# Mechanism of Assembling Isoprenoid Building Blocks 1. Elucidation of the Structural Motifs for Substrate Binding in Geranyl Pyrophosphate Synthase

Zhihong Liu,[†] Jingwei Zhou,[†] Ruibo Wu,* and Jun Xu*

Research Center for Drug Discovery, School of Pharmaceutical Sciences, Sun Yat-sen University, 132 East Circle at University City, Guangzhou 510006, China

S *Supporting Information*
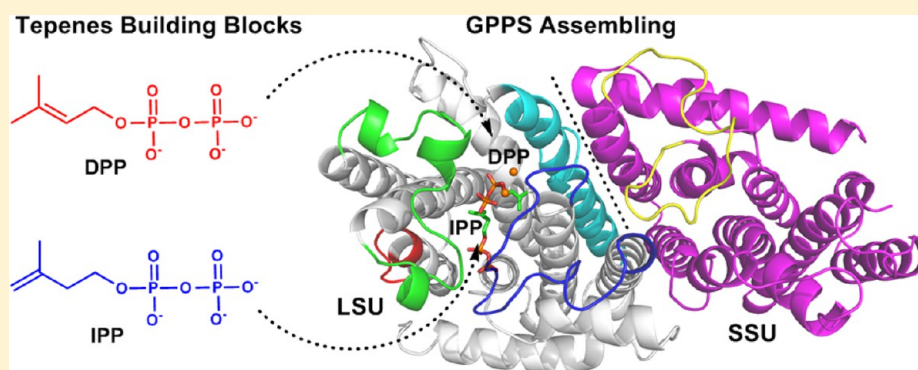
ABSTRACT: Terpenes (isoprenoids) represent the most functionally and structurally diverse group of natural products. Terpenes are assembled from two building blocks, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP or DPP), by prenyltransferases (PTSs). Geranyl pyrophosphate synthase (GPPS) is the enzyme that assembles DPP and IPP in the first step of chain elongation during isoprenoid biosynthesis. The mechanism by which GPPS assembles the terpene precursor remains unknown; elucidating this mechanism will help in development of new technology to generate novel natural product-like scaffolds. With classic and QM/MM MD simulations, an "open-closed" conformation change of the catalytic pocket was observed in the GPPS active site at its large subunit (LSU), and a critical salt bridge between Asp91(in loop 1) and Lys239(in loop 2) was identified. The salt bridge is responsible for opening or closing the catalytic pocket. Meanwhile, the small subunit (SSU) regulates the size and shape of the hydrophobic pocket to flexibly host substrates with different shapes and sizes (DPP/GPP/FPP, $C_5/C_{10}/C_{15}$). Further QM/MM MD simulations were carried out to explore the binding modes for the different substrates catalyzed by GPPS. Our simulations suggest that the key residues (Asp91, Lys239, and Gln156) are good candidates for site-directed mutagenesis and may help in protein engineering.

INTRODUCTION

Terpenes (isoprenoids) exist in almost all life forms, ranging from microorganisms to *Homo sapiens*, and are involved in a number of important biological processes.[1,2] The assembly of structurally diverse isoprenoids from isoprenoid building blocks is catalyzed by large families of enzymes in a series of coupling reactions (Figure 1).[2−4] Initially, two C_5 building blocks, dimethylallyl pyrophosphate (DMAPP, DPP) and isopentenyl pyrophosphate (IPP), are made in the mevalonate phosphate pathway[5−9] (MVA) or in the methylerythritol phosphate pathway (MEP).[6,7,10] The intermediates grow by furthering assembly of the building block (IPP) through elongation steps. For example, geranyl pyrophosphate (GPP, C_{10}) is generated by assembling DPP and IPP in a head-to-tail condensation catalyzed by geranyl pyrophosphate synthase (GPPS). The elongation of GPP with another IPP will produce farnesyl pyrophosphate (FPP, C_{15}), the elongation of FPP with another IPP will produce geranylgeranyl pyrophosphate (GGPP, C_{20}),

and so forth.[11−13] Finally, terpenes are generated from GPP, FPP, or GGPP through rearrangement and or cyclization reactions. Mono-, sesqui-, and diterpenes along with many other terpenes are used as flavors, fragrances (such as menthol), medicines (such as artemisinin), and nutritional supplements (such as vitamin A).[14] Chain elongation is the initial stage of terpene biosynthesis and produces terpene precursors. The catalytic mechanism for elongation was explored by Poulter and colleagues, who established the "ionization-condensation-elimination" mechanism for trans-prenyltransferase.[11,12,15,16] Fujihashi and co-workers reported the first crystal structure of cis-prenyltransferase[17] and articulated the enzyme's architecture and the chain elongation mechanism.[18,19] Quantum mechanics (QM) simulations on farnesyl pyrophosphate synthase (FPPS) suggest a one-step reaction through a carbocationic transition
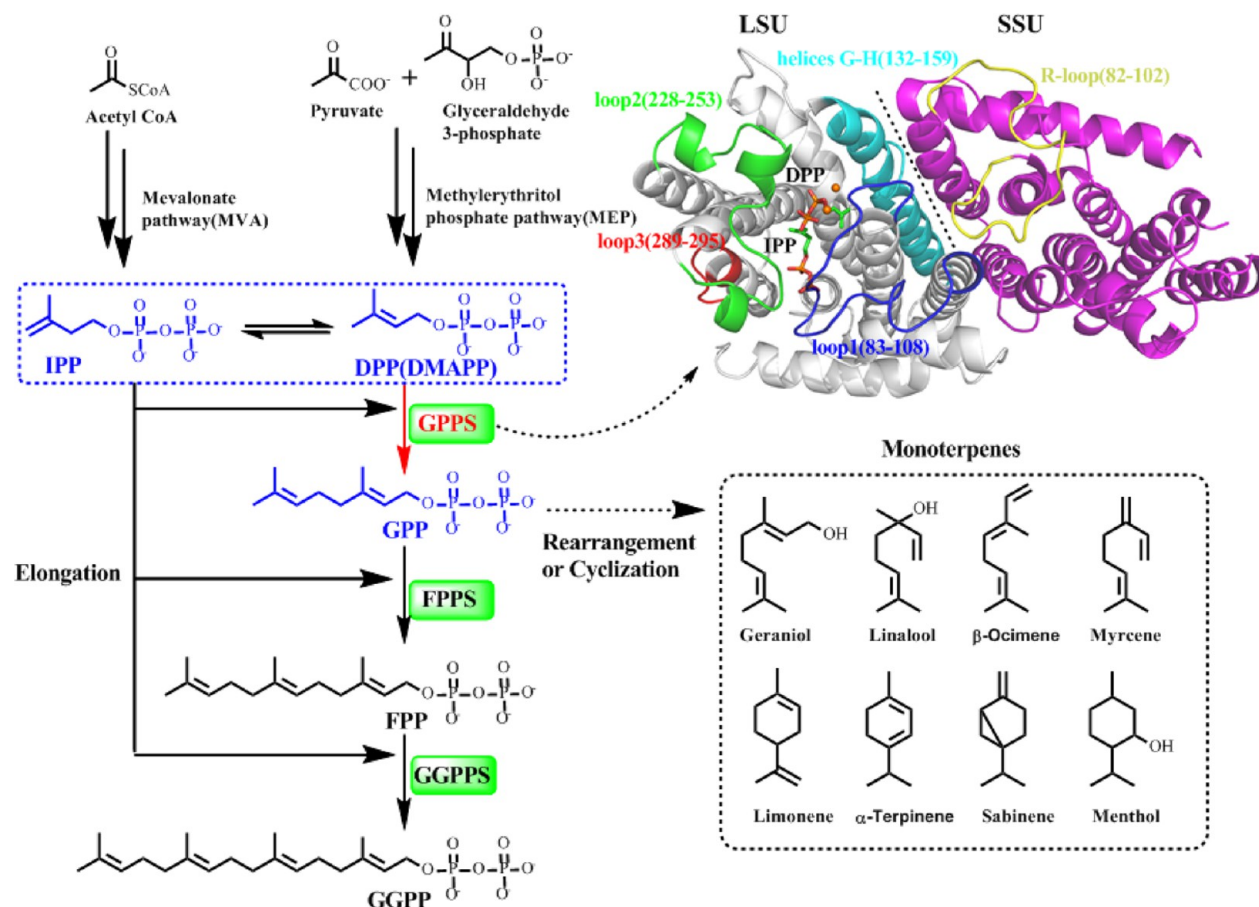
Figure 1. Flowchart for assembling isoprenoid building blocks to produce terpenes. Up right: the crystal structure of *Mp* GPPS (PDB code: 3KRO).

