Published in final edited form as:

J Mol Biol. 2008 August 1; 381(1): 129–140. doi:10.1016/j.jmb.2008.05.060.

The structural basis of chain length control in Rv1086

Wenjian Wang^{*}, Changjiang Dong^{*}, Michael McNeil[†], Devinder Kaur[†], Seababrata Mahapatra[†], Dean C. Crick[†], and James H. Naismith^{*}

^{*}Centre for Biomolecular Science, The University, St Andrews, Scotland, U.K. KY16 9RH

[†]Mycobacteria Research Laboratories, Department of Microbiology, Immunology, and Pathology, Colorado State University, Fort Collins, Colorado 80523

Abstract

In *Mycobacterium tuberculosis* two related Z-prenyl diphosphate synthases, E,Z-farnesyl diphosphate synthase (Rv1086) and decaprenyl diphosphate synthase (Rv2361c) work in series to synthesize decaprenyl phosphate (C₅₀) from isopentenyl diphosphate and E-geranyl diphosphate. Decaprenyl phosphate plays a central role in the biosynthesis of essential mycobacterial cell wall components, such as the mycolyl-arabinogalactan-peptidoglycan complex and lipoarabinomannan; thus, its synthesis has attracted considerable interest as a potential therapeutic target. Rv1086 is a unique prenyl diphosphate synthase in that it adds only one isoprene unit to geranyl diphosphate generating the 15 carbon product (E,Z-farnesyl diphosphate). Rv2361c then adds a further 7 isoprene units to E,Z-farnesyl diphosphate in a processive manner to generate the 50 carbon prenyl diphosphate, which is then dephosphorylated to generate a carrier for activated sugars. The molecular basis for chain length discrimination by Rv1086 during synthesis is unknown. We also report the structure of apo Rv1086 with citronellyl diphosphate bound and with the product mimic E,E-farnesyl diphosphate bound. We report the structures of Rv2361c in the apo form, with isopentyl diphosphate bound and with a substrate analogue, citronellyl diphosphate. The structures confirm the enzymes are very closely related. Detailed comparison reveals structural differences that account for chain length control in Rv1086. We have tested this hypothesis and have identified a double mutant of Rv1086 which makes a range of longer lipid chains.

Keywords

Drug design; enzyme mechanism; tuberculosis; x-ray crystallography; inhibitors

Introduction

Pathogenic bacteria synthesize complex sugar containing polymers which they utilize in their cell wall as structural components including, in some cases, polymers that shield the organism from the human immune system. The biosynthesis of these complex carbohydrate molecules often utilizes sugars which are linked to long chain lipid (polyprenyl phosphate, Pol-P) molecules. The most important example of this is involves peptidoglycan biosynthesis where

Address correspondence to: James Naismith Tel 44-1334-463792; Fax 44-1334-467229; E-mail: naismith@st-and.ac.uk.

Accession Numbers: Coordinates and structure factors have been deposited in the Protein Data Bank with accession number 2vg3, 2vg2, 2vg4, 2vg0, 2vg1 and 2vfw.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Pol-P is an essential carrier for its biosynthesis. Hence, these long-chain lipid carrier molecules are an essential component of cell wall biosynthesis and, as such, they have attracted considerable attention as potential drug targets.

The most common prenol (and therefore Pol-P) structures are generally confined to four main groups: (i) all-*E*-prenol, (ii) di-*E*, poly-*Z*-prenol, (iii) tri-*E*, poly-*Z*-prenol, and (iv) all-*Z*-prenol¹. Eubacteria typically contain a single predominant Pol-P composed of 11 isoprene units (55 carbon atoms), with the di-*E*, poly-*Z* configuration as seen in undecaprenyl phosphate (bactoprenyl phosphate)^{2,3}. The enzymes which carry out the synthesis of these molecules constitute the *Z*-prenyl synthases⁴ and in bacteria these enzymes are highly conserved as shown in the sequence comparisons illustrated in Figure 1a. Generally there is a single *Z*-prenyl synthase found in each organism. Thus, undecaprenyl phosphate is made using single enzyme undecaprenyl diphosphate synthase (UDPS) starting from a 15 carbon *E*,*E*-farnesyl diphosphate (EE-FPP) substrate. UDPS sequentially adds nine isoprene units⁵. In contrast, *Mycobacterium* spp. contain structurally unusual Pol-P molecules, which have 10 isoprene units (50 carbon atoms) with a mono *E*- and poly-*Z*-configuration^{3,6-8}. Thus, the human pathogen *M. tuberculosis* relies on an unusual combination of two related *Z*-prenyl synthetase enzymes, Rv2361c and Rv1086 (Figure 1a), to synthesize this 50 carbon molecule decaprenyl diphosphate⁶⁻⁸, which is subsequently dephosphorylated, generating an activated carrier molecule for sugars. Rv1086 makes only a (relatively) short C₁₅ *E*,*Z*-farnesyl diphosphate (EZ-FPP). Although Rv2361c can use EE-FPP and even the 10 carbon geranyl diphosphate (GPP) as substrates, it has a strong substrate preference for EZ-FPP to which it adds seven isoprene unit⁶ (Figure 1b shows these isoprene structures). Since the discussion of substrate and product binding sites can be confusing for an enzyme which further elongates its products (i.e. a processive enzyme), we designate the binding sites as the isopentenyl diphosphate (IPP) binding site and as the isoprene polymer (polymer) binding site as illustrated in Figure 1c.

The first crystal structure of a *Z*-prenyl synthase was reported for UDPS from *Micrococcus luteus* and described the IPP and isoprene polymer binding sites⁹. The structural characterization of the enzyme from *E. coli* with substrate and substrate analogue complexes identified the key residues involved in recognition and catalysis¹⁰. The mechanism proposed for UDPS (and by extension all *Z*-prenyl transferases)^{9,10} requires binding both IPP and EE-FPP. An essential Mg²⁺ ion catalyses the nucleophilic attack of IPP at the C1 of EE-FPP, displacing diphosphate and making the 20 carbon isoprene polymer (Figure 1c) which now occupies the IPP site. The nascent isoprene polymer then moves from the IPP site to the isoprene polymer site allowing reloading of IPP for the next condensation (Figure 1c). Recently studies on the *M. luteus* UDPS enzyme have identified several residues that are important in regulating the length of the C₅₅ product¹¹. Mutant UDPS enzymes which make both longer and shorter product chain lengths have been reported¹¹.

