biology and life sciences forum

MDPI

Proceedings

Comparative analysis of RuBisCO evolution and intrinsic differences: insights from in silico assessment in cyanobacteria, monocot and dicot plants

Ishita Biswas [1,2] and Debanjan Mitra [1,*]

[1]Dept. of Microbiology, Raiganj University, Raiganj, West Bengal, India, 7331342
[2]Department of Botany, S.B.S. Government College, Hili, West Bengal, India, 733126
* Correspondence: debanjanmitra267@gmail.com Tel.: +917479199988
† Presented at the title, place, and date.

Abstract: RuBisCO is the main photosynthetic enzyme of carbon assimilatory pathways in nature. Despite being the most abundant protein on earth, RuBisCO is still relatively underutilized in the food chain. Although having sequence and structure details in the database, studies on evolutionary relationships have few instances. A bioinformatics and in silico study was conducted to check sequence and structural differences of RuBisCO among different photosynthetic organisms. RuBisCO from *Oryza sativa* showed abundance of charged amino acids, salt-bridges and intra-protein interactions and was more hydrophilic in nature compared to *Nostoc* sp., *Chlamydomonas reinhardtii*, and *Nicotiana tabacum*. From molecular dynamics simulations, lower root mean square deviation and root mean square fluctuation indicate that RuBisCO from *Oryza sativa* was more stable followed by *Nicotiana tabacum* and lower radius of gyrations indicates their tightly packing. From this study, it was clear that some specific evolutions in charged amino acids of RuBisCO of monocot i.e., *Oryza sativa* make it more stable and stronger than other plant groups. The study concludes more stable nature of RuBisCO from monocot *Oryza sativa*.

Keywords: RuBisCO; Evolution; Salt-Bridge; Intra-protein interactions; Molecular Dynamic Simulations

52Citation: To be added by editorial staff during production.

Academic Editor: Firstname Lastname

Published: date

1. Introduction

The most prevalent protein on earth is undoubtedly RubisCO (Ribulose-1,5-bisphosphate carboxylase oxygenase) [1]. Most autotrophic organisms, starting from prokaryotes including cyanobacteria, photosynthetic bacteria, chemoautotrophic bacteria, and archaea to eukaryotes such as algae and higher plants, possess this enzyme. According to estimates, Rubisco can make up to 50% of all soluble proteins found in plant leaves or inside microbes [2]. Its presence in marine phytoplankton, which is projected to contribute more than 45% of yearly world net primary production, is perhaps less evident but nonetheless pervasive [3]. The enzyme RuBisCO (EC 4.1.1.39, Ribulose-1,5-bisphosphate carboxylase oxygenase) catalyzes the primary photosynthetic CO_2 reduction process, which involves the binding of CO_2 to the acceptor molecule Ribulose-1,5-bisphosphate (RuBP) to produce 3-phosphoglycerate [4]. Besides that, it also possesses oxygenase activity through which it binds O_2 with RuBP to form 2-phosphoglycolate. The discovery of the salt bridge's microenvironment is a novel concept in structural biology. Intragenic protein sequence and structural study sheds light on species variety, functions, and evolutionary relationships [5]. Although there are numerous sequence and structure of Rubisco in the database, however, there is no such report on the sequence and structure analysis of this protein in terms of salt bridge and other interactions of proteins inside. The study was conducted to check the evolutionary pattern in RuBisCO protein starting from lower group of photosynthetic

organism to higher group of plants through intra-protein interactions and molecular dynamics simulation studies.

2. Methods

2.1. Dataset:

Sequences of RuBisCO enzyme from *Nostoc* sp., *Chlamydomonas reinhardtii*, *Oryza sativa*, and *Nicotiana tabacum* were retrieved from the UniProt database [6]. Diverse organisms were taken into consideration to check evolutionary significance between them. Structures of those proteins were extracted from RCSB PDB database [7].

