PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Structural studies of prephenate dehydratase from *Mycobacterium tuberculosis* H37Rv by SAXS, ultracentrifugation, and computational analysis

Ana Luiza Vivan,[1,2] Rafael Andrade Caceres,[3] Jose Ramon Beltran Abrego,[4] Júlio César Borges,[5] João Ruggiero Neto,[4] Carlos H.I Ramos,[6] Walter Filgueira de Azevedo Jr.,[2*] Luiz Augusto Basso,[7] and Diógenes Santiago Santos[7*]

[1] Programa de Pós Graduação em Genética e Biologia Molecular, Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre-RS, Brazil

[2] Faculdade de Biociências, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre-RS, Brazil

[3] Programa de Pós Graduação em Biologia Celular e Molecular, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre-RS, Brazil

[4] Departamento de Física, IBILCE/UNESP, São Jose do Rio Preto, São Paulo, Brazil

[5] Departamento de Química e Física Molecular, Instituto de Química de São Carlos, Universidade de São Paulo, São Carlos, São Paulo, Brazil

[6] Instituto de Química, Universidade Estadual de Campinas-UNICAMP, Campinas, São Paulo, Brazil

[7] Centro de Pesquisas em Biologia Molecular e Funcional, Instituto de Pesquisas Biomédicas, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre-RS, Brazil

ABSTRACT

Tuberculosis (TB) is one of the most common infectious diseases known to man and responsible for millions of human deaths in the world. The increasing incidence of TB in developing countries, the proliferation of multidrug resistant strains, and the absence of resources for treatment have highlighted the need of developing new drugs against TB. The shikimate pathway leads to the biosynthesis of chorismate, a precursor of aromatic amino acids. This pathway is absent from mammals and shown to be essential for the survival of Mycobacterium tuberculosis, the causative agent of TB. Accordingly, enzymes of aromatic amino acid biosynthesis pathway represent promising targets for structure-based drug design. The first reaction in phenylalanine biosynthesis involves the conversion of chorismate to prephenate, catalyzed by chorismate mutase. The second reaction is catalyzed by prephenate dehydratase (PDT) and involves decarboxylation and dehydratation of prephenate to form phenylpyruvate, the precursor of phenylalanine. Here, we describe utilization of different techniques to infer the structure of M. tuberculosis PDT (MtbPDT) in solution. Small angle X-ray scattering and ultracentrifugation analysis showed that the protein oligomeric state is a tetramer and MtbPDT is a flat disk protein. Bioinformatics tools were used to infer the structure of MtbPDT. A molecular model for MtbPDT is presented and molecular dynamics simulations indicate that MtbPDT is stable. Experimental and molecular modeling results were in agreement and provide evidence for a tetrameric state of MtbPDT in solution.

Proteins 2008; 72:1352–1362.

Key words: *molecular modeling; small-angle X-ray scattering; molecular dynamics; analytical ultracentrifugation; oligomeric state; bioinformatics; three-dimensional structure; circular dichroism.*

Abbreviations: [Θ], residual molar ellipticity; AUC, analytical ultracentrifugation; CD, circular dichroism; D, diffusion coefficient; $D^0_{20,w}$, standard diffusion coefficient at 0 mg/mL of protein; DLS, dynamic light scattering; D_{max}, maximum distance; MD, molecular dynamics; MDR-TB, multidrug resistant tuberculosis; MM, molecular mass; ns, nanoseconds; PDB, protein data bank; PDT, prephenate dehydratase; $p(r)$, distance distribution function; R_g, radius of gyration; RMSD, root mean square deviation; Rs, Stokes radius; s, sedimentation coefficient; $s^0_{20,w}$, standard sedimentation coefficient at 0 mg/mL of protein; $s_{20,w}$, standard sedimentation coefficient; SAXS, small angle X-ray scattering; TB, tuberculosis; ff_0, frictional ratio.

