

Structural and Molecular Dynamics Studies of UDP Glucose Pyrophosphorylase Dimerization in Rice (*Oryza sativa* L.)

Mamatha Y. S., Redhu S. Neeru and Kumar Sudhir*

Department of Molecular Biology, Biotechnology & Bioinformatics
CCS Haryana Agricultural University, Hisar – 125 004, India

*Corresponding Author E-mail: mammuy06@gmail.com

Received: 10.08.2020 | Revised: 24.09.2020 | Accepted: 29.09.2020

ABSTRACT

UGPase enzyme is involved in sucrose synthesis as a catalyzing agent in the reaction, Mg^{+2} -UTP + Glu-1-P \rightleftharpoons PPi + UDP-Glu. It catalyzes both forward and reverse reaction depending on the metabolic status of the tissue. UGPase contains three domains: N-terminal domain, catalytic domain with nucleotide binding loop at center and the C-terminal domain. UGPase as a monomer is the functionally active in most photosynthetic plants and microorganisms. Comparative structural modelling of UGPase was performed using models predicted from Modeller, Swiss-Model and I-TASSER. All predicted models were further assessed, evaluated and refined through QMEAN, WHATIF and SAVES server to achieve most stable model. To study the interaction between monomers units, homodimers were built through protein-protein docking. Hydrogen bonding between monomer units, solvent accessible area and ligand binding/active sites were predicted with H Bond, NACCESS and CastP server. Solvent accessible area in dimer was approx. 100-200 Å smaller as compared to monomer. This study showed that reduction in solvent accessible area and subsequent smaller active site resulted in loss of activity due to dimerization in UGPase.

Keywords: Molecular Dynamics (MD), Comparative Modelling, UGPase, *Oryza Sativa* L., Phylogenetic analysis.

INTRODUCTION

Rice (*Oryza sativa* L.), belonging to poeaceae family, is one of the most important staple cereals food that is grown extensively all over the world. It is rich in carbohydrates mainly starch and contributes towards 87% of total calorie content in human diet. Carbohydrates are the primary fuel source for human. UDP-

Glucose pyrophosphorylase plays a major role in the synthesis of starch in rice.

UDP-glucose pyrophosphorylase (UGPase) (EC 2.7.7.9) is a vital enzyme producing UDP-glucose through catalyzing, Mg^{+2} -UTP + Glu-1-P \rightleftharpoons PPi + UDP-Glu, reaction (Kleczkowski, 1994).

Cite this article: Mamatha, Y. S., Redhu, S., & Neeru & Sudhir, K. (2020). Structural and Molecular Dynamics Studies of UDP Glucose Pyrophosphorylase Dimerization in Rice (*Oryza sativa* L.), *Ind. J. Pure App. Biosci.* 8(5), 258-267. doi: <http://dx.doi.org/10.18782/2582-2845.8340>

The UDP-Glucose is significant for the synthesis of cell wall polysaccharides as well as in production of carbohydrate moiety of glycolipids, glycoproteins and proteoglycans (Bishop et al., 2002). UDP-Glucose also serves as precursor for the development of sucrose and cell wall polymers like callose, hemicellulose, cellulose etc. Moreover, within both non-photosynthetic tissue and developing leaves UGPase acts as catalyst in sucrose degradation pathway. During sucrose degradation, UGPase produces UTP and glucose-1-P from UDP-Glucose and pyrophosphate. This Glucose-1-P is not only a substrate for ADP-Glucose pyrophosphorylase but also a key player of starch formation in heterotrophic tissues such as oilseed rape embryos and barley endosperm (Schwender et al., 2015).

The rice genome contains two homologous UGPase gene, *OsUgp1* on chromosome 9 (Abe et al. 2002) and *OsUgp2* on chromosome 2 (Mu. 2002). *OsUgp1* gene has six isoforms, named as P1 to P6. These isoforms undergoes conserved but dissimilar N-terminal acetylation. The conserved protein modification might be responsible for the distinct roles of these isoforms during plant development and their respective regulations.

UGPase exists as a mixture of monomer, dimer and higher order oligomers (Martz et al., 2002). In rice, monomer is reported as most active and dimer as inactive. Inactivity of oligomers might be because of their post translational modifications. Comparative structural modeling and molecular dynamic simulation studies might help in revealing conformational differences between UGPase monomer and dimers.

Keeping this in view, the present study was conducted to undertake molecular simulation for dimerization of UDP-Glucose pyrophosphorylase and to compare structural analysis of predicted dimers of UDP-Glucose pyrophosphorylase.

