Communications in Physics, Vol. 36, No. 3 (2026), pp. 333-342
DOI: https://doi.org/10.15625/0868-3166/24093

The structural stability of monomeric glucose-6-phosphate dehydrogenase and its relationship to G6PD deficiency

Nhung T. T. Nguyen[1,2,†], **Le Hoang Phong**[2], **Nguyen Thi Hai Yen**[2], **Phuong Thuy Bui**[3,4], **Mai Thi Lan**[5] **and Trinh Xuan Hoang**[2]

[1]*Graduate University of Science and Technology, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet, Nghia Do, Hanoi 11307, Vietnam*

[2]*Institute of Physics, Vietnam Academy of Science and Technology, 10 Dao Tan, Giang Vo, Hanoi 11108, Vietnam*

[3]*Institute of Theoretical and Applied Research, Duy Tan University, Hanoi, 100000, Vietnam*

[4]*Faculty of Pharmacy, Duy Tan University, Da Nang, 550000, Vietnam*

[5]*Faculty of Engineering Physics, Hanoi University of Science and Technology, 1 Dai Co Viet Road, Hanoi, Vietnam*

E-mail: [†]ntnhung@iop.vast.vn

Received 20 January 2026
Accepted for publication 13 March 2026
Published 15 September 2026

Abstract. *G6PD deficiency, an enzymopathy associated with glucose-6-phosphate dehydrogenase (G6PD), is caused by genetic mutations that reduce the enzyme activity, resulting in hemolytic anemia. Biochemical studies have shown that G6PD deficiency in various variants is associated with reduced thermal stability, reduced catalytic activity, or both. In this study, we investigate the structural stability of the wild-type G6PD monomer at a physiological temperature using molecular dynamics simulations, in the absence and in the presence of its G6P and $NADP^+$ ligands. We find that the G6P ligand has a low affinity for the G6PD monomer, which may result from fluctuations in the size of its binding pocket. This finding helps explain the catalytic inactivity of monomeric G6PD. The binding of $NADP^+$ ligands is found to enhance the structural stability of the protein and the G6P-binding pocket. Our analysis also shows that, with a statistically significant confidence, class I mutations occur preferentially at residues that are more flexible than randomly selected residues, whereas class II mutations tend to occur at residues that are less flexible than randomly selected ones. This result reflects the distinct structural roles of the G6PD mutation sites and further supports the relationship between enzyme activity and thermal stability.*

Keywords: molecular dynamics; protein stability; enzyme activity; mutations.
Classification numbers: 87.15.M-; 87.64.K-.

1. Introduction

Glucose-6-phosphate dehydrogenase (G6PD) is an enzyme in the pentose phosphate pathway (PPP) [1–3]. It catalyzes the chemical reaction that reduces nicotinamide adenine dinucleotide phosphate ($NADP^+$) to NADPH, while oxidizing glucose-6-phosphate (G6P). G6PD protects cells against oxidative damage by maintaining NADPH, which is essential for reducing reactive oxygen species through the production of reduced glutathione. Since PPP is the main source of NADPH in red blood cells, reduced G6PD activity, as found in individuals with G6PD deficiency, can cause acute hemolysis under oxidative stress triggered by infections, certain drugs, or foods such as fava beans. Loss of G6PD activity can also result in chronic hemolytic anemia, neonatal jaundice, and kernicterus.

G6PD deficiency [1–3] is an inherited X-linked disorder affecting approximately 7% of the global population. To date, more than 230 variants associated with G6PD deficiency have been identified [3], 86% of which correspond to single missense mutations that cause a single amino acid substitution. Depending on the enzyme activity level and the associated clinical manifestations, these variants are classified by the World Health Organization (WHO) into four classes [4]. Among these classes, class I represents the most severe deficiency and is associated with chronic nonspherocytic hemolytic anemia. Classes II and III are associated with milder symptoms while class IV is asymptomatic. In both class I and class II mutants, the G6PD activity drops to less than 10% of that of the wild-type enzyme.