Rv1086 unlike all other characterized *Z*-prenyl diphosphate synthases is not a processive enzyme, instead, it condenses IPP and GPP, to synthesize only EZ-FPP. The molecular basis for this tight regulation of chain length is currently unknown. We now report the crystal structures of Rv1086 and Rv2361c, both complexed with an inhibiting substrate analogue citronellyl diphosphate (CITPP). In addition, for Rv1086, we report an EE-FPP complex (product mimic) and for Rv2361c we also report an IPP complex. These structures have allowed detailed molecular insight into chain length regulation in Rv1086. We have engineered a mutant of Rv1086 which is no longer limited to 15 carbon product.

Results

Overall structure

Rv1086 has a 6 stranded β sheet which has four α helices on one face (Figure 2a). Rv2361c has a very similar structure to Rv1086 (Figure 2b) and the root mean square deviation for 203 overlapping C α atoms is 1.4 Å. A comparison of the structure is shown in supplementary Figure 1. Gel filtration studies and the crystal structures suggest both Rv1086 and Rv2361c are dimers. The dimer interface is reminiscent of a four helix bundle, with two helices from each monomer. In addition to the helical bundle, there are contacts from the central β -sheet in one monomer to a loop in the other monomer (Figure 2a, b). Additionally in Rv2361c the C-terminus of one monomer touches the active site of the other (Figure 2b). The nature of this contact varies as the structure of the C-terminus is sensitive to crystal packing. Both Rv1086 and Rv2361c are very similar in structure to *M. luteus*⁹ and *E. coli* UDPS¹⁰. The rmsd is 1.3 Å over 205 C α atoms for a superposition of Rv1086 with UDPS and 1.2 Å over 216 C α atoms for Rv2361c. All Z-prenyl enzymes structurally characterized to date have the same dimeric arrangement^{9,10} but only Rv2361c has an additional contact between the active site and C-terminus (Figure 2b). Rv2361c has an additional fifty residues at the N-terminus that fold up into a compact domain (Figure 2b) that is not seen in Rv1086.

Co-complexes of Rv2361c

In the Rv2361c CITPP complex (Supplementary Figure 2a,b), the inhibitor is bound at the isoprene polymer binding site in all four monomers of the asymmetric unit. The β -phosphate makes salt links to R80, R89 and hydrogen bonds to G79 and R80. G79 is at the N-terminal end of an α helix. The α -phosphate makes salt contacts to R89 and R127. In each subunit, a Mg²⁺ ion is bound to the diphosphate of CITPP. The Mg²⁺ ion is in the same octahedral environment seen in *E. coli*¹⁰ (Figure 2c). In two monomers a phosphate ion is bound at the IPP site and it bridges to the two water molecules which coordinate the Mg²⁺ ion. CITPP has a single bond between C2 and C3 rather than the Z configured double bond. In all four subunits the bond adopts a gauche orientation such that the dihedral angle is around 60° between C4 and C1, this is of course much closer to a Z configuration (0°) than a E configuration (180°). The hydrophobic tail of CITPP binds in channel formed by the side chains of M75, W271, A119, L135, F138 and N139 with the main chain of H93 and G96. The channel is very wide and is not a “tight fit”, this is reflected in the variation of the conformation of the terminal allyl group of CITPP in the four subunits. This is in agreement with the requirement for the growing isoprene polymer to be pushed through this channel; a specific fit would hamper the movement of the elongating chain.

In the IPP structure, we see a variation in active site content. Two of the four monomers have density we interpret as an IPP molecule bound at the IPP binding site (Supplementary Figure 2c,d), the other two are empty. The density is poor. R292 from the C-terminus neighboring subunit binds to diphosphate. The β -phosphate is also held by hydrogen bond and salt contacts with R244, R250 and S252. The α -phosphate makes salt contacts and hydrogen bonds with S121, R127 and N124. The hydrophobic part of IPP sits against the aromatic ring of Y118. There are slight differences in the conformation of the IPP molecules in the two subunits. In both monomers the OD1 atom of the side chain of N124 is within 4.0 Å of C2 of IPP.

Co-complexes of Rv1086

In the Rv1086 CITPP complex, both monomers have CITPP bound at the isoprene polymer site and the IPP site is empty. In contrast to Rv2361c in Rv1086 in both subunits the C2 C3 atoms are disordered leading to a break in the electron density for the molecule (Supplementary Figure 2e,f). This would be consistent with CITPP adopting a range of different conformers around the C2 C3 bond. The diphosphate is bound by a series of salt contacts and hydrogen

bonds to R43, R44 and R92. The β phosphate is also bound to the positively charged N-terminus of an α helix. CITPP is a C_{10} analogue and in this complex it appears to mimic the GPP substrate. We have obtained a complex with Rv1086 and EE-FPP. In this complex a complete molecule of EE-FPP is found only in one monomer (Supplementary Figure 2g, h). EE-FPP is bound at the IPP binding site, mimicking the final EZ-FPP product. Again there is some disorder possibly because the mimic has a different stereochemistry (trans) than the true product (cis). The diphosphate of EE-FPP is bound by R211, R217 and S219. OD1 of N89 is 4.3 Å from C2 of the EE-FPP. The allyl group of both EE-FPP and CITPP overlap and sit in what appears to be specific closed pocket. The pocket is formed by residues C39, A62, I65, L84, I104 and V107 (Figure 3a).