MATERIALS AND METHODS

Sequence of rice UGPase was retrieved from NCBI database. Aliphatic index, instability

index, molecular weight, hydropathicity and theoretical pI values of UGPase were calculated by using ProtParam and Tm predictor. Three 3D models (Model1, Model2 and Model3) of UGPase monomers were generated by using Modeller9.20 (Sali & Blundell, 1993), Swiss-model online server (Guex N. & Schwede M. C. 2009; <https://swissmodel.expasy.org/>) and I-TASSER server (Zhang et al., 2015; <https://zhanglab.ccmb.med.umich.edu/I-TASSER/>) respectively. Modeller9.20 and Swiss-model server uses homology modelling whereas I-TASSER server predicts model through multiple threading alignment method. 5WEG was selected as a template with 94.66% identity for Modeller9.20 and Swiss-Model server. Correspondingly, multiple threading alignment algorithm of I-TASSER also chose 5WEG as a template.

Structure Assessment tool, QMEAN, of Swiss model server was used for estimating quality of models. Models were then refined with locPREFMD (Feig, 2016; <http://feig.bch.msu.edu/locprefmd>) and Galaxy refine server (Ko et al., 2012; <http://galaxy.seoklab.org/>). Refined models were further evaluated and validated through WHATIF, Swiss-structure assessment tool and SAVES server.

Models were subsequently simulated by Gromacs-2019. Steepest descent method for energy minimization of models was opted with <1000 KJ/mol energy cutoff. MD simulation with parameters such as temperature (300K), pressure (1.0 bar) and density (1000 kg m⁻³) were stabilized over a period (100ps) and executed for 20 ns (i.e. $time=nstep*dt$, $nstep=1crore$, $dt=0.002$). RMSD of simulated and unsimulated monomer models were graphically plotted.

Homodimers using predicted models were generated by pyDock and ZDOCK protein-protein docking servers. Afterwards dimers were verified through SAVES and WHATIF server. Generated dimers were then simulated by Gromacs-2019 and analyzed for intra-protein interaction with Dim plot. Active/ligand binding sites in monomers were

predicted by Cast P server. Solvent accessible surface area for residues was calculated in both monomer and dimers by using NACCESS tool (Hubbard, 1996).

RESULT AND DISCUSSION

In silico characterization

Sequence for rice (*Oryza sativa* L) UGPase was retrieved from NCBI with accession number ACA50487.1. Additionally 114 complete UGPase protein sequences were retrieved from NCBI. Monomer of rice UGPase was predicted thermo as well as test tube stable through instability and aliphatic index score of 23.59 and 100.96 respectively by ProtParam. Rice UGPase was characterized with -0.163 GRAVY, 51682.29 molecular weight and 5.43 theoretical pI. Melting point of rice UGPase was estimated by using Tm predictor to be in 55-65^o C range. GRAVY score, aliphatic index and melting point altogether points towards the globular structure of UGPase. Similar results were obtained from other UGPase sequences denoting the conservation in their physiochemical properties during the course of evolution.

Phylogenetic Analysis

A Total of 114 protein sequences from Plants, Fungi and Bacterial UGPase retrieved from NCBI were selected for construction of phylogenetic tree. Maximum parsimony methodology from MEGA7 tool was employed for building phylogenetic tree using selected sequences. Tree depicted closer evolutionary relationship within plantae, algae, fungi and bacterial clades. Poaceae family member were observed to be in same clade highlighted with yellow color and rice UGPase was estimated closer to sugarcane and common millet UGPase (Fig. 1).

Comparative modelling

3D models constructed from Modeller9.20, Swiss Model and I-TASSER named Model1, Model2 and Model3 respectively. 5WEG was used as template for all the three models. The secondary structures of predicted models were studied and compared to template as shown in Table 1. Loop region in all three models were found comparable whereas helix region of

Model1 and sheet region of Model2 were observed to be consistent with template.

The structural quality of models was assessed using QMEAN and Z score. QMEAN score of Model1, Model2 and Model3 were 0.541, 0.695 and 0.718 respectively. Similarly, Z-score were -1.45, -2.54 and -1.85 for Model1, Model2 and Model3 respectively. The scores of models constructed in present study were found closer to optimal Z score of 0 and within optimal QMEAN range of 0-1. Both QMEAN and Z score of models complied with structural quality of solved X-ray crystallography reference structures. The minimum energy calculated for all Model1 (-22420.309KJ/Mol), Model2 (-23439.170KJ/Mol) and Model3 (-24260.711KJ/Mol) further supported that the models were stable and can be taken for further studies.