2.2. Analysis of protein sequences:

All protein sequences were subjected to MSA for block preparation through Clustal omega [8]. Non-block format was used for calculation of amino acid abundance, pI, grand average hydropathy (GRAVY), aliphatic index through the ProtParam server [9]. Block format of sequences were used in calculations of hydropathy, polarity by the ProtScale server [9].

2.3. Analysis of crystal structures:

All those protein structures were minimized through the Chimera 1.15rc with amber forcefield [10-11]. Identification and calculations of intra-protein interactions were done through the PIC server [12]. PDBSum were used to check their type of secondary structures [13].

2.4. Molecular dynamics simulations:

GROMACS [16] and the GROMOS96 43a1 forcefield were used to do molecular dynamic simulations. After equilibration, the energy on the solvated systems was decreased using the steepest descent approach with 5000 steps. The last manufacturing run's molecular dynamic simulations lasted 50 ns at 300 K temperatures. Molecular dynamics simulations were used to estimate the radius of gyration (Rg), solvent accessible surface area (SASA), root mean square deviation (RMSD), root mean square fluctuation (RMSF), and hydrogen bonding.

3. Results and discussion:

3.1. Preferable amino acid abundance

Higher abundance of charged polar residues was observed in *Oryza sativa* followed by *Nicotiana tabacum* (Figure 1). However, the uncharged polar residues showed higher presence in the enzyme of *Nostoc* sp. Hydrophobic amino acid showed highest abundance in *Chlamydomonas reinhardtii*. *Oryza sativa* had higher presence of amino acid Pro. Due to its capacity to reduce the structural entropy of the denatured state, Pro may be able to improve protein stability. Additionally, Pro is often conserved in proteins and frequently contributes significantly to the structure and function of proteins [17].

A common method for identifying hydrophobic areas in proteins is the Kyte-Doolittle scale. Positive values indicate hydrophobic regions. Transmembrane helices predicted using hydropathy plots. The non-polar portion of the lipid membrane contains the transmembrane helices, whereas the loops are in a more polar solution [5]. *Nicotiana tabacum* showed lowest plot followed by *Oryza sativa* which means they were more hydrophilic rather than others.

The specific polarity pattern is crucial to the molecule's structure and functionality. Higher polarity has been shown by protein sequence of *Nicotiana tabacum* followed by *Oryza sativa*. However, it was observed that in some specific positions in *Chlamydomonas*

reinhardtii there are some pick-point for high polarity. Higher polarity drastically increased thermal stability of a protein [19-20].