Grant sponsors: CAPES, Millennium Institutes (CNPq-MCT).
*Correspondence to: Diógenes S. Santos, Centro de Pesquisas em Biologia Molecular e Funcional-PUCRS, Av. Ipiranga, 6681 Prédio 92A, Porto Alegre 90619-900, RS, Brazil. E-mail: diogenes@pucrs.br or Walter F. de Azevedo, Faculdade de Biociências, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre-RS, Brazil. E-mail: walter.junior@pucrs.br.
Received 18 April 2007; Revised 29 January 2008; Accepted 11 February 2008
Published online 2 April 2008 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.22034

INTRODUCTION

Tuberculosis (TB) remains the leading cause of mortality due to a bacterial pathogen, *Mycobacterium tuberculosis*. According to the World Health Organization, TB has reemerged as a world public health problem.[1] Currently one-third of the world population is asymptomatically infected with TB, and approximately eight million people developed TB and three million die every year.[2] The reemergence of TB is basically a consequence of anthropic factors, such as the recent HIV/AIDS pandemic and the development of drug-resistant strains (stemmed from inappropriate treatments and/or patient noncompliance). Another contributing factor is the evolution of multidrug TB (MDR-TB), defined as strains of *M. tuberculosis* resistant to at least isoniazid and rifampicin, two first-line drugs used in the standard "short-course" treatment of TB.[3] The emergence of MDR-TB, the increasing prevalence of TB and the co-infection with HIV, have highlighted the need for the development of new antimycobacterial agents.

The shikimate pathway is present in bacteria, algae, fungi, higher plants, and parasites from the phylum Apicomplexa, but absent from vertebrates.[4,5] This pathway was shown to be essential for *M. tuberculosis* by the disruption of the *arok* gene that codifies for shikimate kinase, the fifth step in the pathway.[6] In Mycobacteria, the product of shikimate pathway, chorismate, is the precursor for aromatic amino acid biosynthesis, menaquinones, and mycobactinas.[7] Accordingly, the enzymes of the amino acid aromatic biosynthetic pathway are promising targets for the development of new antimycobacterial drugs. Homologues of enzymes from the shikimate pathway and phenylalanine biosynthesis have been identified in the genome sequence of *M. tuberculosis* H37Rv,[8] including the gene codifying for prephenate dehydratase (Rv3838c).

Prephenate dehydratase (PDT) is the second enzyme in the pathway leading to the biosynthesis of phenylalanine. It catalyzes the decarboxylation and dehydratation of prephenate to form phenylpyruvate, eliminating water and carbon dioxide.[4]

In *Escherichia coli*, prephenate dehydratase is associated with chorismate mutase, in a bifunctional enzyme called P-protein.[9] In *Amycolatopsis methanolica*, prephenate dehydratase enzyme was characterized as a homotetrameric protein with a subunit molecular weight of 34 kDa.[10] Prephenate dehydratase from *M. tuberculosis* was characterized as a monofunctional enzyme with an allosteric regulation by the aromatic amino acids.[11] As mentioned earlier, PDT is expressed in different organisms and has been characterized biochemically,[9–12] but its structure remains to be resolved.

In this work, we combine the small angle X-ray scattering (SAXS) studies, analytical ultracentrifugation (AUC), molecular dynamics, and molecular modeling to provide the first report of a structural model of prephenate dehydratase from *M. tuberculosis*.

METHODS

Protein sample

Recombinant prephenate dehydratase from *Mycobacterium tuberculosis* (*Mtb*PDT) was expressed and purified as previously described.[13] The protein concentration was determined spectrophotometrically, using the calculated extinction coefficient for native conditions or for denatured proteins.[14]

Hydrodynamic characterization

The diffusion coefficient (D) of the *Mtb*PDT was obtained by dynamic light scattering (DLS) using a DynaPro-MS/X device (Protein Solutions) at 20°C and the protein in 50 mM Tris-HCl (pH 7.8), containing 50 mM NaCl and 1 mM 2-mercaptoethanol. The D is related to the frictional coefficient (f) by the Einstein equation:

$$D = \frac{RT}{N_A f} \tag{1}$$

Where T is the absolute temperature, R is the gas constant, and N_A is the Avogrado's number. The f for a protein of known Stokes radius (Rs) can be obtained applying the Stokes equation:

$$f = 6\pi\eta Rs \tag{2}$$

It is possible to estimate the frictional coefficient for a sphere (f_0) if one uses the Rs for smooth and compact spherical particle (R_0) of molecular mass M_W, which is expressed as:

$$R_0 = \left(\frac{3M_w V_{bar}}{4\pi N_A}\right)^{1/3} \tag{3}$$

Where V_{bar} is the partial specific volume. The ratio between f and f_0 supply the so called frictional ratio (f/f_0) that indicates the shape asymmetry of a hydrated protein, when compared with a globular protein of same MW, being an important parameter for protein structure studies.[15,16] Through the f_0, one can obtain the maximum diffusion coefficient (D_{sphere}), applying the Eq. (1). The ratio $D_{sphere}/D^0_{20,w}$ also supplies directly f/f_0.