Mutants of Rv1086 prenyl diphosphate synthases

The structure identified L84 and V107 in Rv1086 as potentially important in limiting chain length (Figure 3a, b). The single mutant Rv1086_{L84A} was not expressed in a soluble form. Rv1086_{V107G} although expressed and purified had no detectable activity, although the protein appeared to be folded from CD studies. Mutation at this position was not pursued further. Following an observation from other work¹¹ we attempted to make compensating mutations adjacent to L84 to identify a protein which would express in a soluble form. The double point mutant Rv1086_{L84A,L85F} was well expressed and active. Analysis of the products synthesized showed that the mutant protein is capable of synthesizing prenyl diphosphates containing 20 carbon atoms when GPP is used as the allylic substrate and 50 carbon atoms when EE-FPP is used as the allylic substrate (Figure 4), whereas the wild-type enzyme generated product that was restricted to 15 carbon atoms with GPP as substrate and does not react with EE-FPP⁸. In addition, kinetic analysis indicated that Rv1086_{L84A,L85F} has a K_m^{GPP} of 150 $\pm$ 28 μ M (Figure 5) compared to the previously reported value of 38 μ M for wild-type Rv1086¹². When normalized for assay time and protein content the maximal reaction rate of the wild-type enzyme reported earlier⁸ was only 10 fold greater than that of Rv1086_{L84A,L85F}.

Discussion

The crystal structures of Rv2361c and Rv1086 confirm these proteins are related and belong to the Z-prenyl synthetase superfamily. The structural data support the UDPS study which concluded^{9,10} that N89 of Rv1086 (N124 in Rv2361c) act as the catalytic base removing a proton from C2 of IPP. In all complexes where the IPP site is occupied the OD1 atom of this Asn residue is within 4.3 Å of the C2 of IPP where the *pro* S proton has to be abstracted (Figure 1c). The basicity of the Asn is enhanced by its hydrogen bond to the α phosphate of IPP. Superposition of the IPP and CITPP complexes of Rv2361c allows the construction of a model of the enzyme substrate complex. In this model the C4 of IPP is 3.0 Å from C1 of CITPP, more importantly the angle between C4, C1 and the phosphate of CITPP is 140°. This is very close to the 180° required for an S_N2 process which from chemical analysis would seem the likely route of the reaction^{9,10}.

Superposition of Rv2361c, Rv1086 and UDPS co-complexes^{9,10} reveals that the key residues which contact IPP are conserved as are those that interact with diphosphate at the isoprene polymer site. The carbon chain of CITPP and E,E-farnesylthiodiphosphate fill the same space. In Rv2361c CITPP complex it adopts a gauche (close to cis) conformer; in contrast in the E,E-farnesylthiodiphosphate UDPS complex this is trans (Supplementary Figure 3). This structural difference reflects the respective substrate preferences of the two enzymes, however there appears to be no obvious single residue responsible for the difference.

Rv1086 is a unique polyprenyl-diphosphate synthase, all other polyprenyl-diphosphate synthases are capable of adding multiple isopentenyl units to a variety of primers (ranging from C_{10} to C_{40}). Rv1086 adds only one isopentenyl unit to GPP (or Z-neryl diphosphate (Figure

1b,c) *in vitro*⁸). The difference in chain-length discrimination does not appear to lie at the active site which is highly conserved (Figure 1a and Supplementary Figure 4) between Rv1086 and Rv2361c. Rather the differences lie in the “hydrophobic channel”^{9,10} adjacent to the active site (Figure 3a). Comparison of the Rv1086 structure with Rv2361c reveals that in Rv1086 residues I65, L84, I104 and V107 (V100, A119, N139 and V142 in Rv2361c) essentially close the hydrophobic channel seen in Rv2361c (Figure 3a, b). V107 is part of 20 residue α helix which in Rv1086 intrudes into the hydrophobic channel, in Rv2361c the helix is shifted out of the channel. Interestingly mutants of the *M. luteus* enzyme which have lengthened the loop attached the helix, can make products greater than C₅₅¹¹. In Rv1086 the side chain of L84 sits in the channel whilst in Rv2361 the smaller A119 leaves the channel open (Figure 3a). We hypothesized that the bulky side chain of L84 of Rv1086 would be very important in limiting chain length to C₁₅ in Rv1086 and its mutation in L84A should allow longer chains to be made (Supplementary Figure 5). The single mutant Rv1086_{L84A} is not expressed in a soluble form and therefore we tested double mutants to find a soluble enzyme (Table 1). The Rv1086_{L84A,L85F} double mutant is overexpressed in a soluble form. Remarkably the double mutant is active and produces the longer carbon chain molecules (up to 50 carbons detectable) (Figure 4). L85 is not absolutely conserved, in a BLAST analysis of cis prenyl diphosphate synthases, this residue is found as M, C or F. Further L85 in Rv1086 (and F120 in Rv2361c) points away from the hydrophobic tunnel in every structure we have determined (Figure 3a) and in all other cis prenyl diphosphate synthase structures¹⁰. Mutation of the equivalent residue in *M. luteus* UDPS F73A does affect chain length but only because it reduces the overall catalytic efficiency of the processive enzyme as distinct from direct regulation of chain length¹¹. Taken together, we propose that the L85F mutant helps to fold the protein into an active conformation rather than directly affects the chain length regulation. We cannot however exclude a more direct role. The double mutant had approximately a four fold increase in K_m to 150 μ M, relative to the wildtype enzyme⁸, and a reduction in the reaction rate (Figure 5). Our data suggest that L84 is a key residue in restricting the product length of Rv1086 to C₁₅. This complements a study of *M. luteus* UDPS which identified a triple mutant (A72L, F73L, W78L) which produces polyprenyl diphosphates with a range of chain-lengths shorter than C₅₅¹¹. It appears that the Z-prenyl diphosphate synthases have quite simple molecular switches remote from the active site that regulate chain length. This would suggest these are plausible candidates for protein engineering studies, since these approaches are often confounded by difficulty in separating out residues important in recognition from those involved in catalysis.

