INNSPUB

International Journal of Biosciences | IJB |
ISSN: 2220-6655 (Print), 2222-5234 (Online)
http://www.innspub.net
Vol. 10, No. 1, p. 24-41, 2017

RESEARCH PAPER OPEN ACCESS

# Identification of disordered regions and potential active sites from3-hydroxy-3-methylglutaryl-CoA reductase of *Triticum aestivum* L. using theoretical approach

**Utpal Kumar Adhikari, Ferozur Rahman, Mostaq Ahmmed, Razib Chowdhury, M. Mizanur Rahman***

*Department of Biotechnology and Genetic Engineering, Islamic University, Kushtia, Bangladesh*

**Key words:** HMG-CoA reductase, Phosphorylation sites, Disordered regions, Active sites, Comparative modeling

http://dx.doi.org/10.12692/ijb/10.1.24-41 Article published on January 15, 2017

## Abstract

3-hydroxy-3-methylglutaryl CoA reductase is considered as an essential enzyme due to its inevitability in the Mevalonate pathway for the synthesis of isoprenoids in plants. In this study, theoretical investigations were accomplished for comparative protein model and active site analyses of *Triticum aestivum* HMG-CoA reductase (designated as *Ta*HMGR) as there are no three-dimensional structures available for this species in Protein Data Bank. So, to fulfill the necessity of this structure we built the comparative protein model, evaluated using different criteria and finally deposited in Protein Model Database (PMDB). We used different Bioinformatics tools and servers to carry out this research. The selected enzyme contains disordered regions and the residues Glycine, Serine, Lysine, and Proline are mainly responsible for these regions as found by our research. We also found 3 ligand binding sites with the high quantity of Glycine, Valine, and Alanine residues in their binding sites and the significant Z-score ensures these findings. The biochemical function was found as a catalytic and binding activity with the significance score, which confirms the actual function of *Ta*HMGR. The attained data convey necessary fundamental information about this enzyme to pave the way in improving the structure-based drug development using *T. aestivum* species.

* **Corresponding Author:** M. Mizanur Rahman ✉ mmrahmanbtg79@hotmail.com

## Introduction

The HMG-CoA reductase is an important enzyme for its catalyzing activity in the Mevalonate pathway. In plants, the first important step of the Mevalonate pathway is the conversion of HMG-CoA to Mevalonate (Ha *et al*., 2003). This pathway includes different superior branch pathways in comparison with other organisms (Leivar *et al*., 2005). The regulation and the level of Wheat HMG-CoA reductase genes have been trying to explore due to the potential uses of their molecular components. Wheat HMG-CoA reductase genes showed tissue specific gene expression and developmental gene regulation (Aoyagi *et al*., 1993).

The activity of HMG-CoA reductase in plant species is regulated by the level of its mRNA. This regulation can be varied for the different plant HMG-CoA reductase gene family. A large number of HMG-CoA reductase gene families have been reported in wheat (*T. aestivum*) and it includes about three genes namely HMG-CoA reductase 6, 10, 18 (Aoyagi *et al*., 1993). Human HMG-CoA reductase gene was isolated from human fetal adrenal cDNA library (Luskey and Stevens, 1985).The HMG-CoA reductase includes the homogenization strategies (Campos *et al*., 2014), alternative splicing *(Stormo et al*., 2012), RT-PCR (Reverse Transcription polymerase chain reaction) and RACE (Rapid Amplification cDNA Ends) (Zhang *et al*., 2014) as the detection methods. In some species this enzyme unveils double coenzyme specificity (Theivagt *et al*., 2006). The inhibition in the functional site of the protein with the reverse phosphorylation process can control plant HMG-CoA reductase. Protein degradation and calcium element accomplish this inhibition as a post-translational modification (Korth *et al*., 2000). The transcription and translation process of HMG-CoA reductase increase with decreasing the product concentrations of the Mevalonate pathway (Istvan, 2000). On the other hand, the HMG-CoA reductase concentration decreases quickly with increasing the sterol concentrations (Nishikawa *et al*., 1994).

The Bioinformatics approach on HMG-CoA reductase sequences for molecular cloning, multiple sequence alignment, conserved motif analysis, phylogenetic tree construction and trans-membrane helix prediction have already endeavored (Friesen and Rodwell, 2004; Chen *et al.,* 2012; Darabi *et al.,* 2012; Darabi and Seddigh, 2013; Andrade-Pavón *et al.,* 2014; Li *et al*., 2014; Adhikari *et al.,* 2015). But Bioinformatics approaches have not implemented on *Ta*HMGR for regulatory site prediction and functional analysis. There are no protein models of *Ta*HMGR present in PDB (Protein Database) and comparative protein model database PMDB (Protein Model Database). So, there is no way to perform further research on this *Ta*HMGR. The protein structural information provide necessary unknown function, design and improving ligands, studying the catalytic mechanism and also the functional relationship of structural similarity that can easily be done using the *in silico* approach. In general, the three dimensional structure of protein is determined by experimental methods like NMR and X-ray crystallography, which are quite loathsome, highly expensive and time consuming in comparison with the *in silico* modeling approach. In that case, the objectives of our study were to build a comparative protein model and to know some important aspects of this enzyme using *in silico* approach and finally deposition to the PMDB database.

## Materials and methods

*Retrieval of target sequence*

In this study, the protein sequence of *T. aestivum* encoding HMG-CoA reductase (*Ta*HMGR) was retrieved in Fasta format from the UniProt database (www.uniprot.org/uniprot) with accession number W5FJD4 for *in silico* analysis.

*Physico-chemical characters and secondary structure prediction*

The essential and non-essential amino acid compositions and Physico-chemical properties were generated by the ExPASyProtParam server (www.web.expasy.org/protparam) (Gasteiger *et al*., 2005).

It showed the total number of amino acids, PI value (Bjellqvist *et al.*, 1993), molecular weight, total number of negatively and positively charged residues, instability index (II) (Guruprasad *et al.*, 1990), extinction coefficient (EC) (Gill and von Hippel, 1989), aliphatic index (AI) (Ikai, 1980), half-life and GRAVY (grand average hydropathy) (Kyte and Doolittle, 1982) value as the significant physico-chemical properties which are important to characterize a protein model. The Secondary structures with associated features were predicted by PSIPRED server (McGuffin *et al.*, 2000).

*Phosphorylation and intrinsic protein disorder regions identification*

The phosphorylation sites of the protein were predicted by submitting the *Ta*HMGR protein sequence to the Net Phos 2.0 Server (www.cbs.dtu.dk/services/NetPhos). The NetPhos 2.0 server completed its task based on the artificial neural network method. This is widely used method for biological sequence analysis (Blom *et al.*, 1999). The intrinsic protein disorder or unstructured region, domain and globularity region were predicted using the GlobPlot version 2.3 which is mainly common gateway interface (CGI) based web server (www.globplot.embl.de) (Linding *et al.*, 2003). The disordered, domain and globularity regions were predicted by setting the default propensities, disorder prediction, smoothing, globular domain hunting and plot setting parameters.

*Evolutionary relationships prediction*

The sequence homology was identified using NCBI Blastp 2.2.30+ (Altschul *et al.*, 2005) with non-redundant protein sequence and protein blast parameters. The selected sequences were aligned in Clustal W server (Thompson *et al.*, 1994) .The evolutionary relationship was predicted using MEGA 6.0 software (Tamura *et al.*, 2013) based on the neighbor-joining method (Saitou and Nei, 1987) with the statistical reliability, including bootstrap analysis through 1000 replications (Felsenstein, 1985).

*Template identification and creation of protein model*

The template was searched by subjecting the target sequence to Blast using the parameters setting to Protein Data Bank proteins (Pdb) and Blastp (protein-protein BLAST) for homology modeling. The template was selected based on smaller e-value, maximum query coverage and sequence identity (above 50%), lower resolution, gap percentage and elucidating the alignments between a query and subject. The Swiss-Model template library was (Biasini *et al.*, 2014) also confirmed the template selection. The selected template was 1DQ8_A and it showed 55% sequence identity, e-value 3e-163, maximum query coverage, 98%, 2.10 Angstrom resolution, and 0% gaps. After selecting the template the alignment between target and the best template were done using MUSCLE v3.8 (Edgar *et al.*, 2004). The alignment was also performed using 'align2d' function of Modeller 9v14 program. After performing the alignment between target and template sequences the homology modeling of *Ta*HMGR was done using Modeller 9v14 (Eswar *et al.*, 2006). Modeller 9v14 is a python based widely accepted computer program, used to develop protein 3D structures with the expiation of spatial restraint (Eswar *et al.*, 2006). This computer program builds the model using the selected sequence, templates and five python script files (including build a profile.py, compare.py, aligned2d.py, model-single.py and evaluate model.py) incorporating modeler instructions. The Modeller 9v14 program was used to build forty protein models and the best protein model was selected based on DOPE (Discrete Optimized Protein Energy) potential score and the model was visualized using PyMOL v1.3 software.