Biochemical studies have shown that pathogenic G6PD mutations may reduce the enzyme's thermostability, catalytic activity, or both, all of which can lead to a loss of enzymatic activity [1,3]. Some variants also exhibit decreased foldability upon refolding compared to the wild-type enzyme [5]. It is understood that both misfolding, if it occurs *in vivo*, and reduced thermostability make the mutants more prone to aggregation and more susceptible to degradation by protease. Because red blood cells lack a nucleus and the translation machinery, any deficit of active enzymes cannot be compensated, leading to reduced overall enzymatic activity. Although reduced catalytic activity is widely observed, loss of stability prevails in G6PD mutants [3]. A study of biochemical data using principal component analysis indicates that the clinical phenotypes of G6PD variants can be determined by considering a trade-off between protein stability and catalytic activity [6].

Human G6PD is a large protein with a length of 515 amino acids, existing in monomer, dimer or tetramer forms. The fraction of cellular G6PD that exists in each of these oligomer forms within cells is still unknown. The monomeric form of G6PD has no catalytic activity [3] and the enzyme functions primarily as a dimer or tetramer. The tetramer is somewhat more stable and exhibits higher catalytic activity than the dimer [7]. Each protein subunit in the G6PD oligomers can bind up to three ligands: a G6P substrate, a catalytic $NADP^+$ (coenzyme), and a structural $NADP^+$ (Fig. 1). The structural $NADP^+$ does not participate in catalysis, however, its binding is important for both the enzyme's stability and catalysis [5, 8–10].

The molecular mechanism through which the G6PD mutations impair enzyme functionality remains the subject of ongoing research. Wang and Engel [5] demonstrated that class I mutations near the structural $NADP^+$ impair structural $NADP^+$ binding, which affects the enzyme's folding and stability. Garcia *et al.* [11] showed that loss of structural $NADP^+$-binding in class I variants disrupts tetramerization and G6PD binding. Horikoshi *et al.* [12] showed that class I mutations around the structural $NADP^+$-binding site and the dimer interface cause structural deformations

at the interface, that induce long-range defects in the active site through a network of interactions. Hwang *et al.* [13] demonstrated that the class II Canton mutant (R459L) is less thermostable and more susceptible to protease degradation compared to the wild-type enzyme. Furthermore, the Canton mutant loses essential inter-helical interactions, which disrupt the positioning of G6P and $NADP^+$ in their binding pockets, thereby hindering catalysis. The critical role of oligomerization in G6PD activity is underscored by the fact that class I mutations preferentially cluster at the dimer interface and the structural $NADP^+$-binding site, while both class I and class II mutations are found at the tetramer interface [6, 14].

Given the importance of protein oligomerization and thermostability in G6PD deficiency, this study examines the structural stability of monomeric wild-type G6PD in the presence and absence of ligands. This approach may help elucidate why the enzyme lacks catalytic activity in its monomeric form. We further investigate whether pathogenic mutation sites differ from randomly selected residues in terms of structural flexibility, thereby providing insight into the structural roles of mutated residues.

2. Methods

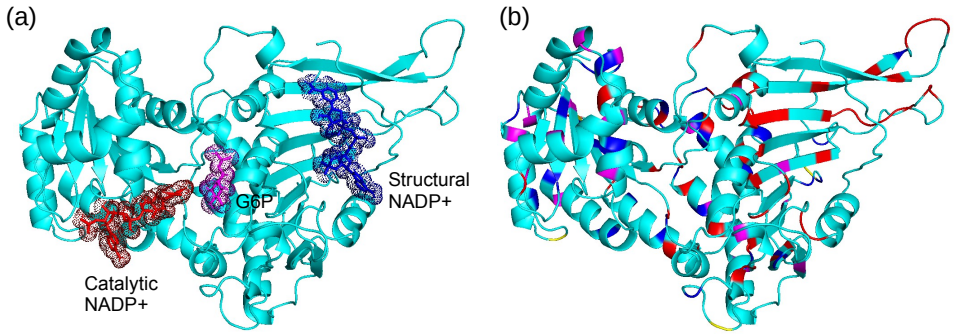

Fig. 1. (Color online) (a) The native structure of the wild-type G6PD monomer (PDB ID: 6e08) [13], comprising residues 28-513, is shown with three bound ligands: the catalytic $NADP^+$ (red), G6P (magenta), and the structural $NADP^+$ (blue). (b) The same native structure is shown with 112 mutation sites highlighted in color, including 53 class I mutation sites (red), 37 class II mutation sites (blue), 18 class III mutation sites (magenta) and 4 class IV mutation sites (yellow).

