Contents lists available at ScienceDirect

Asian Pacific Journal of Tropical Biomedicine

journal homepage:www.elsevier.com/locate/apjtb

Document heading

Novel drug target identification on UDP-Glucose 4-epimerase enzyme in *Catharanthus roseus* by insilico model

Ramachandran M*, Elumalai EK

P.G. and Research Department of Zoology, Voorhees College, Vellore-632001(T.N), India

ARTICLE INFO

Article history:
Received 17 June 2012
Received in revised from 5 July 2012
Accepted 22 August 2012
Available online 28 August 2012

Keywords:
UDP
Model building
Validation
Hydropathicity
Drug target
Model validations

ABSTRACT

Objective: To investigate the definite crystal structure of UDP-glucose 4-epimerase enzymes (EC 5.1.3.2) from *Catharanthus roseus* (*C. roseus*) for further research activities. **Methods:** The structure was modeled using homologous templates. The model validated under PROCHECK and WHAT-IF. **Results:** The model constructed using Modeller9v7 was validated. Moreover, 89% of residues lie in the most favored region. The model was checked for its grand average of hydropathicity and three binding sites were predicted using Molsoft ICM Pro v3.5. **Conclusions:** The model was suggested to be the good model. The constructed model can be used for further pharmacological studies and it can act as potential target against novel inhibitors.

1. Introduction

UDP-glucose/galactose 4-epimerase (named UGE, GalE and Gal10 in various systems) catalyzes the synthesis of UDP-galactopyranose from UDP-glucose. This is a step in the Leloir pathway of galactose metabolism that is highly conserved in species ranging from bacteria to humans[1-3]. D-galactose is widely distributed in combined form in plants, animals and microorganisms as a constituent of oligo- and polysaccharides; it also occurs in galactolipids and as its glucoside. UDP-glucose 4-epimerase requires catalytic amounts of NAD and is inhibited by NADH[4]. This family of proteins utilizes NAD as a cofactor. It contains the NAD (P) - binding domain which is a commonly found domain with a core Rossmann-type fold. The UDP-galactose 4′-epimerase (GalE) catalyzes interconversion of UDP-galactose to UDP-glucose and plays a key role in lipopolysaccharide biosynthesis. This makes it an important virulence determinant, and therefore a potential drug target. This uniqueness necessitated elucidation of its structure and active site[5]. Current therapeutics are inadequate due to toxic side effects, drug resistance, and limited effectiveness, novel therapies are urgently needed. UDP-galactose 4′-epimerase (TbGalE), an enzyme of the Leloir pathway of galactose metabolism, is one promising *Trypanosoma brucei (T. brucei)* drug target[6]. GALE therefore plays key roles in the metabolism of dietary galactose, in the production of endogenous galactose, and in maintaining the ratios of key substrates for glycoprotein and glycolipid biosynthesis[7].

Enzyme over expression in *Thermus thermophilus* (*T. thermophilus)* HB27 led to an increased capacity of biofilm production. Therefore, the galE gene is important to biofilm formation because of its involvement in epimerizing UDP-galactose and UDP-N-acetylgalactosamine for exopolysaccharide biosynthesis[8]. Type III galactosaemia is a hereditary disease caused by reduced activity in the Leloir pathway enzyme, UDP-galactose 4′-epimerase (GALE). Three mutations associated with this intermediate form (S81R, T150M and P293L) were analyzed for their kinetic and structural properties in vitro and their effects on galactose-sensitivity of *Saccharomyces cerevisiae* (*S. cerevisiae*) cells that were deleted for the yeast GALE

*Corresponding author: Mr. M. Ramachandran, PG and Research Department of Zoology, Voorhees College, Vellore-632001(T.N), India.
Tel: 09933238757
E-mail: insilicobrain@gmail.com

homologue Gal10p. All three mutations result in impairment of the kinetic parameters (principally the turnover number, k (cat)) compared with the wild-type enzyme. However, the degree of impairment was mild compared with that seen with the mutation (V94M) associated with the generalized form of epimerase deficiency galactosaemia. None of the three mutations tested affected the ability of the protein to dimerize in solution or its susceptibility to limited proteolysis *in vitro*[9].

