

Research Article

Crystal structure of IspF from *Bacillus subtilis* and absence of protein complex assembly amongst IspD/IspE/IspF enzymes in the MEP pathway

Zhongchuan Liu^{1,2}, Yun Jin^{1,2,4}, Weifeng Liu³, Yong Tao³ and Ganggang Wang^{1,2}

¹Key Laboratory of Environmental and Applied Microbiology, Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu 610041, China; ²Key Laboratory of Environmental Microbiology of Sichuan Province, Chengdu 610041, China; ³Chinese Academy of Sciences Key Laboratory of Microbial Physiological and Metabolic Engineering, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China; ⁴University of Chinese Academy of Sciences, Beijing 100049, China

Correspondence: Ganggang Wang (wanggg@cib.ac.cn)



2-C-Methyl-D-erythritol 2,4-cyclodiphosphate synthase (IspF) is a key enzyme in the 2-C-Methyl-D-erythritol-4-phosphate (MEP) pathway of isoprenoid biosynthesis. This enzyme catalyzes the 4-diphosphocytidyl-2-C-methyl-D-erythritol 2-phosphate (CDPME2P) to 2-C-methyl-D-erythritol 2,4-cyclodiphosphate (MEcDP) with concomitant release of cytidine 5'-diphosphate (CMP). *Bacillus subtilis* is a potential host cell for the production of isoprenoids, but few studies are performed on the key enzymes of MEP pathway in *B. subtilis*. In this work, the high-resolution crystal structures of IspF in native and complex with CMP from *B. subtilis* have been determined. Structural comparisons indicate that there is a looser packing of the subunits of IspF in *B. subtilis*, whereas the solvent accessible surface of its active pockets is smaller than that in *Escherichia coli*. Meanwhile, the protein-protein associations of 2-C-Methyl-D-erythritol-4-phosphatecytidyltransferase (IspD), CDPME kinase (IspE) and IspF from *B. subtilis* and *E. coli*, which catalyze three consecutive steps in the MEP pathway, are analyzed by native gel shift and size exclusion chromatography methods. The data here show that protein complex assembly is not detectable. These results will be useful for isoprenoid biosynthesis by metabolic engineering.

Introduction

Isoprenoid compounds are the most diverse group of natural products with a broad range of biological functions [1]. Two five-carbon compounds, isopentenylidiphosphate (IPP) and dimethylallyldiphosphate (DMAPP), are the universal precursors for isoprenoid molecules. IPP and DMAPP are synthesized by two different biosynthetic pathways, the mevalonate (MVA) pathway and the 2-C-Methyl-D-erythritol-4-phosphate (MEP) pathway. The MVA pathway is dominant in the mammals, plants, and fungi [2,3], whereas the MEP pathway is discovered in many bacteria, algae, cyanobacteria, eubacteria, apicomplexan parasites, and plant chloroplasts [4-7]. Since the MEP enzymes are both highly conserved and absent from humans, all enzymes in MEP pathway have gained interests as potential targets for antimicrobial drug development [8-11]. In addition to the therapeutic value of the enzymes in MEP pathway, the isoprenoid production through MEP pathway has also been widely exploited, because it has several potential advantages compared with the MVA pathway, including a theoretically better stoichiometric yield and lower oxygen consumption during fermentation [12].

The MEP pathway is made of seven consecutive catalytic steps. Amongst them, the enzymes which are encoded by *ispD*, *ispE*, and *ispF* genes, catalyze the third, fourth, and fifth steps to convert MEP into 2-C-methyl-D-erythritol 2,4-cyclodiphosphate (MEcDP). 2-C-Methyl-D-erythritol-4-phosphatecytidyltransferase (IspD) condenses MEP with CTP to give

Received: 12 October 2017
Revised: 11 January 2018
Accepted: 13 January 2018

Accepted Manuscript Online:
15 January 2018
Version of Record published:
21 February 2018

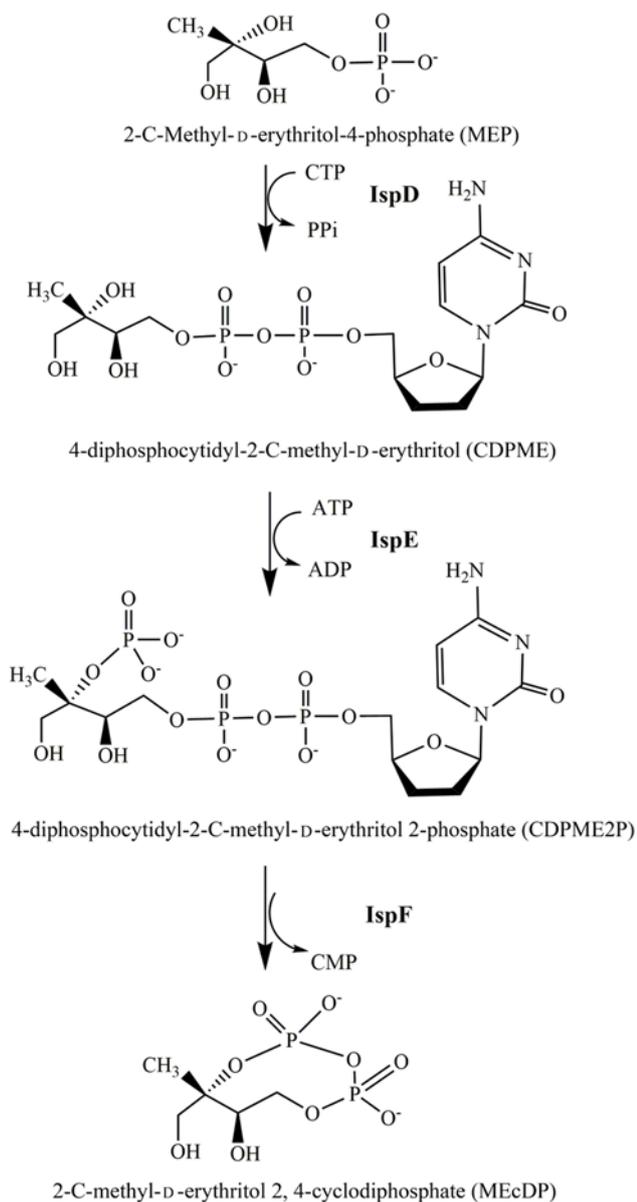


Figure 1. The reaction catalyzed by IspD, IspE, and IspF enzymes in the MEP pathway

4-diphosphocytidyl-2-C-methyl-D-erythritol (CDPME) and PPi [13]. CDPME kinase (IspE) catalyzes the phosphorylation of CDPME to 4-diphosphocytidyl-2-C-methyl-D-erythritol 2-phosphate (CDPME2P) [14]. Finally, the CDPME2P is converted into MEcDP with concomitant release of CMP by MEcDP synthase (IspF) [15] (Figure 1).

Generally, the three genes, *ispD*, *ispE*, and *ispF* encode monofunctional enzymes in most MEP-dependent organisms. However, bioinformatics analysis indicated that *ispD* and *ispF* genes were fused together in several bacterial species, such as *Campylobacter jejuni*, *Sinorhizobium meliloti*, *Brucella melitensis*, *Helicobacter pylori*, and *Mesorhizobium loti* etc. [16-18]. Biochemical analysis confirmed that the IspDF protein from *C. jejuni* was a bifunctional enzyme catalyzing the third and fifth step reactions in MEP pathway [17]. Furthermore, the crystal structure of IspDF from the *C. jejuni* revealed that the protein presented two distinct domains corresponding to the cytidyltransferase IspD and synthase IspF, respectively [17,19]. The structural arrangement of IspDF was uncommon, because it catalyzed non-consecutive steps in the MEP pathway. Sedimentation velocity studies indicated that possible associations could be formed between bifunctional IspDF and monofunctional IspE from *C. jejuni* and *Agrobacterium tumefaciens* [19]. In addition, by the analytical ultracentrifugation analysis, a complex of ~430 kDa was observed

when the individual enzymes IspD, IspE, and IspF from *E. coli* were mixed. This likely corresponded to an assembly of three IspD dimers, three IspE dimers, and two IspF trimers, which could provide 18 catalytic centers and might lead to enhanced metabolic flux in the MEP pathway [19]. However, further research rejected this hypothesis and indicated that there was no evidence for enhancing catalytic rates on the mixture of fused IspDF and individual IspE enzymes from *A. tumefaciens* or any proof of substrate channeling in the mixture of individual IspD, IspE, and IspF proteins [20]. It is controversial whether a multienzyme complex can be formed between IspDF and IspE or amongst the monofunctional enzymes IspD, IspE, and IspF. Further investigations are required to clarify the associations amongst IspD, IspE, and IspF.

