) in G. edulis, while the highest was Gly (10%) in U. lactuca. Additionally, three amino acids—Lys, Ser, and Trp—were found in similar proportions across all three seaweeds. The highest molecular weight (MW) of RuBisCO was predicted as 54.11 kDa in G. edulis. Based on theoretical pI values, all proteins were identified as acidic, with pI values ranging from 5.62 to 5.98. The instability index (II) values were 27.05 for S. fusiforme, 29.11 for G. edulis, and 37.38 for U. lactuca, indicating that the proteins are stable. The grand average of hydropathicity (GRAVY) values, calculated as -0.110, -0.127, and -0.263, suggested that the proteins are hydrophilic and water-soluble. Secondary structure analysis showed that RuBisCO proteins primarily consist of α-helices and β-sheets, which contribute to their structural stability. The 3D structures of these proteins were modeled using the Swiss Model platform and validated using PROCHECK and QMEAN tools. The findings from this study offer valuable insights into the structural and functional characteristics of RuBisCO proteins in different seaweeds, potentially aiding future research in marine biology and protein evolution.*

Keywords: *Seaweeds, RuBisCO, SOPMA, PROCHECK, QMEAN*

INTRODUCTION

Seaweeds are macroalgae (non-flowering, stemless aquatic flora) commonly found attached to solid substrates such as rocks, dead corals, shells, and other organic materials. They thrive in relatively shallow coastal waters, estuaries, intertidal zones, and deep-sea regions up to depths of 180 meters [1]. They are generally classified into three major groups: Chlorophyceae (green algae), Phaeophyceae (brown algae), and Rhodophyceae (red algae). Globally, approximately 10,000 species of seaweeds have been identified, with 271 genera and 1,153 species of marine algae, including forms and varieties, reported from Indian waters [2].

Seaweeds are highly valued worldwide for their rich content of vitamins, minerals, and dietary fiber [1]. Historically, they have been used in traditional medicine, particularly in China, Japan, Europe, and North America, due to their beneficial effects on human health. They were believed to have curative properties for ailments such as tuberculosis, arthritis, colds, and influenza. While earlier uses of seaweeds were limited to traditional and folk medicine [3], the development of new techniques in the 1990s enabled the isolation of bioactive compounds, garnering significant attention from the pharmaceutical industry and modern biological research [4].

Photosynthesis, a fundamental process in plants, involves the reduction of CO_2 in the presence of photosynthetic pigments, light, and water to produce food substances [5]. RuBisCO (ribulose-1,5-bisphosphate carboxylase/oxygenase EC 4.1.1.39) is a key enzyme in this process, facilitating both reductive and oxidative photosynthetic carbon cycles. It directly influences the net photosynthetic rate and is essential for the autotrophy of plant cells [6]. RuBisCO is considered the most abundant protein on Earth and is found in all autotrophic organisms, including bacteria, algae, and plants [7].

Previous studies have successfully utilized bioinformatics tools to characterize RuBisCO proteins, but these efforts have largely focused on terrestrial plants [8,9]. To date, no comprehensive

Table 1. Physicochemical parameters of *RuBisCO*

Parameters	***Ulva lactuca***	***Gracilaria edulis***	***Sargassum fusiforme***
Accession No.	YP_009633165.1	YP_009731956.1	YP_009828273.1
Number of amino acids	474	488	488
Negatively charged residues	61	56	59
Positively charged residues	51	50	50
Molecular weight (KDa)	52.32	54.11	53.99
Theoretical pI	5.82	5.98	5.62
The instability index	37.38	29.11	27.05
Aliphatic index	78.48	86.35	86.76
GRAVY	-0.263	-0.127	-0.110

study has examined the structural and physicochemical properties of RuBisCO proteins in seaweeds. Addressing this gap, the present study analyzed the large subunit of RuBisCO in three seaweed species—*Ulva lactuca*, *Gracilaria edulis*, and *Sargassum fusiforme*. Using bioinformatics tools, the physicochemical characteristics, secondary structures, and 3D structural properties of these proteins were predicted and evaluated.