All models were then validated by WHATIF server. Coarse Packing quality was computed -1.047 for Model1, -0.548 for Model2 and -0.584 for Model3, illustrating all models are of good quality. Omega average, anomalous bond length and anomalous bond angles were other validation programs used to check structural errors. Trans-peptide omega angle in predicted models was approximately +178^o and fulfills Gaussian distribution average for protein structures. The calculated Z-score of anomalous bond angle were 1.011, 0.872 and 0.935; and anomalous bond lengths were 0.665, 0.712 and 0.726 for Model1, Model2 and Model3 respectively. By and large, models were fulfilling protein structural criteria and can be used in subsequent studies.

Models were further analyzed by SAVES (Structural Analysis and Verification Server) using PROCHECK, PROVES, VERIFY3D and ERRAT packages (Table 2). VERIFY3D determined 3D model compatibility with its amino acid sequences and with a >84% score every predicted models/monomers were given a PASS status. Likewise, ERRAT compared predicted models/monomers with refined known structures by calculating non-bonded interactions between atoms and estimated

them as Good with >90% score. Ramachandran plots for assessment of backbone conformation were generated from PROCHECK. Plots showed core region between 85.1 to 94.3% and disallowed region between 0 to 0.5% in all the models. PROVES calculated the atomic volume and Z-score for quality estimation in all models. Overall evaluation and verification deemed all models to fit for further simulation studies.

RMSD, calculated between models and 5WEG through superimposing their backbone, were recorded <1.0Å (Fig. 2). This further emphasizes predicted models/monomer structural quality.

Molecular Dynamic Simulation

Topology file was generated by using force field OPLS-AA/L all-atom force field (Jorgensen L., 1996) with water model SPCE (extended simple point charge model). Models were positioned in cubic box center with 1.0 nm distance from boundary. All models were stabilized with negative potential energy and $F_{max} < 1000.0$ (Fig. 3(a)). Model1 having lowest energy, followed by Model3 and Model2, was considered most stable. Radius of gyration measuring protein compactness for simulated models was around 2.4 nm *viz.* acceptable for the protein folding (Fig. 3(b)). RMSD between simulated and un-simulated models were calculated approx. 1.1 Å over a period (Fig. 3(c)). Additionally, this ascertains that no major change in simulated structure has been taken place as compared to un-simulated models.

Active site prediction

Ligand binding sites for simulated models was predicted using Cast P server (Tian et al., 2018). Active residues such as were also identified from NCBI's Conserved Domain Database. Sixteen amino acids were determined as active residues including K100, P190, G192, K257, G259, E272 etc. The largest pocket of central catalytic domain containing active residues was selected for subsequent analysis. Solvent accessible surface area and volume of the active site for each model was computed (Table 3).

Dimer formation

Homodimers were generated by using protein-protein docking servers (ZDOCK and pyDOCK) using the simulated models. Dimer A, Dimer B and Dimer C were generated by submitting two copies of Model1, Model2 and Model3 respectively to Z-Dock server (Pierce et al., 2014). Correspondingly, Dimer X, Dimer Y and Dimer Z were generated using two copies of Model1, Model2 and Model3 respectively to pyDock server (Cheng et al., 2007). Dimers were subsequently evaluated by VERIFY 3D.

Dimer analysis and verification

Six dimers were then validated with WHATIF server. Coarse Packing quality was computed and found within the range of -1.159 to -1.573, illustrating all dimers are of good quality. Anomalous bond angle and anomalous bond length Z-score were also recorded standard range of 1.943 to 2.029 and Z- 1.486 to 1.554 respectively.

Dimers were also analyzed by SAVES (Structural Analysis and Verification Server) using PROCHECK, VERIFY3D and ERRAT packages (Table 4). PROCHECK results showed core region in the range of 81.5 and 84.2% and disallowed region in the range of 0.2 and 1.0% in all the Dimers. Overall, all dimers were suitable according to structural principles of known protein.

Molecular Dynamic Simulation of Dimer

Topology file was generated by using force field CHARMM27 all-atom force field (MacKerell et al., 1998) with TIP3P water model. Dimers were placed in triclinic box center with the distance 1.0 nm from the edge. All dimers were stable with negative potential energy and $F_{max} < 1000.0$ (Fig. 4). Out of ZDOCK dimers, Dimer A was recorded with lowest energy followed by Dimer B and Dimer C. Similarly out of pyDOCK dimers Dimer X was found having lowest energy followed by Dimer Z and Dimer Y. RMSD calculated between simulated and un-simulated dimers from superimposition was listed approximately <1.0 over a period (Fig. 5).