Rv1086 has a crucial role, unique amongst Z-prenyl synthases, in that it catalyzes a single condensation of IPP and GPP synthesizing EZ-FPP which is the preferred substrate for Rv2361c⁷. We have identified the molecular basis for the unique regulation of chain length in Rv1086. High density transposon mutagenesis experiments predict that Rv2361c is likely to be essential in *M. tuberculosis* H37Rv, suggesting it as an anti-mycobacterial target. The same experiments suggest Rv1086 is not an essential gene for the survival of *M. tuberculosis*¹³, a surprising result given that Rv1086 synthesizes the precursor for Rv2361c^{6,12}. However Rv2361c has been shown to be capable of synthesizing EZ-FPP⁶ although at a greatly reduced catalytic efficiency compared to Rv1086. Rv2361c presumably compensates for the deletion of Rv1086 under the experimental conditions¹³. However, the evolution of two distinct prenyl diphosphate synthases in *M. tuberculosis* to synthesize EZ-FPP could indicate that Rv1086 may well be more important to pathogenesis than indicated by the transposon mutagenesis experiments. Both Rv1086 and Rv2361c may therefore present novel targets for therapeutic intervention and our work has established a molecular basis for inhibitor design.

Materials and methods

Materials

[1-¹⁴C]Isopentenyl diphosphate ([¹⁴C]IPP, 55 mCi/mmol) was purchased from GE Healthcare Life Sciences. Kanamycin, (S)-(-)-β-citronellol, *E,E*-farnesol, geraniol and *E,E,E*-geranylgeraniol were purchased from Sigma-Aldrich. *E,E*-farnesyl diphosphate (FPP), CITPP and geranyl diphosphate (GPP) were synthesized as described by Davisson *et al*¹². Authentic long-chain polyprenols were obtained from the Institute of Biochemistry and Biophysics, Polish Academy of Sciences (Warsaw, Poland).

LKC₁₈F reverse-phase TLC plates were from Whatman, potato acid phosphatase (grade 2) was purchased from Roche Molecular Biochemicals, LB agar and LB broth were from Invitrogen, restriction enzymes NdeI and XhoI, dNTPs and T4 DNA ligase were purchased from New England Biolabs. *Escherichia coli* competent cells (DH5α), pET28a vector and *E. coli* BL21 (DE3)pLysS were purchased from Novagen (Madison, WI). Qiaprep Spin Miniprep kits were purchased from Qiagen Inc. Monoclonal anti-polyhistidine unconjugated antibody from mouse ascites fluid, anti-mouse IgG conjugated to alkaline phosphatase and 5-bromo-4-chloro-3-indoylphosphate tablets were purchased from Sigma-Aldrich. Oligonucleotide synthesis and sequencing were done by Macromolecular Resources (Colorado State University).

Molecular biology

M. tuberculosis open reading frames *Rv1086* and *Rv2361c* were cloned into the pET28a vector giving His tagged construct with a thrombin cleavage site. *E. coli* BL21-DE3(gold) cells were transformed with the plasmids and cultured in tryptone-phosphate medium at 37 °C until OD₆₀₀ nm reached 0.7 - 0.8. Proteins were overexpressed by addition of 1 mM IPTG and further incubation at 20 °C overnight.

For the *Rv1086* protein, the harvested cells were disrupted in 20 mM phosphate pH 7.2, 0.5 M NaCl, 10 mM imidazole by sonication for 6 cycles of 30 second on ice. The disrupted cells were centrifuged at 6000 rpm and the supernatant were loaded on 5 ml of HitrapTM chelating column (Amersham Biosciences). *Rv1086* protein was eluted by a linear gradient elution from 30 mM imidazole to 500 mM imidazole containing 20 mM pH 7.2 phosphate, 0.5 M NaCl. The fractions containing the *Rv1086* protein were pooled and dialyzed against 20 mM pH7.8 Tris buffer overnight. The N-terminal His-tag was removed by incubation overnight with thrombin (50 mg of *Rv1086* per unit of thrombin). The cleaved *Rv1086* was loaded on a 5ml of HitrapTM Q FF column (Amersham Biosciences) and the protein was eluted by a linear gradient from 0 M to 1M NaCl containing 20 mM pH7.8 Tris. The protein was further purified by a gel filtration column HiPrepTM S-100 16/60 (Amersham Biosciences) using 20 mM pH7.2 Tris and 0.3 M NaCl as running buffer. The purified protein was dialyzed against 20 mM pH7.2 Tris, 2 mM MgCl₂, and 5 mM DTT and the protein was concentrated to 7 mg ml⁻¹ for crystallization.

For *Rv2361c* protein, the cells were harvested as above. The supernatant was applied to Ni-NTA (Qiagen), with a loading and washing buffer of 50 mM phosphate pH 7.2, 0.3 M NaCl, 30 mM imidazole. *Rv2361* protein was eluted in single step by buffer containing 50 mM phosphate pH 7.2, 0.3 M NaCl, 250 mM imidazole. After the thrombin cleavage under the same conditions as *Rv1086*, the protein was loaded on a cation column Poros HS20 (Applied Biosystems). It was eluted using a linear gradient from 0 M to 2 M NaCl containing 20 mM pH7.2 Tris, 100 mM NaCl. The *Rv2361c* protein was further purified by the HiPrepTM S-100 16/60 (Amersham Biosciences) as outlined above. Purified *Rv2361* protein was dialyzed against 20 mM pH7.2 Tris containing 0.15 mM MgCl₂ and 3 mM DTT, and concentrated to 6.5 mg ml⁻¹ for crystallization.

