

Active site and structural analysis of sesquiterpene synthase towards synthesis of aromatic compounds related to agarwood formation in *Aquilaria malaccensis*

Aimi Wahidah Aminan (✉ aimiwahidah930415@gmail.com)

Universiti Malaysia Pahang, Lebuhraya Tun Abdul Razak

Azzmer Azzar Abdul Hamid

International Islamic University Malaysia, Pahang Darul Makmur

Saiful Nizam Tajuddin

Universiti Malaysia Pahang, Lebuhraya Tun Abdul Razak

Aizi Nor Mazila Ramli

Universiti Malaysia Pahang, Lebuhraya Tun Abdul Razak

Research Article

Keywords:

Posted Date: July 22nd, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1810009/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Agarwood is a fragrant resinous wood that has been widely used in traditional medicines, incense and perfume. Formation of agarwood in branches and trunks of certain genera including *Aquilaria* and *Gyrinops* is due to activation of defence mechanism that causes by natural phenomena or microbial infection. The main fragrant compounds of agarwood are sesquiterpenes and phenylethyl chromone derivatives, and a great variety of sesquiterpenes are contained in high-quality agarwood. Their diverse cyclic and acyclic structures are known to be synthesized from farnesyl pyrophosphate (FPP) by sesquiterpene synthase catalytic action. Mechanisms of reaction control and structures of sesquiterpene synthases from *Aquilaria* sp are studied by many groups trying to clarify the structure-function relationship in these enzymes, however the molecular recognition between the catalytic residues of sesquiterpene synthases with FPP producing the aroma compound remains unclear. In this study, sesquiterpene synthases isolated from *Aquilaria malaccensis* (AGS3) was predicted for its structure and docking simulation has been performed to determine the affinity of active site residues of sesquiterpene synthases towards FPP molecules. To reveal the unique catalytic mechanism of AGS3, the molecular docking of AGS3 was carried out in two different sites involving active site entrance for initial catalysis and within substrate cavity for final step. It was found that the aspartate and glutamate rich regions interacts with metal complexes at the entrance of active site and aromatic residues was subsequently interact with farnesyl chain of FPP within the substrate binding pocket while the diphosphate moiety of FPP (ppi) detached the substrate. This model will serve to be a framework for exploring the roles of active site residues of AGS3 to explore the enzyme function of sesquiterpene synthase producing aromatic compound sesquiterpene of agarwood in *A. malaccensis*.

Introduction

Aquilaria genus is known as most important source of agarwood production and is well distributed in India, China, Papua New Guinea and South East Asia (Chong *et al.*, 2015). Naturally, *Aquilaria* will take long period in producing high grade aromatic resin between 40 to 50 years. Known as hardwood trees, the aromatic or scented resin can be obtained from the trunk of *Aquilaria* which produced naturally or artificially induced (Zhang *et al.*, 2012). *Aquilaria malaccensis* are known as agarwood producing plants. Among those species, *A. malaccensis* is an industrial crop in Malaysia and known as the major agarwood producer from *Aquilaria* sp. (Elias *et al.*, 2017). Agarwood is produced after the trees are wounded by natural sources such as lightning strikes or munched by insects associated with microbial infection under natural processes that develop slowly over decades. The plant defence mechanism will generate antibody as defensive material during infection. As the antibody successfully prevented the infection from getting worse, the agarwood formation be remote.

Chemical studies revealed that sesquiterpenes and 2-(2-phenethyl) chromones are two main types of aromatic compounds and Constituents of agarwood essential oil were shown primarily to be sesquiterpenes. and a great variety of sesquiterpenes are contained in high-quality agarwood (Yagura *et al.*, 2003). Sesquiterpenes are a class of terpenes composed of three isoprene (C5) units (farnesyl chain)

attached with phosphate (ppi) groups namely farnesyl pyrophosphate (FPP) and usually build a C15 terpene compounds after being catalysed by the responsible enzymes. Terpenes form one of the largest and most diverse groups of natural products, mainly composed of monoterpene (C10), sesqui- terpene (C15), diterpene (C20), and triterpene (C30) compounds catalyze by terpene synthase (TPs), and these compounds and their derivatives include valuable natural products used for pharmaceuticals or perfumes. Numerous studies have been conducted and discovered the responsible sesquiterpene synthase (STs) producing the aromatic agarwood sesquiterpenes at molecular levels. δ -guaiene synthase have been isolated from *Aquilaria* plants (*A. crassna*, *A. sinensis* and *A. microcarpa*), and the translated products of these genes are found to be highly conserved. However, the details upon structural and catalytic mechanism of STs toward formation of aromatic compounds still scarce.

The transcriptomic data of *A. malaccensis* that has been elucidated previously become an insight into study of targeted STs genes that responsible toward agarwood sesquiterpenes production in *A. malaccensis*. The study allows further finding in gene expression analyses of STs from *A. malaccensis* as well upon detectable STs towards agarwood marker biosynthesis pathway. The characterization of protein enables structural studies of the protein by which contribute to the functional study of STs. For example, the elucidation of the three-dimensional structures of 5-epi-aristolochene synthase from *Nicotiana tabaccum* and (+)-bornyl diphosphate synthase from *Salvia ocinalis*, the discovery of STs from other genus revealed that these enzymes each consist of two distinct domains. The active sites of both proteins were located at C-terminal domain by the location of substrate binding pocket and catalytic binding motifs at the entrance of active site which is conserved among other TPs including STs. Current approaches on the protein 3D structure investigation rely on protein crystallography (X-ray diffraction), electron microscopy and nuclear magnetic resonance (NMR) (Dorn *et al.*, 2014).

In this regard, the use of the in-silico protein structure prediction method can serve as an alternative to protein structure related studies. To understand the regulation and function of STs in *A. malaccensis*, the primary sequence of previously isolated STs from *A. malaccensis*, namely AGS3 was utilized for the computational analysis. Protein modelling and molecular docking was implemented to predict the 3D structure of STs from *A. malaccensis* as well as elucidating the molecular recognition between the residues of STs with FPP. The structural analysis reveals the details upon catalytic action of active site residues of AGS3 towards atomic substrate FPP. This study will provide an insight upon understanding STs mechanism toward catalysation of FPP becoming aromatic sesquiterpenes that responsible for agarwood formation in *A. malaccensis*.

Materials And Methods

AGS3 Data mining

The full length of sesquiterpene synthase from *A. malaccensis* was isolated using polymerase chain reaction. The complete nucleotide sequence encoded for sesquiterpene synthase has been deposited in the GenBank database under accession number (QRI93561.1) was analyzed using EXpaSY Translate

database (Gasteiger *et al.*, 2003). The physicochemical properties of AGS3 protein were analyzed using ExPASy protparam tool (Gasteiger *et al.*, 2005). NCBI Open Reading Frame Finder (Wheeler *et al.*, 2003) database was utilized to determine the full-length coding sequence of AGS3. Multiple sequence analysis of AGS3 was performed using nucleotide-to-protein BLASTx (Altschul *et al.*, 1990) against NCBI database. The resulting amino acid sequence (395 residues) was subjected to sequence analyses including BLASTp-PDB and PSI-BLAST (Altschul *et al.*, 1997). SUPERFAMILY HMM (Gough *et al.*, 2001) and Interproscan (Quevillon *et al.*, 2005) was used to identify any possible families or conserved domains in the protein.

Model development and evaluation

In this study, the structural prediction of AGS3 3D model was performed using AlphaFold system (Mirdita *et al.*, 2021). In this program, the protein sequence of AGS3 with length of 315 amino acids was submitted. Five possible models of AGS3 were generated and was ranked according to the best model from highest ranked position to the lowest. Model 5 of AGS3 structure the first ranked and was selected for further evaluation. The 3D structure of AGS3 was refined by minimizing the potential energy of the molecules of the protein via GROMACS using OPLS force field for 50 ps runs with 1.5k steps.

Active site pocket and catalytic residue prediction

Predictions of the active site pocket and of the amino acids that contributed to the pocket were conducted by performing analysis via COACH server (Jayaram *et al.*, 2012; Lipinski, 2004) and PyMOL (Delano, 2002). The catalytic residues and the highly conserved residues of AGS3 were identified using the software NCBI CDD (Marchler-Bauer *et al.*, 2015) by multiple sequence alignment and blind docking.

