J. Plant Biol. (2017) 60:35-47
DOI 10.1007/s12374-016-0211-4

ORIGINAL ARTICLE

Computational Study of Biochemical Properties of Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase (RuBisCO) Enzyme in C_3 Plants

Maryam Darabi[1,*] and Samin Seddigh[2]

[1]*Department of Agronomy and Plant Breeding Sciences, College of Aboureihan, University of Tehran, Tehran, Iran*
[2]*Department of Plant Protection, College of Agriculture, Varamin-Pishva Branch, Islamic Azad University, Varamin, Iran*

Received: April 22, 2016 / Accepted: November 7, 2016

Abstract Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO; EC 4.1.1.39) is a key enzyme in plant photosynthesis that catalyzes the reaction in the photosynthetic assimilation of atmospheric carbon dioxide (CO_2). In this study, bioinformatic characterization of RuBisCO enzyme was performed in protein sequences of its large subunits (rbcL) from 25 C_3 plants of 14 different families. In protein-protein interactions analysis by STRING 10 tool, 7 and 168 relevant datasets were identified in *Arabidopsis thaliana* and other species, respectively. The structural and functional analyses were investigated by ProtParam, SOPMA, Predotar 1.03, SignalP 4.1, TargetP 1.1, and TMHMM 2.0 tools in ExPASy database. Tertiary structure was predicted by Phyre2 and TM-score servers then their qualities were verified by PROCHECK and SuperPose servers. In MEME and MAST analyses, 9 common conserved motifs obtained in all C_3 plants. The protein sequences were aligned with ClustalW algorithm by MEGA 6.06 software and phylogenetic tree was constructed using the Neighbor-joining (NJ) method. According to the results, there is a high identity of RuBisCO in different species of C_3 plants so that they should be derived from a common ancestor. The results provide background of bioinformatic studies for the function and evolution of RuBisCO in other plants.

Key words: Bioinformatics, Phylogenetic analysis, Protein-protein interaction, Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase (RuBisCO), Structural analysis

Introduction

Photosynthesis, the process of alteration of light energy to consumable chemical energy, is the most crucial of the few carbon assimilatory processes, in contrast to most of the carbon dissimilation processes of nature. Ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) (EC 4.1.1.39), the most abundant enzyme on the earth, is responsible for all photosynthetic carbon fixations. The characteristics of this enzyme determine the photosynthetic efficiency and eventually the productivity of photosynthetic organisms however, is often thought of as a highly conserved and sluggish enzyme (Sheth and Thaker 2014).

Therefore, RuBisCO is an enzyme involved in the first main step of carbon fixation, a process by which energy-rich molecules such as glucose is made of atmospheric carbon dioxide by plants. It catalyzes the carboxylation of ribulose-1,5-bisphosphate (RuBP) in the Calvin cycle. The Calvin cycle or C_3 cycle is a series of biochemical redox reactions which take place in the stroma of chloroplasts in photosynthetic organisms (Fig. 1). During carbon fixation, the substrate molecules for RuBisCO are ribulose-1,5-bisphosphate, carbon dioxide (distinct from the "activating" carbon dioxide). RuBisCO also catalyzes a reaction between ribulose-1,5-bisphosphate and molecular oxygen (O_2) instead of carbon dioxide (CO_2) (Feller et al. 2008). In the other word, RuBisCO transforming the carbon dioxide and ribulose-1,5-bisphosphate (RuBP) into two molecular 3-phosphoglyceric acid, catalyzes the first reaction of carbon dioxide fixation in photosynthetic dark reaction. Also, RuBisCO catalyzes the reaction of oxygen and RuBP to phosphoglyceric acid and phosphoglycollic acid, which is the first reaction of photorespiration. Consequently, RuBisCO is the significant enzyme deciding the photosynthetic efficiency by regulating photosynthesis and photorespiration (Andersson et al. 1989).

RuBisCO is a very large and complex protein. In plants, algae, cyanobacteria, and phototropic and *chemoautotropic proteobacteria* the enzyme usually consists of two types of protein subunits, called the large subunit (LSU, about 55,000 Da) and the small subunit (SSU, about 13,000 Da) (Dhingra et al. 2004). The LSU gene is part of the chloroplast DNA

*Corresponding author; Maryam Darabi
E-mail : darabi.ma@yahoo.com

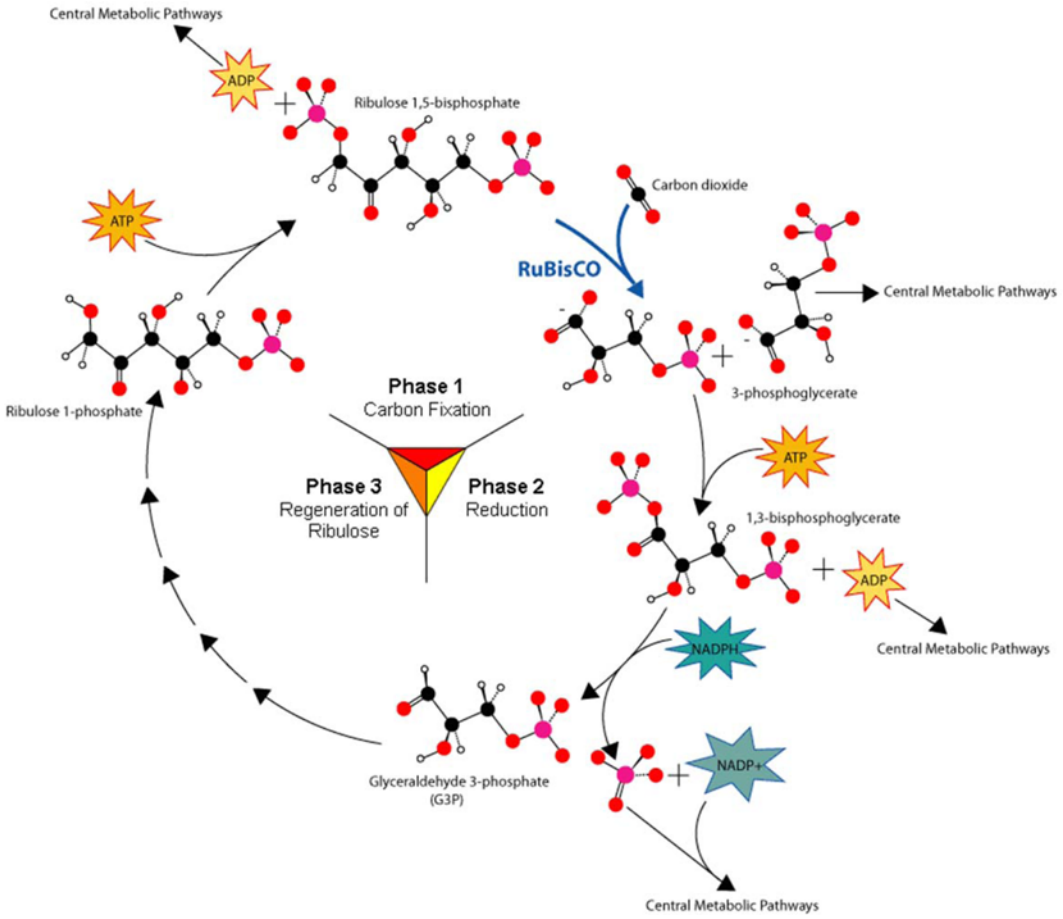

Fig. 1. The major steps in the Calvin cycle.

molecule in plants. There are several related SSU genes in the nucleus of plant cells and the SSU are imported to the stromal compartment of chloroplasts from the cytosol by crossing the outer chloroplast membrane. The enzymatically active substrate (ribulose 1,5-bisphosphate) binding sites are located in the large chains which form dimers in which amino acids from each large chain contribute to the binding sites. A total of eight large chain dimers and eight small chains assemble into a larger complex of about 540,000 Da. Crystallography has been revealed that in some proteobacteria and dinoflagellates, RuBisCO contain only large subunits (Chatterjee and Basu 2011).