state and subsequent hydrogen atom transfer.[20] Merz and colleagues explored the substrate selectivity and product regioselectivity of prenyltransferase in Orf2[21] as well as the conformational activation of farnesyltransferase prior to the catalytic reaction.[22] Tantillo and co-workers performed significant work on theoretical mechanisms of terpene synthases[23,24] and proposed the reaction mechanism of MrTPS2 (a sesquiterpene synthase) based on theoretical and experimental results.[25] To explore the molecular mechanism of the enzymes in terpene biosynthesis, a number of theoretical papers regarding terpene synthase have been published.[26−30]

The GPPS family can be divided into homomeric (*Arabidopsis thaliana*, *Abies grandis*, *Vitis venifera*, *Quercus robur*, and *Citrus sinensis*) and heteromeric proteins (*Humulus lupulus*, *Mentha piperita*, *Clarkia breweri*, and *Antirrhinum majus*).[31−34] The heteromeric *Mentha piperita* GPPS (Mp GPPS) is the only enzyme for which crystal structure data (pdb code: 3KRO) are available;[14] these data demonstrate that GPPS consists of a large subunit (LSU) and a small subunit (SSU). The catalytic site of LSU is surrounded by three loop motifs denoted as loop 1, loop 2, and loop 3. Loop 1 (83−108) contains a $DD(X)_{2−4}D$ motif (X stands for any residue), which is highly conserved in the PTS superfamily and thought to be essential for substrate and/or cofactor binding. Loop 2 (228−253) likely serves as a shutter for the allylic substrate. Loop 3 (289−295) is closely related with IPP binding.[14] The active site is not identified in SSU, which has no $DD(X)_{2−4}D$ motif.[14] It is interesting that Mp GPPS will yield GPP, while producing produce considerable amounts of GGPP and negligible

amounts of FPP in vitro, presenting an unusual bifunctional biosynthetic reactivity. This behavior is not observed in homomeric GPPS.[14] A functional hybrid heterodimer, which is coexpressed between the SSU of Mp GPPS and the GGPPS (a homomeric PTS) from *Taxus canadensis* and *Abies grandis*, yields GPP (C_{10}) instead of $GGPP(C_{20})$ in vitro.[35] GPPS enhances the production of natural products by tobacco (*Nicotiana tabacum*) plants by incorporating the expression of snapdragon (*Antirrhinum majus*) GPPS.SSU.[36] Site-directed mutagenesis indicates that the R-loop (see Figure 1) of SSU plays an important role in the product release.[37] Thus, the role of SSU is of great interest to many research groups.

Although the crystal structure of *Mp* GPPS is available, the mechanism of assembly of the initial isoprenoid building blocks remains unknown. Understanding this mechanism may lead to the development of new technology to synthesize novel natural product-like molecular scaffolds and libraries, which are crucial in the drug lead identification process.[38] Therefore, we performed molecular mechanics molecular dynamics (MM MD) and hybrid quantum mechanics/molecular mechanics (QM/MM) MD simulations to elucidate the structural motifs for substrate binding in geranyl diphosphate synthases and probe the substrate binding modes and bifunctional reactivity of *Mp* GPPS in vitro.

■ METHODS

2.1. Model Preparation. All GPPS crystal structures are listed in Supporting Information Table S1. The crystal structure of *Mp* GPPS (PDB code: 3KRO) was selected for studying the

assembly mechanism of isoprenoid building blocks. The LSU and SSU (chain C and D) were preserved. Several key residues at loop 2 were missed in nonsubstrate bound GPPS (3KRA, 3KRC, AB chains of 3KRF, 3KRO, 3KRP). The sulfur atoms in the allylic substrates were replaced by oxygen atoms and cocrystallized water molecules were kept. Based on the above-mentioned structure, five models were constructed: the first model without substrates was denoted as GPPS-apo; the second model was denoted as GPPS-DPP-IPP with the substrates DPP (C_5) and IPP; the third model denoted as GPPS-GPP-IPP with the substrates GPP (C_{10}) and IPP; the fourth model denoted as GPPS-FPP-IPP with substrates FPP (C_{15}); and the fifth model was denoted as GPPS-LSU-only with only the LSU(without the substrates and SSU). The details of these models are summarized in Table 1. All the models were

Table 1. Computational Models, Their Components, and Simulation Time

model name	protein	ligands	simulation time (ns)
GPPS-apo	GPPS(LSU.SSU)		200
GPPS-DPP+IPP	GPPS(LSU.SSU)	DPP, IPP	100
GPPS-GPP+IPP	GPPS(LSU.SSU)	GPP, IPP	100
GPPS-FPP+IPP	GPPS(LSU.SSU)	FPP, IPP	100
GPPS-LSU-only	GPPS.LSU		200

then modeled using the ff99SB force field[39] and general AMBER force field (gaff) force field[40] for protein and substrates, respectively. Missing protein hydrogen atoms were added using the program LEaP[41] embedded in AMBER12. The partial atomic charges and missing force-field parameters of these ligands were obtained from the restrained electrostatic potential (RESP) charge based on HF/6-31G* calculation with the Gaussian09 package[42] and antechamber suite.[43] The system was then solvated in a TIP3P water box with a minimum distance of 8 Å between any protein atom and the edge of the box. The system was neutralized by adding of Na+ counterions. The amino acid side chain protonation states were assigned using default AMBER protonation states. The protonation

states of the key residues in the catalytic pocket were checked manually. Asp83 and Asp89 were deprotonated to generate negative charge centers for the formation of coordination bonds with magnesium, and His76 was protonated at the epsilon position to form hydrogen bonds with the oxygen of IPP.

2.2. Classical Molecular Dynamics (MD) Simulations and Trajectory Analysis. The MM MD simulations were carried out using the AMBER12 molecular simulation package.[41] First, 4000 cycles of minimization (2000 cycles of steepest descent and 2000 cycles of conjugate gradient) were carried out to relax the solvent, while all protein and substrates atoms were constrained by a potential of 3000 kcal/mol·Å². Second, another minimization stage was conducted with the protein backbone and heavy atoms of the substrates constrained (500 kcal/mol·Å²). Notably, the conformations of GPP and FPP were minimized in the pocket with the protein constrained. Each complex was then submitted to 4000 cycles of energy minimization without any constraint (2000 cycles of steepest descent and 2000 cycles of conjugate gradient). Subsequently, each system was gradually heated from 0 to 300 K over a period of 50 ps, followed by another 100 ps of NPT MD simulations at 300 K. Afterward, at least 100 ns of NVT MD simulation with a target temperature of 300 K were performed for each system to produce trajectories. For the nonsubstrate-bound models, additional 100 ns MD simulations were performed to achieve the final equilibrium (Table 1). The SHAKE algorithm[44] was employed to constrain the bonds involving hydrogen atoms, and a time step of 2.0 fs was used for all simulations.

Cluster analysis is a general unsupervised technique for finding patterns within data.[45−47] Root mean-square deviation (RMSD)-based clustering was performed with ptraj, a simulation analysis tools implemented in AMBER12. As one of the most popular clustering algorithms, the average linkage cluster algorithm was used herein.[48] Each cluster contained a representative structure whose RMSD was equidistant to all other cluster members. Structures were collected by sampling at 200 ps intervals, and cluster analysis was based on the heavy atoms of loop 2 or helices G-H to generate a sum of five clusters, as described in the next sections.
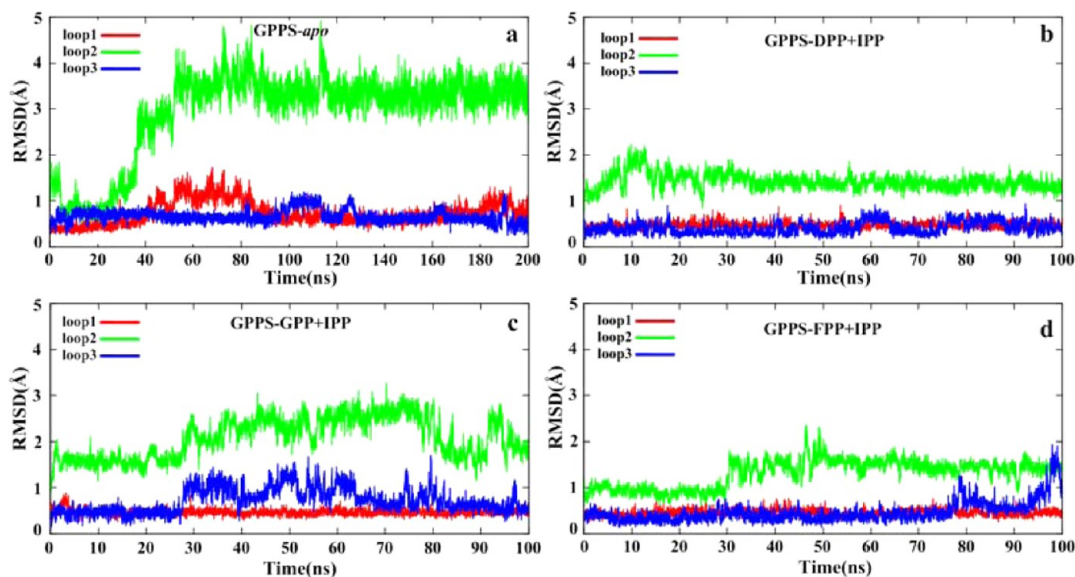

