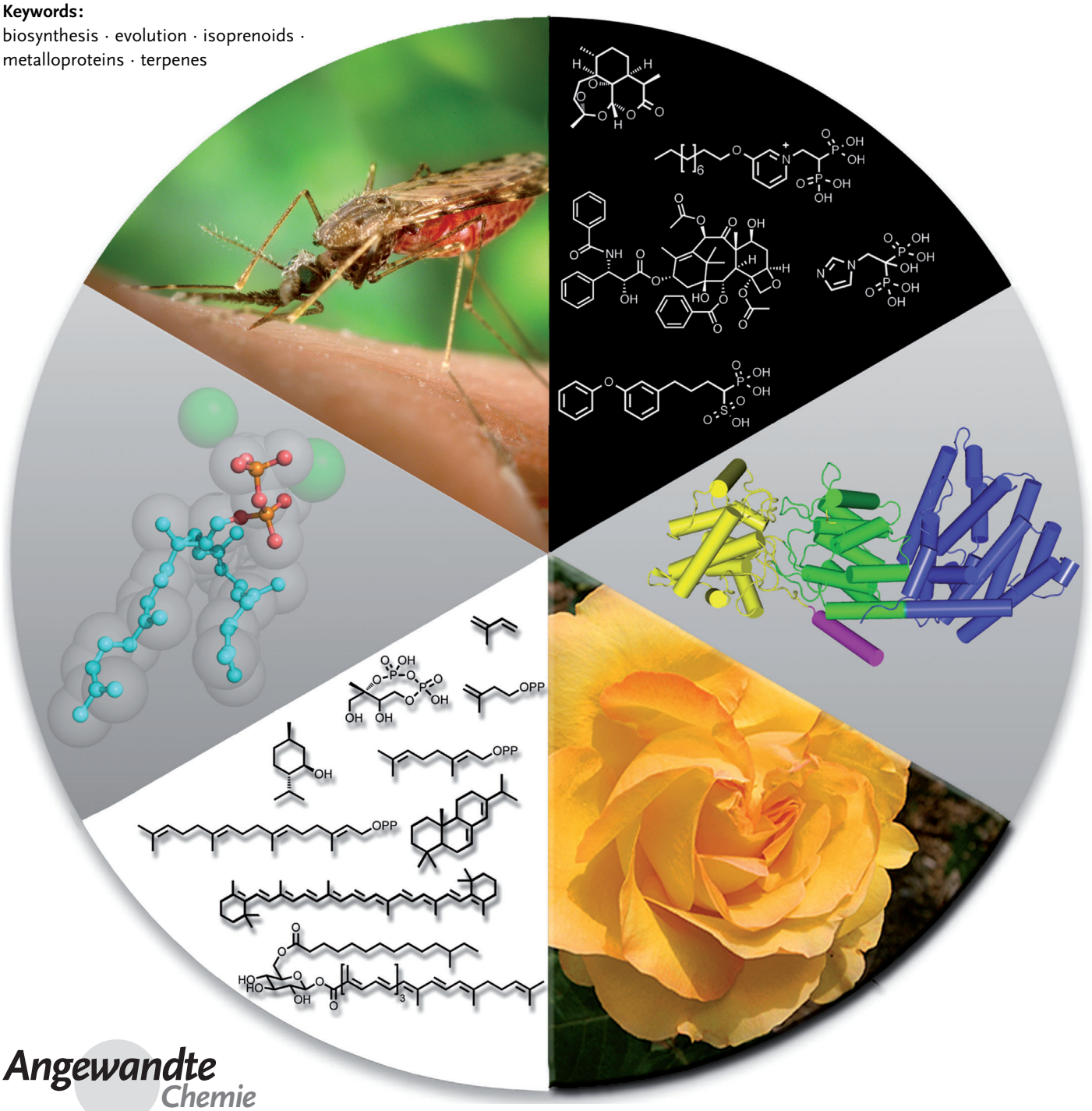
DOI: 10.1002/anie.201103110

Terpenes

Terpene Biosynthesis: Modularity Rules

Eric Oldfield and Fu-Yang Lin*

Keywords:
biosynthesis · evolution · isoprenoids · metalloproteins · terpenes



Angewandte International Edition Chemie

Terpenes are the largest class of small-molecule natural products on earth, and the most abundant by mass. Here, we summarize recent developments in elucidating the structure and function of the proteins involved in their biosynthesis. There are six main building blocks or modules (α, β, γ, δ, ε, and ζ) that make up the structures of these enzymes: the αα and αδ head-to-tail trans-prenyl transferases that produce trans*-isoprenoid diphosphates from C_5 precursors; the ε head-to-head prenyl transferases that convert these diphosphates into the tri- and tetraterpene precursors of sterols, hopanoids, and carotenoids; the βγ di- and triterpene synthases; the ζ head-to-tail cis-prenyl transferases that produce the* cis*-isoprenoid diphosphates involved in bacterial cell wall biosynthesis; and finally the α, αβ, and αβγ terpene synthases that produce plant terpenes, with many of these modular enzymes having originated from ancestral α and β domain proteins. We also review progress in determining the structure and function of the two 4Fe-4S reductases involved in formation of the C_5 diphosphates in many bacteria, where again, highly modular structures are found.*

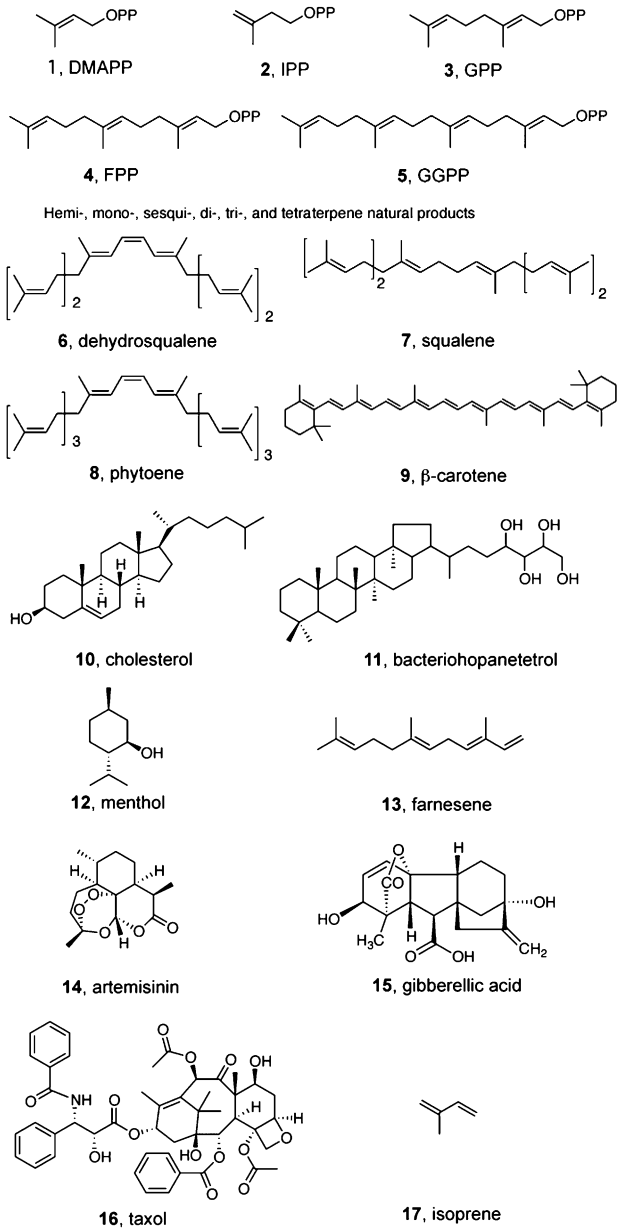
From the Contents

1. Introduction

Terpenes or isoprenoids are the most diverse class of natural products and are of interest since they are found in almost all life forms where they carry out a myriad of functions ranging from primarily structural (cholesterol in cell membranes) to functional (carotenoids in photosynthesis, retinal in vision, quinones in electron transfer).[1] Essentially all originate, at least in part, from the C_5 substrates dimethylallyl diphosphate (DMAPP, **1**; Scheme 1) and isopentenyl diphosphate (IPP, **2**), typically by initially condensing DMAPP with one or more IPP molecules in a 1′–4 or "head-to-tail" fashion to form (C_{10}) geranyl diphosphate (GPP, **3**), (C_{15}) farnesyl diphosphate (FPP, **4**), or (C_{20}) geranylgeranyl diphosphate (GGPP, **5**).[2] FPP and GGPP can then condense in a "head-to-head" fashion,[3] also termed tail-to-tail by some,[4] to form, for example, dehydrosqualene (DHS, **6**), squalene (**7**), or phytoene (**8**), the precursors of carotenoids such as β-carotene (**9**), sterols such as cholesterol (**10**), and hopanoids such as bacteriohopanetetrol (**11**)—some of the most ancient as well as abundant natural products.[1] Isoprenoids can also be used to posttranslationally modify proteins (of importance in cell signaling), or they can be cyclized to form the myriad terpene natural products: (C_{10}) monoterpenes such as menthol (**12**); (C_{15}) sequiterpenes such as farnesene (**13**) and artemisinin (**14**); and (C_{20}) diterpenes that are converted to, for example, gibberellic acid (**15**) and taxol (**16**). In addition, DMAPP is converted by plants to isoprene (**17**) itself at a rate of roughly 100 megatons per year, a reaction that is of current interest as a potential source of "renewable" fuels and other products.[5]

The DMAPP and IPP precursors are made in two different pathways: the mevalonate[6] and methylerythritol phosphate (MEP) pathways.[7] The mevalonate pathway is utilized by most eukaryotes (including humans) as well as archaebacteria,[8] while the MEP pathway is found in most eubacteria. There are of course exceptions. For example, the bacterium *Staphylococcus aureus* uses the mevalonate pathway, while malaria parasites, eukaryotes, use the MEP pathway.[9] In plants, both pathways are found,[7] with the MEP pathway typically operating in plastids while the mevalonate pathway operates in the cytosol: sterols (triterpenes) are produced by means of the mevalonate pathway while hemi-, mono-, and diterpenes, as well as carotenoids (tetraterpenes), are produced by means of the MEP pathway. In the following, we review recent developments in determining the structure and function of many of the key enzymes involved in isoprenoid biosynthesis: the head-to-head and head-to-tail prenyl transferases; the terpene synthases; as well as the 4Fe-4S reductases involved in DMAPP/IPP production in most eubacteria. These structures give important new insights into how the approximately 65 000 terpenoid natural products[10] are made. In particular, we propose that there are six major protein "building blocks" or modules (α, β, γ, δ, ε, and ζ) that are used—often in combination—to make

[*] Prof. Dr. Dr. E. Oldfield, Dr. F.-Y. Lin
Department of Chemistry and
Center for Biophysics and Computational Biology
University of Illinois at Urbana-Champaign
600 South Mathews Avenue, Urbana, IL, 61801 (USA)
E-mail: eo@chad.scs.uiuc.edu

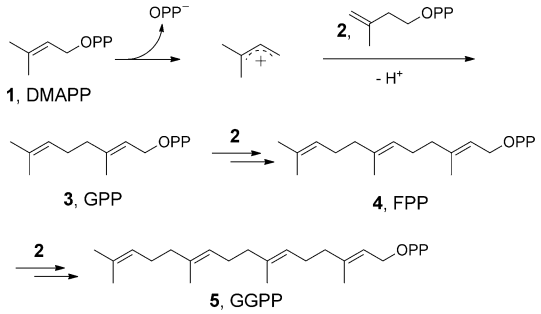


Scheme 1. Isoprenoid biosynthesis: substrates and products.

the enzymes responsible for formation of most known terpenes and isoprenoids.

