Journal of Advanced Bioinformatics Applications and Research
ISSN 0976-2604.Online ISSN 2278–6007
Vol 5, Issue3, 2014, pp197-205
http://www.bipublication.com

Sequence and Structure Analysis of Pyranose Dehydrogenase in *Agaricus campestris* through *Insilico* methods

Nitin Kumar Verma[1*] and Vibha Verma[2]
[1]Department of Biotechnology and Bioinformatics,
Uttaranchal College of Science and Technology, Dehradun, India
[2]Forest-Informatics Division, Forest Research Institute, Dehradun, India
*Corresponding Author: nittinkumarverma@gmail.com

[Received 18/10/2014, Accepted-30/10/2014]

ABSTRACT:

Pyranose dehydrogenase (PDH) is a fungal flavin dependent sugar oxidoreductase which is extremely remarkable for appliance in organic synthesis or electrochemistry. It catalyzes the oxidation of free, non-phosphorylated sugar to the corresponding keto sugars. Therefore, it is important to study the sequence and structural feature of the enzyme shows a wide substrate specificity and intriguing variations in regioselectivity, depending on substrate, enzyme source and reaction environment. In protein databank reporesetory have no evidence of crystal structure of pyranose dehydrogenase. Hence, we developed the structural molecular model of the PDH protein from the *Agaricus campestris* by homology modeling using the server SWISS_MODEL. Based on the sequence similarity to proteins of known structure (template) of *Agaricus meleagris* (PDB ID: 4H7U) was selected from protein databank (PDB) by using DALTA-BLAST. The energy of constructing models was minimized and the qualities of the models were evaluated by PROCHECK, VERRIFY_3D. Resulted Ramachandran plot analysis showed that confirmation of 100.00% amino acid residues are within most favored regions. The protein-protein structure similarity was constructed on iPAB server, they showed 0.49 Å RMSD. Expasy's Prot-param server were used for physiochemical properties of protein. The structure is finally submitted in the Protein Model Database (PMDB).

Keyword: Homology Modeling, Pyanose Dehydrogenase, SWISS-MODEL, BLAST, PMDB

INTRODUCTION:

Pyranose dehydrogenase (EC 1.1.99.29) (PDH) is a glycosylated, extracellular, monomeric flavin-dependent sugar oxidoreductase secreted by several wood degrading fungi and a member of glucose-methanol-choline oxidoreductase family [1, 2]. It is a member of the glucose-methanole-choline (GMC) family together with other sugar oxidoreductase like the catalycically related enzyme glucose oxide, cellobiose dehydrogenase and pyranose-2 oxidase. The enzyme is not capable to use oxygen as an electron acceptor, using substituted benzoquinones and orange metal ion instead. PDH was first introduced in 1997 and isolated

from bacidomycete fungus *Agaricus bisporus* [3]. Later on, the enzyme from the other members of the family of *Agaricaceae* like *Macrelepiota rhacodes* [4], *A. xenthoderma* [5], and *A. meleagris* [1, 2] was investigated. Recently, the crystal structure of *A. meleagris* PDH (*Am*PDH) was determined a two domain structure consisting of the ADP-binding Rossman domain and a sugar-binding domain [6]. The biological function of PDH is still not completely clear. As the enzyme is limited to, litter-decomposing fungi of the family Agaricaceae and is not capable to use molecular oxygen as an electron acceptor, the reduction of quinines radicals formed during lignin degradation were proposed as its natural role [4]. The possible function like a participation in Fenton's reaction or the defense against antimicrobial (quinine) substances produced by plants was reported [7]. Production of PDH appears to be limited to a narrow group of fungi, *Agaricaceae* and *Lycoperdaceae*, which can be ecologically grouped as "litter-decomposing" fungi, growing on lignocelluloses-rich forest litter (leaves, barks, straw-like material [4]. The enzyme displays a wide substrate specificity and intriguing variations in regioselectivity [8]. Depending on the substrate and enzyme source, oxidation at C-1, C-2, C-3 or C-4, as well as double oxidation at C-1,2, C-2,3, or C-3,4 can occur yielding several reaction products that are inaccessible by chemical means or other enzymatic transformation. Similar to P2O [9], PDH from some sources such as the recently characterized enzyme from *Agaricus xanthoderma* [5], display a tendency for double oxidation of D-galactose to 2,3-didehydro-D-galactose. This is determinal for an application of PDH for chemoenzymatic production of D-tagatose via 2-dehydro-D-galactose.

