

# SUBSTRATE AND COFACTOR BINDING INTERACTION STUDIES OF GALACTITOL -1-PHOSPHATE 5- DEHYDROGENASE FROM PEPTOCLOSTRIDIUM DIFFICILE

## Article history

Received

28 February 2016

Received in revised form

2 April 2016

Accepted

15 May 2016

Siti Aisyah Razali<sup>a\*</sup>, Puteri Sarah Diana<sup>a</sup>, Mohd. Shahir Shamsir<sup>a</sup>, Nor Muhammad Mahadi<sup>b</sup>, Rosli Mohd Illias<sup>c</sup>

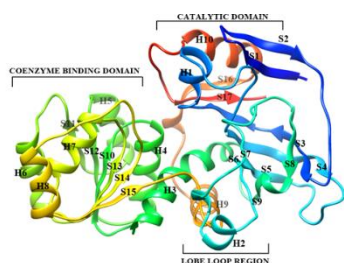
\*Corresponding author  
saisyah56@live.utm.my

<sup>a</sup>Bioinformatics Research Group, Faculty of Biosciences and Medical Engineering, Universiti Teknologi Malaysia, 81310, Johor Bahru, Malaysia

<sup>b</sup>Comparative Genomics and Genetics Research Centre, Malaysia Genome Institute, Kajang, Selangor, Malaysia

<sup>c</sup>Department of Bioprocess Engineering, Faculty of Chemical Engineering, Universiti Teknologi Malaysia, 81310, Johor Bahru, Malaysia

## Graphical abstract



## Abstract

Tagatose is a high value low calorie sweetener that is used as a sugar substitute in the food and pharmaceutical industry. The production of tagatose requires the conversion of galactitol-1-phosphate to tagatose-6-phosphate by galactitol-1-phosphate 5-dehydrogenase (PdGPDH). The objective of this work is to study the protein-ligand interaction between PdGPDH and its ligands; galactitol-1-phosphate, Zn<sup>2+</sup> and NAD<sup>+</sup>. Understanding of this mechanism will provide an insight into the possible catalytic events in these domains, thus providing information for potential protein engineering to improve the tagatose production. A 3D model of PdGPDH was constructed to identify the catalytic and coenzyme binding domains. In order to understand the interaction of PdGPDH with its ligands, a docking analysis of PdGPDH-substrate, PdGPDH-Zn<sup>2+</sup> and PdGPDH-NAD<sup>+</sup> complex was performed using CDOCKER in Discovery Studio 4.0 (DS 4.0). A series of docking events were performed to find the most stable binding interaction for the enzyme and its ligands. This study found that Cys 37, His 58, Glu 59, Glu 142 residues from PdGPDH form an active site pocket similar to known GPDH. A catalytic Zn<sup>2+</sup> binding domain and a cofactor NAD<sup>+</sup> binding domain with strong hydrogen bonding contacts with the substrate and the cofactor were identified. The binding pockets of the enzyme for galactitol-1-phosphate, NAD<sup>+</sup> and Zn<sup>2+</sup> has been defined. The stability of PdGPDH with its ligand was verified by utilizing the molecular dynamic simulation of docked complex. The results from this study will assist future mutagenesis study and enzyme modification work to improve the tagatose production.

**Keywords:** Protein-ligand interaction, galactitol-1-phosphate 5-dehydrogenase, tagatose production, molecular docking

## Abstrak

Tagatose adalah pemanis rendah kalori yang bernilai tinggi dan digunakan sebagai pengganti gula dalam makanan dan industri farmaseutikal. Pengeluaran tagatose memerlukan penukaran galactitol-1-fosfat kepada tagatose-6-fosfat dengan menggunakan galactitol-1-fosfat 5-dehydrogenase (PdGPDH). Objektif projek ini adalah untuk mengkaji interaksi antara protein PdGPDH dan ligand; galactitol-1-fosfat, Zn<sup>2+</sup> dan NAD<sup>+</sup>. Kefahaman mengenai mekanisme ini akan memberikan gambaran tindakan mangkinan dalam domain ini, justeru dapat menyediakan maklumat untuk kejuruteraan protein yang berpotensi dalam meningkatkan pengeluaran tagatose. 3D model PdGPDH

telah dibina untuk mengenal pasti domain mangkinan dan koenzim. Untuk memahami interaksi antara PdGPDH-substrat, PdGPDH-Zn<sup>2+</sup> dan PdGPDH-NAD<sup>+</sup> kompleks, analisis telah dilaksanakan menggunakan CDOCKER dalam perisian Studio 4.0 (DS 4.0). Satu siri mengedok telah dijalankan untuk mencari interaksi mengikat paling stabil untuk enzim dan ligand-ligand tersebut. Kajian ini mendapati bahawa asid amino Cys 37, His 58, Glu 59 dan Glu 142 dari PdGPDH membentuk poket tapak aktif yang sama seperti GPDH yang diketahui. Pengikat domain untuk pemangkin Zn<sup>2+</sup> dan kofaktor NAD<sup>+</sup> berserta ikatan hidrogen yang kukuh dalam substrat dan kofaktor telah dikenal pasti. Poket mengikat enzim untuk galactitol-1-fosfat, NAD<sup>+</sup> dan Zn<sup>2+</sup> telah dijelaskan. Kestabilan PdGPDH dengan ligand telah disahkan dengan menggunakan simulasi dinamik struktur dok. Hasil daripada kajian ini akan membantu mutagenesis kajian dan pengubahsuaian enzim untuk meningkatkan pengeluaran tagatose pada masa hadapan.

**Kata kunci:** Interaksi protein-ligand, galactitol-1-fosfat 5-dehydrogenase, pengeluaran tagatose, mengedok molekul

© 2016 Penerbit UTM Press. All rights reserved

## 1.0 INTRODUCTION

Today, an increasing number of researchers are focusing on tagatose production as an alternative sugar for healthy eating. Tagatose is a ketohexose monosaccharide sugar, which is a C-4 epimer of fructose [1]. Tagatose is rarely found in nature and present in only small amount in cacao, dairy product and fruits [2,3]. As there is no abundant source of tagatose in nature, this sugar is currently become subject of intensive investigation in many aspect.

Because of its unique properties, tagatose has been shown to have numerous health benefits including reduction of risk of type 2 diabetes, prevention of dental carries and treatment of obesity [4-6]. With 92% of the sweetness of sucrose but less than half the calories [7], tagatose can be used as anti-hyperglycemic agent [8]. Compared to other sweetener, tagatose shows the lowest glycemic index (GI) [9]. As health functional food for anti-hyperglycemic effect, Korea Food and Drug Administration (KFDA) approved the safety and function of tagatose for controlling the blood glucose level [10]. The proposed use of tagatose is safe within the term of the Federal Food, Drug and Cosmetic Act and it was approved as a 'generally recognized as safe' (GRAS) by Food and Drug Administration (FDA) [11].

The production of tagatose requires the conversion of galactitol-1-phosphate to tagatose-6-phosphate by galactitol-1-phosphate 5-dehydrogenase [12]. However, the catalytic mechanism of the oxidation of galactitol-1-phosphate is not as well characterized. Understanding this mechanism will provide an insight into the possible catalytic events, thus providing information for potential protein engineering to improve the tagatose production. Galactitol-1 Phosphate 5- Dehydrogenase from *Peptoclostridium difficile* CD196 (PdGPDH) belongs to the medium chain dehydrogenase family. It consist of two domains; a catalytic domain and a nicotinamide cofactor (NAD<sup>+</sup>) binding domain. The 3D structure and the active site of PdGPDH remained to be identified

and the interaction of substrate binding has not been studied in detail at atomic level. The present paper is the first study of the sequences and structural characterization of PdGPDH.