On the basis of amino acid sequences, four known forms of RuBisCO are found in nature as forms I, II, III and IV (Tabita et al. 2007). The most plentiful form of RuBisCO is the form I protein which found among higher plants, eukaryotic algae, cyanobacteria and proteobacteria. This is the classic high molecular weight protein originally found in plants and once called Fraction One Protein (Tabita 1999). Form I RuBisCO comprises large approximately 50 kDa (catalytic) subunits encoded by either the rbcL or the cbbL genes. There are four subclasses of form I RuBisCO, termed IA, IB, IC and ID, found in various organisms (Tabita 1999). Form II RuBisCO is found in different types of proteobacteria with the most studied member being from *Rhodospirillum rubrum* and dinoflagellates which is a group of eukaryotes. It covers only large subunits that share approximately 30% sequence identity with form I large subunits (Tabita 1999). Like form I RuBisCOs, form II catalyzes the carboxylation of RuBP as the first step of the CBB pathway (Tabita et al. 2007; Tabita et al. 2008a; Tabita et al. 2008b). The small subunit that improves catalysis in form I has not been found to be associated with form II RuBisCOs (Warlick 2013). Form III RuBisCO first discovered in 1999 in *hyperthermophilic archaeons*. That was an amazing discovery because these are enzymes which catalyze reactions utilizing either CO_2 or O_2 and they had evolved in the absence of oxygen (Watson et al. 1999). Form III RuBisCOs catalyzes the carboxylation reaction with an extreme sensitivity to O_2 (Warlick 2013). This form is found only in archaea. In these organisms, the enzyme basically serves as a means to remove RuBP, which is produced by the isomerization of ribose 1,5-bisphosphate during purine/pyrimidine metabolism (Finn and Tabita 2004; Sato et al. 2007). Some examinations revealed that the form

IV sequences were missing active site residues crucial to the carboxylation reaction. Lack of these residues implied that these enzymes could not catalyze the carboxylation reaction and they were then named "RuBisCO-like proteins" (Tabita 1999). So, form IV is also called the RuBisCO-like protein (RLP) (Hanson and Tabita 2001). Yet, there are similarities in the primary (Hanson and Tabita 2001) and tertiary (Li et al. 2005; Imker et al. 2007) structures of these proteins that clearly indicate that RLPs are homologues of RuBisCO and are derived from some common ancestor (Tabita et al. 2007).

Recent phylogenetic and bioinformatic analyses of RuBisCO and RLP amino acid sequences provide a useful framework to understand the relationship of the different forms and how they may have evolved from a common ancestor (Tabita et al. 2007). Moreover, structural and functional studies impinge on these analyses as a coherent picture begins to emerge as to how the active site of this protein might have evolved and became adapted to different intracellular milieus. The total conclusion from these studies was that a form III RuBisCO from a methanogenic archaeon ancestor was the most probable source of all RuBisCO and RLP lineages. Undoubtedly, by finding new sequences and models these hypotheses may be re-examined (Tabita et al. 2008a; Tabita et al. 2008b).

Most of the established research on RuBisCO is based on those isolated from *Spinacia oleracea* or *Nicotiania tabacum* (Tabita et al. 2007; Tabita et al. 2008a; Tabita et al. 2008b). The small subunit of RuBisCO has been revealed to be very essential for catalysis, while the exact function is still unknown (Andrews and Ballment 1983, Spreitzer 2003). Spinach RuBisCO has a quaternary structure of eight large subunits and eight small subunits (L_8S_8). Subsequent high resolution crystal structure data of RuBisCO showed an N-terminal β-sheet with three α-helices and a C-terminal $(\beta/\alpha)_8$ barrel (Schneider et al. 1986). As with triose phosphate isomerase and most other $(\beta/\alpha)_8$ barrel proteins, the active site is located at the C-terminal end of the barrel. The RuBisCO active site is also at the interface of the two domains with the acid/base residues coming from the loops of the barrel and substrate specificity provided by N-terminal residues from another RuBisCO molecule (Andersson et al. 1989).

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

Table 1. The list of twenty-five C_3 plants of fourteen different families in the study

	Scientific Name	Type	Family	Abbreviation	Accession Number	Length (aa)
1	*Arabidopsis thaliana* (L.) Heynh.	Thale cress (eudicots)	Brassicaceae	*At*-RuBisCO	NP_051067.1	479
2	*Arachis hypogaea* L.	Peanut (eudicots)	Fabaceae	*Ah*-RuBisCO	AAB67895.1	465
3	*Beta vulgaris* L.	Beet (eudicots)	Amaranthaceae	*Bv*-RuBisCO	AAY68360.1	475
4	*Brassica napus* L.	Rape (eudicots)	Brassicaceae	*Bn*-RuBisCO	AAF78948.1	479
5	*Citrus sinensis* (L.) Osbeck	Sweet orange (eudicots)	Rutaceae	*Csi*-RuBisCO	YP_740483.1	475
6	*Coffea Arabica* L.	Coffee (eudicots)	Rubiaceae	*Ca*-RuBisCO	YP_817490.1	481
7	*Cucumis sativus* L.	Cucumber (eudicots)	Cucurbitaceae	*Csa*-RuBisCO	AAZ94659.1	482
8	*Equisetum hyemale* L.	Dutch rush (ferns)	Equisetaceae	*Eh*-RuBisCO	YP_007374723.1	475
9	*Glycine max* (L.) Merr.	Soybean (eudicots)	Fabaceae	*Gm*-RuBisCO	YP_538747.1	475
10	*Gossypium hirsutum* L.	Upland cotton (eudicots)	Malvaceae	*Gh*-RuBisCO	YP_538943.1	480
11	*Hordeum vulgare* var. nudum Spenn.	Barley (monocots)	Poaceae	*Hv*-RuBisCO	AGP50763.1	479
12	*Medicago sativa* L.	Lucerne (eudicots)	Fabaceae	*Ms*-RuBisCO	CAA94031.1	455
13	*Nicotiana tabacum* L.	Common tobacco (eudicots)	Solanaceae	*Nt*-RuBisCO	NP_054507.1	477
14	*Oryza sativa* subsp. indica Kato	Rice (monocots)	Poaceae	*Os*-RuBisCO	YP_654221.1	484
15	*Phaseolus vulgaris* L.	String bean (eudicots)	Fabaceae	*Pv*-RuBisCO	YP_001122790.1	476
16	*Picea abies* (L.) H.Karst.	Norway spruce (seed plants)	Pinaceae	*Pa*-RuBisCO	CAA53209.1	475
17	*Pisum sativum* L.	Pea (eudicots)	Fabaceae	*Ps*-RuBisCO	YP_003587524.1	475
18	*Populus trichocarpa* Torr. & A.Gray	Black cottonwood (eudicots)	Salicaceae	*Pt*-RuBisCO	YP_001109509.1	475
19	*Ricinus communis* L.	Castor bean (eudicots)	Euphorbiaceae	*Rc*-RuBisCO	AEJ82563.1	475
20	*Secale cereal* L.	Rye (monocots)	Poaceae	*Sc*-RuBisCO	YP_008239180.1	476
21	*Solanum lycopersicum* L.	Tomato (eudicots)	Solanaceae	*Sl*-RuBisCO	YP_008563096.1	477
22	*Solanum tuberosum* L.	Potato (eudicots)	Solanaceae	*St*-RuBisCO	ABB90049.1	477
23	*Triticum aestivum* L.	Bread wheat (monocots)	Poaceae	*Ta*-RuBisCO	NP_114267.1	477
24	*Vicia faba* L.	Fava bean (eudicots)	Fabaceae	*Vf*-RuBisCO	AGS43944.1	475
25	*Vitis vinifera* L.	Wine grape (eudicots)	Vitaceae	*Vv*-RuBisCO	YP_567084.1	475

analysis with the integration of bioinformatics tools and experimental validation will bring out the best results in the RuBisCO research. The present study highlights the correlation between the structural and phylogenetic aspects of the RuBisCO enzymes, between some families of seed plants and the results are discussed.

Results

In the present study, twenty-five RuBisCO protein sequences of fourteen different families in C_3 plants including Amaranthaceae, Brassicaceae, Cucurbitaceae, Equisetaceae, Euphorbiaceae, Fabaceae, Malvaceae, Pinaceae, Poaceae, Rubiaceae, Rutaceae, Salicaceae, Solanaceae, and Vitaceae were retrieved from NCBI with FASTA format. Their characteristics including scientific names, types, families, abbreviations, accession numbers, and length of amino acid sequences were listed in Table 1.

In order to predict the interacting proteins, *At*-RuBisCO was mapped to the STRING 10 tool as a sample of C_3 plants. STRING is a database of known and predicted protein interactions. The interactions include direct (physical) and indirect (functional) associations. According to the results, ten functional partners and thirteen enriched pathways of KEGG were identified in the network analysis (Table 2, Fig. 2).