*Protein model evaluation and submit to PMDB database*

The validation of the protein model is very important for the acceptance to others and for further analysis. In this case, the generated protein model was subjected to the PDB sum generate page server (Laskowski, 2004) for the prediction of Ramachandran plot statistics results.

The RAMPAGE was also used to check the stereochemical quality (Lovell *et al.*, 2003). Then the protein model was subjected to Resprox server (Berjanskii *et al.*, 2012) to know the Molprobity score. After completing the evaluation process the model was then opened by Swiss-Pdb Viewer version 4.1.0 (Guex and Peitsch, 1997) software for energy minimization. Further the energy minimized model was submitted to the SAVES server (www.nihserver.mbi.ucla.edu/SAVES) for verify 3D (Lüthy *et al.*, 1992) and ERRAT (Colovos and Yeates, 1993) result. ProSA (Wiederstein and Sippl, 2007) and ProQ (Benkert *et al.*, 2011) server were also used for evaluating the protein model. After the evaluation and energy minimization the refined protein model of *Ta*HMGR was submitted to Protein Model Database (PMDB) (www.mi.caspur.it/PMDB) (Castrignano, 2006)and the model was assigned the following PMDB id: PM0080186.

*Structure comparison of Ta*HMGR *protein*

The structure comparison was anticipated by using the iPBA web server, a quick and effective protein sequence based structure comparison tool (Gelly *et al.*, 2011). This server performs depends on similarity in the local backbone conformation (Joseph *et al.*, 2010; Gelly *et al.*, 2011).

*Prediction of potential ligand binding sites and function annotation*

The potential protein-ligand interaction sites were anticipated by using the web based MetaPocket 2.0 (www.projects.biotec.tu-dresden.de/metapocket) (Zhang *et al.*, 2011). This server is executed based on a consensus method and work together with eight effective methods, ConCavity (CON), PASS (PAS), Q-site Finder (QSF), LIGSITE$^{cs}$ (LCS), SURFNET (SFN), GHECOM (GHE), POCASA (PCS) and Fpocket (FPR) for increasing the prediction success rate (Zhang *et al.*, 2011). The MetaPocket 2.0 predicts the result after completing the following three steps: calling-based approaches, creating Meta pocket sites and finally representing ligand- binding residues (Zhang *et al.*, 2011). The function annotation was completed by using the ProFunc web server (www.ebi.ac.uk/thorntonsrv/databases/profunc), mostly performs to categorize the biochemical function of the modeled protein using different types of methods (Laskowski, 2004). The ProFunc was established through the Midwest Center for Structural Genomics (MCSG) as an essential portion of the structural Genomics Pipeline (Labarga *et al.*, 2007). ProFunc predicts the function by analyzing the protein sequence and structure using subsisting and innovative methods and identify confining relationships or functional motifs among the functionally active proteins (Laskowski, 2004).

## Results and discussion

*Analysis of physico-chemical properties and secondary structure of the catalytic domain of TaHMGR*

In the present study, the physico-chemical properties computed by the ProtParam server found that the HMG-CoA reductase protein had a molecular weight of 44.5 KDa and extinction co-efficient value 15930. The molecular weight of a typical plant HMG-CoA reductase protein ranges from 60 to 65 KDa but the molecular weight of our selected protein is only 44.5 KDa. So, it is commensurate with the catalytic domain of HMG-CoA reductase protein.

**Table 1.** The phosphorylation sites of HMG-CoA reductase of *T. aestivum* L. predicted by NetPhos 2.0.

| Serine Prediction | | Threonine Prediction | | Tyrosine Prediction | |
|---|---|---|---|---|---|
| Position | Score | Position | Score | Position | Score |
| 39 | 0.998 | 44 | 0.984 | 58 | 0.927 |
| 115 | 0.791 | 97 | 0.607 | 72 | 0.750 |
| 164 | 0.942 | 137 | 0.938 | | |
| 189 | 0.885 | 156 | 0.628 | | |
| 221 | 0.693 | 165 | 0.623 | | |
| 244 | 0.943 | | | | |
| 279 | 0.540 | | | | |
| 337 | 0.839 | | | | |
| 368 | 0.994 | | | | |
| 371 | 0.767 | | | | |
| 410 | 0.994 | | | | |
| 415 | 0.591 | | | | |

The retrieved sequence disclosed the isoelectric point value as 6.94 which demonstrating the weakly acidic nature of the catalytic domain. Thermostability and structural stability of the protein are increased with increasing the aliphatic index value. In this case, the aliphatic index value, 92.40 represent such effectiveness of the protein in a wide range of temperature.

**Table 2.** Residues involved with ligand binding sites, predicted by MetaPocket 2.0.

| Binding Site No. | Residues | | | | |
|---|---|---|---|---|---|
| Site 1 | GLY_9^63^ | GLN_9^64^ | ASN_9^197^ | GLN_9^348^ | GLY_9^345^ |
| | CYS_9^66^ | GLU_9^67^ | MET_9^68^ | PRO_9^69^ | VAL_9^70^ |
| | GLY_9^71^ | ILE_9^61^ | GLY_9^344 | TYR_9^58^ | CYS_9^65^ |
| | MET_9^198^ | PHE_9^56^ | MET_9^196^ | GLY_9^195^ | ALA_9^193^ |
| | TYR_9^72^ | LEU_9^349^ | GLY_9^346^ | MET_9^194^ | VAL_9^343^ |
| | THR_9^190^ | VAL_9^73^ | ASP_9^306^ | LYS_9^230^ | GLN_9^352^ |
| | GLN_9^305^ | GLY_9^304^ | THR_9^347^ | PRO_9^307^ | GLN_9^309 |
| | SER_9^351^ | GLY_9^341^ | ALA_9^308^ | GLN_9^74^ | VAL_9^340^ |
| | ILE_9^301^ | LEU_9^75^ | ALA_9^302^ | ALA_9^350^ | CYS_9^355^ |
| | ALA_9^95^ | VAL_9^77^ | ALA_9^354^ | PRO_9^76^ | LEU_9^51^ |
| | ARG_9^129^ | ASP_9^229^ | SER_9^200^ | GLY_9^55^ | ASP_9^57^ |
| | SER_9^60^ | | | | |
| Site 2 | GLU_9^98^ | GLY_9^99^ | LEU_9^101^ | ASN_9^294^ | CYS_9^100^ |
| | SER_9^104^ | LEU_9^391^ | HIS_9^291^ | GLN_9^402^ | LEU_9^400^ |
| | ALA_9^394^ | HIS_9^399^ | ARG_9^107^ | SER_9^390^ | ALA_9^103^ |
| | VAL_9^401^ | VAL_9^102^ | | | |
| Site 3 | VAL_9^12^ | ILE_9^17^ | VAL_9^21^ | LEU_9^22^ | LYS_9^25^ |
| | LEU_9^26^ | THR_9^24^ | ILE_9^8^ | GLN_9^11^ | GLU_9^3^ |
| | ARG_9^31^ | GLY_9^27^ | GLU_9^7^ | GLU_9^4^ | ASP_9^5^ |
| | PRO_9^2^ | MET_9^1^ | ILE_9^35^ | VAL_9^9^ | |

The Guruprasad *et al.* noted that, the instability index value exceeding 40 indicates the unstable nature while the value less than 40 specify the stable nature of the protein. So, the predicted value 47.05 reveals that the selected catalytic domain is unstable in nature. Kyte and Doolittle stated that, the positive GRAVY value indicates the hydrophobic nature while the negative value indicates the hydrophilic nature of the protein. So, in this study GRAVY value 0.065 (positive) expose the hydrophilic nature of the catalytic domain.

**Table 3.** Result showing the function of the best modeled protein predicted by ProFunc (protein biochemical function prediction server).

| SL. NO. | GO Term | Description | Predicted Score |
|---|---|---|---|
| 1. | GO:0003824 | Catalytic activity | 89.22 |
| 2. | GO:0005488 | Binding activity | 88.57 |
| 3. | GO:0056491 | Oxidoreductase activity | 74.71 |
| 4. | GO:0050662 | Co-enzyme binding | 73.12 |
| 5. | GO:0048037 | Co-factor binding | 73.12 |
| 6. | GO:0009987 | Cellular process | 93.78 |
| 7. | GO:0008152 | Metabolic process | 93.28 |
| 8. | GO:0044237 | Cellular Metabolic process | 93.07 |
| 9. | GO:0044249 | Cellular biosynthetic process | 78.17 |
| 10. | GO:0055114 | Oxidation reduction | 74.71 |

The ProtParam server also predicted the essential and non-essential amino acid composition of the catalytic domain of *Ta*HMGR. The essential amino acids Valine (V), Leucine (L) and Arginine (R) showed the highest composition of 10%, 8.8% and 6.7%, respectively for the catalytic domain of *Ta*HMGR whereas the compositions were 6.6%, 9.0% and 4.7%, respectively, for the catalytic portion of human HMG-CoA reductase.