Figure 2. Flexibility of the loops 1, 2, and 3 of *Mp* GPPS with different complexes (with or without substrate).
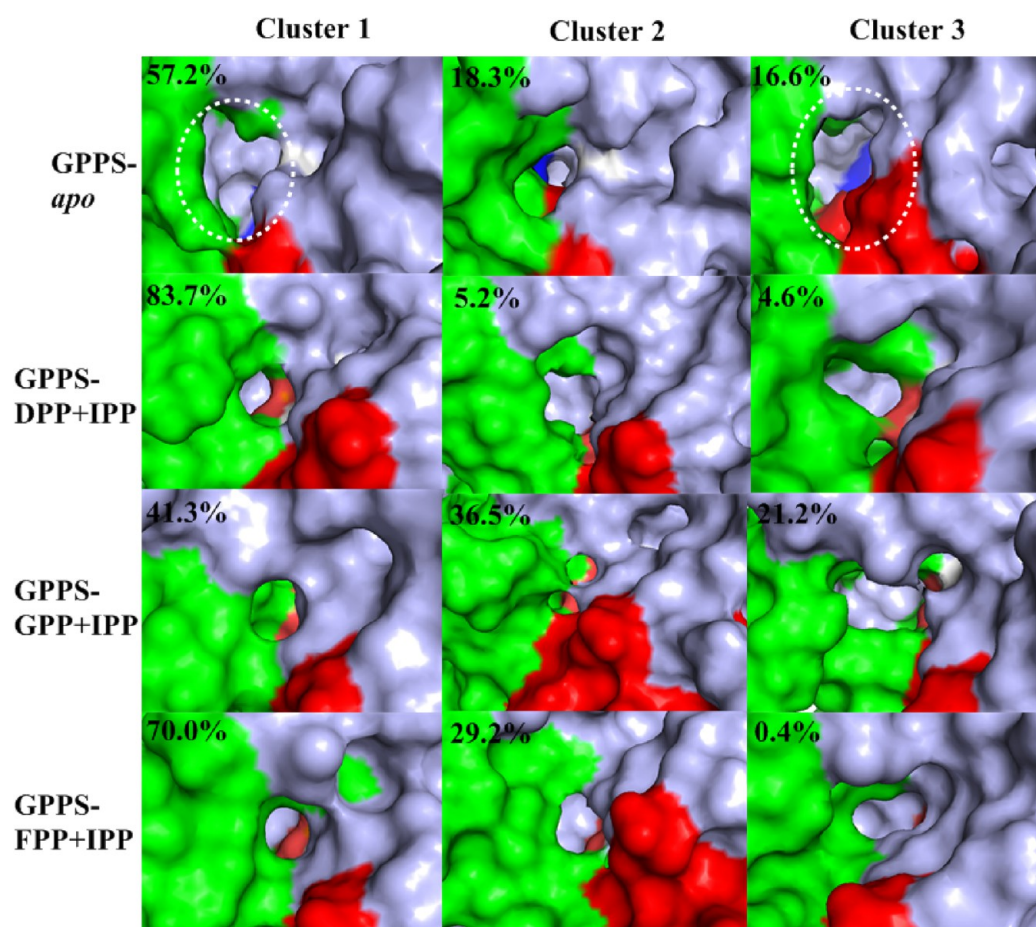
Figure 3. Surface comparison of the active pocket in each model from cluster analysis (loop 1, red; loop 2, green; loop 3, blue). In GPPS-*apo* model, the open conformations of pocket were observed in cluster 1 and 3 (outlined with a dotted circle), while in substrate binding models (GPPS-DPP +IPP, GPPS-GPP+IPP, and GPPS-FPP+IPP), the pockets were all in close state.
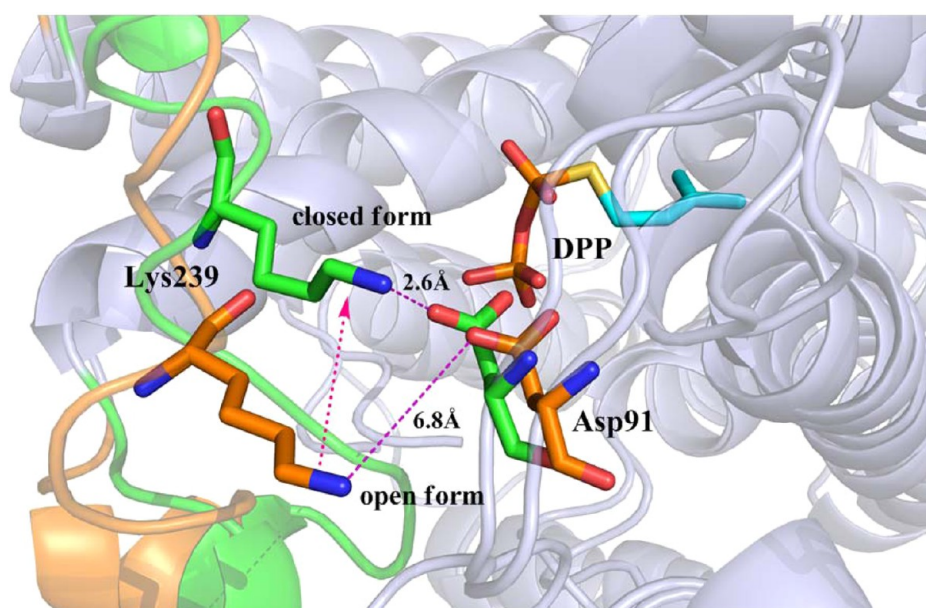

Figure 4. Superposition of the representative open and closed active site structures. Open: orange, refer to cluster 1 of model GPPS-*apo* in Figure 3. Closed: green, refer to cluster 2 of model GPPS-*apo* in Figure 3. The negative charged diphosphate in the allylic substrate DPP would induce the motion of Lys239 (positive charge) to form salt bridge with Asp91, resulting in the closing of the active pocket.

The volume of the substrate binding pocket was measured using POVME, an algorithm for measuring pocket volume.[49]

Structures were collected by sampling at 50 ps intervals and aligned to the initial crystal structure (CD chain of 3KRO). An
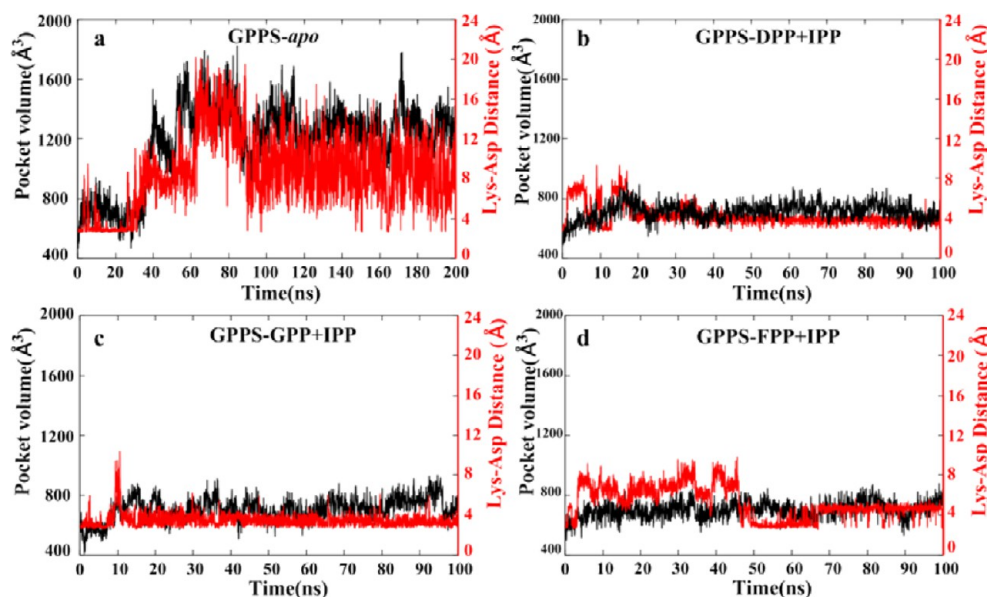
Figure 5. Evolution of the active pocket volume (black, left Y-axis) and the Asp91-Lys239 salt-bridge (red, right Y-axis) along the MM MD trajectories in the selected models. In GPPS-*apo* model (a), the distance change of the salt-bridge closely related with the pocket volume change. In substrate binding models (b, c, and d), the volume and salt bridge remain relatively stable.

