J Mol Model (2010) 16:1061–1073
DOI 10.1007/s00894-009-0615-x

ORIGINAL PAPER

Homology modeling of *Mycobacterium tuberculosis* 2C-methyl-D-erythritol-4-phosphate cytidylyltransferase, the third enzyme in the MEP pathway for isoprenoid biosynthesis

Cristian Obiol-Pardo · Alex Cordero · Jaime Rubio-Martinez · Santiago Imperial

Received: 15 July 2009 / Accepted: 6 October 2009 / Published online: 15 November 2009
© Springer-Verlag 2009

Abstract Tuberculosis is one of the leading infectious diseases in humans. Discovering new treatments for this disease is urgently required, especially in view of the emergence of multiple drug resistant organisms and to reduce the total duration of current treatments. The synthesis of isoprenoids in *Mycobacterium tuberculosis* has been reported as an interesting pathway to target, and particular attention has been focused on the methylerythritol phosphate (MEP) pathway comprising the early steps of isoprenoid biosynthesis. In this context we have studied the enzyme 2C-methyl-D-erythritol-4-phosphate cytidylyltransferase (CMS), the third enzyme in the MEP pathway, since the lack of a resolved structure of this protein in *M. tuberculosis* has seriously limited its use as a drug target. We performed homology modeling of *M. tuberculosis* CMS in order to provide a reliable model for use in structure-based drug design. After evaluating the quality of the model, we performed a thorough study of the catalytic site and the dimerization interface of the model, which suggested the most important sites (conserved and non-conserved) that could be useful for drug discovery and mutagenesis studies. We found that the metal coordination of CDP-methylerythritol in *M. tuberculosis* CMS differs substantially with respect to the *Escherichia coli* variant, consistent with the fact that the former is able to utilize several metal ions for catalysis. Moreover, we propose that electrostatic interactions could explain the higher affinity of the MEP substrate compared with the cytosine 5′-triphosphate substrate in the *M. tuberculosis* enzyme as reported previously.

Keywords 2C-methyl-D-erythritol-4-phosphate cytidylyltransferase (CMS) · Homology modeling · Molecular dynamics · Drug design · Tuberculosis

C. Obiol-Pardo · J. Rubio-Martinez
Dept. de Química Física, Universitat de Barcelona and the Intitut de Recerca en Química Teòrica i Computacional (IQTCUB),
Martí i Franquès 1,
08028 Barcelona, Spain

A. Cordero · S. Imperial (✉)
Dept. de Bioquímica i Biologia Molecular,
Universitat de Barcelona,
Avda. Diagonal 645,
08028 Barcelona, Spain
e-mail: simperial@ub.edu

Introduction

Tuberculosis (TB), caused by *Mycobacterium tuberculosis*, remains one of the leading infectious diseases in humans, with approximately 8 million new cases worldwide and an estimated 2 million deaths annually [1].

The standard first-line treatment against active TB is based on the administration of the drugs rifampicin, isoniazid, pyrizinamide, and ethambutol given in combination over 6–9 months. The combinations are very important to prevent the emergence of multiple drug resistant (MDR) organisms, which can lead to ineffective treatment [2].

There are three basic factors involved in the development of new tuberculosis drugs: to reduce the total duration of treatment, to avoid the emergence of MDR TB, and to provide more effective treatment of latent tuberculosis infection [3, 4].

In *M. tuberculosis*, the synthesis of isoprenoids is an important step in cell wall construction [5–7] and oxidative phosphorylation [8]. It is therefore reasonable to predict that enzymes involved in the early steps of biosynthesis of

essential isoprenoids represent novel and valid drug targets [6].

Despite their diversity of structure and function, all isoprenoids originate from the common five-carbon (C5) building units isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP) [9, 10]. To date, two distinct pathways for the biosynthesis of the two precursors have been discovered: the mevalonate (MVA) pathway and the methylerythritol phosphate (MEP) pathway. Most bacterial pathogens, including *M. tuberculosis*, utilize the MEP pathway exclusively, whereas only the MVA pathway is present in humans [11–13].

The MEP pathway is comprised of seven enzymatic steps starting with the condensation of pyruvate and glyceraldehyde-3-phosphate to produce DXP, which is converted to MEP in reactions catalyzed by DXP synthase [14–16] and DXP reductoisomerase (IspC) [17, 18], respectively. In the third step of the pathway, MEP reacts with cytosine 5′-triphosphate (CTP) to produce CDP-methylerythritol and pyrophosphate in a reaction catalyzed by the enzyme 2C-methyl-D-erythritol-4-phosphate cytidylyltransferase (EC 2.7.7.60; CMS) [19–22]. Not many inhibitors have been described for the enzymes of the MEP pathway [13]. A paradigmatic example is fosmidomycin, an antibiotic that selectively inhibits DXP reductoisomerase [17, 18] that has proven useful against malaria [12].

CMS enzymes from *Escherichia coli* and other bacteria including *M. tuberculosis* have been obtained in recombinant form [19–25]. The steady-state kinetic properties of CMS from *E. coli* [22], from *Streptomyces coelicolor* [23] and from *M. tuberculosis* have been reported [24, 25].

X-ray structural studies have been carried out with the bacterial enzymes from *E. coli* [21], *Thermus thermophilus* [26] and *Thermotoga maritima* [27], among others, but so far no crystal structure is yet available for CMS from *M. tuberculosis*. Moreover, only a weak inhibitor of CMS (erythritol-4-phosphate) has been described [28], but as this compound is a desmethyl analogue of the natural cofactor, it does not provide any structural clue to the design of new active compounds for different binding sites. Structural analysis has shown that CMS is organized as a homodimer, with each enzyme subunit being primarily a single domain α/β structure constructed around a seven-stranded twisted β-sheet, into which is inserted an extended β-arm. Two arms associate to help form a dimer interface that involves numerous inter-subunit hydrogen bonds and several salt bridges. The enzyme active site is created at this interface by seven polypeptide segments, six from one subunit and one from its partner.

Amino acids that are involved in formation of the substrate binding site and the catalytic active site have been identified. A highly basic active site binds and orients the four phosphate groups present in the substrates—two lysine residues in particular (Lys 27 and Lys 213 in *E. coli* CMS) are key to the stabilization of the pentavalent transition state generated following an in-line nucleophilic attack by the MEP phosphate on the alpha phosphate of CTP. A complex network of hydrogen bonds positions MEP so that catalysis can occur.

The reaction catalyzed by CMS, like the rest of the reactions involved in the MEP pathway, represents an excellent potential target site for chemotherapy against *M. tuberculosis*. The properties of the enzyme would provide important information for the design of novel antitubercular agents and any differences observed with CMS from other bacteria would help define the specificity of these agents for *M. tuberculosis*.

M. tuberculosis CMS (the Rv3582c gene product) has approximately 32% identity with *E. coli* CMS, and conserved amino acids are involved in formation of the substrate binding site and the catalytic active site. However, some differences have been reported; for instance, the amino acid corresponding to S88 of the *E. coli* enzyme active site is a threonine in *Rv3582c*. Recombinant *M. tuberculosis* CMS is active over a relatively broad pH range (6.0–9.0) [24, 25] as was previously reported for *E. coli* CMS [22]. A divalent cation, such as Mg^{2+} or Mn^{2+}, is absolutely required for this cytidylyltransferase activity. Zn^{2+} supports *M. tuberculosis* CMS activity [25]. However, *E. coli* CMS activity was not supported by Zn^{2+} suggesting subtle differences in metal coordination [19, 25]. *M. tuberculosis* CMS showed a high degree of specificity for CTP as a substrate [24, 25] over other nucleotide 5-phosphates (GTP, UTP, TTP, or ATP).

Due to the rapid spread of multidrug-resistant TB strains resistant against all major antituberculosis drugs, there is an urgent need for TB drugs with fewer toxic side effects and improved pharmacokinetics properties that would allow a reduction in the total duration of treatment. In this context, a structure of the enzyme from *M. tuberculosis* would be desirable as a starting point for structure-based drug design. Here, we report the first model of CMS from *M. tuberculosis*, using the three-dimensional (3D) structure of the *E. coli* enzyme as a template, performing homology modeling and refining of the model by molecular dynamics (MD) simulations. Homology modeling is a general strategy to obtain a protein structure when no crystal structure is available, and it has been applied to several protein models with interesting results, including *M. tuberculosis* DXP reductoisomerase, the second enzyme of the MEP pathway [29]. Moreover, this methodology can suggest mutagenesis studies and useful sites for drug discovery [30].

