

Characterization of Class IB terpene synthase: the first crystal structure bound with substrate surrogate

Rafaella Stepanova,^a Hayato Inagi,^b Kei Sugawara,^a Kazuya Asada,^a Tomoyuki Nishi,^a Daijiro Ueda,^a Yoko Yasuno,^c Tetsuro Shinada,^c Kunio Miki,^b Masahiro Fujihashi,^{b*} and Tsutomu Sato^{a*}

^a Department of Applied Biological Chemistry, Faculty of Agriculture, and Graduate School of Science and Technology, Niigata University, 8050 Ikarashi-2, Niigata 950-2181 (Japan).

^b Department of Chemistry, Graduate School of Science, Kyoto University, Sakyo-ku, Kyoto 606-8502 (Japan).

^c Graduate School of Science, Osaka City University, 3-3-138 Sugimoto, Sumiyoshi, Osaka 558-8585 (Japan).

* Corresponding Authors

Abstract

Terpene synthases (TS) are classified into two broad types, Class I and II, based on the chemical strategy for initial carbocation formation and motif sequences of the catalytic site. We have recently identified a new class of enzymes, Class IB, showing the acceptability of long (C₂₀-C₃₅) prenyl-diphosphates as substrates and no amino acid sequence homology with known TS. Conversion of long prenyl-diphosphates such as heptaprenyl diphosphate (C₃₅) is unusual and has never been reported for Class I and II enzymes. Therefore, the characterization of Class IB enzymes is crucial to understand the reaction mechanism of the extensive terpene synthesis. Here, we report the crystal structure bound with a substrate surrogate and biochemical analysis of a Class IB TS, using the enzyme from *Bacillus alcalophilus* (BaITS). The structure analysis revealed that the diphosphate part of the substrate is located around the two characteristic Asp rich motifs, and the hydrophobic tail is accommodated in a unique hydrophobic long tunnel where the C₃₅ prenyl diphosphate, the longest substrate of BaITS, can be accepted. Biochemical analyses of BaITS showed that the enzymatic property, such as Mg²⁺ dependency, is similar to those of Class I enzymes. In addition, a new cyclic terpene was identified from BaITS reaction products. Mutational analysis revealed that five of the six Asp residues in the Asp rich motifs and two His residues are essential for the formation of the cyclic skeleton. These results provided a clue to consider the application of the unusual large terpene synthesis by Class IB enzymes.

Introduction

Terpenoids are a large and highly diverse group of natural products consisting of the five-carbon units known as isoprene units. Many terpenoids (>75,000 compounds) have been isolated to date.^{1, 2} The structural diversity is mainly created by different terpene synthases (TS). The terpene biosynthesis is initiated by the isoprene unit synthesis and its linkage to form acyclic terpene diphosphate, which is further transformed into the diverse range of terpene skeletons. TS are classified into two broad classes, Class I and II, depending on the chemical strategy for initial carbocation formation and the characteristic motif residues around the catalytic site.³⁻⁶ Class I enzymes generate an allylic carbocation by the release of the diphosphate from the corresponding isoprenoid diphosphate. The characteristic motifs of Class I enzymes, DDXXD and NSE/DTE, hold the diphosphate part of the substrate through Mg²⁺ bridges and promote the generation of the carbocation. Recent study elucidated that the conserved sensor-linker-effector triad of residues are involved in the induced-fit rearrangement of the enzyme conformation.⁷ Class II has a DXDD motif, which plays a crucial role at the catalytic site in the protonation of a trisubstituted olefin to generate the carbocation (ex. TC shown in Fig. 1a). Class I enzymes are mainly associated with the biosynthesis of relatively smaller C₅-C₂₀ terpenes, whereas Class II enzymes synthesize the larger C₃₀ and C₄₀ terpenes.³⁻⁶ Recently we have identified a new TS (BsuTS; InterPro family: IPR019712) from *Bacillus subtilis* that accepts heptaprenyl (C₃₅) diphosphate (**1**) as a substrate both *in vivo* and *in vitro* to produce tetraprenyl- β -curcumene (**2**) (Fig. 1a).⁸ Although the mode of action seems to be similar to that of Class I, the new enzyme has no homology, including the motif residues with Class I. Moreover, it has been shown that the BsuTS homologs, namely BclTS and BalTS from *B. clausii* and *B. alcalophilus*, respectively, utilize C₂₅-C₃₅ prenyl-diphosphate as natural substrates to generate the β -springene (**6**) type terpenes (**7-9**) (Fig. 1b)⁹⁻¹¹. These two enzymes also accept C₂₀ prenyl-diphosphate (geranylgeranyl diphosphate: **3**) *in vitro* as a substrate in addition to C₂₅-C₃₅ compounds.¹¹ The crystal structure analysis of BalTS suggested that BsuTS homologs belong to a new subclass of TS named Class IB.¹¹ The three-dimensional structure of BalTS is similar to those of Class I TS, despite no similarity in their primary structures.¹¹ Two unprecedented motifs [DYLDNLxD and DY(F,L,W)IDxxED] containing six Asp residues are found at the corresponding positions of the DDXXD and

NSE/DTE motifs of the Class I TS.¹¹ Only the effector glycine residue was found regarding the sensor-linker-effector triad, whereas residues corresponding to neither sensor nor linker were seen. The two Asp-rich motifs were considered to contribute to the catalytic reaction, but insufficient information is available to elucidate the molecular mechanism of the Class IB enzymes. In this study, we report the three-dimensional structures bound with substrate surrogates and the functional studies of BaITS to gain insight into the mechanistic details of Class IB TS. We also found a novel cyclic diterpene by the detailed analysis of the BaITS product. These results promise that investigation of Class IB enzymes would lead to finding various new large ($\geq C_{25}$) terpenes with versatile functions.

