Full length Research Paper

Towards understanding the regulation of rubber biosynthesis: Insights into the initiator and elongator enzymes

Ankita Punetha, Jayaraman Muthukumaran, Anmol Jaywant Hemrom, Nagarajan Arumugam, Mannu Jayakanthan and Durai Sundar*

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

Accepted 7 January, 2010

Natural rubber is an important polymer produced by plants and made up of isoprene units derived from isopentenyl diphosphate (IPP). Although more than 2000 plant species are known to produce natural rubber, currently there are two important commercial sources, Hevea brasiliensis (the Brazilian rubber tree) and Parthenium argentatum Gray (guayule). Natural rubber biosynthesis requires three distinct biochemical processes such as (i) initiation, (ii) elongation and (iii) termination. Computational analyses of the enzymes farnesyl diphosphate (FPP) synthase in P. argentatum and cis-prenyl transferase (CPT) in H. brasiliensis that play a vital role in initiation and elongation stages for biosynthesis of cis-1,4-polyisoprene has been undertaken in this study. Amino acid sequence comparisons of FPP synthase and CPT to their identified similar sequences were carried out to understand the evolutionary relationship among different species. Homology modeling and binding pocket analysis aided in the understanding of structure-function relationship and enzyme-substrate interaction of FPP synthase and CPT. The structural templates farnesyl diphosphate synthase (Source: Gallus gallus) [PDB ID: 1UBX] for FPP synthase and undecaprenyl diphosphate synthase (Source: Micrococcus luteus B-P 26) [PDB ID: 1F75] for CPT were selected for homology modeling. The Ramachandran plots were developed for modeled structures of FPP synthase and CPT, which showed 95.9 and 92.6% of amino acid residues occurring in favored regions. These models were deposited into Protein Model Database [PMDB ID: PM0075218 and PM0075509]. The substrate and cofactor binding site residues of FPP synthase (R103, L149, A184, Y197, L211, H214, E223, T226, D332, K246, Y306, K313) and CPT (Y4, E7, R20, K21, G22, K154, K178, D193, E231, T232, R233) were identified by using binding pocket analysis, which is consistent with available X-ray crystal structure of both the templates. The computational analysis of initiation and elongation for cis-1,4-polyisoprene biosynthesis provided invaluable insights into the identification of putative initiation and elongation factors for FPP synthase and CPT.

Key words: Cis-prenyl transferase, deep view, errat, farnesyl diphosphate synthase, *Hevea brasiliensis*, MODELLER, *Parthenium argentatum*, protein model database.

INTRODUCTION

Natural rubber (polyisoprene) is an important polymer

produced by plants. It is an important raw material used in many products, including hundreds of medical devices. It is obtained from latex, an aqueous emulsion present in the laticiferous vessels (ducts) or parenchymal (single) cells of rubber-producing plants (Puskas et al., 2006). Primarily, due to its molecular structure and high molecular weight, rubber has high performance properties that cannot be easily mimicked by artificially produced polymers. Although more than 2000 plant species are

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

Abbreviations: APP, allylic diphosphate; CPT, cis-prenyl transferase; FPP, Farnesyl diphosphate; GPP, Geranyl diphosphate; IPP, Isopentenyl diphosphate; SG, sub group.

known to produce natural rubber, currently there are two important commercial sources, *Hevea brasiliensis* (the Brazilian rubber tree) and *Parthenium argentatum* Gray (guayule). Prenyl transferase or prenyl diphosphate synthase is an enzyme that catalyze the formation of linear prenyl diphosphate involved in the biosynthesis of various isoprenoid compounds including carotenoids, sterols, terpenes, glycosyl carrier lipids, prenyl proteins, quinones and natural rubber.

Precursors of biosynthesis of cis-1,4-polyisoprene

Cytosolic acetyl CoA is the preliminary substrate of the isoprenoid pathway and for the synthesis of natural rubber. Acetyl CoA is converted into isopentenyl diphosphate through a pathway called mevalonate pathway involving various intermediates such as acetoacetyl CoA, HMG CoA, mevalonate, mevalonate phosphate and mevalonate diphosphate. Isopentenyl diphosphate (IPP) is converted into dimethyl allyl diphosphate (DMAPP) catalyzed by isopentenyl isomerase. DMAPP primes the sequential head to tail condensations of isopentenyl diphosphate molecules by trans-prenyl transferase to form geranyl diphosphate (GPP), farnesyl diphosphate (FPP) and geranyl geranyl diphosphate (GGPP) (van Beilen and Poirier, 2007).

Biosynthesis of cis 1, 4 polyisoprene

Cis 1, 4-polyisoprene formation requires three distinct biochemical processes: (a) initiation, which requires an allylic diphosphate molecule (trans-prenyl transferases catalyze synthesis of allylic diphosphate); (b) elongation, cis-prenyl transferase catalyze cis 1, 4-polymerization of isoprene units from isopentenyl diphosphate; (c) termination, the release of the polymer from the rubber transferase (Cornish, 1993). The comparative sequence analysis is a preliminary step to understand the evolutionary relationship and gene structure. The knowledge of the three-dimensional structure of a protein would be an invaluable aid to understand the functional details of a particular protein. The three-dimensional structures of P. argentatum - FPP synthase and H. brasiliensis - CPT are not available so far. Therefore, in addition to a comparative sequence analysis of rubber biosynthetic genes in Parthenium and Hevea, we have attempted to predict the structural as well as active site information of FPP synthase and CPT, which are helpful to carry out further enzyme-substrate and cofactor interaction studies. The current study aims to get insights into rubber biosynthesis with the help of extended multiple sequence analysis and computational screening of the substrates of both FPP synthase and CPT with existing resources.