The software Sednterp (www.jphilo.mailway.com/download.htm) was used to estimate the PDT hydrodynamic properties from its primary sequence: partial specific volume (V_{bar} = 0.7371 mL/g), the maximum sedimentation coefficient (s_{sphere}), and maximum diffusion coefficient (D_{sphere}) for a globular protein of same molecular mass (MW). These predictions were performed by

applying the Stokes and Svedberg equations as described earlier.[15] The Sednterp also estimated the buffer viscosity ($\eta = 1.0167 \times 10^{-2}$ poise) and density ($\rho = 1.00172$ g/mL).

AUC experiments were performed with a Beckman Optima XL-A analytical ultracentrifuge. *Mtb*PDT sedimentation velocity experiments were done at concentrations ranging from 150 to 1000 μg/mL in 50 m*M* Tris-HCl (pH 7.8) containing 50 m*M* NaCl and 1 m*M* 2-mercaptoethanol. The sedimentation velocity experiments were performed at 20°C, 27,500 rpm (AN-60Ti rotor), and the scan data acquisition was measured at 233 and 236 nm. The software SEDFIT (Version 9.4) was applied in order to fit the absorbance versus cell radius data. This software numerically solves the Fredholm type integral equation to obtain the distribution function of sedimentation coefficients, $c(s)$ which deconvolutes sedimentation and diffusion.[17,18] The D estimated by the DLS experiments was used as prior knowledge for calculation of sedimentation coefficient distributions c(s). SEDFIT performs regularization by floating f/f_0 as a parameter, from which the diffusion coefficient, D, can be calculated. The apparent sedimentation coefficients (s) were found as the maximum of the peaks of the $c(s)$ curves. The apparent s value contains interferences caused by temperature, viscosity, and buffer density, so we calculate the standard sedimentation coefficient at 0 mg/mL of protein concentration ($s^0_{20,w}$), which is an intrinsic parameter of the particle.[16] Changes in the $s^0_{20,w}$ induced by pH, ionic strength, ligands, and temperature are due to conformational changes.[16] The Sednterp software estimated the standard sedimentation coefficients ($s_{20,w}$) at each protein concentration from the s,[15,16] and, by linear regression, we estimated the $s^0_{20,w}$. The M_W of a particle can be calculated by the ratio of the sedimentation to diffusion coefficient by the following equation.

$$M_W = \frac{sRT}{D(1 - V_{bar}\rho)} \quad (4)$$

The SEDFIT (Version 9.4) was also used to estimate the maximum of the peaks in the $c(M)$. This software gives a weight-average frictional ratio, f/f_0, which is the fitting parameter in the regularization process.[16–18] The f/f_0 was also estimated by the ratio $s_{sphere}/s^0_{20,w}$.[16]

The sedimentation equilibrium experiments were carried out at 20°C, at speeds of 6000, 8000, 10,000, and 12,000 rpm with the AN-60Ti rotor, and the scan data acquisition was at 234 nm. *Mtb*PDT in 50 m*M* Tris-HCl (pH 7.8) containing 50 m*M* NaCl and 1 m*M* 2-mercaptoethanol and was tested at the following concentrations: 250, 500, and 750 μg/mL. The analysis involved the fitting of a model of absorbance versus cell radius data using nonlinear regression. The distribution of the protein along the cell, obtained in the equilibrium sedimentation experiments, was fitted by the following equation[15,19]:

$$C = C_0 \exp\left[\frac{M_W(1 - V_{bar}\rho)\omega^2(r^2 - r_0^2)}{2RT}\right] \quad (5)$$

Where C is the protein concentration at radial position r, C_0 is the protein concentration at radial position r_0, and ω is the centrifugal angular velocity. The self-association method was used to analyze the sedimentation equilibrium experiments using several models of association or oligomerization for *Mtb*PDT in order to fit the data obtained at all speeds and concentrations together. The fittings were performed using the Origin software package (Microcal Software) supplied with the instrument.

