J Mol Model (2017) 23:202
DOI 10.1007/s00894-017-3374-0

ORIGINAL PAPER

An in-silico insight into the substrate binding characteristics of the active site of amorpha-4, 11-diene synthase, a key enzyme in artemisinin biosynthesis

Habib Eslami[1] · Seyed Kaveh Mohtashami[2] · Maryam Taghavi Basmanj[3] · Maryam Rahati[2] · Hamzeh Rahimi[3]

Received: 9 July 2015 / Accepted: 17 May 2017

Abstract The enzyme amorphadiene synthase (ADS) conducts the first committed step in the biosynthetic conversion of the substrate farnesyl pyrophosphate (FPP) to artemisinin, which is a highly effective natural product against multidrug-resistant strains of malaria. Due to the either low abundance or low turn-over rate of the enzyme, obtaining artemisinin from both natural and synthetic sources is costly and laborious. In this in silico study, we strived to elucidate the substrate binding site specificities of the ADS, with the rational that unraveling enzyme features paves the way for enzyme engineering to increase synthesis rate. A homology model of the ADS from *Artemisia annua* L. was constructed based on the available crystal structure of the 5-epiaristolochene synthase (TEAS) and further analyzed with molecular dynamic simulations to determine residues forming the substrate recognition pocket. We also investigated the structural aspects of Mg^{2+} binding. Results revealed DDYTD and NDLMT as metal-binding motifs in the putative active site gorge, which is composed of the D and H helixes and one loop region (aa519–532). Moreover, several representative residues including Tyr519, Asp444, Trp271, Asn443, Thr399, Arg262, Val292, Gly400 and Leu405, determine the FPP binding mode and its fate in terms of stereochemistry as well as the enzyme fidelity for the specific end product. These findings lead to inferences concerning key components of the ADS catalytic cavity, and provide evidence for the spatial localization of the FPP and Mg^{2+}. Such detailed understanding will probably help to design an improved enzyme.

Keywords Amorpha-4, 11-diene synthase · Substrate binding site · Artemisinin · Homology modeling · Molecular dynamics

Abbreviations

ADS	amorpha-4,11-diene synthase
FPP	farnesyl pyrophosphate
TEAS	5-epi-aristolochene synthase
TPS	terpene synthase
SQTPS	sesquiterpene synthase

Introduction

The enzyme amorpha-4, 11-diene synthase (ADS) from *Artemisia annua* L. belongs to the sesquiterpene synthase (SQTPS) family of enzymes and catalyzes cyclization of the substrate farnesyl pyrophosphate (FPP) to the amorpha-4, 11-diene (Fig. 1), the main precursor of antimalarial drug artemisinin [1, 2]. Structurally, artemisinin is categorized as one of the members of sesquiterpenoid lactones and has a wide array of biological functions, including, but not limited to, antibacterial, anticancer, antischistosomal and antiviral activity [3]. Currently, among different drugs, artemisinin is the only available one that is highly effective against multidrug-resistant strains of malaria, a life-threatening disease. In fact,

Electronic supplementary material The online version of this article (doi:10.1007/s00894-017-3374-0) contains supplementary material, which is available to authorized users.

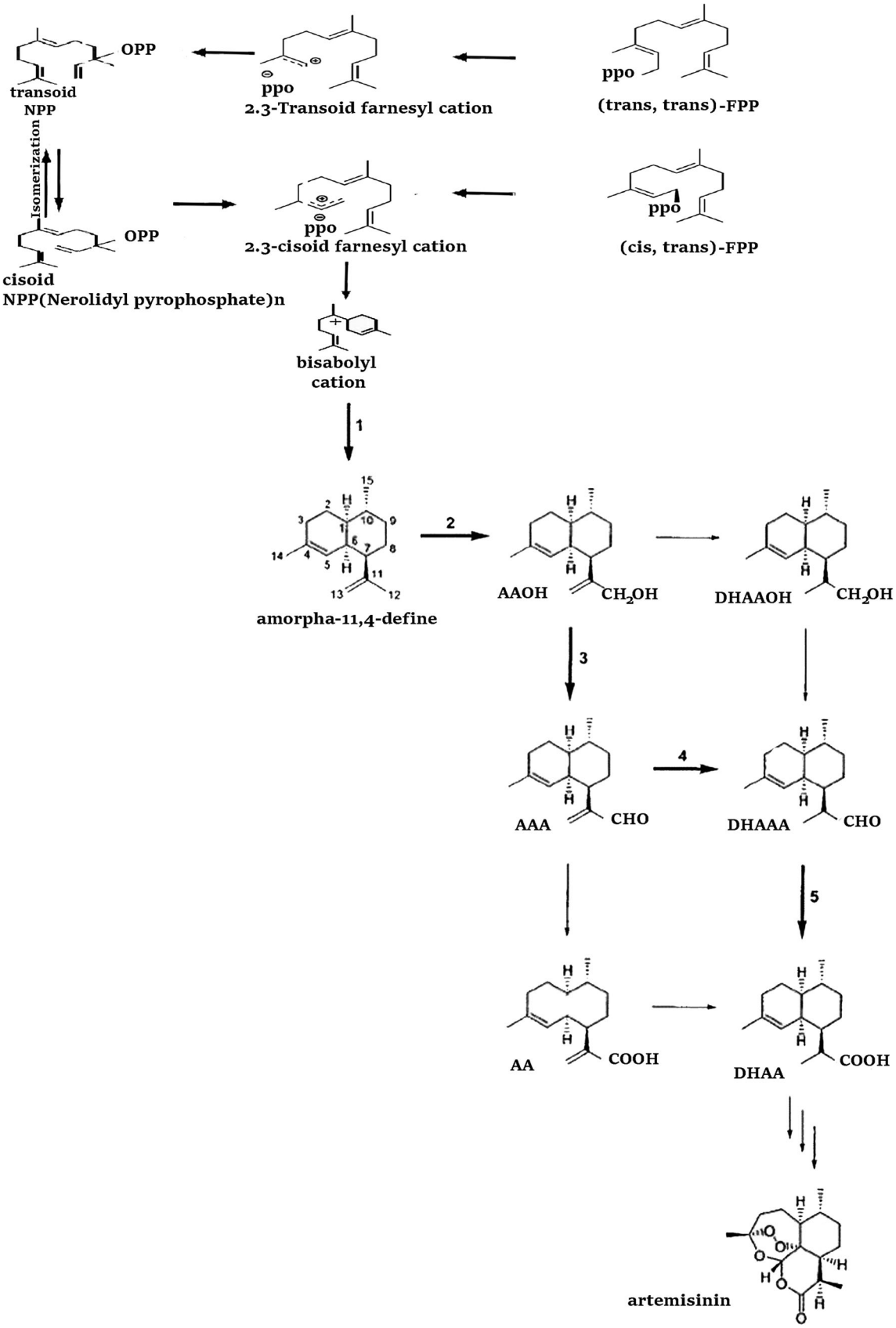
✉ Hamzeh Rahimi
buksanbio@gmail.com

[1] Department of Pharmacology, Molecular Medicine Research Center, Hormozgan Health Institute, School of Pharmacy, Hormozgan University of Medical Sciences, Bandar Abbas, Iran

[2] Crop Genetic Engineering Group- Biofuel Research Team (BRT), Agricultural Biotechnology Research Institute of Iran (ABRII), Karaj, Iran

[3] Biotechnology Research Center, Molecular Medicine Department, Pasteur Institute of Iran, Tehran, Iran

Published online: 15 June 2017


OPP
transoid
NPP
ppo
2.3-Transoid farnesyl cation
ppo
(trans, trans)-FPP
Isomerization
OPP
cisoid
NPP(Nerolidyl pyrophosphate)n
ppo
2.3-cisoid farnesyl cation
ppo
(cis, trans)-FPP
bisabolyl
cation
1
amorpha-11,4-define
2
AAOH
CH2OH
DHAAOH
CH2OH
3
4
AAA
CHO
DHAAA
CHO
5
AA
COOH
DHAA
COOH
artemisinin

◀ **Fig. 1** Schematic view of artemisinin synthesis. The committed step in this pathway is shown by number one and involves conversion of farnesyl pyrophosphate (FPP) to amorpha-4,11-diene by the enzyme amorpha-4,11-diene synthase (ADS). Other numbers indicate important steps in the biosynthesis of artemisinin but are slightly less significant than the first step. AAOH (artemisinic alcohol), DHAA (dihydroartemisin aldehyde), AAA (artemisinic aldehyde), DHAAA (dihydroartemisinic acid), AA (artemisinic acid), DHAA (dihydroartemisin acid). (McCarvey DJ, C.R., TerpenoidMetabolism.Plant Cell 1995.7: p. 1015–1026)

the human and economic burden of this disease is huge. For example, a recent report, released by the World Health Organization (WHO) estimated that approximately 800 thousands deaths worldwide were due to malaria during 2010. Also, the cost of defense against this prevalent parasitic infection was up to 1.8 billion US dollars [4]. Therefore, controlling the burden of malaria is vital both for developed nations and underdeveloped countries, where the disease frequently occurs.

In most malaria-endemic countries, there has been an inclination toward using the artemisinin and artemisinin-based combination therapies (ACTs) over the last decade [5]. However, unfortunately, the high expense of such therapies, particularly for economically disadvantaged peoples, is a serious problem [6]. Besides several contributing factors, the low production level of artemisinin by the biosynthetic pathway constitutes an essential cause of this problem [7, 8]. Indeed, the initial step of the artemisinin biosynthesis involves conversion of FPP to the amorpha-4, 11-diene. This step is catalyzed by the ADS which, like other committed enzymes, is present in very low concentrations in the cytoplasm and catalyzes reactions slowly as well [9, 10]. Hence, artemisinin is produced at low quantities, ranging from 0.01 to 0.8% of dry weight, in the leaves of the plant *A. annua* L. [11, 12]. Similarly, other sesquiterpenes tend to accumulate in low amounts, since the enzymes responsible for their synthesis are present in low abundance or have low turn-over rates [13].