Proteins are the linear chain of amino acids that approve unique three dimensional structure which allow them to carry out the intricate biological function. All of the information required to identify a protein's three dimensional structure is contained within its amino-acid sequences. Homology modeling or comparative modeling builds a three dimensional model of a protein of unknown structure (the target) on the basis of sequence similarity to proteins of known structures (the templates). The prediction process consist of fold assignment, target-template alignment, model building, and model evaluation [10].

2. MATERIAL AND METHODS

2.1. Sequence retrieval and analysis of the sequence

The full length (595) amino acid sequence of the protein (GENBANK ID: AHA85313.1) was retrieved from NCBI database. Primary structure analysis of amino acid sequences was performed *via* Protparam tool on ExPasy proteomics server and secondary structure was predicted by Jpred web server.

2.2. Molecular Homology Modeling

Homology modeling is the method when there is a well-defined relationship of homology between the sequences of a target protein and at least one experimentally definite three-dimensional structure. This computational technique is based on the assumptions that the tertiary structure of two proteins will be similar if their sequences were related, and it is the approach most likely to give an accurate result [11].

Sequence comprising the domain was used to build up the 3-D structures using the comparative protein modeling method of SWISS-MODEL [12]. To search the appropriate templates, DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST) [13] search tool was used against Protein Data Bank (PDB) (http://www.rcsb.org). If the retrieval accuracy and sensitivity towards protein analysis using normal BLAST is more than of DELTA-BLAST, in those cases, we used DELTA-BLAST. To assess the accuracy of template identification process, apart from DELTA-BLAST, pyranose dehydrogenase protein was subject to various

meta-servers like GeneSilico [14], Geno3D [15] and Pcons.net [16] in order to find reliable templates. After observation, target protein shows similarity with a template protein named pyranose dehydrogenase from *Agaricus meleagris* [PDB ID: 4H7U]. The target-template alignment was performed using ClustralW and edited with ESPript [21]. The alignment of the target-template is shown in fig. 1. SWISS-MODEL is an online server automated comparative protein 3D modeling. Total hundred models were constructed using the program SWISS-MODEL [12].

2.3. Analysis of the Model

We used modeller for construction of models and their respective models used for the further study. We used the model with the smallest Modeller Objective Function and PROCHECK statistics. The whole stereochemistry quality of the models pyranaose dehydrogenase protein using Ramachandran plot calculation computed with PROCHECK [17] program, available at NIH (National Institute of Health) server (http://nihserver.mbi.ucla.edu/SAVES_3/Prochec k/) and the final structure was later checked by VERIFY-3D graph accessible through NIH server (http://nihserver.mbi.ucla.edu/SAVES_3/Verify-3D/). The data of all preeminent five models are in Table 1.

2.4. Energy Minimization

The best model with the lowest DOPE score was subject to energy minimization by GROMACS 4.5.6 using minimization protocol. The minimization protocol employs the steepest descent method for the removal of bad van-der-Waals contacts from the model. In this study the calculations were done by utilizing GROMACS 4.5.6 [20] software in the G43a1 force field with a flexible SPS water model in a cubic box of 1.2 Angstrom dimension.

