BIOCHEMISTRY

 Volume 45, Number 23 June 13, 2006

Accelerated Publications

The 2.15 Å Crystal Structure of *Mycobacterium tuberculosis* Chorismate Mutase Reveals an Unexpected Gene Duplication and Suggests a Role in Host−Pathogen Interactions†

Rohini Qamra,‡ Prachee Prakash,§ Bandi Aruna,§ Seyed E. Hasnain,§,||,⊥ and Shekhar C. Mande*,§

Department of Biophysics, University of Delhi South Campus, New Delhi, India, Centre for DNA Fingerprinting and Diagnostics, Hyderabad, India, and Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India

Received April 2, 2006; Revised Manuscript Received May 3, 2006

ABSTRACT: Chorismate mutase catalyzes the first committed step toward the biosynthesis of the aromatic amino acids, phenylalanine and tyrosine. While this biosynthetic pathway exists exclusively in the cell cytoplasm, the *Mycobacterium tuberculosis* enzyme has been shown to be secreted into the extracellular medium. The secretory nature of the enzyme and its existence in *M. tuberculosis* as a duplicated gene are suggestive of its role in host−pathogen interactions. We report here the crystal structure of homodimeric chorismate mutase (Rv1885c) from *M. tuberculosis* determined at 2.15 Å resolution. The structure suggests possible gene duplication within each subunit of the dimer (residues 35−119 and 130−199) and reveals an interesting proline-rich region on the protein surface (residues 119−130), which might act as a recognition site for protein−protein interactions. The structure also offers an explanation for its regulation by small ligands, such as tryptophan, a feature previously unknown in the prototypical *Escherichia coli* chorismate mutase. The tryptophan ligand is found to be sandwiched between the two monomers in a dimer contacting residues 66−68. The active site in the "gene-duplicated" monomer is occupied by a sulfate ion and is located in the first half of the polypeptide, unlike in the *Saccharomyces cerevisiae* (yeast) enzyme, where it is located in the later half. We hypothesize that the *M. tuberculosis* chorismate mutase might have a role to play in host−pathogen interactions, making it an important target for designing inhibitor molecules against the deadly pathogen.

Chorismic acid is the last common precursor in the aromatic amino acid biosynthesis pathway and is a substrate for multiple enzymes. Various enzymes that utilize chorismate as a substrate include chorismate mutase (CM),[1] anthranilate synthase, isochorismate synthase, and *p*-aminobenzoate synthase. Chorismate mutase (EC, 5.4.99.5), responsible for the production of tyrosine and phenylalanine, catalyzes the conversion of chorismate to prephenate. It is generally agreed that this conversion proceeds via a pericyclic (Claisen-type) rearrangement involving a chairlike transition state (*1*). CM is one of the rare enzymes catalyzing the pericyclic isomerization of chorismate to prephanate and is

† The work was supported by grants from the Department of Biotechnology, New Delhi. S.C.M. is a Wellcome Trust International Senior Research Fellow.

* To whom correspondence should be addressed: Centre for DNA Fingerprinting and Diagnostics, ECIL Road, Nacharam, Hyderabad 500 076, India. Phone: +91-40-27171442. Fax: +91-40-27155610. E-mail: shekhar@cdfd.org.in.

‡ University of Delhi South Campus.

§ Centre for DNA Fingerprinting and Diagnostics.

|| Jawaharlal Nehru Centre for Advanced Scientific Research.

⊥ Present address: University of Hyderabad, Gachibowli, Hyderabad, India.

[1] Abbreviations: CM, chorismate mutase; *Mtb*CM, *M. tuberculosis* chorismate mutase.

10.1021/bi0606445 CCC: $33.50
Published on Web 05/18/2006

involved in the shikimate pathway for biosynthesis of aromatic amino acids (*2*). The enzyme channels chorismate toward the biosynthesis of phenylalanine and tyrosine and away from that of tryptophan and therefore represents an important check point for regulating the balance of aromatic amino acids in the cell.

Structures of CMs from *Escherichia coli* (*3*), *Bacillus subtilis* (*4*), *Saccharomyces cerevisiae* (*5*−*7*), and *Thermus thermophilus* (*8*) are known. Structural comparisons show that CMs from different organisms have evolved into two completely unrelated protein folds, suggesting separate evolutionary origins of the enzyme. On the basis of the structural fold adopted by the protein, CMs have been classified into the AroH and AroQ classes. Enzymes of the AroH class adopt a trimeric pseudo-α/β-barrel fold and include the monofunctional CMs from *B. subtilis* and *T. thermophilus* (*4*, *8*). The AroQ class of enzymes, on the other hand, is completely helical and includes the bifunctional enzymes from *E. coli* and the monofunctional yeast CM. Intriguingly, despite the different tertiary structures adopted by the proteins in the two CM classes, the mode of substrate recognition in either of them is similar, suggesting a convergence of the catalytic mechanism (*9*).

On the basis of sequence homologies, it has emerged that among the two CM classes, the AroQ CMs are more prevalent in nature and are widely distributed among the prokaryotes. In most prokaryotes, CMs exist as bifunctional enzymes possessing an associated prephenate dehydrogenase or a prephenate dehydratase activity (*10*). Examples of multiple copies of the CM gene in the same organism are also known, where at least one of the gene products exists as a fusion protein (*11*). The eukaryotic CMs, on the other hand, are generally monofunctional with the enzyme activity being regulated by the aromatic amino acids (*6*, *12*). Thus, substantial variations exist among the eukaryotic and prokaryotic enzymes, even within the AroQ class. The crystal structure reported in this paper corresponds to the AroQ class CM of *Mycobacterium tuberculosis*.

The availability of the *M. tuberculosis* genome sequence and its re-annotation later led to the identification of two putative genes, encoded by Rv1885c and Rv0948c open reading frames, encoding the CM enzyme in the organism (*13*, *14*). Although both these proteins have been shown to possess CM activity, the amino acid sequences of the two are less than 15% identical (*15*, *16*). *M. tuberculosis* CM (*Mtb*CM), encoded by Rv1885c, shares ∼20% sequence identity with the *E. coli* enzyme, while the the level of identity with yeast CM is ∼10%. Despite such a low degree of sequence similarity, however, the Rv1885c gene product has been shown to be a functional enzyme possessing CM activity (*15*, *16*). *Mtb*CM was shown to be monofunctional in nature with no other associated enzyme activity (*15*). Moreover, unlike the other known prokaryotic CMs, the *M. tuberculosis* enzyme exhibited allosteric regulation by aromatic amino acids, a feature limited to the eukaryotic CMs (*7*, *12*, *15*). Interestingly, CM has been shown to be secreted in many bacteria and nematodes such as *Erwinia herbicola*, *Burkholderia fungorum*, *Pseudomonas aeruginosa*, *Rhodococcus equi*, *Salmonella typhi*, *Salmonella typhimurium*, *Yersinia pestis*, *Melidogyne javanica*, and *Heterodera glycines* (*11*, *17*−*19*). Intriguingly, the secreted CMs (designated as *AroQ) from a few of these pathogenic bacteria and the nematodes have been shown to be involved in virulence (*18*−*21*). Indeed, *Mtb*CM when expressed heterologously in *E. coli* is found to be localized in the periplasm (*15*). The secretion of the enzyme has been attributed to the 33 N-terminal residues, deletion of which has been shown to keep the enzyme in the cytoplasm when expressed in *E. coli* (*15*). Thus, the existence of *Mtb*CM in the extracellular space suggests that the secreted enzyme might have evolved to play a similar role in *M. tuberculosis* and aid in its pathogenesis.