Based on the crystal structure of wild-type G6PD (PDB ID: 6e08) [13], we constructed two simulation systems: a monomeric protein without ligands and a monomeric protein with all three ligands bound (Fig. 1a). The simulated protein comprises residues 28-513, as present in the crystal structure. The positions of the G6P and catalytic $NADP^+$ ligands were obtained by alignment with two other crystallized G6PD structures (PDB IDs: 2bh9 and 2bhl) [15]. The systems were simulated in a cubic box with a 1 nm distance between the protein and the box faces. After adding water molecules and ions to achieve a NaCl concentration of 154 mM, each system consisted of approximately 128,000 atoms. The CHARMM36m force field was used for the protein together with the TIP3P water model, and parameters for G6P and $NADP^+$ were obtained

using CGenFF [16,17]. First, the systems were relaxed by energy minimization to eliminate steric clashes, followed by equilibration for 1 ns in the canonical (NVT) ensemble using the V-rescale thermostat at 310 K and for 1 ns in the isothermal–isobaric (NPT) ensemble using the C-rescale barostat at 1 bar. Long-range electrostatic interactions were treated using the particle mesh Ewald (PME) method with a cutoff of 1.2 nm. The cutoff for van der Waals (vdW) interactions was also set to 1.2 nm, and the integration time step was 2 fs. MD simulations were carried out with the GROMACS 2025.3 package [18]. For each system, five independent simulations of 400 ns each were generated. All the simulations were performed at the physiological temperature of 310 K.

3. Results and discussion

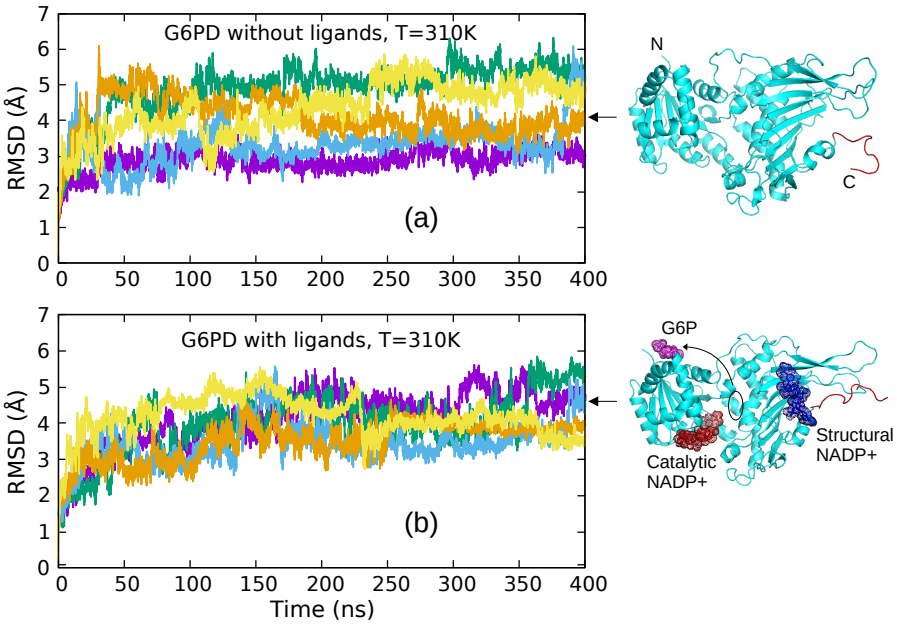

Fig. 2. (Color online) Time dependence of the root mean square deviation (RMSD) from the native state in five independent MD runs at $T = 310$ K for the wild-type G6PD monomer without ligands (a) and with all three ligands (b). The conformations on the right of the panels are the final snapshots obtained at the time $t = 400$ ns in two of the trajectories, with a fluctuating C-terminal tail (residues 501-513) shown in red color. The lower snapshot shows that the G6P ligand has been displaced from its initial binding site (encircled).

