Indian Journal of Biochemistry & Biophysics
Vol. 62, January 2025, pp. 25-37
DOI: 10.56042/ijbb.v62i1.11769

Structural analysis of Glucose-6-Phosphate Dehydrogenase (G6PD) variants from Northeast India and natural antioxidants as a pharmacological agent to alleviate G6PD deficiency associated challenges- An *in silico* approach

Noymi Basumatary[1], Nerswn Basumatary[2], Dipankar Baruah[3], Paresh Kumar Sarma[4] & Jatin Sarmah[2]*

[1]Department of Zoology, Basugaon College, Chirang- 783 372, Assam, India

[2]Department of Biotechnology, Bodoland University, Kokrajhar- 783 370, Assam, India

[3]Department of Pathology, Gauhati Medical College and Hospital, Guwahati-781 032, Assam, India

[4]Department of Medicine, Dhubri Medical College and Hospital, Dhubri-783 325, Assam, India

Received 28 June 2024; *revised 25 August 2024*

The present study focuses on the structural deviation of four Glucose-6-phosphate dehydrogenase (G6PD) variants *viz*., Orissa, Kalyan-Kerala, Mahidol and A^+ detected in Northeast Indian population and to assess the probable efficacy of natural antioxidants to combat G6PD deficiency associated challenges. G6PD deficiency caused by mutations in the *g6pd* gene, results in hemolysis of the Red Blood Corpuscles (RBCs) under oxidative stress. Over sixty years have passed since the identification of G6PD-associated enzymopathy; nonetheless, a treatment for the deficiency remains unavailable. Thus, the potential of natural antioxidants as a remedial agent against these variants were evaluated *in silico*. The three dimensional (3D) structures of the variants were modeled and validated. Molecular docking of variants with the natural antioxidants was performed using AutoDock Vina, wherein the binding affinities with G6PD Orissa, Kalyan-Kerala, Mahidol and A^+ ranged between -5.2 to -9.2 kcal/mol, -5.1 to -9.8 kcal/mol, -5.2 to -9.4 kcal/mol and -5.0 to -10.5 kcal/mol, respectively. Drug-likeness and toxicity analyses were done using SwissADME and ProTox-II, respectively. Molecular dynamics simulation of the variants and the variant-antioxidant complexes done using GROMACS showed best fit in Orissa-Myricetin, Kalyan-Kerala-Apigenin, Mahidol-Catechin, and A^+-Diadzen. Kalyan-Kerala-Apigenin was observed to have the least deviation compared to WT G6PD.

Keywords: G6PD variant, Molecular docking, Molecular dynamics simulation, NADPH, RBC

Glucose-6-phosphate dehydrogenase (G6PD) deficiency is the most common genetic enzymopathy which has globally affected over 400 million people. Approximately 5-20% of total G6PD deficiency is found in Asia[1]. The enzyme G6PD plays a key role in the pentose phosphate pathway (PPP) by catalyzing the conversion of Glucose-6-phosphate to 6-phosphogluconolactone, producing an antioxidant nicotinamide adenine dinucleotide phosphate (NADPH) in the process. NADPH provides protection to the Red Blood Cells (RBC) against harmful oxidative radicals[2]. Mutations in the G6pd gene lead to a deficiency of this enzyme in the RBCs, which results in hemolysis of the RBCs under oxidative stress[3].

In order to attain its active state, G6PD should be present either in the dimer or tetramer form. Each monomeric form has two domains- a catalytic $NADP^+$ binding domain and β+α domain, which contains another structural $NADP^+$ binding site. In between these two domains lies the binding site for glucose 6-phosphate (G6P). Most of the class I (<10% of normal G6PD activity with chronic non-spherocytic hemolytic anemia) and class II variants (<10% of normal G6PD activity with acute hemolytic anemia) are mainly located in those functional regions of the enzyme. These mutations affect the activity and stability of the enzyme[4]. Over 450 genotypic variants for G6PD have been described on the basis of biochemical characteristics. However, only about 230 variants have been characterized at the molecular level[5]. In India, molecular characterization of G6PD variants has revealed the presence of around 22 different variants. Limited studies are available from the Northeastern region of India. The tribal population of Northeast India has been shown to have five variants: Mediterranean, Orissa, Kalyan-Kerala, A^+ and Mahidol[6,7].

*Correspondence:
Phone: +91-9707175220 (Mob)
E-mail: jatinsarmahindia@gmail.com

More than 60 years have passed since the discovery of G6PD-associated enzymopathy; yet, a remedy for the deficiency is not available. Some potential treatments for G6PD deficiency include antioxidant therapy, glutathione biosynthesis enhancement through N-acetyl-cysteine (NAC) administration[8], complementary pathway-generated NADPH[9,10], transcriptional regulator-mediated G6PD upregulation[11], small molecules that directly bind to G6PD to correct enzyme dysfunction[4], small molecules that increase G6PD's transcriptional output[4], and gene therapy[12]. Pharmaceutical companies, however, are less interested in developing a medication against G6PD deficiency because the most severe form of the deficiency is uncommon, and even while Class II and III variants are frequent, they only exhibit symptoms when exposed to an oxidative trigger. Furthermore, the prospect of creating a "one-size-fits-all" treatment is improbable, given the multitude of mutations and their varied effects on protein structure[13].

Molecular docking and molecular dynamics (MD) simulation analyses are efficient approaches in evaluation of structural integrity of proteins and understanding protein-ligand affinities[14-18]. Few studies on G6PD focusing on the protein-ligand affinities and structural integrity for G6PD monomers and dimers have been conducted[19,20]. However, such studies on natural antioxidants as a possible therapeutic agent against G6PD variants are lacking. Based on this background, the present study was carried out a) to comprehend the structural deviation of the variants and b) to assess the possibility of natural antioxidants as a remedial agent against four class III variants *viz*., Orissa, Kalyan-Kerala, Mahidol and A^+ detected among the Northeast Indian population.

Materials and Methods

3D structure modeling of G6PD variants

The three dimensional (3D) structures of the four variants were modeled using the amino acid sequence of the wild type (WT) G6PD (PDB ID: 6e08). This was done by I-TASSER (https://zhanggroup.org/I-TASSER/), an online server for protein structure and function predictions[21-23]. The five models generated by I-TASSER were validated by ProTSAV (http://www.scfbio-iitd.res.in/software/proteomics/protsav.jsp), an online protein structure analysis and validation tool[24]. ProTSAV is a meta-server that incorporates different protein structure assessment tools like Procheck[25], ProSA-Web[26], ERRAT[27], Verify3D[28], dDFire[29], Naccess[30], MolProbity[31], D2N[32] ProQ[33] and PSN-QA[34]. The best model was then selected for further analysis.

Literature search and retrieval of naturally available antioxidant compounds

Literature databases such as Pubmed, Pubmed Central, Google Scholar, Researchgate and Semantic Scholar were searched for natural antioxidant compounds using “natural antioxidants”, “antioxidants”, “phytochemicals”, “natural compounds” as keywords. The two dimensional (2D) conformers of these antioxidants were downloaded from Pubchem (https://pubchem.ncbi.nlm.nih.gov/) in *sdf* format. The 2D conformers were then converted to pdb format using OpenBabel 3.1.1[35].

Molecular docking

The ligands and target proteins were prepared for docking using AutoDockTools-1.5.7 (ADT)[36]. Ligand preparation includes addition of hydrogens, calculation of Gasteiger charges, detection of root and setting number of torsions. For preparation of target, the modeled 3D structures were opened in ADT and steps such as deletion of water, addition of polar hydrogens, check and repair (if any) of missing atoms and addition of Kollman charges were performed. The prepared ligands and target structures were saved in *pdbqt* formats. The grid parameters for docking were set such that the interacting residues of binding pocket are covered. Docking was performed using AutoDock Vina, one of the docking engines of the AutoDock Suite. The binding interactions between the antioxidants and the variants were analyzed using ADT.

Analysis of drug-likeness and toxicity of the antioxidant compounds

The drug-likeness of the selected compounds was studied using SwissADME (http://www.swissadme.ch/), an online tool for studying pharmacokinetics, drug-likeness and medicinal affability of small molecules[37]. The drug-likeness property of the compounds was evaluated based on Lipinski’s rule. Toxicity analysis was done using ProTox-II (https://tox-new.charite.de/protox_II/), an online tool for prediction of toxicity of small molecules[38]. Based on the binding energy, drug-likeness and toxicity analysis, the best complex was selected and visualized using LigPlot+[39].

