J Comput Aided Mol Des (2007) 21:511–522
DOI 10.1007/s10822-007-9132-0

Toward *Mycobacterium tuberculosis* DXR inhibitor design: homology modeling and molecular dynamics simulations

Nidhi Singh · Mitchell A. Avery · Christopher R. McCurdy

Received: 9 January 2007 / Accepted: 28 August 2007 / Published online: 14 September 2007

Abstract *Mycobacterium tuberculosis* 1-deoxy-D-xylulose-5-phosphate reductoisomerase (*Mt*DXR) is a potential target for antitubercular chemotherapy. In the absence of its crystallographic structure, our aim was to develop a structural model of *Mt*DXR. This will allow us to gain early insight into the structure and function of the enzyme and its likely binding to ligands and cofactors and thus, facilitate structure-based inhibitor design. To achieve this goal, initial models of *Mt*DXR were generated using MODELER. The best quality model was refined using a series of minimizations and molecular dynamics simulations. A protein–ligand complex was also developed from the initial homology model of the target protein by including information about the known ligand as spatial restraints and optimizing the mutual interactions between the ligand and the binding site. The final model was evaluated on the basis of its ability to explain several site-directed mutagenesis data. Furthermore, a comparison of the homology model with the X-ray structure published in the final stages of the project shows excellent agreement and validates the approach. The knowledge gained from the current study should prove useful in the design and development of inhibitors as potential novel therapeutic agents against tuberculosis by either de novo drug design or virtual screening of large chemical databases.

Keywords *Mycobacterium tuberculosis* · 1-Deoxy-D-xylulose 5-phosphate reductoisomerase · Tuberculosis · Homology modeling · Molecular dynamics · Drug design

Introduction

Tuberculosis (TB) is the most prevalent infectious disease worldwide and remains the leading cause of mortality, with about 2 million deaths annually [1, 2]. The number of new cases each year (approx. 9 million) is rising, mainly as a result of the increasing burden of human immunodeficiency virus (HIV) infection [1–3]. Today, the threat is most significant for nations in Africa, Latin America, Eastern Europe, and Asia. Tuberculosis can be caused by several species of the pathogenic bacteria *Mycobacterium* (*M.*). Relevant members of the TB complex group include *M. tuberculosis*, *M. bovis*, and *M. microti.* Although *M. tuberculosis* causes over 90% of clinically diagnosed TB in humans, it rarely infects other mammals [4]. The disease is highly contagious and spreads by airborne means from an infected individual through contaminated droplets coughed or sneezed into the air. Upon inhalation by an individual, the bacteria become lodged in the lungs and begin to

Electronic supplementary material The online version of this article (doi:10.1007/s10822-007-9132-0) contains supplementary material, which is available to authorized users.

N. Singh · M. A. Avery · C. R. McCurdy (✉)
Department of Medicinal Chemistry, Laboratory for Applied Drug Design and Synthesis, University of Mississippi, University, MS 38677-1848, USA
e-mail: cmccurdy@olemiss.edu

M. A. Avery · C. R. McCurdy
National Center for Natural Products Research, University of Mississippi, University, MS 38677-1848, USA

M. A. Avery
Department of Chemistry and Biochemistry, University of Mississippi, University, MS 38677-1848, USA

C. R. McCurdy
Department of Pharmacology, School of Pharmacy, University of Mississippi, University, MS 38677-1848, USA

multiply. Subsequently, the bacteria may be transmitted through the lymph system and blood stream to the kidneys, bone marrow, and meninges. Most people have immune systems that are healthy enough to prevent the bacteria from ever causing TB. However, in about 10% of cases, the infection breaks out into active TB disease at some point during the life of the infected person. The incidence of TB in HIV-infected persons is more than 100 times that of the general population since acquired immunodeficiency syndrome (AIDS) robs the body of its natural ability to fight infection, making them more likely to develop TB [5–8]. Consequently, the reemergence of TB as a potential public health threat alone or as an agent of bioterrorism, the high susceptibility of human immunodeficiency virus-infected persons to the disease and the proliferation of multi-drug-resistant (MDR) strains [8] have created much scientific interest in developing novel chemotherapeutic approaches to combat *M. tuberculosis.*

The complete sequencing of the genome of the most thoroughly characterized strain of *M. tuberculosis* H37Rv [9, 10] has provided a better understanding of the biology of this slow growing pathogen and has also unraveled potential targets and biochemical pathways that may be of utility in prophylactic and therapeutic interventions. One of the latter is the 1-deoxy-D-xylulose-5-phosphate/2-C-methyl-D-erythritol-4-phosphate (DXP/MEP) pathway [11–13] for isoprenoid biosynthesis. This pathway is operative in the malarial parasite [14], most eubacteria including several pathogenic species such as *Bacillus anthracis*, *Helicobacter pylori*, *Haemophilus influenzae* [15], algae [16], and in the plastids of plant cells [17]. However, it is absent in mammals, fungi, and the cytosol of plant cells, which alternately use the mevalonate pathway [18] for isoprenoid synthesis. Interestingly, the pathogenic *M. tuberculosis* exclusively relies on the MEP pathway for isoprenoid biosynthesis, which is vital for its growth and survival. For example, prenyl phosphates, a requisite for the biosynthesis of several cell wall components [19–21], and menaquinones, which play an important role in mycobacterial metabolism [22], are formed as a result of this pathway. Given the indispensable nature of the DXP/MEP pathway in these organisms and the absence of this pathway in humans, the constituent enzymes of the DXP/MEP pathway offer the potential for novel therapeutic intervention, perhaps with no toxicity issues to humans.

DXR is the second enzyme in the cascade of the MEP pathway, shown in Fig. 1. It is responsible for the conversion of DXP to MEP in the presence of a cofactor, nicotinamide adenine dinucleotide phosphate (NADPH), and a divalent metal ion (Mg^{2+}, Mn^{2+}, or Co^{2+}), see Fig. 2 [15]. DXP is a precursor not only for isopentenyl diphosphate, the common metabolic precursor of all isoprenoids, but also thiamine (vitamin B1) [23] and pyridoxal (vitamin B6) [24–26]. The importance of this enzyme is underscored by the identification of fosmidomycin, a naturally occurring antibiotic from *Streptomyces lavendulae* [27]. Fosmidomycin acts as a specific, mixed type (competitive and non-competitive) inhibitor of DXR with *K*i of 38 nM against *E. coli*, 600 nM against *Zymomonas mobilis* and an IC_{50} of 28 nM against *Plasmodium falciparum* DXR enzyme [14, 28, 29]. It is found to inhibit *Mt*DXR with an IC_{50} of 310 nM [30]. This compound can cure mice infected with *P. vinckei* [14], and is well tolerated in humans [29, 31]. Knockouts of DXR in *E. coli* [15, 32] and *Bacillus subtilis* [33] are lethal. These findings imply that DXR is a valid target, and inhibitors of this enzyme could serve as potential leads for antitubercular chemotherapy.

