



Publisher homepage: www.universepg.com, eISSN: 2663-6913

American Journal of Pure and Applied Biosciences

Journal homepage: www.universepg.com/journal/ajpab

American Journal of
Pure and
Applied Biosciences



OPEN ACCESS | Research Article



Computational Modeling and Quality Validation of HMG-CoA Reductase for Drug Design Applications

Rubaya¹ , Md. Mizanur Rahman² , and Shohel Mahmud^{1*} 

¹Animal Biotechnology Division, National Institute of Biotechnology, Ganakbari, Ashulia, Savar, Dhaka-1349, Bangladesh

²Department of Biotechnology and Genetic Engineering, Faculty of Biological Science, Islamic University, Kushtia 7003, Bangladesh

*Correspondence: slm.btge@gmail.com (Dr. Shohel Mahmud, Senior Scientific Officer, Animal Biotechnology Division, National Institute of Biotechnology (NIB), Ganakbari, Ashulia, Savar, Dhaka-1349).

Received Date: 20 October 2025 Accepted Date: 22 November 2025 Published Date: 29 November 2025

Abstract

The enzyme 3-hydroxy-3-methylglutaryl CoA reductase (HMG-CoA reductase or HMGR) catalyzes the conversion of HMG-CoA to mevalonate, which is the rate-limiting step in cholesterol and isoprenoid biosynthesis. HMG-CoA reductase (HMGR) is the main target of statins, which are commonly used medicines to lower cholesterol and prevent heart-related diseases. However, the use of statins may lead to harmful side effects, like muscle pain (myopathy) and muscle breakdown (rhabdomyolysis), creating a demand for alternative therapeutic options. This study aims to analyze and predict the 3D structure of HMGR from 12 different species, such as humans, fruit flies, plants, and protists, using a computer-based method to evaluate their potential for future drug design. Protein sequences were retrieved from the NCBI database and analyzed using various bioinformatics tools. The physicochemical properties and amino acid composition were computed by ExPASy's ProtParam tool, and the secondary structural features were determined by SOPMA and PHD servers. ClustalW, CELLO, and MEME, respectively, predicted the phylogenetic tree, subcellular localization, and structural motifs. Homology modeling was performed using the Swiss-Model Workspace with selected templates and visualized with PyMOL software. The predicted models were evaluated with PROCHECK analysis by Ramachandran plot and VERIFY 3D, and ERRAT values using the SAVES server. The aliphatic index and GRAVY values indicated that most HMGRs were thermally stable and hydrophobic. Among the selected HMG-CoA reductase sequences, the alpha helix dominated over the extended strand, beta turn, and random coil, according to the secondary structural properties. The phylogenetic analysis ensured the relationship among experimental species, and the predicted motifs indicated uniformity in the conserved sequences. The result of the predicted model indicated good quality and high reliability of the protein for most species, particularly for *H. sapiens* and *D. melanogaster*. This *in silico* study provides valuable insights into the structural and functional aspects of HMGR across species, laying a strong foundation for the development of safer and more effective cholesterol-lowering medications in the future.

Keywords: HMG-CoA reductase, Computational modeling, Homology modeling, and Phylogenetic analysis.

1. Introduction

HMG-CoA reductase (3-hydroxy-3-methyl-glutaryl-CoA reductase) is considered an important enzyme due to its catalytic activity. It catalyzes the metabolic processes that synthesize cholesterol, isoprenoids, and

other lipids by converting HMG-CoA to mevalonate. Within this metabolic pathway, HMG-CoA reductase controls the rate-limiting step of cholesterol synthesis (Friesen & Rodwell, 2004). Excessive synthesis of cholesterol has a major impact on human morbidity

and mortality, particularly from atherosclerosis, followed by myocardial infarction or stroke (Vance & Van den Bosch, 2000). This metabolic route, commonly known as the mevalonate or isoprenoid pathway, is highly conserved across eukaryotes, archaea, and certain bacteria (Buhaescu & Izzedine, 2007).

Drugs that inhibit HMG-CoA reductase are referred to as HMG-CoA reductase inhibitors, or "statins." Statins primarily target this enzyme and block the conversion of HMG-CoA to mevalonate (Jiang *et al.*, 2018). Statins inhibit the activity of 3-hydroxy-3-methylglutaryl-CoA reductase (HMGCR) by directly binding to its active site and inducing conformational changes (Stancu & Sima, 2001). Owing to this mechanism, statins are the most commonly prescribed drugs for the prevention of cardiovascular diseases (Wang *et al.*, 2015). Additionally, recent research has indicated that statins have protective effects against colorectal cancer, and by inhibiting HMGCR, they can reduce the risk of colorectal cancer by 43% with at least 5 years of use (Bonovas *et al.*, 2007; Lochhead & Chan, 2013). Currently, more than 25 million people worldwide are using statins to lower their risk of cardiovascular disorders (Lipkin *et al.*, 2010). But statins can cause myalgia and rhabdomyolysis, which is characterized by muscular injury, weakness, and dark urine, and severe rhabdomyolysis can cause acute kidney disease (Malani *et al.*, 2024; Petreski *et al.*, 2021). These side effects highlight the need for alternative therapeutic strategies (Gesto *et al.*, 2014; Yousuf *et al.*, 2023).

In this context, computational approaches have become indispensable in structure-based drug design, as they help in understanding the three-dimensional structure of the protein target to which the candidate drugs are intended to bind. Once the special target site of the protein is known, docking methods can be applied for designing drugs (Nishant *et al.*, 2011). Experimental techniques like NMR and X-ray crystallography are typically used to discover the three-dimensional structure of proteins, but these methods are highly expensive, time-consuming, and quite labor-intensive. With the help of various bioinformatics tools, this study examines evolutionary relationships, conserved motifs, and physicochemical parameters of HMGR proteins, thereby providing a UniversePG | www.universepg.com

primary step for designing a target drug for the reduction of cholesterol synthesis in humans.

2. Methodology

Retrieval of target and template sequences

Twelve full-length amino acid sequences of 3-hydroxy-3-methylglutaryl CoA reductase (HMGR) from *Drosophila melanogaster I* (AAF56175.1), *Drosophila melanogaster II* (NP_732900.1), *Homo sapiens* (NP_001124468.1), *Paris fargesii* (AEL-30660.1), *Catharanthus roseus* (AAA33108.1), *Trypanosoma grayi* (XP_009313205.1), *Leishmania donovani* (AF054499.1), *Brevibacillus leterosporus* (WP_022584705.1), *Vibrio alginolyticus* (WP_017820181.1), *Encephalitozoon cuniculi* GB-M1 (CAD25893.2), *Aspergillus oryzae* 100-8 (KDE-78541.1), and *Haloterrigena turkmenica* DSM 5511 (YP_003402108.1) were searched and retrieved from the NCBI database and saved in FASTA format. For clarity, the two *Drosophila melanogaster* sequences were designated as *D. melanogaster I* and *D. melanogaster II*. The sequences were organized according to species group: fruit fly, human, plants, protists, bacteria, fungi, and archaea.

