Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure analysis of geranyl pyrophosphate methyltransferase and the proposed reaction mechanism of SAM-dependent *C*-methylation

Orapin Ariyawutthiphan,[a] Toyoyuki Ose,[b]* Atsushi Minami,[c] Sandip Sinde,[c] Muneya Tsuda,[c] Yong-Gui Gao,[d] Min Yao,[b] Hideaki Oikawa[c] and Isao Tanaka[d]

[a]Graduate School of Life Sciences, Hokkaido University, N10 W8 Kita-Ku, Sapporo 060-0810, Japan, [b]Faculty of Pharmaceutical Sciences, Hokkaido University, N10 W8 Kita-Ku, Sapporo 060-0810, Japan, [c]Division of Chemistry, Graduate School of Science, Hokkaido University, N10 W8 Kita-Ku, Sapporo 060-0810, Japan, and [d]Faculty of Advanced Life Sciences, Hokkaido University, N10 W8 Kita-Ku, Sapporo 060-0810, Japan

Correspondence e-mail: ose@pharm.hokudai.ac.jp

Received 13 July 2012
Accepted 7 September 2012

PDB References: geranyl pyrophosphate methyltransferase, complex with SAM, 4f84; apo, 4f85; complex with GPP and sinefungin, 4f86

In the typical isoprenoid-biosynthesis pathway, condensation of the universal C_5-unit precursors isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) occurs *via* the common intermediates prenyl pyrophosphates (C_{10}–C_{20}). The diversity of isoprenoids reflects differences in chain length, cyclization and further additional modification after cyclization. In contrast, the biosynthesis of 2-methylisonorneol (2-MIB), which is responsible for taste and odour problems in drinking water, is unique in that it primes the enzymatic methylation of geranyl pyrophosphate (GPP) before cyclization, which is catalyzed by an *S*-adenosyl-L-methionine-dependent methyltransferase (GPPMT). The substrate of GPPMT contains a nonconjugated olefin and the reaction mechanism is expected to be similar to that of the steroid methyltransferase (SMT) family. Here, structural analysis of GPPMT in complex with its cofactor and substrate revealed the mechanisms of substrate recognition and possible enzymatic reaction. Using the structures of these complexes, methyl-group transfer and the subsequent proton-abstraction mechanism are discussed. GPPMT and SMTs contain a conserved glutamate residue that is likely to play a role as a general base. Comparison with the reaction mechanism of the mycolic acid cyclopropane synthase (MACS) family also supports this result. This enzyme represented here is the first model of the enzymatic *C*-methylation of a nonconjugated olefin in the isoprenoid-biosynthesis pathway. In addition, an elaborate system to avoid methylation of incorrect substrates is proposed.

1. Introduction

Natural isoprenoids consist of terpenoids, steroids, carotenoids, prenylated quinones (menaquinone) and dolichol phosphate (a sugar carrier), and may include prenylated proteins. These are synthesized by prenylation using the universal C_5-unit precursors isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) *via* common intermediates: the prenyl pyrophosphates (C_{10}–C_{20}). The diversity of isoprenoids originates from chain length and cyclization modes. Further diversification of isoprenoids is effected by oxidative modifications and the introduction of various groups, such as by acylation, glycosylation and alkylation. In addition to these modifications, C1-methylation of the prenyl chain as in the case of the steroid side chain adds further diversity to the isoprenoids. Recently, unprecedented methylation of geranyl pyrophosphate (GPP) was found in the biosynthesis of the bacterial homo-monoterpene 2-methylisoborneol (2-MIB).

2-MIB, which is widely produced by actinomycetes (Medsker *et al.*, 1969), cyanobacteria (Giglio *et al.*, 2011) and

myxobacteria (Dickschat *et al.*, 2007), is responsible for an unpleasant musty odour, and its odorous threshold is below 10 ng l^{-1} (Young *et al.*, 1996). The occurrence of 2-MIB in drinking water is a common and widespread problem in water supplies. In a biosynthetic study of 2-MIB, labelling experiments confirmed that the 'extra' C2 methyl group is derived from *S*-adenosyl-L-methionine (SAM; Bentley & Meganathan, 1981). The involvement of 2-methylgeranyl pyrophosphate (2-methyl-GPP) was proposed, based on feeding experiments in the myxobacterium *Nannocystis exedens*, indicating that the SAM-dependent methylation of GPP proceeds prior to cyclization (Dickschat *et al.*, 2007). Bioinformatics approaches based on hidden Markov models (Eddy, 1998) enabled the identification of a putative operon encoding GPP methyltransferase (GPPMT) and 2-MIB synthase in various actinomycetes (Komatsu *et al.*, 2008). The production of 2-MIB by heterologous expression of these translationally coupled genes from various bacteria using *Streptomyces avermitilis*, which is naturally deficient in 2-MIB biosynthesis, as a host (Komatsu *et al.*, 2008) was confirmed (Fig. 1). Recently, we reported that recombinant GPP methyltransferase (GPPMT) and monoterpene synthase from *S. lasaliensis* convert GPP to 2-MIB *via* 2-methyl-GPP (Komatsu *et al.*, 2008). This conversion by these two proteins (expressed in *Escherichia coli*) from *S. coelicolor* A3(2) (Wang & Cane, 2008) as well as *S. lasaliensis* (Komatsu *et al.*, 2008) has also been shown.

SAM-dependent methyltransferases (MTs) catalyze various C_1-transfer reactions. Among them, enzymatic *C*-methylations of macromolecules such as DNA and RNA have been studied extensively and their detailed mechanisms have been established. SAM-MTs also catalyze *C*-methylation of a number of secondary metabolites, which are biosynthesized *via* methylation of aromatic (tocopherol and novobiocin) and olefin substrates (steroid side chains and fatty acids). Although detailed analysis of cyclopropanation, which is mechanistically related to olefin *C*-methylation, has been reported, the crystal structure of *C*-methyltransferase (C-MT), which is frequently found in the biosynthesis of various natural products, has not been solved to date. GPPMT is the first enzyme observed to catalyze *C*-methylation of the universal isoprenoid precursor prenyl pyrophosphate. Here, we describe the crystal structures of free GPPMT, a binary complex with the cofactor *S*-adenosyl-L-methionine (SAM) and a ternary complex with GPP and sinefungin (SFG, an SAM mimic). GPPMT shows a novel hexameric assembly in the crystal as well as in solution. We discuss the detailed mechanism of the proposed enzymatic reaction based on the ternary crystal structure revealed in complex with the natural substrate GPP and SFG by comparison with the apo form and the binary complex form with SAM.

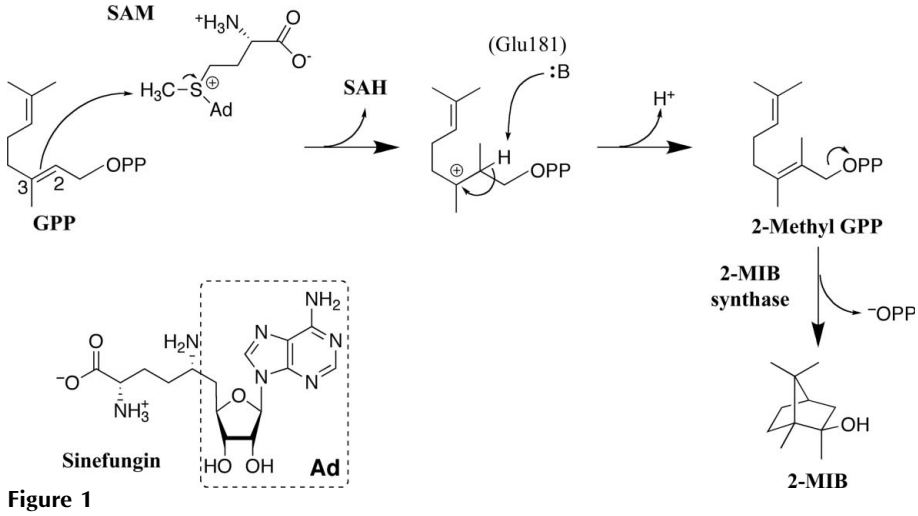

Figure 1
2-MIB biosynthesis pathway from GPP *via* 2-methyl-GPP. The structures of SAM and sinefungin (SFG) are also shown.