Rational inhibitor design relies both on mechanistic and structural information about the target enzyme. The three-dimensional (3-D) structure of a drug target, experimentally determined or theoretically modeled, particularly if it is complexed with a putative ligand, is vital to obtain a comprehensive understanding of the mode of action of a drug at the molecular level, to appreciate the structural consequences of genetic variations, and to construct accurate pharmacophores for drug design [34, 35]. In the absence of a crystallographic structure, a variety of advanced homology modeling methods have been developed, which can provide reliable models of proteins that share 30% or more sequence identity with a known structure [36]. Generally, it is assumed that the tertiary structures of the two proteins will be similar if their sequences are related [37]. The present work details the construction of a 3-D model for the DXR enzyme of *M. tuberculosis* using the X-ray crystal structures of the enzyme DXR from *E. coli* (PDB Id:1Q0L) and *Z. mobilis* (PDB Id:1R0L) as templates that share 43 and 40% sequence identity, respectively, with the *M. tuberculosis* sequence. For such a sequence identity, homology modeling yields a structure that is accurate enough for use in subsequent computational studies [38]. It is anticipated that the refined homology model along with available crystal structures could provide a starting point for the design of novel inhibitors.

Results and discussion

The basic idea of comparative modeling is to develop a 3-D structure of an enzyme making use of the collected body of knowledge from the previously resolved protein crystal structures. To date, several crystal structures of DXR from *E. coli* [39–43] and *Z. mobilis* [44] have been reported by different research groups. DXR is a flexible enzyme that experiences movement of C-terminal domain with respect to the other two domains upon binding of the cofactor and

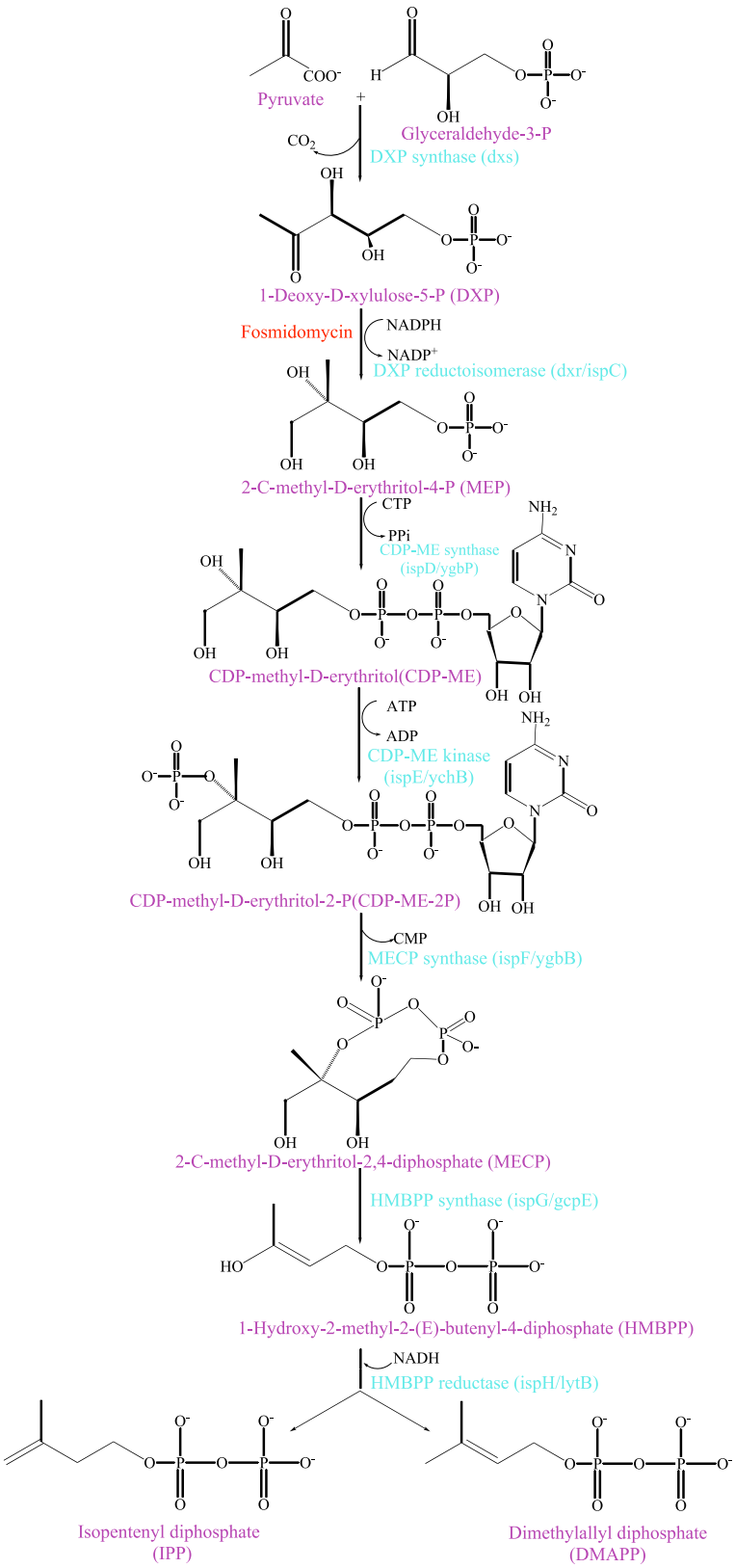

Fig. 1 Biosynthesis of IPP via the MEP pathway. The enzymes responsible for catalyzing each step are indicated while the corresponding genes are within parenthesis. The inhibitor fosmidomycin that blocks the second step of the pathway is highlighted in red

substrate/inhibitor within the binding pocket. Additionally, the active site loop is known to exist in different conformations [39–41, 43]. This flexibility allows the enzyme to adopt wide range of conformations according to the operating environment. The first published *E. coli* DXR (*Ec*DXR) crystal structure was in an apo-form (PDB

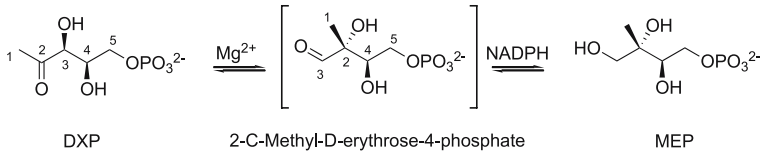

Fig. 2 The reaction catalyzed by DXR

Id:1K5H) [39], the most open form known for this enzyme. Thereafter, several structures with a single substrate or inhibitor in the substrate or cofactor binding sites were reported. The second structure was the *Ec*DXR complexed with NADPH and a sulfate ion (PDB Id:1JVS) [40]. Hereafter, a DXR–ligand complex structure with Mn^{2+} and fosmidomycin bound (PDB Id:1ONP) was reported along with two other structures of DXR: the apoenzyme (PDB Id:1ONN) and with a Mn^{2+} ion (PDB Id:1ONO) [41]. The structure of *Z. mobilis* DXR (*Zm*DXR) complexed with an acetate ion (PDB Id:1R0K) and as a binary complex with NADPH (1R0L) was also solved [44]. Crystallographic structures of DXR with two bisphosphonate inhibitors, a sulfate ion, with (PDB Id:1T1S) or without Mg^{2+} (PDB Id:1T1R) complexed, were published [42]. All these structures adopted an intermediate conformation between the open and closed form. However, once both substrate and cofactor binding site are occupied, the enzyme undergoes a substantial change in the conformation to adopt a closed conformation [43]. This transformation is obvious in the recently published ternary structures of the DXR–NADPH–fosmidomycin (PDB Id:1Q0L), the DXR–NADPH–substrate (PDB Id:1Q0Q) and the selenomethionine-labeled DXR–NADPH–fosmidomycin (PDB Id:1Q0H) complex. These structures are reported in a tight binding closed enzyme conformation. Since the 3-D structure of *Mt*DXR complexed with an inhibitor is not known, but the enzyme is a rational target for structure based molecular design, a homology model was developed.