2.5. Quality assessment and validation

The quality, internal consistency and reliability of the energy minimized pyranose dehydrogenase protein were evaluated by a number of computational tools. PROCHECK [17] was used to check the stereo-chemical of the model which quantifies the residues in the available zones of the Ramachandran plot provide the position of the torsion angles phi *(φ)* and psi (ψ) between C_α-C and N-C_α atoms of the residues contained in a peptide. VERIFY_3D program determines the compatibility of the atomic model (3D) with its own amino acid sequences where a high VERIFY_3D profile [18]. Protein Structure Analysis (ProSA) was employed for the refinement and validation of the modeled structure which checks the native protein folding energy of the modeled by comparing the energy of the model with the potential mean force derived from large sets of known protein structures [19].

3. RESULTS AND DISCUSSION:

3.1. Sequence Analysis

Pyranose dehydrogenase protein (64,240.8 Kda) of *Agaricus campestris* has been extracted from Genbank. The primary sequence analysis of protein was calculated in Table-2 since the isoelectric point (pI), solubility is minimized and mobility in an electro focusing system is zero that's why calculated pI will be useful. Isoelectric point (pI) is the pH at which the surface of the protein is covered by the charge, but a net charge of the protein is zero. At pI, protein is stable and dense. For processing buffer system for purification by isoelectric focusing method, the computed isoelectric point (pI) will be valuable. While Expasy's Protparam computes the extinction coefficient of 276, 278, 279, 280 and 282 nm wavelengths, 280 nm has been elected since proteins absorb light strongly. Extinction coefficient of protein at 280 nm was 67380 M^{-1} cm^{-1}. The computed extinction coefficient can help in the quantitative study of protein-protein and protein-ligand interaction in solution. The instability index provides and determine of the stability of protein in a test-tube. There are

definite dipeptides, the occurrence of which is particularly divergent in the unstable protein compared with those in the stable once. This method assigned a weight value of instability, which is feasible to compute an instability index (II). A protein whose instability index is slighter than 40 is estimated as stable, a value above 40 estimates that the protein may be unstable. The instability index value of the protein was found 32.37, which indicate that, the protein is stable. The aliphatic index (AI) is elucidated as the relative volume of a protein occupied by aliphatic side chains (A, V, I and L) is estimated as a positive factor for the increase of the thermal stability of globular proteins. Aliphatic index (AI) for the protein sequence was 87.29. The very high aliphatic index of the protein sequence indicates that these proteins may be stable for a vast temperature range. The minimal thermal stability of protein was indicative of a more flexible structure when compared to other protein. The Grand Average hydropathy (GRAVY) value for a peptide or protein is calculated as the sum of hydropathy values of all the amino acids, divide by the number of residues in the sequence. A GRAVY index of protein was -0.184. This low value shows the probability of better interaction with water.

3.2. Comparative modelling and energy minimization

Comparative modeling of protein provides a significant hypothesis of homology between the target and template. This approach provides reasonable results based on the assumption that the tertiary structure of the two proteins will be similar if their sequences are related. Absence of the experimentally determined three dimensional structure of protein of *Agaricus campestris* in PDB (Protein Data Bank), comparative modeling method was utilized to construct its theoretical three dimensional structure. DELTA-BLAST scanning results had revealed more identical with crystallographic structure *Agaricus meleagris* (PDB ID: 4H7U at 1.60 Å resolution) while the template was determined on the basis of higher sequence identity. It has been 73 % sequence identity, with 507 conserved residues and 84% sequence similarity. Comparative modeling predicts the 3-D structure of hypothetical model of a given protein sequence (target), based primarily on this alignment to the template. The resulting 3-D structure of protein was sorted according to the scores calculated from discrete optimization, protein energy (DOPE) scoring function. The final model which has lowest root mean square deviation (RMSD), relative to the trace of the crystal structure was selected for further study (Fig. 2).