Many research efforts have been devoted to improve the production level of artemisinin and related sesquiterpene lactones. For instance, Vincent et al. [7] incorporated a mevalonate pathway into *E. coli* and produced up to 24 μg of amorphadiene compound by over expression of the ADS. In another study, the increased expression of ADS in *S. cerevisiae* raised the production of artemisinic acid, a precursor of artemisinin, to a high level (up to 100 mgl^{-1}) [14]. Paddon et al. also reported the discovery of a dehydrogenase that provides an efficient biosynthetic route to produce commercially relevant concentrations of artemisinic acid [15]. As illustrated in the above-mentioned studies, modification of critical enzymes in the artemisinin biosynthetic pathway is an important strategy to elevate production rate in practice [7]. Nevertheless, the lack of understanding the substrate binding characteristics of the active site of ADS yet remains an obstacle for the biotechnological engineering of this enzyme in order to obtain high catalytic activity [11]. The three-dimensional (3D) structural properties of the ADS catalytic cavity at atomistic level is particularly helpful in this regard.

There was no experimental 3D structure data, neither X-ray crystallography nor NMR spectroscopy, available for ADS in the literature. However, the X-ray determined structure of *A. annua* α-bisabolol synthase (AaBOS) with high similarity to ADS is available. Furthermore, homology-modeling of ADS has been performed before. Salmon et al. predicted the 3D structure of ADS to define terpene cyclization mechanism [16, 17], and another theoretical model can be found in a study by Abdin and colleagues [18]. Despite these reports, the precise structural details of the substrate recognition pocket, the network of residues participating in the metal binding and mechanism of catalytic reaction by ADS are still not elucidated. Thus, the purpose of this study was, to computationally analyze substrate binding site determinants, as well as the spatial localization of Mg^{2+} in the active site of ADS. To this aim, we predicted a 3D structure of ADS by homology modeling and further investigated through molecular dynamic simulations.

Methods

Multiple sequence alignment (MSA) and homology modeling

MSA was utilized in the analysis of ADS and functionally similar proteins to identify the possible domains, distinctive motifs, and their degree of conservation in various species. The amino acid sequence of ADS was acquired from the Swiss-Prot database (ID: A2TEY7) [19] and subjected to the BLASTp search to find sequences with high similarity to the query sequence. The retrieved hits with the lowest E-values were selected as input for MSA with ClustalX 2.1 software [20]. This tool sets up with the Blosum62 similarity matrix and clustal algorithm. Secondary structure prediction and visualization of MSA results were performed using Jpred and Jalview software, respectively.

BLASTp tools [21] were used against PDB database to identify homologous protein structures to ADS. Accordingly, the crystal structure of 5-epi-aristolochene synthase (TEAS) from Nicotinia tabacum (PDB ID: 3LZ9 and 3 M00), which has an acceptable sequence similarity (58%) and identity (40%) to ADS, was selected as the suitable template. It should be noted that only the substrate-bound structures were included as potential templates. Therefore, the X-ray resolved structure of α-bisabolol synthase was excluded in spite of the high degree of sequence similarity to ADS. In addition, since the stereochemical attributes of a substrate may exert a significant effect on the binding mode, two crystal

structures of TEAS were chosen that have different isomers of FPP, i.e., cis and trans configuration. As a result, two parallel models were constructed based on two separate templates (3LZ9 and 3 M00). In order to build the models, sequence-structure alignment was carried out with ClustalW and the resulting alignment served as input for MODELLER9.10 software. The program was then executed to generate 10,000 models for the target sequence. Eventually, the best model (with the lowest DOPE score) was selected as recommended by Eswar et al. [22] and further validated using PROCHECK [23], ANOLEA and Verify3D [24].

Molecular dynamics (MD) simulation

The MD simulation was carried out using GROMACS v5.1 [25] with the AMBER force field amber99sb-ildn [26]. Protein molecules were put in the center of a rectangular simulation box with a 1 nm buffer region between protein atoms and box sides. The size of box was 6.72, 7.66 and 5.244 nm and TIP3P water model was used to solvate the protein [27]. The system was neutralized by adding 0.15 M sodium and chloride ions. Afterward, the energy of the system was minimized using steepest descent minimization with a tolerance of 10 kJ mol^{-1} nm^{-1}. Before the MD simulations, the system was equilibrated as follows. LINCS algorithm [28] was employed to constrain all bond lengths. Firstly, the minimized system was heated from 0 to 300 K over a period of 100 ps by V-rescale algorithm [29]. Secondly, the pressure coupling was achieved via the Parrinello–Rahman Barostat for 500 ps [30]. To calculate electrostatics interactions, the particle mesh Ewald (PME) algorithm [31] was used with a distance cut off of 1.4 nm. The PME grid spacing was set at 0.16 nm. Following NVT and NPT system equilibration, 60 nano second (ns) simulation was run with a time step of 2.0 fs, and the frames were saved every 10.0 ps for subsequent analysis.

Analysis of the MD simulations

Gromacs 5.1 and Pymol software were used for analyzing and visualizing the results of MD. Trajectory energy analysis (temperature, pressure, potential energy, and kinetic energy) and the structure convergence were performed via root mean square deviation (RMSD) against the starting structure and the average structure. Protein stability and flexibility were checked by root mean square fluctuation (RMSF). Secondary structure and solvent accessible surface area (SASA) were also investigated at the levels of atom and residue. The global backbone deviation was studied using RMSD, while the local change was examined by RMSF. The average structure was extracted from trajectory of 3000–10,000 frames (step) and energy was minimized by the steepest descent method. This structure was evaluated based on torsion angles analysis of the protein backbone (phi and psi) via PROCHECK software, Z-scor, verify3D, prosaII score, and Molprobity [23, 32]. Structural transitions were analyzed through calculating the RMSD matrix and interatomic distances were presented by distance RMSD plot.

Results

Domain and motif analysis

The structure shown in the figures is the average structure from 60 ns MD simulation. Protein domain analysis by InterProScan [33] revealed that ADS has two domains, including terpene synthase (TPS)-like domain at the N-terminal (20–193) and metal binding domain, which spans from the middle region until the C-terminal (223–491) (Fig. 2). Upon searching for motifs in the ADS sequence by means of the MyHits tool [34], we identified 'DDTYD' (299–303) as metal binding motif (DDXXD). Results of MSA demonstrated that critical motifs of DDXXD (299–303 for ADS), and metal binding site (D/N) XXX are well conserved among numerous SQTPS of different species (Fig. 3a and Fig. S1).

ADS structure

The crystal structure of TEAS was selected as the template for homology modeling (Fig. 4). This enzyme plays a role in cyclization reaction and thus, is functionally homologous to ADS. Results of protein modeling showed that the overall structure of ADS is composed of several alpha-helices, which are oriented in anti-parallel direction (Fig. 5a). The N-terminal domain contains small helices and spans from residues 30–215. Conversely, long helices are observable from the center to the C-terminal part of the ADS structure, i.e., residues 223–546. Therefore, ADS, like other members of TPS families, is also composed of long helices that are situated at the C-terminal end and connected by loops together. The three Mg^{2+} atoms were localized among these long helices (Fig. 5a). The first aspartate-rich motif (DDTYD, 299–303) and the alternative aspartate rich metal-binding motif (NDLMT, 443–447) are located at the end of the helix D and helix H, respectively (Fig. 5a). The C-terminal catalytic domain of ASD is classified as the α-helical class I TPS and composed of six α-helices, with a 38 Å deep active site cleft. The non-catalytic N-terminal domain of ASD adopts the alpha-helical conformation and belongs to class II TPSs. In this respect, it is similar to the N-terminal domain of TEAS [35].

Structural superposition of the predicted model of ADS and homologous proteins, such as TEAS (PDB ID: 1HX9) (rmsd; 0.302 Å), (+)-Delta-cadinene synthase (DCS) (PDB ID: 3G4D) (rmsd; 0.773 Å), limonene synthase (LS) (PDB ID:

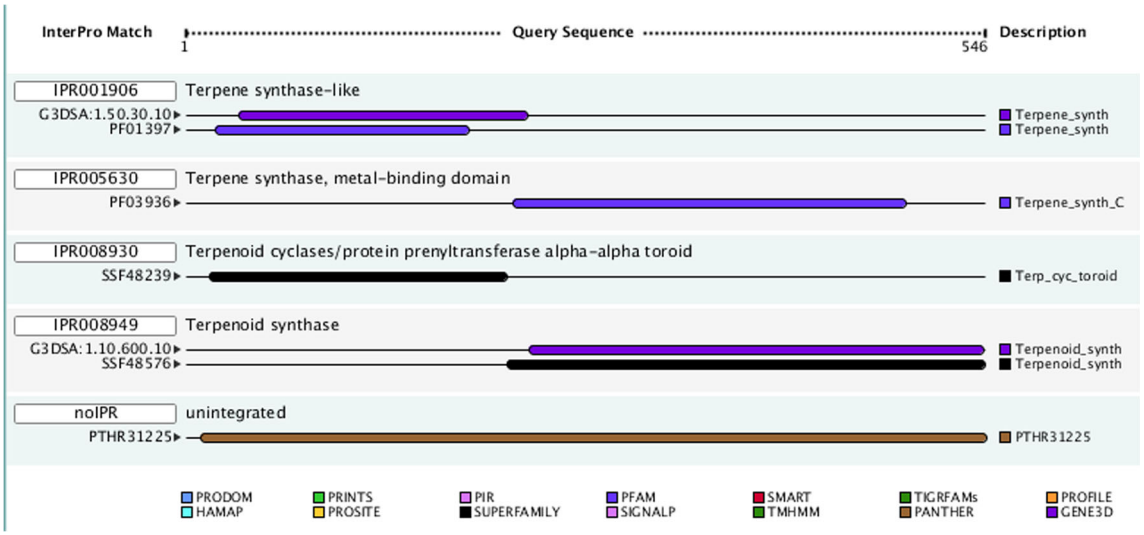


Fig. 2 Domain distribution in the full-length ADS. Domains predicted by Interproscan include TPS-like domain (20–193) and metal-binding domain (223–491). Two blue bars in the first row schematically show the inactive N-terminal domain while the row below indicates the active C-terminal domain of ADS. These results confirm the close match of ADS sequence with TPS superfamily. Interproscan tool, stores signatures of protein families and a sequence is searched against such signature database

2ONH) (rmsd;1.7 Å), *Abies grandis* α-bisabolene synthase (AgBIS) (PDB ID: 3SDR) (rmsd; 1.311 Å), abieta-diene synthase (ABS) (PDB ID: 3S9V) (rmsd;1.416 Å), showed that the overall structure and the catalytic cavity are conserved among TPS family (Fig. 5c). Verify3D profile was indicative of a good quality model with a positive mean force potential (Fig. S2a). The Ramachandran map of the model was excellent with more than 90% of the residues in the most-favored or allowed regions and few residues in the forbidden zones (Fig. S2a). The RMSD between the model and 3LZ9 crystal structure was 0.127 Å (Fig. S3).