Minimum free energy for every monomer and dimer models were deduced and

listed in Table 5. Dimers with lesser free energy were determined as more stable compare to monomer and might be responsible for dimerization.

Dimer analysis

At the end dimers were analyzed for identification of structural changes due to dimerization by Dim plot. It analyzes probable hydrogen bond and other non-covalent interactions between chains of oligomeric proteins. A total of 9, 10, 3, 5, 9 and 7 hydrogen bonds were determined between polypeptide chains of DimerA, DimerB, DimerC, DimerX, DimerY and DimerZ respectively. Residues involved in hydrogen

bond formation are shown in Table 6 and their solvent accessible surface area was calculated. Table 7 shows the total solvent accessible area for monomer and their respective Dimer chain A. The major change in solvent accessible area of amino acids constituting catalytic pocket like L90, K100, S101, S137, N139, T140, P190, G192, L253, K257, G259, E272, I273, F293, N296, K323, A337, I382, I395 etc. was observed. The change in solvent accessible area corresponds to change in binding site due to dimer formation. This might also coincide with the change in biological activity of monomer and dimer.

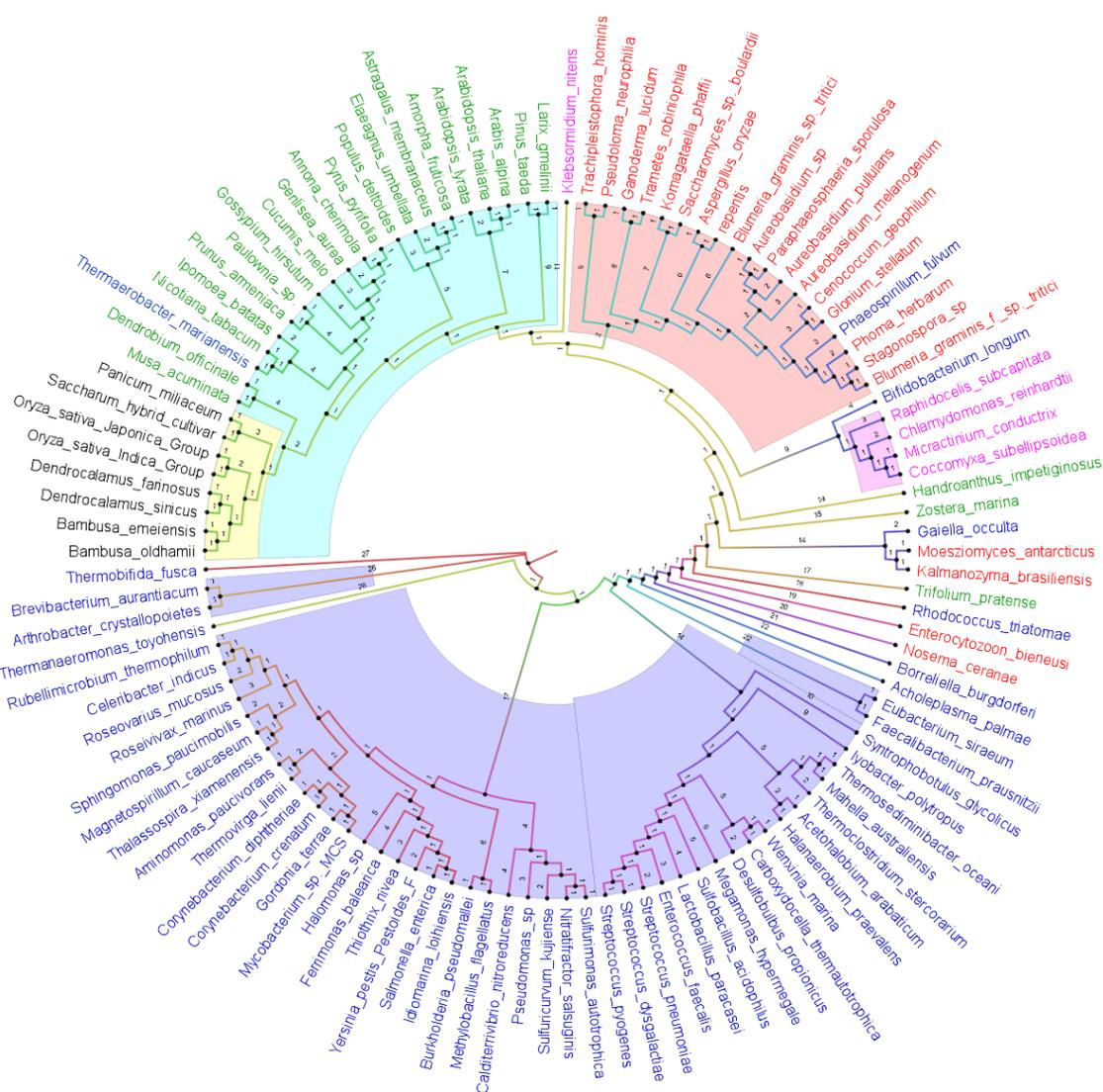


Fig. 1: Phylogenetic tree obtained from MEGA7. Taxa in Blue =Bacteria, Red=fungi, Pink=algae, Green=plants and Black=Poaceae family of plants.