To investigate the structures and functions of RuBisCO proteins, twenty-five sequences of fourteen different families in C_3 plants were analyzed by bioinformatics tools. The primary structure of RuBisCOs was analyzed by ProtParam server. ProtParam computes various physico-chemical properties that can be deduced from a protein sequence including the length, molecular weight, isoelectric point (pI), extinction coefficient, instability index, aliphatic index and grand average of hydropathicity (GRAVY) (Table 3). Secondary structure prediction was performed by SOPMA (Self-Optimized

Table 2. Characteristics of input protein (*At*-RuBisCO), predicted functional partners and KEGG pathways in STRING 10 tool

Input Protein		Information	
RBCL		Ribulose-bisphosphate carboxylases (479 aa)	
Predicted Functional Partners			
	Name	Information	Score
1	RBCS1A	Ribulose bisphosphate carboxylase small chain 1A (180 aa)	0.999
2	RBCS2B	Rubisco small subunit 2B (181 aa)	0.999
3	RBCS1B	Rubisco small subunit 1B (181 aa)	0.999
4	RBCS3B	Rubisco small subunit 3B (186 aa)	0.996
5	PSBA	photosystem II reaction center protein A (353 aa)	0.981
6	RbcX2	Chaperonin-like RbcX protein (203 aa)	0.975
7	ACCD	Acetyl-CoA carboxylase carboxyl transferase subunit beta (488 aa)	0.963
8	PGK1	Phosphoglycerate kinase 1 (481 aa)	0.956
9	PGK	Phosphoglycerate kinase (401 aa)	0.956
10	PRK	Phosphoribulokinase (395 aa)	0.954
KEGG Pathways			
	ID	Term	P-value
1	00710	Carbon fixation in photosynthetic organisms	7.07E-15
2	01200	Carbon metabolism	2.03E-13
3	01120	Microbial metabolism in diverse environments	5.86E-12
4	01100	Metabolic pathways	3.38E-8
5	00630	Glyoxylate and dicarboxylate metabolism	1.31E-7
6	00010	Glycolysis / Gluconeogenesis	1.68E-2
7	01230	Biosynthesis of amino acids	7.28E-2
8	01110	Biosynthesis of secondary metabolites	9.28E-2
9	00640	Propanoate metabolism	1
10	01212	Fatty acid metabolism	1
11	00195	Photosynthesis	1
12	00061	Fatty acid biosynthesis	1
13	00620	Pyruvate metabolism	1

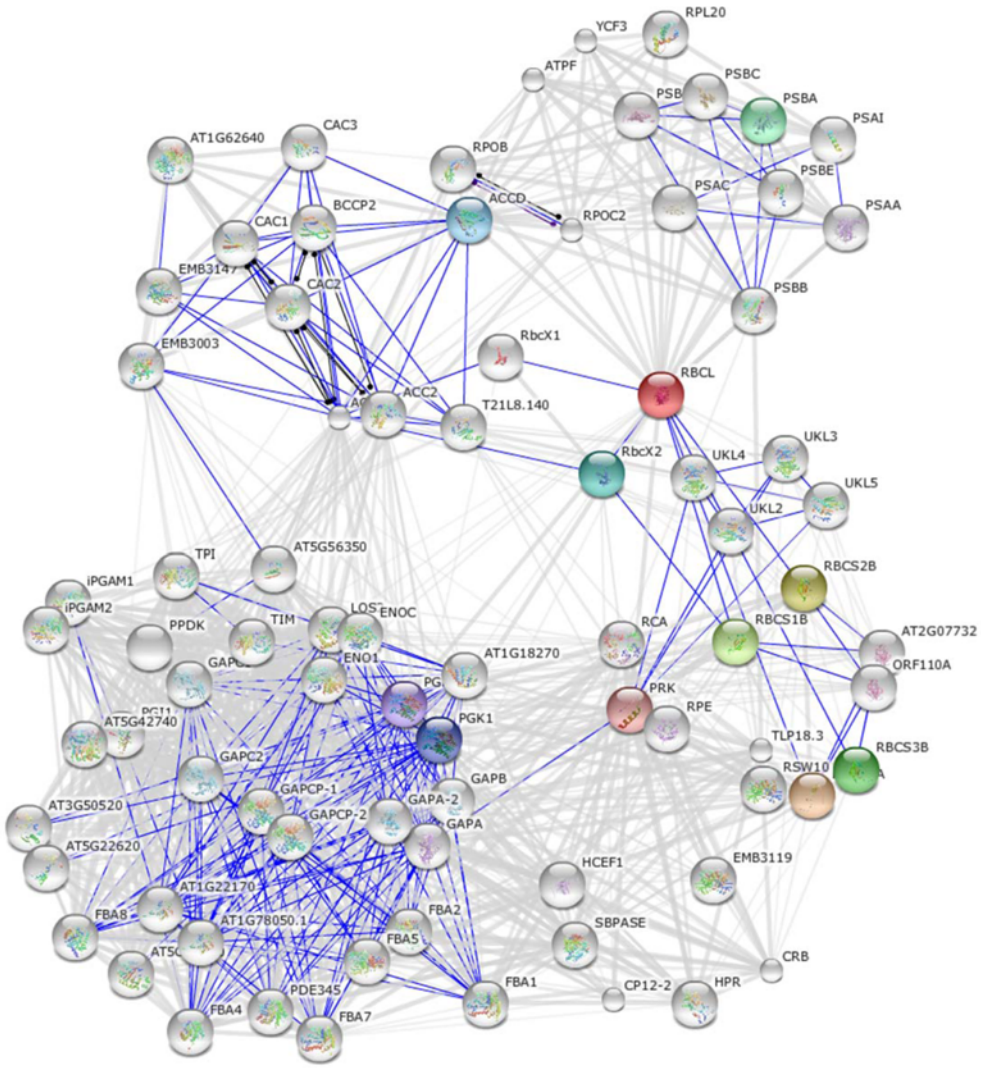

Fig. 2. Protein-protein interactions using STRING 10 tool. In the interactive network view of predicted proteins, nodes are proteins and the edges represent the predicted functional associations. Small nodes: protein of unknown 3D structure; Large nodes: some 3D structure is known or predicted; Colored nodes: query proteins and first shell of interactors; White nodes: second shell of interactors; Blue lines: gene co-occurrence.

Prediction Method with Alignment) server. This server measures the percent of alpha-helix, extended strand, beta-sheet and random coil in protein sequences. Based on the results, the majority parts of the RuBisCOs are alpha helix and random coil (Table 3).

Post-translational modification predictions were analyzed by Predotar 1.03 and SignalP 4.1 servers. Predotar recognizes the N-terminal targeting sequences of classically targeted precursor proteins. For each protein sequence, Predotar provides a probability estimate as to whether the sequence contains a mitochondrial, plastid or ER targeting sequence. The fourth number (elsewhere) is easily the estimated probability that no targeting sequence is present. These estimates assume that the sequence in question was randomly chosen from a proteome in which about 10% of proteins are targeted to mitochondria, 10% to plastids and 20% to the ER. Based on the results of Predotar server, all of the RuBisCO proteins had no targeting sequences (Table 4). SignalP 4.1 server predicts the presence and location of signal peptide cleavage sites in protein sequences. The method incorporates a prediction of cleavage sites and a signal peptide/non-signal peptide prediction based on a combination of several artificial neural networks. A signal peptide is a short (5-30 amino acids long) peptide present at the N-terminus of the majority of newly synthesized proteins that are intended towards the secretory pathway. Based on the results of SignalP analysis, there was no signal peptide in RuBisCO proteins in twenty-five sample of C_3 plants (Table 4).