Circular dichroism spectroscopy

Circular dichroism (CD) measurements were performed in a Jasco J-810 spectropolarimeter equipped with a Peltier-type temperature control system PFD 425S. Measurements were carried out at final concentrations of 200−1000 μg/mL of *Mtb*PDT in 50 m*M* Tris-HCl (pH 7.8) containing 50 m*M* NaCl and 1 m*M* 2-mercaptoethanol. The spectra were collected at a scanning rate of 100 nm/min with a spectral bandwidth of 1 nm and using a 0.1 mm or 1 mm path length cell. The resultant spectra were normalized to residual molar ellipticity ([Θ]) and the *Mtb*PDT secondary structure content was estimated by the CDNN Deconvolution program.[20]

Small angle X-ray studies

Small angle X-ray scattering technique supplies information about a "time averaged conformation" indicating the main conformation presented by the scattering particle, and provides data about the protein structure at low resolution in solution.[21]

X-ray scattering data were collected at SAXS beam line of Brazilian Synchrotron Light Laboratory (LNLS, Campinas, Brazil) with a 1.488 Å wavelength (6–12 KeV) and a capillary sample holder of 1.5 mm diameter.[22] The scattered intensity, $I(s)$, is recorded as a function of the momentum transfer s ($s = 4\pi\sin\theta/\lambda$), where 2θ is the angle between the incident and the scattered radiation and λ is the X-ray wavelength. Contributions of the scattered intensities from the solvent were eliminated by subtracting the intensity curve obtained for the buffer solution. The radius of gyration R_g, a structural parameter related to the overall size of the macromolecule, was determined by Guinier as follows:

$$I(h) = I(0)e[-h^2R_g^2/3] \quad (6)$$

This equation applies to macromolecules in the limits of dilute solutions and a small h value. Information of

the molecular structure can be obtained from the distance distribution function $p(r)$, which is related to SAXS by the equation

$$p(r) = 1/2\pi^2 \int_0^\infty (rh)I(h)\ \sin(rh)\mathrm{d}h \qquad (7)$$

The $p(r)$ function is proportional to the number of pairs of electrons separated by the distance r, which is encountered by combinations between all the elements of the macromolecule. The distance distribution of $p(r)_{\mathrm{exp}}$ has been determined by the inverse Fourier transformation using the GNOM program.[23] The volume (V) and the surface area (S) of the protein were calculated by, respectively, Eqs. (8) and (9), where Q is invariant.[24]

$$V = 2\pi^2\ I\ (0)/Q \qquad (8)$$

$$S = \pi V.\mathrm{lim}[I\ (q)\ q^4]/Q \qquad (9)$$

The SAXS measurements were performed at 25°C using *Mtb*PDT in solution at concentrations of 9 mg/mL, 6 mg/mL, and 3 mg/mL in 50 m*M* Tris-HCl pH 7.2 and 50 m*M* boric acid pH 9.5. The solutions were placed on a vacuum cell with a mica window and irradiated with X-ray beam for 15 min. The protein was monodisperse, which in turn allowed SAXS measurements to be performed. The SAXS measurements were performed within an angular range defined by $0.00894\ \text{Å}^{-1} < q < 0.450\ \text{Å}^{-1}$. The distance between detector and specimen was 831 mm.

The data were analyzed in the programs DAMAVER and SUPCOMB 20.[25,26] The HydroPro software[27] was applied to estimate the $s_{20,w}$, standard diffusion coefficient ($D_{20,w}$), radius of gyration (R_g), Stokes radius (Rs), the maximum distance (D_{max}), and the f/f_0 from the PDT *ab initio* model developed.

Molecular modeling

The molecular modeling method is able to predict the three-dimensional structure of a protein, using its amino acid sequence and homology of known protein. For molecular modeling, we used the program PARMODEL, a web server for automated comparative modeling and evaluation of proteins structures.[28] There is no structure resolved of any prephenate dehydratase in PDB. Therefore, phenylalanine hydroxylase (PDB code: 2PHM) was used as template due to the high similarity between this enzyme and *Mtb*PDT. Although the similarity between phenylalanine hydroxylase and *Mtb*PDT is 35% and the identity is only 15%, phenylalanine hydroxylase is the most similar protein with structure resolved in PDB. The quality of the predicted model determines the information that can be extracted from it.