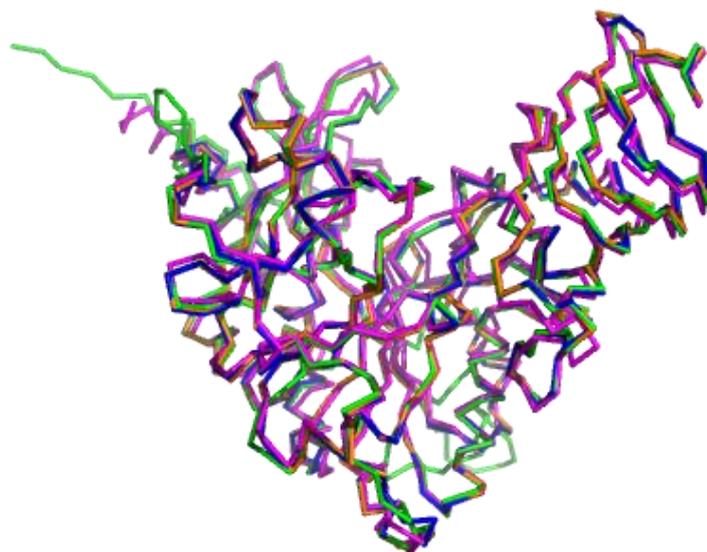


Fig. 2: Backbone superimposition of Model1 (Green), Model2 (Blue), Model3 (Magenta) and 5WEG_A (Orange)