Nowadays, the increasing needs for isoprene products prompt the efforts to develop a bio-based process for isoprene production by engineered microorganisms, since it is efficient and environment friendly [21]. Recent studies showed that overexpression of both MEP and MVA pathways in *E. coli* could improve the isoprene yield ~20-fold and 3-fold, respectively, compared with overexpression of the each pathway alone, which demonstrated that the synergy of the MEP pathway and the MVA pathway could be used to improve the production of isoprene in microorganisms [22–24]. *Bacillus subtilis* is a model organism and of significant industrial importance. It was reported that the *B. subtilis* has potential as an excellent isoprene producer [25–27]. However, only few studies on the key enzymes of MEP in *B. subtilis* had been performed [27–29]. In this study, the high resolution structures of IspF and IspF/CMP complex from *B. subtilis* have been determined. Moreover, the interaction of individual enzymes IspD, IspE, and IspF from *B. subtilis* and *E. coli* are investigated. These results will show useful hints on isoprenoid biosynthesis by metabolic engineering.

Materials and methods

Materials

Restriction enzymes used in the present study were purchased from New England Biolabs. T4 ligase was purchased from TransGen Biotech. Column resins used for protein purification were purchased from GE Healthcare (U.S.A.). Crystallization screening kits were purchased from Hampton Research. All other chemicals used for preparing buffers and solutions were reagent grade and purchased from Merck, Sigma–Aldrich, and local suppliers.

Expression and purification of proteins

The *ispD* and *ispF* genes of *B. subtilis* 168 (DSM 23778, DSMZ, Germany) were amplified by PCR from genomic DNA with the 5'/3' specific primers, which introduced BamHI site and XhoI sites, respectively. The ~699 and ~477 bp amplification products were digested and cloned into pGEX-6P-1; the gene sequence was confirmed by DNA sequencing. The recombinant plasmids were transformed into competent *E. coli* DH5 α . Bacterial cells were grown in LB broth supplemented with 100 μ g/ml ampicillin. The culture was incubated at 37°C with vigorous shaking. At an optical density (600 nm) of 0.5–0.6, IPTG was added to a final concentration of 0.3 mM, and the culture was further incubated at 16°C for 14–16 h. The cells were harvested by centrifugation; the cell pellets were re-suspended in lysis buffer (25 mM Tris/HCl pH 8.0, 50 mM NaCl, 1 mM DTT) and sonicated on ice. The crude fluid was centrifuged at 11800 \times g at 4°C for 40 min to remove cellular debris. The proteins were purified from the supernatant by GST Glutathione Sepharose™ 4 Fast Flow column (GE Healthcare), and the GST tag was removed by Prescission Protease (PPase) at 4°C overnight. The eluted IspD and IspF enzymes were further purified by the combination of the Resource Q (GE Healthcare) anion-exchange column and Superdex 75 (GE Healthcare) size-exclusion column. The protein fractions were pooled and determined to be >95% pure by SDS/PAGE. The protein concentration was determined by Bradford method using BSA as the standard. The purified protein was aliquoted for storage at –80°C.

The *ispE* gene of *B. subtilis* 168 (DSM 23778, DSMZ, Germany) was also cloned into the vector of pGEX-6P-1 expression vector. The protein purification process was similar to IspF of *B. subtilis* above. However, the recombinant proteins without GST tag were insoluble; the fusion proteins were prepared by GSH elution from affinity column directly and purified by subsequent purification steps. If no special instructions, the fusion proteins were named as BsIspE.

To exclude the influence of the tag in the association studies, we also purified the recombinant proteins IspD, IspE, and IspF from *E. coli*. The *ispD*, *ispE*, and *ispF* genes of *E. coli* were PCR amplified from *E. coli* K12 and inserted into the pET-22b expression vector. C-terminal His₆-tagged proteins were expressed in *E. coli* BL21 (DE3) cells. Tagged proteins were purified using an Ni²⁺-NTA (GE Healthcare) column. The histidine tag was removed by thrombin enzyme. The eluted proteins were further purified by the combination of the Resource Q (GE Healthcare) anion-exchange column and Superdex 75 (GE Healthcare) size-exclusion column. Finally, the purified proteins were essentially homogeneous and >95% pure, as analyzed by SDS/PAGE.

Table 1 Data collection and refinement statistics

	Native	CMP complex
Data collection		
Wavelength (Å)	0.97915	0.97915
Space group	<i>P</i> 2 ₁	<i>P</i> 2 ₁
Unit-cell parameters		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	56.96, 89.30, 84.71	57.06, 89.46, 87.99
β (°)	100.85	103.80
Monomers per asymmetric unit	6	6
Resolution (Å)	29.05–1.99 (2.06–1.99) ¹	29.21–1.99 (2.06–1.99) ¹
Number of unique reflections	56504 (5662) ¹	57544 (5799) ¹
CC _{1/2}	0.994 (0.877)	0.995 (0.775)
Redundancy	3.7 (3.7) ¹	3.7 (3.8) ¹
Completeness (%)	99.3 (99.9) ¹	98.2 (99.8) ¹
Mean <i>I</i> /σ(<i>I</i>)	15.0 (3.7) ¹	14.3 (4.8) ¹
<i>R</i> _{merge} (%)	8.2 (41.5) ¹	8.4 (37.4) ¹
Wilson B factor (Å ²)	27.6	26.8
Refinement		
Reflections (working/test)	53728/2755	54704/2819
<i>R</i> _{work} / <i>R</i> _{free} (%)	18.1/23.1	17.4/21.9
<i>Number of residues</i>		
Protein	924	932
CMP/Mg ²⁺	0/6	4/6
Waters	250	256
<i>Average B factor (Å²)</i>		
Main chain/side chain	35.2/42.7	32.4/40.6
CMP/Mg ²⁺	–/32.4	41.6/21.9
Waters	39.7	40.8
<i>Ramachandran plot (%)</i>		
Favored	97.2	97.7
Allowed	2.8	2.3
Clash score	1.76	1.66
Overall score	1.26	0.98
<i>r.m.s.d.</i>		
Bond lengths (Å)	0.013	0.013
Bond angles (°)	1.570	1.652

¹The values in parenthesis means those for the highest resolution shell. Abbreviation: r.m.s.d., root mean square deviation.

Protein crystallization and data collection

Crystals of the native IspF protein of *B. subtilis* (*BsIspF*) proteins were obtained at 18°C by the hanging drop vapor-diffusion method, 1 μl of 10 mg/ml protein in the buffer of 25 mM Tris/HCl pH 8.0, 0.1 M NaCl was added in the reservoir solution containing 0.1 M magnesium formate dehydrate and 12–15% w/v PEG 3350. Crystals of *BsIspF*/CMP complex were incubated with 5 mM CMP on the ice for 2 h before crystallization and obtained as described above. Crystals were grown over a period of 5–7 days and then dehydrated to the reservoir solution plus 15% of glycerol overnight. Then the crystals were picked from the solution using nylon loops and frozen in liquid nitrogen. X-ray diffraction data of *BsIspF* crystals were collected at 100 K using synchrotron radiation at Shanghai Synchrotron Radiation Facilities (SSRF) [30]. Datasets were processed and scaled by HKL2000 [31]. Details are shown in Table 1.

Structure determination and refinement

The structures of *BsIspF* were elucidated by molecular replacement using PHASER [32] from the CCP4 program suite [33]. The starting model was the monomer IspF enzyme from *Shewanella oneidensis* (PDB: 1T0A), which shared 62% sequence identity with the *BsIspF* and served as a good model for the structure solution. Only one solution was evident. Refinement was performed using the maximum likelihood functions implemented in REFMAC5 [34], while model building and improvement were achieved with COOT [35]. Solvent molecules, metal ions and ligands were positioned after a few cycles of refinement. Non-crystallographic symmetry (NCS) restraints were imposed in the early stages of the analyses and then released as the refinements progressed. Isotropic refinement of the atomic displacement parameters was performed for all atoms. The stereochemistry was checked with the program MolProbity [36]. Details of the overall refinement and final quality of the models are shown in Table 1. The program *PyMOL* (<http://www.pymol.sourceforge.net/>) was used to prepare structural figures.

