Journal of Bioinformatics and Sequence Analysis Vol. 2(5), pp. 53-68, October 2010
Available online at http://www.academicjournals.org/JBSA
ISSN 2141-2464 ©2010 Academic Journals

Full length Research Paper

Insights into the key enzymes of secondary metabolites biosynthesis in *Camellia sinensis*

Aditi Sharma, Ankita Punetha, Abhinav Grover and Durai Sundar*

Department of Biochemical Engineering and Biotechnology, Indian Institute of Technology (IIT) Delhi, Hauz Khas, New Delhi 110016, India.

Accepted 18 August, 2010

Tea is one of the most popular beverages consumed throughout the world. It is a source of important secondary metabolites like monoterpenoids, carotenoids and catechins. Monoterpenoids and carotenoids are important constituents of tea aroma. The formation of tea aroma involves synthesis and release of volatile monoterpenoids and carotenoids. On the other hand, catechins are responsible for the beneficial health effects of tea. Detailed *in silico* analysis of enzymes: Phytoene Synthase (PSY), a key enzyme in the carotenoid biosynthetic pathway and β-primeverosidase (BPR), a diglycosidase responsible for release of bound volatile terpenoids, have been undertaken in this study. Similarly, to study catechin biosynthesis, key enzymes in the flavonoid pathway namely, flavanone-3-hydroxylase (F3H), dihydroflavonol-4-reductase (DFR) and leucoanthocyanidin reductase (LAR), have been identified and studied. The comparative sequence analysis of PSY, F3H, DFR and LAR was carried out to identify the consensus and conserved amino acids using multiple sequence alignment. Phylogenetic trees were created to understand the evolutionary relationship of these enzymes present in different species. The three dimensional model structures were obtained for PSY, BPR, F3H, DFR and LAR by homology modeling to gain insights into the structure function relationships of these enzymes. Multiple templates were used to generate more accurate models of the enzymes. The models were further improved by loop refining and energy minimization. Binding pocket analysis was also done to identify the putative substrate binding sites and understand the enzyme-substrate interactions of each of these enzymes. The computational analysis of these key enzymes, PSY, BPR, F3H, DFR and LAR, provided valuable insights into the mechanism of formation of tea aroma and the synthesis of bioactive secondary metabolites like catechins.

Key words: *Camellia sinensis*, catechin, modeller, phytoene synthase, β-primeverosidase, flavanone-3-hydroxylase, dihydroflavonol-4-reductase, leucoanthocyanidin reductase.

INTRODUCTION

Tea is one of the most widely consumed beverages in the world. India is among the major producers of tea. The commercial importance of the tea plant (*Camellia sinensis*) is due to its popularity as a refreshing health drink. It also has great value as a source of important secondary metabolites. The leaves of two varieties of *C. sinensis* are*: assamica* and *sinensis,* are used to manufacture tea. It is classified into three major categories according to the manufacturing process: green (unfermented) tea, oolong (partially fermented) tea, and black (fully fermented) tea.

Here, fermentation is the result of enzymatic action and exposure to atmospheric oxygen during the manufacturing process. The quality of tea can be assessed in terms of two main parameters, namely: flavour and colour of processed tea. Flavour comprises of taste and aroma. The non-volatile constituents are responsible for taste

*Corresponding author. E-mail: sundar@dbeb.iitd.ac.in. Tel: +91-11-26591066. Fax: +91-11-26582282.

Abbreviations: PSY, Phytoene synthase; **BPR,** β-primeverosidase; **F3H,** flavanone-3-hydroxylase; **DFR,** dihydroflavonol-4-reductase; **LAR,** leucoanthocyanidin reductase; **VFC,** volatile flavour compounds; **MEP,** methylerythritol phosphate; **GPP,** geranyl pyrophosphate; **GGPP,** geranylgeranyl pyrophosphate; **EGCG,** epigallocatechin gallate; **EGC,** epi-gallocatechin, **ECG,** epicatechin gallate; **EC,** epicatechin.

while aroma is due to the volatile constituents. A strong attractive aroma is the most important and desirable characteristic of good quality tea, since our sense of smell is much more highly developed than our sense of taste. In recent years, tea has attracted more and more attention because of its reported health benefits particularly as an antioxidant (Luczaj and Skrzydlewska, 2005) and anticarcinogenic (Way et al., 2009). The flavonoids of tea are generally believed to be responsible for these effects. An important class of these flavonoids is catechins which are predominant in black tea.

Tea aroma and tea catechins

Over 500 flavour compounds have been identified in tea (Rawat and Gulati, 2008). The aroma of tea is attributed to the Volatile Flavour Compounds (VFC) in tea. A significant number of these volatile compounds originate from large precursor molecules present in tea leaves. These precursor molecules include products of lipid breakdown, terpenoids and phenolics, which are present as bound glycosides in tea leaves and are released upon the action of enzymes like glucosidases (Rawat and Gulati, 2008).

Tea manufacturing process is known to enhance the release of volatile compounds from bound precursors. Besides the above, volatile compounds synthesised by oxidation of carotenoids are also present among the VFCs (Ravichandran and Parthiban, 1998). VFCs derived from terpenoid related compounds are important components of aroma because of their desirable sweet flowery aroma.

These VFCs include monoterpene alcohols like linalool and its oxides, geraniol and products of oxidation of carotenoids like α-ionone and β-ionone (Ravichandran and Parthiban, 1998). The precursors for the synthesis of monoterpenes and tetraterpenes like carotenoids are provided by the Methylerythritol Phosphate (MEP) pathway in the plastids.

The precursors for monoterpene and carotenoid synthesis are Geranyl Pyrophosphate (GPP) and Geranylgeranyl Pyrophosphate (GGPP) respectively. Tea catechins are, primarily, flavan-3-ols. The major catechins found in tea are epigallocatechin gallate (EGCG), epi-gallocatechin (EGC), epicatechin gallate (ECG) and epicatechin (EC). Catechins are colourless, water-soluble compounds which impart bitterness and astringency to tea. They have been reported to have antioxidative, anticarcinogenic, antiallergenic, anti-inflammatory, and vasodilatory properties. Catechins are synthesised by the flavonoid biosynthetic pathway starting with phenylalanine as the precursor. Almost all of the characteristics of manufactured tea, including its taste, colour and aroma, have been found to be associated directly or indirectly with catechins (Wang et al., 2000).

Key enzymes in aroma formation and catechin biosynthesis

Phytoene synthase (PSY) is a key enzyme in the biosynthetic pathway of carotenoids. Carotenoids are an important group of precursors of volatile flavour compounds present in tea aroma (Ravichandran, 2002). PSY catalyses the first step in the biosynthesis of carotenoids that is the head to head condensation of two GGPP molecules to produce phytoene (a colourless carotenoid). Phytoene thus formed is then converted, through a series of reaction steps, into the volatile carotenoids in tea aroma. Borthakur et al. (2008) studied the relationship of PSY gene expression to the accumulation of carotenoids in tea and found that the carotenoids accumulation showed a strong dependence on the expression of PSY gene (Borthakur et al., 2008). β-primeverosidase (EC 3.2.1.149) (BPR) is a disaccharide-specific glycosidase, which hydrolyzes aroma precursors of β-primeverosides (6-*O*-β-d-xylopyranosyl-β-d-glucopyranosides) to liberate various aroma compounds during the manufacturing process.

Monoterpene alcohols such as linalool and geraniol, and aromatic alcohols such as benzyl alcohol and 2-phenylethanol, are present as glycosidic precursors in fresh leaves of tea plants and are released by the action of BPR. It is known that amongst all the diglycosides present in tea, namely, β-primeverosides, β-acuminosides and β-D-gluconylpyranosides β-primeverosides are the most common that is most of the aroma precursors are present as β-primeverosides (Ijima et al., 1998; Ma et al., 2001).

Therefore, β-primeverosidase is an important enzyme involved the formation of tea aroma. Flavanone 3-hydroxylase (F3H) catalyzes an early step in flavonoid metabolism leading to the formation of dihydroflavonols from flavanones. These dihydroflavonols serve as intermediates for the biosynthesis of flavonols, flavan 3-ols and anthocyanidins. A strong correlation between the concentration of catechins and *CsF3H* expression indicating a critical role of F3H in catechin biosynthesis has been reported (Singh et al., 2008).

Dihydroflavonol 4-reductase (DFR) catalyzes the stereospecific reduction of dihydroflavonols to leucoanthocyanidins (flavan-3,4-diol) using NADPH as a cofactor. The expression of DFR has been reported to be closely related to the concentration of total catechins and polyphenols in various stages of leaf development (Singh et al., 2009). Leucoanthocyanidin reductase (LAR) catalyses the conversion of 3, 4-cis-leucocyanidin to catechin. In Camellia LAR is important in biosynthesis of catechin, epigallocatechin, and anthocyanidins in flavonoid metabolism in tea leaves. It has been reported that catechin accumulation in tea leaves is regulated by the mRNA accumulation of genes involved in their biosynthesis, which include LAR along with F3H and DFR (Eungwanichayapant and Popluechai, 2009).

Table 1. Templates used for basic modeling and multi-template modeling of BPR, PSY, F3H, DFR and LAR.

S.No.	Enzymes	Basic modeling templates	Multiple template modeling templates
1.	PSY	2ZCO:A	1EZF:A, 2ZCO:A
2.	BPR	1CBG:A	1CBG:A, 1H49:A, 1V02:A, 2JF7:A, 3GNP:A
3.	DFR	2C29:D	2C29:D, 2P4H:X
4.	F3H	2BRT:A	1W9Y:A, 1GP6:A
5.	LAR	1QYC:A	2GAS:A, 1QYC:A, 1QYD:A, 2QW8:A

MATERIALS AND METHODS

Comparative sequence analysis

The protein sequences were retrieved from Swissprot database (Stutz et al., 2006). The sequences were then submitted to BLASTP (Altschul et al., 1990). The database chosen for BLASTP was Non-Redundant Data Base (NRDB). The results were used to identify closely related species. Then multiple sequence alignments were performed by using CLUSTAL W web server (Thompson et al., 1994). The results were used to identify consensus and conserved amino acid residues. The output of the multiple sequence alignments were used to construct the phylogenetic trees using PhyloDraw (Choi et al., 2000) and study the relationships among the sequences.

Three dimensional structure prediction

Modeller 9v7 was used to generate basic three dimensional structure models of the enzymes (Eswar et al., 2008). Further, the advanced modeling feature of the Modeller 9v7 was used to improve the model structures. In this, multiple templates were used as opposed to a single template used in basic modelling. Firstly, the script salign.py is used to obtain the multiple sequences alignment of the templates with the target. The templates which were chosen for each of the enzymes have been listed in Table 1. After this, the target sequence was aligned to the template sequences using the align2d_mult.py script of Modeller (Figures 1, 2 and 3). The alignment thus obtained was used to generate the final model by using the model_mult.py script. This generated 5 models for each of the enzymes. The evaluate_model.py script was used to get the DOPE score for each of the models and the model with the lowest DOPE score was selected. Further, loop refinement was done to improve the loops/regions in the model structures with poor DOPE scores. The DOPE profile of the basic model and the multiple templates model were plotted and analyzed to identify such regions. The loop modeling feature was then used to refine the structure of these loops. This was done using the loop_refine.py script of modeller. This generated 10 models whose DOPE scores were then calculated by running the model_energies.py script. The model with the lowest DOPE score was selected and its profile was generated using the evaluate_model.py script. Discovery studio 2.1 was then used to perform energy minimization and stereo chemical quality checks to arrive at the best possible three dimensional structure of the protein. The force field applied was CHARMM and the energy minimization algorithm used was steepest descent with an RMS gradient of 0.1 using a maximum of 1000 steps.