RESEARCH METHODS

The RuBisCO protein sequences of the selected seaweed species were retrieved in FASTA format from the National Center for Biotechnology Information (NCBI) protein database (https://www.ncbi.nlm.nih.gov/protein) on 10.7.2023. The physicochemical properties of the RuBisCO proteins, including isoelectric point (pI), molecular weight (MW), total number of positive (R) and negative (-R) residues, instability index (II), aliphatic index (AI), and GRAVY values, were analyzed using the ExPASy - ProtParam tool (http://web.expasy.org/protparam/) [10].

Putative phosphorylation sites in the RuBisCO proteins were predicted using NetPhos 3.1 (http://www.cbs.dtu.dk/services/NetPhos/) [11]. Secondary structural features of the RuBisCO sequences were calculated using SOPMA (Self-Optimized Prediction Method with Alignment) [12] with default parameter settings (window width: 17, similarity threshold: 8, number of states: 4). Protein domains and conserved motifs were identified using protein BLAST in the MEME Suite (http://meme-suite.org/doc/fasta-format.html) [13].

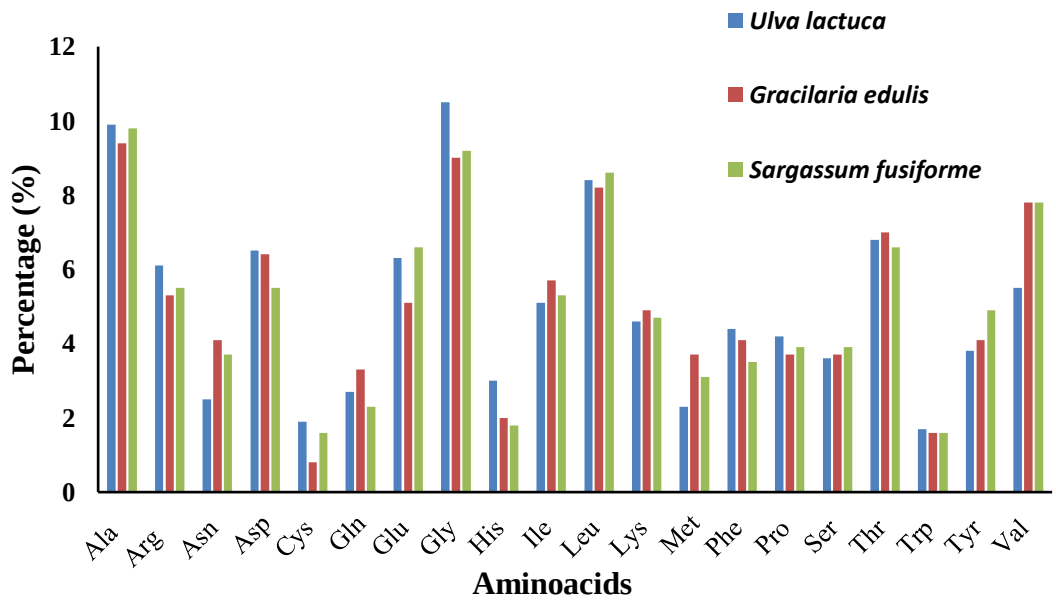

Figure 1. The percentage composition of amino acids of *RuBisCO* in three seaweeds