2. Head-to-Tail trans-Prenyl Transferases: αα- and αδ-Domain Structures

DMAPP and IPP are the C_5 substrates used for terpene biosynthesis. They first condense to form the all-*trans* isoprenoid diphosphates GPP, FPP, and GGPP in reactions catalyzed by the enzymes geranyl diphosphate synthase (GPPS), farnesyl diphosphate synthase (FPPS), and geranylgeranyl diphosphate synthase (GGPPS): **1**+**2**→**3**→**4**→**5** (Scheme 2). The first of these structures to be solved[11] was



Scheme 2. Carbocation mechanism for the biosynthesis of GPP, FPP, and GGPP.

that of FPPS. The structure (Figure 1a) is almost entirely α-helical and there are two highly conserved repeats containing DDXXD residues (Figure 1a, in red). These are used to chelate $3\,Mg^{2+}$ ions[12] that, in turn, are responsible for ionization of the allylic substrate (DMAPP) to form a carbocation (Scheme 2), which then undergoes nucleophilic attack by the olefinic double bond in IPP, followed by H^+ elimination, to form GPP. The process then repeats to form FPP, then (with GGPPS) GGPP. DMAPP (and GPP) bind through a Mg^{2+} ion to the catalytic Asp in the allylic binding site in FPPS, while IPP binds through a cluster of cationic residues (R57, K60 in human FPPS) in the second, homoallylic site (Figure 1b).[12a]

Eric Oldfield received a BSc in Chemistry from Bristol University and a PhD in Biophysical Chemistry from Sheffield University under the direction of Dennis Chapman. He then worked as an EMBO Fellow at Indiana University with Adam Allerhand and at MIT with John Waugh. He joined the Chemistry Department at the University of Illinois at Urbana-Champaign in 1975 and is currently the Alumni Research Scholar Professor of Chemistry. His research interests are in NMR spectroscopy and drug discovery.

Fu-Yang Lin received his BS and MS degrees in Life Science from National Central University, Taiwan, and a PhD in Biophysics and Computational Biology from the University of Illinois at Urbana-Champaign under the direction of Eric Oldfield. He studied the structure, catalytic mechanism of action, and inhibition of prenyl synthases in the Oldfield Group. He is currently a postdoctoral fellow at Harvard Medical School.

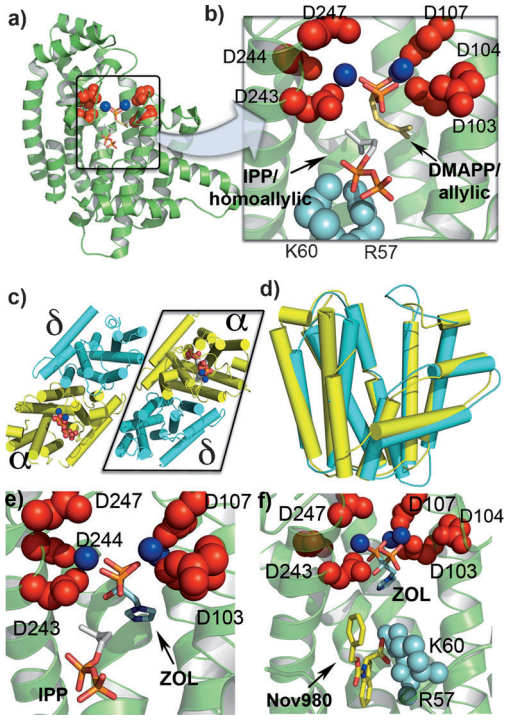


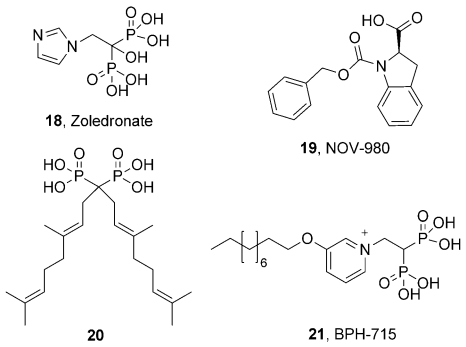
Figure 1. a) Structure of human FPPS (PDB ID code: 1ZW5) showing conserved DDXXD motifs (red), Mg^{2+} (blue), and IPP (bottom) and S-thiolo DMAPP ligands (top, from superposition with PDB ID: 1RQI). b) Expansion of the active-site region in (a), catalytic residues in red and cyan. c) Heterotetramer structure of *M. piperata* GPPS (PDB ID: 3KRF) showing catalytic (α, in yellow) and regulatory (δ) subunits. d) Superposition of the α,δ domains in GPPS. e) Zoledronate and IPP bound to the active site of human FPPS (1ZW5): color code as in (a). f) Zoledronate and **19** (NOV-980) bound to the allylic (ZOL) and allosteric (NOV-980) sites in human FPPS (PDB ID: 3N46).

FPPS and most GGPPS molecules function as homodimers (αα) with, in some cases, residues from both chains making up the catalytic site.[13] However, it has recently been found that the C_{10} isoprenoid synthase, GPPS (which provides the GPP needed for menthol biosynthesis), found in plants such as peppermint and spearmint, is a much more complicated system since it contains not one, but two distinct subunits,[14] both of which are required for activity: a large subunit (α) containing the $DD(X)_nD$ catalytic machinery and a smaller, regulatory subunit (herein called δ) that governs chain elongation. This type of heterodimer organization is absent in GPPS from *Abies grandis*,[15] but is also found in human decaprenyl diphosphate synthase,[16] which produces the C_{50} isoprenoid diphosphate required for CoQ_{10} biosynthesis

Chang et al.[17] have now reported the X-ray crystallographic structure of GPPS from *M.* × *piperita*, a likely prototype for other heterodimeric systems. The structures reveal a novel architecture in which the large and small subunits form a heterodimer that, in turn, dimerizes to form a tetramer ($\alpha_2\delta_2$; Figure 1c). The structures contain Mg^{2+} ions, IPP, S-*thiolo*-DMAPP (a nonhydrolyzable DMAPP analogue), and GPP, all of which bind only to the large, catalytic α subunit (Figure 1c, in yellow). The α-domain fold is quite similar to that found in FPPS, with a root-mean-square deviation (rmsd) at C^α of 2.8 Å. In peppermint GPPS, the regulatory subunit inhibits chain elongation beyond C_{10}, although with FPP as substrate, GGPP can form in vitro. In previous work,[14] it was suggested that these plant GPPS evolved from GGPPS, based on the observation of much larger sequence homology of GPPS with GGPPS than with FPPS (75% versus 25%), a result now supported by the smaller C^α rmsd values for the GPPS α-subunit relative to GGPPS over that relative to FPPS (0.9 Å versus 2.7 Å). What is more surprising about the new GPPS results is that there is also a remarkably close structural similarity between the catalytic (α) and regulatory (δ) subunits in the tetramer, corresponding to a 1.87 Å C^α rmsd (Figure 1d). This strong structural similarity between the α and δ domains, together with a 32% identity and 50% sequence similarity, suggests that such αδ proteins may have originated by means of a gene duplication, just as with the βγ proteins involved in the terpene synthase reactions discussed below. This αδ catalytic/regulatory domain organization has also now been reported in a second system, hexaprenyl diphosphate synthase from *Micrococcus luteus*.[18] The δ domain there is quite small (7 helices versus 17 in the α domain) and there is a 2.1 Å C^α rmsd between the α and δ domains. The small subunit helps stabilize the dimer through hydrophobic interactions, as well as directly regulating product chain length,[18] and based on these results and those with GPPS, it seems likely that similar structures will be found with human DPPS as well.

3. FPPS and GGPPS as Drug Targets

FPPS is of great pharmaceutical interest since it is an important drug target. The bisphosphonates used to treat osteoporosis (and of recent interest in cancer therapy and immunotherapy[19]) such as Zoledronate (**18**, Scheme 3) target the allylic site in FPPS, binding as with DMAPP (Figure 1b) to the $[Mg^{2+}]_3$ cluster (Figure 1e).[12,20] This blocks FPP and GGPP biosynthesis and, consequently, prenylation of proteins such as Ras, resulting in tumor cell killing,[21] inhibition of invasiveness,[22] phenotype switching in macrophages from a tumor-promoting M2 to a tumor-killing M1 phenotype,[23] and also γδ T-cell activation,[24] with activated γδ T cells killing tumor cells.[25] These combined effects are



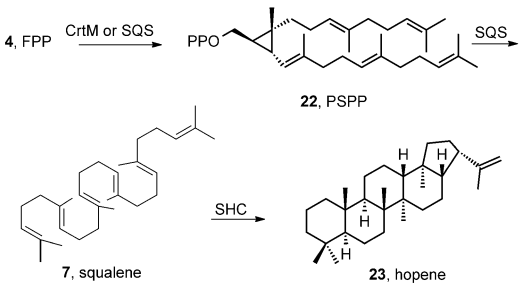
Scheme 3. FPPS and GGPPS inhibitors.

thought to contribute to a relative reduction of 36% in the risk of disease progression in breast cancer patients treated with an aromatase inhibitor plus the bisphosphonate Zoledronate, relative to those treated with aromatase therapy alone.[19] Bisphosphonates are not, however, conventionally druglike, because of their extreme polarity and high bone-binding affinity,[26] so there has recently been considerable interest in developing new, more lipophilic FPPS inhibitors.[27] Jahnke et al.[28] reported the discovery of a third, allosteric site in FPPS, together with a new generation of inhibitors (such as **19**) that bind to this site[28] (Figure 1 f). These inhibitors bind with their polar groups in or close to the IPP diphosphate (PP_i) site (Figure 1 f), and have IC_{50} values as low as 80 nM.[28] Such new-generation non-bisphosphonate FPPS inhibitors lack the structural features needed to bind to bone mineral,[26] so have great potential as anticancer agents.