Our MD simulation was used to analyze the energetics of the entire protein–protein interacting surface as well as the binding site of CDP-methylerythritol, identifying the

role of conserved residues between the *E. coli* and the *M. tuberculosis* variants. In the light of our study, electrostatic interactions could explain the higher affinity of MEP compared to CTP for binding as reported previously in the *M. tuberculosis* enzyme [24]. Moreover, a fundamental difference in the metal ion coordination between the *M. tuberculosis* model and the *E. coli* enzyme is also addressed. Non-conserved regions and identified channels were also analyzed to suggest new binding sites for the future development of selective inhibitors.

Methods

Construction of *M. tuberculosis* CMS

Homology modeling was performed using the MODELLER 8v2 program [31], which is based on comparative protein structure by satisfaction of spatial restraints [32, 33]. Several X-ray structures of CMS are appropriate and can be used as 3D templates: the *E. coli* variant (pdb codes 1INI and 1I52 [21]), the *Thermotoga maritima* variant (pdb code 1VPA [27]), the *Thermus termophilus* variant (pdb code 2PX7 [26]), and other possible templates including those of the bifunctional IspDF from *Campylobacter jejuni* (pdb codes 1W55 and 1W57 [34]), that of *Listeria monocytogenes* (pdb code 3F1C [35]) and those from *Streptococcus pneumoniae* (pdb codes 2VSH and 2VSI [36]). We selected the *E. coli* variant 1INI, which is the solved structure containing CDP-methylerithritol, because this variant has been studied exhaustively in terms of mutagenesis data and reaction kinetics. The other possible templates were considered inappropriate as they presented functional differences, less sequence identity or did not contain the complex with CDP-methylerithritol.

Figure 1 shows the sequence alignment between the *E. coli* and *M. tuberculosis* variants, with a sequence identity of 32% and revealing only four non-close gaps (only two in secondary structure elements). This alignment was performed using the ALIGN automatic server [37] using default parameters of scoring. A multiple sequence alignment including eight variants was examined [24] and no important differences were found with respect to our pairwise alignment.

Moreover, we carried out a superimposition of the six possible species using the Needleman-Wunsch algorithm, including secondary structure scores as defined in Chimera software [38] to determine if the selection of only one of these templates would introduce some bias into the *M. tuberculosis* CMS homology model. As shown in Fig. 2a, the topology and 3D structures of the templates are highly similar, revealing that selecting more than one template would be redundant in this system. Backbone root mean squares (RMS) deviations were 2.4 Å for 216 atom pairs between the *E. coli* and the *Thermotoga maritima* variants, 2.8 Å for 202 atom pairs between the *E. coli* and the *Thermus thermophilus* variants, 5.8 Å for 204 atom pairs between the *E. coli* and the *Campylobaster jejuni* variants, 4.5 Å for 215 atom pairs between the *E. coli* and the *Listeria monocytogenes* variants and 6.1 Å for 211 pairs between the *E. coli* and the *Streptococcus pneumoniae* variants.

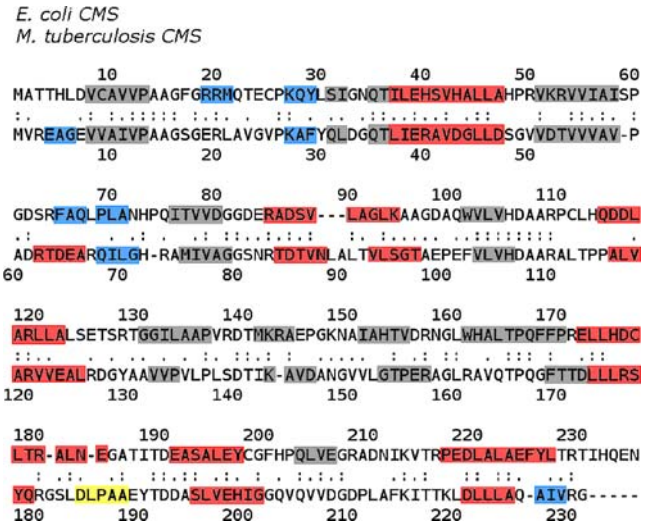

Fig. 1 Sequence alignment between *Escherichia coli* 2C-methyl-D-erythritol-4-phosphate cytidylyltransferase (CMS; *top sequence*) and *Mycobacterium tuberculosis* CMS (*bottom sequence*). Secondary structure prediction using STRIDE is also shown: *red* alpha helix, *gray* beta sheet, *blue* 3–10 helix, *yellow* π-helix

It is worth noting that sequence identity among these templates ranges within 25–32%. Despite this low sequence identity, the 3D structures of CMS are very well conserved among different species. *M. tuberculosis* CMS should be no exception and, for this reason, the sequence identity between this variant and the selected template seemed adequate to perform homology modeling, expecting to obtain a very similar overall structure.

The MODELLER program was applied to generate 30 satisfactory models for each monomer of *M. tuberculosis* CMS. The model with the lowest energy and the lowest restraint violation was selected to construct the homodimeric system. STRIDE software [39] was used to predict secondary elements in both the *E. coli* and the best MODELLER models of *M. tuberculosis* CMS (Fig. 1). The secondary structure of the model was similar to that of the *E. coli* protein, thus confirming the quality of the homology modeling. As regards tertiary structure, Fig. 2b shows the best *M. tuberculosis* model selected superimposed over the crystal structure of *E. coli* CMS. This figure also shows the reaction product CDP-methylerythritol, placed in both monomers of the protein and superimposed between the *E. coli* and the *M. tuberculosis* variants. The superimposition was also performed using the Needleman-Wunsch algorithm,

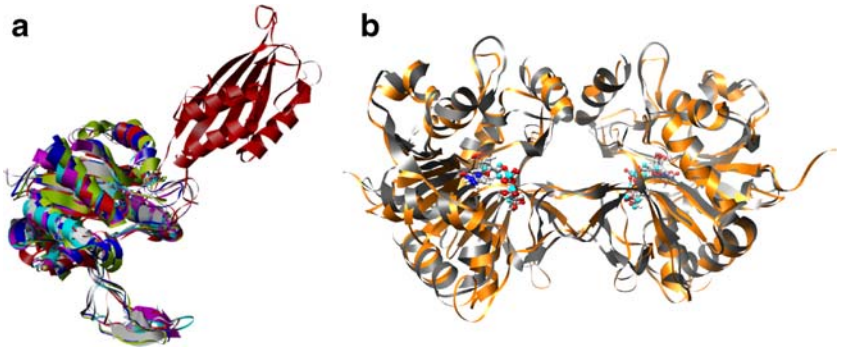

Fig. 2 **a** Superimposition of six X-ray structures that can be used as templates; *gray E. coli* variant (1INI), *purple Thermotoga maritima* variant (1VPA), *cyan Thermus thermophilus* variant (2PX7), *red Campylobacter jejuni* variant (1W55; note that this enzyme has a domain which cannot be used), *dark blue Listeria monocytogenes* variant (3F1C), *green Streptococcus pneumoniae* variant (2VSH). **b** Superimposition between the *E. coli* CMS template (*gray*) and the best model of *M. tuberculosis* generated with MODELLER (*orange*). Superimposed CDP-methylerythritol molecules are also shown. Figures created using Chimera software [38]

including the secondary structure score according to Chimera software [38]. The tertiary structure of both proteins was also very similar, providing a good starting point from which to refine the structure by MD simulation. To reveal the differences among the models generated, the five MODELLER models with the best scores for each monomer of *M. tuberculosis* CMS were compared. The maximum backbone RMS deviation was 2.26 Å. The monomer with the best score was selected to construct the enzyme homodimer (shown in Fig. 2b) taking into account that structures with a maximum deviation of 2.26 Å are expected during a normal MD simulation. Thus, it was considered that performing different MDs with other, lower scored, MODELLER models would not contribute to obtaining substantially different *M. tuberculosis* structures.

The *M. tuberculosis* CMS dimer was constructed by imposing the same symmetry as that of the *E. coli* CMS homodimer.

In the second step, hydrogen atoms were added to the structure by using the LEAP program of the AMBER-7 package [40].

The model was completed with the addition of two CDP-methylerythritol molecules (superimposing their position and conformation from the *E. coli* protein) and the two structural magnesium cations coordinated to the diphosphate fragment of CDP-methylerythritol.