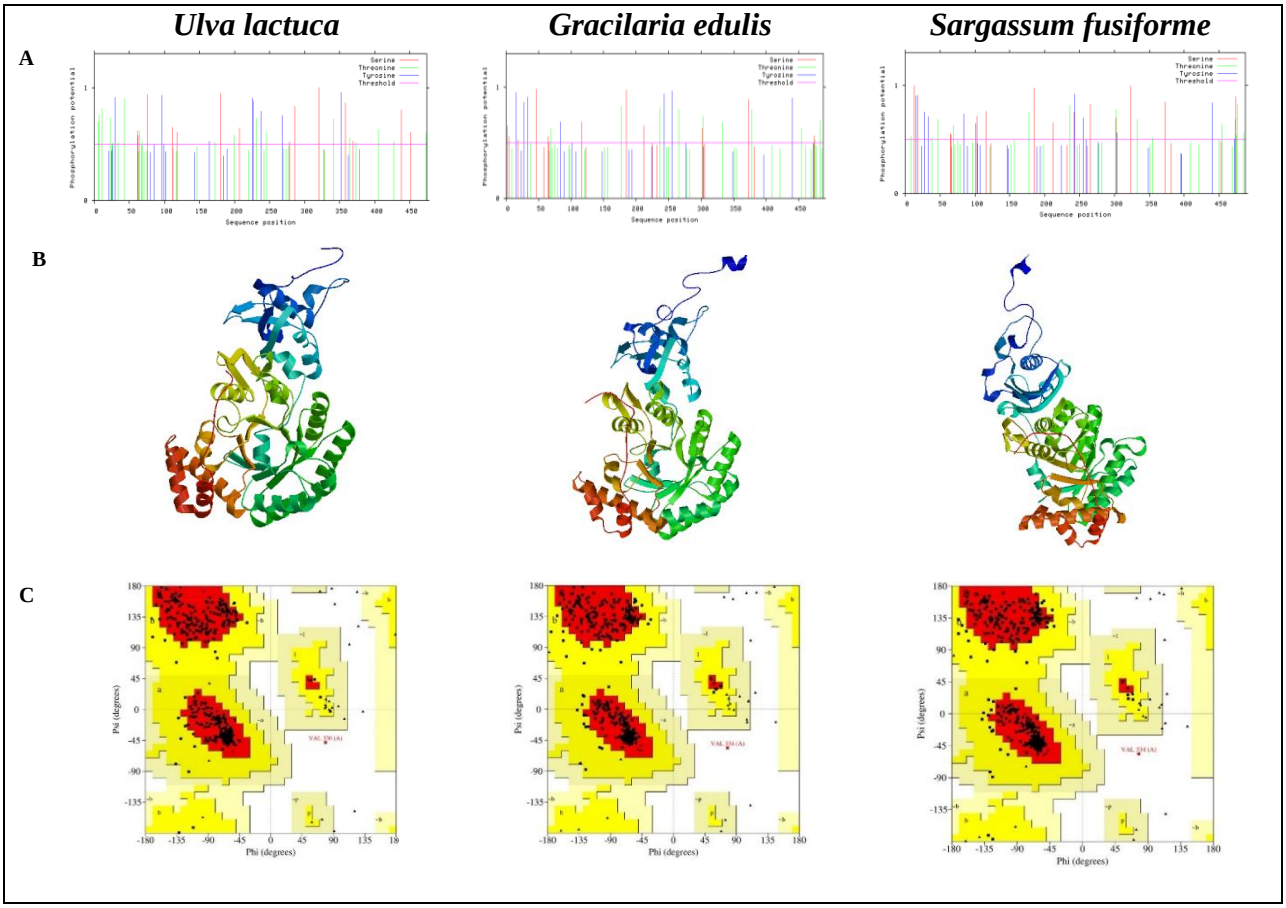

Figure 2. A. Putative phosphorylation sites; B. 3-D modelled structure; C. Ramachandran plot of *RuBisCO*

The three-dimensional structures of the RuBisCO proteins were modeled using the Swiss Model server [14] with preferred parameter settings. Suitable template structures for homology modeling were selected by performing a PDB-BLAST search against the Protein Data Bank (PDB). Among the 50 templates identified, the crystal structure of RuBisCO from *Chlamydomonas reinhardtii* (PDB ID: 2v63.1.A) showed the highest similarity (92.62%) to the target sequences. The Swiss Model server was used to generate a 3D structure of the target protein using 2v63.1.A as the template. The rough model was saved in .pdb format and visualized using the Swiss-PDB Viewer.

The stereochemical quality of the predicted protein models was evaluated using Ramachandran plot analysis in RAMPAGE (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) [15]. The 3D models were further validated using PROCHECK [16], QMEAN (Quantitative Model Energy Analysis) [17], and QMEAN-DisCo [18]. Tertiary structural analyses and visual representations of the models were performed using Swiss-PDB Viewer [19].