M. tuberculosis resides inside the macrophages that play an important role in inducing or regulating the innate and T-cell-mediated adaptive immune responses in the host. Given the extracellular existence of *Mtb*CM and the importance of the *AroQ proteins in virulence in other organisms, secretory *Mtb*CM is likely to mediate interactions with the host macrophage cells and play a role in virulence. We have determined the crystal structure of the unique extracytoplasmic *Mtb*CM in complex with its allosteric ligand, L-tryptophan (Trp), at 2.15 Å resolution. The structure reveals an interesting proline-rich fragment on the protein surface. We believe that the proline-rich surface of the protein might be involved in mediating binding of *Mtb*CM to the macrophage cell surface receptor(s). The overall study offers an understanding of the important role of *Mtb*CM in mediating host−pathogen interactions and the mechanisms involved therein.

MATERIALS AND METHODS

Crystallization and Data Collection. The *M. tuberculosis* gene, Rv1885c, encoding CM was expressed in the heterologous host *E. coli* BL21(DE3) and purified as described previously (*22*). For production of the selenomethionyl (Se-Met) CM, the gene was overexpressed in *E. coli* BL21(DE3) grown in minimal medium, supplemented with the 19 naturally occurring L-amino acids, except methionine, at a concentration of 50 mg/L. Methionine was replaced with seleno-L-methionine at the same concentration, and cells were grown for 16 h after induction with 0.1 mM IPTG at 37 °C. Purification of the Se-Met protein was carried out as described previously (*22*). Crystals of Se-Met CM were grown by vapor diffusion by equilibrating 3 μL of protein at a concentration of 5 mg/mL against a reservoir containing 0.2 M Li_2SO_4 and 22% PEG 8000 in 0.1 M sodium acetate buffer (pH 4.6) over a period of 4−5 days. The crystallization drop also contained 1 mM tryptophan pre-equilibrated with the protein, which corresponds to an approximately 3.5-fold molar excess of the protein concentration. Crystals were soaked in an artificial mother liquor of the same composition but supplemented with 26% (v/v) glycerol for cryoprotection for ∼1 min before freezing them for data collection. Data were collected at tunable beam ID-23 at the European Synchrotron Radiation Facility (ESRF) (Grenoble, France). A fluorescence spectrum was measured around the Se K-absorption edge. Data were measured at the absorption peak (λ = 0.97955 Å; $\Delta\varphi$ = 1°, total φ = 180°), inflection point (λ = 0.97975 Å; $\Delta\varphi$ = 1°, total φ = 180°), and the low-energy remote point (λ = 0.9809 Å; $\Delta\varphi$ = 1°, total φ = 180°) using an ADSC Quantum-4 detector. Data were processed using MOSFLM and SCALA (*23*).

Structure Determination and Refinement. Positions of Se atoms were determined by the anomalous difference Patter-

Table 1: Data Collection Statistics of the Native and Se-Met Crystals[a]

	native	peak	inflection	remote
space group	*C*2	*C*2	*C*2	*C*2
a (Å)	124.5	124.6	124.8	124.9
b (Å)	83.8	84.1	84.3	84.4
c (Å)	62.7	62.9	63.0	63.0
β (deg)	93.4	93.5	93.5	93.6
wavelength (Å)	1.00	0.97955	0.97975	0.9809
resolution range (Å)	70−2.15	30−2.2	30−2.2	30−2.2
no. of unique reflections	36542 (2491)	32649 (4702)	32844 (4736)	32846 (4733)
redundancy	3.6 (1.7)	3.9 (3.9)	3.9 (3.9)	3.9 (3.9)
completeness (%)	96.7 (75.6)	99.2 (98.7)	99.2 (98.7)	99.1 (98.5)
average $I/\sigma(I)$	10.1 (2.8)	18.6 (6.8)	16.8 (4.6)	15.0 (2.7)
R_{merge}[b] (%)	10.9 (28.2)	5.2 (15.7)	5.5 (27.1)	6.3 (43.7)
anomalous completeness (%)		98.9 (98.0)	99.0 (98.0)	98.8 (97.8)
anomalous multiplicity		1.9 (1.9)	1.9 (1.9)	1.9 (1.9)

[a] Values in the parentheses are for the highest-resolution shell (2.19−2.15 Å for native CM and 2.32−2.20 Å for the Se-Met CM crystal). [b] $R_{merge} = \sum|I_{hkl} - \langle I_{hkl}\rangle|/\sum|I_{hkl}|$, where I_{hkl} are the intensities and $\langle I_{hkl}\rangle$ is the average intensity over all observations.

Table 2: Refinement Statistics

no. of protein atoms	3882
no. of subunits per asymmetric unit	3
no. of atoms per asymmetric unit	4167
no. of waters per asymmetric unit	218
resolution range (Å)	50−2.15
R_{work}[a,b] (%)	18.0 (19.6)
R_{free}[b,c] (%)	22.4 (29.3)
rmsd from ideal values	
bond distances (Å)	0.017
bond angles (deg)	1.466
average *B*-factor (Å^2)	
overall	42.4
main chain	40.3
side chain	44.5
water	43.4
sulfate ion	56.6
Trp-ligand	50.5
structure validation	
Procheck Ramachandran plot (%)	
core	96.2
allowed	3.3
Molprobity	
Clashscore (all atoms)	3.81 (99th percentile)
Clashscore ($B < 40$)	0 (100th percentile)
rotamer outliers	1.44%
C^β deviations of >0.25 Å	0
PDB entry	2AO2

[a] $R_{work} = \sum|F_o - F_c|/\sum|F_o|$. [b] Values in parentheses represent those for the highest-resolution shell (2.21−2.15 Å). All the observed reflections were used in the refinement without any σ cutoff. [c] R_{free} is the *R*-factor calculated from a subset of reflections (5%) excluded from refinement.

son method using RSPS (*23*), and the two positions that were identified were used as input for SHARP (*24*). Phases were calculated using SHARP followed by density modification using the solvent flipping option of SHARP. The electron density maps generated with the correct hand were of exceptional quality. A molecular model for 160 residues was constructed using wARPnTRACE (*25*). The molecular model was refined against the native data collected previously at beam line BL41XU of SPring-8 (*22*). Refinement was carried out using the target maximum likelihood function as available in REFMAC (*26*) and CNS (*27*). Models were built in O (*28*) by using σ_A-weighted $2F_o - F_c$ and $F_o - F_c$ electron density maps. Structure quality was assessed with PROCHECK (*29*) and MOLPROBITY (*30*). Figures were prepared using MOLSCRIPT and RASTER3D (*31*, *32*). Structural superpositions were carried out in O (*28*) using C^α atoms.