Sequence alignment and model building procedure

Sequence alignment, a crucial step in homology modeling, is used to establish a one-to-one connection between the amino acids of the reference and the target protein [45]. The accuracy of the alignment is the single most important factor that determines the quality of a model [46]. The alignment of the deduced amino acid sequence of the *Mt*DXR with the sequences from *E. coli* (PDB Id:1Q0L) and *Z. mobilis* (PDB Id:1R0L) obtained from *ClustalW 1.8* [47] is shown in Fig. 3. This alignment shows that the protein sequence of the *Mt*DXR consists of a 22 amino acid long C-terminal tail that is unique to *M. tuberculosis.* However, the exact role of this extension has yet to be determined. The percentage identity between these sequences is approx. 40%, although higher sequence identity exists in the conserved catalytic domain (45–50%). Clearly, the differences in the sequences could be exploited to design improved inhibitors. The final alignment after manual modifications consisted of a total of 7 structurally conserved regions (SCRs) that constituted the major segment (about 96.05%) of total sequence.

The 3-D structure of *Mt*DXR was generated using the MODELER program [48–50] interfaced with *InsightII* software [51]. The recently published *Ec*DXR crystal structure in a tight binding closed conformation (PDB Id:1Q0L; 2.65 Å resolution) served as the template, which should logically be closer to the conformation adopted by the enzyme in a physiological environment. Thus, it will serve as a better tool to understand the critical ligand–protein interactions. The input parameters of MODELER were set to generate 20 models including 10 loop refinements, with the level of optimization set to high. Since it is known that N-terminal and C-terminal does not contribute significantly toward the binding of the ligands [52], some of the overhanging residues in these regions were truncated. Amongst the 20 different models generated, the one showing the lowest energy, restraint violations, and the number of main chain and side chain bad conformations was selected for further refinement (See supporting information; Table S1).

Model refinement and enzyme–inhibitor complex formation

The quality of the initial model was improved by subjecting it to a crude energy minimization protocol as detailed in the Materials and methods section. These minimizations helped relieve any steric clashes or improper geometries in the protein structure to produce a model with correct bond lengths and bond angles and where individual atoms are not too close together.

The binding of a ligand to the active site of a protein is typically associated with local and perhaps, also global structural rearrangement of the receptor (Induced-fit behavior). As a result, structure-based drug design preferentially relies on the structures of protein–ligand complexes containing bound ligands. Keeping this in mind, the next step was to develop a protein–ligand complex that would offer a more detailed and accurate picture of the interactions and structural complementarities between the ligands and the active site. Such a protein–ligand complex

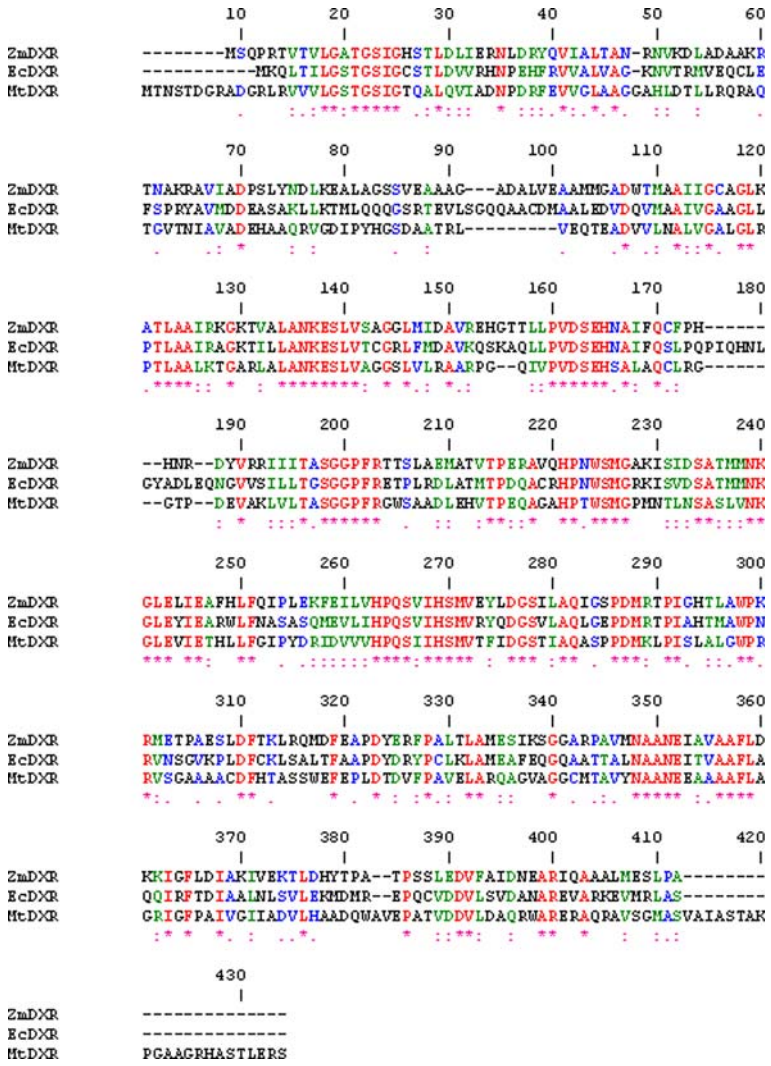

Fig. 3 Multiple sequence alignment using *ClustalW* (Conserved columns are colored red, columns with amino acids belonging to homology groups are colored in green, and columns without particularity in blue). (*) denotes identity; (:) is strongly similar and (.) is weakly similar

may be used more appropriately to perform docking simulation studies and may provide more meaningful structural models.