Molecular dynamics simulation

To evaluate the structural deviation of the variants and validate the docking results, we performed MD simulation using GROMACS 2022 package[40]. A simulation system was configured utilizing GPU-enabled GROMACS 2022 package. The force field GROMOS 54a7 was added to the system and the SPCE water system, along with a Cubic box, was used to model the solvent. To neutralize the systems, 1.5mM of NaCl ions were added to the water-filled box. The model system was relaxed prior to the simulation, and the system was then employed for a simulation time set up for 100 ns under standard NTP conditions, that is, a constant number of particles (N), pressure (P), and temperature (T). The simulation interactive diagram tool further analyzed the simulation data in a graphical way, which revealed protein-ligand complex features during the simulation period. Data on Root Mean Square Distance (RMSD), Root Mean Square Fluctuation (RMSF), Radius of Gyration (RG) and Solvent Accessibility Surface Area (SASA) interaction were analyzed.

Results and Discussion

G6PD is a cytoplasmic enzyme that catalyzes the first and the rate-limiting step of PPP, in which G6P is oxidized to 6-phosphogluconolactone and $NADP^+$ is reduced to NADPH. The enzyme protects the cells from oxidative damages caused by reactive oxygen species. Although G6PD is present in all cells, deficiency of G6PD does not affect those cells as they have alternate source for NADPH generation. However, in RBCs, PPP is the only source for NADPH generation, as a result of which the cells undergo hemolysis if they are G6PD deficient. Recently, there is a lot of interest in the cytoprotective properties of natural antioxidants against oxidative stress and the several defense systems involved. Numerous naturally occurring antioxidants have been isolated from several therapeutic plants. These substances' capacity to combat oxidative stress has expanded their application in the treatment of oxidative stress-related illnesses[41]. In G6PD deficiency, increased oxidative stress results from inadequate NADPH synthesis. So far, research on only few antioxidants including Vitamin C, Vitamin E, Astaxanthin, and α-lipoic acid has focused on treating G6PD deficiency[13]. The process of finding a new drug for a disease involves a long period of time and high financial cost. For accurate identification of pharmacological target and prediction of its efficacy, some novel approaches are required. *In silico* drug design offers a cost-effective way of identifying potential new drugs[42]. However, there is a limitation to this approach, since it is performed computationally for a short duration of time which may not be sufficient to understand the molecular dynamics.

Identifying the binding site is essential for effective structure-based molecular docking[43]. Since the 3D structures of the mutants were not available, the structures were modeled by I-TASSER and validated using ProtSAV (Fig. 1). The WT 3D structure of G6PD was used which consist of 515 amino acids and an atom count of 4430. In the present molecular docking study, the binding pockets were identified and binding affinities of the antioxidant compounds with G6PD Orissa, Kalyan-Kerala, Mahidol and A^+ were explored. The binding affinities with the above four variants were found to be ranging between -5.2 to -9.2 kcal/mol, -5.1 to -9.8 kcal/mol, -5.2 to -9.4 kcal/mol and -5.0 to -10.5 kcal/mol respectively (Table 1). Although not all compounds showed good binding affinity, the average binding affinities of the compounds were observed to be -7.5 kcal/mol, -7.15 kcal/mol, -7.07 kcal/mol and -7.05 kcal/mol with Orissa, Kalyan-Kerala, Mahidol and A^+, respectively. Mutations in the *g6pd* gene decrease the monomers' ability to dimerize, which lowers the binding energy needed to produce homodimers[44]. *In vitro* and *in vivo* studies on efficacy of three antioxidant compounds, namely, astaxanthin, vitamin C and vitamin E have been conducted earlier, but were not completely successful in minimizing G6PD related complications[13]. The current molecular docking investigation further supports this result, as the four G6PD variants showed comparatively decreased binding affinities with Astaxanthin, Vitamin C, and Vitamin E. Following docking, drug-likeness and toxicity analyses were performed.

Lipinski's Rule of Five is the most well-known rule to be applied when determining drug-likeness of a compound[45-47]. According to Lipinski's rule, the compounds that have molecular weight ≤500Da, number of H-bond acceptor ≤10, number of H-bond donor ≤5 and logP value ≤5 are considered to have drug-likeness property. A substance is more likely to exhibit poor absorption or permeation when two or more of the above physico-chemical conditions are not met. Eight compounds, Alpha-carotene, Astaxanthin, Beta-Carotene, Cantaxanthin, Chicoric acid, Lutein, Theaflavin and Zeaxanthin showed good binding affinities with the variants in the present study but lacked the drug-likeness features, with two

to three violations of Lipinski's criterion. However, it has been claimed that 16% of oral drugs do not comply with at least one of Lipinski's Rule of five, and 6% do not meet two or more conditions. For example, several popular medications, like Atorvastatin and Monteplaza, contradict one of the Lipinski guidelines[41]. Furthermore, toxicity analysis is an essential stage in the drug-design process that identifies any potential drawbacks that the compounds may have. The toxicity of the compounds is shown by the LD_{50} values (mg/kg body weight), which are the median fatal doses at which 50% of test subjects die after being exposed to a chemical[48]. Compounds are either classed as dangerous or non-toxic based on their LD_{50} values. According to the globally recognized classification, compounds that have LD_{50} value ≤5 are classified as class I and they are fatal if swallowed, LD_{50} values between 5-50 are classified as class II that are also fatal if swallowed. Class III compounds show LD_{50} values between 50-300 and they are toxic if swallowed, class IV compounds have LD_{50} value between 300-2000 and are harmful if swallowed. LD_{50} value between 2000-5000 is classified as class V and may be harmful if swallowed and compounds with LD_{50} value >5000 are classified as class VI that are non-toxic. Out of all antioxidants considered in the study, Catechin, Canthaxanthin, Epicatechin, Epigallocatechin, Gallocatechin and Lycopene were only found to be non-toxic. The drug-likeness and toxicity analysis of the compounds were analyzed using SwissADME and ProTox-II, respectively. SwissADME analyzes the drug likeness

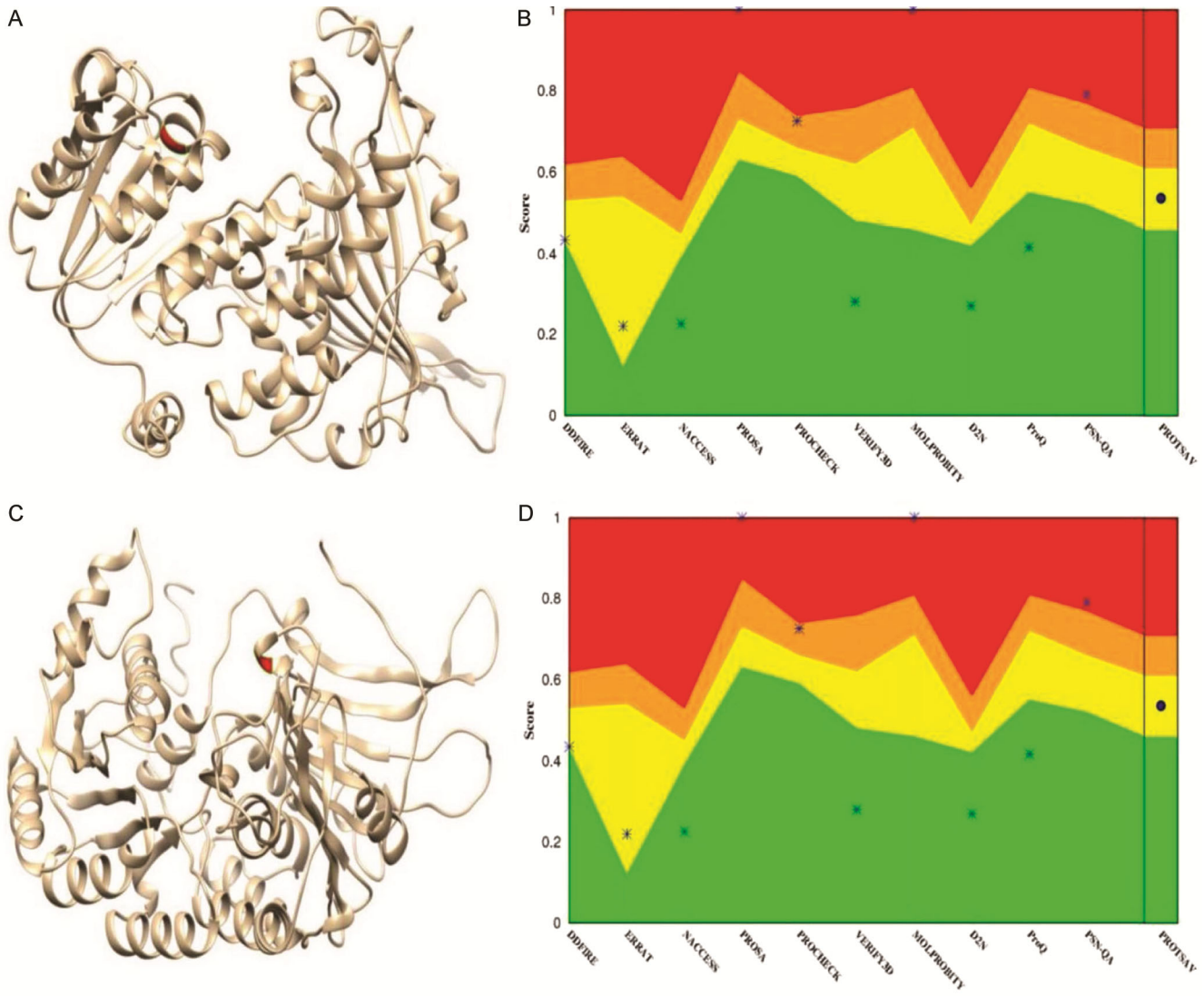

Fig. 1 — 3D structure modeling and validation. (A) G6PD Orissa modeled using I-TASSER; (B) Validation of G6PD Orissa by ProTSAV; (C) G6PD Kalyan-Kerala modeled using I-TASSER; (D) Validation of G6PD Kalyan-Kerala by ProTSAV (*Contd*.)

