Molecular cloning and structural characterization of HMG-CoA reductase gene from *Catharanthus roseus* (L.) G. Donn. cv. Albus

M Z Abdin1*, U Kiran2 and S Aquil1

1Department of Biotechnology, Faculty of Science and 2Faculty of Interdisciplinary Research Studies
Jamia Hamdard, New Delhi 110 062, India

Received 14 September 2010; revised 15 January 2011; accepted 10 August 2011

The 3-hydroxy-3-methyl glutaryl-CoA reductase (HMGR) catalyzes the conversion of HMG-CoA to mevalonate, the first committed step in isoprenoid biosynthesis pathway in plants. HMG-CoA reductase gene was amplified from the *Catharanthus roseus* (L.) G. Donn. cv. Albus by polymerase chain reaction (PCR) with primers designed using published sequence of HMG-CoA reductase c-DNA of *C. roseus* cv. Little Delicata (Acc. No. M96068). PS2 SERVER was used to generate three dimensional (3-D) structure of the enzyme using human HMG-CoA reductase as template. The structure was evaluated at various web interfaced servers, i.e., PROCHECK, Profunc and PDBsum, for checking the stereo interfaced quality of the structure in terms of bonds, bond angles, dihedral angles, structural as well as functional domains. The generated model was visualized using the Rasmol. The results of these studies revealed that HMG-CoA reductase gene from cv. Albus had 99% sequence homology with *hmgr* cDNA of cv. Little Delicata. The amino acid sequence of the HMGR protein of cv. Albus was found closely related to the members of the family Solanaceae and distantly related to the members of the families Euphorbiaceae and Brassicaceae. The enzyme has N-terminal transmembrane domain and a C-terminal catalytic domain with active sites that can bind to HMG-CoA and NADPH2. The fold of the substrate domain is unique and resembles the prism with 28-residue helix forming the central core. The homology model of enzyme generated in the present study, hence, could be used in determining the mechanistic function of this important class of proteins.

**Keywords:** *Catharanthus roseus*, HMG-CoA reductase, isoprenoid biosynthesis, mevalonate pathway, PS2 Server, secondary plant metabolites

**Introduction**

Isoprenoids are produced in all organisms but are especially abundant and diverse in plants. Isoprenoid biosynthetic pathway in plants leads to a number of unique products, including growth regulators (cytokinins, gibberelins and abscisic acid), photosynthetic pigments, phytotoxins and a variety of specialized terpenoids, such as, monoterpenes, sesquiterpenes, polyterpenes etc1. The initial steps of the isoprenoid pathway [acetate/mevalonate (MVA) pathway] involve the fusion of 3 molecules of acetyl-CoA to produce 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA). The key enzyme of the classical mevalonate pathway in plants was identified as HMG-CoA reductase (HMGR, EC 1.1.1.34). It catalyzes the formation of mevalonate by two successive reductions of HMG-CoA, using two entities of NADPH as cofactor. Genome sequencing has identified *hmgr* genes in organisms from all three domains of life, and over 150 HMGR sequences are recorded in public databases. Higher animals, archaea and eubacteria have only a single *hmgr* gene, whereas the plants, which use both MVA-dependent and MVA-independent pathways to synthesize isoprenoids, have multiple HMGR isozymes that appear to have arisen by gene duplication and subsequent sequence divergence2.

Several studies have shown that HMG-CoA reductase is an important control point for the mevalonate synthesis in plants as well as in animals. It regulates the carbon flux from primary to the secondary metabolism for biosynthesis of secondary plant metabolites. The regulation of this enzyme leading to the increased production of isoprenoids has, thus, been reported in a range of plant species3-5.

*Catharanthus roseus* (L.) G. Donn. (Madagascar periwinkle) plants are the sole source of bisindole alkaloids, vinblastine and vincristine, the two medically important antitumor agents used in the chemical treatment of human cancers6,7. Since the concentration of both vinblastine and vincristine present in the plant has been very low (0.0005% dry wt),

*Author for correspondence:*
Tel: 91-11-26059688 Ext. 5583, 5581: Fax: 91-11-26059663
E-mail: mzabdin@rediffmail.com
attempts were made to produce these dimers or their direct monomeric precursors, catharanthine and vindoline, in \textit{C. roseus} tissue cultures\cite{5,6}. However, production of both alkaloids by \textit{de novo} synthesis using the callus or the suspension cultured cell of \textit{C. roseus} could not enhance the alkaloid concentration\cite{9,10}. Research on vinblastine and vincristine metabolic engineering showed that overexpressing a key enzyme in isoprenoid biosynthesis could elevate the level of these alkaloids\cite{11}. On this basis, it is hypothesized that vinblastine and vincristine biosynthesis and their accumulation in \textit{C. roseus} plant can be increased through modulation of HMGR activity and MVA level.