Consequently, a protein–ligand complex was developed from the initially generated homology model of the target protein by including information about known ligands as spatial restraints and optimizing the mutual interactions between the ligands and the binding sites. Assuming that the ligand binding modes are similar in the target and the template protein structures, the coordinates of ligands were transferred from the known DXR crystal structures keeping their orientation as a restraint for the subsequent modeling process. Energy minimization and restrained MD simulations were carried out on the protein–ligand complex in an aqueous environment, as detailed in the Materials and methods section. Care was taken to maintain the overall geometry of the model during the refinement. The stability of the model structure was finally investigated by using an unconstrained MD simulation. Our results show that the total, potential and the kinetic energies remained constant during the simulation. The system remained in equilibrium during the entire simulation. Based on our observations, we concluded that the model is stable at room temperature. The backbone root-mean-square-deviation (RMSD) between the final model and the template crystal structure is 1.76, while that between the starting model and the template is only 1.22 Å; therefore, the adopted minimization and refinement procedures did not appear to cause a substantial distortion in the structure and accomplished the goal of relieving steric clashes and close contacts. The small RMSD can also be interpreted to mean that the two structures share common structural homology and the generated structure is reasonable.

The refined 3-D structure of the resulting protein–ligand complex is displayed in Fig. 4. The final structure consisted of three domains: a larger N-terminal NADPH binding domain (residues 1–150), a connective or linker domain (residues 151–275) and a smaller C-terminal α-helical domain (residues 307–389). The N-terminal domain, a variant of the classical nucleotide-binding domain,

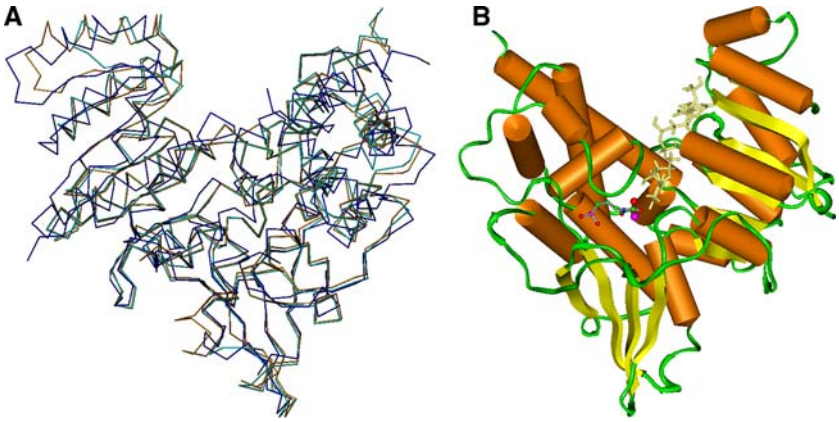

Fig. 4 (**A**) Superposition of the modeled structure of *Mt*DXR (purple) with the X-ray crystal structure of the *Ec*DXR (orange) and *Zm*DXR (blue). Only α-carbons are shown in this picture. (**B**) Schematic view of the final homology-model of *Mt*DXR. (NADPH (in yellow) as stick model; Fosmidomycin (colored by atom type) shown as ball and stick model; Mg^{2+} ion (in purple) and critical binding site water molecule (in red) as CPK models. Helices and sheets are represented as orange cylinders and yellow arrows, respectively

exhibited an α/β topology with a seven-stranded parallel β-sheet and seven α-helices. The connective domain comprises five α-helices and a four-stranded β-sheet with one parallel and two antiparallel alignments. The smaller C-terminal domain consisted of a four-helix bundle and is joined to the connective domain by an extended loop (residues 198–205), which spans the entire central domain.

The highly conserved residues M267 (M276 in *Ec*DXR), M205 (M214 in *Ec*DXR), and W203 (W212 in *Ec*DXR) shield the active site from the solvent, creating a hydrophobic pocket. The nicotinamide ring of NADPH also contributes to the formation of the binding pocket. Figure 5 shows a comparison between the active sites of the target and the template. A representation of the binding pocket in terms of molecular lipophilicity potential (MLP) generated by MOLCAD [53] has also been shown. The MLP demonstrates, in a straightforward way, the existence of such a hydrophobic pocket [54]. This, of course, is due to the fact that hydrophobicity is encoded in lipophilicity. Molecular lipophilicity potential is calculated from atomic lipophilicity contributions. Examining the surface lipophilicity potential can reveal parts of the molecule that are involved in the hydrophobic interactions and provide some hints concerning its mode of action. The active site contains the highly conserved residues D151 (D150 in *Ec*DXR), E153 (E152), E222 (E231), E225 (E234), H200 (H209), and H242 (H257), see Fig. 6. The phosphonate moiety of the inhibitor docked into the active site pocket of the modeled structure interacts with S177 (S186 in *Ec*DXR), S213 (S222), N218 (N227), and K219 (K228). All these residues are conserved in the phosphate/phosphonate-binding site. The hydroxamic acid moiety of fosmidomycin coordinates with the catalytically essential Mg^{2+} ion that in turn, is bound to the highly conserved residues, D151, E153, and E222. The octahedral coordination geometry required to bind the metal ion is completed through hydrogen bonding of a water molecule to another highly conserved residue, K128 (K125). The β-indole of W203 is at a distance of 3.49 Å from the fosmidomycin backbone that is known to provide a key interaction in the *Ec*DXR template.

The dinucleotide moiety of the cofactor NADPH binds to the N-terminal domain in the same fashion as in dinucleotide binding proteins [55] while the adenine ring is found to extend into the active site, participating in the formation of a hydrophobic binding pocket. The 2′ phosphate group of NADPH forms a hydrogen bond with T21 (T10 in *Ec*DXR), which is a highly conserved residue across the DXR family. The cofactor makes several hydrogen bonding interactions with strictly conserved residues G22 (G11), N127 (N124), K128 (K125), E129 (E126), and G206 (G215) that serve to anchors the cofactor into the binding pocket. Additionally, hydrogen bonds are also seen with H50 and N209, which are found exclusively in the amino acid sequence of *Mt*DXR. Other amino acid residues that are found lining the NADPH binding pocket are S23, I24, A103, L104, V105, G106, L108, D115, P208, and M267. Although there are differences in amino acid residues interacting with NADPH when compared with those found in other DXR structures, it does not alter the nicotinamide binding property. These interactions provide evidence that NADPH has a vital role in tight binding of the inhibitor within the enzyme active site.

Protein structure validation

The refined structure was evaluated for overall quality using available analysis procedures. These analyses compare specific properties of the model with those for known