Catharanthus roseus (*C. roseus*) (Apocynaceae) a traditionally used medicinal plants. It is an erected procumbent herb or under shrub containing latex. It is widely growing to 1m tall at subtropical area. It possesses known antibacterial, antifungal, antibiotic, antioxidant, anticancer and antiviral activities[10]. The extracts have demonstrated significant anticancer activity against numerous cell types[11]. However, a 3D structure for *C. roseus* UDP-Glucose 4-Epimerase enzyme is not found in Protein Data Bank (PDB). Moreover constructed homology model of UDP-glucose 4-epimerase enzyme using known structural templates and define its structural features to understand molecular function.

2. Meterials and methods

2.1. Sequence retrieval

Protein sequence of UDP-Glucose 4-epimerase enzyme of *C. roseus* was searched against Uniprot database. The query resulted in a sequence with accession number A5YKE8 with a length of 248 amino acid residues[3].

2.2. Target sequence

Target sequence of *C. roseus* UDP-Glucose 4-epimerase enzyme with Accession Number (A5YKE8) was retrived from EXPASY in FASTA format. The details of Sequence Accession Nunber (A5YKE8) of UDP-Glucose 4-epimerase enzyme of *C. roseus* (Rosy periwinkle, Madagascar periwinkle)- Length was 248 aa.

2.3. Homology search

Retrieved protein sequence was found against Protein Data Bank using BLAST (Basic Local Alignment Search Tool). Blastp was performed which result in nearly five homologous sequences[12]. The best out of the five were selected with 64.77% identities.

2.4. Template sequence

Template sequence for the query was found from NCBI-BLASTp and downloaded in the FASTA format[13] and PDB: 1HZJ as well as sequence length 348 aa.

2.5. Model construction

MODELLER9v7[14] was used for homology modeling of UDP-Glucose 4-epimerase structure. After one hundred models of the UDP-Glucose 4-epimerase was generated, the model was picked by a combination of the MODELLER9v7 objective function value and the Discrete Optimized Protein Energy (DOPE) statistical potential score was -7,411[15].

2.6. Structural alignment and homologies

The differences between two protein structures were expressed by the root mean square deviation (RMSD) of the respective atomic position in the two structures.

$$RMS = \sqrt{\frac{\sum_i d_i}{N}}$$

di is the distance between two corresponding atoms and N is the number of considered atoms. Mostly the differences of the positions of the respective C atoms of the backbones in the two structures were measured[16]. RMS value upto 1Ao show a high similarity between the two structures. The protein structure was described in Cartesian coordinates and each structure had a build-in orientation in its proper coordinate system. To compare the two structures, one structure serves as a reference and the other structure must be superimposed[17,18]. This allows the evaluation of which parts of the structure was showing a good or high similarity and where significant structural deviations were located allowing to examine the patterns of structural conservation and change within a protein family.

2.7. Model validation

Constructed protein model was validated at the next level. The protein structure was validated for errors and the quality of the structure was determined. The quality of the modeled protein depends on how the template structure chosen with high quality data. This validation was done using PROCHECK[19] and WHAT-IF[20]. The conformational localization of Alpha helix and beta sheets was represented by Ramachandran Plot.

2.8. Binding site prediction

The binding site was predicted using MolSoft ICM Pro v3.5. The software clearly picturized binding sites in UDP Glucose 4-epimerase enzyme. This study was carried out which could be used for further research involving the design of

potential inhibitors against this enzyme

3. Results

3.1. Sequence alignment

C. roseus UDP Glucose 4-epimerase enzyme catalyses the reversible epimerization of UDP-glucose and UDP-galactose Hence, it is important to establish the structure-function relationship for *C. roseus* UDP glucose 4-epimerase enzyme. The target protein sequence (248 residues long) for UDP glucose 4-epimerase from *C. roseus* was retrieved from EXPASY in FASTA format [21]. The basic local alignment search tool (BLAST) was performed against Protein Data Bank for homologous structures. UDP-galactose 4-epimerase from human was identified as best template structure. PDB code: 1HZJ_A with 1.50 Å resolutions with 327 maximum score, Query coverage (99%) and E-Value (2e-90) was better obtained with alignment 64.77% sequence identity (Figure 1). Moreover, the 3D structure of protein described 1HZJ_A as a template for model building.

