Identification of amino acids and domains required for catalytic activity of DPPR synthase, a cell wall biosynthetic enzyme of *Mycobacterium tuberculosis*

Hairong Huang,^{1,2} Stefan Berg,^{1†} John S. Spencer,¹ Danny Vereecke,³ Wim D'Haeze,^{3,4} Marcelle Holsters⁴ and Michael R. McNeil¹

Correspondence

Michael R. McNeil

mmcneil@colostate.edu

¹Department of Microbiology, Immunology, and Pathology, Colorado State University, Fort Collins, CO 80523, USA

²Beijing Tuberculosis and Thoracic Tumor Institute, Beijing 101149, China

³Department of Plant Systems Biology, Flanders Interuniversity Institute for Biotechnology, Ghent University, Technologiepark 927, B-9052 Ghent, Belgium

⁴University of Georgia, Complex Carbohydrate Research Center, 315 Riverbend Road, Athens, GA 30602, USA

Decaprenylphosphoryl-D-arabinose (DPA) has been shown to be the donor of the essential D-arabinofuranosyl residues found in the cell wall of *Mycobacterium tuberculosis*. DPA is formed from phosphoribose diphosphate in a four-step process. The first step is the nucleophilic replacement of the diphosphate group with decaprenyl phosphate. This reaction is catalysed by the integral membrane protein 5-phospho- α -D-ribose-1-diphosphate : decaprenyl-phosphate 5-phosphoribosyltransferase (DPPR synthase). The enzyme is essential for growth and thereby an important target candidate for the development of new tuberculosis drugs. Although membrane proteins are an important subset of targets for current antibacterial agents, details about the structures and the active sites of such proteins are often not readily available by X-ray crystallography. To begin a different approach to the issue, homologues from *Mycobacterium smegmatis* and *Corynebacterium glutamicum* were expressed in *Escherichia coli* and shown to be active DPPR synthases. This was followed by bioinformatic analyses of the aligned sequences and then by site-directed mutagenesis of amino acids identified as likely to be important for activity. The results suggested that the enzymic synthesis of decaprenyl-phosphate 5-phosphoribose (DPPR) occurs on the cytoplasmic side of the plasma membrane. Amino acid substitutions showed that the predicted cytoplasmic N-terminal region and two cytoplasmic loops are involved in substrate binding and/or catalysis along with parts of some adjoining inner membrane regions. The enzyme lacks the classical phosphoribose diphosphate (pRpp) binding site found in nucleic acid precursor enzymes of both prokaryotes and eukaryotes but instead contains a conserved NDxxD motif required for enzymic activity. Thus, it is plausible that this DPPR synthase has a pRpp binding site that is different from that of the classical eukaryotic enzymes, and further work to develop inhibitors against this enzyme is thereby encouraged.

Received 25 September 2007

Revised 13 November 2007

Accepted 15 November 2007

INTRODUCTION

The mycobacterial covalently attached cell wall core is made up of two layers, the peptidoglycan layer and a lipid

mycolic acid layer, tethered together by the polysaccharide arabinogalactan. The arabinan region of this polysaccharide is composed of D-arabinofuranosyl residues, which are rarely found in nature. The donor of the D-arabinofuranosyl residues is not a sugar nucleotide but rather decaprenylphosphoryl-D-arabinose (DPA) (Wolucka *et al.*, 1994; Xin *et al.*, 1997; Scherman *et al.*, 1996, 1995). For that reason the biosynthesis of DPA has been studied (Huang *et al.*, 2005; Mikusova *et al.*, 2005; Xin *et al.*, 1997; Scherman *et al.*, 1996), yielding the information that it is

[†]Present address: TB Research Group, Veterinary Laboratories Agency, Weybridge, New Haw, Addlestone, Surrey KT15 3NB, UK.

Abbreviations: CL, cytoplasmic loop; DP, decaprenyl phosphate; DPA, decaprenylphosphoryl-D-arabinose; DPPR, decaprenylphosphoryl-D-5-phosphoribose; pRpp, phosphoribose diphosphate; p[¹⁴C]Rpp, phospho[¹⁴C]ribosyl diphosphate; TM, transmembrane.

synthesized in four steps (Mikusova *et al.*, 2005). The first step of the synthesis is the formation of decaprenylphosphoryl- β -D-5-phosphoribose (DPPR) from decaprenyl phosphate (DP) and phosphoribose diphosphate (pRpp) (Fig. 1). The enzyme responsible for this step, 5-phospho- α -D-ribose-1-diphosphate:decaprenyl-phosphate 5-phosphoribosyltransferase (DPPR synthase), in *Mycobacterium tuberculosis* has recently been cloned and identified (Huang *et al.*, 2005). This enzyme has the potential to be a good drug target since it catalyses a reaction not found in humans. In this study the amino acid sequence of DPPR synthases from several species of *Actinomycetales* and also from an *Azorhizobium* strain were compared and their membrane-spanning domains were predicted. Site-directed mutagenesis of amino acids hypothesized to be involved in substrate binding and/or catalysis was then performed. The results led to the identification of specific domains and amino acids required for substrate binding and catalytic activity.

