Annals of Phytomedicine 6(1): 30-44, 2017
DOI: 10.21276/ap.2017.6.1.5; Print ISSN : 2278-9839 and Online ISSN : 2393-9885
Ann. Phytomed., 6(1):30-44 (2017)

Original article

Comparative protein profile studies and *in silico* structural/functional analysis of HMGR (*Ap*HMGR) *in Andrographis paniculata* (Burm.f.) Wall. ex Nees

Byreddi Bhavani Venkata Bindu, Mote Srinath, Aayeti Shailaja and Charu Chandra Giri
Centre for Plant Molecular Biology (CPMB), Osmania University, Hyderabad-500007, Telangana State, India

Received May 7, 2017: Revised June 5, 2017: Accepted June 10, 2017: Published online June 30, 2017

Abstract

Andrographis paniculata (Burm.f.) Wall. ex Nees (Ap) contains andrographolide, *i.e*., synthesized via cytosolic MVA and plastidial MEP pathways using IPP and DMAPP as precursor molecules. HMGR catalyzes the first committed and rate-limiting step in MVA pathway. Limited information is available on structural and functional aspects of key enzyme HMGR. In a preliminary study, the *in vivo* leaves and light/dark grown callus was used for protein profile analysis. The callus growth and andrographolide content through HPLC analysis was carried out. SDS-PAGE total protein profile of *in vivo* leaves, light, and dark grown callus was investigated. A high intensity band of 62.3 KD protein indicating its proximity with HMGR enzyme was observed in leaves and light callus compared to dark grown callus. On this background, the present communication reports bioinformatics analysis of HMGR enzyme to decipher its functional and structural properties. The isoelectric point (pI), molecular weight and hydrophobicity were calculated using Expasy tools. The secondary (2D) and tertiary (3D) structures were predicted and validated using PROCHECK SAVES algorithm. Our finding on the allowed percentage (88.2%) of amino acid residues in the Ramachandran plot indicated that the simulated 3D structure was reliable. Phylogenetic analysis of *A. paniculata* HMGR (*Ap*HMGR) with different plants and other organisms including humans revealed close relationships and common lineage. The protein-protein interaction studies using STRING 10 tool revealed a close association of HMGR and nine other proteins in MVA and MEP pathways. The present analyzed data will give an insight on the structural and function attributes corresponding to this important HMGR protein. The molecular and biochemical understanding of HMGR protein will be advantageous for andrographolide yield enhancement through metabolic engineering.

Key words: *Andrographis paniculata* (Burm.f.) Wall. ex Nees, *Ap*HMGR, 3D structures, protein profile, phylogenetic analysis, protein-protein interactions

1. Introduction

Andrographis paniculata of family Acanthaceae, popularly known as king of bitters, is widely cultivated in Southern and South eastern Asia (Sareer *et al*., 2014). *A. paniculata* contains main bioactive compounds such as andrographolide (AD), neo-andrographolide (NAD), and 14-deoxy-11, 12-didehydroandrographolide (DAD) (Subramanian *et al*., 2012; Latha *et al*., 2017). These compounds are produced mainly by two distinct biosynthetic pathways, *i.e*., mevalonate (MVA) and methylerythritol 4-phosphate (MEP). MVA pathway occurs in cytosol whereas MEP pathway takes place in plastid (Figure 1). Until date, these two pathways have not been studied thoroughly and very less information is available on the key genes and enzymes of these pathways (Srivastava and Akhila, 2010). Recently, as per Lipko and Swiezewska (2016), MVA is one of the major biosynthetic routes associated with terpenoid synthesis, which provides the key precursors such as isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP). *A. paniculata* has several medicinal properties such as antidiarrhoeal, anti-inflammatory, hepatoprotective, anticancer, antidepressive and anti-HIV (Niranjan *et al.*, 2010; Suriyo *et al*., 2017).

3-hydroxy-3-methylglutaryl-coenzyme A reductase (HMGR) enzyme is found to be coded and regulated by multigene families in plants (Guo *et al*., 2015). The presence of several copies of these genes in plants attributes its relationship with many physiological processes, thus depicts its biological significance and importance. However, the number of structurally characterized proteins in plants is small and limited compared to number of known protein sequences. The 3D structure of proteins is stable and conserved than primary structure (amino acid sequence). Homology modelling has proven as the method of choice to generate reliable 3D model of a protein from its amino acid sequence (Prasada *et al*., 2010). Currently, *in silico* methods and tools have provided numerous advantages over time consuming wet-lab experiments on characterization of proteins (Proctor *et al*., 2016; Singh and Srivastava, 2016). The MVA is synthesized from HMG-CoA by the

Author for correspondence: Dr. Charu Chandra Giri
Professor, Centre for Plant Molecular Biology (CPMB), Osmania University, Hyderabad-500007, Telangana State, India
E-mail: giriccin@yahoo.co.in
Tel.: +91-040-27098087

Copyright @ 2017 Ukaaz Publications. All rights reserved.
Email: ukaaz@yahoo.com; Website: www.ukaazpublications.com

regulation of key rate limiting enzyme HMGR. In addition, HMGR has also shown to be a fundamental key enzyme of terpenoid biosynthesis in *A. paniculata* (Jha *et al*., 2011).

Further, in our laboratory we have a comprehensive research programme on the geographical distribution of *A. paniculata* and related species, haplotyping, screening and yield enhancement of bioactive compounds using biotechnological and molecular approaches (Neeraja *et al*., 2015; Arolla *et al*., 2015; Parlapally *et al*., 2015; Zaheer and Giri, 2015; Neeraja *et al*., 2016; Zaheer and Giri, 2017). Plant tissue culture is an alternative tool for production of bioactive compounds and callus culture system generally used for detection of biologically active molecules in cultured plant tissue *in vitro* (Dias *et al*., 2016; Giri and Zaheer, 2016; Abd El-Aal *et al*., 2016). Meagre information is available on gene and enzymes involved in diterpene lactone biosynthesis with particular reference to HMGR in non-model plant *A. paniculata*. HMGR being a key regulatory rate-limiting enzyme in MVA pathway, the study on *Ap*HMGR gains significance. Therefore, there is a need to understand and elucidate structural/functional properties of HMGR. HPLC analysis of *in vivo* leaves, callus grown in dark/light and andrographolide content was estimated to elucidate possible correlation of the enzyme and specific secondary metabolite synthesis. The protein profiling of *in vivo* leaves and callus using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was carried out to study proteins related to enzymes in andrographolide biosynthetic pathway. *In silico* analysis of HMGR protein is an attempt in this direction to generate further information. The present communication reports protein profile analysis and in detail *in silico* study on the structural, functional, phylogenic analysis and protein-protein interaction of *Ap*HMGR.

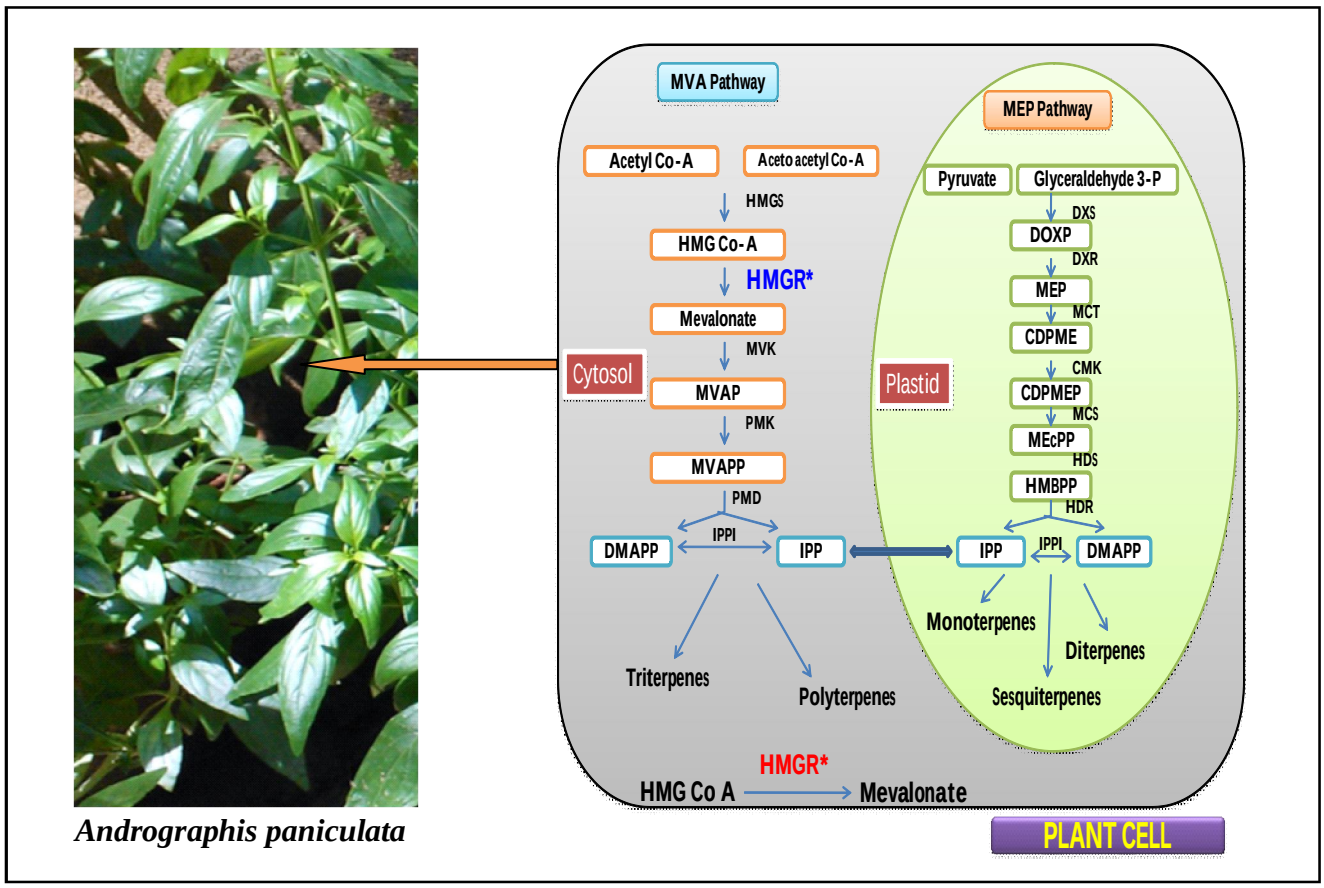


Figure 1: Net house grown *Andrographis paniculata* plant and MVA/MEP pathways showing key HMGR enzyme in *A. paniculata*.

2. Materials and Methods

2.1 Plant tissue culture, HPLC and protein profile analysis

2.1.1 Collection of seeds and establishment of *A. paniculata* plants *in vivo* and *in vitro*

The seeds of *A.paniculata* were sown on soil in net house conditions at Centre for Plant Molecular Biology (CPMB), Osmania University, Hyderabad, India. The plants were established by regular supplement of Hoagland nutrient solution (Hoagland and Arnon, 1938) twice a day. Seeds were collected from these established plants and used for different experiments in the laboratory. The green pods were surface sterilized with 0.1% (w/v) mercuric chloride for 10 min, followed by washing with sterile distilled water 5 times with five minutes for each wash. The surface sterilized pods were inoculated aseptically on MS (Murashige and Skoog, 1962) medium without plant growth regulators (PGRs) and 2.0 mg/l GA_3 supplemented media for germination and establishment of aseptic cultures. Seeds were incubated in dark for one week with subsequent transfer to light conditions and all the cultures were maintained at 25 ± 2ºC.