Proteins for assay were loaded on a cobalt based TALON™ (BD Biosciences Clontech) immobilized metal affinity column (IMAC) which was washed with 50 mM MOPS (pH 8), 0.1% TritonX-100, 7 mM imidazole and 500 mM NaCl eluted using washing buffer containing a linear gradient of 25-200 mM imidazole. Protein containing fractions were analyzed by SDS PAGE and Western blot. Fractions that were estimated to be at least 90% pure by Coomassie Brilliant Blue 250R staining were pooled, desalted and used for further characterization. Protein concentrations were estimated using a BCA protein assay kit (Pierce). Site-specific mutations were generated using overlap extension essentially as described by Sambrook and Russell¹⁴. PCR primers are listed in Table 1. Expression and purification of proteins was as described above.

Crystallography

Both proteins were screened for crystallization using sparse matrix screens and vapor diffusion. The system we employ uses 100 nL protein drop and 100 nL precipitant in equilibrium as 100 μ L precipitant (Rhombix Vision) at 20 °C. Once initial crystallization conditions were found, these were optimized by varying precipitant and protein concentrations using volumes of 2 μ L of protein and 2 μ L precipitant with vapor diffusion at 20 °C. Optimal crystals were those which gave the highest resolution of diffraction. Apo Rv1086 crystals were grown in 0.1 M pH 6.3 MES, 0.2 M (NH₄)₂SO₄, 34% PEGMME-5000. Rv1086 crystals in complex with CITPP were obtained in 0.1 M pH 7.1 MES, 0.2M (NH₄)₂SO₄, 24% PEGMME-5000. Before the crystallization, Rv1086 was first incubated with CITPP at a final concentration of 1 mM over 2 hrs in room temperature. The Rv1086 crystals in complex with EE-FPP were obtained in 0.1 M pH 6.7 MES, 0.2 M (NH₄)₂SO₄, 24% PEGMME-5000. Before the crystallization, the Rv1086 was incubated with 4 mM EE-FPP for 3 hrs at room temperature. Apo Rv2361c crystals were obtained in 0.1 M pH 7.1 HEPES, 24% (v/v) PEG-600, 0.05 M Li₂SO₄, 10% (v/v) glycerol. The Rv2361c crystals in complex with IPP were obtained in 0.1 M pH 7.1 HEPES, 30% PEG1000, 0.05 M Li₂SO₄, 10% glycerol and in presence of 5 mM IPP and 4 mM FPP. It had been hoped to trap the product of the reaction, but only IPP was visualized. The Rv2361c complex with CITPP crystallized from 0.1 M pH 6.9 HEPES, 24% PEG-1000, 0.05 M Li₂SO₄, 10% glycerol. Before crystallization, Rv2361c was incubated with 5 mM CITPP for 3 hrs at room temperature.

Crystals were flash frozen at 100 K prior to and during data collection. Rv1086 crystals were first soaked in the crystallization solution with 15 % (v/v) glycerol. Rv2361c crystals could be frozen straight from the crystallization drop. All data were collected as 0.5° oscillations on ESRF stations ID14-1 and ID29 using ADSC CCD detectors. Data were integrated with MOSFLM¹⁵ and scaled with SCALA in the CCP4 suite¹⁶. Full details on the data collected are given in Table 2. The Rv2361c apo structure was solved using the *E. coli* UDPS structure⁹ (pdb code 1x07) as a search model with PHASER^{17,18}. All Rv2361c co-complexes and apo Rv1086 were solved by molecular replacement using the apo Rv2361c structure as the search model. The two Rv1086 co-complexes were solved using apo Rv1086 as the search model. Both the native and CITPP co-complexes of Rv2361c exhibit merohedral twinning (fractions 0.34 and 0.08 respectively). The native structure was refined using PHENIX-REFINE where the twin fraction is treated as variable and explicit hydrogen were added¹⁹. The native Rv2361c structure is of lower quality than the others but it is ranked by MOLPROBITY²⁰ amongst the best 30 % of structures at that resolution. After manual detwinning using the CCP4 suite¹⁶, the CITPP complex refined smoothly. It and the other structures were refined using REFMAC5²¹. Ligands were added in portions to the model when the Fo-Fc difference electron density map indicated their presence, manual inspection and building used COOT²² (Supporting Figures 2a-h). Most of the ligands give weak density either due to disorder or because they are at less than full occupancy, there could be contaminating phosphate from the preparations of ligand. Low occupancy is possible due to the difficulty in

obtaining high concentrations of the relatively insoluble lipids, we did not refine occupancy for any structure. Dictionaries for the ligands were created using PRODRG server²³. Structures and data (twinned as well as detwinned) have been deposited with the Protein Data Bank²⁴. Structural figures were generated with PYMOL (www.pymol.org).

Prenyl diphosphate synthase assays

Typically, enzyme activity was measured in a 50 μ l reaction mixture containing 50 mM MOPS (pH 7.9), 5.0 mM dithiothreitol, 10 mM sodium orthovanadate, 1 mM MgCl_2 , 0.3% Triton X-100, 30 μ M [^{14}C]IPP, and the indicated concentrations of allylic diphosphate. Reactions were initiated by the addition of 1-10 μ g of recombinant enzyme. The mixture was incubated for 30 min at 37°C and the reaction was stopped with 1 ml of water saturated with NaCl. The products were then extracted with n-butanol saturated with water and an aliquot was taken for liquid scintillation spectrometry. All assays were conducted under conditions where product formation was linear for both time and protein concentration.

To analyze the chain length of the radioactive polyprenyl diphosphate products, the enzymatically synthesized radiolabelled compounds were hydrolyzed to the corresponding alcohols with potato acid phosphatase as described previously²⁵. The alcohols were extracted from the reaction mixture with hexane and analyzed on reverse phase TLC plates developed in 4 parts methanol 1 part water. Authentic standards were visualized with an anisaldehyde spray, and the radioactive compounds were detected using a Bioscan System 200 Imaging System and/or autoradiography. All values reported are averages of duplicate measurements from representative experiments. K_m and values were calculated using non-linear regression analysis (GraFit 5.0.13, Erithacus Software Ltd.; $v = (V_{\max}[\text{S}])/(K_m + [\text{S}])$, initial estimates come from Scatchard re-arrangement).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Robert Eales and Casie Johnson for technical assistance. The work was supported by NIH grants AI057836, AI049151, AI065357, AI018357 and Wellcome Trust.