Here, we describe that the proton and hydride transfer occurs directly without any proton relay mechanism, in contrast to previously described in liver alcohol dehydrogenase [13] and galactitol dehydrogenase [14]. The proposed catalytic mechanism of GPDH for oxidation of L-galactitol-1-phosphate (LG1P) to tagatose-6-phosphate (DT6P) involves the His 58 acting as a general base, abstracting the proton from the C5 hydroxyl of LG1P and driving the transfer of a hydride ion onto C4 nicotinamide ring of NAD<sup>+</sup>. The C5 hydroxyl group of LG1P bound to zinc, thus making a pentacoordinated zinc ion in complex with the substrate, similar to the reported dehydrogenases.

In this study, we present computational studies on the interaction of PdGPDH with its ligands at molecular level. The 3D model of PdGPDH was generated based on the template from BLAST by using MODELER software, and was assessed with different tools for structure validation. The active site, a zinc metal binding domain and cofactor NAD<sup>+</sup> binding domain were identified. A series docking of PdGPDH-substrate, PdGPDH-Zn<sup>2+</sup> and PdGPDH-NAD<sup>+</sup> was performed using CDOCKER in Discovery Studio 4.0 (DS4.0). The docked complex was refined by molecular dynamic simulation to confirm its stable behaviour entire simulation period. Homology modelling did not only generate the desired protein but also helped in substrate identification and molecular docking to analyse the ligand binding interaction [15].

## 2.0 METHODOLOGY

### 2.1 Sequence Analysis

The target sequence was downloaded from GenBank Database under accession no. WP\_009890524.1. PSI-

BLAST and BLAST PDB were used to analyse the amino acid sequence of Galactitol-1-Phosphate 5-dehydrogenase from *Peptoclostridium difficile* (PdGPDH). To identify the conserved domains and the possible families of the protein, SUPER-FAMILY HMM server were utilised.

## 2.2 Model Construction and Evaluation

The three dimensional (3D) structure of NAD<sup>+</sup>-dependent PdGPDH was developed based on the crystal structure of GPDH (PDB ID:4UEO) as the template. The sequence structure alignment was optimised and used as the input to build a 3D model using MODELER from Accelrys Discovery Studio 4.0. The MODELER generated models for each alignment. The model with the lowest PDF and DOPE score were selected. The 3D model then was assessed by different tools including PROCHECK [16], Verify-3D [17], ERRAT [18], and ProsA-web [19].

## 2.3 Molecular Docking

The binding sites of PdGPDH were predicted with a defined receptor molecule using the binding site analysis tool in Accelrys Discovery Studio 4.0 (DS 4.0), which identified several binding sites. These binding sites were compared with the known GPDH. The potential substrates for PdGPH were identified by comparing the sequence and the structural traits of GPDH from previous publications [12,20]. The structural model of galactitol 1-phosphate, Zn<sup>2+</sup> and NAD<sup>+</sup> were extracted from the PUBCHEM server as SDF files. The compounds were then prepared by using Prepare Ligand protocol and were used for docking.

Molecular docking of all compounds was performed by CDOCKER implemented in Discovery Studio 4.0 which is a CHARMM based docking tool using a rigid receptor [21]. To create an explicitly solvated system, the solvation protocol was used. A set of 10 different orientations was randomly generated and placed into the receptor. A MD simulation was performed once the randomized ligand has been docked into the active protein site, starting with a gradual heating phase of 2000 1-fs steps from 300 to 700 K, and followed by a cooling phase of 5000 1-fs steps back to 300 K. Catalytic Zn<sup>2+</sup> and NAD<sup>+</sup> were docked into the Zn<sup>2+</sup> binding domain and coenzyme binding domain, respectively. A short energy minimization containing 100 steps of a smart minimizer was done for structure refinement, followed by 50 steps of a conjugate gradient. The energy minimized structure containing Zn<sup>2+</sup> and NAD<sup>+</sup> was used as the receptor for substrate docking.

For substrate docking, random substrate conformations were generated using high temperature MD. Random rigid-body rotations were used to create candidate poses followed by simulated annealing. The structure of the protein-substrate complex was subjected to energy minimization using the CHARMM force field

implemented in DS4.0. The substrate poses were then refined with a full-potential final minimization. The energy docked conformation of the substrate was retrieved for post-docking analysis based on CDOCKER as described previously [22]. The substrate orientation with the lowest interaction energy was chosen for the docking analysis. The hydrogen bonds of the ligands were defined within a distance 5 Å from the receptor in the binding pocket. The molecular docking or protein complexes were repeated twice to avoid artifacts, however the result of protein binding of these two replicates were same. Therefore, the analysis of only 1 set of docking is described below. The interaction of protein-ligand complexes were displayed and analysed by using 3D and 2D schematic diagrams.

## 2.4 Molecular Dynamic Simulation

Apo PdGPDH and PdGPDH complex were simulated by using GROMACS 4.6.5 package and the GROMOS96 force field for 20 ns in order to examine the molecular dynamic stability of the protein. The starting model of apo PdGPDH was built from the homology modelling of PdGPDH0002.pdb described above. While the initial structure of PdGPDH complex was taken from the highest scored pose of docked protein. The topology and force field parameters for the ligands were generated by using GlycoBioChem PRODRG2 server. The SPCE model was used for water molecules (-water spce) and GROMOS 96 54a7 was selected as force field for the simulation. A simple cubic box with the dimension 1.0 nm<sup>3</sup> was setup for the system. After defining the box dimension, the box was filled with water by using genbox, a program that added the correct number of water molecules needed to solvate the box. Sodium and chloride counter ions were added to preserve neutrality of the system.

The protein was imposed in periodic boundary condition (PBC) and electrostatic interactions were improved by Particle Mesh Ewald (PME) summation method which was used to treat coulomb potential which is the best method for computing long-range electrostatics. The system was energy minimized using 5000 steps from the steepest descent algorithm followed by equilibration for 100 ps pf solute-position – restrained MD. Each system was restrained MD in 2-fs time step using Linear Constraints (LINCS) algorithm for fixing all bond lengths in the system. The system stability and behavior of the proteins were analyzed with use of the tools available in GROMACS 4.6.5 program. To compare the stability and structure folding reliability of apo PdGPDH and PdGPDH complex, the root mean square deviation (RMSD), radius of protein gyration (Rg), and total energy of the proteins were utilized by using g\_rms, g\_gyrate and g\_energy program of GROMACS.

### 3.0 RESULTS AND DISCUSSION

#### 3.1 Sequence Analysis

Multiple sequence alignment (MSA) is a key role in comparative structure and function analysis of biological sequences which help the analysis of the the fundamental biology of sequence-structure-function relationships of protein sequence families [19]. The results from the ClustalW online server showed that Galactitol 1-phosphate 5-dehydrogenase from *Peptoclostridium difficile* (PdGPDH) has 34.3% sequence identity and 58.5% sequence similarity with Galactitol 1-phosphate Dehydrogenase (GPDH) from *E.coli*. The PSI-BLAST and SUPERFAMILY HMM library result identified that catalytic domain of PdGPDH as Medium Chain Dehydrogenase/Reductase (MDR). MDR contains zinc-dependent alcohol dehydrogenase (ADH-Zn) and related protein.

From the sequence alignment in Figure 1, PdGPDH and the other dehydrogenase enzymes [20,29-31] are strictly conserved for the catalytic zinc at active site (green stars) and structural zinc (yellow stars). The active site zinc is coordinated by a cysteine (position 37), histidine (position 58) and glutamine (position 59), while the second structural ion is coordinated by four cysteines (position 88, 91, 94 and 102). Notably, PdGPDH shows conserved a glycine in the GHE motif. Tiwari and co-workers (2012) [32] demonstrated that conserved the glycine in GHE motif plays a role in maintaining the metal binding affinity and the electronic state of the catalytic zinc ion during catalysis of the MDR superfamily enzymes. Other highly conserved regions include the GXGXXG dinucleotide binding motif of the Rossmann fold (formed by Gly 166, Gly 168 and Gly 171).