Physicochemical characterization

Physicochemical properties of the protein, which include the theoretical isoelectric point (PI value) (Bjellqvist *et al.*, 1993), total number of negatively (Asp+Glu) and positively (Arg+Lys) charged residues, and extinction coefficients (Gill & von Hippel, 1989), instability index (Guruprasad *et al.*, 1990), aliphatic index (Ikai, 1980; Saikat *et al.*, 2020), and grand average of hydropathicity (GRAVY) (Kyte and Doolittle, 1982), were computed using ExPasy's Protparam server (Wilkins *et al.*, 1999). The molecular weight and amino acid composition (essential and non-essential residues) were also calculated using the same server.

Secondary structure prediction

Secondary structural features of amino acid sequences were predicted by the Self-Optimized Prediction Method Alignment (SOPMA) server (Geourjon & Deleage, 1995) and PHD server (Rost *et al.*, 1994). These tools classified regions of the proteins into alpha helices, beta strands, and random coils. The secondary structures from amino acid sequences were predicted

by setting default parameters in the web server (number of states: 3, similarity threshold: 8, and window width: 17).

Subcellular localization prediction

Subcellular localization was predicted to provide insights into the biological functions of the proteins (Gillani & Pollastri, 2024). For fruit fly, plant, and fungal sequences, localization was predicted using WoLF PSORT (Horton *et al.*, 2007). Subcellular localization of human, protists, and bacteria was determined by CELLO (Yu *et al.*, 2004), while archaeal sequences were analyzed by PSORT-b (Gardy *et al.*, 2003).

Multiple sequence alignment and construction of a phylogenetic tree

A multiple sequence alignment was conducted using ClustalW (Thompson, Higgins, & Gibson, 1994) software to compare sequences of HMG-CoA reductase from human and other species. A phylogenetic tree was constructed using the UPGMA method (Schlee, 1975) and were aligned with each other by using ClustalW software (Thompson *et al.*, 1994).

Motif analysis

Protein sequence motifs were predicted using Multiple EM for Motif Elicitation (MEME) version 4.10.0 (Bailey *et al.*, 2006) and Motif Alignment and Search Tool (MAST) version 4.10.0 (Bailey & Gribskov, 1998). The parameters were set as follows: the minimum and maximum motif widths were 6 and 50, respectively; the maximum number of motifs was limited to 10; and each sequence was allowed zero or one occurrence of a motif.

Homology modeling and visualization

Three-dimensional structures of the HMGR proteins were generated using the SWISS-MODEL workspace (<https://swissmodel.expasy.org>) (Arnold *et al.*, 2006). The predicted models were visualized and analyzed using PyMOL version 1.3 (DeLano, 2002).

Protein 3D model quality assessment

The stereochemical quality of the predicted protein models was assessed using PROCHECK (Laskowski *et al.*, 1996) by Ramachandran plot analysis (Spencer *et al.*, 2019), through PDBsum (Laskowski *et al.*, 2005). Additionally, ERRAT (Colovos & Yeates, UniversePG | www.universepg.com

1993), and VERIFY 3D (Bowie *et al.*, 1991; Eisenberg *et al.*, 1997) scores by using the Structure Analysis and Verification (SAVES) Server version 4 (<https://saves.mbi.ucla.edu/>) were employed for further evaluation of the HMG-CoA reductase protein models.

3. Results

Physico-chemical characterization

The physicochemical properties of each of the proteins were determined by the ProtParam server (Wilkins *et al.*, 1999). These characteristics include the molecular weight, theoretical pI, and extinction coefficient (Gill & von Hippel, 1989), instability index (Guruprasad *et al.*, 1990), aliphatic index (Ikai, 1980), grand average of hydropathicity (GRAVY) (Kyte & Doolittle, 1982) and the absolute number of negatively and positively charged residues, which are shown in **Table 1**, and the essential amino acid compositions are shown in **Fig. 1**. Understanding the theoretical pI, which is the pH at which a specific molecule or surface has no net electrical charge, helps understand the charge stability of proteins (Enany, 2014). The calculated isoelectric points (pI) of HMG-CoA reductase proteins ranged from 4.43 to 9.01. Determination of the Instability index results in an understanding of the stability of the protein. An Instability index value exceeding the value of 40 indicates an unstable nature of the protein, while the protein whose Instability index value is less than 40, predicted as stable (Guruprasad *et al.*, 1990). The instability index value indicated that HMG-CoA reductase proteins from *H. sapiens* (51.39) and *C. roseus* (52.25), *P. fargesii* (44.59), *D. melanogaster I and II* (41.89) are unstable; on the other hand, proteins of the rest of the species were stable (instability index < 40), indicating better robustness.

A protein that has a high aliphatic index is thermostable over a broad temperature range (Ikai, 1980). Value of AI, an indicator of thermostability, was highest in *H. sapiens* (97.27), closely followed by *C. roseus* (96.74), *D. melanogaster* (96.58), and *P. fargesii* (95.48), while the lowest was in *E. cuniculi* (87.66), indicating reduced thermal stability. Proteins with positive or negative grand average hydropathy (GRAVY) values are considered hydrophobic or hydrophilic, respectively (Kyte & Doolittle, 1982). GRAVY (Grand Average of Hydropathicity) analysis

suggested that most HMG-CoA reductase proteins, including *H. sapiens*, were hydrophobic. Exceptions included *B. laterosporus* (-0.013), *E. cuniculi* (-0.108), and *H. turkmenica* (-0.084), indicating hydrophilicity.

The percentages of essential amino acids show that all species contain higher amounts of leucine and valine compared to other essential amino acids.

Table 1: Parameters computed using ExPASy's ProtParam tools.

Species	No. of amino acids	Molecular weight	Total no. of atoms	Theoretical I (P ^I)*	EC* at 280nm M ⁻¹ Cm ⁻¹	Instability index	Aliphatic index	GRAVY
<i>D. melanogaster I</i>	920	98317.1	13831	6.18	61445-59820	41.89	96.58	0.147
<i>D. melanogaster II</i>	920	98317.1	13831	6.18	61445-59820	41.89	96.58	0.147
<i>H. sapiens</i>	835	92020.6	12991	6.57	66695-65320	51.39	97.27	0.057
<i>P. fargessii</i>	575	60966.6	8592	6.64	36745-35870	44.59	95.48	0.213
<i>C. roseus</i>	601	64106.9	9065	5.93	35130-34380	52.25	96.74	0.107
<i>T. grayi</i>	435	45961.9	6492	8.37	30660-29910	34.57	93.29	0.036
<i>L. donovani</i>	434	45827.0	6467	7.93	26190-25440	38.46	94.19	0.127
<i>B. laterosporus</i>	432	46640.2	6546	6.38	36245-35870	32.57	88.73	-0.013
<i>V. alginolyticus</i>	420	44633.8	6271	6.37	24910-24410	37.89	93.88	0.031
<i>E. cuniculi</i>	398	43461.9	6083	8.36	32610-31860	35.90	87.66	-0.108
<i>A. oryzae</i>	1044	11226.0	15848	9.01	11156-11081	38.94	94.35	0.115
<i>H. turkmenica</i>	408	42125.7	5868	4.43	14565-14440	27.74	90.07	-0.084