3.3. Model quality assessment

The detailed residue-by-residue stereo-chemical quality of the modelled protein structure was evaluated by the Ramachandran plot (Fig. 3) using Procheck tool. The reliability of the backbone torsion angle Φ and Ψ distribution of the protein and the template was evaluated by the Ramachandran plot in Procheck tool. The perceived Ramachandran plot (Phi-Phi) pairs had 86.6% residues in most favored regions, 13.8% core residues in additional allowed regions, 0.6% residues in generously allowed regions and 0.0% residues in disallowed regions Table-1. This value indicates a good quality model. Whereas the crystal structure of *Agaricus meleagris* PDB ID: 4H7U shows 86.60% residues in most favored regions. In order to characterize the model, structural motif and mechanically important loops were assigned to build a final 3D model of a protein. The packing quality of each residue of the model was assessed by Verify_3D program where the compatibility of the model residues with their environment is assessed by a score function. Residues with a score over 0.2 should be considered reliable. As shown in Table-2; the score of the refined model maximally was above 0.2 which corresponds to the acceptable side chain environment. PorSA revealed a Z-score of -11.38 for modelled protein,

where the template has a Z-score of -10.74 reflecting the overall quality of the model.

RMSD (root mean square distance) between the equivalent Cα atoms pair (target and template) was measured to check the degree of structural similarity. We examined the best modelled structure for fitting into the template (crystal structure), the prepare model and its closest relative was superimposed based on Cα and backbone atom pairs. A pairwise 3D alignment search of the template protein with the modelled structure through iPAB web server showed the identity of enormous 98.90% for 570 aligned residues (Fig. 4) with massive RMSD of 0.49 Å on their backbone atom. The iPAB web server's results conclude that protein and its structural homologues share strong structural conservation and similarity in the structural folding. It also signifies that the generated model is reasonably good for further studies.

4. CONCLUSIONS

In this study, we have evaluated both sequence and structural function of protein through homology modeling and comparative genomics approach. An aggregation of Bioinformatics tools, focused not only on sequence analysis, but also structural information, guided us to suggest the function of protein in *Agaricus campestris*. First of all, we determined their physiochemical characteristics of protein by different parameter like as isoelectric point, molecular weight, total number of positive and negative residue, extinction coefficient, instability index, aliphatic index and grand average hydropathy (GRAVY). Secondly, we have also constructed a 3D model of a protein using the comparative modeling approach. Computational analysis was carried out to distinguish its structural and impulsive features, which was an advance validated by the SAVES, WHAT IF. We also calculate the structure-structure alignment between best model constructed throughout swiss-model and template they have approximate 99% similarity they showed reliable structure. We have also submitted this structure into Protein Model Database (PMDB) and PMDB-ID is PM0079746. The presented model will be potentiate and facilitate structural and functional investigation to use structure-based drug designing.

ACKNOWLEDGEMENTS

We are thankful to Director FRI University, Dehradun and Uttaranchal College of Scieince and Technology, Dehradun for providing the necessary facilities to this pursue the above research.

REFERENCES:

[1] C. Sygmund, R. Kittl, J. Volc, P. Halada, E. Kubátová, D. Haltrich, et al., Characterization of pyranose dehydrogenase from Agaricus meleagris and its application in the C-2 specific conversion of D-galactose., J. Biotechnol. 133 (2008) 334–42. doi:10.1016/j.jbiotec.2007.10.013.

[2] R. Kittl, C. Sygmund, P. Halada, J. Volc, C. Divne, D. Haltrich, et al., Molecular cloning of three pyranose dehydrogenase-encoding genes from Agaricus meleagris and analysis of their expression by real-time RT-PCR., Curr. Genet. 53 (2008) 117–27. doi:10.1007/s00294-007-0171-9.

[3] E. Kubátová, D.A. Wood, Pyranose 2-dehydrogenase , a novel sugar oxidoreductase from the basidiomycete fungus Agaricus bisporus, (1997) 119–125.

[4] J. Volc, E. Kubátová, G. Daniel, P. Sedmera, D. Haltrich, Screening of basidiomycete fungi for the quinone-dependent sugar C-2/C-3 oxidoreductase, pyranose dehydrogenase, and properties of the enzyme from Macrolepiota rhacodes, Arch. Microbiol. 176 (2001) 178–186. doi:10.1007/s002030100308.