2.1.2 Induction of callus and growth study

The callus was induced, established using cotyledon and hypocotyl explants on MS media containing 1.0 mg/1 2, 4-dichlorophenoxy acetic acid (2, 4-D) and 0.5 mg/1 Kinetin (Kn). The callus induced and maintained both in dark/light conditions with same combination of MS medium. These cultures were maintained by sub-culturing at a regular interval of four weeks. The growth study of callus cultures were performed by taking fresh weight (FW) and dry weight (DW) of cultures every week until the sub-culturing time (up to 4 weeks). The fresh weight was taken, samples were kept for drying in an incubator for 72 h. at 60^0C and dry weight data was scored.

2.1.3 Statistical analysis

All the data on germination of seeds and callus growth in *A. paniculata* was recorded for analysis after four weeks of culture. Statistical analysis was carried out to calculate the mean and the standard error.

2.1.4 Extraction of andrographolide from leaf, callus and HPLC analysis

Samples of leaf and callus were oven-dried and ground using mortar and pestle. Dry powdered material of leaf and callus samples weighing 50 mg were transferred to vials. About 5.0 mL of HPLC-grade methanol was added to each sample and incubated for 24 h. at ambient temperature. Final extract was ultra-sonicated using an ultrasonic cleaning bath (Spectra Lab., Model UCB 30, India) for 30 min. The final extract was first filtered using Whatman filter paper No. 41 subsequent to sonication. The filtrate was further passed through 0.45 µm disposable hydrophilic PVDF membrane filters (Millex-HV, Millipore, Ireland) before its injection into HPLC system for analysis. The HPLC analysis was performed as per the protocol reported earlier from our laboratory (Zaheer and Giri, 2015).

2.1.5 Protein isolation and SDS-PAGE

Total protein was isolated from *in vivo* leaf, dark and light grown callus using acetone precipitation method as per Talei *et al*. (2014). The quantitative analysis was performed by Bradford method (Bradford, 1976). The spectrophotometric analysis of protein was done at 595 nm. Bovine serum albumin (BSA) was used as a standard protein for all the experiments. The qualitative analysis was performed using SDS-PAGE (Laemmli, 1970). The protein samples were run on 5% stacking gel and 12% separating gel at 90 volts per cm. The staining was done using 0.25% (w/v) coomassie brilliant blue G-250 in glacial acetic acid, methanol and water at 1:2:2 ratio. The bands were observed on image viewer and data was collected.

2.2 *In silico* study

2.2.1 Primary sequence collection and analysis of biochemical properties

*Ap*HMGR sequence with accession number AAL28015.2 was obtained from NCBI and biochemical properties were analyzed. Estimation of pI and molecular weight was carried out using Expasy tools website: Compute pI/Mw tool (Kyte and Doolittle, 1982). Hydropathy-plot analysis was performed using the ProtScale tool, available from the Expasy online server.

2.2.2 Prediction of secondary, tertiary (3D) structure and validation

The secondary structure of protein was predicted using CFSSP prediction server (Chou and Fasman, 1974; Dor *et al*., 2016). Analysis of protein sequence was carried out to predict the alpha helix, extended beta sheet and random coil structures contribution at each position based on 17 amino acid sequence window. SWISS-MODEL server was used for protein 3D structure prediction (Kiefer *et al*., 2009). Structure databases and protein sequence was identified for homology modelling. Energy levels and protein stabilization was taken into account to optimise model structure. Visualization of 3D structure was carried out using WebLab Viewer Lite 4. PROCHECK analysis was performed for the evaluation of 3D model stereo chemical qualities. 3D structure stereo chemical reliability was analyzed using Ramchandran plot (Laskowski *et al*., 1996). All 20 amino acids were analyzed considering most allowed and disallowed regions.

2.2.3 BLAST analysis and multiple sequence alignment (MSA)

BLASTp (Basic Local Alignment Search Tool- protein) was performed to compare the *Ap*HMGR protein sequence with other plant non-redundant protein data available in the NCBI. To find out the percentage of similarity and conserved regions between the *Ap*HMGR, *Homo sapiens* (*Hs*HMGR) and *Mus musculus* (*Mm*HMGR)multiple sequence alignment (CLUSTAL W2) was performed. The default parameters were considered for running the programme and analysing the data.

2.2.4 Motifs and domain analysis

The conserved motifs were analyzed among different plants of 30 species which are highly similar with *Ap*HMGR and other organisms like *Hs*HMGR and *Mm*HMGR using the program of multiple Em for motif elicitation (MEME; version 4.10.2). The parameters of MEME analysis were applied as follows: minimum width for each Motif: six; maximum width for each motif: fifty; maximum number of motifs to find is three and the numbers of repetitions are zero or one per sequence (Bailey *et al*., 2006). TMHMM server V.2.0 (CBS; Denmark) guided by hidden Markov model (HMM) was used to analyze trans-membrane helices in the *Ap*HMGR as per Krogh *et al*. (2001). The elucidation of *Ap*HMGR structural orientation was performed using Phyre2 (Kelley *et al*., 2015).

2.2.5 Analysis of ligand binding site, phylogenetic and protein-protein interaction studies

The evaluation of predicted 3D structure for ligand binding was carried out using bioinformatics tool for site prediction (Wass *et al*., 2010). The identification of homologous structures with bound ligands from structural library was assisted by predicted 3D structure. The elucidation of a ligand-binding site was accomplished by superimposing ligands onto predicted protein structure. Neighbor-Joining method was used to analyze the evolutionary trend. The optimal tree was depicted with the sum of branch length value (1.80027160). The clustering of associated taxa together visible as percentage of replicate trees was performed using bootstrap test (1000 replicates) and is being portrayed beside the branches. The tree is drawn to scale, with branch lengths (next to the branches) in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the

number of amino acid substitutions per site. The analysis involved 30 amino acid sequences. All positions containing gaps and missing data were eliminated. There were 66 positions in the final dataset. Evolutionary analyzes were conducted in MEGA6 (Saitou *et al*., 1987; Kumar and Gadagkar, 2000; Tamura *et al*., 2013). STRING 10 was used to find out the protein-protein interactions. *Arabidopsis thaliana* was used as reference sequence to draw protein-protein interactions (Szklarczyk *et al*., 2015).

3. Results and Discussion

3.1 Plant tissue culture, HPLC and protein profile analysis

3.1.1 Establishment of *A. paniculata* seedlings from pods *in vitro*

Germination of *A. paniculata* seeds was observed after three days of inoculation. These germinated seedlings were used as explants for induction of callus. The highest germination frequency (46.5 ± 5.5) was obtained on MS medium without hormone when compared to 37 ± 4.84 MS media supplemented with GA_3 (2.0 mg/l) and is depicted in Figure 2.

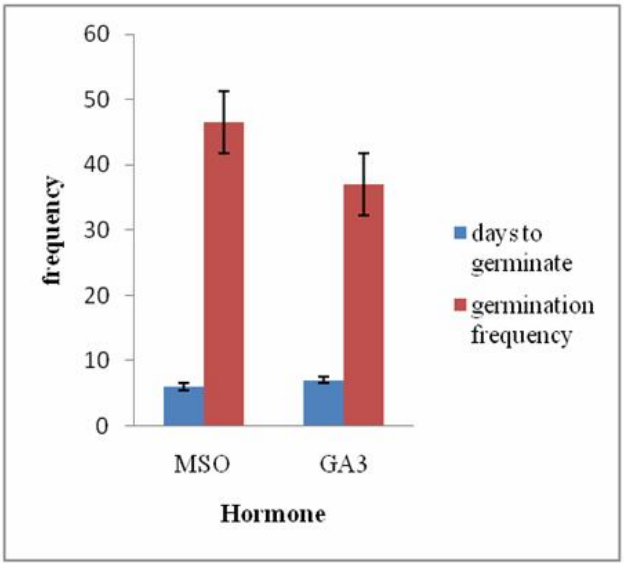


Figure 2: Comparative study for germination frequency of seeds in MSO and MS+GA_3.

3.1.2 Induction of callus, growth and HPLC analysis

The highest percentage of callus induction (100 ± 0) was observed from cotyledon explants and initiation within 4-5 days on MS medium containing 1.0 mg/1 2, 4-D and 0.5 mg/1 Kn. The callus induction frequency of 97±0 was observed within 5-6 days in hypocotyl explants on same callus induction medium (Figure 3, Figures 4 A, B). Callus cultures are generally used for detection of secondary metabolites *in vitro* and their further exploitation for production (Giri and Zaheer, 2016; Elenaet *et al*., 2016). The growth study of callus cultures up to four weeks revealed that the fresh and dry weights of callus cultures gradually increased (Figure 5).

The HPLC analysis of callus revealed that the andrographolide content was increased up to four weeks. There was increase in andrographolide content with increase in callus age. The andrographolide content of leaves was up to 1.667% DW following HPLC analysis. The andrographolide content of callus cultures was less compared to *in vivo* leaves (25 fold). The light callus andrographolide content was 1.46 fold higher than dark callus (Figure 6, Figures 7 A, B). The study on andrographolide content of callus in both light and dark is reported first time in *A. paniculata*. This plant contains two different pathways, *i.e.*, MEP and MVA for the synthesis of andrographolides. The MEP pathway takes place in plastid and MVA takes place in cytosol. Light has a direct influence on plant development, morphogenic potential and synthesis of active compounds in medicinal herbs (Fazal *et al*., 2016). Likewise light may perhaps have some influence in enhancing the andrographolide content. The molecular analysis of light callus can provide an insight into study the MEP pathway in detail.

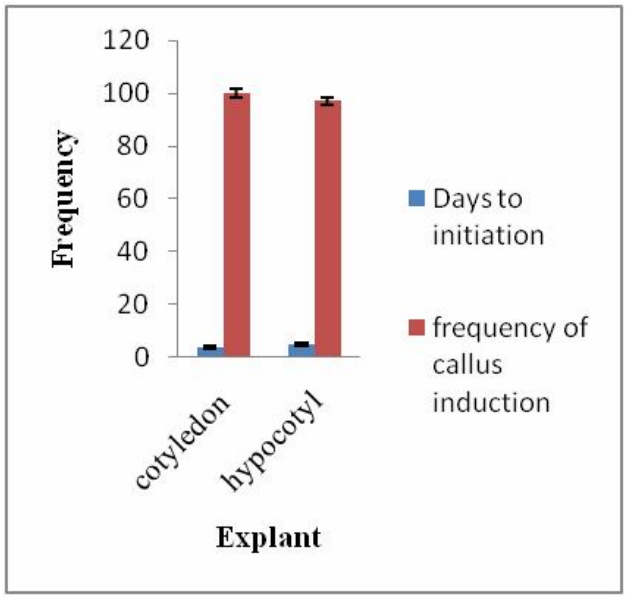
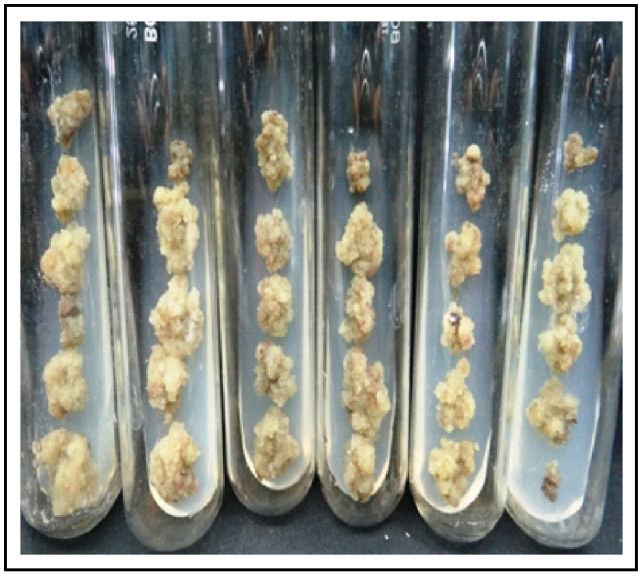
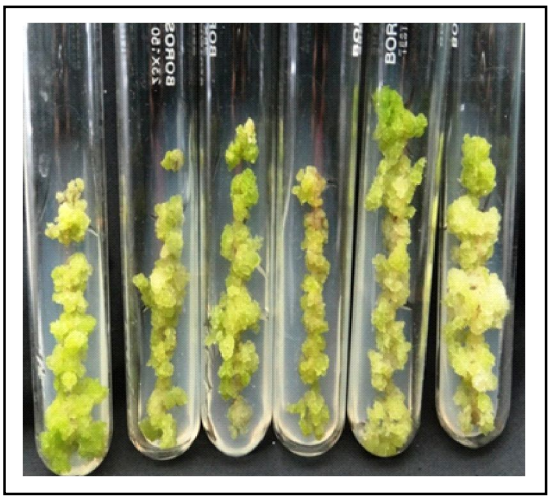


Figure 3 : Callus induction from cotyledon and hypocotyl explants.