Figure 2 shows significant variations, from ∼3 to 6 Å, in the root mean square deviation (RMSD) of the wild-type G6PD monomer from its native structure, both within individual trajectories and among different trajectories obtained from the simulations, in both the absence (Fig. 2a) and presence (Fig. 2b) of ligands. These variations arise not only from the fluctuating C-terminal tail (shown in red color in the final snap shots of the protein conformations in Fig. 2), but also from various loops connected to the C-terminal β-sheet. Notably, the RMSD variations between

trajectories are somewhat smaller in the presence of ligands, indicating a stabilizing effect of the ligands on the protein structure. This observation is consistent with previous studies showing that structural $NADP^+$ binding enhances the enzyme thermostability [5,8,9], even though it is shown here at the level of the monomeric protein. Our simulations show that both the structural and the catalytic $NADP^+$ ligands essentially remain in their binding pockets throughout all trajectories, indicating a significant binding affinity. In contrast, the G6P ligand is quickly displaced from its binding site and can eventually dissociate from the monomer (Fig. 2b, snapshot of the protein conformation). The median escape time of G6P from its binding pocket is found to be about 10 ns from our simulations (see Fig. 4). This result helps explain why monomeric G6PD lacks enzymatic activity, as reported in biochemical studies [3]. Note that the weak binding of G6P is also indicated by a number of crystal structures of G6PD in which G6P is absent. The only available crystal structure that contains G6P is that of the G6PD dimer (PDB ID: 2bhl) [15], in which $NADP^+$ molecules are absent. Thus, our finding that G6P binds weakly, or does not bind significantly, to the G6PD monomer is consistent with these experimental observations.

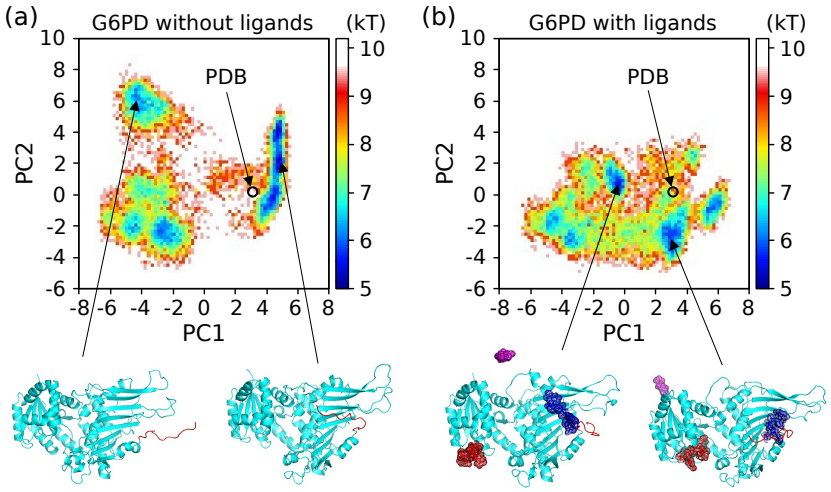

Fig. 3. (Color online) Free energy landscapes of G6PD projected onto the first two principal components (PC1 and PC2) in absence (a) and presence (b) of ligands. The landscapes are constructed from MD trajectories shown in Fig. 2, excluding the first 150 ns of the simulations, and represented as two-dimensional free energy surfaces (in units of kT, color scale). The principal components are calculated from combined data of the apo and holo simulations, therefore PC1 and PC2 are identical in the two panels. The position of the experimental PDB structure is indicated by an open circle. Representative conformations corresponding to major free energy minima are displayed below each landscape.

We have carried out a principal component analysis (PCA) for the G6PD monomer using combined data from the trajectories shown in Fig. 2. The free energy landscapes of G6PD projected onto the first two principal components (Fig. 3) reveal marked differences between the apo and holo states. In the absence of ligands (Fig. 3a), the landscape is more rugged and characterized by multiple well-separated low-energy basins. In contrast, binding of $NADP^+$ reshapes the landscape, resulting in a more compact and less fragmented free energy surface (Fig. 3b). The

experimental PDB structure lies not exactly at, but close to, the deepest free energy minimum in both cases. Overall, ligand binding reduces conformational diversity and smooths the free energy surface, supporting a mechanism in which $NADP^+$-binding stabilizes the enzyme structure [5,8,9] by suppressing large-scale structural fluctuations.