Although genes encoding HMGR have been isolated and characterized from several plant species including tomato\cite{12}, potato\cite{13,14}, raddish\cite{15}, \textit{Arabidopsis}\cite{16}, tobacco\cite{17}, \textit{C. roseus} cv. Little Delicata\cite{18}, wheat\cite{19} and the \textit{Hevea} rubber tree\cite{20}, no attempts were made to clone and characterize \textit{hmgr} gene from \textit{C. roseus} cv. Albus (CrAHMGR). In view of the potential medicinal and insecticidal importance of this species, an attempt has been made to isolate and elucidate the 3-D structure of HMGR of \textit{C. roseus} by homology modeling. Knowledge gained from its 3-D structure with functionally important domains and structural features would give much more insight to understand the function and regulatory mechanism of the enzyme at molecular level as well as to identify the target for metabolic engineering to enhance biosynthesis of alkaloids.

**Materials and Methods**

**Chemicals and Plant materials**

\textit{C. roseus} cv. Albus plants were obtained from the herbal garden of Jamia Hamdard, New Delhi. Bacterial strain used in present study was \textit{Escherichia coli} DH5-\alpha. All chemicals were purchased from M/s Sigma Chemicals Company (USA) and enzymes from M/s MBI Fermentas, unless otherwise specified.

**Gene Isolation**

Genomic DNA was isolated from leaves of \textit{C. roseus} cv. Albus using modified Doyle and Doyle method\cite{31} and \textit{hmgr} gene was amplified by polymerase chain reaction (PCR) using gene-specific primers. The primers were designed using the published c-DNA sequence of \textit{C. roseus} cv. Little Delicata (Acc. No. M96068)\cite{18}. The sequences of PCR primers were as follows: forward primer- 5'GGGGATCCATGGACTCTCGCCGCGATC3' and reverse primer- 5'GGTCGACTCATCAC TCTCTAACTGAGAG3'. PCR was performed in a reaction mixture (50 \(\mu\)L) consisting of 1\(\times\) reaction buffer, 0.2 mM dNTPs, 20 pmoles of each primer DNA, 0.5 \(\mu\)g template DNA and 1 unit of \textit{Taq} DNA polymerase. Reaction mixture was heated at 94°C for 4 min for melting of template DNA, followed by 32 cycles at 94°C for 1 min, annealing at 60°C for 1 min and extension at 72°C for 1 min. At the end of 32 cycles, additional final extension at 72°C for 5 min was carried out to extend any premature synthesis of DNA. Following electrophoresis, the desired DNA fragment of 3 kb from \textit{C. roseus} cv. Albus was isolated and purified using QIA quick® Gel Extraction Kit (Qiagen).

**Cloning and Sequencing**

The isolated product eluted from gel was ligated with pBluescript SK\(^+\) plasmid (Stratagene). The competent cells of \textit{E. coli} DH5-\alpha prepared by CaCl\(_2\) treatment were transformed with the recombinant vector. Recombinants were selected through blue-white screening on Luria agar containing 1 \(\mu\)g/mL ampicillin. Presence of the insert in single white colonies was confirmed by PCR with the same primer combination and sequenced using T7 primer at the automated DNA sequencing service, Bangalore Genei, India. The sequence was submitted to NCBI (Acc. No. AY623812).

**Similarity Search and Physiochemical Characterization**

The mRNA and amino acid sequences for \textit{hmgr} genes from various plant species available in the databank of National Centre for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov) were used to locate conserved boxes by multiple sequence alignment through CLUSTAL W 1.8.

Theoretical analysis of the nucleotide sequence and deduced amino acid sequence was carried out using various algorithms available on the ExPasyServer (http://www.expasy.org/tools). Physicochemical properties of the selected proteins were determined using the ProtParam tool\cite{22}. The hydropathy profile of the predicted \textit{C. roseus} cv. Albus (CrA) HMGR was calculated using TMHMM 2.0 (http://www.cbs.dtu.dk/services/TMHMM-2.0/).