*Extinction coefficient (EC), *Isoelectric Point (P^I), *Grand average of hydropathicity (GRAVY)

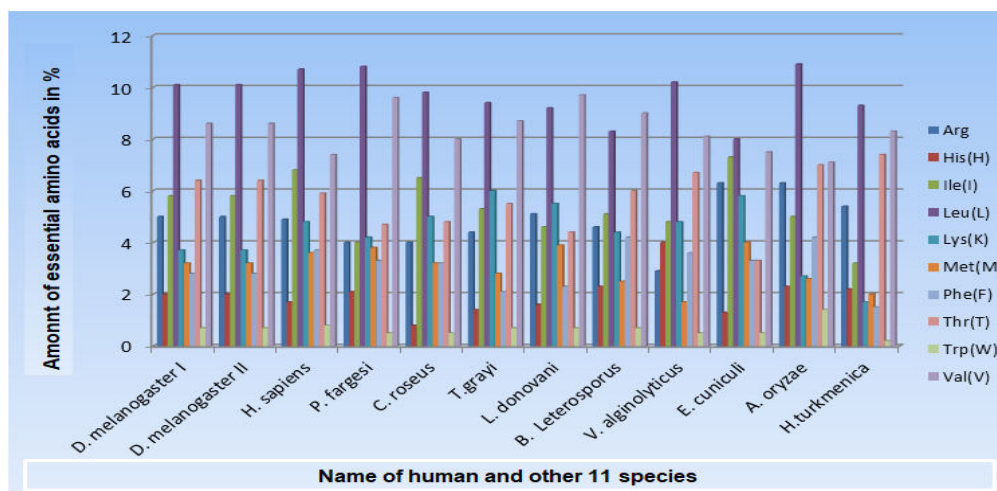


Fig. 1: Percentages of essential amino acids of HMG-CoA reductase of human and 11 other species.

Protein secondary structure prediction

There remains a close relationship between protein structure and function, so secondary structure analysis is a necessity for understanding protein function (Mamun et al., 2024). Predicting the secondary structure of proteins can also help predict the tertiary structure by minimizing the gap between the primary sequence and the tertiary structure (Zhang et al., 2018). The secondary structure implies whether the

given amino acids lie in a helix, strand, or coil. Secondary structural features analyzed by SOPMA (Geourjon & Deleage, 1995) and PHD server (Rost et al., 1994) are shown in **Table 2**. The results revealed that the alpha helix dominated over the secondary structures, followed by the random coil and extended strand. *P. fargessi*, *C. roseus*, *T. grayi*, *V. alginolyticus*, and *A. oryzae* random coil outnumbered alpha helix. In all of the proteins, extended strands were less

prevalent than other secondary structural features. Similar types of output were also observed when validating this result by using the PHD server (Rost et al., 1994).

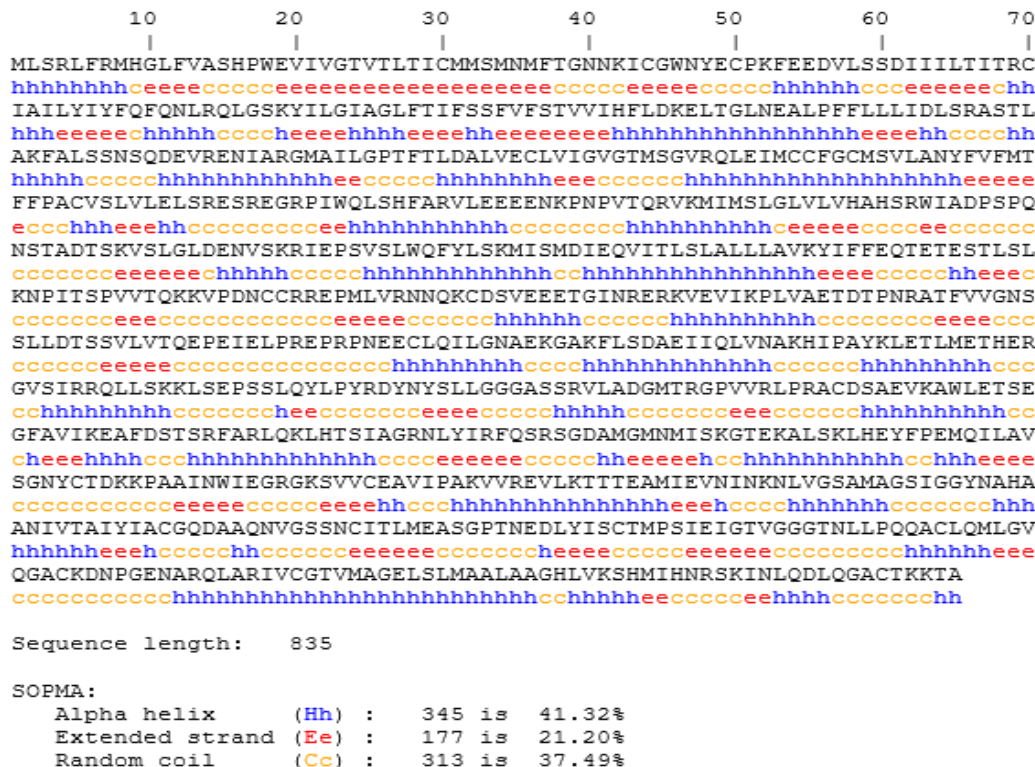


Fig. 2: Analysis of Secondary structures of HMG-CoA reductase of human using SOPMA server.

Table 2: Secondary structural features of HMG-CoA reductase predicted by the PHD Server and SOPMA Server.

Name	SOPMA			PHD		
	Alpha helix (%)	Extended strand (%)	Random coil (%)	Alpha helix (%)	Extended strand (%)	Random coil (%)
<i>D. melanogaster I</i>	41.41	18.37	40.22	35.65	20.54	43.80
<i>D. melanogaster II</i>	41.41	18.37	40.22	35.65	20.54	43.80
<i>H. sapiens</i>	41.32	21.20	37.49	43.35	15.57	41.08
<i>P. fargessi</i>	38.96	19.48	41.57	35.48	18.43	46.09
<i>C. roseus</i>	38.60	19.80	41.60	41.43	13.48	45.09
<i>T. grayi</i>	37.70	21.61	40.69	31.49	24.14	44.37
<i>L. donovani</i>	51.15	15.44	33.41	42.86	18.89	38.25
<i>B. leteroporus</i>	42.59	21.06	36.34	47.22	21.76	31.02
<i>V. alginolyticus</i>	37.86	22.38	39.76	36.67	21.43	41.90
<i>E. cuniculi</i>	42.96	22.61	34.42	36.68	31.41	31.91
<i>A. oryzae</i>	40.13	17.62	42.24	44.73	14.37	40.90
<i>H. turkmenia</i>	51.47	10.78	37.75	51.23	16.18	32.60

MSA and phylogenetic analysis

Multiple sequence alignments of amino acid sequences of HMG-CoA reductases from diverse species reveal that all sequences originated from a common ancestral gene. Two major clusters were identified in the phylogenetic tree (Fig. 3), generated by the UPGMA UniversePG | www.universepg.com

method (Schlee, 1975). Cluster I included 11 species, further divided into 2 subclusters, and one of these subclusters contained *Homo sapiens* and *D. melanogaster*. Cluster II contained only *Encephalitozoon cuniculi*, a fungus, indicating its distant evolutionary relationship with others.