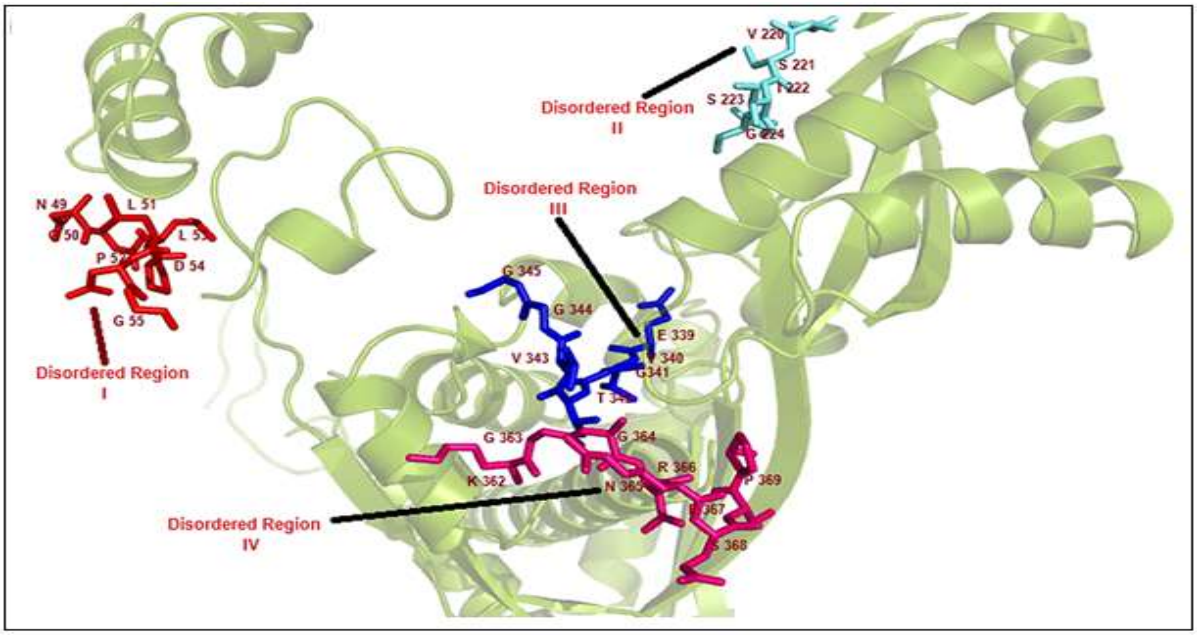


**Fig. 1.** The disordered regions of the catalytic domain of *Ta*HMGR. The Red, Cyan, Blue, and Hot pink color sticks indicate the first, second, third and fourth disordered regions, respectively.

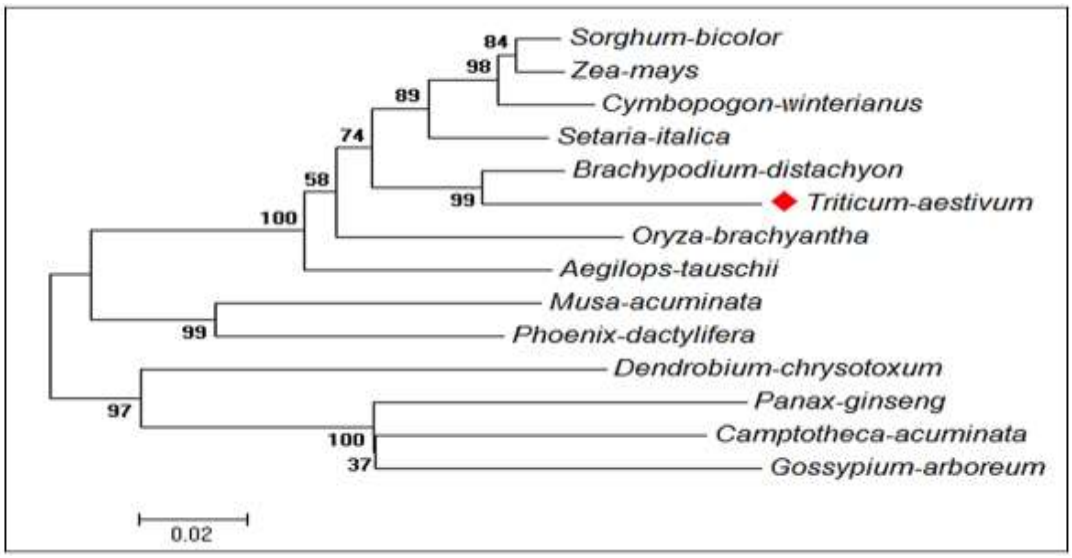


**Fig. 2.** Evolutionary relationship among HMG-CoA reductase sequences from different plant species were conducted in MEGA6. The following phylogenetic tree was deduced by using the Neighbor-Joining method and this finest tree are consisted with the total branch length of 0.81483066. The bootstrap test was conducted using 1000 replications and has been shown next to the branches. The same units such as length have been used throughout the phylogenetic tree and poisson correction method was used to compute the evolutionary distances. This analysis included 14 amino acid sequences from different plant species. Entire locations enclosing gaps and missing data were removed. There were a total of 417 positions in the final data set.

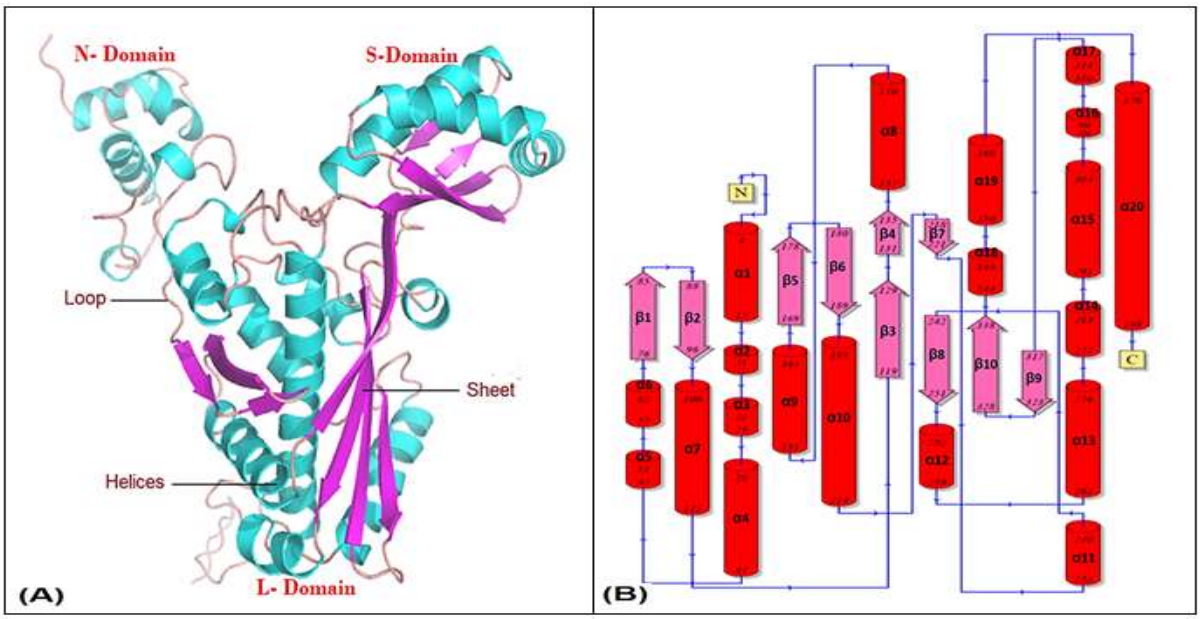


**Fig. 3.** (A) Three dimensional structure of the catalytic domain of *Ta*HMGR protein. Helices, sheets and coils are colored in cyan, magenta and orange, respectively. (B) Topology of the predicted *Ta*HMGR protein and it clearly displays the presence of α1-α20 (alpha helices) and β1- β10 (beta sheets) in the predicted protein model.

The non-essential amino acids Glycine (G), Alanine (A) and Serine (S) showed the highest composition of 10%, 10.5% and 8.6%, respectively for *Ta*HMGR and 9.0%, 10.7%, and 7.1%, respectively, for human HMG-CoA reductase. After analyzing the amino acid compositions, it can be revealed that the enzyme comprises a high composition of Leucine (L), Glycine (G), Alanine (A) and Serine (S) and these compositions showed very little differences between *Ta*HMGR and human HMG-CoA reductase.