2. Materials and methods

2.1. Structure determination

The procedures used to prepare the proteins and crystals have been described previously (Ariyawutthiphan *et al.*, 2011). We used three types of crystals in this study, *i.e.* apo, a complex with SAM and a tertiary complex with both SFG and GPP (hereafter called the substrate complex), to analyze how this enzyme recognizes the substrates and its reaction mechanism. Methods for solving the structures have previously been published (Ariyawutthiphan *et al.*, 2011); the data-collection, processing and structure-refinement statistics are summarized in Table 1. All data sets were integrated, merged and scaled using *HKL*-2000 (Otwinowski & Minor, 1997). Molecular replacement was performed using *MOLREP* (Vagin & Teplyakov, 2010). A molecular-replacement solution for the apoprotein crystal was found using the structure of RebM (PDB entry 3bus; Singh *et al.*, 2008) as a search model. The sequence identity between these proteins is 22%, with an additional 15% strong similarity. Final structure refinement was carried out using *CNS* (Brünger *et al.*, 1998). NCS restraints were introduced for refinement of the substrate-complex form. The resolution limits of these structures were 2.2, 2.2 and 3.0 Å, respectively. In all crystal forms, the molecule assembles with point group 32; the molecules are crystallographically related in the $P6_322$ forms, but have local symmetry in the *P*1 form. SAM in the SAM-complex structure and SFG/GPP in the substrate-complex structure were easily modelled from residual electron density around the active site. In the later stages of refinement, residual electron density close to the pyrophosphate moiety of GPP was present in all chains of the substrate-complex structure. All terpene synthases require a divalent metal ion, usually Mg^{2+} (Davis & Croteau, 2000), suggesting the presence of Mg^{2+} in our GPPMT structure. Assigning Mg^{2+} at this position in all of the protein chains improved the *R* values. Thus, the position of Mg^{2+} in the

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

	Apo GPPMT	GPPMT–SAM	GPPMT–GPP–SFG
Data collection			
Wavelength (Å)	1.000	1.000	1.000
Space group	$P6_322$	$P6_322$	$P1$
Unit-cell parameters (Å, °)	$a = b = 143.9$, $c = 66.7$	$a = b = 147.9$, $c = 66.5$	$a = 78.9$, $b = 87.7$, $c = 160.2$, $\alpha = 100.0$, $\beta = 96.6$, $\gamma = 90.9$
Resolution range (Å)	50–2.20 (2.28–2.20)	50–2.20 (2.28–2.20)	50–3.0 (3.05–3.00)
Total No. of observations	317808	161781	290328
Unique reflections	21196 (2059)	22017 (2174)	82669 (4153)
Multiplicity	15.0 (12.7)	7.4 (7.3)	3.5 (3.5)
Completeness (%)	99.9 (100)	98.8 (99.6)	96.3 (95.8)
Mean $I/\sigma(I)$	33.9 (4.7)	28.6 (2.7)	5.90 (1.7)
R_{merge}	0.07 (0.421)	0.06 (0.424)	0.118 (0.359)
Wilson B factor† (Å^2) [resolution range used]	33.9 [4.0–2.2 Å]	52.8 [4.0–2.2 Å]	39.5 [4.0–3.0 Å]
Refinement			
R_{work}‡/R_{free}§	0.218/0.243 (0.275/0.304)	0.261/0.294 (0.504/0.515)	0.279/0.298 (0.435/0.434)
No. of atoms			
Protein	1925	1937	25514
Ligand/ion	0	20	563
Water	116	5	—
B factors (Å^2)			
Protein	30.6	64.2	50.9
Ligand/ion	—	71.0	51.3
Water	36.7	48.1	—
R.m.s. deviations			
Bond lengths (Å)	0.008	0.009	0.004
Bond angles (°)	1. 4	1.4	0.9
Ramachandran analysis¶ (%)			
Favoured	98.8	97.2	94.1
Outliers	0.0	0.0	0.6

† The Wilson B factor was calculated by *TRUNCATE* (French & Wilson, 1978). ‡ $R_{work} = \sum_{hkl} ||F_{obs}| - |F_{calc}|| / \sum_{hkl} |F_{obs}|$. § R_{free} was calculated using 5% of reflections excluded from refinement. ¶ The Ramachandran plot was calculated using *MolProbity* (Chen *et al.*, 2010).

substrate-complex GPPMT structure was confirmed next to the pyrophosphate moiety when considered together with the enzymatic assay results (described below). The final R and R_{free} values of the crystal structures were 0.218 and 0.243 for the apo form, 0.261 and 0.294 for the SAM complex and 0.279 and 0.298 for the substrate complex. The volume of the cavities was calculated using *POCASA* (Yu *et al.*, 2010) with a 2 Å probe size and a 1 Å unit grid. All of the figures showing protein structures were generated using *PyMOL* (DeLano, 2002). The atomic coordinates and crystal structure factors of geranyl pyrophosphate methyltransferase from *S. lasaliensis* in the apo form, in complex with *S*-adenosyl-L-methionine and in complex with sinefungin, geranyl pyrophosphate and magnesium ion were deposited in the PDB as entries 4f85 (free form), 4f84 (SAM complex) and 4f86 (substrate complex).

2.2. Site-directed mutagenesis

GPPMT mutants were generated by a two-step PCR method (Kammann *et al.*, 1989) using the wild-type cDNA in an expression plasmid as the template. The first PCR reactions were carried out using the general forward primer and each reverse primer, as well as the general reverse primer and each forward primer, for mutagenesis. The sequences of the general primers were as follows: forward, CATG**CCA-TGG**GCAGCAGCCATCATC; reverse, **CTCGAG**CTACACCCGGTCGGCCG (bases in bold indicate *Nco*I and *Xho*I restriction sites, respectively). The second PCR was carried out using the forward and reverse general primers and the products of the first PCR as substrates, which were harvested after agarose gel electrophoresis. The products from the second round of PCR were harvested after agarose gel electrophoresis and digested with the restriction enzymes *Nco*I and *Xho*I before being introduced into the plasmid pET-28b(+) (Novagen, San Diego, USA) such that 6×His was fused at the C-terminal ends. The primer sequences used for site-directed mutagenesis are listed in Supplementary Table S1[1]. The amino-acid coding regions of all mutants were sequenced using an ABI 3100 sequencer (Applied Biosystems, Foster City, USA). The expression and solubility of the recombinant proteins were confirmed by SDS–PAGE and Western blotting.

2.3. Enzymatic assay

200 µl reaction mixtures consisting of 1.1 µ*M* GPPMT, 100 µ*M* GPP and 100 µ*M* SAM in reaction buffer [50 m*M* Tris–HCl pH 8.0, 100 m*M* NaCl, 20%(*v*/*v*) glycerol, 10 m*M* $MgCl_2$, 2 m*M* EDTA, 0.2 m*M* β-mercaptoethanol] were incubated for 2 h. 25 µl 0.5 *M* EDTA pH 8.0 was then added to terminate the reaction. The reaction mixture was hydrolyzed with an excess of apyrase and acid phosphatase at 303 K for 2 h. The hydrolyzed products were extracted with *n*-hexane and analyzed by gas chromatography–mass spectrometry (GC-MS).

GC-MS analysis was performed using a GC-MS QP2010 apparatus (Shimadzu, Kyoto, Japan) with a DB-1 MS capillary column (0.32 mm × 30 m, 0.25 µm film thickness; J&W Scientific, Folsom, USA). Each sample was injected onto the column at 353 K in the splitless mode. After an isothermal hold at 353 K for 3 min, the column temperature was increased by 20 K min^{-1} to 513 K with an isothermal hold at 513 K for 3 min. The flow rate of the helium carrier gas was 1 ml min^{-1}.

Enzyme-activity assays of GPPMT with other divalent cations, including Zn^{2+}, Ca^{2+} and Mn^{2+}, were also carried out.

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: MH5072). Services for accessing this material are described at the back of the journal.

In each case, $ZnCl_2$, $CaCl_2$ or $MnCl_2$ was used instead of $MgCl_2$. Enzyme activity was only observed in the presence of Mg^{2+} or Mn^{2+}. Therefore, we used Mg^{2+} in the structure refinement.

2.4. Gel-filtration chromatography

Gel-filtration chromatography to check the assembly of GPPMT in solution was carried out using an S200 10/300 column (GE Healthcare, Fairfield, USA). The gel-filtration buffer consisted of 50 m*M* Tris–HCl pH 8.0, 100 m*M* NaCl, 20%(*v*/*v*) glycerol, 0.2 m*M* β-mercaptoethanol, 2 m*M* EDTA. Gel-filtration standards (Bio-Rad Laboratories, Hercules, USA) consisting of bovine thyroglobulin (670 kDa), bovine γ-globulin (158 kDa), chicken ovalbumin (44 kDa) and horse myoglobin (17 kDa) were used (Supplementary Fig. S1).