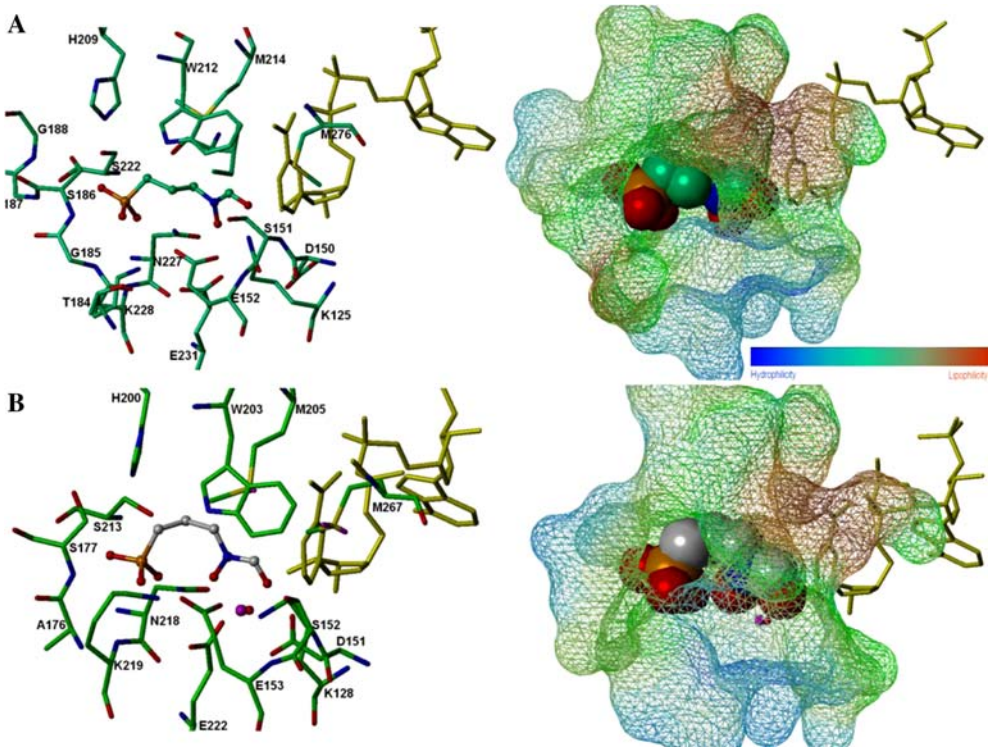

Fig. 5 Representation of the differences within the active sites of (**A**) *Ec*DXR and (**B**) *Mt*DXR. A depiction of the respective binding pockets in terms of MLP has also been shown. The color spectrum shown in the right shows the gradation of lipophilicity potential in the respective enzymes

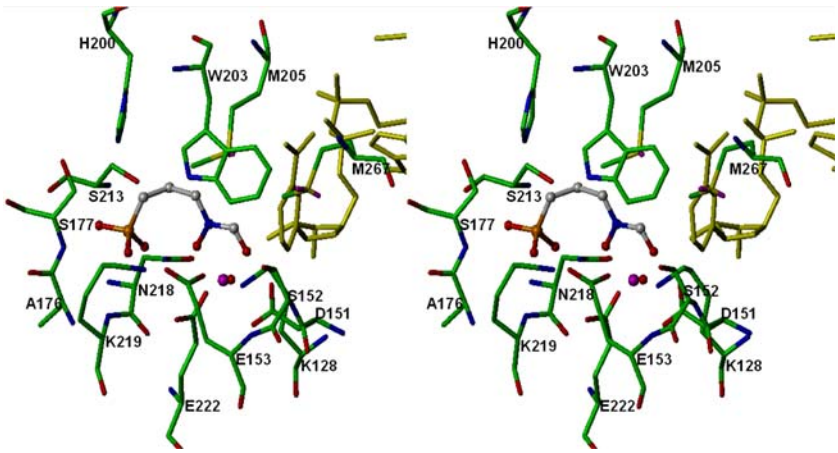

Fig. 6 Stereoscopic view of the binding mode of fosmidomycin within the active site of *Mt*DXR with a prominent hydrophilic cavity: fosmidomycin represented as ball and stick model (colored by atom types); NADPH indicated as capped stick model (colored yellow); metal ion and water molecule indicated as CPK models (in purple and red, respectively)

high quality protein structures. For this purpose, three protein analysis programs were used: PROCHECK [56], PROSTAT, and Profile-3-D.

PROSTAT was used to assess the stereochemical quality of the model. This program verifies the accuracy of parameters such as bond lengths, bond angles, torsion angles, and correctness of amino acid chirality. The cut-off used was 10 standard deviations from the reference value. No spurious angle or bond length was detected in our model. The results are listed at the bottom of Table 1. Using these geometric criteria, the *Mt*DXR model compares well with the X-ray structure of the *Ec*DXR structure used as a template.

Another important indicator of the stereochemical quality of the model is the distribution of the main chain torsion angles phi and psi which may be examined in a Ramachandran plot. The phi–psi plots are shown in Fig. 7, while the more detailed results are listed in Table 1. The plot clearly shows the vast majority of the amino acids are in a phi–psi distribution consistent with right-handed

Table 1 Results of protein structure check by PROCHECK and PROSTAT

	*Ec*DXR	*Mt*DXR
Residues in most favored regions	332 (93.0%)	273 (84.3%)
Residues in additional allowed regions	24 (6.7%)	43 (13.3%)
Residues in generously allowed regions	1 (0.3%)	8 (2.5%)
Residues in disallowed regions	0 (0.0%)	0 (0.0%)
Number of non-glycine and non-proline residues	357	324
Number of end-residues (excl. Gly and Pro)	2	2
Number of glycine residues (shown as triangles)	21	34
Number of proline residues	18	20
Total number of residues	398	380
Overall PROCHECK score[a]	0.19	–0.28
Number of bond distances with significant deviations[b] (by PROSTAT)	0	0
Number of bond angles with significant deviations[b] (by PROSTAT)	0	0
Number of dihedral angles with significant deviations[b] (by PROSTAT)	2	2

[a] Recommended value ≥0.50 and investigation is needed for ≤1.0

[b] Number of instances for which the property differs more than 10 standard deviations from the reference value

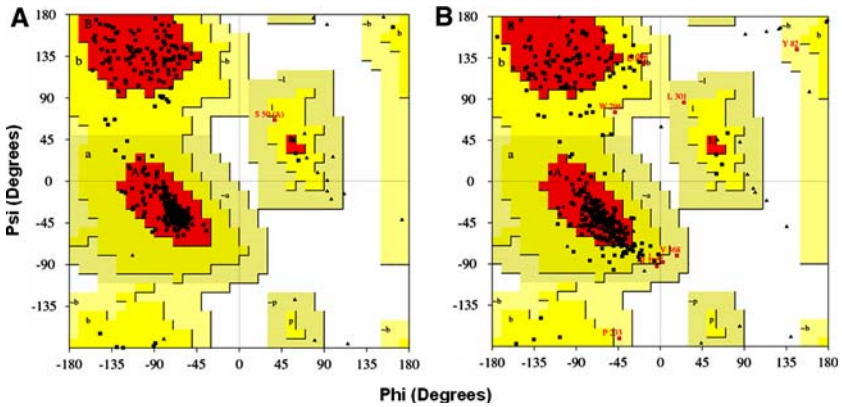

Fig. 7 Ramachandran plot of (**A**) the X-ray structure of *Ec*DXR (**B**) and the homology-modeled structure of *Mt*DXR. The different colored areas indicate "disallowed" (white), "generously allowed" (light yellow), "additional allowed" (yellow), and "most favored" (red) regions (refer Table 1)

α-helices. The remaining residues that fall into the random or beta configuration geometries are very short segments and are primarily in the loop regions of the protein. A quick check of amino acid side chain integrity did not reveal any amino acid side chain clashes with one another or with the backbone. No residues are outside the allowed regions. Figure 7 shows that the structure is reasonable overall because the spot distribution for the homology-modeled structure was similar to the standard X-ray structure of *Ec*DXR. The results showed that our modeled structure was comparable with the standard crystal structures.