#### 3.2 Three Dimensional Model Development

The 3D features of the PdGPDH model were based on the crystal structure of GPDH (PDB ID: 4UEO). To build the 3D model, the alignment between GPDH and PdGPDH was submitted to the MODELER program from Discovery Studio 4.0. MODELER is a program used to compare protein structure models based on the satisfaction of spatial restraint. The restraints are designed from the alignment between the input and the template sequence. Twenty models were generated by MODELER. PdGPDH0002.pdb was chosen as the best model structure for PdGPDH because it has the lowest PDF score and DOPE with 1740 and -36319, respectively. The molecular PDF (molpdf) is the sum of restraint violation while Discrete Optimized Protein Energy (DOPE) score is an atomic distance-dependent statistical potential used to assess the homology models in protein structure prediction [33].

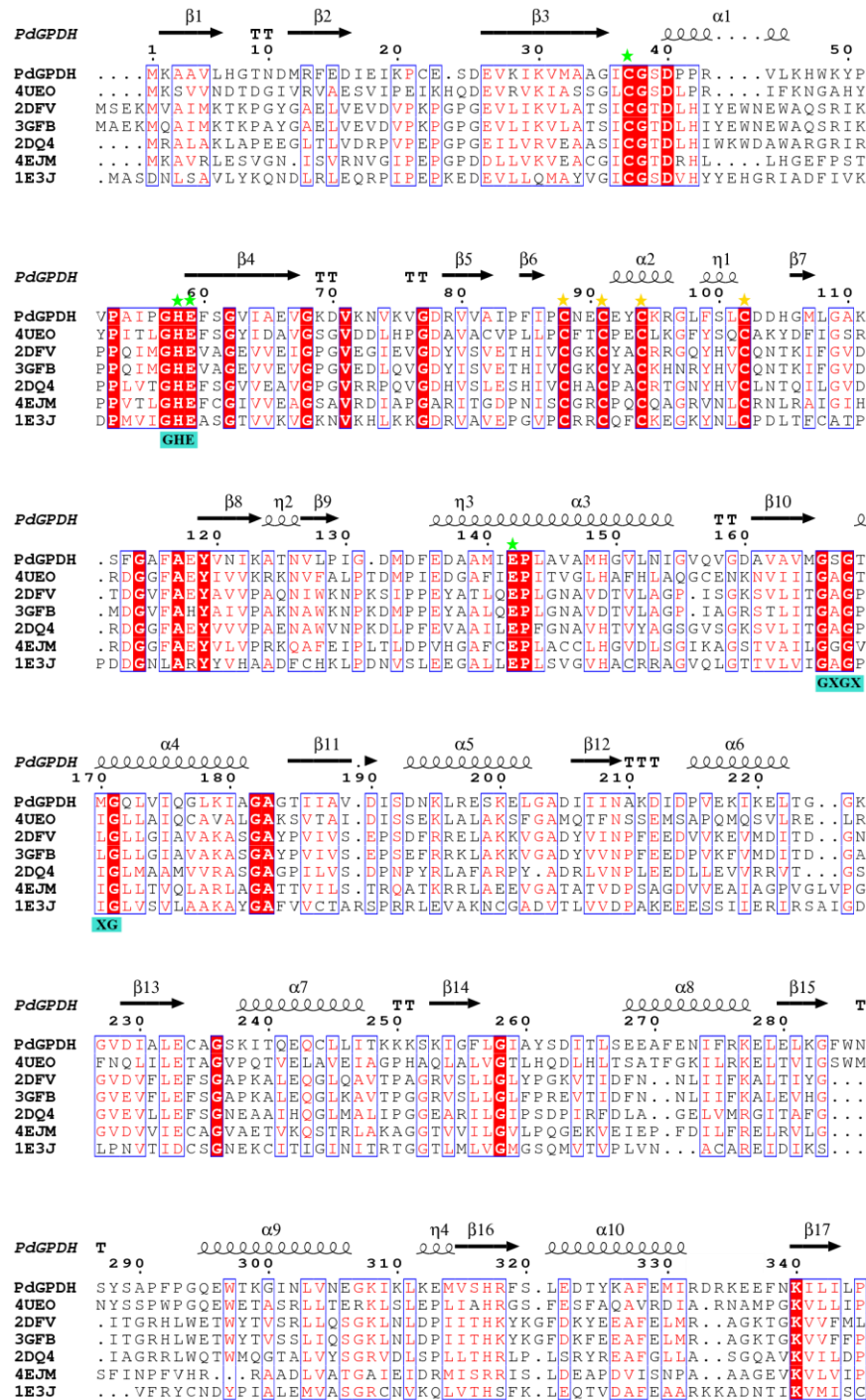
The 3D model of PdGPDH is represented in Figure 2. The structural features of the model are similar to other members in the MDR family which generally

have two tightly bound zinc atoms per subunit, a catalytic zinc at the active site and a structural zinc ion. Cysteine (Cys 37), histidine (His 58) and glutamine (Glu 59) are coordinated to the active site zinc and four cysteine (Cys 88, Cys 91, Cys 94, Cys 102) are coordinated to the second zinc, located in the lobe loop region. The structure of PdGPDH contains ten  $\alpha$ -helices (H1-H10) and 17  $\beta$ -strands (S1-S17). PdGPDH is folded into two domains; a catalytic domain (residues 1-146 and 285-350) and a coenzyme binding domain (residues 147-284) separated by a deep cleft. The coenzyme domain contains the  $\alpha/\beta$  Rossmann fold dinucleotide binding protein [34] that consist a six stranded parallel  $\beta$ -sheet, flanked by five  $\alpha$ -helices.

Figure 3 shows a representation of the superimposition between PdGPDH 3D model and GPDH as the selected template. The superimposition of these two proteins showed a good structure alignment with RMSD of 0.25 Å and a 98% coverage of the backbone atoms. PdGPDH and GPDH show a high degree of structural similarity despite their low sequence similarity. The residues in the catalytic and structural region of PdGPDH that are highlighted in the grey boxes had similar orientations and positions to the residues in the template GPDH.

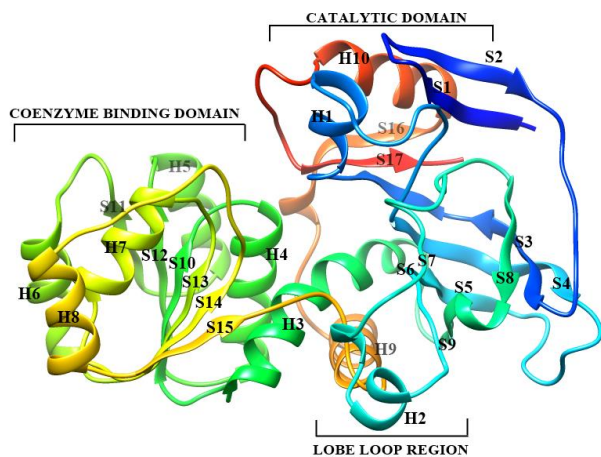
#### 3.3 Structural Model Validation

Several model validation tools such as PROCHECK [16], VERIFY 3D [17], ERRAT [18], and ProSA-web [19] were used to assess the quality of the 3D model. From the PROCHECK analysis, the backbone conformation was evaluated based on the Psi/Phi Ramachandran plot. The Ramachandran plot showed that the Psi/Phi angle of 92.6% are in the most favoured region, 7% in additional and generously allowed region and only 0.3% amino acid residue are in disallowed region. Therefore, 99.60% residues are in allowed regions. The result of VERIFY-3D showed 80% of the residues had an average 3D-1D score above 0.2 and the rest of the residues (20%) obtained the lower score. A VERIFY-3D above 80% is considered as satisfactory quality of a predicted model. The model assessment was also performed by using ERRAT to calculate the overall quality score for non-bonded atomic interaction by a comparison of the statistics with highly refined structures. According to Chaitanya and co-workers [35], the normally accepted range of ERRAT score for a high quality model is greater the 50%. The ERRAT score of PDGPDH model was 68.36% which is acceptable in the normal range.