METHODS

Amino acid sequence comparison and phylogenetic tree construction

The protein sequence of FPP synthase and CPT were retrieved from Genpept database and their accession numbers were CAA57893 and BAB71777. The proteins with significant amino acid sequence similarity to FPP synthase or CPT were collected by a BLAST (Altschul et al., 1997) search of the non-redundant protein sequence database at the NCBI site (http://www.ncbi.nlm.nih.gov/ BLAST). A multiple sequence alignment of these sequences was produced using Multalign web server (Corpet, 1988).

Based on the alignment, an unrooted molecular phylogenetic tree was constructed by the neighbour joining method (Saitou and Nei, 1987). The statistical significance of the NJ tree topology was evaluated by a bootstrap analysis (Felsentein, 1985) with 1000 iterative tree constructions using the software package MEGA 3.1 (Kumar et al., 2008).

Secondary structure prediction

The secondary structures of FPP synthase and CPT were predicted from the amino acid sequence by the method of GOR IV, based on information-theoretical ideas that are essential for function prediction, protein classification and understanding the structural changes (Garnier et al., 1996).

Three-dimensional structure prediction

An effort was made to find a suitable structural template protein or homolog for the modeling of FPP synthase and CPT in *P. argentatum* and *H. brasiliensis* respectively. Initially, a structural template was obtained from BLASTP (Altschul et al., 1990) with the aid of Protein Data Bank (Berman et al., 2003). Alternatively, the threading method was also employed by 3D PSSM web server (Kelley et al., 2000) to recognize a fold, based on secondary structure of proteins. The amino acid sequence of these two key enzymes and their templates was aligned by using align-2D python script (sequence alignment module in MODELLER9V2) that required two essential parameters such as target sequence (Protein Information Resource format) and structural template (Protein Data Bank format).

The three dimensional structure of FPP synthase and CPT was predicted by using another python script file (model-single.py) in MODELLER9V2 (Eswar et al., 2008). The theoretical model of these two enzymes was subjected into Swiss-PDB Viewer (Kaplan and Littlejohn, 2001) for energy minimization with a harmonic constraint of 100 kJ/mol/Å2, applied for all protein atoms, using the steepest descent and conjugate gradient technique to correct the stereochemistry of the model. Computational analysis was carried out *in vacuo* with the GROMOS96 43B1 parameters set, without a reaction field.

Finally, the refined models were subjected to a series of tests for testing its internal stability and reliability. Backbone conformation was evaluated by examining the Psi/Phi Ramachandran plot obtained from RAMPAGE web server (Lovell et al., 2003). Errat web server (Colovos and Yeates, 1993) was used to investigate the statistics of non-bonded interactions between different atom types and plotting the graph. Finally, the evaluated reliable models were deposited into Protein Model Database [PMDB].