Finally, the 3-D homology model was verified using the Profile-3-D program in *InsightII* software and then compared with the *Ec*DXR crystal structure, shown in Fig. 8. A smoothing window size of 10 residues was used. The analysis yielded an overall score of 174.78 similar to the typical score of 188.87 for a native protein of equivalent size and well above 84.99, a score that would indicate an incorrect structure. For comparison, analysis for the *E. coli* crystal structure used as a template yielded a score of 181.51, while the range for the structure was 81.68–188.59, indicating the reliability of the model.

In summary, the above-mentioned analyses indicate that the model structure is consistent with our current understanding of the protein structure.

Comparison with crystal structure of *Mt*DXR

Just before submission of this manuscript, a crystal structure of *Mt*DXR (PDB Id:2C82) complexed with a sulfate ion was published [57]. Comparison of the X-ray structure to our homology model of *Mt*DXR offers another way of independently validating our results. With a Cα RMSD of 1.98 Å, the *Mt*DXR homology model is very similar to the X-ray structure with only a few differences. The crystal structure is in an intermediate conformation between the closed and open form of the enzyme. However, our model represents *Mt*DXR enzyme in a closed, tight binding conformation. It has been shown that in order to assume a tight

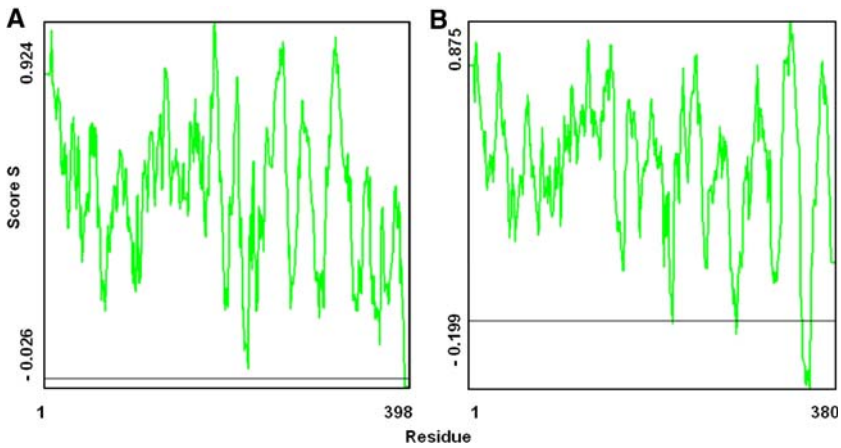

Fig. 8 The evaluation of the DXR modeled structure by Profile-3-D program. Profiles for (**A**) the template, *Ec*DXR (**B**) the *Mt*DXR comparative model

binding closed conformation, it is necessary for the protein to have both the cofactor and the ligand placed appropriately within the active pocket. The overall secondary and tertiary structures are very similar for the model and the reported X-ray structure [57]. The active site residues are highly conserved in both structures and show similar alignment within the binding pocket although the side chains of the acidic residues coordinating with the metal ion seem to be disoriented against their position in the model where the presence of the metal ion allows it to adopt the requisite conformation, a difference obvious in Fig. 9.

Another conformational difference between our model and the X-ray structure of *Mt*DXR was found in the active site loop, where amino acid residues H200 and M205 lie far from the putative active site in the crystal structure.

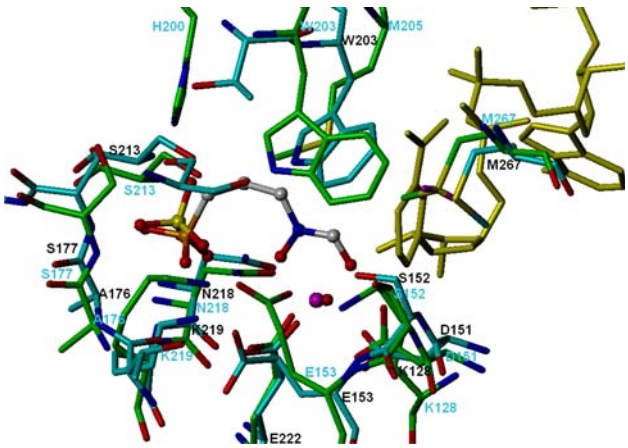

Fig. 9 The superimposed active sites of the published *Mt*DXR crystal structure with a sulfate ion (as ball and stick model; colored by atom types) and the comparative model with fosmidomycin (as ball and stick model; colored by atom types), NADPH (as stick model; colored yellow), active site water molecule and Mg^{2+} ion (as CPK models; colored red and purple, respectively). The published crystal structure is shown in cyan and the modeled enzyme-inhibitor complex is shown in green

Residue H200 has been implicated in binding the substrate/inhibitor in the correct orientation for catalysis [39]. However, in the active site of the *Mt*DXR crystal structure this residue appears to be oriented away from the active pocket. This result differs from our model where this residue is found within the active pocket making a hydrogen bond with S177 to lock the protein in a closed conformation. Interestingly, in the *Ec*DXR crystal structure, this residue does not show hydrogen bonding with the inhibitor [43]. However, in our previous study with *P. falciparum* DXR homology model, a hydrogen bond between histidine and phosphonate moiety of fosmidomycin was observed [58]. It has been shown in *Ec*DXR that H209Q mutation results in a 5200-fold decrease in k_{cat}/K_m, stressing the importance of this residue [59]. Despite the differences, the position of the sulfate ion present in the binding pocket of the crystal structure corresponds to that of the phosphonate moiety of the inhibitor and is found to interact similarly, see Fig. 9. Therefore, analyses of the available crystal structure and homology model reveals that the overall shape of the active site is nearly identical. Since our modeled protein–ligand complex is thought to represent a conformation closer to that existing under physiological conditions, it seems logical to use it for future structure-based studies.

Materials and methods

Computational resources

Molecular modeling was carried out on a Silicon Graphics Octane2 R12000 dual processor workstation running on Irix 64 (SGI, 1600 Amphitheater Parkway, Mountain View, CA 94043). For sequence alignment *CLUSTALW 1.8* in *InsightII 2005*/HOMOLOGY module were used. Comparative modeling was performed using *InsightII*/MODELER. Energy minimization and molecular dynamic simulations

were carried out using *InsightII*/DISCOVER module. Ramachandran plots and statistics were generated through the use of PROCHECK 3.4.