METHODS

Reagents. All reagents were from Sigma-Aldrich Chemical unless indicated otherwise. pRpp (MP Biomedicals or Sigma) was purified by ion-exchange HPLC as described elsewhere for radiolabelled pRpp (Scherman *et al.*, 1996).

Cloning and expression of genes encoding DPPR synthase. To search for homologues of the DPPR synthase of *M. tuberculosis*, the amino acid sequence of Rv3806c was used as a query in a homology search of the TIGR database (<http://tigrblast.tigr.org/cmr-blast/>). The

forward primers used to amplify the DPPR synthase genes of *M. tuberculosis* (Rv3806c), *Mycobacterium smegmatis* (MSMEG_6401), *Corynebacterium glutamicum* (cg3189) and *Azorhizobium caulinodans* (noeC) (Mergaert *et al.*, 1996) included an *NdeI* restriction site (underlined) while the reverse primers had a *BamHI* restriction site (underlined) downstream of the stop codon. The primer sequences were: *M. tuberculosis* (Fw) 5'-CATATGAGTGAAGATGTGGTGAC-TCAACC-3', (Rev) 5'-GGATCCCTAGCCGAAGGCAACAGCGGC-3'; *M. smegmatis* (Fw) 5'-CATATGGATGCCACCCACATGAGTG-3', (Rev) 5'-GGATCCTCAGCTGAAATAGATGGCCGC-3'; *C. glutamicum* (Fw) 5'-CATATGAGCGAACACGCCGCTGAAC-3', (Rev) 5'-GGATCCTCAAAACATCGGCATGATGTAC-3'; *A. caulinodans* (Fw) 5'-CATATGTGGAATAAGGAATGGCCCG-3', (Rev) 5'-GGATCC-TCAACCAATCGATTGCGACCG-3'.

Vent DNA polymerase (New England Biolabs) was used in the PCR reactions. Purified DNA fragments were cloned into pSTBlue-1 plasmid (blunt vector) using the Perfectly Blunt cloning kit (Novagen). DH5 α competent cells (Invitrogen) were then transformed with 1 μ l of the ligation mixture. Single colonies were isolated and plasmids carrying the correct insert were identified by DNA sequencing. The genes were then cloned into pET16b, in-frame with the coding sequence of the histidine tag at the 5' end, by using the restriction sites *NdeI* and *BamHI*. The resulting expression vectors were transformed into *Escherichia coli* ER2566 (New England Biolabs) and protein expression was induced by the addition of 0.4 mM IPTG followed by incubation at room temperature for 4 h. The cells were harvested by centrifugation and resuspended in buffer A [50 mM MOPS (pH 8.0), 10 mM MgCl₂, 5 mM β -mercaptoethanol]. The cell suspension was disrupted either by probe sonication on ice with a Sanyo Soniprep 150 (10 cycles of 60 s on and 90 s off) or by French pressure cell FC020K (Cell Scientific) (two presses at 12 600 p.s.i.=8687 kPa). The suspension was centrifuged at 20 000 *g* for 30 min, to generate supernatant and pellet. The supernatant was then ultracentrifuged at 100 000 *g* for 1 h, and the resulting membrane-enriched pellet was removed and homogenized in a small volume of buffer A to a concentration of approximately 2 mg protein ml⁻¹, as determined by Coomassie Plus Protein Assay reagent (Pierce).

Site-directed mutagenesis. Mutagenesis was accomplished by overlapping extension as described by Ho *et al.* (1989), using two internal overlapping primers (one forward and one reverse with the same desired mutation) in combination with forward and reverse primers (as above). The resulting PCR products, with specific pre-determined mutations, were cloned into the plasmid pSTBlue-1 using the Perfectly Blunt cloning kit, and the presence of the correct mutations was confirmed by sequencing. Subsequently, the genes modified by site-directed mutagenesis were cloned into pET16b and transformed into and expressed in *E. coli* ER2566, as described above.