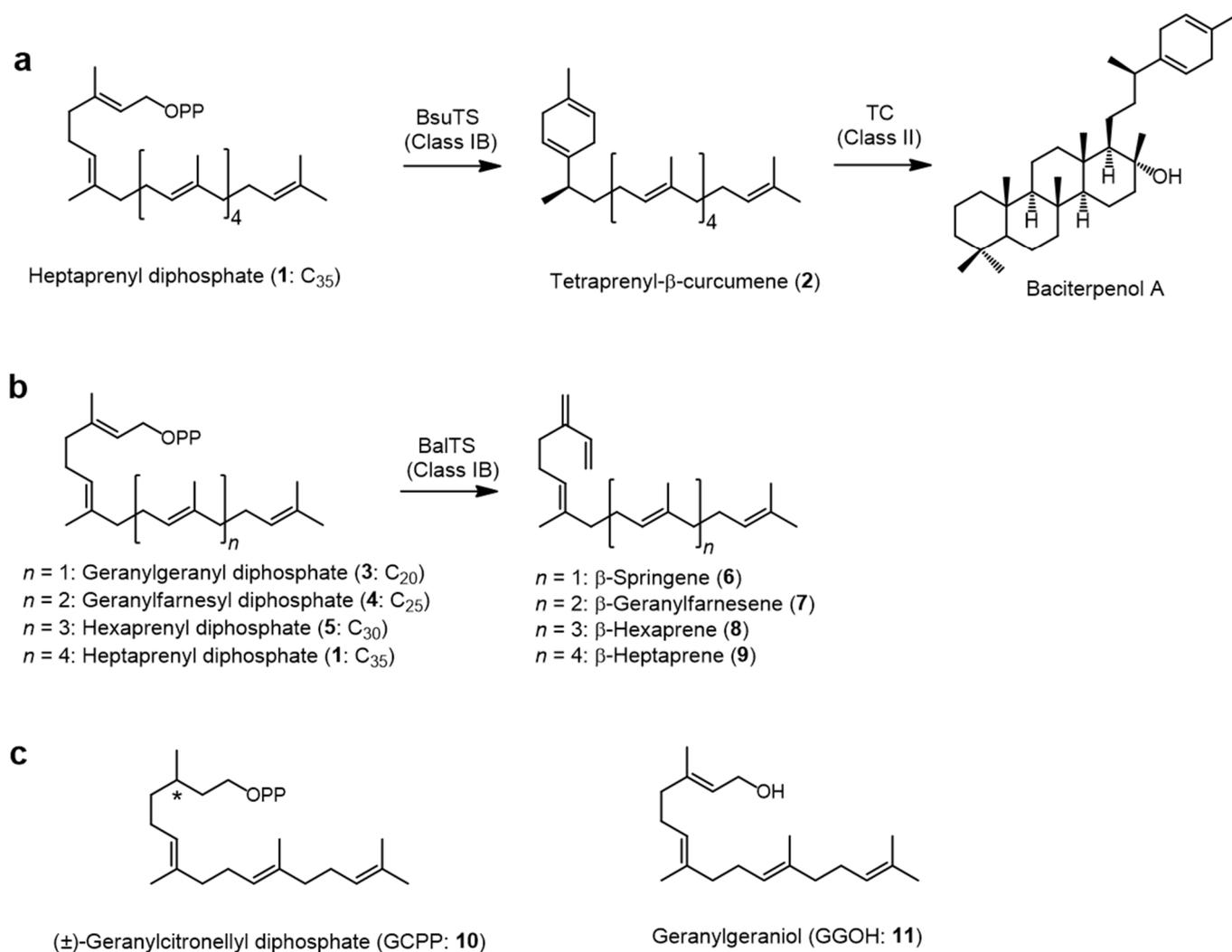


Fig. 1 Biosynthetic pathway and chemical structure of terpenoids mentioned in this paper. (a) Biosynthetic pathway of sesquiterpenes in *B. subtilis* (BsuTS). BsuTS (Class IB) catalyzes the cyclization of **1** into monocyclic **2**. **2** is converted to pentacyclic baciterpenol A by TC (Class II). (b) The reaction catalyzed by the BsuTS homolog from *B. alcalophilus* (BaITS: Class IB). (c) Substrate surrogates used for analyzing substrate-binding mode of BaITS. The asterisk indicates the chiral center generating racemate.

Results and Discussion

Structure analysis of the substrate-binding mode of Class IB terpene synthase

Our previous report¹¹ showed the three-dimensional structure of a Class IB terpene synthase (TS). However, because the structure was determined as a substrate-free form, the substrate-binding mode of this class is still unknown. In this study, we attempted to determine the crystal structure of BalTS bound with a substrate surrogate (\pm)-geranylcitronellyl diphosphate (GCPP, **10**, Fig. 1c) or geranylgeraniol (GGOH, **11**, Fig. 1c) to elucidate the catalytic mechanism. **10** was expected to mimic geranylgeranyl diphosphate (GGPP; **3**) since they could not form allylic cation in the enzyme like the case of 2,3-dihydrofarnesyl diphosphate used for the crystallographic analysis of selinadiene synthase.⁷ **11** was adopted to investigate the role of the diphosphate head in binding. The crystallographic structure of the BalTS-**10** complex was reasonably determined at 1.99 Å resolution (Supporting Table S1).¹²⁻¹⁷ As shown in Fig. 2a, the electron density corresponding to **10** was seen around the Asp-rich motif, and in the cleft of subunit A. In contrast, only noisy densities can be seen around the cleft of subunit B (Supporting Figure S1a). Thus, we discuss the binding mode of **10** in subunit A in detail. The difference between the two subunits may be related to the crystal packing. The density for the bound molecule in subunit A corresponds to the mixture of both stereoisomers of (\pm)-**10** (Fig. 2a). The diphosphate part is located around the Asp-rich motif of BalTS. Though Mg^{2+} , which is essential for the BalTS reaction, was expected to bridge between the Asp-rich motif and the diphosphate part of **10**, it could not be assigned. However, these results did not directly exclude the possibility of Mg^{2+} existence in the complex. One of the reasonable interpretations is that the two sets of Mg^{2+} coordination corresponding to each (*R* and *S*) isomer could have existed in the crystal, hence reducing the occupancy and peak density heights of Mg^{2+} ions corresponding to each isomer through the mixed binding of the racemate. Also, the mixed binding makes it hard to detect the characteristic octahedral coordination of Mg^{2+} .