**Homology Modeling**

The structural homologue to CrAHMGR protein sequence was found using BLAST-p against PDB database at www.ncbi.nlm.nih.gov/BLAST. Human HMG-COA reductase was taken as a reference
template (PDB ID: 1DQ8) after BLAST-p analysis. The similarity between the two sequences was calculated using CLUSTAL W. The 3-D structure was generated by submitting the deduced amino acid sequences of CrAHMGR protein to the PS2 Server. Thus generated 3-D structure was visualized by Rasmol. The functional domains of CrAHMGR were obtained by submitting 3-D structure to Profunc server at http://www.ebi.ac.uk. Prediction of secondary structure using 3-D model was done at Pdbsum, http://www.ebi.ac.uk. The stereochemical quality of the structure was analyzed using PROCHECK and WHAT IF serve. The Ramchandran map was generated to check peptide bond planarity, bond lengths, bond angles, and hydrogen bond geometry and side chain conformations of protein structures as a function of atomic resolution.

Results

Isolation, Cloning and Sequencing of HMGR

HMGR gene (3 Kb) was amplified from genomic DNA isolated from the leaves of C. roseus cv. Albus (CrA), cloned in pBluescript SK+ plasmid and subsequently confirmed by sequencing (Fig. 1a). The sequence was submitted to NCBI (Acc. No. AY623812). The complete sequence of the gene was found to be of 3095 bases with 58.90% AT and 41.10% GC content. The ORF search result showed that CrAHMGR gene contained a 1803 bp-ORF encoding a 601 amino acid protein. The exons are from 1-1019 bp, 1638-1813 bp, 1915-2261 bp and 2832-3095 bp (Fig. 1b).

Characterization of Deduced CrAHMGR Protein

The comparison between gene sequences of HMGR from the two varieties of C. roseus (Acc. Nos AY623812 & M96068) showed difference in 8 nucleotides (at 258, 295, 976, 999, 1014, 1042, 1092, 1534, 1549 & 1774). The comparison between the predicted translation products of the two varieties of C. roseus (Acc. Nos AAT52222.1 & Q03163) showed 99% homology with difference in 6 amino acids (at 99, 326, 348, 512, 517 & 592). The theoretical pI value and theoretical mol wt of CrAHMGR was 5.93 and 64 kDa, respectively.

Protein-protein BLAST and sequence alignment analysis showed that the deduced CrAHMGR amino acid sequence had high homology with HMGR sequences from other plant species, such as, C. roseus cv. Little Delicata (99%), Artemisia annua (68%), Solanum tuberosum (80%), Datura stramonium (83%), Camptotheca acuminata (73%), Hevea brasiliensis (74%), Nicotiana tabacum (77%), Lycopersicon esculentum (74%), Ginkgo biloba (64%). The N-terminal end of plant HMGRs were quite diverse in both length and composition, whereas the C-terminal catalytic domain showed high similarity (Fig. 2, Table 1).

Homology Modeling

BLAST-p analysis of CrAHMGR amino acid sequence against PDB database revealed the maximum sequence identity to human HMGR (Class I) (56%; PDB: 1DQ8A) and minimal to HMGR from Pseudomonas mevalonii (Class II) (17%; PDB: 1R71A), suggesting that it belongs to class I HMGRs.

The 3-D structure was generated by submitting the deduced amino acid sequences of CrAHMGR protein to the PS2 Server, using human HMGR crys...
structure as template. Comparison of CrAHMGR protein with human HMGR revealed 56% identity, which was good for homology-based modeling. The amino acid sequence of CrAHMGR showed the presence of two major domains: N-terminal transmembrane domain and C-terminal catalytic domain. The primary sequence of the initial 179 amino acid residues present at the N-terminus did not show any significant sequence homology to the template protein (1DQ8) sequence deposited at the PDB data bank; therefore, these sequence were excluded during the modeling. This initial sequence however, showed the presence of two transmembrane regions (Pro41-Leu63 & Leu84-Val106).

The PDB Sum results of 3-D structure of CrAHMGR showed the predominance of \( \alpha \)-helix (19) and \( \beta \)-strands (10) arranged in two sheets with 3 \( \beta \)-hairpins, 5 \( \beta \)-bulges, 20 helix-helix interfaces, 27 \( \beta \)-turns and 1 \( \gamma \)-turn in CrAHMGR. The Profunc results suggest the presence of two major domains, substrate (HMG-CoA) binding domain and co-factor (NADPH) binding domain. The fold of substrate domain showed the presence of three domains; N-terminal ‘N-domain’, large ‘L-domain’ and small ‘S-domain’. The N-domain is the smallest of the three domains and is \( \alpha \)-helical. The fold of the L-domain (residues 242-304 & 408-585) is unique to HMGRs and its architecture resembles a prism, with 28-residue \( \alpha \)-helix forming the central element. Results were obtained from Profunc server and visualized with RASMOL (cartoon).

The models validated by PROCHECK, essentially satisfy the stereo-chemical parameters with well-refined structures, at similar resolutions. The distribution of residues in the most favoured regions of the Ramachandran plot for CrAHMGR is 94%.