Native gel shift

According to different shift distance of different components, the native gel shift could show whether a stable complex could be formed by different individual proteins. The control groups were only one component protein and corresponding buffer. The experimental groups were carried out using two or three proteins amongst IspD, IspE, and IspF (the concentration of these proteins were 2.5 $\mu\text{g}/\mu\text{l}$, respectively), 10% glycerol, 25 mM Tris/HCl pH 8.0, 100 mM NaCl in final volume 12 μl ; meanwhile, adding 1 mM MgCl_2 , 1 mM MEP, and 1 mM CTP if the IspD was present in the reaction system. The reaction was incubated at 30°C for 30 min. Finally, the samples were analyzed by a 8% native page in 1 \times TBE buffer for 2.5 h at 200 V at 4°C. The gel was photographed by an imaging system.

Size exclusion chromatography

Size exclusion chromatography is also applied to investigate whether a stable complex could be formed by different individual proteins. The interactions amongst the IspD, IspE, and IspF were studied using an AKTA purifier and a 24 ml (1 \times 30 cm) Superdex75 10/300 GL or Superdex200 10/300 GL size exclusion chromatography column. Typically, size exclusion chromatography was carried out at buffer 100 mM NaCl, 25 mM Tris/HCl, pH 8.0. The analytical standards were BSA (67 kDa), ovalbumin (43 kDa), ribonuclease (13.7 kDa), Aprotinin (6.5 kDa), vitamin B₁₂ (1.3 kDa) of Superdex75 and Thyroglobulin (669 kDa), Ferritin (440 kDa), BSA (67 kDa), β -lactoglobulin (3 kDa), ribonuclease (13.7 kDa), Aprotinin (6.5 kDa), vitamin B₁₂ (1.3 kDa) of Superdex200. The control groups were only one-component protein in elution buffer (25 mM Tris/HCl pH 8.0, 100 mM NaCl). The experimental groups were carried out at three proteins of IspD, IspE, and IspF (0.2 mg), in a final volume of 250 μl ; meanwhile, adding the 1 mM MgCl_2 , 1 mM MEP, and 1 mM CTP if the IspD was present in the reaction system. The reaction was incubated at 30°C for 30 min.

Other methods

The cavity analysis was performed using CASTp server [37], surface calculations were performed by ArealMol from CCP4 programs suite [33].

Accession numbers

The atomic co-ordinates and structure factors have been deposited in the Protein Data Bank with accession codes 5IWX (native) and 5IWY (CMP complex), respectively.

Results

Overall structure of *BsIspF*

The *BsIspF* crystallizes in space group $P2_1$, with unit-cell parameters $a = 56.96$, $b = 89.30$, $c = 84.71$ Å, $\beta = 100.85^\circ$. The crystallographic asymmetric unit contains six monomers, which are arranged as two trimers of closely associated subunits. The monomer displays an α/β fold constructed with a four-strand β -sheet of β_1 , β_2 , β_3 , and β_4 , with β_2 antiparallel to the others on one side, and a three-helix bundle of α_1 to α_3 on the other. In *BsIspF* trimer, the four-strand β -sheet of each subunit lay roughly parallel to the threefold NCS axis, forming a compact bell-shaped trimer of approximately 45 \times 60 Å in the axial and equatorial dimensions, respectively (Figure 2). Overall, the two trimers are extremely similar with a root mean square deviation (r.m.s.d) value of 0.20 Å over all C α atoms. The largest C α r.m.s.d differences (~ 1.5 Å) between the two trimers occur in two loop regions, the first one connects β_1 to α_1 with the residues 34–39, and the second one links α_1 to α_2 , which comprises residues 62–72. Both loops are parts of the active pocket and involved in the binding of 2-C-methyl-D-erythritol (MEPP) moiety.

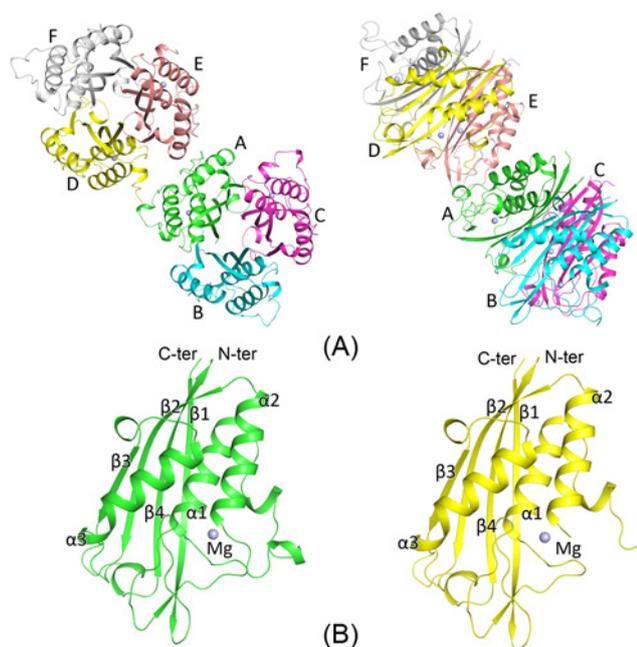


Figure 2. Overall structure of the BsIspF trimer and monomer

(A) Molecule packing in the asymmetric unit, with six monomers arranged into two trimers. (B) The monomer structure. Secondary structures and the magnesium ions are labeled.

The BsIspF shares a sequence identity of $\sim 62\%$ with IspF from *E. coli* (*EcIspF*), and overlay between the BsIspF and *EcIspF* (PDB: 1KNK) trimer give an r.m.s.d value of 0.86 \AA , indicating a high structural similarity. The superimposition of structures of BsIspF and *EcIspF* are shown in Supplementary Figure S1. However, although the accessible surface area (ASA) of a BsIspF monomer (7584 \AA^2) is nearly the same as ASA of *EcIspF* monomer (7681 \AA^2), the ASA of the BsIspF trimer (17966 \AA^2), is larger than that of *EcIspF* trimer (16293 \AA^2), which demonstrates a looser packing of the subunits of BsIspF. Compared with the structure of *EcIspF*, there is also a hydrophobic cavity along the three-fold symmetry axis of BsIspF homotrimer (Figure 4A). Several aliphatic amino acids from three subunits line inside of the cavity (Phe⁸, Ile¹⁰⁰, and Phe¹⁴⁰). The bottom of the cavity is closed by the hydrogen bonds involving six glutamines (Gln¹⁵⁰ and Gln⁶), and the cap of the cavity is narrowed to a 5-\AA diameter hole by the residues Arg¹⁴³ and Glu¹⁴⁵ from three subunits. Sequence alignment reveals that all the residues mentioned above are highly conserved between *EcIspF* and BsIspF enzymes, except the residues Gln¹⁵⁰ and Gln⁶ in BsIspF (corresponding residues Glu¹⁴⁹ and His⁵ in *EcIspF*) (Figure 3B). The volume of the cavity of BsIspF (1024 \AA^3) is larger than that of *EcIspF* (718 \AA^3), which is consistent with the looser packing of the subunits in BsIspF. In other reported ispF structures, such as *E. coli*, *S. oneidensis* and *C. jejuni* etc., electron density had been found inside the cavities, which was interpreted as isoprenoid species such as geranyl- or farnesyl-phosphosphate, with a possible feedback-regulatory role through a yet undetermined mechanism, which might be related to the trimer stability [15,19,38-42]. However, not all ispF enzymes have isoprenoid species binding in the cavity. For the structure of IspF from *Haemophilus influenzae* [43], the density inside the cavity was interpreted as a sulphate ion, for the structure of IspF from *Burkholderia cenocepacia* [44], a molecule of di(hydroxyethyl)ether, a likely decomposition product or impurity of PEG3350 used in crystallization, occupies the cavity, and for the structure of IspF from *Plasmodium falciparum*, only a hollow cavity is observed. In addition, such a cavity is absent from the structure of IspF from *Thermus thermophilus* [45], where the packing of subunits is tighter and likely more stable according to ultracentrifugation data. Besides, in the IspF structure from *Arabidopsis thaliana*, the cavity is unsuited for binding a diphosphate moiety, which suggests a different regulatory mechanism of IspF enzymes between bacteria and plants [46]. In BsIspF, no ligand binding in the cavity is observed because no obviously continuous electron density is found here, and only a hollow cavity is presumed in both native and BsIspF/CMP complex, but regarding the highly conserved residues involved in the cavity formation between *B. subtilis* and *E. coli* and the size or dimension of the cavity in BsIspF, potential ligands of isoprenoid species could bind here. Till now, no accurate reports are proposed to elucidate the mechanism of the possible feedback-regulatory