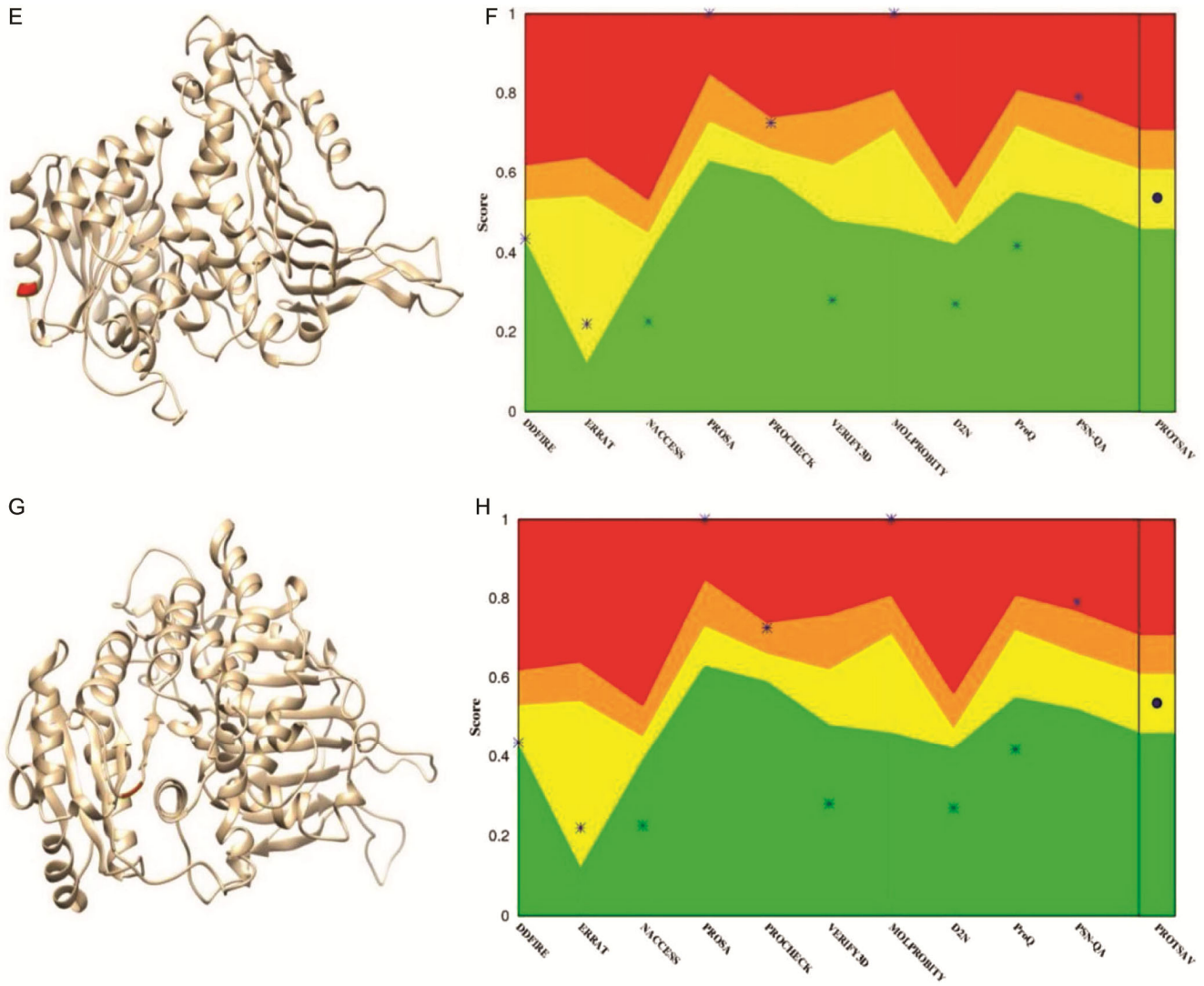

Fig. 1 — 3D structure modeling and validation. (E) G6PD Mahidol modeled using I-TASSER; (F) Validation of G6PD Mahidol by ProTSAV; (G) G6PD A^+ modeled using I-TASSER; and (H) Validation of G6PD A^+ by ProTSAV

properties of small molecules following Lipinski rule of five. The drug-likeness properties of all the compounds are presented in (Table 2). Among 43 compounds, 34 were found to possess drug-likeness properties. The TPSA values of the compounds were also less than 140 $Å^2$, indicating that the compounds have higher permeability through the cell membrane. ProTox-II is a tool for virtual prediction of toxicity of compounds which gives us the LD_{50} values (mg/kg body weight). In the present study, 28 out of 43 compounds were found to be class V and VI compounds, indicating that they may be harmful if swallowed or non-toxic (Table 2).

Based on the binding affinity, number of residues interacting with the protein, drug-likeness property and toxicity analysis, four complexes were selected and visualized using LigPlot+. Since hydrogen and hydrophobic contacts are thought to be crucial to molecular docking, LigPlot+ was used to investigate both kinds of interactions. This hydrogen bond most likely has a major impact on the binding mode selection and docking mechanism[49]. Fig. 2A represents the detailed interaction between the variant Orissa and Myricetin showing binding affinity of -8.0kcal/mol. Myricetin showed hydrogen bonding interaction with Glu239, Arg365, Pro434 and Ala436, respectively, and hydrophobic interactions with His201, Lys205, Glu206, Val394, Gln395, Glu398 and Asp435. Kalyan-Kerala showed a binding affinity of -7.5kcal/mol with Apigenin. Figure 2B depicts the interaction of Kalyan-Kerala with Apigenin, where Apigenin showed hydrogen bonding with Thr243 and

Table 1 — Binding affinities of the antioxidant compounds with the four G6PD variants

Ligand	Binding affinity (kcal/mol)			
	Orissa	Kalyan-Kerala	Mahidol	A^+
Alphacarotene	-8.3	-6.8	-7.6	-8.0
Apigenin	-7.8	-7.5	-7.3	-7.6
Astaxanthin	-9.2	-9.8	-9.4	-10.5
Betacarotene	-8.1	-7.5	-7.9	-8.4
Canthaxanthin	-8.4	-7.7	-7.8	-8.1
Catechin	-7.8	-7.3	-7.1	-7.0
Chicoric acid	-8.3	-7.8	-6.7	-7.2
Chlorogenic acid	-6.9	-6.6	-6.6	-6.6
Cinnamic acid	-5.7	-5.6	-5.2	-6.0
Cyanidin	-7.8	-7.5	-7.2	-6.9
Daidzein	-7.4	-7.1	-7.6	-7.4
Delphinidin	-7.7	-7.7	-7.2	-7.3
Ellagic acid	-7.8	-7.7	-7.4	-7.5
Epicatechin	-7.9	-7.2	-7.3	-7.1
Epigallocatechin	-7.8	-7.0	-7.4	-7.3
Eriodictyol	-8.1	-7.5	-7.4	-7.6
Gallic acid	-5.7	-5.9	-5.7	-5.4
Gallocatechin	-7.8	-7.0	-7.2	-6.9
Genistein	-7.5	-7.0	-7.3	-7.0
Glycitein	-7.6	-6.9	-6.9	-6.9
Hesperetin	-8.1	-7.4	-7.7	-7.4
Isorhamnetin	-7.9	-7.2	-7.3	-7.0
Kaempferol	-7.8	-7.4	-7.2	-7.3
Lutein	-8.2	-8.8	-8.3	-7.6
Luteolin	-8.0	-7.7	-7.6	-7.1
Lycopene	-6.9	-6.6	-6.5	-6.8
Malvidin	-7.5	-6.5	-7.5	-6.8
Melatonin	-6.0	-6.2	-5.9	-6.3
Myricetin	-8.0	-7.4	-7.5	-7.3
Naringenin	-7.7	-7.2	-7.2	-7.3
Pelargonidin	-7.6	-7.2	-7.2	-6.8
Peonidin	-7.8	-7.0	-7.3	-6.8
Petunidin	-7.7	-7.7	-7.4	-7.4
Pterostilbene	-6.6	-6.4	-6.6	-6.2
Quercetin	-8.1	-7.7	-7.5	-7.2
Resveratrol	-5.9	-5.9	-6.1	-6.0
Rosmarinic acid	-7.4	-7.3	-5.9	-6.7
Salicyclic acid	-5.2	-5.1	-4.8	-5.1
Tangeritin	-6.4	-6.9	-6.4	-6.4
Theaflavin	-8.9	-9.1	-8.5	-8.4
Vitamin C	-5.4	-5.1	-5.2	-5.0
Vitamin E	-7.2	-6.6	-6.2	-6.5
Zeaxanthin	-8.0	-7.5	-7.7	-7.2