inclusion sphere with a radius of 12 Å was defined that entirely encompassed the binding pocket. A written shell script was used to automatically calculate the volumes of all collected structures automatically. All plots were generated with Gnuplot, and molecular figures were created using PyMOL.[50]

2.3. Born−Oppenheimer Ab Initio QM/MM MD Simulations. Three substrate binding system (GPPS-DPP+IPP, GPPS-GPP+IPP, and GPPS-FPP+IPP) from the above MD simulations were cut into a sphere by removing the solvent water molecules beyond 30 Å of the C1′ atom (as shown in Supporting Information Figure S1) of DPP/GPP/FPP in the active site. The resulting systems were then partitioned into QM and MM subsystems. The DPP/GPP/FPP, IPP, D83, D89, K180, and Mg^{2+} were chosen as the QM subsystem (Figure S1), which was treated by B3LYP[51,52] with the 6-31+G* basis set for the oxygen atoms of DPP/GPP/FPP/IPP and the 6-31G* basis set for all other QM atoms. The QM/MM boundaries were described by the pseudobond approach[53−56] with improved pseudobond parameters. All the remaining atoms were described by the same MM force field used in the previous classical MD simulations. For all QM/MM calculations, the spherical boundary condition[53,56] was applied, and atoms more than 22 Å away from the C1′ atom of DPP/GPP/FPP were fixed. Cutoffs of 18 and 12 Å were employed for the electrostatic and van der Waals interactions, respectively. There was no cutoff for the electrostatic interactions between QM and MM regions. After minimization of the prepared QM/MM systems, 25 ps QM/MM MD simulations were performed using a time step of 1 fs and the Beeman algorithm[57] to integrate the Newton equations of motion. The Berendsen thermostat method[58] controlled the system temperature at 300 K. The configurations of the last 20 ps were collected for data analysis. All ab initio QM/MM calculations were performed in modified QChem-Tinker programs.[59,60]
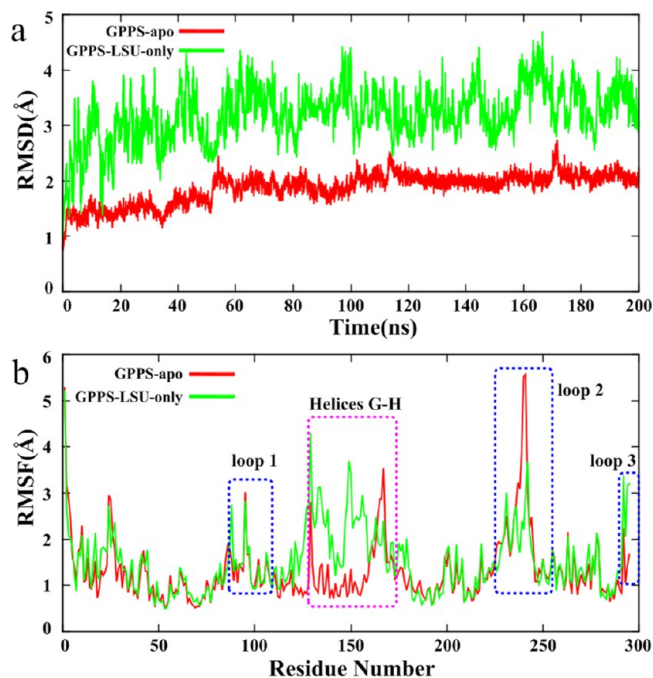

Figure 6. RMSD (a) and RMSF (b) analysis of LSU in model GPPS-*apo* and GPPS-LSU-only. The residue numbers of loops 1, 2, 3 and helices G-H are identified and highlighted.

■ RESULTS AND DISCUSSION

3.1. Opening and Closing of the Active Pocket Is Modulated by the Asp91-Lys239 Salt Bridge. As shown in Supporting Information Figure S2, the RMSD of each system is reasonable and achieves equilibrium within the simulation time, indicating that the force field and simulation protocol used here are adequate for investigating the current systems. The GPPS-DPP+IPP, GPPS-GPP+IPP, and GPPS-FPP+IPP models achieve equilibrium at 20 ns, while the GPPS-*apo* model equilibrates at 100 ns. This phenomenon is due to the more flexible active pocket in the *apo*-state GPPS. To identify the
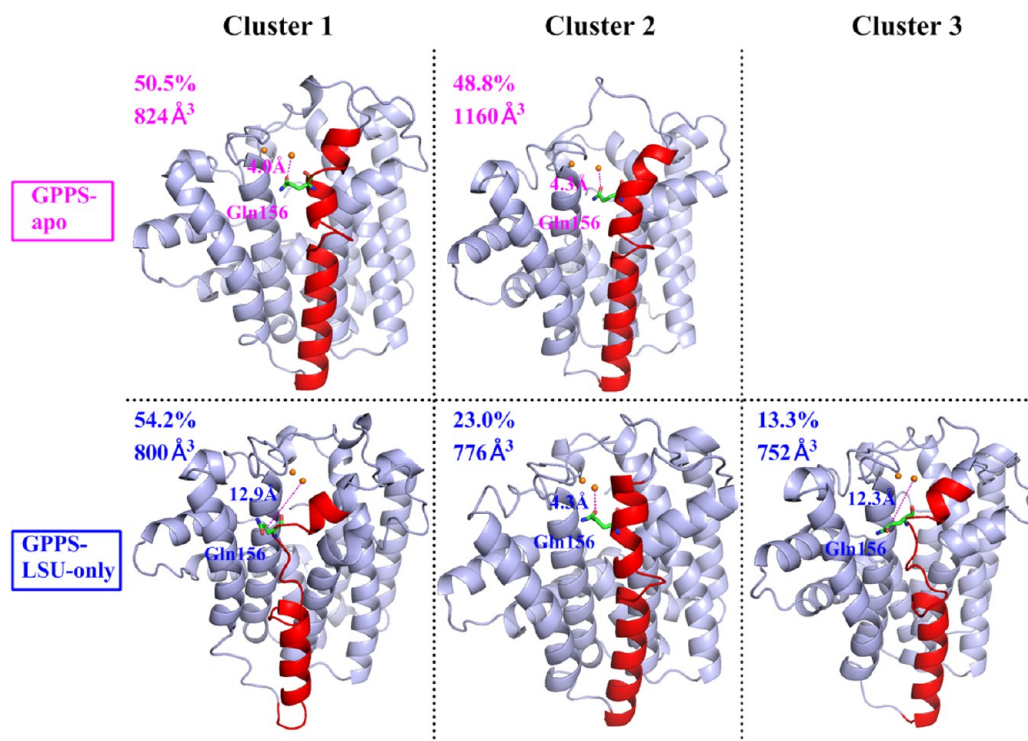
Figure 7. Cluster analysis for the LSU in apo-state (SSU included) and without-SSU state. The helices G-H and Mg ions are shown as red and orange, respectively. It indicates that the secondary structure of helices G-H is partially destroyed in cluster 1 and cluster 3. The population, pocket volume, residue Gln156 and its distance to the closest Mg ion were highlight in each structure. There are only two clusters for GPPS-*apo* model with a very small population for cluster 3 (0.7%).

detailed structural motif that confers the catalytic active pocket flexibility, the RMSDs of loops 1, 2, and 3 were also calculated (Figure 2). Both loop 1 and 3 are very stable during the simulations in each model, while loop 2 is stable in the substrate (DPP/GPP/FPP+IPP) binding models. However, a dramatic conformational change of loop 2 in the GPPS-*apo* model was observed after 40 ns (Figure 2), indicating that the active pocket in the *apo*-state is different from that in the binding state by the essential modulator loop 2.

