Published in final edited form as:

ACS Catal. 2018 May 4; 8(5): 4299–4312. doi:10.1021/acscatal.8b00543.

Catalytic Role of Conserved Asparagine, Glutamine, Serine, and Tyrosine Residues in Isoprenoid Biosynthesis Enzymes

Satish R. Malwal^{†,+}, Jian Gao^{†,+}, Xiangying Hu^{†,+}, Yunyun Yang^{§,+}, Weidong Liu[‡], Jian-Wen Huang[‡], Tzu-Ping Ko[Ⓛ], Liping Li[§], Chun-Chi Chen[‡], Bing O'Dowd[†], Rahul L. Khade^{\$\$}, Yong Zhang^{\$\$}, Yonghui Zhang^{§,*}, Eric Oldfield^{†,*}, and Rey-Ting Guo^{†,*}

[†]Department of Chemistry, University of Illinois, Urbana, IL 61801, USA

[‡]Industrial Enzymes National Engineering Laboratory, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China

[Ⓛ]Institute of Biological Chemistry, Academia Sinica, Taipei 11529, Taiwan

[§]School of Pharmaceutical Sciences; MOE Key Laboratory of Bioorganic Phosphorus Chemistry & Chemical Biology, Tsinghua University, Beijing 100084, China

^{\$\$}Department of Chemistry and Chemical Biology, Stevens Institute of Technology, 1 Castle Point Terrace, Hoboken NJ 07030, USA

Abstract

We report the results of an investigation into the catalytic role of highly conserved amide (asparagine, glutamine) and OH-containing (serine, tyrosine) residues in several prenyltransferases. We first obtained the X-ray structure of cyclolavandulyl diphosphate synthase containing two molecules of the substrate analog dimethylallyl (*S*)-thiolodiphosphate (DMASPP). The two molecules have similar diphosphate group orientations to those seen in other ζ -fold (*cis*-head-to-tail and head-to-middle) prenyltransferases with one diphosphate moiety forming a bidentate chelate with Mg^{2+} in the so-called S1 site (which is typically the allylic binding site in ζ -fold proteins) while the second diphosphate binds to Mg^{2+} in the so-called S2 site (which is typically the homoallylic binding site in ζ -fold proteins) *via* a single P1O1 oxygen. The latter interaction can facilitate direct phosphate-mediated proton abstraction *via* P1O2, or more likely by an indirect mechanism in which P1O2 stabilizes a basic asparagine species that removes H^+ , which is then eliminated *via* an Asn-Ser shuttle. The universal occurrence of Asn-Ser pairs in ζ -fold proteins leads to the idea that the highly conserved amide (Asn, Gln) and OH-containing (Tyr) residues seen in many “head-to-head” prenyltransferases such as squalene and dehydrosqualene

*Corresponding Author: E.O.: eo@chad.scs.uiuc.edu, R.T.G.: guo_rt@tib.cas.cn, Y.Z.: zhangyonghui@tsinghua.edu.cn.

[†]Author Contributions

These authors contributed equally to this work.

Notes

The authors declare no competing financial interest.

Supporting Information

Tables containing LLDF, B-factor, dihedral angle and distance information for DMASPP ligands bound to CLDS, and SCORECONS scores and rank-ordering for the top 14 most highly conserved residues in CrtM and CrtM-like proteins, are included in the Supporting Information.

This information is available free of charge on the ACS Publications website

synthase might play similar roles, in H^+ elimination. Structural, bioinformatics and mutagenesis investigations indeed indicate an important role of these residues in catalysis, with the results of density functional theory calculations showing that Asn bound to Mg^{2+} can act as a general (imine-like) base, while Gln, Tyr and H_2O form a proton channel that is adjacent to the conventional (Asp-rich) “active site”. Taken together, our results lead to mechanisms of proton-elimination from carbocations in numerous prenyltransferases in which neutral species (Asn, Gln, Ser, Tyr, H_2O) act as proton shuttles, complementing the more familiar roles of acidic groups (in Asp and Glu) that bind to Mg^{2+} , and basic groups (primarily Arg) that bind to diphosphates, in isoprenoid biosynthesis.

Graphical abstract

Highly conserved amide and hydroxy-containing amino-acid residues are involved in proton elimination in proteins catalyzing isoprenoid biosynthesis

Keywords

Isoprenoid biosynthesis; X-ray crystallography; *cis*- and *trans*-prenyltransferases; proton shuttle; dehydrosqualene synthase; quantum chemistry

INTRODUCTION

There is considerable interest in the structure, function and inhibition of the enzymes involved in the biosynthesis of isoprenoids and terpenes since these small molecules are the most diverse species on Earth with ~80,000 compounds being known.^{1,2} Most such molecules are synthesized by proteins that have just three major “folds”: α , $\beta\gamma$ and ζ .^{2,3} The α -fold proteins catalyze the *trans* “head-to-tail” and “head to head”⁴ formation of prenyl diphosphates, the cyclization of prenyl diphosphates, or the alkylation of aromatic species and typically contain one or two catalytic “DDXXD”-like domains that bind to three Mg^{2+} , leading to diphosphate ionization, carbocation formation, condensation/cyclization, followed by H^+ abstraction and product release. The $\beta\gamma$ fold proteins are comprised of two highly α -helical domains that surround a central catalytic domain typically containing a DXDD motif that can protonate alkene or oxirane species, leading to complex electrocyclization reactions. The third major class of proteins are those that are employed in the “head-to-tail”⁴ formation of *cis*-prenyl diphosphates such as *cis*-farnesyl diphosphate and *cis*-undecaprenyl diphosphate, involved in bacterial cell wall biosynthesis. These protein have the ζ -fold³ and utilize just a single Mg^{2+} bound to a single aspartate to initiate diphosphate charge separation, condensation, H^+ abstraction, and product formation. In both the α and ζ -fold proteins there are thus three basic steps in catalysis, as illustrated in Scheme 1a. First,

diphosphate substrates bind to active site Mg^{2+} (coordinated to Asp or Glu), and to basic residues (typically Arg but also Lys or His), plus, there are hydrophobic interactions with non-polar residues. Second, there is a Mg^{2+} -initiated charge separation or ionization step,⁵ followed by condensation, forming a carbocation species. Third, an H^+ in the carbocation is then eliminated (i.e. removed) from the active site region by a base (to prevent the back-reaction), leading to product formation. This third reaction is arguably the least well studied or understood.

In recent work, it has now been found that some ζ -fold proteins are involved in more unusual reactions, unlike the familiar head-to-head and head-to-tail prenyl synthases noted above. For example, there are three enzymes that are involved in formation of “head-to-middle” isoprenyl diphosphates: lavandulyl diphosphate (**1**, Scheme 1b) synthase (LPPS),⁶ isosesquilandulyl diphosphate (**2**) synthase (also known as Mcl22⁷), and cyclolavandulyl diphosphate (**3**) synthase (CLDS).⁸ The chemical structures of their substrates (dimethylallyl diphosphate **4**, and geranyl diphosphate **5**) and products (**1–3**) are shown in Scheme 1b, and the X-ray structures of LPPS, CLDS and Mcl22 have recently been reported.^{9–11} The mechanisms of action of LPPS and Mcl22 involve ionization (or at least partial charge separation if the mechanism is concerted⁴) of a dimethylallyl diphosphate (DMAPP, **4**) in the allylic S1 site followed by condensation with either **4** or **5** in the S2 site to form **1** or **2**, and this mechanism is quite similar to that reported¹² for undecaprenyl diphosphate synthase, UPPS. The mechanism of action of CLDS which forms *cyclo*-LPP (**3**) and not LPP is more complex since it involves an intermediate cyclization reaction, but in very recent work, Tomita et al.¹⁰ proposed an elegant mechanism involving a proton shift leading to a carbocation rearrangement, Scheme 2, based on isotope labeling, crystallography and computational docking. However, the actual locations of the DMAPP molecules were not determined crystallographically since the DMAPP used in crystallization hydrolyzed to diphosphate, PPI. The mechanisms of action of the *trans*- head-to-head prenyltransferases are even more complex since they involve not only a cyclization, but also ring-opening and in the case of squalene synthase, a final NADPH-mediated reduction.