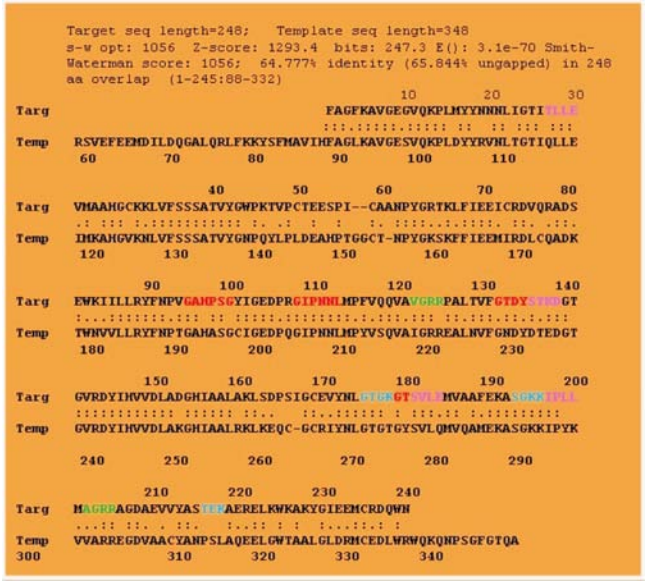

Figure 1. Pairwise Alignment between target and template sequences. Rose color sequence indicating Casein kinase II phosphorylation site, Red color N-myristoylation site, Green color Amidation site and Blue color Protein kinase C phosphorylation site was found in the target sequence.

3.2. Model construction

The template structural model PDB ID: 1HZJ_A [13], the sequence similarity was 64.77% with the template and reliability of the predicted model thus generated using MODELLER9v7 software. The discrete optimized protein energy (DOPE) scores -25321.44 and RMSD value 0.006 was predicted[14]. The consequence of model development towards establishing the specific function of using predicted model, Compositions of modeled protein UDP-Glucose 4-Epimerase Enzyme sequence Ala =24 (9.7%), Cys=6 (2.4%), Asp = 11 (4.4%), Glu=17 (6.9%), Phe=8 (3.2%), Gly=24 (9.7%), His=4 (1.6%), Ile=14 (5.6%),Lys = 17 (6.9%), Leu =18 (7.3%), Met =6 (2.4%), Asp = 11 (4.4%), Pro=13 (5.2%), Gln=5 (5.2%), Arg =12 (4.8%), Ser =12 (4.8%), Thr=13 (5.2%), Val=20 (8.1%), Trp=4 (1.6%) and Tyr=11 (4.4%) was computed using the tool BioEdit v7.0.

3.3. Model validation

The constructed model of UDP-Glucose 4-Epimerase enzyme (Figure 2) from *C. roseus* was examined and validated using different criteria. The stereochemical quality of the predicted model was evaluated using PROCHECK. The Ramachandran plot of phi/psi distribution in the model was determined using PROCHECK for checking non-Gly residues at the disallowed regions.

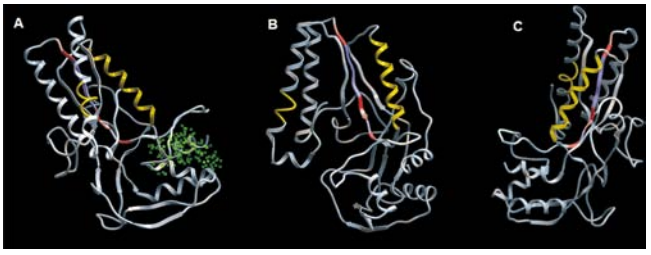

Figure 2. Modeled structure of UDP-Glucose 4-epimerase of *C. roseus*.

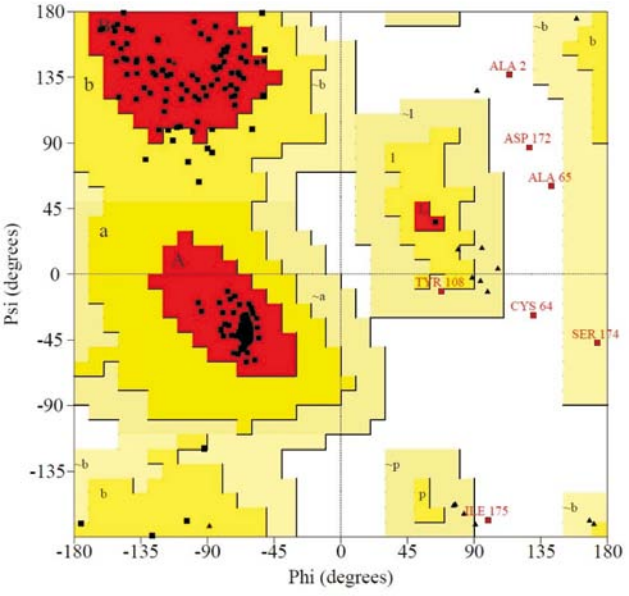

Figure 3. Ramachandran Plot amino acids clustering determinations of the modeled protein structure.