In addition to FPPS, GGPPS is also a drug target. Both FPPS as well as GGPPS have αα structures and there is only a 2.4 Å C^{α} rmsd between human FPPS and GGPPS (Figure 2 a[20,29]). Surprisingly, however, bisphosphonates such as Zoledronate do not inhibit human (or yeast) GGPPS, due to the absence of one Asp residue in the second Asp-rich cluster (DDXXN, instead of DDXXD). This absence inhibits binding of the third Mg^{2+} ion.[30] Zoledronate does, however, bind to a GGPPS that has the extra Asp, from the malaria parasite *Plasmodium vivax* (Figure 2 b[30]). More-lipophilic bisphosphonates such as **20** and **21** (Scheme 3) bind to yeast, human, and *P. vivax* GGPPS,[30,31] as illustrated, for example, in pink in Figure 2 c,[31] where the long, hydrophobic side chain binds in the same site[29] as does the GGPP product (Figure 2 c, in cyan). These lipophilic bisphosphonates are expected to exhibit better cell/tissue penetration and weaker bone binding[26] than do conventional bisphosphonates, and indeed, they are far more effective in killing tumor cells[26a] and malaria parasites[32] than is, for example, Zoledronate both in vitro and in vivo.

4. The ε Head-to-Head Prenyl Transferases

The isoprenoid diphosphates produced by GPPS, FPPS, and GGPPS can be cyclized by a wide variety of terpene synthases (see the following sections), and the C_{15} and C_{20} diphosphates can also be condensed in a 1′–2,3 or "head-to-head" fashion[3] to form C_{30} and C_{40} hydrocarbon species. These are the precursors of sterols, carotenoids, and hopanoids, whose diagenetic products are among the most abundant small-molecule organic compounds on the planet (ca. 10^{12} tons present, in sediments).[33] With FPP, the initial condensation product (Scheme 4) is the C_{30} diphosphate



Scheme 4. Converting FPP to cyclic products.

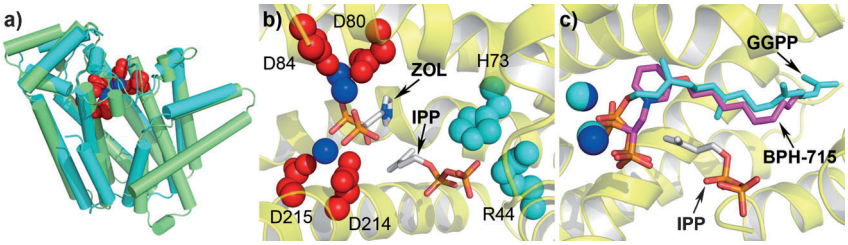


Figure 2. Structures of GGPPS. a) Superposition of human FPPS (green; PDB ID: 2F8C) and human GGPPS (cyan; PDB ID: 2Q80) with the Asp-rich domain (red) and Mg^{2+} ions (blue) highlighted as spheres. b) Zoledronate (ZOL) and IPP bound to the active site in *P. vivax* GGPPS (PDB ID: 3LDW). c) Compound **21** (BPH-715, pink), IPP, Mg^{2+} bound to *S. cerevisiae* GGPPS (PDB ID: 2Z4V) superimposed on GGPP (cyan; PDB ID: 2Q80) bound to the product site. Human GGPPS has a very similar local structure and is potently inhibited by **21**, but not by Zoledronate.

presqualene diphosphate (PSPP, **22**), formed in a reaction catalyzed by either squalene synthase (SQS) or dehydrosqualene synthase (CrtM).

In plants, animals, fungi, and some bacteria, PSPP then undergoes a Mg^{2+}-dependent ionization and loss of PP_i, ring opening, and reduction (by NADPH) to form squalene, the precursor for sterols such as sitosterol, cholesterol, and ergosterol, as well as many hopanoids such as hopene (**23**). In the bacterium *S. aureus*, the reductive step is missing and the product is dehydrosqualene (**6**), the precursor of the carotenoid virulence factor staphyloxanthin, a target for anti-infective development.[34] In plants, the C_{20} diphosphate GGPP condenses in a similar manner to form (C_{40}) prephytoene diphosphate and thence, phytoene, the precursor of carotenoids such as β-carotene.[35]

Given the abundance and importance of sterols, carotenoids, and hopanoids, it is surprising that, until very recently, the only known structure of a head-to-head prenyl transferase was that of human SQS[36] (Figure 3 a, in orange). As with FPPS, the structure is highly α-helical (with a 3.5 Å C^{α} rmsd versus FPPS for 189 residues). However, the SQS structure gave relatively little mechanistic information since the inhibitor used was not obviously substrate- or product-like. More recently, the structure of the *S. aureus* dehydrosqualene synthase enzyme, CrtM (in the presence of the nonreactive, FPP-substratelike inhibitor, S-*thiolo*-farnesyl diphosphate, FSPP), was reported.[34] The overall fold (which we will call ε) is similar to that seen in SQS (2.7 Å C^{α} rmsd; Figure 3 a)

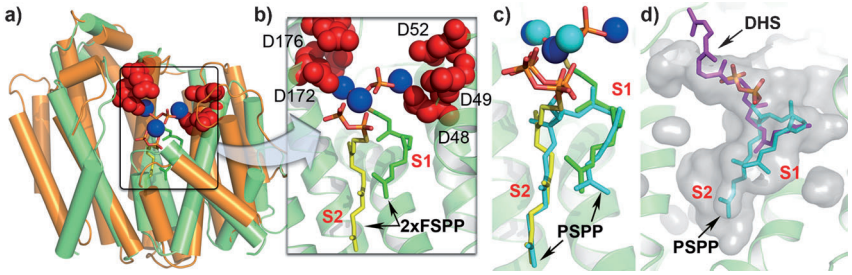


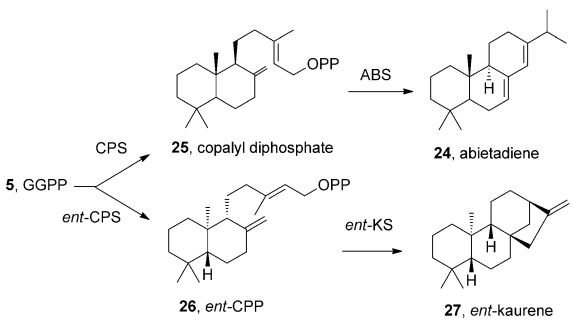
Figure 3. Structures of CrtM and SQS. a) *S. aureus* CrtM (green; PDB ID: 2ZCP) with FSPP, Mg^{2+} superimposed on human SQS (orange; PDB ID: 1EZF); essential Asp residues and Mg^{2+} ions colored as in Figure 1. b) Active-site region in CrtM + FSPP (green, yellow), Mg^{2+}. c) PSPP, Mg^{2+} (all in cyan; PDB ID: 3NPR) bound to CrtM, superimposed on the FSPP/Mg^{2+} structure (in green/yellow/blue; PDB ID: 2ZCP). d) Dehydrosqualene product (pink; PDB ID: 3NRI) bound to CrtM, superimposed on PSPP structure (cyan). S1 = allylic binding site; S2 = homoallylic binding site.

and there are two FSPP ligands and three Mg^{2+} ions (Figure 3b). But which FPP ionizes to form the farnesyl cation, and which acts as the nucleophile that reacts with the carbocation? This is not clear by inspection of the FSPP X-ray structure (Figure 3b) since the two sets of possible cation–C(olefin) distances are both approximately 5.5 Å.[34]

Fortunately, the X-ray crystallographic structure of the PSPP intermediate bound to CrtM has now been reported.[37] The results obtained (Figure 3c, in cyan), show that FPP in the so-called S1 site is likely to ionize and then react with the double bond in the FPP in the S2 site to form the cyclopropyl carbinyl diphosphate (PSPP, Figure 3c). The PSPP diphosphate then "flips" back to the $[Mg^{2+}]_3$ cluster and undergoes a second ionization, ring opening, and H^+ loss, forming dehydrosqualene, which has now been detected in a surface pocket (Figure 3d, in purple). This mechanism is supported by the results of site-directed mutagenesis[37] and the observation that superimposing the FSPP/Mg^{2+} CrtM structure on that of prenyl synthases (FPPS, GGPPS) whose mechanisms are known, places the S1 site in the "allylic" position found in those enzymes, as well as in terpene cyclases, whose mechanisms are also known.[37] In addition, when CrtM, FSPP, and FPP are mixed, S-*thiolo*-PSPP (but no dehydrosqualene) is produced, consistent with FPP ionizing in S1 and FSPP being a good nucleophile, in S2. S-*thiolo*-PSPP is unable to ionize in the allylic site, just as with S-*thiolo*-diphosphate inhibitors of other prenyl synthases.[38] The observation of the FPP substrate and PSPP intermediate binding sites, as well as the observation that potent SQS inhibitors (of interest as anti-infectives) also inhibit CrtM and have large hydrophobic interactions in both S1 and S2 sites,[37] opens up new routes to developing anti-infective drug leads that target sterol biosynthesis,[39] as well as targeting virulence factor formation in *S. aureus*.[34] However, work still remains to be done to solve where and how the NADPH reduction step occurs, in SQS.

5. Diterpene Cyclases: The "αβγ-Fold" Hypothesis

Most terpenes contain ring structures and are made by terpene synthases that are generally referred to as terpene cyclases. There was a burst of activity in this area several years ago when the structures of the (C_{30}) triterpene cyclases squalene–hopene cyclase[40] (SHC) and oxidosqualene cyclase,[41] the (C_{15}) sesquiterpene cyclases epi-aristolochene synthase (EAS[42]) and pentalene synthase,[43] and the (C_{10}) monoterpene synthase bornyl diphosphate synthase (BS[44]) were reported. However, the structures of the (C_{20}) diterpene cyclases have been much more difficult to obtain, but are of interest since they are involved in, for example, taxol and gibberellin biosynthesis. To try and circumvent this lack of direct structural information, bioinformatics and mutagenesis experiments aimed at elucidating some of the key features of diterpene cyclase structure and function were recently reported.[45] This work was stimulated by an earlier genomics study[46] which indicated that an ancestral diterpene cyclase might be the progenitor of modern plant terpene cyclases, as well as by the observation[47] that there were structural similarities between a triterpene cyclase and a sesquiterpene cyclase, and similar observations[48] that there were sequence similarities between the bifunctional diterpene cyclase abietadiene (**24**) synthase (whose structure has not been reported), and that of the sesquiterpene cyclase epi-aristolochene synthase, whose structure is known. The diterpene cyclases catalyze two types of reactions. In class II cyclases, GGPP (**5**) is protonated to form a carbocation which then cyclizes to form, for example, copalyl diphosphate (**25**; Scheme 5). This reaction is known to be catalyzed by a DXDD (not a DDXXD) catalytic motif and is chemically similar to the protonation/cyclization reaction catalyzed by, for example, squalene–hopene cyclase (**7**→**23**; Scheme 4)—which also has a highly conserved DXDD catalytic domain. In the class I cyclases, catalysis is fundamentally different and involves the same type of DDXXD/$[Mg^{2+}]_3$ domains[12b] as seen in the head-to-head and *trans*-head-to-tail prenyl transferases: the products are the very diverse range of monoterpenes, sesquiterpenes, and diterpenes found in plants. The third class of terpene cyclases are the "mixed" class I + II cyclases such as abietadiene synthase and levopimaradiene synthase, which can carry out both protonation-initiated as



Scheme 5. Formation of diterpenes from GGPP.

well as ionization-initiated reactions. But what might the structures of these, or indeed any other, diterpene cyclases, be?