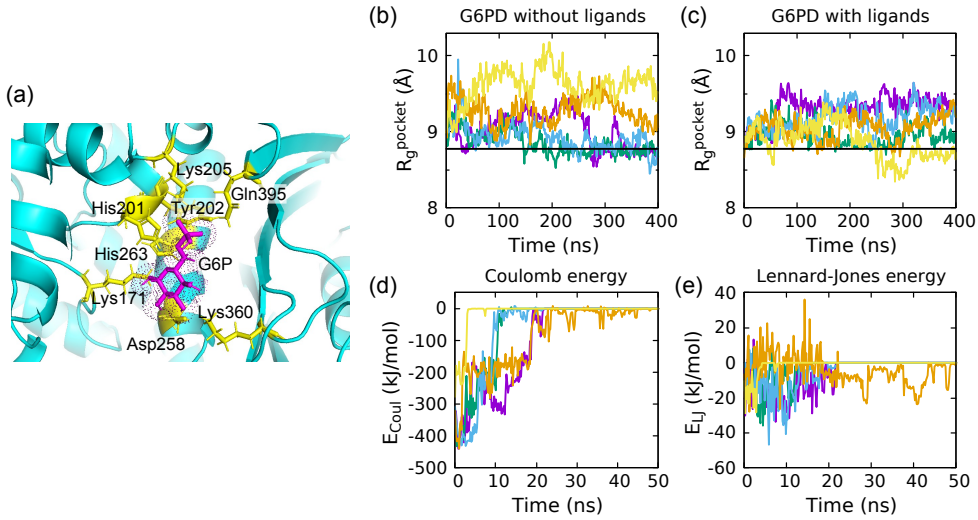

Fig. 4. (Color online) (a) Close-up view of the G6P-binding pocket in G6PD comprised by residues having significant contact with the ligand: His201, Tyr202, Lys205, Lys171, His263, Asp258, Lys360, and Gln395 (yellow). (b and c) Time dependences of the gyration radius, R_g^{pocket}, of the G6P-binding pocket in the trajectories shown in Fig. 2 for the simulations without ligands (b) and with ligands (c). The PDB value of R_g^{pocket} is indicated by black horizontal line. (d and e) Time dependences of the Coulomb (d) and Lennard-Jones (e) interaction energies between the G6P ligand and the binding pocket during the first 50 ns of the trajectories shown in (c).

In order to understand why G6P has a low affinity for the G6PD monomer, we took a closer look at its binding pocket in the initially energy-minimized structure of the protein-ligands complex. We found that, in addition to previously noted residues—His201, Tyr202, Lys205, Lys171, His263, Lys360, Gln395 [10,12]—having significant contact with G6P, G6P also forms two hydrogen bonds with Asp258 (Fig. 4a). Figs. 4b and 4c show that the binding pocket size, as indicated by the its gyration radius R_g^{pocket}, strongly fluctuates and often increases above its PDB value. These significant size deviations weaken the ligand-pocket interactions are likely to inhibit G6P binding, although in the bound state the ligand forms substantial Coulombic and vdW interactions with the binding pocket (Figs. 4d and 4e). Fig. 4c shows that binding of $NADP^+$ somewhat stabilizes the pocket size compared to that shown in Fig. 4b. However, this stabilization is insufficient to maintain G6P-binding. Figs. 4d and 4e indicate that the G6P ligand leaves the binding pocket within 20 ns, as the electrostatic and vdW energies become negligible. It is expected that, in the G6PD dimer, the G6P binding pocket is more stable due to the enhanced overall structural stability conferred by dimerization.

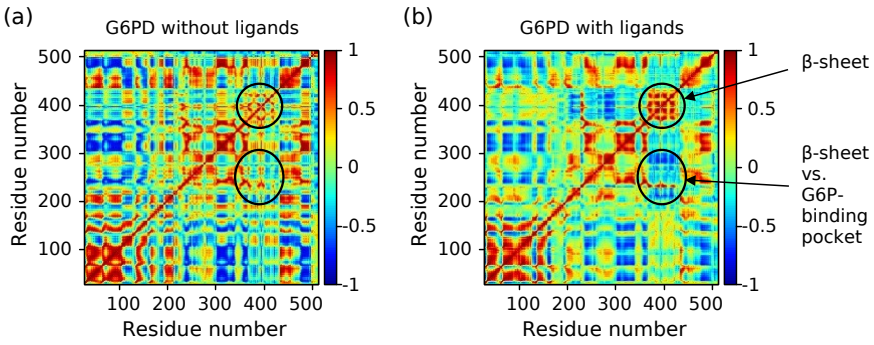

Fig. 5. (Color online) Dynamic cross-correlation maps (DCCMs) of G6PD in absence (a) and presence (b) of ligands. The DCCMs were calculated from the covariance matrix obtained by the principle component analysis. The correlation values range from -1 to 1, as shown by the color scale. Marked differences between the two maps indicated by encircled regions correspond to correlations within the C-terminal β-sheet and those between this β-sheet and the G6P-binding pocket.