Topology prediction was carried out by TargetP 1.1 and TMHMM 2.0 servers. TargetP 1.1 predicts the subcellular

Table 3. The results of primary structure analysis and secondary structure prediction in twenty-five different species of C_3 plants

No.	Name	Tools										
		ProtParam							SOPMA			
		Length (aa)	Molecular weight (Mw)	Isoelectric point (pI)	Extinction coefficient	Instability index	Aliphatic index	GRAVY[a]	Alpha helix (%)	Extended strand (%)	Beta turn (%)	Random coil (%)
1	*Ah*-RuBisCO	465	51525.6	6.19	69330-69830	37.71	82.90	-0.228	37.20	18.06	13.55	31.18
2	*At*-RuBisCO	479	52955.0	5.87	69330-69830	42.31	79.44	-0.272	33.82	18.79	12.73	34.66
3	*Bn*-RuBisCO	479	52956.0	5.87	69330-69830	41.07	78.83	-0.278	35.91	17.95	12.32	33.82
4	*Bv*-RuBisCO	475	52519.6	6.33	72310-72810	42.00	80.32	-0.245	34.32	20.42	13.47	31.79
5	*Ca*-RuBisCO	481	53353.6	6.13	69330-69830	40.57	79.13	-0.295	38.05	17.88	14.14	29.94
6	*Csa*-RuBisCO	482	53418.7	6.00	69330-69830	39.79	78.57	-0.270	39.00	17.43	13.28	30.29
7	*Csi*-RuBisCO	475	52518.7	6.29	70820-71320	37.47	80.95	-0.240	37.47	18.11	14.95	29.47
8	*Eh*-RuBisCO	475	52494.6	5.86	67840-68215	38.87	80.95	-0.234	38.53	18.53	13.89	29.05
9	*Gh*-RuBisCO	480	53281.6	6.00	69330-69955	40.77	79.27	-0.283	39.17	17.71	13.96	29.17
10	*Gm*-RuBisCO	475	52609.7	6.00	69330-69830	35.42	79.71	-0.265	40.21	17.26	13.05	29.47
11	*Hv*-RuBisCO	479	53078.4	6.22	73340-73965	44.03	76.81	-0.275	35.07	18.16	13.36	33.40
12	*Ms*-RuBisCO	455	50317.2	6.22	67840-68215	39.16	82.13	-0.223	38.24	20.00	13.41	28.35
13	*Nt*-RuBisCO	477	52898.2	6.41	70820-71320	38.10	82.62	-0.241	37.53	19.92	14.26	28.30
14	*Os*-RuBisCO	484	53701.0	6.33	69330-69955	44.59	76.22	-0.298	36.36	18.18	13.02	32.44
15	*Pa*-RuBisCO	475	52482.6	6.13	66350-66850	43.03	78.06	-0.272	39.37	16.42	13.26	30.95
16	*Ps*-RuBisCO	475	52763.1	6.55	72310-72810	39.90	80.53	-0.272	37.47	18.53	13.05	30.95
17	*Pt*-RuBisCO	475	52600.7	6.00	69330-69830	38.32	80.53	-0.266	37.47	17.89	14.53	30.11
18	*Pv*-RuBisCO	476	52800.9	5.96	69330-69830	38.36	80.57	-0.270	38.66	17.86	13.45	30.04
19	*Rc*-RuBisCO	475	52637.8	6.13	69330-69830	42.10	79.52	-0.269	35.58	18.74	14.11	31.58
20	*Sc*-RuBisCO	476	52749.9	6.04	73340-73965	43.63	76.68	-0.275	34.66	18.07	13.45	33.82
21	*Sl*-RuBisCO	477	52954.3	6.55	69330-69830	37.79	82.41	-0.240	35.64	21.17	14.68	28.51
22	*St*-RuBisCO	477	52944.3	6.55	69330-69830	37.95	81.61	-0.247	38.36	19.71	14.05	27.88
23	*Ta*-RuBisCO	477	52851.1	6.22	73340-73965	44.03	76.52	-0.278	35.01	17.82	13.00	34.17
24	*Vf*-RuBisCO	475	52720.9	6.19	70820-71320	41.90	79.92	-0.271	37.68	18.11	12.63	31.58
25	*Vv*-RuBisCO	475	52518.7	6.33	67840-68340	42.23	80.13	-0.243	37.68	17.05	13.26	32.00

[a]Grand average of hydropathicity

location of eukaryotic proteins. The location assignment is based on the predicted presence of any of the N-terminal presequences: chloroplast transit peptide (cTP), mitochondrial targeting peptide (mTP) or secretory pathway signal peptide (SP). As can be seen in Table 4, none of the RuBisCOs of C_3 plants were predicted to be targeted to chloroplast, mitochondrion or secretory pathway signal peptide. TMHMM 2.0 server is for prediction of transmembrane helices in proteins. Transmembrane helices are to be found in structures of membrane proteins determined by X-ray diffraction. They may also be predicted on the basis of hydrophobicity scales. Because the interior of the bilayer and the interiors of most proteins of studied structure are hydrophobic, it is presumed to be a requirement of the amino acids that span a membrane that they are hydrophobic as well. However, membrane pumps and ion channels also contain numerous charged and polar residues within the generally non-polar transmembrane segments. According to the results of TMHMM 2.0 server, there was not any transmembrane helix in RuBisCO proteins (Table 4).

The tertiary structure prediction of RuBisCO proteins were carried out by Phyre2 (Protein Homology/analogY Recognition Engine) server. It uses the alignment of hidden Markov models via HHsearch (Söding 2005) to significantly evaluate the accuracy of alignment and detection rate. It also incorporates a new *ab initio* folding simulation called Poing (Jefferys et al. 2010) to model regions of the proteins with no detectable homology to known structures. In this analysis, it was tried to consider high confidence in all samples. So, all the samples tertiary structure were estimated using "c1rcxH" model (PDB accession code: 1rcx), with 100% confidence, but *Gh*-RuBisCO which used c1bwvA model with this high confidence (Table 5; Fig. 3).

Verification of the stereochemical quality of *At*-RuBisCO structure was performed by PROCHECK (Programs to Check the Stereochemical Quality of Protein Structures) server and gave a result in the form of Ramachandran plot. The Ramachandran plot shows the phi-psi torsion angles for

Table 4 The results of post-translational modification and topology prediction by different tools in twenty-five different species of C_3 plants

		Tools											
		Predotar 1.03					SignalP 4.1	TargetP 1.1					TMHMM 2.0
No.	Name	Mitochondrial	Plastid	ER[a]	Elsewhere	Prediction	Signal peptide	cTP[b]	mTP[c]	SP[d]	other	Loc[e]	TMhelix[f]
1	*Ah*-RuBisCO	-	-	-	-	-	no	0.078	0.194	0.114	0.834	-	no
2	*At*-RuBisCO	0.01	0.00	0.00	0.99	none	no	0.069	0.137	0.132	0.870	-	no
3	*Bn*-RuBisCO	0.01	0.00	0.00	0.99	none	no	0.072	0.125	0.138	0.870	-	no
4	*Bv*-RuBisCO	0.01	0.00	0.00	0.99	none	no	0.076	0.137	0.141	0.840	-	no
5	*Ca*-RuBisCO	0.01	0.00	0.00	0.99	none	no	0.074	0.135	0.123	0.870	-	no
6	*Csa*-RuBisCO	0.01	0.00	0.00	0.99	none	no	0.092	0.210	0.091	0.851	-	no
7	*Csi*-RuBisCO	0.01	0.00	0.00	0.99	none	no	0.066	0.144	0.132	0.869	-	no
8	*Eh*-RuBisCO	0.01	0.00	0.00	0.99	none	no	0.062	0.172	0.111	0.877	-	no
9	*Gh*-RuBisCO	0.01	0.00	0.00	0.99	none	no	0.078	0.133	0.124	0.852	-	no
10	*Gm*-RuBisCO	0.01	0.00	0.00	0.99	none	no	0.066	0.141	0.119	0.884	-	no
11	*Hv*-RuBisCO	0.01	0.00	0.00	0.99	none	no	0.083	0.145	0.109	0.878	-	no
12	*Ms*-RuBisCO	-	-	-	-	-	no	0.077	0.190	0.143	0.816	-	no
13	*Nt*-RuBisCO	0.01	0.00	0.00	0.99	none	no	0.043	0.185	0.147	0.876	-	no
14	*Os*-RuBisCO	0.01	0.00	0.00	0.99	none	no	0.057	0.396	0.064	0.804	-	no
15	*Pa*-RuBisCO	0.02	0.00	0.00	0.98	none	no	0.058	0.185	0.106	0.832	-	no
16	*Ps*-RuBisCO	0.02	0.00	0.00	0.97	none	no	0.055	0.163	0.111	0.879	-	no
17	*Pt*-RuBisCO	0.01	0.00	0.00	0.99	none	no	0.086	0.130	0.114	0.887	-	no
18	*Pv*-RuBisCO	0.01	0.00	0.00	0.99	none	no	0.071	0.137	0.135	0.881	-	no
19	*Rc*-RuBisCO	0.01	0.00	0.00	0.99	none	no	0.075	0.137	0.130	0.874	-	no
20	*Sc*-RuBisCO	0.01	0.00	0.00	0.99	none	no	0.089	0.144	0.103	0.875	-	no
21	*Sl*-RuBisCO	0.01	0.00	0.00	0.99	none	no	0.043	0.185	0.147	0.876	-	no
22	*St*-RuBisCO	0.01	0.00	0.00	0.99	none	no	0.043	0.186	0.145	0.877	-	no
23	*Ta*-RuBisCO	0.01	0.00	0.00	0.99	none	no	0.081	0.132	0.116	0.885	-	no
24	*Vf*-RuBisCO	0.02	0.00	0.00	0.98	none	no	-	0.090	0.067	0.905	-	no
25	*Vv*-RuBisCO	0.01	0.00	0.00	0.99	none	no	0.080	0.138	0.119	0.867	-	no

[a]Endoplasmic reticulum; [b]Chloroplast transit peptide; [c]Mitochondrial targeting peptide; [d]Signal peptide; [e]Prediction of localization; [f]Transmembrane helix

all residues in the structure (except those at the chain termini) (Morris et al. 1992). The Ramachandran plot for *At*-RuBisCO indicated that 89.9% of the main-chain dihedral angles are found in the most-favored regions, 9.3% in the additionally allowed, 0.4% in the generously allowed and 0.4% in the disallowed regions (Fig. 4).