To begin to investigate this question, Cao et al.[45] followed up on earlier observations that many terpene cyclases (such as EAS) contain a highly α-helical catalytic domain (DDXXD/Mg^{2+}) linked to a vestigial N-terminal region, a pattern found in many other proteins including bornyl diphosphate synthase and more recently, isoprene synthase.[49] The DDXXD-containing catalytic domain we call α, since there is considerable three-dimensional structural similarity between the α-domain protein FPPS and this domain in such terpene cyclases, for example, a 3.4 Å C^{α} rmsd between human FPPS and IS. In the hemi-, mono-, and sesquiterpene cyclases there is also, in general, a second helical domain we call β that itself has structural homology with the barrel structure found in squalene–hopene cyclase[45] (Figure 4a). This suggested that plant diterpene cyclases might contain not only α (Figure 4b) and β domains but also—since the β domain in the mono- and sesquiterpene cyclases has structural similarity to the β domain in SHC and the plant diterpene cyclases are very large, a third helical γ domain as well—since SHC itself contains two β-barrel domains.[40] The diterpene cyclases could then have originated by fusion of the genes of α- and βγ-domain proteins,[45] as illustrated in the hypothetical αβγ structure shown in Figure 4c. This "structure" (obtained from a SHC/EAS/FPPS alignment) lacks a covalent bond between the α and βγ domains, but the C terminus of SHC is only about 2.5 Å from the N terminus of FPPS, as highlighted in orange in Figure 4c.

In class I diterpene cyclases such as taxadiene synthase (TXS), just the conserved DDXXD α domain (blue) would be functional, even though the βγ domains would be present (and would likely be important for folding). In the class II diterpene cyclases, the β (green, Figure 4c) and γ (yellow) domains would be functional, but the α domain would not be—except, again, for a likely role in folding—since the DDXXD domain is absent. In the bifunctional class I + II diterpene cyclases, all three domains would be present and involved in catalysis, with the βγ domain catalyzing cyclization of GGPP and the α domain processing the product of the first reaction.

These structural ideas received support from the observation that many bacteria[50] produce gibberellins (diterpenes), but in bacteria their biosynthesis is catalyzed by two separate enzymes: a class II diterpene cyclase, *ent*-copalyl diphosphate synthase (*ent*-CPPS; GGPP → *ent*-CPP, **26**) and a class I diterpene cyclase, kaurene synthase (KS; *ent*-CPP → *ent*-kaurene, **27**); in *Bradyrhizobium japonicum*, the two open reading frames coding for these different proteins overlap by a single nucleotide. What is of interest with this *ent*-CPPS is that it contains not only the DXDD catalytic motif, but also two "QW" motifs or foldons, characteristic of a β barrel. A β barrel has six inner and six outer helices, so there should be 24 helices for a βγ structure; using JPRED[51] and COUDES[52] bioinformatics computer programs, 23 of these were detected.[45] But are these structural ideas correct?

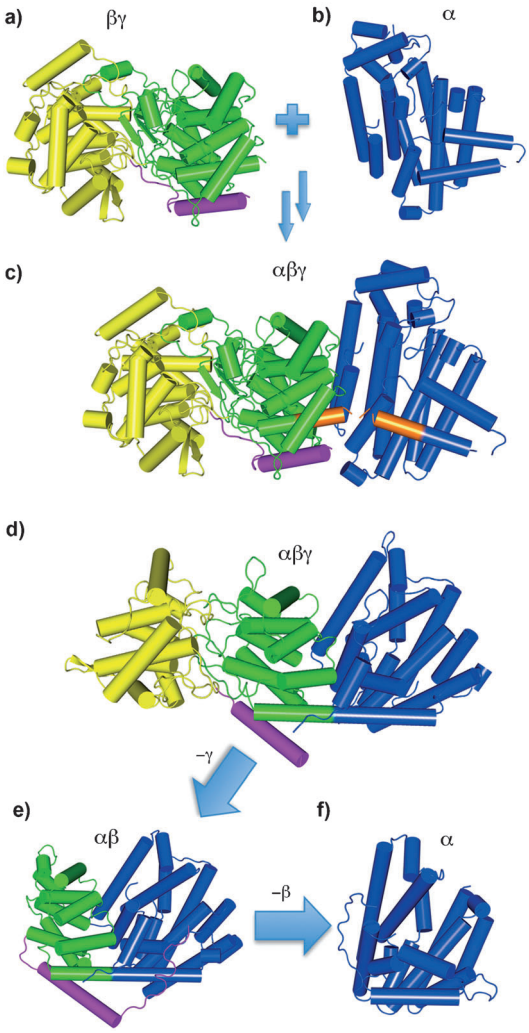


Figure 4. Genesis and evolution of terpene cyclases. a) Genes for ancestral βγ-domain proteins (like SHC; PDB ID: 3SQC) fuse with genes for ancestral α-domain species like FPPS (PDB ID: 1ZW5); b) a diterpene cyclase with three helical domains (αβγ) is generated. c) Orange shading indicates the close proximity (ca. 2.5 Å) of the SHC C terminus and the FPPS N terminus (from an α/αβ/βγ FPPS/EAS/SHC alignment). d) Structure of an actual diterpene cyclase, taxadiene synthase[53] (PDB ID: 3P5P). e) Loss of the γ domain yields an αβ protein, for example, the sesquiterpene cyclase isoprene synthase (PDB ID: 3N0F). f) Further loss of the β domain yields other cyclases such as pentalenene synthase (PDB ID: 1PS1), an α-domain cyclase. Ancestral α- and βγ-domain species presumably produced the FPP, GGPP, and squalene used to produce lipids in archaebacteria; the αβγ-derived families are much later arrivals. Note the N-terminal helix (magenta) portion is conserved in αβ, βγ, and αβγ proteins and is known to be required for activity.

6. Taxadiene Synthase: Structure of an αβγ Fold, and Evolution to the αβ-Domain Proteins

The very recent solution[53] of the first single-crystal X-ray crystal structure of a diterpene cyclase, taxadiene synthase (TXS), supports the structural proposals described above. As predicted, TXS does in fact contain a three-helical domain, αβγ structure (Figure 4d). In TXS, only the DDXXD motif is present since TXS is a class I cyclase, and the more ancestral

DXDD catalytic motif is absent. It is thus remarkable that the αβγ structure is still preserved, even though the β and γ domains play no role in catalysis per se, though of course may be important for folding. Indeed, it was recently shown using chimeras[45] of a plant (αβγ) CPS that while the α domain is required for activity, it has no effect on the stereochemical outcome of the actual βγ-domain-catalyzed reaction, and is thus only likely to be important for protein folding/stability.

These TXS structural results support the evolutionary proposal put forth previously[45] that an ancestral (class II) βγ triterpene cyclase (like SHC) may have evolved to a more modern bacterial class II diterpene cyclase, which then fused with an ancestral class I cyclase to form a bifunctional, abietadiene synthase-like diterpene cyclase, the progenitor (after exon loss and recombination[46]) of many modern mono-, sequi-, and diterpene cyclases, in addition to isoprene synthase itself.[49] In TXS, the α domain is quite similar to that found in FPPS (a 3.4 Å C^{α} rmsd), but the γ domain clearly has fewer helices present than expected for a "complete" βγ-barrel structure. In many plant terpene cyclases, as well as in isoprene synthase, the γ domain is completely absent but the β domain remains, even though it does not play a direct role in catalysis.

These are the αβ-domain proteins. They contain a very long helix "bridge" that forms part of both the α and β domains (Figure 4d). And, as noted above, it appears likely that this bridge may have arisen by fusion of the C terminus of a βγ-domain protein with the N terminus of an α-domain protein (orange, in Figure 4c). In the case of the αβ protein isoprene synthase, there is positive cooperativity which has been attributed to formation of a dimeric, quaternary structure: $\alpha_2\beta_2$. This dimer is present both in solution as well as in the solid state. Strikingly, the X-ray structure of isoprene synthase as well as two monoterpene cyclases, limonene synthase and bornyl diphosphate synthase, have almost identical $\alpha_2\beta_2$ quaternary structures,[49] as can be seen in Figure 5. Mechanistically, it has been proposed that in isoprene synthase, the diphosphate group acts a general base, abstracting one of the methyl protons in the DMAPP (**1**) substrate to form isoprene (**17**; Scheme 6). This elimination step would then be analogous to that yielding farnesene, a potentially important diesel-fuel substitute, from FPP, a reaction catalyzed by farnesene synthase, another (predicted) αβ protein. The molecular basis of the cooperativity found in isoprene synthase remains, however, to be elucidated. After loss of the β domain, the α-domain cyclases such as pentalenene synthase[43] form (Figure 4f), as proposed earlier.[46]

The solution of the TXS structure is thus a major development since it strongly supports previous genomics- and bioinformatics-based ideas[45,46] about the genesis, as well as the evolution, of many modern plant terpene synthases, in addition to giving some confidence in the use of bioinformatics tools to correctly predict structure and function. Moreover, in more recent work, the structure of an ent-copalyl diphosphate synthase has been reported.[54] As with TXS, it is an αβγ three-helical-domain protein, but in this case, has the active site at the interface between the β and γ domains.