### Discussion

Plants synthesize over 22,000 known isoprenoid compounds\(^{26}\), which are involved in their growth and development. The enzyme HMG-CoA reductase catalyzes the conversion of HMG-CoA to mevalonate. There are several studies showing a strong correlation between HMG-CoA reductase and biosynthesis of isoprenoid compounds, as HMG-CoA reductase

---

**Table 1**—Comparison of C-terminal (a) the catalytically active residues and the phosphorylation (b) site residue

<table>
<thead>
<tr>
<th>Acc. No.</th>
<th>Plant name</th>
<th>Catalytically active residue site</th>
<th>Phosphorylation site</th>
</tr>
</thead>
<tbody>
<tr>
<td>gi</td>
<td>AAT52222</td>
<td>Catharanthus roseus cv. Albus</td>
<td>TEGCLVAS FCSDKPA</td>
</tr>
<tr>
<td>gi</td>
<td>167488</td>
<td>C. roseus cv. Little Delicata</td>
<td>TEGCLVAS FCSDKPA GQDPAQN HMKY NRSSRD</td>
</tr>
<tr>
<td>gi</td>
<td>169485</td>
<td>Solanum tuberosum</td>
<td>TEGCLVAS FCSDKPA GQDPAQN HMKY NRSSRD</td>
</tr>
<tr>
<td>gi</td>
<td>2654423</td>
<td>Nicotiana tabacum</td>
<td>TEGCLVAS FCSDKPA GQDPAQN HMKY NRSSRD</td>
</tr>
<tr>
<td>gi</td>
<td>2246450</td>
<td>Lycopersicum esculentum</td>
<td>TEGCLVAS FCSDKPA GQDPAQN HMKY NRSSRD</td>
</tr>
<tr>
<td>gi</td>
<td>12082120</td>
<td>S. tuberosum</td>
<td>TEGCLVAS FCSDKPA GQDPAQN HMKY NRSSRD</td>
</tr>
<tr>
<td>gi</td>
<td>4704615</td>
<td>Capsicum annuum</td>
<td>TEGCLVAS FCSDKPA GQDPAQN HMKY NRSSRD</td>
</tr>
<tr>
<td>AAQ11737</td>
<td>Datura stramonium</td>
<td>TEGCLVAS FCSDKPAA -</td>
<td>-</td>
</tr>
<tr>
<td>AAB69726</td>
<td>Camptotheca acuminata</td>
<td>TEGCLVAS FCSDKPA GQDPAQN HMKY NRSSRD</td>
<td></td>
</tr>
<tr>
<td>gi</td>
<td>52858428</td>
<td>Ginkgo biloba</td>
<td>TEGCLVAS FCSDKPA GQDPAQN HMKY NRSSRD</td>
</tr>
<tr>
<td>gi</td>
<td>6554470</td>
<td>Arabidopsis italiana</td>
<td>TEGCLVAS FCSDKPA GQDPAQN HMKY NRSSRD</td>
</tr>
<tr>
<td>gi</td>
<td>51599128</td>
<td>Hevea brasiliensis</td>
<td>TEGCLVAS FCSDKPA GQDPAQN HMKY NRSSRD</td>
</tr>
<tr>
<td>gi</td>
<td>75221430</td>
<td>Artemisia annua</td>
<td>TEGCLVAS FCSDKPA GQDPAQN HMKY NRSSRD</td>
</tr>
<tr>
<td>gi</td>
<td>75266529</td>
<td>A. annua</td>
<td>TEGCLVAS FCSDKPA GQDPAQN HMKY NRSSRD</td>
</tr>
<tr>
<td>gi</td>
<td>75221431</td>
<td>A. annua</td>
<td>TEGCLVAS FCSDKPA GQDPAQN HMKY NRSSRD</td>
</tr>
</tbody>
</table>

---

Fig. 3—Major domains of catalytic domain of CrAHMGR: N domain in red, L domain in green and S domain in purple. The L domain resembles prism with 27-residue \( \alpha \)-helix forming the central element. Results were obtained from Profunc server and visualized with RASMOL (cartoon).
regulates the carbon flux from primary to secondary metabolism leading to the synthesis of these compounds. The sequence of the amplified 3 Kb CrAHMGR gene (Acc. No. AY623812) was found to be of 3095 bases with more of adenine and thymine bases. The ORF search result showed that CrAHMGR gene contained 1803 bp-ORF with four exons, encoding a 601 amino-acid protein (Fig. 1b). The gene and the translated amino acid sequences of HMGR from the two varieties of *C. roseus* differed only by 8 nucleotides and 6 amino acids, respectively. The physicochemical analysis of CrAHMGR suggest a theoretical pI value of 5.93 and mol wt of 64 kDa, which was very similar to the previously reported *C. roseus* cv. Little Delicata.