Protein-ligand docking simulation

The 3D ligands molecule; farnesyl pyrophosphate (FPP) was retrieved from PUBCHEM (Kim *et al.*, 2019) server. BABEL (O'Boyle *et al.*, 2011) was used to convert the ligand structures from MDL mol to PDBQT files. Protein-ligand docking studies were carried out based on homology-modelled structures of AGS3 receptor associated with Mg^{2+} ions to FPP targeted ligand. The preparation of AGS3 receptor and FPP ligand were done Autodock version 4.2.6 and AutoTools 1.5.6 software (Morris *et al.*, 2009). The docking simulation and substrate docking calculations for the enzyme with each substrate were performed via Autodock VINA (Eberhardt *et al.*, 2021). The predicted binding affinity analysis and consistency of results were then combined to identify the most favored pair using PyMOL. The best result for each substrate was taken as the largest conformation cluster with the lowest binding energy and further submitted to PLIP (Adasme *et al.*, 2021) server for protein ligand type of interaction.

Results And Discussion

Primary structure analysis via protein sequence of AGS3

The nucleotide sequence of AGS3 from *A. malaccensis* was obtained from NCBI (accession number:). ORF finder reveals the translated sequence of AGS3 consists of 948 bp nucleotides that encoded for 315 amino acids. The physicochemical properties of AGS3 shows the calculated molecular weight of AGS3 protein is 36.89 kDa with isoelectric point (pI) value of 4.91. The full length coding sequence of AGS3 was considered shorter in length (315 aa) as compared to the average length of other STs that comprise of amino acids range between 553 to 556 residues (Durairaj *et al.*, 2018). The shorter length of AGS3 was further analysed for its confirmation. Protein domain analysis was carried out and it was found that other STs genes composed of two domains; namely N-terminal domain and C-terminal domain whereas AGS3 full length sequence covers only C-terminal domain (Figure 1). both large domains considered to be conserved in TP's superfamily including monoterpene synthase (MTs), sesquiterpenes (STs) and diterpenes (DTs). Each domain carries totally distinct function. In general, C-terminal domain responsible as catalytic domain to perform the catalytic action occurs in TP's protein family (Jia, 2016). The actual function of N-terminal domain is not well elucidated in TP's enzymes except for monoterpene synthase (MTs) as it is involved in catalytic activity of α -limonene synthase of *Mentha Spicata* toward its substrate Geranyl pyrophosphate (GPP) producing cyclizable intermediate such as linalyl cation (Ashaari *et al.*, 2021).

In addition, it has also been reported that N-terminal domain plays role in stabilizing the structure of catalytically active site located in C-terminal domain for (Ee *et al.*, 2014). Domain swapping analysis of N-terminal domain between sesquiterpene synthase vetispiradiene synthase from *Hyoscyamus muticus* with N-terminal domain of tobacco 5-epi-aristolochene synthase (TEAS) from *nicotiana tabacum* found out that vetispiradiene synthase was fully functioned. it has been suggested to involve in interaction between these two large domains by facilitate the proper folding of the structure. s indicates that this domain acts as a scaffold facilitating proper folding of the catalytically active C-terminal domain (Köllner *et al.*, 2004).

However, it might be not contributed to any important role towards the enzyme. The available protein crystal structures suggest that most of plant diterpene and some sesquiterpene synthases possess all these domains but only one domain is functional. In case of AGS3 protein, the absence of N-terminal domain possibly not affected the functionality of the enzyme in catalysing FPP into sesquiterpene since the active site of AGS3 was predicted to locate at C-terminal domain. Moreover, previous study upon chemical profiling of similar sample of *A. malaccensis* revealed several potential sesquiterpene compounds that related to high quality of agarwood production as high-quality agarwood is believed to be enriched with aromatic and terpene compounds especially sesquiterpenes such as α -guaiene, β -selinene, aromadendrene, α -bulnesene and agarospirol. Thus, it is clearly proved that the AGS3 from *A. malaccensis* was functioned in producing these aromatic compounds of agarwood resin. Also, previous study upon transcriptomic data of *A. malaccensis* found out that transcript encodes for AGS3 was highly upregulated in term of gene expression. It shows that incomplete presence of both terminal domains did not affects the expression of STs. This is due to not all genes with complete domains were functional,

and some functional ones may lose activity in either one domain although its presence in the protein sequence (Zhang *et al.*, 2022).

Subsequence analysis was to elucidate motifs available in AGS3 from *A. malaccensis* (Figure 2). Multiple sequence alignment upon primary sequence of AGS3 with other STs available in NCBI database from different species of *Aquilaria* shows the presence of conserved motifs of STs of *Aquilaria*; the aspartate rich motifs; DDTYD and DDLQTYKxE and arginine-rich motif RDR; and revealed the lack of N-terminal domain from the primary sequence of AGS3 compared to other STs from *Aquilaria* sp as another conserved motif of TP enzymes superfamily RRx8W located at N-terminal domain was absent in AGS3 sequence (Figure 3). These three conserved motifs located at C-terminal domain of STs. Both DDTYD and DDLQTYKxE flanked the entrance of substrate binding pocket while the RDR motif located opposite to the entrance. Previous study discovered the function of aspartate rich motifs in STs as binding location for metal ions required in initiation step of STs catalytic mechanism; while arginine rich motif RDR helps in complexation of the diphosphate moiety (ppi) group near the entrance of active site. As for RRx8W motif that supposedly located at the N-terminal domain and plays role in monoterpene synthase cyclization (Bohlmann *et al.*, 1998; Trapp and Croteau, 2001)

Furthermore, the deduced amino acid of AGS3 showed no signal peptide is present which is consistent with other sesquiterpene synthases and is shorter than monoterpene synthases (600–650 amino acids). Sesquiterpene synthases responsible for sesquiterpenes (C₁₅) production are localized in the cytosol, whereas monoterpene synthases that catalyze the production of monoterpenes (C₁₀) are present in the plastids. The N-terminus signal peptide is important for targeting the protein to plastid such as in monoterpene synthase (MTs) because the monoterpene biosynthesis occurs in plastid. The absence of an N-terminal plastid targeting signal peptide suggests that AGS3 is localized to the cytosol where FPP is found and where sesquiterpene biosynthesis takes place.

3D model building and refinement

In order to study the structural properties of STs in *A. malaccensis*, the 3D model of AGS3 was generated via AlphaFold program. Among five models generated, the first ranked model according to the algorithm of AlphaFold was selected and was refined by minimizing energy of the protein to the lowest state. The goal of refinement is to improve the predicted model in a way that it has accurate side chain orientations and high stereo-chemical quality so that it might get closer to the native structure (Carlsen & Røgen, 2015). Minimum energy arrangements of the atoms correspond to stable states of the system and subsequently the model will be evaluated. The minimized energy obtained after process was -17800.85 kJ/mol from -9531.31 kJ/mol.

Quality of preliminary and refined AGS3 model using were then evaluated using ERRAT, Verify3D and PROCHECK program as shown in Table 1. ERRAT program analyzed the relative frequencies of non-covalent interactions between atoms of various types; as overall quality factor which higher scores resembles good quality (Omar *et al.*, 2018). An ERRAT score value of ~95% conferring high resolution

(Khor *et al.*, 2014). The AGS3 3D model scores a low ERRAT with 97.39% for refined structure. While ERRAT calculated the non-bonded atomic interactions within predicted model, Verify3D estimated the compatibility of an atomic model with their own protein sequences (Omar *et al.*, 2018). The satisfactory model was is assumed to have score of more than 80% of estimated 3D/1D profile of amino acids of predicted structure (Kaushik, 2019; Khor *et al.*, 2014). In this study, the Verify3D score for AGS3 predicted 3D structure was higher with 95.87% for refined 3D model. Even though the ERRAT score was slightly lower than confidence value, the Verify 3D score is much higher indicated that the AGS3 predicted 3D model is in good quality and expected to be satisfactory.