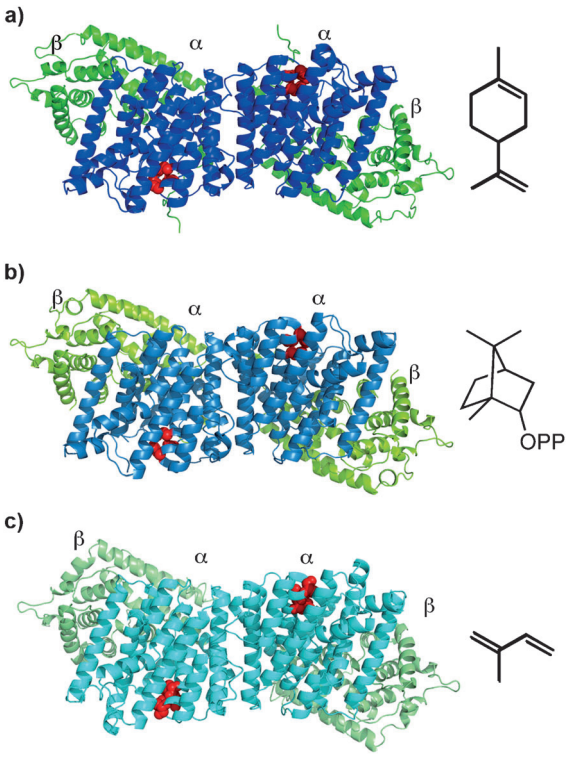
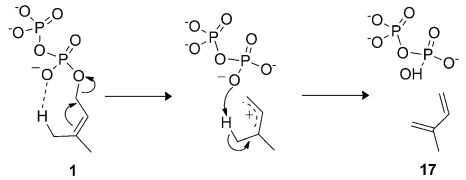


Figure 5. Dimeric quaternary structure of three $\alpha_2\beta_2$ terpene synthases. a) Limonene synthase and product limonene. b) Bornyl diphosphate synthase and product bornyl diphosphate. c) Isoprene synthase and product isoprene. The catalytic, α, or C-terminal domains are in blue, the β or N-terminal domains are in green, and the catalytic DDXXD domains are in red. The buried surface areas that comprise the dimerization interface are large, (1148 ± 88) A^2. The C^{α} rmsd between the three structures is 1.4 Å. This figure is adapted from Figure 8 in Ref. [49] and was constructed from the Protein Data Bank entries 2ONG, 1N1B, and 3N0F.



Scheme 6. Diphosphate acts as a general base in the conversion of DMAPP to isoprene, catalyzed by isoprene synthase.

7. The ζ (Z or cis) Prenyl Diphosphate Transferases and Tuberculosinol Synthase

So far, we have only considered the structures and function of the trans-prenyl transferases and some terpene cyclases. There is, however, another important class of prenyl transferases, the Z or cis-prenyl transferases, which catalyze formation of isoprenoid diphosphates containing primarily *cis* double bonds. These enzymes are essential for cell wall biosynthesis in bacteria, and as such are potentially important targets for the development of anti-infectives. The protein fold, herein called the ζ fold (for Z), is completely different to

that found in the "FPPS-like" α, δ, and ε prenyl transferases.[55]

In undecaprenyl (C_{55}) diphosphate synthase (UPPS) from *E. coli* there is a central β sheet with six parallel strands and seven surrounding α helices (Figure 6a).[56] The FPP and IPP substrates bind as shown in Figure 6b and, unlike the trans-prenyl transferases and terpene cyclases, there are no conserved DDXXD motifs and no $[Mg^{2+}]_3$ cluster, although

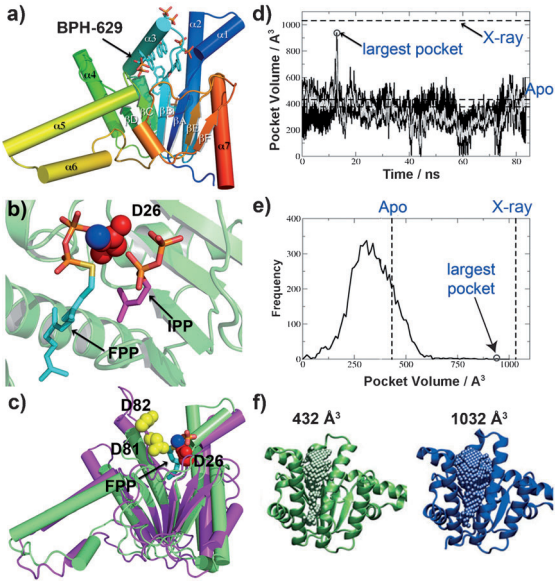
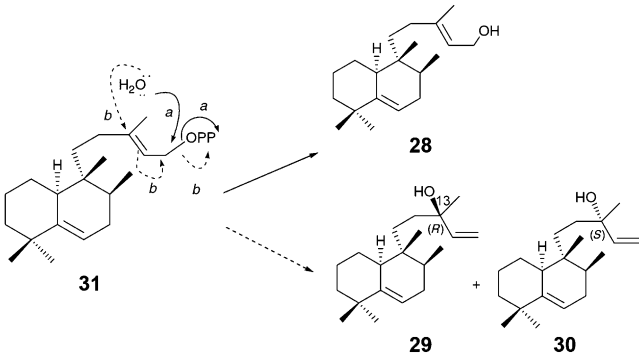


Figure 6. Structures and dynamics of the ζ prenyl transferase UPPS. a) Overall structure of a bisphosphonate-bound UPPS monomer from *E. coli* (PDB ID: 2E98). b) Substrates (FPP and IPP) bound to UPPS active site (PDB ID: 1X06). c) Structural alignment of the predicted structure of Rv3378c with UPPS. d) Molecular dynamics simulation of UPPS; data recorded every 10 ps in black, and every 100 ps in gray. e) Frequency of pocket occurrence versus pocket volume. The apo structure has a small pocket volume; the largest volume is close to that occupied by a large inhibitor. d)–f) from Ref. [65].

Mg^{2+} is required for catalysis.[57] These results suggest a mechanism for UPPS catalysis different from the sequential ionization—condensation–elimination mechanism observed in the trans-prenyl synthases. In recent work, Lu et al.[57] have shown that with IPP as substrate, there is no evidence for formation of a farnesyl carbocation intermediate (no formation of [^{3}H]-farnesol formation from [^{3}H]-FPP,) with either the trans-prenyl transferase octaprenyl diphosphate synthase (OPPS) or with UPPS, but when the reaction rate is decreased by using 3-Br IPP, [^{3}H]-farnesol forms with OPPS, but not with UPPS.[57] Since 3-Br-IPP slows down the UPPS reaction, it was proposed that cationic character develops on C3 of IPP after condensation, a concerted mechanism in which IPP attacks FPP without accumulation of a farnesyl carbocation. These results are consistent with the observation that while some of the most potent UPPS inhibitors are bisphosphonates, there is no cationic feature in the UPPS inhibitor pharmacophore,[56] unlike the situation with FPPS. Indeed, the presence of a cationic feature actually reduces activity by about one order of magnitude.[56] The key inhibitor features are thus the presence of multiple hydrophobic features, in addition to the polar group, with hydrophilic bisphosphonate drugs such as risedronate having essentially no activity (IC_{50} ≈ 660 μM[56]).

The question then arises as to whether the ζ fold is unique, being restricted to UPPS and closely related systems seen in mycobacteria, or whether it might occur in other systems as well. Using the SSM program[58] to find similar folds revealed no hits. Likely candidates would be other prenyl synthases that use Mg^{2+}, but whose structures are unknown, and the Rv3378c gene product of *Mycobacterium tuberculosis*, a target for anti-infective therapy,[59] appears to be a likely candidate. This protein catalyzes formation of the diterpene virulence factors tuberculosinol (**28**) and the isotuberculosinols (**29**, **30**) from tuberculosinol diphosphate (**31**; Scheme 7), and in this case, H_2O acts as the nucleophile, attacking either the C1 or C3 sites in the allylic substrate.[60] Using three structure prediction programs: I-TASSER,[61] SWISS-MODEL,[62] and Phyre,[63] we found that Rv3378c has distant sequence homology to UPPS (19% identity by Phyre). The top-scoring predicted folds from each program are very similar and one is shown in Figure 6c, superimposed on the UPPS structure, where there is a 1.93 Å C^{α} rmsd between the Rv3378c prediction and that found in *E. coli* UPPS. What is particularly interesting about the models is that the DDXXD motif, known to be essential for catalysis,[64] is located at the entrance to the main (UPPS) ligand-binding site, adjacent the essential D26 and Mg^{2+} in the UPPS structure (Figure 6c). This finding supports the idea that this protein also adopts the ζ fold—though as with the diterpene cyclases, X-ray structures are desirable to confirm these predictions.

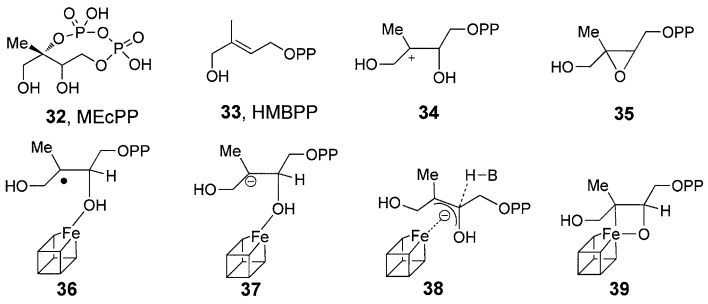


Scheme 7. Formation of tuberculosinol virulence factors in *M. tuberculosis*.

8. The Dynamic Structure of a Prenyl Transferase, UPPS

From a drug-discovery perspective, with UPPS, as well as with other proteins, it is of interest to consider how ligand binding affects protein structure. In some cases, the structure of a protein may be known, but there are no substrate-, product-, or inhibitor-bound structures, which makes discovering inhibitors using virtual screening difficult since there may be no obvious ligand-binding pocket that can be

targeted. A ligand-free protein must, however, expand to accommodate substrates and products, and it is this more "open" structure that is likely to enable inhibitor discovery. One approach to finding such structures is to use molecular dynamics simulations.[65] Starting with an "open" form of UPPS (the structure shown in Figure 6a, but with the bound ligand removed), a molecular dynamics trajectory (Figure 6d) shows that the volume (ca. 1000 Å^3) originally occupied by the bound ligand rapidly decreases, then stays constant for most of the trajectory, and this volume (ca. 430 Å^3, Figure 6f) is very similar to the volume of roughly 330 Å^3 seen in a crystal structure of the ligand-free protein.[56,65] There is, however, a transient opening of the protein to form the substrate/product/inhibitor-binding site, as shown in Figure 6d–f. Remarkably, use of the rarely sampled conformational state structure enables (with the Glide program[66]) much tighter ligand-bound poses and better corrrelations between the IC_{50} values and the docking scores for a series of bisphosphonate inhibitors of UPPS than found with the closed form of the enzyme.[65] These results suggest that using MD methods to sample rare "expanded-pocket" states is a potentially significant new approach to facilitate inhibitor discovery using virtual screening: this approach is likely to be applicable to most prenyl transferases and terpene synthases in which large pocket volumes are needed to accommodate large ligands.



Scheme 8. Formation of HMBPP: substrate, product, and possible reactive intermediates.