3. Results

3.1. Overall structure

The final models of both apo GPPMT and GPPMT–SAM comprised 246 amino acids and 248 amino acids, respectively, of the 300 residues of the full-length protein (Fig. 2*a*). The first 52 residues in GPPMT–SAM and the first 54 residues in apo GPPMT could not be modelled because of a lack of significant

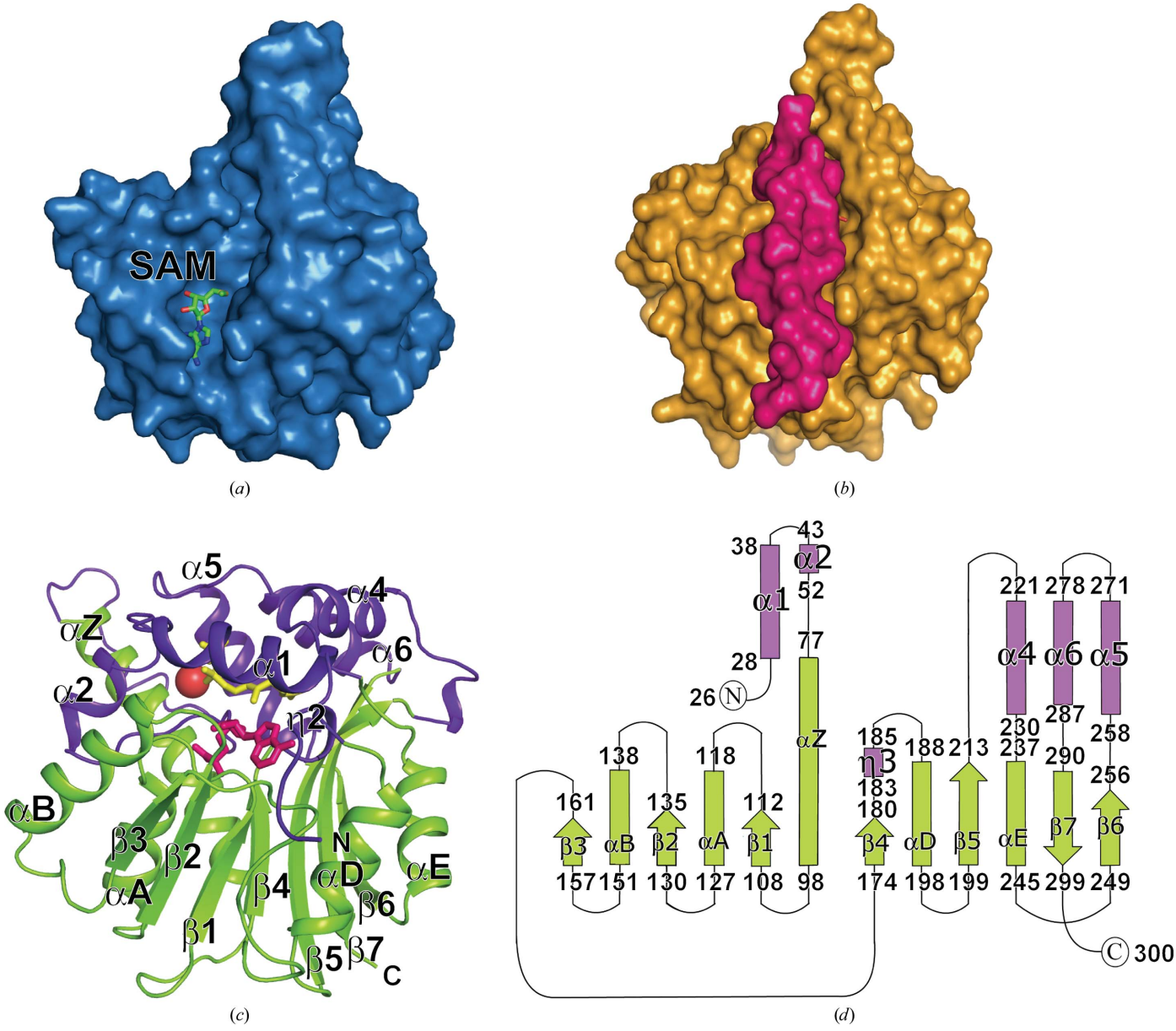

Figure 2
Overall structure of GPPMT. (*a*) Surface representation of the SAM-bound form. SAM is shown as a green stick model. (*b*) Surface representation of substrate-bound GPPMT. The structure of the apo form is shown in orange and the newly appearing part that is only observed in the substrate-complex structure is shown in magenta. (*c*) Ribbon diagram of substrate-bound GPPMT. The typical Rossmann-fold core is shown in light green and additional regions are shown in purple. SFG and GPP are shown as magenta and yellow stick models, respectively. The Mg^{2+} ion is shown as a red sphere. (*d*) Topological diagram of the substrate-bound form. The colouring is the same as that used in (*c*).

electron density. However, the electron density corresponding to residues 26–52 was sufficiently clear to build the model of the GPPMT–GPP–SFG structure (substrate-complex form; Fig. 2*b*). There is one GPPMT molecule in each asymmetric unit of apo GPPMT and GPPMT–SAM, whereas 12 molecules are present in the asymmetric unit of the substrate-complex form. Serious clashes between crystallographically related hexamers occurred if we attempted to model the N-terminal (residues 26–52) part of the apo GPPMT or GPPMT–SAM structures based on the substrate-complex structure. This also supports the flexibility of the N-terminal region.

The overall structure of GPPMT consists of a mixed α/β topology with a core region containing seven-stranded β-sheets and five α-helices (Figs. 2*c* and 2*d*), which is consistent with the highly conserved Rossmann-fold core typical of SAM-MTs (Martin & McMillan, 2002; Kozbial & Mushegian, 2005; Schubert *et al.*, 2003). The β-strands lie in the order 3–2–1–4–5–7–6, with only strand 7 lying antiparallel (Fig. 2*d*). Additional flanking helical regions are inserted before Val51 (only in the substrate-complex structure), between Asn213 and Arg237 and between Asp256 and Phe290 (Fig. 2*d*). No significant structural differences were detected between the apo form and the SAM-complex form of GPPMT (r.m.s.d. of 0.34 Å for C^{α} atoms 55–300). Comparison with the substrate-complex form showed a slightly increased r.m.s.d. value (0.57 Å for C^{α} atoms 55–300). This difference mainly arises from residues around the substrate and the Mg^{2+} ion in the substrate-complex structure as well as from the difference in crystal packing.

3.2. Hexamer formation

The structure of GPPMT is unique in terms of forming a hexameric assembly with point group 32 in both crystal forms (Figs. 3*a* and 3*b*). The buried surface area between neighbouring protomers is rather extensive (1010 and 796 $Å^2$ in the apo crystal structure) along two independent twofold axes as calculated using *PISA* (Krissinel & Henrick, 2007). There are two types of crystallographic twofold relationship and one threefold relationship in the hexamer. Specific hydrogen bonds and many specific van der Waals contacts can be observed between monomers. The inter-chain interactions of one pair include hydrogen bonds between Val254 and Val254′ (3.0 Å, β-sheet-like interaction between crystallographically twofold-related monomers, a pair of main-chain carbonyl and amide atoms; a prime sign after the residue number indicates a symmetry-related mate), Gln252′ and Asp256 (2.9 Å, side chain), Gln252′ and Gln291 (3.3 Å, side chain), and Arg237 and Asp298′ (2.7 Å, side chain) (Supplementary Fig. S2*a*). In

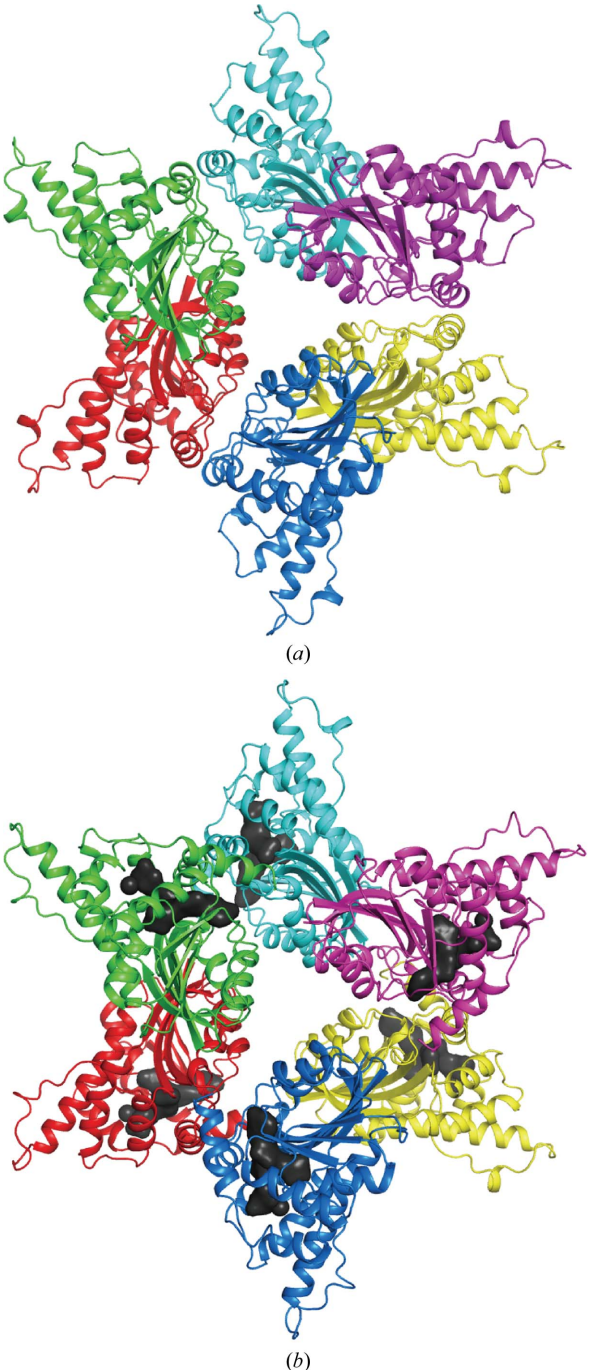

Figure 3
Ribbon diagrams of the hexameric assembly of the apo form (*a*) and the substrate-bound form (*b*). Each monomer is shown with different colouring. A side view of the substrate-bound form is shown in (*c*). Sinefungin and GPP are presented as grey surface models in (*b*) and (*c*).