Binding pocket analysis

Binding sites and active sites of proteins are often associated with

	1	10	20	30	40	50	60	70	80	90	100	110	120	130
Bankhanium	I	четотремс	א ודמציים ו		nnconut		сси мост сч	TREVOLUX-CC		OCOL CUCUEI	IL OOVEL 11 DOTE	DECUTODCOOL	1000 0010	ытооност
Betanicia	несс	NSTOLKSKE	LATIDIEN LEVYDTEN	SILLIND FAFEFE	INDERMALL		GOM NDGLOT	TO STOLL K-00	EL SONETEL	CCOL CULTER	ALQUITELYLODII	IDE SITT NAVORAL	JERNER NYU MERI RAVU	HILINNOUL HTRONDCT
Hatricaria	naaa E	ISTVDE KSKE		SELTIND PREFE	INNSROUVE	KHLDYNVPF	raki NRGI SV	VDSYOLLK-00	FLITODETEL	SSAL GUCTEL	I DAYEL VI GOTH	IDESHTRRGOP	MERI PKVE	HTAANDGT
Hal i ant hue		HASON MENT	I HUYOTI K	SELLNDPREEF			REM NOGLES	UDSYNLLK-66	EL TODETEL	ACAL CULTER	JI OBYET VI DDTH	INCENTRACION	UFPL PYUS	HTRANDG
Panav		HSDI KTRF	I FVYSVI K	SELLINDPAFEET	nnspouve	RHL DY NVPE	GEN NRGI SV	TDSYKLLKEG	ELSODETEL	SSAL GUCTEL	JI ORYFI VI DDTH	Insshtrrgop	WERI PKVE	NTRYNDGT
Centel la		HSDI KTRF	I EVYSVI K	SOLL NO PAFEET	INNSPORVE	RHI DYNVPP	GEKLINRGL SV	TDSYKLLKEG	FLSDDETVI	SSAL GUCTEL	I DAYFI VI DDTH	IDESHTRREOPE	MERTPKVE	NTATNOGT
Halus		HADE KSKE	I KVYSVI K	SELLED PAEDED	NOSROWYF	RHL DY NVPE	SEKI NRGI SV	TOSYOL LOOGE	EL TENETEL	ASALGUCTE	I GAFFI VI DDTH	INGSHTRRGOP	WER PKVP	NTRANDGY
Hentha	MANLNE	AASOLRKTF	LGVYSTLK	SELLNDPREENT	DGSROHVE	RHLDYNVPO	GEKLINRGLSV	IDSYOLLKEG	OLTODEVEL	ASALGHCIE	LORYFLYLDDIN	DNSHTRRGOPO	HEKVPKVE	MININDGI
Hevea		HADLKSTE	LKVYSVLK	DELLEDPREENT	POSROAVE	RHLDYNVPC	GRUNRGLSV	IDSYKLLKEG	ELTEFETFL	ASALGHCTE	LOAYFLYLDDTH	DSSHTRRGOPC	HERVPKVE	LTAANDGT
Gossypium		HADLRSAF	LNVYSQLK	SELLOOPSFEL	DESROAVE	RHLDYNVPO	GGKLNRGLSV	IDSYRLLKOGK	ELTQUEIFL	TSALGHCIE	LORYFLYLDDI	DSSHTRRGOPO	HFRLPKVE	HERVNDGV
Ÿit.is		HSETKSKF	LEVYSYLK	SELLNDPAFEF1	DOSROHYE	RHLDYNYPE	GGKLNRGLSV	VDSYKLL-QGR	QUTODEVFL	ACYLGLCIEN	ILGAYFLYLDDT H	IDNSHTRRGOP(HFRYPKYE	INTR <mark>r</mark> ndgy
Humulus		HSGLRSKF	HEVYSILK	SELLNDPAFEF1	DOSROHVE	RHLDYNVPO	GGKLNRGLSV	IDSYOLLKGG	ELTEEEIFL	TSALGHCIE	ILGAYFLYLDDI N	IDNSVTRRGOPC	HFRVPKVO	LIAANDGI
Arabidopsis		HADLKSTF	LDVYSVLK	SIDLLQDPSFEF1	HESROALE	RHLDYNVRG	GGKLNRGLSV	VDSYKLLKOGO	DLTEKETFL	SCALGHCIE	ILQAYFLYLDDIN	IDNSVTRRGQPC	MFRKPKVO	HIRINDGI
Consensus			L.VY <mark>s</mark> .LK	S#L1#DPaFEft	d#SRQAv+	THEOY NVp6	GGKLNRGLSV	!IDSYqLLk.G	#LL#dEiFL	.sALGHCIE	ILQAYFLYLDDIN	ID.Shtrrgqpc	INFR. PKV6	#IRaNDG!
	131	140	150	160	1/0	180	190	200	210	220	230	240	250	260
Parthenium	TIPNHU		GKPYYVNI	IN ENEVEENTE	сконто т	TTI VGEKDI	SKYSI STHR	PTUOYKTRYYS	EYI PUACAL	HEGEDLEN	IVEVICINU VENGT	VEOVODOVI DO	FROPENTO	*TGTDTED
Artenicia	I I PNHU		GREAMAN	VIDLENE VEFOTE	ISCONTOL T	TTI VGEKDI	SKYSI STHR	PTUNYKTRYYS	EYI PVACAL	HEGEDLDKI	IVEVKNU VEHGT	YEOVODDYL DO	FGOPENTO	KTGTDIED
Matricaria	I I RNHV	PRTLIKNHER	GKPYYVDL	VIDI ENEVERATE	SCONTOL T	TTI VGEKDI	SKYSI SVHR	RTVOYKTRYYS	FYL PVAXAL	I NEGEDI OKI	IVEVKNVI VENGT	YEOVONOYI DO	FGTPEVTE	KTGTOTED
Hel tanthus	TLRNH	PRTLIKKHER	GKPYYVDI	VIDI ENEVEENTE	SCONTOL T	TTI VGEKDI	SKYSI STHR	RTVOYKTRYYS	FYL PVACAL	HEGEDI DNI	IVEVKHVI VEHGT	YEOVODOYI DO	FROPE VTC	KTOTOTEO
Рапаж	LLRNHI	PRILIKKHER	OKPYYYDL	DLFNEVEFOTA	SCONTOLI	TTLVGEKDL	SKYSLPIHR	RIVOYKTRYYS	FYLPVACAL	LHSGEDLEKI	ITHYKDILIENGT	YFOVODDYLDO	FGAPEVIC	KIGTDIED