Furthermore, the structural similarities and differences between *At*-RuBisCO and other models were analyzed using TM-score server. TM-score is a metric for measuring the structural similarity of two protein models. It is designed to solve two major problems in the traditional metrics like RMSD (Root Mean Square Deviation). (1) TM-score measures the global fold similarity and is less sensitive to the local structural variations; (2) magnitude of TM-score for random structure pairs is length-independent. TM-score has the value in (0,1], where 1 indicates a perfect match between two structures. Following strict statistics of structures in the PDB, scores below 0.17 corresponds to randomly chosen unrelated proteins whereas with a score higher than 0.5 assume generally the same fold. According to the obtained results, the least similarity of plant sample species to *A. thaliana* belonged to *Arachis hypogaea* with TM-score = 0.2581 (RMSD = 3.72) and *Medicago sativa* with TM-score = 0.2782 (RMSD = 3.74). Because tertiary structure is much more conserved than sequence, its comparisons let us to look even further back into biological prehistory (Maiti et al. 2004). The most common method for tertiary structure comparison is called structure superposition (or superimposition). So, in order to further evaluation of *At*-RuBisCO and two other models, superposition of their structures was performed by SuperPose web server and no significant difference were observed (Fig. 5).

The analyses of protected motifs of RuBisCO proteins were performed by MEME and MAST programs. A conserved

Table 5. The results of tertiary structure prediction of twenty-five different species of C_3 plants

		Tools					
		Phyre2				TM-score	
No.	Name	Template	Confidence (%)	Coverage (%)	Identity (%)	TM-score	RMSD (A°)[a]
1	*Ah*-RuBisCO	c1rcxH	100	100	95	0.2581	3.72
2	*At*-RuBisCO	c1rcxH	100	97	93	-	-
3	*Bn*-RuBisCO	c1rcxH	100	97	92	1	0
4	*Bv*-RuBisCO	c1rcxH	100	98	98	1	0
5	*Ca*-RuBisCO	c1rcxH	100	97	93	1	0
6	*Csa*-RuBisCO	c1rcxH	100	97	94	0.3428	3.04
7	*Csi*-RuBisCO	c1rcxH	100	98	95	1	0
8	*Eh*-RuBisCO	c1rcxH	100	98	91	1	0
9	*Gh*-RuBisCO	c1bwvA	100	97	58	0.9723	1.01
10	*Gm*-RuBisCO	c1rcxH	100	98	94	1	0
11	*Hv*-RuBisCO	c1rcxH	100	97	92	1	0
12	*Ms*-RuBisCO	c1rcxH	100	100	96	0.2782	3.74
13	*Nt*-RuBisCO	c1rcxH	100	98	93	1	0
14	*Os*-RuBisCO	c1rcxH	100	96	92	0.3428	3.04
15	*Pa*-RuBisCO	c1rcxH	100	98	93	1	0
16	*Ps*-RuBisCO	c1rcxH	100	98	94	1	0
17	*Pt*-RuBisCO	c1rcxH	100	98	95	1	0
18	*Pv*-RuBisCO	c1rcxH	100	98	95	1	0
19	*Rc*-RuBisCO	c1rcxH	100	98	96	1	0
20	*Sc*-RuBisCO	c1rcxH	100	98	93	1	0
21	*Sl*-RuBisCO	c1rcxH	100	98	93	1	0
22	*St*-RuBisCO	c1rcxH	100	98	94	1	0
23	*Ta*-RuBisCO	c1rcxH	100	98	93	1	0
24	*Vf*-RuBisCO	c1rcxH	100	98	94	1	0
25	*Vv*-RuBisCO	c1rcxH	100	98	94	1	0

[a]Root Mean Square Deviation (Angstrom)

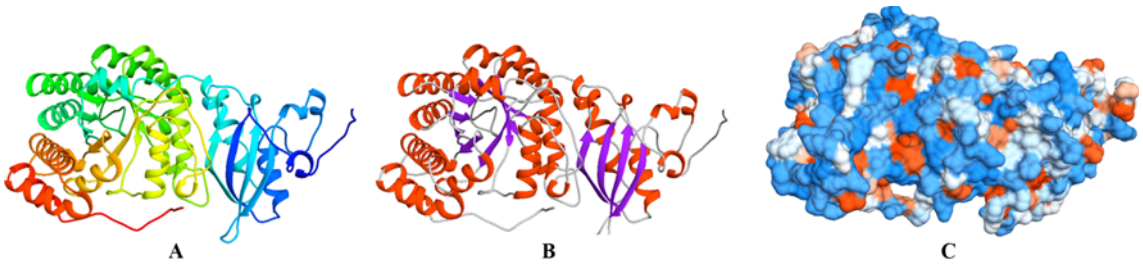

Fig. 3. Tertiary structure prediction of RuBisCO protein in *Arabidopsis thaliana* (Accession number: NP_051067.1; PDB accession code: 1rcx), established by Phyre2 server. (A) Ribbons model, (B) Secondary structure model (the α-helix in red, the beta sheet in purple, and the random coil in gray), and (C) Hydrophobicity surface model (from dodger blue for the most hydrophilic, to white, to orange red for the most hydrophobic). Images were shown by Chimera 1.10.1.

motif is a sequence pattern that occurs repetitively in a group of related protein sequences. MEME represents motifs as position-dependent letter-probability matrices which describe the probability of each possible letter at each position in the pattern and motifs in MAST are represented as position-dependent scoring matrices which describe the score of each possible letter at each position in the pattern. Based on the results, ten conserved motifs of RuBisCOs were identified (Table 6; Fig. 6). As can be seen in Fig. 6, just motif 9 was not common in twenty-five species.

The multiple sequence alignment of RuBisCO proteins was performed with ClustalW algorithm implemented in MEGA 6.06 with default parameters. The phylogenetic tree was constructed using the Neighbor-joining (NJ) method and the bootstrap test carried out with 1000 replicates (Fig. 7). Based on the phylogenetic relationships, RuBisCOs were

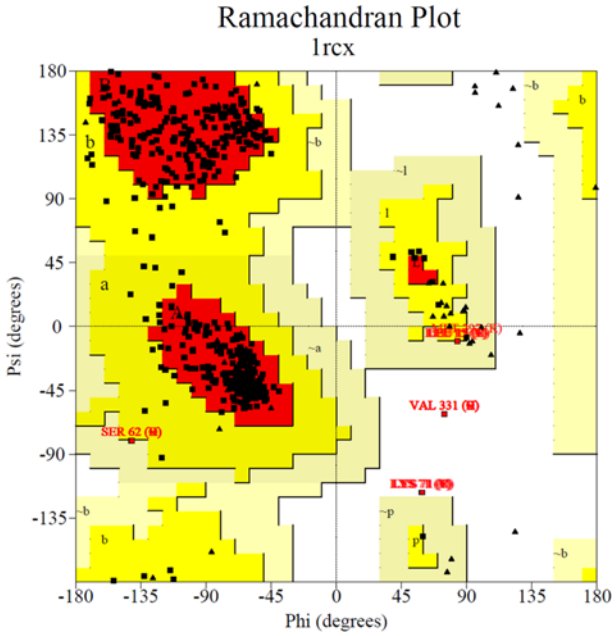

Fig. 4. Ramachandran plot of RuBisCO protein in *Arabidopsis thaliana* (Accession number: NP_051067.1; PDB accession code: 1rcx), obtained through PROCHECK server. Glycine residues are separately identified by triangles. The coloring / shading on the plot represents the different regions. The darkest areas (here shown in red) correspond to the "core" regions representing the most favorable combinations of phi-psi values. The regions are labelled as follows: A-Core alpha, L-Core left-handed alpha, a-Allowed alpha, l-Allowed left-handed alpha, ~a-Generous alpha, ~l-Generous left-handed alpha, B-Core beta, p-Allowed epsilon, b-Allowed beta, ~p-Generous epsilon, ~b-Generous beta.