addition to these polar interactions, significant hydrophobic interactions were observed between twofold-related monomers around the C atoms of residues such as Pro214, Arg238, Leu241, Ala245, Val250, Pro251, Val254 and Val300 (Supplementary Fig. S2*b*). With the other twofold mate (the other monomer related by the second crystallographic twofold axis), there are hydrogen bonds between Gln218 and Asn233′ (2.8 Å, side chain; Supplementary Fig. S2*c*) and Glu231 (side-chain carboxylate O atom) and Lys221 (3.0 Å, main-chain amide), as well as hydrophobic interactions involving C atoms from residues such as Gln218, Pro219, Lys221, Ala228 and Asn233 (Supplementary Fig. S2*d*). Hydrogen bonds can also be observed between threefold-related monomers; the side chains of Asp190 and Arg248′ are almost in the same plane, with a proper hydrogen-bonding distance (2.9 Å; Supplementary Fig. S2*e*).

These interactions, taken together with the observation that GPPMT forms a hexamer with local point group 32 as a substrate-complex structure, are intriguing and suggest that this enzyme uniquely acts as a hexamer. Interestingly, the N-terminal region, which consists of residues 22–53 (which are only observed in the substrate-complex structure), makes hydrophobic interactions with a twofold-related monomer, as well as hydrogen bonds (Supplementary Fig. S2*f*). The residues used in these interactions are Thr27, Pro28 and Tyr29, located at the N-terminus of helix α1, and Tyr216′, Gly217′, Gln218′ and Pro219′ from the neighbouring monomer. Hydrogen bonds can be observed between the amide N atom of Tyr29 and the carbonyl O atom of Tyr216′ (3.2 Å) and between the side-chain carboxyl O atom of

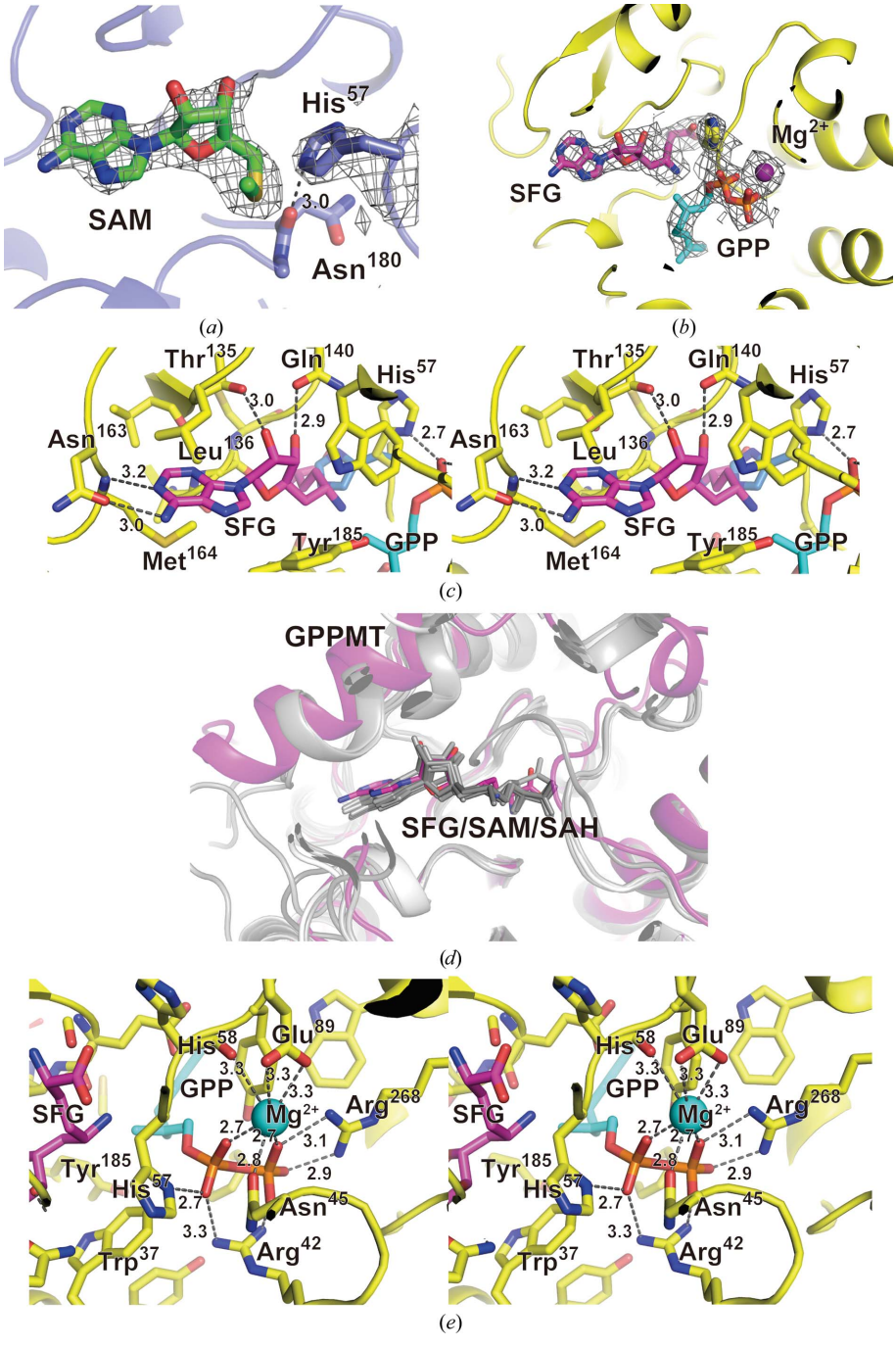

Figure 4
(*a*) Composite OMIT map ($2F_o - F_c$) for SAM in the SAM-complex structure. The contour level of the electron-density map is 1.2σ and the resolution is 2.3 Å. SAM is shown with O atoms in red, N atoms in blue, C atoms in green and S atoms in yellow. The protein is also shown as a light blue cartoon model with the side chain of His57 (ball-and-stick representation) interacting with the main-chain carbonyl O atom of Asn180. (*b*) Composite OMIT map ($2F_o - F_c$) for SFG, GPP and Mg^{2+}. The contour level of the electron-density map is 1.4σ and the resolution is 3.0 Å. SFG is shown with O atoms in red, N atoms in blue and C atoms in magenta. GPP is shown with O atoms in red, C atoms in light blue and P atoms in orange. Mg^{2+} is shown as a purple sphere. The protein is also shown as a light yellow cartoon model with the side chain of His57 in ball-and-stick representation. (*b*) is presented from the same angle as (*a*). (*c*) Active sites of the substrate-bound form. The colours of SFG and GPP are as in (*b*). GPPMT is shown with O atoms in red, N atoms in blue and C atoms in yellow. The side chain of His57 of apo GPPMT is also shown as a ball-and-stick model with N atoms in blue and C atoms in light blue. (*d*) Comparison around the active sites of GPPMT (substrate-bound form), CmaA1 (PDB entry 1kph), CmaA2 (PDB entry 1kpi), PcaA (PDB entry 1l1e) and Hma (PDB entry 2fk8). SFG and the protein structure of substrate-bound GPPMT are shown with O atoms in red, N atoms in blue and C atoms in magenta. The other proteins are presented as grey cartoon models. Cofactors (SAM or SAH) of the other proteins are also shown in grey. (*e*) Active-site architecture around the pyrophosphate moiety of GPP in the substrate-bound complex. Atoms are coloured as in (*c*). Mg^{2+} is shown as a light blue sphere.

Asp32 and the hydroxyl O atom of Tyr216′ (2.8 Å). These interactions increase the interface between the monomers (an increment of 170 Å^2 as calculated using *PISA*; Krissinel & Henrick, 2007) and would fix the hexamer.

The elution volume of apo GPPMT from gel-filtration chromatography using an S200 10/300 column (GE Healthcare, Fairfield, Connecticut, USA) was between that of standard proteins with molecular weights of 670 kDa (bovine thyroglobulin) and 158 kDa (bovine γ-globulin) (Supplementary Fig. S1); therefore, we concluded that GPPMT forms a hexamer even in solution (hexamer molecular weight of 210 kDa). Although some SAM-MTs in the same fold class act as homodimers or tetramers, the majority are monomeric proteins to the best of our knowledge. In the recently reported structure of GPPMT from *S. coelicolor* (*Sc*GPPMT), the corresponding residue Glu173 is proposed to play a role in stabilizing the carbocation intermediate (Köksal *et al.*, 2012).