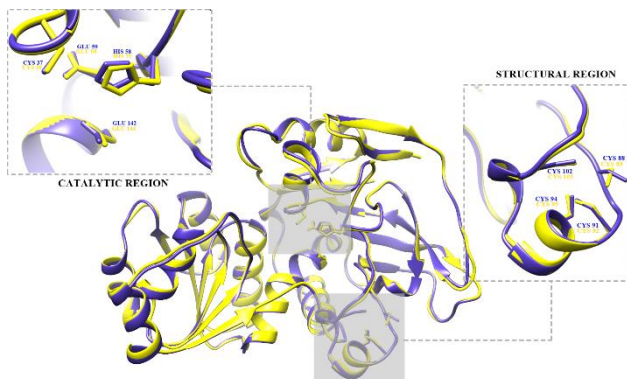


**Figure 1** Sequence alignment for PdGPDH with the closest structural homologues; galactitol-1-phosphate 5-dehydrogenase from *E.coli* (PDB ID: 4UEO), threonine dehydrogenase from *P.horikoshii* (PDB ID:2DFV), L-threonine dehydrogenase form *T.kodakaraensis* (PDB ID:3GFB), threonine 3-dehydrogenase from *thermophilus* PDB ID: 2DQ4), zinc-binding dehydrogenase from *S.melloti* (PDB ID:4EJM), and sorbitol dehydrogenase from silverleaf whitefly (PDB ID:1E3J). The predicted secondary structure elements for PdGPDH were shown above the sequence. The conserved positions are shown in red characters. Residues that coordinate the catalytic zinc are in green stars and four cysteine residues that coordinate the structural zinc ion are within yellow stars. The cyan boxes denote to the conserved glycine in GHE and GXGXXG motif. The presentation of the sequence alignment is made using EsPript

Lastly, ProSA-web was used to analyse the overall quality score for the model. If this score is in the normal range for native protein, the model most likely contain no error. The z-score of the structure model PdGPDH was calculated to be -7.98 which is considered to be within the normal range of scores for native proteins of a similar size. Table 1 shows the summary of the evaluation result. Overall, the values of predicted PdGPDH model obtained from the different validation tools are considered reasonable.



**Figure 2** 3D-model of PdGPDH representing the secondary structure elements including  $\alpha$ -helices (H1-H10) and  $\beta$ -strands (S1-S17). The structure PdGPDH comprises two domains, a coenzyme binding domain and a catalytic domain



**Figure 3** Superimposition of PdGPDH (yellow) and its template GPDH (blue) in a cartoon representation. The grey boxes donate to two domain region including the catalytic region and structural region

**Table 1** Summary of model validation using different tools

Model evaluation tools	Evaluation scheme	score	Normal range of the score
PROCHECK	The number of residues in allowed region based on Psi/Phi Ramachandran plot	99.60%	>90%

Model evaluation tools	Evaluation scheme	score	Normal range of the score
VERIFY 3D	The number of residues having an average 3D-1D score above 0.2	80%	>80%
ERRAT	The overall quality for nonbonded atomic interaction	68.36%	>50%
ProSA-web	Model evaluation by calculating an overall quality score (z-score)	-7.98	NPS

NPS is native protein size check whether the Z-score of the input structure is within the range of scores for native proteins of similar size.

### 3.4 Molecular Docking

#### 3.4.1 Location and Binding of $Zn^{2+}$ Ions

With a broad range of activities, the members of the MDR superfamily are currently a subject of intensive investigation [36,37]. Generally, MDR have two tetrahedrally coordinated zinc ions per subunit; (1) a catalytic zinc at the active site with three residues, one histidine, and two cysteine residues and (2) a structural zinc that interacts with four protein residues [38]. In the polyol dehydrogenase family, the third catalytic residue is varied which can either be the residue that is adjacent to the His or another Glu that is 85 residues away [36].

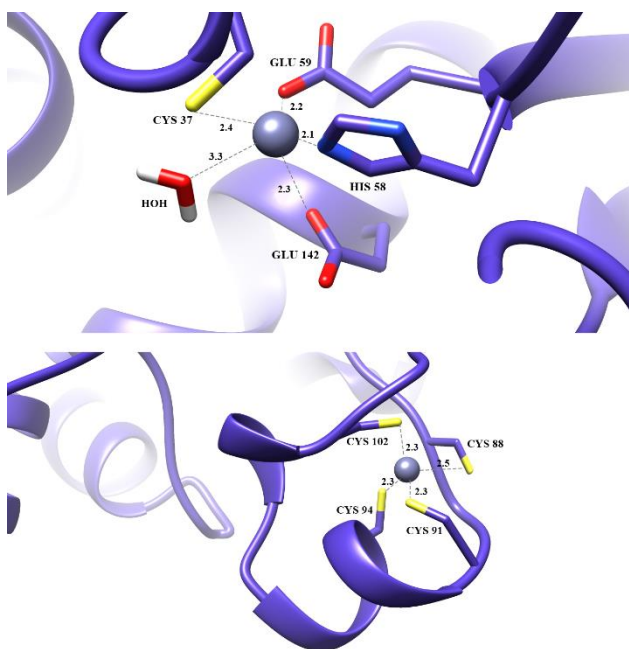
The structure of PdGPDH reveals a catalytic  $Zn^{2+}$  binding site coordinates by four protein residues (Cys 37, His 58, Glu 59, and Glu 142) with distances of 2.4Å, 2.1Å, 2.2Å, and 2.3Å, respectively, along with a water molecule (Figure 4). The position of Glu 59 was adjacent to the principle residue His 58, while another Glu 142 was 84 residues away. The structural zinc ion coordinated by Cys 88 (2.5Å), Cys 91 (2.3Å) Cys 94 (2.3Å) Cys 102 (2.3Å). This zinc ion stabilizes a long loop extension from the sheet structure.

#### 3.4.2 Coenzyme Binding Domain

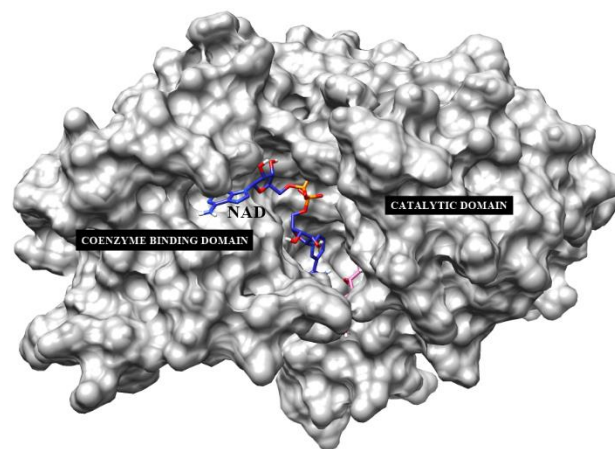
Most of the MDR proteins are zinc-dependent proteins which catalyze the oxidation of primary or secondary alcohols to the corresponding aldehydes or ketones using  $NAD(P)^+$  as a cofactor [39].  $NAD$  consists of two nucleotides joined through their phosphate group; one nucleotide contains an adenine base and the other is nicotinamide [40]. In the present research, the binding of  $NAD^+$  to the coenzyme binding domain was analysed. The adenosine half bound in a cleft at the surface of the domain, and the nicotinamide half bound deep in the protein at the active site cleft. The adenine base is positioned in a hydrophobic cleft with the N6 amino group pointing out towards the surface (Figure 5).