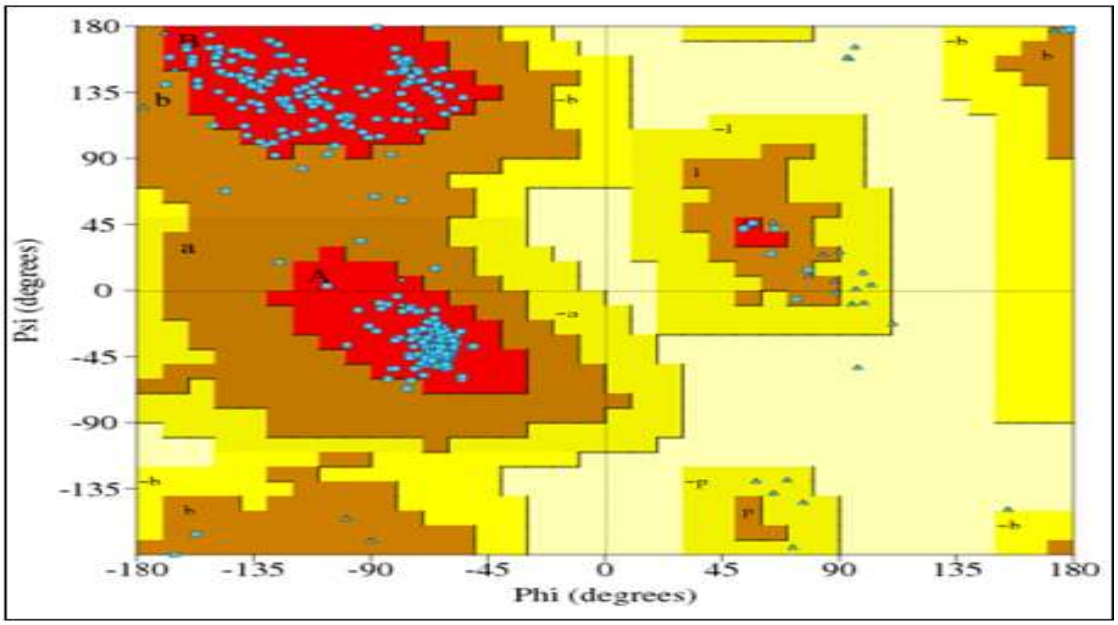


**Fig. 4.** Ramachandran plot for the modeled *Ta*HMGR. Here, total 362 residues present in the non-glycine and non-proline regions in the model. 337 residues (93.1%) lies in most favored regions, 25 residues (6.9%) in additional allowed regions, 0 residues (0.0%) in generously allowed regions and 0 residues (0.0%) in disallowed regions. It also contains 2 end-residues (excl. Gly and Pro), 42 Glycine residues and 14 Proline residues.

The amino acid Valine (V) and Leucine (L) raise the serum levels of HDL-cholesterol and in this case they can be used as valuable components for the reduction of coronary artery diseases like atherosclerosis (Cojocaru *et al.*, 2010).

So, the regulation of Valine (V) and Leucine (L) residues in this *Ta*HMGR enzyme may bring a remarkable effect on the research as this enzyme is also involved in cholesterol synthesis.

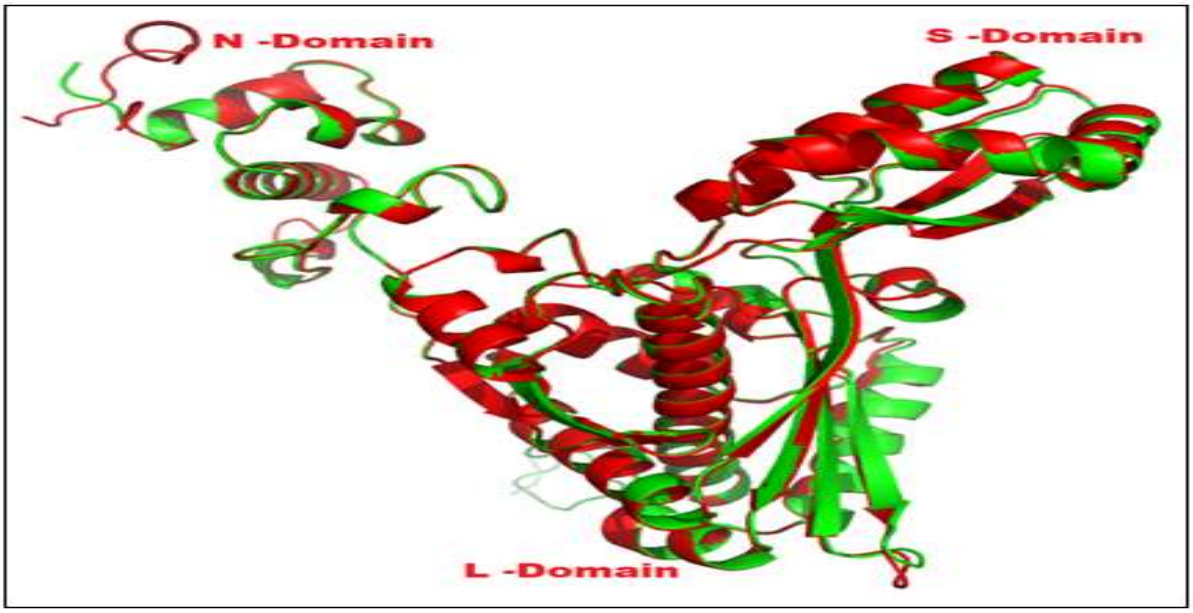


**Fig. 5.** The superposition of the three dimensional model of *Ta*HMGR using iPBA server.

The annotations of secondary structure by PSIPRED showed that the *Ta*HMGR contained 41.90% (176 AA) alpha helices (Hh), 42.14% (177 AA) random coils (Cc), and 14.95% (67 AA) extended strand (Ee) while the selected template contains 37.04% (173 AA) alpha helices (Hh), 50.53% (236 AA) random coils (Cc), and 12.42% (58 AA) extended strand (Ee). So, the secondary structural topographies exposed that, the random coils (Cc) structure of this enzyme is dominated to over alpha helices (Hh), and extended strands (Ee) in both targets (catalytic domain of *Ta*HMGR) and template sequence (catalytic portion of Human HMGR) which also indicate structural similarity between target and template sequence.

*Analysis of phosphorylation sites*

In eukaryotes, phosphorylation is responsible for the regulation of cellular compartments such as regulation of RNA synthesis, DNA repair mechanism, ecofriendly stress response, metabolism and cellular diversity (Zhang and Johnson, 2000; Wang *et al.*, 2010; Trost and Kusalik, 2011).

The NetPhos 2.0 server predicted the 19 phosphorylation sites in HMG-CoA reductase encoding protein sequence, including 12 Serine (S), 5 Threonine (T) and 2 Tyrosine (Y) sites with significant score. The predicted phosphorylation sites with score are shown in Table 1. The phosphorylation site of HMG-CoA reductase has been determined experimentally in mammals and it appears to be conserved in plant HMG-CoA reductase protein (Dale *et al.*, 1995). The phosphorylation of the protein containing amino acid (Serine, Threonine and Tyrosine) residues affects the cellular signaling process and may also prevent the activity of other amino acid residues (Blom *et al.*, 1999). So, the results reveal that, identified phosphorylation sites can be used as regulatory sites and may also affect the cellular and molecular functions of the protein.

*Analysis of intrinsic protein disorder regions*

The protein sequence which consists of both disordered and ordered regions is called intrinsically disordered protein (Obradovic *et al.*, 2005).

The disordered regions are not the useless regions of the protein sequence because the disordered regions can contribute the significant role in the molecular-interaction linkage such as transcription and translation process, cell signaling and alternative splicing (Dyson and Wright, 2005; Ishida and Kinoshita, 2008). Disordered regions identification of a protein is essential for useful function prediction and structural determination to remove the complications in crystallization and refinement of the protein (Oldfield *et al*., 2005; Ishida and Kinoshita, 2008). In eukaryotes, the disordered regions are mainly found in the protein's functional sites (Dunker and Obradovic, 2001). The GlobPlot v2.3 predicts four intrinsic disorder regions in *Ta*HMGR. The resulting analysis showed that 27 residues are involved in the formation of the disordered region, which is mainly disseminated in the N-terminal and C-terminal region of the protein. The potential globular domain (GlobDom) by Rusell/Linding definition is located in the position 2-420. There are no low complexity regions in the selected sequence. The first disordered region is 'NGLPLDG' between the sequence positions 49-55, second is 'VSISG' between the positions 220-224, third is 'EVGTVGG' between the positions 339-345 and fourth is 'KGANRESP' between the sequence positions 362-369 in *Ta*HMGR. The disordered regions have shown in the three dimensional structure as Fig. 1. The disordered regions do not occur in the same position of all proteins such as the disordered regions are seen in position 4-42, 72-77, 201-205, 209-214 and 223-228 in Camel milk kappa casein protein (Wadood *et al*., 2014) and in rice Chitinase I the positions are 1-2, 56-70, 80-88, and 319-323 (Sarma *et al*., 2012).