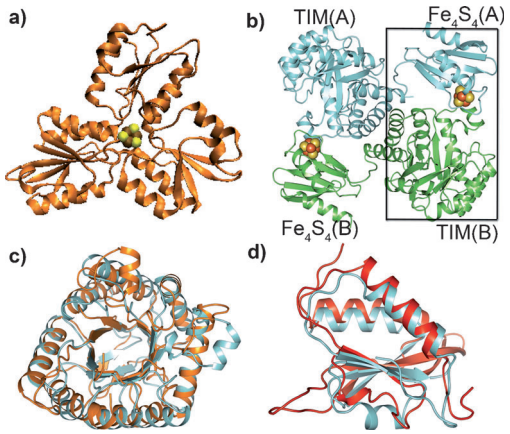


Figure 7. Modular structures of the 4Fe-4S-cluster-containing proteins IspH and IspG. a) IspH from *A. aeolicus* (PDB ID: 3DNF) showing the three helix/sheet domains surrounding the 4Fe-4S cluster in the "closed" form (which buries the Fe/S cluster). b) IspG from *A. aeolicus* (PDB ID: 3NOY) showing an "open" structure. The 4Fe-4S cluster from one chain is thought to interact with the TIM barrel in the second chain to form the active site (black box). c) Superposition of the TIM barrel in *A. aeolicus* IspG (orange) with *B. anthracis* dihydropteroate synthase (cyan; PDB ID: 1TWS). d) Superposition of the 4Fe-4S cluster domain in IspG with that in spinach nitrite reductase. The C^α rmsd values in (c, d) are 2.4 and 2.1 Å, respectively.

9. The 4Fe-4S Reductases: Progress and Puzzles with IspG and IspH

Finally, we consider the question of how, in plant plastids and in many bacteria, the DMAPP and IPP terpene precursors are made. DMAPP and IPP biosynthesis involves the initial condensation of pyruvate with glyceraldehyde phosphate to form 1-deoxyxylulose-phosphate which, after four additional reactions, forms 2-*C*-methyl-D-erythritol-2,4-*cyclo*-diphosphate (MEcPP, **32**; Scheme 8). MEcPP is then converted by (*E*)-1-hydroxy-2-methyl-but-2-enyl 4-diphosphate (HMBPP) synthase (IspG; also known as GcpE) to form HMBPP (**33**), which is then reduced by HMBPP reductase (IspH; also known as Lyt B) to form IPP and DMAPP, in a roughly 5:1 ratio. But what are the structures of these enzymes? How do they carry out these remarkable reactions?

Based on chemical analysis, bioinformatics, and EPR spectroscopy, both IspG and IspH have been shown to contain 4Fe-4S clusters akin to those found in ferredoxins, but with an unusual coordination—a non-Cys residue at the unique fourth Fe atom,[67] results now confirmed by X-ray structures.[68] The structure of IspH (Figure 7a) is unusual in that it consists of a modular, cloverleaf- or trefoil-like structure in which three distinct helical/sheet domains surround a central FeS cluster. This fold is now, however, seen to be very similar to that of two other 4Fe-4S proteins, quinolinate synthase[69] and diphthamide synthase,[70] with, on average, a 2.5 Å C^α rmsd amongst each of the domains; this suggests again, in analogy to the other prenyl synthases, gene duplication and condensation. Mechanistically, it appears that the HMBPP substrate first binds to the fourth Fe center of the 4Fe-4S cluster, then is reduced to an allyl intermediate.[68b,71] An intermediate lacking the HMBPP O1 can be seen crystallographically[68b] and has Fe–C bond lengths of 2.6–2.7 Å, which suggests a metal–ligand interaction (since the sum of van der Waals radii for Fe and C is ca. 3.6 Å[72]). A detailed discussion of the IspH structure and mechanism has recently been reported in this journal.[73] The situation with IspG is, however, more complex.

There are two main mechanisms for IspG catalysis that seem plausible. In one, the *cyclo*-diphosphate ring in the MEcPP substrate opens to form a carbocation (**34**) that is then reduced to form an anion, which is converted to the HMBPP product. In the other mechanism, the *cyclo*-diphosphate first isomerizes to form an epoxide (**35**), which is then deoxygenated by the 4Fe-4S cluster. Support for the latter mechanism is based on precedent: epoxides are known to be reduced to olefins by reduced 4Fe-4S clusters in model

systems.[74] In addition, HMBPP epoxide is reduced by IspG to HMBPP with similar kinetics to that found with MEcPP.[75] However, it has now been reported that the rate of the MEcPP→epoxide reaction catalyzed by IspG is very slow[76] and is inconsistent with the k_{cat} seen with both MEcPP as well as HMBPP-epoxide[77] as substrates, suggesting parallel rather than consecutive reactions, and a common reaction intermediate. Plus, there is now evidence that a carbocation forms with MEcPP + IspG.[77] What, then, might the reaction intermediate be?

When either MEcPP or HMBPP-epoxide are added to reduced IspG, the same reactive intermediate "X"[78] forms,[78,79] as observed by EPR, ENDOR, or HYSCORE spectroscopy. On incubation, "X" converts to the HMBPP product, which, as with IspH, then binds to the 4Fe-4S cluster,[79] and there have been several structures (e.g. **36**–**39**) considered for "X". A radical (**36**) is unlikely since no radical-like signals are seen in EPR spectra; plus, the resonance that is seen broadens on ^{57}Fe labeling.[79b] A carbanion (**37**) is unlikely since it would be very reactive, and a π/η complex (**38**) is unlikely since not only is it not an η^3-oxaallyl (because the oxygen is protonated), but H3 is retained during isoprenoid biosynthesis, as evidenced by ^{2}H-labeling studies.[80] The ^{13}C hyperfine coupling observed (ca. 16 MHz) is similar to that found[81] for an Fe–C bond in a Fe-Fe hydrogenase (17 MHz) as well as that computed[82] for a Mo–C single bond in a model formaldehyde–xanthine oxidase complex (ca. 16 MHz), but is roughly three times smaller than the "transannular" (through two bonds) hyperfine coupling seen (and computed) in the square-pyramidal formaldehyde-inhibited xanthine oxidase complex.[82] These results favor, then, the presence of an Fe–C bond, as in ferraoxetanes such as **39**, plus, ferraoxetane itself is known to undergo a [2+2] reaction to form ethylene.[83] But what is the structure of IspG, and how might it catalyze such reactions?

In recent work, the first single-crystal X-ray crystallographic structure of an IspG, from *Aquifex aeolicus,* was reported.[84] The structure, Figure 7b, is of interest in that it is again modular and contains two distinct domains. The large N-terminal domain consists of a triose phosphate isomerase (TIM) barrel that is highly homologous to the structure of *Bacillus anthracis* dihydropteroate synthase (Figure 7c; 2.4 Å C^α rmsd), while the C-terminal domain (which houses the 4Fe-4S cluster) is highly homologous to spinach nitrite reductase (Figure 7d; 2.1 Å C^α rmsd)). The crystallographic results also show the presence of three Cys residues in the 4Fe-4S cluster, together with a highly conserved Glu, coordinated to the fourth Fe atom. The structure of the *Thermus thermophilus* protein[85] is very similar. Based on the crystallographic structures, it appears unlikely that the two domains function independently in a monomer since the DHPS and 4Fe-4S cluster regions are separated by approximately 45 Å. However, if IspG functions as a dimer—as suggested by the observation that it crystallizes as a dimer—then the C terminus (4Fe-4S cluster) of one molecule in the dimer is situated close to the N terminus (TIM barrel) of the second molecule in the dimer (Figure 6b), and Lee et al.[84] proposed that these two domains can form a "closed" conformation. This would be reminiscent of the movement of one of the three domains in IspH to form the "closed" structure that protects the reactive intermediates during catalysis,[68b] as well as the closing of two domains around the 4Fe-4S cluster in acetyl-CoA synthase/carbon monoxide dehydrogenase.[86]

In the closed conformation, the substrate would be sandwiched between the TIM barrel of one molecule and the 4Fe-4S cluster in the second molecule in the dimer, forming a single catalytic center in which the *cyclo*-diphosphate fragment in MEcPP binds to a highly conserved patch of basic residues in the TIM barrel.[84] Other highly conserved residues then catalyze ring opening, while the 4Fe-4S domain (in the closed conformation) carries out the $2H^+/2e^-$ redox reaction. This "hybrid" catalytic center would of course be reminiscent of the $\alpha_2\delta_2$ module interactions in GPPS (Figure 1c), and is supported by the observation that there are no highly conserved basic residues in the 4Fe-4S cluster to which a diphosphate group can bind. Further support for the "open-and-closed" model comes from the second IspG structure (from *T. thermophilus*) in which Rekittke et al.[85] report an even more open, "open" structure, as well as a closed structure model in which the two domains come together to form the catalytic center. The closed structure is generated by a hinge motion between the two domains. It is, however, the TIM barrel of one molecule in the dimer that interacts with the 4Fe-4S cluster in the second molecule, as shown in the box in Figure 7b. Closed structures with inhibitors/substrates are eagerly awaited.

10. Summary and Outlook

There have recently been numerous major developments in our understanding of the structure, function, evolution, and inhibition of many of the enzymes involved in terpene and isoprenoid biosynthesis. These results are important not only from an academic perspective, they are also of practical significance because many of these proteins are targets for drug discovery. The (αα) structures of FPPS and GGPPS are of interest as anticancer and anti-infective drug targets, with numerous new drug leads now identified. The structures of a GPPS have been reported: they are remarkable in that GPPS (from *M. piperata*) consists of an $\alpha_2\delta_2$ heterotetramer with both catalytic (α) and regulatory (δ) subunits. The two subunits have very similar three-dimensional structures, though neither domain alone has catalytic activity, and a catalytic/regulatory modular structure appears to be present in C_{35} and C_{50} prenyltransferases as well.[18] The new head-to-head synthase structures (of CrtM and SQS) are of interest since they help illuminate the first committed steps in sterol and carotenoid biosynthesis, formation of presqualene diphosphate, again of importance in drug discovery. Bioinformatics predictions about (plant) diterpene cyclase structures in which there are three domains (α, β, and γ) have been confirmed experimentally. This leads to added confidence in the genomics and bioinformatics proposals that many plant terpene cyclases derive from ancestral αβγ proteins, which themselves appear to have originated by fusion of α- and βγ-domain proteins. A schematic illustration of the different

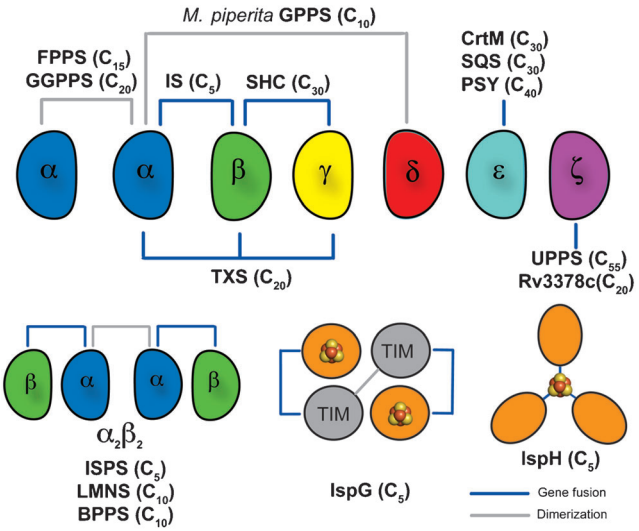


Figure 8. Schematic illustration summarizing the modular nature of many terpene/isoprenoid biosynthesis enzymes. FPPS and GPPS are αα dimers (human GGPPS is a trimer of dimers); GPPS forms a heterotetramer, $\alpha_2\delta_2$; plant diterpene cyclases (like TXS) are αβγ; many other plant terpene cyclases have lost γ and are αβ, others lack both β and γ and are purely α.

structural arrangements found with the α, β, γ, and δ modules is shown in Figure 8. The structures of several cis-prenyl transferases, some with bound inhibitors, have also been reported. These adopt the ζ fold, and based again on bioinformatics, it appears likely that this fold may also be more widespread. The structures of the two 4Fe-4S proteins involved in C_5-diphosphate production in most eubacteria have also now been solved. Both have unusual 4Fe-4S clusters with a unique Fe center which appears to be involved in Fe–C bond formation during catalysis, and, again, both are modular proteins. And finally, the structures of several αβ proteins, including limonene synthase and isoprene synthase, have been solved. Their three-dimensional structures are remarkably similar, an observation that extends to their essentially identical quaternary structures, and the structure of isoprene synthase itself is of interest in the context of alternative fuel development. Also of general interest is the observation that while "there is no substitute" for X-ray crystallography, the use of bioinformatics tools helped correctly predict the three-helical model for TXS.