In this study, the  $NAD^+$  requires a glycine rich, highly conserved GXGXXG sequence motif. This motif was also present within the sequence of GPDH and comprised the following amino acids GAGTIG. The Gly

38, Gly 168, Thr 169, Met 170 and water molecules were hydrogen bonded with oxygen of the pyrophosphate. A similar hydrogen bond was observed between a water molecule and nitrogen at the adenine base. The presence of water molecules implicated a significant component in dinucleotide recognition [41]. The Ala 235 (2Å) formed hydrogen bond with C3 hydroxyl group of ribose (Figure 7b). The Asn 286 (2.5Å) and Leu 257 (1.9Å) were hydrogen bonded with nitrogen and oxygen at C1 of the nicotinamide ring respectively. The C2 and C3 hydroxyl group of the other ribose which was associated with adenine moiety, were hydrogen bonded to Asp 192 (1.9Å) and Lys 195 (1.8Å). The presence of an aspartic acid (Asp) constitute a common feature of dehydrogenases with preference of NAD<sup>+</sup> over NADP<sup>+</sup> [34]. The Asp would create repulsion of the extra phosphate group of NADP due to charge and space, thus explaining the coenzyme dependency.



**Figure 4** The binding of zinc at a) catalytic region and b) structural region. The catalytic zinc is coordinated by Cys 37, His 58, Glu 144 and a water molecule. The structural zinc is coordinated by 4 cysteine in the lobe loop region



**Figure 5** The binding of NAD<sup>+</sup> to the coenzyme binding domain. NAD<sup>+</sup> is shown in the blue stick representation, with a transparent molecular surface. The adenosine half bound in a cleft at the surface of the domain, and the nicotinamide half bound deep in the protein was at the active site cleft. The presentation of the structure is made by using Chimera software

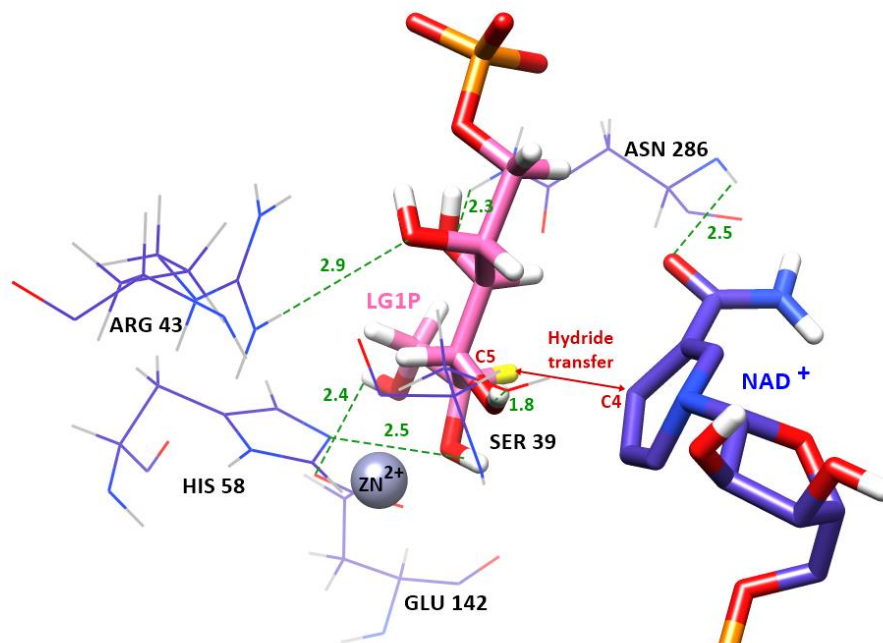
### 3.4.3 Substrate Binding

The possible interaction between PdGPDH and galactitol-1-phosphate (G1P) has been studied by molecular docking of substrate G1P into the active site of the complex PdGPDH, Zn<sup>2+</sup> and NAD<sup>+</sup>. The binding mode of Zn<sup>2+</sup> and the substrate in the catalytic site of PdGPDH is illustrated in Figure 6a. All of the hydroxyl groups of the substrate were within the distance for hydrogen bonding. The C2 hydroxyl group of G1P formed hydrogen bonding with Arg 43 (2.9 Å). The C3 OH group was also hydrogen bonded (2.3 Å) with the Asn 286 side chain. The C4 OH group interacted with the Ser 39 side chain (1.8 Å). The C5 and C6 OH group of G1P interacted via hydrogen bonding with His 58 (2.5 Å) and Glu 142 (2.4 Å), respectively. The 2D diagrams of the interaction in catalytic and coenzyme domain were generated in Figure 7 (a) and (b).

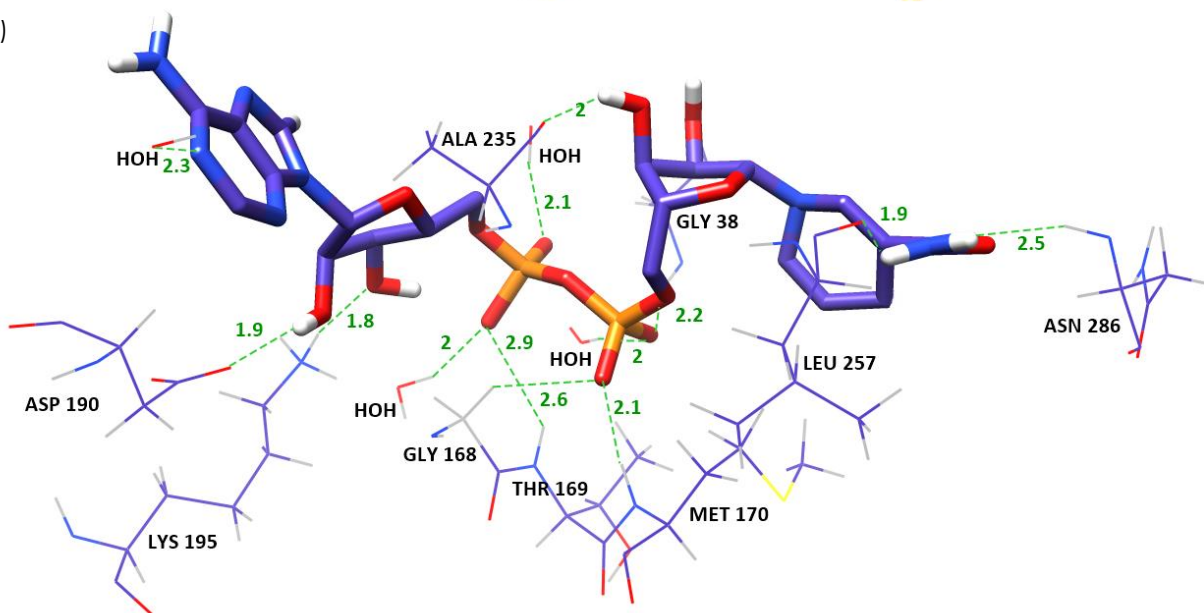
### 3.4.4 Direct Proton and Hydride Transfer Mechanism

The study suggests that the hydride transfer mechanism is different from previously studied dehydrogenases [13,14]. Overall, the mechanism for most of the alcohol dehydrogenase (ADHs) is ordered. The coenzyme binds before substrate binding occurs. The positive charge of the NAD<sup>+</sup> nicotinamide ring contributed to the deprotonation of the substrate alcohol and promoted binding to the active site Zn<sup>2+</sup> ion. The Zn<sup>2+</sup> atom was coordinated by four proteins side chain (Cys 37, His 58, Glu 59 and Glu 142) and a water molecule at the active site of the apo enzyme. This metal bound water molecule is displaced by the hydroxyl group of the LG1P on substrate binding, to leave the zinc ion coordinated to the substrate and the same four protein ligands as the apo enzyme.

a)