Metal-binding

One of the properties of TPS family is presence of a metal in the active site. Metal-binding prediction by using the FINDSITE-metal software showed that ADS contains three Mg^{2+} ions. The geometry of three Mg^{2+} ions were analyzed by Chimera software and the results are summarized in Table 1. MSA demonstrated that Mg^{2+} binding residues are conserved in TPS family (Fig. S4). The motif DDTYD (299–303), which is located on helix D binds to Mg^{2+}I and Mg^{2+}II atoms. In addition, the alternative motif "NDLMT (443–447)" on helix H binds to Mg^{2+}III atom (Fig. 5b).

Insight into the active site of ADS

The active site of the ADS predicted structure was analyzed by the preset plug-in of Pymol software. The results indicated that the residues including Trp271, Tyr519, Asn443, Arg440, Glu451, Glu377, Thr399, Gly400, Arg262, Asp444, Ilu295, Val292, and Leu405 form the active site (Fig. 6 c, d). The structural superposition of TPS family showed that the predicted active site is highly conserved among the individuals of the family (Fig. 5b). The highly similar structures were comprehensively aligned together, and interestingly they possess similar amino acids in the active site (Table 2). In all structures, contiguous electron density stretches from the DDTYD motif through the diphosphate moiety into the NDLMT motif enshrouding the catalytically essential Mg^{2+} ions (Fig. 5a).

The binding pocket is composed mainly of the D and H helixes and one loop region (519 to 532) (Fig. 6a & b). Coordination of the loop and the helixes affects the binding of residues with the FPP and magnesium ions (Fig.6 c & d) and further changes of the allylic carbocation. The coordination of transoid and cisoid conformations of ligand (FPP) is illustrated in Fig. 6e. To gain deeper insight into the important structural features that dictate ligand selectivity, we proceeded to monitor binding of FPP along a molecular dynamics. Results revealed that ADS and FPP retained their stability throughout MD (Fig. 7a and b). The structural superposition of B-factor and the average structure from simulation indicated that the structure was stable during simulation (Fig. 7c and d). Interestingly, the active site is more stable than the N-terminal region. As depicted in Fig. 7c, the loop region (519–532) in the active site has some flexibility. This result was confirmed by calculation of the RMSF. However, NDLMT region has the lowest RMSF, and therefore, it is stable (Fig. 7d). Furthermore, it has been reported that the loop has the flexibility to make a lid by which the active site cleft is protected from the hydrous outside environment [36]. This specificity of the loop facilitates formation of a solvent inaccessible area, keeping away water from the allylic carbocation [37]. The lid causes the residues Tyr519 and Trp271 to come

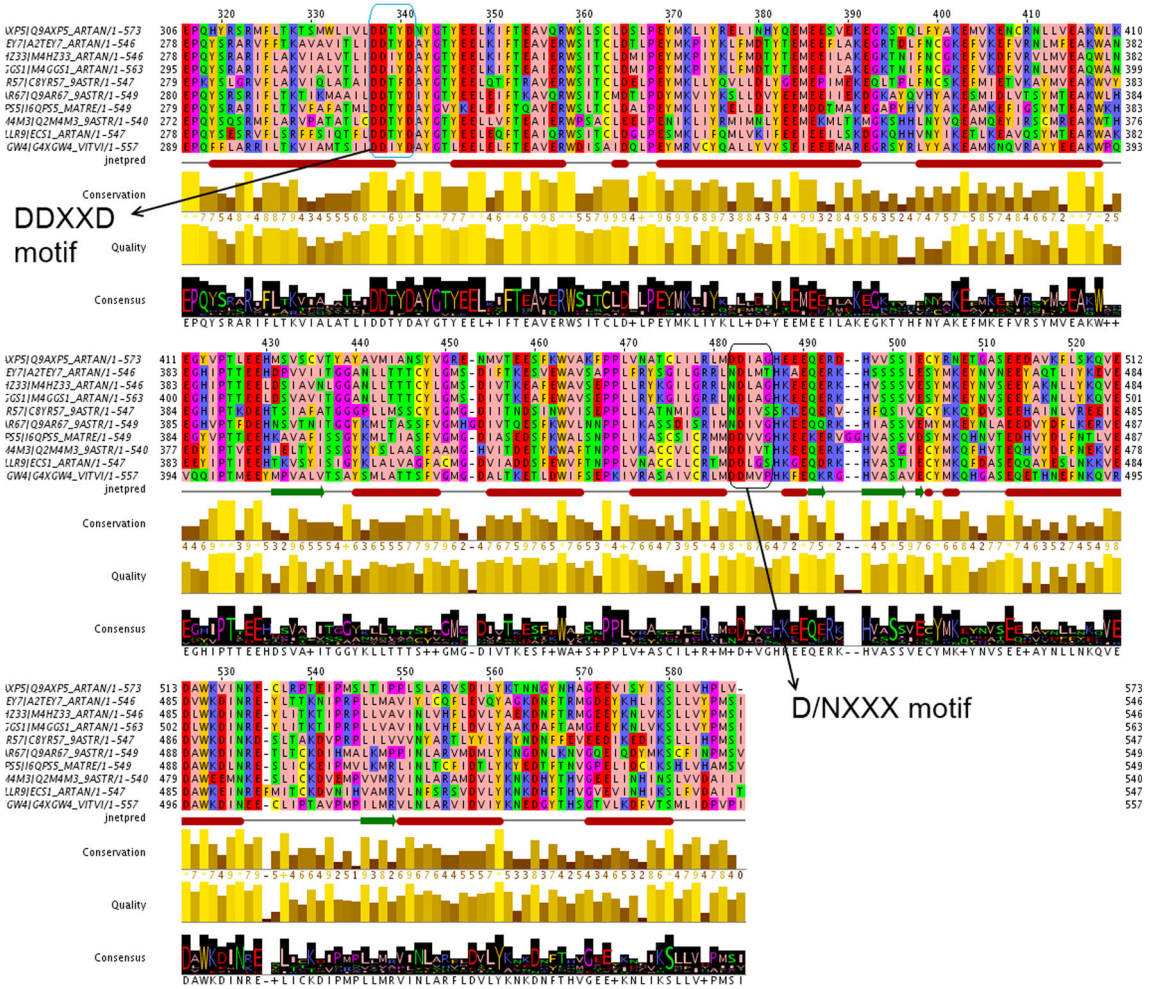


Fig. 3 Multiple sequence alignment of the C-terminal region of Terpene synthase. DDXXD and metal-binding motifs are conserved among ADS from different resources. The secondary structure was predicted by Jpred. Aligned sequences include Sesquiterpene cyclase (Q9AXP5, *Artemisia annua*), ADS (A2TEY7, *Artemisia annua*), Alpha-bisabolol synthase (M4HZ33, *Artemisia annua*), ADS (M4GGS1, *Artemisia annua*), Beta-caryophyllene synthase (C8YR57, *Mikania micrantha*), Germacrene A synthase (Q9AR67, *Solidago canadensis*), TPS5 (I6QPS5, *Matricaria recutita*), SQTPS (Q2M4M3, Ixeridium dentatum), Epi-cedrol synthase (Q9LLR9, *Artemisia annua*), and (E)-beta-caryophyllene synthase (G4XGW4, *Vitis vinifera*). Looking at the first stack of rows, the most conserved one is DDTYD motif (the row called "consensus"). Moreover, conservation is higher in helix regions than others as clear in the yellow "conservation" line. The most conserved amino acids are Glu (E), Asp (D), Arg (R). Their significance and role in the catalytic reaction is discussed in the text

close to each other in an opposite direction, making a hydrophobic box along with Leu405 and Val292 (Fig. 7e).

The binding dynamic of FPP was studied during MD and the distances between donors and acceptors for the hydrogen bonds were plotted for both cis (Fig. 8a) and trans-FPP model (Fig. 8b). The number of hydrogen bonds was higher for cis model (six, compared to three for trans). Considering non-bonded interactions such as *van der Waals* and electrostatics forces, the Lennard-Jones short-range (LJ-SR) and Coulombic short-range (Coul-SR) potential energy were calculated between ADS and FPP. As shown in Fig. 9a & b, the sum of Lennard-Jones and Coulombic energy was more negative for cis (−1900 kJ mol^{-1}) than trans (−1700 kJ mol^{-1}). From these results, it can be deduced that cisoid form has high binding strength in comparison with trans-isomer. Throughout the simulation, the conformation of ADS was stable. To detect the statistically independent conformational regions of ADS in the MD simulation trajectories, we analyzed the entire pairwise backbone RMSD matrix for the entire structure of both cis (Fig. 9c) and trans model (Fig. 9d) over the 60 ns MD simulations. The red represented low RMSD, whereas green represented high RMSD, and each red square along the diagonal represented an independent conformational region.