Binding pocket analysis

The predicted three-dimensional structures were then used to locate the structural pockets and cavities in the structures. This helped in identification of the putative substrate binding sites. CASTp server was used to identify the putative binding pockets and protein substrate binding sites in the generated models (Binkowski et al., 2003). It uses the weighted Delaunay triangulation and the alpha complex for shape measurements, which provides identification and measurements of surface accessible pockets as well as interior inaccessible cavities for proteins. Identification of binding sites and active sites in an enzyme structure can give insights into enzyme-substrate and enzyme cofactor interaction.

RESULTS AND DISCUSSION

The number of amino acids present in PSY, BPR, F3H DFR and LAR sequences obtained from Swissprot database are 329, 507, 347, 342 and 368 respectively and their molecular weights are 37518.9, 57041.2, 41464.8, 38674.3 and 37517.9 Da respectively as calculated by ProtPram web server (Gasteiger et al., 2003).

Comparative sequence analysis and phylogenetic trees

Closely related sequences were identified by using the $BLAST_P$ (Altschul et al., 1990) against non-redundant protein sequence database (Tables 2, 3, 4 and 5). The multiple sequence alignment of these related sequences was done to identify the consensus sequences and conserved amino acid residues. These conserved domains were later compared to the results of the binding site analysis which was carried out using $CAST_P$. However, the $BLAST_P$ results for β-primeverosidase did not contain any result with this enzyme in other species. Therefore, due to lack of deposited data a comparative sequence analysis and phylogenetic analysis for β-primeverosidase could not be done.

Two major groups were obtained from phylogenetic analysis of phytoene synthase (Figure 4). Group 1 has one subgroup namely SG1 and two mono taxa: *Helianthus annus* and *Elaeagnus umbellate*. SG1 contains two clades, taxa *Diospyros kaki* and *Actinidia deliciosa* occur in clade A while *Coffea canephora* and *Gentiana lutea* occur in clade B. *C. sinensis* and *Solanum lycopersicum* exit as mono taxae in clade A and B respectively. Group 2 contains 1 clade which has *Carica*

A

```
_aln.pos          10        20        30        40        50        60
1ezfA     --NSLKTCYKYLNQTS---RSFAAVIQALDGEMRNAVCIFYLVLRALDTLEDDMTISVEKKVPLLHNF
2zcoA     MTMMDMNFKYCHKIMKKHSKSFSYAFDLLPEDQRKAVWAIYAVCRKIDDSIDVYGD-----IQFLNQI
PSY       MSAALLWVVSPNSEVS---SGFGFLESVREGNSLLDSSKFSPRERTLICHGRFKKSRNKATRYRRKAI
 _consrvd                          *                   *

 _aln.p   70        80        90       100       110       120       130
1ezfA     HSFLYQ---------PDWRFMESKEKDRQVLEDFPTISLEFRNLAEKYQTVIADICRRMGIGMAEFLD
2zcoA     KEDIQSIEKYPYEYHHFQS-------------DRRIMMALQHVAQHK-NIAFQSFYNLIDTVYKDQH
PSY       FLYFQAW-----LQTLLENWQSHLNKRFTMLFLKQAALVKKHLKSEEEYDVKPDIVLPGTSSLLSEAY
 _consrvd

 _aln.pos  140       150       160       170       180       190       200
1ezfA     -KHVTSEQEWDKYCHYVAGLVGIGLSRLFSASEFEDPLVGEDTERANSMGLFLQKTNIIRDYLEDQQG
2zcoA     FTMFETDAELFGYCYGVAGTVGEVLTPIL-----SDHETHQTYDVARRLGESLQLINILRDVGEDFEN
PSY       DRCGEVCAEYAKTFYLGTLLMTPERRRAIWAIYVWCRRTDELVDGPNASHITPTALDRWESRLEDLFR
 _consrvd          *                                                      **

 _aln.pos    210       220       230       240       250       260       270
1ezfA     GREFWPQEVWSRYVKKLGDFAKPENIDLAVQCLNELITNALHHIPDVITYLSRLRNQSVFNFCAIPQV
2zcoA     ERIYFSKQRLKQYEVDIAEVYQNGVNNHYIDLWEYYAAIAEKDFRDVMDQIKVFSIE-AQPIIELAAR
PSY       GRPFDMLDAALSDTVTKFPVDIQPFKDMIEGMRLDLKKSRYKNFDEL--YLYCYYVAGTVGLMSVPVM
 _consrvd  *

 _aln.pos      280       290       300       310       320       330       340
1ezfA     MAIATLAACYNNQQVFKGAVKI----------DATNMPAVKAIIYQYMEEIYHRIPDSDPSSSKTRQI
2zcoA     IYIEILDEVRQANYTLHERVFVEKRKKAKLFHEINSKY------------------------------
PSY       GIAPESQATTESVYNAALALGIANQLTNILRDVGEDARRGRVYLPQDELAQA---------GLSDEDI
 _consrvd

 _aln.pos
1ezfA     ISTIRTQN
2zcoA     --------
PSY       FAGKVTEK
 _consrvd
```

B

```
_aln.pos          10        20        30        40        50        60
1cbgA     --------------------FKPLPISFDDFSDLNRSCFAPGFVFGTASSAFQYEGAAFEDGKGPSIW
1h49A     ---------------------------MLSPSEIPQRDWFPSDFTFGAATSAYQIEGAWNEDGKGESNW
1v02A     ----------------------------RLSPWEIPRRDWFPPSFLFGAATSAYQIEGAWNEDGKGPSTW
2jf7A     ----------------------------------VVHRRDFPQDFIFGAGGSAYQCEGAYNEGNRGPSIW
3gnpA     ----------------------------------GLTRGSFPEGFVFGTASAAYQYEGAVKEDGRGQTIW
BPR       MMAAKGSVVVGVLAIVAYALVVSEVAIAAQISSFNRTSFPDGFVFGAASSAYQFEGAAKEGGKGPNIW
 _consrvd                                   *  *   * **    * * ***  *   *   *

 _aln.p   70        80        90       100       110       120       130
1cbgA     DTFTHKYPEKIKDRTNGDVAIDEYHRYKEDIGIMKDMNLDAYRFSISWPRVLPKGKLSGGVNREGINY
1h49A     DHFCHNHPERILDGSNSDIGANSYHMYKTDVRLLKEMGMDAYRFSISWPRILPKGTKEGGINPDGIKY
1v02A     DHFCHNFPEWIVDRSNGDVAADSYHMYAEDVRLLKEMGMDAYRFSISWPRILPKGTLAGGINEKRVEY
2jf7A     DTFTQRSPAKISDGSNGNQAINCYHMYKEDIKIMKQTGLESYRFSISWSRVLPGGRLAAGVNKDGVKF
3gnpA     DTFAH-TFGKITDFSNADVAVDQYHRFEEDIQLMADMGMDAYRFSIAWSRIYPNGV--GQVNQAGIDH
BPR       DTFTHEFPGKISNGSTGDVADDFYHRYKEDVKVLKFIGLDGFRMSISWARVLPRGKLSGGVNKEGIAF
 _consrvd * *        *            **    *                * ** * *  * *      *

 _aln.pos  140       150       160       170       180       190       200
1cbgA     YNNLINEVLANGMQPYVTLFHWDVPQALEDEYRGFLG---RNIVDDFRDYAELCFKEFGDRVKHWITL
1h49A     YRNLINLLLENGIEPYVTIFHWDVPQALEEKYGGFLDKSHKSIVEDYTYFAKVCFDNFGDKVKNWLTF
1v02A     YNKLIDLLLENGIEPYITIFHWDTPQALVDAYGGFLD---ERIIKDYTDFAKVCFEKFGKTVKNWLTF
2jf7A     YHDFIDELLANGIKPSVTLFHWDLPQALEDEYGGFLS---HRIVDDFCEYAEFCFWEFGDKIKYWTTF
3gnpA     YNKLIDALLAKGIQPYVTLYHWDLPQALEDKYKGWLD---RQIVDDFAAYAETCFREFGDRVKHWITL
BPR       YNNVINDLLSKGIQPFITIFHWDLPQALEDEYGGFLS---PHIVNDFRDFAELCFKEFGDRVKHWITM
 _consrvd *   *   *  *  *  *  *** ****   * * *      *  *     *  **  **   * * *

 _aln.pos    210       220       230       240       250       260       270
1cbgA     NEPWGVSMNAYAYGTFAPGRCSDWLKLNCTGGDSGREPYLAAHYQLLAHAAAARLYKTKYQASQNGII
1h49A     NDPQTFTSVSYGTGVFAPGRCSPGLDCAYPTGNSLVEPYTAGHNILLAHAEAVDLYNKHYKRD-DTRI
1v02A     NEPETFCSVSYGTGVLAPGRCSPGVSCAVPTGNSLSEPYIVAHNLLRAHAETVDIYNKYHKGA-DGRI
2jf7A     NEPHTFAVNGYALGEFAPGRGGK-----GDEGDPAIEPYVVTHNILLAHKAAVEEYRNKFQKCQEGEI
3gnpA     NEPHTVAIQGYDAGLQAPGRCSVLLHLYCKAGNSGTEPYVVAHHFILAHAAAASIYRTKYKATQNGQL
BPR       NEPWSYSYGGYDAGLLAPGRCSAFMA-FCPKGNSGTEPYIVTHNLLLSHAAAVKLYKEKYQAYQKGQI
 _consrvd * *        *  *  ****            *     ***   *     *      *

 _aln.pos      280       290       300       310       320       330       340
1cbgA     GITLVSHWFEPASKEKADVDAAKRGLDFMLGWFMHPLTKGRYPESMRYLVRKRLPKFSTEESKELTGS
1h49A     GLAFDVMGRVPYGTSFLDKQAEERSWDINLGWFLEPVVRGDYPFSMRSLARERLPFFKDEQKEKLAGS
1v02A     GLALNVFGRVPYTNTFLDQQAQERSMDKCLGWFLEPVVRGDYPFSMRVSARDRVPYFKEKEQEKLVGS
2jf7A     GIVLNSMWMEPLSDVQADIDAQKRALDFMLGWFLEPLTTGDYPKSMRELVKGRLPKFSADDSEKLKGC
3gnpA     GIAFDVMWFEPMSNTTIDIEAAKRAQEFQLGWFADPFFFGDYPATMRARVGERLPRFTADEAAVVKGA
BPR       GITLVTYWMIPYSNSKADKDAAQRALDFMYGWFIEPLSFGEYPKSMRRLVGKRLPRFTKEQAMLVKGS
 _consrvd *        *  *  *       ***  *   * **  **      * * *          *

 _aln.pos        350       360       370       380       390       400
1cbgA     FDFLGLNYYSSYYAAKAPRIPNA--RPAIQTDSLINATFEH-NGKPLGPMAASSWLCIYPQGIRKLLL
1h49A     YNMLGLNYYTSRFSKNIDISPNYS-PVLNTDDAYASQEVNGPDGKPIGPPMGNPWIYMYPEGLKDLLM
1v02A     YDMIGINYYTSTFSKHIDLSPNNS-PVLNTDDAYASQETKGPDGNAIGPPTGNAWINMYPKGLHDILM
2jf7A     YDFIGMNYYTATYVTNAV-------KLSYETDDQVTKTFER-NQKPIGHALYGGWQHVVPWGLYKLLV
3gnpA     LDFVGINHYTTYYTRHNNTNIIGTLLNNTLADTGTVSLPFK-NGKPIGDRANSIWLYIVPRGMRSLMN
BPR       FDFLGLNYYIANYVLNVPTS--NSVNLSYTTDSLSNQTAFR-NGVAIGRPTGVPAFFMYPKGLKDLLV
 _consrvd     * * *                  *                 *              * *

 _aln.p  410       420       430       440       450       460       470
1cbgA     YVKNHYNNPVIYITENGRNEFNDP--TLSLQESLLDTPRIDYYYRHLYYVLTAIG-DGVNVKGYFAWS
1h49A     IMKNKYGNPPIYITENGIGDVDTKETPLPMEAALNDYKRLDYIQRHIATLKESID-LGSNVQGYFAWS
1v02A     TMKNKYGNPPMYITENGMGDIDKG--DLPKPVALEDHTRLDYIQRHLSVLKQSID-LGADVRGYFAWS
2jf7A     YTKETYHVPVLYVTESGMVEENKT--KILLSEARRDAERTDYHQKHLASVRDAID-DGVNVKGYFVWS
3gnpA     YVKERYNSPPVYITENGMDDSNNP--FISIKDALKDSKRIKYHNDYLTNLAASIKEDGCDVRGYFAWS
BPR       YTKEKYNDPVIYITENGMGDNNN----VTTEEGIKDPQRVYFYNQHLLSLKNAIA-AGVKVKGYFTWA
 _consrvd   *  *  *  * ** *                  *  *              *   *  * *** *
```