The overall stereochemical quality of the final model was assessed by the program PROCHECK.[29] The 3D profile measures the compatibility of a protein model with its sequence, and was calculated using the program Verify3D.[30,31] The root mean square deviation (RMSD) was calculated using the program LSQKAB from CCP4 suite.[32]

To obtain the dimeric structure of PDT, the monomer structure generated by PARMODEL was docked against its own structure, resulting in 10 dimeric models. The docking was performed in the program GRAMM-X.[33] The best model for a dimer was docked against itself, generating a homotetrameric structure. The algorithm is based on the correlation between the digitized molecular images, using the fast Fourier transformation (FFT).[33]

Molecular dynamics

The initial structure for *Mtb*PDT was obtained from molecular modeling by homology. The simulation was performed by a time period of six nanoseconds (ns) with GROMACS[34] package using the Gromacs 96.1 (43A2) force field.

The all-atom model of the protein contains 2973 atoms and a net molecular charge of −6. Hence, six sodium counterions were added using *Genion* Program of the GROMACS simulation suite to neutralize the negative charge density of the system, which was then immersed in a cubic box containing a total of 28163 SPC/E water molecules. The initial simulation cell dimensions were 56.41 Å, 49.63 Å, 73.72 Å and had the protein solvated by a layer of water molecules of at least 10 Å in all directions. The simulation cell contained a total of 87,480 atoms.

The simulations were carried out using explicit solvent water molecules (described by the simple point charge, SPC/E) and periodic boundary conditions (cubic). In the molecular dynamics (MD) protocol, all hydrogen atoms, ions, and water molecules were first subjected to 200 steps of energy minimization by steepest descent to remove close van der Waals contacts. The system was then submitted to a short molecular dynamics with position constrains for a period of 1 ps and afterwards performed a full molecular dynamics cycle without constrains. The temperature of the system was then increased from 50 to 310 K in 5 steps (50–100 K, 100–150 K, 150–200 K, 200–250 K, and 250–310 K), and the velocities at each step were reassigned according to the Maxwell-Boltzmann distribution at that temperature and equilibrated for 2 ps.

Energy minimization and MD were carried out under periodic boundary conditions. The simulation was computed in the isobaric-isothermal (NPT) ensemble at 310 K and 1 atm with the Berendsen temperature coupling and constant pressure of 1 atm with isotropic molecule-based scaling.[35] The LINCS algorithm, with a 10^{-5} Å

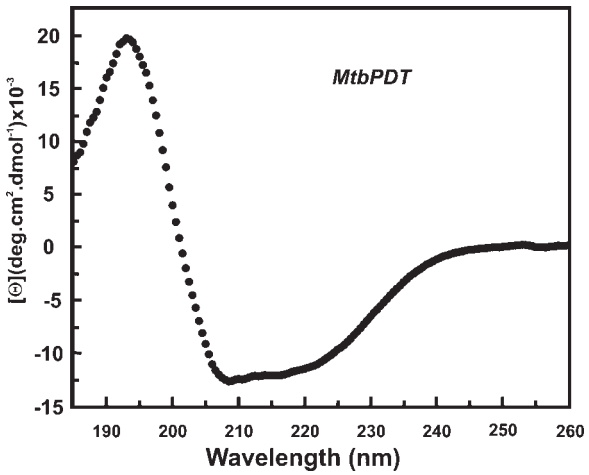

Figure 1

MtbPDT presents high α-helix content. Residual molar ellipticity [Θ] of MtbPDT were measured from 185 to 260 nm in tris-HCl 50 mM (pH 7.8), NaCl 50 mM, and 2-mercaptoethanol 1 mM at 20°C and in a 0.1 mm pathlength cell.

tolerance, was applied to fix all bonds containing a hydrogen atom, allowing the use of a time step of 2.0 fs in the integration of the equations of motion. No extra constrains were applied after the equilibration phase. The electrostatic interactions between nonligand atoms were evaluated by the particle-mesh Ewald method[36] with a charge grid spacing of approximately of 1.0 Å, and the charge grid was interpolated on a cubic grid with the direct sum tolerance set to 1.0×10^{-5}. The Lennard-Jones interactions were evaluated using a 10.0 Å atom-based cutoff.[37]

All analysis were performed on the ensemble of system configurations extracted at 0.5-ps time intervals from the simulation and MD trajectory collection was initiated after 1 ns of dynamics to ensure a completely equilibrated evolution. The MD simulation and results analysis were performed on a workstation Intel Xeon Duo Core 1.67 GHz and 2 Gb RAM.