To further elucidate the relationship between the conformation change of loop 2 and the active pocket state of LSU in GPPS, cluster analysis based on loop 2 residues was performed. The binding pockets of the representative structures for the top three most dominant clusters are depicted in Figure 3. The binding pockets are apparently in closed form in the substrate (DPP/GPP/FPP+IPP) binding complexes, while they are mostly open in the *apo*-GPPS with high population (57.2% of cluster 1 and 16.6% of cluster 3, with only 18.3% in the closed form). Therefore, the open-close transformation of the binding pocket is directly related to the conformational change of loop 2; that is, the loop 2 is likely to participate in the substrate binding kinetics. This has also been shown directly by several previous experimental results. In the *Mp* GPPS crystal structure data (see Supporting Information Table S1), the loop 2 structure cannot be observed without allylic substrate binding. A structural comparison between the opened and closed LSU indicates that loop 2 is flexible.[14] Herein, our MD simulations reveal significant conformational changes in loop 2 that can be induced by the allylic substrate. Hence, loop 2 is responsible for opening and closing the catalytic pocket.

Nevertheless, the mechanism on the conformational change of loop 2 at the atomic level is still unclear. The structural

alignment of the open and closed form are illustrated in Figure 4. A 2.6-Å salt bridge between Asp91 of loop 1 and Lys239 of loop 2 is observed in the closed state. The salt bridge is broken when it is stretched to 6.8 Å in the open state. This conclusion is further validated by checking the evolution of the salt bridge distance and pocket volume along the trajectory (see Figure 5). In the GPPS-*apo* model, the salt bridge distance and the pocket volume increased at 30 ns, and they reach a stable state after 90 ns. An shown in Figure 5a, a similar trend in salt bridge distance and pocket volume was observed. In contrast, the salt-bridge length and pocket volume of the substrates (closed forms) remain at 4 and 800 Å³, respectively. Based on the above-mentioned results, we conclude that the salt bridge controls the pocket status (open or closed). A similar active pocket open-close mechanism also exits in influenza neuraminidase.[61]

The locations of the allylic substrates and distances between diphosphate and Lys239 were identified by aligning the substrate binding models (GPPS-DPP+IPP, GPPS-GPP+IPP, and GPPS-FPP+IPP); details can be found in Supporting Information Figure S3. The diphosphate of the allylic substrates points to the outside of the pocket, while the carbon chain points toward the inside of the pocket. This suggests that the carbon chain of the allylic substrates enters the pocket as the head, while the diphosphate of the allylic substrates is the tail. The distances between the diphosphate of the allylic substrates and Lys239 indicate electrostatic interaction between the allylic substrates and GPPS. Based on the above information, we proposed the probable mechanism for the entry of the allylic substrates (Figure 4). The carbon chain of the allylic substrate enters the pocket as the head. The negatively charged pyrophosphate of the allylic substrate acts as the tail and subsequently attracts the NH_3^+ group of the loop 2 Lys239
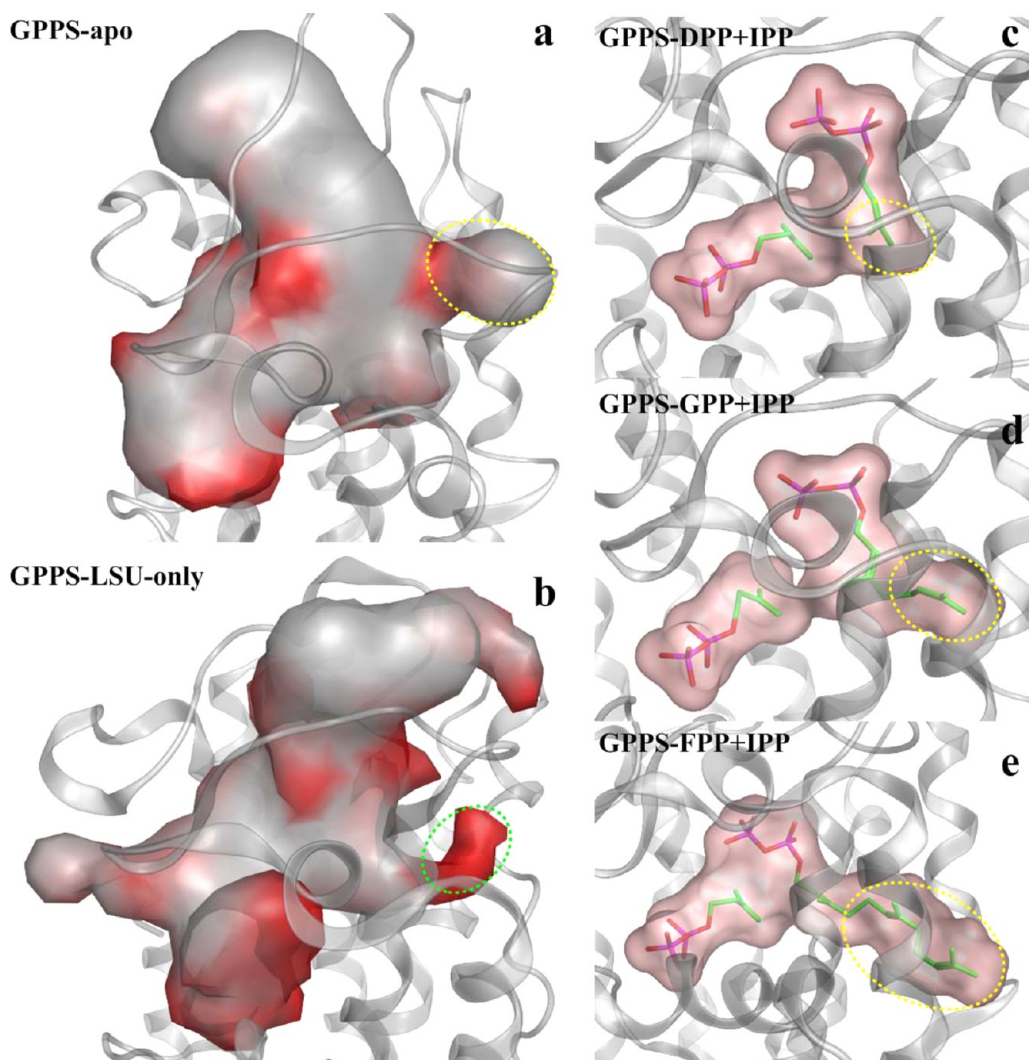
Figure 8. Comparison of the binding pocket in the selected models. The cavity becomes much smaller in model GPPS-LSU-only (b, green dashed circle). The DPP/GPP/FPP substrate binding cavity (namely, the elongation cavity) (c, d, and e) is highlighted by a yellow dashed circle.

through electrostatic interaction. This leads to the inward motion of Lys239 to form a salt bridge with Asp91 of loop 1. The stable salt bridge maintains the closed state of the pocket to avoid exposure to solvent for the subsequent catalytic reaction.

3.2. Active Pocket Hydrophobicity and Size/Shape Is Regulated by SSU. To investigate the role of SSU in substrate binding, a GPPS-LSU-only model was constructed without SSU. As shown in Figure 6, the RMSD of LSU presents a large fluctuation (3.20 ± 0.49 Å) in the GPPS-LSU-only model due to the absence of SSU, while it is stable (1.87 ± 0.27 Å) in the GPPS-*apo* model due to the presence of SSU. This indicates that the SSU stabilizes the LSU to some extent. RMSF analysis demonstrates that the values for each residue of the two models are in agreement, with the exception of helices G-H (Pro132-Gly151). This indicates that the local conformation change of helices G-H occurs spontaneously without the SSU. Particularly, helix H is close to the catalytic pocket and interacts with the Mg-bound water through hydrogen bonding, thus playing a major role in substrate binding; this phenomenon has been demonstrated in a previous biological study, which showed that individually expressed LSU from *Mentha piperita* shows no catalytic activity.[14] This is further confirmed by the cluster

analysis shown in Figure 7, which demonstrates that the secondary structure of helices G-H retain a large pocket volume (average $\sim$990 Å³) during the MD simulations when SSU is present (50.5% for cluster 1 and 48.8% for cluster 2). However, the secondary structure of helices G-H is partially destroyed during the MD simulations when SSU is absent, exhibiting a smaller pocket volume (average $\sim$785 Å³).