The AMBER force field parameters of CDP-methylerythritol were calculated with the General Amber Force Field (GAFF) program [41] using a charge population derived from an adjustment to electrostatic potential (RESP) implemented in the GAUSSIAN package [42].

Finally, a cubic box of approximately 22,000 TIP3P water molecules was added to the system. Molecular dynamics was performed in an explicit solvent model, using periodic boundary conditions within the particle mesh Ewald (PME) context [43]. The system was neutralized by adding an appropriate number of sodium counterions.

All calculations were carried out at the molecular mechanics level using the force field of Cornell et al. [44] with the AMBER-7 suite of programs [40].

Minimization and molecular dynamics

The *M. tuberculosis* CMS homodimer was energy-minimized, using the steepest descent algorithm, to remove bad contacts derived from the homology modeling and to achieve a good starting structure with which to perform MD.

In the first step (1,000 iterations), only water and counterion molecules were allowed to move. In the second step (1,500 iterations), water, counterions and protein side chains were allowed to move. Finally, 50,000 iterations were calculated with the restriction-free system up to an energy gradient lower than 1 kcal mol^{-1}.

The minimized structure was heated by coupling the system to a thermal bath at 300 K using Berendsen's algorithm [45] and at a constant rate of 30 K/10 ps; all atoms except water and counterions were maintained frozen. The time step was set to 2 fs and the list of nearest neighbor atoms was updated every 15 ps. A cut-off distance of 9 Å was chosen to compute non-bonded interactions and the SHAKE [46] algorithm was imposed in order to constrain all bonds involving hydrogen atoms.

Once the system was heated, a second step of 40 ps at constant pressure was calculated to increase the density of the system. Restrictions on solute main chains and cofactors were then relaxed gradually until the system was free. Finally, 7.6 ns of unrestrained MD were calculated within the NVT ensemble at a constant temperature of 300 K.

Binding free energy calculation

The molecular mechanics generalized Born surface area (MMGBSA) protocol [47] was applied to estimate the binding free energy of CDP-methylerythritol in complex

with the *M. tuberculosis* CMS model as well as to evaluate the energetics of the dimer interface. This protocol was used within the one-trajectory approximation, and computes the binding free energy as follows:

$$\Delta G_{binding} = G_{complex} - \left(G_{receptor} + G_{ligand}\right) \quad (1)$$

$$G = E_{MM} + G_{Polar\ Solv} + G_{Non-Polar\ Solv} - TS \quad (2)$$

$$E_{MM} = E_{int} + E_{vdw} + E_{ele} \quad (3)$$

$$G_{Non-Polar\ Solv} = aSASA \quad (4)$$

Where E_{MM} denotes bond, angle and dihedral force field energies, and E_{vdw} and E_{ele} are the van der Waals and electrostatic energies. An external dielectric constant of 80 was set to treat the system with implicit solvation. The generalized Born methodology was applied to solve the polar contribution to solvation ($G_{Polar\ Solv}$), using the parameterization of Tsui and Case [48].

Finally, $G_{Non\text{-}Polar\ Solv}$ is the non-polar contribution to solvation, related linearly with the solvent accessible surface area (SASA), while a takes the value of 0.0072 kcal mol^{-1} $Å^{-2}$.

SASA was computed by means of the linear combination of pairwise overlap (LCPO) method [49]. Entropic effects (TS) were computed through a normal mode analysis using the nmode module of AMBER-7. The minimization step before normal mode analysis was carried out up to an energy gradient of 10^{-4} kcal mol^{-1}. It is important to note that, due to the limitations of the method, both intrinsic and based on the MD simulation [50], the absolute numerical values of the binding free energy are not relevant for these studies, being in general larger than the experimental data [51]. Only their relative values are important and informative and, consequently, the numerical value of the total binding free energy was not found to be close to the experimental data.

MMGBSA was applied to 100 equidistant snapshots extracted from the production time, except for the entropic contribution, which is computationally intensive, and therefore is usually based on only a few snapshots. For this contribution, ten snapshots were extracted by limiting the system to only those residues located at a distance of 9 Å from the CDP-methylerythritol molecules.

Results

Assessment of the *M. tuberculosis* CMS model

The total energy of the system versus time was examined in order to verify the stability of the model; the final 2 ns, from a total of 7.6 ns, was considered as the production time, and was used to extract structural and energetic results. In addition, the average backbone RMS deviation of the full system, which can be considered as a measure of the quality and stability of the model, was 4.3 Å. This value decreased to 3.9 Å by not accounting for the first 10 residues at both ends.

Energy and RMS convergence was achieved only after more than 5 ns of MD, due to the slow conformational changes of side chains together with the low sequence identity of the two variants. As can be seen, a long MD simulaiton is necessary to improve the initial MODELLER structure, and in general to refine a homology model with low sequence similarity.

As regards CDP-methylerythritol, the average all-atom RMS deviation of the first CDP-methylerythritol molecule with respect to the conformation in the *E. coli* variant was 2.4 Å (similar results were obtained for the second CDP-methylerythritol molecule). This is also a small deviation, in agreement with a good binding site description.

As a final test of the quality of the model, four structural evaluation programs were used; PROCHECK [52], Verify-3D [53], ProSA [54] and ERRAT [55]. PROCHECK was applied to quantify the residues in available zones of the Ramachandran plot, selecting the MODELLER model of *M. tuberculosis* CMS. Thus, 88.3% of residues were located in the most favored zones, 9.6% in allowed regions, 1% in generously allowed regions, and 1% in disallowed regions. Similar results were found in the *E. coli* template: 81.8% of residues in the most favored zones, 17.7% in allowed regions, no residues in generously allowed regions, and 0.5% of residues in disallowed regions.

Verify-3D uses a score function to assess the quality of the model. Figure 3 shows the Verify-3D profile of the complete homodimer, selecting the initial model and the final MD structure; residues with a score over 0.2 should be

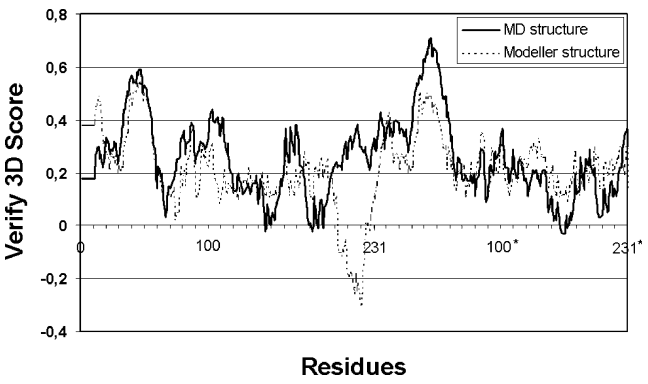

Fig. 3 Verify 3D score profiles [53] for the generated homodimer of *M. tuberculosis* extracted from MODELLER (*dashed lines*) or refined through molecular dynamics (MD). Scores greater than 0.2 indicate a high-quality structure. *Asterisks* Residues belonging to the second monomer

considered reliable. Thus, 51.6% of residues have a score over 0.2 in the initial model and 63.6% in the final MD structure since the structure was substantially improved, specifically in the sequence 202–220. This fragment corresponds to a structure with no secondary elements in *M. tuberculosis* CMS (Fig. 1), and which was therefore difficult to model; nevertheless, the subsequent MD improved this fragment of the protein as indicated by the Verify-3D profile score.

Additionally, Fig. 4 shows the energy profiles of the MODELLER model (middle) and the final MD structure (bottom) using the ProSA score, compared with the experimental structure of *E. coli* CMS (top). This program evaluates the energy of the structure using a distance-based pair potential. Residues with negative ProSA energies confirm the reliability of the model. As can be seen, the homology model shows a similar ProSa profile with respect to the template, and MD additionally decreased the positive peaks of the middle zone to better compare this profile with the *E. coli* structure.

Finally, an overall quality factor of 50.2% for the MODELLER structure was obtained using the ERRAT function, which is based on statistics of non-bonded interactions. This score was highly improved in the MD structure, giving an overall quality factor of 79.3 and 86.6% for monomers one and two, respectively. These results are very similar to the overall quality factor obtained for the *E. coli* template, which was 89.2%.

The main differences before and after MD, as indicated by these scoring functions, were found in the loops of *M. tuberculosis* CMS, particularly the loops formed by residues 14–26, 147–152 and 202–220. Figure 5a shows the positions of these three loops in the MODELLER structure, and Fig. 5b focuses on the short alpha helix obtained in loop 14–26 after MD, corresponding to a short helix found in the template structure (Fig. 1). This was the most marked structural change in the backbone during the simulation.

