ABSTRACT

Glucose-6-Phosphate Dehydrogenase (G6PD) is a cytoplasmic enzyme involved in Pentose Phosphate Pathway. It is encoded by the gene *g6pd* located on the X chromosome's distal long arm (locus q28). With 13 exons and 12 introns, this enzyme has a total of 515 amino acids. Mutations in the *g6pd* gene lead to reduced enzyme activity, causing G6PD deficiency. It is a hereditary condition that frequently manifests in areas of the world where malaria is endemic. People with G6PD deficiency often have no symptoms, unless they are exposed to specific triggers that induce haemolysis. As of now, there hasn't been any treatment developed to address the challenges linked to G6PD deficiency. The World Health Organization (WHO) has classified G6PD deficiency into five classes based on enzyme activity: Class I and II display <10% of enzyme activity with chronic non-spherocytic haemolytic anaemia and acute haemolytic anaemia, Class III has 10-60%, Class IV between 60-150%, and Class V has >150%.

Over 400 million people worldwide or roughly 4.9% of the world's population are affected by G6PD deficiency. The Sub-Saharan Africa holds the record of highest prevalence of G6PD deficiency, followed by Arabian subcontinent, Central and Southeast Asia, Mediterranean, Europe and Latin American countries. In India, the deficiency prevalence ranges between 0-30% across various population groups, with notably higher rates observed among Scheduled Tribes and Scheduled Caste. Amidst all variants, Mediterranean, Orissa and Kalyan-Kerala stood out as the most prevalent variants observed in G6PD deficiency cases in India. In the Northeastern region of India, the Angami Nagas (27.0%) from Nagaland displayed the highest frequency, followed by the Rabha (15.8%) and Mikir (15.6%) populations from Assam. To add, G6PD Mahidol, Acores, A^{+}, A^{-202} are the recorded variants reported from Northeast Indian population. Thus, the present work was conducted to study the prevalence of G6PD deficiency and its correlation with haemoglobinopathies (if any), within the tribal population of Assam, while identifying specific mutations prevalent in the region. Further, the naturally available antioxidants were investigated as potential therapeutic solution to alleviate the challenges linked with G6PD deficiency. To accomplish the work, it was planned in four different phases-

i. Prevalence of G6PD deficiency and haemoglobinopathies.

ii. Haematological studies.

iii. Variant analysis using Polymerase chain reaction – Restriction Fragment Length Polymorphism (PCR-RFLP) based method.

iv. *In-silico* study using molecular docking and molecular dynamics simulation.

Random sampling was conducted among 2310 tribal adolescents from the districts of Kokrajhar, Chirang, Baksa and Udalguri in Bodoland Territorial Region, Assam to assess both G6PD status and the percentage of haemoglobin (Hb). Blood samples, approximately, 2ml each, were collected from the individuals showing abnormalities in values of G6PD, Hb% or both. Haematological parameters were clinically evaluated using a Hematology Analyzer, while High-performance liquid chromatography and electrophoretic methods were utilized for Hb-typing for identification of abnormal Hb variant. PCR-RFLP was done to identify the G6PD mutations using restriction endonucleases in the exonic regions of *g6pd*. Furthermore, molecular docking of naturally available antioxidants, using AutoDock Vina was performed with the modeled three-dimensional (3D) structures of the detected variants. The drug-likeness and toxicity analysis of the ligands were performed and the best complex for each of the detected variants were selected for molecular dynamics (MD) simulation using GROMACS.

The study reported that haemoglobinopathies were present in 42.7% of the individuals studied, and that the incidence of G6PD deficiency was 6.2%. Solely one instance showed both G6PD deficiency and HbE. Additionally, two uncommon occurrences were noted: the simultaneous presence of HbS and HbE, along with a notably high percentage of foetal Hb (HbF). The Red blood cell parameters, viz., Hb%, Red Blood Cell count and Mean Corpuscular Haemoglobin Concentration showed significant positive co-relation with G6PD. PCR-RFLP analysis showed the presence of five mutations in the study population, viz., Orissa, Mediterranean, Kalyan-Kerala, Mahidol and A^+.G6PD Mediterranean is a class II variant (less than 10% enzyme activity) and MD simulation studies are conducted by earlier researchers; consequently, we omitted this particular variant from subsequent *in-silico* analysis. Following assessments of binding affinity, drug-likeness and toxicity, we chose four ligands, assigning one to each of the four variants: Orissa-Myricetin, Kalyan-Kerala-Apigenin, Mahidol-Catechin and A^+-Diadzen. The MD simulation study indicated a marginal decrease in stability for three G6PD variants viz., Kalyan-Kerala, Mahidol and A^+, in comparison to the wild type (WT) G6PD.

Whereas, Orissa showed stability similar to that of the WT protein. The MD simulation analysis of all four complexes showed that G6PD Kalyan-Kerala-Apigenin exhibited higher stability in contrast to the mutant variant.

The significant findings of the study were:

i. High prevalence of haemoglobinopathies was observed.
ii. A case exhibiting both G6PD deficiency and HbE co-occurrence was identified.
iii. Two cases of compound heterozygosity involving HbS and HbE trait were identified.
iv. High HbF was observed in 86 cases of HbE and HbS disease.
v. One class II G6PD variant (Mediterranean) and four class III G6PD variants (Orissa, Kalyan-Kerala, A^+ and Mahidol) were identified.
vi. G6PD Orissa was exclusively identified among the Proto-Australoids, while G6PD Mahidol was prevalent among Mongoloids. Conversely, other variants were found in both Proto-Australoids and Mongoloids.
vii. MD simulation revealed G6PD Kalyan-Kerala-Apigenin to have increased stability in comparison to the mutant variant.

The present study had certain limitations attributed to restricted financial resources and time constraints. *Firstly*, the study was limited to the tribal population in four districts, hence, the reported prevalence of deficiencies relates only to this group of people. *Secondly*, PCR-RFLP was performed solely for the commonly reported mutations from India and neighbouring countries. *Thirdly*, numerous antioxidants were found to have good binding affinities, drug-likeness and toxicity results, but only one complex underwent MD simulation.