Life Science Informatics Publications

Research Journal of Life Sciences, Bioinformatics, Pharmaceutical and Chemical Sciences

Journal Home page http://www.rjlbpcs.com/

RJLBPCS

ISSN 2454-6348

Original Research Article **DOI - 10.26479/2017.0206.04**

STRUCTURAL MODELING AND MOLECULAR DYNAMIC SIMULATION OF A NOVEL THER MOPHILLIC RUBISCO ENZYME FROM CYANOBACTERIUM *THERMOSYNECHOCOCCUS* SP. STRAIN NK55A

A I Matinja*[1, 2], A Abubakar [1,3], A Attahiru [1], M Ali Dau [1], A Y Saadatu[1,4], M S Shamsir[1]

[1] Faculty of Bioscience and Medical Engineering, Universiti Teknologi Malaysia

[2] Biochemistry Department, Bauchi State University Gadau, Nigeria.

[3]Biological Sciences Department, Federal University Kashere, Nigeria.

[4]Microbiology and Biotechnology Department, Federal University Dutse, Nigeria.

ABSTRACT: Ribulose 1,5-bisphosphate carboxylase/oxygenase (Rubisco) catalyzes the carboxylation of ribulose-1,5-bisphosphate during carbon fixation and it has been named as the most abundant protein on earth, which serves as linkage between non-living inorganic carbon and living organic carbon. Despite its abundance in nature, temperature affects the activity of this enzyme. Therefore, understanding, modeling and validation of this enzyme are of great importance. The enzyme was isolated from novel unicellular cyanobacterium *Thermosynechococcus* sp. strain NK55a. Four different species comprising of *Oryza sativa subsp japonica, Rhodopseudomonas palustris, Thermosynechococcus elongates* and *Burkholderia fungorum,* were used for sequence alignment. The Homology modelling was done to obtain the 3D structure of the enzyme using Raptor X. The model stereo chemical quality is assessed using PROCHECK which generate the Ramachandran plot and GROMACS was used for molecular dynamic simulation. The PROCHECK and GROMACS (RMSD and RMSF) result indicates the stable structure of the modeled enzyme.

KEYWORDS: Rubisco, GROMACS, Molecular Dynamic Simulations, Thermophiles

***Corresponding Author: Dr. A. I. Matinja** Ph.D.

Faculty of Bioscience and Medical Engineering, Universiti Teknologi Malaysia

* Email Address: adamumatinja@gmail.com

Peer review under responsibility of Life Science Informatics Publications

2017 March- April RJLBPCS 2(6) Page No.44

1. INTRODUCTION

Thermophiles are those microorganisms that live and grow at temperature above 45ºC. To survive at this harsh condition, thermophilic organisms have developed some modifications at cellular as well as molecular level different from the rest of their counterparts [1]. These microorganisms use two basic strategies to resist thermal denaturation of their enzyme's extrinsic stabilization which is conferred by certain molecules, and also intrinsic stabilization conferred by specific structure and the conformation of the enzymes itself [2, 3]. Proteins in thermophiles have highly organized hydrophobic interactions, more hydrogen bonds and non-covalent bonds which impart them heat stability [4]. Thermophiles show adaptations in nucleic acid structure with high GC content [5] etc. Thermophilic bacteria are of extreme commercial importance because of a large number of thermostable enzymes having been isolated from these organisms which are used in the biochemical processes that are carried out at higher temperature. The use of *Taq* polymerase for DNA amplification by PCR is excellent example of importance of thermophilic bacteria. Enzymes from thermophiles are supposed to be intrinsically stable and active at high temperatures as well as more rigidity than the rest of their counterparts. Hence, enzymes of thermophilic and hyperthermophilic origin have more biotechnological advantages over its psychrophilic and mesophilic counterparts [6]. When thermophilic and hyperthermophilic enzymes are expressed in mesophiles organisms, they are easily purified through heat treatment. This thermal stability is associated with chemical denaturants resistivity. Therefore, at high temperatures, enzyme reactions have few risks of microbial contamination. Rubisco (Ribulose 1, 5-bisphosphate carboxylase) is the key enzyme of Kelvin cycle. Rubisco (EC: 4.1.1.39) catalyses the carboxylation of ribulose-1,5-bisphosphate and has been named as the most abundant protein on earth. The enzymes relative abundance draws from its importance in nearly all plants and selected bacterium for its role in the fixation of carbon. Nearly all the biomass on earth has had contact with this very inefficient molecule [7]. The molecules inefficiency and abundance drives the bioscience world to Rubisco's possible improvement. Since all plants use this molecule that performs the carbon fixing function, improvement of Rubisco and its active site which is non-competitive would greatly reduce the amount of energy and nitrogen that a plant needs to build biomass [7]. Rubisco present in every "green" plant can now be extracted as a protein ingredient for the food market. Netherlands Institute for Dairy Research (NIZO) has developed an extraction method resulting in a colourless protein isolate having an excellent solubility *[8]*, hence interest on extraction of Rubisco increased in the world of biotechnology because of its potential application in generating cellulose transgenic plant which can be powerful tool for use in the production of commercial biomass conversion [9].Nearly hundred solved NMR and X-ray 3D crystal structures of Rubisco are available in protein data bank *[10]* due to its abundance *[11-13]*. These enzymes optimal temperatures range from mesophilic *[14]* to moderately thermophilic *[15]*. However, psychrophilic