(A)



(B)

Figure 4 : (A) Callus induced in dark and (B) Callus induced in light

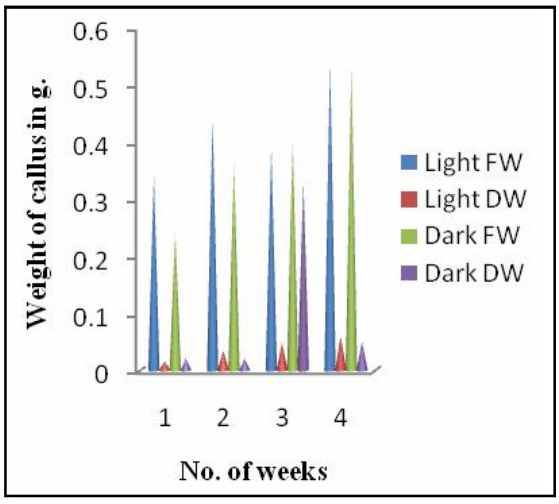


Figure 5 : Growth study of callus cultures in light and dark

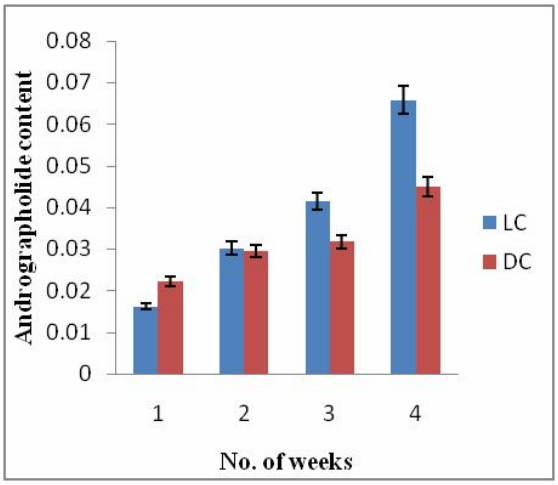
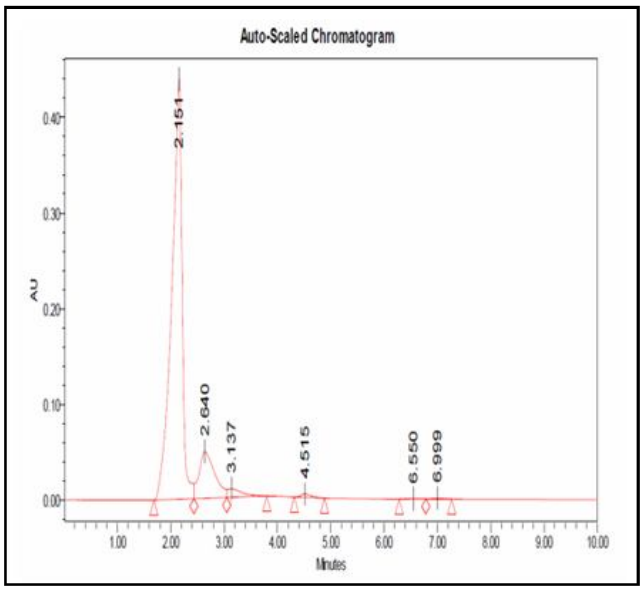
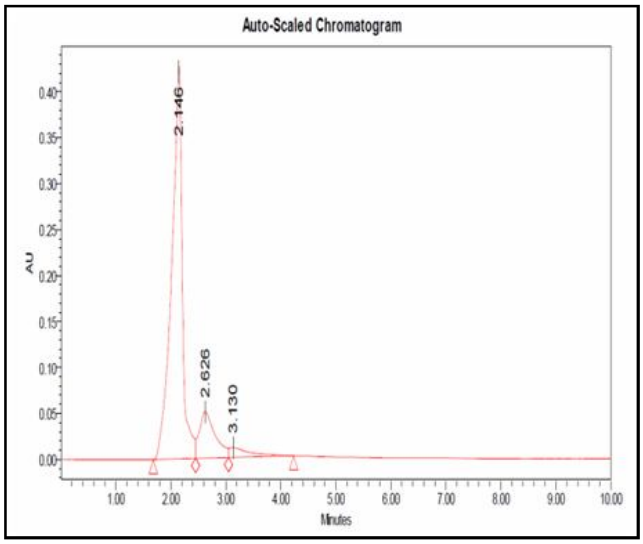


Figure 6 : HPLC analysis of callus cultures (Light and Dark)



(A)



(B)

Figure 7 : (A) HPLC analysis of Dark callus and (B) HPLC analysis of Light callus.

3.1.3 Protein profile analysis

The Bradford estimation determines that the total protein content in leaves of *A.paniculata* is 0.588±0.22. In callus cultures, it was observed that the total protein content in both light and dark induced callus was almost same, *i.e.*, 0.251±0.152 and 0.256±0.10, respectively. The qualitative analysis by SDS-PAGE showed that leaves are having 22 bands, though the protein content is almost same in light and dark callus, the number of bands varied (Figure 8). In SDS-PAGE protein profile analysis, 13 bands were observed with callus grown in light compared to 8 bands in callus grown in dark (Figure 9). The banding pattern was almost similar in light and dark callus samples. The extra bands found in light induced callus were also coinciding with *in vivo* leaf samples which are absent in dark callus. The intensity of bands indicated abundance of proteins

influenced by presence of light in both callus and *in vivo* leaves. The protein band intensity was higher in leaves followed by light callus and low in dark callus. A high intensity band of 62.3 KD protein indicating its proximity with HMGR enzyme was found in *in vivo* leaves and light callus compared to callus grown in dark. The similar studies were also reported in *Artemisia vulgaris* (Kumar *et al*., 2009), *Glycine max* L. Merr (Radhakrishnan and Ranjitha, 2009), and *Plumbago zeylanica* L. (Rout *et al*., 2010). Protein and enzyme analysis of stem in *Tinospora cordifolia* was also reported (Sharma and Batra, 2015). Some other medicinal plant species were too examined for their protein profile *in vivo* and *in vitro*, *viz*., *Bacopa monnieri* (Mohapatra and Rath, 2005).

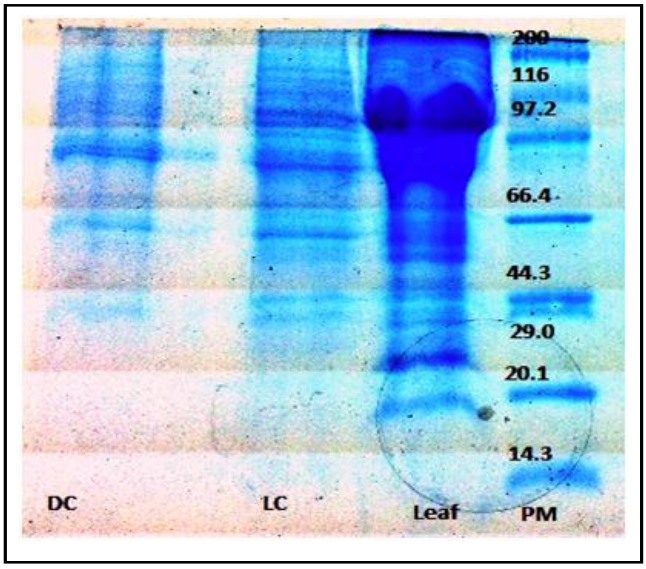


Figure 8 : Qualitative analysis was performed using SDS-PAGE DC-Dark callus; LC-Light callus; PM-Protein marker; 62.3 KD protein similar to HMGR.

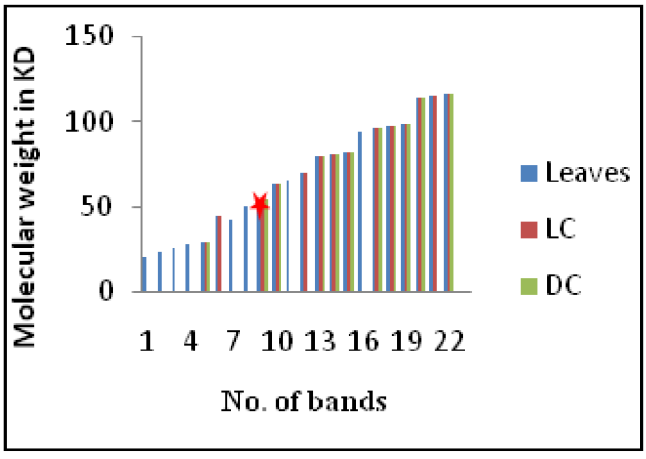


Figure 9 : Qualitative protein profile analysis showing presence and absence of bands with specific molecular weight in leaf, Light callus (LC) and dark callus (DC). The gaps between the bars in LC and DC indicates the absence of specific bands in SDS-PAGE profile.Lane 10 indicated with star represents 62.3 KD protein similar to HMGR

3.2 *In silico* study

3.2.1 *Ap*HMGR primary sequence and analysis of biochemical properties

The collected primary structures of *Ap*HMGR with Accession number AAL28015.2 comprise 595 amino acid residues. It indicated the gene size, which is approximately 1.7 kb (1785 base pairs). Elucidation of protein properties such as solubility and mobility are based on theoretical pI and being used to develop buffer system or salt bridge in expression studies (Dewald *et al*., 2015). The *Ap*HMGR protein has theoretical pI value of 5.75 and molecular weight of 63268.59 daltons (~63 KD). The theoretical molecular weight obtained from the present study will be helpful in isoenzyme studies and their purification. Generally, hydropathy plots are used for visualization of hydrophobicity of a peptide sequence length (Fung *et al.*, 2010). The hydrophobicity plot for *Ap*HMGR showed the percentage of buried amino acid residues inside protein core. The highest percentage is of Val (12.9%), which is mostly hydrophobic, with least of 0.5% in both Arg and Lys, which are mostly hydrophilic (Table 1). It indicated that the *Ap*HMGR N-terminus is highly hydrophobic with two highest hydrophobic peaks (Figure 10). Generally, scales of hydrophobicity were used to envisage the genetic code preservation. In an earlier study, the hydrophobicity profiles of the N-terminus showed two hydrophobic sequences in *Arabidopsis thaliana* (Uozumi *et al.*, 1997). This was similar to our finding where we identified residues, which were long enough to cover the membrane bilayer in *ApHMGR*.