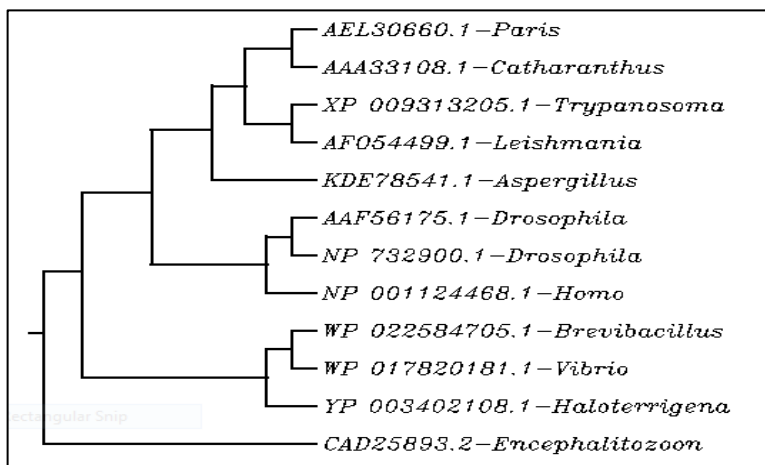


Fig. 3: Phylogenetic tree of the HMG-CoA reductase sequence of human and 11 other species.

Motif analysis

Motifs are a set of conserved amino acid sequences that provide information about function. Motifs also provide information and help to identify the binding sites (Bailey et al., 2006). Motifs predicted using MEME (Bailey et al., 2006) and MAST (Bailey & Gribskov, 1998) showed the distribution of 10

conserved motifs among these HMG-CoA reductase protein sequences. The widths of those predicted 10 motifs range from 21 to 50 amino acids. Pfam (Mistry et al., 2021) analysis ensures that most of the predicted motifs are associated with the HMG-CoA reductase family (Table 3).

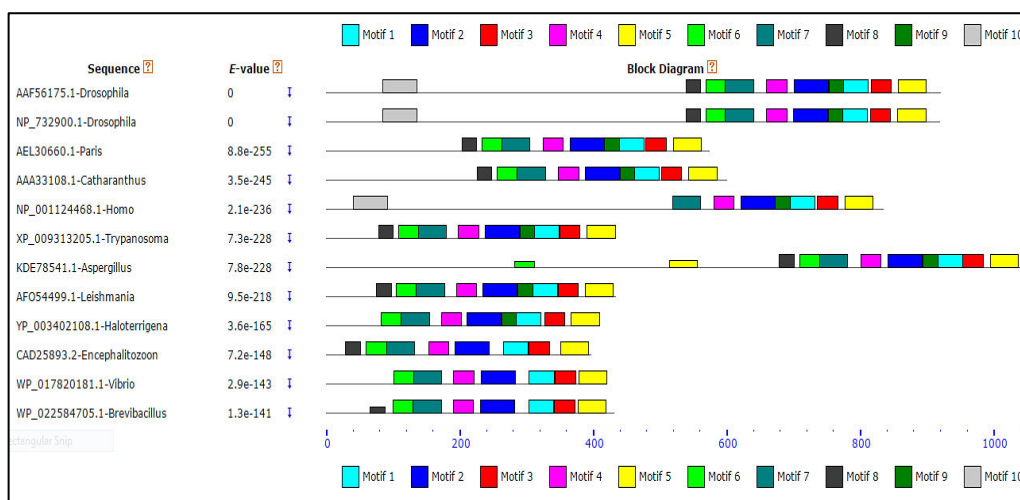


Fig. 4: Predicted motif of HMG-CoA reductase sequences by MAST.

Table 3: Showing protein family of conserved motifs determined by Pfam analysis.

Motif	Width	Best possible match	Protein family
1	41	QHHPDPMQIIFLSGNFCCDKKPAAINWIEGRGKSVVCEATI	HMG-coA reductase
2	34	FNAHAANIVAAIFIATGQDPAQNVESSNCITTME	HMG-coA reductase
3	29	DLYISCTMPSEIVGTVGGGTGLPSQKACL	HMG-coA reductase
4	29	HCAMDGRQVYIRFQYYTGDMGMNMITKG	HMG-coA reductase
5	41	PGANARQLARIVCATVLAGELSLMSALAAAGHLVKSHEMKHNR	HMG-coA reductase
6	29	DYYIPMATTEGALVASTNRGCKAINQCGG	HMG-coA reductase
7	21	YNQCCENIIGYVPIPVGYAGP	HMG-coA reductase

8	31	VLKDG MTRAPCFRFP SFKEAEFKFWCEEDF	HMG-coA reductase
9	29	VRKVLKTTVDALVECNMKNL VGSAMAGS	HMG-coA reductase
10	50	RPCHGWSQSCDGLAEYNAADVILMTIVRCTAVLYCYYQFCSLHR LGSKY	Uncharacterized

Predicting the subcellular location of proteins

Understanding the subcellular location of proteins is essential for genome analysis and determining protein function. The position of a protein within a cell is known as subcellular location, which may be the cytoplasm, periplasm, extracellular space, outer membrane, or inner membrane (Mamun *et al.*, 2024). Depicting the protein as a target for a drug or vaccine is the significance of the localization study. Proteins found in the cytoplasm and surface membranes can be utilized as drug targets and vaccine targets, respectively (Barh *et al.*, 2011). Subcellular localization analysis suggested that HMGRs are mostly found in the cytoplasm and endoplasmic reticulum across a variety of species. In addition to intracellular localizations, *B. leterosporus* and *V. alginolyticus* showed minor localization of these proteins to the extracellular space, determined by CELLO (Yu *et al.*, 2004). These results suggest that cytoplasmic and extracellular HMG-CoA reductase proteins could serve as promising targets for drug and vaccine development.

Tertiary structure prediction

In this study, the three-dimensional (3D) structures of HMG-CoA reductase (HMGR) proteins from various species were predicted using SWISS-MODEL (Arnold *et al.*, 2006), an automated server for comparative protein modeling. SWISS-MODEL demonstrates reliable performance when the sequence identity between the target and template exceeds 30% (Arnold *et al.*, 2006). High-confidence models were generated for *Drosophila melanogaster* II (NP_732900.1) and *Homo sapiens* (NP_001124468.1), supported by template identities greater than 60% (see **Table 4**). Other HMGR proteins - *D. melanogaster* I (AAF-56175.1), *P. fargesii* (AEL30660.1), *Catharanthus roseus* (AAA33108.1), *Trypanosoma grayi* (XP_009313205), *Leishmania donovani* (AF054499.1), and *Aspergillus oryzae* 100-8 (KDE78541.1) - showed template identities near 50%, enabling the construction of reliable models. Additionally, acceptable structures were obtained for *Brevibacillus laterosporus* (WP_022584705.1), *Haloterrigena turkmenica* DSM 5511 (YP_003402108.1), *Encephalitozoon cuniculi* GB-M1 (CAD25893.2), and *Vibrio alginolyticus* (WP_017820181.1), with template identities exceeding the 30% threshold.