[12] and hyperthermophilic [16] enzymes have been isolated. Rubisco thermophilic activity is useful to enhance photosynthesis and increase agricultural productivity [17]. Rubisco activity and photosynthesis are affected with increase in temperature as a result of global warming [11]. Despite its abundance and important to plant [18], Rubisco from diverse microorganism including bacteria and some archaea such as *Thermosynechococcus elongatus* BP-1 [15], *Thermococcus kodakaraensis* [19], *Rhodopseudomonas palustris [20]* and *Burkholderia fungorum* [21] have been studied. Molecular dynamics (MD) simulation is widely used to study the dynamics and folding of large macromolecules such as proteins, this is achieved through the understanding of the interactions between atoms [22]. This is of great advantage because you can understand the behaviour of a molecule without synthesizing it, and it reduces the high cost of using NMR or X-ray crystallography in determining the structure of a molecule [23]. Recent genome studies have demonstrated the presence of Rubisco genes in unicellular cyanobacterium *Thermosynechococcus* sp (*T. NK55a*) suggesting a general, despite unknown modeled structure of the gene. The present study provides the first structure of a thermophilic unicellular cyanobacterium *T. NK55a* Rubisco, isolated from the Nakabusa hot spring, this research uses the information provided in the amino acids sequence and comparison with its mesophiles and psychrophiles counterparts, hence discuss the predicted Rubisco structure and its features, which might provide more details on its thermostability.

2. MATERIALS AND METHODS

2.1. Sequence Retrieval, Template Selection and Analysis

The Rubisco gene is part of genome of thermophilic cyanobacterium *T.* NK55a isolated from hot spring in Nakabusa [24], Sequence of the gene was obtained from Uniprot [25] with accession number V5V441. Template with highest sequence identity was selected through NCBI Blast server and was used as reference to build 3D model structure of the enzyme. Four other sequences from organisms of diverse environment with known structure were selected as shown in Table 1 from RCSB protein data bank [26] and aligned using Jalview [27].

2.2. Homology Modeling and Model validation

I-Tasser [28], Raptor X [29], and Swiss model [30] were used for homology modeling. Out of the three models generated, PROCHECK [31] was used to construct Ramachandran plot and validate the structure and the model with the highest PROCHECK G-score was selected.

2.3 Molecular dynamics study

GROMACS version 4.5.1 with Gromos96 53a6 force field [32] was used for protein stability simulation. The MD simulation of the model Rubisco was carried out at 10 ns at 67^0C (340K) after neutralizing the model system with Na^+ in a simple cubic box. Active site was visualized using visual discovery studio [33].

3. RESULTS AND DISCUSSION

3.1 Sequence Retrieval and template analysis

The sequence retrieved has a length of 475 amino acids similar to other Rubisco large subunit from different organisms with average length of 380 to 490 amino acids residues. In Table 1, the protein data bank identities of four different organisms were shown together with their adaptive environment, which ranges from psychrophiles to thermophiles. The closet template to the query sequence 3ZXW [34] from *Thermosynechococcus elongatus* have the highest identity with similarity index of 99.0%, E-value of 0.0 and score of 2,512. This shows that the organism have high quality of the match and descendent from common evolutionary ancestors. Multiple sequence alignment constructed using Jalview, the sequence was aligned with four other species of known structures, in Figure 1, the result from the alignment indicates active site amino acids that involve in the enzyme thermostability and conserved regions, mostly between thermophile and its mesophilic counterpart. This shows a greater link between the mesophiles than the psychrophiles [35]. *Thermosynechococcus NK55a* Rubisco like most thermostable enzymes, they have an increased number of salt bridge networks and interactions through hydrogen bonds. Instead of Lysine or Glycine found in the mesophiles and Phenylalanine in psychrophiles, Glutamate residue is found at position 261 which forms a salt bridge with lysine 258 and gives greater thermostability to the enzyme [36]. However, Rubisco from this *T.* NK55a have this glutamate at it position 261. It is also reported [37] that serine is normally replaced by the Cysteine at position 247 of the psychrophilic Rubisco, which increases their thermolability. The Rubisco from *T.* NK55a have cysteine in that position, this enable it to form salt bridge, thereby increasing it thermostability. Further salt bridge was found between glutamate at position 470 and Lysine at residue 131in the *T.* NK55a Rubisco which is not found in psychrophiles and mesophiles. Raptor X as shown in Table 2, which according to the score is the best model that nearly satisfies the criteria for the stereochemical restraints with PROCHECK G-score of 97.3%, the Ramachandran plot in Figure 4, 97.3% residues are in the favored region, 2.3% residues are in allowed region and 0.4% residues are in disallowed region. The 3D structure of the Rubisco from *T. NK55a* in Figure 2, was constructed using Raptor X, the result of the modeled structure is composed of alpha helices (which is shown in pink colour) and beta sheet (in golden colour), and loops in gray colour. The ribbon diagram of the Rubisco structure indicates the presence of TIM-barrel where the active pocket or catalytic domain is usually found.