Table 1: Percentage of hydrophobicity along each amino acid of *Ap*HMGR protein*

S.No.	Amino acid	Percentage (%)	S.No.	Amino acid	Percentage (%)
1.	Val	12.9	11	Asp	2.90
2.	Gly	11.8	12	Pro	2.70
3.	Leu	11.7	13	Tyr	2.60
4.	Ala	11.2	14	Trp	2.20
5.	Ile	8.6	15.	His	2.00
6.	Ser	8.0	16.	Met	1.90
7.	Phe	5.1	17.	Glu	1.80
8	Thr	4.90	18.	Gln	1.60
9	Cys	4.10	19.	Arg	0.5
10	Asn	2.90	20.	Lys	0.5

***** Data presented as per Hydropathy-plot analysis using the ProtScale tool

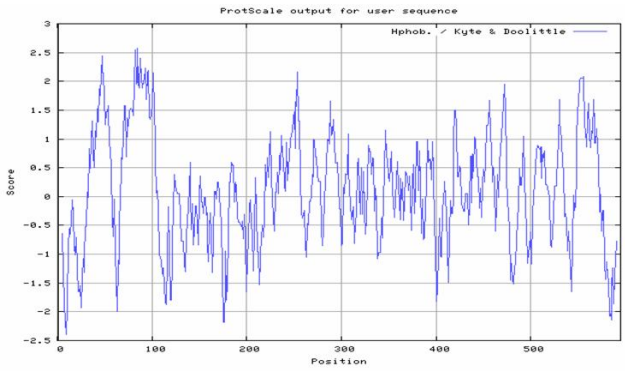


Figure 10 : The hydrophobicity plot of *Ap*HMGR protein showing peaks for each amino acid.

3.2.2 Prediction of secondary, tertiary (3D) structure and validation

The secondary structure analysis showed that 455 residues possible for the α-helix (H); extended sheet (E) 355 residues and turns (T) 66 residues. Hence, *Ap*HMGR consists of 76.5% helices, 59.7% extended sheets and 11.1% turns (Figure 11). α- helices and extended sheets are the most abundant structural elements of *Ap*HMGR penetrating through most part of the secondary structure. On the other hand, beta turns and random coils were found to be distributed intermittently in trans-membrane protein. The multiple sequence alignment by simultaneous assessment of many homologous sequences, the evolutionary conservation of secondary structures can be exploited. This can be achieved by calculating the net secondary structure propensity of an aligned column of amino acids. Development of related homologous protein three-dimensional structure requires homology or comparative modelling through the involvement of an atomic-resolution protein model from its primary structure. In general, structures are more conserved compared to sequences of homologous proteins. However, it has been found that, the sequence identity below 20% can have different structure and it will be functionally different (Xiang *et al*., 2006). Predicted 3D structures of proteins are of great importance while planning wet lab biological experiments in the absence of experimental 3D structures (Bhagavathi *et al*., 2014).

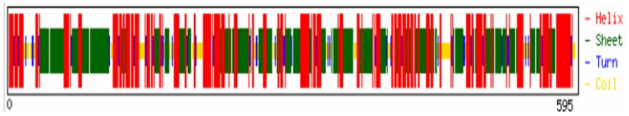


Figure 11 : Secondary structure of *Ap*HMGR showing 76.5% helices, 59.7% extended sheets and 11.1% turns.

The tertiary structure prediction using homology-modeling programme SWISS-MODEL revealed that, the template having maximum identity, the lowest e-value was selected and three models were created. The model having lowest energy was considered for further study and searched against the Uni-prot database. Forty templates found to match the target sequence out of 250-plant protein database. This database list then further filtered and 18 templates were finally selected. It has given three most suitable predicted 3D models with co-factors such as 2 CoA and 6 $NADP^+$ binding onto it. Model-I, II has 56.97% and Model-III 55.61% sequence identity, respectively. Model-I was visualized by WebLab Viewer Lite 4 depicting 3D view (Figure 12).The *Ap*HMGR is a homo-tetramer and the interaction between the enzyme and co-factor was non-covalent.

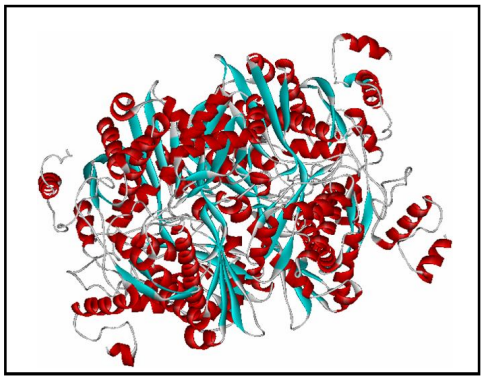


Figure 12 : The 3D structure of *Ap*HMGR established by homology-based modelling. The 3D structure shown as Schematic model by WebLab viewer Lite 4.

The quality of predicted 3D structure and its stereochemistry analysis by Ramchandran plot showed that its structure has good chain of 88.2% acceptable stereochemistry with least disallowed region 0.6% (Figure 13, Table 2). The Ramchandran plot for 20 different amino acids is shown in Figure 14. The result obtained was in credence with the 3D structure of *Centella asiatica* (84.9%) HMGR (Kalita *et al*., 2015), *Withania somnifera* (80.1%) HMGR (Sanchita *et al*., 2014), *Cyanobium* sp. (Siqueira *et al*., 2016), *Musa paradisiaca* (Hemmati, 2016). Recently, similar kind of studies is also reported in *A. paniculata* (Alexander *et al*., 2015) and *Cryptococcus laurentii* strain RY1 (Sarkar *et al*., 2017).

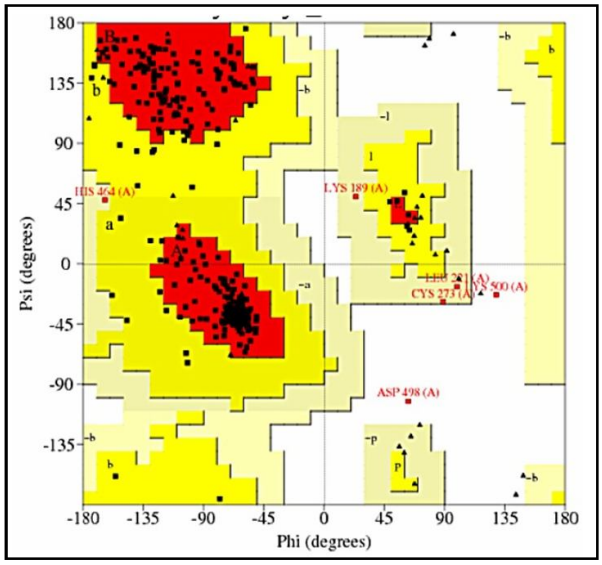


Figure 13 : Ramchandran plot analysis of *Ap*HMGR Red (core); Dark yellow (allowed); Light yellow (generously allowed); Triangles (Glycine).

Table 2 : Validation of modelled structure of *Ap*HMGR using various tools with respect to templates*

Tools	Parameter	*Ap*HMGR
Procheck Saves	Most favoured region (A, B, I in residues)	88.2%
	Additional allowed region (a, b, i, p)	10.1%
	Generously allowed region (~a,~b, ~i, ~p)	1.1%
	Disallowed region (empty white spaces)	0.6%

* Data presented for validation as per Procheck Saves

3.2.3 BLAST analysis and multiple sequence alignment (MSA)

Blast analysis of *Ap*HMGR showed similarity with 420 plant species. Identification of conserved domain was based on plants with query cover greater than 95% (Figure 15, Table 3). It contains 402 amino acid residues. 97 tetramerization sites, 9 inhibitor binding sites, 18 substrate binding sites, 20 NADPH binding sites and 4 catalytic residues.

The MSA (CLUSTAL W2) analysis of *Ap*HMGR showed 43.03% homology with *Homo sapiens* and 42.86% homology with *Mus musculus* HMGR (Figure 16). It indicated that the *H. sapiens* HMGR was matching with *Ap*HMGR. But the *H. sapiens* HMGR is very less identical (15.3%) compared to *Oryza sativa* (*Os* HMGR) in monocot (Darabi *et al*., 2012).

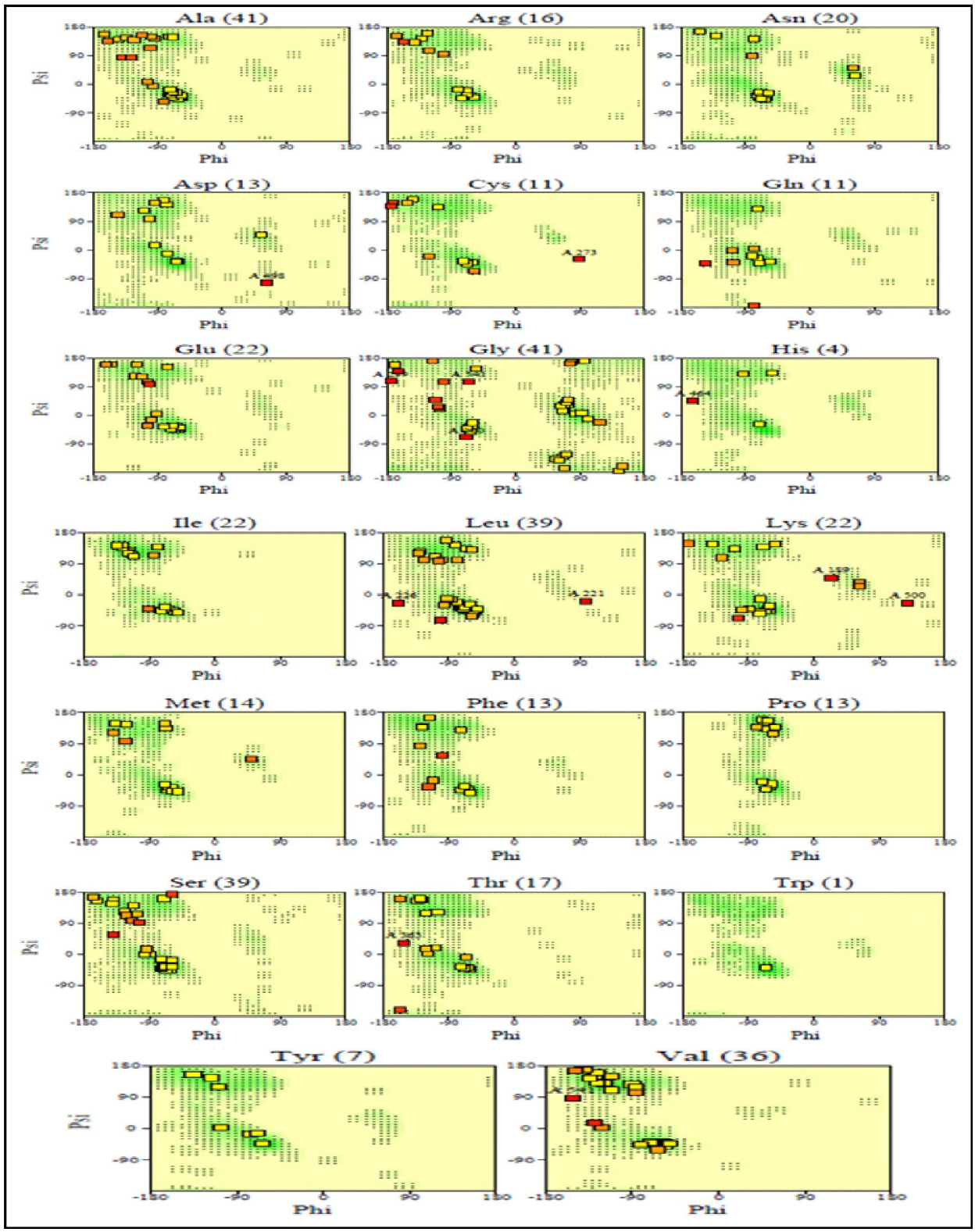


Figure 14: Ramachandran plot of *Ap*HMGR. The plot shows separate Ramachandran plots for each of the 20 different amino acid types. Numbers of amino acid residues are shown in brackets. Those in unfavourable conformations (Z score<-3.00) are labelled. Green shading on the background of yellow shows favourable conformations as obtained from an analysis of 163 structures at resolution 2.0A or better.