Here, we report the first structure of CLDS with two substrate-like ligands, dimethylallyl (*S*)-thiolodiphosphate (DMASPP, **6**), an unreactive analog of the normal CLDS substrate, DMAPP. We then investigate using site-directed mutagenesis how protons are abstracted during catalysis, the results obtained leading to new proposals for the mechanism of action of α -fold head-to-head prenyltransferases such as dehydrosqualene synthase in which there are two half-reactions, one or both involving H^+ -elimination. In both the α and ζ -fold proteins, we test the hypothesis that amide and OH-containing residues may be involved in the H^+ -elimination that drives product formation.

MATERIALS AND METHODS

Protein Production

The gene encoding full-length CLDS (*Streptomyces* sp. CL190, genbank: BAO66170.1) was chemically synthesized by GENEray Biotech Co., Ltd. (Shanghai, China) and ligated into the pET46Ek/LIC vector. The forward primer was 5'-GACGACGACAAGATGATGACCACCCTGATGCTG-3', the reverse primer was 5'-

GAGGAGAAGCCCGGTTATTATGCCGGATAACCACCAAA-3'. The recombinant plasmids were verified by DNA sequencing. Each mutant was prepared by using a Quik Change site-directed mutagenesis kit (Agilent Technologies, Santa Clara, CA, USA) following the manufacturer's instructions with the pET46Ek/LIC-clds plasmid as a template and the following oligonucleotides: for S54A, forward: 5'-CTGTATATTACCGCAAGCGCAGCAGCAAATCATGGTCGT-3', reverse 5'-ACGACCATGATTTGCTGCTGCGCTTGCGGTAATATACAG-3'; for N57A, forward: 5'-ACCGCAAGCAGCGCAGCAGCACATGGTCGTCCGGAAGCA-3', reverse: 5'-TGCTTCCGGACGACCATGTGCTGCTGCGCTGCTTGCGGT-3'. The recombinant plasmids were verified by DNA sequencing.

Plasmids were transformed into *E. coli* BL21(DE3) cells and protein expression was induced by 0.3 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) at 16 °C for 24 hours. CLDS protein purification was carried out at 4 °C as follow. Cells were harvested by centrifugation at $5,000 \times g$ for 15 minutes, then re-suspended in lysis buffer containing 25 mM Tris-Cl, pH 7.5, 150 mM NaCl and 20 mM imidazole, followed by disruption with a French Press. Cell debris was removed by centrifugation at $17,000 \times g$ for 1 hour. The supernatant was then applied to a Ni-NTA column FPLC system (GE Healthcare). The target proteins eluted at ~100 mM imidazole when using a 20–250 mM imidazole gradient. Proteins were dialyzed against buffer containing 25 mM Tris-Cl and loaded onto a DEAE Sepharose column. Target proteins were eluted at ~200 mM NaCl when using a 0–500 mM NaCl gradient. The purified proteins were then concentrated to 10 mg/mL in buffer containing 25 mM Tris-Cl and 150 mM NaCl. Protein purity was verified by SDS-PAGE analysis.

Activity measurements

Protein activity was measured by using a PPi release assay. CLDS protein concentration was measured by using the BCA method (Thermo Fischer Scientific), according to the manufacturer's instructions. The protein concentration used in the activity assays was 0.1 mg/mL. PPi release was determined by using a continuous spectrophotometric assay in 96-well plates with 200 μ L reaction mixtures containing: 140 μ M 2-amino-6-mercapto-7-methyl purine ribonucleoside (MESG), 4 μ g/mL purine nucleoside phosphorylase and inorganic diphosphatase, 100 μ M DMAPP, 50 mM Tris-HCl, 0.1 mM sodium azide and 1 mM $MgCl_2$. Absorption at 360 nm was measured as a function of time to determine activity, reported as relative activity for the WT and mutant proteins. Samples were assayed in triplicate in each independent experiment.

Crystallization, Data Collection, Structure Determination and Refinement

CLDS crystals were obtained from 0.3 M NaCl, 27 % v/v polyethylene glycol 3350, 0.1 M HEPES pH 7.5 by using the sitting-drop vapor diffusion method. In general, 1 μ L protein (10 mg/mL) was mixed with 1 μ L of reservoir solution in 48-well Cryschem Plates, and equilibrated against 100 μ L of the reservoir at 25°C. Within one month, crystals reached a size suitable for X-ray diffraction data collection. CLDS/DMASPP crystals were obtained by soaking apo-CLDS crystals in 0.4 M NaCl, 28 % v/v polyethylene glycol 3350, 0.15 M HEPES, pH 7.5 containing 10 mM DMASPP for 6 hours. Crystals were mounted in a cryo-loop and flash-cooled in liquid nitrogen. Data sets were collected at beam lines BL13B1 and

TPS05A of the National Synchrotron Radiation Research Center (NSRRC, Hsinchu, Taiwan) and processed by using the HKL2000¹³ program. Prior to structure refinement, 5% randomly selected reflections were set aside for calculating R_{free} ¹⁴ as a monitor of model quality. Phase information was obtained by using the molecular replacement program Phaser¹⁵ in the CCP4i suite¹⁶ using isosesquilandulyl diphosphate synthase (McI22; PDB ID code 5XK3) as the template. Structure refinements were carried out by using Phenix¹⁷ and Coot.¹⁸ Figures were prepared by using the PyMOL program (<http://pymol.sourceforge.net/>).

Computational Methods

All calculations were carried out by using the Gaussian 09.1 program.¹⁹ Full geometry optimizations were conducted using the B3LYP2²⁰ functional and a uniform 6–311++G(2d, 2p) basis set. All Cartesian coordinates are given in Table S1. Atomic charges were calculated by using the Natural Population Analysis (NPA) method in Gaussian 09.²¹

RESULTS AND DISCUSSION

Structure of CLDS with DMASPP

We first obtained the X-ray structure of CLDS with two bound DMASPP ligands to see if we could determine the mechanism of H^+ elimination (**4d** in Scheme 2). Full crystallographic data acquisition and refinement details for the *apo*-protein, as well as CLDS containing DMASPP (by soaking), are given in Table 1. The *apo*-protein crystallized in the C2 space group (PDB ID code 5YGJ) but the crystals transformed to P2₁ on incubation with DMASPP. A stereo-view of Chains A and B in the dimeric protein containing DMASPP (PDB ID code 5YGK) is shown in Figure 1a, where it can be seen that there are two DMASPP molecules in the active site, Figure 1b. One occupies the so-called S1 (“allylic”) site seen in LPPS and UPPS, while the second binds to the S2 site (the “homoallylic” site in UPPS, or the second allylic site seen in LPPS). Ligand electron densities are shown in Figure 1b. There was good agreement between our *apo*-CLDS and CLDS/DMASPP structures with the *apo*-CLDS and CLDS/PPi structures reported by Tomita et al. (PDB ID codes 5GUK, 5GUL; Ca rmsds, root mean square deviations, range from 0.318 to 0.498 Å). In the PPi/CLDS structure¹⁰ (PDB ID code 5GUL), there are two Tyr residues in the active site region, Y26 and Y215* (the asterisk indicates the residue is from the second chain in the dimer), and these residues are involved in diphosphate release in the initial ionization step, based active site region in *apo*-CLDS,¹⁰ and is also not seen in our liganded structure, presumably because we used a DMASPP soak with *apo*-crystals.