The allosteric effect of the structural $NADP^+$ ligand on the G6P-binding pocket can be observed in the dynamic cross-correlation maps (DCCMs), which display the correlated motions between pairs of residues. Fig. 5 shows that the binding of this ligand to the C-terminal β-sheet enhances the correlations within this β-sheet (around residue 400) and also those between this β-sheet and a region related to the G6P-binding pocket (between residues 200-300). These enhanced correlations can be associated with improved structural stability, consistent with various other studies on the allosteric role of structural $NADP^+$ [5, 11–13].

Given that the structural stability of G6PD is crucial for its enzymatic activity, it is interesting to examine whether the flexibility of a residue in the protein structure is related to the likelihood that this residue corresponds to a mutation associated with G6PD deficiency. For doing this, we calculated the root mean square fluctuations (RMSF) of amino acid residues in the G6PD monomer from the simulation data in the absence and in the presence of ligands. Fig. 6 shows that class I mutations tend to occur at the residues with high RMSF values, notably at the RMSF peaks in C-terminal part of the protein. In contrast, class II mutations are more likely to occur at residues with low or minimal RMSF values. Class III and class IV mutation sites are few in number and do not exhibit any clear tendency.

To assess the statistical significance of the above observation, we compare the mean RMSF value of the residues at mutation sites with those of randomly selected sites. Fig. 7 shows the distributions of mean RMSF values for sets of 53, 37 and 18 randomly selected sites, along with the corresponding values (indicated by arrows in the figure) for the sets of class I, class II and class III mutation sites. Note that the number of sites in each random set is equal to the number of mutation sites in the corresponding class of G6PD deficiency, ensuring unbiased comparison. Furthermore, the random sites were selected in the range of residues 28–500, thereby excluding the flexible C-terminal tail, where no mutations are found. Figs. 7(a–d) clearly demonstrate that the mean RMSF values of class I and class II mutation sites lie at opposite tails of the corresponding distributions, indicating their statistical significance. In particular, we find that, for class I mutation

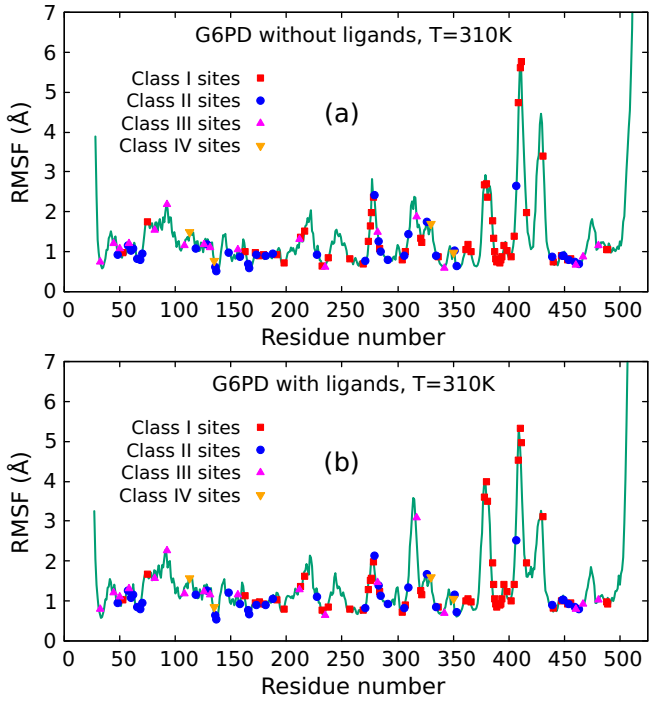

Fig. 6. (Color online) Root mean square fluctuations (RMSF) of amino acid residues in the monomeric G6PD without ligands (a) and with ligands (b). The RMSF values were calculated for the C_α atoms over the last 200 ns of the trajectories shown in Figs. 1, and averaged across the trajectories. The RMSF values of 112 mutation sites are indicated by data points: class I (squares), class II (circles), class III (upward triangles) and class IV (downward triangles).

sites, the mean RMSF value is higher than that of more than 98% of the random site sets for the G6PD monomer without ligands. In contrast, for class II mutation sites, the mean RMSF value is lower than that of 99% of the random site sets. Similar results are observed for the G6PD monomer in the presence of ligands. Figs. 7(e and f) show that the mean RMSF value of class III mutation sites is indistinguishable from that of typical random sites, lying close to the peak of the distribution.