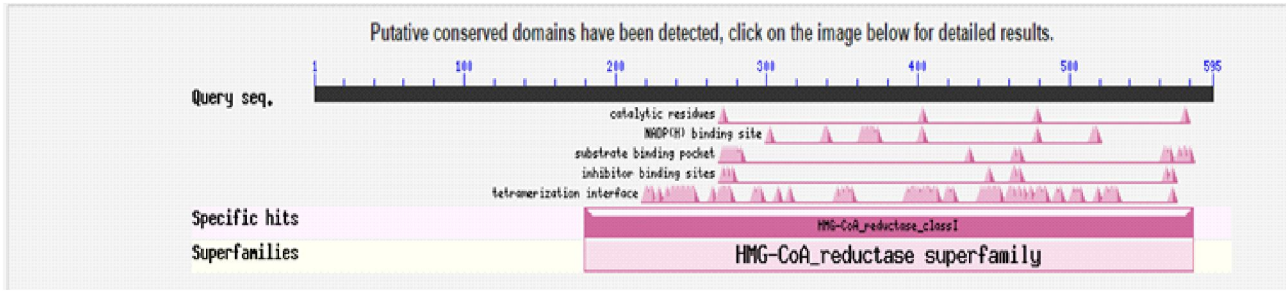


Figure 15 : BLAST analysis of *Ap*HMGR representing putative conserved domain with different binding sites for ligands, substrate, inhibitors, co-factors and amino acid sequence of residues length.

Table 3 : Blast analysis showing percentage query cover > 95% for *Ap*HMGR*

S.No	Plant Name	Family	Length	Accession no.
1	*Andrographis paniculata*	Acanthaceae	595	AAL28015.2
2	*Gossypium raimondii*	Malvaceae	585	gi763764699
3	*Pyrus pyrifolia*	Rosaceae	609	ACE80213.1
4	*Arabidopsis lyrata*	Brassicaceae	635	gi297333483
5	*Brassica napus*	Brassicaceae	581	gi674912770
6	*Bacopa monneri*	Plantaginaceae	589	gi323145669
7	*Solanum tuberosum*	Solanaceae	601	gi370344588
8	*Nicotiana sylvestris*	Solanaceae	604	gi700584204
9	*Panax ginseng*	Araliaceae	573	gi298370743
10	*Camellia sinensis*	Theaceae	590	gi345748131
11	*Lavandula augustifolia*	Lamiaceae	565	gi521953401
12	*Curcumis sativus*	Cucurbitaceae	102	gi1199472
13	*Jasminum sambac*	Oleaceae	594	gi724470974
14	*Ricinus communis*	Euphorbiaceae	586	gi255546682
15	*Dioscorea zinziberensis*	Dioscoreaceae	588	gi510031326
16	*Artmesia annua*	Asteraceae	567	gi5712672
17	*Camptotheca accuminata*	Cornaceae	593	gi289881
18	*Glycine soja*	Fabaceae	353	gi734307337
19	*Glycyrrhizia uralensis*	Fabaceae	573	gi306416887
20	*Medicago truncatula*	Fabaceae	583	gi357517535
21	*Pisum sativum*	Fabaceae	577	gi17224608
22	*Theobroma cacao*	Malvaceae	584	gi508707045
23	*Salvia miltiorrhiza*	Lamiaceae	565	gi187884600
24	Litchi chinensis	Sapindaceae	568	gi357240553
25	Morus notabilis	Moraceae	586	gi587850260
26	Cyanotis arachinoidea	Commelinaceae	589	gi641457656
27	Oryza sativa indica grp	Poaceae	576	gi1171364
28	Triticum aestivum	Poaceae	150	gi545473
29	Homo sapiens	Hominidae	888	gi110644584
3				

***** Data presented depicts >95% similarity with *Ap*HMGR

3.2.4 Motifs and domain analysis

A mathematical model is helpful in deriving motif sequences in biology and is similar to a defined group of organisms. In our study, three conserved motifs were found through MEME tool and are depicted in Figure 17, Figure 18 and Table 4. Domain prediction is helpful to find out the evolutionary relationship. Further, domain alignments convey potential information in simpler manner (Fong *et al*., 2008). The domain structure analysis of *Ap*HMGR using TMHMM server showed two transmembrane domains at 36 (lys)-58 (lys) and 78 (Ile)-100 (val) positions (Figure 19). Similar result was also found with *Os*HMGR (Darabi *et al.*, 2012).

PHYRE2 analysis revealed that *Ap*HMGR contains three transmembrane regions, namely S1 (39-57), S2 (95-70) and S3 (553-568). The N-terminal is present towards extracellular side and the C-terminal is present towards cytoplasmic side and was connected through linker sequences (Figure 20). The C-terminal portion of the protein, which included the catalytic domain, was highly conserved (74 to 98%) across all plant species as well as animal and yeast HMGRs with 65% identity. The linker region is highly divergent in both size and sequence among all HMGRs in plants and animals. The most striking difference is in the size of the N-terminal domain in plants, which contains the putative membrane-spanning region. This is highly conserved among plant species, which may be a reflection of functional attributes (Saur *et al*., 2015).

SeqA	Name	Length	SeqB	Name	Length	Score
1	gi|25990288|gb|AAL28015.2|AF389879_1	595	2	gi|160358778|ref|NP_032281.2|	887	42.86

SeqA	Name	Length	SeqB	Name	Length	Score
1	gi|25990288|gb|AAL28015.2|AF389879_1	595	2	gi|10644584|gb|AAG21343.1|	888	43.03

Figure16: Multiple sequence alignment (MSA) of amino acid sequences in *Ap*HMGR, *Hs*HMGR and *Mm*HMGR using CLUSTAL W2 programme.

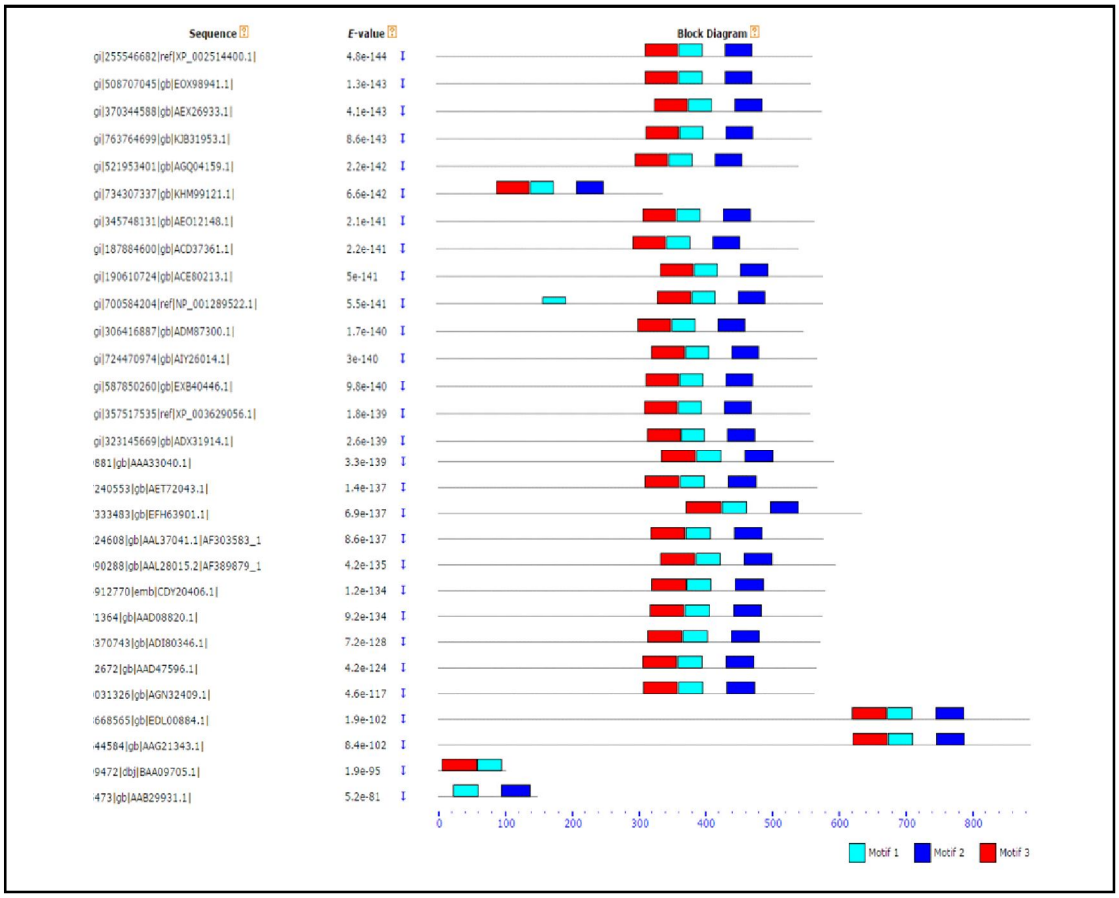
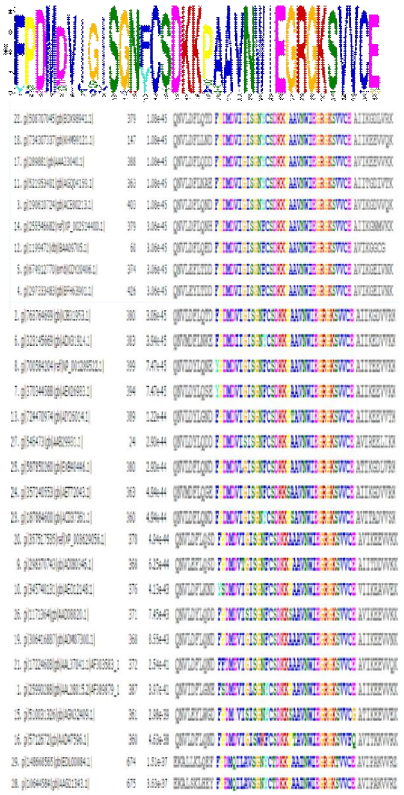
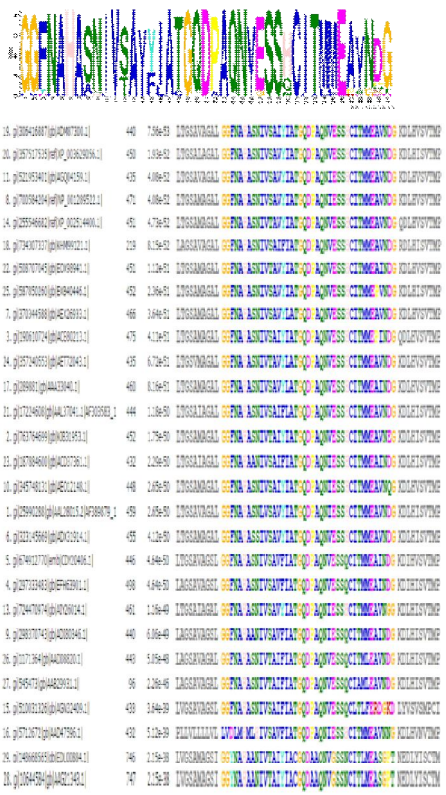


Figure 17 : Common motifs among the different plants and other species.