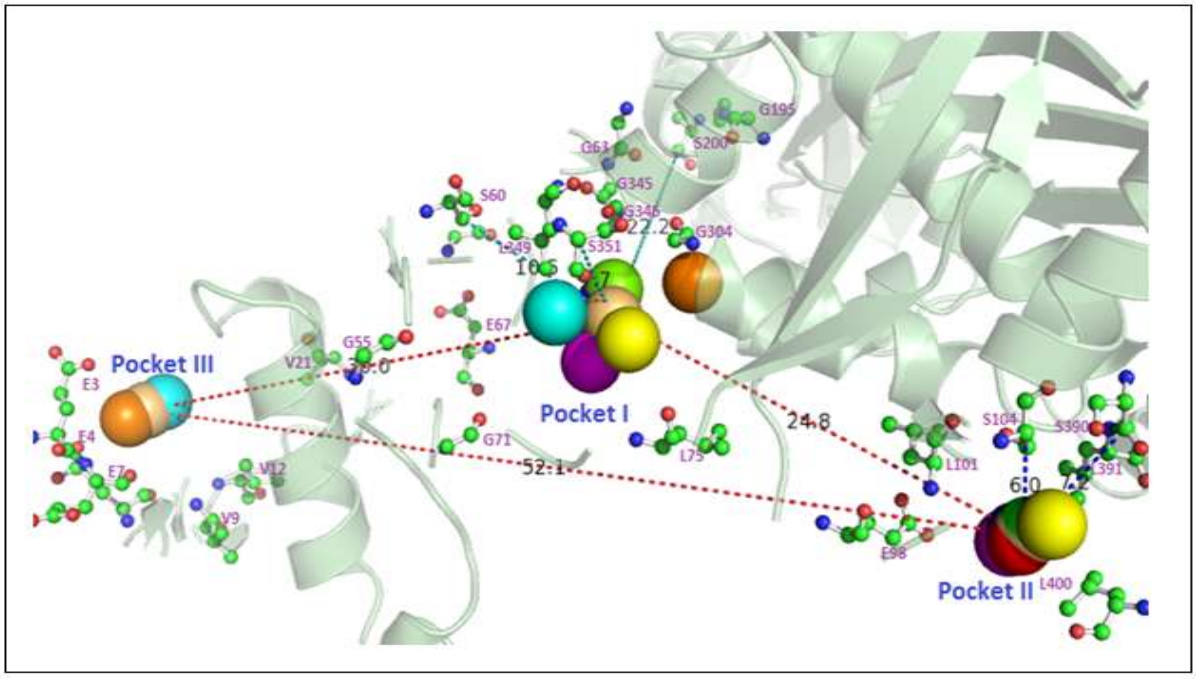


**Fig. 6.** Three important pockets of the catalytic domain of *Ta*HMGR protein, predicted by MetaPocket 2.0.The distances are shown among the pockets and involved Serine residues in Ångstroms unit.

The amino acid compositions of a protein, namely aliphatic (Alanine, Glycine), basic (Arginine, Lysine), acidic (Glutamine, Glutamic acid), hydroxyl-containing (Serine) and cyclic (Proline) are responsible for creating the disorder regions in the protein sequence (Orosz and Ovádi, 2011). On the other hand, the aliphatic amino acid residues, namely Isoleucine, Leucine and Valine and aromatic amino acids, namely Phenylalanine, Tyrosine and Tryptophan reduce the intrinsically disordered regions of a protein (Dunker and Obradovic, 2001; Orosz and Ovádi, 2011). The results disclose that, the composition of amino acids, namely Glycine, Serine, Lysine, Proline, Arginine,

Alanine and Glutamic acid are present in the disordered regions and that amino acids may be responsible for the formation of intrinsic disorder regions in this selected protein sequence. The predicted disordered regions will not be visible by UV spectrophotometry due to the absence of Tryptophan, Tyrosine or Cysteine in the regions.

*Analysis of evolutionary relationships*

The protein sequences predicted by Blastp showed 94%, 90%, 89%, 89%, 89%, 89%, 88%, 83%, 82%, 81%, 79%, 79% and 78% sequence identity with *Brachypodium distachyon* (XP_003578341.1), *Setaria italica* (XP_004957224.1), *Oryza brachyantha* (XP_006660783.1), *Sorghum bicolor* (XP_002460459.1), *Zea mays* (DAA62078.1), *Cymbopogon winterianus* (AIZ09174.1), *Aegilops tauschii* (EMT29532.1), *Phoenix dactylifera* (XP_008796235.1), *Musa acuminate* (XP_009387500.1), *Dendrobium chrysotoxum* (AHI88624.1), *Camptotheca acuminate* (AAB69726.1), *Panax ginseng* (AIX87980.1) and *Gossypium arboretum* (KHG10514.1), respectively. The phylogenetic tree constructed by MEGA 6.0 is illustrated in Fig. 2. The phylogenetic tree result showed that the selected sequences of HMG-CoA reductase from different plant source, including *Ta*HMGR separated in two different clusters. These clusters showed more than 50 bootstrap values in the branches of the phylogenetic tree except in the branch between *C. acuminate* and *G. arboretum*. The separation of cluster maintains the order level order.

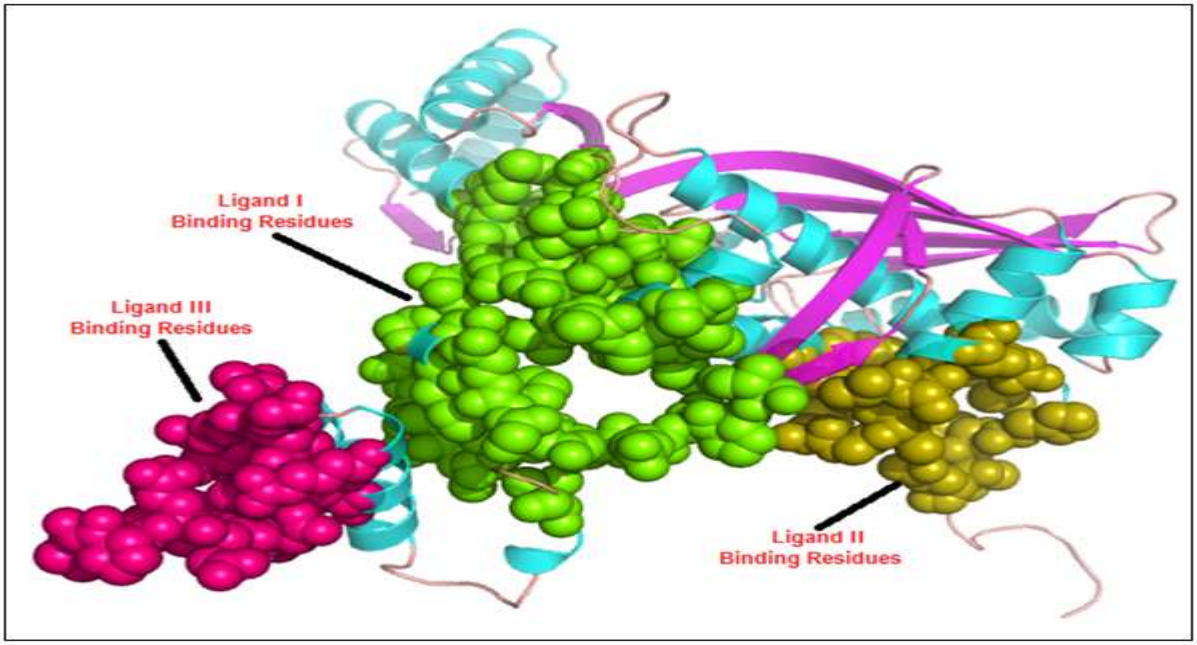


**Fig. 7.** Three important ligand binding residues of the catalytic domain of *Ta*HMGR protein. Helices, sheets and coils are colored in cyan, magenta and orange, respectively.

The cluster I consists of ten species which were further distributed into two sub-clusters. Sub-cluster I comprises eight species, namely *S. bicolor*, *Z. mays*, *C. winterianus*, *S. italica*, *B. distachyon*, *T. aestivum*, *O. brachyantha* and *A. tauschii* which have come from Poales order and Poaceae family while Sub-cluster II includes two species, namely *M. acuminata* and *P. dactylifera* which are from Zingiberales and Areacales order, respectively.

Cluster II consists of four species, namely *D. chrysotoxum*, *C. acuminate*, *P. ginseng*, and *G. arboretum* which have come from four different orders, namely Asparagales, Cornales, Apiales and Malvales, respectively. So, the results proposed to that the selected plant species containing HMG-CoA reductase have come from different orders and evolved into different groups, but there is strong similarity exists among these species and probably have come from a common ancestor gene.

This result will be helpful for creating postulates about protein function, especially intend for biochemical and molecular studies.