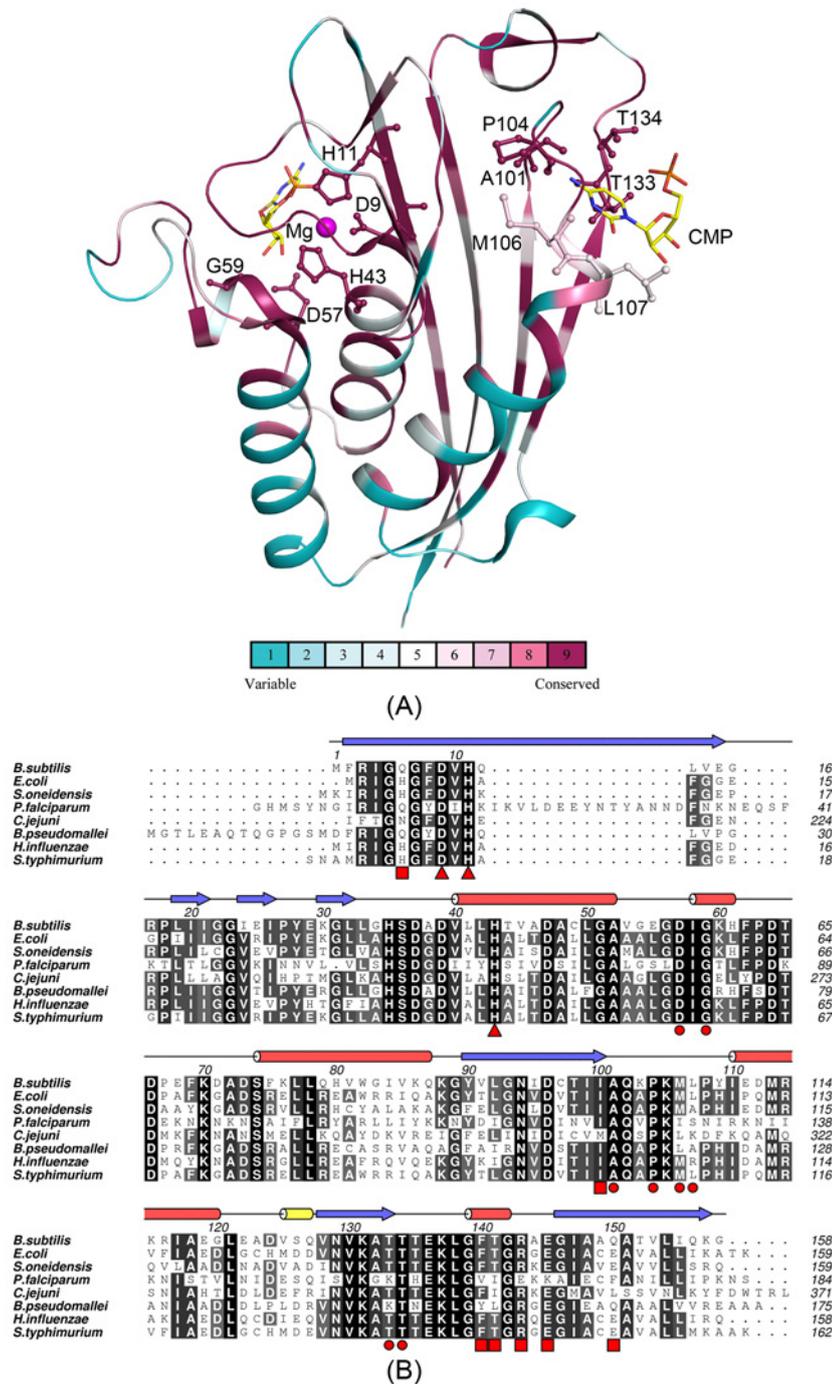


Figure 3. Sequence conservation of BslSpF structure

(A) Sequence-conservation scores of BslSpF monomer [50]. Residues that contribute to the interactions with the CMP are shown as sticks; the most variable residues are shown in blue and the most conserved residues are shown in dark purple. (B) Multiple sequence alignment IspF protein. Sequences of IspF from *B. subtilis*, *E. coli*, *S. oneidensis*, *P. falciparum*, *C. jejuni*, *B. pseudomallei*, *H. influenzae*, *S. typhimurium* (UniProt identifiers: Q06756, P62617, Q8EBR3, P62368, Q9PM68, Q3JRA0, P44815, Q8ZMF7, respectively) were aligned with MUSCLE [51] and edited by hand to match the structural similarity where appropriate by using ALINE [52]. Identical and similar residues were highlighted in black and gray, respectively. The secondary structure elements base on the BslSpF monomer, α -helix and β -strand are marked by red pillar and blue arrow, respectively. Three residues that co-ordinate the magnesium ions are marked with red triangles. Residues that bind CMP are marked with red circles. Residues that form the hydrophobic cavity are highlighted with red squares.

role by the hydrophobic cavity in the IspF enzymes, further investigations are required to clarify and that is what we will do.

Active pockets

BsIspF trimer have three active pockets which are formed by the interface between two adjacent subunits, and here the magnesium ions deriving from the magnesium formate dehydrate used in crystallization are co-ordinated by the side chains of two histidines and an aspartate (Asp⁹, His¹¹, and His⁴³). Generally, the IspF enzymes require the divalent cation to position the substrate and stabilize the charge in the transition state during the catalytic reaction. In other structures of IspF proteins, the manganese or zinc ions were commonly bonded in the active site [15,38], meanwhile the magnesium ions were also substituted in the IspF structure from *T. thermophilus* [45]. In both native and CMP bonded structures of *BsIspF*, magnesium ions were refined in the active site, with the ingredient of magnesium ions used in the crystallization cocktail. Moreover, in our structures, the B-factors of magnesium ions were refined well in both native and complex structures. Structure-based sequence alignment reveals that residues around the active pocket are well conserved in all the IspF enzymes (Figure 3). And, the overlay of each chain between the native and *BsIspF*/CMP complex reveal r.m.s.d values ranging from 0.12 to 0.24 Å. The active sites comprise a nucleotide binding pocket, a cation binding site, and two loops for the binding of the MEPP moiety. The nucleotide binding pocket and cation binding site are relatively rigid, with a main-chain B-factor of 37.1 Å², which is basically consistent with the average B-factor of the main-chain protein atoms, 35.2 Å². Moreover, structural comparisons show that only a small Cα r.m.s.d differences (~0.1 Å) occur in this region between *BsIspF* and *BsIspF*/CMP complex. However, the two loops binding for the MEPP moiety are rather flexible. The first loop (residues: 34–39) displays elevated main-chain B-factors of 45.1 Å² compared with 42.2 Å² in *BsIspF* and *BsIspF*/CMP complex, respectively. The electron density of the second loop (residues: 62–72) in *BsIspF* and *BsIspF*/CMP complex is of insufficient quality to allow this region to be modeled, indicating a high degree of flexibility. CMP acts as the leaving group to give the cyclic diphosphate reaction product MEcDP. In the *BsIspF*/CMP complex, only four CMP moieties can be modeled in the six active pockets of the two trimers in the asymmetric unit with the occupancy higher than 0.7, one trimer fully loads CMP substrates, and the other captures only one. Although the two trimers in the asymmetric unit display extremely structural similarity (r.m.s.d: 0.20 Å), the B-factors of the active pocket in the two trimers are quite different, which may be the possible cause for this asymmetric binding (Supplementary Figure S2). The CMP moieties reveal similar interactions with the protein. The cytidine group of CMP interacts with the residues Ala¹⁰¹, Pro¹⁰⁴, Met¹⁰⁶, and Leu¹⁰⁷ through main-chain atoms, and the ribose hydroxyl groups form hydrogen bonds with the carboxylate of Asp⁵⁷ and the amide of Gly⁵⁹ from the adjacent subunits. Residues Thr¹³³ and Thr¹³⁴ hydrogen bond to the α-phosphate of CMP, waters act as a bridge between the α-phosphate and the magnesium ion (Figure 4B). All these residues are highly conserved in the IspF protein (Figure 3).