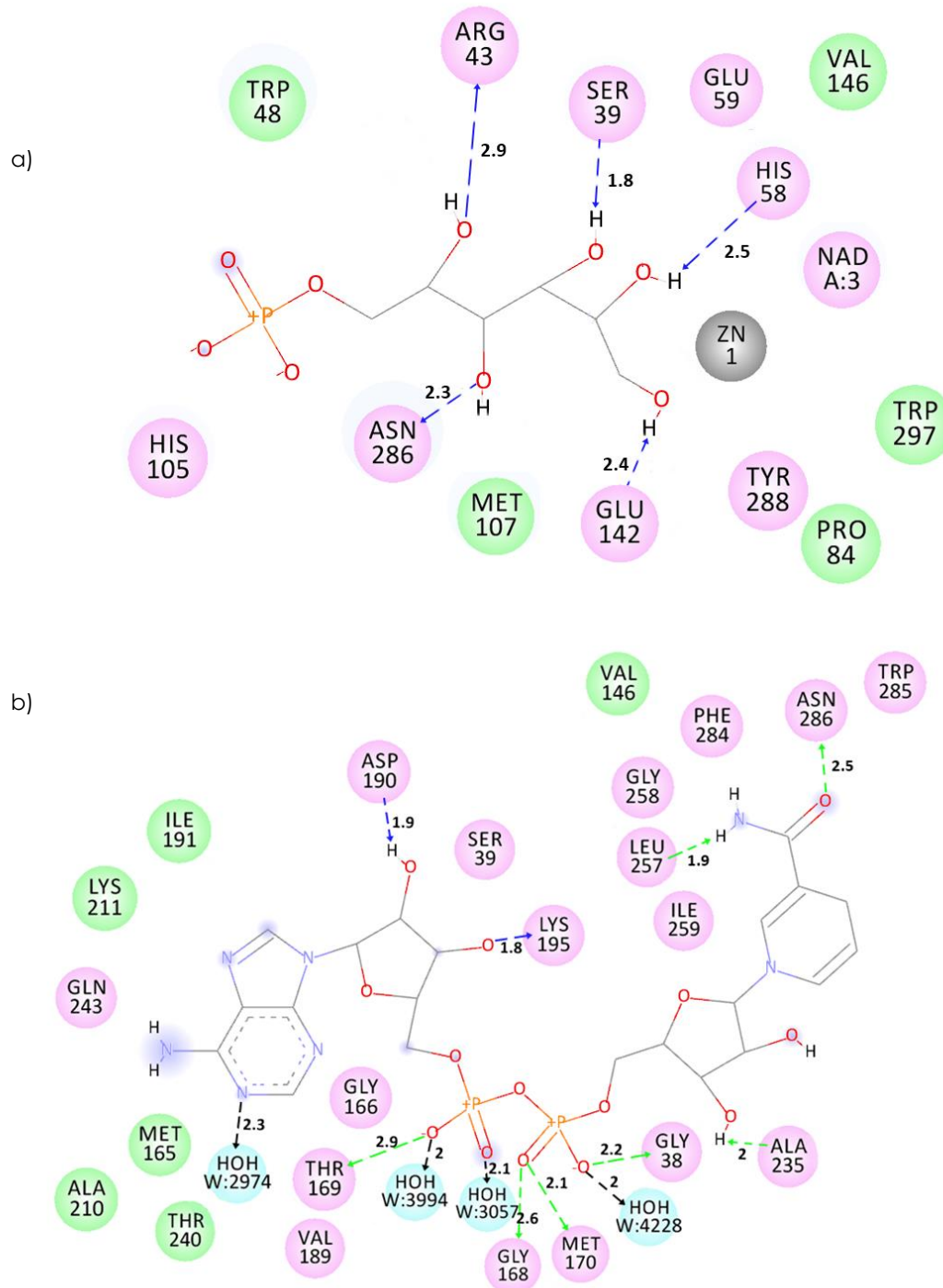


b)

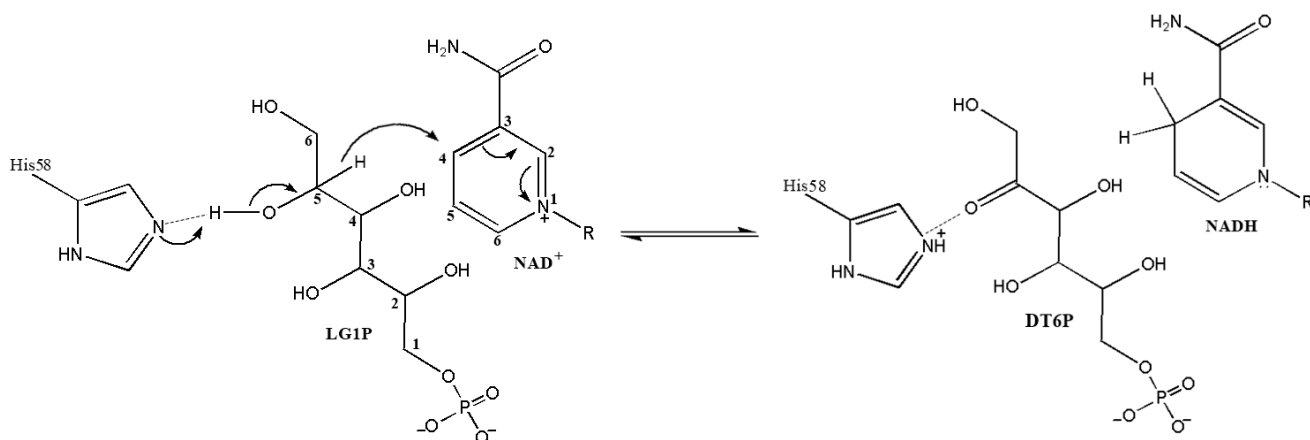


**Figure 6** (a) The binding mode of  $Zn^{2+}$  and substrate in catalytic site. (b) The interaction of PdGPDH with  $NAD^+$  in the coenzyme binding domain. The grey ball indicates the metal ion and the residues involved in the interaction are represented in wire form. Galactitol-1-phosphate (G1P) is shown in the pink stick model while  $NAD^+$  is shown in the blue stick model. Hydrogen bonds are the green dashed line and the distances shown are in Å





**Figure 7** The interactions of PdGPDH with (a)  $Zn^{2+}$  and substrate in catalytic site; (b)  $NAD^+$  in the coenzyme binding domain. The residues involved in various events are represented as follows. Hydrogen bond or polar interactions (magenta-colored circles), van der Waals (green circles), metal atoms (grey circles), water molecules (aquamarine circles), hydrogen-bond interactions with non-amino acid residues (black dashed line), charge interactions (pink dashed line), amino acid main chains (green dashed line), and amino acid side chains (blue dashed line). The distances shown are in Å. The 2D schematic diagrams were made by using Discovery Studio 4.0 (DS 4.0)



**Figure 8** Proposed catalytic mechanism of GPDH for oxidation of L-galactitol-1-phosphate (LG1P) to tagatose-6-phosphate (DT6P). In the conversion of LG1P to DT6P, it is proposed that His 58 acts as general base, abstracting the proton from the C5 hydroxyl of LG1P and driving the transfer of a hydride ion onto C4 nicotinamide ring of NAD<sup>+</sup>. The mechanism diagram is made using ChemDraw

The C5 hydroxyl group of LG1P bound to zinc, thus making a pentacoordinated zinc ion in complex with the substrate.