3.3. SAM-binding site

A SAM-binding site with the same fold type as in SAM-MTs is present in the N-terminal region of the β-sheet and is formed by residues from the loops following β-strands 1, 2 and 3. Some residues that recognize SAM are conserved in SAM-MTs with the same fold type (Supplementary Fig. S3). The location of the SAM cofactor in the GPPMT–SAM structure was easily defined by the significant electron density corresponding to the adenosyl and sulfonium moieties (Fig. 4*a*), but lacked clear electron density around the methionine moiety of SAM. Therefore, we deleted this part from the refinement and the final model. In the substrate-complex structure, electron density for both SFG and GPP is unambiguous and is sufficient for discussion of the details of the substrate-recognition mechanism (Fig. 4*b*). The orientation of SFG is strictly determined by both hydrogen bonding and van der Waals contacts. Essentially, the recognition mode of SAM and SFG in these structures is almost identical for both the adenine and the ribose groups (Fig. 4*c*). The adenine group is placed in the hydrophobic milieu formed by Val134, Leu136, Met164, Tyr185 and Val186. The exocyclic N6 atom is within hydrogen-bonding distance of the Asn163 carboxamide O atom (3.0 Å; hereafter, distances are given for the *A* chain of the substrate-complex structure). The adenine ring N1 makes a hydrogen bond to the main-chain amide N atom of Met164 (3.2 Å). The ribose 2′-hydroxyl O atom is hydrogen-bonded to the side-chain hydroxyl O atom of Thr135 (3.0 Å), whereas the ribose 3′-hydroxyl O atom is hydrogen-bonded to the carboxamide O atom of Gln140 (2.9 Å). Surprisingly, the side chain of His57 is flipped in the presence of GPP to incorporate the amino-acid tail of SFG (Fig. 4*c*). The original orientation of the His57 side chain observed in the apo form is the same as that in the SAM-complex form. In the substrate-complex form, the carboxylate of the SFG methionine moiety is fixed through interaction with the imidazole-ring N atom of His58 (2.9 Å) and the main-chain amide N atom of His57 (3.3 Å).

Superimposing cofactor-bound structures of RebM (PDB entry 3bus; Singh *et al.*, 2008), CmaA1 (PDB entry 1kph; Huang *et al.*, 2002), CmaA2 (PDB entry 1kpi; Huang *et al.*, 2002), PcaA (PDB entry 1l1e; Huang *et al.*, 2002), Hma (MmaA4; PDB entry 2fk8; Boissier *et al.*, 2006) and GPPMT (this work) showed that the orientation and conformation of SAM [or *S*-adenosylhomocysteine (SAH) or SFG] are well conserved, as is the main-chain trace of the protein (Fig. 4*d*). In a hexameric assembly of GPPMT, six active sites are located around the outer rim and have access to the outer solvent region (Fig. 3*c*). Therefore, the hexameric arrangement does not seem to interfere directly with the enzymatic activity. However, interaction with GPP seems to make the hexameric assembly more stable (see §4). In the substrate-complex structure, the N-terminal region formed by helices α1, α2 and loop α2–αZ blocks access to the SAM binding site.

3.4. Substrate recognition

One of the distinct characteristics that are shared by all small-molecule SAM-MTs is the intrinsic N-terminal addition, which usually includes two helices (Martin & McMillan, 2002). In the structure of substrate-bound GPPMT the N-terminal segments are composed of two helices named α1 and α2, which are connected to the structurally conserved αZ helix by a long loop (residues 46–79; Fig. 2). These N-terminal helices, which are disordered in the apo and the SAM-bound forms, should undergo a structural rearrangement for incorporation of the substrate GPP and subsequent catalytic reaction. Helices that act as a lid for the active-site pocket have been reported in several small-molecule SAM-MTs, including RebM (Singh *et al.*, 2008) and cyclopropane mycolic acid synthase (CmaA1; Huang *et al.*, 2002). This region has been suggested to be flexible and to be fixed by the presence of the cofactor SAM. Notably, these N-terminal helices could not be modelled in either the apo or cofactor-bound GPPMT structures, suggesting that this part is flexible even in the presence of the cofactor SAM. From the extensive interaction with GPP, this N-terminal region is likely to be important for prenyl pyrophosphate substrate recognition. In contrast to SAM binding, substrate binding differs markedly among the small-molecule SAM-MTs (Martin & McMillan, 2002).

The two helices form a complete lid over the active site (Fig. 2*b*). The residues in this N-terminal region, which include a few residues from a long loop between α2 and αZ, interact with both SAM and GPP (Figs. 4*c* and 4*e*). The residues involved in interaction with the pyrophosphate moiety of GPP are mainly from the N-terminal region, such as Arg42, Asn45 and His57. Arg268, which is located in the α5 helix near the C-terminus, also contributes to recognition of the pyrophosphate moiety. Additionally, Tyr59 is in close proximity to the C2 atom of GPP (3.1 Å). We also found a magnesium ion in the active site coordinated by two O atoms of the pyrophosphate moiety (2.7 Å each). Another coordination site is occupied by the O atom of the Asn45 carboxamide (2.8 Å). Two carboxylate O atoms from Glu89 are located at 3.3 Å (Fig. 4*e*). A main-chain carbonyl O atom of His58 is also located within 3.3 Å distance. In contrast, several hydrophobic residues recognize the geranyl moiety of GPP by van der

Waals interactions, such as Tyr36, Trp37, Tyr59, Glu181 (alkyl part), Met184, Tyr185, Ile226, Phe230, Cys232, Ile234, Trp265, Phe281 and Phe290 (Fig. 4*e*).

3.5. Effects of divalent cations and selected mutations

Enzymatic activity was monitored by capillary GC-MS after enzymatic hydrolysis of the reaction mixture (Komatsu *et al.*, 2008; Supplementary Fig. S4). To confirm which divalent ion is essential for GPPMT catalysis, enzymatic activity assays were carried out under the following conditions: with 2 m*M* EDTA and in the presence of divalent cations such as Mg^{2+}, Zn^{2+}, Ca^{2+} or Mn^{2+} at 10 m*M*. GPPMT can convert GPP into 2-methyl-GPP in buffer containing 10 m*M* $MgCl_2$ (Supplementary Fig. S4), while no methylated product was detected on incubation with only 2 m*M* EDTA (Supplementary Fig. S4). The enzymatic activity of GPPMT with other divalent cations, including Zn^{2+}, Ca^{2+} or Mn^{2+}, was also tested and activity was only observed in the presence of Mg^{2+} or Mn^{2+}. The Mg^{2+} ion is located close to the pyrophosphate moiety of GPP in almost the opposite direction to the geranyl moiety (Fig. 4*e*) where methylation takes place; therefore, it is unlikely that Mg^{2+} participates directly in the methylation reaction. Judging from its location, Mg^{2+} is at least important in fixing GPP *via* the pyrophosphate moiety during the enzymatic reaction.

To investigate the roles of the residues around the GPP substrate, we generated mutants of Tyr59 and Glu181 because these two amino acids are closest to the C2 atom of the substrate GPP in our tertiary complex. We obtained Y59F and E181A mutants; however, an E181D mutant was insoluble. The Y59F mutant showed nearly comparable activity to the wild-type enzyme, whereas no activity was observed for the E181A mutant. The elution volume of the E181A mutant from the gel-filtration column was the same as