Surprisingly, we find that the DMASPP side-chain conformations are very varied. In the unit cell there are 16 DMASPP ligands, 8 in the S1 site (in Chains A-H) and 8 in the S2 sites. Ligand structures and electron densities together with several amino-acid residues are shown in Figure 2. The B-factors for the 16 ligands are 48 $\pm$ 7 Å², the real-space refinement (RSR) values are $\sim$ 0.1 and the LLDF (local ligand density function) values are -0.79 ± 0.25 , Table 1 and Table S2. Thus, all ligands are well defined. However, the dimethylallyl side-chains pack in very different ways, as can be seen in the DMASPP S-C1-C2-C3 torsion angles shown in Table S2, and in Figure 2. In some chains, the double bonds in the two

ligands are appropriately oriented (stacked) so as to facilitate the proton shift/cyclization reaction proposed,¹⁰ resulting in formation of intermediates **4b-d** (Scheme 2), but in other chains they are not. However, the C1 (S1 DMASPP) to C2' (S2 DMASPP) as well as the C3 (S1 DMASPP) to C5' (S2 DMASPP) distances are quite similar (3.69 $\pm$ 0.53 Å and 4.35 $\pm$ 0.52 Å), Table S2, so it appears that there is simply a low barrier for ligand side-chain rotation, with only some conformers leading to cyclization. The question then arises: what is the mechanism for proton abstraction used by CLDS?

Proton elimination in CLDS and other ζ -fold prenyl transferases

To begin to answer the proton elimination question, we compared the sequences and structures of CLDS and a series of other ζ -fold proteins. Figure 3 shows a sequence alignment of CLDS together with six other ζ -fold proteins: CLDS, Mcl22, LPPS, a tomato *cis*-farnesyl diphosphate synthase (FPPS), *E. coli* undecaprenyl diphosphate synthase (UPPS), *Mycobacterium tuberculosis* FPPS (Rv1086c), and *M. tuberculosis* decaprenyl diphosphate synthase (Rv2361c). Figure 4 shows structural alignments between CLDS and these 6 proteins^{9-11, 22,23} and Figure 5a shows an alignment of the ligands and residues of mechanistic interest. All proteins have similar ζ -folds³ with, on average, a 1.842 Å Ca rmsd between CLDS and the six other proteins. A single Asp (D9 in CLDS) binds to Mg²⁺ in each of the proteins, and this Mg²⁺ is essential for ionization/charge separation of the S1 diphosphate, initiating condensation with the olefinic double bond in the substrate in the S2 site. Why, then, do diphosphates (allylic or homoallylic) that bind to the S2 site not ionize? What is clear from the ligand and residue super-positions shown in Figures 3 and 5a is that in each and every protein, there are a totally conserved pair of Arg residues (R163, R169 in CLDS) that prevent the terminal, β phosphate group (P2) from binding to Mg²⁺, by forming strong Arg-diphosphate electrostatic interactions. This Arg pair thus acts to 'hoist' the β phosphate away from Mg²⁺, Figure 5a, preventing ionization. Since only a single oxygen (P1O1, Scheme 2) binds to Mg²⁺ in the S2 site, the second oxygen (P1O2) can then act as a base, removing H⁺ from any carbocation formed on condensation or condensation/cyclization. Remarkably, this diphosphate-binding pattern is found even in the absence of any ligand side-chain, as can be seen in the CLDS-PPi structure shown in Figure 5b (PDB ID code 5GUL). It thus appears that at least bidentate chelation of a diphosphate to Mg²⁺ is required in order to permit diphosphate ionization, and this only happens in S1. But how are protons removed from the active site?

In our earlier mechanistic investigations of *E. coli* UPPS¹², we proposed that protons were removed by a relay involving an Asn-Ser pair, S71 and N74, and on inspection of the sequence alignment (Figure 3) as well as on inspection of CLDS and the other ζ -fold structures and structure alignments shown in Figures 4 and 5a it is clear that these Asn-Ser residues are: i) totally conserved, and ii) have very similar locations/conformations in all seven proteins. These observations imply a common functional role. We also find that mutagenesis of these residues to Ala in CLDS, Figure 5c, as well as in UPPS¹² and Mcl22,¹¹ decreases activity and the largest effects are seen with the Asn mutants. While the variability in the DMASPP side-chain conformations in CLDS make it difficult to draw firm structural conclusions, in LPPS, Mcl22 and UPPS we proposed (based on distance measurements) that the P1O2 diphosphate groups play a key role in H⁺-elimination. However, there could

actually be direct (H^+ -abstraction by P1O2) as well as indirect phosphate involvement, as shown in Scheme 3, or perhaps both mechanisms might be involved.

In the direct mechanism, Schemes 3a, 3b, the sequence of reactions involved in proton abstraction (i.e. elimination) would be: carbocation (XH^+) > PO^- > Asn ($CONH_2$) > Ser (OH) > H_2O . That is, an H^+ is abstracted by phosphate, then, this acidic proton is transferred to Asn, then to Ser, in a proton shuttle. This shuttle could involve zwitterionic (Scheme 3a) or perhaps imidic acid tautomers of the amide side-chain (Scheme 3b), as reported in a neutron diffraction study of the mechanism of action of a cellulase in which imine species are part of a proton relay.²⁴ The second possibility is that there is *indirect* phosphate-mediated H^+ elimination, Schemes 3c, 3d. In these mechanisms, phosphate acts to stabilize enolate or imine forms of the amide (e.g. *via* a $PO^- \dots NH_2(+)C(R)O^-$ electrostatic interaction), facilitating direct removal of H^+ by Asn. We measured donor-acceptor distances between the putative carbocations (CH^+), $(P)O^-$, Asn (CO/NH_2 are generally indistinguishable in X-ray structures), and the Ser OH , in Mcl22, LPPS, UPPS and a plant-derived *cis*-FPPS²⁵ finding a mean value of $3.8 \text{ \AA} \pm 0.62 \text{ \AA}$ for the 12 distances involved. Rv1086c, Rv2361c and CLDS structures were not used because they lacked appropriate ligands or had variable ligand conformations. Hydrogen bonds with donor-acceptor distances in the $2.2\text{--}2.5 \text{ \AA}$ range are classified as strong, mostly covalent;²⁶ $2.5\text{--}3.2 \text{ \AA}$ distances are classified as moderate, mostly electrostatic, and $3.2\text{--}4 \text{ \AA}$ are weak, electrostatic.²⁶ The distances we measure are therefore consistent with a proton-abstraction mechanism involving the intermediacy of weak, electrostatic hydrogen bonds. However, it is also possible that the indirect route is involved since the mean distances for the PO^- -Asn; Asn- (CH^+) and Asn-Ser interactions are almost the same, $3.5 \pm 0.73 \text{ \AA}$ for the 12 distances involved. In the absence of any neutron structures of ζ -fold proteins (which would enable determination of the positions of exchangeable deuterons) it is not possible to prove which mechanism dominates since NH , NH_2 , O , O^- and OH forms of the Asn amide group are very difficult to be differentiated by protein X-ray crystallography, but PO^- involvement does seem to be important since a neutral amide side-chain has little basicity, and it is likely that there will be strong interactions between the P1O2 and Asn residues because the mean PO^- -amide (closest atom) distance is $3.0 \pm 0.2 \text{ \AA}$. Indeed, the $P1O2 \dots$ amide distance is $3.0 \text{ \AA}$ and the Asn-Ser distance is $3.3 \text{ \AA}$ even in the CLDS-PPi structures (PDB ID code 5GUL), again suggesting a common PPi-Asn-Ser core structure, and function.

Since, it is difficult to experimentally determine the nature (O or N) or protonation state of the Asn amide group, we reasoned that it might be possible to answer the questions raised above computationally, by using density functional theory (DFT). Of particular interest was a determination of a phosphate-amide interaction energy since a very high energy would tend to rule out a role for that structure and thus, mechanism. We thus evaluated the energies of the 6 phosphate-amide models shown in Figure 6, using full geometry optimization. Model 1, containing a $PO^- \dots NH_2$ amide interaction had the lowest energy and a $2.8 \text{ \AA}$ $PO^- \dots N$ distance and Model 6 optimized to a similar structure but had a slightly higher (10 kJ) energy, Figure 6. All of the other species had $\sim 40\text{--}95 \text{ kJ}$ higher energies. These results thus support a phosphate... NH_2^+ zwitterionic amide interaction, consistent with the indirect mechanism for H^+ -elimination shown in Scheme 3c for the ζ -fold prenyltransferases.