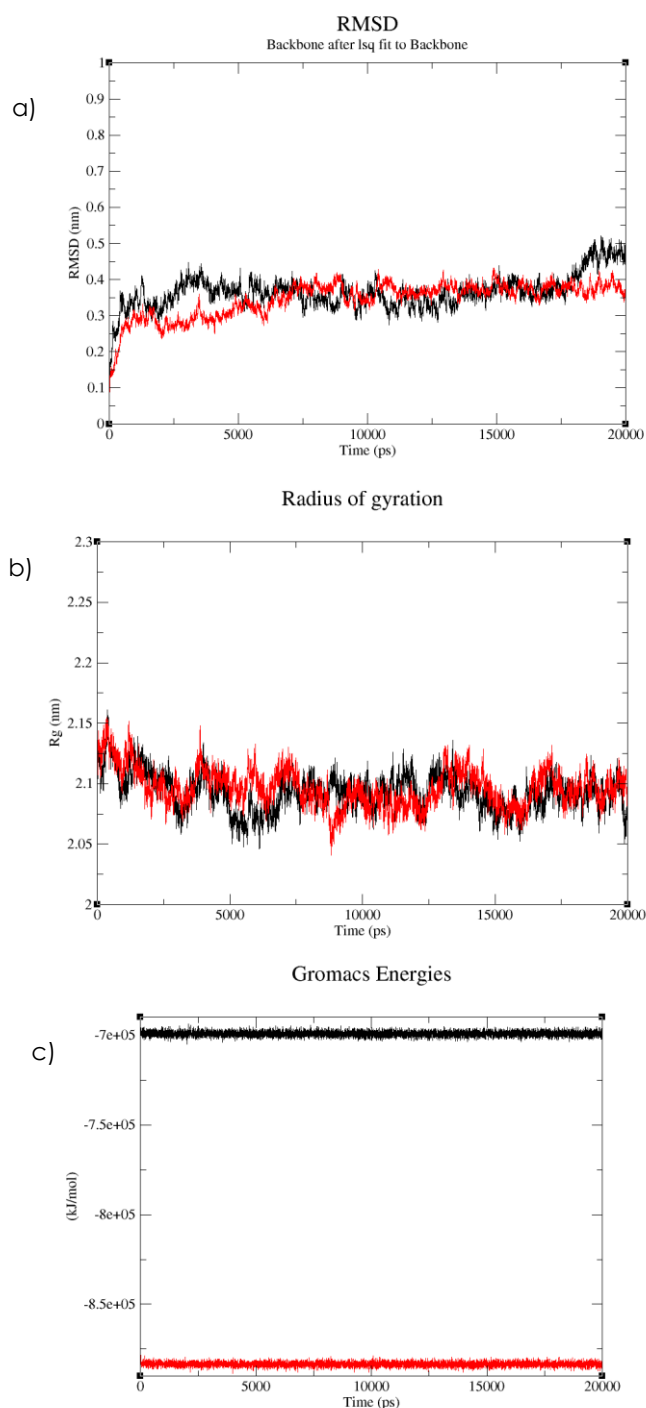
This study found that the C5 hydroxyl of galactitol-1-phosphate (G1P), which was oxidized by the enzyme to a tagatose-6-phosphate (T6P) was hydrogen bonded with catalytic His 58 (Figure 8). To facilitate hydride transfer, the substrate and nicotinamide ring were placed close to each other. The nicotinamide moiety is bound in the active site, and is positioned to donate a hydride ions to the C5 atom of the substrate during the catalysis. The C5 carbon was 3.5 Å from the nicotinamide C4 which was a suitable distance for hydride transfer. This suggests that the proton and hydride transfer occurs directly without any proton relay mechanism previously described in liver alcohol dehydrogenase [13] and galactitol dehydrogenase [14].

### 3.4 Molecular Dynamic Simulation

To determine stability of conformations among apo PdGPDH and PdGPDh complex MD simulations, we utilized root mean square deviation (RMSD), radius of protein gyration, and total energy. The overall stability of the protein throughout the molecular dynamics simulations was monitored by the root-mean-square deviation (RMSD) of backbone which measure of the average distance between two conformations of a protein. The RMSD values of apo PdGPDH and PdGPDH complex in the entire MD simulation trajectory were shown in Figure 9(a). The average RMSD of apo PdGPDH was 0.37 nm whereas the average RMSD of PdGPDH complex was 0.35 nm. From this graph, it can be seen that the apo PdGPDH RMSD became stable after initial deviation. Further, it maintained stable conformation before slightly fluctuate during the last 3 ns. For the complex of PdGPDH, the protein was equilibrated with no obvious RMSD fluctuations observed over the 20 ns simulation period. The RMSD gradually increased in the first 7 ns

and then converged in the time frame from 7 to 20 ns. The RMSD curves of PdGPDH complex show less deviation compared to apo protein which indicate the stable dynamic behaviour during the 20ns simulation period.

The compactness of both proteins were described by calculating the radius of gyration. The apo PdGPDH and PdGPDh complex exhibited a similar pattern of Rg value, which stabilize over the entire simulation time, at average gyration distance (Rg) of approximately 2.09 nm (Figure 9b). This result indicates that both Apo and PdGPDH complex kept the fold of their original design, and the maintained their compactness during 20 ns simulation time. Both simulations show constant total energy during the MD run (Figure 9c). The total energy calculated for apo PdGPDH is higher than PdGPDH complex with average of  $-6.99216 \times 10^5$  kJ mol<sup>-1</sup> and  $-8.83403 \times 10^5$  kJ mol<sup>-1</sup> respectively.



**Figure 9** (a) Proposed catalytic Root mean square deviation (RMSD) (b) radius of gyration (Rg) and (c) total energy of apo PdGPDH and PdGPDH complex during 20 ns MD simulation. Apo PdGPDH is represented in black line, while PdGPDH complex is depicted in red line

#### 4.0 CONCLUSION

To date, this is the first known report on the structural analysis and protein interaction of G1P-PdGPDH

complex using molecular docking and molecular dynamic simulation. The binding pocket of the enzyme for G1P, NAD<sup>+</sup> and Zn<sup>2+</sup> has been defined. The residues involved in the catalytic domain and coenzyme binding domain could potentially contribute to the protein dynamic that influences the catalytic mechanism. The molecular dynamic simulation of PdGPDH with its ligand verified stability of the docked complex. An understanding of this mechanism will provide an insight into the possible catalytic events in these domains, thus providing the information for future protein engineering. The interaction between the enzyme and substrate proposed in this study would also assist future mutagenesis study and enzyme modification work to improve tagatose production.

#### Acknowledgement

The work is supported by the Malaysian Genome Institute and Ministry of Science, Technology and Innovation for funding and research support. We thank Peter Klappa (University of Kent) for useful comments on the manuscript and the research.

#### References

- [1] Bertelsen, H., Jensen, B. B., Buemann, B. 1999. D-Tagatose-a Novel Low-Calorie Bulk Sweetener With Prebiotic Properties. 98-109.
- [2] Mendoza, M. R., Olano, A., Villamiel, M. 2005. Chemical Indicators Of Heat Treatment In Fortified And Special Milks. *Journal Of Agricultural And Food Chemistry*. 53: 2995-2999.
- [3] Vastenavond, C. M., Bertelsen, H., Hansen, S. J., Laursen, R. S., Saunders, J., et al. 2011. Tagatose (d-tagatose). Alternative sweeteners. 197e221.
- [4] Wong, D. 2000. Sweetener Determined Safe In Drugs, Mouthwashes, And Toothpastes. *Dentistry Today*. 19(32): 34-35.
- [5] Moore, M. C. 2006. Drug Evaluation: Tagatose In The Treatment Of Type 2 Diabetes And Obesity. *Current Opinion In Investigational Drugs (London, England: 2000)*. 7: 924-935.
- [6] Lu, Y., Levin, G., Donner, T. 2008. Tagatose, A New Antidiabetic And Obesity Control Drug. *Diabetes, Obesity And Metabolism*. 10: 109-134.
- [7] Levin, G. V., Zehner, L. R., Saunders, J. P., Beadle, J. R. 1995. Sugar Substitutes: Their Energy Values, Bulk Characteristics, And Potential Health Benefits. *The American Journal Of Clinical Nutrition*. 62: 1161S-1168S.
- [8] Zehner, L. R., Levin, G. V., Saunders, J. P., Beadle, J. R. 1995. d-Tagatose As Anti-hyperglycemic Agent. *Google Patents*.
- [9] Seri, K., Sanai, K., Negishi, S. 1997. Prophylactic And Remedial Preparation For Diseases Attendant On Hyperglycemia, And Wholesome Food. EP, 0560284.
- [10] Park, C., Lee, J-S. 2013. Mini Review: Natural Ingredients For Diabetes Which Are Approved By Korean FDA. *Biomedical Research*. 24: 164-169.
- [11] Levin, G. V. 2002. Tagatose, The New GRAS Sweetener And Health Product. *Journal Of Medicinal Food*. 5: 23-36.
- [12] Esteban-Torres M, Álvarez Y, Acebrón I, de las Rivas B, Muñoz R, et al. 2012. The Crystal Structure Of Galactitol-1-Phosphate 5-Dehydrogenase From *Escherichia Coli* K12 Provides Insights Into Its Anomalous Behavior On IMAC Processes. *FEBS Letters*. 586: 3127-3133.