Centella	LLRNHI	(PRILIKKHF <mark>r</mark>	ÓKPYYYDLI	LOLFNEVEFÖTF	ICGÓHIDLI	TTLYGEKDL	LSKYSLP <mark>T</mark> HR	RIVÓYKTRYYS	FYLPVACAL	LINAGEDLEKI	ITNYKDILIENGT	YFOYODYLDO	FGAPE VIC	KIGTDIED
Malus	VLRNHI	IPRILIRKYFR	EKPYYVDLI	LIDLENEVEFOTE	SGOMIDLI	TT <mark>IE</mark> GEKDL	lskysls <mark>t</mark> hr	RIVQYKTRYYS	FYLSVACAL	LINSGEELEKI	IIDVKHILVENGI	YFQVQDDYLDO	FGOPETIC	KIGTDIED
Hent-ha	ILRNHI	(PRILIKKHF <mark>r</mark>	SKPYYYDLI	LIDLFNEVEFQTF	ISGQHIDLI	TTTEGEKDL	LSKYSLPLHR	RIVQYKTRYYS	FYLPVACAL	LHRGENLEN	IPTYKDYLIDHGI	YFQVQDDYLDO	FGEPEKI O	KIGTDIED
Hevea	LLRNHI	(PRILIKKHF <mark>r</mark>	GKAYYYDL	LDLFNEVEFQTF	ISGQHIDLI	TTLEGEKDL	LSKYTLSLHR	RIVQYKTRYYS	FYLPVACAL	LINGENLON	IIVVKDILVQHG1	YFQVQDDYLDO	FGDPETIC	KIGTDIED
Gossypium	ILRNHI	ITRILIKHHFR	GKPYYVDLI	LIDLFNEVEFQTF	SGOMIDLI	TTLEGEKDL	LSKYSLOOHR	RIVQYKTRYYS	FYLPVACAL	.VICGENLONI	IIDYKHILYDHGI	YFQVQODYLDO	FGNPETI 0	KIGTDIEN
Vit.is	ILRNOI	IPRILIKHHFK	GKPYYYDL	DLFNEVEFQTF	ISGQHIDLI	. TTIEGEKDL	LSKYSLPLHR	RIVQYKTRYYS	FHLPVACAL	LHAGENLON	IT SYKDIL YQMGI	YFQVQODYLDO	FGDPQVIC	KIGTDIED
Humulus	LLRNHI	IPRILIKKHFK	GKSYYVDL	LOLFNEVEFOTF	SGQHIDLI	TTIEGEKDL	LSKYSIPLHH	RIVQYKTRYYS	FYLPVACAL	VHRGENLON	IVDVKHVLIENGI	YFQVQODYLDO	FGHPDVIC	KIGTDIED
Hrabidopsis	LLKNH		ENPTYPE	VULFNEVEFUTF	ICGUNIDEI	I IFUGEKUL	LSKTSLUTHK	RIVETRINTS	FTEPWHCHE	LANGENLEN	IT DYK TYLYDHG.	TEQYQUUTLU	THUPETLE	KTRIDTED
Consensus	LINNHE	PKILIKKHEP	SKDITANT.	LUCHNEVERUTE	SEAUTION	. I I LEGEKUL	LSKTSL.1HK	RTABLELINE	FTEPVHCHE	.Inage#L#ni	1YKN (L) #NG1	TEQYQUUTLU	3F8*L#A16	KTRIDTED
	261	270	280	290	300	310	320	330	340	349				
Parchenium	FRUSHL	YYKHLELHN	EEUKRYLH	ENTERNUPSPYF	INVINELTINI	LNLQGYFEL	UTERISTAL	TISTEBHPSKH	VUHYERSPE	ALTERNO				
HFCEN181a	FREEME	VYKHILELHN		ENTOKK UPHSYN	INVINE VITU		UTERISTAAL	TTCTCW DCM	VORVERSEE	CHINKUK				
Hatricaria Naliantkan	FINCORE FR-CUE	VUNDE CLON	CCUNNTLA	CHIUNNUL DECUC	INYNEL I DI WIWEI WIT		UI CRI ƏLINAL	TTCTCCUDCK0	VOOUL VCEL					
Param	EVECUL	VUVOI EL CN	CENNEL U	CHIUN-FROMF CNYCZNOPOCUC	INYNEL I NI IVWEL VNT		UT COTOTINAL EVECKOVINI	TRETEORIESNO	UNDUL POEL	CUTYUPOU				
Centel la	FKCSUL	VVKALEL CN	FEORKELH	ENYGKEDPORVE	KTKELYKI	I KLODVERE	FYESKSYEKI	TRETERHPHOS	VOAVI KSEL	GKTYKROK				
Halus	EVECUL	VUVOI EL CN	ECONNYL U		INTRACT NO	I DTEGNEOF	INCONCYPTI	TCUTECUDOVO	UNCUL MORT	CALANDOR				
Ment ha	FKCSHL	VVKAL FLON	FEOKKTLE	ENYGKENPADVE	KTKALYNN	TNI OGHFAT	FESKSYEKT	TSSTERHPSKS	VORVERSEL	GKTYKROK				
Hessea	FRESH	VVKALET CN	FERKKVLY	ENYGKBOPBSVE	IKVKVI YNF	I KLOGVETE	EVENESYKKI	VISTERHPSKP	VORVERSEL	RKTYKROK				
Gossupi un	FKCSHL	VYKAL FECK	EEHNKYLY	ENTGETREANVE	IKYKALYNE	LNLKGVFEL	ITESKSTERL	VTSTERHPSKP	V08VLKSEL	GKTYKROK				
Vitis	FKCSHL	IVKALETCH	EEQKKTLY	ENYGKROPANYE	KYKALYKO	LOLOGYFLE	PESKSYETL	VSSTEAHPSKA	VOBVLKSFL	GKTYKROK				
Hunglus	FKCSHL	VYKALETAT	EEQKKHLFI	EHYGKGDERSVI	KYKELYKA	LOLEGYFAC	DYENASYOKL	IKSIEAHPKEE	VORVLKSFL	RKIYKROK				
Arabidopsis	FKCSHL	VVKALERCS	EEQTKILY	ENYGKREPSNVF	KYKALY KE	LOLEGAFHE	EYEKESYEKL	TKLIERHQSKR	TORVLKSFL	RKIYKROK				
Consensus	FKCSHL	VVKALE1cn	EEQKK L	EnYGK. dPa. VF	K!K.LY.	L.L#gvF.F	WEs SY kL	sIEaHpska	IQRVLKSFL	RIYKROK				