Role of conserved residues in *M. tuberculosis* CMS

Shi et al. [24] performed an accurate study of *E. coli* CMS mutants and identified invariant residues. A role in catalytic activity was suggested for these residues, and two signature motifs of CMS were found. We used our homology model of *M. tuberculosis* CMS to suggest the role of some of these conserved residues. In addition, this analysis was also utilized to shed light on the experimental affinities of the MEP and CTP substrates reported by Shi et al. [24] in *M. tuberculosis* CMS (Fig. 6). The MEP molecule, being smaller than CTP and therefore maintaining less contacts with the protein, has 2-fold more affinity for CMS (43 μM vs 92 μM).

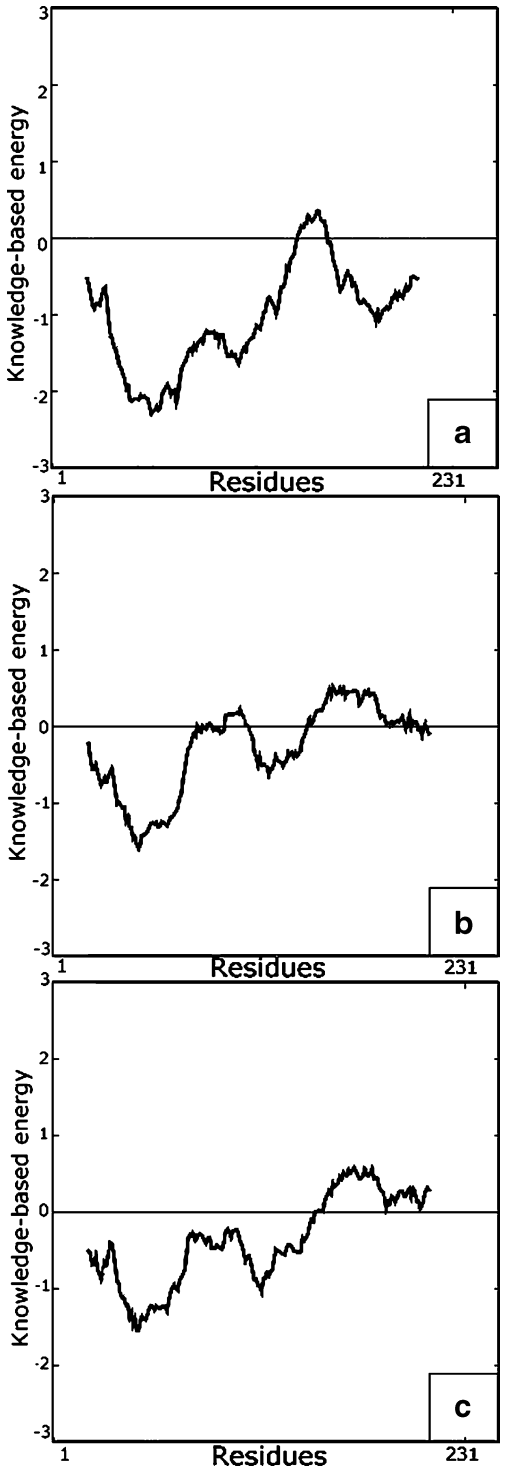

Fig. 4 ProSa energy scores [54] for the experimental *E. coli* CMS used as template (**a**), the MODELLER model of *M. tuberculosis* CMS (**b**) and the refined model of *M. tuberculosis* CMS (**c**). Negative scores indicate a high-quality structure

The similarity between CMS from different species is very high in the binding site, which is adapted to highly charged substrates such as CTP. Molecules able to interact in this zone should present a diphosphate or analogue

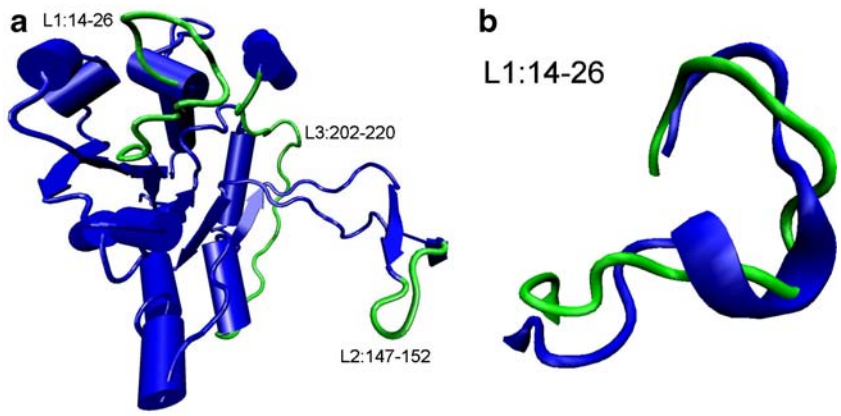

Fig. 5 **a** MODELLER structure of *M. tuberculosis* CMS (only one monomer shown for clarity) marking three loops (in *green*) whose structures were refined as suggested by the scoring functions applied. **b** Detailed structure of loop 14–26 before (*green*) and after (*blue*) MD. Figures created with VMD software [61]

group; nevertheless, some authors have described nonphosphate inhibitors to target isoprenoid biosynthesis [56]. An MMGBSA 'per residue' protocol [57] was performed to identify residues that stabilize CDP-methylerythritol. The interaction energies of residues of the whole protein were calculated with respect to the CDP-methylerythritol molecule present in the first monomer, averaging over 100 snapshots of the converged MD. Table 1 presents the results of this analysis, showing the ΔGB_{TOT} contribution (which includes van der Waals and electrostatic interactions, polar solvation contribution and non-polar solvation contribution) of the CMS residues that offered the main contributions to binding. Important residues of *M. tuberculosis* CMS were P13, A14, A15 and G16, which form an invariant motif among different CMS species called the glycine-rich loop. This cluster of residues contributes −7.96 kcal mol^{-1} to the total binding energy (Table 1).

Residue T86, which corresponds to the similar S88 in the *E. coli* variant, is much less important but has been postulated to sequester the CDP-methylerythritol in the *E. coli* enzyme [24]. The conserved A108 also interacts weakly and similarly to the same residue in *E. coli* (A107) [22]. R110 and K219, which also play a role in dimerization, contribute −8.30 and −2.46 kcal mol^{-1}, respectively (Table 1). In fact, R110 is the most critical residue for CDP-methylerythritol stabilization and its mutation in the *E. coli* enzyme (R109A) has been reported to cause a drastic inhibition [22]. Since this residue interacts with a hydroxyl of the MEP moiety of CDP-methylerythritol, we propose that R110 is in part responsible of the higher affinity of the MEP substrate (43 μM) compared to CTP (92 μM) reported previously in the *M. tuberculosis* protein [24]. As shown in the MMGBSA analysis (Table 1), R110 offers the most energetic stability per residue to CDP-methylerythritol binding. Additionally, we also decomposed the CDP-methylerythritol into chemical groups and calculated the average van der Waals and electrostatic energies from the force field to shed light on the reported affinities (Fig. 6). The methylerythritol

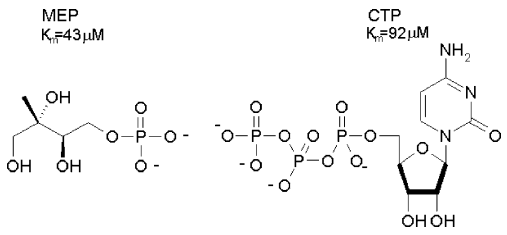

Fig. 6 Structures of the substrates methylerythritol phosphate (MEP) and cytosine 5′-triphosphate (CTP) showing the calculated affinities in *M. tuberculosis* CMS as reported in [24]

Table 1 Molecular mechanics generalized Born surface area (MMGBSA) 'per residue' decomposition showing the main residues that contribute to the CDP-methylerythritol binding. *CDP-ME* CDP-methylerythritol molecule

Residue	ΔGB_{TOT} (kcal/mol)
P13	−2.00
A14	−2.34
A15	−1.57
G16	−2.05
T86	−1.07
A108	−1.82
R110	−8.30
D193	4.39
K219	−2.46
Mg^{2+}	−38.81
CDP-ME	−63.38
T141[a]	−1.55

[a] Residue of the second monomer

fragment had a van der Waals interaction energy of −9.14 kcal mol^{-1}, whereas the ribose and the pyrimidine groups showed van der Waals interactions of −9.03 and −9.23 kcal mol^{-1}, respectively. Nevertheless, experimental affinities can be correlated taking into account the electrostatic interaction energy too. In this case, the amino group attached at the pyrimidine has an average electrostatic energy of −5.68 kcal mol^{-1} and the two hydroxyl groups of the ribose offer −12.23 kcal mol^{-1}, but the three hydroxyl groups of the methylerythritol maintain a much higher electrostatic interaction energy with the protein, i.e. −39.47 kcal mol^{-1}. All in all, although hydrophobic recognition is identified most in the moiety coming from the CTP, the much higher electrostatic energies, which can explain the different affinity of the substrates, are focused only on the hydroxyl groups of the methylerythritol fragment.