3.2 MD Simulation

RMSF analysis in Figure 4, revealed that most of the residues in this Rubisco model fluctuated between 0.1nm to 0.3nm throughout the simulation period and usual presence of high fluctuation in the N-terminal due to loop and turn structure as it is loosely bound to the protein structure. The

amino acids found between 200-210 residues have the highest fluctuation up to 0.3nm. The active amino acid Lysine [38] is found at the position 210 is likely the reason why the position has this high fluctuation. The average RMSD results according to C-alpha, visible atoms and backbone was 0.23±0.04 in 10ns simulation time which is in an acceptable range. In Figure 5, the potential energy remains in equilibrium during the simulation. This result indicates the protein is stable.

3.3 Active site identification

According to [38], the residues that form the active pockets are His 302, His 335, Ser387, Gln409, Thr182, and Lys210, in the Figure 6, the active pocket is shown in red colour indicating the active site of the enzyme.

4. CONCLUSION

In this study, the geometrical quality of the C alpha backbone-backbone conformation indicates the reliability of the enzyme structure like most other thermostable enzymes. Molecular dynamic simulation (RMSD and RMSF) generated during simulation indicates the stability of the enzyme structure. The data is however providing a good foundation for experimentally derived crystal structure of this enzyme. Therefore the thermostability of this enzyme could be considered in reengineering plants with resistivity to high temperature as well as preventing the decrease in the plant enzymatic activity due to the rapid increase in the global warming.

CONFLICT OF INTEREST

The authors have no conflict of interest.

ACKNOWLEDGEMENT

We are grateful to Bioinformatics Research Group of Universiti Teknologi Malaysia.

REFERENCES

1. Falcicchio P, Levisson M, Kengen SM, Koutsopoulos S, van der Oost J. (Hyper) thermophilic Enzymes: Production and Purification, in Protein Downstream Processing. Humana Press, 2014. 487-496.
2. Fields PA, Wahlstrand BD, Somero GN. Intrinsic versus extrinsic stabilization of enzymes. Eur J Biochem. 2001; 268(16):4497-4505.
3. Grogan DW. Hyperthermophiles and the problem of DNA instability*:* Molecular Microbiology. J Tro Eco. 1998; 28(6): 1043-1049.
4. Elleuche S, Schröder C, Sahm K, Antranikian G. Extremozymes—Biocatalysts with unique properties from extremophilic microorganisms. Curr Opin Biotechno. 2014; 29:116-123.
5. Satyanarayana T, Raghukumar C, Shivaji S. Extremophilic microbes: Diversity and perspectives. Curr Sci. 2005; 89(1):78-90.
6. Vieille C, Zeikus GJ. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. Microbiol Mol Biol Rev. 2001; 65(1):1-43.