- [13] Agarwal, P. K., Webb, S. P., Hammes-Schiffer, S. 2000. Computational Studies Of The Mechanism For Proton And Hydride Transfer In Liver Alcohol Dehydrogenase. *Journal of the American Chemical Society*. 122: 4803-4812.
- [14] Benavente, R., Esteban-Torres, M., Kohring, G-W, Cortés-Cabrera, Á., Sánchez-Murcia, P. A., et al. 2015. Enantioselective Oxidation Of Galactitol 1-phosphate by galactitol-1-phosphate 5-dehydrogenase from *Escherichia coli*. *Acta Crystallographica Section D: Biological Crystallography*. 71: 1540-1554.
- [15] Sudí, I. Y., Wong, E. L., Joyce-Tan, K. H., Shamsir, M. S., Jamaluddin, H., et al. 2012. Structure Prediction, Molecular Dynamics Simulation and Docking Studies of D-Specific Dehalogenase from *Rhizobium* sp. RC1. *International Journal Of Molecular Sciences*. 13: 15724-15754.
- [16] Laskowski, R. A., MacArthur, M. W., Moss, D. S., Thornton, J. M. 1993. PROCHECK: A Program To Check The Stereochemical Quality Of Protein Structures. *Journal Of Applied Crystallography*. 26: 283-291.
- [17] Eisenberg, D., Lüthy, R., Bowie, J. U. 1997. VERIFY3D: Assessment Of Protein Models With Three-Dimensional Profiles. *Methods In Enzymology*. 277: 396-404.
- [18] Colovos, C., Yeates, T. O. 1993. Verification Of Protein Structures: Patterns Of Nonbonded Atomic Interactions. *Protein Science*. 2: 1511-1519.
- [19] Wiederstein, M., Sippl, M. J. 2007. Prosa-Web: Interactive Web Service For The Recognition Of Errors In Three-Dimensional Structures Of Proteins. *Nucleic Acids Research*. 35: W407-W410.
- [20] Benavente, R., Esteban-Torres, M., Kohring G-W, Cortés-Cabrera, Á., Sánchez-Murcia, P. A., et al. 2015. Enantioselective Oxidation Of Galactitol 1-phosphate by galactitol-1-phosphate 5-dehydrogenase from *Escherichia coli*. *Acta Crystallographica Section D: Biological Crystallography*. 71: 1540-1554.
- [21] Wu, G., Robertson, D. H., Brooks, C. L., Vieth, M. 2003. Detailed analysis Of Grid-Based Molecular Docking: A Case Study Of CDOCKER—A Charrm-Based MD Docking Algorithm. *Journal of Computational Chemistry*. 24: 1549-1562.
- [22] Tiwari, M., Lee, J-K. 2010. Molecular Modeling Studies Of L-Arabinitol 4-Dehydrogenase Of *Hypocrea Jecorina*: Its Binding Interactions With Substrate And Cofactor. *Journal of Molecular Graphics and Modelling*. 28: 707-713.
- [23] Van der Spoel, D., Lindahl, E., Hess, B. 2013. The GROMACS Development Team, GROMACS User Manual, version 4.6. 5.
- [24] Schuler, L. D., Daura, X., Van Gunsteren, W. F. 2001. An Improved GROMOS96 Force Field For Aliphatic Hydrocarbons In The Condensed Phase. *Journal of Computational Chemistry*. 22: 1205-1218.
- [25] Schüttelkopf, A. W., Van Aalten, D. M. 2004. PRODRG: A Tool For High-Throughput Crystallography Of Protein-Ligand Complexes. *Acta Crystallographica Section D: Biological Crystallography*. 60: 1355-1363.
- [26] Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., et al. 1995. A Smooth Particle Mesh Ewald Method. *The Journal of Chemical Physics*. 103: 8577-8593.
- [27] Hess, B., Bekker, H., Berendsen, H. J., Fraaije, J. G. 1997. LINCS: A Linear Constraint Solver For Molecular Simulations. *Journal Of Computational Chemistry*. 18: 1463-1472.
- [28] Edgar, R. C., Batzoglou, S. 2006. Multiple Sequence Alignment. *Current Opinion In Structural Biology*. 16: 368-373.
- [29] Ishikawa, K., Higashi, N., Nakamura, T., Matsuura, T., Nakagawa, A. 2007. The First Crystal Structure Of L-Threonine Dehydrogenase. *Journal Of Molecular Biology*. 366: 857-867.
- [30] Bowyer, A., Mikolajek, H., Stuart, J., Wood, S., Jamil, F., et al. 2009. Structure And Function Of The L-Threonine Dehydrogenase (TktDh) From The Hyperthermophilic Archaeon *Thermococcus kodakaraensis*. *Journal Of Structural Biology*. 168: 294-304.
- [31] Banfield, M. J., Salvucci, M. E., Baker, E. N., Smith, C. A. 2001. Crystal structure Of The NADP (H)-Dependent Ketose Reductase From *Bemisia argentifolii* at 2.3 Å Resolution. *Journal Of Molecular Biology*. 306: 239-250.
- [32] Tiwari, M. K., Singh, R. K., Singh, R., Jeya, M., Zhao, H., et al. 2012. Role Of Conserved Glycine In Zinc-Dependent Medium Chain Dehydrogenase/Reductase Superfamily. *Journal of Biological Chemistry*. 287: 19429-19439.
- [33] Shen, My, Sali, A. 2006. Statistical Potential For Assessment And Prediction Of Protein Structures. *Protein Science*. 15: 2507-2524.
- [34] Lesk, A. M. 1995. NAD-binding Domains Of Dehydrogenases. *Current Opinion In Structural Biology*. 5: 775-783.
- [35] Chaitanya, M., Babajan, B., Anuradha, C., Naveen, M., Rajasekhar, C., et al. 2010. Exploring The Molecular Basis For Selective Binding Of Mycobacterium Tuberculosis Asp Kinase Toward Its Natural Substrates And Feedback Inhibitors: A Docking And Molecular Dynamics Study. *Journal Of Molecular Modeling*. 16: 1357-1367.
- [36] Persson, B., Hedlund, J., Jörnvall, H. 2008. Medium-And Short-Chain Dehydrogenase/Reductase Gene And Protein Families. *Cellular And Molecular Life Sciences*. 65: 3879-3894.
- [37] Gonzalez-Duarte, R., Albalat, R. 2005. Merging Protein, Gene And Genomic Data: The Evolution Of The MDR-ADH Family. *Heredity*. 95: 184-197.
- [38] Nordling, E., Jörnvall, H., Persson, B. 2002. Medium-Chain Dehydrogenases/Reductases (MDR). *European Journal of Biochemistry*. 269: 4267-4276.
- [39] Eklund, H., Plapp, B., Samama, J., Brändén, C. 1982. Binding Of Substrate In A Ternary Complex Of Horse Liver Alcohol Dehydrogenase. *Journal of Biological Chemistry*. 257: 14349-14358.
- [40] Baker, P. J., Britton, K. L., Rice, D. W., Rob, A., Stillman, T. J. 1992. Structural consequences Of Sequence Patterns In The Fingerprint Region Of The Nucleotide Binding Fold: Implications For Nucleotide Specificity. *Journal Of Molecular Biology*. 228: 662-671.
- [41] Bottoms, C. A., Smith, P. E., Tanner, J. J. 2002. A Structurally Conserved Water Molecule In Rossmann Dinucleotide-Binding Domains. *Protein Science*. 11: 2125-2137.