Future work may focus on the structures of the enzymes involved in carotenoid biosynthesis: the dehydrogenases that convert, for example, dehydrosqualene and phytoene into conjugated polyenes, as well as systems such as lycopene cyclase which catalyze ring formation. The latter is of interest since it is one of the nonredox flavoproteins in which, apparently, an anionic reduced flavin cofactor (FAD) stabilizes a cationic intermediate or transition state,[87] which would be formally similar to the situation found in the class II terpene synthases. A new structure in which FAD plays a key redox role is the FAD-catalyzed reduction of GGPP chains by geranylgeranyl reductase from *Thermoplasma acidophilum*.[88] This enzyme catalyzes the reduction of geranylgeranyl side chains to phytanyl side chains in lipids in archaeabacteria. However, how such C_{20} side chains couple through their terminal methyl groups to form the C_{40} lipids that span the lipid bilayer is still a mystery. Interest in these systems is again not purely academic since dehydrogenase inhibitors could act as antivirulence factors for staph infections; carotenoid biosynthesis inhibitors are targets for bleaching herbicides; and it may be possible to use structure-based design to engineer reductases to convert, for example, β-farnesene to farnesane, a biofuel, or to produce lower-molecular-weight compounds such as dimethyloctane, with GPP.

We are, therefore, near the end of the beginning: the structures of many of the major proteins directly involved in terpene/isoprenoid biosynthesis are now known, and the stage is set for developing novel inhibitors that can be turned into new drugs as well as, potentially, developing new platforms for renewable fuels, and other materials.

This work was supported by the United States Public Health Service (NIH grants GM65307, GM073216, AI074233 and CA158191). We thank Drs. Rong Cao and Ke Wang for help with the structures and graphics, and Professor Robert Coates for helpful suggestions.

Received: May 6, 2011
Published online: November 21, 2011

[1] J. Chappell, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1995**, *46*, 521–547.
[2] a) P. H. Liang, T. P. Ko, A. H. Wang, *Eur. J. Biochem.* **2002**, *269*, 3339–3354; b) H. V. Thulasiram, H. K. Erickson, C. D. Poulter, *Science* **2007**, *316*, 73–76.
[3] C. D. Poulter, *Acc. Chem. Res.* **1990**, *23*, 70–77.
[4] J. M. Berg, L. Stryer, *Biochemistry*, 5th ed., W. H. Freeman, New York, **2002**.
[5] J. Bohlmann, C. I. Keeling, *Plant J.* **2008**, *54*, 656–669.
[6] H. M. Miziorko, *Arch. Biochem. Biophys.* **2011**, *505*, 131–143.
[7] M. Rohmer, *Lipids* **2008**, *43*, 1095–1107.
[8] a) Y. Koga, H. Morii, *Microbiol. Mol. Biol. Rev.* **2007**, *71*, 97–120; b) J. Lombard, D. Moreira, *Mol. Biol. Evol.* **2011**, *28*, 87–99.
[9] M. Eberl, M. Hintz, A. Reichenberg, A. K. Kollas, J. Wiesner, H. Jomaa, *FEBS Lett.* **2003**, *544*, 4–10.
[10] J. Buckingham, *Dictionary of Natural Products on DVD*, CRC, Boca Raton, FL, **2007**.
[11] L. C. Tarshis, M. Yan, C. D. Poulter, J. C. Sacchettini, *Biochemistry* **1994**, *33*, 10871–10877.
[12] a) D. J. Hosfield, Y. Zhang, D. R. Dougan, A. Broun, L. W. Tari, R. V. Swanson, J. Finn, *J. Biol. Chem.* **2004**, *279*, 8526–8529; b) J. A. Aaron, D. W. Christianson, *Pure Appl. Chem.* **2010**, *82*, 1585–1597.
[13] T. H. Chang, R. T. Guo, T. P. Ko, A. H. Wang, P. H. Liang, *J. Biol. Chem.* **2006**, *281*, 14991–15000.
[14] C. Burke, R. Croteau, *J. Biol. Chem.* **2002**, *277*, 3141–3149.
[15] C. Burke, R. Croteau, *Arch. Biochem. Biophys.* **2002**, *405*, 130–136.
[16] R. Saiki, A. Nagata, T. Kainou, H. Matsuda, M. Kawamukai, *FEBS J.* **2005**, *272*, 5606–5622.
[17] T. H. Chang, F. L. Hsieh, T. P. Ko, K. H. Teng, P. H. Liang, A. H. Wang, *Plant Cell* **2010**, *22*, 454–467.
[18] D. Sasaki, M. Fujihashi, N. Okuyama, Y. Kobayashi, M. Noike, T. Koyama, K. Miki, *J. Biol. Chem.* **2011**, *286*, 3729–3740.
[19] M. Gnant, B. Mlineritsch, W. Schippinger, G. Luschin-Ebengreuth, S. Postlberger, C. Menzel, R. Jakesz, M. Seifert, M.

Hubalek, V. Bjelic-Radisic, H. Samonigg, C. Tausch, H. Eidtmann, G. Steger, W. Kwasny, P. Dubsky, M. Fridrik, F. Fitzal, M. Stierer, E. Rucklinger, R. Greil, C. Marth, *N. Engl. J. Med.* **2009**, *360*, 679 – 691.

[20] J. M. Rondeau, F. Bitsch, E. Bourgier, M. Geiser, R. Hemmig, M. Kroemer, S. Lehmann, P. Ramage, S. Rieffel, A. Strauss, J. R. Green, W. Jahnke, *ChemMedChem* **2006**, *1*, 267 – 273.

[21] S. G. Senaratne, G. Pirianov, J. L. Mansi, T. R. Arnett, K. W. Colston, *Br. J. Cancer* **2000**, *82*, 1459 – 1468.

[22] C. Denoyelle, L. Hong, J. P. Vannier, J. Soria, C. Soria, *Br. J. Cancer* **2003**, *88*, 1631 – 1640.

[23] M. Coscia, E. Quaglino, M. Iezzi, C. Curcio, F. Pantaleoni, C. Riganti, I. Holen, H. Monkkonen, M. Boccadoro, G. Forni, P. Musiani, A. Bosia, F. Cavallo, M. Massaia, *J. Cell Mol. Med.* **2009**, *14*, 2803 – 2815.

[24] V. Kunzmann, H. Jomaa, J. Feurle, E. Bauer, M. Herderich, M. Wilhelm, *Blood* **1997**, *90*, 2558 – 2558.

[25] V. Kunzmann, S. Eckstein, A. Lindner, H. P. Tony, M. Wilhelm, *Blood* **2002**, *100*, 309B – 309B.

[26] S. Mukherjee, C. Huang, F. Guerra, K. Wang, E. Oldfield, *J. Am. Chem. Soc.* **2009**, *131*, 8374 – 8375.

[27] a) J. M. Sanders, Y. Song, J. M. Chan, Y. Zhang, S. Jennings, T. Kosztowski, S. Odeh, R. Flessner, C. Schwerdtfeger, E. Kotsikorou, G. A. Meints, A. O. Gomez, D. Gonzalez-Pacanowska, A. M. Raker, H. Wang, E. R. van Beek, S. E. Papapoulos, C. T. Morita, E. Oldfield, *J. Med. Chem.* **2005**, *48*, 2957 – 2963; b) S. Cotesta, J. F. Glickman, W. Jahnke, A. Marzinzik, S. Ofner, J. Rondeau, T. Zoller (Novartis AG), PCT/EP2009/063257, **2010**.

[28] W. Jahnke, J. M. Rondeau, S. Cotesta, A. Marzinzik, X. Pelle, M. Geiser, A. Strauss, M. Gotte, F. Bitsch, R. Hemmig, C. Henry, S. Lehmann, J. F. Glickman, T. P. Roddy, S. J. Stout, J. R. Green, *Nat. Chem. Biol.* **2010**, *6*, 660 – 666.

[29] K. L. Kavanagh, J. E. Dunford, G. Bunkoczi, R. G. Russell, U. Oppermann, *J. Biol. Chem.* **2006**, *281*, 22004 – 22012.

[30] J. D. Artz, A. K. Wernimont, J. E. Dunford, M. Schapira, A. Dong, Y. Zhao, J. Lew, R. G. Russell, F. H. Ebetino, U. Oppermann, R. Hui, *J. Biol. Chem.* **2011**, *286*, 3315 – 3322.

[31] Y. Zhang, R. Cao, F. Yin, M. P. Hudock, R. T. Guo, K. Krysiak, S. Mukherjee, Y. G. Gao, H. Robinson, Y. Song, J. H. No, K. Bergan, A. Leon, L. Cass, A. Goddard, T. K. Chang, F. Y. Lin, E. Van Beek, S. Papapoulos, A. H. Wang, T. Kubo, M. Ochi, D. Mukkamala, E. Oldfield, *J. Am. Chem. Soc.* **2009**, *131*, 5153 – 5162.

[32] A. P. Singh, Y. Zhang, J. H. No, R. Docampo, V. Nussenzweig, E. Oldfield, *Antimicrob. Agents Chemother.* **2010**, *54*, 2987 – 2993.

[33] G. Ourisson, P. Albrecht, *Acc. Chem. Res.* **1992**, *25*, 398 – 402.

[34] C. I. Liu, G. Y. Liu, Y. Song, F. Yin, M. E. Hensler, W. Y. Jeng, V. Nizet, A. H. Wang, E. Oldfield, *Science* **2008**, *319*, 1391 – 1394.

[35] F. X. Cunningham, E. Gantt, *Annu. Rev. Plant Biol.* **1998**, *49*, 557 – 583.