7. Andersson I, Catalysis and regulation in Rubisco. J Exp Bot. 2008; 59(7):1555-1568.
8. Pouvreau L, Smit B, van de Velde F. Securing Food Proteins: From by-Products to Functional Ingredient. Gums and Stabilisers for the Food Industry 17: The Changing Face of Food Manufacture: The Role of Hydrocolloids 2014, 346: 46.
9. Sticklen M. Plant genetic engineering to improve biomass characteristics for biofuels. Curr Opin Biotechno. 2006; 17(3):315-319.
10. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat T, Weissig H, Shindyalov IN, Bourne PE. The protein data bank. Nucleic Acids Res. 2000; 28(1): 235-242.
11. Sage RF, Way DA, Kubien DS. Rubisco, Rubisco activase, and global climate change. J Exp Bot. 2008; 59(7): 1581-1595.
12. Devos N, Ingouff M, Loppes R, Matagne RF. RUBISCO adaptation to low temperatures: a comparative study in psychrophilic and mesophilic unicellular algae. J Phycol. 1998; 34(4): 655-660.
13. Lyon BR, Mock T. Polar Microalgae: New Approaches towards Understanding Adaptations to an Extreme and Changing Environment. Biol. 2014; 3(1): 56-80.
14. Matsumura H, Mizohata E, Ishida H, Kogami A, Ueno T, Makino A, Inoue T, Yokota A, Mae T, Kai Y. Crystal Structure of Rice Rubisco and Implications for Activation Induced by Positive Effectors NADPH and 6-Phosphogluconate. J Mol Biol. 2012; 422(1): 75-86.
15. Nakamura Y, Kaneko T, Sato S, Ikeuchi M, Katoh H, Sasamoto S, Watanabe A, Iriguchi M, Kawashima K, Kimura T, Complete genome structure of the thermophilic cyanobacterium Thermosynechococcus elongatus BP-1. DNA Res. 2002; 9(4): 123-130.
16. Fukui T, Atomi H, Kanai T, Matsumi R, Fujiwara S, Imanaka T. Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with Pyrococcus genomes. Genome Res. 2005; 15(3): 352-363.
17. Spreitzer RJ, Salvucci ME. Rubisco: structure, regulatory interactions, and possibilities for a better enzyme. Annu Rev Plant Biol 2002; 53: 449-75.
18. Raven JA, Rubisco: still the most abundant protein of Earth? New Phytol. 2013; 198(1): 1-3.
19. Nohara K, Orita I, Nakamura S, Imanaka T, Fukui T. Genetic Examination and Mass Balance Analysis of Pyruvate/Amino Acid Oxidation Pathways in the Hyperthermophilic Archaeon Thermococcus kodakarensis. J Bacteriol. 2014; 196(22): 3831-3839.
20. Larimer FW, Chain P, Hauser L, Lamerdin J, Malfatti S, Do L, Land ML, Pelletier DA, Beatty JT, Lang AS. Complete genome sequence of the metabolically versatile photosynthetic bacterium Rhodopseudomonas palustris. Nat Biotechnol. 2003; 22(1): 55-61.
21. Nakano T, Saito Y, Yokota A, Ashida H. His267 is involved in carbamylation and catalysis of RuBisCO-like protein from *Bacillus subtilis*. Biochem Biophys Res Commun. 2013;

431(2):176-180.

22. Oyedotun KS, Lemire BD. The quaternary structure of the saccharomyces cerevisiae succinate dehydrogenase Homology Modeling, Cofactor Docking, and Molecular Dynamics Simulation Studies. J Biol Chem. 2004;. 279(10): 9424-9431.
23. Fowler SB, Best RB, Toca Herrera JL, Rutherford TJ, Steward A, Paci E, Karplus M, Clarke J. Mechanical Unfolding of a Titin Ig Domain: Structure of Unfolding Intermediate Revealed by Combining AFM, Molecular Dynamics Simulations, NMR and Protein Engineering. J Mol Biol. 2002; 322(4):841-849.
24. Stolyar S, Liu Z, Thiel V, Tomsho LP, Pinel N, Nelson WC, Lindemann SR, Romine MF, Haruta S, Schuster SC. Genome sequence of the thermophilic cyanobacterium *Thermosynechococcus* sp. strain NK55a. Genome announc. 2014; 2(1): e01060-13.
25. Consortium U., Activities at the Universal Protein Resource (UniProt). Nucleic Acids Res. 2014; 42(D1): D191-D198.
26. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res. 2000; 28(1): 235-242.
27. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinform. 2009; 25(9):1189-1191.
28. Zhang Y. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 2008; 9(1): 40.
29. Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, Xu J. Template-based protein structure modeling using the RaptorX web server. Nat Protocols. 2012; 7(8): 1511-1522.
30. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L, Schwede T. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 2014.
31. Lovell SC, Davis IW, Arendall WB, de Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. Proteins: Struct Funct Bioinf. 2003; 50(3): 437-450.
32. van der Spoel D, Lindahl E, Hess B, Van Buuren A, Apol E, Meulenhoff P, Tieleman D, Sijbers A, Feenstra K, van Drunen R. *GROMACS user manual version 3.3.* 2008.
33. GLANCE NFAA. *WHAT'S NEW IN DISCOVERY STUDIO® 3.5.*
34. Nakamura Y, Kaneko T, Sato S, Ikeuchi M, Katoh H, Sasamoto S, Watanabe A, Iriguchi M, Kawashima K, Kimura T, Kishida Y, Kiyokawa C, Kohara M, Matsumoto M, Matsuno A, Nakazaki N, Shimpo S, Sugimoto M, Takeuchi C, Yamada M, Tabata S. Complete Genome Structure of the Thermophilic Cyanobacterium *Thermosynechococcus* elongatus *BP-1*. DNA Res. 2002; 9(4): 123-130.

35. Kulbachinskiy A, Bass I, Bogdanova E, Goldfarb A, Nikiforov V. Cold sensitivity of thermophilic and mesophilic RNA polymerases. J Bacteriol. 2004; 186(22): 7818-7820.
36. Gubernator B, Bartoszewski R, Kroliczewski J, Wildner G, Szczepaniak A. Ribulose-1, 5-bisphosphate carboxylase/oxygenase from thermophilic cyanobacterium *Thermosynechococcus elongatus*. Photosyn. Res. 2008; 95(1): 101-109.
37. Dolhi J, Maxwell D, Morgan-Kiss R. Review*:* the Antarctic *Chlamydomonas raudensis*: an emerging model for cold adaptation of photosynthesis. Extremophiles 2013; 17(5): 711-722.
38. Stec B. Structural mechanism of RuBisCO activation by carbamylation of the active site lysine. Proc Natl Acad Sci. 2012; 109(46): 18785-18790.