Motif-1



Motif-2



Motif-3

10. gi|345748131|gb|AEO12148.1|	323	2.38e-62	PHNFETLSLH FNRSSRFGRLQNIKCAIAGKNLYMRFSCSTGDAMGMNMVSKGVQNVLDFL KNDYSDMDVI
7. gi|370344588|gb|AEX26933.1|	341	1.70e-61	PLNFETLSLM FNKSSRFARLQGIQCAIAGKNLYMRFSCSTGDAMGMNMVSKGVQNVLDYL QSEYPDMDVI
16. gi|5712672|gb|AAD47596.1|	307	2.43e-61	GVNFDTLSVV FNKSSRFARLQNIQCSIAGKNLYIRFTCSTGDAMGMNMVSKGVQNVLDFL QNDFPDMDVI
2. gi|7637646999|gb|KJB31953.1|	327	2.43e-61	PDNFETLSVV FNRSSRFGRLQGIKCAIAGKNLYIRFTCSTGDAMGMNMVSKGVQNVLDFL QTDFPDMDVI
23. gi|187884600|gb|ACD37361.1|	307	3.23e-61	PLNFDTLSVV FNKSSRFARLQNVQCAIAGKNLYIRFSCSTGDAMGMNMVSKGVQNVLDFL QNDFPDMDVI
14. gi|255546682|ref|XP_002514400.1|	326	4.58e-61	PSNFDVISSA FNKSSRFGRLQNIKCAIAGKNLYLRFCCSTGDAMGMNMVSKGVQNVLDFL QNHFPDMDVI
22. gi|508707045|gb|EOX98941.1|	326	1.20e-60	PDNSETLAVV FNRSSRFARLQGIKCAIAGKNLYLRFSCSTGDAMGMNMVSKGVQNVLDFL QTDFPDMDVI
13. gi|724470974|gb|AIY26014.1|	336	2.20e-60	PANYETLCVV FNKSSRFARLQGIQCAIAGKNLYIRFNCSTGDAMGMNMVSKGVQNVLDYL GNDFPDMDVI
19. gi|306416887|gb|ADM87300.1|	315	3.47e-59	PLNFDSLSVV FNKSSRFARLQSIQYAIAGKNLYIRFRCSTGDAMGMNMVSKGVQNVLDFL QNDFPDMEVI
11. gi|521953401|gb|AGQ04159.1|	310	5.84e-59	PLNFETISLV FNSSSRFAKLQNIKCAIAGKNLYIRFSCSTGDAMGMNMISKGVQNVLDFL NAEFPDMDVI
12. gi|1199472|dbj|BAA09705.1|	7	8.61e-59	KLLVLV FTKSSRFARLQNIKCAIAGKNLYTRFSCSTGDAMGMNMVSKGVQNVLDFL QEDFPDMDVI
3. gi|190610724|gb|ACE80213.1|	350	1.11e-58	PNNYDTLSAV FNRSSRFGRLQTIKCAIAGKNLYMRFTCSTGDAMGMNMVSKGAQNVLDFL QNDFPDMDVI
8. gi|700584204|ref|NP_001289522.1|	346	1.63e-58	PVKFETLAAV FNQSSRFARLQRIQCAIAGKNLYMRFVCSTGDAMGMNMVSKGVQNVLDYL QNEYPDMDVI
25. gi|587850260|gb|EXB40446.1|	327	1.85e-58	TANFETLAMV FNRSSRFARLQSIKCAIAGKNLYTRFCCGTGDAMGMNMISKGVQNVLDFL QNDFPDMDVL
6. gi|323145669|gb|ADX31914.1|	330	1.85e-58	PLNFESLSLI FNSSSRFGRLQSIKCAVAGKNLYMRFTCSTGDAMGMNMVSKGVQNVMDFL NKEFPDMDII
18. gi|734307337|gb|KHM99121.1|	94	3.92e-58	PLNFDTLSLV FNRSSRFARLQGIQCAMAGKNAYLRFTCSTGDAMGMNMVSKGVQNVLDFL LNDFPDMDVI
1. gi|25990288|gb|AAL28015.2|AF389879_1	334	4.43e-58	PLNFETLSMV FNSSSRFGRLQSIKCAVAGKNLYIRFSCSTGDAMGMNMVSKGVQNVIDFL GNEFSDMEVI
26. gi|1171364|gb|AAD08820.1|	318	3.03e-57	PANFDTLAMV FNRSSRFARLQRVKCAVAGRNLYMRFSCSTGDAMGMNMVSKGVQNVLDYL QDDFPDMDVI
20. gi|357517535|ref|XP_003629056.1|	325	3.03e-57	PLNFDTLAVT FNKSSRFARLQSLQYTIAGKNLYIRFRCSTGDAMGMNMVSKGVQNVLDFL QSDFPDMDVI
17. gi|289881|gb|AAA33040.1|	335	3.83e-57	PLNFETLAAV FNSSSRFGKLQNIKCAIAGKNLYMRYSCSTGDAMGMNMISKGVQNVLDFL QDDFPDMDVI
24. gi|357240553|gb|AET72043.1|	310	6.10e-57	PANFETLSVV FNKSSRFGRLQNIKCAIAGKNLYIGFCCSSGDAMGMNMVSKGVQNVMDFL QGEFPDMDVI
4. gi|297333483|gb|EFH63901.1|	373	5.20e-56	PENFDTLTVV FNRSSRFARLQSVKCTIAGKNAYVRFCCSTGDAMGMNMVSKGVQNVLEYL TDDFPDMDVI
5. gi|674912770|emb|CDY20406.1|	321	1.12e-53	PENFDTLAVV FNRSIRFARLQSVKCTLAGKNAYVRFSCSTGDAMGMNMVSKGVQNVLEFL TDDFPDMDVI
15. gi|510031326|gb|AGN32409.1|	308	5.02e-53	PANFDTLATV FNRSSRFGRLQGIYCSLAGRNVYLRFSCRTGDAMGMNMVSKGVQNVLEYL WGDFPDMHVI
9. gi|298370743|gb|ADI80346.1|	315	2.25e-49	PFNFDTLAVV SKKSSRFARLQSIQCSMAGKNLYIRLCCSTVDVMGMNMVSKGVQNVLEFL QSDFPDMDVT
29. gi|148668565|gb|EDL00884.1|	621	1.38e-41	PEGFAVIKEA FDSTSRFARLQKLYVTMAGRNLYIRFQSRTGDAMGMNMISKGTEKALLKL QEFFPDMQIL

Figure 18: All motifs and their sequences for HMGR proteins in selected plants. The MEME motifs are shown as different-coloured boxes. (colour figure online generated) and amino acid residues of motif-1, 2 and 3.

Table 4: Motifs sequence showing width, e-value and similarity for HMGR*

Motif	Width	Sequence	e-value	Similarity		
				1	2	3
1	35	F(Y)P(SF)D(E)MD(EQH)V(I)I(LT)G(SA)I(VL) SG(W)N(K)Y (F)CS (T)DKKP(SA)A(TS)AV(I) NWIEGRGKS(T)VVC(F)E(QG)	1.1e-896	-	0.13	0.12
2	41	G(L)G(V)F(YD)N(L)A(M)HA(M)S(AL)N(SH)I(L)VS(T)AI(V)Y (F)I(L)AT(C)GQDP(A)AQNV(I)E (G)SSH(QN)CI(L)T(A)M(L) M(LF)EA(PR)V(ISD)N(G)D(PQNKGE)G(TD)	4.6e-958	0.13	-	0.13
3	50	F(S)N(DTK)K(RSQ)S(T)S(I)RFA(G)R(K)LQ(SGRKTN)I(VL) Q(KH) C(PVT)A(TS)I(VML)AGK(R)NL(AV)YI(MLVT)R(G) F(YL)C(STRQVN) C(S)S(RG)T(S)G(V)DA(V)MGMNMV(I)SK GV(TA)Q(E)N(K)V(A) L(MI)D(ESL)F(YK)L	1.1e-117	0.12	0.13	-

*Data presented as per MEME; version 4.10.2

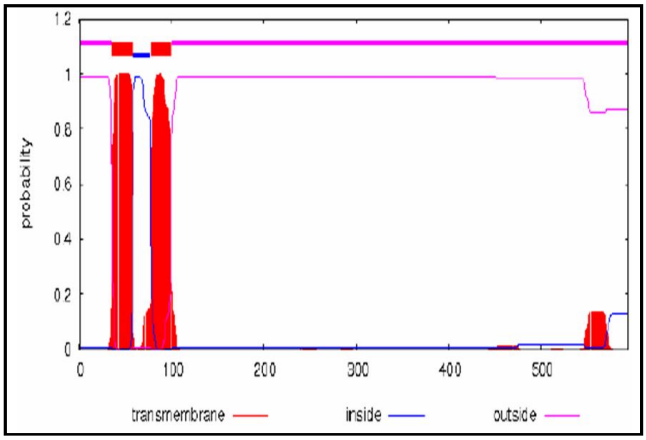


Figure 19 : Graphical representation of transmembrane domain prediction in *Ap*HMGR using TMHMM server.

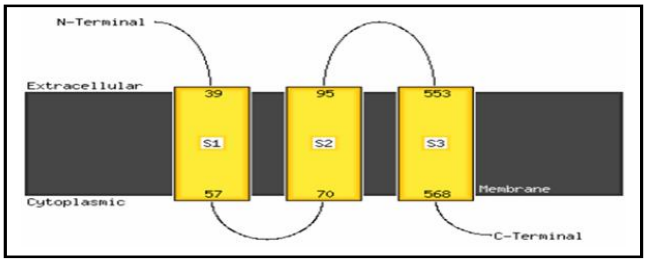


Figure 20 : Structural representation of transmembrane helices in *Ap*HMGR using PHYRE2 online tool.

3.2.5 Analysis of ligand binding site, phylogenetic and protein-protein interaction studies

Ligand a small signal-triggering molecule binds to proteins by feeble interactions. Drug design, drug side effects and elucidation of protein function is facilitated by protein-ligand binding site prediction studies (Krivák *et al.*, 2015). The ligand binding site of *Ap*HMGR contains amino acids Tyr (193), Glu (240) and Met (241) at the active site (Figure 21, Table 5). Two molecules of CoA and six molecules of ADP were found as ligands for *Ap*HMGR. These bind to enzyme non-covalently and provide energy to process the enzymatic reaction. The phylogenetic analysis usually compares the sequences of proteins and finds out identities among different organisms. Phylogenetic tree was constructed by amino acid sequences of *Ap*HMGR involving 28-plant species with higher similarity (query cover > 95%). Further, other organisms like *H. sapiens* and *Mus musculus* was also analyzed to find out the evolutionary relationship among different organisms. The result showed that HMGRs are derived from common ancestor gene of plants and animals. However, plants are categorized into monocots and dicots. Phylogenetic analysis revealed that both are closely related evolutionarily with bootstrap 21. In animals, it is again diverged with bootstrap 99. *Ap*HMGR was closely related to *Bacopa monneri* HMGR with bootstrap 33 (Figure 22). Further, in our study the monocot group included *O. sativa, Triticum aestivum, C arachinoidea* and *D. zinziberensis*. In this group, dicot plant *C. accuminata* also present with bootstrap 21 indicated close relationship with monocots. The similar studies were carried out in the starch-binding domain family CBM41 (Janecek *et al*., 2017). The biosynthesis of isoprenoids *via* MVA and MEP pathways is well studied in plants compared to other organisms. However, the regulation of *Ap*HMGR a key enzyme in MVA pathway still needs attention. Recently, the revised data of *Ap* chloroplast genome (150,249 bp) provided information regarding MEP pathway genes in chloroplast and is a positive endeavour in this direction (Ding *et al*., 2015). Likewise the complete chloroplast genome was annotated in *Scrophularia dentate* (Mehrotra *et al.*, 2016)

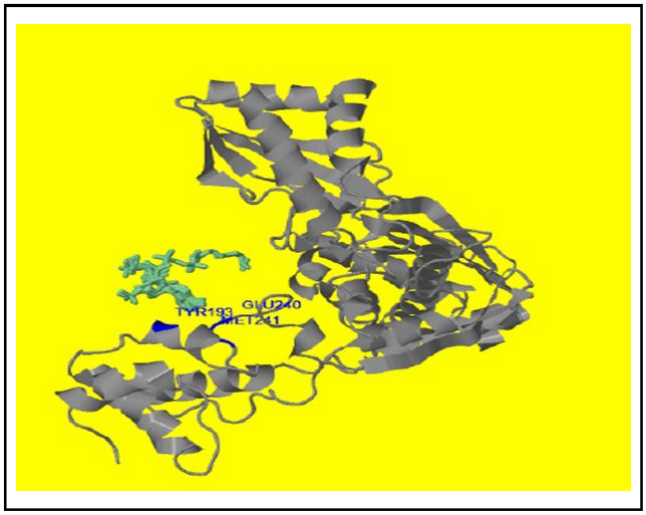


Figure 21: Ligand (green on the background of yellow) binding to *Ap*HMGR showing active sites with amino acids TYR, MET and GLU at 193, 241 and 240 positions.