RESULTS AND DISCUSSIONS

The physicochemical properties of RuBisCO proteins from the three seaweeds were analyzed using ExPASy - ProtParam, and the results are presented in Table 1. The amino acid sequence length was 474 in *U. lactuca* and 488 in both *G. edulis* and *S. fusiforme*. The amino acid compositions are shown in Figure 1. The lowest percentage of amino acids was Cys (0.8%) in *G. edulis*, while the highest was Gly (10%) in *U. lactuca*. Lys, Ser, and Trp were present in similar proportions across all three seaweeds. Notably, negatively charged amino acids were more abundant in *U. lactuca* compared to the other two species. The highest molecular weight (MW) was 54.11 kDa in *G. edulis*.

Table 2. Secondary structure characters of *RuBisCO*

Parameter	***Ulva lactuca***	***Gracilaria edulis***	***Sargassum fusiforme***
Alpha helix	40.71%	38.93%	29.30%
Extended strand	18.98%	22.33%	22.74%
Beta turn	8.43%	8.19%	9.63%
Random coil	31.85%	30.53%	26.43%

Table 3. Conserved motifs and their dis-similarity index of *RuBisCO*

Motif number	Motif width	1	2	3	4	5	6
1	49	--					
2	50	0.14	--				
3	50	0.11	0.17	--			
4	50	0.14	0.17	0.17	--		
5	49	0.13	0.21	0.19	0.13	--	
6	50	0.15	0.18	0.13	0.13	0.20	--
7	41	0.19	0.19	0.22	0.20	0.20	0.24