Table 4: Percentages of sequence identity, resolution, and E-value of HMG CoA-reductase protein sequences.

Name	Template	Sequence identity (%)	Resolution	E-Value
<i>D. melanogaster I</i>	2q6bA	57.21	2.00 Å	0.00e-1
<i>D. melanogaster II</i>	1hw8D	60.00	2.10 Å	2.13e-123
<i>H. sapiens</i>	2q6bA	87.27	2.00 Å	0.00e-1
<i>P. fargesii</i>	2q6bA	55.8	2.00 Å	0.00e-1
<i>C. roseus</i>	2q6bA	56.54	2.00 Å	0.00e-1
<i>T. grayi</i>	2q6bA	49.29	2.00 Å	0.00e-1
<i>L. donovani</i>	2q6bA	49.53	2.00 Å	0.00e-1
<i>B. leterosporus</i>	2q6bA	32.35	2.00 Å	0.00e-1
<i>V. alginolyticus</i>	2q6bA	33.41	2.00 Å	0.00e-1
<i>E. cuniculi</i>	2q6bA	31.03	2.00 Å	0.00e-1
<i>A. oryzae</i>	2q6bA	50.67	2.00 Å	0.00e-1
<i>H. turkmenica</i>	2q6bA	40.2	2.00 Å	0.00e-1

The predicted 3D models were visualized using PyMOL software (DeLano, 2002). Structural comparison revealed that while the overall 3D

architectures were not identical, they exhibited notable similarities in general shape and domain organization. **Fig. 5** presents the modeled structures, where helices,

sheets, and loops are highlighted in blue, red, and violet, respectively.

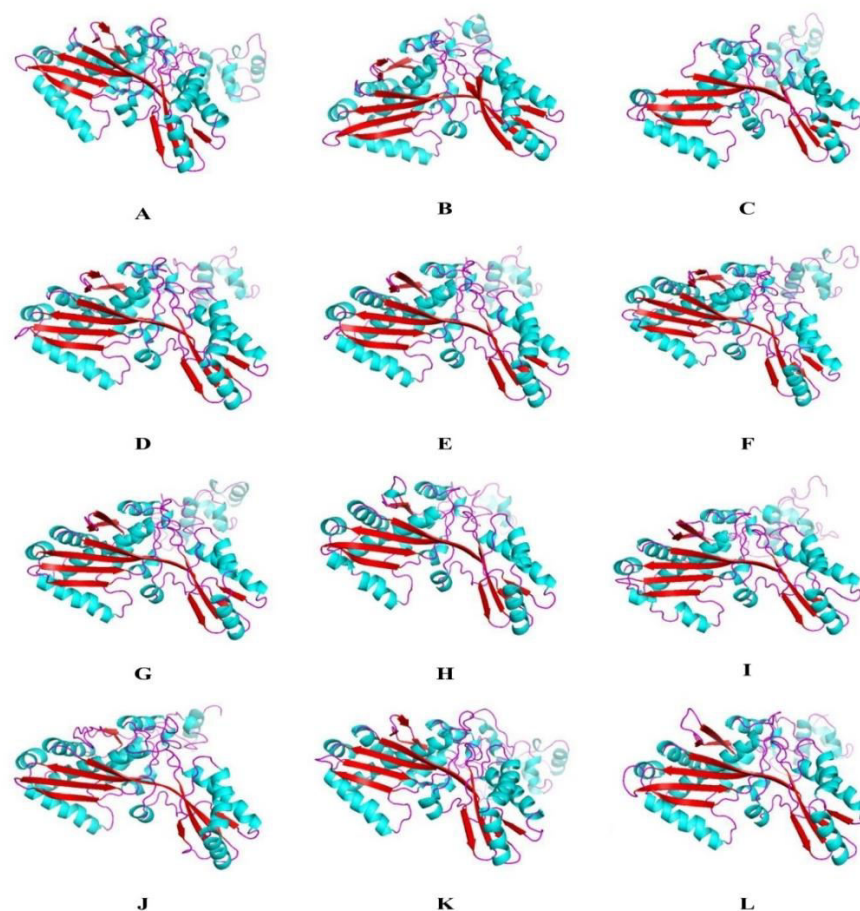


Fig. 5: Predicted 3D structures of HMG-CoA reductase proteins. **A**= *D. melanogaster-I*, **B**= *D. melanogaster-II*, **C**= *H. sapiens*, **D**= *P. fargesii*, **E**= *C. roseus*, **F**= *T. grayi*, **G**= *L. donovani*, **H**= *B. laterosporus*, **I**= *V. alginolyticus*, **J**= *E. cuniculi*, **K**= *A. oryzae* 100-8, **L**= *H. turkmenica* DSM 5511

Quality assessment of the model

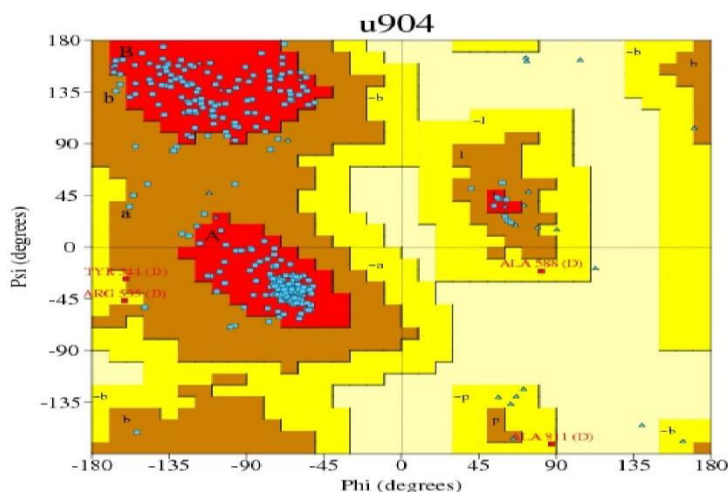
Quality assessment and validation of proteins were done by several quality assessment steps. Accuracy and quality of the 3D model were assessed with PROCHECK (Laskowski *et al.*, 1996) by Ramachandran plot analysis. It was found from the analysis that most of the residues of the three-dimensional models belonged to the most favored regions, ranging from 83-89% (shown in **Table 5**, **Fig. 6**). VERIFY 3D (Bowie *et al.*, 1991; Eisenberg *et al.*, 1997) analysis revealed that the 3D scores of all the experimental

protein models were above 80% except for the proteins from *B. laterosporus* (77.81%) and *V. alginolyticus* (73.46%). ERRAT (Colovos & Yeates, 1993) was used to evaluate the reliability of the protein models. It was found that the ERRAT overall quality factor values exceeded 80 for *H. sapiens*, *D. melanogaster I*, *D. melanogaster II*, *P. fargesii*, *C. roseus*, *T. grayi*, *L. donovani*, and *H. turkmenica* DSM 5511 (**Table 6**), while the ERRAT overall quality factor values for the remaining species were below 80.