[5] M. Kujawa, J. Volc, P. Halada, P. Sedmera, C. Divne, C. Sygmund, et al., Properties of pyranose dehydrogenase purified from the litter-degrading fungus Agaricus xanthoderma., FEBS J. 274 (2007) 879–94. doi:10.1111/j.1742-4658.2007.05634.x.

[6] M.M.H. Graf, U. Bren, D. Haltrich, C. Oostenbrink, Molecular dynamics simulations give insight into D-glucose dioxidation at C2 and

C3 by Agaricus meleagris pyranose dehydrogenase., J. Comput. Aided. Mol. Des. 27 (2013) 295–304. doi:10.1007/s10822-013-9645-7.

[7] C.K. Peterbauer, J. Volc, Pyranose dehydrogenases: biochemical features and perspectives of technological applications., Appl. Microbiol. Biotechnol. 85 (2010) 837–48. doi:10.1007/s00253-009-2226-y.

[8] P. Sedmera, P. Halada, C. Peterbauer, J. Volc, A new enzyme catalysis: 3,4-dioxidation of some aryl β-d-glycopyranosides by fungal pyranose dehydrogenase, Tetrahedron Lett. 45 (2004) 8677–8680. doi:10.1016/j.tetlet.2004.09.149.

[9] J. Volc, P. Sedmera, M. Kujawa, P. Halada, E. Kubátová, D. Haltrich, Conversion of lactose to β-d-galactopyranosyl-(1 → 4)-d-arabino-hexos-2-ulose-(2-dehydrolactose) and lactobiono-1,5-lactone by fungal pyranose dehydrogenase, J. Mol. Catal. B Enzym. 30 (2004) 177–184. doi:10.1016/j.molcatb.2004.05.006.

[10] M.A. Martí-Renom, A.C. Stuart, A. Fiser, R. Sánchez, F. Melo, A. Sali, Comparative protein structure modeling of genes and genomes., Annu. Rev. Biophys. Biomol. Struct. 29 (2000) 291–325. doi:10.1146/annurev.biophys.29.1.291.

[11] N.K. Verma, B. Singh, Insight from the structural molecular model of cytidylate kinase from Mycobacterium tuberculosis., Bioinformation. 9 (2013) 680–4. doi:10.6026/97320630009680.

[12] N. Guex, M.C. Peitsch, T. Schwede, SWISS-MODEL□: an automated protein homology-modeling server, 31 (2003) 3381–3385. doi:10.1093/nar/gkg520.

[13] G.M. Boratyn, A.A. Schäffer, R. Agarwala, S.F. Altschul, D.J. Lipman, T.L. Madden, Domain enhanced lookup time accelerated BLAST, Biol. Direct. 7 (2012) 12.

[14] M.A. Kurowski, J.M. Bujnicki, GeneSilico protein structure prediction meta-server., Nucleic Acids Res. 31 (2003) 3305–3307.

[15] C. Combet, M. Jambon, G. Deléage, C. Geourjon, Geno3D: automatic comparative molecular modelling of protein., Bioinformatics. 18 (2002) 213–214.

[16] J. Lundström, L. Rychlewski, J. Bujnicki, A. Elofsson, Pcons: a neural-network-based consensus predictor that improves fold recognition., Protein Sci. 10 (2001) 2354–2362.

[17] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, J. Appl. Crystallogr. 26 (1993) 283–291

[18] D. Eisenberg, R. Luthy, J.U. Bowie, VERIFY3D: assessment of protein models with three-dimensional profiles, Methods Enzymol. 277 (1997) 396-404.

[19] M. Wiederstein, M.J. Sippl, PorSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, Nucleic Acids Res. 35 (2007) W407-W410.

[20] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, H.J.C. Berendsen, GROMACS: fast, flexible, and free., J. Comput. Chem. 26 (2005) 1701–1718.

[21] P. Gouet, E. Courcelle, D.I. Stuart, F. Métoz, BIOINFORMATICS ESPript□: analysis of multiple sequence alignments in PostScript, 15 (2000) 305–308.