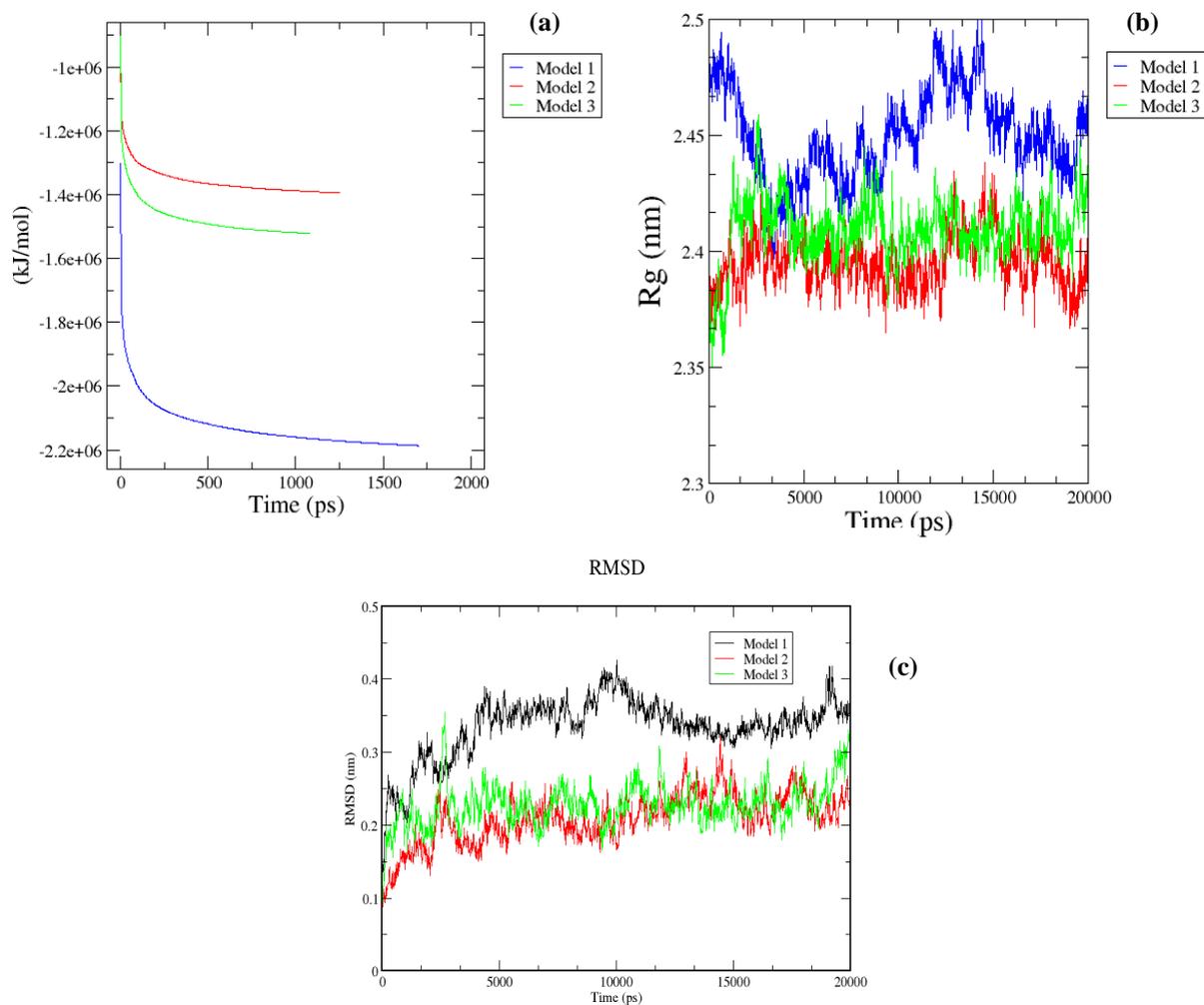


Fig. 3: (a) Potential energy graph for all models (b) Radius of gyration of simulated models (c) RMSD for simulated and un simulated models