RESULTS

Crystallization, Structure Determination, and Refinement. In an attempt to understand the structural basis of the immunomodulatory nature of *Mtb*CM, the crystal structure of the enzyme was determined. Se-Met *Mtb*CM was crystallized in space group *C*2 in the presence of Trp. Cell parameters and scaling with the diffraction amplitudes of the native crystals suggested that the Se-Met *Mtb*CM crystals were isomorphous with the native crystals (Table 1). Anomalous Patterson maps led to the identification of two strong peaks corresponding to the Se positions, which were refined using SHARP. The initial map allowed the construction of a molecular model with 160 residues in each chain. Unexpectedly, wARPnTRACE constructed three chains in the asymmetric unit, whereas only two molecules were anticipated from the Matthews coefficient (*33*). The asymmetric unit therefore contains one and one-half dimers. The final structure contains, in addition to the three polypeptide chains, 218 water molecules. Five sulfate ions were also clearly visible in the σ_A-weighted $2F_o - F_c$ and $F_o - F_c$ density maps and were included in the later parts of the refinement. Moreover, the $F_o - F_c$ difference maps showed large connected density at the $>5\sigma$ level, which water molecules could not account. Since the crystallization buffers contained 1 mM Trp, and the crystals could not be grown in its absence, two Trp molecules were modeled in this density in the later stages of refinement. The Trp ligands were seen to be sandwiched symmetrically between two monomers in a dimer, contacted by residues Pro66, Ile67, and Glu68. The *R*-factor and R_{free} of the final model are 0.180 and 0.224, respectively (Table 2).

Overall Structure of M. tuberculosis Chorismate Mutase. The final refined structure consists of residues 35−199 in three chains designated A−C. We did not observe electron density for the 34 N-terminal residues. These residues are probably disordered or cleaved during heterologous expression in *E. coli*. It has previously been shown that the recombinant protein when expressed in *E. coli* is secreted from the cytoplasmic space (*15*, *16*). Secretion of *Mtb*CM is facilitated by the presence of a signal sequence comprising 33 residues at the N-terminus, which is cleaved after secretion (*16*). The absence of electron density for the first 34 residues at the N-terminus hence might suggest cleavage of the signal peptide from the polypeptide chain prior to crystallizations.

The *M. tuberculosis* enzyme is an all-helical protein, similar to the *E. coli* and yeast CMs. The polypeptide chain is folded into eight α-helical segments, numbered 1−8 (Figure 1a). Helix 3, spanning residues 90−119, forms the longest helix and is involved in dimerization across the two subunits. This helix also contributes to the allosteric site for the binding of Trp residues involved in enzyme regulation, as discussed below. Helix 7 (residues 159−176) is juxtaposed with respect to helix 8 (residues 179−190) in such a manner

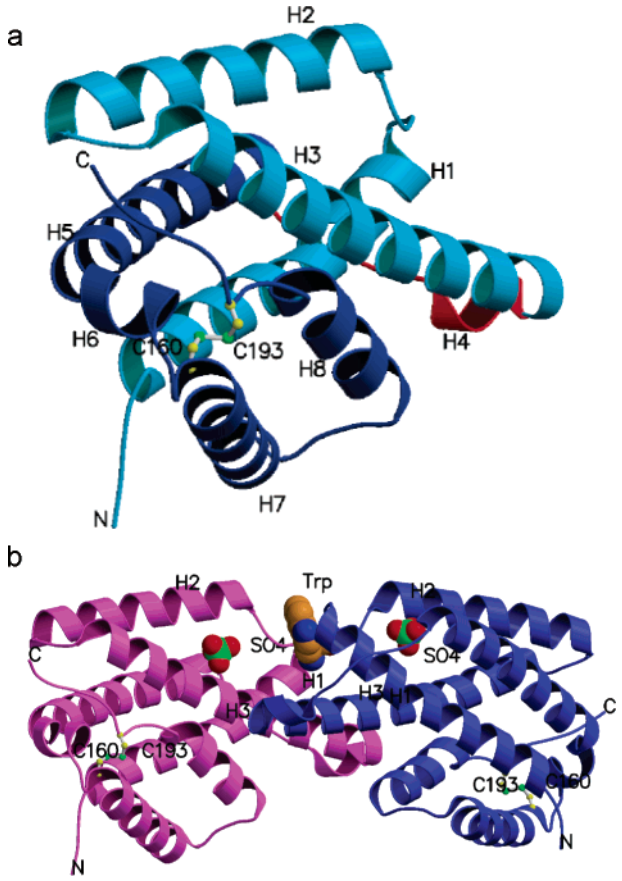

FIGURE 1: Overall structure of *M. tuberculosis* CM. (a) The eight helices shown in the monomer are numbered serially. Two parts of the structure that arise out of gene duplication are colored in different shades of blue. The Pro-rich region connecting these two halves is colored red. The disulfide linkage between residues 160 and 193 is also shown. (b) The dimer is shown in the same orientation with the two chains colored differently. The allosteric modulator, tryptophan, shown in CPK, binds at the interface of the dimeric CM. Positions of sulfate ions, to mark the active site in each monomer, are indicated.

that Cys160 forms a disulfide linkage with Cys193 (Figure 1a). Helix 8 leads to a segment of polypeptide, which appears to be an unwound helix, but cannot be classified into a defined secondary structure. Cys193 is placed just after helix 8 in the unwound part of the polypeptide. Formation of the disulfide linkage between Cys160 and Cys193 would not have been possible if the geometry of this stretch was maintained as a typical α-helix.

The three polypeptide chains can be superimposed very well with a root-mean-square deviation (rmsd) between chains A and B of 0.62 Å for 165 C^{α} atoms and between chains A and C of 0.50 Å for 165 C^{α} atoms. The only deviant stretch in the three polypeptide chains is the region spanning residues 119−130. Electron density at the loop region spanning these residues is poorly defined in all the three chains. This region is a proline-rich fragment, which connects two distinct segments of the polypeptide. Thus, despite minor differences in residues 119−130, the overall conformation of the protein is essentially the same in all the three chains.

Mode of Dimerization. Size exclusion results have previously shown that *Mtb*CM exists as a dimer (*15*, *16*). Since the crystal asymmetric unit contains three CM molecules, calculation of the buried surface area across different monomers was used to identify the physiological dimer. Surface area calculations reveal that two of the three molecules in the asymmetric unit bury an area of 1591 $Å^2$, suggesting that these molecules form a biological dimer (Figure 1b). The third molecule in the asymmetric unit forms the biological dimer across the crystallographic 2-fold symmetry axis. The two dimers are essentially identical in the crystallographically independent subunits, and the subsequent discussion pertains to chains B and C, which are related through noncrystallograpic 2-fold symmetry.