Proton elimination in head-to-head prenyl transferases

The results discussed above lead to the idea that similar (amide/OH-containing species) proton shuttle pathways might be involved in other prenyl transferases, such as the *trans*-head-to-head prenyl transferases dehydrosqualene synthase (CrtM)²⁷ and squalene synthase (SQS), where there are actually two sequential reactions, as shown in Scheme 4. This possibility is of interest because in addition to the “canonical” prenyl transferase active-site residues Asp binding to Mg²⁺ and Arg binding to diphosphates a previous bioinformatics (SCORECONS²⁸) analysis of *S. aureus* dehydrosqualene synthase, SaCrtM, based on the sequences of SaCrtMs and other proteins with similar sequences, revealed that there are actually five amide or OH-containing residues (Y41, Y129, Y183, Q165 and N168) that are very highly conserved, Table S3. A sequence alignment of 5 structure of sequence-related proteins: SaCrtM, human squalene synthase (HsSQS), *Alicyclobacillus acidocaldarius* 12-hydroxysqualene synthase (AaHpnC), *Neisseria meningitidis* presqualene diphosphate synthase (NmHpnD), which makes presqualene diphosphate²⁹ and *Arabidopsis thaliana* phytoene synthase (AtPHY), is shown in Figure 7 in which the conserved neutral residues of interest (Asn, Gln and Tyr) are shown in pink. Based on the results with the *cis*-head-to-tail and head-to-middle prenyltransferases, we hypothesized that some of these highly conserved amide (Asn, Gln) and OH-containing (Tyr) residues might play a similar role to that of the conserved Asn and Ser in the ζ -fold proteins in H⁺-elimination.

In order to test this hypothesis (and cognizant of the difficulty of actually proving a proton relay mechanism), we posed a series of questions: Are these residues essential for catalysis? Are these residues close to the active site? Do the X-ray structures reveal a likely proton channel? Are the distances between the atoms involved in any such channel “reasonable”, for example, are they similar to the distances involved in weak, electrostatic hydrogen bonds? Is there evidence for the presence of water molecules in any putative channel? The conserved Asn168, a neutral residue, binds to Mg²⁺: why is this, and might this residue play a role in H⁺-abstraction? Do the structural results suggest other, more obvious roles for these residues, in catalysis? Might a diphosphate be involved in H⁺ elimination, as in the ζ -fold proteins? And finally, do the X-ray structures indicate how ligands (and H⁺) enter and exit these proteins?

We first consider the roles of the highly conserved (non-Asp, non-Arg) residues: Y41, Y129, Y183, Q165 and N168, in SaCrtM. These are 5 of the 14 most highly conserved residues, Table S1, the others all being either Asp or Arg (that bind to Mg²⁺, or diphosphate). On inspection of the X-ray structures of CrtM bound to farnesyl *S*-thiolodiphosphate (FSPP; PDB ID code 3W7F), it is apparent that Y41 interacts with a S2 diphosphate, while N168 binds to a Mg²⁺ and appears to play a structural role—a point we return to in the following section. On the other hand, Y129, Y183 and Q165 have no obvious role. As can be seen in Figure 8, Y129 and Y183 and Q165 form a distinct cluster of residues (shown in yellow in Figure 8b) adjacent to the canonical active site DXXXD, DXXED or Arg-rich regions. One possibility is that these residues are directly involved in catalysis (which we define here as involvement in ionization, condensation, cyclization, rearrangement, or H⁺-elimination reactions). A second possibility is that these residues might play a purely “structural” role and are not directly involved in catalysis.

As to the possibility that the most highly conserved residue in CrtM, Y129, plays a purely structural role, this seems unlikely since in earlier work we reported²⁶ the X-ray structures of both WT SaCrtM as well as that of the inactive Y129A mutant, finding a C α rmsd of only 0.26 Å (over 279 residues). In the region (+/- 5 residues) surrounding Y129, the C α rmsd is only 0.16 Å, so there are no major global or even local structural changes caused by the mutation. We conclude that a purely structural role for Y129 is unlikely. Notably, Y41, Y129 and Q165 (in CrtM) are the top 3 most essential residues (excluding structural Gly and Pro) in the SCORECONS analysis, Table S3,²⁷ and they have even higher SCORECONS scores than those of the 9 Asp/Arg residues that are generally accepted to be directly involved in catalysis.²⁷ Moreover, the two dehydrosqualene mutants that we investigated (Y129A and Q165A) had little or no catalytic activity in both 1st half (farnesyl diphosphate, FPP, to presqualene diphosphate, PSPP) and second half (PSPP to dehydrosqualene) reactions.²⁷ The analog of Y129 in rat SQS is Y171, and in very early work,³⁰ Gu et al. showed that Y171 was also essential for activity. Thus, Q165, Y129 and Y183 (SaCrtM numbering) form a cluster of very highly conserved residues that are adjacent to the active site and represent a potential path for H⁺-elimination. We also see, Figure 8b, that there is a hydrogen-bonded water (O440) between Q165 and Y129, suggesting the following pathway for proton elimination: Gln (Q165) > H₂O (O440) > Tyr (Y129) > Tyr (Y183) with on average, donor-acceptor distances of ~3.8 +/- 1.0 Å, comparable to that seen in the ζ -fold prenyltransferases (3.8 Å +/- 0.62 Å) discussed above. And since N215, Q212 and Y171 in HsSQS have similar locations to the corresponding residues (N168, Q165 and Y129) in CrtM (sequence alignments with the SaCrtM numbering scheme are shown in Figure 7), these residues might also provide a pathway for H⁺ removal from the SQS active site (in the first half-reaction). The second Tyr seen in CrtM, Y183, is absent in HsSQS, but the corresponding Tyr is present in AaHpnC (Figure 7) as well as in HpnD (Figure 7), making it likely that this residue is also involved in catalysis in these systems, which are involved in hopanoid biosynthesis. In addition, as noted above, all 4 residues are present in *Arabidopsis thaliana* phytoene synthase meaning that the same basic structure and mechanism of action is used by bacteria, plants and animals although the substrates and products differ. In HsSQS, it has also been shown that Q212L and Q212N mutants produce only 4% and 8% of WT squalene using FPP as the substrate while a Q212E mutant produces an essentially normal (85%) amount of squalene.³¹ These results support a role of Q212 and E212 in proton elimination since the glutamine and glutamic acids have the same side-chain lengths and a carboxyl/carboxylate group can transport protons, while the leucine in Q212L cannot and apparently, the asparagine in Q212N is too short to be effective. But what is the base involved in the initial stage of H⁺-elimination in SaCrtM? Unlike in LPPS, Mcl22 and UPPS where there are close-by P1O2 groups that can mediate proton abstraction, there do not seem to be such diphosphate groups present in the CrtM structure that are near the Gln (Q165) > H₂O (O440) > Tyr (Y129) > Tyr (Y183) cluster that might serve as a base, and the highly conserved asparagine N168 (N215 in HsSQS) is actually coordinated to a Mg²⁺ (Mg2, Figure 8). However, since a PPi group (in S1) is generated on ionization of the FPP in S1, this PPi might facilitate H⁺ removal and transfer to the close-by N168, in particular if N168 acted is a base. But is there any reason to think that N168 might function as a base? Indeed-why does it even bind to Mg²⁺?

How does an Asn amide side-chain bind to Mg²⁺?

The nature of the Asn-Mg²⁺ interaction is of general interest due to the frequent occurrence of the “NSE” (asparagine-serine-glutamic acid) motif in terpene cyclases.² However, the presence of an ion-dipole (Mg²⁺-Asn amide) interaction would be expected to be much weaker than that of the ion-ion interaction that would be found with an Asp. This suggests the interaction might be stronger than anticipated for an ion-dipole interaction. A possible reason is Mg²⁺ stabilization of the Asn H₂N⁺=C-O⁻ dipolar resonance form, Schemes 5a, 5b. And if deprotonated, 5b would yield an imine species, Scheme 5c, which could act as a base and participate in H⁺-elimination, “bridging” the PPI formed on ionization and the Q165/H₂O/Y129/Y183 cluster. Such isomeric forms of amides are known as imidic acids^{32,24} and are involved in proton relays in cellulase, as observed by neutron crystallography.²⁴ Also, the pK_a for the protonation/deprotonation of the nitrogen in a series of imidic esters and ethers is computed to be in the ~5.3–7.4 range,³² supporting the idea that the imidate form of N168 could be involved in proton abstraction.