Sequence alignment and comparative model

The amino acid sequences of the target protein, *Mt*DXR (GenBank Accession No. P64012), *Ec*DXR (GenBank Accession No. P45568) and *Zm*DXR (GenBank Accession No. Q9X5F2) were extracted from the NCBI protein sequence database (http://www.ncbi.nlm.nih.gov/). The *Mt*DXR sequence is composed of 413 residues. Sequence alignment was derived with the *CLUSTALW 1.8* package and *Align123* in *InsightII*/HOMOLOGY module using the BLOSUM matrices [49] for scoring the alignments. Procuring a high quality multiple sequence alignment of the target with the templates is a central step in protein homology modeling in view of the fact that only small local errors made in alignment can be corrected in the subsequent steps. Openings and extension GAP penalty were changed systematically, and the obtained alignment was checked for deletions and insertions in structurally conserved regions and finally fine-tuned manually. The PHD algorithm [60] was used to predict the secondary structure of *Mt*DXR. Although the X-ray crystal structures of DXR from both *Z. mobilis* (PDB Id:1R0L) and *E. coli* (PDB Id:1Q0L) are available in Protein Data Bank [61], the 3-D structure of *Ec*DXR was used as a template since it is the only structure known in a ternary complex with NADPH and fosmidomycin that reveals a tight binding closed conformation.

The resulting alignment was used as an input for the automated homology modeling program MODELER interfaced with the *InsightII* software suite. This program assigns atomic coordinates to regions structurally aligned with the template, builds intervening loops, optimizes the rotamers of amino acid side chains, and performs an initial energy optimization of the structure. MODELER generates protein 3-D structures by satisfying spatial restraints imposed by the sequence alignment with the template structure. These spatial restraints consist of homology-derived restraints on the distances and dihedral angles in the target sequence based on its alignment with template structures, stereochemical restraints such as bond length and bond angle terms obtained from CHARMm forcefield [62], and statistical preferences for dihedral angles and non-bonded inter-atomic distances obtained from a representative set of protein structures. These restraints are expressed in terms of probability density functions (PDF). A 3-D protein model is obtained by optimizing the probability density function with the variable target function procedure in cartesian space that employs methods of conjugate gradients and molecular dynamics with simulated annealing. To guarantee sufficient conformational sampling of each active site residue, several homology models are generated in this step.

Initial model refinement

Hydrogen atoms were added to the selected model. By explicitly specifying the pH as 7.0, the protonation states for all the amino acid side chains are set in accordance with their typical pK_a values. Consistent Valence ForceField (CVFF) parameters, a distance dependent dielectric constant and a non-bond cutoff distance of 10.0 Å were employed. The initial model was energy-minimized, employing the *InsightII*/DISCOVER module, in a stepwise fashion following a standard procedure consisting of 500 steps of steepest descent and 1,000 steps of conjugate gradient minimization with an rms gradient of the potential energy of 0.001 kcal/mol Å in each step. In the first step, all newly added hydrogen atoms were minimized followed by energy minimization of the loop side chains, loop main chains, all core side chains, and finally the core main chains.

Enzyme–inhibitor complex formation

For the purpose of complex protein formation, coordinates of fosmidomycin, metal ion, and a tightly bound water molecule in the binding site were extracted from the *Ec*DXR crystal structure (PDB Id:1ONP). Bond orders and the correct protonation states were assigned upon visual inspection. The coordinates for the cofactor, NADPH were transferred from PDB Id:1Q0L, and the previously stated procedure was carried out. A subsequent minimization of the transferred cofactor with the CVFF force field, as implemented in biopolymer module of *InsightII*, revealed no significant movements. The metal ion, Mn^{2+} present in the crystal structure (PDB Id:1ONP), was converted to Mg^{2+} since it is the relevant divalent cation in vivo [52] and also because the software used in this study are better parameterized for the magnesium ion.

The resulting complex was embedded in a 5 Å shell of 1,799 water molecules to imitate aqueous solvent conditions. The assembled molecules were subjected to energy minimization by DISCOVER module included in *InsightII*. Steepest descents were carried out until maximum derivative of 1.0 kcal/mol Å, and conjugate gradients were followed until maximum derivative of 0.1 kcal/mol Å were obtained. This minimized solvated system was used as the starting structure for the following MD simulations. In order to relax the position of water molecules, an MD

simulation for 20 ps by fixing the solute structure was performed. Following this, an equilibrium MD simulation for 600 ps at 300 K was carried out. It was observed that temperature, pressure, density, and total energy were well equilibrated. In order to assess the time-dependent behavior of our molecular system, 1.2 ns of production run was performed. In all the simulations, bonds involving hydrogen were constrained by applying the SHAKE algorithm to the system, and a time step of 1 fs was used.

Protein structure validation

The quality of the final refined model was assessed by subjecting it to a series of tests for its internal consistency and reliability. The PROSTAT module of *InsightII* was used to analyze the properties of bonds, angles, and torsions. The Profile-3-D program in *InsightII* was used to check the structure and sequence compatibility, while the PROCHECK suite of programs was used for assessing the "stereochemical quality" of the modeled protein structure. Also, RMSD values were calculated for the minimized homology model and compared to the template structures. RMSDs were also calculated for the structures before and after minimization to show there were no gross movements, indicating the suitability of the template used.

Conclusions

In summary, the 3-D structure of *Mt*DXR was developed using homology modeling and validated by bioinformatics techniques. A model structure of the enzyme–cofactor–inhibitor complex was also obtained, which can be used as a significant tool to augment our understanding of interaction of DXR inhibitors with the protein at the atomic level. This information could also aid in the understanding of its catalytic mechanism. The generated model of *Mt*DXR is very similar in structure to other known DXRs from *E. coli* and *Z. mobilis*, having the same features of overall fold and disposition of active site loop. Analyses of currently available crystal structures and the generated model revealed that a high homology exists in the active site region of the known DXRs and the essential amino acid residues are highly conserved and interact similarly with the ligands within the binding pocket. However, distinct differences are observed in the NADPH binding site residues.

The homology model was found to closely reproduce the crystal structure that became publicly available in the final stages of this work. However, our model has an advantage of being representative of the enzyme in the ligand-bound tight binding closed conformation believed to exist under physiological conditions. This model should prove useful in the design and development of inhibitors as potential novel therapeutic agents against TB by either de novo drug design or virtual screening of large chemical databases.

Furthermore, the optimization of the interactions in the NADPH binding site and with small differences within the active site taken into consideration could counteract the weaker metal binding of a monodentate ligand, providing alternatives to the currently used hydroxamic acids binding units. Thus, understanding the mechanism and possible reasons for enzymatic selectivity will further the development of drugs, with the potential to make a major impact against TB.

Acknowledgment This work is funded by the CDC Cooperative agreement number U50/CCU423310-02.