SDS-PAGE and immunodetection. Purified cell membranes were incubated with loading buffer for 5 min at 45 °C (rather than boiled), as is required to visualize the DPPR synthase protein (Huang *et al.*, 2005). Proteins were separated by 12 % SDS-PAGE and visualized by using GelCode Blue Stain reagent (Pierce). For immunodetection, the proteins were transferred to PVDF membranes, and monoclonal anti-polyhistidine (mouse IgG2a isotype) and conjugated anti-mouse-IgG-alkaline phosphatase were used as primary and secondary antibodies, respectively.

DPPR synthase assay. The radiolabelled substrate, phospho[¹⁴C]ribosyl diphosphate (p[¹⁴C]Rpp), was prepared enzymically from [¹⁴C]glucose (American Radiolabelled Chemicals) as described by Scherman *et al.* (1996). DP (Indofine Chemical) was dissolved in 7.5 % (w/v) CHAPS to a concentration of 1 mg DP ml⁻¹. *M. smegmatis* mc²155 membranes (Scherman *et al.*, 1996) were used as the positive control and *E. coli* membranes purified from ER2566,

Fig. 1. The pathway for DPA biosynthesis. The reaction catalysed by DPPR synthase and the subsequent reactions that form DPA, the donor of arabinosyl residues in the cell wall of *M. tuberculosis*, are shown.

which harboured empty pET16 vector, were used as the negative control. A standard reaction mixture of 30 μ l contained 0.18 nmol [45 nCi (1665 Bq)] HPLC-purified p[14 C]Rpp, 4 μ g DP, 5 μ g membrane protein and 0.25 % (w/v) CHAPS in buffer A. After a 15 min incubation at 37 °C, the enzymic reaction was stopped by adding chloroform and methanol to give a final chloroform/methanol/water ratio of 8:4:3 (by vol.). After centrifugation to generate two phases, an aliquot of the lower organic layer was then dried and radioactivity was measured on a Trilux MicroBeta scintillation counter (Perkin-Elmer/Wallac). Periodically, to confirm the identity of the product, TLC analysis was performed. In this case, the remaining organic phase was evaporated to dryness and the pellet was resuspended in chloroform/methanol/1 M ammonium hydroxide [65:25:4 (by vol.)]. Samples were spotted on a Silica gel 60 F₂₅₄ TLC plate (Merck) and developed in chloroform/methanol/water/concentrated ammonium hydroxide/1 M ammonium acetate (180:140:23:9:9, by vol.). Autoradiography was carried out by exposing the TLC plate to X-OMAT AR film (Kodak) for 2 days at -80 °C. For the determination of K_m values for pRpp and DP, the reaction time was reduced to 5 min and various concentrations of non-radioactive pRpp and/or DP were used. After separation of the organic layer from the aqueous layer by centrifugation, the radioactivity in a known volume of the organic phase was determined by liquid scintillation counting (as above). The data were then analysed by non-linear regression using the program Grafit 5.0 (Leatherbarrow, 2001).

RESULTS

Identification of DPPR synthases

In a search of the TIGR database (see Methods), the query sequence of the previously characterized DPPR synthase (Rv3806c) of *M. tuberculosis* (Huang *et al.*, 2005) hit one strong homologue in each of the complete sequenced genomes of mycobacteria, corynebacteria and nocardiae (data not shown). This was the expected result due to the presence of arabinose in their cell walls (Dover *et al.*, 2004). A BLAST search against the human database at NCBI showed the absence of any significant homologue to Rv3806c. The DPPR synthase candidates from *M. smegmatis*, *C. glutamicum* and *A. caulinodans* were selected, cloned and expressed in *E. coli*, essentially as done previously for DPPR synthase from *M. tuberculosis* (Huang *et al.*, 2005). Western blot analysis using anti-His₆-antibody showed that all these orthologues were expressed in the plasma membrane fraction (data not shown). The DPPR synthases of *M. smegmatis* and *C. glutamicum* were also shown to be capable of converting pRpp and DP to DPPR, as shown by a radioactive assay (Fig. 2a) and by confirmation of the identity of the products by TLC (Fig. 2b). However, the putative DPPR synthase from *A. caulinodans*, with 22 % amino acid identity to the *M. tuberculosis* homologue, showed no enzymic activity (Fig. 2a) despite being successfully expressed (see Discussion).