Dimerization is mediated through residues spanning helix 3 in both subunits, with the two helices placed antiparallel to one another. A majority of the interactions are mediated by association between side chains of the two adjacent subunits through salt bridges and hydrogen bonding. One of the interactions is formed by an extensive hydrogen bonding network through the guanidino group of Arg103. The guanidino group of Arg103 forms a hydrogen bond with the main chain carbonyl oxygen of Lys117 in the other subunit, and with the side chain amide of Gln190 of the same subunit. Another crucial network brings the side chain carboxyl oxygens of Glu68 and Glu106 into close contact with the N^{ζ} group of Lys117. Moreover, this interaction appears to be crucial in relaying the allosteric effects of Trp, since the Trp forms a hydrogen bonding network (Trp-Tyr110-Glu68-Lys117-Glu106), whereas Glu106 is a part of the active site (Figures 3 and 4). Thus, intimate side chain−side chain and side chain−main chain interactions contribute to the association of the monomers in the biological dimer.

Structural Analysis and Comparison with E. coli CM Suggest Possible Gene Duplication in MtbCM. The *Mtb*CM structure can be thought to consist of two distinct segments, 35−119 and 130−199, as mentioned above. Comparison of the two halves of *Mtb*CM reveals a weak degree of similarity. Superposition between the two halves, i.e., residues 35−119 and residues 130−199, yielded a structural alignment with a rmsd of 2.13 Å for 51 C^{α} atoms. Structural similarity between the two halves of *Mtb*CM is therefore suggestive of a possible gene duplication event in *Mtb*CM. The two halves are connected via a highly flexible proline-rich region occurring between helices 3 and 5. This intervening proline-rich loop appears to form an "intergenic" region connecting the gene-duplicated polypeptide of *Mtb*CM.

E. coli CM exists as a homodimer with each polypeptide adopting an all-helical fold consisting of three helices in each subunit (*3*). Sequence comparisons of *Mtb*CM with *E. coli* CM reveal a low level of pairwise sequence identity of <20%. Moreover, *E. coli* CM is a small protein of 90 amino acids, while *Mtb*CM is larger and consists of 199 amino acid residues. Despite these differences, however, structural comparisons of *Mtb*CM with the *E. coli* monomer yield a superposition with a rmsd of 1.97 Å for 78 C^{α} positions (Figure 2a). This superposition aligns helices 1−3 of *E. coli* CM with helices 1−3 of *Mtb*CM. Moreover, the same rigid body transformation also aligns the second monomer of *E. coli* CM with helices 5, 7, and 8 of *Mtb*CM. Thus, a single polypeptide chain of *Mtb*CM can simultaneously be superimposed onto both chains of the *E. coli* CM dimer. Such a superposition results in an overall rmsd of 1.42 Å for 117 C^{α} atoms. Since the two halves of *Mtb*CM are structurally homologous to each other and both halves can be simultaneously superimposed with the *E. coli* CM dimer, *Mtb*CM