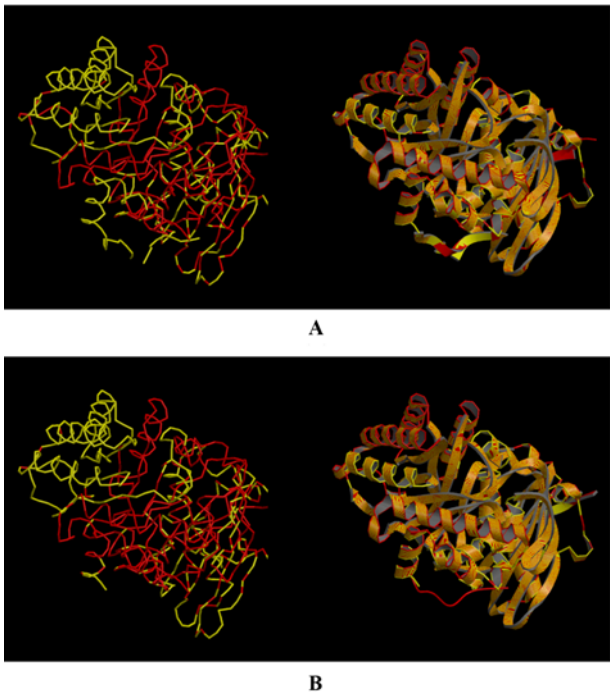

Fig. 5. Superposition of (A) *At*-RuBisCO (red) and *Ah*-RuBisCO (yellow), and (B) *At*-RuBisCO (red) and *Ms*-RuBisCO (yellow) models by SuperPose web server. Backbone (left) and Ribbon (right) superposed structures were shown in both figures.

divided into five groups, designated from Group 1 to Group 5.

Discussion

RuBisCO is the most abundant protein on earth (Ellis 1979), and can comprise up to 50% of the total soluble protein found in leaf tissue or within specific microbes (Tabita 1995; Tabita 1999). It is an extremely important enzyme ecologically, agriculturally, and industrially. Most terrestrial carbon has been through the active site of this enzyme (Phillips and Milo 2009).

In this study, bioinformatics methods were used to investigate the characteristics of RuBisCO in twenty-five species of fourteen different families of C_3 plants. At first, a protein-protein interaction network was gathered to find which genes are possibly regulated by RuBisCO protein or pathway (Table 2; Fig. 2). Genes involved in related biological pathways are usually expressed cooperatively for their functions, and

Table 6. Discovered conserved motifs by MEME tool in twenty-five different species of C_3 plants

Motif	E-value[a]	Width	Sites	Best possible match
1	1.1e-1278	50	25	PIVMHDYLTGGFTANTSLAHYCRDNGLLLHIHRAMHAVIDRQKNHGMHFR
2	2.5e-1274	50	25	CTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDENVNSQPFMRWRDRFLFC
3	3.9e-1241	50	25	FTQDWVSMPGVLPVASGGIHVWHMPALTEIFGDDSVLQFGGGTLGHPWGN
4	1.2e-1180	50	25	NMFTSIVGNVFGFKALRALRLEDLRIPPAYSKTFQGPPHGIQVERDKLNK
5	9.3e-1178	50	25	ILAAFRVTPQPGVPPEEAGAAVAAESSTGTWTTVWTDGLTSLDRYKGRCY
6	8.2e-1132	50	25	APGAVANRVALEACVQARNEGRDLAREGNEIIREACKWSPELAAACEVWK
7	6.7e-1068	50	25	VLAKALRMSGGDHIHSGTVVGKLEGERDITLGFVDLLRDDFIEKDRSRGI
8	9.2e-693	31	25	QAETGEIKGHYLNATAGTCEEMIKRAVFARE
9	2.9e-632	31	23	MSPQTETKASVGFKAGVKDYKLTYYTPEYQT
10	7.4e-458	22	25	EENQYICYVAYPLDLFEEGSVT

[a]The statistical significance of the motif. MEME usually finds the most statistically significant (low E-value) motifs first. The E-value of a motif is based on its log likelihood ratio, width, sites, the background letter frequencies, and the size of the training set.

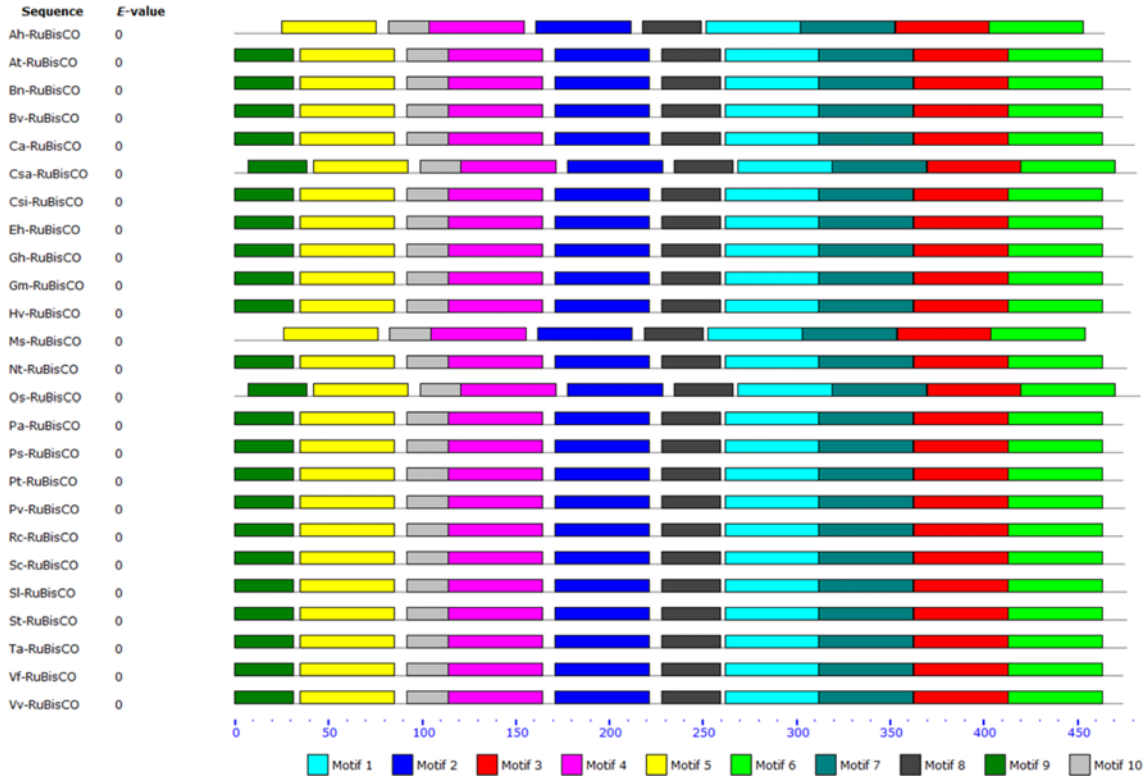

Fig. 6. Discovered conserved motifs for RuBisCO proteins in C_3 plants by MEME tool. The motifs were shown as different-colored boxes.

thus information on their interaction is a key to understand the biological systems at the molecular level (Eisen et al. 1998). For this reason *A. thaliana* was selected as the representative of C_3 plants and examined by STRING 10 tool. As it goes on the Table 2, ten functional partners were identified. Partners RBCS1A, RBCS2B and RBCS1B had the highest score (0.999), while partner PRK had the lowest score (0.954). Other obtained partners in this analysis had scores between 0.996 and 0.956. These functional partners have interactions with each other in biological process like reductive pentose-phosphate cycle, carbon fixation, photosynthesis, photorespiration and single-organism biosynthetic process. Additionally, thirteen pathways of KEGG, including "Carbon fixation in photosynthetic organisms", "Carbon metabolism", "Microbial metabolism in diverse environments", "Metabolic pathways", "Glyoxylate and dicarboxylate metabolism", "Glycolysis/Gluconeogenesis", "Biosynthesis of amino acids", "Biosynthesis of secondary metabolites", "Propanoate metabolism", "Fatty acid metabolism", "Photosynthesis", "Fatty acid biosynthesis" and "Pyruvate metabolism" were identified in the network analysis with their P-values.