Figure 1. Multiple sequence alignment between (A) Phytoene synthase from *Camellia sinensis,* Squalene synthase from *Homo sapiens* (PDB Id - 1EZF:A) and Dehydrosqualene synthase from *Staphylococcus Aureus* (PDB Id – 2ZCO:A) (B) β-primeverosidase from *Camellia sinensis,* Cyanogenic Beta-glucosidase from *Trifolium repen* (PDB Id – 1CBG:A), Beta-glucosidase mutant from *Zea mays* (PDB Id – 1H49:A), Dhurrinase from *Sorghum bicolour* (PDB Id – 1V02:A), Strictosidine-o-beta-D-glucosidase from *Rauvolfia serpentine* (PDB Id – 2JF7:A) and Hydrolase from *Oryza sativa subsp. japonica* (PDB Id – 3GNP:A). "*" indicates a conserved region between two sequences and "-" indicates gaps.

```
_aln.pos          10        20        30        40        50        60
2gasA     ---------TENKILILGPTGAIGRHIVWASIKAGNPTYALVRKTITAANPETKEELIDNYQSLGVIL
1qycA     --------GSRSRILLIGATGYIGRHVAKASLDLGHPTFLLVRESTASSNS-EKAQLLESFKASGANI
1qydA     --------DKKSRVLIVGGTGYIGKRIVNASISLGHPTYVLFRPEVVSN--IDKVQMLLYFKQLGAKL
2qw8A     --------GMKSKILIFGGTGYIGNHMVKGSLKLGHPTYVFTRPNS------SKTTLLDEFQSLGAII
LAR       MTVLESVSAAGGGVLIVGASGFIGQFIAEASLHADRPTYLLVRSVG--------SKTNKTLQDKGAKV
 _consrvd               *  *  * **      *     **    *                    *

 _aln.p   70        80        90       100       110       120       130
2gasA     LEGDINDHETLVKAIK--QVDIVICAAGR----LLIEDQVKIIKAIKEAGNVKKFFPSEFGLDVDR-H
1qycA     VHGSIDDHASLVEAVK--NVDVVISTVGS----LQIESQVNIIKAIKEVGTVKRFFPSEFGNDVDN-V
1qydA     IEASLDDHQRLVDALK--QVDVVISALAGGVLSHHILEQLKLVEAIKEAGNIKRFLPSEFGMDPDIME
2qw8A     VKGELDEHEKLVELMK--KVDVVISALAF----PQILDQFKILEAIKVAGNIKRFLPSDFGVEEDR-I
LAR       IPGVVKDQAFMEKILKEHKIDIVISAIGG----ANILDQLTLVHAIKAVGTIKRFLPSEFGHDVDR-A
 _consrvd                *    * **           *  *     ***  *  * * ** **   *

 _aln.pos  140       150       160       170       180       190       200
2gasA     DAVEPVRQVFEEKASIRRVIEAEGVPYTYLCCHAFTGYFLRNLAQLD-ATDPPRDKVVILGDGNVKGA
1qycA     HAVEPAKSVFEVKAKVRRAIEAEGIPYTYVSSNCFAGYFLRSLAQAG-LTAPPRDKVVILGDGNARVV
1qydA     HALQPGSITFIDKRKVRRAIEAASIPYTYVSSNMFAGYFAGSLAQLDGHMMPPRDKVLIYGDGNVKGI
2qw8A     NALPPFEALIERKRMIRRAIEEANIPYTYVSANCFASYFINYLL----RPYDPKDEITVYGTGEAKFA
LAR       NPVEPGLTMYNEKRRVRRLIEECGVPYTYICCNSIASWPYYDNTHPS-EVIPPLDEFQIYGDGSVKAY
 _consrvd     *       *   ** **    ****                       * *     * *

 _aln.pos    210       220       230       240       250       260       270
2gasA     YVTEADVGTFTIRAANDPNTLNKAVHIRLPKNYLTQNEVIALWEKKIGKTLEKTYVSEEQVLKDIQES
1qycA     FVKEEDIGTFTIKAVDDPRTLNKTLYLRLPANTLSLNELVALWEKKIDKTLEKAYVPEEEVLKLIADT
1qydA     WVDEDDVGTYTIKSIDDPQTLNKTMYIRPPMNILSQKEVIQIWERLSEQNLDKIYISSQDFLADMKDK
2qw8A     MNYEQDIGLYTIKVATDPRALNRVVIYRPSTNIITQLELISRWEKKIGKKFKKIHVPEEEIVALTKEL
LAR       FVAGSDIGKFTIKTVDDIRTLNKSVHFRPSCNFLNINELASLWEKKIGRTLPRVTVSENDLLAAAAVN
 _consrvd      * *  **    *   **     *   *     *    **

 _aln.pos      280       290       300       310       320       330       340
2gasA     SFPHNYLLALYHSQQIKGDAVYEIDP-AKDIEASEAYPDVTYTTADEYLNQFV---------------
1qycA     PFPANISIAISHSIFVKGDQTNFEIG-PAGVEASQLYPDVKYTTVDEYLSNFV---------------
1qydA     SYEEKIVRCHLYQIFFRGDLYNFEIG-PNAIEATKLYPEVKYVTMDSYLERYV---------------
2qw8A     PEPENIPIAILHCLFIDGATMSYDFK-ENDVEASTLYPELKFTTIDELLDIFVH--------D-----
LAR       IIPRSVVASFTHDIFIKGCQINFSIEGPNDVEVCSLYPDESFRTVGECFDDFVVKMNGKNFTDETDGN
 _consrvd                 *            *    **     *        *

 _aln.pos        350
2gasA     ----------------
1qycA     ----------------
1qydA     ----------------
2qw8A     -------PPPPASAAF
LAR       TAQNHVVEVLPITMCA
 _consrvd
```

Figure 2. Multiple sequence alignment between (A) Flavanone-3-hydroxylase from *Camellia sinensis,* Leucoanthocyanidin dioxygenase from *Arabidopsis thaliana* (PDB Id - 1GP6:A) and 1-aminocyclopropane-1-carboxylate-oxidase 1 from *Petunia Hybrida* (PDB Id - 1W9Y:A) (B) Dihydroflavonol 4-reductase from *Camellia sinensis,* Dihydroflavonol 4-reductase from *Vitis Vinifera* (PDB Id - 2C29:A) and Chaperone from *Paracoccus denitrificans* pd1222. Hydrolase from *Actinomadura sp.r39* (PDB Id - 2P4H:X).

```
_aln.pos          10        20        30        40        50        60
2gasA     ---------TENKILILGPTGAIGRHIVWASIKAGNPTYALVRKTITAANPETKEELIDNYQSLGVIL
1qycA     --------GSRSRILLIGATGYIGRHVAKASLDLGHPTFLLVRESTASSNS-EKAQLLESFKASGANI
1qydA     --------DKKSRVLIVGGTGYIGKRIVNASISLGHPTYVLFRPEVVSN--IDKVQMLLYFKQLGAKL
2qw8A     --------GMKSKILIFGGTGYIGNHMVKGSLKLGHPTYVFTRPNS------SKTTLLDEFQSLGAII
LAR       MTVLESVSAAGGGVLIVGASGFIGQFIAEASLHADRPTYLLVRSVG--------SKTNKTLQDKGAKV
 _consrvd               *  *  * **      *     **    *                    *

 _aln.p   70        80        90       100       110       120       130
2gasA     LEGDINDHETLVKAIK--QVDIVICAAGR----LLIEDQVKIIKAIKEAGNVKKFFPSEFGLDVDR-H
1qycA     VHGSIDDHASLVEAVK--NVDVVISTVGS----LQIESQVNIIKAIKEVGTVKRFFPSEFGNDVDN-V
1qydA     IEASLDDHQRLVDALK--QVDVVISALAGGVLSHHILEQLKLVEAIKEAGNIKRFLPSEFGMDPDIME
2qw8A     VKGELDEHEKLVELMK--KVDVVISALAF----PQILDQFKILEAIKVAGNIKRFLPSDFGVEEDR-I
LAR       IPGVVKDQAFMEKILKEHKIDIVISAIGG----ANILDQLTLVHAIKAVGTIKRFLPSEFGHDVDR-A
 _consrvd                *    * **           *  *     ***  *  * * ** **   *

 _aln.pos  140       150       160       170       180       190       200
2gasA     DAVEPVRQVFEEKASIRRVIEAEGVPYTYLCCHAFTGYFLRNLAQLD-ATDPPRDKVVILGDGNVKGA
1qycA     HAVEPAKSVFEVKAKVRRAIEAEGIPYTYVSSNCFAGYFLRSLAQAG-LTAPPRDKVVILGDGNARVV
1qydA     HALQPGSITFIDKRKVRRAIEAASIPYTYVSSNMFAGYFAGSLAQLDGHMMPPRDKVLIYGDGNVKGI
2qw8A     NALPPFEALIERKRMIRRAIEEANIPYTYVSANCFASYFINYLL----RPYDPKDEITVYGTGEAKFA
LAR       NPVEPGLTMYNEKRRVRRLIEECGVPYTYICCNSIASWPYYDNTHPS-EVIPPLDEFQIYGDGSVKAY
 _consrvd     *       *   ** **    ****                       * *     * *

 _aln.pos    210       220       230       240       250       260       270
2gasA     YVTEADVGTFTIRAANDPNTLNKAVHIRLPKNYLTQNEVIALWEKKIGKTLEKTYVSEEQVLKDIQES
1qycA     FVKEEDIGTFTIKAVDDPRTLNKTLYLRLPANTLSLNELVALWEKKIDKTLEKAYVPEEEVLKLIADT
1qydA     WVDEDDVGTYTIKSIDDPQTLNKTMYIRPPMNILSQKEVIQIWERLSEQNLDKIYISSQDFLADMKDK
2qw8A     MNYEQDIGLYTIKVATDPRALNRVVIYRPSTNIITQLELISRWEKKIGKKFKKIHVPEEEIVALTKEL
LAR       FVAGSDIGKFTIKTVDDIRTLNKSVHFRPSCNFLNINELASLWEKKIGRTLPRVTVSENDLLAAAAVN
 _consrvd      * *  **    *   **     *   *     *    **

 _aln.pos      280       290       300       310       320       330       340
2gasA     SFPHNYLLALYHSQQIKGDAVYEIDP-AKDIEASEAYPDVTYTTADEYLNQFV---------------
1qycA     PFPANISIAISHSIFVKGDQTNFEIG-PAGVEASQLYPDVKYTTVDEYLSNFV---------------
1qydA     SYEEKIVRCHLYQIFFRGDLYNFEIG-PNAIEATKLYPEVKYVTMDSYLERYV---------------
2qw8A     PEPENIPIAILHCLFIDGATMSYDFK-ENDVEASTLYPELKFTTIDELLDIFVH--------D-----
LAR       IIPRSVVASFTHDIFIKGCQINFSIEGPNDVEVCSLYPDESFRTVGECFDDFVVKMNGKNFTDETDGN
 _consrvd                  *             *    **     *        *

 _aln.pos        350
2gasA     ----------------
1qycA     ----------------
1qydA     ----------------
2qw8A     -------PPPPASAAF
LAR       TAQNHVVEVLPITMCA
 _consrvd
```

Figure 3. Multiple sequence alignment between Leucoanthocyanidin reductase from *Camellia sinensis, isoflavone reductase from Medicago sativa* (PDB Id - 2GAS:A), Phenylcoumaran benzylic ether reductase from *Pinus taeda* (PDB Id - 1QYC:A), pinoresinol-lariciresinol reductase from *Thuja plicata* (PDB Id - 1QYD:A) and Eugenol synthase 1 from *Osimum basilicum* (PDB Id - 2QW8:A).

papaya and *Citrus unshui* while *Momor charantia* exists as mono taxa in it. Phytoene Synthase from *C. sinensis* is closest to *D. kaki* and *A. deliciosa*. The most primitive forms of the enzyme include the ones in *Elaegnus umbellate* and *Momor charantia*.