The high RMSF values of class I mutation sites are related to the fact that class I mutations cluster around the structural $NADP^+$-binding site and the dimer interface [6, 14], which contain loops that are more flexible than other parts of the enzyme. Class I mutations are known to impair both structural $NADP^+$-binding and oligomerization [5, 11], leading to decreased protein stability and catalysis. The low RMSF values of class II mutation sites are more intriguing. Class II mutations are mainly located at the tetramer interface and within the interior of the protein's N-terminal domain [6]. Residues at the tetramer interface are important for dimer–dimer interactions and therefore contribute to tetramer stability. Residues in the protein interior typically exhibit lower RMSF values and are more likely to be critical for monomer stability. Thus, the extremely

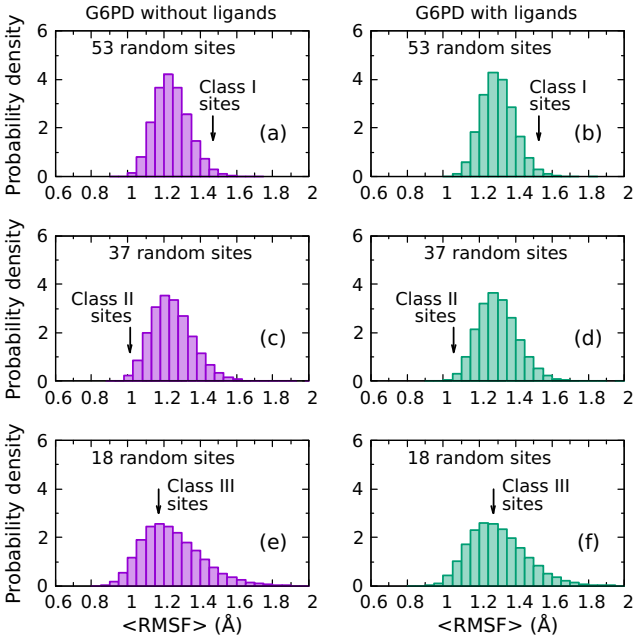

Fig. 7. (Color online) Distributions of the mean RMSF values (boxes) for various sets of random sites taken from the RMSF data of G6PD without ligands (a,c,e) and with ligands (b,d,f). The numbers of random sites in the selected sets are 53 (top panels), 37 (middle panels) and 18 (bottom panels), corresponding to the numbers of sites in the class I, class II and class III mutations, respectively. The mean RMSF values for the mutation sites in each class are indicated by arrows.

high and low RMSF values observed for class I and class II mutations, respectively, are consistent with their structural roles.

4. Conclusion

We have simulated the wild-type G6PD monomer with and without ligands. The results show that the protein exhibits limited structural stability, both in the presence and in the absence of ligands, as indicated by a flexible C-terminal tail and fluctuations in the loops connected to the C-terminal β-sheet. It is found that the structural and catalytic $NADP^+$ ligands essentially remain in their binding pockets, while G6P becomes displaced and detached from its binding site during the simulations. This result helps explain why monomeric G6PD has no catalytic activity, as reported by biochemical studies. We show that the weak binding of G6P may be associated with significant fluctuations in the size of its binding pocket, as measured by the gyration radius of the binding pocket residues. Our analysis suggests that, to some extent, the binding of structural $NADP^+$ allosterically enhances the stability of the G6P-binding pocket. Interestingly, we found a statistical link between residue flexibility, as measured by RMSF, and the locations of class I and class II mutations. This relationship reflects distinct structural roles of the mutation sites and clearly differentiates them from randomly selected residues. Our study demonstrates that, even at

the monomer level, an intrinsic relationship between protein stability and G6PD deficiency can be observed.

Acknowledgments

This work was funded by Vietnam Academy of Science and Technology under grant number KHCBVL.01/24-25. The simulations were conducted on the HPC cluster at IOP-VAST. The authors thank Prof. Mai Suan Li for discussions.