The modeled protein was validated under PROCHECK and obtained a Ramachandran Plot (Figure 3). The 89% of residues were found to be localized by the core right handed alpha helices (A), beta sheets (B) and left-handed alpha helix (L). The 7.7% of residues were found to be localized with allowed right-handed alpha helix (a), beta sheets (b) and left handed alpha helix. The 1.0% of the residues were in the generous alpha helices (~a), beta sheets (~b), left-handed alpha helices (~l) and epsilon (~p). The 2.4% of the

residues was found to be localized at the disallowed regions. The residues with the generously allowed region can be treated as disallowed regions. The triangles present in the plot glycine residues, which was 24 in number.

Standard bond lengths and bond angles of the model were determined using WHAT-IF, the predictions were revealed RMS Z-scores for bond lengths 6.989, bond angles 1.285, Omega angle restraints 1.737, side chain planarity 1.291, improper dihedral distribution 5.852 and inside-outside distribution 1.108. The values are almost equal to the suggested high model quality. Template structure, Modeled structure and superimpose secondary structures (Figure 4) were generated, compared and visualized by Molsoft ICMpro Program.

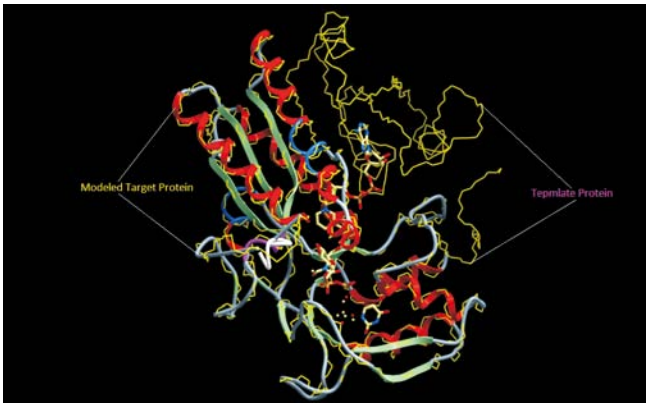

Figure 4. Superposed structure of template and the modeled protein structure.

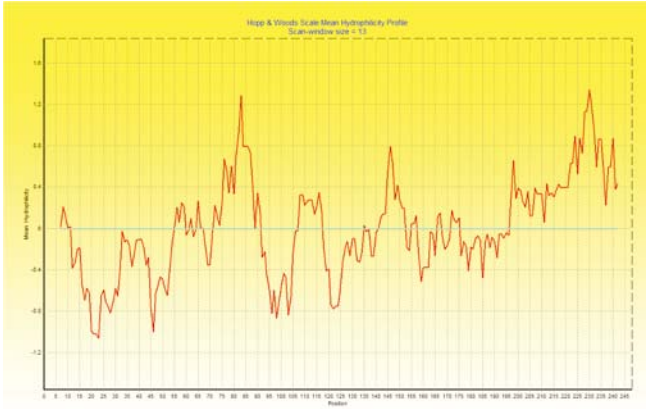

Figure 5. Hydrophilicity profile of the Modeled protein structure and position is 227(Ala) residue & mean hydrophilicity 0.8.

3.4. Physiochemical characterization

The Physio chemical parameters, such as theoretical Isoelectrical point (pI)-7.73, Molecular weight(M.Wt)-, total number of positive residue (Arg, Lys)-29, negative residues (Asp,Glu)-28, Extinction Coefficient(EC)-38765, Half life, Instability Index-31.88[22,23] and Grand average of hydropathicity (GRAVY)-0.160 were computed using Expasy proteomics server and ProtParam prediction server[21]. Hydrophobicity and Hydrophilicity plots were predicted by BioEdit Software v7.0 to modeled protein structure for hydrophilcity environment maximum peak redidue (Ala) position was 227 and Mean value-0.8 in the Hopp & Woods Scale Mean Hydrophilicity profile plot (Figure 5). Hydrophobicity environment maximum peak residue(Gly) Position was 36 and Mean value-0.9 in the kyte & Doolittle Scale Mean Hydrophilicity profile plot (Figure 6) .