Supplementary Files

Table 1: Accession of Various Organisms with Their Adaptive Environment

	Pdb ID	**Organism**	**Adaptive environment**
1.	1WDD	*Oryza sativa subsp japonica*	Mesophile
2.	4LF1	*Rhodopseudomonas palustris*	Psychrophile
3.	3ZXW	*Thermosynechococcus elongatus* BP-1	Thermophile
4.	3NWR	*Burkholderia fungorum*	Halophile

Table 2: Result From Three Different Homology Modeling Servers

SERVER NAME	**PROCHECK (G-SCORE)** **Expected score 98.0%**
I-Tasser	90.5%
Swiss Model	97.2%
Raptor X	97.3%

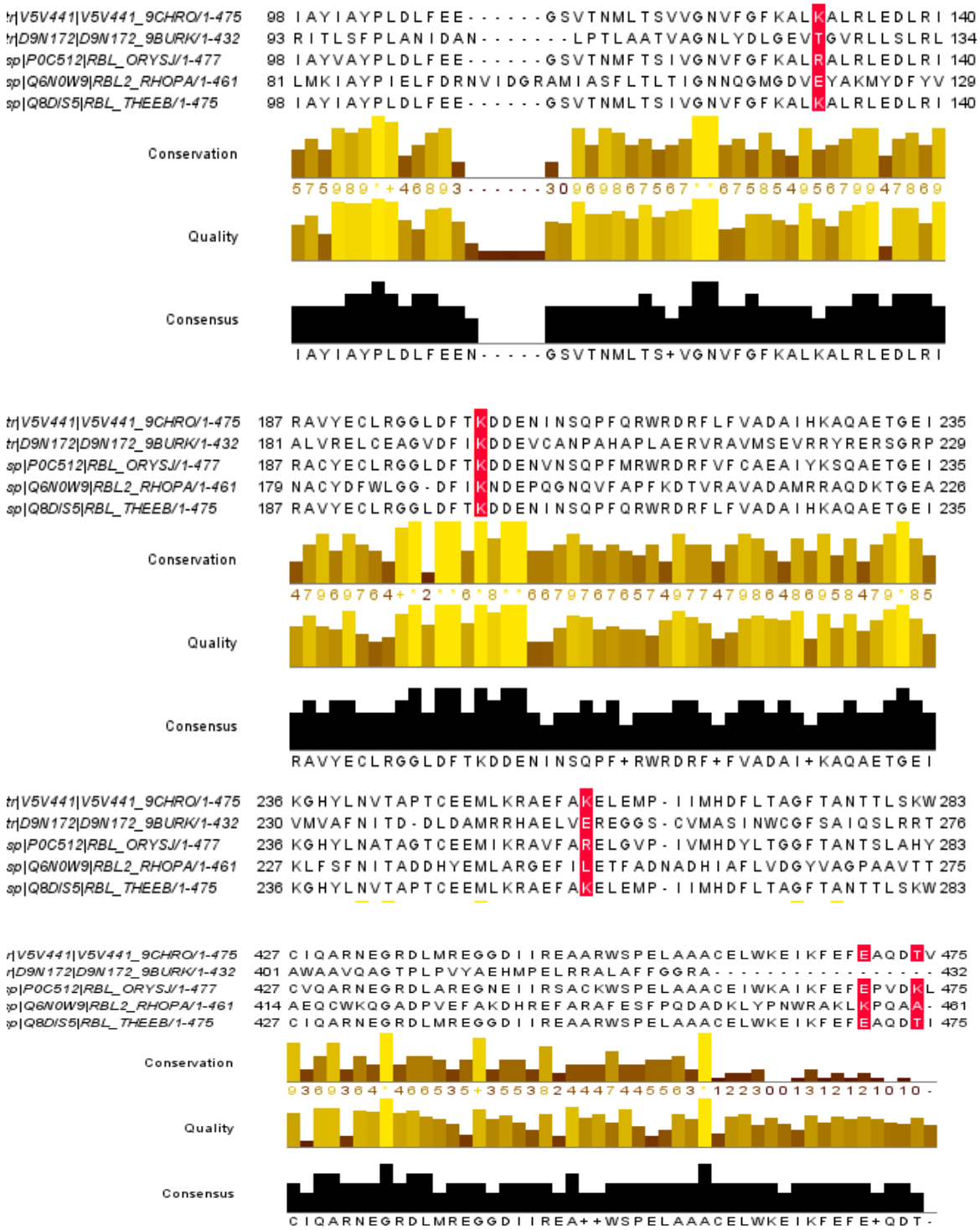


Figure 1: Multiple sequence alignment and Active site amino acids. The golden and yellow bars are showing the highly conserved regions of the enzymes, the amino acids shaded with red indicate the active site amino acids from the literature.