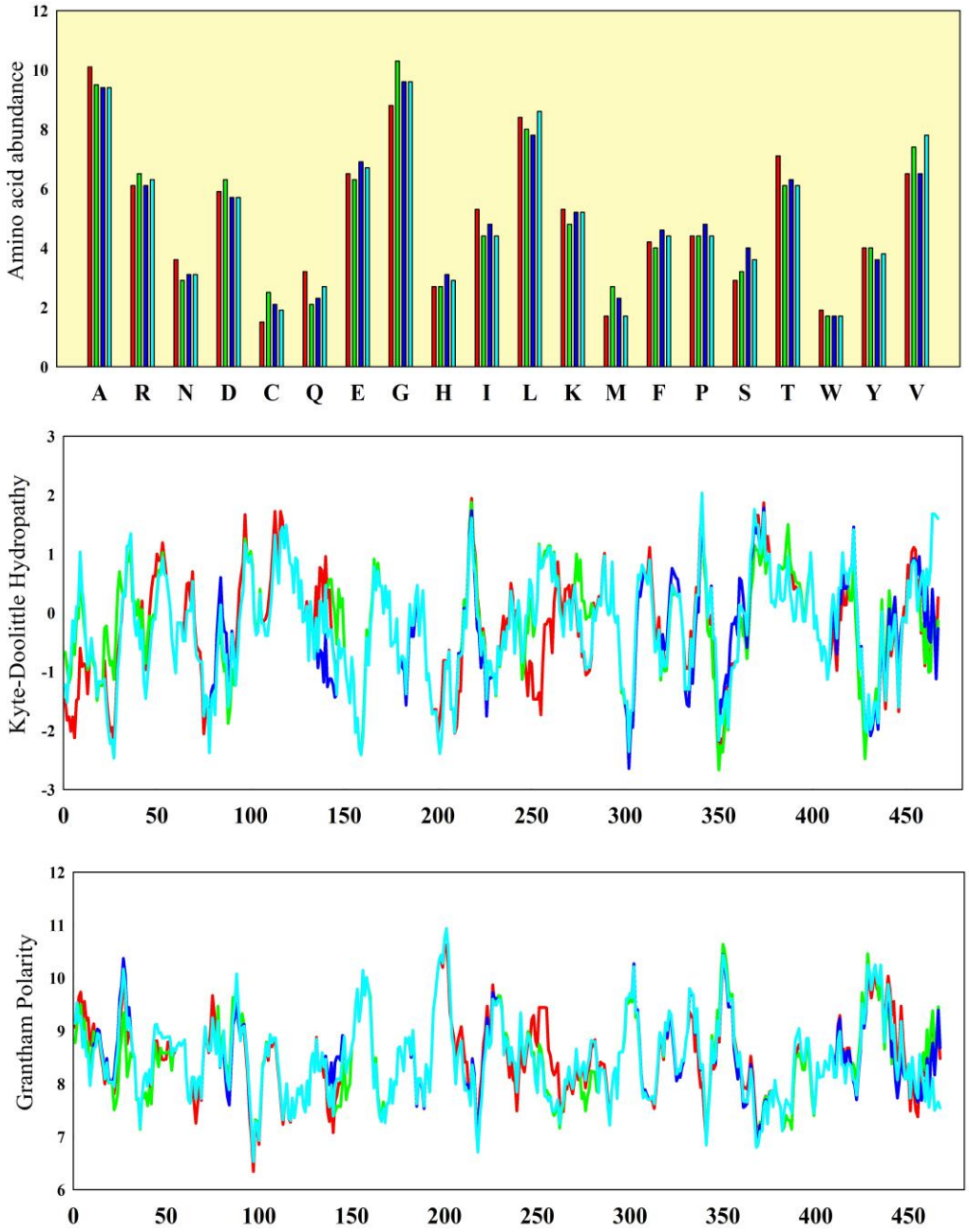

Figure 1. Amino acid abundance, Kyte-Dolittle hydropathy, Grantham polarity for RuBisCO enzyme from *Nostoc* sp. (red), *Chlamydomonas reinhardtii* (green), *Oryza sativa* (blue), and *Nicotiana tabacum* (cyan).

3.2. Secondary structure assessment:

Nostoc sp., and *Chlamydomonas reinhardtii* both showed 4 sheets, 4 beta-alpha-beta units, 2 beta hairpins, 3 beta bulges and 16 strands. Difference has been showed in *Nostoc* sp., where there are 22 helix, 23 helix-helix intraces, 22 beta turns, 5 gamma turns and 1 disulphide structure. However, in those places, in *Chlamydomonas reinhardtii*, there are 21 helix, 21 helix-helix intraces, 31 beta turns, 4 gamma turns. RuBisCO of *Oryza sativa* possess 5 sheets, 4 beta-alpha-beta units, 2 beta hairpins, 4 beta bulges, 16 strands, 23 helix, 28 helix-helix intraces, 23 beta turns, 3 gamma turns whereas *Nicotiana tabacum* had 3 sheets, 5 beta-alpha-beta units, 2 beta hairpins, 4 beta bulges, 15 strands, 22 helix, 25 helix-

helix intraces, 23 beta turns, 1 gamma turns and 1 disulphide bond. Increasing the amount of helix in protein secondary structure ultimately increase the protein stability [21].

3.3. Intra-protein interactions

Intra-protein interactions are crucial interactions to enhance the protein stability. Salt-bridges, aromatic-aromatic interactions, aromatic-sulfur interactions, cation-pi interactions have significant contributions in this field. Generally, they act as single pair called isolated, however sometime multiple isolated are connected to each other to make network formation.