[36] J. Pandit, D. E. Danley, G. K. Schulte, S. Mazzalupo, T. A. Pauly, C. M. Hayward, E. S. Hamanaka, J. F. Thompson, H. J. Harwood, Jr., *J. Biol. Chem.* **2000**, *275*, 30610 – 30617.

[37] F. Y. Lin, C. I. Liu, Y. L. Liu, Y. Zhang, K. Wang, W. Y. Jeng, T. P. Ko, R. Cao, A. H. Wang, E. Oldfield, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 21337 – 21342.

[38] R. M. Phan, C. D. Poulter, *J. Org. Chem.* **2001**, *66*, 6705 – 6710.

[39] J. C. Fernandes Rodrigues, J. L. Concepcion, C. Rodrigues, A. Caldera, J. A. Urbina, W. de Souza, *Antimicrob. Agents Chemother.* **2008**, *52*, 4098 – 4114.

[40] K. U. Wendt, A. Lenhart, G. E. Schulz, *J. Mol. Biol.* **1999**, *286*, 175 – 187.

[41] R. Thoma, T. Schulz-Gasch, B. D'Arcy, J. Benz, J. Aebi, H. Dehmlow, M. Hennig, M. Stihle, A. Ruf, *Nature* **2004**, *432*, 118 – 122.

[42] C. M. Starks, K. Back, J. Chappell, J. P. Noel, *Science* **1997**, *277*, 1815 – 1820.

[43] C. A. Lesburg, G. Z. Zhai, D. E. Cane, D. W. Christianson, *Science* **1997**, *277*, 1820 – 1824.

[44] D. A. Whittington, M. L. Wise, M. Urbansky, R. M. Coates, R. B. Croteau, D. W. Christianson, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 15375 – 15380.

[45] R. Cao, Y. Zhang, F. M. Mann, C. Huang, D. Mukkamala, M. P. Hudock, M. E. Mead, S. Prisic, K. Wang, F. Y. Lin, T. K. Chang, R. J. Peters, E. Oldfield, *Proteins Struct. Funct. Genet.* **2010**, *78*, 2417 – 2432.

[46] S. C. Trapp, R. B. Croteau, *Genetics* **2001**, *158*, 811 – 832.

[47] K. U. Wendt, G. E. Schulz, *Structure* **1998**, *6*, 127 – 133.

[48] R. J. Peters, R. B. Croteau, *Biochemistry* **2002**, *41*, 1836 – 1842.

[49] M. Köksal, I. Zimmer, J. P. Schnitzler, D. W. Christianson, *J. Mol. Biol.* **2010**, *402*, 363 – 373.

[50] D. Morrone, J. Chambers, L. Lowry, G. Kim, A. Anterola, K. Bender, R. J. Peters, *FEBS Lett.* **2009**, *583*, 475 – 480.

[51] J. A. Cuff, M. E. Clamp, A. S. Siddiqui, M. Finlay, G. J. Barton, *Bioinformatics* **1998**, *14*, 892 – 893.

[52] P. F. Fuchs, A. J. Alix, *Proteins Struct. Funct. Genet.* **2005**, *59*, 828 – 839.

[53] M. Köksal, Y. Jin, R. M. Coates, R. Croteau, D. W. Christianson, *Nature* **2011**, *469*, 116 – 120.

[54] M. Köksal, H. Hu, R. M. Coates, R. J. Peters, D. W. Christianson, *Nat. Chem. Biol.* **2011**, *7*, 431 – 433.

[55] M. Fujihashi, Y. W. Zhang, Y. Higuchi, X. Y. Li, T. Koyama, K. Miki, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 4337 – 4342.

[56] R. T. Guo, R. Cao, P. H. Liang, T. P. Ko, T. H. Chang, M. P. Hudock, W. Y. Jeng, C. K. M. Chen, Y. H. Zhang, Y. C. Song, C. J. Kuo, F. L. Yin, E. Oldfield, A. H. J. Wang, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 10022 – 10027.

[57] Y. P. Lu, H. G. Liu, P. H. Liang, *Biochem. Biophys. Res. Commun.* **2009**, *379*, 351 – 355.

[58] E. Krissinel and K. Henrick. *PDBeFold*, http://www.ebi.ac.uk/msd-srv/ssm.

[59] a) K. Pethe, D. L. Swenson, S. Alonso, J. Anderson, C. Wang, D. G. Russell, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 13642 – 13647; b) T. Hoshino, C. Nakano, T. Ootsuka, Y. Shinohara, T. Hara, *Org. Biomol. Chem.* **2011**, *9*, 2156 – 2165.

[60] T. Hoshino, C. Nakano, T. Ootsuka, Y. Shinohara, T. Hara, *Org. Biomol. Chem.* **2011**, *9*, 2156 – 2165.

[61] A. Roy, A. Kucukural, Y. Zhang, *Nat. Protoc.* **2010**, *5*, 725 – 738.

[62] F. Kiefer, K. Arnold, M. Kunzli, L. Bordoli, T. Schwede, *Nucleic Acids Res.* **2009**, *37*, D387 – 392.

[63] L. A. Kelley, M. J. Sternberg, *Nat. Protoc.* **2009**, *4*, 363 – 371.

[64] C. Nakano, T. Ootsuka, K. Takayama, T. Mitsui, T. Sato, T. Hoshino, *Biosci. Biotechnol. Biochem.* **2011**, *75*, 75 – 81.

[65] W. Sinko, C. de Oliveira, S. Williams, A. Van Wynsberghe, J. D. Durrant, R. Cao, E. Oldfield, J. A. McCammon, *Chem. Biol. Drug Des.* **2011**, *77*, 412 – 420.

[66] R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, P. S. Shenkin, *J. Med. Chem.* **2004**, *47*, 1739 – 1749.

[67] a) P. Adam, S. Hecht, W. Eisenreich, J. Kaiser, T. Grawert, D. Arigoni, A. Bacher, F. Rohdich, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12108 – 12113; b) S. Hecht, W. Eisenreich, P. Adam, S. Amslinger, K. Kis, A. Bacher, D. Arigoni, F. Rohdich, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 14837 – 14842.

[68] a) I. Rekittke, J. Wiesner, R. Rohrich, U. Demmer, E. Warkentin, W. Xu, K. Troschke, M. Hintz, J. H. No, E. C. Duin, E. Oldfield, H. Jomaa, U. Ermler, *J. Am. Chem. Soc.* **2008**, *130*, 17206 – 17207; b) T. Grawert, I. Span, W. Eisenreich, F. Rohdich, J. Eppinger, A. Bacher, M. Groll, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 1077 – 1081.

[69] H. Sakuraba, H. Tsuge, K. Yoneda, N. Katunuma, T. Ohshima, *J. Biol. Chem.* **2005**, *280*, 26645 – 26648.

[70] Y. Zhang, X. Zhu, A. T. Torelli, M. Lee, B. Dzikovski, R. M. Koralewski, E. Wang, J. Freed, C. Krebs, S. E. Ealick, H. Lin, *Nature* **2010**, *465*, 891 – 896.
[71] W. Wang, K. Wang, Y. L. Liu, J. H. No, J. Li, M. J. Nilges, E. Oldfield, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 4522 – 4527.
[72] S. S. Batsanov, *Inorg. Mater.* **2001**, *37*, 871 – 885.
[73] T. Gräwert, I. Span, A. Bacher, M. Groll, *Angew. Chem.* **2010**, *122*, 8984 – 8991; *Angew. Chem. Int. Ed.* **2010**, *49*, 8802 – 8809.
[74] T. Itoh, T. Nagano, M. Sato, M. Hirobe, *Tetrahedron Lett.* **1989**, *30*, 6387 – 6388.
[75] R. L. Nyland II, Y. Xiao, P. Liu, C. L. Freel Meyers, *J. Am. Chem. Soc.* **2009**, *131*, 17734 – 17735.
[76] Y. L. Xiao, R. L. Nyland, C. L. Freel Meyers, P. H. Liu, *Chem. Commun.* **2010**, *46*, 7220 – 7222.
[77] Y. Xiao, D. Rooker, Q. You, C. L. Freel Meyers, P. Liu, *ChemBioChem* **2011**, *12*, 527 – 530.
[78] D. Adedeji, H. Hernandez, J. Wiesner, U. Kohler, H. Jomaa, E. C. Duin, *FEBS Lett.* **2007**, *581*, 279 – 283.
[79] a) W. Wang, J. Li, K. Wang, C. Huang, Y. Zhang, E. Oldfield, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 11189 – 11193; b) W. Y. Xu, N. S. Lees, D. Adedeji, J. Wiesner, H. Jomaa, B. M. Hoffman, E. C. Duin, *J. Am. Chem. Soc.* **2010**, *132*, 14509 – 14520.
[80] J. F. Hoeffler, A. Hemmerlin, C. Grosdemange-Billiard, T. J. Bach, M. Rohmer, *Biochem. J.* **2002**, *366*, 573 – 583.
[81] A. Silakov, B. Wenk, E. Reijerse, S. P. Albracht, W. Lubitz, *J. Biol. Inorg. Chem.* **2009**, *14*, 301 – 313.
[82] M. Shanmugam, B. Zhang, R. L. McNaughton, R. A. Kinney, R. Hille, B. M. Hoffman, *J. Am. Chem. Soc.* **2010**, *132*, 14015 – 14017.
[83] Z. H. Kafafi, R. H. Hauge, W. E. Billups, J. L. Margrave, *J. Am. Chem. Soc.* **1987**, *109*, 4775 – 4780.
[84] M. Lee, T. Grawert, F. Quitterer, F. Rohdich, J. Eppinger, W. Eisenreich, A. Bacher, M. Groll, *J. Mol. Biol.* **2010**, *404*, 600 – 610.
[85] I. Rekittke, T. Nonaka, J. Wiesner, U. Demmer, E. Warkentin, H. Jomaa, U. Ermler, *FEBS Lett.* **2011**, *585*, 447 – 451.
[86] C. Darnault, A. Volbeda, E. J. Kim, P. Legrand, X. Vernede, P. A. Lindahl, J. C. Fontecilla-Camps, *Nat. Struct. Biol.* **2003**, *10*, 271 – 279.
[87] Q. Yu, P. Schaub, S. Ghisla, S. Al-Babili, A. Krieger-Liszkay, P. Beyer, *J. Biol. Chem.* **2010**, *285*, 12109 – 12120.
[88] Q. Xu, T. Eguchi, Mathews, II, C. L. Rife, H. J. Chiu, C. L. Farr, J. Feuerhelm, L. Jaroszewski, H. E. Klock, M. W. Knuth, M. D. Miller, D. Weekes, M. A. Elsliger, A. M. Deacon, A. Godzik, S. A. Lesley, I. A. Wilson, *J. Mol. Biol.* **2010**, *404*, 403 – 417.