With regard to the lid covering the enzyme active site, the structures of two states "closed" and "open" were superimposed and depicted in Fig. 10a, in which the green color displays the open state while the blue one indicates the closed position of the lid in the protein structure. The two closed/open structures were obtained through analysis of the RMSD profile curve and RMSF curve. The higher peaks of

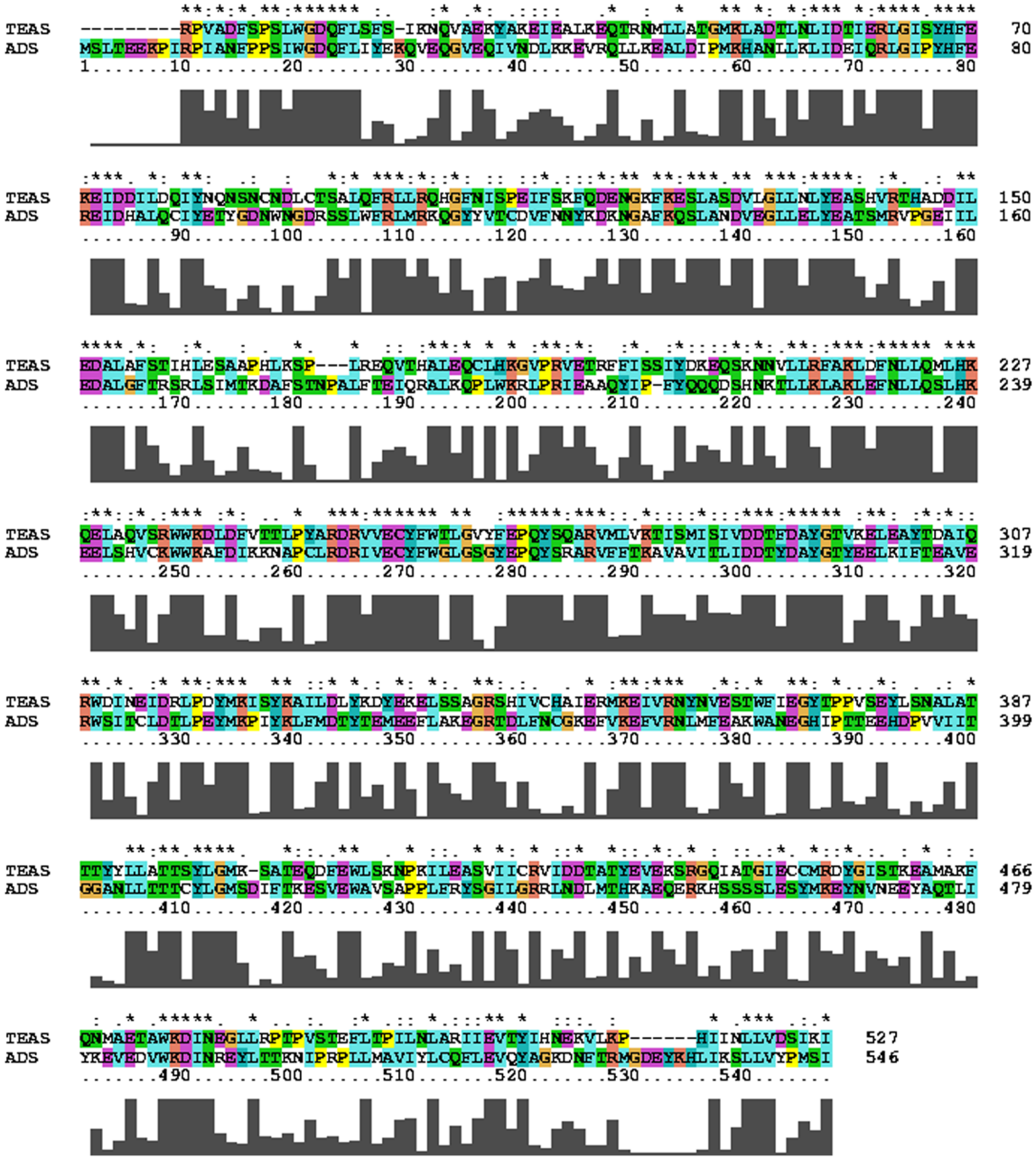


Fig. 4 Pairwise sequence alignment of ADS and TEAS. This analysis was done to confirm the suitable similarity between target and template sequences

fluctuations were observed in lid region and confirmed the movement of the lid. Then, pdb files of structures with lid in two positions (closed and opened) were superimposed (Fig. 10a). Analyses were based on one trajectory in one simulation. The distance between the lid and the active site (Fig. 10b), the distance between Tyr519 and Trp271 (Fig. 10d) and the solvent accessible surface area (SASA) of the active site (Fig. 10c) were calculated. Results indicated that lid region has flexibility and the flexibility of the loop influences the opening of the helixes and thus, the accessibility of solvent accessible area of the hydrophobic amino acids as well as hydrophobic FPP itself.

Interaction between magnesium ions and residues were also investigated. Analyzing the docking of trans form of FPP showed that Asn443, Glu451, Asp303, Asp299, Thr296, Arg262, Tyr519, Gly400, Leu405, Val292, and Trp271 interact with FPP. In the binding pocket, Asn443 and Glu451 were indirectly bound to pyrophosphate via interaction to MgIII. Additionally, Asp299 and Asp303 interact with MgII and Asp303 interacts with MgI (Fig. 6c). Two-dimensional representation of cis-FPP binding by LigPlot + showed the most favorable interactions, as follows: Asp303 and Asp299 interact with MgII, Asp299 and Glu377 interact with MgIII, and Glu451, Asp444, and Asn443 make bonds

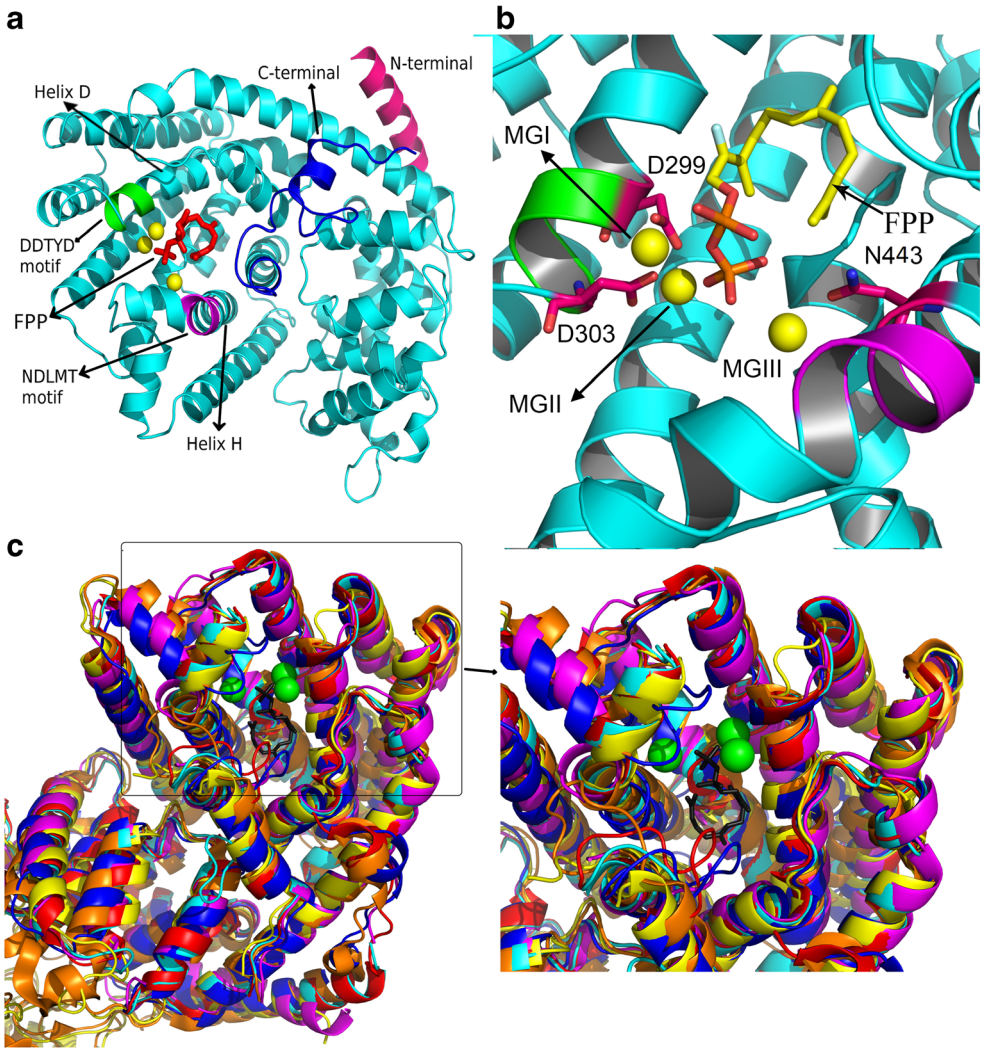


Fig. 5 Global fold and surface properties of the predicted structure of ADS rendered by PyMol. a) The overall structure contains so many alpha-helices that the small ones are seen in N-terminal but longer helices were arranged in C-terminal. Conserved motifs (DDTYD and NDLMT) are illustrated around the active site. DDTYD motif is located in the helix D and NDLMT on helix H. The Mg^{2+} and FPP have been colored in yellow and red, respectively. b) Mg^{2+} (shown as MG in the figure) geometry and binding. MGI & II were bound to D299 and D303, respectively, in DDTYD (299–303) motif. The third MG was bound to N443 in NDLMT (443–447) motif. c) Structural superposition of TPS family members including TEAS (red), DCS (magneta), LS (blue), 3SDR (orange), ABS (yellow) and ADS (cayn). Clearly, there is high conservation in the Mg^{2+} binding coordination and secondary structure among the selected family members of the TPSs

with MgI. Moreover, Arg440, Gly439, Gln518, Gly400, Tyr519, Ile295, and Thr399 bind to ADS directly through non-polar interaction (Fig. 6d). The average distance between three magnesium ions and the residues at the active site were plotted for cis (Fig. 10e) and trans-FPP model (Fig. 10f). The distances for all residues are roughly similar when compared one by one in two models.