Finally, conserved T141 of the second monomer, which is also involved in dimerization, interacts with −1.55 kcal mol^{-1} (Table 1), and is also important for *E. coli* CMS. Mutation of this residue in *E. coli* (T140V) was shown to decrease enzyme activity greatly [22], presumably by a steric effect. The result of the *E. coli* CMS mutation T165V [22] is more ambiguous. This residue (also T165 in *M. tuberculosis* CMS) makes a weak hydrogen bond with the methylerythritol moiety in the second monomer of our model. Nevertheless, this contact is lost in the first subunit but replaced with another hydrogen bond with R110. This could be an effect of protein flexibility, which effectively exchanges a hydrogen bond between R110 and T165. Other mutations were also important in affecting the *E. coli* CMS activity: K27A, K27S and K213S. These two residues were postulated to affect the transition state of the reaction by a mechanism of charge stabilization. In our analysis by the MMGBSA protocol, no special role for these residues (K27 and K215) in the product stabilization of *M. tuberculosis* CMS was observed. We modeled a system with the reaction product and thus some residues should show different orientations with respect to the reactive or transition state structures. Another possibility is that these two residues are not effectively involved in direct interactions with CDP-methylerythritol as our model suggests. Therefore, we propose to test the effect of the K27A, K27S and K215S mutations on k_{cat}/K_m in *M. tuberculosis* CMS to shed some light on the role of these two residues. As regards D193, this residue makes an unfavorable contact with CDP-methylerythritol (Table 1) due to its negatively charged side chain, which is in close contact with the diphosphate group of CDP-methylerythritol, but this residue is coordinated to the structural magnesium and thus the unfavourable interaction is corrected. This MMGBSA analysis does not take into account this effect. Obviously, the magnesium ion and CDP-methylerythritol make the most important contribution to the total binding free energy for the complexation process (Table 1).

To complement these results, Fig. 7 shows a visual representation of the most important hydrogen bonds between the protein and CDP-methylerythritol, being similar in the *E. coli* variant (residues with van der Waals contacts are not shown for simplicity). The magnesium coordination is also shown. The carbonyl group of P13 forms a hydrogen bond with a hydroxyl group of the ribose moiety, and the amino groups of A14, A15 and G16 form three hydrogen bonds with the carbonyl of the pyrimidine base (the former is weak) similarly as described in the *E. coli* variant. In our model, the methyl group of T86 interacts by van der Waals forces with the pyrimidine base; the methyl of A108 interacts in the same manner with the ribose ring; the side chain of R110 interacts by electrostatic forces with the phosphate groups, with a hydroxyl of methylerythritol offering the most important contribution; the side chain of K219 also forms contacts with a hydroxyl group of methylerythritol. Finally, the methyl group of T141 of the second monomer makes a van der Waals contact with the methyl group of methylerythritol. Other residues involved in hydrogen bonds to the product are N82, T84, A109, T217, T218 from the first monomer and D140 from the second. Two strong hydrogen bonds localized in the erythritol fragment are formed with charged residues, D140 and R110, thus explaining the strong electrostatic interaction energy of the hydroxyl groups of this moiety discussed above.

The structural magnesium ion is stabilized by contact with two oxygen atoms of the phosphate group of CDP-methylerythritol, with both side chain oxygen atoms and the carbonyl group of D193 and with the side chain hydroxyl of T84 (Fig. 7). The structural magnesium ion located in the second subunit shows a slightly different

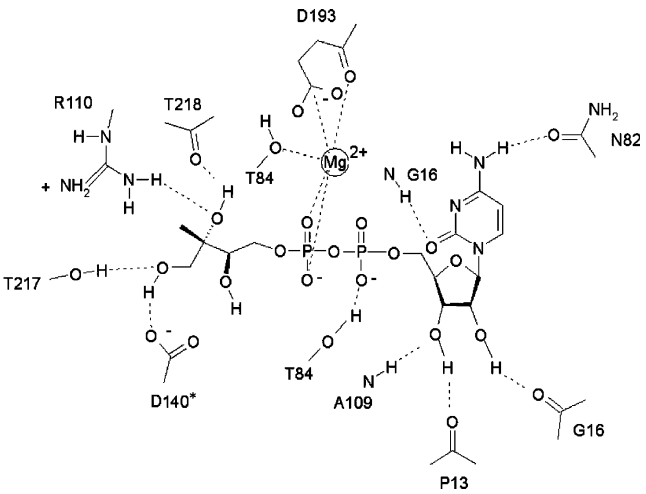

Fig. 7 Detailed view of the most important residues implicated in hydrogen bonds involved in CDP-methylerythritol binding and magnesium coordination. *Asterisk* Residue belonging to the second monomer

coordination in the final MD structure. Thus, the carbonyl of D193 that makes a contact with magnesium in the first subunit is replaced by the carbonyl of T84 in the second subunit. Both carbonyl fragments of T84 and D193 exhibit dynamic behavior throughout the simulation, but the same average coordination is always maintained. We suggest that D193A and T84A mutations should distort the ion coordination and affect the catalytic activity of *M. tuberculosis* CMS. Interestingly, the magnesium ion is not coordinated to any residue of the protein in the crystal structure of *E.coli* CMS [21] so its coordination is not as tight as shown in our model of *M. tuberculosis* CMS. This fundamental difference may explain why the *M. tuberculosis* variant can accept a zinc ion (which has an ionic radius slightly larger than that of magnesium and so different polarizability) whereas the *E. coli* variant cannot [19] and is more specific for magnesium. A fundamental difference in metal coordination between the *E. coli* and the *M. tuberculosis* enzymes was earlier suggested by Eoh et al. [25] based on their affinities for different metal ions. The presented model offers a structural explanation, involving D193 and T84, for these experimental data.

In order to elucidate the driving force of binding, we estimated the binding free energy of CDP-methylerythritol in complex with the protein, also including magnesium as part of the receptor as suggested in our previous study on a similar metal-coordinated cofactor [58]. Results of this analysis are shown in Fig. 8. The hydrophobic balance upon binding (VDW+ΔGB_{SUR}) was approximately three-fold lower than the electrostatic balance (ELE+ΔGB). According to these results, binding seems to be driven by electrostatics forces; this agrees with our previous MMGBSA studies carried out on similar ligands with the phosphate group [58].

Inclusion of entropy by a normal mode analysis decreased the final energy by 19.27 kcal mol^{-1}, which is not decisive for binding.

It is worth remarking here that the final binding free energy is unrealistic because the high electrostatic energy extracted from the force field is not well balanced by the desolvation penalty using the GB approach. Overestimated values of the binding free energy are usually found in the literature [51, 58], and for this reason MMGBSA, while useful for studying the relative contributions to binding, must be interpreted with caution when discussing absolute values.

Description of the dimerization interface by the MMGBSA protocol

An MMGBSA 'per residue' protocol [57] was also applied to calculate the binding free energy contribution of all residues of the first monomer with respect to the whole second monomer. In this approach, the most important residues, which show the strongest energies, should contribute to the stability of the homodimer. Moreover, small molecules that could bind to these critical residues could be used to disrupt dimerization and to inhibit the protein as Shi et al. [24] suggested for an active homodimer for *M. tuberculosis* CMS.