Table 1. Intra-protein interactions in RuBisCO of *Nostoc* sp., *Chlamydomonas reinhardtii*, *Oryza sativa*, and *Nicotiana tabacum***.**

Protein	Salt bridge		Aromatic-aromatic		Aromatic-sulfur		Cation-pi	
	Isolated	Network	Isolated	Network	Isolated	Network	Isolated	Network
Nostoc sp.	15	7	7	3	3	0	8	2
Chlamydomonas reinhardtii	19	9	5	3	8	1	7	2
Oryza sativa	23	6	5	6	7	0	8	2
Nicotiana tabacum	11	10	5	5	5	0	8	4

Strong hydrogen bonds that are formed by the interaction of two charged residues are known as salt bridges or ion pairs. In contrast to surface salt bridges, a subsurface salt bridge destabilizes by 3–4 kcal/mol when one partner is removed. Nowadays, beside the isolated and network salt bridge, a special salt-bridge i.e., cyclic salt bridge has been discovered [22]. Highest number of isolated salt bridges was found in RuBisCO of *Oryza sativa* (23) whereas *Nicotiana tabacum* showed highest number of network salt bridges (10) (Table 1). *Nostoc sp.* showed 15 isolated and 7 network salt bridges whereas *Chlamydomonas reinhardtii* had 19 isolated and 9 network salt bridges. Higher formation of salt bridges in monocot and dicot makes them more stable than other two species. Important non-covalent interactions in proteins involve aromatic-aromatic interactions between several aromatic amino acids (Phe, Tyr, and Trp). *Nostoc sp.* showed highest number of isolated aromatic-aromatic interactions i.e., 7 with 3 network aromatic-aromatic interactions. *Chlamydomonas reinhardtii* showed 5 isolated and 3 network aromatic-aromatic interactions. RuBisCO of monocot *Oryza sativa* had 5 isolated and 6 network formation which was highest formation of network aromatic-aromatic interactions. *Nicotiana tabacum* showed equal number of isolated and network aromatic-aromatic interactions. Aromatic-sulphur interactions was higher in *Chlamydomonas reinhardtii* in the form of isolated bonds. However, the only formation of network aromatic-sulphur interactions was observed here. Formation of isolated cation-pi interactions was almost same in every species except *Chlamydomonas reinhardtii.* However, the network formation was higher in *Nicotiana tabacum* whereas, others showed equal number i.e., 2 network formations. Formation of higher number of intra-protein interactions in RuBisCO of dicot and monocot gives advantage to them to gain more stability over the algae and cyanobacteria.

3.4. Stability through simulation Study

50 ns molecular dynamics simulations provide the details of RMSD, RMSF, Rg and SASA. From the RMSD, it was observed that *Nostoc sp.* had highest RMSD than others. It

started to deviate from 0.2 nm and ended at o.6 nm. On other hand, *Oryza sativa* and *Nicotiana tabacum* showed lower and almost equal trajectory throughout the 50 ns RMSD analysis. From the starting at 0.2 nm become stabilized and ended at almost same range. RMSF analysis revealed all proteins showed almost similar trajectory throughout the path. The Rg plot was also showed similarity with plot to RMSD. *Oryza sativa* and *Nicotiana tabacum* showed lowest Rg than others which means they had tightest packing RuBisCO [23]. SASA was also high in *Oryza sativa* and *Nicotiana tabacum*. Increasing the value of SASA enhance the stability and protein folding [24].

