MD simulations of G6PD and potential risks using chloroquine/hydroxychloroquine based treatments

Natasha Sanabria[1], Özlem Tastan Bishop[2]

1 Department of Toxicology and Biochemistry, National Institute for Occupational Health, National Health Laboratory Service (NatashaS@nioh.ac.za)
2 Director of Research Unit in Bioinformatics (RUBi), Department of Biochemistry and Microbiology, Rhodes University

Abstract: Human diseases associated with the Glucose-6-phosphate dehydrogenase (G6PD) enzyme coincide with certain geographical distributions (e.g. malaria areas). Chloroquine, hydroxychloroquine (CQ/HCQ) and other aminoquinolines have pharmacogenomic associations with G6PD, where these antimalarial drugs cause hemolysis after administration. In most cases, the diseases are due to reduced enzyme activity. The enzyme deficiency is caused by changes in the protein, originating from genetic mutations. Some suggest that mutations causing non-spherocytic haemolytic anaemia are clustered near the C-terminal, whereas mutations causing milder forms are located at the N-terminal of the sequence. However, it is unknown exactly how the genetic sequence variation influences the protein amino acid sequence, the 3D structure or even the stability of the altered protein. In this study, the aim was to use *in silico* structural bioinformatics tools, i.e. Molecular Dynamic (MD) simulations to investigate the effects of these mutations, i.e. the substrate and ligand interactions were analysed to determine how substitutions at codon 163 play a role in the clinical manifestations of G6PD deficiency (i.e. between Class I and Class III respectively). Data was retrieved from the various databases for each variant, i.e. Plymouth (Gly163Asp) and Mahidol mutation (Gly163Ser). The predictions confirmed that the mutations were disease causing (deleterious), resulting from a single point mutation in the nucleotide sequence (missense), which leads to a change in the residue amino acid (nonsynonymous). Thereafter, MD simulations were performed for 100ns and included the G6P substrate, as well as both co-enzyme and structural $NADP^+$ ligands, where previous studies have used one or the other. The trajectories were analysed and focussed on the specific changes (i.e. RMSD, RG, RMSF, H-bonds). In addition, the broader analyses of the overall changes caused by the mutations were also investigated (e.g. Ligplots, distance pairs). The communication between the residues within the structure of the protein was investigated via network analyses, e.g. MD-TASK, where the residue interaction network (RIN) result was visualised by contact maps. The significant RIN contact map findings support the idea that changes at residues Ser29/Asp30, Gly131/Ser132 and/or Arg136 may influence the enzyme activity between Class I and Class III. This is important to assess potential clinical applications, where previously identified clinically important residues arose from studies performed at half the simulation time and with missing substrate or co-enzyme / structural ligands. The results obtained in this study are important because CQ/HCQ have been proposed as potential treatments for COVID-19, where physicians should be alerted to a possible correlation between infection and countries with high prevalence of G6PD deficiency. The findings recently reported by da Rocha *et al*. showed significant G6PD sequence variation in sub-Saharan Africa, with large allele frequency differences between sub-populations in South Africa. They recommended that G6PD gene variations could affect risk of adverse effects of

CQ/HCQ based treatments and be a significant interaction factor for COVID-19 clinical trials in Africans.