RESULTS AND DISCUSSION

Secondary structure

The *Mtb*PDT secondary structure content was estimated by circular dichroism method. The [Θ] of *Mtb*PDT (see Fig. 1) indicates that the secondary structure content estimated by the CDNN Deconvolution software[20] was: 33% in α-helix, 18% in β-sheet, 17% in turns, and 32% in coils (standard deviation was less than 5%). The secondary structure content of the *Mtb*PDT model was analyzed by the Procheck software,[29] and estimated at 38% in α-helix and 13% in β-sheet, which is in good agreement with the experimental data.

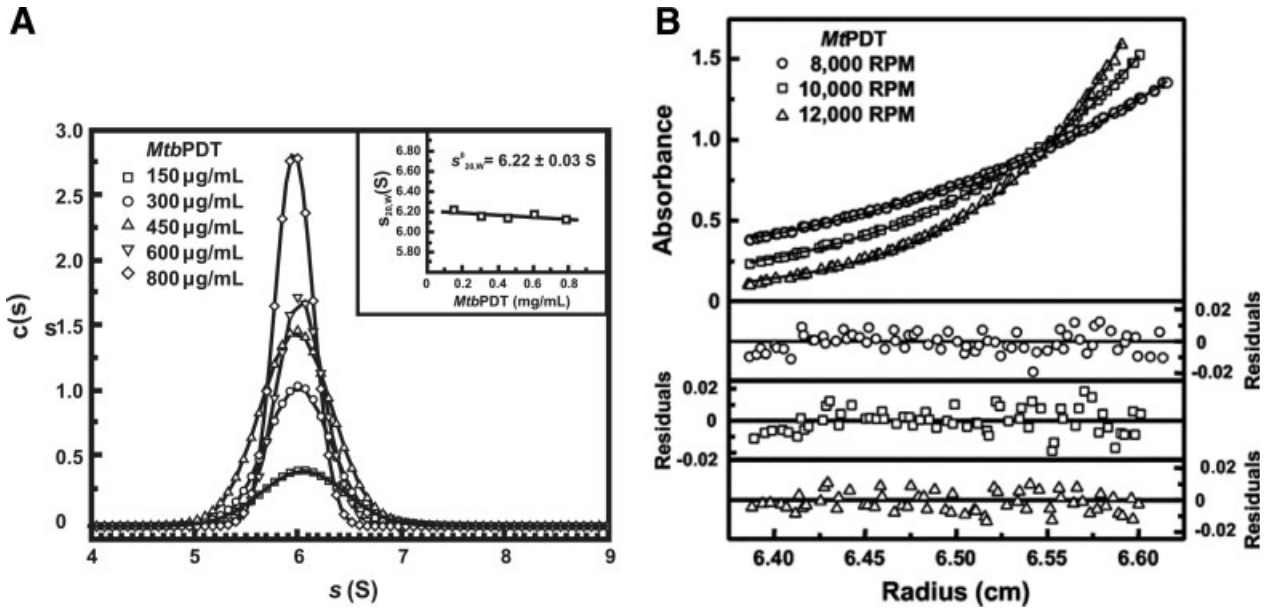

Figure 2

*AUC experiments show MtbPDT as a tetramer. (**A**) Sedimentation velocity experiments were carried out at 20°C, with rotor (AN-60Ti) at the velocity of 27,500 rpm. The figure shows experiments with 150, 300, 450, 600, and 800 µg/mL of the MtbPDT in tris-HCl 50 mM (pH 7.8) containing NaCl 50 mM. (**B**) Sedimentation equilibrium experiments were carried out in 250, 500, and 750 µg/mL at 20°C and in 8000, 10,000, and 12,000 rpm.*

Table I
Structural and Hydrodynamic Parameters Derived from SAXS, DLS, and Estimated to the PDT Models by HydroPro Software