To investigate the Asn-Mg²⁺ interaction in more detail, we carried out a series of DFT calculations on three systems that model the structures shown in Scheme 5: acetamide, acetamide coordinated to a neutral [Mg(dimethylphosphate)(acetate)(H₂O)₃] cluster (forming a 6-coordinate species), and a six coordinate [Mg(acetamide)(dimethylphosphate)(acetate)(H₂O)₃]⁻¹ anionic complex, in addition to the acetamide anion (imine form) itself, for purposes of comparison. Geometry-optimized structures are shown in Figure 9 and computed geometric and charge properties are shown in Table 2. The results shown in Table 2 indicate that the CO bond length increases slightly, from 1.217 Å in acetamide to 1.240 Å in the acetamide neutral complex, then significantly (to 1.308 Å) in the acetamide anionic complex. The CN bond length decreases from 1.365 Å in acetamide to 1.338 Å in the acetamide neutral complex, then decreases further (to 1.290 Å) in the acetamide anionic complex (that contains the imine base). These structural changes indicate an increase in single-bond character in the acetamide carbonyl group, as well as a major increase in CN double-bond character from amide to iminium to imine. The changes are also reflected in increased oxygen atomic charges (from -0.636 e, to -0.702 e and to -0.924 e), but perhaps of most interest here, the short CN bond length is indicative of formation of an imine group that could act as a base in initiating the H⁺ elimination cascade. The 1.308 Å CO bond length in the acetamide anionic complex is even longer than the 1.268 Å in the acetamide anion (imine form), and the oxygen in the anionic complex also carries a closer-to-unity negative charge (-0.924 e) than that in the acetamide anion alone, -0.836 e, indicating stabilization of CO single bond character by the magnesium ion. Moreover, the 1.290 Å CN bond length in the acetamide anionic complex is even shorter than the 1.321 Å in the acetamide anion itself, again meaning increased CN double-bond character in the imine form in the complex. We also find that there is resonance over the oxygens in the O=P-O-group bonded to Mg²⁺, based on the very small differences in P1O1 and P1O2 bond lengths, and very similar O1 and O2 charges, in both the neutral and anionic complexes, Table 2.

We then computed the difference in Gibbs free energy between the neutral complex and the anionic complex plus a solvated H⁺ since a relatively small value would be required for catalytic purposes, that is, the complexes need to “recycle” for each turnover. The computed

total Gibbs free energy of the neutral complex was -997744.03 kcal/mole and that of the anionic (imine-containing) complex was -997469.04 kcal/mole. When the latter value is added to the experimental proton energy value (-270.89 kcal/mole³³) we obtain a total Gibbs free energy (anion complex plus proton) of -997739.93 kcal/mole which is 4.10 kcal/mole higher than that of the neutral acetamide complex. This ΔG corresponds, approximately, to a $pK_a \sim 3$ —slightly more acidic than the ~ 5 – 7 range for the protonation/deprotonation of the nitrogen in imidate esters and ethers.³² This is as expected and is due to metal coordination and indicates that at $pH \sim 7$, the anionic complex can serve as a base (an imine) that can accept a proton, during catalysis. Based on all of the above observations we thus propose the proton elimination mechanism in CrtM shown in Scheme 6. After ionization of the FPP in S1 and condensation, an XH^+ carbocation forms, together with a PPi bound to Mg²⁺, in S1. A diphosphate oxygen then abstracts the XH^+ proton and transfers it to the N165 imine. Then, this H^+ is shuttled out of the active site via Glu, H₂O and Tyr.

Thus, in answer to the questions posed above: the highly conserved Asn/Gln/Tyr residues in the head-to-head prenyltransferases are highly conserved and appear to be essential for catalysis and are either part of, or are adjacent to, the active site. They can provide both a “bridging” base (Asn) for proton abstraction (from XH^+ and PPi) to the proton channel with distances that are consistent with the presence of hydrogen bond involvement in the proton transfers. A water is found between Q165 and Y129 (the 2 most conserved residues) in CrtM, and the distances between the conserved Q165, Y129, Y183 and either Mg²⁺ or the substrates are large (~ 7 – 8 Å), consistent with the absence of any obvious “direct” interactions of these residues with Mg²⁺ or substrates.

As to other possible mechanisms of action, we considered three alternative roles for Gln, H₂O and the Tyr residues in catalysis although each would remove the H^+ -elimination pathway proposed (and not replace it with an alternative one). One possibility is involvement in diphosphate (PPi) product release, shuttling PPi, but this seems unlikely since the Coulombic interaction between PPi and Mg²⁺ is expected to be very strong due to the $+2$ (Mg) and ~ -3 (PPi) charges involved, and the chelate effect; there would be no role for H₂O, and any interaction with Tyr would involve just a single hydrogen bond. That is, elimination of PPi as a Mg²⁺-PPi complex seems more likely. A second possibility is a role in Mg²⁺ influx/efflux, but again such interactions would be weak when compared to the multiple Mg²⁺-Asp charge-charge (Coulomb) interactions. The third possibility is a role in substrate/product isoprenoid diphosphate influx/efflux. In support of this, we do see that Tyr (Y41) can interact with a diphosphate (in S2), and similar interactions are seen in CLDS, but a role for Y129, Y183 in substrate/product diphosphate uptake/release would involve opening up a completely different pathway (capable of housing products with volumes of ~ 500 Å³) to that involving the Mg²⁺/Asp groups (pink residues in Figure 8b) in which Mg²⁺ can interact strongly with diphosphate groups *via* chelate formation. Thus, all three alternative mechanisms have drawbacks, including not providing an alternative H^+ -elimination pathway. Nevertheless, we next examined in more detail the structures of SaCrtM with diverse ligands to see if we might deduce ligand and proton entry and exit points that might support the mechanisms discussed above.

Topological models for FPP, dehydrosqualene, squalene, PPi, NAD(P)H, NAD(P)⁺ and H⁺ exits and entrances

The results we discussed above focused on the “microscopic” aspects of H⁺ elimination in CrtM/SQS but we did not address the more global or topological questions: How does FPP enter? How or where do dehydrosqualene and squalene exit? How or where do the two PPis leave? Where do H⁺ leave? What about NAD(P)H and NAD(P)⁺? To help answer some of these topological questions we re-examined the structures of CrtM bound to FSPP (PDB ID code 3W7F), PSPP (PDB ID code 3NPR) and DHS (PDB ID code 3NRI). Figures 10a–c show these FSPP, PSPP and DHS structures in which the ligands are shown as red/orange spheres (diphosphate) and the ligand carbon atoms are shown as cyan spheres. Mg²⁺ is green; the conserved Asp-rich domains are in dark blue, and the Asn/Gln/Tyr cluster residues are cyan. As observed from the “front” of the protein (Figures 10a–c), it is clear that the diphosphate groups are quite solvent exposed (Figures 10a,b) while the prenyl side-chains are buried in the hydrophobic interior. In the DHS structure, Figure 10c, the ligand is partially buried and in overlays with the FSPP or PSPP structures, it partially overlaps with their isoprenoid side-chains. There is no solvent exposure of any ligand (or Mg²⁺) on the back-side of the protein, as shown in the 180° rotations about the vertical axis, Figures 10d–f. However, we now see in these back-side views that there is a cyan patch which arises from partial exposure of Y129, the most essential residue in CrtM—a result that suggests that Y183 may simply be involved in stabilizing O440, rather than in direct H⁺ elimination.