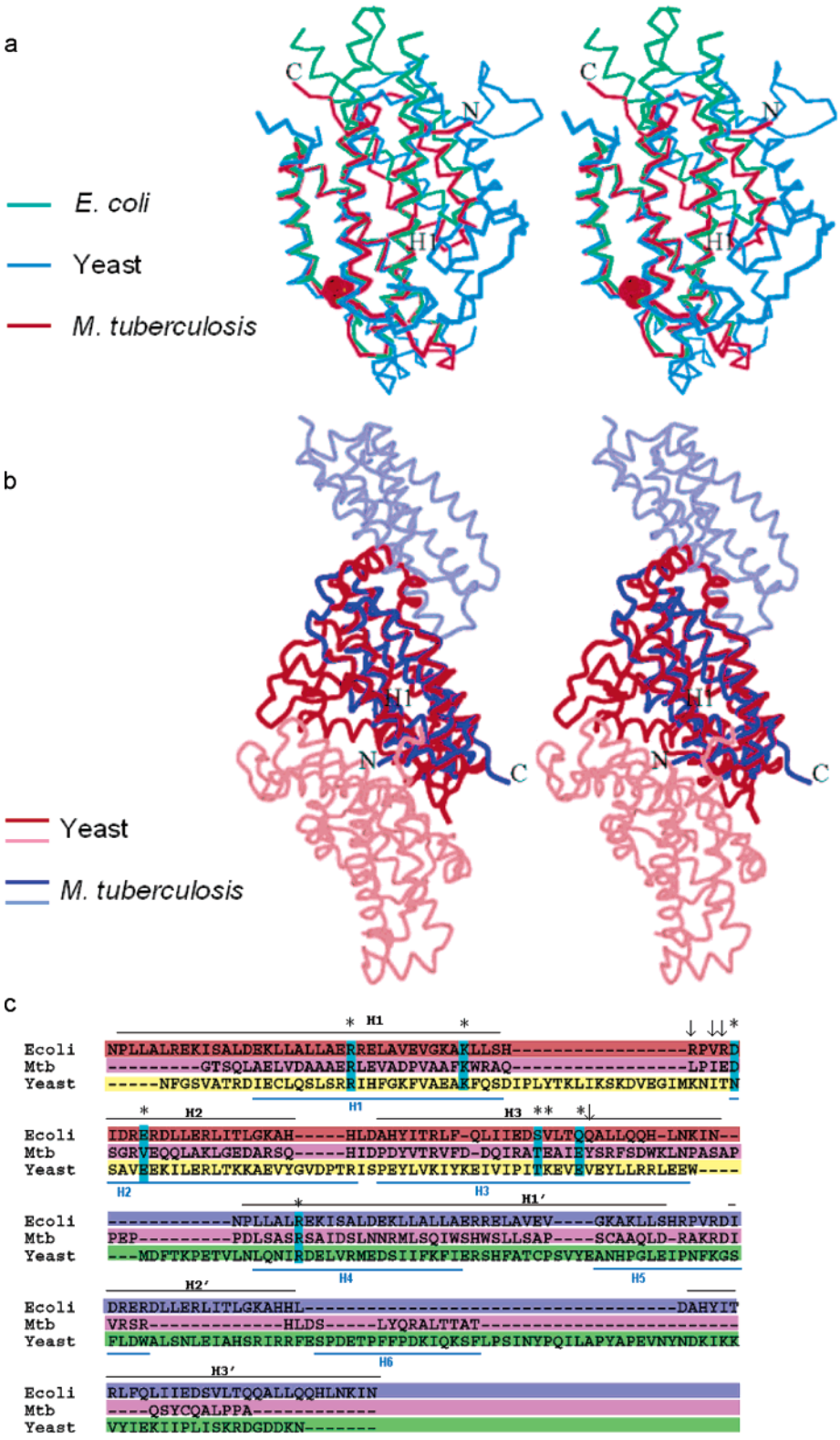

FIGURE 2: Comparisons of the overall structures of *M. tuberculosis*, *E. coli*, and yeast CMs. (a) Stereoview of the superposed CMs from *M. tuberculosis* (red), *E. coli* (green), and yeast (blue). *E. coli* CM is shown in its dimeric state, while the yeast and *M. tuberculosis* CMs are shown in their monomeric form. A remarkable conservation of the four-helix folds is seen. Positions of the active site sulfate ions in *Mtb*CM are shown. (b) Stereoview of the different modes of dimerization in the yeast CM and *Mtb*CM. The two chains of yeast CM are colored in different shades of red, while those of *M. tuberculosis* are colored in different shades of blue. Despite similar tertiary structures, the two proteins form dimers in a very different manner. (c) Structure-guided sequence alignment of CMs from *M. tuberculosis*, *E. coli*, and yeast. The *Mtb*CM sequence is colored pink. The two identical polypeptides of *E. coli* CM are colored orange and blue. The yeast CM sequence is colored yellow (C-terminus) and green (N-terminus). The conserved active site residues are colored cyan. The C-terminus of yeast CM aligns well with the N-terminus of *Mtb*CM. Arrows indicate residues that interact with the allosteric Trp, while asterisks denote the active site residues. Positions of α-helices in *E. coli* (black bar above the sequences) as well as *Mtb*CM (blue bar below the sequences) are also shown for comparison.

might indeed be a result of gene duplication. Gene duplication within the *Mtb*CM cannot be discerned by sequence analysis between the two halves of *Mtb*CM, or between the *M. tuberculosis* and *E. coli* sequences, but is strikingly evident from comparisons of the two structures.

Despite significant similarity between the *E. coli* and *Mtb*CM structures, the most noticeable deviation occurs in the shortening of helices 1 and 5 of *Mtb*CM, which correspond to helix 1 of both subunits of *E. coli* CM. These two helices in *Mtb*CM appear to be two-thirds as long as the corresponding helices in the *E. coli* enzyme. While helix 1 of *Mtb*CM appears to be truncated at the N-terminus, helix 5 is truncated at its C-terminus. Similarly, larger deviations toward the C-terminal region of the polypeptide, such as unwinding of helix 8 of *Mtb*CM, are also observed.

Comparison with the Yeast Chorismate Mutase. The level of pairwise sequence identity between the *M. tuberculosis* and yeast CMs is less than 10%, yet the tertiary structure of the *Mtb*CM superposes well with that of yeast CM, with a rmsd of 1.52 Å for 114 C^{α} atoms. Structural comparisons between yeast CM and *Mtb*CM using C^{α} atoms reveal that the N-terminal half of *Mtb*CM (residues 35−119) superposes very well with the C-terminal half of yeast CM (residues 141−256). This resulting alignment matches helices 1−3 of the *M. tuberculosis* protein with helices 8, 11, and 12 of yeast CM (Figure 2b). Interestingly, the second half of the *M. tuberculosis* protein, i.e., helices 5, 7, and 8, aligns well with helices 2, 4, and 7 of yeast CM, respectively. It is interesting to note that the structural alignment matches the C-terminal domain of yeast CM with the N-terminal half of *Mtb*CM and vice versa (Figure 2c).

An important difference in the yeast and *M. tuberculosis* CMs is the mode of dimerization of the two subunits (Figure 2b). The yeast CM exists as a strong dimer that is evident from a large buried surface area of 4078 $Å^2$ between the two monomers. The site of dimerization in yeast CM lies at the interface comprised of helices 2, 4, 8, and 11 and connecting loop regions 50s and 80s (nomenclature as described in ref *7*). Dimerization of *Mtb*CM, on the other hand, involves contributions from helix 3 and the loop spanning residues 64−69 and buries a much smaller area of 1591 $Å^2$. Thus, despite similar tertiary structures among the different CMs, the quaternary associations between the individual monomers are very different (Figure 2b).

Active Site Residues of MtbCM. The active site of *E. coli* CM is made up by contributions from both its subunits. On the other hand, the active site in *Mtb*CM lies within each polypeptide chain and is contributed by residues in helices 1−3 and 5. The active site is therefore made up of contributions from both the N-terminal and C-terminal halves of the polypeptide. This is very similar to the case for yeast CM where a single polypeptide folds into a functional unit. Since the yeast (*7*) and *M. tuberculosis* CMs (this work) are believed to have evolved from duplication of an ancestral CM gene, formation of the active site in a single polypeptide chain appears to be a consequence of this duplication event.

Although the overall level of sequence identity among the *E. coli*, yeast, and *M. tuberculosis* CM sequences is very low, the structure-guided sequence alignment shows a remarkable conservation of the active site residues in *Mtb*CM (Figure 2c). Of the seven residues contributing to the active site of the AroQ class of CMs, four residues are conserved

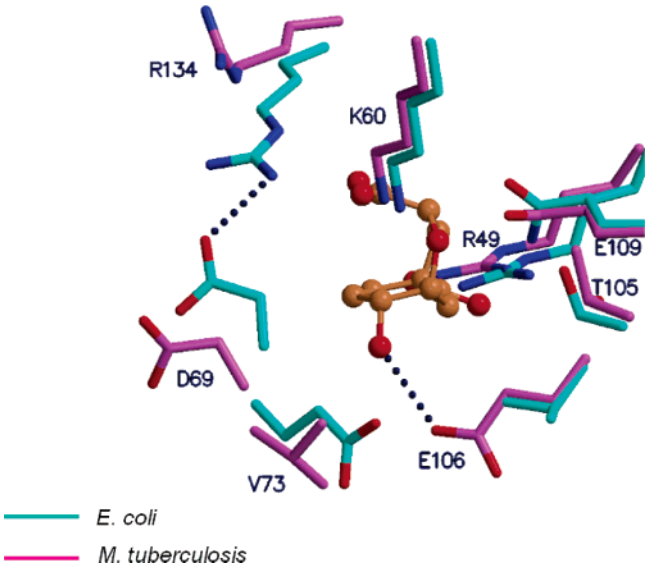

FIGURE 3: Comparison of the *E. coli* and *M. tuberculosis* CM active sites. Superposition of the *E. coli* (cyan) and *M. tuberculosis* (magenta) CMs shows a remarkable conservation of the substrate binding residues. The absence of a crucial glutamate at position 73 in *Mtb*CM is compensated by Glu106, which is now positioned for interaction with the substrate through H-bonding. The endo-oxabicyclic transition state analogue colored brown is a part of the *E. coli* CM structure (PDB entry 1ECM).

in *Mtb*CM, while two are conservative replacements, i.e., Ser84Thr and Gln88Glu (numbers corresponding to the *E. coli* sequence; Figure 2c). The major variation is the replacement of a glutamate, Glu52 in *E. coli* CM, with a valine, Val73, in *Mtb*CM. Glu52 in *E. coli* CM has been shown to be crucial, where its side chain carboxyl is important in stabilizing the transition state. The corresponding Glu198 in yeast CM is also crucial for the enzyme activity as demonstrated through site-directed mutagenesis studies (*34*). Intriguingly, despite this important difference, *Mtb*CM seems to have retained its activity (*15*, *16*). This is most likely due to the presence of Glu106, which occurs at a structurally analogous position, thereby compensating for the loss of a carboxyl group at position 73 (Figure 3). The side chain of Glu106 in *Mtb*CM is appropriately placed to mediate the necessary interactions with the substrate. Therefore, despite the alteration at a crucial residue in *Mtb*CM, the active site geometry is maintained by a compensating mutation.