Table 5 : Amino acids present at the active site of *Ap*HMGR protein for ligand binding*

S.No.	Amino acid	Contact sites	Position	Avg. distance
1	TYR	4	193	0.29
2	GLU	8	240	0.53
3	MET	8	241	0.11

***** Data presented as per 3D ligand site prediction tool

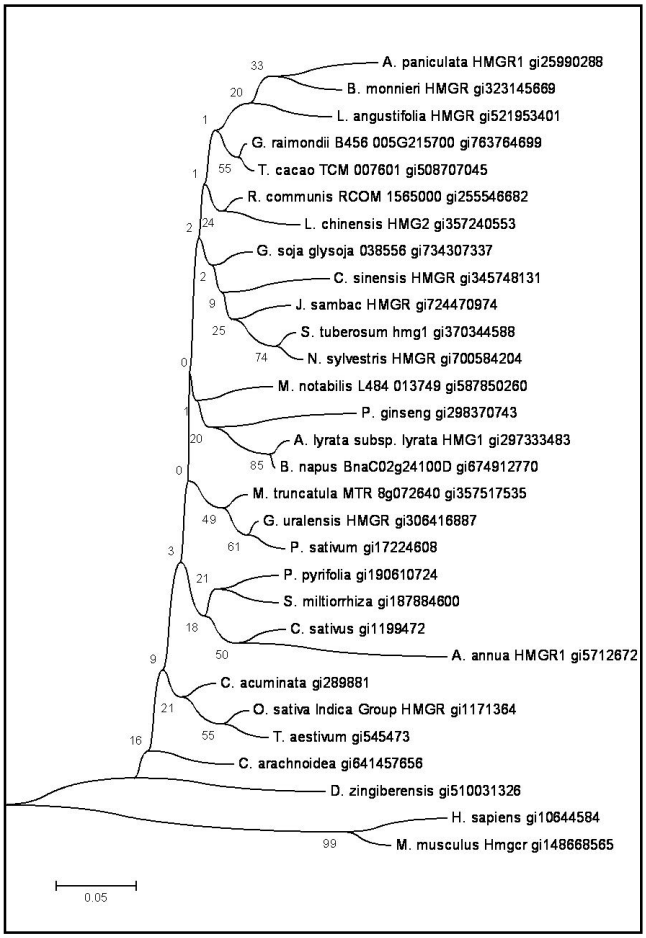


Figure 22 : The phylogenetic tree of *Ap*HMGR along with other HMGRs from selected plants and animals using the CLUSTAL-W (MEGA 6) program and Neighbor-Joining (NJ) method.

Protein-protein interactions study revealed that along with HMGR, and nine other proteins a close interaction was found which are involved in MVA and MEP pathways (Figure 23). Not only MVA and MEP proteins, other proteins involved in secondary metabolite pathway, cell cycle and signalling were studied. Co-expression analysis of 10 proteins with reference to *A. thaliana* and other organisms showed highest score for HMGS 0.999 and the least at 0.771. Remaining proteins have shown scores in between 0.998 to 0.834. The intensity of the spot indicated the expression levels shown on scale in Figure 24. These proteins carry out important functions like phosphorylation, condensation and decarboxylation, *etc.*

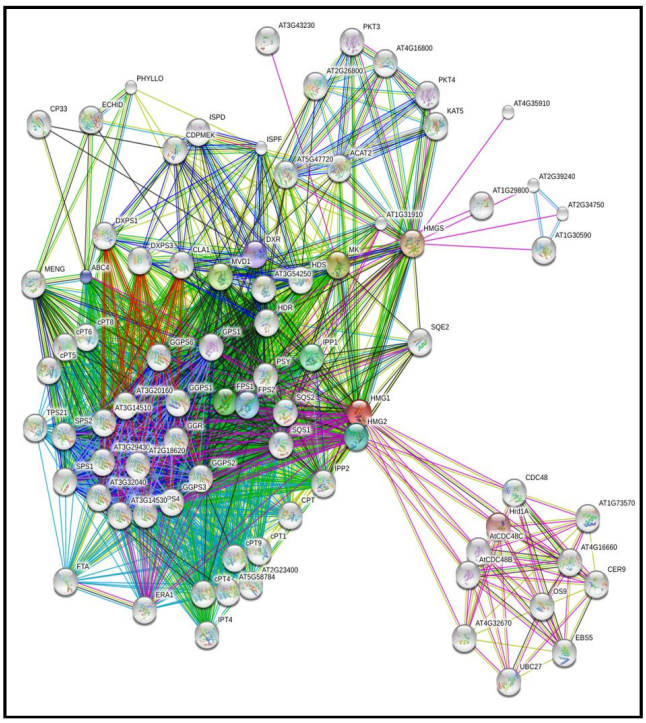


Figure 23 : Protein-protein interactions using STRING 10. *A. thaliana* used as reference sequence. Coloured nodes indicate direct interaction with the input.

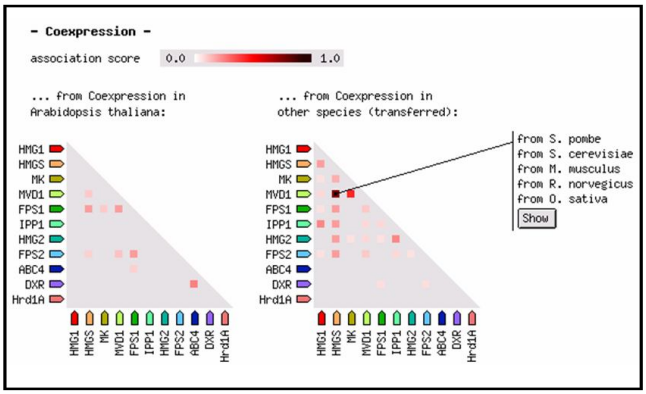


Figure 24 : Co-expression analysis of different proteins with reference plant and other microorganisms HMGS-hydroxy methyl glutaryl Co-A synthase;MK-mevalonate kinase; MVD-mevalonate diphosphate decarboxylase; FPS1-farnesyl diphosphate synthase; IPP-isopentenyl diphosphate isomerase; HMGR-3hydroxy3methylglutaryl Co-A reductase; FPS2- farnesyl diphosphate synthase; ABC-aberrant chloroplast development; DXR-1deoxy-D-xylulose 5-phosphate reductoisomerase; HRD-1-ubiquitin ligase.

4. Conclusion

The light grown callus promoted increased growth and secondary metabolite content in *A. paniculata*. More number of proteins was expressed in light grown callus with increased number and higher intensity of specific protein bands compared to dark grown callus. The *in vitro* studies on callus growth in both light/dark, HPLC and protein profile analysis of *in vivo* leaves and callus may be useful

to understand the possible role of light (MEP pathway in chloroplast) on andrographolide synthesis in *A. paniculata*. This study may help to find the genes and proteins responsible for the production of bioactive compounds such as andrographolide.

*Ap*HMGR structural analysis in the present study will be useful in functional prediction, regulation and location of specific proteins across organisms. Functional aspects of *Ap*HMGR can find out the ligand binding sites and active site information, for their use in docking studies. The phylogenetic analysis revealed the gene structure and protein architecture of all plant HMGRs were considerably conserved. This is the first report describing the structural, physicochemical and functional information of *Ap*HMGR, which will provide a good foundation and can be exploited for primer designing and isolation of genes from this important medicinal plant. The protein-protein interactions will be helpful to understand cell-cell interactions for the elucidation of biological mechanisms at the molecular level and drug-target relationships. These findings may also lead to the possibilities of metabolic engineering of isoprenoids, for increased synthesis of diterpene lactone andrographolide there by overcoming metabolic bottlenecks.

Acknowledgements

Authors would like to thank OU-UGC-CPEPA programme sponsored by University Grants Commission (UGC), New Delhi for financial support. Ms. B. B. V. Bindu, Mr. M. Srinath and Ms. A. Shailaja thank UGC, New Delhi for Research Fellowships.

Conflict of interest

We declare that we have no conflict of interest.

References

Abd El-Aal, S.M.; Rabie, K.A.E. and Manaf, H.H. (2016). The effect of UV-C on secondary metabolites production of *Echinacea purpurea* culture *in vitro*. J. Biol. Chem. Environ. Sci., **2**:465-483.

Alexander, R.A.; Sankari, S.; Ramaraj, K. and Rajaram, S.K. (2016). *In silico* analysis and molecular interactions studies of selected phytoconstituents from *Andrographis paniculata* as potential inhibitors of monoamine oxidase b. Asian J. Pharm. Clinical Res., **2**:239-242.

Arolla, R.G.; Cherukupalli, N.; Rao, K.V. and Reddy, V.D. (2015). DNA barcoding and haplotyping in different species of *Andrographis*. Biochem. Syst. Ecol., **62**:91-97.

Bailey, T.L.; Williams, N.; Misleh, C. and Li, W.W. (2006). MEME: Discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res., **34**:369-373. MEME; version 4.10.2) at website (http://meme.nbcr.net/meme/meme.html)

Bhagavathi, S.; Prakash, A. and Gulshan, W. (2014). An insight to virtual ligand screening methods for structure-based drug design and methods to predict protein structure and function in lung cancer: approaches and progress. J. Crit. Rev., **1**:10-24.

Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem., **72**: 248-254.

Chou, P.Y. and Fasman, G.D. (1974). Prediction of protein information. Biochem., **13**:222-245. CFSSP prediction server (http://cib.cf.ocha.ac.jp/bitool/MIX/)

Darabi, M.; Izadi-Darbandi, A.; Masoudi N.A. and Nemat, Z.G. (2012). Bioinformatics study of the 3-hydroxy-3-methylglotaryl-coenzyme A reductase (HMGR) gene in Gramineae. Mol. Biol. Rep., **39**:8925-35.

Daryush, T.; Alireza, V.; Mohd Y.R. and Mahmood, M. (2014). Proteomic analysis of the salt-responsive leaf and root proteins in the anticancer plant, *Andrographis paniculata* nees. PLoS One., **9**:1-10.

Dewald, I.; Isakin, O.; Schubert, J.; Kraus, T. and Chanana, M. (2015). Protein identity and environmental parameters determine the final physicochemical properties of protein-coated metal nanoparticles. J. Phys. ChemC., **119**:25482-25492.

Dias, M.I.; Sousa, M.J.; Alves, R.C. and Ferreira, I.C. (2016). Exploring plant tissue culture to improve the production of phenolic compounds: A review. Indust. Crops Prod., **82**:9-22.

Ding, P.; Shao, Y.; Li, Q.; Gao, J. Zhang, R.; Lai, X. and Zhang, H. (2015). The complete chloroplast genome sequence of the medicinal plant *Andrographis paniculata*. Mitochond. DNA, **4**:1-2.

Dor, O.; Zhou, Y. and Zhou. (2006). Achieving 80% tenfold cross-validated accuracy for secondary structure prediction by large-scale training. Proteins, **66**:838-845.