Further structural analysis indicates that the side-chain of Gln156 points inward toward the binding pocket when SSU is present, while facing outward when SSU is absent (Figure 7). As shown in Supporting Information Figure S4, hydrogen bonds are formed among Gln156, Asp83, and a water molecule. These hydrogen bonds are crucial for the substrate binding and magnesium coordination. Meanwhile, the orientation is somewhat changed for classical MD simulation because the accuracy of coordination interactions is not as good as in the state-of-the-art ab initio QM/MM method; the QM/MM calculation can accurately reproduce the coordination shell and hydrogen bond network around Mg^{2+}, as evidenced by comparison with the crystal structure. Therefore, we chose the QM/MM method to further identify the substrate binding modes (vide infra, see Section 3.3). The justification for using B3LYP was verified by the benchmark test on B3LYP/6-31G*, B3LYP/6-
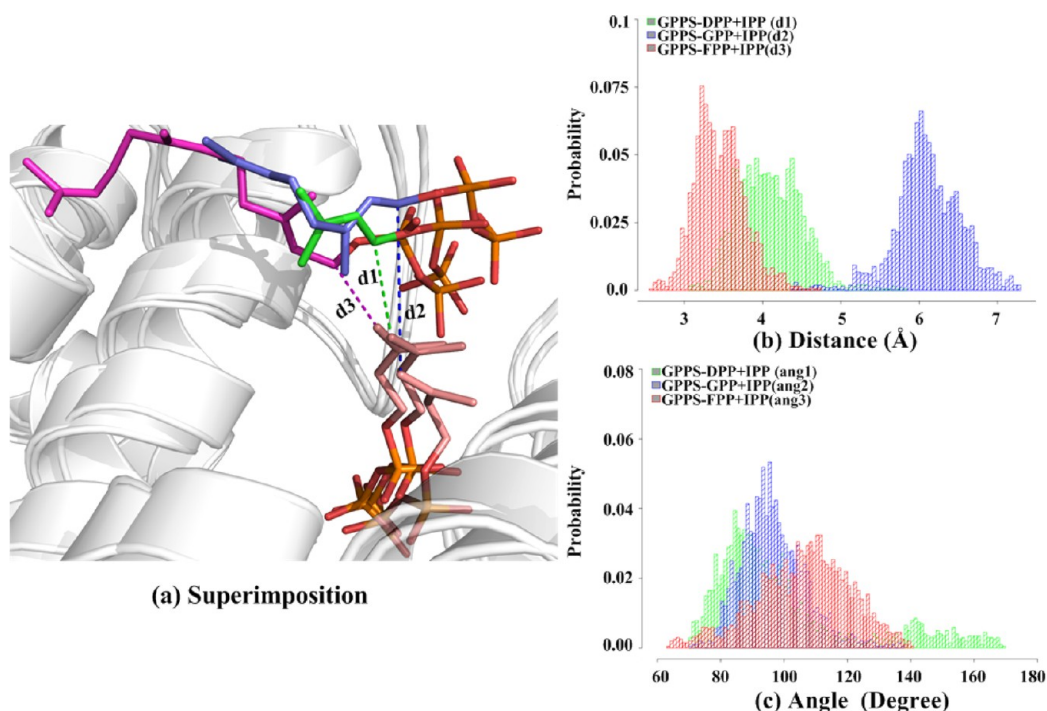
Figure 9. (a) Superimposition of the substrates binding models (GPPS-DPP+IPP, GPPS-GPP+IPP, and GPPS-FPP+IPP). (b) Distributions of the electrophilic attack distance (d1/d2/d3) in each model from ab initio QM/MM MD simulation.(c) Distributions of the electrophilic attack angle (ang1/ang2/ang3) in each model from ab initio QM/MM MD simulation. Herein, ang1, ang2, and ang3 mean the angle of C4–C1′–O7′, which refer to Supporting Information Figure S1.

311G*, and M06-2X/6-31G* (see Supporting Information Figure S5).

To explain the inward and outward orientations of Gln56, the average number of water in the pocket is summarized in Supporting Information Figure S6. These numbers suggest that the solvent molecules move into the pocket and destroy the hydrogen bond network around the substrate binding site, since the active pocket is exposed to the solvent water without the protection of SSU. The hydrophilic side chain of Gln156 is then solvated by water molecules. Indeed, Gln156 is conserved in the PTS family based on sequence alignment analysis (Supporting Information Figure S7). The substrate binding cavity for each model is depicted in Figure 8. Apparently, the elongation cavity as well as its flexibility are observed to accommodate the different length of substrates when SSU is present. The cavity becomes much smaller when SSU is absent. Thus, we conclude that the size and shape of the catalytic pocket are regulated by the SSU. The role of SSU is to stabilize the LSU by protecting the structure of helices G-H, maintaining the inward orientation of Gln56, and adjusting the shape or size of the active cavity for substrate binding. This mechanism is in agreement with the previous structural analysis of hexaprenyl diphosphate synthase (a C_{30} synthase), another heteromeric type of trans-PTS; the SSU of this enzyme was found to stabilize the LSU through hydrophobic interaction and to directly regulate the product chain length.[62]

3.3. Bifunctional Reactivity Is Related to the Substrate Binding Modes. To precisely treat with the metal coordination interaction, QM/MM MD were performed to construct the substrate binding modes. As shown in Supporting Information Figure S4, the GPPS-DPP+IPP model contains an Mg^{2+} with three coordination bonds to water molecules, two coordination bonds to Asp83 and Asp89, and one coordination

bonds to DPP, along with another Mg^{2+} with two coordination bonds to water molecules, two coordination bonds to Asp83 and Asp89, and two coordination bonds to DPP. The distance between the two Mg^{2+} ions is 3.4 ± 0.1 Å (close to the value measured in the crystal structure). Since the nature of elongation reaction between the building block (IPP and DPP/GPP/FPP) is SN_2 or SN_1 nucleophilic attack reaction, the distance between IPP and DPP/GPP/FPP (C4–C1′) and the C4–C1′–O7′ angle are two key factors (Figure 9). The distance of the nucleophilic attack for the GPPS-GPP+IPP system is 6.10 ± 0.44 Å, which is greater than the distance for the GPPS-DPP+IPP system (4.10 ± 0.41 Å) or GPPS-FPP +IPP system (3.46 ± 0.32 Å; Figure 9). *Mp* GPPS shows bifunctional reactivity and yields GPP, GGPP, and negligible FPP in vitro. Studies of the reaction kinetics show that the k_{cat}/K_m of the second step is 1000-fold lower than that of the first step and 31-fold lower than that of the third step. These results suggest that the second step (GPPS-GPP+IPP to generate FPP) is the slowest among the three elongation steps.[14] The nucleophilic attack distances are consistent with the kinetic data, indicating that the lowest reactivity corresponds to the assembly of GPP to produce FPP catalyzed by *Mp* GPPS, which can catalyze the assembly of both DPP and FPP in vitro to yield GPP and GGPP.[14]

The nucleophilic attack angle and crucial atomic charges were also calculated. The C4–C1′-O7′ angle is approximately 90° for GPP and 115° for FPP. The C4–C1′–O7′ angle is approximately 80° or 150° for DPP (Figure 9). In the GPP +IPP system (Supporting Information Figure S8), IPP-C4 is more negatively charged, GPP-C1′ is more positively charged, and GPP-O7′ is less negatively charged. This probably due to the large distance between C4 and C1′ (~6.1 Å), resulting in negligible charge transfer between GPP and IPP. In the FPP

+IPP system, IPP-C4 is negatively charged, GPP-C1′ is almost neutral, and GPP-O7′ is negatively charged. This is most likely due to the IPP attack and C1′-O7′ breakage caused by the charge transfer between FPP and IPP. In the DPP+IPP system, the distributions of the C4−C1′−O7′ angle and the charge on C1′ or C7′ are very dispersed for nucleophilic attack. Nevertheless, structural based mechanistic arguments are arbitrary and the definite mechanism will be determined by further QM/MM MD simulations, which will be reported later.