We analyzed the ASA of active pockets in the native IspF structures which had been deposited in the PDB bank. The IspF enzymes from pathogens basically display a small ASA of its active pockets. The IspF from *Salmonella typhimurium* (PDB: 3GHZ) and *H. influenzae* [43] revealed minimum-size active pocket with an ASA of ~610 Å², whereas the IspF from *Plasmodium falciparum* [44], *Burkholderia pseudomallei* [47], and bifunctional IspDF from *C. jejuni* [19] show a medium-size active pocket with an ASA ranging from ~900 to 950 Å², the rest IspF enzymes show a large-size active pocket, like IspF from *S. oneidensis* [39], *E. coli* [38], *B. subtilis* etc. The diversity of the primary sequence in difference organisms may lead to the variation on the active pocket. Furthermore, we focussed our study on the two model organisms, *B. subtilis* and *E. coli*. The primary sequences of the active pockets are strictly conserved in the *BsIspF* and *EcIspF*, except the residue His⁶² in the *BsIspF* (corresponding residue Leu60 in *E. coli*, which have an van der Waals interaction between the methyl group on MEcDP in the *EcIspF*/MEcDP complex [38]). However, cavity analysis revealed that the ASA of active pocket in native *BsIspF* is approximately 1020 Å² compared with 1150 Å² in native *EcIspF* (PDB:1KNK), whereas the ASA of active pocket in *BsIspF*/CMP complex is approximately 820 Å² compared with 910 Å² in *EcIspF*/CMP complex (PDB:1H47) [38,41]. Although a looser packing of the subunits of *BsIspF*, the active pockets of *BsIspF* are tighter than that of *EcIspF* in native or in complex with CMP.

Protein associations

Protein associations amongst IspD, IspE, and IspF enzymes are investigated by native gel shift. In *B. subtilis*, no stable complexes of *BsIspD*/*BsIspE*, *BsIspD*/*BsIspF*, and *BsIspD*/*BsIspE*/*BsIspF* have been observed (Figure 5A). Otherwise, similar results are observed in *E. coli*, referring to *EcIspD*/*EcIspE*, *EcIspE*/*EcIspF*, and *EcIspD*/*EcIspE*/*EcIspF* (Figure 5A). However, the dual mixture of *BsIspE* and *BsIspF* seems to merge into a single band. Similar results

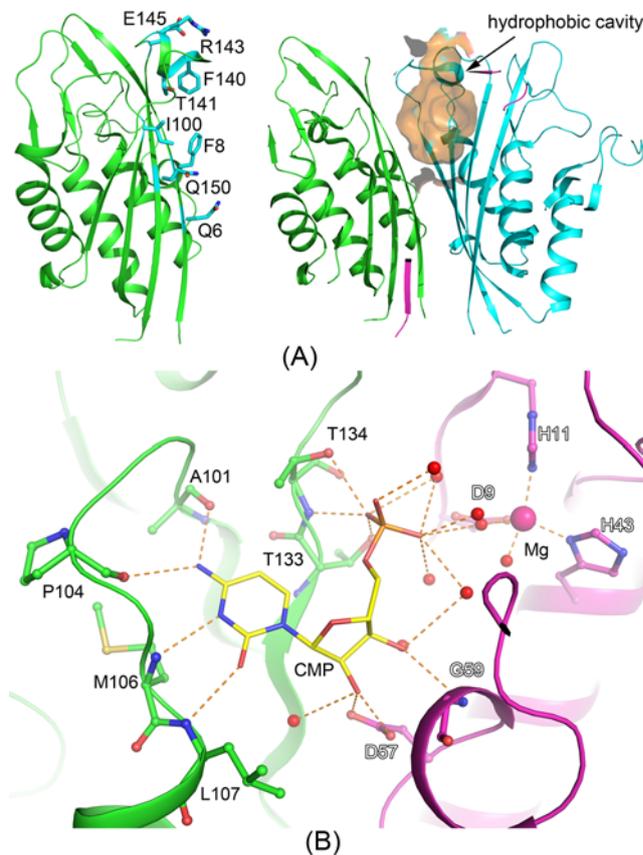


Figure 4. Representative of the hydrophobic cavity and the interactions between CMP and BsIspF

(A) Representative of the hydrophobic cavity in the *BsIspF* trimer. Cartoon representations of *BsIspF* subunits (left) and a section of the overall structure of *BsIspF* trimer (right). The residues involved in the hydrophobic cavity formation are shown as stick model and labeled. The hydrophobic cavity is shown in surface model and labeled. (B) Representative of the interactions between CMP and *BsIspF*. The CMP molecule is shown as stick model in yellow. The magnesium ion and the water molecules are shown as sphere. The relevant residues from only two subunits are shown as stick model, with C positions colored green and purple, respectively. Hydrogen and co-ordination bonds are shown as orange dotted lines.

are observed in the dual mixture of *EcIspD* and *EcIspF* (Figure 5A). Therefore, the size exclusion chromatography method is performed to verify whether a stable complex could be formed.

BsIspD, *BsIspE*, and *BsIspF* are dimer (51.6 kDa), monomer (57.7 kDa, a fusion protein of GST-*BsIspE*) and trimer (51.3 kDa), respectively. As shown in Figure 5B, *BsIspD*, *BsIspE*, and *BsIspF* are eluted at ~11.25, 9.11, and 11.58 ml by a Superdex75 column, respectively. However, the mixture of triple enzymes is eluted at two peaks, ~9.1 and 11.4 ml. Compared with the profile of each single protein, the peak1 (~9.1 ml) exhibits a monomeric *BsIspE* and peak2 (~11.4 ml) may represent mixture of a dimer *BsIspD*, a trimer *BsIspF*, or the complex of *BsIspD/BsIspE*. Therefore, no complex formation of *BsIspE/BsIspF* is detectable. Furthermore, the samples from peak2 are analyzed by native gel shift, but no *BsIspD/BsIspF* complex is observed. So in the *B. subtilis*, no dual complex or triple complex is formed between/amongst the *IspD*, *IspE*, and *IspF* *in vitro* (Figure 5B).

EcIspD, *EcIspE*, and *EcIspF* are dimer (51.4 kDa), monomer (30.9 kDa) and trimer (50.7 kDa), respectively. As shown in Figure 5C, *EcIspD*, *EcIspE*, and *EcIspF* are eluted at ~14.3, 15.5, and 14.9 ml, respectively. The mixture of triple enzymes can be resolved by a Superdex200 column into two singlet peaks, ~14.6 and 15.4 ml. Compared with the profile of the single protein, the peak1 (~14.6 ml) may represent mixture of a dimer *EcIspD*, a trimer *EcIspF* or the complex between them, the peak2 (~15.4 ml) reveals the monomer *EcIspE*. Samples from peak1 are re-analyzed by native gel shift, no *EcIspD/EcIspF* complex is observed. Therefore, in the *E. coli*, no dual complex or triple complex is formed between/amongst the *IspD*, *IspE*, and *IspF* *in vitro*, either (Figure 5C).