Based on these structural observations, we propose the following “topological mechanism” for CrtM: 1) FPPs enters the active site from the front-side of the protein where there is clearly a deep cleft, via initial insertion of their hydrophobic farnesyl tails which then become completely buried in the interior of the protein, Figure 10a. 2) After ionization, condensation and H⁺ elimination, the S1 diphosphate exits from the front of the protein, perhaps by interacting with Mg²⁺ in the Mg²⁺/Asp/Glu cluster. There is no large pocket on the back of the protein. 3) H⁺ is eliminated from the back of the protein (via PPi/Asn/Gln/H₂O/Tyr), Figure 10d. Basically the same residues are again involved in the second ionization-elimination, of PSPP (Figures 10b,e) resulting in formation of dehydrosqualene, Figure 10c. In summary: FPP enters from the front-side; DHS and PPi exit from the front-side and H⁺ exits from the back-side, as shown in cartoon form in Figure 11. Initial penetration of the “detergent-like” or amphiphilic FPP isoprenyl-group-first seems most reasonable because if the diphosphate end entered first, it would block the cavity opening. PPi release from the front-side also seems very logical since the PPi moieties are close to the front-side surface, and exit may be facilitated via interactions with Mg²⁺. The observation that the very large dehydrosqualene molecule is seen in the front pocket also means that this is the likely entry/exit port for the isoprenoid substrates, and products. Very similar residue, ligand and Mg²⁺ locations are seen in HsSQS but in SQS, the back-side “exit” is not observable, perhaps because the CrtM and SQS reactions differ in one important aspect. In CrtM, PPi and H⁺ are eliminated in both the first and second half-reactions, Figure 4. However, in SQS, PPi is eliminated together with H⁺ in the first half-reaction (forming PSPP), but in the second half-reaction, although PPi is still released, there is no H⁺ elimination since a hydride (H[−]) from NADPH/NADH is used to reduce the (rearranged) carbocation to form the saturated (−CH₂CH₂−) central group in squalene, as

opposed to the central double bond that is formed in dehydrosqualene (where an H^+ is eliminated, rather than being retained). Since NADPH binding is known to speed-up the first half reaction in SQS³⁴ but is not actually involved in any first-half chemistry, it has been proposed³⁴ that it causes a conformational change, and this might expose the Y129 and facilitate H^+ elimination. And by default, since there are no pockets on the back-side of SQS, we propose that NAD(P)H and NAD(P)⁺ enter and exit from the front-side, Figure 11, consistent with the computational docking and mutagenesis results of Liu et al.³¹

CONCLUSIONS

The results we have presented above are of broad general interest since they lead to new insights into the mechanisms of action of ζ -fold head-to-tail, ζ -fold head-to-middle as well as α -fold head-to-head prenyltransferases, molecules that play key roles in the biosynthesis of e.g. bacterial cell walls, sterols, and carotenoids. We began by solving the structure of CLDS with two bound DMASPP ligands, finding a range of side-chain conformations, some of which can lead to cyclization. We then used site-directed mutagenesis and comparisons with related systems to show that an Asn-Ser pair is involved in catalysis in all ζ -fold prenyltransferases. Ionization is only feasible in the S1 site, due to bidentate chelation to Mg^{2+} . In S2, only P1O1 binds to the metal, enabling phosphate-mediated H^+ abstraction by P1O2, most likely (based on the results of DFT calculations) *via* an indirect mechanism involving Asn-Ser. The observations on Asn-Ser in ζ -fold proteins then led to the hypothesis that similar clusters of highly conserved (and where measured, essential for activity) amide (Gln, Asn) and OH-containing (Tyr) residues (or water) play a similar role in catalysis in “head-to-head” *trans*-prenyltransferases such as dehydrosqualene and squalene synthase. In these systems, Asn side-chains bind to Mg^{2+} , expected (based on DFT calculations) to promote formation of basic imidic acid species (imines) that may also be involved in H^+ elimination, analogous to the role of Asn in cellulase. We also found (in CrtM) that there is a potential Gln>H₂O>Tyr proton transfer network adjacent to the active site with (heavy atom) donor-acceptor distances of $\sim 3.8 \pm 1.0$ Å, about the same as in the ζ -fold proteins (3.8 ± 0.6 Å). These distances are consistent with the intermediacy of weak, electrostatic hydrogen bonds in the proton transfer mechanism. We thus propose that there are 3 major classes of groups involved in catalysis in many prenyltransferases: 1) the acidic groups (Asp, Glu) that bind to Mg^{2+} ; 2) the basic groups (Arg, His, Lys) that bind to diphosphate, and 3) neutral species (Asn, Gln, Ser, Tyr and H₂O) that in combination with diphosphate groups are involved in the 3rd step in isoprenoid biosynthesis: the H^+ -elimination that drives product formation. And finally, based on crystal structures and the other results discussed above, we propose topological models for FPP, dehydrosqualene, squalene, PPi, NAD(P)H, NAD(P)⁺ and H^+ exits or entrances in both CrtM and SQS.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grants 31700057, 31470240 and 31570130); CAS Interdisciplinary Innovation Team; the United States Public Health Service (National Institutes of

Health Grants GM065307 and CA158191 to EO, and GM085774 to YZ); by the Taiwan Protein Project (Grants MOST106-0210-01-15-04 and MOST107-0210-01-19-02) and by grants to Y.Z. from the National Key R&D Program of China (2017YFA1014000, 2017YFA1014001) and the National Natural Science Foundation of China (81573270).

ABBREVIATIONS

LPPS	lavandulyl diphosphate synthase
Mcl22	isosesquilavandulyl diphosphate synthase
CLDS	cyclolavandulyl diphosphate synthase
UPPS	undecaprenyl diphosphate synthase

References

1. Buckingham J. Dictionary of Natural Products on DVD, CRC. Boca Raton, FL: 2007.
2. Christianson DW. Structural and chemical biology of terpenoid cyclases. *Chem Rev.* 2017; 117(17): 11570–11648. [PubMed: 28841019]
3. Oldfield E, Lin FY. Terpene biosynthesis: modularity rules. *Angew Chem Int Ed Engl.* 2012; 51(5): 1124–1137. [PubMed: 22105807]
4. Poulter CD. Biosynthesis of non-head-to-tail terpenes. Formation of 1'-1 and 1'-3 linkages. *Acc Chem Res.* 1990; 23(3):70–77.
5. Lu YP, Liu HG, Liang PH. Different reaction mechanisms for *cis*- and *trans*-prenyltransferases. *Biochem Biophys Res Commun.* 2009; 379(2):351–355. [PubMed: 19103164]
6. Demissie ZA, Erland LA, Rheault MR, Mahmoud SS. The biosynthetic origin of irregular monoterpenes in lavandula: isolation and biochemical characterization of a novel *cis*-prenyl diphosphate synthase gene, lavandulyl diphosphate synthase. *J Biol Chem.* 2013; 288(9):6333–6341. [PubMed: 23306202]
7. Teufel R, Kaysser L, Villaume MT, Diethelm S, Carbullido MK, Baran PS, Moore BS. One-pot enzymatic synthesis of merochlorin A and B. *Angew Chem Int Ed Engl.* 2014; 53(41):11019–11022. [PubMed: 25115835]
8. Ozaki T, Zhao P, Shinada T, Nishiyama M, Kuzuyama T. Cyclolavandulyl skeleton biosynthesis via both condensation and cyclization catalyzed by an unprecedented member of the *cis*-isoprenyl diphosphate synthase superfamily. *J Am Chem Soc.* 2014; 136(13):4837–4840. [PubMed: 24640943]
9. Liu M, Chen CC, Chen L, Xiao X, Zheng Y, Huang JW, Liu W, Ko TP, Cheng YS, Feng X, Oldfield E, Guo RT, Ma Y. Structure and function of a “head-to-middle” prenyltransferase: lavandulyl diphosphate synthase. *Angew Chem Int Ed.* 2016; 55(15):4721–4724.
10. Tomita T, Kobayashi M, Karita Y, Yasuno Y, Shinada T, Nishiyama M, Kuzuyama T. Structure and mechanism of the monoterpene cyclolavandulyl diphosphate synthase that catalyzes consecutive condensation and cyclization. *Angew Chem Int Ed Engl.* 2017; 56(47):14913–14917. [PubMed: 28922556]
11. Gao J, Ko TP, Chen L, Malwal SR, Zhang J, Hu X, Qu F, Liu W, Huang JW, Cheng YS, Chen CC, Yang Y, Zhang Y, Oldfield E, Guo RT. New views of head-to-middle and *cis*- head-to-tail prenyltransferases: structure of isosesquilavandulyl diphosphate synthase. *Angew Chem Int Ed Engl.* 2017; 57:683–687. [PubMed: 29215779]
12. Guo RT, Ko TP, Chen AP, Kuo CJ, Wang AH, Liang PH. Crystal structures of undecaprenyl pyrophosphate synthase in complex with magnesium, isopentenyl pyrophosphate, and farnesyl thiopyrophosphate: roles of the metal ion and conserved residues in catalysis. *J Biol Chem.* 2005; 280(21):20762–20774. [PubMed: 15788389]
13. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 1997; 276:307–326.