Figure 1. A multiple alignment of farnesyl diphosphate synthase with its homologs from various organisms. Multalign program was used for alignment. Conserved region color code: High (red in color), low (blue) and neutral (black).

structural pockets and cavities. The accurate identification of substrate binding sites in enzyme structures can be valuable in determining their substrate or cofactor interaction. CASTp server (Binkowski et al., 2003) was used to identify the binding pocket or potential substrate-binding site of the predicted models. 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 of proteins.

RESULTS AND DISCUSSION

The number of amino acids present in two enzymes FPP synthase and CPT were 342 and 284 respectively and their molecular weight was 39385.2 and 32786 Da calculated by ProtParam web server (http://www.expasy.org/tools). Closely related sequences or homologs were identified by using BLASTP against non-redundant protein sequence database (Additional File 1 and 2).

Multiple sequence alignment, conserved sequences and the phylogenetic tree of FPP synthase and CPT

Numerous consensus regions and potential conserved amino acid residues were identified from this study for the two enzymes FPP synthase and CPT (Figures 1 and 2). The consensus sequences are essential for binding to its specific substrate or cofactor (Fujihashi et al., 2001), which is similar in all the related protein sequences. Koyama et al. (1993) had earlier reported the conserved amino acid residues from the set of isopentenyl diphosphate synthases (Koyama et al., 1993). The results of our comparative sequence alignment showed that there are four significant sites present in the two enzymes, namely, conserved, variable, singleton and parsimony and their statistics were 42, 308, 226, 80 for CPT and 97, 148, 59, 88 for FPP synthase. The multiple sequence alignment results showed that, an elongating enzyme has fewer conserved sites and more variable sites with their homologs compared with initiating enzyme. The overall



Figure 2. A multiple alignment of cis-prenyl transferase with its homologs from various organisms. Gaps for insertions and/or deletions are indicated by hyphens.

distance average of two enzymes was 0.88 and 0.17. Two major groups were obtained from the phylogenetic analysis of FPP synthase (Figure 3). Group 1 has two sub-groups namely SG1 and SG2. SG1 contains two clades, taxa Artemisia tridentata and Artemisia annua occur in clade A and clade B consisting of P. argentatum and H. annus showing 94% identity. Matricaria recutita exist as mono taxa, but it is closer to clade 1. SG2 contains only one clade consisting of Panax ginseng and Panax notoginseng. Centella asiatica exist as mono taxa, and it is nearer to the same clade. Group 2 did not contain any sub groups and had two clades. It consisted of Mentha x piperita, and Vitis vinifera occurring in first clade and Hevea brasiliensis and Gossypium arboreum occurring in second clade respectively. The most primitive branch was a functional segregation of FPP synthase of Arabidopsis thaliana from the remaining Group 1 and 2 of FPP synthases.

An initiating enzyme, FPP synthase was found in second clade of SG1, which is closely related with

Helianthus annus and its alignment score and pairwise distance was 93 and 0.06 respectively (Additional File 1). The other probable similar organisms were *Artemisia tridentata*, *Artemisia annua* and *Matricaria recutita*, which were found in neighboring clade.

Two Groups were obtained from molecular phylogeny of CPT (Figure 4), and it contains sub groups and clades. The first group contains only one sub group and a clade. The SG of Group 1 has three clades namely clade A (Mus musculus, Rattus norvegicus), clade B (Aedes quinquefasciatus) aegypti. Culex and clade С (Neurospora cressa, Saccharomyces cerevisae). Homo sapiens and Danio rerio exist as mono taxa, and it is closer to clade A. As seen in Additional File 2, the alignment score of Arabidopsis thaliana was 43 revealing that it is more similar to Hevea brasiliensis, while it is found in a separate clade of Group 1 with Hevea brasiliensis (Kharel and Koyama, 2003). Group 2 did not contain any sub groups, which has only one clade (Thermoanaerobacter pseudethanolicus, Thermoanaerobacter



Figure 3. An unrooted phylogenetic tree of farnesyl diphosphate synthase with their homologs. Tree was constructed by the neighbor-joining method. Topology was also evaluated by bootstrap analysis (MEGA 3.4.1). The numerical values in the tree represent bootstrap results.



Figure 4. An unrooted phylogenetic tree of cis-prenyl transferase with their homologs.

tengcongensis MB4). Escherichia coli str. K-12, Micrococcus luteus B-P 26 and Clostridium thermocellum are mono taxa, which is present in the same clade. The bootstrap value of *Hevea brasiliensis* and *Arabidopsis* thaliana was 100 and other probable similar organisms were *Mus musculus*, *Rattus norvegicus*, *Aedes aegypti*, *Culex quinquefasciatus*, *Neurospora crassa* and *Saccharomyces cerevisae*, which were found in clade A, B and C respectively.

The Pairwise distance of CPT from *H. brasiliensis* and FPP synthase from *P. argentatum* with other species are represented in Additional Files 1 and 2. The Pairwise distance between *Arabidopsis thaliana* and *Hevea brasiliensis* was 0.780, which was very low in comparison to others indicating that less divergence has occurred.

Secondary structure prediction

Comparative sequence analysis of FPP synthase and CPT can lead to identify the evolutionary relationship alone. In order to find out the structural class of these two enzymes, proclass web server (http://www.imtech.res.in/ raghava/proclass/) was used. The result of this tool is based on a statistical approach, which reveals that FPP synthase and CPT belong to all α , or α/β class. We have adopted Garnier-Osguthorpe-Robson also (GOR) approach for this prediction. The percentage of alpha helix, extended strand and random coil were 54.86, 30.42 and 14.72 in CPT and 52.98, 24.91 and 22.11 in FPP synthase. From these results, it is confirmed that alpha helices was dominated in protein secondary structure of two enzymes.