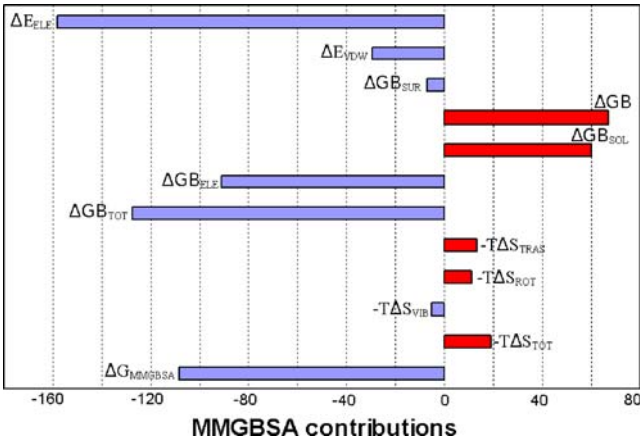

Fig. 8 Estimated binding free energy components, averaged over 100 snapshots, for the CDP-methylerythritol bound to the *M. tuberculosis* model (kcal mol^{-1} units). ΔE_{ELE} and E_{VDW} Electrostatic and van der Waals in vacuo binding enthalpic contributions, respectively; ΔGB_{SUR} non-polar contribution to solvation; ΔGB polar contribution to solvation; ΔGB_{SOL} $\Delta GB_{SUR}+\Delta GB_{CAL}$; ΔGB_{ELE} $\Delta E_{ELE}+\Delta GB_{CAL}$; ΔGB_{TOT} total enthalpic contribution, including solvation, to binding free energy using the Generalized Born framework; $-T\Delta S_{TRAS}$ translational entropic contribution; $-T\Delta S_{ROT}$ rotational entropic contribution; $-T\Delta S_{VIB}$ vibrational entropic contribution; $-T\Delta S_{TOT}$ total entropic balance; ΔG_{MMGBSA} final estimation of the theoretical binding free energy at 300 K. All entropic components are calculated truncating the receptor to only those atom within a cutoff of 9 Å to the ligand and using 10 snapshots

This methodology describes not only the interactions 'in vacuo' but also adds the solvation contribution to the final binding free energy. Figure 9 shows the results of this analysis, plotting the ΔGB_{TOT} contribution (which includes van der Waals and electrostatic interactions, polar solvation contribution and non-polar solvation contribution) versus each important CMS residue of the first monomer, averaged over 100 snapshots extracted from MD. Also included are CDP-methylerythritol and the structural magnesium. It can be seen that binding between subunits is driven by small clusters or "hot spots" that play the major role in protein–protein recognition. Eleven residues that contribute with the highest binding energies were identified: T141, I142, D146, R161, V163, D193, A194, S195, I200, K219 and I228. All these residues are located in the middle and C-terminal regions of the protein. The first five residues are located within the extended β-arm domain and contribute with the strongest interactions, while the rest are located within the globular domain. We suggest that T141A and I142A mutations should have important effects on the dimer stability

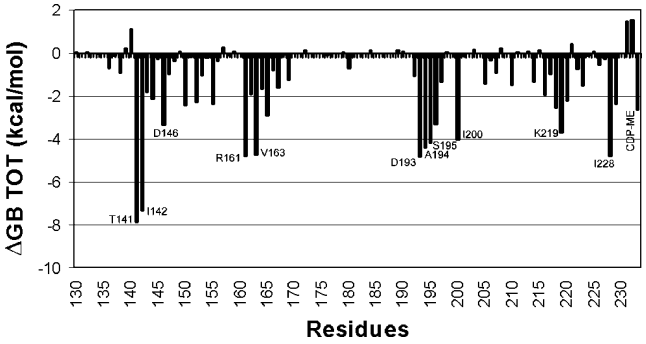

Fig. 9 MMGBSA 'per residue' decomposition showing only those residues that contribute most to dimer stabilization. *CDP-ME* CDP-methylerythritol

of CMS. T141 also contributes to CDP-methylerythritol stabilization (as discussed above) and therefore its mutation could disrupt dimer formation and at the same time distort the catalytic site.

It is worth remarking that T141, A194 and S195 are conserved residues and that I142, V163, D193 are similar residues in *E. coli* CMS, therefore the dimerization interface is clearly conserved. CDP-methylerythritol contributes slightly (−3 kcal mol^{-1}) to dimer stability (Fig. 9).

Role of non-conserved residues in *M. tuberculosis* CMS

Seven zones in which the sequence identity between *M. tuberculosis* CMS and the *E. coli* variant is substantially decreased were identified and analyzed: L21–V25 (0% identity), D47–T53 (14% identity), R62–I76 (7% identity), A111–L117 (0% identity), E123–136 (0% identity), V145–G153 (0% identity) and L175–Y190 (19% identity). A potential compound able to interact with these zones would be of great interest for selective inhibition of *M. tuberculosis* CMS. Some of these non-conserved zones include loops or sequences without secondary structure, and are located entirely within the globular domain, except V145–G153, which is in the middle zone of the β-arm domain. These zones presented high mobility during the MD

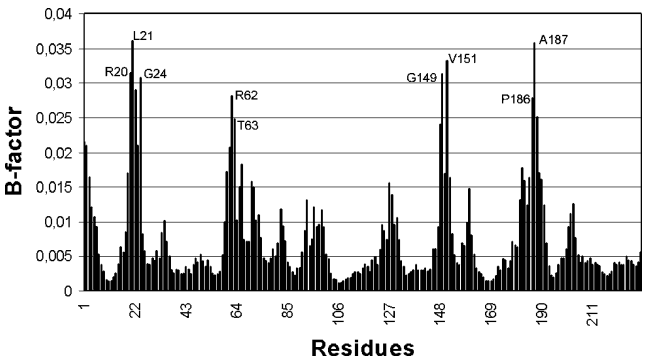

Fig. 10 Calculated B-factors for the first monomer of *M. tuberculosis* CMS refined through MD

Table 2 MolAxis [60] results for the homodimer of *M. tuberculosis* CMS, showing the ranked channels found and the bottleneck residues involved in these channels

Ranked channel	"Bottleneck" amino acids
1	L220, L222[b], P211[a], K215[a,b]
2	Q164, K143[a,b], Q164[a]
3	T141[b], Q164, Q164[a]
4	A213, D221[a,b]
5	K215[a,b], T217[a,b]
6	T217[b], D140[a,b], T141[a,b]
7	V118, V135, P211
8	D61[b], T63, D64
9	T141[b], P166[a,b], K219[a]
10	K219, L220, L138[a]

[a] Residues belonging to the second monomer

[b] Residues conserved with respect to the *E. coli* variant

simulation, and docking procedures aimed at finding molecules recognizing them should take into account these flexible structures. Thus, a docking procedure with induced-fit effects would be preferred.

B-factors (Fig. 10) using normal mode analysis were carried out using the elNémo server [59], corroborating the flexibility of these sequences. Interestingly, the most flexible loop was found between residues R20 and K27; as suggested by Richard et al. [21], this loop is the one implicated in catalysis in *E. coli* CMS. The fourth sequence (A111–L117), which is not flexible, is close to the catalytic site and should be a potential new site for drug development, focusing on selectivity. On the other hand, the R62–I76 sequence is enclosed in the channel 8 structure discussed in the following

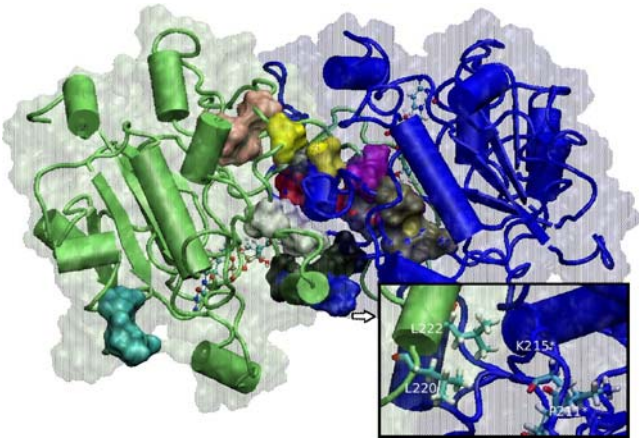

Fig. 11 Most important gating residues involved in channels as defined by MolAxis [60]. Channels: *blue* 1, *red* 2, *gray* channel 3, *yellow* 4, *tan* 5, *white* 6, *pink* 7, *cyan* 8, *purple* 9, *black* 10. CDP-methylerythritol molecules are shown in *ball and stick* representation. *Inset* Detailed view of the best scored channel showing the gating residues; *asterisks* residues belonging to the second monomer. Figure created with VMD software [61]

section, and so also seems suitable for docking studies to find selective drug candidates.