The phylogenetic tree for flavanone-3-hydroxylase can be divided into 2 groups and one clade (Figure 5). The clade contains *C. sinensis* and *Actinidia chinensis*. Group 1 has one clade which contains *Rubus coreanus* and *Pyrus communis* while *Epimedum sagittatum* exists as mono taxa near it. Group 2 contains 1 subgroup SG1 and clade which has *Gossypium hirsutum* and Dimocarpus longan. The subgroup contains *Solanum tuberosum* and *Nicotiana tabacum* in a clade while *Eustoma grandiflorum*

Table 2. Closely related protein sequences for PSY obtained using $BLAST_P$. The sequences with high identities with the target were used for multiple sequence alignment and phylogenetic analysis. The alignment score was obtained from multiple sequence alignment using ClustalW webserver and the distances were obtained using PhyloDraw by the neighbour joining method.

S.No	Accesssion number	Source	Identity (%)	Alignment score	Distance (s)
1	ACM44688	*D. kaki*	75	79	0.22
2	ACO53104	*A. deliciosa*	73	76	0.22
3	ABU40771	*S. lycopersicum*	73	75	0.25
4	ABA43898	*Coffea canephora*	70	74	0.25
5	ABG72805	*C. papaya*	70	73	0.26
6	AAF33237	*C. unshiu*	71	72	0.28
7	AAR86104	*M. charantia var. abbreviate*	72	71	0.27
8	ACU29637	*E. umbellata*	69	72	0.28
9	CAC27383	*H. annuus*	71	72	0.27
10	BAE45299	*G. lutea*	68	71	0.31

Table 3. Closely related protein sequences for flavanone-3-hydroxylase (F3H) obtained using $BLAST_P$ which were used for comparative sequence analysis.

S.No	Accesssion number	Organism	Identity (%)	Alignment score	Distances
1	ACL54955	*A. chinensis*	89	89	0.1
2	ABW74548	*R. coreanus*	86	86	0.15
3	ABM64799	*G. hirsutum*	86	85	0.13
4	ABO48521	*D. longan*	86	86	0.13
5	BAD34459	*E. grandiflorum*	84	84	0.15
6	AAX89399	*P.communis*	84	85	0.16
7	Q7XZQ7	*P. crispum*	84	82	0.18
8	ABY63660	*E. sagittatum*	84	82	0.17
9	AAM48289	*S. tuberosum*	84	85	0.15
10	BAF96938	*N. tabacum*	83	82	0.15

Table 4. Closely related protein sequences for dihydroflavonol-4-reductase (DFR) obtained using $BLAST_P$.

S.No	Accesssion number	Organism	Identity (%)	Alignment score	Distances
1	CAC88859	*R. simsii*	88	87	0.13
2	ACK57789	*C. maculosa*	81	77	0.21
3	AAL89715	*V.macrocarpon*	85	79	0.2
4	ABU93477	*H.annuus*	80	76	0.23
5	ABQ97018	*S. medusa*	80	77	0.22
6	P14721	*A. majus*	77	74	0.25
7	ACB56920	*H. pilosella*	81	76	0.24
8	BAA12736	*G. triflora*	79	75	0.24
9	AAD56578	*D. carota*	79	73	0.27
10	ACN82380	*V. amurensis*	77	75	0.24

exists as mono taxa near this clade. *C. sinensis* is closest to *Actinidia chinensis*. The most primitive forms of the enzyme include *C. sinensis* and *A. chinensis,* which exit as an outgroup to Group 1 and 2, and *Petroselinum crispum* which exists as a segregated branch from Group 2.The molecular phylogenetic analysis of dihydroflavonol-4-reductase gave 2 groups (Figure 6). Group 1 has one subgroup namely SG1 and one clade which contains *Gentiana triflora* and *Antirrhinum majus*. SG1 contains 1 clade, taxa Rhododendron simsii and *Vaccinium macrocarpon* occur in it while *C. sinensis* exists as mono taxae near this clade. Group 2 contains 1 clade which

Table 5. Closely related protein sequences for leucoanthocyanidin reductase (LAR) obtained using $BLAST_P$.

S.No	Accesssion number	Organism	Identity (%)	Alignment score	Distances
1	CAI56319	*G. arboreum*	70	69	0.31
2	XP_002314885	*P. trichocarpa*	69	68	0.34
3	CAI56326	*Vi. shuttle worthii*	70	66	0.34
4	ABB77696	*P. communis*	66	64	0.33
5	ABC71328	*L. corniculatus*	65	63	0.35
6	CAI56327	*M.turcatula*	63	61	0.4
7	CAI56322	*P. coccineus*	63	61	0.39
8	CAI56328	*O.sativa*	60	56	0.43
9	CAI56320.1	*H. vulgare*	62	57	0.43
10	CAI56321	*P. taeda*	59	52	0.47

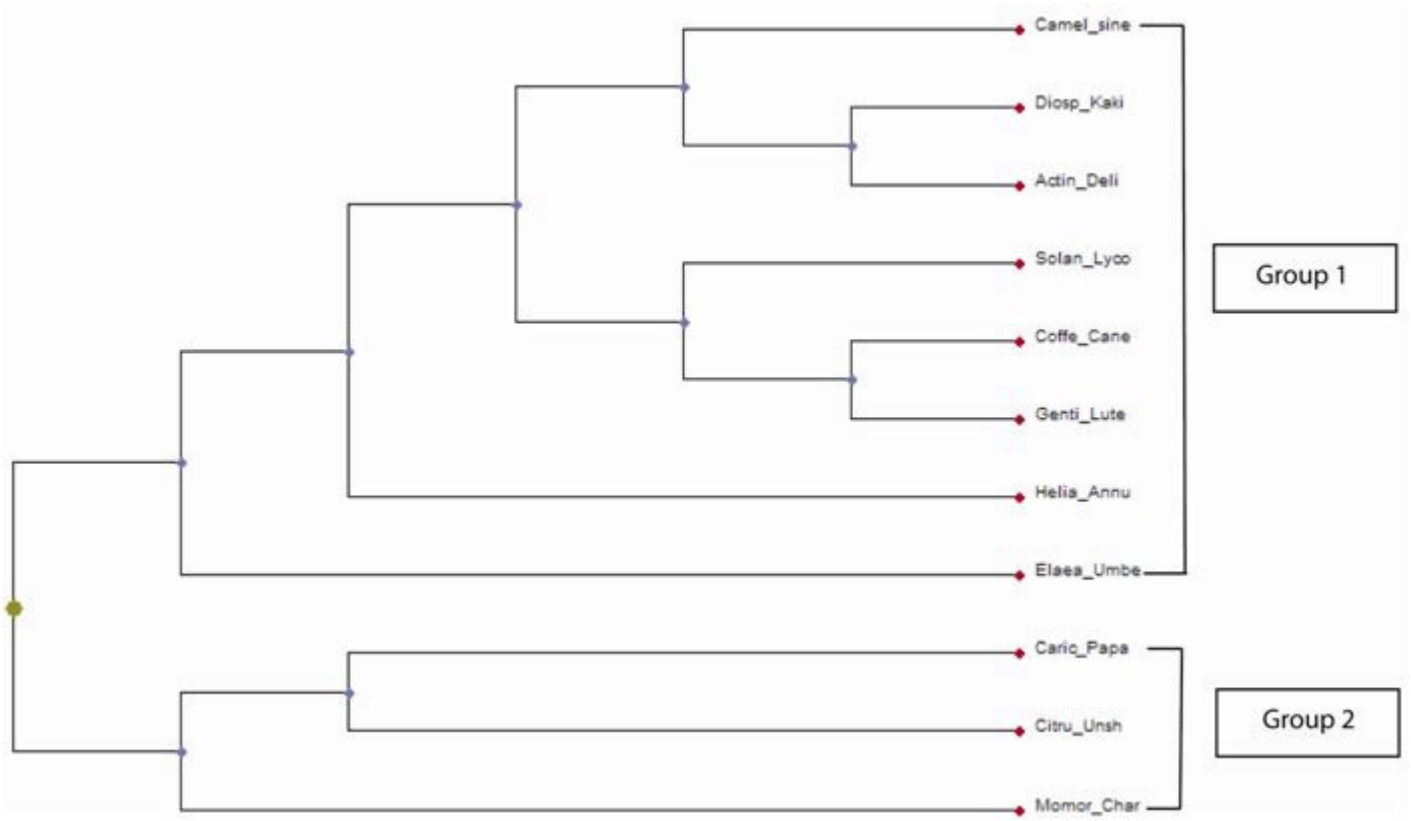


Figure 4. The phylogenetic tree of phytoene synthase with its homologs. The tree were constructed by neighbour joining method using PhyloDraw.

has *Centaurea maculosa* and *Saussurea medusa* while *Hieracium pilosella* exists as mono taxa near this clade. Group 2 also has *Helianthus annus, Daucus carota* and *Vitis amurensis* as mono taxae. DFR from *C. sinensis* is closest to *Rhododendron simsii* and *Vaccinium macrocarpon*. The most primitive branch is the enzyme from *Vitis amurensis*.

Two groups were obtained from the molecular phylogeny of LAR (Figure 7). Group 1 contains 1 clade and 1 subgroup. The clade has 2 taxae: *C. sinensis* and *Gossypium arboretum* while the subgroup contains *Oryza sativa* and *Hordeum vulgae* in 1 clade while *Pinus taeda* exists as a mono taxa near this clade. Group 2 contains two subgroups. Subgroup 1 contains a clade with *Populus trichocarpa* and *Pyrus communis* while *Vitis shuttleworthi* exists a mono taxa near it. Similarly subgroup 2 contains *Medicago truncatula* and *Phaseolus coccineus* in a clade with mono taxa *Lotus corniculatus* close to it. LAR from *C. sinensis* is closest to *G. arboretum* with which it shares the maximum identity.