Prediction of homology models

The protein sequence of FPP synthase was submitted into BLASTP (Altschul et al., 1990) and picked out its homologs with the help of Protein Data Bank (Berman et al., 2003). The template crystal structure of FPP synthase (PDB ID: 1UBX) from Gallus gallus showed 48% sequence identity with the given target sequence. A threading method was employed for predicting the threedimensional model of CPT because BLASTP provided the low sequence identity structural templates. The 3D-PSSM web server suggested that the high-resolution Xray crystallography structure of Micrococcus luteus B-P 26 - undecaprenyl diphosphate synthase (PDB ID: 1F75) could be a suitable template because of its secondary structural identity of 86%. It could be validated by comparing the composition of secondary structural elements present in template and CPT sequences. The percentage of alpha helices, beta pleated sheets and coils were similar in both the cases which infer the accuracy of the threading results. The next step of comparative modeling was to align the target sequence with template, which

was done by the align-2D python script (Sequence alignment module in MODELLER), used to identify the conserved regions or motifs (Figure 5). Once targettemplate alignment was completed, the three dimensional structure was predicted using the program MODELLER9V2, which produced several different conformations of each enzyme. In general, the best model could be the one, which has the lowest value of the MODELLER objective function, and hence the model of FPP synthase with 2011.1 and CPT with 1909.5 were selected for energy minimization and validation of models.

Several programs were employed to check the stereo chemical properties of the predicted homology models. Initially, the models were subjected to energy minimization in Swiss-PDB Viewer and the optimal conformers of the two enzymes were again processed into RAMPAGE and Errat web server for verification and reliability. The Ramachandran plot of FPP synthase and CPT showed that more than 90% of amino acid residues fall into the most favoured region for valid models (Lovell et al., 2003). The guality factor of any high-resolution structures should be around 90% or higher. Here, the over all quality factor of refined models were 92.526 and 90.060, which was predicted from Errat (Colovos and Yeates, 1993). The evaluated reliable models were deposited in to PMDB and are now publicly accessible. The three dimensional model structures of FPP synthase and CPT are shown in Figure 6.

Binding site analysis

The predicted conserved residues, secondary and threedimensional structure can serve as an important factor for locating putative binding sites for their substrate and cofactor of FPP synthase and CPT. CASTp server (Binkowski et al., 2003) was used to predict the binding pockets (pocket size: CPT – 4312.7, FPP synthase – 3183.3) (Figure 7). The substrate and cofactor binding site residues of FPP synthase (R103, L149, A184, Y197, L211, H214, E223, T226, D332, K246, Y306, K313) and CPT (Y4, E7, R20, K21, G22, K154, K178, D193, E231, T232, R233) were identified. These binding site residues were well correlated with the experimental results of crystal structure of undecaprenyl diphosphate and farnesyl diphosphate synthase from *Micrococcus luteus* B-P 26 (Fujihashi et al., 2001).

Conclusion

The study of plant rubber biosynthesis is an essential part for production of useful natural rubber. An attempt to produce synthetic rubber of good quality comparable to natural rubber is yet to be materialized. This study on enzymes involved in cis-1,4-polyisoprene biosynthesis A aln.pos 10 20 30 40 50 60 SPWVEREREEFVGFFPQIVRDLTEDGIGHPEVGDAVARLKEVLQYNAPGGKCNRGLTVVAAYRELSG 1UBX FPP MSTDIRSKFLOVYDTLKSELINDP&FEFD----DDSRQVIEKMLDYNVPGGKLNRGLSVIDSYQLLKconscvd aln.p 70 80 90 100 110 120 130 PGQKDAESLRCALAVGWCIELFQAASLVADDINDQSLTRRGQLCWYKKEGVGLDAINDSFLLESSVYR 1UBX GGKLTDDEIFHASALGNCVENLOAYFLVLDDINDESHTRRGOPCUFRLPKVGNIAANDGIILRNHVPR FPP. * * *** * ** ** ***** * ***** _consrvd * 150 aln.pos 140 160 170 180 190 2:00 VLKKYCRORPYYVHLLELFLOTAYOTELGONLDLITAPVSKVDLSHFSEERYKAIVKYKTAFYSFYLP 1UBX ILKKHFRGKPYYVDLVDLFNEVEFQTASGQNIDLITTLVGEKDLSKYSLSIHRRIVQYKTAYYSFYLP FPP _consrvd *** * **** * ** ** *** **** * *** * 210 220 230 240 250 2.60270 aln.pos 1UBX VAAAMYMVGIDSKEEHENAKAILLEMGEYFQIQDDYLDCFGDPALTGKVGTDIQDNKCSWLVVQCLQR VACALLMPGED-LEKHVEVKNVLVEMGTYPQVQDDYLDCFGAPEVIGKIGTDIEDFKCSULVVKALEL FPP. 280 290 300 310 320 330 3:40 aln.pos VTPEQRQLLEDNYGRKEPEKVAKVKELYEAVGNRAAFQQYEESSYRRLQELIEKHSNRLPKEIFLGLA 1UBX FPP ANEEQKKVLHENYGKKDPSPVAKVKELYNTLNLQGVFEDYENTSYKKLITSIEGHPSKAVQAVLKSFL consrvd ** * *** * * ******** * ** ** * ** * aln.pos QKIYKRQK 1UBX FPP GKIYRRQK consrvd *** *** aln.pos 10 20 30 40 В 50 60 1F75 NINAAQIPKHIALINDGNGRUAKQKKNPRIKGHYEGNQTVRKITRYASDLGVKYLTLYAFNYLMKLPG CPT ----NELVNGERPSVFRLLGKVMRKGLVSILTQGPI * ** consrvd aln.p 70 80 90 100 110 120 130 1F75 DFLNTFLPELIEKNVKVETIGFIDDLPDHTKKAVLEAKEKTKHNTGLTLVFALNVGGRKEIISAVQLI CPT $\texttt{PTHIAFILDGNRRFARKHELPEGGGHKAGFLALLNVLTYCYELGVKYATIYAFSIDNFRRKPHEVQYV$ consrvd t . ÷. また 160 170 aln.pos 140 150 180 190 200 1F75 AERYKSGEISLDEISETHFNEVLFTANMPDPELLIRTSGEERLSNFLINQCSYSEFVFIDEFWPDFNE CPT HDLHLE-KIEGHIHEESIINAYDICVRFVGNLKLLSEPVKTAADKIMRATANNSKFVLLLAVCVTSTD _consrvd * * * * * * ** 250 230 240 220 2 60 210 aln.pos 270 1F75 ESLAQCISIYQNR/QIPKHIAIIMDGNGRWAKQKKMPRIKGHVEGMQTVRKITRVASDLGVKVLTLVA CPT EIVHAVEESSKDK-LKSNEICNDGNGDCVIKIEEMEPYSEIKLVELERNTYINPYPDVLIRTSGETRL ± _consrvd t ± t. **t t** 290 300 310 280 320 330 340 aln.pos 1F75 FNYLHKLPGDFLNTFLPELIEKNVKVETIGFID DLPDHTKKAVLEAKEKTKHNTGLTLVFALNVGGRK CPT SNYLLVQTTNCILYSPHALVPEIGLRHVVVSVINFQRHYSYLEKHKEYLK------consrvd *** * 370 380 350 360 390 400 aln.pos 1F75 EIISAVQLIAERYKSGEISLDEISETHFNEYLFTANMPDPELLIRTSGEERLSNFLIWQCSYSEFVFI CPT consrvd aln.p 410 420 43.0 1F75 DEFUPDFNEESLAQCISIYONR CPT _____ consrvd