The abbreviations used

(S)-(-)- β -citronellyl diphosphate
CITPP

ω ,E,Z-farnesyl diphosphate
EZ-FPP

ω ,E,E-farnesyl diphosphate
EE-FPP

isopentyl diphosphate
IPP

undecaprenyl phosphate synthase
UDPS

References

1. Cornishbowden A. Prenol Nomenclature - Recommendations 1986. *Eur J Biochem* 1987;167:181–184. [PubMed: 3113948]
2. Bugg TDH, Brandish PE. From Peptidoglycan to Glycoproteins - Common Features of Lipid-Linked Oligosaccharide Biosynthesis. *FEMS Microbiol Lett* 1994;119:255–262. [PubMed: 8050708]
3. Crick DC, Mahapatra S, Brennan PJ. Biosynthesis of the arabinogalactan-peptidoglycan complex of *Mycobacterium tuberculosis*. *Glycobiology* 2001;11:107r–118r.
4. Seiji Takahashi TK. Structure and function of *cis*-prenyl chain elongating enzymes. *The Chemical Record* 2006;6:194–205.
5. Shimizu N, Koyama T, Ogura K. Molecular cloning, expression, and purification of undecaprenyl diphosphate synthase - No sequence similarity between E- and Z-prenyl diphosphate synthases. *J Biol Chem* 1998;273:19476–19481. [PubMed: 9677368]
6. Crick DC, Schulbach MC, Zink EE, Macchia M, Barontini S, Besra GS, Brennan PJ. Polyprenyl phosphate biosynthesis in *Mycobacterium tuberculosis* and *Mycobacterium smegmatis*. *J Bacteriol* 2000;182:5771–5778. [PubMed: 11004176]
7. Kaur D, Brennan PJ, Crick DC. Decaprenyl diphosphate synthesis in *Mycobacterium tuberculosis*. *J Bacteriol* 2004;186:7564–7570. [PubMed: 15516568]
8. Schulbach MC, Brennan PJ, Crick DC. Identification of a short (C-15) chain Z-isoprenyl diphosphate synthase and a homologous long (C-50) chain isoprenyl diphosphate synthase in *Mycobacterium tuberculosis*. *J Biol Chem* 2000;275:22876–22881. [PubMed: 10816587]
9. Fujihashi M, Zhang YW, Higuchi Y, Li XY, Koyama T, Miki K. Crystal structure of *cis*-prenyl chain elongating enzyme, undecaprenyl diphosphate synthase. *Proc Natl Acad Sci U S A* 2001;98:4337–4342. [PubMed: 11287651]
10. Guo RT, Ko TP, Chen APC, Kuo CJ, Wang AHJ, Liang PH. Crystal structures of undecaprenyl pyrophosphate synthase in complex with magnesium, isopentenyl pyrophosphate, and farnesyl thiopyrophosphate - Roles of the metal ion and conserved residues in catalysis. *J Biol Chem* 2005;280:20762–20774. [PubMed: 15788389]
11. Kharel Y, Takahashi S, Yamashita S, Koyama T. Manipulation of prenyl chain length determination mechanism of *cis*-prenyltransferases. *FEBS J* 2006;273:647–657. [PubMed: 16420487]
12. Davisson VJ, Woodside AB, Poulter CD. Synthesis of Allylic and Homoallylic Isoprenoid Pyrophosphates. *Meth Enzymol* 1985;110:130–144. [PubMed: 2991702]
13. Sassetti CM, Boyd DH, Rubin EJ. Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* 2003;48:77–84. [PubMed: 12657046]
14. Sambrook J, Russell DW. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory; Cold Spring Harbor, NY: 2001.
15. Leslie AGW. Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 and ESF-EAMCB newsletter on protein crystallography* 1992;No 26:1–10.
16. CCP4. The CCP4 suite: Programs for Protein Crystallography. *Acta Crystallogr, Sect D: Biol Crystallogr* 1994;50:760–763. [PubMed: 15299374]
17. McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ. Likelihood-enhanced fast translation functions. *Acta Crystallogr, Sect D: Biol Crystallogr* 2005;61:458–464. [PubMed: 15805601]
18. Storoni LC, McCoy AJ, Read RJ. Likelihood-enhanced fast rotation functions. *Acta Crystallogr, Sect D: Biol Crystallogr* 2004;60:432–438. [PubMed: 14993666]
19. Adams PD, Gopal K, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Pai RK, Read RJ, Romo TD, Sacchettin JC, Sauter NK, Storoni LC, Terwilliger TC. Recent developments in the PHENIX software for automated crystallographic structure determination. *J Synchrotron Radiat* 2004;11:53–55. [PubMed: 14646133]
20. Davis IW, Murray LW, Richardson JS, Richardson DC. MolProbity: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res* 2004;32:W615–W619. [PubMed: 15215462]
21. Murshudov GN, Vagin AA, Lebedev A, Wilson KS, Dodson EJ. Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr, Sect D: Biol Crystallogr* 1999;55:247–255. [PubMed: 10089417]

22. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr, Sect D: Biol Crystallogr* 2004;60:2126–2132.
23. Schuttelkopf AW, van Aalten DMF. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr, Sect D: Biol Crystallogr* 2004;60:1355–1363. [PubMed: 15272157]
24. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook JD, Zardecki C. The Protein Data Bank. *Acta Crystallogr, Sect D: Biol Crystallogr* 2002;58:899–907. [PubMed: 12037327]
25. Schulbach MC, Brennan PJ, Crick DC. Identification of a short (C₁₅) chain Z-isoprenyl diphosphate synthase and a homologous long (C₅₀) chain isoprenyl diphosphate synthase in *Mycobacterium tuberculosis*. *J Biol Chem* 2000;275:22876–22881. [PubMed: 10816587]
26. Dunphy PJ, Kerr JD, Pennock JF, Whittle KJ, Feeney J. The plurality of long chain isoprenoid alcohols (polyprenols) from natural sources. *Biochem Biophys Acta* 1967;136:136–147. [PubMed: 6032033]