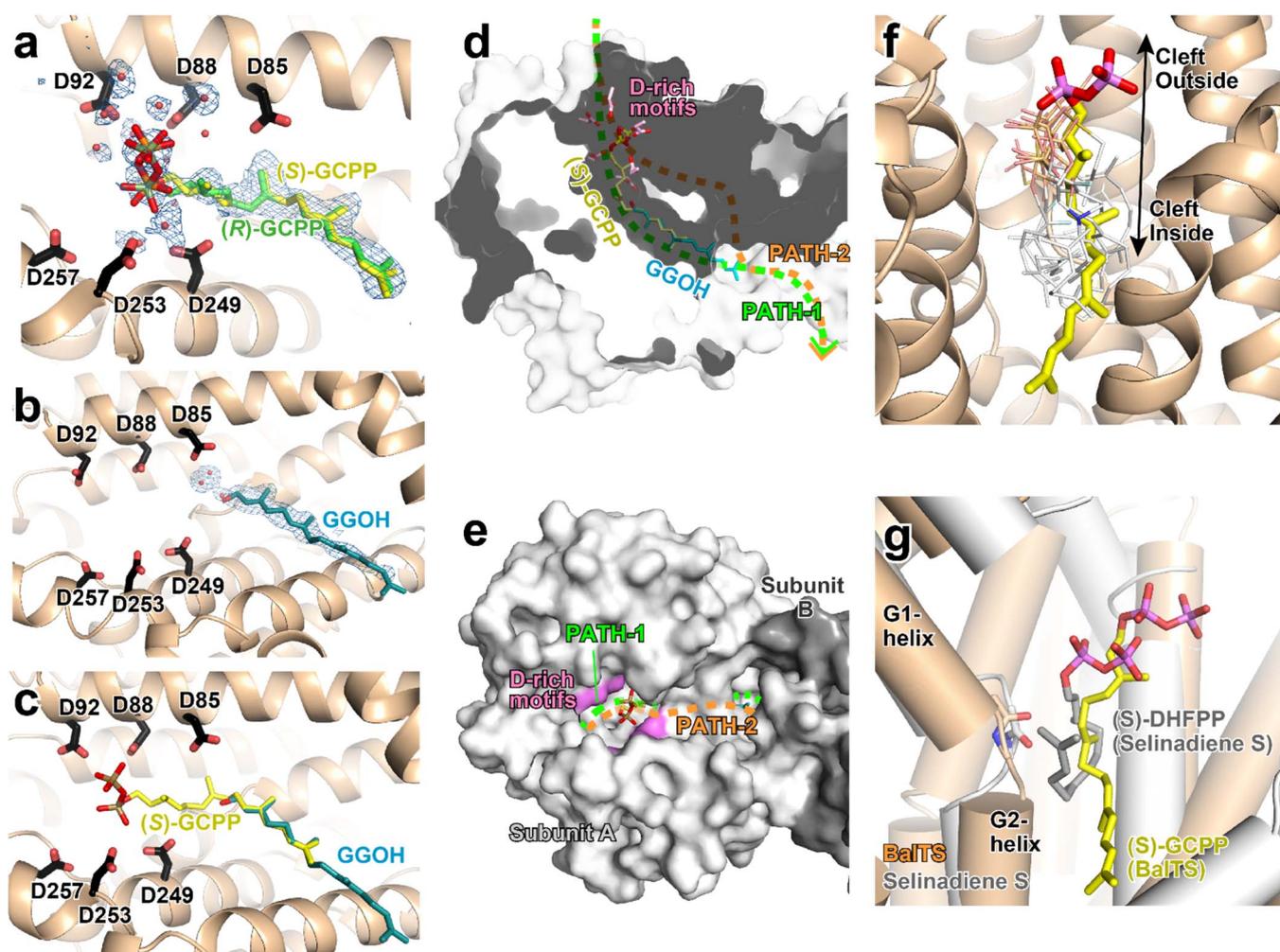


Fig. 2 Crystal structures of BalTS bound with substrate surrogates. The blue mesh indicates the omit electron density map drawn using the Program Polder.¹⁷ Asp residues in these panels belong to Asp rich motifs in Class IB terpene synthases. Yellow, green, and blue-cyan sticks show *S* and *R* isomer of **10**, and **11**, respectively. (a) Subunit A of BalTS bound with **10** (GCPP). Small red spheres show water molecules located within 5 Å from oxygens in Asp residues or **10**. The contour of the omit electron density map is at 3.5 σ . (b) Subunit A of BalTS bound with **11** (GGOH). Small red spheres show water molecules located around the OH group of **11**. The contour of the omit electron density map is at 4.0 σ . (c) Superposition of BalTS **10** and **11** complexes. BalTS molecule structure was drawn based on the BalTS-**10** complex. (d and e) Two paths on subunit A of BalTS, which are possible substrate-binding sites. The surface of the BalTS molecule is represented. Pink sticks and surfaces are Asp residues in Asp rich motifs. The compound **10/11** were found in path-1. The corresponding figures drawing subunit B is shown in Supporting Fig. S1c, and d. (f) Comparison of substrate(surrogate)-binding position between BalTS and representative Class I enzymes. The cartoon shows subunit A of the BalTS-**10** complex. Skinny sticks indicate the substrates or their surrogates bound in representative Class I enzymes. Carbons, oxygens, and phosphorous atoms in the skinny stick molecules are colored white, salmon, and ocher, respectively. Details of the superposed Class I enzymes¹⁸⁻²⁴ were described in supporting file. (g) Carbonyl groups of glycine at the G helix kink (between G1 and G2 helices). Subunit A of BalTS-**10** complex (wheat) is superposed on selinadiene synthase (Selinadiene S, gray) from *Streptomyces pristinaespiralis* in complex with (*S*)-dihydro farnesyl diphosphate ((*S*)-DHFPP) (PDB: 4OKZ).⁷

Polder electron density map clearly indicates the two phosphorous atom positions in **10** (Supporting Fig S2). In contrast, no clear peaks corresponding to oxygens were observed. The oxygens were modeled based on the phosphorous positions and ideal bond distances/angles of the diphosphate part. Some oxygens in Asp88, Asp 92, and Asp249 are located 3.1~3.4 Å from oxygens in **10**. The distances are within a typical range between the two oxygens bridged by an Mg²⁺ ion. One of the oxygens in Asp 253 locates only from 2.7 Å from the closest oxygen in the diphosphate. The distance is a little bit closer than the typical value, suggesting the overall model reliability is acceptable but not perfect. Distances from Asp85 and Asp257 to the phosphate is >5 Å, suggesting the two residues are not directly involved in the diphosphate recognition.

We also determined the crystal structure of BaITS in complex with geranylgeraniol (GGOH, **11**, Fig. 1c) at 1.91 Å resolution (Supporting Table S1). The conformations of **11** in subunits A and B are different from each other (Fig. 2b and Supporting Fig. S1b). The difference might be related to crystal packing like the complex with **10**. The superposition of the BaITS complex with **10** on that with **11** revealed well superpositioning of the subunit A of each complex (Table 1). The subunit B of the two complexes was also closely superposed. In contrast, the RMSD values of cross-superpositions, i.e., superpositions of a subunit on the other subunit of any complexes, are relatively high (Table 1). RMSD values between substrate-free structure and **10/11**-complex structures are also relatively high (Table 1). Note that the cell constants and crystal packing of substrate-free crystals¹¹ are different from those of **10/11**-complexes. These aspects suggest that the BaITS structure is flexible, and the protein conformation is a little bit restricted by the crystal packing. The binding site of **11** was, as expected, found in the interior of the cleft, which is quite different from that of **10**. The Asp-rich motif was far from **11** (Fig. 2b). The difference of the binding positions between **10** and **11** suggested the existence of an apparent interaction between the Asp-rich motif and the diphosphate part of the substrate probably through Mg²⁺. It is consistent with our previous mutational analysis that substitution of any one of Asp residues into Ala completely suppresses the catalytic activity of BaITS.¹¹

Table 1. Root-mean-square-distances (RMSD) between the determined structures.

	10-B ^[a]	11-A ^[a]	11-B ^[a]	free-A ^[b]	free-B ^[b]
10-A ^[a]	0.60 (345) ^[c]	0.14 (345) ^[c]	0.60 (345) ^[c]	0.51 (341) ^[c]	0.68 (340) ^[c]
10-B ^[a]		0.64 (347) ^[c]	0.16 (347) ^[c]	0.43 (343) ^[c]	0.66 (341) ^[c]
11-A ^[a]			0.63 (347) ^[c]	0.56 (343) ^[c]	0.73 (342) ^[c]
11-B ^[a]				0.43 (343) ^[c]	0.65 (342) ^[c]
free-A ^[b]					0.62 (345) ^[c]

[a] Subunit A or B of BaITS in complex with GCPP **10** or with GGOH **11**. [b] Subunit A or B of previously determined substrate-free BaITS.¹¹ [c] RMSD in Å between the corresponding subunits. Values in parenthesis are the number of superposed C α atoms.