Figure 5. Pairwise sequence alignment. (A) A pairwise sequence alignment between farnesyl diphosphate synthase (1UBX) from *Gallus gallus* and farnesyl diphosphate synthase (FPP) from *Parthenium argentatum.* "*" indicates a conserved region (conserved) between two sequences and "-"indicates gaps. (B) A pairwise sequence alignment between undeca prenyl diphosphate synthase (1F75) from *Micrococcus lotus* B-P 26 and cis-prenyl transferase (CPT) from *Hevea brasiliensis*.



Figure 6. Three-dimensional structural representation (Pymol display) of farnesyl diphosphate (A) and cis-prenyl transferase (B).

Table 1. Closely related protein sequences, identity, sequence alignment score and distances of farnesyl diphosphate synthase (FPP) in *Parthenium argentatum* with their homologs are represented. Homologs, alignment score and distances were predicted from Protein BLAST (BLAST_P), clustal weight (ClustalW) and molecular evolutionary Genetics Analysis 3.1 (MEGA) respectively.

S. No	Accession number	Source	Protein name	Identity (%)	Alignment score	Distance (s)
1.	AAP74720	Artemisia tridentata	farnesyl diphosphate synthase	93	92	0.07
2.	AAD17204	Artemisia annua	farnesyl diphosphate synthase	93	92	0.07
3.	AAC78557	Helianthus annuus	farnesyl pyrophosphate synthase	94	93	0.06
4.	ABS11699	Matricaria recutita	farnesyl diphosphate synthase	92	92	0.07
5.	AAY87903	Panax ginseng	farnesyl diphosphate synthase	85	85	0.15
6.	AAY53905	Panax notoginseng	farnesyl pyrophosphate synthase	85	85	0.15
7.	AAV58896	Centella asiatica	farnesyl diphosphate synthase	82	82	0.18
8.	AAK63847	Malus x domestica	farnesyl diphosphate synthase	81	82	0.19
9.	AAM98379	Mentha x piperita	farnesyl diphosphate synthase	81	81	0.19
10.	AAB49290	Hevea brasiliensis	farnesyl diphosphate synthase	75	81	0.19
11.	AAK58594	Gossypium arboreum	farnesyl diphosphate synthase	80	80	0.20
12.	AAX76910	Humulus lupulus	farnesyl diphosphate synthase	79	80	0.21
13.	AAM08927	Vitis vinifera	farnesyl pyrophosphate synthase	82	80	0.22
14.	CAA72793	Arabidopsis thaliana	farnesyl pyrophosphate synthase	80	79	0.26

S. No	Accession number	Source	Protein name	Identity (%)	Alignment score	Distance(s)
1.	AAM65193	Arabidopsis thaliana	undecaprenyl diphosphate synthase	42	43	0.780
2.	XP_001657391	Aedes aegypti	dehydrodolichyl diphosphate synthase	34	29	0.979
3.	NP_080420	Mus musculus	dehydrodolichyl diphosphate synthase	36	32	0.968
4.	NP_995583	Homo sapiens	dehydrodolichyl diphosphate synthase isoform b	37	34	0.991
5.	NP_001011978	Rattus norvegicus	dehydrodolichyl diphosphate synthase	36	31	1.003
6.	XP_001843572	Culex quinquefasciatus	dehydrodolichyl diphosphate synthase	35	30	1.009
7.	NP_998352	Danio rerio	dehydrodolichyl diphosphate synthase	36	31	1.015
8.	YP_001037426	Clostridium thermocellum	undecaprenyl pyrophosphate synthase	34	32	1.039
9.	CAD21109	Neurospora crassa	cis prenyl transferase	35	30	0.979
10.	YP_001665210	Thermoanaerobacter pseudethanolicus	undecaprenyl diphosphate synthase	35	34	0.991
11.	NP_623022	Thermoanaerobacter tengcongensis	undecaprenyl pyrophosphate synthase	42	34	1.039
12.	BAA36577	Saccharomyces cerevisae	cis prenyl transferase	34	29	1.027
13.	AAC73285	Escherichia coli K12	undeca prenyl diphosphate synthase	36	24	1.323
14.	BAA31993	Micrococcus Luteus B-P 26	undeca prenyl diphosphate synthase	37	30	1.143