Protein-protein BLAST and sequence alignment analysis showed that the deduced CrAHMGR amino acid sequence had high homology with HMGR sequences from other plant species, suggesting that the predicted protein belongs to the HMGR family. The N-terminal end of plant HMGRs were quite diverse in both length and composition, whereas the C-terminal catalytic domain showed high similarity (Fig 2, Table 1). The alignment results further suggest that CrAHMGR amino acid sequence is closely related to the members of family Solanaceae and is distinct and distantly related to the groups comprising of families Euphorbiaceae (*Hevea*) and Brassicaceae (*Arabidopsis*).

According to the sequence, HMGRs are classified into two distinct classes: eukaryotic HMGRs (class I) and prokaryotic HMGRs (class II). BLAST-p analysis of CrAHMGR amino acid sequence against PDB database revealed that it belongs to class I HMGRs and has highest sequence identity to human HMGR. The domain structure of plant, animal and fungal HMGRs is made up of a C-terminal catalytic domain and N-terminal transmembrane domain, connected by linker region. The amino acid sequence of CrAHMGR showed the presence of two major domains: N-terminal transmembrane domain that anchors the protein molecule to the endoplasmic reticulum membrane and C-terminal catalytic domain, which contains the active site that resides in the cytosol. The results suggested the presence of two transmembrane regions and it is in accordance of earlier reports where plant HMGRs were shown to belong class I HMGRs and contained two membrane spanning helices.

The 3-D structure of CrAHMGR showed 3 domains; N-terminal ‘N-domain’, large ‘L-domain’ and small ‘S-domain’. The N-domain is the smallest of the three domains and is α-helical. It connects the catalytic portion of HMGR to the membrane domain in full length protein. The fold of the L-domain (residues 242-304 & 408-585) is unique to HMGRs. Its architecture resembles a prism, with 28-residue helix forming the central core (Fig. 3). The S-domain (residues 306-381) is inserted into the L-domain and probably forms the binding site for NADP(H). Detailed analysis shows four catalytically active residues E273GC, DK405K, GQD481 and H579. It is proposed that the catalytic lysine form a hydrogen bond network with aspartate and glutamate that interacts with the carbonyl group of HMG-CoA. The active-site aspartate participates in both reductive stages of the overall reaction is central to the hydrogen bond network, and may be part of a proton shuttle. The active-site glutamate participates in the second reductive stage of the reaction and the active-site lysine appears to stabilize the mevaldyl-CoA intermediate. Further, it is proposed that the histidine protonates the departing CoA thioanion.

HMG-CoA reductase activity is regulated at transcriptional and post-transcriptional levels. In addition, HMG-CoA reductase enzyme activity is modulated directly by reversible phosphorylation of the protein by the AMP-activated protein kinase. In human HMGR-CoA, serine is located close to the catalytically important residue histidine at phosphorylation site in the primary structure and its phosphorylation reduces the activity of the enzyme. The presence of S in the conserved sequence NRS suggests that *C. roseus* HMGR may also be regulated in a similar manner.

Based on the results obtained in this study, it can be concluded that the nucleotide sequence of CrAHMGR has 99% homology with the *C. roseus* cv. Little Delicata HMG-CoA reductase cDNA. Phylogenetic analysis of the CrAHMGR amino acid sequence showed that it is closely related to the members of family Solanaceae and is distinct and distantly related to the groups comprising of the members of families Euphorbiaceae (*Hevea*) and Brassicaceae (*Arabidopsis*).

The analysis of structure build through homology modeling reveals that the *C. roseus* cv. Albus HMGR protein consists of two major domains: N-terminal transmembrane domain that probably anchors the
protein molecule to the endoplasmic reticulum membrane and C-terminal catalytic domain, which contains the active sites and probably resides in the cytosol where the sterol biosynthesis occurs. Further, the 3-D structure generated shows the presence of a substrate (HMG-CoA) binding domain and co-factor (NADPH) binding domain. The fold of substrate domain is unique and resembles a prism, with 28-residue helix forming the central core. The homology model, thus generated in this study, could aid in determining the mechanistic function of this important class of proteins.

Acknowledgement

The authors are thankful to the Jamia Hamdard, New Delhi, India, for providing the infrastructure and financial assistance. We are also thankful to Dr P Ananda Kumar, NRC on Plant Biotechnology, Indian Agricultural Research Institute, New Delhi, India, for providing laboratory facilities to clone the HMGR gene from C. roseus cv. Albus.

References