In primary structure analysis, sequence length, molecular weight, theoretical isoelectric point (pI value), extinction coefficient, instability index, aliphatic index, and GRAVY were computed (Table 3). In the study of twenty-five RuBisCO proteins, it was found that *Os*-RuBisCO and *Ms*-RuBisCO had the most and the least of molecular weights, respectively. In addition, the computed isoelectric point for RuBisCOs was between 5.86 and 6.55. Isoelectric point is a pH at which a protein carries no net charge. It is of importance in protein purification as it is the pH at which solubility is commonly minimal and mobility in an electrofocusing system is zero. The extinction coefficient indicates how much light a protein absorbs at a certain wavelength. It is useful to have an estimation of this coefficient for following a protein which a spectrophotometer when purifying it. Two values are produced by ProtParam, both for proteins measured in water at 280 nm. The first one shows the computed value based on the assumption that all cysteine residues appear as half cystines, and the second one assuming that no cysteine appears as half cystine. A protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable. The aliphatic index of a protein is defined as the relative volume occupied by aliphatic side chains (alanine, valine, isoleucine, and leucine). It may be regarded as a positive factor for the increase of thermostability of globular proteins. As it was shown in Table 3, the most aliphatic index belonged to *Ah*-RuBisCO (82.90), while the least was observed in *Os*-RuBisCO (76.22). As a result, high aliphatic index in RuBisCOs indicated structural stability. In the obtained values of GRAVY, only negative values were

detected in all samples which show that RuBisCO is a hydrophilic protein. The GRAVY value for a protein is calculated as the sum of hydropathy values of all the amino acids, divided by the number of residues in the sequence. In secondary structure analysis of twenty-five selected species, alpha helix was the most percent (33.82-40.21), while the least percent structure belonged to beta turn (12.32-14.95) (Table 3). In general, according to the analysis of the primary and secondary structure, there is a little difference among obtained values. This would be a reason for their biological similarities' function.

Two servers including Predotar 1.03 and SignalP 4.1 were used in post-translational modification prediction (Table 4). Based on the results of Predotar server, none of the twenty-five species were predicted to be targeted to mitochondria, Plastid and endoplasmic reticulum. In the signal peptide analysis, it was shown that none of twenty-five selected species had it, so RuBisCO is not a secretory protein. In addition, two other servers including TargetP 1.1 and TMHMM 2.0 were studied in topology prediction. In TargetP 1.1 analysis, none of the RuBisCOs were predicted to be targeted to chloroplast, mitochondrion or secretory pathway signal peptide. While based on some previous researches, it has been proved that RuBisCO seems to be localized in the chloroplast (Borkhsenious et al. 1998). One of the causes of this contradiction is probably the protein sequence length in selected plants. Also, in TMhelix analysis, there were no transmembrane domains in RuBisCO proteins.

The 3D structure is the final goal of protein structure prediction and it is necessary to fully understand protein function. In this analysis, twenty-four species of RuBisCOs were estimated using "c1rcxH" model (PDB accession code: 1rcx), with 100% confidence. *Gh*-RuBisCO was estimated using "c1bwvA" model (PDB accession code: 1bwv). Also, the values of TM-score were computed between *At*-RuBisCO model and the others. According to the results of TM-score server, the highest and lowest structural similarity of *A. thaliana* to other plant samples belonged to *Gh*-RuBisCO and *Ah*-RuBisCO, respectively (Table 5; Fig. 3). Also, PROCHECK server was used for evaluating tertiary structure and the tertiary structure was validated.

The MEME and MAST analyses of RuBisCO proteins were performed in order to find patterns of conserved motifs. As can be seen in Table 6 and Fig. 6, among ten conserved motifs in twenty-five different species, only motif 9 (31 aa, E-value = 2.9e-632) was not common in species so that it was not present in two species including *Ah*-RuBisCO and *Ms*-RuBisCO. This motif is involved in some functions like integral component of membrane, ribulose-bisphosphate carboxylase activity, monooxygenase activity and reductive pentose-phosphate cycle. It seems lack of this motif in two *Ah*-RuBisCO and *Ms*-RuBisCO sequences is due to the shorter sequences than the other plants (Table 1).

In phylogenetic analysis among twenty-five RuBisCO proteins, five main groups were categorized. Group 4 contained eight members while Group 5 had two members constituted as the largest and smallest groups, respectively. *At*-RuBisCO located in Group 4, which made a subgroup with *Bn*-RuBisCO (bootstrap 100). According to the phylogenetic tree, RuBisCOs were derived from an ancestor and evolved into different groups.

Bioinformatics can play a vital role in the analysis and interpretation of genomic and proteomic data. It uses methods and technologies from mathematics, statistics, computer sciences, physics, biology, and medicine (Romano et al. 2011). Proteomics analyses generate mountains of data, expression information for hundreds or thousands of proteins in a single experiment (Wojcik and Schächter 2000). Therefore, data analysis (bioinformatics) is a vital part of some surveys. Currently proteomics and bioinformatics provide new insights into the processes in living cells and organisms (Darabi et al. 2012; Darabi and Farhadi-Nejad 2013; Seddigh and Darabi 2014; Darabi and Seddigh 2015; Seddigh and Darabi 2015; Seddigh and Darabi 2016; Darabi et al. 2017). In this research, bioinformatic analyses of RuBisCOs in C_3 plants exhibited similarities of this protein in different families and the obtained data provide background of bioinformatic studies for the function and evolution of RuBisCO in seed plants.

Material and Methods

In this study, a total of twenty-five RuBisCO protein sequences of fourteen different families in C_3 plants were collected from National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov).

The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING 10) database (http://string-db.org/) was used to foresee the interacting proteins (Szklarczyk et al. 2014). The database contains information from numerous sources, including experimental repositories, computational prediction methods and public text collections.

Various online web services and software were used for analyses of RuBisCO proteins in C_3 plants. Comparative and bioinformatic analyses were carried out online at the website ExPASy (http://expasy.org/tools). Physico-chemical parameters of RuBisCO proteins were analyzed by ProtParam (http://web.expasy.org/protparam) (Gasteiger et al. 2005). The secondary structure prediction was analyzed by SOPMA (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.tml) (Geourjon and Deleage 1995).

Prediction of mitochondrial and plastid targeting sequences was accomplished by Predotar 1.03 server (https://urgi.versailles.inra.fr/predotar/predotar.html) and prediction of signal peptide cleavage sites was achieved by SignalP 4.1 server (http://www.cbs.dtu.dk/services/SignalP) (Petersen et al. 2011). Also, identification of subcellular locations and transmembrane helices in proteins were performed by TargetP 1.1 and (www.cbs.dtu.dk/services/TargetP/) (Emanuelsson et al. 2000) and TMHMM 2.0 (http://www.cbs.dtu.dk/services/TMHMM-2.0/) (Moller et al. 2001) servers, respectively.

The tertiary structure prediction analysis of RuBisCO proteins were performed by Phyre2 server (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) (Kelley and Sternberg 2009) using profile-profile matching and secondary structure. Chimera 1.10.1 was used for 3D structure visualization of *Arabidopsis thaliana* as the model of C_3 plants (https://www.cgl.ucsf.edu/chimera/). Backbone similarities and differences of obtained models were estimated by TM-score server (http://zhanglab.ccmb.med.umich.edu/TM-score/) (Zhang and Skolnick 2004, Xu and Zhang 2010). Stereochemical quality and accuracy of the model was evaluated with PROCHECK 3.5 (http://www.ebi.ac.uk/thornton-srv/software/PROCHECK) by Ramachandran plot analysis (Laskowski et al. 1993). In addition, RuBisCO protein in *A. thaliana* was superposed on two plant samples which had the least TM-score values by SuperPose web server (http://wishart.biology.ualberta.ca/SuperPose/) (Maiti et al. 2004).