Sequence and topology analysis of the DPPR synthases

The amino acid sequences of the DPPR synthases of *M. tuberculosis*, *M. smegmatis* and *C. glutamicum* along with

Fig. 2. Assay of DPPR synthase activities of the putative DPPR synthases from *M. smegmatis*, *C. glutamicum* and *A. caulinodans* expressed in *E. coli*. (a) Conversion of radioactively labelled substrate from water-soluble form (i.e. [14 C]-Rpp) to organic-soluble form (i.e. DPPR[14 C]-R) by membranes isolated from *E. coli* expressing the DPPR synthase from *M. tuberculosis* (*M. tb.*), *M. smegmatis* (*M. smeg.*), *C. glutamicum* (*C. glu.*) and *A. caulinodans* (*A. caul.*) and vector control. (b) TLC analysis of the radioactive products produced by the DPPR synthases from *M. tuberculosis*, *M. smegmatis* and *C. glutamicum* confirming their identity as DPPR. The R_F of the reference lipid (DPPR) is indicated by an arrow.

the related NoeC protein from *A. caulinodans* were aligned as shown in Fig. 3. In addition, the membrane-spanning regions for each of these proteins were predicted by Phobius (Kall *et al.*, 2004) and are included in the figure. The topology pattern was very similar among the four proteins, but the first transmembrane (TM) domain of NoeC was not predicted. These proteins are obviously integral membrane proteins, largely embedded in the plasma membrane, but they carry several larger soluble domains i.e. the N terminus and three cytoplasmic loops (CL-I, CL-II, CL-IV and CL-V in Fig. 3). No regions other than small connecting peptides were predicted to be located on the extracellular side of the membrane. Among the four aligned homologues in Fig. 3, 54 invariant amino acids were found and they tended to cluster in three regions. These regions are the cytoplasmic N terminus and the beginning of TM domain 1 (TM-1), several residues of TM-2 and the subsequent loop CL-II, and parts of TM-6 and CL-IV. The last approximately 50–70 amino acids in the C-terminal region, including CL-V, showed lower

Fig. 3. Amino acid sequence alignment and analysis. DPPR synthase of *M. tuberculosis* (mtDPPRS) was aligned by CLUSTAL W (Chenna *et al.*, 2003) with its homologues of *M. smegmatis* (msDPPRS), *C. glutamicum* (cgDPPRS) and *A. caulinodans* (acNoeC). Strictly conserved residues are in bold type. Amino acids underlined with solid lines indicate a TM domain predicted by Phobius (Kall *et al.*, 2004) and are labelled as TM followed by an Arabic number. Underlining with dashed lines signifies that amino acids of that sequence region are located on the cytosolic side of the plasma membrane, based on topology prediction, and these loops are labelled as CL (cytoplasmic loop) followed by a Roman numeral. The position of the last residue on each line of the alignment is indicated for each protein. The arrows indicate the residues selected for site-directed mutagenesis. See Tables 1 and 2 for further information.

sequence conservation than the other regions of these four proteins.

The DPPR synthase homologues of all *Actinomycetales* species investigated contain an NDxxD motif located in CL-II, beginning, in the case of the *M. tuberculosis* enzyme, at amino acid 73 (Fig. 3). This motif, in closely related forms (where N can be substituted for D or E), is found in UbiA phenyltransferases (Hemmi *et al.*, 2004) and *E*-prenyl diphosphate synthases (Wang & Ohnuma, 1999). It is also found in prenyl diphosphate transferases (Brauer *et al.*, 2004; Saiki *et al.*, 1993) such as CyoE, which prenylates haem O (Saiki *et al.*, 1993). Finally it is found in phosphoribosyl transferases (Craig & Eakin, 2000). In all these instances a pyrophosphate group is present on one of the substrates. Indeed, in the *E*-prenyl diphosphate synthases two of these motifs are present (Wang & Ohnuma, 1999), presumably one for each diphosphate substrate (isopentyl diphosphate and the growing allylic diphosphate). In the case of the substrates for DPPR synthase, the diphosphate-leaving group is part of pRpp while the DP isoprene is present as a monophosphate

rather than diphosphate. Thus, the NDxxD motif is most likely to be important for pRpp binding and subsequent dissociation of the PP_i instead of (or perhaps in addition to) DP binding.

We also searched the sequences of the DPPR synthases for homology to pRpp and DP binding sites that have been described in enzymes known to use these substrates. However, using the conserved pRpp binding motif of pyrR of *M. tuberculosis* (Kantardjieff *et al.*, 2005), no corresponding motif was found among the DPPR synthase homologues. In a search for a binding region for DP, three polyisoprene-recognition sequences from NeuE, Dpm1 and Alg7 (Zhou & Troy, 2005) were aligned with the DPPR synthases. Interestingly, a region of Alg7 (L⁸¹FVMFIYIPFIFY), weakly aligned with residues 58–70 of *M. tuberculosis* DPPR synthase, with the first two phenylalanine residues (F59 and F62) and the second tyrosine residue (Y70) being present in both Alg7 and all four sequences studied here (Fig. 3). This sequence analysis of the DPPR synthase homologues gave us clues about where functional sites may occur, since amino acid

residues/motifs directly involved in specific functions are most often well conserved.