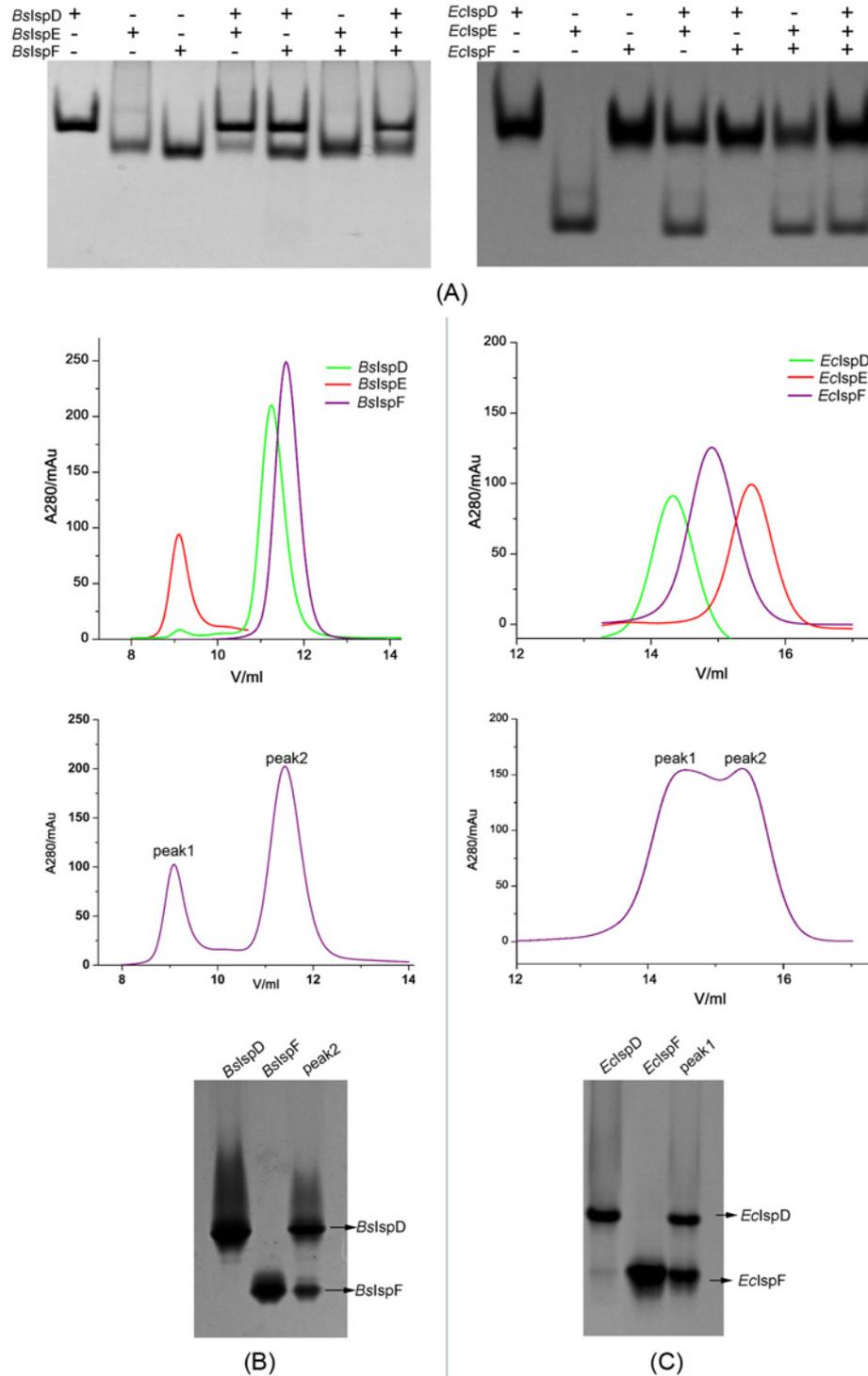


Figure 5. Analysis on the associations of IspD, IspE, and IspF

(A) Three individual proteins (IspD, IspE, and IspF), dual mixture (IspD/IspE, IspD/IspF, and IspE/IspF) and triple mixture (IspD/IspE/IspF) from *B. subtilis* (left) and *E. coli* (right) are analyzed by native gel shift, respectively. (B) Three individual proteins (green, red, and purple corresponding to BslspD, BslspE, and BslspF, respectively) (upper), and triple mixture (BslspD/BslspE/BslspF) (middle) are analyzed by size exclusion chromatography, and the samples from the peak2 are re-analyzed by native gel shift (lower). (C) Three individual proteins (green, red, and purple corresponding to EclspD, EclspE, and EclspF, respectively) (upper) and triple mixture (EclspD/EclspE/EclspF) (middle) are analyzed by size exclusion chromatography, the samples from the peak1 are re-analyzed by native gel shift (lower).

Discussion

Nowadays, structural and biochemical studies of IspF and a number of ligands complexes from several species have been reported, which had produced an excellent understanding of substrate recognition and mechanism of catalysis by IspF enzyme [15,38-41,43-46,48]. In this study, crystal structures of *BsIspF* in native and complexed with CMP have been determined both at 1.99-Å resolutions. *BsIspF* reveals a trimeric enzyme with three active pockets. Residues around the active pocket are well conserved. The parts of active pocket, nucleotide and cation binding sites are relatively rigid, whereas the other part of pocket that binds the methylerythritol-phosphate portion of the substrate is relatively flexible. These flexible loops may assist in the successive conformational change of the active pocket, and benefit the substrates binding, intermediate forming, and products releasing [28]. The *BsIspF* trimer has high structural similarity with the *EcIspF*, only differences are observed in the packing of the subunits and the solvent accessible surface of active pockets. Sequence alignment reveals that the primary sequences of active pockets in *B. subtilis* and *E. coli* are highly conserved (except one residue). It is likely that the active pockets in *EcIspF* are more exposed to the solvent than in *BsIspF*. These surface variations may mainly result from the packing in the interface between two adjacent subunits, and lead to the difference on the active pockets, such as the size and the shape, which may affect the binding of the substrates and relate to the efficiency of enzyme [49]. In our previous work, the enzyme IspD from *B. subtilis* presents a higher CTP hydrolytic activity and also a tighter active pocket than that in *E. coli* [28]. Perhaps, the specific property of active pocket in *BsIspF* may relate to the catalytic efficiency of *BsIspF* in MEP pathway.

Previous studies had focussed on the IspDF bifunctional enzyme that catalyzed the third and fifth steps in MEP pathway steps and proposed possible associations could be formed between bifunctional IspDF and monofunctional IspE or the individual enzymes IspD, IspE, and IspF [17,19], however, further studies rejected this hypothesis [20]. In our work, whether a multienzyme complex can be formed *in vitro* amongst the monofunctional enzymes IspD, IspE, and IspF from *B. subtilis* and *E. coli* are analyzed by the traditional methods, native gel shift and size exclusion chromatography methods. Initially, we conceived if a multienzyme could be stably formed amongst the IspD, IspE, and IspF proteins, using structural biology methods, such as crystallography or cryo-EM, to investigate the protein complex assembly was both attractive and challenging. However, in our study, no protein complex assembly amongst IspD, IspE, and IspF is observed in *E. coli* and *B. subtilis in vitro*, which shows the structural biology methods are unsuited for further analysis for the proteins complex assembly of individual IspD, IspE, and IspF. Furthermore, a comparison between the structures of the monofunctional protein, *EcIspD*, *EcIspF*, *BsIspD*, *BsIspF*, and the fused protein IspDF from *C. jejuni* indicate that there have been some differences amongst them. Structure of *CjIspDF* reveals two distinct domains, corresponding to the cytidyltransferase IspD (N-terminal domain) and synthase IspF (C-terminal domain), respectively [19]. The two domains are connected by a linker loop comprising eight residues. The interface between the two domains in *CjIspDF* is mainly hydrophobic, and consists of several aliphatic amino acids, such as leucines, isoleucines, phenylalanines etc. [19]. These hydrophobic interactions may have an advantage on the stabilization of the fused protein. However, the corresponding region of the individual *EcIspD* (*BsIspD*), *EcIspF* (*BsIspF*) is dominated by polar amino acids, such as the glutamates and asparagines. These features may be unfavorable for the stable interaction between IspD and IspF. It must be pointed out that the native gel shift and size exclusion chromatography methods are limited tools for the analysis of proteins assembly *in vitro*, especially for the weak affinity by a complex formation. Perhaps, weak interactions may still be relevant in the crowded solutions found in the bacterial cytoplasm or the associations amongst IspD, IspE, and IspF might be transient in the bacterial cytoplasm, which is undetectable using the current methods *in vitro*. In fact, further investigations are required to clarify and the suitable methods, like NMR, CO-IP, SPR, etc. can be used in the weak interactions study, and that is what we will do in the next.

B. subtilis is a potential host cell for the production of isoprenoid. Several studies suggested that the enzymes of *B. subtilis* functioned more efficiently compared with that of other microorganisms [25-27,29]. In this work, the crystal structures of *BsIspF* and *BsIspF*/CMP complex have been determined. Structural comparisons reveal that the *BsIspF* has high structural similarity with the *EcIspF*, only differences are observed in the packing of the subunits and the solvent accessible surface of active pockets. In addition, the results show that the proteins complex assembly by IspD, IspE, and IspF enzymes in *B. subtilis* and *E. coli* are undetectable. The present study can supplement the high-resolution structures of the key enzyme and advance the understanding in the MEP pathway of *B. subtilis*. Probably, with advances in metabolic engineering, systems biology, and protein engineering, the *B. subtilis* strain can be optimized as an efficient Gram-positive host for the production of isoprenoids.

Acknowledgements

The Shanghai Synchrotron Radiation Facilities (SSRF), China is thankful for the provision of synchrotron radiation facilities and efficient support.

Funding

This work was supported by the 100 Talents Program of the Chinese Academy of Sciences ; the open fund from Key Laboratory of Microbial Physiological and Metabolic Engineering, Institute of Microbiology, Chinese Academy of Sciences(NO.KLIM-201303).