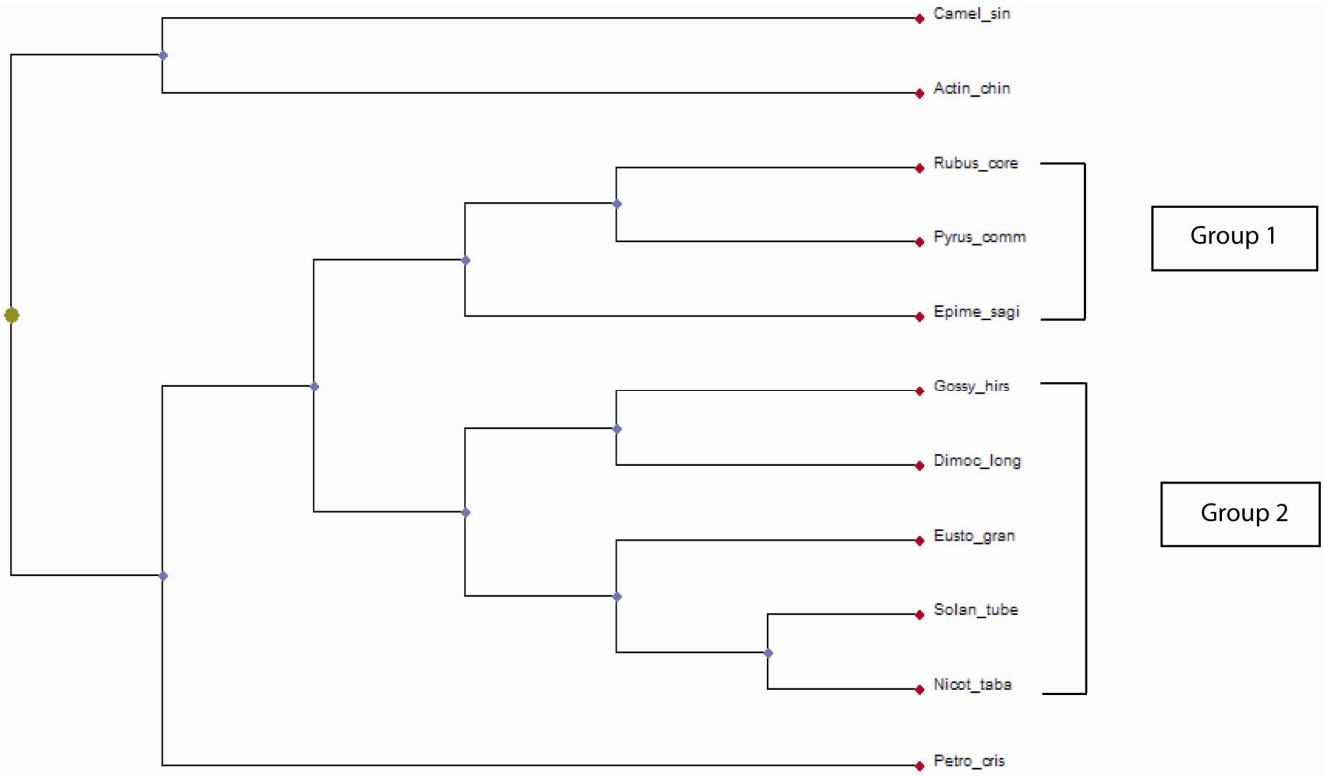


Figure 5. The phylogenetic tree of flavanone-3-hydroxylase with its homologs.

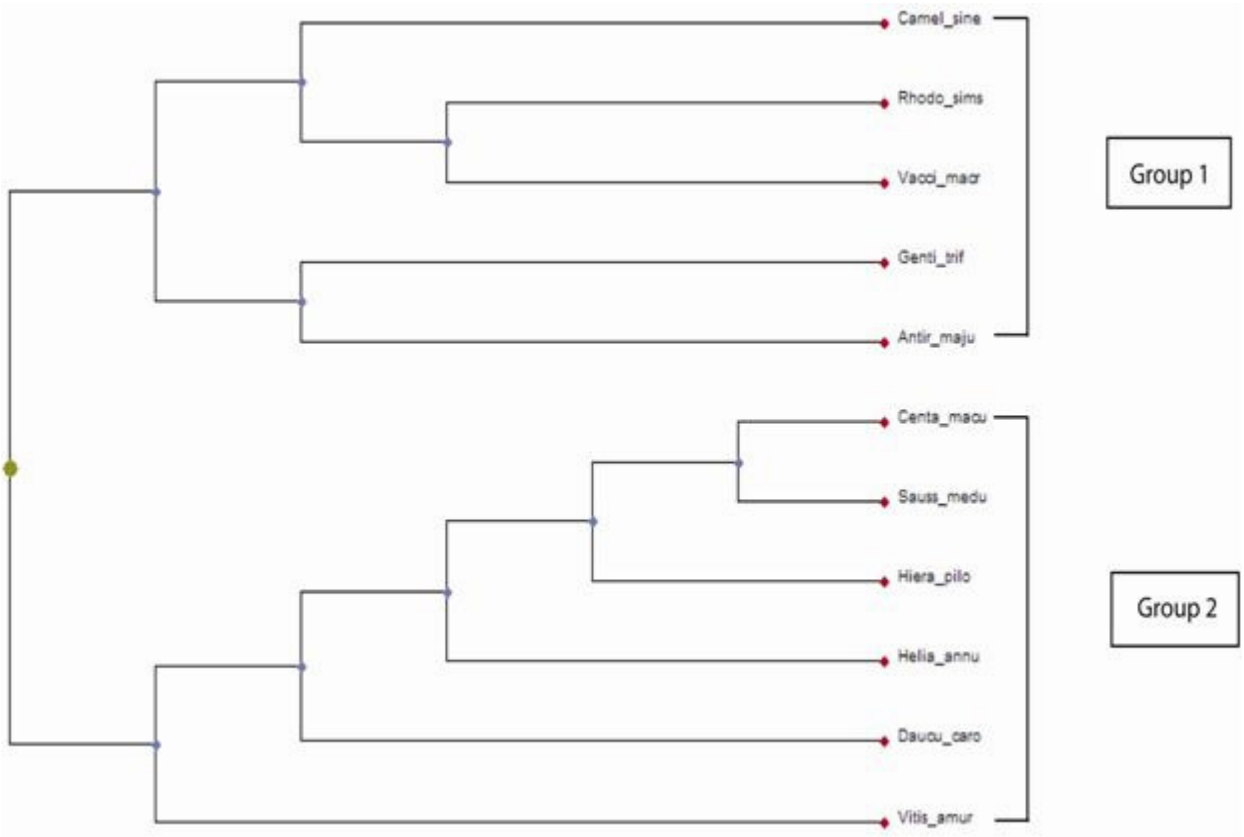


Figure 6. The phylogenetic tree of dihydroflavonol-4-reductase with its homologs.

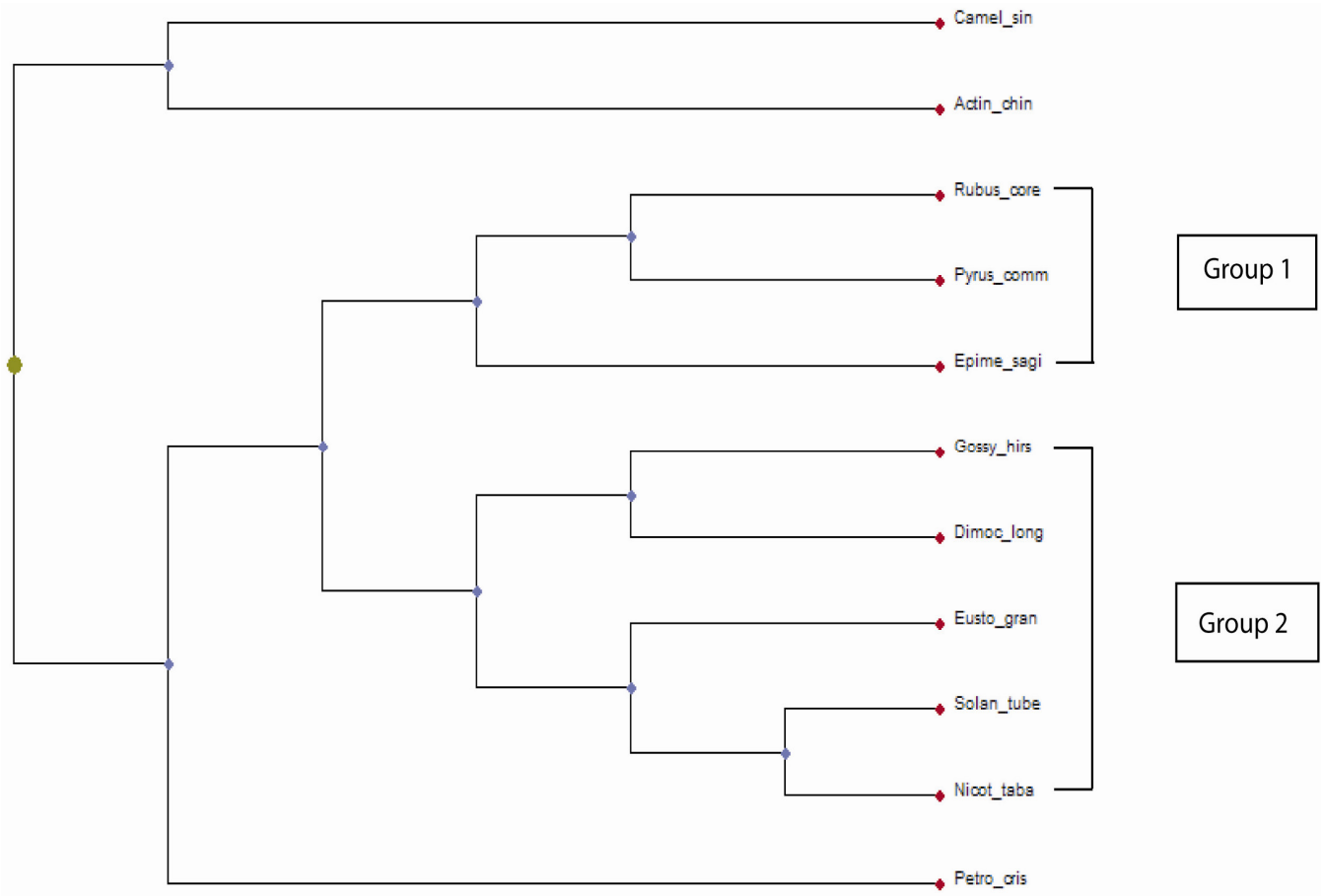


Figure 7. The phylogenetic tree of leucoanthocyanidin reductase with its homologs.

Table 6. Regions/loops of poor DOPE profile.

S.No.	Enzyme model	No. of regions	Residues
1	β-primeverosidase	2	418 to 428, 1 to 7
2	Phytoene Synthase	2	47 to 55, 103 to 108
3	Dihydroflavonol-4-reductase	3	73 to 77, 142 to 149, 161 to 167
4	Flavanone-3-hydroxylase	6	1 to 7, 22 to 28, 105 to 130, 220 to 220, 230 to 242, 340 to 345
5	Leucoanthocyanidin reductase	3	8 to 13, 308 to 321, 330 to 335

Prediction of homology models

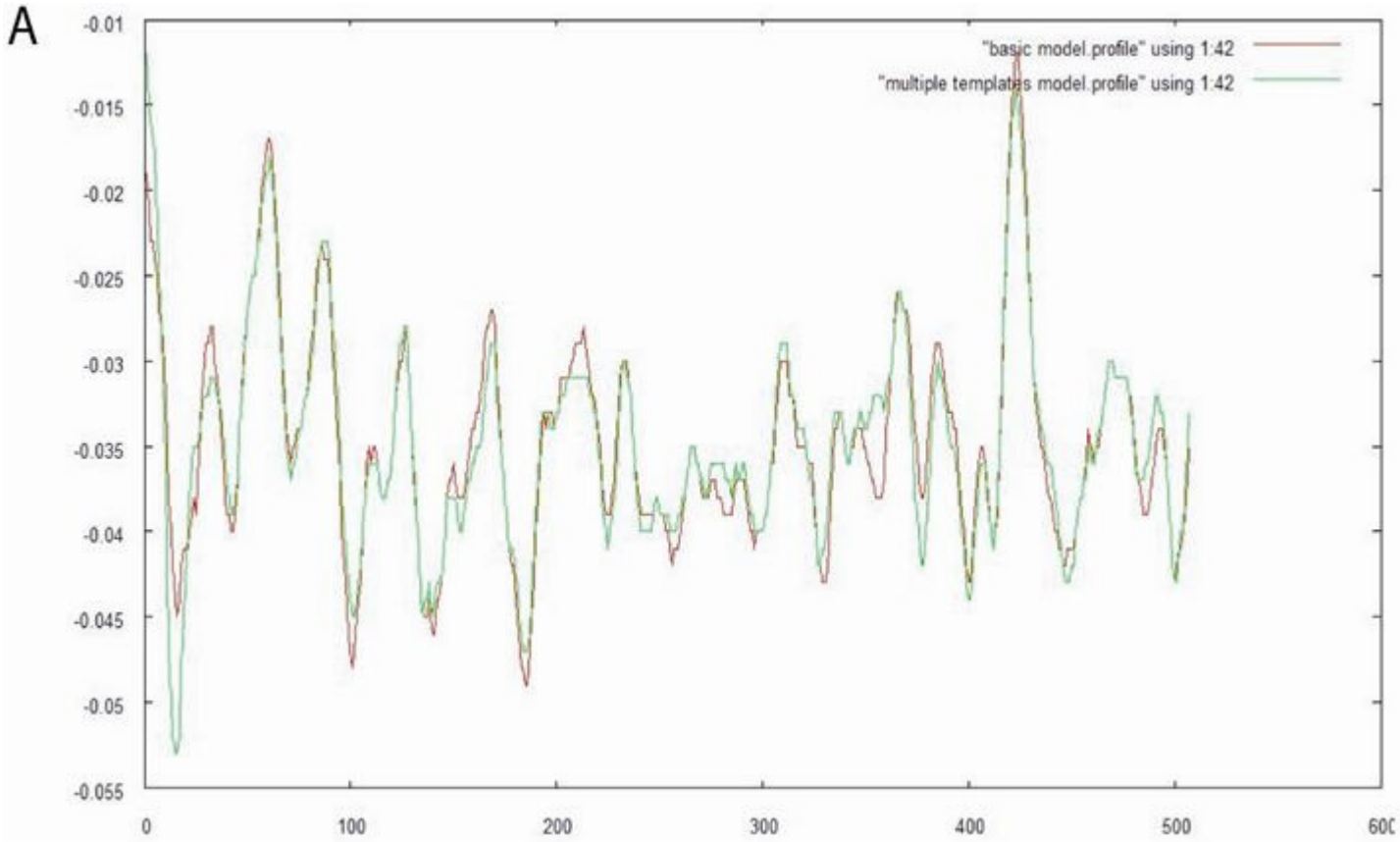
The structures used as templates for the respective target enzymes to develop the basic and multiple-template models have been give in Table 1. The multiple templates were aligned using the align2d_mult.py script.