Table 1: Comparison of Ramachandran plot statistics of PDH model with closest homologue structure 4H7U

Ramachandran Plot Statistics	Modeled structure of PDH		Template (4H7U)	
	Residue	Percentage (%)	Residue	Percentage (%)
Residue in most favored region	423	86.6	441	89.5
Residue in additional allowed region	68	13.8	49	9.9
Residues in generously allowed region	3	0.6	3	0.6
Residue in disallowed region	0	0.0	0	0.0
Number of non-glycine and non-proline residues	494	100	493	100
Number of end residues	2	-	2	-

Number of glycine residues	45	-	49	-
Number of proline residues	33	-	33	-

Table 2: Comparison of model validation scores from different server between modeled PDH and its closest structural homologue 4H7U (A chain).

Target/ template	Verify_3D	Errat	ProSA
PDH	96.52%	83.68	-11.38
4H7U	98.27%	94.89	-10.74

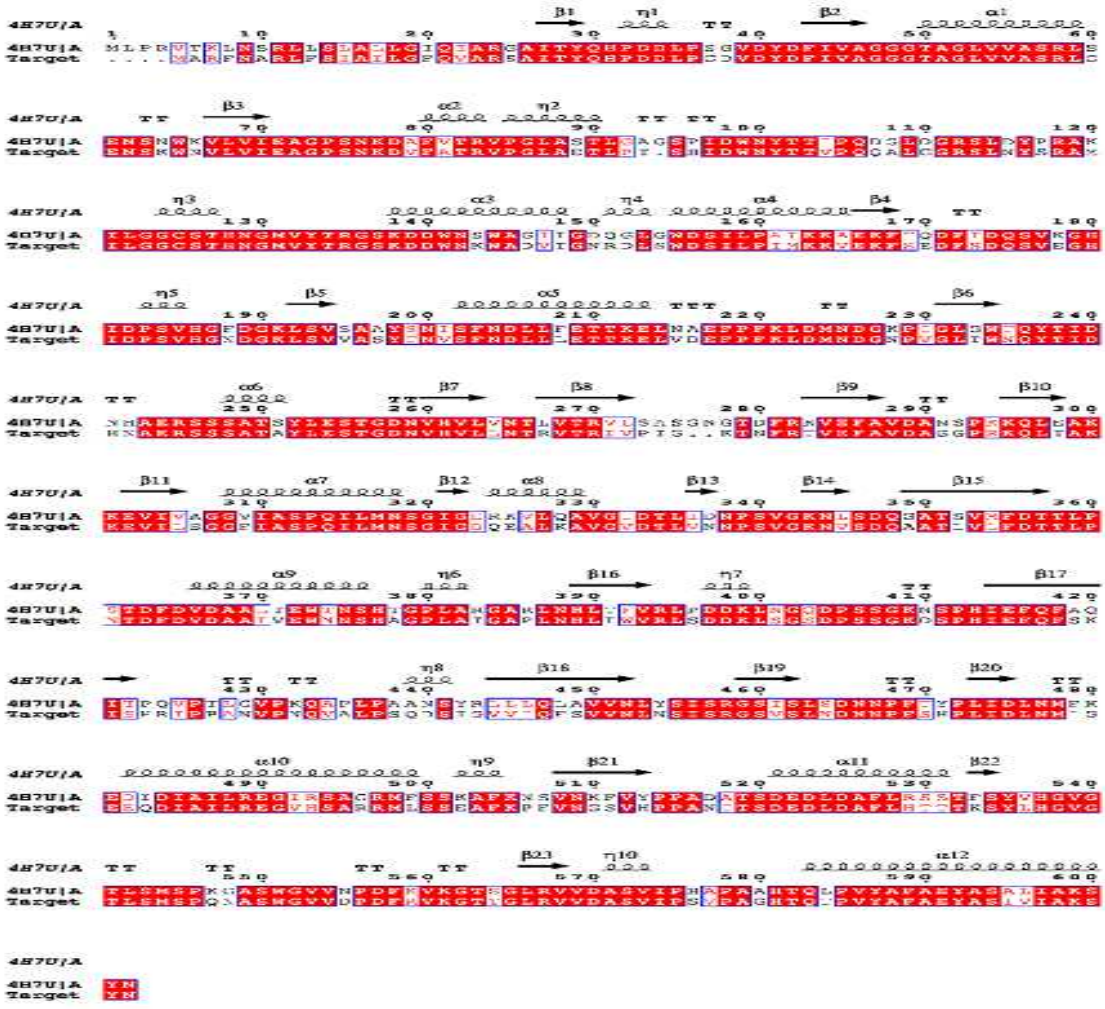

Figure 1: The pairwise sequence alignment of the target Pyrenees dehydrogenase and template 1H6A were constructed using ClustralX and ESPript.

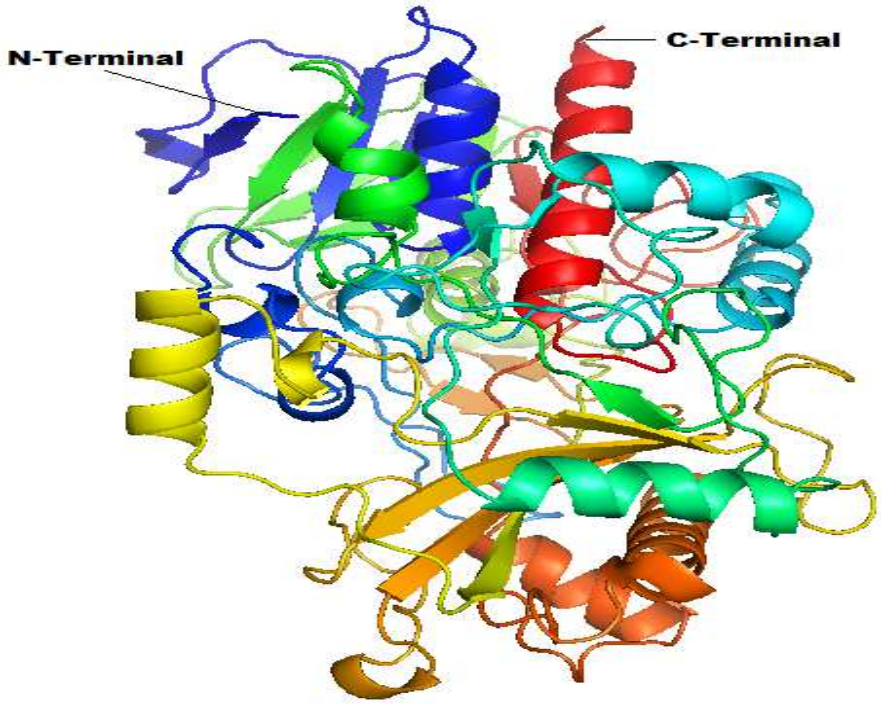

Figure 2: Solid ribbon representation of pyranose dehydrogenase model (GI: AHA85313.1) from *Agaricus campestris* constructed through homology modeling.