Table 5: Percentages of different regions in the Ramachandran plot.

Name	Most favoured regions (%)	Most favoured regions (%)	Generously allowed regions (%)	Disallowed regions (%)
<i>D. melanogaster I</i>	86.7	11.4	1.1	0.8
<i>D. melanogaster II</i>	89.8	9.80	1.3	0.8

<i>H. sapiens</i>	87.9	11.2	0.0	0.9
<i>P. fargesii</i>	88.9	10.5	0.3	0.3
<i>C. roseus</i>	89.3	9.50	0.9	0.3
<i>T. grayi</i>	85.6	12.4	1.7	0.3
<i>L. donovani</i>	85.8	12.9	1.1	0.3
<i>B. leterosporus</i>	83.3	12.2	3.3	1.2
<i>V. alginolyticus</i>	85.0	12.1	2.0	0.8
<i>E. cuniculi</i>	83.0	11.9	2.7	0.3
<i>A. oryzae</i>	86.5	11.9	1.1	0.5
<i>H. turkmenica</i>	85.6	12.9	0.9	0.6



Ramachandran Plot statistics

		No. of residues	%-tage
Most favoured regions	[A,B,L]	282	89.0%*
Additional allowed regions	[a,b,l,p]	31	9.8%
Generously allowed regions	[~a,~b,~l,~p]	4	1.3%
Disallowed regions	[XX]	0	0.0%
Non-glycine and non-proline residues		317	100.0%
End-residues (excl. Gly and Pro)		3	
Glycine residues		35	
Proline residues		15	
Total number of residues		370	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20.0 good quality model would be expected to have over 90% in the most favoured regions [A,B,L].

Fig. 6: Ramachandran plot of HMG-CoA reductase protein of *D. melanogaster* generated by PROCHECK.

Table 6: Percentages of Verify 3D and ERRAT overall quality factor of HMG-CoA reductase protein models.

Name	ERRAT	Verify 3D%
<i>D. melanogaster I</i>	82.082	92.07
<i>D. melanogaster II</i>	86.610	88.62
<i>H. sapiens</i>	88.483	88.95
<i>P. fargesii</i>	87.339	91.04
<i>C. roseus</i>	85.052	90.05
<i>T. grayi</i>	81.886	90.65
<i>L. donovani</i>	82.878	84.32
<i>B. leterosporus</i>	67.391	77.81
<i>V. alginolyticus</i>	70.886	73.46
<i>E. cuniculi</i>	62.032	86.16
<i>A. oryzae</i> 100-8	74.408	91.38
<i>H. turkmenica</i> DSM5511	83.551	87.47

4. Discussion

In this study, the amino acid sequences were retrieved and saved in FASTA format from the NCBI database. Sequence analysis showed that the HMG-CoA reductase of *D. melanogaster I*, *D. Melanogaster II*, *H. sapiens*, *P. fargesii*, *C. roseus*, *T. grayi*, *L. donovani*, *B. leterosporus*, *V. alginolyticus*, *E. cuniculi* GB-M1, *A. oryzae* 100-8, and *H. turkmenica* DSM 5511 consisted of amino acids 920, 920, 835, 575, 601, 435, 434, 432, 420, 398, 1044, and 408 amino acids, respectively. Analysis of the instability index (Guruprasad *et al.*, 1990) indicated that most of the evaluated proteins were unstable, except proteins from humans, fruit flies, and plants, which showed values below the threshold of 40. In this *in silico* analysis, aliphatic indices were calculated to assess the structural stability of the selected enzyme. The results showed that the aliphatic index (Ikai, 1980) of the protein sequences ranged from 87.66 to 97.27, suggesting the stability of these enzymes across a wide range of temperatures. Positive GRAVY (Kyte & Doolittle, 1982) values indicated the hydrophobic nature of the protein from different species, with an exception in *B. leterosporus* (-0.013), *E. cuniculi* (-0.108), and *H. turkmenica* DSM 5511 (-0.084). Overall, the computational analysis suggested that HMG-CoA reductase proteins from *D. melanogaster*, *P. fargesii*, and *C. roseus* share strong structural and physicochemical similarities with the human enzyme.

Essential amino acids, particularly leucine and valine, are known to play a crucial role in reducing the synthesis of cholesterol (Adhikari *et al.*, 2013). By lowering serum cholesterol, the essential amino acids leucine and valine may help in reducing the risk of atherosclerosis (Cojocaru *et al.*, 2010). In this study, the high level of leucine and valine in every species suggested that these enzymes could serve as potential sources of HMG-CoA reductase for reducing the human blood cholesterol levels. In the case of secondary structure analysis, joint prediction with SOPMA and a neural network method (PHD) correctly predicts secondary structure with 82.2% accuracy (Rost *et al.*, 1994). Both prediction servers indicated that α -helices were the predominant structural elements, followed by random coils and extended strands.

Multiple sequence alignment of HMG-CoA reductases from different species indicated that they likely originated from a common ancestral gene. A motif is a sequence pattern that occurs repeatedly in a group of related protein or DNA sequences. It may also have a role in the development of the protein's binding sites (Sanchita *et al.*, A motif analysis by MEME (Bailey *et al.*, 2006) and the MAST server (Bailey & Gribskov, 1998) indicated the uniformity in the conserved sequences of all the species. Proteins found in the cytoplasm and surface membranes can act as potential drug or vaccine targets (Barh *et al.*, 2011). Subcellular localization analysis using PSORT-b (Gardy *et al.*, 2003), WoLF PSORT (Horton *et al.*, 2007), and CELLO (Yu *et al.*, 2004) suggested that most of the proteins were cytoplasmic and could serve as promising drug target candidates. To predict the three-dimensional structure of HMG-CoA reductase protein for fruit flies, humans, plants, protists, bacteria, archaea, and fungi by homology modelling, appropriate templates were selected by subjecting the sequences of proteins to BLASTp against the PDB database, considering several parameters like query coverage, lower resolution, lowest E-value, and above 30% sequence identity. SWISS-MODEL server (Arnold *et al.*, 2006) provided suitable templates, and all sequences were found to have identity values above 30%, suggesting their suitability for reliable 3D modeling.