Another residue that has been shown to play an important role in CM catalysis is Glu109 (Gln88 in *E. coli* and Glu246 in yeast). This glutamate is involved in interactions with the ether O of the substrate and thereby has been shown to restrict the enzyme activity to acidic pH (*34*). Comparison of enzyme activities at different pHs shows that the yeast CM has a pH optimum in the acidic range of pH 5−6 while the *E. coli* enzyme shows maximal activity at the physiological pH of 7.4. This effect of pH on enzyme activity has been attributed to the presence of glutamate in the yeast (Glu246) and glutamine in the *E. coli* (Gln88) enzymes (*34*). In *Mtb*CM, the residue at the corresponding position is a glutamate (Glu109) and is therefore expected to behave like the yeast enzyme, yet paradoxically, the *M. tuberculosis* enzyme has a pH optimum of 7.5 (*15*). The difference in the pH profile of *Mtb*CM cannot, therefore, be explained easily, and perhaps detailed electrostatics calculation can shed more light on it.

The active site of *Mtb*CM is seen to be blocked due to the presence of a sulfate ion in the structure. The presence of the ion is attributed to its inclusion in the crystallization buffers. Side chains of residues Arg72, Arg134, Lys60, and Gln76 interact with the ion, making a strong ionic association. Interaction with the sulfate ion places the side chains of these residues so that the active site is completely closed for entry to the substrate. It therefore appears that sulfate ion acts as an inhibitor of the enzyme by blocking entry of the substrate into the active site.

Interactions of the Allosteric Regulator with MtbCM. Difference electron density maps showed the presence of a Trp at the dimer interface of the enzyme. This site is occupied in both dimers in the crystal asymmetric unit. The binding site of Trp overlaps with the 2-fold symmetry axis; hence, two "half-Trp's" have been modeled at this site in different orientations with each tryptophan being modeled with half-occupancy. Similarly, the Trp at the interface of the noncrystallographically related dimers is also present at the noncrystallographic 2-fold axis and has therefore been modeled with half-occupancy in two different conformations. The conformation of Trp modeled on the noncrystallographic 2-fold axis is the same as that modeled on the crystallographic 2-fold axis. Within a dimer, the two half-Trp's make identical interactions with both monomers, and interactions of one of the Trp's are depicted in Figures 1b and 4. The Trp at the dimer interface interacts extensively with residues from both subunits (Figure 4). While interactions with one subunit are mostly of the hydrogen bonding type, it makes van der Waals interactions with the other subunit. The carboxyl group of tryptophan bound at the dimer interface interacts with the main chain amide group of residues Ile67 and Glu68. The amino group of Trp forms a close contact with the hydroxyl OH group of Tyr110, while the indole NH group is involved in interaction with the main chain carbonyl oxygen of Leu65. The indole ring also forms several van der Waals interactions with residues from another subunit in the dimer. Binding of Trp at the dimer interface creates a network of interactions involving the allosteric Trp, Tyr110, Glu68, Glu106, and Lys117 (Figure 4).

The site of binding of the allosteric regulators, Trp and Tyr, in yeast CM also lies at the dimer interface. However, this allosteric site is placed very far from the substrate-binding pocket and therefore enacts a large helical movement during the allosteric transition in yeast CM (*7*). Interestingly, in *Mtb*CM, while one face of helix 3 contributes to residues involved in monomer−monomer contacts and the allosteric site, the other face contributes to residues involved in the interaction with the substrate. As a result, the allosteric site in *Mtb*CM is close to its substrate binding site. It is therefore likely that helix 3 might play an important role in allostery and hence the regulatory effect observed in the *M. tuberculosis* enzyme.

Pro residues dominate the peptide segment between the apparently duplicated regions of the *Mtb*CM polypeptide (residues 120−128). This region adopts a left-handed poly-Pro-II conformation. Since such conformational regions are known to bind to cell signaling proteins, such as the SH3 domains (*35*), it is likely that the secreted *Mtb*CM binds cell surface signaling proteins.

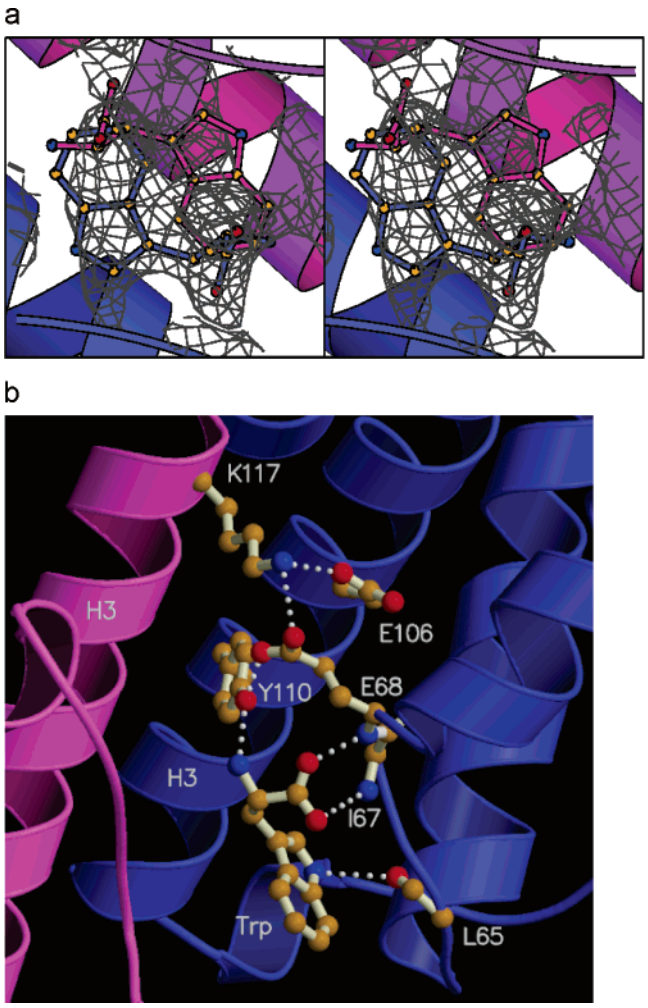

FIGURE 4: Interactions of the allosteric ligand, tryptophan, at the dimer interface. (a) The $F_o - F_c$ difference map shows large density at the interface of the two monomers. Two Trp molecules could account for the density when they were constrained symmetrically. Two Trp molecules were therefore modeled with half-occupancy each as mentioned in the text in detail. (b) Tryptophan interacts with one of the CM polypeptides through main chain and side chain interactions, facilitating a hydrogen bonding network across the two subunits.