The PROCHECK analysis was performed in order to validate the predicted 3D model’s reliability (Bhavsar *et al.*, 2019). The AGS3 3D model exhibited a good fit with the reference geometry with no residues were found in disallowed region in both preliminary and refined AGS3 3D model. Meanwhile 93.1% of residues located in the most favored region for refined AGS3 3D model. The other residues were found to reside in the additional allowed regions. As good quality model to be expected to have greater than 90% residues in most favored region plus none residues in disallowed region, thus the overall model evaluation indicated that the AGS3 structure predicted exhibited a good quality of modelled protein.

Table 1: Model evaluation analysis via verify3D, ERRAT and PROCHECK programs.

Model	Energy minimization (EM) state	Energy (kJ/mol)	ERRAT	VERIFY3D (%)	PROCHECK (%)
AGS3	Before	-9531.31	98.04	92.06	Favoured region : 94.1 Allowed region :5.8 Outlier region :0.0
	After	-17800.85	94.12	97.46	Favoured region :91.3 Allowed region 8.0 Outlier region :0.0

3D protein structure prediction

The structure consisted of 16 α-helices with short turning loops (Table 2). This model generated was similar to the secondary structure analysis, by which AGS3 protein enzyme was predicted to consists of 16 α-helices (Helix A to helix K) where 77.14% (243 aa) of amino acid residues of AGS3 folded into α-

helices and 22.86% (72 aa) of amino acids folded into coils. The catalytic binding motifs of terpene synthases that occupy the entrance of active site were located on helix D (ASP67 to ASP71) and helix H2 (ASP212 to GLU220). Another conserved arginine motif, RDR, was also found in C-terminal domain of terpene synthase which is located at the upstream region of DDTYD motif and was important in the complexing of the diphosphate group, after the ionization of farnesyl pyrophosphate (Davis & Croteau, 2000). The RDR motif was located at turning loop between Helix B and Helix C (ARG30 to ARG32). This results also supported by protein folding prediction analysis of AGS3 that was performed using SUPERFAMILY HMM library Both of the conserved aspartate-rich regions, the DDTYD, and NSE/DTE motifs were found in the C-terminal domain demonstrating the importance of the C-terminal domain, which contains the active site, for the catalysis of the substrate FPP. Basically, the DDTYD motif involved in coordination of divalent metal ions for substrate binding while the second motif, DDLQTYKxE, is less conserved and appear mostly in microbial and fungal terpene cyclases (NxxSxxxE) and in most plant terpene cyclases (DxxTxxxE) (Durairaj *et al.*, 2018).

Table 2: The helices structure of AGS3 model with their respective amino acids. The catalytic motif of sesquiterpene synthase located in bold helices structure.

AGS3 16-helices	
Name	Amino acid
HA	2-20
HB	22-25
HC	33-43
HD	50-73
HE1	76-88
HE2	91-96
HE3	100-121
HF	125-152
HG1	158-169
HG2	171-182
HH1	188-196
HH2	199-223
HI1	229-237
HI2	241-265
HJ	269-288
HK	300-309

Multiple sequence alignment of AGS3 with other sesquiterpene synthase from *Aquilaria* sp. is performed using CLUSTALW program (Figure 3). The sesquiterpene synthase primary sequences used in this alignment was choose according to their highest percentage of similarity from BLASTx analysis. The conserved motif DDTYD, DLQTYKAE and RDR was determined with all residues lie on these motifs are similar to other sesquiterpene synthase from other species of *Aquilaria* indicated that these amino acid residues are highly conserved in *Aquilaria* sp. sesquiterpene synthase proteins (Figure 4).

Active site pocket and catalytic residue analysis

The well-known catalytic residues of sesquiterpene synthase (DDTYD) and (DDLQTYKxE) was predicted to locate at the entrance of active site. The active site residues however are not been fully elucidated. Analysis of catalytic residues within substrate binding pocket of AGS3 was performed via multiple

sequence alignment, blind docking and server prediction methods. It was found that 10 residues appeared in the three different analyses of active site prediction suggested that these residues occupied the substrate cavity of AGS3 structure with another 14 conserved catalytic residues located at the entrance of the pocket as shown in Table 3.

Table 3: The catalytic residues of AGS3 at the entrance of active site and prediction of catalytic residues within substrate binding pocket.

CATALYTIC RESIDUES OF AGS3					
Active site residues prediction				Entrance	
Multiple sequence alignment	Blind docking	COACH server	Summarised binding pocket residues		
30arg	39trp	30arg	30arg	asp67	asp212
39trp	60ile	39trp	39trp	asp68	asp213
60ile	174phe	60ile	60ile	thr69	leu214
62ile	208ala	63thr	63thr	tyr70	gln215
63thr	294tyr	64ser	64ser	asp71	thr216
64ser		169glu	174phe		tyr217
143tyr		170glu	208ala		lys218
209arg		174phe	209arg		ala219
210tyr		208ala	288tyr		glu220
288tyr		209arg	294tyr		
292asp		288tyr			
294tyr					

Molecular docking of AGS3

To gain further insight into the active site of the enzyme investigated here, the model structure of AGS3 was carried out with molecular docking using FPP. Docking of FPP substrate yielded multiple docking positions. The criteria for choosing the best docking position were based on the lowest docking score and the number of hydrogen bonds between the substrate and the amino acid residues (Ashaari *et al.*, 2021). A docking position with the least docking score has the highest affinity towards the ligand, and hence is

the best docked conformation as shown in Figure 6. Docking analysis was performed between AGS3 model with its substrate FPP. The catalytic mechanisms of sesquiterpene synthases can be divided into two phases; the metal ion dependent binding of the substrate and cleavage of the C–O bond to generate an ion pair between diphosphate; and a farnesyl cation is followed by a precise carbocation reaction cascade to the final reaction products. To investigate further upon catalytic mechanism of AGS3, the analysis was carried out in two different locations which is at the entrance of active site (to predict the molecular recognition between magnesium cluster required for substrate binding to the pocket) and within substrate binding pocket (to predict the molecular recognition between catalytic residues of the pocket with farnesyl chain of FPP). Different condition of AGS3 protein (receptor) preparation is that, the AGS3 model contained three magnesium ions located at the entrance of the active site.

Docking of AGS3 with FPP at the entrance of active site

Docking of FPP substrates confirmed that the active site of AGS3 was located at the C-terminal domain, proximate to the location of the Mg²⁺ cofactor. The diphosphate (PPi) moiety of the FPP substrates interacted with the highly conserved aspartate-rich DDTYD and DDLQTYKAE motifs located on helix H2 and helix D respectively via magnesium cluster as shown in Table 4.

Table 4: The complexes of magnesium cluster with its respective binding interactions with AGS3 and FPP molecules.

Magnesium cluster	Binding interaction
Mg ²⁺ A	asp213
	glu220
Mg ²⁺ B	asp71
	phosphate atoms
Mg ²⁺ C	asp67

From the analysis, it reveals the magnesium cluster was initially attached to the catalytic residues (as shown in Figure 7) of AGS3 protein. The Asp213 and Glu220 were coordinated to Mg²⁺A. The aspartate and glutamate residues coordinated to Mg²⁺A were identical to the avian FPP synthase (Tarshis *et al.*, 1994), taxadiene synthase (Koksai *et al.*, 2011), and *M. spicata* limonene synthase (Hyatt *et al.*, 2007). The Mg²⁺C however was the only metal ion that solely interacting with the Asp67. The second metal-binding catalytic residue Asp71 was coordinated to the Mg²⁺B and the ion interacted with phosphate moiety (PPi) of FPP. Similar metal ion coordination by the corresponding residues was also observed in

trichodiene synthase (Rynkiewicz, 2001),5-epi-aristolochene synthase (Starks, 1997) and taxadiene synthase (Koksal *et al.*, 2011).

The ideal distance for metal ion coordination was between 2.0 and 2.2 Å, which was more typically observed in higher-resolution structures (Koksal *et al.*, 2010). It was revealed that the coordination distance with the metal ion for AGS3 was within the range of 2.0–2.75 Å, which was longer than what was expected for Mg²⁺ coordination (Ashaari *et al.*, 2021). In addition to metal coordination interactions, the PPI moiety of FPP were also predicted to accept hydrogen bonds from Ser168 and Glu146 (Figure 8). The coordination of three metal ions and hydrogen bond interactions with basic residues presumably triggered the ionization of the substrate to yield carbocation intermediates that led to the production of terpenoids (Huang *et al.*, 2011). The metal-dependent ionization of the substrate resulted in the generation of a negatively charged PPI that was stabilized by Mg²⁺ ions and polar amino acid residues which created a positively charged region that drew the PPI away from the carbocations in the hydrophobic active pocket (Ashaari *et al.*, 2021).