Elena, A.; Günter and Oxana V.P. (2016). Calcium pectinate gel beads obtained from callus cultures pectins as promising systems for colon-targeted drug delivery. Carbohydr. Polym., **147**:490-499.

Fazal, H.; Abbasi, B.H.; Ahmad, N.; Ali, M. and Ali, S. (2016). Sucrose induced osmotic stress and photoperiod regimes enhanced the biomass and production of antioxidant secondary metabolites in shake-flask suspension cultures of *Prunella vulgaris* L. Plant Cell Tiss. Organ Cult., **3**:573-581.

Fong, J.H. and Marchler, B.A. (2008). Protein subfamily assignment using the Conserved Domain Database. BMC. Res. Notes, **1**:114-120.

Fung, P.K.; Krushkal, J and Weathers, P.J. (2010). Computational analysis of the evolution of 1 deoxy d xylulose 5 phosphate reductoisomerase, an important enzyme in plant terpene biosynthesis. Chem. Biodivers., **7**:1098-1110.

Giri, C.C. and Zaheer, M. (2016). Chemical elicitors versus secondary metabolite production *in vitro* using plant cell, tissue and organ cultures: Recent trends and a sky eye view appraisal. Plant Cell Tiss. Organ Cult., **126**:1-18.

Guo, D.; Yi, H.Y.; Li, H.L.; Liu, C.; Yang, Z.P. and Peng, S.Q. (2015). Molecular characterization of HbCZF1, a *Hevea brasiliensis* CCCH-type zinc finger protein that regulates hmg1. Plant Cell Tiss. Organ Cult., **34**:1569-1578.

Hemmati, S. (2016). Predicting the functionality of major intrinsic proteins: An *in silico* analysis in *Musa*. Trends Phram. Sci., **2**:139-150.

Hoagland, D.R. and Arnon, D.I. (1938) The water culture method for growing plants without soil. California Agri. Exp. Station Circul., **347**: 32

Janecek, S.; Majzlová, K.; Svensson, B. and MacGregor, E.A. (2017). The starch-binding domain family CBM41-an *in silico* analysis of evolutionary relationships Proteins (Online First).

Jha, Z.; Sharam, S.N. and Sharma, D.K. (2011). Differential expression of 3-hydroxy- 3-methylglutaryl-coenzyme A reductase of *Andrographis paniculata* in andrographolide accumulation. J. Chem. Pharm. Res., **3**:499-504.

Kalita, R.; Patar, L.; Shasany, A.K.; Modi, M.K. and Sen, P. (2015). Molecular cloning, characterization and expression analysis of 3-hydroxy-3-methylglutaryl coenzyme A reductase gene from *Centella asiatica* L. Mol. Biol. Rep., **42**:1431-1439.

Kelley, L.A.; Mezulis, S.; Yates, C.M.; Wass, M.N. and Sternberg, M.J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc., **10**:845-58. Phyre2 An online tool (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index

Kiefer, F.; Arnold, K.; Künzli, M.; Bordoli, L. and Schwede, T. (2009). The SWISS-MODEL repository and associated resources. Nucleic Acids Res., **37**:387-392. SWISS-MODEL server (http://swissmodel.expasy.org/interactive/BYrhz7/models/)

Krivák, R. and Hoksza, D. (2015). Improving protein-ligand binding site prediction accuracy by classification of inner pocket points using local features. J. Cheminformatics, **1**:1-25.

Krogh, A.; Larsson, B; Von Heijne, G. and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J. Mol. Biol., **3**:567-580.

Kumar, P.S. and Kumari R.B.D. (2009). *In vitro* and *in vivo* identification of variation in protein expression in *Artemisia vulgaris* L. Adv. Biol. Res., **5-6**:63-70.

Kumar, S. and Gadagkar, S.R. (2000). Efficiency of the neighbor-joining method in reconstructing deep and shallow evolutionary relationships in large phylogenies. J. Mol. Evol., **51**:544-553.

Kyte, J. and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol., **1**:105-132. Expasy tools website (http://web.expasy.org/cgi-bin/compute_pi/pi_tool), (http://web.expasy.org/protscale/).

Laemmli, U.K. (1970). Cleavage of structural proteins during assembly of the head of bacteriophage T4. Nature, **227**:680-685.

Laskowski, R.A.; Rullmannn, J.A.; MacArthur, M.W.; Kaptein, R. and Thornton, J.M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR, **4**:477-486. PROCHECK (http://services.mbi.ucla.edu/SAVES/)

Latha, B.C.; Ahalya,.S.; Naidu, P.D.; Mounica, K. and Kumar, A.R. (2017). Phytochemical evaluation of *Andrographis paniculata*, *Cassia angustifolia* and *Eclipta alba*. Ind. J. Res. Pharm. Biotechnol., **5**:160-163

Lipko, A. and Swiezewska, E. (2016). Isoprenoid generating systems in plants-a handy toolbox how to assess contribution of the mevalonate and methylerythritol phosphate pathways to the biosynthetic process. Prog. Lipid Res., **63**:70-92

Mehrotra, P.; Ramakrishnan, G.; Dhandapani, G.; Srinivasan, N. and Madanan, M.G. (2016). The Complete Chloroplast Genome of Ye-Xing-Ba (*Scrophularia dentata*; Scrophulariaceae), an Alpine Tibetan Herb Mol. Biosyst., **13**:883-891

Mohapatra, H.P. and Rath, S.P. (2005) *In vitro* studies of *Bacopa monnieri*-an important medicinal plant with reference to its biochemical variations. Ind. J. Exp. Biol., **4**:373-376.

Murashige, T. and Skoog, F. (1962). A revised medium for rapid growth and bioassays with tobacco tissue cultures. Physiol. Plant., **15**:473-497.

Neeraja, C.; Krishna, P.H.; Reddy, C.S.; Giri, C.C.; Rao, K.V. and Reddy, V.D. (2015). Distribution of *Andrographis* species in different districts of Andhra Pradesh. Proc. Natl. Acad. Sci. India Sect. B. Biol. Sci., **85**:601-606.

Neeraja, C.; Mayur, D.; Mittapelli, S.R.; Rao, K.V. and Reddy, V.D. (2016). De novo assembly of leaf transcriptome in the medicinal plant *Andrographis paniculata*. Front. Plant Sci., **7**:1-13.

Niranjan, A.; Tewari, S.K. and Lehri, A. (2010). Biological activities of Kalmegh (*Andrographis paniculata* Nees) and its active principles: A review. Ind. J. Natl. Prod. Resources, **2**:125-135.

Uozumi, N.; Nakamura, T.; Schroeder, J.I. and Muto, S. (1997). Determination of transmembrane topology of an inward-rectifying potassium channel from *Arabidopsis thaliana* based on functional expression in *Escherichia coli*. Proc. Natl. Acad. Sci., **95**:9773-9778.

Parlapally, S.; Cherukupalli, N.; Bhumireddy, S.R.; Sripadi, P.; Anisetti, R.; Giri, C.C. and Reddy, V.D. (2015). Chemical profiling and anti-psoriatic activity of methanolic extract of *Andrographis nallamalayana* J.L. Ellies. Nat. Prod. Res., **30**:1256-1261.

Prasada, R.K.; Sharmaa, R. and Prajapati, G.L. (2010). Homology modeling and evaluation of human TEK tyrosine kinase using SWISS-MODEL workspace. J. Chem. Pharm. Res., **2**:440-451.

Proctor, E.A. and Dokholyan, N.V. (2016). Applications of discrete molecular dynamics in biology and medicine. Curr. Opin. Struct. Biol., **37**:9-13.

Radhakrishnan, R. and Ranjitha, K.B.D. (2009). Changes in protein content in micropropagated and conventional soybean plants (*Glycine max* (L.) Merr.). World J. Agric. Sci., **2**:186-189.

Rout, J.R.; Kanungo, S.; Das, R. and Sahoo, S.L. (2010). *In vivo* protein profiling and catalase activity of *Plumbago zeylanica* L. Nat. Sci., **1**:87-90.

Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol., **4**:406-425.

Sanchita, S.S. and Sharma, A. (2014). Bioinformatics approaches for structural and functional analysis of proteins in secondary metabolism in *Withania somnifera*. Mol. Biol. Rep., **41**:7323-7330.

Sareer O, Ahmad S, Umar S (2014). *Andrographis paniculata*: A critical appraisal of extraction, isolation and quantification of andrographolide and other active constituents. Nat. Prod. Res., **23**:2081-2101.

Sarkar, S.; Gupta, S.; Chakraborty, W.; Senapati, S. and Gachhui, R. (2017). Homology modeling, molecular docking and molecular dynamics studies of the catalytic domain of chitin deacetylase from *Cryptococcus laurentii* strain RY1. Int. J. Biol. Macromol. (Online First).

Saur, I.M.L.; Conlan, B.F and Rathjen, J.P. (2015). The N-Terminal domain of the tomato immune protein Prf contains multiple homotypic and Pto kinase interaction sites. J. Biol. Chem., **290**:11258-11267.

Sharma, A. and Batra, A. (2015). *In vivo* and *in vitro* protein profiling in *Tinospora cordifolia*-an antidiabetic plant species. Anal. Int. Daily J. Species, **51**:88-96.

Singh, S. and Srivastava, A.K. (2016). *In silico* and wet lab study revealed cadmium is the potent inhibitor of HupL in *Anabaena* sp. PC C 7120. Arch. Microbiol., **198**:27-34.

Siqueira, A.S.; Lima, A.R.J.; DallAgnol, L.T.; de Azevedo, J.S.N.; Vianezm J.L.D.S.G. and Gonçalves, E.C. (2016). Comparative modeling and molecular dynamics suggest high carboxylase activity of the *Cyanobium* sp. CACIAM14 RbcL protein. J. Mol. Model., **3**:1-8.

Srivastava, N. and Akhila, A. (2010). Biosynthesis of andrographolide in *Andrographis paniculata*. Phytochem., **71**:1298-1304.

Subramanian, R.; Asmawi, M.Z. and Sadikun, A. (2012). A bitter plant with a sweet future? A comprehensive review of an oriental medicinal plant: *Andrographis paniculata*. Phytochem. Rev., **11**:39-75.

Suriyo, T.; Pholphana, N.; Ungtrakul, T.; Rangkadilok, N.; Panomvana, D.; Thiantanawat, A.; Pongpun, W. And Satayavivad, J. (2017). Clinical Parameters following multiple oral dose administration of a standardized *Andrographis paniculata* capsule in healthy Thai subjects. Planta Med. (Online first).

Szklarczyk, D.; Franceschini, A.; Wyder, S.; Forslund, K.; Heller, D.; Huerta-Cepas, J.; Simonovic, M.; Roth, A.; Santos, A.; Tsafou, K.P.; Kuhn, M.; Bork, P.; Jensen, L.J. and von Mering, C. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res., **43**:447-452.

Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A. and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol. Biol. Evol., **30**:2725-2729.

Wass, M.N.; Kelley, L.A. and Sternberg, M.J. (2010). 3DLigandSite: predicting ligand-binding sites using similar structures. Nucleic Acids Res., **38**:469-73. Three dimensional (3D) ligand site (http://www.sbg.bio.ic.ac.uk/3dligandsite/)

Xiang, Z. (2006) Advances in homology protein structure modeling. Curr. Protein Pept. Sci., **7**:217-227.

Zaheer, M. and Giri, C.C. (2015). Multiple shoot induction and jasmonic versus salicylic acid driven elicitation for enhanced andrographolide production in *Andrographis paniculata*. Plant Cell Tiss. Organ Cult., **122**:553-563.

Zaheer, M. and Giri, C.C (2017) Enhanced diterpene lactone (andrographolide) production from elicited adventitious root cultures of *Andrographis paniculata*. Res. Chem. Intermed., **43**:2433-2444.