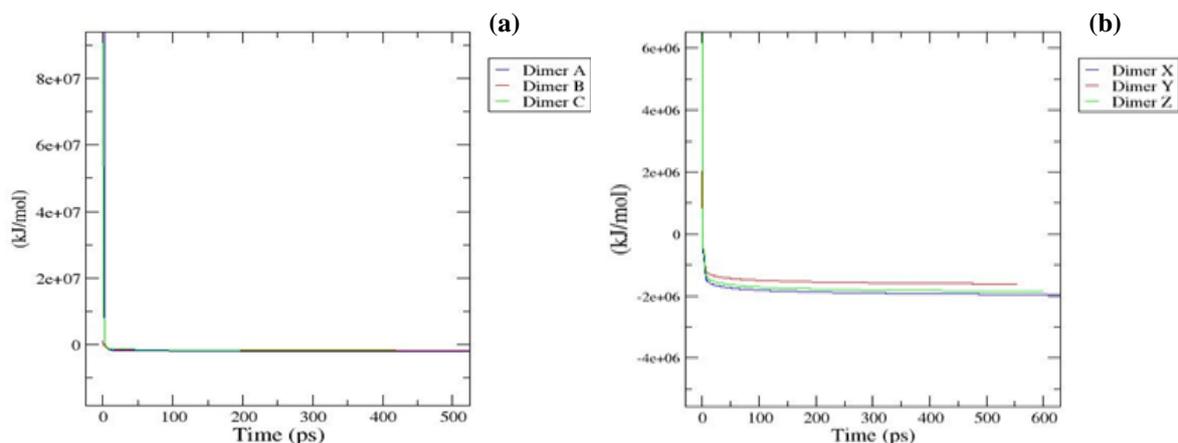


Fig. 4: Potential energies of all dimers obtained from (a) ZDOCK (b) pyDOCK

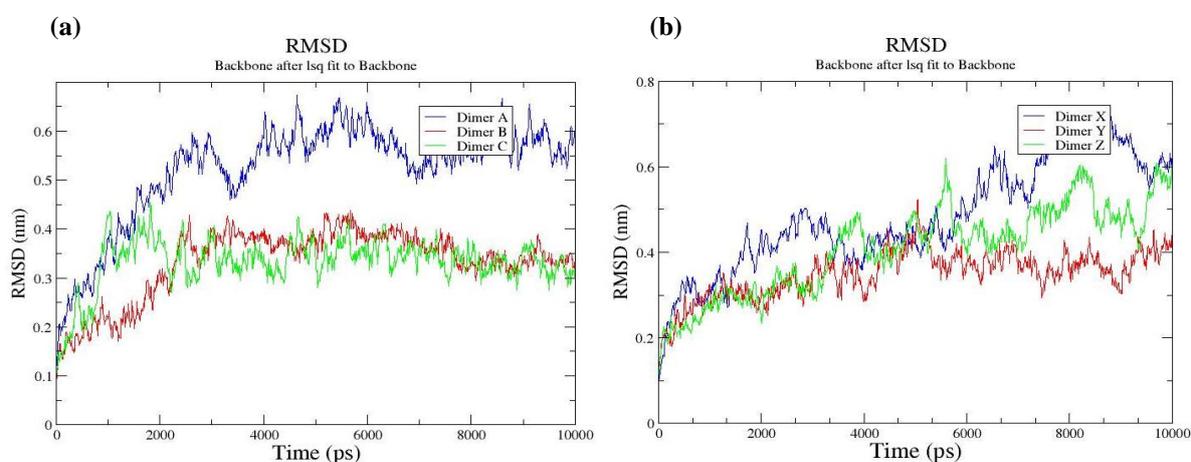


Fig. 5: RMSD for simulated and un-simulated obtained from (a) ZDOCK (b) pyDOCK

Table 1: Secondary structure details of UGPase Models and Template

Sr. No.	Model	Helix	Sheet	Loop
1	Model1	25	38	54
2	Model2	22	40	50
3	Model3	17	33	53
4	Template	24	32	52

Table 2: Analysis of models by SAVES (Structural Analysis and Verification Server)

Sr No.	Model	PROCHECK (%)				ERRAT	VERIFY 3D (%)	PROVES (%)
		Favorable region	Allowed region	General allowed region	Disallowed region			
1	Model1	85.1	14.1	0.2	0.5	90.5077	85.93	4.5
2	Model2	94.3	5.4	0.2	0.0	97.1047	98.70	2.8
3	Model3	91.0	8.3	0.2	0.5	95.6616	96.16	3.5

Table 3: Solvent accessible surface area and volume of the active site

Sr.No.	Models	SASA(Å ²)	Volume (Å ³)
1	Model1	2979.972	2418.407
2	Model2	1979.276	2603.030
3	Model3	1739.610	02620.732

Table 4: Structural details results of dimers from WHATIF server

Sr.No.	Dimer	PROCHECK (%)				ERRAT	VERIFY 3D (%)
		Favorable region	Allowed region	General allowed region	Disallowed region		
1	DimerA	81.5	16.1	1.5	1.0	A:91.49 B:91.49	90.72
2	DimerB	84.2	15.3	0.2	0.2	A:86.55 B:86.57	99.57
3	DimerC	82.9	16.6	0.2	0.2	A:85.34 B:85.34	98.51
4	DimerX	81.5	16.0	1.5	1.0	A:90.63 B:90.65	91.03
5	DimerY	84.1	15.4	0.2	0.2	A:86.24 B:86.41	99.13
6	DimerZ	82.8	16.7	0.2	0.2	A:85.91 B:85.94	98.08

Table 5: Minimum Energy for predicted models

Sr.No.	Model	Energy (KJ/Mol)	Dimer	Energy (KJ/Mol)	Dimer	Energy (KJ/Mol)
1	Model1	-3307.542	DimerA	-6589.254	DimerX	-5249.548
2	Model2	-4787.671	DimerB	-8941.594	DimerY	-9536.165
3	Model3	-4981.171	DimerC	-9259.852	DimerZ	-8550.889

Table 6: H-bond between peptide chains of Dimer

Sr.No.	Dimer	No. of H-bonds	Residues involved in H-bond
1	DimerA	9	T677-L134, D676-Y151, D676-K150, D489-R309, E618-K75, K619-N123, Y620-M135, K674-Q212, E493-D207
2	DimerB	10	K465-D415, K859-D56, K903-E67, K881-E63, K639-E282, E781-N345, D630-A248, D630-L247, K786-K245, E629-N386
3	DimerC	3	H749-T252, D747-N392, K798-D372
4	DimerX	5	D611-Y61, D610-S355, D846-K124, K592-D378, Q613-D62
5	DimerY	9	S484-A305, E497-R7, E497-S29, S495-S29, S491-S33, S491-E35, R469-E35, R469-E35, S488-K309
6	DimerZ	6	K592-R106, K593-R106, N540-D372, N575-N155, K618-S152, S620-S152

Table 7: solvent accessible surface area for atoms involved in binding interaction

Sr.No.	Model	SASA (Å)	Dimer	SASA (Å)	Dimer	SASA (Å)
1	Model1	3436.975	DimerA	3216.412	DimerX	3246.164
2	Model2	1959.265	DimerB	1843.426	DimerY	1806.349
3	Model3	1802.512	DimerC	1695.492	DimerZ	1689.642