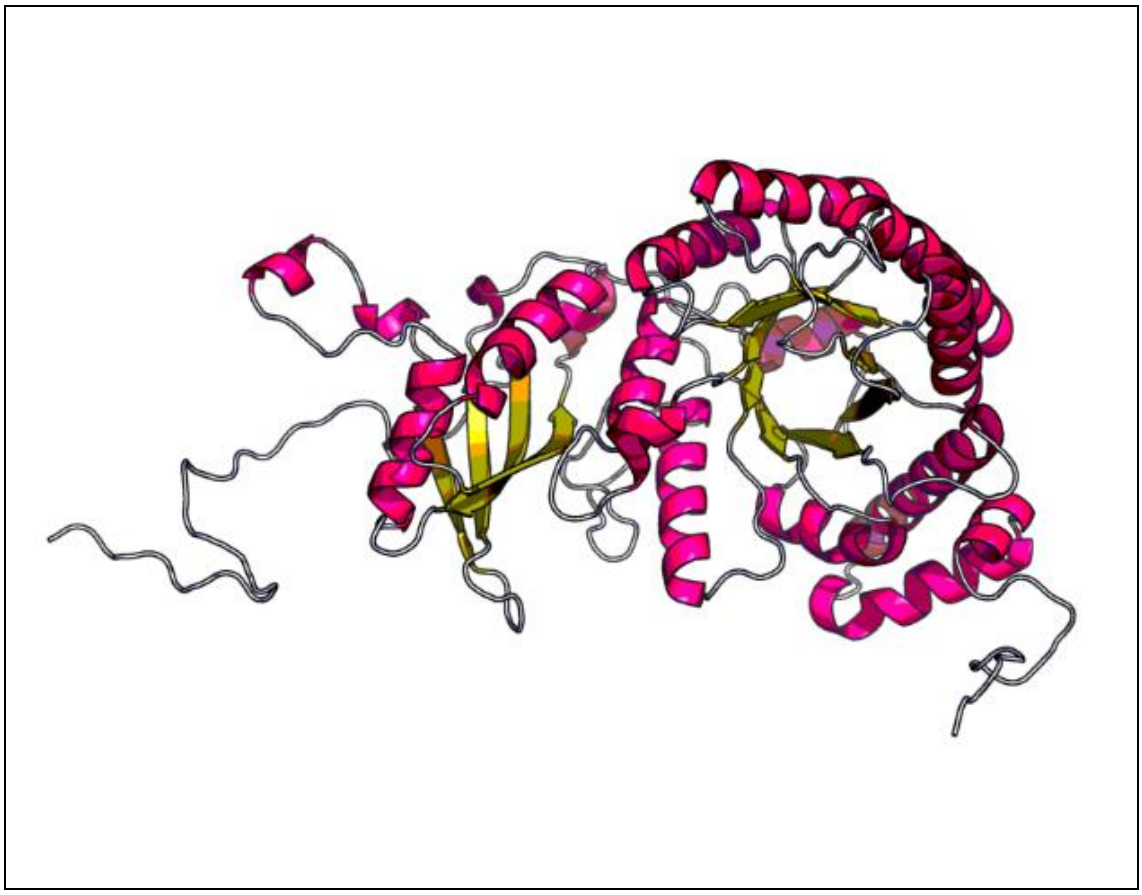


Figure 2: The ribbon diagram of the Rubisco structure from Raptor X

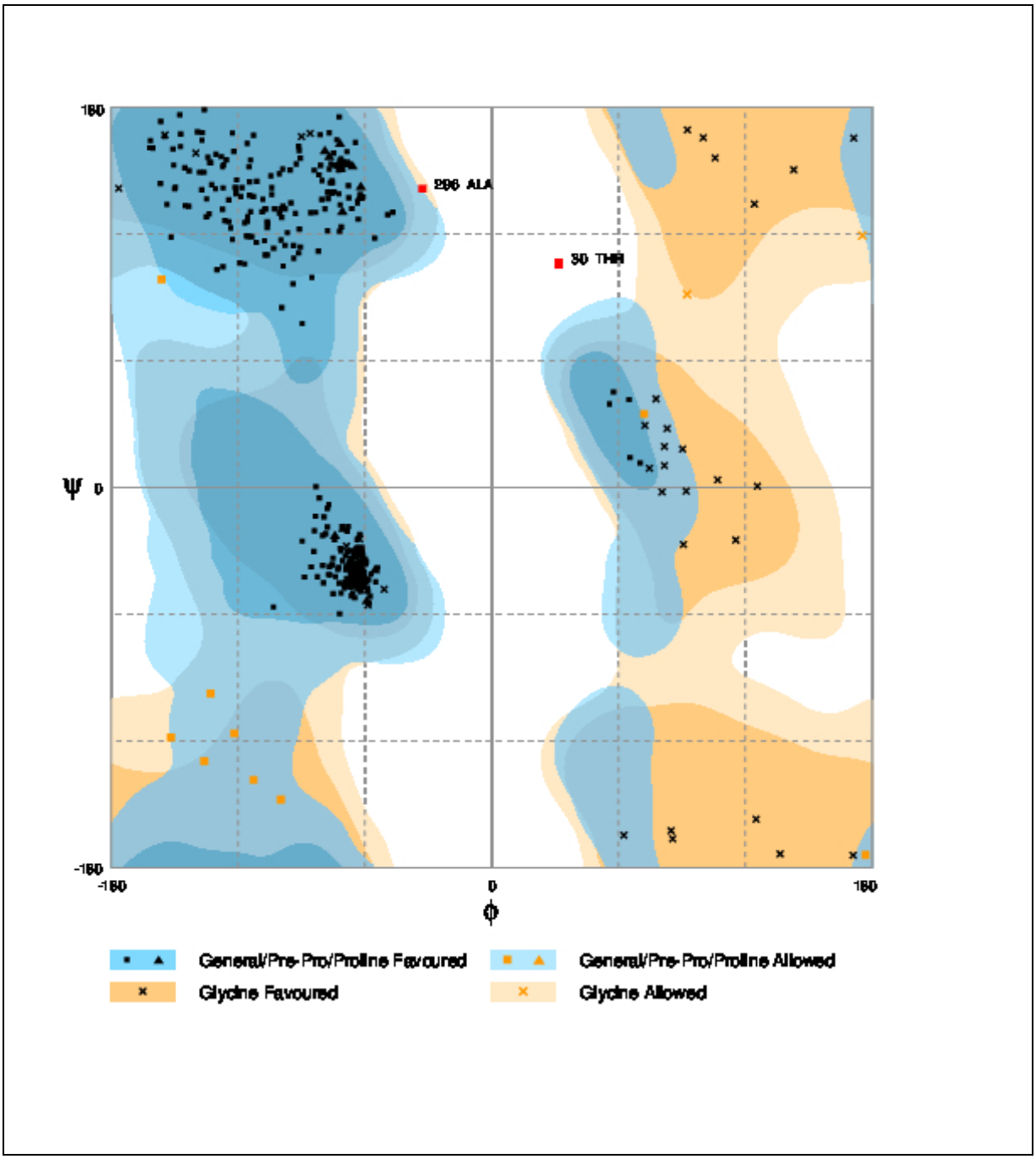
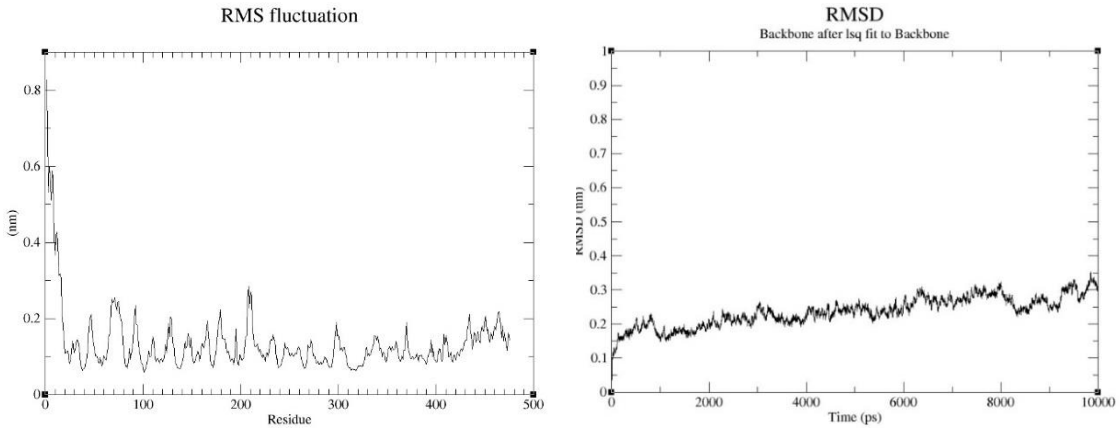


Figure 3: Ramachandran plot from Raptor X model



A. B.

Figure 4: A. Graph showing the RMSF of the modeled structure B. graph showing the C-alpha backbone-backbone RMSD simulation after 10ns