1 80

Rv1086 MEII PPRLKEPLYR LYELRLRQGL AASKSDLPRH IAVLCDGNRR

Rv2361c MARDARKRTS SNFPQLPPAP DDYPTFPDTS TWPVVFPELP AAPYGGPCRP PQHTSKAAAP RIPADRLPNH VATVMDGNCR

M. luteus MFP IKKRKAIKNN NINAAQIPKH IATIMDGNCR

E. coli M MLSATQPLSE KLPAHGC-RH VATIMDGNCR

81 160

Rv1086 WARSAGYDDV SYGYRMCAAK TAEMLRWCHF AGIELATVYL LSTENLQDLP DELAALIEII TDVVEEICAP ANHWSVR--T

Rv2361c WATQRGLART -EGHKMGEAV VIDIAACGAIE LGIKWLSLYA FSTENWKRSF EEVRRLMGFN RDVVRRRRDT LKKLGVRIRW

M. luteus WAKQKKMPRI -KGHYEGMOT VKKITRYASD LGVKYLTLYA FSTENWSRPF DEVNYLMKLP GDFLNTFLPE LIEKNVKVET

E. coli WAKKQGKIRA -FGHKAGAKS VRRAVSFAAN NGIEALTLYA FSSNNWRPA QEVSAIMELF VWALDSEVKS LHRHNVRLRI

161 240

Rv1086 VGDGLGIGEE PARRLRGAVE STPEVASFHV NVAVGYGGR EIVDAVRALL SKELANGATA EELVDAVTVE GISENLYTSG

Rv2361c VGSRLRLWRS VINELAVAE MTKSNDVITI NYCVMYGGRT EITEATREIA REVAAGRLNP ER----ITES TIARHLQRPD

M. luteus IGFIDDLDPH TKKAVLEAKE KTKHNTGLTL VFALNYGGRK EIIISAVQLIA ERYKSGEISL DE----ISET HFNEYLFTAN

E. coli IGDTSRFNSR LQERIRKSEA LTAGNTGLTL NIAANYGGRW DIVQGVRLA EKVVQGNLQP DO----IDEE MLNQHVCMHE

241 308

Rv1086 QPDFDLVIRT SGEQRLSGFL IWQSAYSEMW FTEAHWPAFR HVDFLRALRD YSARHRSYGR

Rv2361c IPDVDFLIRT SGEQRSSNFM LWQAAYAEYI FQDKLWPDYD RRDWLWACEE YASRTRRFGS A

M. luteus MPDFELLIRT SGEERLSNFL IWQCSYSEFV FIDEFWPDFN EESLAQCISI YQNRHRRFGG L

E. coli LAPVDLVIRT GGEHRISNFL LWQIAYAEFY FTDVLWPDFD EQDFEGALNA FANRRRFGG TEPGDETA

Figure 1. Rv1086 and Rv2361c with the reactions they catalyze

(a) Multiple sequence alignment of open reading frames Rv1086 and Rv2361c from *Mycobacterium tuberculosis* H37Rv with undecaprenyl diphosphate synthases from *Micrococcus luteus* (SWISS-PROT O82827) and *Escherichia coli* (SWISS-PROT Q47675). The alignment was generated using the Multalin interface. The N-terminal extension of 2361c is evident.

(b) The molecules discussed in the paper. The product of Rv1086, EZ-FPP, is the preferred substrate for Rv2361c, we employed EE-FPP in our crystallographic studies to mimic this molecule. CITPP is an inhibitor of both enzymes and is mimic of prenyl diphosphates. The phosphates are labeled as α and β to distinguish them.

(c) Schematic diagram of the reaction catalyzed by Z-prenyl synthases. In Rv1086 is not processive as it stops after one cycle; Rv2361c continues for a further 7 cycles to make a 50 carbon product.

Figure 2. Structural biology of Rv1086 and Rv2361c

(a) The structure of Rv1086 dimer is shown as a cartoon, the EE-FPP ligand is shown as spheres (oxygen red, phosphorous orange, carbon yellow).

(b) The structure of Rv2361c dimer is shown as a cartoon, the CITPP (both subunits) and IPP ligand (one subunit). The ligands are derived from separate structures and are positioned based on a superposition of the Ca protein atoms. The ligand atoms shown as spheres (oxygen red, phosphorous orange, carbon yellow). The bound Mg²⁺ ion is shown as a grey sphere. The well ordered N-terminal module unique to Rv2361c is shown circled by a dotted line. In Rv2361c

the C-terminus (pink sphere) from subunit A (green) reaches into the active site of subunit B (teal).

(c) Stereo diagram of the interaction between CITPP, IPP and Rv2361c. The ligands are derived from separate structures are positioned based on a superposition of the C α protein atoms. Nitrogen atoms are colored blue, oxygen atoms colored red, protein carbon atoms are colored salmon, except for the three residues (G294, S295 and A296 denoted by *) from the other monomer where the carbon atoms are colored purple. Active site residues from the CITPP Rv2361c complex are shown and labeled. The octahedral co-ordination of the Mg²⁺ ion (colored green) is highlighted by dashed black lines; water molecules shown as red balls. The contact between the ND1 atom N124 and C2 atom of IPP is shown as a black dashed line. N124 has been proposed to act as the catalytic base. The active site residues are conserved in Rv1086.

Figure 3. The crucial role of L84 in limiting chain length in Rv1086

(a) Superposition of the active site of EE-FPP Rv1086 complex (protein carbon atoms are colored white, (all other atom are colored as in Figure 2b) and Rv2361c (protein carbon atoms colored salmon). The double mutant Rv1086_{L84A, L85F} can be expressed in a soluble (folded) form. L85 in Rv1086 points away from product. We suggest the L85F mutation role assists only in folding protein.