Discussion

Although the experimental findings have shed light on many of the biochemical aspects of ADS, the most important enzyme in artemisinin biosynthesis [10], the structural details of the substrate-binding profile and catalytic reaction in the active site have remained elusive until now. Given this, we initially predicted ADS structure by homology modeling. Subsequently, MD simulation was implemented for model verification as well as to find out the key components of the ADS catalytic cavity. Our results explain the structural details of chemical reaction in ADS active site and the role of Mg^{2+} binding motifs in cyclization of the substrate FPP.

Regarding metal binding motif in the active site of ADS, it was found that NDLMT (443–447) motif has evolved as a replacement to 'NSE/DTE (Fig. 5). The results of MSA showed that this motif and DDTYD motif are conserved in TPS family (Fig. 3). In general, C-terminal of the class-I of the plant TPSs contain two metal-binding motifs. The first one is 'DDXXD' and the second one is 'NSE/DTE', which is known as (N, D) D (L, I, V) X (S, T) XXXE and specifically binds to Mg^{2+} ions [35]. Recently, studies reported that the DDXXD motif of the enzyme AaBOS, which has highly similar structure to ADS, is located in helix D and binds to metal ions [17]. Therefore, it seems that the divalent metal binding site in ADS is composed of the most common DDTYD that is situated on helix-D and the substituted motif of NDLMT on helix-H (Fig. 5a). Unlike the case of TEAS, there are two common expected aspartate rich motifs, including DDXXD [36] and

Table 1 Metal binding geometry in ADS active site. MG indicates Mg^{2+} with numbering and the amino acids bonded with MG are in three letter code (FPP: farnesyl pyrophosphate) under the coordinator column. The analysis was done with the chimera tool. The three letter code after the amino acid residue three letter codes and FPP indicates the position of oxygen atom on the molecule in metal bond with MG

Metal	Coordinator	Distance (A)	Distance RMSD (A)
MGI	Asp299 OD1	2.00	0
	FPP O2B	2.15	0.117
	Asp303 OD2	2.32	0.155
	Asp303 OD1	2.35	0.397
	Glu377 OE1	3.56	0.238
MGII	Asp303 OD2	2.14	0
	Asp299 OD2	2.26	0.098
	FPP O3A	2.79	0.082
	FPP O1A	3.45	0.332
	FPP O2B	3.69	0.702
	Arg262 NH2	3.72	0.775
MGIII	Glu451 OE2	2.08	0
	Asn443 OD1	2.65	0.015
	FPP O1B	3.16	0.244
	Thr447 OG1	3.54	0.266

NSE/DTE [38] that bind to Mg^{2+} ions. The coordination of two Mg^{2+} binding sites in both ADS and TEAS are in opposition to each other like those of FPP synthase (Fig. 5a) [36]. Replacement in the Mg^{2+} binding sites is not unique to ADS, since another member of the TPS family, DCS, has a replacement of DDVAE (451–455) [39]. Another case for substitution in aspartate rich motifs was observed in the (+)-Germacrene-D synthase (GDS) from *Solidago Canadensis* in initial D to N residue causing the first DDXXD motif to take the form of NDTYD [40].

It has been proposed that ADS catalysis proceeds through the bisabolyl intermediate (Fig. 1). Interestingly, NDXXX site is highly conserved among the enzymes whose intermediate carbocation is bisabolyl. These enzymes were diverged from an ancestral gene prior to angiosperm speciation and have conserved motifs and functional domains. This conservation shows the importance of this domain in cell function and maintenance [40]. A site-directed mutagenesis study of the DCS indicated that mutations of the first and last Asp of/in the DDTYD motif and Glu in the DDVAE motif to Ala adversely affected the activity of enzyme. Consequently, the Mg^{2+}-binding residues play a critical role in catalytic and specificity of the enzyme [39]. In addition, by using site-directed mutagenesis, the role of the second Asp in the enzymatic activity has been studied for the following enzymes: aristolchene synthase [36], yeast farnesyl-diphosphate synthase [41], trichodiene synthase [41, 42], and pentalenene synthase [43, 44]. In this study of ADS structure, Asn443 in NDLMT motif seems to be critical. The reason is that this residue directly bonds with MgIII and contributes to the stability of metal tirade in the enzyme active site (Fig. 6c). As shown in Fig. 7e, the flexibility of the residues (aa32–253) of ADS in the loop region, but in the case of TEAS residues 36–255 [36], causes the Arg262 to come near the C1 and C2 of the FPP to form hydrogen bonds with them. The Asp residue along with the three Mg^{2+} creates a positively charged area driving the PPi to be removed from the farnesyl chain.

The interesting point here is that the ADS structure contains only one arginine residue, the Arg262 (Fig. 6c). On the contrary, in TEAS, there are two arginine (Arg264 and Arg266) in one side of the active site cleft, which forms a bond with hydroxyl group of C1 and NH2 terminal section of FPP, respectively [36]. This binding mode of Arg with FPP is crucial in cis catalytic reaction in TEAS [45]. In this case, the farnesyl tail of FPP goes deep into the conserved hydrophobic area of the active site cleft making a region with dense hydrogen bonding. In this study, Arg262 was bound to C1 and O1A of trans-FPP (Fig. 6c).

Even though SQTPS family of enzyme contains widely varied members, each member produces the same C15H24 skeleton as the general form of product. Therefore, it is their unique form of catalysis which gives them the ability to use the chemical property of the substrate, FPP, to make products with diverse specificity and efficiency. In fact, sharing minimum class I topology has enabled them to become diversified. Diversity in cyclization chemistry and TPS product array occurred by variation of amino acids composing metal binding motifs [38, 40] such as DDVAE (451–455) replacement in DCS [39]. Change of initial D to N residue occurring in the first Asp rich motif, NDTYD, in the GDS may be another example of amino acid variation in the metal binding motifs [46]. The results of site-directed mutagenesis showed that substitution of threonine for serine could increase the release of product and accelerate the enzyme activity. This observation confirmed the critical role of the plasticity of active site residue in the fate of final product [16, 17]. The substitution in Mg^{2+} binding site is a kind of evolutionary process to create variation in cyclization chemistry. Such variation contributes to the increment of the community fitness, e.g., as a response to biotic as well as abiotic stresses [47], and may also constitute an explanation for the mutation in the second metal-binding site. Previous studies demonstrated that substitutions in MgIII binding site (NDLMT, aa443–447) increase the affinity of binding with Mg^{2+} [36].

The second major issue addressed in the present work was the structural details of chemical reaction in ADS active site. Residues such as Leu445, Thr447 in ADS; Cys440, Thr519 in TEAS; Trp753 and Phe835 in taxadiene synthase are indicative of the size of active site in TPS family of enzymes. The more extended size of active site exhibits wider spectrum of carbocation intermediates [36]. There is a common platform