14. Brunger AT. Assessment of phase accuracy by cross validation: the free R value. *Methods and applications. Acta Crystallogr D Biol Crystallogr.* 1993; 49:24–36. [PubMed: 15299543]
15. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Crystallogr.* 2007; 40(4):658–674. [PubMed: 19461840]
16. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Pottertton EA, Powell HR, Read RJ, Vagin A, Wilson KS. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr.* 2011; 67:235–242. [PubMed: 21460441]
17. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. PHENIX: a comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr.* 2010; 66:213–221. [PubMed: 20124702]
18. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr.* 2004; 60:2126–2132. [PubMed: 15572765]
19. Frisch MJT, GWSchlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda YKO, Nakai H, Vreven T, Montgomery JA, JrPeralta JE, Ogliaro F, Bearpark MJ, Heyd J, Brothers EN, Kudin KN, Staroverov VN, Kobayashi R, Normand J, Raghavachari K, Rendell AP, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam NJ, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Farkas Ö, Foresman JB, Ortiz JV, Cioslowski J, Fox DJ. Gaussian 09. Gaussian, Inc; Wallingford, CT: 2013.
20. Becke AD. Density-functional thermochemistry. III. The role of exact exchange. *J Chem Phys.* 1993; 98(7):5648–5652.
21. Glendening EDR, AECarpenter JE, Weinhold F. NBO, version 3.1 ed. Theoretical Chemistry Institute, University of Wisconsin; Madison, WI:
22. Zhu W, Zhang Y, Sinko W, Hensler ME, Olson J, Molohon KJ, Lindert S, Cao R, Li K, Wang K, Wang Y, Liu Y-L, Sankovsky A, de Oliveira CAF, Mitchell DA, Nizet V, McCammon JA, Oldfield E. Antibacterial drug leads targeting isoprenoid biosynthesis. *Proc Natl Acad Sci U S A.* 2013; 110(1):123–128. [PubMed: 23248302]
23. Wang W, Dong C, McNeil M, Kaur D, Mahapatra S, Crick DC, Naismith JH. The Structural Basis of Chain Length Control in Rv1086c. *J Mol Biol.* 2008; 381(1):129–140. [PubMed: 18597781]
24. Nakamura A, Ishida T, Kusaka K, Yamada T, Fushinobu S, Tanaka I, Kaneko S, Ohta K, Tanaka H, Inaka K, Higuchi Y, Niimura N, Samejima M, Igarashi K. “Newton’s cradle” proton relay with amide–imidic acid tautomerization in inverting cellulase visualized by neutron crystallography. *Sci Adv.* 2015; 1(7):1500263.
25. Chan Y-T, Ko T-P, Yao S-H, Chen Y-W, Lee C-C, Wang AHJ. Crystal structure and potential head-to-middle condensation function of a z,z-farnesyl diphosphate synthase. *ACS Omega.* 2017; 2(3): 930–936. [PubMed: 30023621]
26. Jeffrey GA. *An Introduction to Hydrogen Bonding.* Oxford University Press; New York, USA: 1997.
27. Lin FY, Liu CI, Liu YL, Zhang Y, Wang K, Jeng WY, Ko TP, Cao R, Wang AH, Oldfield E. Mechanism of action and inhibition of dehydrosqualene synthase. *Proc Natl Acad Sci U S A.* 2010; 107(50):21337–21342. [PubMed: 21098670]
28. Valdar WS. Scoring residue conservation. *Proteins.* 2002; 48(2):227–241. [PubMed: 12112692]
29. Schwalen CJ, Feng X, Liu W, BOD, Ko TP, Shin CJ, Guo RT, Mitchell DA, Oldfield E. Head-to-head prenyl synthases in pathogenic bacteria. *Chembiochem.* 2017; 18(11):985–991. [PubMed: 28340291]
30. Gu P, Ishii Y, Spencer TA, Shechter I. Function-structure studies and identification of three enzyme domains involved in the catalytic activity in rat hepatic squalene synthase. *J Biol Chem.* 1998; 273(20):12515–12525. [PubMed: 9575210]

31. Liu C-I, Jeng W-Y, Chang W-J, Shih M-F, Ko T-P, Wang AH-J. Structural insights into the catalytic mechanism of human squalene synthase. *Acta Crystallogr D Biol Crystallogr*. 2014; 70(2):231–241. [PubMed: 24531458]
32. p*K*_a values were estimated with the www.chemicalize.org server (www.chemaxon.com).
33. Liptak MD, Shields GC. Accurate p*K*(a) calculations for carboxylic acids using complete basis set and Gaussian-n models combined with CPCM continuum solvation methods. *J Am Chem Soc*. 2001; 123(30):7314–9. [PubMed: 11472159]
34. Radisky ES, Poulter CD. Squalene synthase: steady-state, pre-steady-state, and isotope-trapping studies. *Biochemistry*. 2000; 39(7):1748–1760. [PubMed: 10677224]

Figure 1. Structure of DMASPP (the non-reactive S-thiolo analog of the normal CLDS substrate dimethylallyl diphosphate, DMAPP) bound to CLDS, together with a ligand electron density map. (a) Stereo-view of dimer structure of CLDS+DMASPP (PDB ID code 5YGK; Chains A and B). (b) Stereo-view of active site region in CLDS showing some residues of interest, and ligand electron densities. PDB ID code 5YGK, Chain A. All 8 chains are shown in Figure 2.

Figure 2.

Ligand electron densities ($2F_o - F_c$; 1σ ; $RSR=0.108 \pm 0.015$; $LLDF = -0.78 \pm 0.25$) for the 8 chains in CLDS (PDB ID code 5YGK). (a)-(h) correspond to Chains A-H.

Figure 3.

Sequence alignments of CLDS, Mc122, LPPS, a tomato *cis*-FPPS, *E. coli* UPPS, the *cis*-FPPS Rv1086c, from *M. tuberculosis*, and Rv2361c, the *cis*-decaprenyl diphosphate synthase from *M. tuberculosis*. Highly conserved residues are in black; α -helices in red, β -sheets in green. Residues discussed in the Text are colored pink.

Figure 4.

Structure superpositions of CLDS (PDB ID code 5GUL) with 6 other ζ -fold prenyl transferases. (a) Mcl22 (PDB ID code 5XK9) rmsd = 0.862 Å. (b) LPPS (PDB ID code 5HC7) rmsd = 1.659 Å. (c) UPPS (PDB ID code 4H8E) rmsd = 1.667 Å. (d). Z-FPPS (PDB ID code 5HXT) rmsd = 2.009 Å. (e) Rv1086c (PDB ID code 2VG0) rmsd = 1.866 Å. (f) Rv2361c (PDB ID code 2VG3) rmsd = 2.069 Å. Ligands are sticks and Mg²⁺ are green spheres.

Figure 5.

Structural and mutagenesis results. (a) Structure superpositions of Mg^{2+} , ligands and selected residues in CLDS, Mcl22, LPPS, UPPS, Rv1086c and Rv2361c. The CLDS numbering system was used. (b) Structure of the diphosphate-bound form of CLDS (PDB ID code 5GUL) showing that inorganic diphosphates bind in the same way that the esterified substrates do, and that the conserved S54 and N57 have similar conformations to those found in the other ζ -fold proteins. (c) Effects of mutations of S54 and S57 in CLDS to Ala, on catalytic activity as determined by PPi release assay.