Table 2. Comparative sequence analysis results of cis-prenyl transferase in Hevea brasiliensis.

provided invaluable insights into the identification of putative initiation and elongation factors for FPP synthase and CPT. A further study of comparative sequence analysis could be extended to multiple species to derive the ancestral details of FPP synthase and CPT. The three-dimensional structure details of proteins are of major importance in providing insights into their molecular functions. Further analysis of threedimensional structure and binding pockets of FPP synthase and CPT will aid in protein-lipid and protein-ligand interaction studies and this approach will help in identifying possible factors essential for initiation and elongation of natural rubber using high throughput virtual screening technique. By adopting both the extended comparative sequence analysis and virtual screening approach, one would be able to achieve the above task significantly.

Competing interests

The authors declare that they have no competing

interests.

Authors' contributions

AP, AJ, NA, JM and DS designed the methods and experimental setup. AJ, NA and AP carried out the implementation of the various methods. JM, AP, MJ and DS interpreted the results and wrote the manuscript. All authors have read and approved the final manuscript.

ACKNOWLEDGEMENTS

Some of the described protocols were developed and implemented by JM, AJH, NA and MJ at the Centre for Bioinformatics, Pondicherry University, Pondicherry, India. This work was supported partly by a grant from the Industrial Research and Development (IRD) Unit of IIT Delhi to D.S. The authors acknowledge the Bioinformatics Facility at the DBT-funded Distributed Information Sub Centre Centre at IIT Delhi.

REFERENCES

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. J. Mol. Biol., 215(3): 403-410.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997). Gapped BLAST and PSI-BLAST: a new genera-tion of protein database search programs. Nucl. Acids Res., 25(17): 3389-3402.
- Berman H, Henrick K, Nakamura H (2003). Announcing the worldwide Protein Data Bank. Nat. Struct. Biol. 10(12): 980.
- Binkowski TA, Naghibzadeh S, Liang J (2003). CASTp: Computed Atlas of Surface Topography of proteins. Nucl. Acids Res., 31(13): 3352-3355.
- Colovos C, Yeates TO (1993). Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci., 2(9): 1511-1519.
- Cornish K (1993). The Separate Roles of Plant Cis and Trans Prenyl Transferases in Cis-1,4-Polyisoprene Biosynthesis. Eur. J. Biochem., 218(1): 267-271.
- Corpet F (1988). Multiple sequence alignment with hierarchical clustering. Nucl. Acids Res. 16(22): 10881-10890.
- Eswar N, Eramian D, Webb B, Shen MY, Sali A (2008). Protein structure modeling with MODELLER. Methods Mol. Biol., 426145-159.

- Felsentein J (1985). Confidence limits on phylogenies: an approach using the bootstrap. Evolution, 39: 783-791.
- Fujihashi M, Zhang YW, Higuchi Y, Li XY, Koyama T, Miki K (2001). Crystal structure of cis-prenyl chain elongating enzyme, undecaprenyl diphosphate synthase. Proc. Natl. Acad. Sci. U. S. A., 98(8): 4337-4342.
- Garnier J, Gibrat JF, Robson B (1996). GOR method for predicting protein secondary structure from amino acid sequence. Methods Enzymol., 266:540-553.
- Kaplan W, Littlejohn TG (2001). Swiss-PDB Viewer (Deep View). Brief Bioinform, 2(2): 195-197.
- Kelley LA, MacCallum RM, Sternberg MJ (2000). Enhanced genome annotation using structural profiles in the program 3D-PSSM. J. Mol. Biol., 299(2): 499-520.
- Kharel Y, Koyama T (2003). Molecular analysis of cis-prenyl chain elongating enzymes. Nat. Prod. Rep. 20(1): 111-118.
- Koyama T, Obata S, Osabe M, Takeshita A, Yokoyama K, Uchida M, Nishino T, Ogura K (1993). Thermostable farnesyl diphosphate synthase of *Bacillus stearothermophilus*: molecular cloning, sequence determination, overproduction, and purification. J. Biochem., 113(3): 355-363.

- Kumar S, Nei M, Dudley J, Tamura K (2008). MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief. Bioinform., 9(4): 299-306.
- Lovell SC, Davis IW, Arendall WB, 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC (2003). Structure validation by Calpha geometry: phi, psi and Cbeta deviation. Proteins, 50(3): 437-450.
- Puskas JE, Gautriaud E, Deffieux A, Kennedy JP (2006). Natural rubber biosynthesis - A living carbocationic polymerization? Prog. Polym. Sci., 31(6): 533-548.
- Saitou N, Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4(4): 406-425.
- van Beilen JB, Poirier Y (2007). Establishment of new crops for the production of natural rubber. Trends Biotechnol., 25(11): 522-529.