References

[1] P. J. Mason, J. M. Bautista and F. Gilsanz, *G6pd deficiency: the genotype-phenotype association*, Blood Rev. **21** (2007) 267.
[2] M. D. Cappellini and G. Fiorelli, *Glucose-6-phosphate dehydrogenase deficiency*, Lancet **371** (2008) 64.
[3] L. Luzzatto, M. Ally and R. Notaro, *Glucose-6-phosphate dehydrogenase deficiency*, Blood **136** (2020) 1225.
[4] WHO working group, *Glucose-6-phosphate dehydrogenase deficiency*, Bull. World Health Organ. **67** (1989) 601.
[5] X.-T. W. Wang and P. C. Engel, *Clinical mutants of human glucose 6-phosphate dehydrogenase: Impairment of $NADP^+$ binding affects both folding and stability*, Biochim. Biophys. Acta **1792** (2009) 804.
[6] A. D. Cunningham, A. Colavin, K. C. Huang and D. Mochly-Rosen, *Coupling between protein stability and catalytic activity determines pathogenicity of g6pd variants*, Cell Rep. **18** (2017) 2592.
[7] A. T. Ranzani and A. T. Cordeiro, *Mutations in the tetramer interface of human glucose-6-phosphate dehydrogenase reveals kinetic differences between oligomeric states*, FEBS Lett. **591** (2017) 1278.
[8] S. Gómez-Manzo, J. Terrón-Hernández, I. De la Mora-De la Mora, A. Gonzalez-Valdez, J. Marcial-Quino, I. Garcia-Torres *et al.*, *The stability of g6pd is affected by mutations with different clinical phenotypes*, Int. J. Mol. Sci. **15** (2014) 21179.
[9] S. Gómez-Manzo, J. Marcial-Quino, D. Ortega-Cuellar, H. Serrano-Posada, A. González-Valdez, A. Vanoye-Carlo *et al.*, *Functional and biochemical analysis of glucose-6-phosphate dehydrogenase (g6pd) variants: elucidating the molecular basis of g6pd deficiency*, Catalysts **7** (2017) 135.
[10] X. Wei, K. Kixmoeller, E. Baltrusaitis, X. Yang and R. Marmorstein, *Allosteric role of a structural $NADP^+$ molecule in glucose-6-phosphate dehydrogenase activity*, Proc. Natl. Acad. Sci. USA **119** (2022) e2119695119.
[11] A. A. Garcia, I. I. Mathews, N. Horikoshi, T. Matsui, M. Kaur, S. Wakatsuki *et al.*, *Stabilization of glucose-6-phosphate dehydrogenase oligomers enhances catalytic activity and stability of clinical variants*, J. Bio. Chem. **298** (2022) 101610.
[12] N. Horikoshi, S. Hwang, C. Gati, T. Matsui, C. Castillo-Orellana, A. G. Raub *et al.*, *Long-range structural defects by pathogenic mutations in most severe glucose-6-phosphate dehydrogenase deficiency*, Proc. Natl. Acad. Sci. USA **118** (2021) e2022790118.
[13] S. Hwang, K. Mruk, S. Rahighi, A. G. Raub, C.-H. Chen, L. E. Dorn *et al.*, *Correcting glucose-6-phosphate dehydrogenase deficiency with a small-molecule activator*, Nat. Commun. **9** (2018) 4045.
[14] A. D. Cunningham and D. Mochly-Rosen, *Structural analysis of clinically relevant pathogenic g6pd variants reveals the importance of tetramerization for g6pd activity*, Matters **2017** (2017) 10.19185.
[15] M. Kotaka, S. Gover, L. Vandeputte-Rutten, S. W. Au, V. M. Lam and M. J. Adams, *Structural studies of glucose-6-phosphate and $NADP^+$ binding to human glucose-6-phosphate dehydrogenase*, Acta Cryst. D **61** (2005) 495.
[16] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. De Groot *et al.*, *Charmm36m: an improved force field for folded and intrinsically disordered proteins*, Nat. Methods **14** (2017) 71.
[17] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim *et al.*, *Charmm general force field: A force field for drug-like molecules compatible with the charmm all-atom additive biological force fields*, J. Comp. Chem. **31** (2010) 671.
[18] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess *et al.*, *Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers*, SoftwareX **1** (2015) 19.