Superposition of the subunit A of **10** complex on that of **11** complex led us to construct a hexaprenyl (C₃₀) diphosphate **5** binding model on BaITS (Fig. 2c). The third and fourth methyl groups from the diphosphate head of **10** are very well superposed on the first and second methyl groups from the alcohol group of **11**. The residues in the two complexes around the substrate-binding site are also very well superposed on each other (Supporting Fig. S3). These phenomena suggest the high reliability of this model. It should be noted that no \geq C₃₀ prenyl diphosphate binding model on TS have been determined so far. The C₃₀ substrate-mimic model is accommodated in the tunnel of subunit A. It was observed that each subunit of BaITS has a tunnel and a cavity (path-1 and path-2, respectively) that potentially contributed to substrate-binding (Fig. 2d, e). Both paths connected to the Asp-rich motif, and their wall was highly conserved and hydrophobic. The crystal structures of **10**- and **11**-complexes indicated that the tunnel (path-1) as the substrate-binding site. Because the tunnel is very narrow around Ile306 in the substrate-free structure (Supporting Fig. S4a and b), the substrate was previously assumed to be accommodated in the cavity (path-2).¹¹ Comparison of the **10/11**-complex and substrate-free structures indicated that the tunnel is enlarged with the movement of Ile306 by the induced fit with **10/11**-binding (Supporting Fig. S4). The proposed binding-site for the 7th isoprene unit is around the exit of the tunnel, i.e., the tunnel-length corresponds to C₃₅ prenyl-diphosphate (Supporting Fig. S4e). This feature may be related to the maximum (C₃₅) substrate size of

BalTS.¹¹ The two complex structures suggested that the hydrophobic tail of the substrate may initially enter the substrate-binding tunnel from the Asp-rich motif side because the narrow tunnel does not allow the passage of the diphosphate group. The hydrophobic feature of the tunnel also prevents the hydrophilic diphosphate group from passing through the tunnel.

Comparison between the substrate(surrogate)-bound structures of BalTS and Class I enzymes revealed several features of Class IB enzymes. The volume of the substrate-binding site is related to the size of the substrate molecule of each enzyme,¹⁸ and substrate-binding “tunnel” has never been found in Class I enzymes. The tunnel existence is highly consistent that all known Class IB enzymes, including BalTS, handle long (C₂₀~C₃₅) prenyl-diphosphate.¹¹ The diphosphate bound position of BalTS was located outside (the other direction of the hydrocarbon chain of the substrate(surrogates)) comparing to those of representative Class I enzymes (Fig. 2f).¹⁸⁻²⁴ The difference of the diphosphate binding sites corresponds to one or two phosphate diameter(s). In Class I enzymes, NSE or DTE motifs cooperatively recognize the diphosphate part of the substrate through one Mg²⁺ ion.¹⁸⁻²³ The difference of the motif sequences (i.e., NSE/DTE in class I and Asp rich in class IB) may affect the diphosphate position difference. The binding-position difference is subtle in the overall structure of these enzymes but may severely affect the scheme of the catalytic reaction and the chemical structure of the product because the difference is very large in the scale of substrate -molecule size.

Enzymatic Characterization of a Class IB terpene synthase

Next, for the first time, we performed enzymatic characterization of BalTS in detail. To this end, we analyzed the chain-length preference of BalTS using C₂₀-C₃₅ prenyl-diphosphate as substrates. The results showed that the shorter the carbon chain length, the more product was formed (Fig. 3a). We decided to use GGPP (C₂₀; **3**) as a substrate for this study to analyze the BalTS activity because the difference of the reaction rate that depends on the substrate chain length might be related to the facility of the product release. The product from the shorter substrate was considered to be less hydrophobic and easier to release to the solvent water after the reaction.

The optimal temperature and pH of BalTS were estimated to be 37°C and 10.0, respectively (Fig. 3b, c). The high pH preference could be due to the acceleration of the

diphosphate release under the high pH, wherein the preference could be related to the optimal environment for the source organism, *B. alcalophilus*,²⁵ thus may not be a general feature for the Class IB enzyme. Our previous study revealed that Mg^{2+} is essential for the activity of BalTS,¹¹ but no information was available regarding its optimal concentration. Here we investigated the divalent cation dependency (Fig. 3d, e) and observed that BalTS requires > 5 mM Mg^{2+} for its full activity (Fig. 3d). No divalent cations could substitute the roles of Mg^{2+} (Fig. 3e). It was observed that the Na^+ and K^+ do not affect BalTS activity (Fig. 3f, g), unlike the reports regarding K^+ of some Class I TS.²⁶ We have added dithiothreitol (DTT) while measuring the reported activity of BalTS,¹¹ but the compound was found to inhibit the BalTS reaction (Fig. 3h).

Specific activity under the optimal condition [25 mM 3-(Cyclohexylamino)-2-hydroxy-1-propanesulfonic acid (CAPSO) buffer (pH 10.0), and 10 mM $MgCl_2$] is 48.0 ± 0.2 nmol/min/mg using 130 μM of the compound **3** as a substrate was estimated (Fig. 3i). The value was ca. 3 times higher than that reported previously (14.1 ± 1.0 nmol/min/mg)¹¹ under the condition of [25 mM Tris-HCl buffer (pH 8.5), 10 mM DTT, and 1 mM $MgCl_2$]. The K_m and k_{cat} values for BalTS were calculated to be 37 ± 3 μM and 2.7 ± 0.1 min^{-1} , respectively (Fig. 3i), under the optimal conditions. These values are comparable to those obtained for the previously characterized bacterial Class I TS (ex. DtcycA from *Streptomyces* sp. SANK 60404: K_m 94 ± 8 μM and k_{cat} 2.8 min^{-1} ; DtcycB from *Streptomyces* sp. SANK 60404: K_m 42 ± 7 μM and k_{cat} 1.3 min^{-1}).²⁷ In total, the enzymatic properties of Class IB TS are similar to those of Class I enzymes except for the capability of BalTS to synthesize much larger terpenes than the Class I enzymes.