Identification of channels in *M. tuberculosis* CMS

The MolAxis algorithm [60], which is based on α-shape theory and on the geometrical concept of the median axis was used to identify channels in the refined model of *M. tuberculosis* CMS. Channels that connect inside cavities with the outside of the protein can be used as reference sites for drug discovery. Table 2 shows ten identified channels in the homodimer model, and the gating residues (or "bottleneck" residues) involved in these channels. Interestingly, all channels, except channels 5, 7 and 8, are composed of residues contributed by both monomers. All channels are well conserved except channels 7 and 10.

Figure 11 shows these gating residues highlighted in the complete homodimer structure. As can be seen, gating residues of channels 6 and 10 (white and black, respectively) are close to the catalytic site, and therefore any compound that could bind to these proposed sites would probably inhibit *M. tuberculosis* CMS allosterically. In this regard, the inset in Fig. 11 shows a detailed view of the best scored channel and its gating residues. L220 and L222 of the first monomer constitutes a hydrophobic cluster, while carbonyl groups of P211 and K215 of the second monomer make a polar cluster. Interestingly, the proximity of this channel to the catalytic pocket suggests that this site is the entrance of the natural substrates of CMS. The analysis shown in Fig. 10 also indicates that this sequence is highly rigid, accounting for the fine degree of selectivity of the enzyme for its substrates. Therefore, the gating residues of this channel could be used as a reference site to carry out a virtual screening, focused on finding new inhibitors by blocking CMS substrates.

Conclusions

Tuberculosis remains one of the leading infectious diseases in humans and isoprenoid biosynthesis in *M. tuberculosis* has been validated as a new target pathway for the future development of specific drugs against tuberculosis. The third enzyme involved in this pathway is 2C-methyl-D-erythritol-4-phosphate cytidylyltransferase (CMS). Since no structure of this enzyme from *M. tuberculosis* has yet been reported, we presented a homology model refined through molecular dynamics using the resolved *E. coli* CMS crystal structure as template. We analyzed the quality of the model in detail, in terms of secondary and tertiary structure, as well as the energetics of the model by applying different assessment programs and the MMGBSA protocol. We concluded that, in light of the scoring functions applied, the reported *M. tuberculosis* homology model seems reliable and could be used in drug design strategies. The model was further analyzed to address the role of conserved and non-conserved residues. The former were located in the substrate binding site and in the dimerization interface, whereas non-conserved residues were localized in loops with high mobility during the simulation. Overall, our model provided the following main insights: (1) the most important residues for dimer stability were conserved T141 and I142; (2) the higher affinity of the MEP substrate compared with the CTP substrate in *M. tuberculosis* CMS could be attributed to hydrogen bonding with conserved R101, and to the much higher electrostatic interactions of the hydroxyl groups of the MEP moiety; (3) there is a fundamental difference in metal binding between the *M. tuberculosis* model and the *E. coli* structure, as the metal ion was coordinated by residues D193 and T84 in our reported model, which are absent in the crystal structure of *E. coli* CMS. This finding is in agreement with the ability of *M. tuberculosis* CMS to support several metal ions for catalysis, whereas the *E. coli* variant is more specific; (4) the identified channel with the gating residues L220, L222, K215 and P211 seems an appropriate target with which to carry out a virtual screening protocol for identifying molecules able to block the entrance of the natural substrates of the protein.

In the absence of the crystal structure of the third enzyme of the MEP pathway in the *M. tuberculosis* organism, we have presented an accurate model to be used in future drug design protocols.

Acknowledgments The Spanish Ministry of Science and Technology supported this work through the projects CTQ2006-06588/BQU and BIO2002-04419-C02-02. The 'Generalitat de Catalunya' also supported this work through grants 2005SGR00914 and 2009SGR1308. We are also grateful to the 'Departament d'Universitat, Recerca i Societat de la informació de la Generalitat de Catalunya i del Fons Social Europeu'.

References

1. Global tuberculosis control: surveillance, planning, financing (2007) (WHO/HTM/TB/2007376)
2. Vinícius M, de Souza N (2006) Promising drugs against tuberculosis. Recent Patents Anti-Infect Drug Disc 1:33–44
3. O'Brien RJ, Nunn PP (2001) The need for new drugs against tuberculosis. Obstacles, opportunities and next steps. Am J Respir Crit Care Med 162:1055–1058
4. Mukherjee JS, Rich ML, Socci AR, Joseph JK, Viru FA, Shin SS, Furin JJ, Becerra MC, Barry DJ, Kim JY, Bayona J, Farmer P, Fawzi MCS, Seung KJ (2004) Programs and principles in treatment of multidrug-resistant tuberculosis. Lancet 363:474–481
5. Wolucka BA, McNeil MR, Hoffmann E, Chojnacki T, Brennan PJ (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