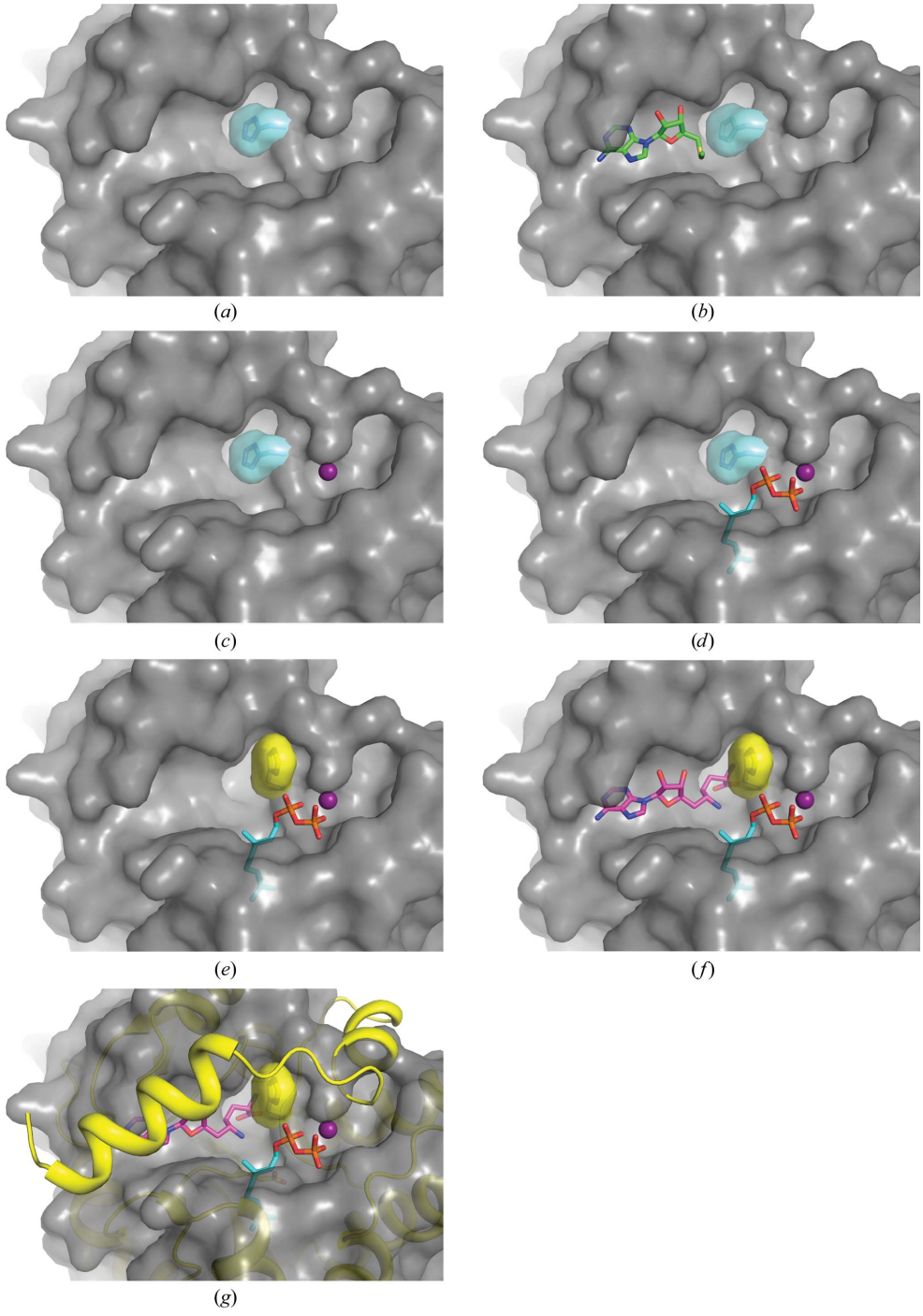

Figure 5
The active site of GPPMT as a protein-surface representation (shown in grey). (*a*) The side chain of His57 is shown in transparent light blue in the apo form. (*b*) The SAM-complex form. SAM is shown with O atoms in red, N atoms in blue, C atoms in green and S atoms in yellow. The side chain of His57 is also shown. (*c*) The location of Mg^{2+} is shown as a magenta sphere in the SAM-bound structure from superposition of the SAM-bound and substrate-bound forms. (*d*) The locations of Mg^{2+} and GPP are shown in the SAM-bound structure from superposition of the SAM-bound and substrate-bound forms. GPP is shown with O atoms in red, C atoms in light blue and P atoms in orange. (*e*) The locations of Mg^{2+} and GPP are shown with the orientation of His57 (shown in transparent yellow) based on the substrate-bound crystal structure. (*f*) The locations of Mg^{2+}, GPP and SFG are shown with the His57 side chain in the substrate-bound form. SFG is shown with O atoms in red, N atoms in blue and C atoms in green. (*g*) The structure of the bound form of GPPMT is added to (*f*) as a yellow cartoon representation.

those of the wild type and the Y59F mutant (data not shown). Therefore, we conclude that Glu181 is a key residue for the enzymatic activity of GPPMT.

4. Discussion

4.1. Ordered substrate incorporation

We obtained crystal structures in three different states, which allows discussion of how this enzyme incorporates the two substrates (GPP and SAM) with the magnesium ion and readily catalyzes the methyltransferase reaction. The apo crystal structure has a deep cavity as the active site (Fig. 5*a*). The volume of the cavity for GPP and magnesium ion in the apo-crystal structure is 111 Å^3; the cavity for SAM is 59 Å^3. Cocrystal structures with SAM showed little conformational change at the active site or in the whole protein (Fig. 5*b*). Attempts to generate cocrystals with only GPP failed and crystals from a mixture of GPPMT and GPP gave only the apo-type structure (Ariyawutthiphan *et al.*, 2011). However, the crystal structure of the substrate-bound form provides many insights into the incorporation of the substrate into the active site. As described previously, the methionine moiety of SAM could not be modelled. This would mean that this part is not properly placed without GPP (Figs. 5*b* and 5*d*). Surprisingly, we discovered that the side chain of His57 in the substrate-complex structure is flipped to open the space for the methionine moiety of SFG (a competitive inhibitor of SAM; Figs. 5*b* and 5*f*). This indicates that the binding of GPP to the active site in the presence of Mg^{2+} ion triggers a conformational change of the His57 side chain that permits the active site to accept SAM binding at the correct position (Figs. 5*c*–5*f*). This process is accompanied by capping of the active site with the α-helical N-terminal region (residues 22–53; Fig. 5*g*). After incorporating SAM and GPP into the active site of GPPMT, the incoming methyl group of SAM interacts with the olefin and forms a new C—C bond at C2, giving a carbocationic intermediate at C3 (Fig. 1).

The newly appearing N-terminal region, which consists of residues 22–53 and is only observed in the substrate-complex structure, has many roles in the recognition of GPP (described in §3.4). However, as this part acts as a lid to the active site, neither GPP nor SFG (or SAM) can access the active site after this lid is completely closed. Therefore, the incorporation and arrangement of these components (GPP, SAM, Mg^{2+} ion and the lid) to the appropriate locations seems to be initiated by Mg^{2+} and GPP, followed by SAM and the lid (Fig. 5).

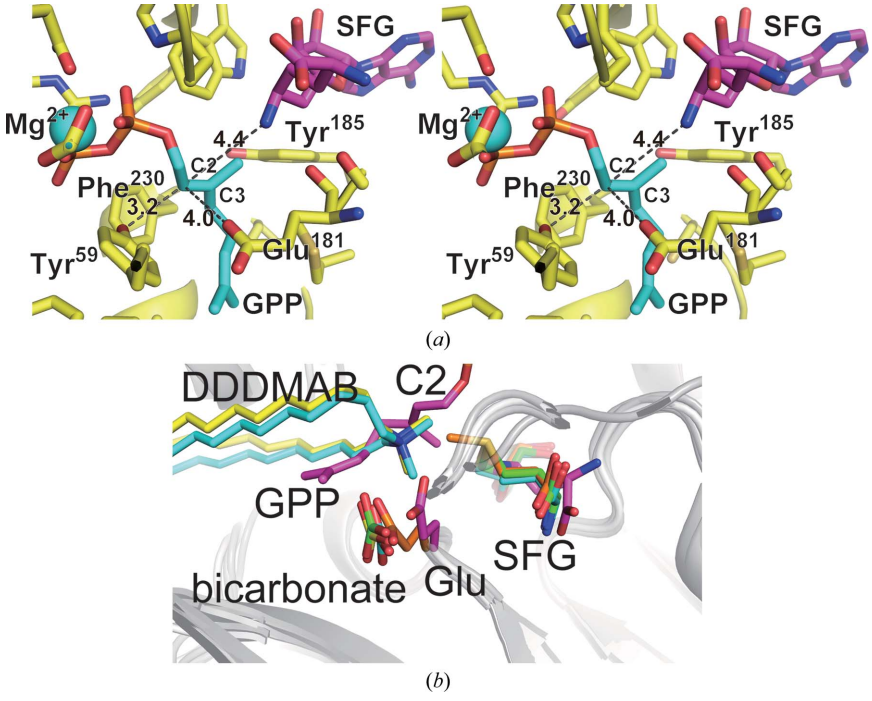

Figure 6
(*a*) Configuration of SFG, GPP and Mg^{2+} in the active site as a stereoview. Coordination of Tyr59 and Glu181 around the GPP C2 atom as well as Tyr59, Tyr185 and Phe230 facing the GPP C3 atom are also shown. The C2 and C3 atoms are labelled. The numbers next to dotted lines indicate the distance between atoms. The same colouring is used as in Fig. 4(*e*). (*b*) Conserved bicarbonate ion and Glu residue in the active sites of GPPMT (substrate-bound form), CmaA1 (PDB entry1kph), CmaA2 (PDB entry 1kpi), PcaA (PDB entry 1l1e; Huang *et al.*, 2002) and Hma (PDB entry 2fk8; Boissier *et al.*, 2006). All proteins are shown as grey cartoons. C atoms of SFG, GPP and Glu181 of the substrate-bound GPPMT is shown in magenta. C atoms of SAH, didecyldimethylammonium bromide (DDDMAB) and the bicarbonate ion of CmaA1 are shown in light blue. C atoms of SAH, DDDMAB and the bicarbonate ion of CmaA2 are shown in yellow. The C atom of SAH and the bicarbonate ion of PcaA are shown in green. The C atoms of SAM and Glu146 of Hma are shown in orange.