Glu244, and hydrophobic interactions with Phe241, Arg246, Tyr249, Phe250, Phe253 and Asp258. The interaction of Catechin with Mahidol is shown in (Fig. 2C). Hydrogen bonding interactions of Catechin were observed with His201, Glu206, Pro434 and Tyr437, hydrophobic interactions were observed with Lys47, Lys205, Gln395, Glu398 and Ala436. These interactions were involved in forming the complex with a binding affinity of -7.1kcal/mol. Figure 2D represents the interaction of A^+ with Daidzen. Hydrogen bond was formed with Leu142 and hydrophobic interaction with Leu43, Ala141, Pro144, Glu170, Lys171, Tyr249, Phe250, Phe253 and Asp258.

The stability of both the WT G6PD enzyme and four different mutant variants were assessed by analyzing their RMSD values. These RMSD values were then plotted over time as presented in (Fig. 3A). It was observed that three systems (Orissa, Mahidol and A^+) reached an equilibrium state after 65ns, with the RMSD values fluctuating in the range of approximately 0.35 nm to 0.5 nm. Table 3 represents the Mean ± SD values of RMSD, RMSF, RG and SASA of G6PD variants. The WT and the Orissa variant showed similar state of equilibrium with average RMSD of 0.41 ± 0.06. The A^+ variant had a slightly higher deviation from the WT with an average RMSD of 0.46 ± 0.05. Mahidol variant reached equilibrium at 75ns but further deviated from 80 ns. Kalyan-Kerala variant, on the other hand, showed highest deviation from the WT enzyme with an average RMSD of 0.57 ± 0.05. The fact that mutations located in the functional area of the enzyme impair its functionality and stability may provide evidence for the decreased stability of these variants[50]. The RMSD of G6PD Orissa was found to be comparable to that of the WT protein, suggesting that both proteins have comparable stability.

MD simulations of the docked complexes were also carried out and compared with the WT protein and the respective variants without the ligand in order to confirm or further validate the complexes' stability. Receptor and ligand flexibility is critical for precise drug binding prediction as well as for the prediction of related thermodynamic and kinetic parameters[49]. MD simulation examines the atomic and molecular movements of a ligand-target model system to provide an image of the structural flexibility[51]. The MD production trajectories were superimposed on the initial structures to calculate the RMSD of these structures, which allowed us to evaluate the stability of the ligands and variant C-α atoms in the ligand-variant complexes over time. The stability of the mutants complexed with ligand was also evaluated by analyzing their RMSD values. These RMSD values were then plotted over time in (Fig. 3A). The results showed that A^+ variant complexed with Daidzen

Table 2 — Drug-likeness and toxicity analysis of the compounds

Ligand	MW (Da)	HbA	HbD	logP	TPSA ($Å^2$)	Lipinski violation	Drug likeness	Toxicity	
								LD_{50} (mg/kg)	Toxicity class
Alpha carotene	536.87	0	0	8.96	0	2	No	1510	4
Apigenin	270.24	5	3	0.52	90.9	0	Yes	2500	5
Astaxanthin	596.84	4	2	5.09	74.6	2	No	4600	5
Beta carotene	536.87	0	0	8.96	0	2	No	1510	4
Catechin	290.27	6	5	0.24	110.38	0	Yes	10000	6
Canthaxanthin	564.84	2	0	6.82	34.14	2	No	10000	6
Chicoric acid	474.37	12	6	0.14	208.12	2	No	5000	5
Chlorogenic acid	354.31	9	6	-1.05	164.75	1	Yes	5000	5
Cinnamic acid	148.16	2	1	1.90	37.3	0	Yes	2500	5
Cyanidin	287.24	6	5	0.32	114.29	0	Yes	5000	5
Daidzein	254.24	4	2	1.08	70.67	0	Yes	2430	5
Delphinidin	338.7	7	6	0.03	134.52	1	Yes	5000	5
Ellagic acid	302.19	8	4	0.14	141.34	0	Yes	2991	4
Epicatechin	290.27	6	5	0.24	110.38	0	Yes	10000	6
Epigallocatechin	306.27	7	6	-0.29	130.61	1	Yes	10000	6
Eriodictyol	288.25	6	4	0.16	107.22	0	Yes	2000	4
Gallic acid	170.12	5	4	-0.16	97.99	0	Yes	2000	4
Gallocatechin	306.27	7	6	-0.29	130.61	1	Yes	10000	6
Genistein	270.24	5	3	0.52	90.9	0	Yes	2500	5
Glycitein	284.26	5	2	0.77	79.9	0	Yes	2500	5
Hesperetin	302.28	6	3	0.41	96.22	0	Yes	2000	4
Isorhamnetin	316.26	7	4	-0.31	120.36	0	Yes	5000	5
Kaempferol	286.24	6	4	-0.03	111.13	0	Yes	3919	5
Lutein	568.87	2	2	6.96	40.46	2	No	10	2
Luteolin	286.24	6	4	-0.03	111.13	0	Yes	3919	5
Lycopene	536.87	0	0	9.21	0	2	No	5700	6
Malvidin	331.3	7	4	0.28	112.52	0	Yes	5000	5
Melatonin	232.28	2	2	0.97	54.12	0	Yes	963	4
Myricetin	318.24	8	6	-1.08	151.59	1	Yes	159	3
Naringenin	272.25	5	3	0.71	86.99	0	Yes	2000	4
Pelargonidin	271.24	5	4	0.86	94.06	0	Yes	3919	5
Peonidin	301.27	6	4	0.57	103.29	0	Yes	5000	5
Petunidin	317.27	7	5	0.03	123.52	0	Yes	5000	5
Pterostilbene	256.3	3	1	2.76	38.69	0	Yes	1560	4
Quercetin	302.24	7	5	-0.56	131.36	0	Yes	159	3
Resveratrol	228.24	3	3	2.26	60.69	0	Yes	1560	4
Rosmarinic acid	360.31	8	5	0.90	144.52	0	Yes	5000	5
Salicyclic acid	138.12	3	2	0.99	57.53	0	Yes	1034	4
Tangeritin	372.37	7	0	0.63	76.36	0	Yes	5000	5
Theaflavin	564.49	12	9	-0.79	217.6	3	No	2500	5
Vitamin C	176.12	6	4	-2.60	107.22	0	Yes	3367	5
Vitamin E	430.71	2	1	6.14	29.46	1	Yes	5000	5
Zeaxanthin	568.87	2	2	6.96	40.46	2	No	10	2

could not reach the equilibrium state within the 100ns trajectory. The RMSD of Orissa-Myrecetin complex, Kalyan-Kerala-Apigenin complex and A^+-Catechin entered the equilibrium state from 75ns and showed a consistent state.