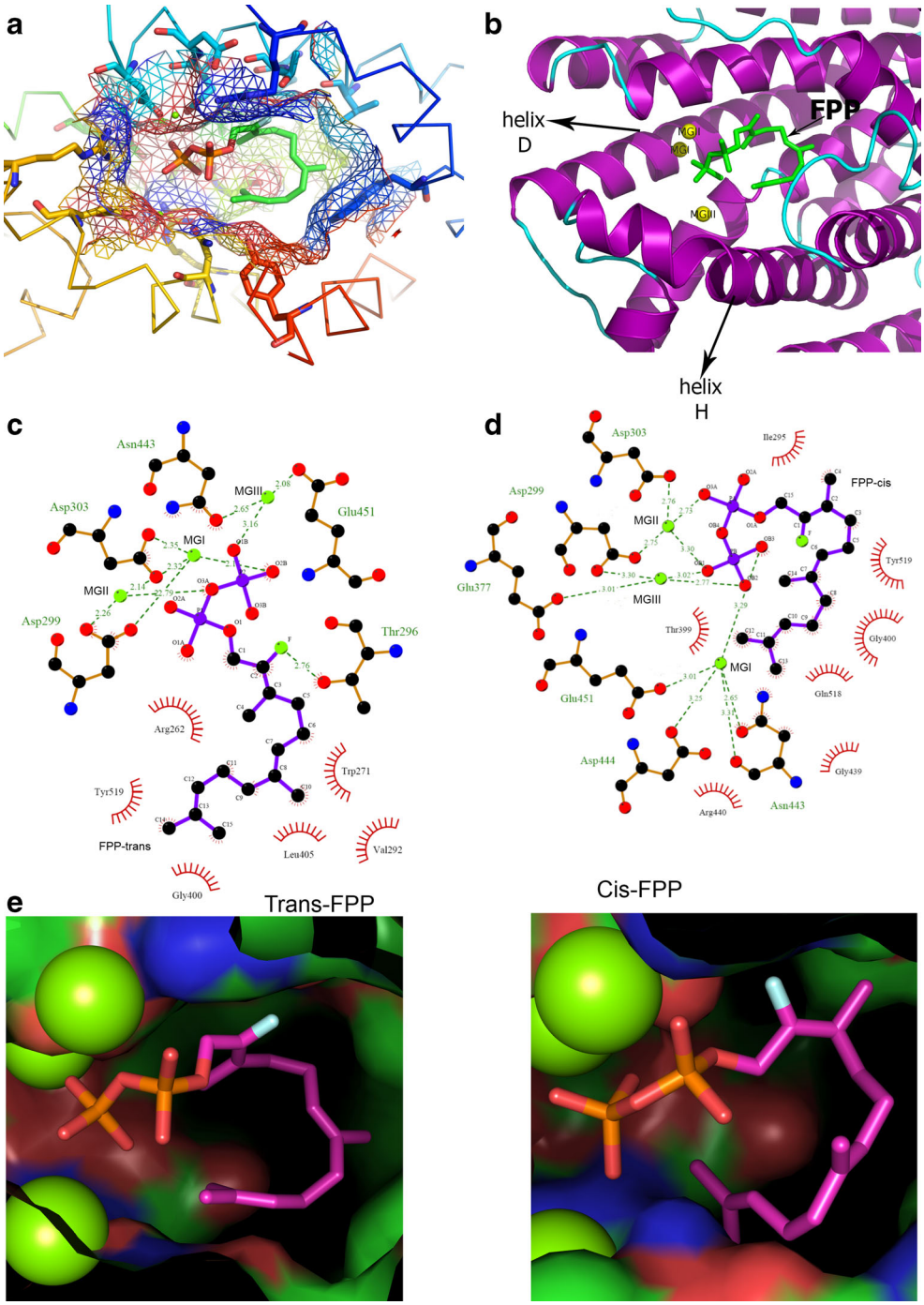


Fig. 6 Active site and binding pocket of FPP. a) The residues composing the FPP binding pocket are shown in stick with mesh surface around FPP. FPP molecule is depicted in two color coding. The farnesyl chain is shown in green and PP head in orange-pink. The stick and mesh around the FPP represents the side chains of the amino acids. b) Cartoon representation of 3D structure of binding pocket. Binding pocket is composed of two helices and one loop region. In addition, Mg^{2+} ions are shown in yellow and FPP is drawn in green. c) 2D representation of trans-FPP binding to ADS. The dashed lines indicate the hydrogen bonds between relevant amino acid side chains and FPP as well as Mg^{2+}. Other bonds with the farnesyl chains are shown with red radiated semicircular drawing. d) Cis-FPP binding to ADS. Polar bindings are illustrated as green line and non-polar interacting residues as semi-circles. e) 3-D conformation of cis and trans-FPP in the active site

for binding of FPP with aid of three Mg^{2+}, which upon binding facilitates ionization and initiation of the cyclization cascade. During cyclization reaction, the hydrophobic side chains of Leu and Met act as a barrier to water and stabilizes carbocationic intermediate transitions. Thus, the aromatic quadrupoles stabilize catalytic reaction.

In the ADS structural model of the present study, Mg^{2+} ions form ionic bridges between the negatively charged pyrophosphate group of FPP and negatively charged residues (Asp299, Asp303) in the active site of the enzyme (Fig. 6c and d) [37]. Our results showed that the parts of the active site that bind to the isoprenoid tail of the substrate are largely made up of aliphatic and aromatic residues (Gly400, Gly439, Trp519, Arg440) (Fig. 6d). These residues allow additional stabilization of the carbocations of the reaction cascades through cation–π or charge–quadrupole interactions [48]. This is the reason why ADS has evolved to incorporate the unique motif of NDLMT and, consequently, has aliphatic and hydrophobic residues in the active site. This motif makes a suitable hydrophobic area for the allylic carbocation in cyclization

Table 2 Active site residues in TPS family. Conserved residues in the active site of several members of TPS family are shown (right). The residues were detected based on multiple sequence alignment

amorpha-4,11-diene synthase(ADS)	Trp271, Tyr519, Asn443, Arg440, Glu451, Arg454, Glu377, Phe370, Thr399, Gly400, Arg262, Asp303, Asp300, Ile295, Thr296, Val292 and Leu405
(+)-delta-cadinene synthase(DCS)	Lys294, Lys248, Glu249, Leu245, Ser290, Glu252,
5-epi-aristolochene synthase (TEAS)	Tyr520, Val516, Ile515, Leu407, Tyr404, Thr403, Cys440, Arg441, Asp444, Glu452, Glu379, Asp305, Asp301, Tyr376, Thr401, Ile297, Ile294, Arg264, Thr528, Tyr527,
α-bisabolene synthase(AgBIS)	Asp570, Asp567, Tyr641, Asp713, Arg710, Ile667
limonene synthase(LS)	Asn345, Met458, Ser454, Ile453, Ser452, Tyr427, Glu430, Asp509, Lys512, Asp356, Asp353, Arg315, Thr349, Ile348, Asp352, His579, Trp324, Tyr573, Asp577, Leu492, Arg492, Asp497, Thr500

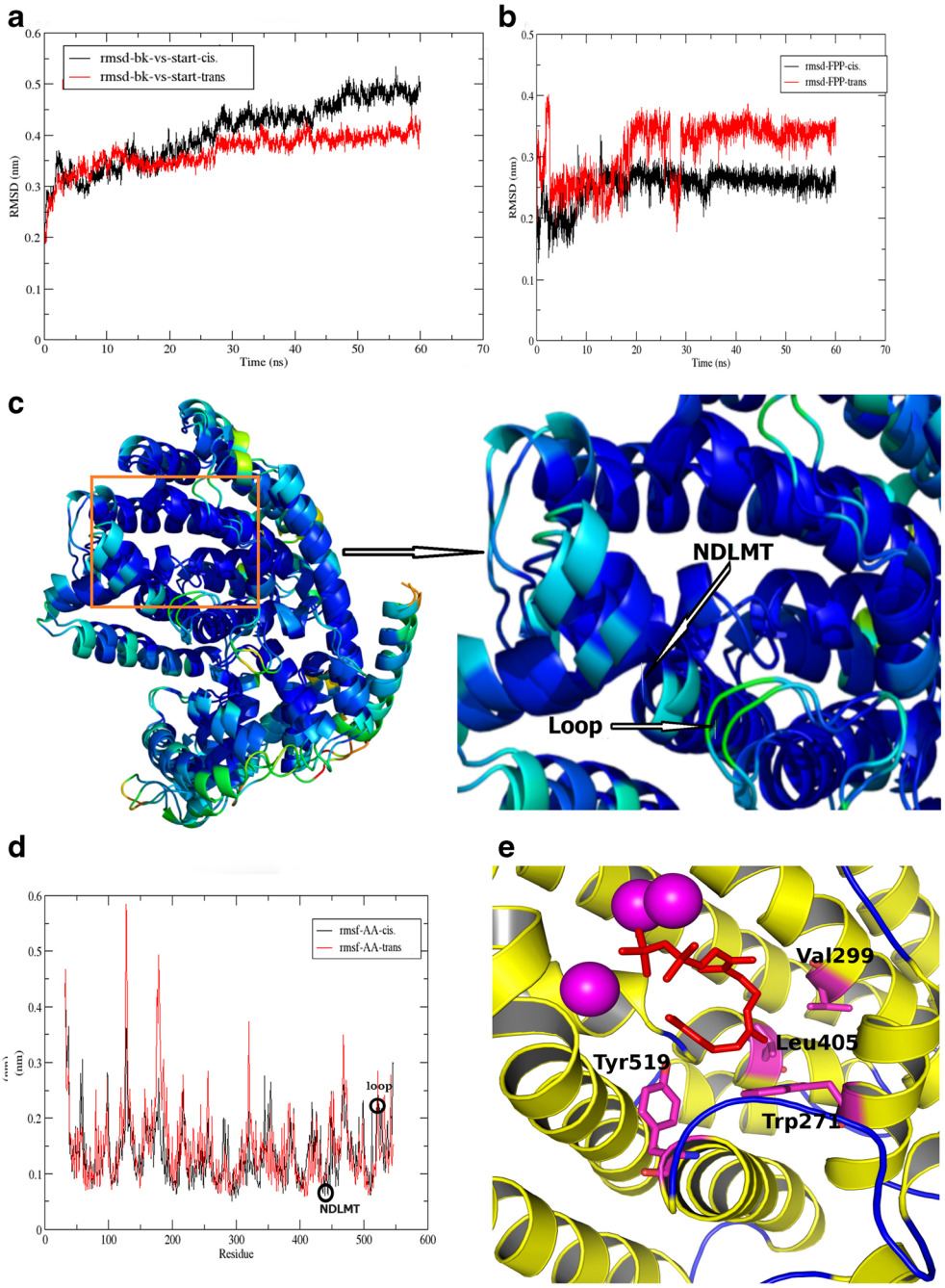


Fig. 7 Structure validation of the ADS model. RMSD of ADS (a) and FPP (b) showed that the predicted model is stable. c) Color coded cartoon representation of fluctuations during MD simulations. The light colors indicate highly flexible regions and the dark blue indicates regions with little flexibility. In the active site, loop region has flexibility that has been magnified in the right panel. d) RMSF plot demonstrated that the ADS structure is stable during MD simulation, only in N-terminal, there are some fluctuations in comparison with other regions. However, NDLMT region is stable. e) Conformation of loop region in active site