Site-directed mutations in *M. tuberculosis* DPPR synthase

Invariant amino acids can be found along nearly the entire alignment (Fig. 3), but the three main clusters of conserved residues (in CL-I, CL-II and CL-IV), including the NDxxD motif discussed above, stand out. The topology prediction suggested that these clusters are predominantly found in the cytoplasmic domains or as part of the neighbouring TM domains, and led us to believe that many of these conserved residues may form functional site(s) in the native structure of the protein. Thus, selected amino acids from these three clusters were changed by site-directed mutagenesis in the *M. tuberculosis* DPPR synthase and the enzymic activity for each point-mutated version was investigated (see Methods). The results are shown in Table 1. In all cases but two (D77A and A66F) the mutated protein was expressed in the membrane fraction, in a fashion indistinguishable from that of the wild-type enzyme, as shown in Fig. 4.

Most point mutations introduced in the DPPR synthase had a profound effect on the enzymic activity, suggesting the successful prediction by our bioinformatic analysis of

Fig. 4. (a) Immunodetection, using anti-His antibody, of the DPPR synthase of *M. tuberculosis* and four mutants that retained detectable activity. Equal amounts of membrane proteins (5 µg) were loaded to each lane. Lanes: 1, wild-type DPPR synthase; 2, E195A; 3, R201A; 4, R22L; 5, N29A; 6, protein standard. (b) Immunodetection of the DPPR synthase A66F. Lanes: 1, membrane fraction; 2, soluble fraction; 3, protein standard. (c) Immunodetection of two DPPR synthase mutants that were completely inactive. Lanes: 1, D74A; 2, R192A; 3, protein standard. Very similar results were obtained from Western blot experiments of the remaining DPPR synthase mutants prepared in this study, except for mutant D77A, where DPPR synthase was expressed primarily in inclusion bodies.

Table 1. Site-directed mutagenesis of DPPR synthase of *M. tuberculosis*

Position	Codon change	Amino acid exchange	Activity (% of wild type)*
Wild-type	None	None	100
22	CGC to CTG	R to L	10
29	AAC to GCC	N to A	18
59	TTC to GCC	F to A	94
62	TTC to GCC	F to A	72
66	GCC to TTC	A to F	0†
70	TAC to GCC	Y to A	0
73	AAC to GCC	N to A	0
73	AAC to CAG	N to Q	15
74	GAT to GCC	D to A	0
77	GAC to GCC	D to A	0‡
77	GAC to GAG	D to E	15
81	GAC to GCC	D to A	0
84	CAC to CTG	H to L	15
192	CGC to GCC	R to A	0
195	GAG to GCC	E to A	17
201	CGC to GCC	R to A	1

*Under standard assay conditions; performed in duplicate on two separate clones after transformation with plasmid carrying gene with mutation.

†Expressed (His-tag-positive) protein was detected in the soluble but not in the membrane fraction for this mutant.

‡Protein was only weakly present in the membrane fraction (mostly in the insoluble fraction).

residues/domains involved in its function. Beginning with the cluster in the N-terminal region, two amino acids were exchanged, R22L and N29A. In both cases, the activity was strongly diminished, although neither seemed to be essential. CL-II and its upstream region (residues 58–93) contain many invariant residues, including the NDxxD motif. Starting with this motif, the substitutions of N73 or D74 to alanine resulted in non-functional proteins. However, slight residual activity could be restored when N73 and D77 were exchanged with amino acids with similar properties; N73Q and D77E gave both an activity of 15%, showing the importance of this conserved motif and its plausible involvement in catalysis. An additional mutant of this motif, D77A, was not expressed in the membrane and hence yielded no information. Several invariant residues are to be found just downstream of the NDxxD motif. As part of the same loop, substitutions D81A and H84L resulted in 0% and 15% retained activity, respectively, suggesting that the whole of CL-II is highly involved in the catalytic function.

It is reasonable to suggest that the TM region upstream of the NDxxD motif is part of a pocket for the lipid substrate DP. Point mutations in that highly conserved region would, in that case, affect the enzymic activity and binding of DP. Mutations F59A and F62A had only small effects on the activity, while Y70A led to a complete loss of activity, suggesting a critical role for this aromatic amino acid predicted to be located at the plasma membrane interface. The conversion A66F resulted in a protein that was not expressed in the membrane (Table 1).