Docking of AGS3 with farnesyl chain of FPP in the substrate binding pocket

Docking of AGS3 with farnesyl chain of FPP (without PPI moiety) was done assuming that the PPI moiety was carried away by the magnesium cluster; leaving only farnesyl chain in the pocket after activating the ionization of carbon atom of farnesyl chain (Durairaj *et al.*, 2018). Previous study upon catalytic mechanism of sesquiterpene synthase in other plant discovered that the addition of water to the cationic intermediate, followed by deprotonation, results in the formation of acyclic sesquiterpene products (Durairaj *et al.*, 2018). Thus, in this study, the hypothesis is tested to verify the similarity or differences of catalytic mechanism of sesquiterpene product within substrate binding pocket of *A. malaccensis* sesquiterpene synthase AGS3 to other genus or Aquilaria species. Results showed that six important residues lie within substrate binding pocket contributed to the formation of sesquiterpene product (Table 5).

Table 5: The interaction of AGS3 residues lies within substrate binding pocket to the farnesyl chain of FPP.

AGS3 catalytic residues	LIGAND ATOM
Trp39	5
Phe174	8
Phe280	4
Tyr294	1
Tyr294	13

The residues interacted with different atoms of farnesyl chain after initial step of ionization is done (the PPi moiety was carried away by the magnesium cluster) and catalysed the formation of carbocation. It was found that the farnesyl atoms interacted with all aromatic residues. Trp39 formed an interaction with C5, Phe280 and Phe174 interacted with C14 and C8 respectively; and Tyr294 forms double interaction with C1 and C13 with suggested that the residues might play the role of final product formation (Figure 9). Quenching of the final carbocation either by proton loss or by nucleophilic capture generates the sesquiterpene-type product.

Insight into catalytic residues of AGS3

The unique formation of various type of aromatic compound sesquiterpene drove an urge study upon its biosynthesis mechanism. The two phases catalytic mechanism of sesquiterpene synthase were illustrated in Figure 10 involving ionization of carbon atom of FPP after detaching of phosphate groups from the molecules via magnesium cluster cofactor; series of hydration and rearrangement of ionized carbon atom (carbocation intermediate); and deprotonation of final step led to product formation (Köllner *et al.*, 2006). The initial phase (Phase I) of catalytic reaction that begins with the metal interaction between magnesium cluster to AGS3 and PPi moiety of FPP which activates the carbocation formation of FPP causes the detachment of PPi moiety from FPP as well as magnesium cluster from AGS3 leaving farnesyl chain within the substrate binding pocket. Docking of AGS3-FPP complex from both analyses reveals the catalytic residues responsible in formation of aromatic compound sesquiterpene. It was found that the catalytic motifs **DDTYD** and **DDLQTYKxE** by which the residues in bold involved in this phase suggested the importance of acidic residues in metal interaction of sesquiterpene synthase catalytic activity. The Phase II of sesquiterpene synthase mechanism that take place within the active site pocket was further analysed.

From both analyses, several residues that interacts with farnesyl chain within substrate binding pockets was determined. Results from different docking site analyses were compared and it was found that four residues Trp39, Phe174 and Tyr294 were likely to be appear in both analyses as shown in Table 6. From both results, the aromatic residue Tyr294 was noted to form similar protein-ligand interaction to C13 suggested that this residue might play important role in the catalysis.

Table 6: Predicted residues of AGS3 that interacts with farnesyl chain of FPP within substrate binding pockets from different site of docking analysis results. The aromatic residue appeared the similar binding interaction was bold.

Catalytic residues of AGS3	Ligand atom	
	Entrance	Substrate binding pocket
Trp39	9	5
Phe174	12	8
Tyr294	13	1 & 13

The detachment of PPI moiety from FPP activates the formation of carbocation intermediate farnesyl cation (if undergo isomerization; nerolidyl cation) followed by the closure of ring structure by the nucleophile attack of C10 to C1. The series of hydrate shift begins with the C13 followed by C6 attack to C2 forming another ring structure led to possible formation of sesquiterpene δ -guaiene. This suggested that the AGS3 from *A. malaccensis* might catalysed the formation of aromatic sesquiterpene δ -guaiene. The prediction is possible as Tyr294 interacts with C1 and C13 suggested that the residue plays role in nucleophilic attack to C1 forming ring structure and C13 deprotonation led to second ring structure formation giving the formation of δ -guaiene product.

The nature and position of aromatic amino acid residue at the active site determined the docking orientation of the intermediate prenyl cation and therefore product specificity (Ashaari *et al.*, 2021). In amorphadiene synthase, the aromatic phenylalanine residue was similarly involved in positioning of the FPP substrate in the active site, which subsequently stabilized the carbocation intermediates (Abdallah *et al.*, 2016). A similar observation was also reported by Zhang et al with Nicotiana tabacum 5-epi-aristolochene synthase (TEAS) that catalysed the cyclisation of FPP into bicyclic 5-epi-aristolochene (Zhang *et al.*, 2016). Mutational analysis of the aromatic amino acids proved the essential role of these residues in the active site for stabilization of the carbocation intermediates (Abdallah *et al.*, 2018; Zhang *et al.*, 2016). Positioning of FPP in the AGS3 active site involving the action of aromatic residues (tyrosine, phenylalanine and tryptophan) as well as aspartate and glutamate rich regions at the entrance of active site proved that this docking analysis was rational and compatible with other crystal structures of sesquiterpene synthases.

Conclusion

The AGS3 model structure was successfully constructed from its amino acid sequences and this model will serve as the structural insights of AGS3 c-terminal domain catalytic site as a sesquiterpene synthase. The *A. malaccensis* sesquiterpene synthase exhibited features of a class I terpene synthase fold made up of α -helices consisting catalytic c-terminal domain with large substrate cavity. Based on the prior knowledge of the reaction mechanisms of other sesquiterpene synthases, it is hypothesized that AGS3 reaction mechanism begins with the metal-dependent ionization of the PPI moiety of respective FPP to form farnesyl cation. Assisting the metal ions in PPI complexation are the basic residues of Asp213,

Glu220, Asp67 and Asp71 that direct the PPI away from the active site after ionization. The addition of water to the cationic intermediate, followed by deprotonation by Tyr294 results in the possible formation of aromatic sesquiterpene δ -guaiene that gives the odour for agarwood resin in *A. malaccensis*. Thus, model will serve as a framework for exploring more the roles of aromatic residues of active site of sesquiterpene synthase that enables the formation of aromatic compound of high-quality agarwood sesquiterpene from *A. malaccensis*.

Declarations

Acknowledgments

The authors acknowledge the financial assistance provided by Universiti Malaysia Pahang.

Funding

The authors thank the funding from Universiti Malaysia Pahang under grant RDU182207-1.

Conflict of interest

The authors declare that they have no conflict of interest.

Author contributions

AAW - conceptualization, investigation & writing original draft. AHAA - data curation, methodology, validation. TSN - supervision, investigation. RANM - supervision, writing, review & editing.

Data availability

The authors confirm that the data underlying the findings are available from the corresponding author upon reasonable request.

References

1. Adasme *et al.* PLIP 2021: expanding the scope of the protein–ligand interaction profiler to DNA and RNA. *NAR* 2021
2. Ashaari, N. S., Mohd, M. H., Sabri, S., Lai, K. S., Song, A. A. L., Abdul Rahim, R., & Ong Abdullah, J. (2021). Kinetic studies and homology modeling of a dual-substrate linalool/nerolidol synthase from *Plectranthus amboinicus*. *Scientific Reports*, 11(1), 1–16. <https://doi.org/10.1038/s41598-021-96524-z>
3. Davis, E. M., & Croteau, R. (2000). *Cyclization Enzymes in the Biosynthesis of Monoterpenes, Sesquiterpenes, and Diterpenes*. 209.
4. Division, B., & Ehsan, S. D. (2015). *Agarwood Inducement Technology: A Method for Producing Oil Grade Agarwood in Cultivated Aquilaria malaccensis Lamk*. 6.