*Comparative protein modeling and evaluation*

The protein three dimensional structures and its function are important for establishing the controlling mechanism at the molecular level. The crystal structure of plant HMG-CoA reductase has not been determined, but only the homology model has been published for some plant species (Liao *et al.*, 2004; Jiang *et al.*, 2006; Kiran *et al.*, 2010; Kalita *et al.*, 2015). The protein sequence of the catalytic domain of *Ta*HMGR having accession id W5FJD4 has no three dimensional structure present in the PDB (Protein Data Bank) or PMDB (Protein Model Database). In that case, the template (Pdb id: 1DQ8 chain A) was identified with 55% sequence identity through the NCBI blast against the Protein Data Bank for comparative protein modeling. Total five models were constructed after performing comparative modeling with Modeller 9v14. The lowest DOPE score of the constructed models was deemed to select the best model. The predicted protein model was functional as a monomer because the selected sequence was a catalytic domain of *Ta*HMGR and only a single chain "Chain A" of template "1DQ8" was used for the comparative protein modeling. The selected template is a catalytic portion of human HMG-CoA reductase enzyme (a homo-tetramer) (Istvan *et al.*, 2000) and our selected protein sequence is also a catalytic domain sequence of *Ta*HMGR and only this catalytic domain is conserved between plant and human HMG-CoA reductase enzyme. The result revealed that the monomer of the catalytic domain of *Ta*HMGR consisted of three domains, including the small helical N-terminal N-domain, the largest central L-domain and the smallest helical S-domain (Fig. 3A.) and the topology of our predicted protein model shows twenty alpha helices and ten beta sheets (Fig. 3B.) which are also comparable to human HMG-CoA reductase (Istvan *et al.*, 2000) and different plant species, namely, *Artemisia annua* (Kiran *et al.*, 2010), *Eucommia ulmoides* (Jiang *et al.*, 2006), and *Taxus media (Liao et al.*, 2004) HMG-CoA reductase reported previously.

The analysis also exposed that the three dimensional structure of *Ta*HMGR extremely resembled with the catalytic portion of human HMG-CoA reductase, which designate the possible catalytic similarities between them.

The validation of the predicted protein models was carried out based on the Ramachandran plot statistics, which shows the complete symmetrical study along with the stereochmical exactness of the modeled protein structure through evaluating residue by residue geometry and comprehensive structural geometry. The predicted model with a lower DOPE score showed that 93.1% of residues were located in the most favored regions, 6.9% in additional allowed regions and there were no residues in the generously allowed and disallowed regions (Fig. 4.) whereas the template showed 89.5% in most favored, 9.8% in additional allowed, 0.6% in generously allowed and 0.1% in disallowed regions. The RAMPAGE also ensures the model quality by maintaining the criteria. The predicted model showed the good Molprobity value in bad bond length (-0.33), bad bond angles 0.45) and also in the Ramachandran plot outliers (-0.11). The Verify 3D score above 80% is considered as an acceptable score for the model. In the present study, the Verify 3D analysis showed 85.00% for the predicted model, whereas for the template showed 97.79% residues in the 3D-1D score of greater than 0.2. ERRAT is used to take the decision about the reliability of a protein model. The ERRAT value for the template and predicted model was 96.873 and 84.250, respectively. The ProSA is mainly used to find out the potential errors in protein model. The positive Z-score indicates the presence of problematic region in the protein model. In our study, the ProSA showed the negative Z-score (-9.18) for predicted model while the Z-score value for the template was -8.6 in the NMR region, designating the acceptability of the predicted model. The energy plot ensures the finest compatibility between target and template sequence. The evaluation process using Ramachandran plot statistics, Verify 3D, ERRAT quality factor and ProSA Z-score confirms the quality of the predicted model in comparison with the template.

The energy of the predicted model before energy minimization and after energy minimization was -5558.672 and -22260.557, respectively. The ProQ (Protein Quality Predictor) server predicted the LG score (4.147) and MaxSub score (0.348). The LG and MaxSub score of ProQ server confirms the quality of the model as an extremely good model and good model, respectively. The evaluated protein model was deposited in PMDB database and assigned with a PMDB id: PM0080186. The three dimensional model of *Ta*HMGR, using the best template 1DQ8 chain A was visualized using PyMOL software that contains helices, sheets and coils.

*Structure comparison assessment of HMG-CoA reductase enzyme*

Comparative structural analysis of *Ta*HMGR enzyme was performed using the iPBA server. The structural comparison of the model helps to comprehend the essential character of protein and their functions. The functions of newly identified proteins can also be assessed by observing the structural co-relationships (Ingale and Goto, 2014). The energy minimized protein model was superimposed with the selected template. The superimposed model of *Ta*HMGR with a template at zero degree angles is exhibited in Fig. 5. The result predicted by the iPBA server showed the lowest RMSD score of 0.40 Angstrom with the normalized score of 400.30 and 91.10% of its residues is aligned in the predicted model. The predicted GDT_TS score was 89.63 for the predicted model. The GDT_TS score ranges from 0 to 100 but the score 50 or more than 50 is deliberated as significant for structural relationship. So, the predicted score 89.63 demonstrating the significant structural similarity between target and template sequence. Finally, the superimposed model of HMG-CoA reductase with RMSD and GDT_TS score specifies the same fold of the predicted and template model.

*Analysis of potential ligand binding sites and function*

The protein surface contains the cavities or pockets where small molecules or ligands can bind directly and contribute to structure based drug design (Huang, 2009).

After completing the cluster technique, the MetaPocket 2.0 clusters the top 3 files from various algorithmic methods. This server finds out the six metapocket clusters for the submitted protein model, but the best three pockets sites are considered for presenting the results (Fig. 6).

The result showed the first metapocket site (MPT 1) consists of seven pockets. These seven pockets include the first pocket from Concavity (CON-1, Z-score 1.00), the second pocket from Fpocket (FPK-2, Z-score 1.53), the first pocket from GHECOM (GHE-1, Z-score 4.60), the first pocket from Ligisitecs (LCS-1, Z-score 7.38), the second pocket from PASS (PAS-2, Z-score 0.87), the first pocket from POCASA (PCS-1, Z-score 1.99), and the first pocket from SURFNET (SFN-1, Z-score 6.82) with total Z-score 24.18.

The second MetaPocket site (MPT-2) comprises of five pockets, from CON-2 (Z-score -1.00), GHE-2 (Z-score 0.97), LCS-2 (Z-score 1.63), PAS-1 (Z-score 1.33) and PCS-2 (Z-score -0.34) with total Z-score 2.60. The third and final MetaPocket site (MPT-3) contains two pockets, from the third pocket FPK-3 (Z-score 1.42) and the third pocket SFN-3 (Z-score 0.32) with total Z-score 1.74. The residues involved in the potential three ligand binding sites of *Ta*HMGR is shown in Table 2 and also in the three dimensional structure (Fig. 7). The potential ligand binding sites are consisted of total 94 residues which are divided into three sites. The site 1 contains 56 residues; site 2 contains 17 residues while the site 3 contains 19 residues. The site 1 contains a high composition of Glycine, Alanine, Valine, Glutamine and Aspartic acid residues and site 2 contains Alanine, Serine and Valine residues while the site 3 contains Valine, Glutamic acid, Isoleucine and Leucine residues, but the amino acid residues Glycine (G), Valine (V), Leucine (L), Arginine (R), Glutamic acid (E) and Glutamine (Q) were common in all the three binding sites.

In *Ta*HMGR, the first and third binding sites contain Lys25 and Lys230; respectively, while the Lysine residue (Lys735) is present in human HMG-CoA reductase and it helps to clamp the substrate to enzyme (Istvan *et al.*, 2000) and

other residue Lys692 makes the ion-dipole contact with the C1 carboxylate of statins (Istvan and Deisenhofer, 2001). The amino acid residues Alanine (Ala751) and Leucine (Leu853) makes hydrophobic interactions with the HMG part of human HMG-CoA reductase(Istvan *et al*., 2000) and in *Ta*HMGR the residues Ala95, Ala193, Ala302, Ala308, Ala350, and Ala354 present in the first binding site and Leucine is present in all the three binding sites (Fig. 6). In human HMG-CoA reductase, Ser684 stabilizes the substrate by creating a hydrogen bond (Istvan *et al*., 2000) and in *Ta*HMGR 'Ser60, Ser200, Ser351' and 'Ser104, Ser390' are involved in the formation of first and second binding sites, respectively (Fig. 6.) and maybe these residues also acts like as human HMG-CoA reductase.

The Glutamic acid (Glu559) act as a proton donor in human HMG-CoA reductase (Istvan *et al*., 2000) while the Glu67 and Glu98 are involved in the formation of first and second binding site, respectively, and the residues Glu3, Glu4, and Glu7 are involved in the third binding site formation (Fig. 6). Arginine is present in all the three binding sites of *Ta*HMGR and the Arginine (Arg590) makes an ion-dipole interaction in the statins (Istvan and Deisenhofer, 2001). So, the results demonstrated that the predicted three important ligand binding sites or pockets and involved residues may be used as an initial target point for further functions prediction and improvement of structures based medicine.
The protein structure is important for the function prediction, but the low availability of structural information is the biggest limitation for predicting function from structure. Furthermore, structure prediction methods are better at predicting the core of proteins than the loops.

The important biochemical functions predicted by the ProFunc server are exhibited in Table 3. The results showed some biological process, including metabolic process, cellular metabolic process, cellular biosynthetic process and primary metabolic process with significant score. The biochemical function such as catalytic activity and binding ability of other molecules with the significance score reveals the actual functional site of the *Ta*HMGR.