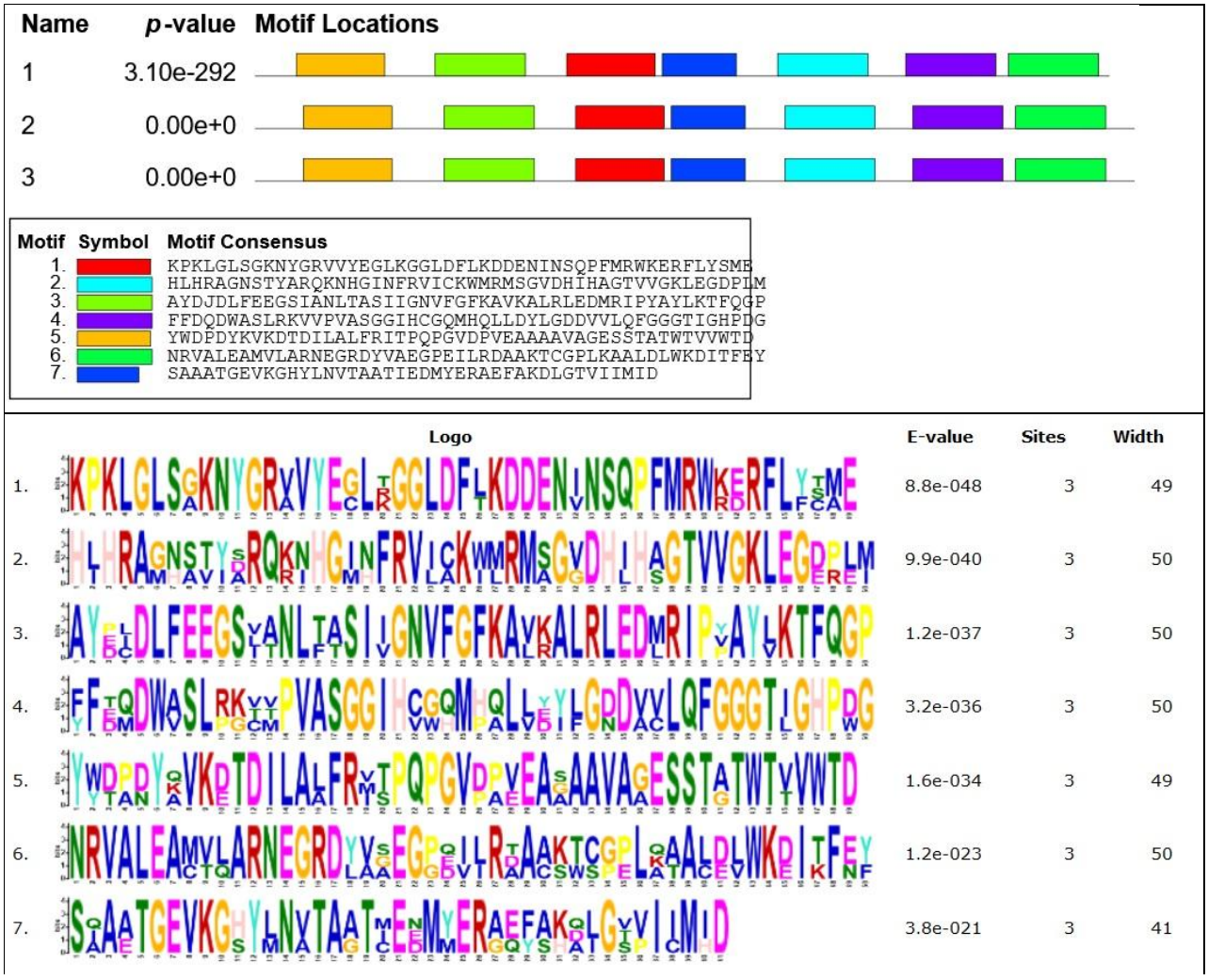

Figure 3. Conserved motif structure of *RuBisCO* in three seaweeds

The theoretical pI values indicated that RuBisCO proteins in all three seaweeds were acidic, with values ranging from 5.62 to 5.98. The instability index (II) values were 27.05 (*S. fusiforme*), 29.11 (*G. edulis*), and 37.38 (*U. lactuca*), confirming the stability of these proteins. The aliphatic index (AI), which indicates protein thermostability based on the abundance of Ala, Ile, Leu, and Val, ranged from 78.48 to 86.76 (Table 1). The grand average of hydropathicity (GRAVY) values, which reflect hydrophilicity (- values) or hydrophobicity (+ values), were -0.110 (*U. lactuca*), -0.127 (*G. edulis*), and -0.263 (S. fusiforme), confirming that RuBisCO proteins in these seaweeds are hydrophilic and water-soluble.

Phosphorylation sites were predicted using NetPhos 3.1, with results depicted in Figure 2A. Phosphorylation is a common post-translational modification crucial for protein adaptation and function [22, 23]. Among the three species, *S. fusiforme* had the most phosphorylated residues (75), while *U. lactuca* had the least (67).

Table 4. Ramachandran plot calculation of the Swiss model computed with PROCHECK

Residues in plot	***Ulva lactuca***	***Gracilaria edulis***	***Sargassum fusiforme***
In the most favoured region	94.7%	92.8%	92.0
In additionally allowed region	5.1%	7.0%	7.7%
In disallowed region	0.3%	0.2%	0.2%

Table 5. Structural validation by QMEAN

Seaweeds	***Ulva lactuca***	***Gracilaria edulis***	***Sargassum fusiforme***
GMQE	0.94	0.94	0.94
QMEAN Disco Global	0.89 ± 0.05	0.89 ± 0.05	0.88 ± 0.05
Function terms	**QMEAN Z-score value**		
C-beta	-0.0863	-0.0749	-0.380
Interaction	-0.0925	-0.3061	-0.555
Packing	-0.8236	-0.2226	-0.1249
Torsion	-0.7681	-0.4335	-0.5310
SS-Agreement	-0.8675	-0.2936	-0.5765
QMEAN4	-0.9840	-0.5215	-0.6495

Secondary structural features of RuBisCO were predicted using SOPMA, and the results are summarized in Table 2. The secondary structure of this protein was predominantly composed of α-helices and β-sheets, indicating structural stability. In *U. lactuca*, 41.71% of residues formed α-helices, 18.98% formed extended structures, 8.43% formed β-turns, and 31.85% formed random coils. In *G. edulis*, 38.93% formed α-helices, 22.33% formed extended structures, 8.19% formed β-turns, and 30.53% formed random coils. For *S. fusiforme*, 29.30% formed α-helices, 22.74% formed extended structures, 9.63% formed β-turns, and 26.43% formed random coils.