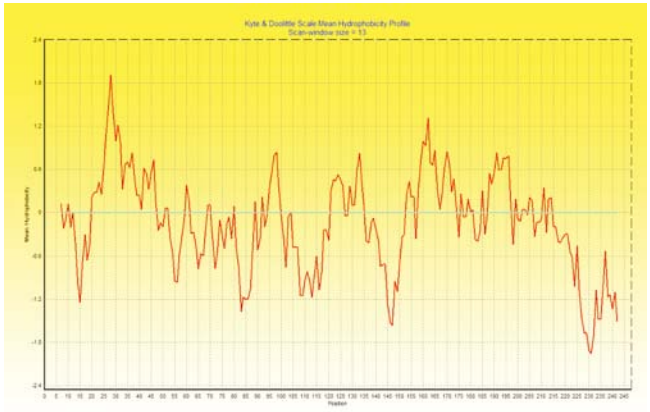

Figure 6. Hydrophobicity profile of the Modeled protein structure and Position=36(Gly) residue Mean value=0.9.

3.5. Drug target detail

Three binding sites were predicted in the modeled protein of UDP-glucose 4-epimerase enzyme by MolSoft ICM Pro 3.5 software. First binding site responsibility amino acid (ILE-109), second binding site (PHE-1) and third binding site (GLN-84) respectively (Figure 7).

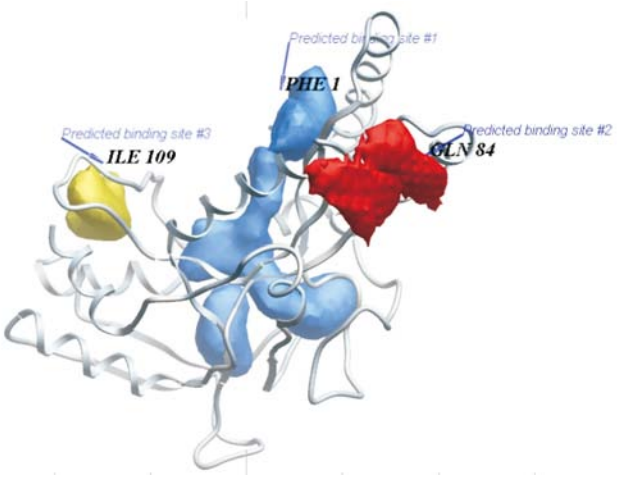

Figure 7. Modeled UDP-Glucose 4-Epimerase enzyme binding sites.

4. Discussion

In present study, the best homology 3D structure of protein model was generated and validated of *C. roseus* UDP-glucose 4-epimerase enzyme using MODELLER9v7 software and obtained a refined model after energy minimization. Because, three-dimensional structures not available in protein data bank until now. The structure of *C. roseus* UDP-glucose 4-epimerase enzyme very essential for establishing and detect molecular function. The modeler homology-

modeling algorithms were demonstrated excellent accuracy in blind predictions. Modeler models built with as little as 35% identities can be accurate enough to be successfully utilized drug target identifications and drug docking. Moreover, the sequence similarity of *C. roseus* UDP-glucose 4-epimerase enzyme and identity was 64.77% with the template sequence alignment. The discrete optimized protein energy (DOPE) scores-25 321.44 and RMSD value was 0.006. The final refined model was assessed by PROCHECK, WHAT-IF program and Prot Param. Totally three-ligand-binding sites were predicted. The molecular structure of the potential active site and responsible residues first site (ILE-109), second site (GLN-84) and third site (PHE-1) were predicted by the Molsoft ICM Pro software. The result shows that the model was stable and reliable. The stable model may further be used for virtual drug docking. The ligand binding sites in the predicted model will provide valuable insights towards inhibitors design and developments.

Conflict of interest statement

We declare that we have no conflict of interest.

References

[1] Allard ST, Giraud MF, Naismith JH. Epimerases: structure, function and mechanism. *Cell Mol Life Sci* 2001; **58**: 1650-1665.

[2] Amira M. El-Ganiny, Inder Sheoran, David A.R. Sanders, Susan G.W. Kaminskyj. Aspergillus nidulans UDP-glucose-4-epimerase UgeA has multiple roles in wall architecture, hyphal morphogenesis, and asexual development. *Fungal Genet Biol* 2010; **47**(7): 629-635.

[3] Zu Y, Wang M. Characteristics of UDP-glucose 4-epimerase in *C. roseus*. EMBL/GenBank/DDBJ databases 2007.