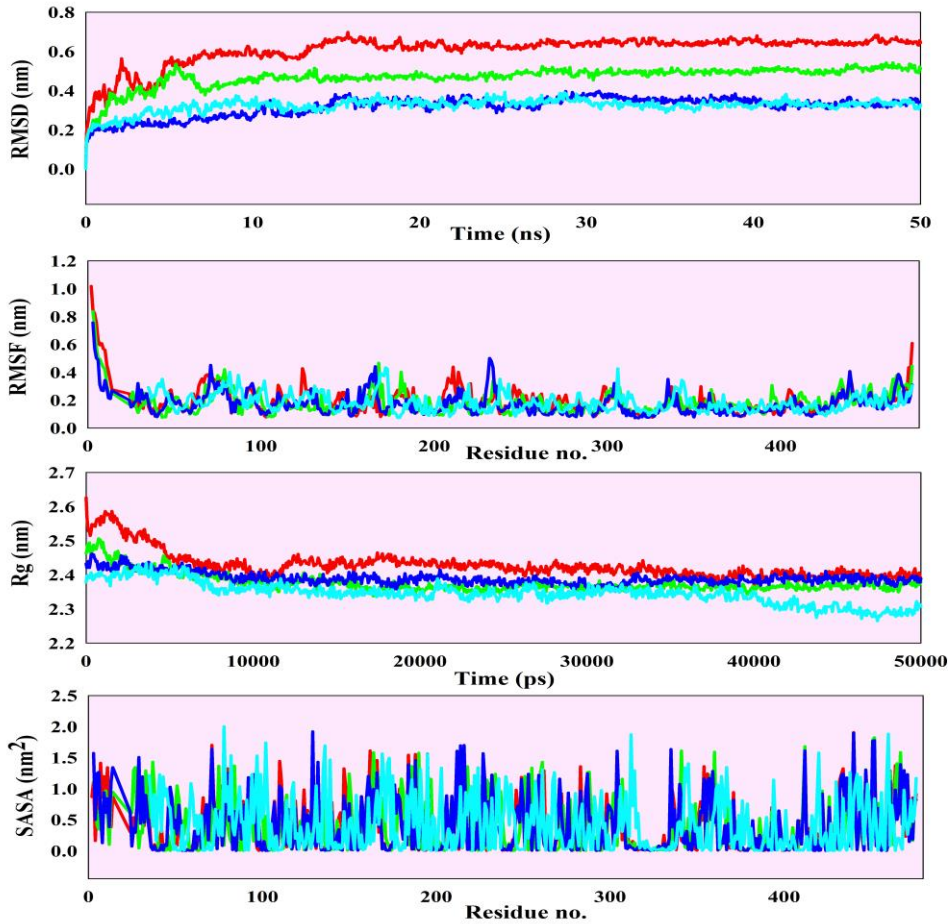

Figure 2. RMSD, RMSF, Rg, SASA for RuBisCO enzyme from *Nostoc* sp. (red), *Chlamydomonas reinhardtii* (green), *Oryza sativa* (blue), and *Nicotiana tabacum* (cyan).

4. Conclusion:

Insilico investigation on different RuBisCO revealed how the amino acid evolutions make significant changes to gain more stability and flexibility in higher group of plant. Charged amino acid residues abundance was found mainly in *Nicotiana tabacum* and *Oryza sativa* plant. Moreover, higher hydrophilicity and higher polarity enhance the stability, functionality, and flexibility. Increase of helix in secondary structure further boost the stability. Molecular dynamics simulations revealed higher stability, flexibility, and folding patterns of RuBisCO from *Oryza sativa* and *Nicotiana tabacum*. This study will be helpful to understand the protein evolution and play role for protein engineering.

Author Contributions: Conceptualization, D.M. and I.B.; methodology, I.B.; software, D.M.; validation, D.M.; formal analysis, I.B.; investigation, D.M.; resources, D.M.; data curation, I.B.; writing—original draft preparation, I.B.; writing—review and editing, D.M.; visualization, I.B.; supervision, D.M.

Funding: None.