The RMSF of WT and the mutant variants fluctuated the most at the structural $NADP^+$ binding site in residues 310-440 with the largest (1 nm) observed in the WT protein and Mahidol variant (Fig. 3B). In the catalytic $NADP^+$ binding domain, fluctuations were observed in residues 70-140 with largest (0.5 nm) observed in Orissa, Kalyan-Kerala and Mahidol. The gyration (RG) plots in Figure 3C showed that the mutant variants have slightly higher RG values compared to the WT protein indicating that the mutations affected their protein

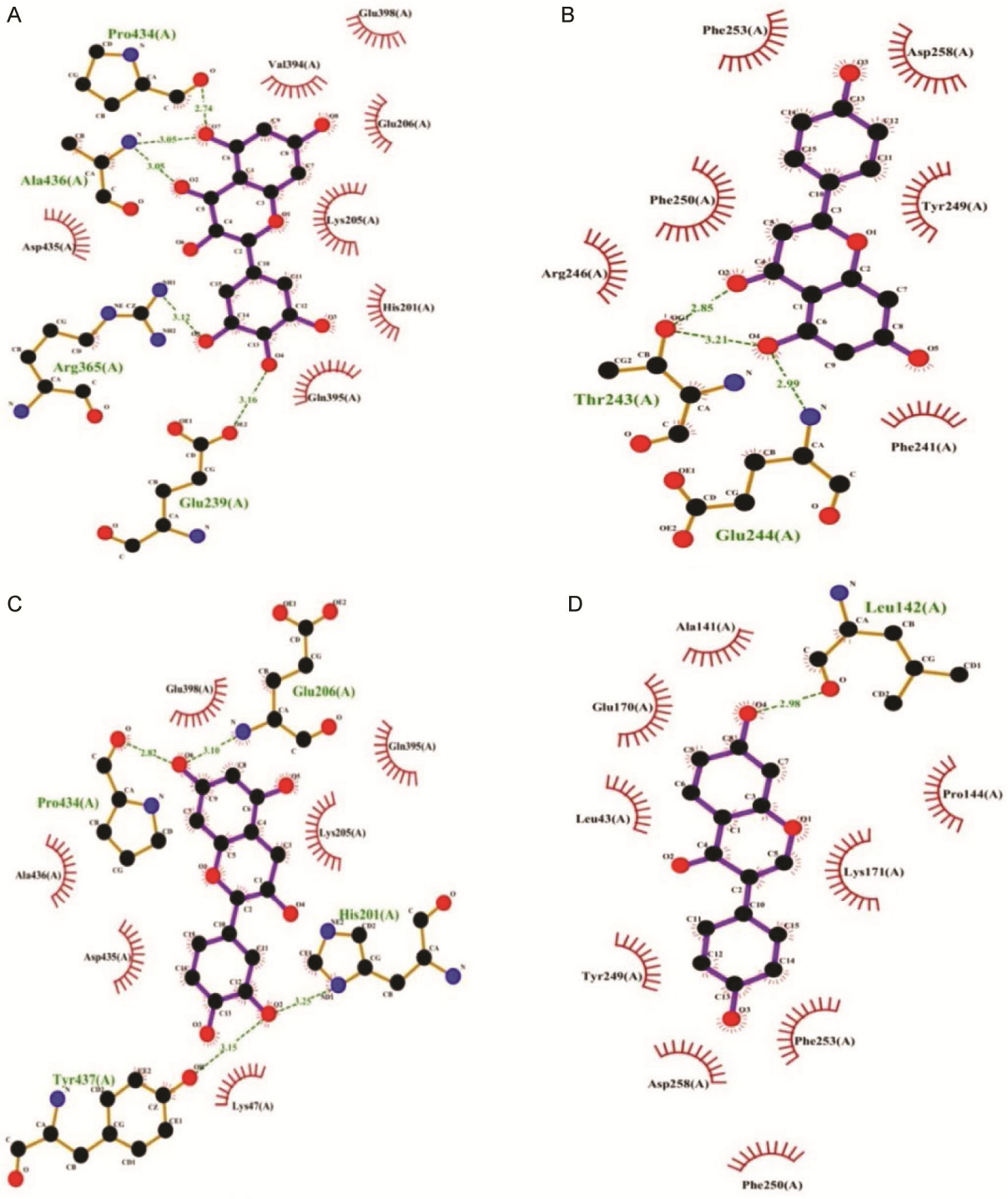

Fig. 2 — Visualization of complexes using LigPlot+. (A) 2D interaction between Orissa and Myricetin; (B) 2D interaction between Kalyan-Kerala and Apigenin; (C) 2D interaction between Mahidol and Catechin; and (D) 2D interaction between A^+ and Daidzen

folding. The Orissa variant had an average RG value of 2.45 nm which is almost similar to that of the WT protein having average RG value 2.44 nm. Figure 3D illustrates the temporal evolution of the WT and the mutants' SASA changes. Throughout the 100 ns trajectory the WT protein has lower SASA value compared to the mutant protein variants indicating that the mutation has affected protein folding of the variants.

Comparison of the MD simulation analysis of the four variants along with their respective complexes

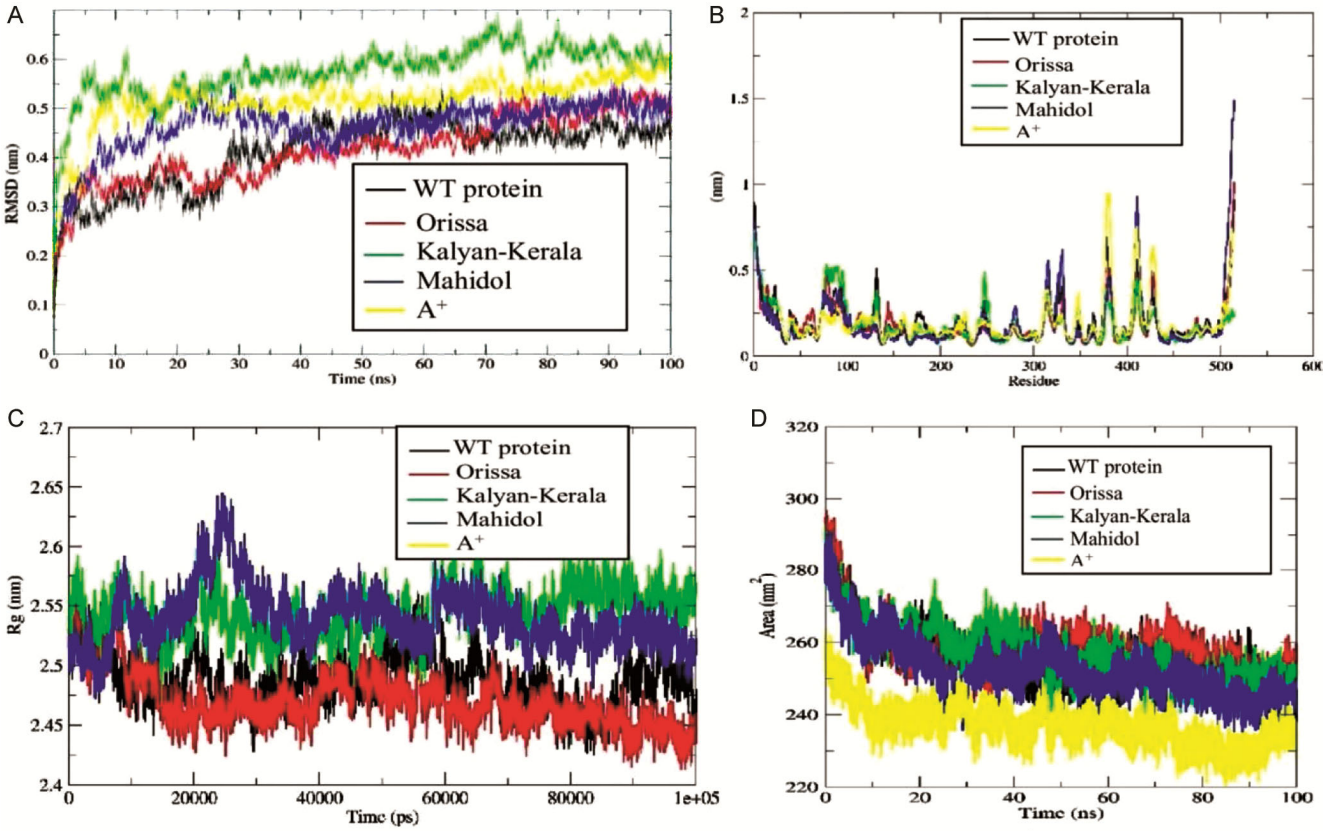

Fig. 3 — Molecular dynamics (MD) simulation of WT type and G6PD variants. (A) RMSD of WT G6PD and Orissa, Kalyan-Kerala, Mahidol and A^+; (B) RMSF of WT G6PD and Orissa, Kalyan-Kerala, Mahidol and A^+; (C) RG of WT G6PD and Orissa, Kalyan-Kerala, Mahidol and A^+; and (D) SASA of WT G6PD and Orissa, Kalyan-Kerala, Mahidol and A^+

Table 3 — Mean ± SD values of RMSD, RMSF, RG and SASA of G6PD variants

Parameter	WT G6PD	Orissa	Kalyan-Kerala	Mahidol	A^+
RMSD (nm)	0.41 ± 0.06	0.41 ± 0.06	0.57 ± 0.05	0.46 ± 0.05	0.51 ± 0.05
RMSF (nm)	0.18 ± 0.12	0.19 ± 0.12	0.17 ± 0.11	0.18 ± 0.10	0.19 ± 0.17
RG (nm)	2.44 ± 0.01	2.45 ± 0.02	2.52 ± 0.02	2.50 ± 0.02	2.47 ± 0.02
SASA (nm^2)	237.31 ± 5.98	255.25 ± 7.89	257.81 ± 7.45	256.78 ± 7.95	252.46 ± 8.34

has been presented in (Figs 4-7). The Mean ± SD values of RMSD, RMSF, RG and SASA of G6PD variants complexed with the ligands are also presented in (Table 4). RMSD of the Orissa-Myricetin complex was increased (Fig. 4A). Least deviations were observed in the Kalyan-Kerala-Apigenin complex (Table 4). Orissa-Myricetin (Fig. 4C) and Kalyan-Kerala-Apigenin (Fig. 5C) showed lower RG values compared to Mahidol-Catechin (Fig. 6C) and A^+- Daidzen (Fig. 7C), displaying compactness of the complex protein structure. The mutant variants complexed with ligands showed similar fluctuations to that of the variants in the structural NADP binding site in residues 310-440 with the largest (1 nm) observed in Kalyan-Kerala-Apigenin complex. In the catalytic $NADP^+$ binding domain, the complexes showed fluctuation in residues 70-100 with the largest (0.5 nm) observed in A^+-Daidzen (Fig. 7B).