CONCLUSION

The presence of rice UGPase in monomer and various oligomeric states was found to be associated with its function. The rice UGPase is reported biologically active and dimer as biologically inactive. In absence of detailed structural and functional information, the link between oligomeric state and biological activity is yet to investigate. This link was explored through computational modeling and subsequent comparative analysis of monomer and dimer of rice UGPase. The present study

involves first UGPase model generation via homology modeling and threading approach then their optimization through MD simulations. This was followed by subsequent generation and simulation of dimers. All predicted models satisfied the ponametus of best possible computational modeling and had high value of correctness. The active residues of UGPase catalytic domain were identified with Cast P and CDD. Based on solvent accessible surface area calculation of monomer vs dimers and corresponding change

in cavity of UGPase catalytic domain the difference in functional status of UGPase was determined. It was concluded that reduction in size of cavity/ligand binding site in catalytic domain leads to the lose UGPase activity due to dimerization.

REFERENCES

- Abe, T., Niiyama, H., & Sasahara, T. (2002). Cloning of cDNA for UDP-glucose pyrophosphorylase and the expression of mRNA in rice endosperm. *Theoretical and Applied Genetics*, 105, 216–221.
- Bishop, J. D., Moon, B. C., Harrow, F., Ratner, D., Gomer, R. H., Dottin, R. P., & Brazill, D. T. (2002). A second UDP-glucose pyrophosphorylase is required for differentiation and development in *Dictyostelium discoideum*. *Journal of Biological Chemistry*, 277(36), 32430-32437.
- Cheng, T. M. K., Blundell, T. L., & Fernandez-Recio, J. (2007). pyDock: electrostatics and desolvation for effective scoring of rigid-body protein–protein docking. *Proteins: Structure, Function, and Bioinformatics*, 68(2), 503-515.
- Feig, M. (2016). Local protein structure refinement via molecular dynamics simulations with locPREFMD. *Journal of chemical information and modeling*, 56(7), 1304-1312.
- Guex, N., Peitsch, M. C., & Schwede, T. (2009). Automated comparative protein structure modeling with SWISS-MODEL and Swiss-Pdb:A historical perspective. *Electrophoresis*, 30(S1), S162-S173.
- Hubbard, S. J., & Thornton, J. M. (1996). NACCESS, version 2.1. 1 computer program. *Department of Biomolecular Sciences, UMIST, Manchester, UK*.
- Jorgensen, W. L., Maxwell, D. S., & Tirado-Rives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, 118(45), 11225-11236.
- Kleczkowski, L. A. (1994a). Glucose activation and metabolism through UDP-glucose pyrophosphorylase in plants. *Phytochemistry* 37, 1507–1515.
- Ko, J., Park, H., Heo, L., & Seok, C. (2012). GalaxyWEB server for protein structure prediction and refinement. *Nucleic acids research*, 40(W1), W294-W297.
- MacKerell, A. D. (1998, August). Developments in the CHARMM all-atom empirical energy function for biological molecules. In *Abstracts of papers of the American Chemical Society* (216, pp. U696-U696). 1155 16TH ST, NW, WASHINGTON, DC 20036 USA: AMER CHEMICAL SOC.
- Martz, F., Wilczynska, M., & Kleczkowski, L. A. (2002). Oligomerization status, with the monomer as active species, defines catalytic efficiency of UDP-glucose pyrophosphorylase. *Biochemical Journal*, 367(1), 295-300.
- Mu, H. (2002). Screening of Genes Related to Pollen Development in a Thermo-Sensitive Male Sterile Rice (*Oryza sativa* L.). Cloning and Characterization of UDP-Glucose Pyrophosphorylase. PhD dissertation. (Hong Kong, China: University of Hong Kong).
- Pierce, B. G., Wiehe, K., Hwang, H., Kim, B. H., Vreven, T., & Weng, Z. (2014). ZDOCK server: interactive docking prediction of protein–protein complexes and symmetric multimers. *Bioinformatics*, 30(12), 1771-1773.
- Šali, A., & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial

- restraints. *Journal of molecular biology*, 234(3), 779-815.
- Schwender, J., Hebbelmann, I., Heinzl, N., Hildebrandt, T. M., Rogers, A., & Naik, D. (2015). Quantitative multilevel analysis of central metabolism in developing oilseeds of oilseed rape during *in vitro* culture. *Plant Physiol.* 168, 828–848.
- Tian, W., Chen, C., Lei, X., Zhao, J., & Liang, J. (2018). CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic acids research*, 46(W1), W363-W367.
- Zhang, Y., Yang, J., Yan, R., Roy, A., Xu, D., & Poisson, J. (2015). The I-TASSER Suite: protein structure and function prediction. *Nature methods*, 12(1), 7.