Model validation using PROCHECK (Laskowski *et al.*, 1996) demonstrated that most species had 85% to 89.9% of residues in the most favored regions of the Ramachandran plot (Spencer *et al.*, 2019). These findings displayed a higher percentage of residues occupying the favorable regions within the plot, signifying good stereochemical quality and structural reliability. No species in the experimental sequences exceeds the value of most favored regions above 90%. For the quality of the protein model to be considered satisfactory, it is expected to have a Verify3D score greater than 80% (Khor *et al.*, 2014). In this study, for most species, a value of more than 80% indicates that at least 80% of the amino acids have scored > 0.2 in the 3D/1D profile, which is considered passed, suggesting that most protein model quality in this analysis is satisfactory. ERRAT (Colovos & Yeates,

1993) predicted an overall quality factor value above 80 for most of the protein, which indicates the high reliability of the predicted protein model (Colovos & Yeates, 1993). But according to Messaoudi (Messaoudi *et al.*, 2011), a generally acceptable model range is >50 for a high-quality model. It is found from analysis by ERRAT that, overall quality factor value for all the species is >60, with a value of *H. sapiens*, along with *D. melanogaster I*, *D. melanogaster II*, *P. forgesii*, *C. roseus*, *T. grayi*, *L. donovani*, and *H. turkmenica* DSM 5511 (**Table 5**), above 80. The overall quality factor obtained through ERRAT indicated a high-quality protein model.

5. Conclusion

This study successfully performed a comprehensive in silico characterization and comparative analysis of HMG-CoA reductase (HMGR) enzymes from twelve diverse species to investigate their properties, subcellular localization, functional motifs, and phylogenetic relationships using various bioinformatics tools and servers. Additionally, this study aimed to construct a 3D model of the HMG-CoA reductase enzyme, as understanding its three-dimensional structure is crucial for gaining insights into its functional characteristics. The findings provide comprehensive information on the molecular properties, architecture, and functional roles of the HMG-CoA reductase enzyme. The results suggest that HMG-CoA reductase enzymes from all the studied species may serve as potential sources for developing targeted cholesterol-lowering drugs in humans.

6. Author Contributions

R: Conducted data collection, performed bioinformatics analyses, and prepared the initial draft of the manuscript. M.M.R.: Supervised methodological design, guided phylogenetic and motif analyses, and critically reviewed and revised the manuscript for important intellectual content. S.M.: Conceived and designed the research, interpreted the results, and finalized the manuscript for submission and publication. All authors read and approved the final version of the manuscript.

7. Acknowledgment

The authors gratefully acknowledge Mr. Utpal Kumar Adhikari of the School of Medicine, Western Sydney UniversePG | www.universepg.com

University, Australia, for his kind cooperation during this research.

8. Conflicts of Interest

The authors declare that there is no conflict of interest.

9. References

- Adhikari, U. K., Islam, K., & Technology. (2013). A Computational Study on 3-Hydroxy-3-Methylglutaryl-CoA Reductase in *Lacto-bacillus spp.* **9**(1), 1-7.
- Arnold, K., Bordoli, L., & Schwede, T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, **22**(2), 195-201. <https://doi.org/10.1093/bioinformatics/bti770>
- Bailey, T. L., & Gribskov, M. (1998). Combining evidence using p-values: application to sequence homology searches. *Bioinformatics*, **14**(1), 48-54. <https://doi.org/10.1093/bioinformatics/14.1.48>
- Bailey, T. L., Williams, N., & Li, W. W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res*, **34**(Web Server issue), W369-373. <https://doi.org/10.1093/nar/gkl198>
- Barh, D., Tiwari, S., & Kumar, A. J. D. D. R. (2011). In silico subtractive genomics for target identification in human bacterial pathogens. *72*(2), 162-177.
- Bjellqvist, B., Hughes, G. J., & Hochstrasser, D. (1993). The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis*, **14**(10), 1023-1031. <https://doi.org/10.1002/elps.11501401163>
- Bonovas, S., Filioussi, K., & Sitaras, N. M. (2007). Statins and the risk of colorectal cancer: a meta-analysis of 18 studies involving more than 1.5 million patients. *J Clin Oncol*, **25**(23), 3462-3468. <https://doi.org/10.1200/JCO.2007.10.8936>
- Bowie, J. U., Luthy, R., & Eisenberg, D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. *Science*, **253**(5016), 164-170. <https://doi.org/10.1126/science.1853201>

- Buhaescu, I., & Izzedine, H. (2007). Mevalonate pathway: a review of clinical and therapeutical implications. *Clin Biochem*, **40**(9-10), 575-584. <https://doi.org/10.1016/j.clinbiochem.2007.03.016>
- Cojocaru, E., Zamfir, C. L., & Cotuțiu, C. J. R. M.-c. a. S. d. M. s. N. d. I. (2010). The effects of some nonpolar aminoacids--valine, leucine--administration on the arterial wall already exposed to a hypercholesterolemic diet. *114*(2), 504-509.
- Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci*, **2**(9), 1511-1519. <https://doi.org/10.1002/pro.5560020916>
- DeLano, W. L. J. h. w. p. o. (2002). The PyMOL molecular graphics system.
- Eisenberg, D., Luthy, R., & Bowie, J. U. (1997). VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol*, **277**, 396-404. [https://doi.org/10.1016/s0076-6879\(97\)77022-8](https://doi.org/10.1016/s0076-6879(97)77022-8)
- Enany, S. (2014). Structural and functional analysis of hypothetical and conserved proteins of *Clostridium tetani*. *J Infect Public Health*, **7**(4), 296-307. <https://doi.org/10.1016/j.jiph.2014.02.002>
- Friesen, J. A., & Rodwell, V. W. (2004). The 3-hydroxy-3-methylglutaryl coenzyme-A (HMG - CoA) reductases. *Genome Biol*, **5**(11), 248. <https://doi.org/10.1186/gb-2004-5-11-248>
- Gardy, J. L., Spencer, C., & Brinkman, F. S. (2003). PSORT-B: Improving protein subcellular localization prediction for Gram-negative bacteria. *Nucleic Acids Res*, **31**(13), 3613-3617. <https://doi.org/10.1093/nar/gkg602>
- Geourjon, C., & Deleage, G. (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl Biosci*, **11**(6), 681-684. <https://doi.org/10.1093/bioinformatics/11.6.681>
- Gesto, D. S., Cerqueira, N. M., & Fernandes, P. A. (2014). Discovery of new druggable sites in the anti-cholesterol target HMG-CoA reductase by computational alanine scanning mutagenesis. *J Mol Model*, **20**(4), 2178. <https://doi.org/10.1007/s00894-014-2178-8>
- Gill, S. C., & von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem*, **182**(2), 319-326. [https://doi.org/10.1016/0003-2697\(89\)90602-7](https://doi.org/10.1016/0003-2697(89)90602-7)
- Gillani, M., & Pollastri, G. (2024). Protein subcellular localization prediction tools. *Comput Struct Biotechnol J*, **23**, 1796-1807. <https://doi.org/10.1016/j.csbj.2024.04.032>
- Guruprasad, K., Reddy, B. V., & Pandit, M. W. (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 Eng*, **4**(2), 155-161. <https://doi.org/10.1093/protein/4.2.155>
- Horton, P., Park, K. J., & Nakai, K. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Res*, **35**(Web Server issue), W585-587. <https://doi.org/10.1093/nar/gkm259>
- Ikai, A. J. T. J. o. B. (1980). Thermostability and aliphatic index of globular proteins. **88**(6), 1895-1898.
- Jiang, S. Y., Li, H., & Song, B. L. (2018). Discovery of a potent HMG-CoA reductase degrader that eliminates statin-induced reductase accumulation and lowers cholesterol. *Nat Commun*, **9**(1), 5138. <https://doi.org/10.1038/s41467-018-07590-3>
- Khor, B. Y., Tye, G. J., Lim, T. S., Noordin, R., & Choong, Y. S. (2014). The structure and dynamics of BmR1 protein from *Brugia malayi*: in silico approaches. *Int J Mol Sci*, **15**(6), 11082-11099. <https://doi.org/10.3390/ijms150611082>
- Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol*, **157**(1), 105-132. [https://doi.org/10.1016/0022-2836\(82\)90515-0](https://doi.org/10.1016/0022-2836(82)90515-0)
- Laskowski, R. A., Chistyakov, V. V., & Thornton, J. M. (2005). PDBsum more: new summaries and analyses of the known 3D structures of