5. Durairaj, J., Di, A., Bouwmeester, H. J., Ridder, D. De, Beekwilder, J., & Van, A. D. J. (2018). *An Analysis of Characterized Plant Sesquiterpene Synthases*. 1–13.
6. Ee, S., Othman, R., Shaharuddin, N. A., Ismail, I., & Zainal, Z. (2014). *Functional Characterization of Sesquiterpene Synthase from Polygonum minus*. February. <https://doi.org/10.1155/2014/840592>
7. Elias, M. F., Ibrahim, H., & Mahamod, W. R. W. (2017). A Review on the Malaysian Aquilaria species in Karas Plantation and Agarwood Production. *International Journal of Academic Research in Business and Social Sciences*, 7(4), 1021–1029. <https://doi.org/10.6007/ijarbss/v7-i4/2911>
8. Jia, Q. (2016). *Computational Identification of Terpene Synthase Genes and Their Evolutionary Analysis*.
9. Köllner, T. G., Maille, P. E. O., Gatto, N., Boland, W., Gershenzon, J., & Degenhardt, J. (2006). *Two pockets in the active site of maize sesquiterpene synthase TPS4 carry out sequential parts of the reaction scheme resulting in multiple products* . 448, 83–92. <https://doi.org/10.1016/j.abb.2005.10.011>
10. Zhang, A., Xiong, Y., Fang, J., Jiang, X., Wang, T., Liu, K., Peng, H., & Zhang, X. (2022). *Diversity and Functional Evolution of Terpene Synthases in Rosaceae*.
11. Kelley LA, Sternberg MJE (2009) Protein structure prediction on the web: a case study using the Phyre server. *Nat Protoc* 4(3): 363–371
12. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410
13. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl Acids Res* 25(17):3389–3402
14. Gough J, Karplus K, Hughey R, Chothia C (2001). Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J Mol Biol* 313(4):903–919
15. Quevillon E., Silventoinen V., Pillai S., Harte N., Mulder N., Apweiler R., Lopez R. *Nucleic Acids Research* (2005), Vol. 33, Issue suppl 2, PMID: 15980438
16. Sillitoe I, Bordin N, Dawson N, Waman VP, Ashford P, Scholes HM, Pang CSM, Woodridge L, Rauer C, Sen N, Abbasian M, Le Cornu S, Lam SD, Berka K, Varekova IH, Svobodova R, Lees J, Orengo CA. *Nucleic Acids Res*. 2021 Jan Pubmed: 33237325
17. Eswar N, Webb B, Marti-Renom M, Madhusudhan MS, Eramian D, Shen M-Y, Pieper U, Sali A (2007) Comparative protein structure modeling using MODELLER. *Curr protoc protein sci* 2(1):1–30).
18. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* 4(3): 435–447
19. Guex N, Peitsch M (1997) SWISS-MODEL and the Swiss-Pdb viewer: an environment for comparative protein modeling. *Electrophoresis* 18(15):2714–2723).

20. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26(2):283–291.
21. Luthy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. *Nature* 356(6364):83–85
22. Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci* 2(9): 1511–1519.
23. Lipinski CA (December 2004). "Lead- and drug-like compounds: the rule-of-five revolution". *Drug Discovery Today: Technologies* 1 (4): 337–341.
24. B. Jayaram, Tanya Singh, Goutam Mukherjee, Abhinav Mathur, Shashank Shekhar, and Vandana Shekhar, "Sanjeevini: a freely accessible web-server for target directed lead molecule discovery", *BMC Bioinformatics*, 2012, 13, S7.
25. Delano WL (2002) The PyMOL molecular graphics system on World Wide Web
26. Marchler-Bauer A *et al.* (2015), "*CDD: NCBI's conserved domain database.*", **Nucleic Acids Res.**43(D)222-6.
27. Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J., & Bolton, E. E. (2019). PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.*, 49(D1), D1388–D1395.
28. N M O'Boyle, M Banck, C A James, C Morley, T Vandermeersch, and G R Hutchison. "Open Babel: An open chemical toolbox." *J. Cheminf.* (2011), 3, 33.
29. Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*, 30(16), 2785–2791.
30. Eberhardt, J., Santos-Martins, D., Tillack, A.F., Forli, S. (2021). AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *Journal of Chemical Information and Modeling*.

Figures

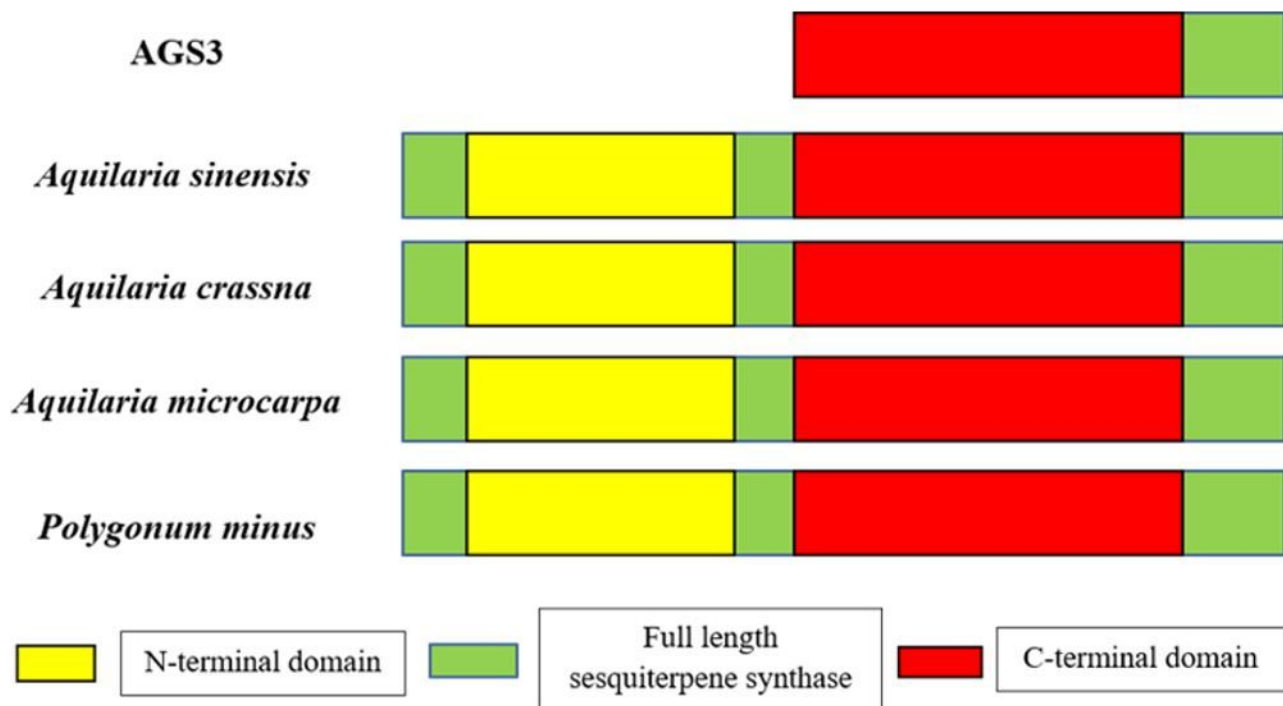


Figure 1

Terpene synthase domains found in plant sesquiterpene synthases. from *A. malaccensis* (QRI93561.1), *A. sinensis* (AIT75876.1), *A. crassna* (AEG77020.1), *A. microcarpa* (AHH25146.1) and *Polygonum minus* (JX025008) respectively.