Among the four complexes of the study, Kalyan-Kerala-Apigenin was observed to have the least deviation compared to the WT G6PD. The other complexes, namely, Orissa-Myricetin, Mahidol-Catechin and A^+-Daidzen did not much improve the stability of the variants. This could be because only one compound, Myricetin, Catechin, and Daidzen was used for the simulation study for Orissa, Mahidol, and

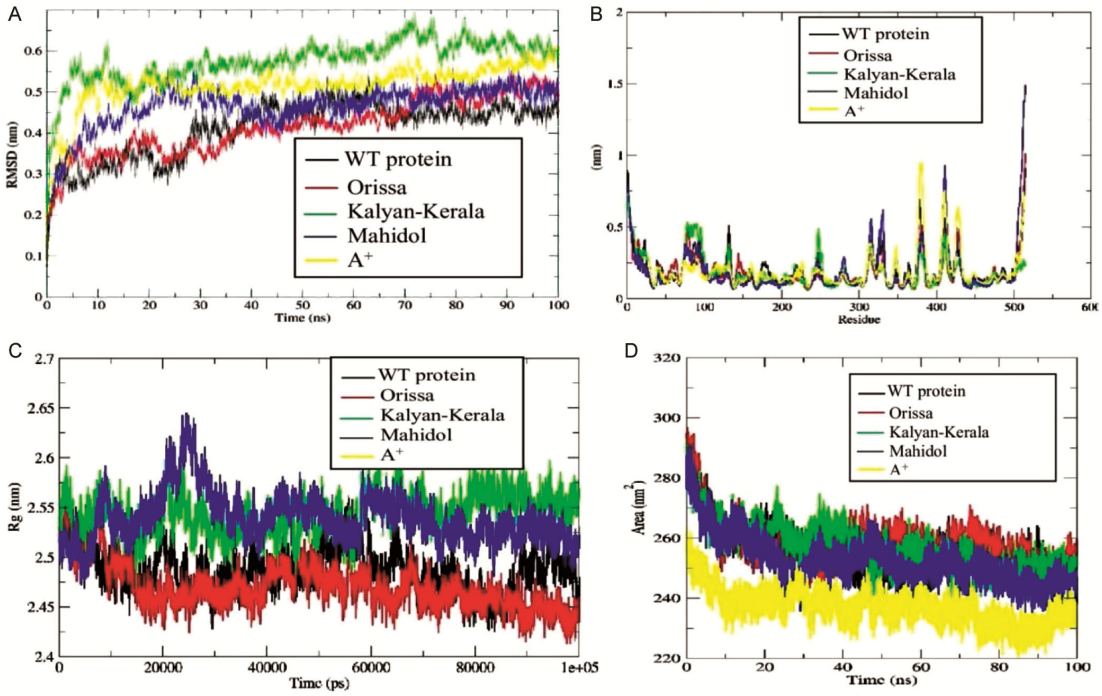

Fig. 4 — Molecular dynamics (MD) simulation of WT type, G6PD Orissa and Orissa-Myricetin. (A) RMSD of WT G6PD, Orissa and Orissa-Myricetin; (B) RMSF of WT G6PD, Orissa and Orissa-Myricetin; (C) RG of WT G6PD, Orissa and Orissa-Myricetin; and (D) SASA of WT G6PD, Orissa and Orissa-Myricetin

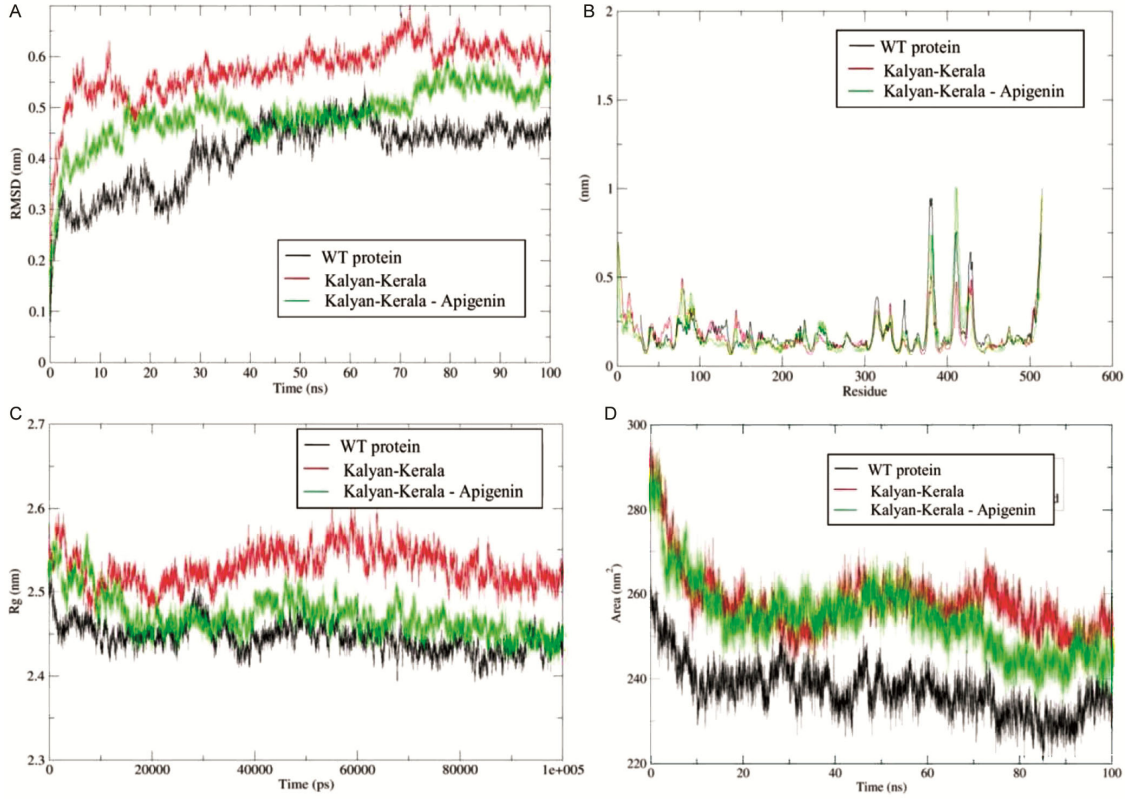

Fig. 5 — Molecular dynamics (MD) simulation of WT type, G6PD Kalyan-Kerala and Kalyan-Kerala-Apigenin. (A) RMSD of WT G6PD, G6PD Kalyan-Kerala and Kalyan-Kerala-Apigenin; (B) RMSF of WT G6PD, G6PD Kalyan-Kerala and Kalyan-Kerala-Apigenin; (C) RG of WT G6PD, G6PD Kalyan-Kerala and Kalyan-Kerala-Apigenin; and (D) SASA of WT G6PD, G6PD Kalyan-Kerala and Kalyan-Kerala-Apigenin