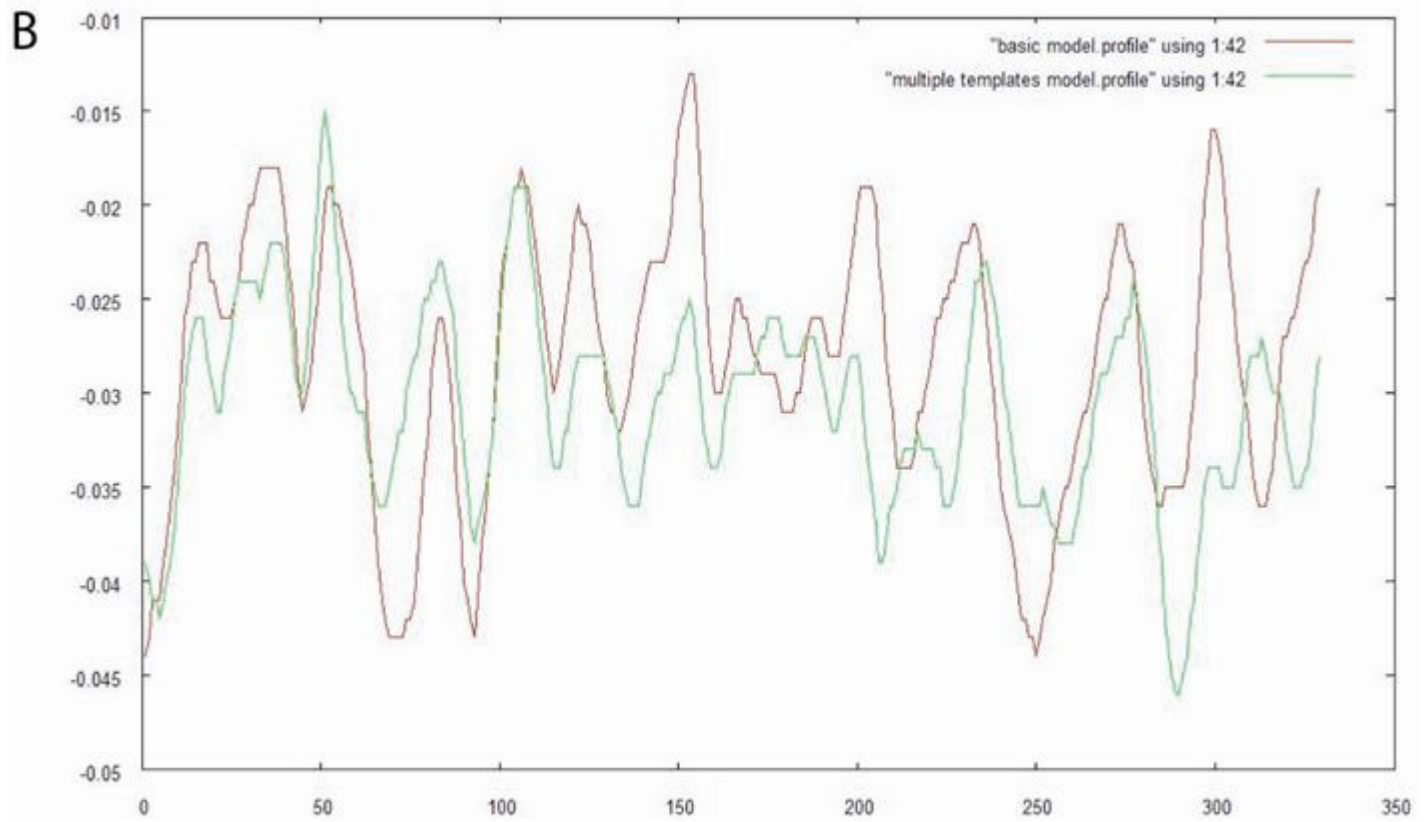
The models were generated using the model_mult.py script and the model with the lowest DOPE score was selected. The DOPE profiles of the basic models were plotted along with the multiple templates model using GNUPLOT (Figures 8 A, B, C and D). It can be seen that the plots for the improved models are better than the basic models.

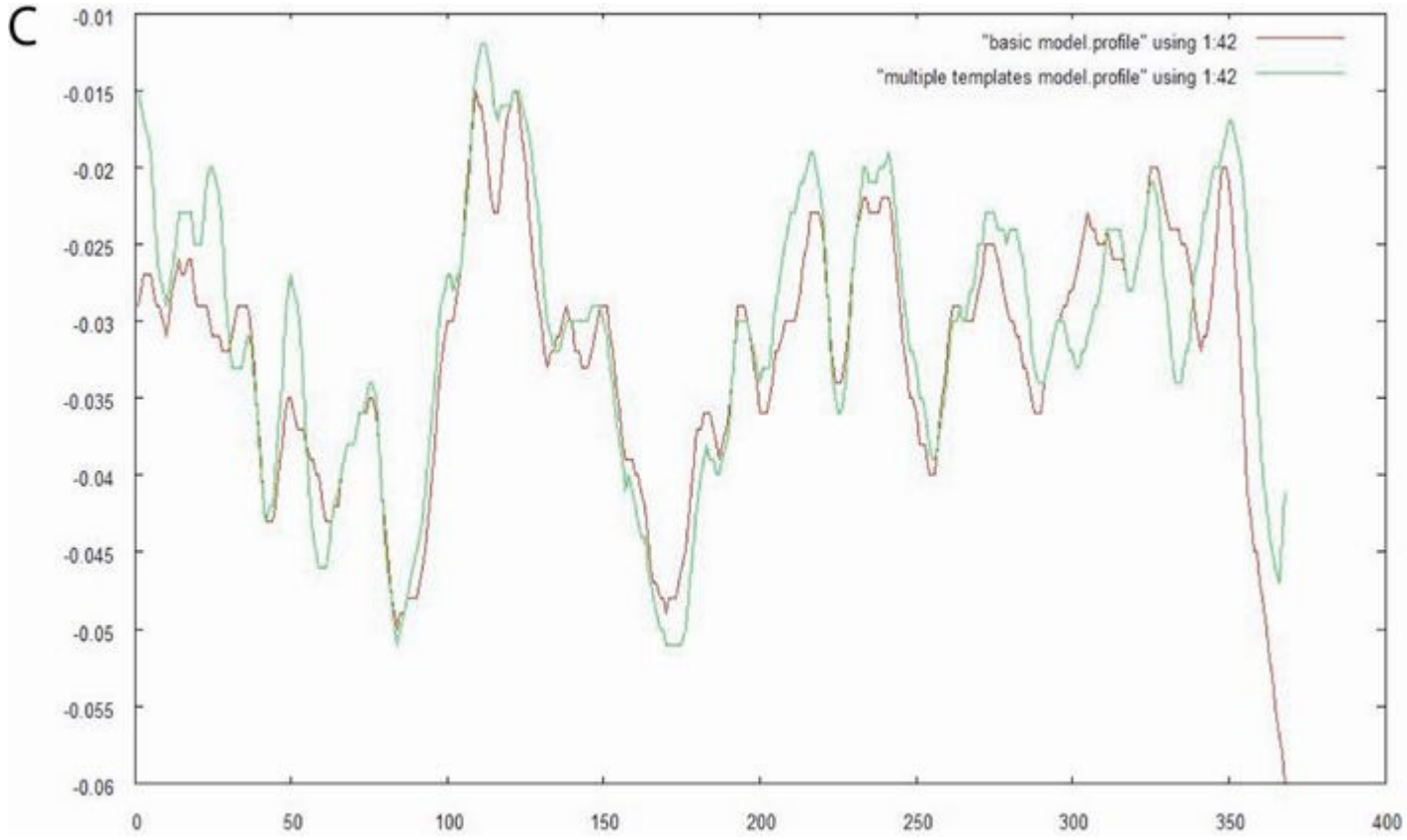
Loop refining

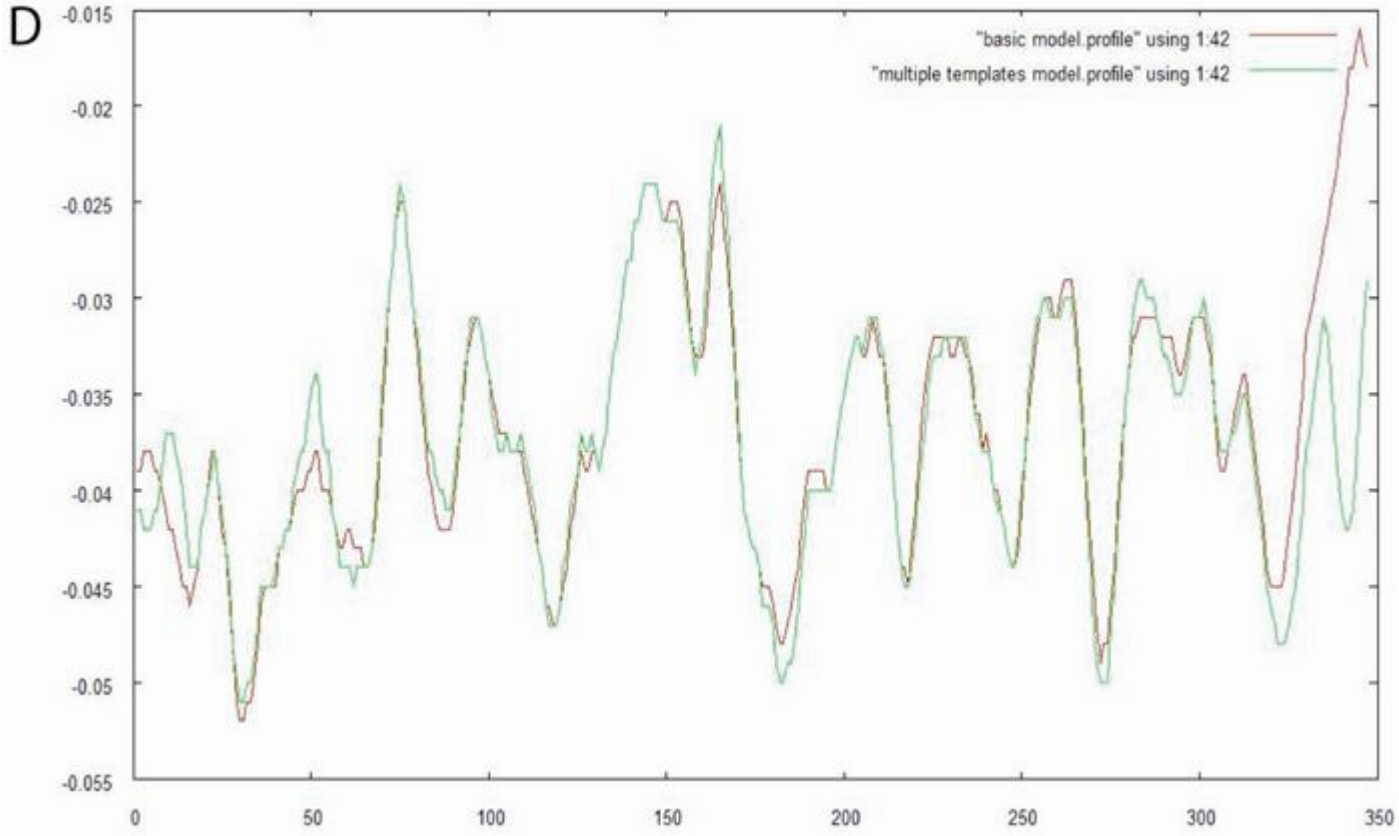
The DOPE score profiles were used to identify the loops with poor DOPE score in the multiple template models. The regions/loops which were poorly modelled have been given in Table 6. The loop_refine.py script of Modeller 9v7 was used to improve the structure of these regions. This resulted in further decrease in the DOPE score. The DOPE scores of the models generated by basic modeling and advanced modeling as well as the final models obtained after loop refining have been given in Table 7.