4.2. Proposed mechanism of the methyltransferase

GPPMT catalyzes olefin methylation at the C2 position of GPP. Related C-MTs are frequently involved in biosynthesis of secondary metabolites with aromatic and olefinic moieties. According to *BRENDA* (a web-based enzyme-information system; Scheer *et al.*, 2011), approximately 30 members of the SAM-MTs that catalyze the *C*-methylation reaction have been identified, of which only a few have been analyzed with regard to their structural basis. To date, two reaction mechanisms for SAM-MT-catalyzed *C*-methylation have been reported. The first example is DNA methylation. A two-step direct-displacement mechanism (conjugate addition–methylation of the α,β-unsaturated carbonyl system) was reported in the reaction of *C*-methyltransferase, a C5-cytosine DNA MT (Wu & Santi, 1987). This mechanism appears to be limited to the case of substrates such as RNA and DNA, in which the olefin is conjugated with a carbonyl group (Schubert *et al.*, 2003). The second mechanism is olefin methylation (methylation–deprotonation), which is involved in the biosynthesis of natural products such as steroids and triterpenoids. This stepwise

mechanism *via* a carbocationic intermediate is more generally applicable to substrates with an isolated olefin, such as GPP.

Using SFG, a mimic of SAM, we can illustrate the moment immediately before the transfer of the methyl group. The locations of the amino-group N atom of SFG and the GPP C2 atom are shown in Fig. 6(*a*). The distance between these atoms is around 4.4 Å. GPPMT-mediated methylation is plausibly initiated by electrophilic attack of SAM on the GPP C2═C3 double bond to form a carbocation intermediate at C3 (Fig. 1). Several aromatic planes of side chains, such as Tyr59, Tyr185 and Phe230, face the C3 atom of GPP (Fig. 6*a*), suggesting a role in stabilization of the carbocation intermediate through cation–π interactions. A similar situation has been observed in the active sites of mycolic acid cyclopropane synthases (MACSs; Huang *et al.*, 2002). The O atoms of the side chains of Tyr59 and Glu181 are located close to the C2 atom, with distances of 3.1 and 4.0 Å, respectively (Fig. 6*a*). As subsequent deprotonation of the GPP C2 atom is required to complete the reaction, these residues might act as a general base. Mutation experiments on these two residues (Y59F and E181A) indicated that Glu181 was responsible for the final deprotonation (Supplementary Fig. S4). In addition, three sterol SAM-MTs (SMTs) are known, the substrates of which are nonconjugated olefins from the sterol-synthesis pathway: sterol 24-*C*-methyltransferase (EC 2.1.1.41), cycloartenol 24-*C*-methyltransferase (EC 2.1.1.142) and 24-methylenesterol *C*-methyltransferase (EC 2.1.1.143). Amino-acid sequence alignment of GPPMT with these SMTs, as well as with IPPMT (described in §4.3), showed that this glutamate residue is completely conserved (Supplementary Fig. S5). These results suggest that Glu181 is the preferred general base critical for abstraction of the C2 proton.

Other mechanistically related examples have been studied in the cyclopropanation of fatty acids based on crystal structures (Huang *et al.*, 2002; Boissier *et al.*, 2006). In these cases, methyl-group transfer from SAM to the olefin initially occurs to generate a carbocation intermediate, followed by deprotonation of the resultant methyl group to give a cyclopropane ring (Fig. 7). The difference between MT and cyclopropanase is the position of deprotonation: at the α-proton from the carbocation in the MT reaction (Fig. 1) and at the β-proton from the carbocation in the cyclopropanase reaction (Fig. 7). Therefore, the location of the catalytic base and carbocation intermediate for deprotonation is important. In the cyclopropane-forming enzymes CmaA1, CmaA2 and PcaA, residual electron density around the active sites was observed and was assigned as a bicarbonate ion located close to the cyclopropanation site (Huang *et al.*, 2002), suggesting that the corresponding general base is a bicarbonate ion. A similar interaction of a bicarbonate ion was observed in *E. coli* cyclopropane fatty-acid synthase (CFAS; Iwig *et al.*, 2005; Courtois & Ploux, 2005). In contrast, the crystal structure of Hma (MmaA4; Boissier *et al.*, 2006), which catalyzes the introduction of a methyl branch together with an adjacent hydroxyl group, does not contain a bicarbonate ion; however, the carboxylate group of residue Glu146 is instead located in the appropriate position. Based on these structural data, the authors speculated that Glu146 is a potential base for deprotonation (Boissier *et al.*, 2006).

In our structures of GPPMT, the side chain of a potential base, Glu181, is located at the position corresponding to the bicarbonate ion or glutamate side chain in the loop connecting β4 and η3 (Fig. 6*b*). In the structures of CmaA1, CmaA2 and PcaA, the amino acid corresponding to the glutamate residue in GPPMT and Hma is glycine (Supplementary Fig. S3): a bicarbonate ion seems to be incorporated into the cavity created by the small glycine residue. Taken together, we propose that these invariant residues play similar catalytic roles (where Glu181 is equivalent to the general base) in each of these MTases for nonconjugated olefin substrates. Although crystal structures of the mycolic acid cyclopropane synthase (MACS) family have been reported, the substrate-bound structure has not been obtained because of difficulty in obtaining the actual substrate appended to an acyl carrier protein (ACP). Therefore, the structure of the tertiary complex GPPMT is most likely to reflect the actual reaction intermediate of the methyltransferase compared with that of MACS-family enzymes. In the recently reported structure of GPPMT from *S. coelicolor* (*Sc*GPPMT), the corresponding residue Glu173 is proposed to play a role in stabilizing the carbocation intermediate (Köksal *et al.*, 2012). This suggestion came from the fact that the side chain of Glu173 is on the same face of the C2═C3 double bond as the cofactor SAM and would be on the wrong face for deprotonation (Köksal *et al.*, 2012). *Sc*GPPMT was analyzed in complex with either SAH and GPP (PDB entry 3vc2) or with SAH and geranyl-*S*-thiolodiphosphate (GSSP; PDB entry 3vc1). In the structure of the complex with GPP (PDB entry 3vc2), only four monomers (*A*, *D*, *H* and *J*) among the 12 chains in the structure contained GPP in their active site. In the structure of the complex with GSSP (PDB entry 3vc1), all 12 monomers contained GSSP. Superimposing each monomer which harbours GPP or GSSP showed that the conformations of the prenyl side chains of each GPP and GSSP were diverse. Therefore, it is intriguing to think that GPP could move or change its conformation after accepting a methyl group to be depronated. The possibility of artifacts in the GPP conformation because of the use of SAH has also been mentioned (Köksal *et al.*, 2012).

4.3. GPPMT homologues in the terpenoid-modification pathway

Although SAM-MTs that catalyze the methylation of steroid or triterpenoid side chains have been reported, there have been no reports of their being involved in catalysing the methylation of universal prenyl pyrophosphate intermediates such as IPP, DMAPP or GPP. Methylation of these intermediates could provide incorrect final products after processing. As a result, these methylated products may become inadvertent substrates for many modification enzymes because of their low substrate specificity (Koyama & Seto, 1977). The finding that the side chain of His57 flips in the presence of the substrate GPP to appropriately accommodate

SAM may represent the presence of a safety lock. Because His57 is one of the three consecutive His residues (56–58) that are only observed in GPPMTs (Fig. 8), the mechanism of this switching may be uniquely conserved in GPPMTs. Because of this safety lock, this enzyme would not catalyze incorrect substrates. The location of the His57 side chain which is fixed through hydrogen bonding between the imidazole N atom and the main-chain carbonyl O atom of Asn180 (Fig. 4*a*; 3.0 Å) in both the apo and the SAM-complex structures prevents the amino moiety of SAM being incorporated. In fact, in the SAM-complex structure no electron density was observed for the methyl moiety of SAM because of a plausible block of His57. In the substrate-complex structure, the His57 side chain interacts with an α-phosphate O atom of GPP. It may be that only substrates with a phosphate moiety can release this safety lock.

To the best of our knowledge, GPPMT is the first enzyme that has been reported to catalyze this unusual methylation of prenyl pyrophosphate compounds. A putative IPPMT (termed Lon23) that could use IPP as a substrate was reported in the biosynthesis pathway of KS-505a (longestin; Hayashi *et al.*, 2007; Supplementary Fig. S6). Its catalytic activity has recently been confirmed (unpublished work). Pairwise sequence comparison between *S*/GPPMT and Lon23 indicated that they share 38% sequence identity and 33% similarity at the amino-acid level (Fig. 8). Among the residues that interact with the pyrophosphate moiety in the substrate-complex structure, Asn45, His57 and Arg268 are conserved. The residues that form the hydrophobic pocket to incorporate the geranyl moiety are also conserved in the Lon23 sequence. These residues are Tyr36, Trp37, Tyr59 (found as Phe in Lon23), Glu181 (alkyl part), Tyr185, Ile226, Phe230 (Tyr), Cys232, Ile234 (Thr), Trp265 (Phe), Phe281 (Tyr) and Phe290 (Ile). Importantly, Glu181, which has been suggested to act as a general base in this study, is also conserved in IPPMT. Furthermore, His57, which plays a key role in our safety-lock proposal, is also conserved in Lon23 (Fig. 8). *GPPMT* and putative *IPPMT* genes can be translationally coupled with genes encoding monoterpene cyclase and GGPP synthase, respectively. In general, polyprenyl synthases show a broad substrate specificity and accept artificial modified substrates. These observations indicate that methylation of common building blocks produces various unnatural methylated prenyl pyrophosphate intermediates, which may be incorporated into biologically important isoprenoids, thus increasing the risk of malfunction of living cells. Rapid trapping of methylated prenyl pyrophosphates with downstream enzymes that transform them into a specific product would avoid the formation of undesired isoprenoid derivatives. Although it is surprising that only two enzymes have been discovered in the terpenoid-biosynthesis pathway, the observations from the present study could be used to annotate new SAM-MTs that use prenyl phosphate compounds as substrates.