DISCUSSION

The shikimate pathway involved in the biosynthesis of many important aromatic compounds such as siderophores, menaquinones, and aromatic amino acids has been shown to be indispensable in *M. tuberculosis* (*36*, *37*). CM constitutes one of the key enzymes at the branch point in this pathway and is involved in the biosynthesis of aromatic amino acids. Recently, two independent studies have shown that the product of the gene, Rv1885c, from *M. tuberculosis* possesses a monofunctional CM activity (*15*, *16*).

Biosynthesis of aromatic amino acids in most organisms is carried out exclusively in the cytoplasm, and hence, the role of CM is believed to be confined to the cytosol. Intriguingly, however, *Mtb*CM has been shown to be secreted in the extracellular medium (*15*, *16*). CMs from several other bacteria, including *E. herbicola*, *S. typhimurium*, and *P. aeruginosa*, have also been shown to exist in the periplasm (*11*, *17*). The periplasmic nature of the various CMs has however remained perplexing. Earlier studies suggest the role of the secretory CMs in virulence among a few of these

pathogens such as *S. typhimurium* and *Rhodococcus equi* (*20*, *21*). It is interesting to note that products of the genes upstream and downstream of *Mtb*CM are secretory in nature (*13*). It therefore appears that the entire region on the *M. tuberculosis* genome flanking Rv1885c might be involved in pathogenesis and might represent an important virulence determinant. It is pertinent to note that while a duplicate set of CMs occurs in the *M. tuberculosis* genome, only one of the gene products is secreted into the extracellular space. This observation strongly suggests that while the cytosolic enzyme might have retained its role in the aromatic amino acid synthesis pathway, the secreted enzyme is likely to have evolved a distinct role, such as the one aiding the bacterium in its pathogenesis.

Structural analysis suggests that evolution of the CM gene might have undergone a gene duplication event as suggested by the similarity of the two halves of *M. tuberculosis* CM. Such evidence was not discernible from prior sequence comparisons. The possible gene duplication event in *Mtb*CM suggests that the enzyme would constitute two active sites. Intriguingly, however, only one active site is observed in the *M. tuberculosis* enzyme, reasons for which are elegantly apparent from its crystal structure. Loss of one of the two possible active sites is a consequence of an apparent truncation of helices 1, 5, and 8 of *Mtb*CM. Truncation of the N-terminal part of helix 1 results in the loss of the crucial Arg11, which contributes to the active site in *E. coli* CM. Similarly, C-terminal truncation of helix 5 and consequent loss of the region corresponding to the loop joining helices 1 and 2 (nomenclature of *E. coli* CM) also severely affects the active site. Thus, one of the active sites in *Mtb*CM appears to have been abolished postduplication.

Another possible reason for the observed gene duplication event might be the necessity for allosteric regulation (*38*). Indeed, the yeast and *M. tuberculosis* CMs are allosterically regulated by aromatic amino acids, unlike the *E. coli* enzyme (*15*, *39*). However, molecular details of the allostery appear to be distinctly different in the two organisms. These differences may be attributed to the dissimilar modes of dimerization, the diverse location of the binding site of the allosteric ligand, and the conformational changes occurring thereof in the two proteins. In yeast CM, the binding sites for the allosteric regulators and the substrate are placed far from each other. As a result, the allosteric regulators effect large conformational changes across the two monomers (*6*). In *Mtb*CM on the other hand, the allosteric site is close to the active site of the enzyme. Placement of Trp at the dimer interface and its proximity to the main chain NH group of Glu68 appear to be vital for relaying the regulatory effect at the active site in the enzyme. Moreover, the side chain of Glu68 is in contact with the crucial carboxyl of Glu106 through a bridging Lys117. Thus, allostery in *Mtb*CM presumably arises from alterations in the side chain conformations of a few residues, such as Glu106, Tyr110, and Glu68. Major main chain conformational changes, on the other hand, are unlikely to occur in *Mtb*CM.

The structure of *Mtb*CM reported in this work offers several new insights into its function. The unexpected gene duplication possibly leads to a different allosteric regulation mechanism than that is known for other CMs. Moreover, its secretory nature and the presence of a Pro-rich region suggest that it might be involved in host−pathogen interactions. Clearly, cell signaling experiments and those probing its possible immunomodulatory nature will enhance our understanding of its function.

NOTE ADDED IN PROOF

We have noticed the structure of *Mtb*CM reported in a different crystal form in the absence of the allosteric ligand, Trp. Ökvist, M., Dey, R., Sasso, S., Grahn, E., Kast, P., and Krengel, U. (2006) 1.6 Å crystal structure of the secreted chorismate mutase from mycobacterium tuberculosis: Novel fold topoly revealed, *J. Mol. Biol.*, *357*, 1483−1499.

ACKNOWLEDGMENT

The kind help of the beamline staff at SPring-8 and ESRF is gratefully acknowledged. We thank Fred Vellieux for help during data collection at ESRF and Amit Sharma and ICGEB for access to the computer graphics facility. We are also grateful to Edward Baker for useful discussions.