[4] Qisen Zhang, Maria Hrmova, Neil J Shirley, Jelle Lahnstein, Geoffrey B Fincher. Gene expression patterns and catalytic properties of UDP-D-glucose 4-epimerases from barley (*Hordeum vulgare* L.). *Biochem J* 2006; **394**: 115-124.

[5] Agarwal S, Mishra N, Agarwal S, Dixit A. Characterization of the active site and coenzyme binding pocket of the monomeric UDP- galactose 4′- epimerase of *Aeromonas hydrophila*. *BMB Rep* 2010; **43**: 419-426.

[6] Durrant JD, Urbaniak MD, Ferguson MA, McCammon JA. Computer-aided identification of *Trypanosoma brucei* uridine diphosphate galactose 4′-epimerase inhibitors: toward the development of novel therapies for African sleeping sickness. *J Med Chem* 2010; **8**: 5025-5032.

[7] Sanders RD, Sefton JM, Moberg KH, Fridovich-Keil JL. UDP-galactose 4′ epimerase (GALE) is essential for development of *Drosophila melanogaster*. *Dis Model Mech* 2010; **3**: 628-638.

[8] Niou YK, Wu WL, Lin LC, Yu MS, Shu HY, Yang HH, et al. Role of galE on biofilm formation by Thermus spp. *Biochem Biophys Res Commun* 2009; **390**: 313-318.

[9] Chhay JS, Vargas CA, McCorvie TJ, Fridovich-Keil JL, Timson DJ. Analysis of UDP-galactose 4′-epimerase mutations associated with the intermediate form of type III galactosaemia. *J Inherit Metab Dis* 2008; **31**: 108-116.

[10] Jaleel CA, Gopi R, Paneerselvam R. Alterations in non-enzymatic antioxidant components of *C. roseus* exposed to paclobutrazol, gibberellic acid and *Pseudomonas fluorescens*. *Plant Omics J* 2009; **2**:30-40.

[11] Jayanthi M, Sowbala N, Rajalakshmi G, Kanagavalli, Sivakumar V. Study of antihyerglycemic effect of *Catharanthus roseus* in alloxan induced diabetic rats. *Int J Pharm Pharm Sci* 2010; **2**(4): 114-116.

[12] Taliah Mittler, Marcel Levy, Chad Feller, Karen Schlauch. MULTBLAST: A web application for multiple BLAST searches. *Bioinformation* 2010; **5**(5): 224-226.

[13] Thoden JB, Wohlers TM, Fridovich-Keil JL, Holden HM. Human UDP-galactose 4-epimerase. Accommodation of UDP-N-acetylglucosamine within the active site. *J Biol Chem* 2001; **276**: 15131-15136.

[14] Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, et al. A comparative protein structure modeling using MODELLER. *Curr Protoc Protein Sci* 2007 Chapter 2: Unit 2.9.

[15] Fernandez-Recio J, Totrov M, Skorodumov C, Abagyan R. Optimal docking area: A new method for predicting protein-protein interaction sites. *Proteins* 2005; **57**: 400-413.

[16] Lazaros Mavridis. 3D-Blast: 3D Protein structure alignment, comparison, and classification using spherical polar fourier correlations. *Pac symp biocomput* 2010; **15**: 281-292.

[17] Sippl M, Wiederstein M. A note on difficult structure alignment problems. *Bioinformatics* 2008; **24**(3): 426-427.

[18] Gabriele Mayr, Francisco S Domingues, Peter Lackner. Comparative analysis of protein structure alignments. *BMC Struct Biol* 2007; **7**: 50.

[19] Archna Sahay, Madhvi Shakya. Structure prediction of dihydroflavonol 4-reductase and anthocyanidin synthase from Spinach. *Bioinformation* 2010; **5**(6): 259-263.

[20] Hekkelman ML, Te Beek TA, Pettifer SR, Thorne D, Attwood TK Vriend G. WIWS: a protein structure bioinformatics Web service collection. *Nucleic Acids Res* 2010; **38**(2): 719-723.

[21] Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, et al. Protein identification and analysis tools on the ExPASy server. In: John M, Walker. (ed). *The Proteomics Protocols Handbook*. New York: Humana Press; 2005, p. 571-607.

[22] Archna Sahay, Madhvi Shakya. In silico analysis and homology modelling of antioxidant proteins of Spinach. *J Proteomics Bioinf* 2010; **3**: 148-154.

[23] Ikai AJ. Thermostability and aliphatic index of globular proteins. *J Biochem* 1980; **88**: 1895-1898.