Substrates binding analysis revealed the key residues (Lys44, Arg47, His76, and Arg95 for IPP binding; Arg94, Lys180, and Lys235 for DPP binding) for substrate binding. All these residues are positively charged basic amino acids that have strong electrostatic interactions with the diphosphate of the substrates. Furthermore, our simulations suggest that the Asp91 and Lys239 residues are crucial for allylic substrate entry and that the conserved residue Gln156 is important for providing hydrophobic environment for catalytic reaction. Although they are not directly related to substrate binding, these residues play important roles in the catalytic reactions. Mutation analyses were performed and further confirmed the crucial role of the residues. Details on mutation analysis can be seen in the Supporting Information Figures S9 and S10. Thus, these residues are good candidates for site-directed mutagenesis in protein engineering.

CONCLUSIONS

In this study, classic and hybrid QM/MM MD were employed to investigate the flexibility of the catalytic pocket and the role of SSU in *Mp* GPPS. Our results suggest that the salt bridge between Asp91 and Lys239 controls the state of the pocket (open or closed) and that the allylic substrate induces the salt bridge formation. This mechanism can explain how the allylic substrate enters the pocket. We also reveal that the SSU protects the catalytic pocket through maintaining hydrophobic environment and modifying the pocket size or shape to accommodate different size of substrates ($C_5/C_{10}/C_{15}$) to be assembled into precursors of terpene products. Finally, the distinct binding modes of different substrates are revealed by hybrid QM/MM MD simulations. Our simulations suggest that Asp91 and Lys239 are important for substrate entry, while Gln156 is important for providing a hydrophobic environment for the catalytic reaction. These residues are good candidates for site-directed mutagenesis and may help in protein engineering.

ASSOCIATED CONTENT

Ⓢ Supporting Information

Figure S1. Detailed QM/MM partition schemes for ab initio QM/MM MD simulations. Figure S2. RMSD analyses of LSU, SSU chains, and the whole system. Figure S3. Alignment of substrates binding models. Figure S4. The comparison of the binding pocket. Figure S5. The benchmark test on B3LYP/6-31G*, B3LYP/6-311G*, and M06-2*X*/6- 31G*. Figure S6. Comparative analysis of the number of water molecules in binding pocket for GPPS-*apo* and GPPS-LSU-only models. Figure S7. Multiple sequence alignments of GPPS with FPPS and GGPPS. Figure S8. The ESP charges distributions of the selected atoms in each model from ab initio QM/MM MD simulation. Figure S9. Mutation analysis of the salt bridge. Figure S10. Mutation analysis of the Gln156. Table S1. The crystal structures of wild-type GPPS from *Mentha piperita*. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors
*Tel: +86-020-39943023. E-mail: xujun9@mail.sysu.edu.cn.
*Tel: +86-020-39943074. E-mail: wurb3@mail.sysu.edu.cn.
Author Contributions
†Liu and Zhou contributed equally to this work.
Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS

REFERENCES

(1) Gershenzon, J.; Dudareva, N. The function of terpene natural products in the natural world. *Nat. Chem. Biol.* **2007**, *3*, 408−414.
(2) Oldfield, E.; Lin, F. Y. Terpene biosynthesis: Modularity rules. *Angew. Chem., Int. Ed.* **2012**, *51*, 1124−1137.
(3) Gao, Y.; Honzatko, R. B.; Peters, R. J. Terpenoid synthase structures: A so far incomplete view of complex catalysis. *Nat. Prod. Rep.* **2012**, *29*, 1153−1175.
(4) Vandermoten, S.; Haubruge, E.; Cusson, M. New insights into short-chain prenyltransferases: Structural features, evolutionary history, and potential for selective inhibition. *Cell. Mol. Life Sci.* **2009**, *66*, 3685−3695.
(5) Miziorko, H. M. Enzymes of the mevalonate pathway of isoprenoid biosynthesis. *Arch. Biochem. Biophys.* **2011**, *505*, 131−143.
(6) Bouvier, F.; Rahier, A.; Camara, B. Biogenesis, molecular regulation and function of plant isoprenoids. *Prog. Lipid Res.* **2005**, *44*, 357−429.
(7) Kumari, S.; Priya, P.; Misra, G.; Yadav, G. Structural and biochemical perspectives in plant isoprenoid biosynthesis. *Phytochem. Rev.* **2013**, *12*, 255−291.
(8) Vinokur, J. M.; Korman, T. P.; Cao, Z.; Bowie, J. U. Evidence of a novel mevalonate pathway in Archaea. *Biochemistry* **2014**, *53*, 4161−4168.
(9) Korman, T. P.; Sahachartsiri, B.; Li, D.; Vinokur, J. M.; Eisenberg, D.; Bowie, J. U. A synthetic biochemistry system for the in vitro production of isoprene from glycolysis intermediates. *Protein Sci.* **2014**, *23*, 576−585.
(10) Rohmer, M. From molecular fossils of bacterial hopanoids to the formation of isoprene units: Discovery and elucidation of the methylerythritol phosphate pathway. *Lipids* **2008**, *43*, 1095−1107.
(11) Kellogg, B. A.; Poulter, C. D. Chain elongation in the isoprenoid biosynthetic pathway. *Curr. Opin. Chem. Biol.* **1997**, *1*, 570−578.
(12) Ogura, K.; Koyama, T. Enzymatic aspects of isoprenoid chain elongation. *Chem. Rev.* **1998**, *98*, 1263−1276.
(13) Vranova, E.; Coman, D.; Gruissem, W. Structure and dynamics of the isoprenoid pathway network. *Mol. Plant* **2012**, *5*, 318−333.
(14) Chang, T. H.; Hsieh, F. L.; Ko, T. P.; Teng, K. H.; Liang, P. H.; Wang, A. H. Structure of a heterotetrameric geranyl pyrophosphate synthase from mint (*Mentha piperita*) reveals intersubunit regulation. *Plant Cell* **2010**, *22*, 454−467.

(15) Dale Poulter, C.; R, H. C. The prenyl transfer reaction. Enzymic and mechanistic studies of the 1′-4 coupling reaction in the terpene biosynthetic pathway. *Acc. Chem. Res.* **1978**, *11*, 307−313.

(16) Poulter, C. D.; Satterwhite, D. M. Mechanism of the prenyl-transfer reaction. Studies with (E)- and (Z)-3-trifluoromethyl-2-buten-1-yl pyrophosphate. *Biochemistry* **1977**, *16*, 5470−5478.

(17) Fujihashi, M.; Zhang, Y. W.; Higuchi, Y.; Li, X. Y.; Koyama, T.; Miki, K. Crystal structure of cis-prenyl chain elongating enzyme, undecaprenyl diphosphate synthase. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 4337−4342.

(18) Kharel, Y.; Koyama, T. Molecular analysis of cis-prenyl chain elongating enzymes. *Nat. Prod. Rep.* **2003**, *20*, 111−118.

(19) Liang, P. H. Reaction kinetics, catalytic mechanisms, conformational changes, and inhibitor design for prenyltransferases. *Biochemistry* **2009**, *48*, 6562−6570.

(20) Sanchez, V. M.; Crespo, A.; Gutkind, J. S.; Turjanski, A. G. Investigation of the catalytic mechanism of farnesyl pyrophosphate synthase by computer simulation. *J. Phys. Chem. B* **2006**, *110*, 18052−18057.

(21) Cui, G.; Li, X.; Merz, K. M., Jr. Understanding the substrate selectivity and the product regioselectivity of Orf2-catalyzed aromatic prenylations. *Biochemistry* **2007**, *46*, 1303−1311.

(22) Cui, G.; Merz, K. M., Jr. Computational studies of the farnesyltransferase ternary complex. Part II: The conformational activation of farnesyldiphosphate. *Biochemistry* **2007**, *46*, 12375−12381.

(23) Pemberton, R. P.; Tantillo, D. J. Lifetimes of carbocations encountered along reaction coordinates for terpene formation. *Chem. Sci.* **2014**, *5*, 3301−3308.

(24) Tantillo, D. J. Biosynthesis via carbocations: Theoretical studies on terpene formation. *Nat. Prod. Rep.* **2011**, *28*, 1035−1053.

(25) Hong, Y. J.; Irmisch, S.; Wang, S. C.; Garms, S.; Gershenzon, J.; Zu, L. S.; Kollner, T. G.; Tantillo, D. J. Theoretical and experimental analysis of the reaction mechanism of MrTPS2, a triquinane-forming sesquiterpene synthase from chamomile. *Chem.—Eur. J.* **2013**, *19*, 13590−13600.

(26) van der Kamp, M. W.; Mulholland, A. J. Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. *Biochemistry* **2013**, *52*, 2708−2728.