REFERENCES

1. Haslam, E. (1993) in *Shikimic acid: Metabolism and metabolites*, Wiley, New York.
2. Weiss, U., and Edwards, J. M. (1980) in *The biosynthesis of aromatic amino acids*, John Wiley and Sons, New York.
3. Lee, A. Y., Karplus, P. A., Ganem, B., and Clardy, J. (1995) Atomic structure of the buried catalytic pocket of *Escherichia coli* chorismate mutase, *J. Am. Chem. Soc. 117*, 3627−3628.
4. Chook, Y. M., Gray, J. V., Ke, H., and Lipscomb, W. N. (1994) The monofunctional chorismate mutasefrom *Bacillus subtilis*, *J. Mol. Biol. 240*, 476−500.
5. Xue, Y., Lipscomb, W. N., Gray, R., Schnappauf, G., and Braus, G. (1994) The crystal structure of allosteric chorismate mutase at 2.2-Å resolution, *Proc. Natl. Acad. Sci. U.S.A. 91*, 10814−10818.
6. Strater, N., Hakansson, K., Schnappauf, G., Braus, G., and Lipscomb, W. N. (1996) Crystal structure of the T state of allosteric yeast chorismate mutase and comparison with the R state, *Proc. Natl. Acad. Sci. U.S.A. 93*, 3330−3334.
7. Strater, N., Schnappauf, G., Braus, G., and Lipscomb, W. N. (1997) Mechanisms of catalysis and allosteric regulation of yeast chorismate mutase from crystal structures, *Structure 5*, 1437−1452.
8. Helmstaedt, K., Heinrich, G., Merkl, R., and Braus, G. H. (2004) Chorismate mutase of *Thermus thermophilus* is a monofunctional AroH class enzyme inhibited by tyrosine, *Arch. Microbiol. 181*, 195−203.
9. Lee, A. Y., Stewart, J. D., Clardy, J., and Ganem, B. (1995) New insight into the catalytic mechanism of chorismate mutases from structural studies, *Chem. Biol. 2*, 195−203.
10. Llewellyn, D. J., Daday, A., and Smith, G. D. (1980) Evidence for an artificially evolved bifunctional 3-deoxy-D-arabinoheptulosonate-7-phosphate synthase-chorismate mutase in *Bacillus subtilis*, *J. Biol. Chem. 255*, 2077−2084.
11. Calhoun, D. H., Bonner, C. A., Gu, W., Xie, G., and Jensen, R. A. (2001) The emerging periplasm-localized subclass of AroQ chorismate mutases, exemplified by those from *Salmonella typhimurium* and *Pseudomonas aeruginosa*, *Genome Biol. 2*, 1−16.
12. Goers, S. K., and Jensen, R. A. (1984) The differential allosteric regulation of two chorismate-mutase isoenzymes of *Nicotiana silvestris*, *Planta 162*, 117−124.
13. Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S. V., Eiglmeier, K., Gas, S., Barry, C. E., III, et al. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence, *Nature 393*, 537−544.
14. Camus, J. C., Pryor, M. J., Medigue, C., and Cole, S. T. (2002) Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv, *Microbiology 148*, 2967−2973.
15. Prakash, P., Aruna, B., Sardesai, A. A., and Hasnain, S. E. (2005) Purified recombinant hypothetical protein coded by open reading frame Rv1885c of *Mycobacterium tuberculosis* exhibits a mono-

functional AroQ class of periplasmic chorismate mutase activity, *J. Biol. Chem. 280*, 19641−19648.
16. Sasso, S., Ramakrishnan, C., Gamper, M., Hilvert, D., and Kast, P. (2005) Characterization of the secreted chorismate mutase from the pathogen *Mycobacterium tuberculosis*, *FEBS Lett. 272*, 375−389.
17. Xia, T., Song, J., Zhao, G., Aldrich, H., and Jensen, R. A. (1993) The aroQ-encoded monofunctional chorismate mutase (CM-F) protein is a periplasmic enzyme in *Erwinia herbicola*, *J. Bacteriol. 175*, 4729−4737.
18. Lambert, K. N., Allen, K. D., and Sussex, I. M. (1999) Cloning and characterization of an esophageal-gland-specific chorismate mutase from the phytoparasitic nematode *Meloidogyne javanica*, *Mol. Plant−Microbe Interact. 12*, 328−336.
19. Bekal, S., Niblack, T. L., and Lambert, K. N. (2003) A chorismate mutase from the soybean cyst nematode *Heterodera glycines* shows polymorphisms that correlate with virulence, *Mol. Plant−Microbe Interact. 16*, 439−446.
20. Takai, S., Hines, S. A., Sekizaki, T., Nicholson, V. M., Alperin, D. A., Osaki, M., Takamatsu, D., Nakamura, M., Suzuki, K., Ogino, N., Kakuda, T., Dan, H., and Prescott, J. F. (2000) DNA sequence and comparison of virulence plasmids from *Rhodococcus equi* ATCC 33701 and 103, *Infect. Immun. 68*, 6840−6847.
21. Bumann, D. (2002) Examination of *Salmonella* gene expression in an infected mammalian host using the green fluorescent protein and two-colour flow cytometry, *Mol. Microbiol. 43*, 1269−1283.
22. Qamra, R., Prakash, P., Aruna, B., Hasnain, S. E., and Mande, S. C. (2005) Crystallization and preliminary X-ray crystallographic studies of *Mycobacterium tuberculosis* chorismate mutase, *Acta Crystallogr. F61*, 473−475.
23. Collaborative Computational Project, Number 4 (1994) *Acta Crystallogr. D50*, 760−763.
24. de La Fortelle, E., and Bricogne, G. (1997) Maximum-likelihood heavy atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods, *Methods Enzymol. 276*, 472−494.
25. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement, *Nat. Struct. Biol. 6*, 458−463.
26. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum likelihood method, *Acta Crystallogr. D53*, 240−255.
27. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., et al. (1998) Crystallography and NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr. D54*, 905−921.
28. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) Improved method for building protein models in electron density maps and the location of errors in these models, *Acta Crystallogr. A47*, 110−119.
29. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr. 26*, 283−291.
30. Lovell, S. C., Davis, I. W., Arendall, W. B., III, de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) Structure validation by C-α geometry: ϕ, ψ, and C-β deviation, *Proteins: Struct., Funct., Genet. 50*, 437−450.
31. Kraulis, P. J. (1991) MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures, *J. Appl. Crystallogr. 24*, 946−950.
32. Merritt, E. A., and Bacon, D. J. (1997) Raster3D: Photorealistic molecular graphics, *Methods Enzymol. 277*, 505−524.
33. Matthews, B. W. (1968) Solvent content of protein crystals, *J. Mol. Biol. 33*, 491−497.
34. Schnappauf, G., Strater, N., Lipscomb, W. N., and Braus, G. H. (1997) A glutamate residue in the catalytic center of the yeast chorismate mutase restricts enzyme activity to acidic conditions, *Proc. Natl. Acad. Sci. U.S.A. 94*, 8491−8496.
35. Lim, W. A., Richards, F. M., and Fox, R. O. (1994) Structural determinants of peptide-binding orientation and of sequence specificity in SH3 domains, *Nature 372*, 375−379.
36. Parish, T., and Stoker, N. G. (2002) The common aromatic amino acid biosynthesis pathway is essential in *Mycobacterium tuberculosis*, *Microbiology 148*, 3069−3077.
37. Sassetti, C. M., Boyd, D. H., and Rubin, E. J. (2003) Genes required for mycobacterial growth defined by high-density mutagenesis, *Mol. Microbiol. 48*, 77−84.
38. Monod, J., Wyman, J., and Changeux, J. P. (1965) On the nature of allosteric transitions: A plausible model, *J. Mol. Biol. 12*, 88−118.
39. Xue, Y., and Lipscomb, W. N. (1995) Location of the active site of allosteric chorismate mutase from *Saccharomyces cerevisiae*, and comments on the catalytic and regulatory mechanisms, *Proc. Natl. Acad. Sci. U.S.A. 92*, 10595−10598.

BI0606445