We see that the model quality has improved by following the advanced modeling procedure followed by


A
"basic model.profile" using 1:42
"multiple templates model.profile" using 1:42
-0.01
-0.015
-0.02
-0.025
-0.03
-0.035
-0.04
-0.045
-0.05
-0.055
0
100
200
300
400
500
600


B
"basic model.profile" using 1:42
"multiple templates model.profile" using 1:42
-0.01
-0.015
-0.02
-0.025
-0.03
-0.035
-0.04
-0.045
-0.05
0
50
100
150
200
250
300
350


C
"basic model.profile" using 1:42
"multiple templates model.profile" using 1:42
-0.01
-0.015
-0.02
-0.025
-0.03
-0.035
-0.04
-0.045
-0.05
-0.055
-0.06
0
50
100
150
200
250
300
350
400


D
"basic model.profile" using 1:42
"multiple templates model.profile" using 1:42
-0.015
-0.02
-0.025
-0.03
-0.035
-0.04
-0.045
-0.05
-0.055
0
50
100
150
200
250
300
350

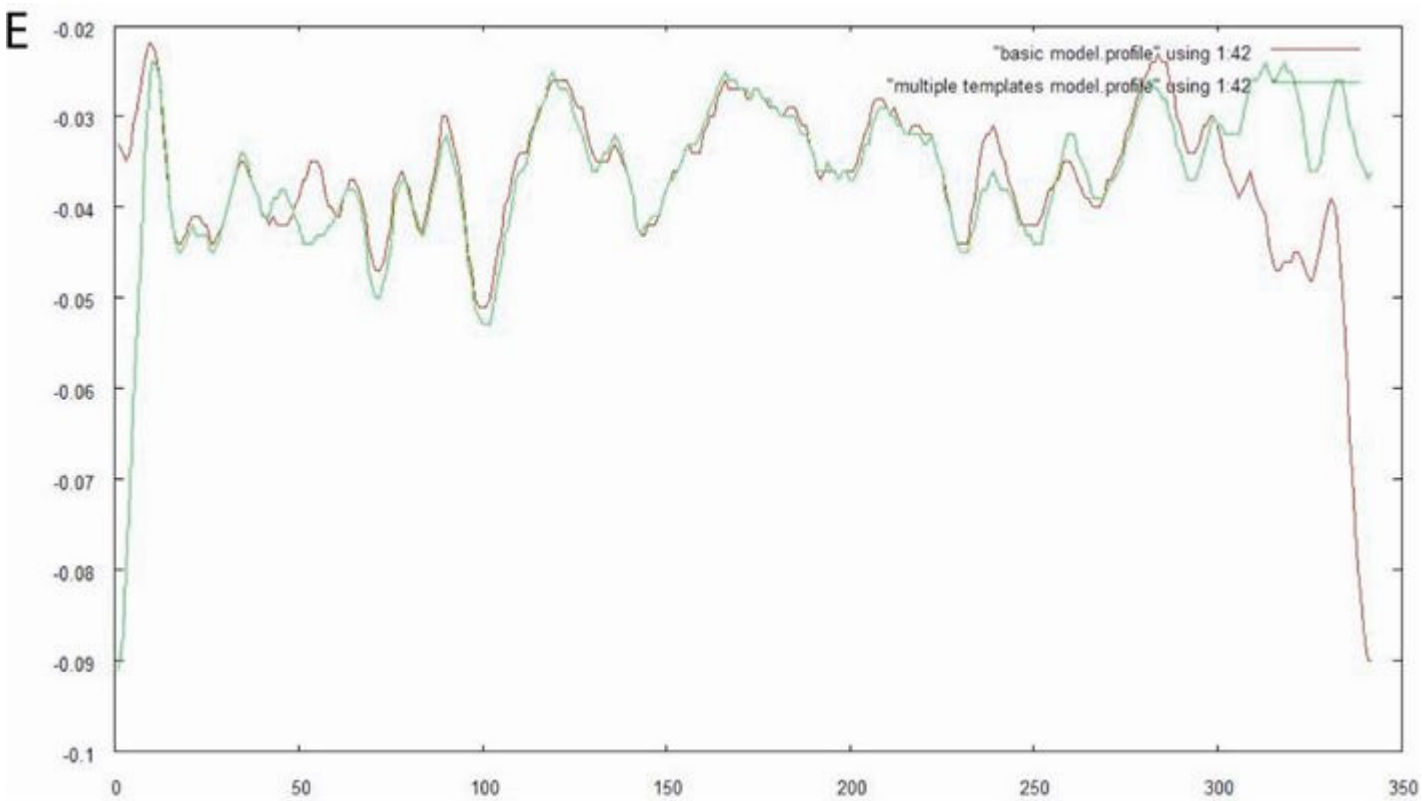


Figure 8. DOPE score profile of the basic model and multiple templates model for (A) BPR (B) PSY (C) DFR (D) F3H (E) LAR.

Table 7. DOPE score of basic models, multiple templates models and final models.

0	Enzyme model	Basic model DOPE score	Multiple templates model DOPE score	Final models DOPE score (after loop refining)
1	β-primeverosidase	-64611.58984	-66076.109375	-66212.117188
2	Phytoene Synthase	-30988.27148	-31887.291016	-32060.505859
3	Dihydroflavonol-4-reductase	-41902.91797	-41907.683594	-42024.601563
4	Flavanone-3-hydroxylase	-36218.65625	-36421.832031	-36973.710938
5	Leucoanthocyanidin reductase	-35788.60156	-38037.937500	-38435.492188

Table 8. Energy of models generated before and after minimization.

S.No.	Enzyme model	Initial potential energy (Kcal/mol)	Energy after minimization (Kcal/mol)
1	β-primeverosidase	1207309.9693	-29290.03714
2	Phytoene Synthase	715697.94340	-70384.83381
3	Dihydroflavonol-4-reductase	23983.39479	-19322.86169
4	Flavanone-3-hydroxylase	52848.11603	-20668.01549
5	Leucoanthocyanidin reductase	12939.21057	-18415.83381

loop refining. These improved final models were subjected to energy minimizations using discovery studio 2.1. The energy minimization results are shown in Table 8. The force field applied was CHARMM and the energy minimization algorithm used was Steepest Descent with an RMS gradient of 0.1 using a maximum of 1000 steps. The structures of the models obtained after energy minimization have been given in Figure 9.

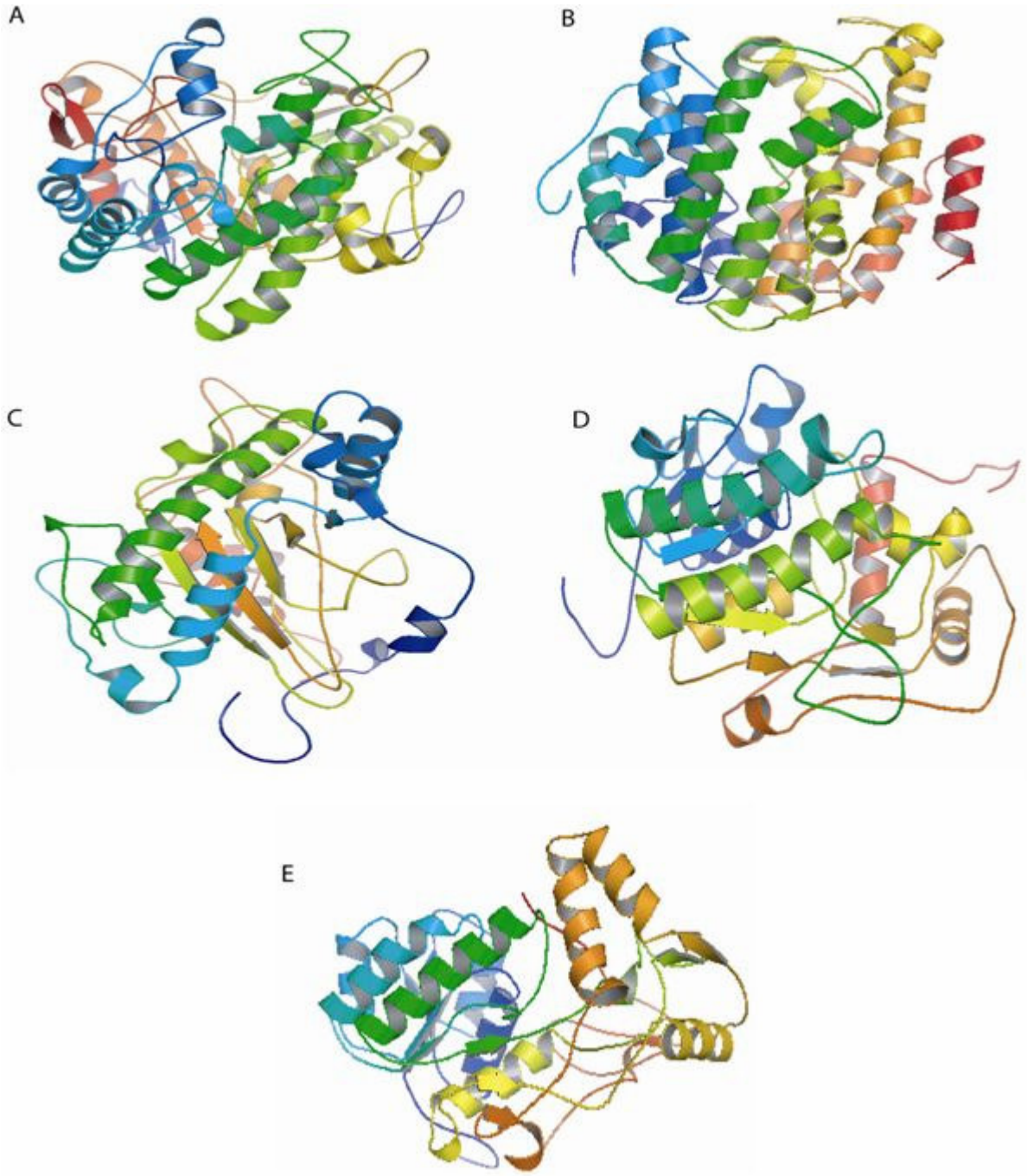


Figure 9. Models of (A) Phytoene synthase, (B) β-primeverosidase, (C) Flavanone-3-hydroxylase, (D) Dihydroflavonol 4-reductase, (E) Leucoanthocyanidin reductase.