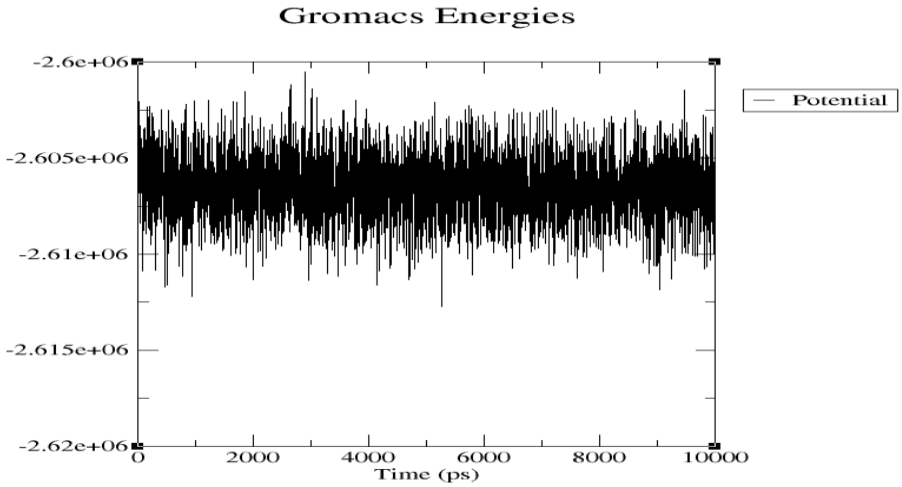


Figure 5: Potential energy of the molecular dynamics simulation

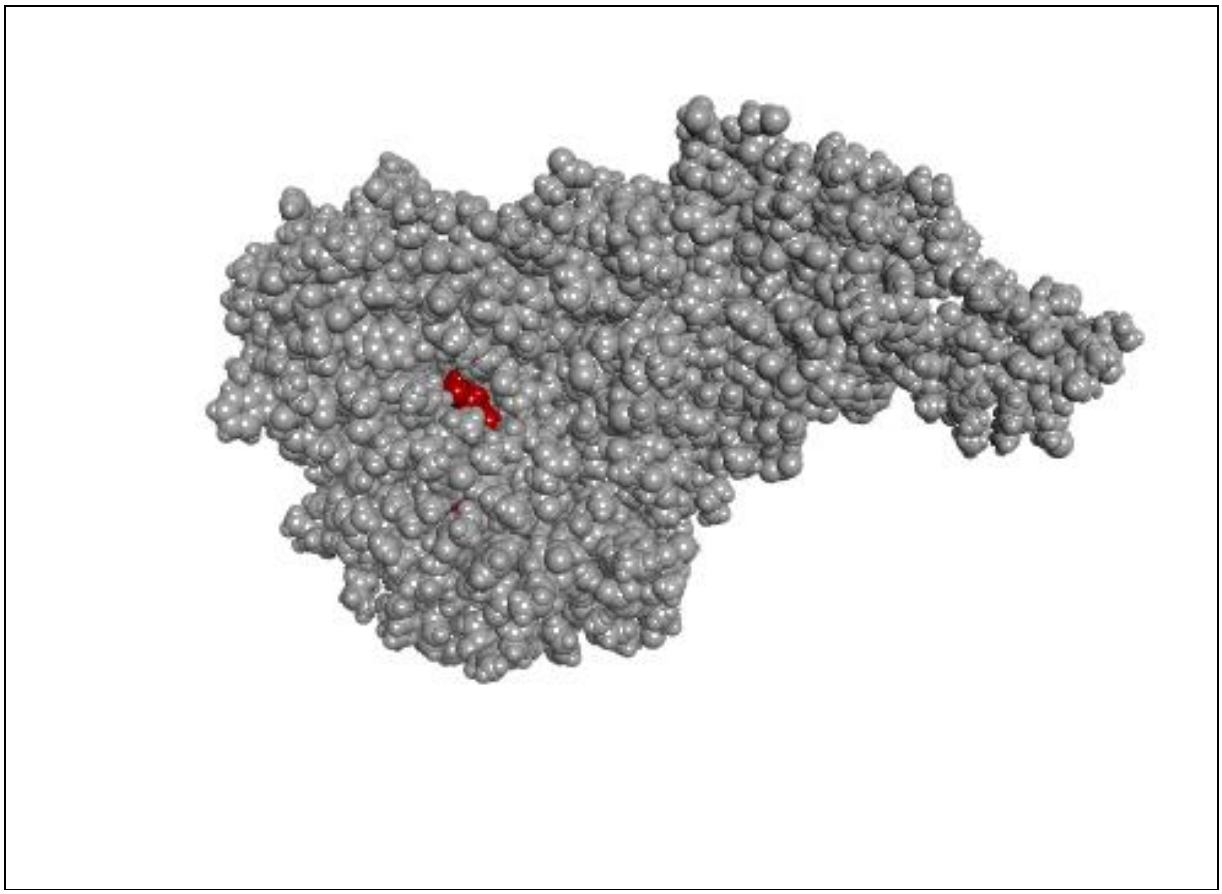


Figure 6: Active site pocket in cartoon grey format; the red colors are showing the residues that make up the active pocket.