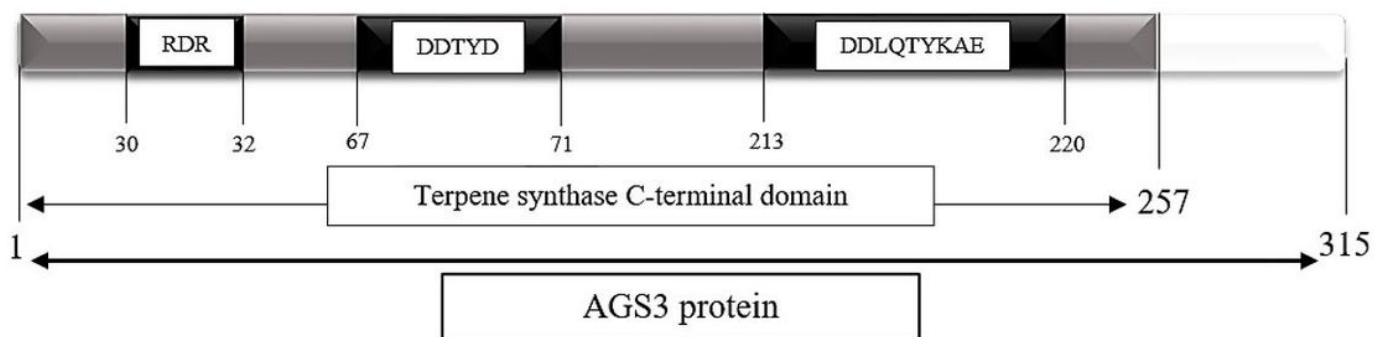


Figure 2

Domain and motifs analysis of AGS3. The large terpene synthase C-terminal domain was located at amino acid 1 to residue 257; the arginine rich RDR motif was located at residue 30 to residue 32; the first aspartate-rich DDTYD motif was located at amino acid 67 to 71; and the second aspartate-rich DDLQTYKAE motif was located at amino acid 213 to 220.

marked with asterisk [*].The universally conserved terpene synthase motifs RRx8W, DDTYD, RDR and DDLQTYKxE are highlighted in boxes. motifs RRx8W is absence in AGS3.

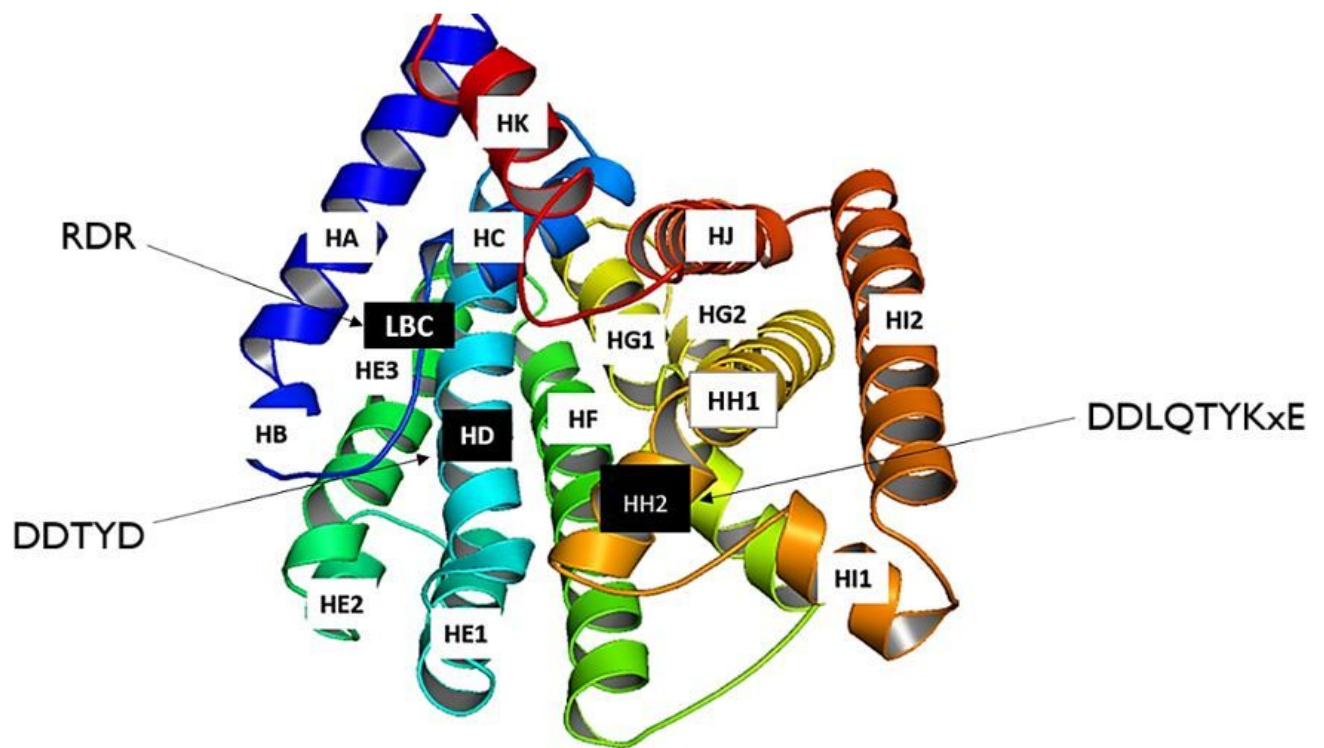


Figure 4

Predicted 3D model of AGS3 generated by the AlphaFold server. The 16 α -helices composing AGS3 are labelled. Helix D, Helix H and Loop BC were the location for conserved motifs DDTYD, DDLQTYKxE and RDR respectively.

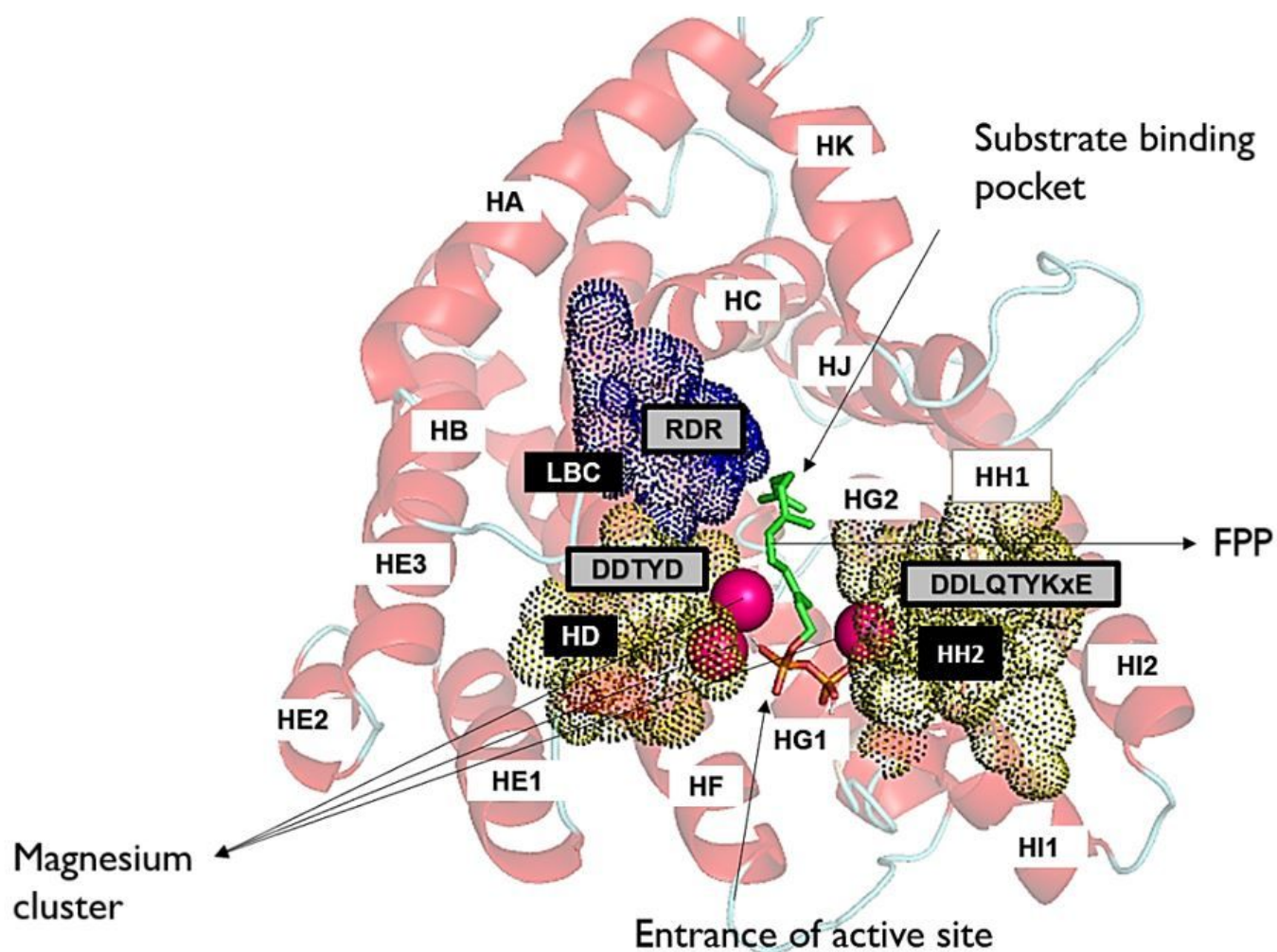


Figure 5

Best docked position of AGS3 associated with Mg^{2+} ions and FPP substrate via Autodock VINA. FPP is displayed in licorice with farnesyl chain (green) located in the substrate binding pocket and PPi moiety (orange) located at the entrance of active site. The three associated Mg^{2+} (magnesium cluster) were illustrated in purple sphere attached to the catalytic binding motifs DDTYD and DDLQTYKxE located at Helix D and Helix H2. Motif RDR was predicted to locate at the upper site of the entrance.