(b) Space filling view of the side chains that limit the growing polymer in Rv1086. EE-FPP is shown as yellow space filling spheres. It can be seen that the side chain of L84 covers the polymer. Carbon atoms are in grey for the protein and yellow for the ligand.

Figure 4. TLC analysis of [¹⁴C]IPP radiolabeled products synthesized by wildtype Rv1086 or Rv1086_{L84A, L85F}

(a) TLC analysis of products generated by wildtype Rv1086 in the presence of [¹⁴C]IPP and GPP. Radioactive compounds were located using a Bioscan System 2000 imager. There is only the C₁₅ product. (b) TLC analysis of products generated by Rv1086_{L84A, L85F} in the presence of [¹⁴C]IPP and GPP. Radioactive compounds were located using a Bioscan System 2000 imager, the presence of the C₂₀ product is visible. (c) TLC analysis of products generated by Rv1086_{L84A, L85F} (Lane 1) and Rv2361c (Lane 2) in the presence of [¹⁴C]IPP and EE-FPP. Radioactive compounds were located using autoradiography. Prenyl diphosphate synthase activity was assayed in mixtures containing 50 mM MOPS (pH 7.9), 10 mM sodium orthovanadate, 1 mM MgCl₂, 5.0 mM dithiothreitol, 0.3% Triton X-100, 30 μM [¹⁴C]IPP, and 10 μg protein in a final volume of 50 μl. The allylic diphosphate in the reactions were used at 100 μM. Reactions were incubated for 30 min and stopped by addition of 1 ml water saturated with NaCl and extracted with n-butanol saturated with water. The extracted prenyl diphosphates were dephosphorylated with potato acid phosphatase as described in Experimental Procedures. Equal amounts of radioactivity were spotted on reverse-phase TLC plates. The plates were developed in methanol:acetone (8:2, v/v) and radioactive compounds were visualized by autoradiography. Standard polyprenols were located with anisaldehyde spray reagent²⁶. The chain-lengths of the products are indicated with arrows.

Figure 5. The effect of substrate concentration on the rate of product formation by Rv1086_{L84A, L85F}

Reaction mixtures contained 50 mM MOPS (pH 7.9), 5.0 mM dithiothreitol, 10 mM sodium orthovanadate, 1 mM MgCl₂, 0.3% Triton X-100, 10 μg protein and the indicated amount of GPP in a final volume of 50 μl. The reactions were incubated at 37°C for 30 min and stopped with 1ml of water saturated with NaCl, extracted with 1 ml of butanol saturated with water, and an aliquot was taken for liquid scintillation spectrometry. The K_m (inset) was calculated using nonlinear regression analysis (GraFit 5.0.13).

Table 1**Primers used in this study**

Primer	Sequence
Rv1086UP [*]	CAT ATG GAG ATC ATC CCG CCG CGG CTC A
Rv1086RP [*]	AAG CTT TCA CCT GCC GTA GCT GCG ATG CCT
Rv1086 _{L84A} UP [†]	CGT CTA TGC TCT GTC CAC CGA AAA CCT GC
Rv1086 _{L84A} RP [†]	TGG ACA GAG CAT AGA CGG TGG CCA GTT
Rv1086 _{L84A,L85F} UP [†]	ACC GTC TAT GCT TTC TCC ACC GAA AAC CTG C
Rv1086 _{L84A,L85F} RP [†]	TTC GGT GGA GAA AGC ATA GAC GGT GGC CAG TTC

* Genomic DNA was used as the template

[†] Plasmid containing *Rv1086* was used as the template

Table 2

X-ray data

		Rv2361		Rv1086	
Data collection		CTTPP*	IPP	CTTPP	EE-FPP
λ (Å)		0.9762		0.9330	0.976
Resolution		75 - 1.8	75 - 1.95	27 - 1.7	60 - 1.7
Last shell (Å)		1.85- 1.8	2.0 - 1.95	1.8 - 1.7	1.7 - 1.7
Spacegroup		P32	P32	P2 ₁ -2 ₁ -2 ₁	P2 ₁ -2 ₁ -2 ₁
Cell (Å)		a = 86.9 c = 184.1	a = 87.0 c = 183.7	a = 58.5 b = 73.5 c = 102.6	a = 58.8 b = 74.1 c = 103.3
Unique refl's		134070	111969	46922	50127
Average redundancy		2.8	3.5	7.5	3.9
I/ σ		12.8 (2.0)	12.1 (1.8)	24.5 (6.6)	15.7 (2.4)
Complete (%)		93 (91.7)	98(95)	100 (100)	99.6(100)
R _{merge}		0.074(0.42)	0.094(0.47)	0.055 (0.48)	0.059(0.52)
Refinement					
R %		20.4(24.8)	18.7(22.9)	17.6(16.2)	18.8(24.4)
R _{free} %		23.1(29.0)	21.1(27.6)	21.5(23.5)	22.2(32.3)
rmsd bonds (Å) /angles (°)		0.010 /1.3	0.009 /1.2	0.010 / 1.2	0.009 / 1.1
Ramach'n Core (%)		94	93	95	96
No of protein chains in asymmetric unit		4	4	2	2
PDB accession code		2vg3	2vg2	2vg0	2vg1
* These data were twinned, for the APO the twin fraction was refined for CTTPP the data were detwinned. The twinned data have been deposited as well as the de-twinned.					
† The completeness was limited by radiation damage. Thus the resolution limits for these data sets are nominal.					

Table 3

Summary of protein activity

Construct	Protein	Mutations	Expression	Purification	Activity	Product chain length	
						GPP	Substrate
Rv1086	Rv1086		Yes	Yes	Yes	C ₁₅	EE-FPP
Rv1086 ^{L84A}	Rv1086	L84A	No	N/A	N/A		None
Rv1086 ^{L84A,L85F}	Rv1086	L84A, L85F	Yes	Yes	Yes	C ₁₅ and C ₂₀	C ₂₀ -C ₅₀