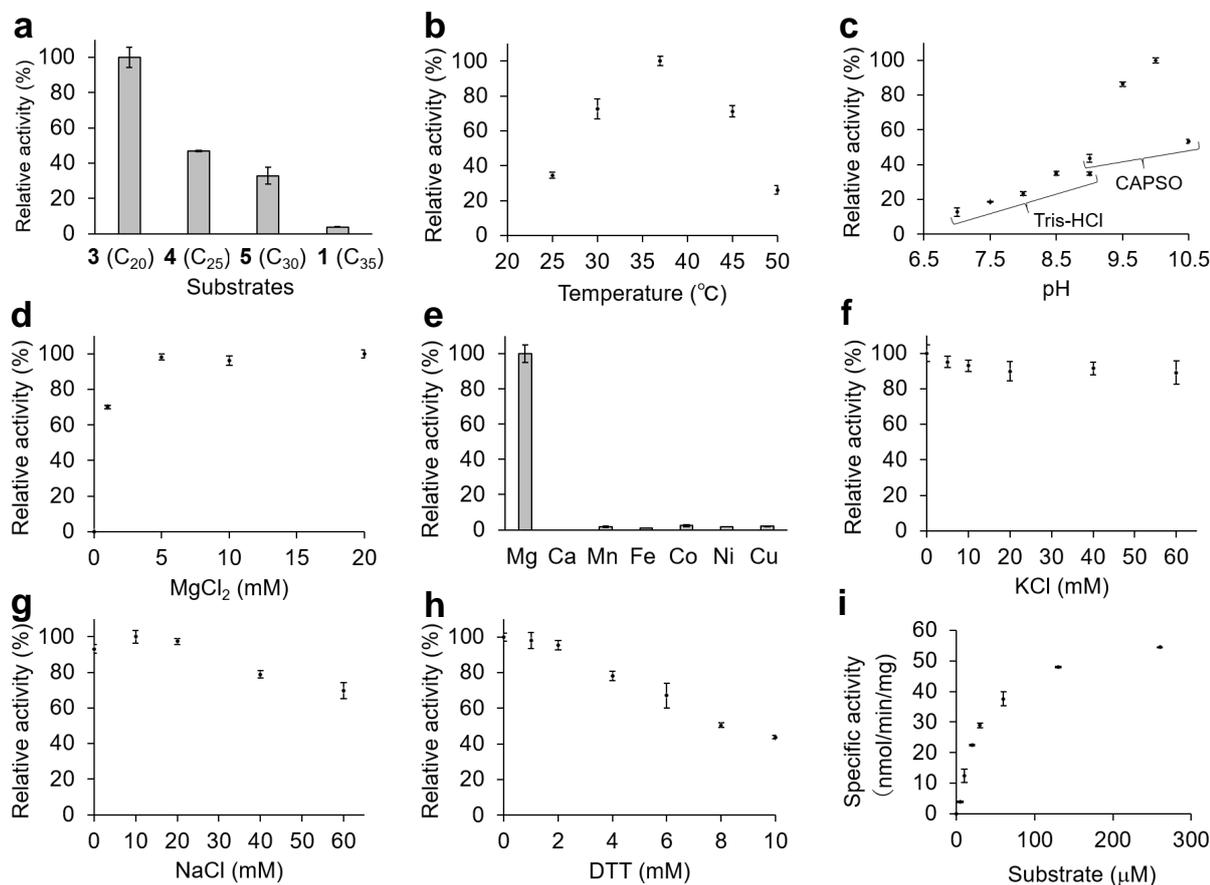


Fig. 3 Characterization of BaITS. Error bars indicate the standard error. (a) Analysis of substrate specificity for 130 μM substrates (**1**, **3-5**). (b) Analysis of optimal temperature (25, 30, 37, 45, 50°C). (c) Analysis of optimal pH. The buffer solutions used in the experiment were as follows: 25 mM Tris-HCl buffer for pH 7.0–9.0; and 25 mM CAPSO buffer for pH 9.0–10.5. Except for the buffer reagent and/or pH, the standard reaction mixture was used for the assays. (d) Analysis of optimal Mg^{2+} concentration (0, 1, 5, 10, 20 mM). Only MgCl_2 concentration was changed from the standard reaction mixture. (e) Analysis of requirements for divalent metal ions. Concentrations of all ions were 10 mM. These ions substitute MgCl_2 in the standard reaction mixture. (f and g) Analysis of optimal K^+ (f) and Na^+ (g) concentrations. These monovalent ions were added to the standard reaction mixture. (h) Analysis of optimal DTT concentration. DTT was added to the standard reaction mixture. (i) Specific activity of BaITS with substrate **3** (0, 5, 10, 29, 30, 60, 130, 260 μM).

A new cyclic product formed by BaITS

As shown in Fig. 4, BaITS produces a considerable amount of the minor product (**12**) in addition to the main product **6**. The production ratio of **6** and **12** was almost constant (**6**:**12** = ca. 8:1) under various reaction conditions, including temperature, pH, as well as Mg^{2+} and monovalent ion concentrations. The chemical structure of **12** was analyzed by electron ionization mass spectrometry (EIMS) and nuclear magnetic resonance spectroscopy (NMR) (Supporting Fig. S5–S12). The molecular formula of **12** was assigned as $\text{C}_{20}\text{H}_{32}$

based on high-resolution EIMS (HREIMS). Although $C_{20}H_{32}$ was identical with that of β -springene (**6**), further NMR analysis suggested that **12** would be a new cyclic diterpene (Fig. 5). 1H -NMR data of **12** showed six olefinic protons, four methyls, six methylenes, and two methine protons. One of the four methyls was observed as a doublet. The other three were singlets. Two of the six olefinic protons are putatively assigned as those of an exo-olefin. The ^{13}C -NMR analysis supported the presence of four olefins. Two dimensional NMR analysis, the HMBC correlation from H-16 (δ_H 4.92, s; $J=4.96$ Hz) to C-4 (δ_C 130.1), H-16 (δ_H 4.92, s; $J=4.96$ Hz) to C-3 (δ_C 143.8), and H-4 (δ_H 6.34, *dd*; $J=10.0$ Hz, 2.8 Hz) to C-3 (δ_C 143.8), and the 1H - 1H COSY correlation of H-4 (δ_H 6.34, *dd*; $J=10.0$ Hz, 2.8 Hz) with H-5 (δ_H 5.73, *brd*; $J=10.0$ Hz) suggested the characteristic cyclohexene structure with a conjugated diene moiety of **12**. (Fig. 5). The position of the methyl group observed as a doublet was placed adjacent to the cyclohexene ring. The new diterpene **12** showed a levorotation value ($[\alpha]^{25}_D -0.02$ (c 0.42, EtOH)), indicating that compound **12** could be a chiral product. The compound **12** seems not to be racemate because GC-MS analysis with chiral column used for the optical resolution of manool²⁸ showed only one peak (Supporting Fig. S12). However, the stereochemistry of the chiral center of **12** could not be determined at this time because only a small amount of **12** (0.38 mg) was isolated. BalTS product **12** is a novel diterpene and named as geranyl- β -phellandrene according to the structural similarity with β -phellandrene (monoterpene) and β -sesquiphellandrene (sesquiterpene).^{29, 30}

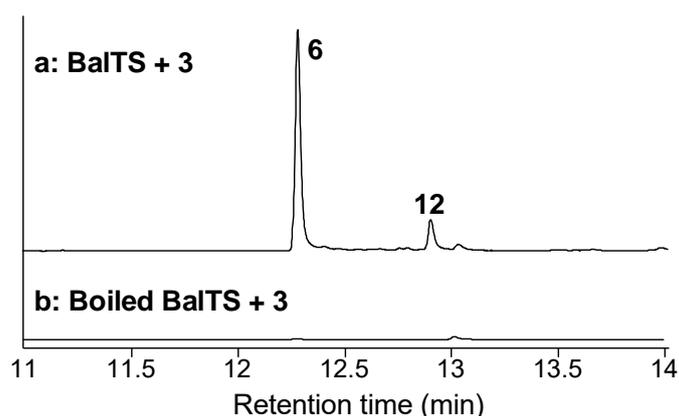


Fig. 4 The GC analysis of the reaction products formed by the incubation of **3** with (a) BalTS and (b) boiled BalTS

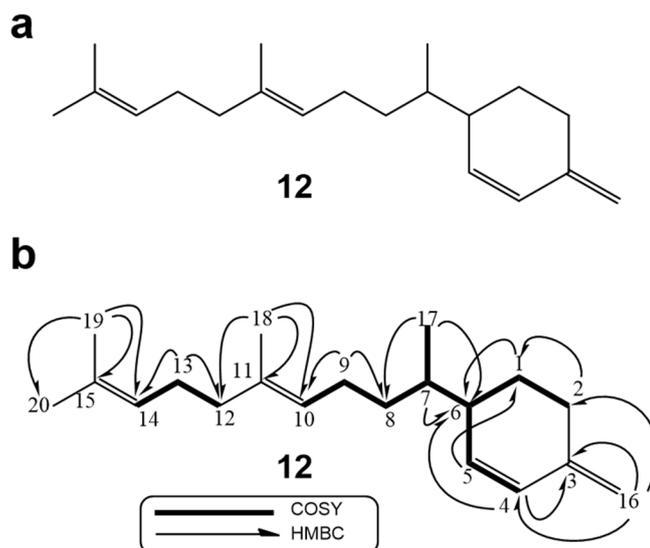


Fig. 5 Chemical structure of **12** (a) and key correlations in the 2D NMR data for **12** (b).