- proteins and nucleic acids. *Nucleic Acids Res*, **33**(Database issue), D266-268.
<https://doi.org/10.1093/nar/gki001>
- Laskowski, R. A., Rullmann, J. A., & Thornton, J. M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR*, **8**(4), 477-486.
<https://doi.org/10.1007/BF00228148>
- Lipkin, S. M., Chao, E. C., & Gruber, S. B. (2010). Genetic variation in 3-hydroxy-3-methylglutaryl CoA reductase modifies the chemopreventive activity of statins for colorectal cancer. *Cancer Prev Res (Phila)*, **3**(5), 597-603.
<https://doi.org/10.1158/1940-6207.CAPR-10-0007>
- Lochhead, P., & Chan, A. T. (2013). Statins and colorectal cancer. *Clin Gastroenterol Hepatol*, **11**(2), 109-118; quiz e113-104.
<https://doi.org/10.1016/j.cgh.2012.08.037>
- Malani, S. K., Chigullapalli, S., Sujanyal, S., & Sharma, V. (2024). Rosuvastatin-Induced Myopathy: A Case Series. *Cureus*, **16**(8), e66180. <https://doi.org/10.7759/cureus.66180>
- Mamun, T. I., Bourhia, M., & Sitotaw, B. (2024). Structure based functional identification of an uncharacterized protein from *Coxiella burnetii* involved in adipogenesis. *Sci Rep*, **14**(1), 16789. <https://doi.org/10.1038/s41598-024-66072-3>
- Messaoudi, A., Belguith, H., & Ben Hamida, J. (2011). Three-Dimensional Structure of *Arabidopsis thaliana* Lipase Predicted by Homology Modeling Method. *Evol Bioinform Online*, **7**, 99-105.
<https://doi.org/10.4137/EBO.S7122>
- Mistry, J., Chuguransky, S., & Bateman, A. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Res*, **49**(D1), D412-D419.
<https://doi.org/10.1093/nar/gkaa913>
- Nishant, T., Sathish Kumar, D., & VVL, P. K. A. J. J. P. B. (2011). Computational methods for protein structure prediction and its application in drug design. *I*, **2**.
- Petreski, T., Piko, N., & Bevc, S. (2021). Statin-Associated Necrotizing Myopathy Leading to Acute Kidney Injury: A Case Report. *Case Rep Nephrol Dial*, **11**(2), 129-135.
<https://doi.org/10.1159/000515584>
- Rost, B., Sander, C., & Schneider, R. (1994). PHD--an automatic mail server for protein secondary structure prediction. *Comput Appl Biosci*, **10**(1), 53-60.
<https://doi.org/10.1093/bioinformatics/10.1.53>
- Saikat ASM, Islam R, and Uddin ME. (2020). Structural and Functional Annotation of Uncharacterized Protein NCGM946K2_146 of *Mycobacterium tuberculosis*: An In-Silico Approach. *Proceedings*, **66**(1),13.
<https://doi.org/10.3390/proceedings2020066013>
- Sanchita, Chauhan, R., & Sharma, A. (2013). Docking and molecular dynamics studies of peptide inhibitors of ornithine decarboxylase: a rate-limiting enzyme for the metabolism of *Fusarium solani*. *J Biomol Struct Dyn*, **31**(8), 874-887.
<https://doi.org/10.1080/07391102.2012.718526>
- Schlee, D. (1975). Numerical taxonomy. The principles and practice of numerical classification. In: JSTOR.
- Spencer, R. K., Butterfoss, G. L., & Zuckermann, R. N. J. B. (2019). Stereochemistry of polypeptoid chain configurations. **110**(6), e23266.
- Stancu, C., & Sima, A. (2001). Statins: mechanism of action and effects. *J Cell Mol Med*, **5**(4), 378-387.
<https://doi.org/10.1111/j.1582-4934.2001.tb00172.x>
- Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, **22**(22), 4673-4680.
<https://doi.org/10.1093/nar/22.22.4673>
- Yousuf, M.; Saikat, A.S.M.; Uddin, M.E. (2023). Computational Approaches for Molecular Characterization and Functional Annotation of an Uncharacterized Protein of *V. cholerae*

- Applying Bioinformatics Tools and Databases. *Eng. Proc.* 2023, **37**, 1-6.
<https://doi.org/10.3390/ECP2023-14644>
- Vance, D. E., & Van den Bosch, H. (2000). Cholesterol in the year 2000. *Biochim Biophys Acta*, **1529**(1-3), 1-8.
[https://doi.org/10.1016/s1388-1981\(00\)00133-5](https://doi.org/10.1016/s1388-1981(00)00133-5)
- Wang, H. J., Park, J. Y., & Kang, E. S. (2015). Chronic HMGCR/HMG-CoA reductase inhibitor treatment contributes to dysglycemia by upregulating hepatic gluconeogenesis through autophagy induction. *Autophagy*, **11**(11), 2089-2101.
<https://doi.org/10.1080/15548627.2015.1091139>
- Wilkins, M. R., Gasteiger, E., & Hochstrasser, D. F. (1999). Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol*, **112**, 531-552.
<https://doi.org/10.1385/1-59259-584-7:531>
- Yu, C. S., Lin, C. J., & Hwang, J. K. (2004). Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Sci*, **13**(5), 1402-1406.
<https://doi.org/10.1110/ps.03479604>
- Zhang, B., Li, J., & Lu, Q. (2018). Prediction of 8-state protein secondary structures by a novel deep learning architecture. *BMC Bioinformatics*, **19**(1), 293.
<https://doi.org/10.1186/s12859-018-2280-5>

Citation: Rubaya, Rahman MM, and Mahmud S. (2025). Computational modeling and quality validation of HMG-CoA reductase for drug design applications. *Am. J. Pure Appl. Sci.*, 7(6), 485-498.
<https://doi.org/10.34104/ajpab.025.04850498>

Copyright: © The Author(s), 2025. Published by UniversePG. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, and provided the original work is properly cited. 