The final conserved region among the DPPR synthases is found in domain TM-6 and the subsequent CL-IV (Fig. 3). It is of interest to note that the proteins of the MenA family (enzymes which add a polyisoprene unit to dihydroxynaphthoic acid) have a similar predicted membrane topology to that proposed here for the DPPR synthases, including a loop corresponding to CL-IV (Fig. 3). In the MenA family, this loop contains many conserved residues (alignment not shown). However, the sequence homology between CL-IV of the MenA family and the DPPR synthases is very low, suggesting a similar overall structure of the two protein families (DPPR synthase and MenA) but with diverse structural features in this loop due to recognition of different substrates. To determine the relevance of this loop in the DPPR synthases, three conserved residues in loop IV were changed (R192A, E195A and R201A). All three substitutions had a profound effect, with R192A and R201A more or less abolishing the activity, while E195A substantially reduced the activity. Such strong effects on activity from all three point mutations suggest that this loop makes a large contribution to the structure of a binding/active site.

Kinetic analysis of mutants

Kinetic analysis of the DPPR synthase was difficult due to the instability of pRpp and the low solubility of DP. Multiple analyses of the wild-type enzyme gave the K_m values reported in Table 2, albeit with relatively large standard deviations. Hence subtle changes in K_m were difficult to observe for mutants with the various amino acid changes. However, for several of the mutants,

significant changes in K_m were seen, as shown in Table 2. The exchange N29A gave a fourfold increase in K_m for pRpp while the K_m value for DP remained unaffected. Similarly, the mutation N73Q had a larger effect on the K_m value for pRpp than on that for DP. In the case of substitutions D77E and E195A, the effects were associated with both the substrates, as indicated by strong increase in K_m values for both pRpp and DP. None of the mutations that were part of the kinetic study had an effect on the K_m for DP without having an effect on the K_m for pRpp.

DISCUSSION

Identification of additional DPPR synthase enzymes

This study confirmed that the protein products of the *M. smegmatis* gene (MSMEG_6401) and *C. glutamicum* gene (cg3189), both of which showed strong homology to DPPR synthase of *M. tuberculosis* (Fig. 3), in fact show DPPR synthase activity. Surprisingly, NoeC of *A. caulinodans* was inactive under our assay conditions, even though the protein was expressed and detected in the *E. coli* membrane fraction. This was especially unexpected since the sequence of NoeC was initially used to identify the DPPR synthase gene in *M. tuberculosis* (Huang *et al.* 2005). The inactivity of the recombinant protein is supported by the fact that we have failed in efforts to show DPPR synthase activity in crude *A. caulinodans* membranes (data not presented). It is possible that *A. caulinodans* utilizes a slightly different substrate, such as 5-phosphoribose 1-phosphate or ribose 1-diphosphate instead of pRpp. Alternatively, the lack of enzymic activity of NoeC may reflect improper folding or insertion into the membrane. A region perhaps especially liable to not being inserted properly into the membrane is the first TM domain of the *A. caulinodans* homologue. Indeed in our computer topology study, this transmembrane region was not even predicted. We do think that this transmembrane region is very likely to occur in the natural context of *A. caulinodans*; otherwise the whole topology of the enzyme would be different and specific regions that we have now shown to be required for activity would be on different sides of the membrane. Hence it is a real possibility that the first TM domain does not insert properly into the membrane when the *A. caulinodans* protein is expressed in *E. coli*. Ultimately, further experimentation will be required to resolve the reason for the lack of activity of the *A. caulinodans* homologue.

Identification of amino acid residues required for substrate binding and/or catalysis

The two substrates of the DPPR synthase, the sugar donor pRpp and the acceptor lipid DP, are components of the cytosol and the plasma membrane, respectively. It is

Table 2. Kinetic parameters for selected mutations

Mutation	K_m pRpp (μ M) (SE)*	K_m DP (μ M) (SE)*
Wild-type	43† (17)	28 (6)
N29A	170 (17)	24 (8)
N73Q	$\geq 162\ddagger$	62 (20)
D77E	$\geq 162\ddagger$	≥ 207
E195A	$> 330\S$	183 (36)

*All analyses were repeated at least twice. The SE was determined by the Grafit 5.0 program (Leatherbarrow, 2001), which calculated the K_m values by linear regression. When K_m was greater than 80 % of the highest substrate concentration used, only a lower limit of the K_m could be determined, as indicated by the sign $\geq$.

†Substrate inhibition for pRpp was seen at 500 μ M for wild-type but no data above 150 μ M were used to calculate K_m . The K_m value of 43 μ M differs from the previously published value of 120 μ M (Huang *et al.*, 2005). The present study used pRpp purified by HPLC (Scherman *et al.*, 1996), while in our previous work it was less pure, leading to an incorrect calculation of the concentration of pRpp.

‡Substrate inhibition for pRpp at ~ 125 μ M made it possible to determine a minimum K_m value only.

$\S K_m$ was much higher than the highest amount of pRpp that could be added and substrate inhibition was not seen.