(27) van der Kamp, M. W.; Sirirak, J.; Zurek, J.; Allemann, R. K.; Mulholland, A. J. Conformational change and ligand binding in the aristolochene synthase catalytic cycle. *Biochemistry* **2013**, *52*, 8094−8105.

(28) Major, D. T.; Freud, Y.; Weitman, M. Catalytic control in terpenoid cyclases: Multiscale modeling of thermodynamic, kinetic, and dynamic effects. *Curr. Opin. Chem. Biol.* **2014**, *21C*, 25−33.

(29) Major, D. T.; Weitman, M. Electrostatically guided dynamics—The root of fidelity in a promiscuous terpene synthase? *J. Am. Chem. Soc.* **2012**, *134*, 19454−19462.

(30) Rajamani, R.; Gao, J. Balancing kinetic and thermodynamic control: The mechanism of carbocation cyclization by squalene cyclase. *J. Am. Chem. Soc.* **2003**, *125*, 12768−12781.

(31) Burke, C. C.; Wildung, M. R.; Croteau, R. Geranyl diphosphate synthase: Cloning, expression, and characterization of this prenyl-transferase as a heterodimer. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 13062−13067.

(32) Kang, J. H.; Gonzales-Vigil, E.; Matsuba, Y.; Pichersky, E.; Barry, C. S. Determination of residues responsible for substrate and product specificity of *Solanum habrochaites* short-chain cis-prenyltransferases. *Plant Physiol.* **2014**, *164*, 80−91.

(33) Burke, C.; Croteau, R. Geranyl diphosphate synthase from *Abies grandis*: cDNA isolation, functional expression, and characterization. *Arch. Biochem. Biophys.* **2002**, *405*, 130−136.

(34) van Schie, C. C. N.; Ament, K.; Schmidt, A.; Lange, T.; Haring, M. A.; Schuurink, R. C. Geranyl diphosphate synthase is required for biosynthesis of gibberellins. *Plant J.* **2007**, *52*, 752−762.

(35) Burke, C.; Croteau, R. Interaction with the small subunit of geranyl diphosphate synthase modifies the chain length specificity of geranylgeranyl diphosphate synthase to produce geranyl diphosphate. *J. Biol. Chem.* **2002**, *277*, 3141−3149.

(36) Orlova, I.; Nagegowda, D. A.; Kish, C. M.; Gutensohn, M.; Maeda, H.; Varbanova, M.; Fridman, E.; Yamaguchi, S.; Hanada, A.; Kamiya, Y.; Krichevsky, A.; Citovsky, V.; Pichersky, E.; Dudareva, N. The small subunit of snapdragon geranyl diphosphate synthase modifies the chain length specificity of tobacco geranylgeranyl diphosphate synthase in planta. *Plant Cell* **2009**, *21*, 4002−4017.

(37) Hsieh, F. L.; Chang, T. H.; Ko, T. P.; Wang, A. H. Enhanced specificity of mint geranyl pyrophosphate synthase by modifying the R-loop interactions. *J. Mol. Biol.* **2010**, *404*, 859−873.

(38) Xu, J.; Gu, Q.; Liu, H. B.; Zhou, J. J.; Bu, X. Z.; Huang, Z. S.; Lu, G.; Li, D.; Wei, D. Q.; Wang, L.; Gu, L. Q. Chemomics and drug innovation. *Sci. China: Chem.* **2013**, *56*, 71−85.

(39) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* **2006**, *65*, 712−725.

(40) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general Amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.

(41) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668−1688.

(42) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*; Gaussian. Inc: Wallingford, CT, 2009.

(43) Wang, J. M.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247−260.

(44) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.

(45) Estacio, S. G.; Moreira, R.; Guedes, R. C. Characterizing the dynamics and ligand-specific interactions in the human leukocyte elastase through molecular dynamics simulations. *J. Chem. Inf. Model.* **2011**, *51*, 1690−1702.

(46) Amaro, R. E.; Swift, R. V.; Votapka, L.; Li, W. W.; Walker, R. C.; Bush, R. M., Mechanism of 150-cavity formation in influenza neuraminidase. *Nat. Commun.* **2011**, *2*.

(47) Colombo, G.; Morra, G.; Meli, M.; Verkhivker, G. Understanding ligand-based modulation of the Hsp90 molecular chaperone dynamics at atomic resolution. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 7976−7981.

(48) Shao, J. Y.; Tanner, S. W.; Thompson, N.; Cheatham, T. E. Clustering molecular dynamics trajectories: 1. Characterizing the performance of different clustering algorithms. *J. Chem. Theory Comput.* **2007**, *3*, 2312−2334.

(49) Durrant, J. D.; de Oliveira, C. A.; McCammon, J. A. POVME: An algorithm for measuring binding-pocket volumes. *J. Mol. Graphics Modell.* **2011**, *29*, 773−776.

(50) *The PyMOL Molecular Graphics System*, Version 1.6; Schrödinger LLC: New York, 2013.

(51) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle−Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B Condens Matter* **1988**, *37*, 785−789.

(52) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098−3100.

(53) Zhang, Y. K. Pseudobond ab initio QM/MM approach and its applications to enzyme reactions. *Theor. Chem. Acc.* **2006**, *116*, 43−50.

(54) Zhang, Y. K. Improved pseudobonds for combined ab initio quantum mechanical/molecular mechanical methods. *J. Chem. Phys.* **2005**, *122*, 024114.

(55) Zhang, Y. K.; Liu, H. Y.; Yang, W. T. Free energy calculation on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined ab initio QM/MM potential energy surface. *J. Chem. Phys.* **2000**, *112*, 3483−3492.

(56) Zhang, Y. K.; Lee, T. S.; Yang, W. T. A pseudobond approach to combining quantum mechanical and molecular mechanical methods. *J. Chem. Phys.* **1999**, *110*, 46−54.

(57) Beeman, D. Some multistep methods for use in molecular dynamics calculations. *J. Comput. Phys.* **1976**, *20*, 131−139.

(58) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.

(59) Shao, Y.; Fusti-Molnar, L.; Jung, Y.; Kussmann, J. O. C.; Brown, S. T.; Gilbert, A. T.; Slipchenko, L. V.; Levchenko, S. V.; O'Neill, D. P.; DiStasio, R. A.; Lochan, R. C.; Wang, T.; Beran, G. J.; Besley, N. A.; Herbert, J. M.; Lin, C. Y.; Van Voorhis, T.; Chien, S. H.; Sodt, A.; Steele, R. P.; Rassolov, V. A.; Maslen, P. E.; Korambath, P. P.; Adamson, R. D.; Austin, B.; Baker, J.; Byrd, E. F.; Dachsel, H.; Doerksen, R. J.; Dreuw, A.; Dunietz, B. D.; Dutoi, A. D.; Furlani, T. R.; Gwaltney, S. R.; Heyden, A.; Hirata, S.; Hsu, C. P.; Kedziora, G.; Khalliulin, R. Z.; Klunzinger, P.; Lee, A. M.; Lee, M. S.; Liang, W.; Lotan, I.; Nair, N.; Peters, B.; Proynov, E. I.; Pieniazek, P. A.; Rhee, Y. M.; Ritchie, J.; Rosta, E.; Sherrill, C. D.; Simmonett, A. C.; Subotnik, J. E.; Woodcock, H. L.; Zhang, W.; Bell, A. T.; Chakraborty, A. K.; Chipman, D. M.; Keil, F. J.; Warshel, A.; Hehre, W. J.; Schaefer, H. F.; Kong, J.; Krylov, A. I.; Gill, P. M.; Head-Gordon, M. *Q-Chem, version 3.0*; Q-Chem, Inc.: Pittsburgh, PA, 2006.

(60) Ponder, J. W., *TINKER, Software Tools for Molecular Design*, Version 4.2; 2004.

(61) Amaro, R. E.; Swift, R. V.; Votapka, L.; Li, W. W.; Walker, R. C.; Bush, R. M. Mechanism of 150-cavity formation in influenza neuraminidase. *Nat. Commun.* **2011**, *2*, 388.

(62) Sasaki, D.; Fujihashi, M.; Okuyama, N.; Kobayashi, Y.; Noike, M.; Koyama, T.; Miki, K. Crystal structure of heterodimeric hexaprenyl diphosphate synthase from *Micrococcus luteus* B-P 26 reveals that the small subunit is directly involved in the product chain length regulation. *J. Biol. Chem.* **2011**, *286*, 3729−3740.