The motifs of protein sequences were discovered using the program of Multiple Em for Motif Elicitation (MEME; version 4.10.2) (Bailey et al. 2009) and Motif Alignment and Search Tool (MAST; version 4.9.1) (Bailey and Gribskov 1998) at website http://meme.nbcr.net/meme. The parameters of MEME analyses were applied as follows: distribution of motif occurrences, zero or one per sequence; number of different motifs, ten; minimum motif width, six and maximum motif width, fifty.

The multiple sequence alignment of RuBisCO proteins was performed with ClustalW algorithm implemented in Molecular Evolutionary Genetic Analysis (MEGA 6.06) (http://www.megasoftware.net) (Tamura et al. 2013) with default parameters. The phylogenetic tree was constructed using the Neighbor-joining (NJ) method and the bootstrap test carried out with 1000 replicates.

Author's Contributions

MD designed and performed bioinformatics analyses of RuBisCO enzyme in C3 plants and wrote the manuscript; SS helped in some analyses and some parts of discussion and revised the manuscript. The authors agreed on the contents of the paper and post no conflicting interest.

References

Andersson I, Knight S, Schneider G, Lindqvist Y, Lundqvist T, Brändén C-I, Lorimer GH (1989) Crystal structure of the active site of ribulose-bisphosphate carboxylase. Nature 337: 229–234

Andrews T, Ballment B (1983) The function of the small subunits of ribulose bisphosphate carboxylase-oxygenase. J Biol Chem 258:7514–7518

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 (Web Server issue): W202–208

Bailey TL, Gribskov M (1998) Combining evidence using p-values: application to sequence homology searches. Bioinformatics 14:48–54

Borkhsenious ON, Mason CB, Moroney JV (1998) The intracellular localization of ribulose-1, 5-bisphosphate carboxylase/oxygenase in Chlamydomonas reinhardtii. Plant Physiol 116:1585–1591

Chatterjee A, Basu A (2011) Rubisco: limitations and re-engineering for a better enzyme. INT J PLANT SCI 2:22–24

Darabi M, Farhadi-Nejad H (2013) Study of the 3-hydroxy-3-methylglotaryl-coenzyme A reductase (HMGR) protein in Rosaceae by bioinformatics tools. Caryologia 66:351–359

Darabi M, Masoudi-Nejad A, Nemat-Zadeh G (2012) Bioinformatics study of the 3-hydroxy-3-methylglotaryl-coenzyme A reductase (HMGR) gene in Gramineae. Mol Biol Rep 39:8925–8935

Darabi M, Seddigh S (2015) Bioinformatic characterization of aspartic protease (AP) enzyme in seed plants. Plant Syst Evol 301:2399–2417.

Darabi M, Seddigh S, Abarshahr M (2017) Structural, functional, and phylogenetic analysis of the cytochrome P450 (CYP) enzyme in seed plants by bioinformatics tools. Caryologia. DOI: 10.1080/00087114.2016.1274866

Dhingra A, Portis AR, Daniell H (2004) Enhanced translation of a chloroplast-expressed RbcS gene restores small subunit levels and photosynthesis in nuclear RbcS antisense plants. Proc Natl Acad Sci USA 101:6315–6320

Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 95:14863-14868

Ellis RJ (1979) The most abundant protein in the world. Trends Biochem Sci 4:241–244

Emanuelsson O, Nielsen H, Brunak S, von Heijne G (2000) Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol 300:1005–1016

Feller U, Anders I, Mae T (2008) Rubiscolytics: fate of Rubisco after its enzymatic function in a cell is terminated. J Exp Bot 59: 1615–1624

Finn MW, Tabita FR (2004) Modified pathway to synthesize ribulose 1, 5-bisphosphate in methanogenic archaea. J Bacteriol 186: 6360–6366

Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A (2005) Protein identification and analysis tools on the ExPASy server. The proteomics protocols handbook, Springer: 571–607

Geourjon C, Deleage G (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Computer applications in the biosciences: CABIOS 11:681–684

Hanson TE, Tabita FR (2001) A ribulose-1, 5-bisphosphate carboxylase/oxygenase (RubisCO)-like protein from Chlorobium tepidum that is involved with sulfur metabolism and the response to oxidative stress. Proc Natl Acad Sci USA 98:4397–4402

Imker HJ, Fedorov AA, Fedorov EV, Almo SC, Gerlt JA (2007) Mechanistic diversity in the RuBisCO superfamily: the "enolase" in the methionine salvage pathway in Geobacillus kaustophilus. Biochemistry 46:4077–4089

Jefferys BR, Kelley LA, Sternberg MJ (2010) Protein folding requires crowd control in a simulated cell. J Mol Biol 397:1329–1338

Kelley LA, Sternberg MJ (2009) Protein structure prediction on the Web: a case study using the Phyre server. Nature protocols 4: 363–371

Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26:283–291

Li H, Sawaya MR, Tabita FR, Eisenberg D (2005) Crystal structure of a RuBisCO-like protein from the green sulfur bacterium Chlorobium tepidum. Structure 13:779–789

Maiti R, Van Domselaar GH, Zhang H, Wishart DS (2004) SuperPose: a simple server for sophisticated structural superposition. Nucleic Acids Res 32(suppl 2):W590–W594

Moller S, Croning MD, Apweiler R (2001) Evaluation of methods for the prediction of membrane spanning regions. Bioinformatics 17:646–653

Morris AL, MacArthur MW, Hutchinson EG, Thornton JM (1992) Stereochemical quality of protein structure coordinates. Proteins: Structure, Function, and Bioinformatics 12:345–364

Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785–786

Phillips R, Milo R (2009) A feeling for the numbers in biology. Proc Natl Acad Sci USA 106:21465–21471

Romano P, Giugno R, Pulvirenti A (2011) Tools and collaborative environments for bioinformatics research. Brief Bioinform 12:549–561

Sato T, Atomi H, Imanaka T (2007) Archaeal type III RuBisCOs function in a pathway for AMP metabolism. Science 315:1003–1006

Schneider G, Lindqvist Y, Brändén C-I, Lorimer G (1986) Three-dimensional structure of ribulose-1, 5-bisphosphate carboxylase/oxygenase from Rhodospirillum rubrum at 2.9 Å resolution. EMBO J 5:3409

Seddigh S, Darabi M (2014) Comprehensive analysis of beta-galactosidase protein in plants based on *Arabidopsis thaliana*. TURK J BIOL 38:140–150

Seddigh S, Darabi M (2015) Structural and phylogenetic analysis of α-glucosidase protein in insects. Biologia 70:812–825

Seddigh S, Darabi M (2016) Proteomics comparison of aspartic protease enzyme in insects. TURK J BIOL 40:69–83

Sheth BP, Thaker VS (2014) In silico analyses of Rubisco Enzymes from different classes of Algae. INT J BIOL SCI 3:11–17

Söding J (2005) Protein homology detection by HMM–HMM comparison. Bioinformatics 21:951–960

Spreitzer RJ (2003) Role of the small subunit in ribulose-1, 5-bisphosphate carboxylase/oxygenase. Arch Biochem Biophys 414:141–149

Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP (2014) STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Res: gku1003

Tabita FR (1995) The biochemistry and metabolic regulation of carbon metabolism and CO2 fixation in purple bacteria. Anoxygenic photosynthetic bacteria, Springer: 885–914

Tabita FR (1999) Microbial ribulose 1, 5-bisphosphate carboxylase/oxygenase: a different perspective. Photosynthesis Res 60:1–28

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, Hanson TE, Satagopan S, Witte BH, Kreel NE (2008a) Phylogenetic and evolutionary relationships of RubisCO and the RubisCO-like proteins and the functional lessons provided by diverse molecular forms. Philosophical Transactions of the Royal Society of London B: Biological Sciences 363:2629–2640

Tabita FR, Satagopan S, Hanson TE, Kreel NE, Scott SS (2008b) 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

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30:2725–2729

Warlick B (2013) Functional discovery and promiscuity in the RuBisCO superfamily, University of Illinois at Urbana-Champaign

Watson GM, Yu J-P, Tabita FR (1999). Unusual ribulose 1, 5-bisphosphate carboxylase/oxygenase of anoxic Archaea. J Bacteriol 181: 1569–1575

Wojcik J, Schächter V (2000) Proteomic databases and software on the web. Brief Bioinform 1:250–259

Xu J, Zhang Y (2010) How significant is a protein structure similarity with TM-score= 0.5?. Bioinformatics 26:889-895

Zhang Y, Skolnick J (2004) Scoring function for automated assessment of protein structure template quality. Proteins: Structure, Function, and Bioinformatics 57:702–710