Fig. 5. Topology model. The nine predicted transmembrane domains, also shown in the alignment in Fig. 3, were used for modelling of DPPR synthase of *M. tuberculosis*. Amino acids subjected to site-directed mutagenesis in this study are labelled with their type and position number, and the effects on the activity of the mutations are indicated by different colours.

therefore logical to suggest that the catalytic reaction forming DPPR (Fig. 1) occurs at the membrane interface, and towards the cytosol. Here the topology prediction and the sequence analysis of the DPPR synthases both pointed in the same direction, with a polarization of conserved residues in loops facing the cytoplasm. Many of these residues are likely to be involved in the function of the enzyme. In addition, the results of the site-directed mutagenesis presented here strongly confirm this. As summarized in Table 1, we clearly have identified residues involved in substrate binding and/or catalysis. Not surprisingly, all conserved residues in the NDxxD motif were essential for activity, as were the neighbouring residues Y70 and D81, suggesting that CL-II carries key structural elements for the activity of this enzyme. This is also the case for CL-IV, since amino acid substitutions R192 and R201 both essentially abolished the activity of the enzyme. This suggests that amino acids from CL-II and CL-IV are likely to form the major structures of the active site. The predicted TM-2 and TM-6 domains are directly related to these loops and they carry several conserved residues. In this regard it seems possible that the polyisoprene chain of DP interacts with either or both of these TM domains. However, the amino acid substitutions F59A and F62A made in TM-2 failed to significantly affect the activity of the enzyme. This suggests that the two phenylalanines exchanged here are not involved in a critical hydrophobic interaction. The kinetic analysis (Table 2) showed that N29 is specific for pRpp binding as its K_m increased dramatically when this residue was mutated to an alanine while the K_m for DP was not affected. In contrast, the K_m for pRpp and DP both increased in the mutants N73Q, D77E and E195A.

All proposed mechanisms for reactions related to that of DPPR synthase involve nucleophilic substitution (either S_N1 or S_N2). In both instances it is likely that both substrates need to be bound to the enzyme simultaneously. In the process of obtaining K_m values we consistently noted that higher concentrations of pRpp inhibited the formation of DPPR, both for the wild-type enzyme and for all of the

mutants except those with very high K_m values for pRpp. A likely explanation for the pRpp inhibition is that DP must bind to the enzyme before pRpp. Therefore DP binding is inhibited at higher pRpp concentrations, where pRpp binds to the apoenzyme and thereby inhibits the binding of DP and catalytic activity.

In total, the data presented in this study revealed an active site with contribution from residues of the cytoplasmic N terminus, part of TM-2, and CL-II and CL-IV (Fig. 5). The N-terminal region is clearly involved in pRpp binding, given the effect of N29A on its K_m . The NDxxD motif in loop II is essential for activity and, by homology to analogous enzymes, is likely to be involved in stabilizing the loss of PPi from pRpp. The perturbation D77E had a profound effect on the K_m values for both pRpp and DP, suggesting fundamental contributions to the active site. In a similar fashion, all three residues mutated in CL-IV had profound effects on the enzymic activity; kinetic studies on the E195A mutant protein revealed that this region, much like the NDxxD motif, is fundamentally involved in the active site, including substrate binding. These results are a step towards the design of effective inhibitors against this enzyme, which is a potential target for new tuberculosis drugs (Huang *et al.*, 2005). The lack of a classical pRpp binding site, as found in eukaryotic nucleic acid monomer synthetic enzymes, is encouraging in this regard.

ACKNOWLEDGEMENTS

We acknowledge and appreciate suggestions from Dr Peter Tonge regarding the analysis of the kinetic data. This work was supported by US Public Health Service Grant NIH-NIAID AI 33706 and R03TW006237

REFERENCES

- Brauer, L., Brandt, W. & Wessjohann, L. A. (2004). Modeling the *E. coli* 4-hydroxybenzoic acid oligoprenyltransferase (*ubiA* transferase) and characterization of potential active sites. *J Mol Model* **10**, 317–327.

- Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G. & Thompson, J. D. (2003). Multiple sequence alignment with the CLUSTAL series of programs. *Nucleic Acids Res* **31**, 3497–3500.
- Craig, S. P. & Eakin, A. E. (2000). Purine phosphoribosyltransferases. *J Biol Chem* **275**, 20231–20234.
- Dover, L. G., Cerdano-Tarraga, A. M., Pallen, M. J., Parkhill, J. & Besra, G. S. (2004). Comparative cell wall core biosynthesis in the mycolated pathogens, *Mycobacterium tuberculosis* and *Corynebacterium diphtheriae*. *FEMS Microbiol Rev* **28**, 225–250.
- Hemmi, H., Shibuya, K., Takahashi, Y., Nakayama, T. & Nishino, T. (2004). (S)-2,3-Di-O-geranylgeranylglycerol phosphate synthase from the thermoacidophilic Archaeon *Sulfolobus solfataricus* – molecular cloning and characterization of a membrane-intrinsic prenyltransferase involved in the biosynthesis of archaeal ether-linked membrane lipids. *J Biol Chem* **279**, 50197–50203.
- Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**, 51–59.
- Huang, H., Scherman, M. S., D'Haese, W., Vereecke, D., Holsters, M., Crick, D. C. & McNeil, M. R. (2005). Identification and active expression of the *M. tuberculosis* gene encoding, 5-phospho- α -D-ribose-1-diphosphate: decaprenyl-phosphate 5-phosphoribosyltransferase, the first enzyme committed to decaprenylphosphoryl-D-arabinose synthesis. *J Biol Chem* **280**, 24539–24543.
- Kall, L., Krogh, A. & Sonnhammer, E. L. (2004). A combined transmembrane topology and signal peptide prediction method. *J Mol Biol* **338**, 1027–1036.
- Kantardjiev, K. A., Vasquez, C., Castro, P., Warfel, N. M., Rho, B. S., Lakin, T., Kim, C. Y., Segelke, B. W., Terwilliger, T. C. & Rupp, B. (2005). Structure of *pyrR* (Rv1379) from *Mycobacterium tuberculosis*: a persistence gene and protein drug target. *Acta Crystallogr D Biol Crystallogr* **61**, 355–364.
- Leatherbarrow, R. J. (2001). *GraFit Version 5*, 5th edn. Horley, UK: Erithacus Software Ltd.
- Mergaert, P., D'Haese, W., Fernandez-Lopez, M., Geelen, D., Goethals, K., Prome, J. C., Van Montagu, M. & Holsters, M. (1996). Fucosylation and arabinosylation of Nod factors in *Azorhizobium caulinodans*: involvement of *nolK*, *nodZ* as well as *noeC* and/or downstream genes. *Mol Microbiol* **21**, 409–419.
- Mikusova, K., Huang, H., Yagi, T., Holsters, M., Vereecke, D., D'Haese, W., Scherman, M. S., Brennan, P. J., McNeil, M. R. & Crick, D. C. (2005). Decaprenylphosphoryl arabinofuranose, the donor of the D-arabinofuranosyl residues of mycobacterial arabinan, is formed via a two-step epimerization of decaprenylphosphoryl ribose. *J Bacteriol* **187**, 8020–8025.
- Saiki, K., Mogi, T., Hori, H., Tsubaki, M. & Anraku, Y. (1993). Identification of the functional domains in heme O synthase. Site-directed mutagenesis studies on the *cyoE* gene of the cytochrome *bo* operon in *Escherichia coli*. *J Biol Chem* **268**, 26927–26934.
- Scherman, M., Weston, A., Duncan, K., Whittington, A., Upton, R., Deng, L., Comber, R., Friedrich, J. D. & McNeil, M. (1995). The biosynthetic origin of the mycobacterial cell wall arabinosyl residues. *J Bacteriol* **177**, 7125–7130.
- Scherman, M. S., Kalbe-Bournonville, L., Bush, D., Xin, Y. & McNeil, M. (1996). Polyphosphorylphosphate-pentoses in mycobacteria are synthesized from phosphoribose pyrophosphate. *J Biol Chem* **271**, 29652–29658.
- Wang, K. & Ohnuma, S. (1999). Chain-length determination mechanism of isoprenyl diphosphate synthases and implications for molecular evolution. *Trends Biochem Sci* **24**, 445–451.
- Wolucka, B. A., McNeil, M. R., de Hoffmann, E., Chojnacki, T. & Brennan, P. J. (1994). Recognition of the lipid intermediate for arabinogalactan/arabinomannan biosynthesis and its relation to the mode of action of ethambutol on mycobacteria. *J Biol Chem* **269**, 23328–23335.
- Xin, Y., Lee, R. E., Scherman, M. S., Khoo, K. H., Besra, G. S., Brennan, P. J. & McNeil, M. (1997). Characterization of the in vitro synthesized arabinan of mycobacterial cell walls. *Biochim Biophys Acta* **1335**, 231–234.
- Zhou, G. P. & Troy, F. A. (2005). NMR study of the preferred membrane orientation of polyisoprenols (dolichol) and the impact of their complex with polyisoprenyl recognition sequence peptides on membrane structure. *Glycobiology* **15**, 347–359.

Edited by: W. Bitter