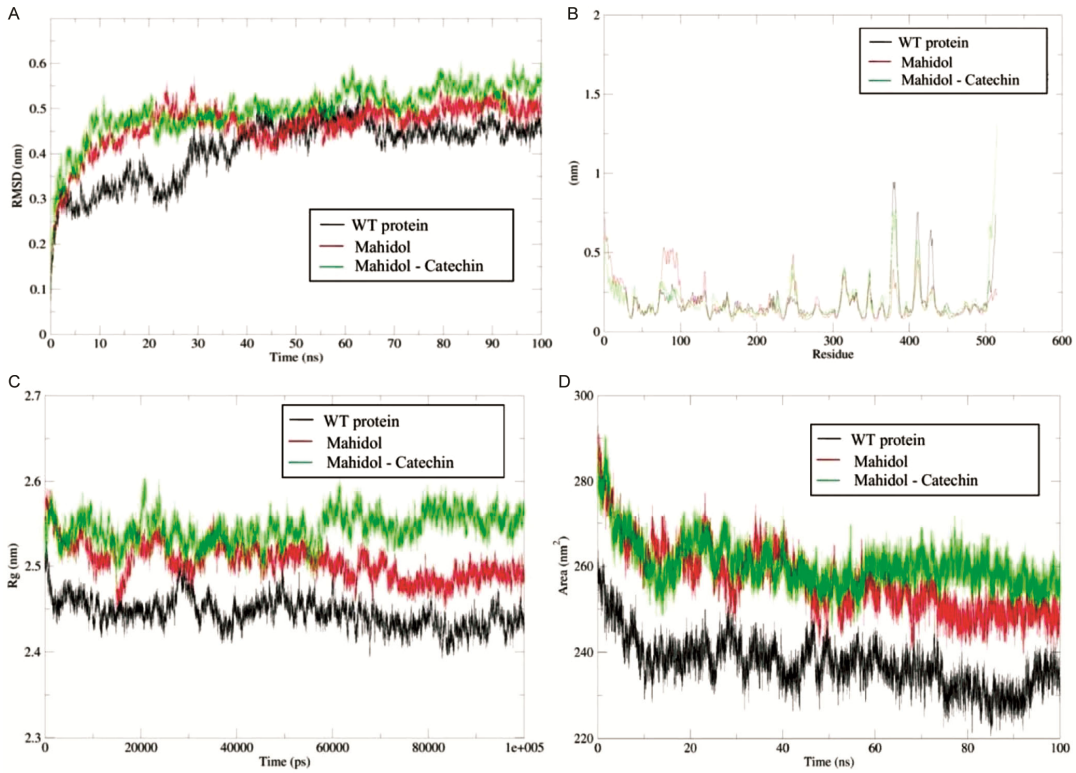

Fig. 6 — Molecular dynamics (MD) simulation of WT type, G6PD Mahidol and Mahidol-Catechin. (A) RMSD of WT G6PD, G6PD Mahidol and Mahidol-Catechin; (B) RMSF of WT G6PD, G6PD Mahidol and Mahidol-Catechin; (C) RG of WT G6PD, G6PD Mahidol and Mahidol-Catechin; and (D) SASA of WT G6PD, G6PD Mahidol and Mahidol-Catechin

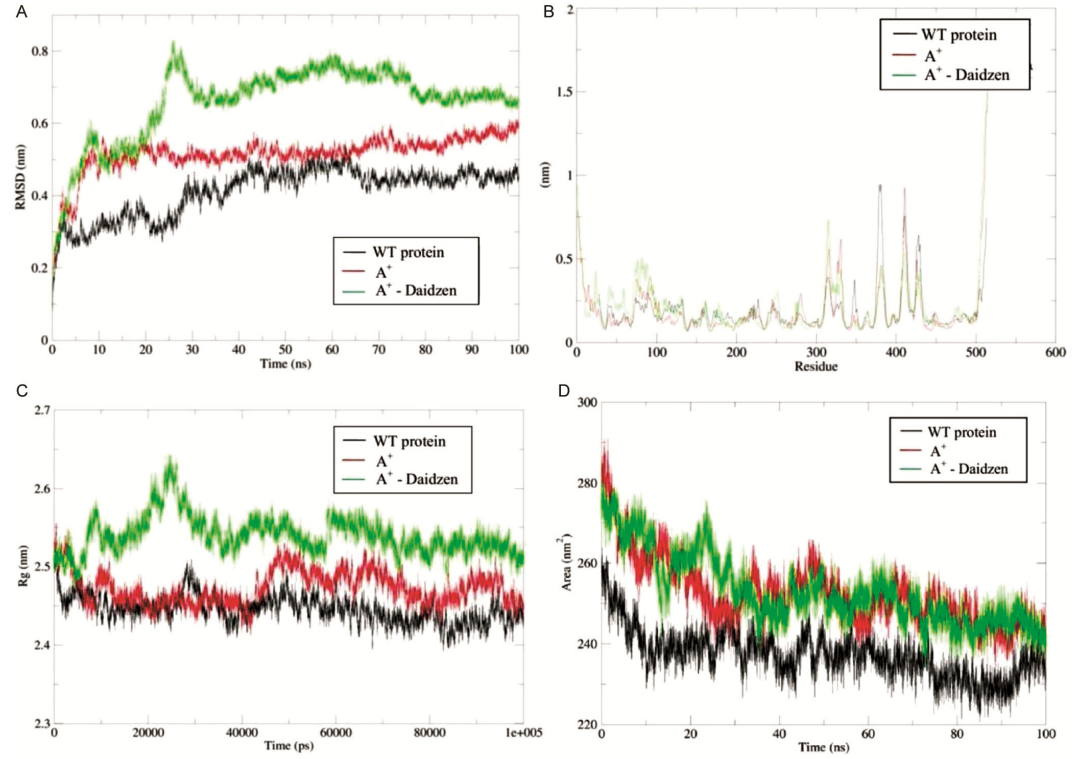

Fig. 7 — Molecular dynamics (MD) simulation of WT type, G6PD A^+ and A^+-Diadzen. (A) RMSD of WT G6PD, A^+ and A^+-Diadzen; (B) RMSF of WT G6PD, A^+ and A^+-Diadzen; (C) RG of WT G6PD, A^+ and A^+-Diadzen; and (D) SASA of WT G6PD, A^+ and A^+-Diadzen

Table 4 — Mean ± SD values of RMSD, RMSF, RG and SASA of G6PD variant complexed with the ligand.

Parameter	Orissa - Myricetin	Kalyan-Kerala -Apigenin	Mahidol - Catechin	A^+ - Diadzen
RMSD (nm)	0.51 ± 0.08	0.48 ± 0.05	0.49 ± 0.05	0.65 ± 0.10
RMSF (nm)	0.19 ± 0.13	0.17 ± 0.12	0.19 ± 0.14	0.21 ± 0.18
RG (nm)	2.48 ± 0.02	2.47 ± 0.02	2.54 ± 0.01	2.53 ± 0.02
SASA (nm^2)	256.50 ± 8.56	254.29 ± 8.65	260.37 ± 5.74	253.09 ± 8.84

A^+, respectively. A more thorough understanding of the antioxidants' effectiveness in these variations might have been obtained through MD simulation of a greater number of complexes containing ligands that also violate Lipinski's rule with one or two violations. Moreover, experimental validation using *in vitro* or *in vivo* approach is required to comprehend the efficacy of the selected antioxidants. These aspects form the future scopes of the study.

Acknowledgement

The authors gratefully acknowledge the Lady Tata Memorial Trust, Mumbai, India for providing Senior scholarship to Noymi Basumatary from August, 2020 to July, 2023. The authors are also grateful to the DBT-Govt. of India sponsored Bioinformatics Infrastructure Facility at Bodoland University, Kokrajhar, Assam for providing necessary facilities.

Conflict of interest

All authors declare no conflict of interest.

References

1 Howes RE, Piel FB, Patil AP, Nyangiri OA, Gething PW, Dewi M, Hogg MM, Battle KE, Padilla CD, Baird JK & Hay SI, G6PD deficiency prevalence and estimates of affected populations in malaria endemic countries: a geostatistical model-based map. *PLoS Med*, 9 (2015) e1001339.

2 Kuhn V, Diederich L, Keller TCS, Kramer, CM, Luckstadt W, Panknin C, Suvorava T, Isakson BE, Kelm M & Cortese-Krott MM, Red blood cell function and dysfunction: Redox regulation, nitric oxide metabolism, anemia. *Antioxid Redox Signal*, 26 (2017) 718.

3 Luzzatto L, Ally M & Notaro R, Glucose-6-Phosphate dehydrogenase deficiency. *Blood*, 136 (2020) 1225.

4 Hwang S, Mruk K, Rahighi S, Raub AG, Chen CH, Dorn LE, Horikoshi N, Wakatsuki S, Chen JK & Mochly-Rosen D, Correcting Glucose-6-phosphate dehydrogenase deficiency with a small-molecule activator. *Nat Commun*, 9 (2018) 4045.

5 Luzzatto L, Nannelli C & Notaro R, Glucose-6-phosphate dehydrogenase deficiency. *Hematol Oncol Clin North Am*, 30 (2016) 373.

6 Basumatary N, Baruah D, Sarma PK, Wary KK & Sarmah J, The first report of three Glucose-6-phosphate dehydrogenase (G6PD) variants: Mediterranean, Orissa and Kalyan-Kerala from Northeast India. *Indian J Hematol Blood Transfus*, 40 (2023) 175.