The minor product formation could be due to the flexibility of the BaITS structure (Table 1). It was suggested that the volume of the substrate-binding cleft might be responsible for catalytic fidelities in addition to the substrate size by controlling the number of allowed conformation of the substrate in the enzyme.^{18, 31} Thus, the flexible substrate-binding site of BaITS could allow several conformations. The cyclization reaction may be supported by G-helix kink (Fig. 2g). The kink is found in both the Class I and Class IB enzymes,^{7, 11} and a glycine residue is located at the kinked position in the two classes. The carbonyl oxygen of the glycine is suggested to stabilize the carbocation intermediate formed by the diphosphate release in selinadiene synthase, a Class I sesquiterpene synthase.⁷ The glycine corresponded to Gly208 in BaITS, and the carbonyl oxygen of Gly208 directs the 2nd isoprene unit of **10** (Fig. 2g). This feature showed that the carbonyl group of the glycine at the G-helix kink might not be related to the diphosphate release in BaITS. Instead, the glycine might play a pivotal role in the cyclization forming the minor product **12**. The roles of the carbonyl oxygen at G-helix kink in Class IB enzyme may be different from that of Class I enzymes.

Role of residues and reaction mechanism of Class IB terpene synthases

The differences of the substrate-binding position of the diphosphate and the distribution of the residues around the substrate between the Class IB (BaITS) and Class I enzymes, led us to analyze the roles of BaITS residues. The establishment of the optimal condition for the enzymatic assay facilitated mutational analysis in order to elucidate the roles of characteristic residues of Class IB enzymes. First, we targeted six Asp residues in the Asp rich motifs, which are conserved in more than 97.5% of Class IB enzymes.¹¹ The BaITS-**10** complex structure showed that the diphosphate group of **10** is located near the Asp rich motifs (Fig. 6a). Our previous work elucidated that mutations of any one of these residues into alanine completely suppress the product formation,¹¹ and reanalysis in this study under the optimal conditions led to the same results (data not shown). However, alternation from Asp into Ala might be too drastic to analyze the role of these residues. Thus, Asp to Asn mutations was introduced to investigate these residues. As shown in Fig. 6b, the amounts of compound **6** formation by any of the mutants (Asp85/88/92/249/253/257 to Asn) was about 5% of wild-type enzymes. Gas-chromatography mass-spectroscopy (GC-MS) analysis (Figs. 6c and S13) showed that only Asp92Asn mutant produces the minor product **12**, whereas the other five Asp mutants (Asp85/88/249/253/257 to Asn) lost the ability to form **12**. The product profiles indicate that the five Asp residues in the Asp rich motifs were essential for the formation of the cyclic skeleton. The difference could be attributed to the distances from these residues to the isoprene units of the substrate. Asp92 is very far from the isoprene units (~12 Å from the double bond in the 2nd isoprene unit), thus may not contribute to the cyclization of the substrate.

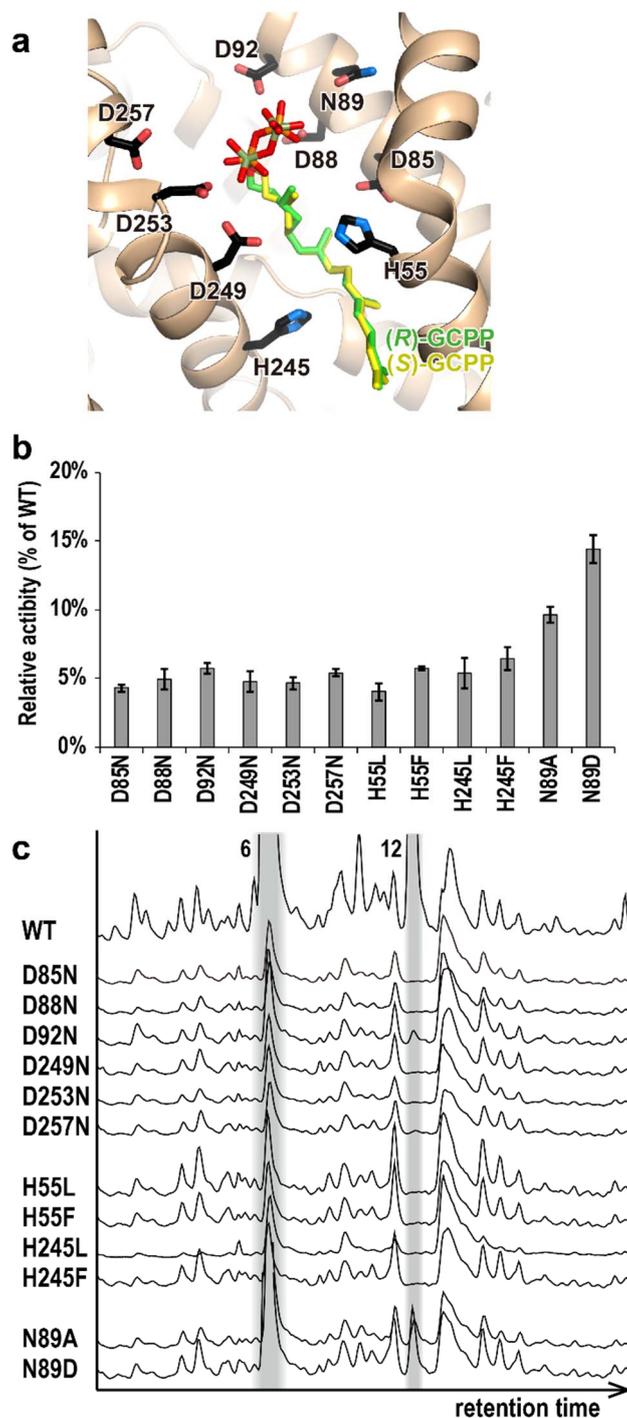


Fig. 6 Mutational analysis. (a) Location of investigated residues. (b) Amounts of β -springene **6** (main product of BalTS) formation by mutants. Relative amounts comparing to that by wild-type BalTS are shown. Error bars represent standard errors. (c) Product analysis of wild-type and mutants of BalTS by GC-MS. Peaks corresponding to compounds **6** and **12** are highlighted.