6. Mikusova K, Mikus M, Besra GS, Hancock I, Brennan PJ (1996) Biosynthesis of the linkage region of the mycobacteria cell wall. region. J Biol Chem 271:7820–7828
7. Mahapatra S, Yagi T, Belisle JT, Espinosa BJ, Hill PJ, McNeil MR, Brennan PJ, Crick DC (2005) Mycobacterial lipid II is composed of a complex mixture of modified muramyl and peptide moieties linked to decaprenyl phosphate. J Bacteriol 187:2747–2757
8. Collins MD, Pirouz T, Goodfellow M, Minnikin DE (1977) Distribution of menaquinones in actinomycetes and corynebacteria. J Gen Microbiol 100:221–230
9. Sacchettini JC, Poulter CD (1997) Creating isoprenoid diversity. Science 277:1788–1789
10. Bochar DA, Freisen JA, Stauffacher CV, Rodwell VW (1999) In: comprehensive natural products chemistry. Pergamon, Oxford
11. Testa CA, Brown MJ (2003) The methylerythritol phosphate pathway and its significance as a novel drug target. Curr Pharm Biotechnol 4:248–259
12. Rodríguez-Concepcion M (2004) The MEP pathway: a new target for the development of herbicides, antibiotics and antimalarial drugs. Curr Pharm Des 10:2391–2400
13. Eoh H, Brennan PJ, Crick DC (2008) The Mycobacterium tuberculosis MEP (2C-methyl-D-erythritol 4-phosphate) pathway as a new drug target. Tuberculosis 89:1–11
14. Sprenger GA, Schörken U, Wiegert T, Grolle S, De Graaf AA, Taylor SV, Begley TP, Bringer-Meyer S, Sahm H (1997) Identification of a thiamin-dependent synthase *Escherichia coli* required for the formation of 1-deoxy-D-xylulose 5-phosphate precursor to isoprenoids, thiamin and piridoxol. Proc Natl Acad Sci USA 94:12857–12862
15. Lois LM, Campos N, Putra SR, Danielsen K, Rohmer M, Boronat A (1998) Cloning and characterization of a gene from *Escherichia coli* encoding a transketolase-like enzyme that catalyzes the synthesis of D-1-deoxyxylulose 5-phosphate, a common precursor for isoprenoid, thiamin, and pyridoxol biosynthesis. Proc Natl Acad Sci USA 95:2105–2110
16. Lange BM, Wildung MR, McCaskill D, Croteau R (1998) A family of transketolases that directs isoprenoid biosynthesis via a mevalonate-independent pathway. Proc Natl Acad Sci USA 95:2100–2104
17. Kuzuyama T, Takahashi S, Watanabe H, Seto H (1998) Direct formation of 2-C-methyl-D-erythritol 4-phosphate from 1-deoxy-D-xylulose 5-phosphate by 1-deoxy-D-xylulose 5-phosphate reductoisomerase, a new enzyme in the Non-Mevalonate pathway to isopentenyl diphosphate. Tetrahedron Lett 39:4509–4512
18. Proteau PJ (2004) 1-Deoxy-D-xylulose 5-phosphate reductoisomerase: an overview. Bioorg Chem 32:483–493
19. Rohdich F, Wungsintaweekul J, Fellermeier M, Sagner S, Herz S, Kis K, Eisenreich W, Bacher A, Zenk MH (1996) Cytidine 5′-triphosphate-dependent biosynthesis of isoprenoids: YgbP protein of *Escherichia coli* catalyzes the formation of 4-diphosphocytidyl-2-C-methylerythritol. Proc Natl Acad Sci USA 96:11758–11763
20. Kuzuyama T, Takagi M, Kaneda K, Dairi T, Seto H (2000) Formation of 4-(cytidine 5′-diphospho)-2-C-methyl-D-erythritol from 2-C-methyl-D-erythritol 4-phosphate by 2-C-methyl-D-erythritol4-phosphate cytidylyltransferase, a new enzyme in the nonmevalonate pathway. Tetrahedron Lett 41:703–706
21. Richard SB, Bowman ME, Kwiatkowski W, Kang I, Chow C, Lillo AM, Cane DE, Noel JP (2001) Structure of 4-diphosphocytidyl-2-C-methylerythritol synthetase involved in mevalonate-independent isoprenoid biosynthesis. Nature Struct Biol 8:641–648
22. Richard S, Lillo A, Tetzlaff C, Bowman M, Noel J, Cane D (2004) Kinetic analysis of *Escherichia coli* 2-C-Methyl-D-erythritol-4-phosphate Cytidyltransferase, wild type and mutants, reveals roles of active site amino acids. Biochemistry 43:12189–12197
23. Cane DE, Chow C, Lillo A, Kang I (2001) Molecular cloning, expression and characterization of the first three genes in the mevalonate-independent isoprenoid pathway in Streptomyces coelicolor. Bioorg Med Chem 9:1467–1477
24. Shi W, Feng J, Zhang M, Lai X, Xu S, Zhang X, Wang H (2007) Biosynthesis of isoprenoids: characterization of a functionally active recombinant 2-C-methyl-D-erythritol 4-phosphate cytidyl-transferase (IspD) from *Mycobacterium tuberculosis* H37Rv. Biochem Mol Biol 40:911–920
25. Eoh H, Brown AC, Buetow L, Hunter WH, Parish T, Kaur D, Brennan PJ, Crick DC (2007) Characterization of the *Mycobacterium tuberculosis* 4-diphosphocytidyl-2-C-methyl-D-erythritol synthase: potential for drug development. J Bacteriol 189:8922–8927
26. Chen L, Tsukuda M, Ebihara A, Shinkai A, Kuramitsu S, Yokoyama S, Chen LQ, Liu ZJ, Lee D, Chang SH, Nguyen D, Rose JP, Wang BC. Crystal structure of 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase from *Thermus thermophilus* HB8. Structure Summary MMDB, http://www.ncbi.nlm.nih.gov/Structure/mmdb/mmdbsrv.cgi?uid=2PX7
27. Joint Center for Structural Genomics (JCSG). Joint Center for Structural Genomics (JCSG) Crystal structure of 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (TM1393) from *Thermotoga maritima* at 2.67 A resolution. Structure Summary MMDB, http://www.ncbi.nlm.nih.gov/Structure/mmdb/mmdbsrv.cgi?uid=1VPA
28. Lillo AM, Tetzlaff CN, Sangari FJ, Cane DE (2003) Functional expression and characterization of EryA, the erythritol kinase of Brucella abortus, and enzymatic synthesis of L-erythritol-4-phosphate. Bioorg Med Chem Lett 13:737–739
29. Singh N, Avery MA, McCurdy CR (2007) Towards *Mycobacterium tuberculosis* DXR inhibitor design: homology modeling and molecular dynamics simulation. J Comput Aided Mol Des 21:511–522
30. Hillisch A, Pineda LF, Hilgenfeld R (2004) Utility of homology models in the drug discovery process. Drug Discov Today 9:659–669
31. Eswar N, Mari-Renom MA, Webb B, Madhusudhan MS, Eramian D, Shen M, Pieper U, Sali A (2006) Comparative protein structure modeling with MODELLER. Current protocols in Bioinformatics. Supplement 15, Wiley, New York, pp 5.6.1–5.6.30, 200
32. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234:779–815
33. Fiser A, Do RK, Sali A (2000) Modeling of loops in protein structures. Protein Sci 9:1753–1773
34. Gabrielsen M, Bond CS, Hallyburton I, Hecht S, Bacher A, Einsenreich W, Rohdich F, Hunter WN (2004) Structure of the bifunctional ISPDF from *Campylobacter jejuni*. J Biol Chem 279:52753–52761
35. Patskovsky Y, Ho J, Toro R, Gilmore M, Miller S, Groshong C, Sauder MJ, Burley SK, Almo SC. Crystal structure of 2-C-methyl-D-erythritol-4-phosphate cytidyltransferase from *Listeria monocytogenes*.
36. Baur S, Marles-Wright J, Buckenmaier S, Lewis RJ, Vollmer W (2009) Synthesis of CDP-activated ribitol for teichoic acid precursors in *Streptococcus pneumoniae*. J Bacteriol 191:1200–1210
37. http://xylian.igh.cnrs.fr/bin/align-guess.cgi
38. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612
39. Heinig M, Frishman D (2004) STRIDE: a web server for seconday structure assigment from known atomic coordinates of proteins. Nucleic Acids Res 32:W500–W502
40. Case DA, Pearlman DA, Caldwell JW, Cheathan TE III, Wang J, Ross WS, Simmerling CL, Darden TD, Merz KM, Stanton RV,

Cheng AL, Vincent JJ, Crowley M, Tsui V, Gohlke H, Radmer RJ, Duan Y, Pitera J, Massova I, Seibel GL, Sligh UC, Weiner PK, Kollman PA (2002) AMBER 7. Univ California, San Francisco
41. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) Development and testing of a general amber force field. J Comput Chem 25:1157–1174
42. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery JA Jr, Stratmann RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Baboul AG, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Gomperts R, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Gonzalez C, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Andres JL, Head-Gordon M, Replogle ES, Pople JA (1998) Gaussian 98. Revision A.6. Gaussian, Pittsburgh PA
43. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an W log(N) method for Ewald sums in large systems. J Chem Phys 98:10089–10092
44. Cornell WD, Cieplak P, Bayly CI, Goud IR, Mertz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) A second generation of force fields for the simulation of proteins, nucleic acids and organic molecules. J Am Chem Soc 117:5179–5197
45. Berendsen HJC, Postman JPM, Van Gunsteren WF, DiNola A, Haak JA (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684–3690
46. Ryckaert JP, Ciccotti G, Berendsen HJ (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. J Comput Chem 23:327–341
47. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Srivasan J, Case DA, Cheatam TE III (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Acc Chem Res 33:889–897
48. Tsui V, Case DA (2001) Theory and applications of the generalized born solvation model in macromolecular simulations. Nucleic Acids Sci 56:275–291
49. Weiser J, Shemkin PS, Still WC (1999) Approximate atomic surfaces from linear combinations of pairwise overlaps. J Comput Chem 20:217–230
50. Pearlman DA (2005) Evaluating the molecular mechanics poisson-boltzmann surface area free energy method using a congeneric series of ligands to p38 MAP kinase. J Med Chem 48:7796–7807
51. Zang X, Li X, Wang R (2009) Interpretation of the binding affinities of PTP1B inhibitors with the MM-GB/SA method and the X-Score scoring function. J Chem Inf Model 49:1033–1048
52. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 26:283–291
53. Eisenberg D, Luthy R, Bowie JU (1997) VERIFY3D: assessment of protein models with three-dimensional profiles. Methods Enzymol 277:396–404
54. Wiederstein M, Sippl MJ (2007) ProsaWeb: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 35:407–410
55. Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci 2:1511–1519
56. Hirsch AKH, Lauw S, Gersbach P, Schweizer WB, Rohdich F, Eisenreich W, Bacher A, Diederich F (2007) Nonphosphate inhibitors of IspE protein, a kinase in the non-mevalonate pathway for isoprenoid biosynthesis and a potential target for antimalarial therapy. Chemmedchem 2:806–810
57. Gohlke H, Kiel C, Case DA (2003) Insights into protein-protein binding by binding free energy calculations and free energy decomposition for Ras-Raf and Ras-RaIGDS complexes. Mol Biol 330:891–913
58. Obiol-Pardo C, Rubio-Martinez J (2008) Homology modeling of human transketolase: description of critical sites useful for drug design and study of the cofactor binding mode. J Mol Graph Model 27:723–734
59. Suhre K, Sanejouand YH (2004) ElNemo: a normal mode web-server for protein movement analysis and the generation of templates for molecular replacement. Nucleic Acids Res 32:W610–W614
60. Yaffe E, Fishelovitch D, Wolfson HJ, Halperin D, Nussinov (2008) MolAxis: efficient and accurate identification of channels in macromolecules. Prot Struct Funct Bioinform 73:72–86
61. Humprey W, Dalke A, Schulten K (1996) VMD visual molecular dynamics. J Mol Graph 14:33–38