Model #	Input structures	Optimized Geometry	O-X (Å)	ΔE (kJ/mol)
1			2.8	0
2	<i>cis</i>		2.5	69
3	<i>trans</i>		2.6	50
4	<i>cis</i>		3.1	95
5	<i>trans</i>		2.9	43
6			2.9	10

Figure 6.

Density functional theory results for phosphate-amide interactions. Input chemical structures, geometry-optimized structures, O–X distances and interaction energies (relative to 1) are shown. *Cis* and *trans* refer to the CH/NH group geometries.

Figure 7. Sequence alignment of SaCrtM, human squalene synthase (HsSQS), *Alicyclobacillus acidocaldarius* 12-hydroxysqualene synthase (AaHpnC), *Neisseria meningitidis* presqualene diphosphate synthase (NmHpnD) and *Arabidopsis thaliana* phytoene synthase (AtPHY). Highly conserved residues are in black and the Asn, Gln and Tyr residues discussed in the Text are shown in pink. Orange = loop; red = helix.

Figure 8. Structure of SaCrtM and residues of interest. (a) Structure of SaCrtM (PDB ID code 3W7F) with bound farnesyl S-thiodiphosphate ligands. (b) Zoomed-in view of the proposed proton channel (Gln (Q165) > H₂O (O440) > Tyr (Y129) > Tyr (Y183) and the highly conserved asparagine, N168, that is coordinated to a Mg²⁺. All the distances shown (in black) are in Angstrom units.

Figure 9.

Optimized structures. (a) Acetamide. (b) Acetamide coordinated to a neutral complex $[\text{Mg}(\text{dimethylphosphate})(\text{acetate})(\text{H}_2\text{O})_3]$. (c) Acetamide anionic complex $[\text{Mg}(\text{acetamide})(\text{dimethylphosphate})(\text{acetate})(\text{H}_2\text{O})_3]^{-1}$ (d) Acetamide anion.

Figure 10.

Front-side and back-side views of SaCrtM with ligands and residues of interest. (a) FSP, front, PDB ID code 3W7F. (b) PSPP, front, PDB ID code 3NPR. (c) DHS, front, PDB ID code 3NRI. (d) FSP, back, PDB ID code 3W7F. (e) PSPP, back, PDB ID code 3NPR. (f) DHS, back, PDB ID code 3NRI. FSP=thiolodiphosphate-analog of FPP; PSPP=presqualene diphosphate; DHS=dehydrosqualene. Asp-rich domains are colored dark blue; the Asn/Gln/Tyr cluster residues are cyan; ligands are red/orange spheres for diphosphate, cyan spheres for carbon, Mg^{2+} are green.

Figure 11.

Schematic illustration of the entrances or exits of FPP, PPi, H⁺, dehydrosqualene, squalene, NADPH and NADP⁺ in SaCrtM or HsSQS.

Scheme 1.

General over-view of some of the species involved in isoprenoid biosynthesis. (a) Residues/species that are involved in substrate binding (pink), ionization/condensation/cyclization (cyan), and residues that may be involved in H⁺-elimination (green). (b) Structure of some compounds discussed in the Text. OPP = diphosphate fragment. SPP = thiolodiphosphate fragment.

Scheme 2.

Proposed mechanism for the CLDS-catalyzed reaction.

Scheme 3.

Four possible mechanisms for proton elimination in a ζ -prenyl transferase. (a) A direct phosphate-mediated enolate pathway. (b) A direct phosphate-mediated imine pathway. (c) An indirect phosphate-mediated enolate pathway. (d) An indirect, imine-mediated pathway.

Scheme 4.

Reactions catalyzed by dehydrosqualene synthase and squalene synthase. (a) Formation of presqualene diphosphate occurs in both systems in the first half-reaction. (b) Second half-reaction in dehydrosqualene synthase. (c) Second half-reaction in squalene synthase.

Scheme 5.

Asn- Mg^{2+} binding can lead to formation of an imine species. (a) Amide $\text{C}=\text{O}$ coordinates to Mg^{2+} . (b) Iminium/enolate/amidic acid binds tightly to Mg^{2+} , due to strong $\text{Mg}^{2+} \cdots \text{O}^-$ Coulombic interaction. (c) Loss of H^+ results in formation of the basic imine species. The pK_a values for imidic acid esters are 6 ± 1.5 .

Scheme 6.

A possible proton elimination mechanism in SaCrtM. (a) Proton abstraction from a carbocation XH^+ by a PPI in S1 is followed by H^+ transfer to N168 bound to Mg^{2+} , then is shuttled out of the protein *via* a relay. Y129 and Q165 are two of the three most conserved residues in CrtM-like proteins and Y129A and Q165A mutants are inactive. These residues, together with Y41, N168 and Y183 represent 5 of the 14 most highly conserved residues, the other 9 being Asp or Arg (that presumably interact with Mg^{2+} or diphosphates). A role for N168 in H^+ -elimination is based on DFT results indicating imine stabilization; a role for Q165 in H^+ -elimination is based on the observation that the corresponding Q>L mutant in SQS is inactive while the Q>E mutant (in which the Glu carboxyl can tautomerize) has very high activity. A role for diphosphate is proposed since it provides a bridge between the cationic center and the Q165 base. Y183 is likely to stabilize the bound H_2O (O440). (b) Protonation states after H^+ -removal (to H_2O).

Table 1

Data Collection and Refinement Statistics for CLDS

	<i>apo</i> -CLDS	CLDS/DMASPP
PDB code	5YGJ	5YGK
<i>Data collection</i>		
Space group	C2	P2 ₁
<i>a</i> , <i>b</i> , <i>c</i> [Å]	92.5, 118.4, 88.9	90.1, 119.3, 93.1
β (°)	116.35	117.35
Resolution [Å]	25-2.65 (2.74-2.65)	25-2.05 (2.12-2.05)
Unique reflections	24803 (2416)	109264 (10800)
Redundancy	3.5 (3.4)	5.3 (5.0)
Completeness [%]	99.1 (96.3)	99.7 (99.0)
Average $I/\sigma(I)$	25.8 (4.1)	28.9 (2.5)
R _{merge} ^a [%]	5.3 (36.4)	5.5 (49.9)
<i>Refinement</i>		
No. of reflections	24781 (2534)	109181 (10336)
R _{work} (95 % of data)	0.208 (0.322)	0.184 (0.248)
R _{free} (5 % of data)	0.268 (0.426)	0.235 (0.302)
r.m.s.d. bonds [Å]	0.012	0.009
r.m.s.d. angles [°]	1.7	1.4
Ramachandran plot		
Most favored [%]	96.6	98.0
Allowed [%]	3.3	2.0
Disallowed [%]	0.1	0
No. of non-H atoms/Average B [Å ²]		
Protein	6477/76.8	13262/41.7
Water	35/57.9	851/42.9
Ion		8/41.7
Ligand		224/48.1

Values in parentheses are for the highest resolution shell.

$$^a R_{\text{merge}} = \frac{\sum_{hk\ell} \sum_i |F_i(hk\ell) - \langle F(hk\ell) \rangle|}{\sum_{hk\ell} \sum_i F_i(hk\ell)}.$$

Table 2

Computed Geometric and Charge Properties for Amide Systems^a

System	R _{MgO} (Å)	R _{OC} (Å)	R _{NC} (Å)	R _{PI01} (Å)	R _{PI02} (Å)	Q _O (e)	Q _C (e)	Q _N (e)	Q _{O1} (e)	Q _{O2} (e)
Acetamide		1.217	1.365			-0.636	0.675	-0.812		
Acetamide anion		1.268	1.321			-0.836	0.583	-0.942		
Acetamide neutral complex	2.093	1.240	1.338	1.505	1.518	-0.702	0.708	-0.776	-1.172	-1.197
Acetamide anion complex	1.997	1.308	1.290	1.500	1.511	-0.924	0.600	-0.814	-1.165	-1.200

^a Bond lengths R between pairs of atoms are given in Å; Q values are elementary charges, e.