Domains and conserved motifs of RuBisCO were analyzed using MEME, as shown in Table 3 and Figure 3. Seven conserved motifs were identified across the sequences, with motif widths ranging from 41 to 50. Among these, motifs 3 and 7 exhibited greater dissimilarity, while motifs 1 and 3 were highly conserved.

The three-dimensional structure of RuBisCO was modeled using Swiss Model, with visualization performed in Swiss-PDB Viewer (Figure 2B). Homology modeling identified Chlamydomonas reinhardtii RuBisCO (PDB ID: 2v63.1.A) as the most suitable template, with 92.62% similarity. The modeled structures were validated using the Ramachandran plot and PROCHECK (Table 4, Figure 2C). In *U. lactuca*, 94.7% of residues were in the favored region, 5.1% in the additionally allowed region, and 0.3% in the outlier region. For *G. edulis*, 92.8% were in the favored region, 7% in the additionally allowed region, and 0.2% in the outlier region. In *S. fusiforme*, 92% were in the favored region, 7.7% in the additionally allowed region, with the remainder in the outlier region.

The amino acid Gly, known for its structural flexibility due to its small size, was the most abundant residue in *U. lactuca* (50) and least abundant in *G. edulis* (44). Cysteine provides strong antioxidant power through thiol-based ROS scavenging, disulfide bond formation, and its role in glutathione. Phenylalanine offers indirect antioxidant benefits through its aromatic ring's electron-donating properties and by affecting the peptide's structural stability. So these bioactive dipeptides (Phe and Cys), which contribute to antioxidant properties, were also detected in the RuBisCO proteins of all three seaweeds as reported earlier [24, 25].

The quality of the 3D structures was evaluated using QMEAN, with negative z-scores confirming high structural accuracy (Table 5). Previous studies have used bioinformatics tools to analyze RuBisCO proteins in terrestrial plants, including Sideritis species [9], *Triticum aestivum* [26], and Asteraceae species [27]. These studies highlight the utility of bioinformatics tools in understanding protein function, evolutionary relationships, drug discovery, and structural validation.

CONCLUSION

Bioinformatics tools have significantly simplified the understanding of protein structures and functions by leveraging advanced software applications. In this study, bioinformatic analyses of the RuBisCO protein in three seaweed species were conducted using tools such as ExPASy - ProtParam and NetPhos 3.1. The three-dimensional structure of RuBisCO was modeled using the Swiss Model, and the accuracy of the model was validated using PROCHECK and QMEAN tools.

Experimental validation of these findings, alongside further in-silico studies with additional algal species, could provide deeper insights into the evolutionary adaptations of RuBisCO structures. The observed hydrophilicity and thermostability of RuBisCO in these seaweeds suggest potential applications in algae-based biofuel production and other biotechnological industries.

ACKNOWLEDGEMENT

The author is grateful Dean and Director, CAS in Marine Biology and Authorities of Annamalai University for providing the necessary facilities and encouragement. The author thank the authorities for providing fund in the form of research project (DRD/RUSA 2.0/R&I/Project Proposal/Filed 5/8/2021) to carry out this work.

REFERENCES

[1] Pati MP, Sharma SD, Nayak L, Panda CR (2016). Uses of seaweeds and its applications to human welfare: A review. *Int J Phar Pharmaceut Sci*., 8(10): 12-20.

[2] Subba Rao PV, Mantri A. (2006). Indian seaweed resources and sustainable utilization: scenario at the dawn of a new century. *Curr Sci*., 91:164-174.

[3] Lincoln RA, Strupinski K, Walker JM. (1991). Bioactive compounds from algae. Life Chem. Rep., 8:97-183.

[4] Ireland CM, Copp BR, Foster MP, Mcdonald LA, Radisky DC, Swersey JC (1993). Biomedical potential of marine natural products. Marine Biotechnology, Pharmaceutical, and Bioactive Natural Products: Plenum Press NY, 1:1-43.