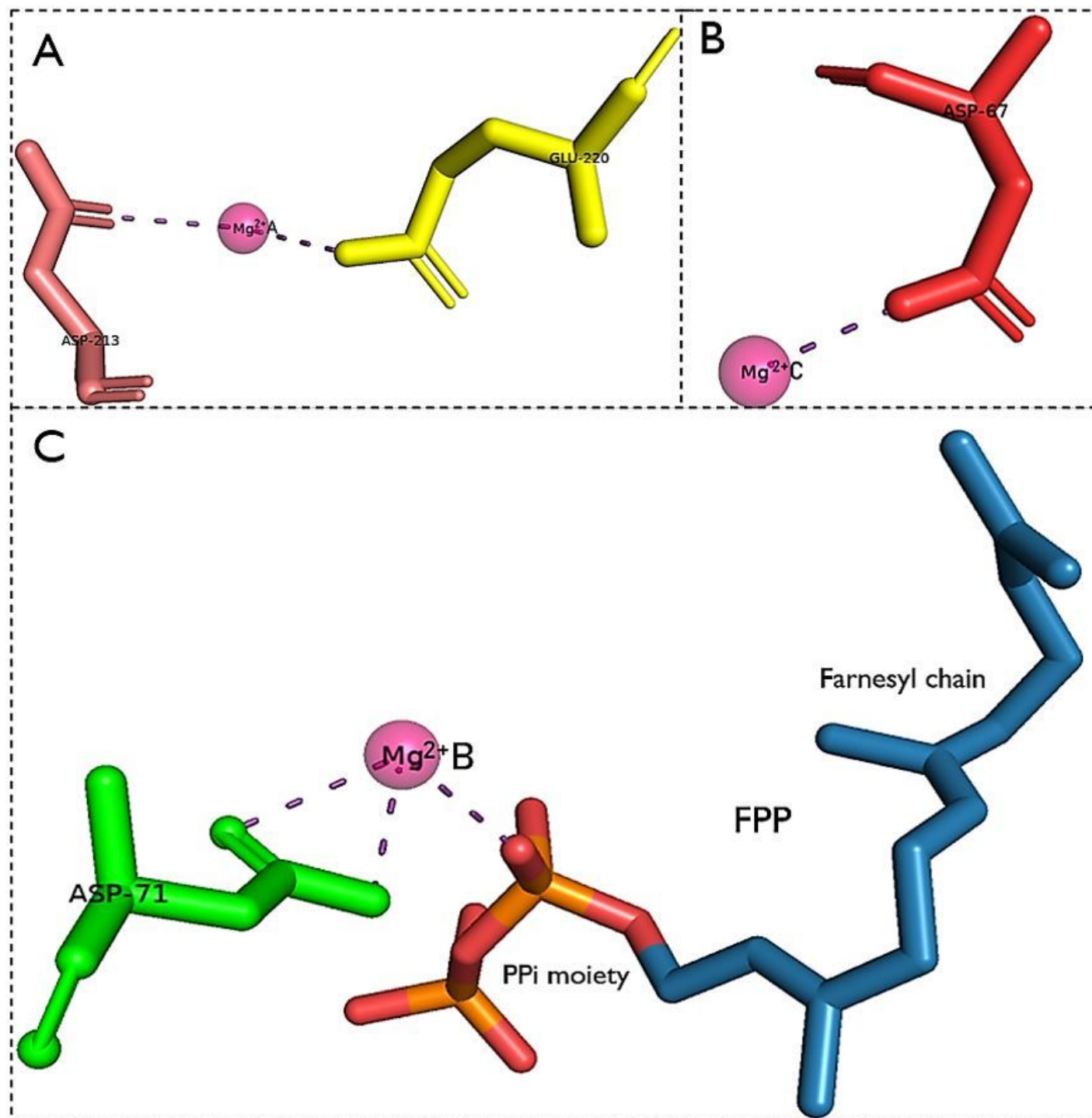


Figure 6

Representation of Mg^{2+} interactions located at the entrance A) $Mg^{2+}A$ attached to Asp213 and Glu220 of AGS3 structure; B) $Mg^{2+}C$ attached to Asp67 of AGS3; and C) $Mg^{2+}B$ with Asp71 of AGS3 and PPI moiety of FPP. The metal ion clusters that attached to catalytic residues of AGS3 interacts with PPI moiety of FPP before detaching the PPI moiety from FPP leaving only farnesyl chain in substrate pocket.

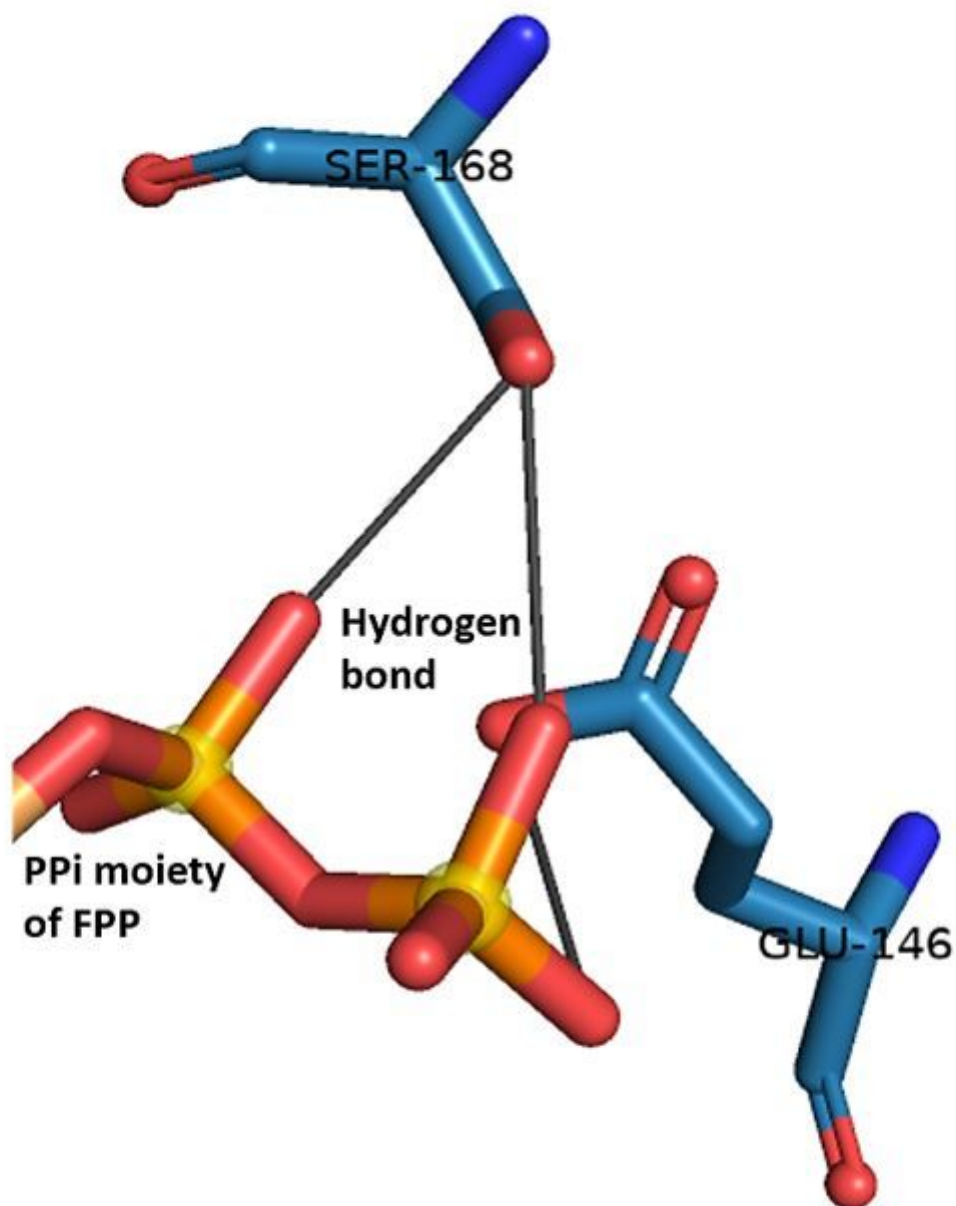


Figure 7

Representation of interaction between Ser168 and Glu146 to PPI moiety of FPP via hydrogen bonds.