## Conclusion

In this *in silico* study, various Bioinformatics tools and servers have been used for the analyses and protein model building of *Ta*HMGR. The selected protein sequence is a catalytic domain and is weakly acidic, unstable, high temperature sensitive and containing phosphorylation sites, regulatory amino acid residue Serine, Threonine and Tyrosine. There are four disordered regions existing and amino acid residues mainly Proline, Lysine, Glycine and Serine are involved in the formation of disordered regions in this protein sequence. The predicted catalytic domain model is functional as a monomer and is containing small helical N-domain, the central L-domain and the smallest helical S-domain and also having three important ligand binding sites.

This catalytic domain model is involved in binding and oxidoreductase activity and also in cellular and metabolic processes as its vital functions. Finally, we hope and expect that these findings will afford primary necessary information to initiate a wet lab research in improving the structure based drug design using the *Ta*HMGRin near future.

## Acknowledgement

The authors reply to thanks to the Professor A.K.M Akhtarul Islam, Department of English, Islamic University, Kushtia for editing and checking the spelling and grammatical error of this manuscript.

## References

**Adhikari UK, Hossain S, Rahman MM.** 2015. Investigations on 3-hydroxy-3-methylglutaryl CoA reductase enzyme from different source organisms. Journal of Nature Science and Sustainable Technology **9,** 725–735.

**Altschul SF, Wootton JC, Gertz EM.** 2005. Protein Database Searches Using Compositionally Adjusted Substitution Matrices. FEBS Journal **272,** 5101–5109.
http://dx.doi.org/10.1111/j.1742-4658.2005.04945.x

**Andrade-Pavón D, Sánchez-Sandoval E, Rosales-Acosta B, Ibarra JA, Tamariz J, Hernández-Rodríguez C, Villa-Tanaca L.** 2014. The 3-hydroxy-3-methylglutaryl coenzyme-A reductases from fungi: a proposal as a therapeutic target and as a study model. Revista Iberoamericana de Micología **31**, 81–5. http://dx.doi.org/10.1016/j.riam.2013.10.004

**Aoyagi K, Beyou A, Moon K, Fang L, Ulrich T.** 1993**.** lsolation and Characterization of cDNAs Encoding Wheat 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase. Plant physiology **102**, 623–628.

**Benkert P, Biasini M, Schwede T.** 2011. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics **27**, 343–350. http://dx.doi.org/10.1093/bioinformatics/btq662

**Berjanskii M, Zhou J, Liang Y, Lin G, Wishart DS.** 2012. Resolution-by-proxy: a simple measure for assessing and comparing the overall quality of NMR protein structures. Journal of biomolecular NMR **53**, 167–180. http://dx.doi.org/10.1007/s10858-012-9637-2

**Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Schwede T.** 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Research **42,** W252–8. http://dx.doi.org/10.1093/nar/gku340

**Bjellqvist B, Hughes GJ, Pasquali C, Paquet N, Ravier F, Sanchez JC, Hochstrasser D.** 1993. The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. Electrophoresis **14,** 1023–1031.

**Blom N, Gammeltoft S, Brunak S.** 1999. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. Journal of molecular biology **294,** 1351–1362. http://dx.doi.org/10.1006/jmbi.1999.3310

**Campos N, Arro M, Ferrer A, Boronat A.** 2014. Determination of 3-hydroxy-3-methylglutaryl CoA reductase activity in plants. Plant Isoprenoids: Methods and Protocols **1153**, 21–40. http://dx.doi.org/10.1007/978-1-4939-0606-2_3

**Castrignano T.** 2006. The PMDB Protein Model Database. Nucleic Acids Research **34,** D306–D309. http://dx.doi.org/10.1093/nar/gkj105

**Chen X, Wang X, Li Z, Kong L, Liu G, Fu J, Wang A.** 2012. Molecular cloning, tissue expression and protein structure prediction of the porcine 3-hydroxy-3-methylglutaryl-Coenzyme A reductase (HMGR) gene. Gene **495,** 170–7. http://dx.doi.org/10.1016/j.gene.2011.12.051

**Cojocaru E, Zamfir CL, Lupuşoru CE, Cotuţiu C.** 2010. The effects of some nonpolar aminoacids--valine, leucine--administration on the arterial wall already exposed to a hypercholesterolemic diet. Revista medico-chirurgicala a Societatii de Medici si Naturalisti din Iasi, **114**, 504—509.

**Colovos C, Yeates TO.** 1993. Verification of protein structures: patterns of nonbonded atomic interactions. Protein Science **2,** 1511–1519. http://dx.doi.org/10.1002/pro.5560020916

**Dale S, Arró M, Becerra B, Morrice NG, Boronat A, Hardie DG, Ferrer A.** 1995. Bacterial expression of the catalytic domain of 3-hydroxy-3-methylglutaryl-CoA reductase (isoform HMGR1) from Arabidopsis thaliana, and its inactivation by phosphorylation at Ser577 by Brassica oleracea 3-hydroxy-3-methylglutaryl-CoA reductase kinase. European Journal of Biochemistry **233,** 506–513. http://dx.doi.org/10.1111/j.1432-1033.1995.506_2.x

**Darabi M, Izadi-Darbandi A, Masoudi-Nejad A.** 2012. Bioinformatics study of the 3-hydroxy-3-methylglotaryl-coenzyme A reductase (HMGR) gene in Gramineae. Molecular biology reports **39**, 8925—8935. http://dx.doi.org/10.1007/s11033-012-1761-2

**Darabi M, Seddigh S.** 2013. Conserved motifs identification of 3-hydroxy-3-methylglotaryl-coenzyme A reductase (HMGR ) protein in some different species of drosophilidae by bioinformatics tools. Annals of Biological Research **4,** 158–163.

**Dunker AK, Obradovic Z.** 2001. The protein trinity--linking function and disorder. Nature Biotechnology **19,** 805—806. http://dx.doi.org/10.1038/nbt0901-805.

**Dyson HJ, Wright PE.** 2005. Intrinsically unstructured proteins and their functions. Nature Reviews Molecular Cell Biology **6**, 197—208. http://dx.doi.org/10.1038/nrm1589.

**Edgar RC, Drive RM, Valley M.** 2004. MUSCLE : multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research **32,** 1792–1797. http://dx.doi.org/10.1093/nar/gkh340

**Eswar N, Webb B, Marti-Renom MA.** 2006. Comparative Protein Structure Modeling Using Modeller. Current protocols in bioinformatics **5,** 5–6. http://dx.doi.org/10.1002/0471250953.bi0506s15

**Felsenstein J.** 1985. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. Evolution (N Y) **39,** 783–791.

**Friesen JA, Rodwell VW.** 2004. The 3-hydroxy-3-methylglutaryl coenzyme-A (HMG-CoA) reductases. Genome Biology **5,** 248. http://dx.doi.org/10.1186/gb-2004-5-11-248

**Gasteiger E, Hoogland C, Gattiker A, Duvaud SE, Wilkins MR, Appel RD, Bairoch A.** 2005. Protein Identification and Analysis Tools on the ExPASy Server. Proteomics Protoc Handb, Humana Press, 571-607. http://dx.doi.org/10.1385/1-59259-890-0:571

**Gelly J-C, Joseph AP, Srinivasan N, de Brevern AG.** 2011. iPBA: a tool for protein structure comparison using sequence alignment strategies. Nucleic Acids Research **39**, W18–W23. http://dx.doi.org/10.1093/nar/gkr333

**Gill SC, von Hippel PH.** 1989. Calculation of protein extinction coefficients from amino acid sequence data. Analytical biochemistry **182,** 319–326.

**Guex N, Peitsch MC.** 1997. SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. Electrophoresis **18**, 2714–2723. http://dx.doi.org/10.1002/elps.1150181505

**Guruprasad K, Reddy B V, Pandit MW.** 1990. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Protein engineering **4,** 155–161. http://dx.doi.org/10.1093/protein/4.2.155

**Ha SH, Kim JB, Hwang YS, Lee SW.** 2003. Molecular characterization of three 3-hydroxy-3-methylglutaryl-CoA reductase genes including pathogen-induced Hmg2 from pepper (Capsicum annuum). Biochimica *et* Biophysica Acta (BBA)-Gene Structure and Expression **1625,** 253—260. http://dx.doi.org/10.1016/s0167-4781(02)00624-3

**Huang B.** 2009. MetaPocket: a meta approach to improve protein ligand binding site prediction. OMICS **13**, 325—330. http://dx.doi.org/10.1089/omi.2009.0045\

**Ikai A.** 1980. Thermostability and aliphatic index of globular proteins. Journal of biochemistry **88,** 1895–1898.

**Ingale AG, Goto S.** 2014. Prediction of CTL epitope, in silico modeling and functional analysis of cytolethal distending toxin (CDT) protein of Campylobacter jejuni. BMC Research Notes **7**, 92. http://dx.doi.org/10.1186/1756-0500-7-92

**Ishida T, Kinoshita K.** 2008. Prediction of disordered regions in proteins based on the meta approach. Bioinformatics **24**, 1344—1348. http://dx.doi.org/10.1093/bioinformatics/btn195