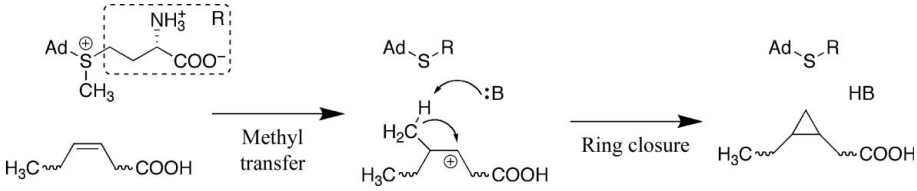

Figure 7
Proposed reaction mechanism for cyclopropane-ring introduction by MACS-family enzymes.

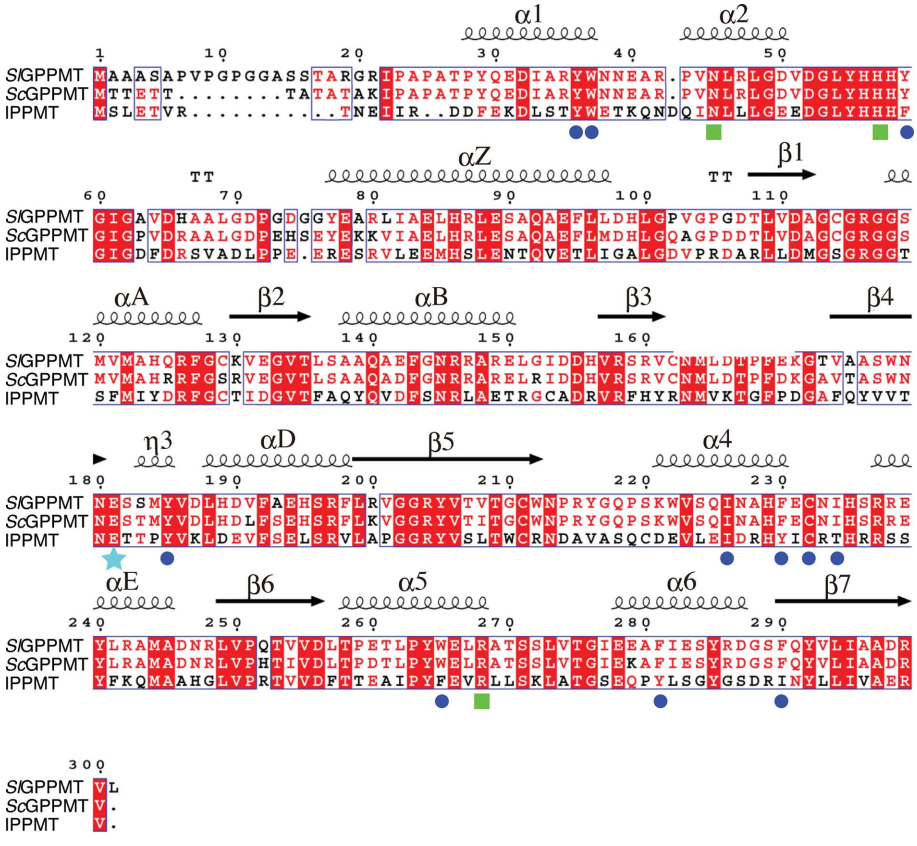

Figure 8
Amino-acid sequence alignment of GPPMT from *S. lasaliensis* (*Sl*GPPMT), GPPMT from *S. coelicolor* (*Sc*GPPMT) and Lon23 (IPPMT) from *S. argenteolus* (Hayashi *et al.*, 2007) presented using *ESPript* (Gouet *et al.*, 2003). Labels show the secondary structure of *Sl*GPPMT (substrate-bound form). Identical residues are highlighted in red and similar residues are framed in blue. Among the conserved residues, residues that recognize the pyrophosphate moiety of GPP are marked with green squares, residues that interact with the geranyl moiety of GPP are marked with blue circles and the glutamate residue proposed as a general base is marked with a light blue star.

We are grateful to Satoshi Omura and Haruo Ikeda (Kitasato Institute) for providing sinefungin and for helpful discussions. We also thank the beamline staff of SPring-8 (Hyogo, Japan) for technical help during the X-ray data collection. This work was supported by Grants-in-Aid for Scientific Research (22108002 to HO) from Japan Society for the Promotion of Science (JSPS). OA is supported by a scholarship from JSPS.

References

Ariyawutthiphan, O., Ose, T., Tsuda, M., Gao, Y., Yao, M., Minami, A., Oikawa, H. & Tanaka, I. (2011). *Acta Cryst.* F**67**, 417–420.
Bentley, R. & Meganathan, R. (1981). *FEBS Lett.* **125**, 220–222.
Boissier, F., Bardou, F., Guillet, V., Uttenweiler-Joseph, S., Daffé, M., Quémard, A. & Mourey, L. (2006). *J. Biol. Chem.* **281**, 4434–4445.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Courtois, F. & Ploux, O. (2005). *Biochemistry*, **44**, 13583–13590.
Davis, E. M. & Croteau, R. (2000). *Top. Curr. Chem.* **209**, 53–95.
DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.
Dickschat, J. S., Nawrath, T., Thiel, V., Kunze, B., Müller, R. & Schulz, S. (2007). *Angew. Chem. Int. Ed. Engl.* **46**, 8287–8290.
Eddy, S. R. (1998). *Bioinformatics*, **14**, 755–763.
French, S. & Wilson, K. (1978). *Acta Cryst.* A**34**, 517–525.
Giglio, S., Chou, W. K. W., Ikeda, H., Cane, D. E. & Monis, P. T. (2011). *Environ. Sci. Technol.* **45**, 992–998.
Gouet, P., Robert, X. & Courcelle, E. (2003). *Nucleic Acids Res.* **31**, 3320–3323.
Hayashi, Y., Onaka, H., Itoh, N., Seto, H. & Dairi, T. (2007). *Biosci. Biotechnol. Biochem.* **71**, 3072–3081.1
Huang, C. C., Smith, C. V., Glickman, M. S., Jacobs, W. R. & Sacchettini, J. C. (2002). *J. Biol. Chem.* **277**, 11559–11569.
Iwig, D. F., Uchida, A., Stromberg, J. A. & Booker, S. J. (2005). *J. Am. Chem. Soc.* **127**, 11612–11613.
Kammann, M., Laufs, J., Schell, J. & Gronenborn, B. (1989). *Nucleic Acids Res.* **17**, 5404.
Köksal, M., Chou, W. K. W., Cane, D. E. & Christianson, D. W. (2012). *Biochemistry*, **51**, 3003–3010.
Komatsu, M., Tsuda, M., Omura, S., Oikawa, H. & Ikeda, H. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 7422–7427.
Koyama, T., Ogura, K. & Seto, S. (1977). *J. Am. Chem. Soc.* **99**, 1999–2000.
Kozbial, P. K. & Mushegian, A. R. (2005). *BMC Struct. Biol.* **5**, 19.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Martin, J. L. & McMillan, F. M. (2002). *Curr. Opin. Struct. Biol.* **12**, 783–793.
Medsker, L. L., Jenkins, D., Thomas, J. F. & Koch, C. (1969). *Environ. Sci. Technol.* **3**, 476–477.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Scheer, M., Grote, A., Chang, A., Schomburg, I., Munaretto, C., Rother, M., Söhngen, C., Stelzer, M., Thiele, J. & Schomburg, D. (2011). *Nucleic Acids Res.* **39**, D670–D676.
Schubert, H. L., Blumenthal, R. M. & Cheng, X. (2003). *Trends Biochem. Sci.* **28**, 329–335.
Singh, S., McCoy, J. G., Zhang, C., Bingman, C. A., Phillips, G. N. & Thorson, J. S. (2008). *J. Biol. Chem.* **283**, 22628–22636.
Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Wang, C.-M. & Cane, D. E. (2008). *J. Am. Chem. Soc.* **130**, 8908–8909.
Wu, J. C. & Santi, D. V. (1987). *J. Biol. Chem.* **262**, 4778–4786.
Young, W. F., Horth, H., Crane, R., Ogden, T. & Arnott, M. (1996). *Wat. Res.* **30**, 331–340.
Yu, J., Zhou, Y., Tanaka, I. & Yao, M. (2010). *Bioinformatics*, **26**, 46–52.