Two of the Asp residues Asp85 and Asp249 in the Asp rich motifs are closely located to C4 (methyl group of the 1st isoprene unit from the diphosphate) rather than the diphosphate group of **10** (Fig. 6a). The methyl group is also near (4-5 Å) to His55 (~98% conserved in Class IB). These residues may contribute to deprotonate C4 to terminate the reaction producing the compound **6**. In order to evaluate this hypothesis, His55 was mutated into Leu and Phe. Replacement from His to Leu or Phe disabled the proton transfer, and the sizes of these two residues were not so different from that of His. The activities of the mutants were ~5% of the wild-type enzyme (Fig. 6b), which were almost equivalent to those of the His245Leu and His245Phe mutants (Fig. 6b). Contribution of His245 to C4 deprotonation is unlikely, because, the residue contacts to second and third isoprene units of **10** (Fig. 6a). The equivalent impacts on enzymatic activities by Asp85, Asp249, His55, and His245 mutations imply that these four residues do not significantly contribute to the C4 deprotonation to form **6**. Regarding the cyclization to form **12**, the corresponding peak was not detected in the GC-MS spectrums for products by both His55 and His245 mutants like Asp85/249Asn (Fig. 6c). These results showed that His55 and His245 were also essential for the formation of the cyclic skeleton. These His mutants may change the conformation of the bound substrate, which prevents the cyclization forming **12**. Previously, it was reported that the C-terminal region of BsuTS is lacked in BalTS and is important for the formation of a cyclic structure.¹¹ In this study, we additionally found seven residues (Asp85/88/249/253/257 and His55/245) contributing to the cyclic skeleton formation in BalTS based on both the crystal structure and point mutations. Since these seven residues are also conserved in BsuTS, they may also be required for cyclization in BsuTS in addition to the C-terminal region.

Superposition of BalTS structure on Class I enzymes indicated that Asp88 and Asp92 correspond to the first and fifth Asp residues in DDXXD motifs, respectively (Supporting Fig. S14). The second Asp corresponds to Asn89 in BalTS, which is >95% conserved in Class IB enzymes.¹¹ The residue was mutated into Asp in order to construct an artificial DDXXD motif in the BalTS structure. As shown in Fig. 6b, Asn89Asp indicated 14% activity of wild-type enzymes. The activity is comparative to the Asn89Ala mutant. GC-MS spectrum showed that the product profiles of these two mutants are mostly the same as that of the wild-type (Fig. 6c). The non-enhancement of activity and no-alternation of the products of the Asn89Asp mutant showed that the DDXXD motif in Class I enzymes is not compatible with the Asp rich motif in Class IB enzyme.

One of the conceivable catalytic mechanisms of the Class IB enzyme is shown in Supporting Fig. S15. The mechanism would be rationalized by the same as that of β -phellandrene produced by Class I enzymes.³² Further studies such as isotope incorporation assay and/or computational analysis are required to unveil the reaction mechanism of Class IB enzymes.

GC-MS analysis in our previous work¹¹ showed peaks possibly corresponding to phellandrene derivatives synthesized from substrates **4** (C₂₅) and **5** (C₃₀) by BaITS. The considerable overlap of **11** on **10** in the substrate-binding site (Fig. 2c) suggested that the conformations of **3** and **5** in the enzyme are equivalent, especially regarding the reacting atoms around the diphosphate. The equivalent binding mode may lead to the formation of the same ring. **12** is a second cyclic skeleton synthesized by Class IB enzymes.⁸

In conclusion, we determined the first crystal structures of Class IB terpene synthase in complex with substrate surrogate. The hydrophobic long carbon chain is bound in a unique long hydrophobic tunnel, which has never been seen in Class I enzymes. The tunnel length is consistent with the maximum carbon chain length of the substrates for the enzyme. The diphosphate group of the substrate surrogate was found to be located between the two Asp rich motifs. The two motifs corresponded to the NSE/DTE and DDXXD motifs of Class I enzymes on the three-dimensional structures, but the diphosphate bound position was shifted approximately one or two phosphate size as compared to that for Class I enzymes. The different binding positions may strongly affect the catalysis, including the chemical structure of the product. Biochemical characters of Class IB enzymes are basically similar to those of Class I enzymes except for the capability to convert the large (C₂₀-C₃₅) prenyl-diphosphate. We also identified a novel cyclic diterpene formed by the Class IB enzyme. Mutational analysis revealed that five of the six Asp residues in the Asp rich motifs and two His residues play pivotal roles to form the cyclic skeleton. Because there are more uncharacterized Class IB enzymes, further genome-mining and modification of this family promise the finding of more novel terpenoids with valuable functions, especially \geq C₂₅ rare terpenes, which cannot be formed by conventional Class I enzymes.

Methods

All methods used in this study are available in Supporting Information.

Associated content

The Supporting Information is available free of charge on the ACS Publications website at DOI:???: general procedure and materials, expression and purification of BalTS, enzymatic assay, isolation of compound **12**, chemical structure analyses of compound **12**, crystallization and crystallographic analysis, site-directed mutagenesis, hypothetical production mechanism of BalTS and BsuTS products

AUTHOR INFORMATION

Corresponding Author

*E-mail: mfuji@kuchem.kyoto-u.ac.jp, satot@agr.niigata-u.ac.jp

ORCID

Daijiro Ueda: 0000-0002-9870-7690, Tetsuro Shinada: 0000-0001-9145-1533, Kunio Miki:0000-0002-8381-0871, Masahiro Fujihashi: 0000-0002-6882-9697, Tsutomu Sato: 0000-0002-1130-9025

Author Contributions

M.F. and T.Sa. designed the whole project, analyzed the results. R.S., T.N. and D.U. analyzed the new product. K.S., R.S., K.A. and D.U. performed enzymatic characterization. K.A., R.S., T.N. and D.U. conducted mutational analysis. H.I. and M.F. determined the crystal structure of BalTS. Y.Y. and T. Sh. synthesized the compound **10**. T. Sh., K.M., M. F. and T. Sa. administrated and supervised the project. The manuscript was written by M.F., R.S. and T.Sa, edited by T.Sh. and K.M., and approved by all authors.

Notes

The authors declare no competing financial interest.

Conflicts of interest

There are no conflicts to declare.

Acknowledgments

The authors are grateful to the staff members at the beamlines of the Photon Factory and SPring-8 for their help with data collection. This work was supported in part by JSPS KAKENHI (Grant# 17H05439 and 19H04652 to MF, 18H02145, 19K22273, and 19H04648 to T. Sa, and 19H04661 to T. Sh) as well as, by Life Sciences fellowships from the Takeda Science Foundation (to MF). The use of beamlines at the Photon Factory and SPring-8 was approved by the Photon Factory Advisory Committee (2017G696 and 2019B530) and by the Japan Synchrotron Radiation Research Institute (JASRI) (2017B2718, 2017B2711, and 2019B2716). Usage of the beamlines was also supported by the Platform Project for Supporting in Drug Discovery and Life Science Research from Japan Agency for Medical Research and Development (AMED) (BINDS1447, 2018RP-27, 2019RP-31).