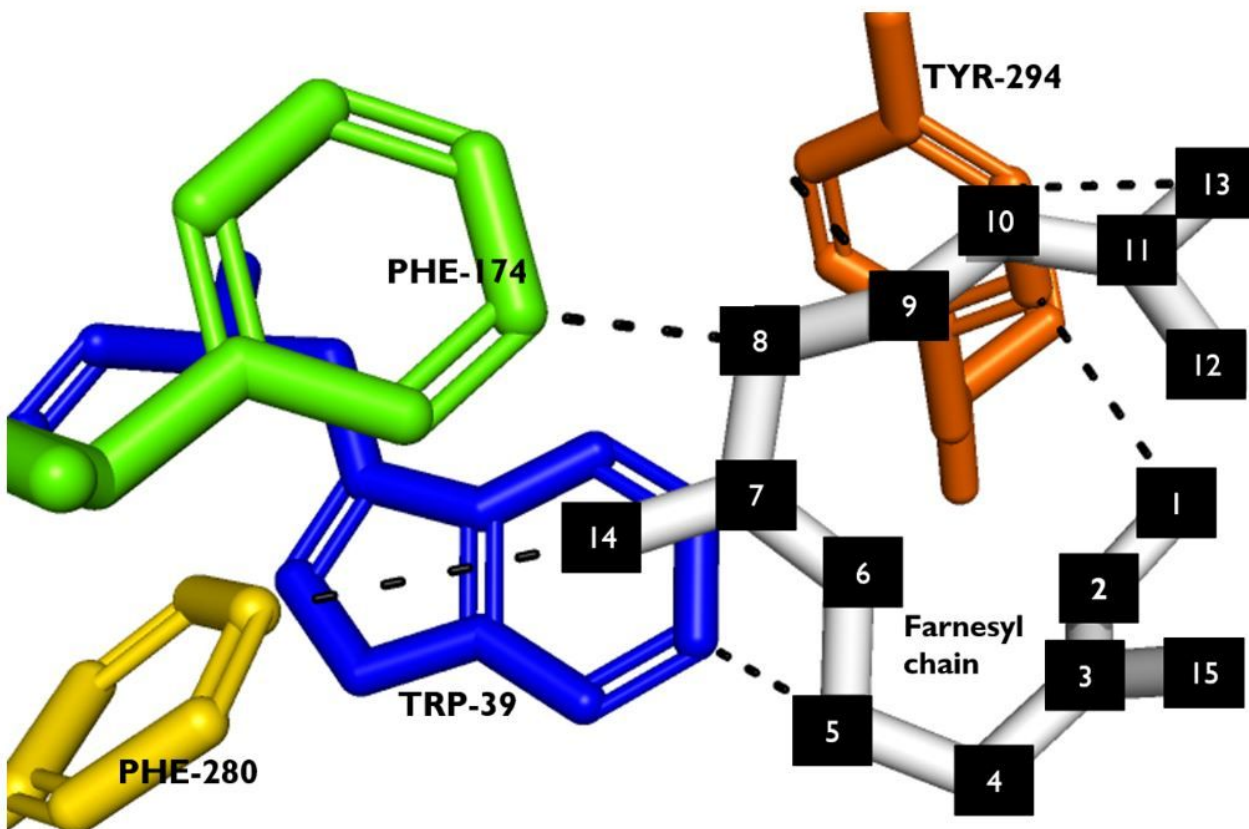


Figure 8

Representation of interactions between farnesyl chain of FPP with substrate binding pocket residues. Trp39 formed an interaction with C5; Phe280 interacted with C14 and Phe174 interacted with C8.

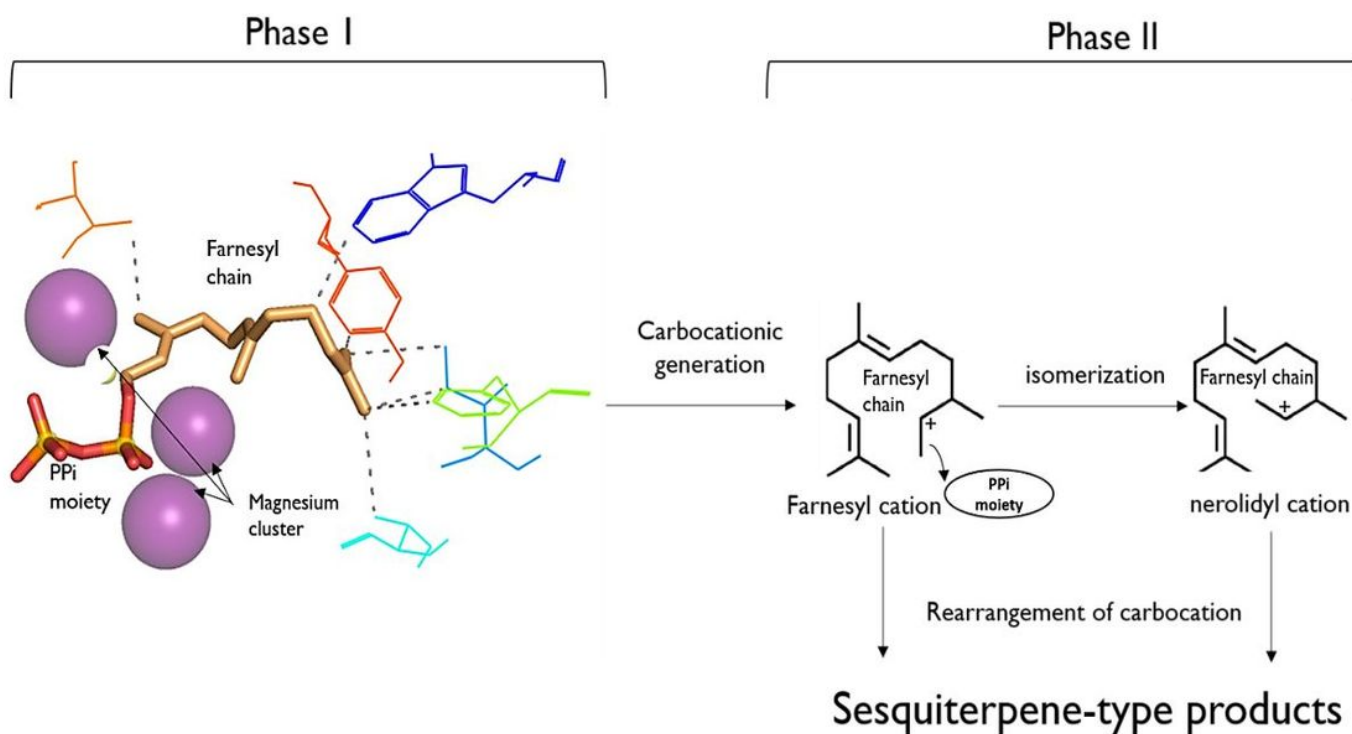


Figure 9

Illustration of catalytic mechanism of sesquiterpene synthase forming aromatic sesquiterpene. Phase I occurs at the entrance of active site between magnesium cluster and PPi moiety of FPP. Phase II occurs within substrate binding pocket between farnesyl chain of FPP and AGS3 structure.