Binding site analysis

$CAST_P$ server was used to locate the putative binding sites in each of the enzyme models with a probe radius of 1.4 A. The proposed binding sites have been shown in Figure 10. The binding site residues predicted by $CAST_P$ lie in the conserved regions obtained by the multiple sequence alignment of PSY, F3H, DFR and LAR. The binding site residues for phytoene synthase were: N12, V15, S16, S17, G18, F19, G20, L22, E23, F37, S38, P39, E41, R42, L44, I45, C46, H47, G48, R49, F50, K51, S53, R54, N55, T58, R62, N86, K87, R88, I115, V116, G119, L123, E126, A127, D129, T141, L144, G145, T146, L148, M149, R153, S177, H178, P181, T182, D185, R186, E188, S189, E192, R196, G197, R198, F200, L243, Y244, L245, Y248, L256, M257, V259, P260, V161,

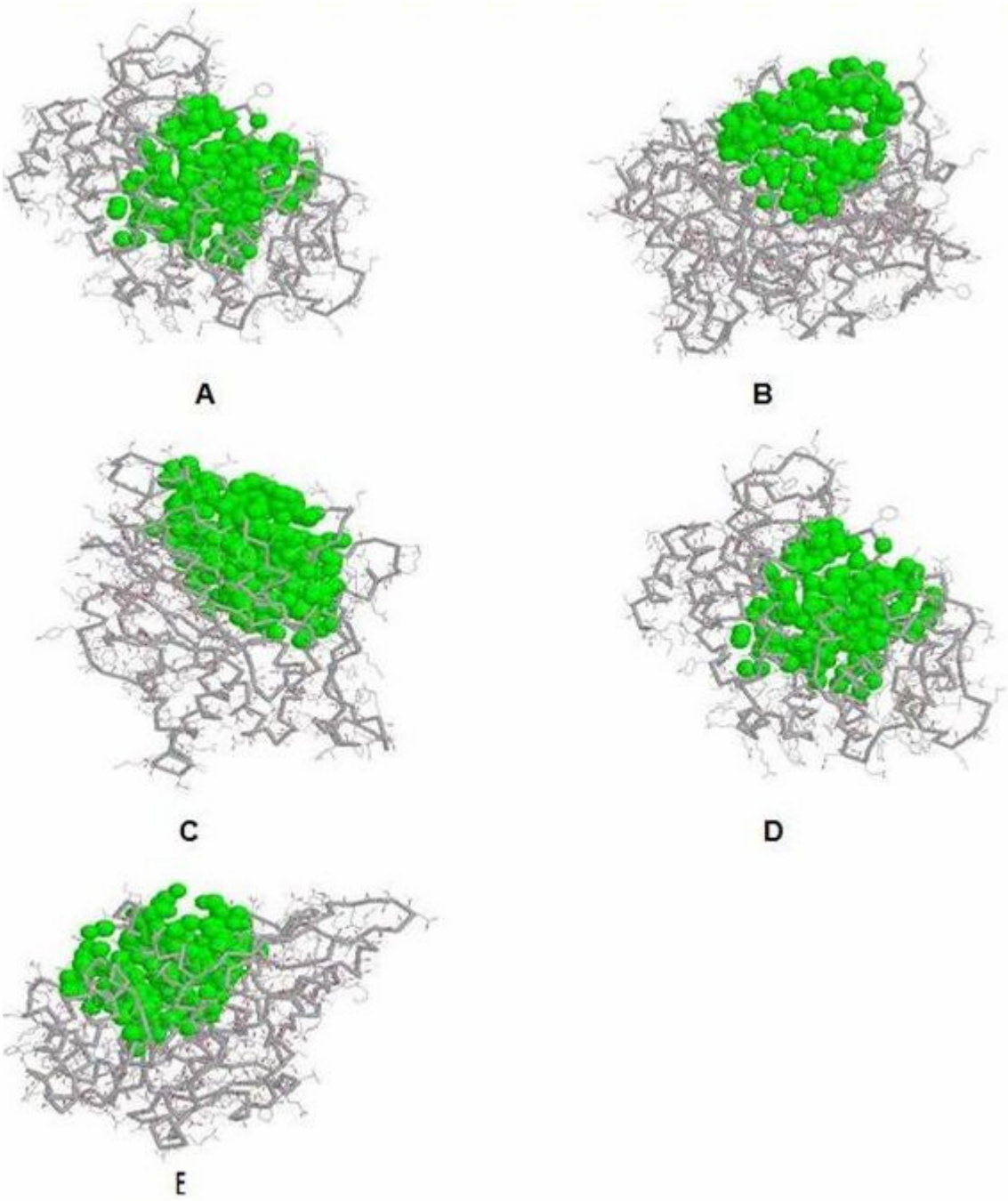


Figure 10. Potential substrate binding pocket for the modeled structures of (A) Phytoene synthase, (B) β-primeverosidase, (C) Flavanone-3-hydroxylase, (D) Dihydroflavonol 4-reductase, (E) Leucoanthocyanidin reductase

G263, I264, E267, A281, L282, G283, I284, N286, T289 and L292.

The binding site residues identified for β-prime-verosidase were: M1, M2, A3, K5, G6, S7. V8, V9, G11, V12, L13, A14, I15, V16, A17, Y18, A19, L20, V21, W276, M277, I278, P279, S281, N282, S283, K284, K287, D288, A290, Q291, L294, F307, R375, Y392, K394, K397, D398, V401, Y402, K404, E405, K406, N408, A453 and V455.

The binding site residues for F3H were: M1, A2, P3, E12, E13, K14, S15, L16,Q17, Q18, K19, F20, F93, F94, P98, K101, L102, F104, D105, M106, S107, G108, G109, F114, I115, V116, S117, S118, H119, L120, Q121, G122, E123, A124, V125, Q126, D127, W128, R129, E130, I131, V132, T133, Y134, F135, S145, R146, D195, K197, V199, F202, P204, T212, L213, L215, K216, R217, H218,

T219, D220, P221, G222, R239, D261, Y265, Q277, F292, N294, P297, Y324, R326, M328, S329, D331, I332, E333, L334, A335, K336, K338, D352, I353, E354, K355, A356, L358, E359, I360, K361, S362, T363, E365, I366, F367 and A368.

The binding site residues for DFR were: T19, G20, A22, G23, F24, I25, G26, T43, V44, R45, K51, K52, A71, D72, L73, N74, F79, V92, A93, T94, P95, M96, F98, E99, T134, S135, S136, A137, G138, N141, V142, Q143, E144, Q146, F150, F160, K164, K165, M166, T167, Y171, F172, K175, P198, T199, L200, V201, M207, T209, P212, S213, I215, T216, R223, E225, G226, H227, Y228,S229, I230, I231, K232, Q233, G234, Q235, I268, P285, E287, F288, K289, I291, L295 and V298.

The binding site residues for LAR were: V17, G18, A19, S20, G21, F22, I23, L41, V42, R43, V45, S47, T49, N50, L53, G63, V64, V65, A86, I87, G88, G89, A90, N91, I92, D94, S114, E115, F116, G117, H118, V120, M132, Y133, K136, C155, N156, S157, I158, S160, W161, P162, Y164, D165, T167, P169, S170, E171, Q180, Y182, I197, L251, A254, A255, N258, P261, R262, S263, V264, V265, A266, F268, T269, I272, F273, P337, I338, T339, M340, C341 and A342.

Conclusion

Tea is an extremely important crop because of its popularity as a beverage and as a source of beneficial secondary metabolites. However, due to the long periods involved in conventional crop breeding, it is not really an option to improve crop varieties. So, computational studies of key biosynthetic enzymes in tea can provide valuable insights into the mechanism of action of these enzymes aiding in the ultimate aim of improving tea quality. The study of Volatile Flavour Compounds (VFC) and the enzymes involved in their synthesis and release is required to improve the quality of tea. Similarly, to enhance the beneficial health properties of tea the study of flavonoids like catechins is essential.

ACKNOWLEDGEMENTS

Research in the laboratory of DS is supported by grants from Department of Biotechnology (DBT), Department of Science and Technology (DST) and Department of Information Technology (DIT), Government of India, New Delhi, India.

REFERENCES

Altschul SF, Gish W, Mille W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. J. Mol. Biol., 215(3): 403-410.

Binkowski TA, Naghibzadeh S, Liang J (2003). CASTp: Computed Atlas of Surface Topography of proteins. Nucleic Acids Res., 31(13): 3352-3355.

Borthakur D, Lu JL, Chen H, Lin C, Du YY, Liang YR (2008). Expression of phytoene synthase (psy) gene and its relation with accumulation of carotenoids in tea [Camellia sinensis (L) O Kuntze]. Afr. J. Biotechnol., 7(4): 434-438.

Choi JH, Jung HY, Kim HS, Cho HG (2000). PhyloDraw: a phylogenetic tree drawing system. Bioinformatics, 16(11): 1056-1058.

Eswar N, Eramian D, Webb B, Shen MY, Sali A (2008). Protein structure modeling with modeller. Methods Mol Biol 426145-159.

Eungwanichayapant PD, Popluechai S (2009). Accumulation of catechins in tea in relation to accumulation of mRNA from genes involved in catechin biosynthesis. Plant Physiol. Biochem., 47(2): 94-97.

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res., 31(13): 3784-3788.

Ijima Y, Ogawa K, Watanabe N, Usui T, Ohnishi-Kameyama M, Nagata T, Sakata K (1998). Characterization of beta-primeverosidase, being concerned with alcoholic aroma formation in tea leaves to be processed into black tea, and preliminary observations on its substrate specificity. J. Agric. Food Chem., 46(5): 1712-1718.

Luczaj W, Skrzydlewska E (2005). Antioxidative properties of black tea. Prev. Med., 40(6): 910-918.

Ma SJ, Mizutani M, Hiratake J, Hayashi K, Yagi K, Watanabe N, Sakata K (2001). Substrate specificity of beta-primeverosidase, a key enzyme in aroma formation during oolong tea and black tea manufacturing. Bioscience Biotechnology and Biochemistry 65(12): 2719-2729.

Ravichandran R (2002). Carotenoid composition, distribution and degradation to flavour volatiles during black tea manufacture and the effect of carotenoid supplementation on tea quality and aroma. Food Chem., 78(1): 23-28.

Ravichandran R, Parthiban R (1998). The impact of processing techniques on tea volatiles. Food Chem., 62(3): 347-353.

Rawat R, Gulati A (2008). Seasonal and clonal variations in some major glycosidic bound volatiles in Kangra tea (Camellia sinensis (L.) O. Kuntze). European Food Res. Technol., 226(6): 1241-1249.

Singh K, Kumar S, Yadav SK, Ahuja, P.S. (2009). Characterization of dihydroflavonol 4-reductase cDNA in tea [Camellia sinensis (L.) O. Kuntze]. Plant Biotechnol. Reports, 3(1): 95-101.

Singh K, Rani A, Kumar S, Sood P, Mahajan M, Yadav SK, Singh B, Ahuja PS (2008). An early gene of the flavonoid pathway, flavanone 3-hydroxylase, exhibits a positive relationship with the concentration of catechins in tea (Camellia sinensis). Tree Physiol., 28(9): 1349-1356.

Stutz A, Bairoch A, Estreicher A, Grp SP (2006). UniProtKB/Swiss-Prot: the protein sequence knowledgebase. Febs J., 27362-62.

Thompson JD, Higgins DG, Gibson TJ (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22(22): 4673-4680.

Wang HF, Provan CJ, Helliwell K (2000). Tea flavonoids: their functions, utilisation and analysis. Trends Food Sci. Technol., 11(4-5): 152-160.

Way TD, Lin HY, Hua KT, Lee JC, Li WH, Lee MR, Shuang CH, Lin JK (2009). Beneficial effects of different tea flowers against human breast cancer MCF-7 cells. Food Chem., 114(4): 1231-1236.