References

- (1) Buckingham, L. (2002) *Dictionary of Natural Products, online edition*, Chapman & Hall, London.
- (2) Christianson, D. W. (2007) Chemistry. Roots of biosynthetic diversity, *Science* 316, 60-61.
- (3) Baunach, M., Franke, J., and Hertweck, C. (2015) Terpenoid Biosynthesis Off the Beaten Track: Unconventional Cyclases and Their Impact on Biomimetic Synthesis, *Angew. Chem., Int. Edit.* 54, 2604-2626.
- (4) Christianson, D. W. (2006) Structural biology and chemistry of the terpenoid cyclases, *Chem. Rev.* 106, 3412-3442.
- (5) Gao, Y., Honzatko, R. B., and Peters, R. J. (2012) Terpenoid synthase structures: a so far incomplete view of complex catalysis, *Nat. Prod. Rep.* 29, 1153-1175.
- (6) Rudolf, J. D., and Chang, C.-Y. (2020) Terpene synthases in disguise: enzymology, structure, and opportunities of non-canonical terpene synthases, *Nat. Prod. Rep.* 37, 425-463.
- (7) Baer, P., Rabe, P., Fischer, K., Citron, C. A., Klapschinski, T. A., Groll, M., and Dickschat, J. S. (2014) Induced-fit mechanism in class I terpene cyclases, *Angew. Chem., Int. Edit.* 53, 7652-7656.
- (8) Sato, T., Yoshida, S., Hoshino, H., Tanno, M., Nakajima, M., and Hoshino, T. (2011) Sesquiterpenes (C₃₅ terpenes) biosynthesized via the cyclization of a linear C₃₅ isoprenoid by a tetraprenyl-beta-curcumene synthase and a tetraprenyl-beta-curcumene cyclase: identification of a new terpene cyclase, *J. Am. Chem. Soc.* 133, 9734-9737.
- (9) Sato, T., Yamaga, H., Kashima, S., Murata, Y., Shinada, T., Nakano, C., and Hoshino, T. (2013) Identification of novel sesterterpene/triterpene synthase from *Bacillus clausii*, *Chembiochem* 14, 822-825.

- (10) Ueda, D., Yamaga, H., Murakami, M., Totsuka, Y., Shinada, T., and Sato, T. (2015) Biosynthesis of Sesterterpenes, Head-to-Tail Triterpenes, and Sesquiterpenes in *Bacillus clausii*: Identification of Multifunctional Enzymes and Analysis of Isoprenoid Metabolites, *Chembiochem* 16, 1371-1377.
- (11) Fujihashi, M., Sato, T., Tanaka, Y., Yamamoto, D., Nishi, T., Ueda, D., Murakami, M., Yasuno, Y., Sekihara, A., Fuku, K., Shinada, T., and Miki, K. (2018) Crystal structure and functional analysis of large-terpene synthases belonging to a newly found subclass, *Chem. Sci.* 9, 3754-3758.
- (12) Kabsch, W. (2010) Xds, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 66, 125-132.
- (13) Vagin, A., and Teplyakov, A. (1997) MOLREP: an Automated Program for Molecular Replacement, *J. Appl. Crystallogr.* 30, 1022-1025.
- (14) Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 66, 486-501.
- (15) Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 53, 240-255.
- (16) Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* 66, 12-21.
- (17) Liebschner, D., Afonine, P. V., Moriarty, N. W., Poon, B. K., Sobolev, O. V., Terwilliger, T. C., and Adams, P. D. (2017) Polder maps: improving OMIT maps by excluding bulk solvent, *Acta Crystallogr. D Struct. Biol. Acta Crystallogr., Sect. D: Struct. Biol.* 73, 148-157.
- (18) Koksai, M., Jin, Y., Coates, R. M., Croteau, R., and Christianson, D. W. (2011) Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis, *Nature* 469, 116-120.
- (19) Tomita, T., Kim, S. Y., Teramoto, K., Meguro, A., Ozaki, T., Yoshida, A., Motoyoshi, Y., Mori, N., Ishigami, K., Watanabe, H., Nishiyama, M., and Kuzuyama, T. (2017) Structural Insights into the CotB2-Catalyzed Cyclization of Geranylgeranyl Diphosphate to the Diterpene Cyclooctat-9-en-7-ol, *ACS Chem. Biol.* 12, 1621-1628.
- (20) Chen, M., Al-lami, N., Janvier, M., D'Antonio, E. L., Faraldos, J. A., Cane, D. E., Allemann, R. K., and Christianson, D. W. (2013) Mechanistic insights from the binding of substrate and carbocation intermediate analogues to aristolochene synthase, *Biochemistry* 52, 5441-5453.
- (21) Koo, H. J., Vickery, C. R., Xu, Y., Louie, G. V., O'Maille, P. E., Bowman, M., Nartey, C. M., Burkart, M. D., and Noel, J. P. (2016) Biosynthetic potential of sesquiterpene synthases: product profiles of Egyptian Henbane premnaspirodiene synthase and related mutants, *J. Antibiot.* 69, 524-533.
- (22) Whittington, D. A., Wise, M. L., Urbansky, M., Coates, R. M., Croteau, R. B., and Christianson, D. W. (2002) Bornyl diphosphate synthase: structure and strategy for carbocation manipulation by a terpenoid cyclase, *Proc. Natl. Acad. Sci. U. S. A.* 99, 15375-15380.
- (23) Karupiah, V., Ranaghan, K. E., Leferink, N. G. H., Johannissen, L. O., Shanmugam, M., Ni Cheallaigh, A., Bennett, N. J., Kearsley, L. J., Takano, E., Gardiner, J. M., van der Kamp, M. W., Hay, S., Mulholland,

- A. J., Leys, D., and Scrutton, N. S. (2017) Structural Basis of Catalysis in the Bacterial Monoterpene Synthases Linalool Synthase and 1,8-Cineole Synthase, *ACS Catal.* 7, 6268-6282.
- (24) Koksai, M., Zimmer, I., Schnitzler, J. P., and Christianson, D. W. (2010) Structure of isoprene synthase illuminates the chemical mechanism of teragram atmospheric carbon emission, *J. Mol. Biol.* 402, 363-373.
- (25) Vedder, A. (1934) *Bacillus alcalophilus* n. sp., *Antonie van Leeuwenhoek* 1, 141-147.
- (26) Green, S., Friel, E. N., Matich, A., Beuning, L. L., Cooney, J. M., Rowan, D. D., and MacRae, E. (2007) Unusual features of a recombinant apple alpha-farnesene synthase, *Phytochemistry* 68, 176-188.
- (27) Meguro, A., Tomita, T., Nishiyama, M., and Kuzuyama, T. (2013) Identification and characterization of bacterial diterpene cyclases that synthesize the cembrane skeleton, *Chembiochem* 14, 316-321.
- (28) Hoshino, T., Nakano, C., Ootsuka, T., Shinohara, Y., and Hara, T. (2011) Substrate specificity of Rv3378c, an enzyme from *Mycobacterium tuberculosis*, and the inhibitory activity of the bicyclic diterpenoids against macrophage phagocytosis, *Org. Biomol. Chem.* 9, 2156-2165.
- (29) Wallach, O. (1905) Zur Kenntniss der Terpene und der ätherischen Oele, *Justus Liebigs Ann. Chem.* 340, 1-16.
- (30) Connell, D. W., and Sutherla.Md. (1966) Terpenoid Chemistry .11. (-)-Beta-Sesquiphellandrene, *Aust. J. Chem.* 19, 283-&.
- (31) Lopez-Gallego, F., Agger, S. A., Abate-Pella, D., Distefano, M. D., and Schmidt-Dannert, C. (2010) Sesquiterpene synthases Cop4 and Cop6 from *Coprinus cinereus*: catalytic promiscuity and cyclization of farnesyl pyrophosphate geometric isomers, *Chembiochem* 11, 1093-1106.
- (32) Lafever, R. E., and Croteau, R. (1993) Hydride Shifts in the Biosynthesis of the p-Menthane Monoterpenes α -Terpinene, γ -Terpinene, and β -Phellandrene, *Arch. Biochem. Biophys.* 301, 361-366.