**Istvan E.** 2000. The structure of the catalytic portion of human HMG-CoA reductase. Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids **1529**, 9–18.
http://dx.doi.org/10.1016/S1388-1981(00)00134-7

**Istvan ES, Deisenhofer J.** 2001. Structural mechanism for statin inhibition of HMG-CoA reductase. Science **292**, 1160–1164.
http://dx.doi.org/10.1126/science.1059344

**Istvan ES, Palnitkar M, Buchanan SK, Deisenhofer J.** 2000. Crystal structure of the catalytic portion of human HMG-CoA reductase: insights into regulation of activity and catalysis. EMBO Journal **19,** 819–830.
http://dx.doi.org/10.1093/emboj/19.5.819

**Jiang J, Kai G, Cao X, Chen F, He D, Liu Q.** 2006.Molecular cloning of a HMG-CoA reductase gene from Eucommia ulmoides Oliver. Bioscience Reports **26**, 171–81.
http://dx.doi.org/10.1007/s10540-006-9010-3

**Joseph AP, Agarwal G, Mahajan S, Gelly JC, Swapna LS, Offmann B, Schneider B.** 2010. A short survey on protein blocks. Biophysical Reviews **2**, 137—147.
http://dx.doi.org/10.1007/s12551-010-0036-1

**Kalita R, Patar L, Shasany AK, Modi MK, Sen P.** 2015. Molecular cloning, characterization and expression analysis of 3-hydroxy-3-methylglutaryl coenzyme A reductase gene from Centella asiatica L. Molecular biology reports **42**, 1431-1439.

**Kiran U, Ram M, Mather Ali Khan SK, Jha P, Alam A, Abdin MZ.** 2010.Structural and functional characterization of HMG-COA reductase from Artemisia annua. Bioinformation **5**, 146–149.
http://dx.doi.org/10.6026/97320630005146

**Korth KL, Jaggard DAW, Dixon RA.** 2000. Developmental and light-regulated post-translational control of 3-hydroxy-3-methylglutaryl-CoA reductase levels in potato. The Plant Journal **23,** 507–516.
http://dx.doi.org/10.1046/j.1365-313x.2000.00821.x

**Kyte J, Doolittle RF.** 1982. A simple method for displaying the hydropathic character of a protein. Journal of molecular biology **157,** 105–132.
http://dx.doi.org/10.1016/0022-2836(82)90515-0

**Labarga A, Valentin F, Anderson M, Lopez R.** 2007. Web services at the European bioinformatics institute. Nucleic Acids Research **35,** W6—11.
http://dx.doi.org/10.1093/nar/gkm291

**Laskowski RA.** 2004. PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids. Nucleic Acids Research **33**, D266–D268.
http://dx.doi.org/10.1093/nar/gki001

**Leivar P, González VM, Castel S, Trelease RN, López-Iglesias C, Arró M, Fernandez-Busquets X.** 2005. Subcellular Localization of Arabidopsis 3-Hydroxy-3-Methylgluraryl-Coenzyme A Reductase. Plant physiology **137**, 57–69.
http://dx.doi.org/10.1104/pp.104.050245.far

**Li W, Liu W, Wei H, He Q, Chen J, Zhang B, Zhu S.** 2014.Species-specific expansion and molecular evolution of the 3-hydroxy-3-methylglutaryl coenzyme A reductase (HMGR) gene family in plants. PLoS One **9,** e94172.
http://dx.doi.org/10.1371/journal.pone.0094172

**Liao Z, Tan Q, Chai Y, Zuo K, Chen M, Gong Y, Tang K.** 2004. Cloning and characterisation of the gene encoding HMG-CoA reductase from Taxus media and its functional identification in yeast. Functional Plant Biology **31**, 73.
http://dx.doi.org/10.1071/FP03153

**Linding R, Russell RB, Neduva V, Gibson TJ.** 2003. GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic Acids Research **31**, 3701–3708.
http://dx.doi.org/10.1093/nar/gkg519

**Lovell SC, Davis IW, Arendall WB, de Bakker PI, Word JM, Prisant MG, Richardson DC.** 2003. Structure validation by Cα geometry: ϕ,ψ and Cβ deviation. Proteins: Structure, Function, and Bioinformatics **50**, 437–450.
http://dx.doi.org/10.1002/prot.10286

**Luskey KL, Stevens B.** 1985. Human 3-Hydroxy-3-methylglutaryl Coenzyme A Reductase. Journal of Biological Chemistry **260**, 10271–10277.

**Lüthy R, Bowie JU, Eisenberg D.** 1992. Assessment of protein models with three-dimensional profiles. Nature **356**, 83–85.
http://dx.doi.org/10.1038/356083a0

**McGuffin LJ, Bryson K, Jones DT.** 2000. The PSIPRED protein structure prediction server. Bioinformatics **16,**404–405.
http://dx.doi.org/10.1093/bioinformatics/16.4.404

**Nishikawa S, Hirata A, Nakano A.** 1994. Inhibition of endoplasmic reticulum (ER)-to-Golgi transport induces relocalization of binding protein (BiP) within the ER to form the BiP bodies. Molecular biology of the cell **5,** 1129–1143.

**Obradovic Z, Peng K, Vucetic S, Radivojac P, Dunker AK.** 2005. Exploiting heterogeneous sequence properties improves prediction of protein disorder. Proteins: Structure, Function, and Bioinformatics **61,** 176-182.
http://dx.doi.org/10.1002/prot.20735

**Oldfield CJ, Ulrich EL, Cheng Y, Dunker AK, Markley JL.** 2005. Addressing the intrinsic disorder bottleneck in structural proteomics. Proteins: Structure, Function, and Bioinformatics **59**, 444-453.
http://dx.doi.org/10.1002/prot.20446

**Orosz F, Ovádi J.** 2011. Proteins without 3D structure: definition, detection and beyond. Bioinformatics **27,** 1449—1454.
http://dx.doi.org/10.1093/bioinformatics/btr175

**Saitou N, Nei M.** 1987. The neighbour-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evo **4**, 406–425.

**Sarma K, Dehury B, Sahu J, Sarmah R, Sahoo S, Sahu M, Barooah M.** 2012. A comparative proteomic approach to analyse structure, function and evolution of rice chitinases: a step towards increasing plant fungal resistance. Journal of Molecular Modeling **18,** 4761—4780.
http://dx.doi.org/10.1007/s00894-012-1470-8

**Stormo C, Kringen MK, Grimholt RM, Berg JP, Piehler AP.** 2012. A novel 3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMGCR) splice variant with an alternative exon 1 potentially encoding an extended N-terminus. BMC molecular biology **13**, 29.
http://dx.doi.org/10.1186/1471-2199-13-29

**Tamura K, Stecher G, Peterson D.** 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular biology and evolution **30,** 2725–2729.
http://dx.doi.org/10.1093/molbev/mst197

**Theivagt AE, Amanti EN, Beresford NJ, Tabernero L, Friesen JA.** 2006. Characterization of an HMG-CoA Reductase from *Listeria monocytogenes* That Exhibits Dual Coenzyme Specificity. Biochemistry **45,** 14397–14406.
http://dx.doi.org/10.1021/bi0614636

**Thompson JD, Higgins DG, Gibson TJ.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research **22,** 4673–4680.
http://dx.doi.org/10.1093/nar/22.22.4673

**Trost B, Kusalik A.** 2011. Computational prediction of eukaryotic phosphorylation sites. Bioinformatics **27**, 2927—2935.
http://dx.doi.org/10.1093/bioinformatics/btr525

**Wadood A, Riaz M, Shams S.** 2014. Structural Modeling and Molecular Dynamics Simulation Studies of Camel Milk Kappa Casein Protein. International Journal of Computational Bioinformatics and In Silico Modeling **3**, 483–490.

**Wang Y, Chen S, Li H.** 2010. Hydrogen peroxide stress stimulates phosphorylation of FoxO1 in rat aortic endothelial cells. Acta pharmacologica Sinica **31,** 160—164.
http://dx.doi.org/10.1038/aps.2009.201

**Wiederstein M, Sippl MJ.** 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Research **35,** W407–W410.
http://dx.doi.org/10.1093/nar/gkm290

**Zhang J, Johnson GV.** 2000. Tau protein is hyperphosphorylated in a site-specific manner in apoptotic neuronal PC12 cells. Journal of neurochemistry **75,** 2346—2357.
http://dx.doi.org/10.1046/j.14714159.2000.0752346.x

**Zhang L, Wang JT, Zhang DW, Zhang G, Guo SX.** 2014. Molecular characterization of a HMG-CoA reductase gene from a rare and endangered medicinal plant, Dendrobium officinale. Yao xue xue bao= Acta pharmaceutica Sinica **49**, 411—418.

**Zhang Z, Li Y, Lin B, Schroeder M, Huang B.** 2011. Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. Bioinformatics **27**, 2083—2088.
http://dx.doi.org/10.1093/bioinformatics/btr331