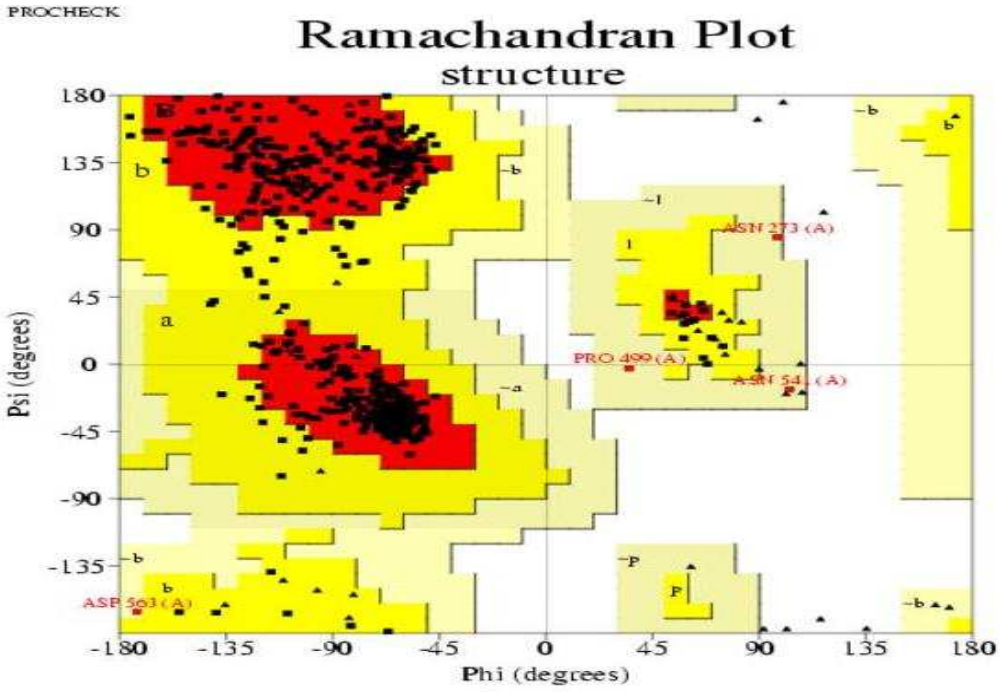

Figure 3: Ramachandran plot of the modelled pyranose dehydrogenase protein (GI: AHA85313.1). The plot was calculated by PROCHECK program.

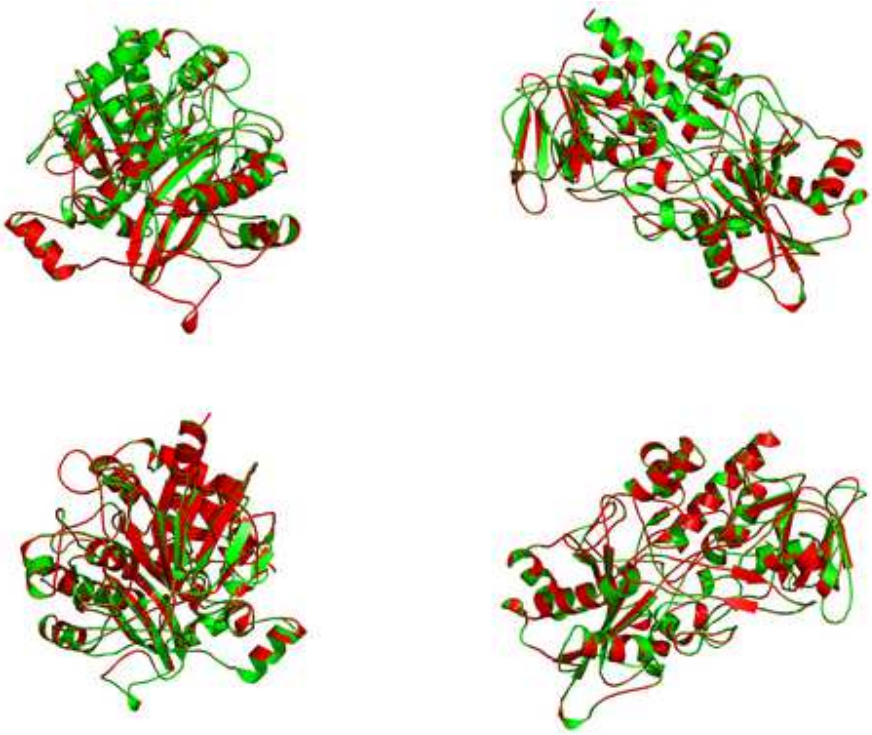

Figure 4: Superimposition of the Target sequence (GI: AHA85313.1) and with Template (PDB-ID: 1H6A) by iPAB web server.