[5] Türk M, Çelik N. (2006). CO2 Özümlemesinde C-3 ve C-4 Tipi Bitkilerde Fotosentez-Solunum Denge Noktalarının Belirlenmesi. *SDÜ Fen Bil Enst Der*., 10(1):48-51.

[6] Mei Y, Li HL, Xie J, Luo HY. (2007). Ribulose-1,5-bisphosphate Carboxylase/oxygenase (Rubisco). *Plant Physiol Commun*., 43:363-368.

[7] Udenigwe CC, Gong M, Wu S. (2013). *In silico* analysis of the large and small subunits of cereal RuBisCO as precursors of cryptic bioactive peptides. *Process Biochem*., 48(11): 1794-1799.

[8] Thangaraj M. (2023). *In silico* analysis and homology modeling of ribulose-1,5-bisphosphate carboxylase/oxygenase in green seaweed, *Ulva fasciata*. *J Biol Nature*, 15(1):1-6.

[9] Sevindik E. (2019). Comparative and phylogenetic analysis of rubisco large subunit (rbcl) proteins in some Sideritis L. (Lamiaceae) species: a bioinformatic approach. *Genetika*, 51(1): 69-80.

[10] Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. (2005). Protein identification and analysis tools on the ExPASy server, In: John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press, 571–607.

[11] Blom N, Gammeltoft S, Brunak S. (1999). Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J Mol Biol*., 294(5): 1351-1362.

[12] Geourjon C, Deléage G. (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl Biosci*., 11:681-684.

[13] Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res*., 37(2): 202-208.

[14] Arnold K, Bordoli L, Kopp J, Schwede T. (2006). The SWISS-Ma web-based environment for protein structure homology modelling. *Bioinformatics*, 22: 195-201.

[15] Lovell SC, Davis IW, Arendall WB, Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC. (2003). Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. *Proteins: Structure, Function, and Bioinformatics*, 50(3): 437-450.

[16] Laskowski, RA, Rullmannn JA, Macarthur MW, Kaptein R, Thornton JM. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR*., 8:477-486.

[17] Benkert P, Biasini M, Schwede T. (2011). Toward the estimation of the absolute quality

of individual protein structure models. *Bioinformatics*, 27: 343-350.

[18] Studer G, Rempfer C, Waterhouse AM, Gumienny G, Haas J, Schwede, T. (2020). QMEANDisCo - distance constraints applied on model quality estimation. *Bioinformatics*, 36: 1765-1771.

[19] Guex N, Peitsch MC. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, 18(15): 2714-2723.

[20] Idicula-Thomas S, Balaji PV. (2005). Understanding the relationship between the primary structure of proteins and their amyloidogenic propensity: clues from inclusion body formation. *Prot Eng Des Sel*., 18:175-180.

[21] Kyte J, Doolittle RF. (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol*., 157:105-142.

[22] Vener AV. (1990). Protein phosphorylation: a motive force for adaptive evolution. *BioSystems*, 24:53-59.

[23] Fangru N, Yuxin H, Xudong L, Jia F, Junping L, Qi L, Shulian X. (2020). Analysis of adaptive evolution and coevolution of rbcL gene in the genus Hildenbrandia (Rhodophyta). *Evol Bioinfo*., 16: 1-7.

[24] Yan BX, Sun YQ. (1997). Glycine Residues Provide Flexibility for Enzyme Active Sites. *J Biol Chem*., 272: 3190- 3194.

[25] Je JY, Cho YS, Gong M, Udenigwe CC. (2015). Dipeptide Phe-Cys derived from *in silico* thermolysin hydrolysed RuBisCO large subunit suppresses oxidative stress in cultured human hepatocytes. *Food Chem*., 171: 287-291.

[26] Naeem MK, Rauf S, Iqbal H, Nawaz Shah MK, Mir A. (2012). *In silico* studies of C3 metabolic pathway proteins of wheat (*Triticum aestivum*). *BioMed Res Int*., 2013: 1-7.

[27] Avci MK, Tezcan E, Sevindik E. (2016). Genome-wide identification and comparative structural analysis of RuBisCO proteins in the asteraceae. *Hortic Environ Biotech*., 57: 404-414.