References

1. Kaufmann SH, McMichael AJ (2005) Nat Med 11:S33
2. WHO World Health Organization (2006) Fact sheet on tuberculosis. http://www.who.int/mediacentre/factsheets/fs104/en/print.html
3. Corbett EL, Watt CJ, Walker N, Maher D, Williams BG, Raviglione MC, Dye C (2003) Arch Intern Med 163:1009
4. Singh SK, Verma R, Shah DH (2004) J Vet Sci 5:331
5. Bloom BR, Murray CJ (1992) Science 257:1055
6. Heym B, Honore N, Truffot-Pernot C, Banerjee A, Schurra C, Jacobs WR Jr, van Embden JD, Grosset JH, Cole ST (1994) Lancet 344:293
7. Perlman DC, El Sadr WM, Heifets LB, Nelson ET, Matts JP, Chirgwin K, Salomon N, Telzak EE, Klein O, Kreiswirth BN, Musser JM, Hafner R (1997) AIDS 11:1473
8. Rattan A, Kalia A, Ahmad N (1998) Emerg Infect Dis 4:195
9. Cole ST, Barrell BG (1998) Novartis Found Symp. 217:160
10. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV et al (1998) Nature 393:537
11. Rohmer M, Knani M, Simonin P, Sutter B, Sahm H (1993) Biochem J 295:517
12. Schwarz MK (1994) Ph.D. Dissertation, Eidgeno ssischen Technischen Hochschule, Zurich
13. Broers STJ (1994) Ph.D. Dissertation, Eidgeno ssischen Technische Hochschule Zurich, 1994
14. Jomaa H, Wiesner J, Sanderbrand S, Altincicek B, Weidemeyer C, Hintz M, Turbachova I, Eberl M, Zeidler J, Lichtenthaler HK, Soldati D, Beck E (1999) Science 285:1573
15. Takahashi S, Kuzuyama T, Watanabe H, Seto H (1998) Proc Natl Acad Sci U S A 95:9879
16. Proteau PJ (2004) Bioorg Chem 32:483
17. Lange BM, Croteau R (1999) Proc Natl Acad Sci U S A 96:13714
18. Bochar DA, Friesen JA, Stauffacher CV, Rodwell VW (1999) Comprehensive natural product chemistry. Pergamon, Oxford, p 1544
19. Wolucka BA, McNeil MR, de Hoffmann E, Chojnacki T, Brennan PJ (1994) J Biol Chem 269:23328
20. Mahapatra S, Yagi T, Belisle JT, Espinosa BJ, Hill PJ, McNeil MR, Brennan PJ, Crick DC (2005) J Bacteriol 187:2747
21. Mikusova K, Mikus M, Besra GS, Hancock I, Brennan PJ (1996) J Biol Chem 271:7820

22. Minnikin DE (1982) In: Ratledge C, Stanford J (eds) The biology of Mycobacteria. Academic Press, London, England, p 95
23. Sprenger GA, Schorken U, Wiegert T, Grolle S, de Graaf AA, Taylor SV, Begley TP, Bringer-Meyer S, Sahm H (1997) Proc Natl Acad Sci U S A 94:12857
24. Mueller C, Schwender J, Zeidler J, Lichtenthaler HK (2000) Biochem Soc Trans 28:792
25. Schwender J, Seemann M, Lichtenthaler HK, Rohmer M (1996) Biochem J 316:73
26. Qureshi AA, Porter JW (eds) (1981) Biosynthesis of isoprenoids. John Wiley, New York, p 47
27. Lois LM, Campos N, Putra SR, Danielsen K, Rohmer M, Boronat A (1998) Proc Natl Acad Sci U S A 95:2105
28. Kuzuyama T, Shimizu T, Takahashi S, Seto H (1998) Tetrahedron Lett 39:7913
29. Lell B, Ruangweerayut R, Wiesner J, Missinou MA, Schindler A, Baranek T, Hintz M, Hutchinson D, Jomaa H, Kremsner PG (2003) Antimicrob Agents Chemother 47:735
30. Dhiman RK, Schaeffer ML, Bailey AM, Testa CA, Scherman H, Crick DC (2005) J Bacteriol 187:8395
31. Missinou MA, Borrmann S, Schindler A, Issifou S, Adegnika AA, Matsiegui PB, Binder R, Lell B, Wiesner J, Baranek T, Jomaa H, Kremsner PG (2002) Lancet 360:1941
32. Rodriguez-Concepcion M, Campos N, Maria Lois L, Maldonado C, Hoeffler JF, Grosdemange-Billiard C, Rohmer M, Boronat A (2000) FEBS Lett 473:328
33. Kobayashi K, Ehrlich SD, Albertini A, Amati G, Andersen KK, Arnaud M, Asai K et al (2003) Proc Natl Acad Sci U S A 100:4678
34. Whittle PJ, Blundell TL (1994) Annu Rev Biophys Biomol Struct 23:349
35. Blundell TL (1996) Nature 384:23
36. Burley SK (2000) Nat Struct Biol 7 Suppl:932
37. Kroemer RT, Doughty SW, Robinson AJ, Richards WG (1996) Protein Eng 9:493
38. Baker D, Sali A (2001) Science 294:93
39. Reuter K, Sanderbrand S, Jomaa H, Wiesner J, Steinbrecher I, Beck E, Hintz M, Klebe G, Stubbs MT (2002) J Biol Chem 277:5378
40. Yajima S, Nonaka T, Kuzuyama T, Seto H, Ohsawa K (2002) J Biochem (Tokyo) 131:313
41. Steinbacher S, Kaiser J, Eisenreich W, Huber R, Bacher A, Rohdich F (2003) J Biol Chem 278:18401
42. Yajima S, Hara K, Sanders JM, Yin F, Ohsawa K, Wiesner J, Jomaa H, Oldfield E (2004) J Am Chem Soc 126:10824
43. Mac Sweeney A, Lange R, Fernandes RP, Schulz H, Dale GE, Douangamath A, Proteau PJ, Oefner C (2005) J Mol Biol 345:115
44. Ricagno S, Grolle S, Bringer-Meyer S, Sahm H, Lindqvist Y, Schneider G (2004) Biochim Biophys Acta 1698:37
45. Lee KW, Briggs JM (2004) Proteins 54:693
46. Krieger E, Nabuurs SB, Vriend G (2003) Methods Biochem Anal 44:509
47. Thompson JD, Higgins DG, Gibson TJ (1994) Nucleic Acids Res 22:4673
48. Sali A, Blundell TL (1993) J Mol Biol 234:779
49. Henikoff S, Henikoff JG (1992) Proc Natl Acad Sci U S A 89:10915
50. Fiser A, Do RK, Sali A (2003) Protein Sci 9:1753
51. InsightII Molecular Modeling Program Package, Accelrys: 9685 Sranton Road, San Diego, CA 92121-3752, USA, 2005
52. Argyrou A, Blanchard JS (2004) Biochemistry 43:4375
53. Brickmann J (1997) MOLCAD-MOLecular Computer Aided Design, Darmstadt University of Technology. The major part of the MOLCAD program is included in the SYBYL package of TRIPOS associates, St. Louis, MI, USA
54. Heiden W, Moeckel G, Brickmann J (1993) J Comput Aided Mol Des 7:503
55. Bellamacina CR (1996) FASEB J 10:1257
56. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) J Appl Crystallogr 1993:283
57. Henriksson LM, Bjorkelid C, Mowbray SL, Unge T (2006) Acta Crystallogr D Biol Crystallogr 62:807
58. Singh N, Cheve G, Avery MA, McCurdy CR (2006) J Chem Inf Model 46:1360
59. Kuzuyama T, Takahashi S, Takagi M, Seto H (2000) J Biol Chem 275:19928
60. Rost B, Sander C (1999) J Mol Biol 232:584
61. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) Nucleic Acids Res 28:235
62. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) J Comp Chem 4:187