Author contribution

Z.L., Y.J., and G.W. conceived and designed the experiments. Z.L., Y.J., W.L., Y.T., and G.W. performed the experiments. Z.L., Y.J., Y.T., and G.W. analyzed the data, wrote, and revised the manuscript. Manuscript was reviewed, discussed, and approved by all the authors.

Competing interests

The authors declare that there are no competing interests associated with the manuscript.

Abbreviations

ASA, accessible surface area; *BslspF*, IspF protein of *B. subtilis*; CDPME, 4-diphosphocytidyl-2-C-methyl-D-erythritol; CDPME2P, 4-diphosphocytidyl-2-C-methyl-D-erythritol 2-phosphate; DMAPP, dimethylallyldiphosphate; *EclspF*, IspF from *E. coli*; GST, Glutathione-S-Transferase; IPP, isopentenyl diphosphate; IspD, 2-C-methyl-D-erythritol-4-phosphatecytidyltransferase; IspE, CDPME kinase; IspF, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; MEP, 2-C-methyl-D-erythritol-4-phosphate; MECDP, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate; MEPP, 2-C-methyl-D-erythritol; MVA, mevalonate; NCS, non-crystallographic symmetry; r.m.s.d, root mean square deviation; SSRF, Shanghai Synchrotron Radiation Facilities; TBE, Tris Boric acid EDTA.

References

- Sacchettini, J.C. and Poulter, C.D. (1997) Creating isoprenoid diversity. *Science* **277**, 1788–1789, <https://doi.org/10.1126/science.277.5333.1788>
- Bloch, K. (1992) Sterol molecule: structure, biosynthesis, and function. *Steroids* **57**, 378–383, [https://doi.org/10.1016/0039-128X\(92\)90081-J](https://doi.org/10.1016/0039-128X(92)90081-J)
- Bach, T.J. (1995) Some new aspects of isoprenoid biosynthesis in plants - a review. *Lipids* **30**, 191–202, <https://doi.org/10.1007/BF02537822>
- Rohmer, M., Knani, M., Simonin, P., Sutter, B. and Sahm, H. (1993) Isoprenoid biosynthesis in bacteria - a novel pathway for the early steps leading to isopentenyl diphosphate. *Biochem. J.* **295**, 517–524, <https://doi.org/10.1042/bj2950517>
- Rohmer, M. (1999) The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Nat. Prod. Rep.* **16**, 565–574, <https://doi.org/10.1039/a709175c>
- Lichtenthaler, H.K. (2000) Non-mevalonate isoprenoid biosynthesis: enzymes, genes and inhibitors. *Biochem. Soc. Trans.* **28**, 785–789, <https://doi.org/10.1042/bst0280785>
- Hunter, W.N. (2007) The non-mevalonate pathway of isoprenoid precursor biosynthesis. *J. Biol. Chem.* **282**, 21573–21577, <https://doi.org/10.1074/jbc.R700005200>
- Chaturvedi, D., Goswami, A., Saikia, P.P., Barua, N.C. and Rao, P.G. (2010) Artemisinin and its derivatives: a novel class of anti-malarial and anti-cancer agents. *Chem. Soc. Rev.* **39**, 435–454, <https://doi.org/10.1039/B816679J>
- Kunfermann, A., Witschel, M., Illarionov, B., Martin, R., Rottmann, M., Hoffken, H.W. et al. (2014) Pseudilins: halogenated, allosteric inhibitors of the non-mevalonate pathway enzyme IspD. *Angew. Chem.* **53**, 2235–2239, <https://doi.org/10.1002/anie.201309557>
- Thelemann, J., Illarionov, B., Barylyuk, K., Geist, J., Kirchmair, J., Schneider, P. et al. (2015) Aryl bis-sulfonamide inhibitors of IspF from *Arabidopsis thaliana* and *Plasmodium falciparum*. *Chem. Med. Chem.* **10**, 2090–2098, <https://doi.org/10.1002/cmdc.201500382>
- Masini, T. and Hirsch, A.K. (2014) Development of inhibitors of the 2C-methyl-D-erythritol 4-phosphate (MEP) pathway enzymes as potential anti-infective agents. *J. Med. Chem.* **57**, 9740–9763, <https://doi.org/10.1021/jm5010978>
- Zhao, L., Chang, W.C., Xiao, Y., Liu, H.W. and Liu, P. (2013) Methylerythritol phosphate pathway of isoprenoid biosynthesis. *Annu. Rev. Biochem.* **82**, 497–530, <https://doi.org/10.1146/annurev-biochem-052010-100934>
- Richard, S.B., Bowman, M.E., Kwiatkowski, W., Kang, I., Chow, C., Lillo, A.M. et al. (2001) Structure of 4-diphosphocytidyl-2-C-methylerythritol synthetase involved in mevalonate-independent isoprenoid biosynthesis. *Nat. Struct. Biol.* **8**, 641–648, <https://doi.org/10.1038/89691>
- Miallau, L., Alphey, M.S., Kemp, L.E., Leonard, G.A., McSweeney, S.M., Hecht, S. et al. (2003) Biosynthesis of isoprenoids: crystal structure of 4-diphosphocytidyl-2C-methyl-D-erythritol kinase. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9173–9178, <https://doi.org/10.1073/pnas.1533425100>
- Kemp, L.E., Bond, C.S. and Hunter, W.N. (2002) Structure of 2C-methyl-D-erythritol 2,4- cyclodiphosphate synthase: an essential enzyme for isoprenoid biosynthesis and target for antimicrobial drug development. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 6591–6596, <https://doi.org/10.1073/pnas.102679799>
- Testa, C.A., Lherbet, C., Pojer, F., Noel, J.P. and Poulter, C.D. (2006) Cloning and expression of IspDF from *Mesorhizobium loti*. Characterization of a bifunctional protein that catalyzes non-consecutive steps in the methylerythritol phosphate pathway. *Biochim. Biophys. Acta* **1764**, 85–96, <https://doi.org/10.1016/j.bbapap.2005.08.006>