Data Availability Statement: NA.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tabita FR, Hanson TE, Li H, Satagopan S, Singh J, Chan S. Function, structure, and evolution of the RubisCO-like proteins and their RubisCO homologs. Microbiology and Molecular Biology Reviews. 2007 Dec;71(4):576-99.
2. Ellis RJ. The most abundant protein in the world. Trends in biochemical sciences. 1979 Nov 1;4(11):241-4.
3. Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. Primary production of the biosphere: integrating terrestrial and oceanic components. science. 1998 Jul 10;281(5374):237-40.
4. Andersson I, Backlund A. Structure and function of Rubisco. Plant Physiology and Biochemistry. 2008 Mar 1;46(3):275-91.
5. Biswas I, Mitra D, Bandyopadhyay AK, Mohapatra PK. Contributions of protein microenvironment in tannase industrial applicability: An in-silico comparative study of pathogenic and non-pathogenic bacterial tannase. Heliyon. 2020 Nov 1;6(11).
6. UniProt Consortium. UniProt: a hub for protein information. Nucleic acids research. 2015 Jan 28;43(D1):D204-12.
7. Kouranov A, Xie L, de la Cruz J, Chen L, Westbrook J, Bourne PE, Berman HM. The RCSB PDB information portal for structural genomics. Nucleic acids research. 2006 Jan 1;34(suppl_1):D302-5.
8. Sievers F, Higgins DG. Clustal omega. Current protocols in bioinformatics. 2014 Dec;48(1):3-13.
9. Gasteiger E, Hoogland C, Gattiker A, Duvaud SE, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. Humana press; 2005.
10. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera—a visualization system for exploratory research and analysis. Journal of computational chemistry. 2004 Oct;25(13):1605-12.
11. Mitra D, Das Mohapatra PK. In silico comparative structural and compositional analysis of glycoproteins of RSV to study the nature of stability and transmissibility of RSV A. Systems Microbiology and Biomanufacturing. 2023 Apr;3(2):312-27.
12. Tina KG, Bhadra R, Srinivasan N. PIC: protein interactions calculator. Nucleic acids research. 2007 Jul 1;35(suppl_2):W473-6.
13. Laskowski RA. PDBsum: summaries and analyses of PDB structures. Nucleic acids research. 2001 Jan 1;29(1):221-2.
14. Mitra D, Paul M, Thatoi H, Mohapatra PK. Study of potentiality of dexamethasone and its derivatives against Covid-19. Journal of Biomolecular Structure and Dynamics. 2022 Nov 29;40(20):10239-49.
15. Bajaj K, Madhusudhan MS, Adkar BV, Chakrabarti P, Ramakrishnan C, Sali A, Varadarajan R. Stereochemical criteria for prediction of the effects of proline mutations on protein stability. PLoS computational biology. 2007 Dec;3(12):e241.
16. Mitra D, Das Mohapatra PK. Cold adaptation strategy of psychrophilic bacteria: an in-silico analysis of isocitrate dehydrogenase. Systems Microbiology and Biomanufacturing. 2021 Oct;1(4):483-93.
17. Mitra D, Pal AK, Das Mohapatra PK. Intra-protein interactions of SARS-CoV-2 and SARS: a bioinformatic analysis for plausible explanation regarding stability, divergency, and severity. Systems Microbiology and Biomanufacturing. 2022 Oct;2(4):653-64.
18. Mitra D. Deep Dive Into the DNA Polymerase Through Insilico Analysis: An Information to Get Better PCR Enzyme From the Ancient One. Preprint. 2023.
19. Mitra D, Das Mohapatra PK. Discovery of novel cyclic salt bridge in thermophilic bacterial protease and study of its sequence and structure. Applied Biochemistry and Biotechnology. 2021 Jun;193(6):1688-700.
20. Mitra D, Afreen S, Das Mohapatra PK, Abdalla M. Threat of respiratory syncytial virus infection knocking the door: a proposed potential drug candidate through molecular dynamics simulations, a future alternative. Journal of Molecular Modeling. 2023 Apr;29(4):91.
21. Ausaf Ali S, Hassan I, Islam A, Ahmad F. A review of methods available to estimate solvent-accessible surface areas of soluble proteins in the folded and unfolded states. Current Protein and Peptide Science. 2014 Aug 1;15(5):456-76.