7 Basumatary N, Baruah D, Sarma PK & Sarmah J, Identification of Glucose-6-phosphate dehydrogenase variants by utilizing polymerase chain reaction – Restriction fragment length polymorphism based method. *Gene Rep*, 35 (2024) 101911.

8 Tang HY, Ho HY, Wu PR, Chen SH, Kuypers FA, Cheng ML & Chiu DT, Inability to maintain GSH pool in G6PD-deficient red cells causes futile AMPK activation and irreversible metabolic disturbance. *Antioxid Redox Signal*, 20 (2015) 744.

9 Liu L, Shah S, Fan J, Park JO, Wellen KE & Rabinowitz JD, Malic enzyme tracers reveal hypoxia-induced switch in adipocyte NADPH pathway usage. *Nat Chem Biol*, 12 (2016) 345.

10 Herrero-Mendez A, Almeida A, Fernandez E, Maestre C, Moncada S & Bolanos JP, The bioenergetic and antioxidant status of neurons is controlled by continuous degradation of a key glycolytic enzyme by APC/C-Cdh1. *Nat Cell Biol*, 11 (2009) 747.

11 Zhang L, Chen Y, Jiang Q, Song W & Zhang L, Therapeutic potential of selective histone deacetylase 3 inhibition. *Eur J Med Chem*, 162 (2019) 534.

12 Ludwig LS, Khajuria RK & Sankaran VG, Emerging cellular and gene therapies for congenital anemias. *Am J Med Genet C Semin Med Genet*, 172 (2016) 332.

13 Garcia AA, Koperniku A, Ferreira JCB & Mochly-Rosen D, Treatment strategies for Glucose-6- phosphate dehydrogenase deficiency: past and future perspectives. *Trends Pharmacol Sci*, 42 (2021) 829.

14 Aier I, Varadwaj P & Raj U, Structural insights into conformational stability of both wild-type and mutant EZH2 receptor. *Sci Rep*, 6 (2016) 34984.

15 Daghestani M, Purohit R, Daghestani M, Daghistani M & Warsy A, Molecular dynamic (MD) studies on Gln233Arg (rs1137101) polymorphism of leptin receptor gene and associated variations in the anthropometric and metabolic profiles of Saudi women. *PLoS One*, 14 (2019) e0211381.

16 Cakmak S & Erdogan T, Some bis (3-(4-nitrophenyl) acrylamide derivatives: Synthesis, characterization, DFT, antioxidant, antimicrobial properties, molecular docking and molecular dynamics simulation studies. *Indian J Biochem Biophys*, 60 (2023) 209.

17 Matilda JJ, Reji TA, Design, structural characterization, biological evaluation and molecular docking studies of methylindole bearing thiocarbamoylpyrazole moieties. *Indian J Biochem Biophys*, 61 (2024) 418.

18 Karamarathodi N, Binukumar SM, Das S, Sundararajan S, Karunakaran K & Muniyan R, Computational screening, docking and simulation analysis of phytochemicals from Senna auriculata against multiple targets of *Mycobacterium tuberculosis*. *Indian J Biochem Biophys*, 61 (2024) 688.

19 Doss CG, Alasmar DR, Bux RI, Sneha P, Bakhsh FD, Al-Azwani I, Bekay RE & Zayed H, Genetic epidemiology

of Glucose-6-phosphate dehydrogenase deficiency in the Arab world. *Sci Rep*, 6 (2016) 37284.

20 Nguyen H, Nguyen TTK & Le L, Computational study of Glucose-6-phosphate-dehydrogenase deficiencies using molecular dynamics simulation. *South Asian Journal of Life Sciences (SAJLS)*, 4 (2016) 32.

21 Zhang Y, I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, 9 (2008) 40.

22 Roy A, Kucukural A & Zhang Y, I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc*, 5 (2010) 725.

23 Yang J, Yan R, Roy A, Xu D, Poisson J & Zhang Y, The I-TASSER Suite: protein structure and function prediction. *Nat Methods*, 12 (2015) 7.

24 Singh A, Kaushik R, Mishra A, Shanker A & Jayaram B, ProTSAV: A protein tertiary structure analysis and validation server. *Biochim Biophys Acta*, 1864 (2016) 11.

25 Laskowski RA, MacArthur MW, Moss DS & Thornton JM, PROCHECK - a program to check the stereochemical quality of protein structures. *J App Cryst,* 26 (1993) 283.

26 Wiederstein M & Sippl MJ, ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res*, 35 (2007) 407.

27 Colovos C & Yeates TO, Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci*, 2 (1993) 1511.

28 Luthy R, Bowie JU & Eisenberg D, Assessment of protein models with three-dimensional profiles. *Nature*, 356 (1992) 83.

29 Yang Y & Zhou Y, Specific interactions for *ab initio* folding of protein terminal regions with secondary structures. *Proteins*, 72 (2008) 793.

30 Lee B & Richards FM, The interpretation of protein structures: estimation of static accessibility. *J Mol Biol*, 55 (1971) 379.

31 Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB 3rd, Snoeyink J, Richardson JS & Richardson DC, MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res*, 35 (2007) 375.

32 Mishra A, Rana PS, Mittal A & Jayaram B, D2N: Distance to the native. *Biochim Biophys Acta*, 1844 (2014) 1798.

33 Wallner B & Elofsson A, Can correct protein models be identified? *Protein Sci*, 12 (2003) 1073.

34 Ghosh S & Vishveshwara S, Ranking the quality of protein structure models using sidechain based network properties. *F1000Res*, 3 (2014) 17.

35 O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T & Hutchison GR, Open Babel: An open chemical toolbox. *J Cheminform*, 3 (2011) 33.

36 Trott O & Olson AJ, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*, 31 (2010) 455.

37 Daina A, Michielin O & Zoete V, SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci Rep*, 7 (2017) 42717.

38 Banerjee P, Eckert AO, Schrey AK & Preissner R, ProTox-II: a webserver for the prediction of toxicity of chemicals. *Nucleic Acids Res*, 46 (2018) W257.

39 Laskowski RA & Swindells MB, LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J Chem Inf Model*, 51 (2011) 2778.

40 Bekker H, Berendsen HJC, Dijkstra EJ, Achterop S, van Drunen R, van der Spoel D, Sijbers A & Keegstra H, Gromacs, A parallel computer for molecular dynamics simulations. In: *Physics Computing '92*, (Ed. by de Groot RA & Nadrchal J; World Scientific, Singapore), 1993, 252.

41 Mehta J, Rayalam S & Wang X, Cytoprotective effects of natural compounds against oxidative stress. *Antioxidants*, 7 (2018) 147.

42 Kiewhuo K, Jamir E, Priyadarsinee L, Nagamani S & Sastry GN, Screening of phytochemicals for potential breast cancer targets BRCA1 and BARD1: A network pharmacology approach. *Indian J Biochem Biophys*, 60 (2023) 393.

43 Ferreira LG, Dos Santos RN, Oliva G & Andricopulo AD, Molecular docking and structure-based drug design strategies. *Molecules*, 20 (2015) 13384.

44 Wijk RV, Huizinga EG, Prins I, Kors A, Rijksen G, Bierings M & Solinge WWV, Distinct phenotypic expression of two de novo missense mutations affecting the dimer interface of Glucose-6-phosphate dehydrogenase. *Blood Cells Mol Dis*, 32 (2004) 112.

45 Lipinski CA, Lombardo F, Dominy BW & Feeney PJ, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*, 46 (2001) 3.

46 Katturajan R, Medha T, Karra S, Vidya R & Prince SE, A comparative computational approach on the most deleterious missense variant in Connexin 43 protein and its potent inhibitor analysis. *Indian J Biochem Biophys*, 60 (2023) 7.

47 Balakrishnan M, Bhavana V, Sharma T, Soam SK, Supriya P, Vyshnavi M & Rao SCH, Homology modeling and molecular docking studies of INF1 protein of *Phytophthora infestans* in Potato. *Indian J Biochem Biophys*, 61 (2024) 335.

48 Sinnarkar S, Kumbhar GM, Vaibhav L, Bhawalkar J & Mahajan J, Evaluating the anti-cancer activity of myricetin in the management of oral cancer using in silico analysis. *Indian J Biochem Biophys*, 61 (2024) 659.

49 Fikrika H, Ambarsari L & Sumaryada T, Molecular docking studies of Catechin and its derivatives as anti-bacterial inhibitor for Glucosamine-6-Phosphate Synthase. *IOP Conf Ser Earth Environ Sci*, 31 (2016) 012009.

50 Cunningham AD, Colavin A, Huang KC & Mochly-Rosen D, Coupling between protein stability and catalytic activity determines pathogenicity of G6PD variants. *Cell Rep*, 18 (2017) 2592.

51 Fischer M, Coleman RG, Fraser JS & Shoichet BK, Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery. *Nat Chem*, 6 (2014) 575.