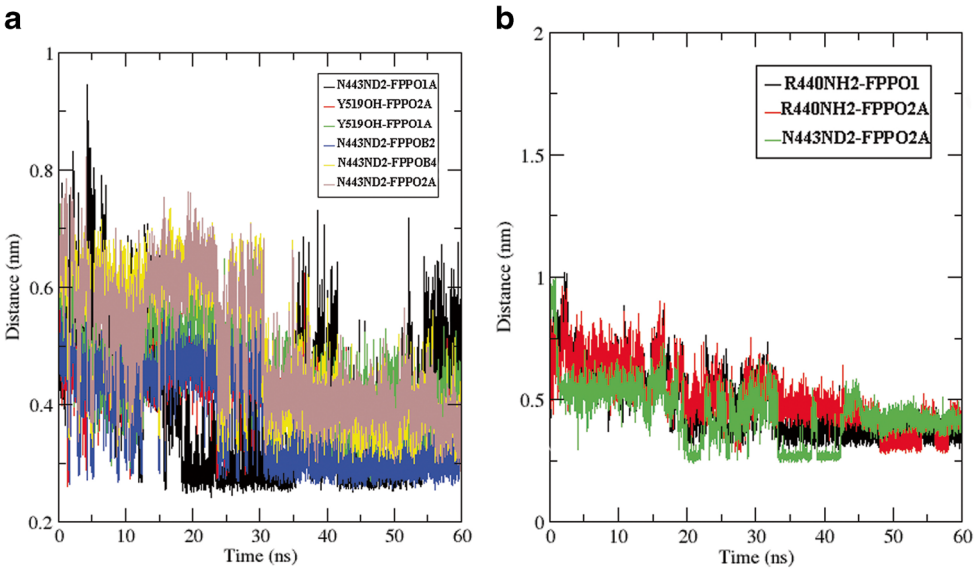


Fig. 8 Hydrogen-binding between FPP and ADS. The average distances between donors and acceptors for the hydrogen bonds are shown for cis (a) and trans-FPP model (b)

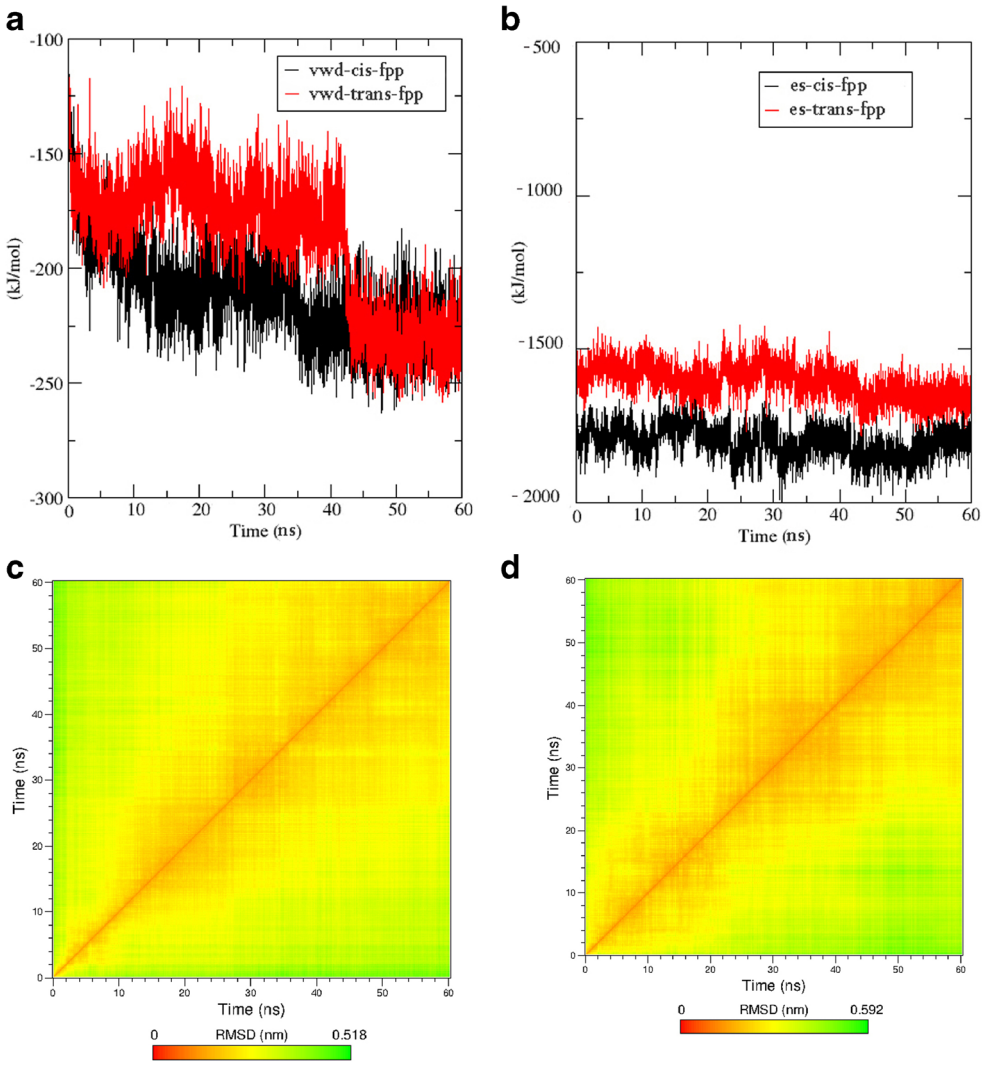


Fig. 9 Non-bonded interactions between FPP and ADS. The Lennard-Jones short-range (LJ-SR) for *van der Waals* interaction (a) and Coulombic short-range (Coul-SR) (b) for electrostatics interactions were calculated between ADS and FPP. The sum of Lennard-Jones and Coulombic energy was more negative for cis (-1900 kJ mol^{-1}) than trans (-1700 kJ mol^{-1}). RMSD matrix was calculated over the backbone atoms for cis (c) and trans (d) model over the 60 ns MD simulation. The red represents low RMSD, whereas green represents high RMSD

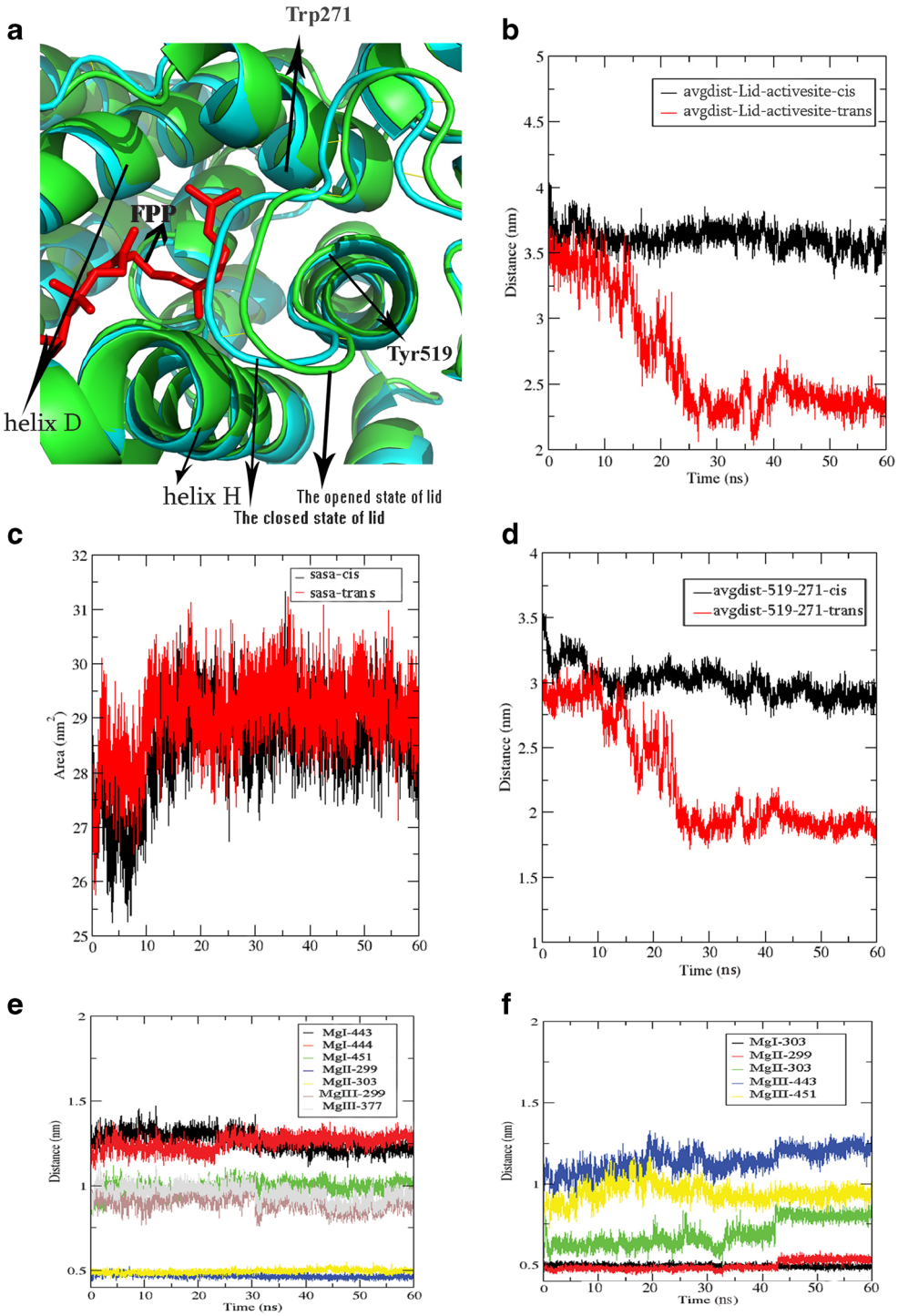


Fig. 10 A structurally flexible lid covers the ADS active site. The structures of two states of the lid "closed" and "open" were superimposed (a). The distances between the lid and the active site for both cis and trans model (b) were calculated and compared with each other. c) The solvent accessible surface area (SASA) of the active site. d) The distance between Tyr519 and Trp271 is depicted during 60 ns MD simulation. The distances between three Mg^{2+} ions and protein residues in the active site were calculated for cis (e) and trans-FPP model (f) during simulation

mechanisms [36]. Regarding the role of aliphatic amino acids in the active site of SQTPSs, substitution of the Tyr92 with smaller aliphatic and aromatic residues (e.g, Lue, Val, Ala, and Tyr) leads to an increased amount of acyclic products. Interestingly, a linear relationship was observed between the van der Waals volume of residues at position92 and the amount of cyclic products [49].