- 17 Gabrielsen, M., Rohdich, F., Eisenreich, W., Grawert, T., Hecht, S., Bacher, A. et al. (2004) Biosynthesis of isoprenoids: a bifunctional IspDF enzyme from *Campylobacter jejuni*. *Eur. J. Biochem.* **271**, 3028–3035, <https://doi.org/10.1111/j.1432-1033.2004.04234.x>
- 18 Perez-Gil, J., Bergua, M., Boronat, A. and Imperial, S. (2010) Cloning and functional characterization of an enzyme from *Helicobacter pylori* that catalyzes two steps of the methylerythritol phosphate pathway for isoprenoid biosynthesis. *Biochim. Biophys. Acta* **1800**, 919–928, <https://doi.org/10.1016/j.bbagen.2010.06.008>
- 19 Gabrielsen, M., Bond, C.S., Hallyburton, I., Hecht, S., Bacher, A., Eisenreich, W. et al. (2004) Hexameric assembly of the bifunctional methylerythritol 2,4-cyclodiphosphate synthase and protein-protein associations in the deoxy-xylulose-dependent pathway of isoprenoid precursor biosynthesis. *J. Biol. Chem.* **279**, 52753–52761, <https://doi.org/10.1074/jbc.M408895200>
- 20 Lherbet, C., Pojer, F., Richard, S.B., Noel, J.P. and Poulter, C.D. (2006) Absence of substrate channeling between active sites in the *Agrobacterium tumefaciens* IspDF and IspE enzymes of the methyl erythritol phosphate pathway. *Biochemistry* **45**, 3548–3553, <https://doi.org/10.1021/bi0520075>
- 21 Li, Y. and Wang, G. (2016) Strategies of isoprenoids production in engineered bacteria. *J. Appl. Microbiol.* **121**, 932–940, <https://doi.org/10.1111/jam.13237>
- 22 Yang, C., Gao, X., Jiang, Y., Sun, B.B., Gao, F. and Yang, S. (2016) Synergy between methylerythritol phosphate pathway and mevalonate pathway for isoprene production in *Escherichia coli*. *Metab. Eng.* **37**, 79–91, <https://doi.org/10.1016/j.ymben.2016.05.003>
- 23 Morrone, D., Lowry, L., Determan, M.K., Hershey, D.M., Xu, M. and Peters, R.J. (2010) Increasing diterpene yield with a modular metabolic engineering system in *E. coli*: comparison of MEV and MEP isoprenoid precursor pathway engineering. *Appl. Microbiol. Biotechnol.* **85**, 1893–1906, <https://doi.org/10.1007/s00253-009-2219-x>
- 24 Yoon, S.H., Park, H.M., Kim, J.E., Lee, S.H., Choi, M.S., Kim, J.Y. et al. (2007) Increased beta-carotene production in recombinant *Escherichia coli* harboring an engineered isoprenoid precursor pathway with mevalonate addition. *Biotechnol. Prog.* **23**, 599–605, <https://doi.org/10.1021/bp070012p>
- 25 Zhao, Y.R., Yang, J.M., Qin, B., Li, Y.H., Sun, Y.Z., Su, S.Z. et al. (2011) Biosynthesis of isoprene in *Escherichia coli* via methylerythritol phosphate (MEP) pathway. *Appl. Microbiol. Biot.* **90**, 1915–1922, <https://doi.org/10.1007/s00253-011-3199-1>
- 26 Kuzma, J., Nemecek-Marshall, M., Pollock, W.H. and Fall, R. (1995) Bacteria produce the volatile hydrocarbon isoprene. *Curr. Microbiol.* **30**, 97–103, <https://doi.org/10.1007/BF00294190>
- 27 Wagner, W.P., Helmig, D. and Fall, R. (2000) Isoprene biosynthesis in *Bacillus subtilis* via the methylerythritol phosphate pathway. *J. Nat. Prod.* **63**, 37–40, <https://doi.org/10.1021/np990286p>
- 28 Jin, Y., Liu, Z., Li, Y., Liu, W., Tao, Y. and Wang, G. (2016) A structural and functional study on the 2-C-methyl-d-erythritol-4-phosphate cytidyltransferase (IspD) from *Bacillus subtilis*. *Sci. Rep.* **6**, 36379, <https://doi.org/10.1038/srep36379>
- 29 Julsing, M.K., Rijpkema, M., Woerdenbag, H.J., Quax, W.J. and Kayser, O. (2007) Functional analysis of genes involved in the biosynthesis of isoprene in *Bacillus subtilis*. *Appl. Microbiol. Biotechnol.* **75**, 1377–1384, <https://doi.org/10.1007/s00253-007-0953-5>
- 30 Wang, Q.-S., Yu, F., Huang, S., Sun, B., Zhang, K.-H., Liu, K. et al. (2015) The macromolecular crystallography beamline of SSRF. *Nucl. Sci. Tech.* **26**, 12–17
- 31 Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326, [https://doi.org/10.1016/S0076-6879\(97\)76066-X](https://doi.org/10.1016/S0076-6879(97)76066-X)
- 32 McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674, <https://doi.org/10.1107/S0021889807021206>
- 33 Collaborative Computational Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Cryst.* **D50**, 760–763, <https://doi.org/10.1107/S0907444994003112>
- 34 Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst.* **D53**, 240–255, <https://doi.org/10.1107/S0907444996012255>
- 35 Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Cryst.* **D60**, 2126–2132, <https://doi.org/10.1107/S0907444904019158>
- 36 Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J. et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Cryst.* **D66**, 12–21, <https://doi.org/10.1107/S0907444909042073>
- 37 Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y. and Liang, J. (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.* **34**, W116–W118, <https://doi.org/10.1093/nar/gkl282>
- 38 Richard, S.B., Ferrer, J.L., Bowman, M.E., Lillo, A.M., Tetzlaff, C.N., Cane, D.E. et al. (2002) Structure and mechanism of 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase. An enzyme in the mevalonate-independent isoprenoid biosynthetic pathway. *J. Biol. Chem.* **277**, 8667–8672, <https://doi.org/10.1074/jbc.C100739200>
- 39 Ni, S., Robinson, H., Marsing, G.C., Bussiere, D.E. and Kennedy, M.A. (2004) Structure of 2C-methyl-D-erythritol-2,4-cyclodiphosphate synthase from *Shewanella oneidensis* at 1.6 Å: identification of farnesyl pyrophosphate trapped in a hydrophobic cavity. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 1949–1957, <https://doi.org/10.1107/S0907444904021791>
- 40 Buetow, L., Brown, A.C., Parish, T. and Hunter, W.N. (2007) The structure of Mycobacteria 2C-methyl-D-erythritol-2,4-cyclodiphosphatesynthase, an essential enzyme, provides a platform for drug discovery. *BMC Struct. Biol.* **7**, <https://doi.org/10.1186/1472-6807-7-68>
- 41 Kemp, L.E., Alphey, M.S., Bond, C.S., Ferguson, M. A.J., Hecht, S., Bacher, A. et al. (2005) The identification of isoprenoids that bind in the intersubunit cavity of *Escherichia coli* 2C-methyl-D-erythritol-2,4-cyclodiphosphate synthase by complementary biophysical methods. *Acta Cryst.* **D61**, 45–52, <https://doi.org/10.1107/S0907444904025971>
- 42 Sgraja, T., Kemp, L.E., Ramsden, N. and Hunter, W.N. (2005) A double mutation of *Escherichia coli* 2C-methyl-D-erythritol-2,4-cyclodiphosphate synthase disrupts six hydrogen bonds with, yet fails to prevent binding of, an isoprenoid diphosphate. *Acta Crystallogr. Sect. F Struct. Biol. Commun.* **61**, 625–629, <https://doi.org/10.1107/S1744309105018762>

- 43 Lehmann, C., Lim, K., Toedt, J., Krajewski, W., Howard, A., Eisenstein, E. et al. (2002) Structure of 2C-methyl-D-erythrol-2,4-cyclodiphosphate synthase from *Haemophilus influenzae*: activation by conformational transition. *Proteins* **49**, 135–138, <https://doi.org/10.1002/prot.10182>
- 44 O'Rourke, P.E., Kalinowska-Tluszcik, J., Fyfe, P.K., Dawson, A. and Hunter, W.N. (2014) Crystal structures of IspF from *Plasmodium falciparum* and *Burkholderia cenocepacia*: comparisons inform antimicrobial drug target assessment. *BMC Struct. Biol.* **14**, 1, <https://doi.org/10.1186/1472-6807-14-1>
- 45 Kishida, H., Wada, T., Unzai, S., Kuzuyama, T., Takagi, M., Terada, T. et al. (2003) Structure and catalytic mechanism of 2-C-methyl-D-erythritol 2,4-cyclodiphosphate (MECDP) synthase, an enzyme in the non-mevalonate pathway of isoprenoid synthesis. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **59**, 23–31, <https://doi.org/10.1107/S0907444902017705>
- 46 Calisto, B.M., Perez-Gil, J., Bergua, M., Querol-Audi, J., Fita, I. and Imperial, S. (2007) Biosynthesis of isoprenoids in plants: structure of the 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase from *Arabidopsis thaliana*. Comparison with the bacterial enzymes. *Protein Sci.* **16**, 2082–2088, <https://doi.org/10.1110/ps.072972807>
- 47 Begley, D.W., Hartley, R.C., Davies, D.R., Edwards, T.E., Leonard, J.T., Abendroth, J. et al. (2011) Leveraging structure determination with fragment screening for infectious disease drug targets: MECP synthase from *Burkholderia pseudomallei*. *J. Struct. Funct. Genomics* **12**, 63–76, <https://doi.org/10.1007/s10969-011-9102-6>
- 48 Steinbacher, S., Kaiser, J., Wungsintaweekul, J., Hecht, S., Eisenreich, W., Gerhardt, S. et al. (2002) Structure of 2C-methyl-D-erythritol-2,4-cyclodiphosphate synthase involved in mevalonate-independent biosynthesis of isoprenoids. *J. Mol. Biol.* **316**, 79–88, <https://doi.org/10.1006/jmbi.2001.5341>
- 49 Williams, D.H., Stephens, E. and Zhou, M. (2003) Ligand binding energy and catalytic efficiency from improved packing within receptors and enzymes. *J. Mol. Biol.* **329**, 389–399, [https://doi.org/10.1016/S0022-2836\(03\)00428-5](https://doi.org/10.1016/S0022-2836(03)00428-5)
- 50 Ashkenazy, H., Erez, E., Martz, E., Pupko, T. and Ben-Tal, N. (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **38**, W529–W533, <https://doi.org/10.1093/nar/gkq399>
- 51 Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797, <https://doi.org/10.1093/nar/gkh340>
- 52 Bond, C.S. and Schuttelkopf, A.W. (2009) ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **65**, 510–512, <https://doi.org/10.1107/S0907444909007835>