Studying the catalytic dynamic of ADS also contributes to disciplines other than antimalarial drugs. For example, bisabolene is recognized as the precursor of bisabolane which is an alternative fuel candidate. An interesting point here is that in both ADS and the *Abies grandis* α-bisabolene synthase (AgBIS), bisabolyl carbocation is the intermediate bound compound in the active site [50].

Conclusions

ADS catalyzes the first committed step in the biosynthetic conversion of the substrate FPP to artemisinin, which is a highly effective drug against malaria, but is in short supply and unaffordable to most malaria sufferers. In this study, homology modeling and several steps of MD simulations were performed to unravel the mechanisms behind the stereochemical influence of ADS active site conformation on FPP and Mg^{2+} binding. Results revealed DDYTD and NDLMT as metal-binding motifs in the putative active site, which is comprised of the D and H helixes and one loop region (aa519–532). Additionally, residues including Tyr519, Asp444, Trp271, Asn443, Thr399, Arg262, Val292, Gly400, and Leu405, are involved in the FPP binding and determining the fate of the allylic carbocation intermediate. Structural details might have implications for designing an enzyme with improved activity. Such an achievement will help to raise artemisinin production to a level high enough to reduce the expense of ACTs below their current prices.

References

1. Olsson ME, Olofsson LM, Lindahl A-L, Lundgren A, Brodelius M, Brodelius PE (2009) Localization of enzymes of artemisinin biosynthesis to the apical cells of glandular secretory trichomes of *Artemisia annua* L. Phytochemistry 70(9):1123–1128
2. Wang H, Ma C, Ma L, Du Z, Ye H, Li G, et al (2009) Secondary metabolic profiling and artemisinin biosynthesis of two genotypes of *Artemisia annua*. Planta Med 75(15):1625–1633
3. Harborne JB (1991) Recent advances in the ecological chemistry of plant terpenoids. In: Harborne JB, Tomas-Barberan FA (eds) Ecological chemistry and biochemistry of plant terpenoids. Clarendon Press, Oxford, pp. 399–426
4. (WHO) WHO World Malaria Report 2010
5. WHO, (WHO) WHO (2010) Global report on antimalarial drug efficacy and drug resistance: 2000–2010. WHO, Geneva-Switzerland
6. Abdin MZ, Rehman RU, Jain SK, Israr M (2003) Artemisin, a novel antimalarial drug: biochemical and molecular approaches for enhanced production. Planta med. 69:289–299
7. Martin VJ, Pitera DJ, Withers ST, Newman JD, Keasling JD (2003) Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. Nat. Biotechnol. 21(7):796–802
8. Tsuruta H, Paddon CJ, Eng D, Lenihan JR, Horning T, Anthony LC, et al (2009) High- level production of amorpha-4, 11-diene, a precursor of the antimalarial agent artemisinin, in *Escherichia coli*. PLoS One 4(2):e4489
9. McCarvey DJ, Croteau R (1995) Terpenoid metabolism. Plant Cell 7:1015–1026
10. Bouwmeester HJ, Wallaart TE, Janssen MHA, van Loo B, Jansen BJM, Posthumus MA, et al (1999) Amorpha-4, 11-diene synthase catalyses the first probable step in artemisinin biosynthesis. Phytochemistry 52(5):843–854
11. Olofsson L, Engström A, Lundgren A, Brodelius PE (2011) Relative expression of genes of terpene metabolism in different tissues of *Artemisia annua* L. BMC Plant Biol 11(1):45
12. Croteau R, Cane DE (1985) Monoterpene and sesquiterepene cyclases. Methods Enzym 110:383–405
13. Chappell J (1995) The Biochemistry and molecular biology of Isoprenoid metabolism. Plant Physiol. 107(1):1–6
14. Ro DK, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, et al (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature 440(7086):940–943
15. Paddon CJ, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, McPhee D, et al (2013) High-level semi-synthetic production of the potent antimalarial artemisinin. Nature 496(7446):528–532
16. Salmon M, Laurendon C, Vardakou M, Cheema J, Defernez M, Green S, et al (2015) Emergence of terpene cyclization in *Artemisia annua*. Nat Commun 6:6143
17. Jian-Xu L, Xin F, Qin Z, Ju-Xin R, Chang-Qing Y, Ling-Jian W, et al (2013) Rational engineering of plasticity residues of sesquiterpene synthases from *Artemisia annua*: product specificity and catalytic efficiency. Biochem J 451(3):417–426
18. Alam P, Kiran U, Ahmad MM, Kamaluddin KMA, Jhanwar S, et al (2010) Isolation, characterization and structural studies of amorpha - 4, 11-diene synthase (ADS(3963)) from *Artemisia annua* L. Bioinformation 4(9):421–429
19. Bairoch A, Apweiler R (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res 28(1):45–48
20. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23(21):2947–2948
21. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J. Mol. Biol. 215(3):403–410
22. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, et al (2006) Comparative protein structure modeling using Modeller. Curr Protoc Bioinforma Chapt 5:Unit 5.6. doi:10.1002/0471250953.bi0506s15
23. Laskowski RA, Moss DS, Thornton JM, Macarthur MW (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26:283–291
24. Melo F, Devos D, Depiereux E, Feytmans E (1997) ANOLEA: a www server to assess protein structures. Proc Int Conf Intell Syst Mol Biol 5:187–190
25. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC (2005) GROMACS: fast, flexible, and free. J Comput Chem 26(16):1701–1718
26. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, et al (2010) Improved side chain torsion potentials for the Amber ff99SB protein force field. Proteins Struct Funct Bioinforma 78(8):1950–1958
27. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79(2):926–935
28. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: a linear constraint solver for molecular simulations. J Comput Chem 18(12):1463–1472
29. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81(8):3684–3690
30. Parrinello M, Rahman A (1980) Crystal structure and pair potentials: a molecular- dynamics study. Phys Rev Lett 45(14):1196
31. Merlino A, Mazzarella L, Carannante A, Di Fiore A, Di Donato A, Notomista E, et al (2005) The importance of dynamic effects on the enzyme activity X-ray structure and molecular dynamics of Onconase mutants. J Biol Chem 280(18):17953–17960
32. Lüthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. Nature 356(6364):83–85
33. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics 30(9):1236–1240

34. Pagni M, Ioannidis V, Cerutti L, Zahn-Zabal M, Jongeneel CV, Hau J, et al (2007) MyHits: improvements to an interactive resource for analyzing protein sequences. Nucleic Acids Res 35:W433–W437
35. Chen F, Tholl D, Bohlmann J, Pichersky E (2011) The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. Plant J. 66(1):212–229
36. Starks CM, Back K, Chappell J, Noel JP (1997) Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. Science 277(5333):1815–1820
37. Picaud S, Olofsson L, Brodelius M, Brodelius PE (2005) Expression, purification, and characterization of recombinant amorpha-4,11-diene synthase from *Artemisia annua* L. Arch Biochem Biophys 436(2):215–226
38. Christianson DW (2006) Structural biology and chemistry of the terpenoid cyclases. Chem Rev 106(8):3412–3442
39. Gennadios HA, Gonzalez V, Di Costanzo L, Li A, Yu F, Miller DJ, et al (2009) Crystal structure of (+)-δ-cadinene synthase from *Gossypium arboreum* and evolutionary divergence of metal binding motifs for catalysis. Biochemistry 48(26):6175–6183
40. Mau CJ, West CA (1994) Cloning of casbene synthase cDNA: evidence for conserved structural features among terpenoid cyclases in plants. Proc Natl Acad Sci 91(18):8497–8501
41. Vedula LS, Jiang J, Zakharian T, Cane DE, Christianson DW (2008) Structural and mechanistic analysis of trichodiene synthase using site-directed mutagenesis: probing the catalytic function of tyrosine-295 and the asparagine-225/serine-229/glutamate-233–motif. Arch Biochem Biophys 469(2):184–194
42. Baas B-J, Zandvoort E, Wasiel AA, Quax WJ, Poelarends GJ (2011) Characterization of a newly identified mycobacterial tautomerase with promiscuous dehalogenase and hydratase activities reveals a functional link to a recently diverged cis-3-chloroacrylic acid dehalogenase. Biochemistry 50(14):2889–2899
43. Augustin JM, Kuzina V, Andersen SB, Bak S (2011) Molecular activities, biosynthesis and evolution of triterpenoid saponins. Phytochemistry 72(6):435–457
44. Lesburg CA, Zhai G, Cane DE, Christianson DW (1997) Crystal structure of pentalenene synthase: mechanistic insights on terpenoid cyclization reactions in biology. Science 277(5333):1820–1824
45. Noel JP, Dellas N, Faraldos JA, Zhao M, Hess Jr BA, Smentek L, et al (2010) Structural elucidation of cisoid and transoid cyclization pathways of a sesquiterpene synthase using 2-fluorofarnesyl diphosphates. ACS Chem Biol 5(4):377–392
46. Levskaya A, Chevalier AA, Tabor JJ, Simpson ZB, Lavery LA, Levy M, et al (2005) Synthetic biology: engineering *Escherichia coli* to see light. Nature 438(7067):441–442
47. Dercole F, Rinaldi S (2008) Analysis of evolutionary processes: the adaptive dynamics approach and its applications: the adaptive dynamics approach and its applications. Princeton University Press, Princeton
48. Imhof M, Schlötterer C (2001) Fitness effects of advantageous mutations in evolving *Escherichia coli* populations. Proc Natl Acad Sci 98(3):1113–1117
49. Allemann RK (2008) Chemical wizardry? The generation of diversity in terpenoid biosynthesis. Pure Appl Chem 80(8):1791–1798
50. McAndrew RP, Peralta-Yahya PP, DeGiovanni A, Pereira JH, Hadi MZ, Keasling JD, et al (2011) Structure of a three-domain sesquiterpene synthase: a prospective target for advanced biofuels production. Structure 19(12):1876–1884