				HydroPro[b] (using models)		
*Mtb*PDT hydrodynamic properties	Predicted[a] (tetramer)	DLS	AUC	Homology	*Ab initio*	SAXS
M_W (kDa)	134.5	125	128 ± 4[c] 133 ± 2[d] 130 ± 2[e]	—	—	—
s (S)	9.20	—	6.22 ± 0.03	6.93 ± 0.07	5.90 ± 0.01	—
D (10^{-7}cm^2/seg)	6.30	4.5 ± 0.1	4.4 ± 0.1	4.6 ± 0.1	3.9 ± 0.1	—
Rs (Å)	34.2	47 ± 1	50[f]	46.5	54.6	—
f/f_0	1.0	1.41	1.43 ± 0.03[c] 1.46 ± 0.01[e]	1.33	1.56	—
D_{max} (Å)	—	—	142[e]	146	147	130
R_g (Å)	—	—	—	38.2 ± 0.1	42.3 ± 0.1	45.8

[a]Values predicted for PDT as a tetrameric globular particle (predicted by Sednterp software).
[b]Obtained from the homology model developed and from the *ab initio* model developed from the SAXS data assuming the *Mtb*PDT as a tetramer of 134.5 kDa (data predicted for water and temperature of 20°C).
[c]Calculated from SedFit $c(M)$ fitting of the sedimentation velocity data.
[d]From equilibrium sedimentation data.
[e]From s/D ratio [Eq. (4)].
[f]Data from extrapolation of the Sednterp software assuming the protein as a hydrated tetramer with an oblate shape.

Analytical ultracentrifugation experiments

Sedimentation velocity experiments were carried out in order to obtain information about the oligomeric state and shape of the *Mtb*PDT. The Figure 2(A) shows the sedimentation coefficient distribution $c(s)$ of the *Mtb*PDT at different protein concentration, where the maximum of the peak was taken as the s-value. The results suggest that the protein behaved as a monodisperse system without aggregates or other species with different oligomeric states. To avoid interferences caused by the solution viscosity and density,[16] we determined the $s^0_{20,w}$ [Fig. 2(A), inset and Table I]. The value extrapolated to 0 mg/mL was 6.22 ± 0.03 S which is larger for *Mtb*PDT as a monomer (s_{sphere} = 3.65 S) or even as a dimer (s_{sphere} = 5.75 S). The $D^0_{20,w}$ value extracted from the SEDFIT fitting routine was 4.4 ± 0.1 × 10^{-7} cm^2/seg, which was similar to the experimental D obtained from DLS data (Table I). The experimental $D^0_{20,w}$ estimated for *Mtb*PDT was smaller than that predicted ones [Eq. (1), see Methods for details] for a globular protein of same MtbPDTs M_w as either monomer (D_{sphere} = 10 × 10^{-7} cm^2/seg) or dimer (D_{sphere} = 7.9 × 10^{-7} cm^2/seg). We also estimated the M_W for *Mtb*PDT using the $c(M)$ evaluated by the SEDFIT software (data not shown). The values obtained indicate that the *Mtb*PDT posses a M_W of about 128 ± 4 kDa and a f/f_0 of about 1.43 ± 0.03 (Table I). The s/D ratio [Eq. (4)] gave M_W of about 130 ± 2 kDa for *Mtb*PDT, a close value for the protein as a tetramer (Table I).

The values for $s^0_{20,w}$ and $D^0_{20,w}$ estimated for *Mtb*PDT (Table I) suggest that the protein presented an asymmetric shape. Both the s_{sphere} and D_{sphere} values for a spherical particle of 134.5 kDa (tetrameric *Mtb*PDT) estimated by the software Sednterp were larger than the ones determined experimentally (Table I). These data suggest that the *Mtb*PDT had an f/f_0 of about 1.47, as estimated by the Sednterp program (Table I). The DLS data suggest that the *Mtb*PDT has a Stokes radius (Rs) of about 47 ± 1 Å and a f/f_0 of about 1.41 (Table I), as evaluated from Stokes-Einstein equations [Eqs. (1) and (2)]. The ratios $s_{sphere}/s^0_{20,w}$ and $D_{sphere}/D^0_{20,w}$ showed similar values (data

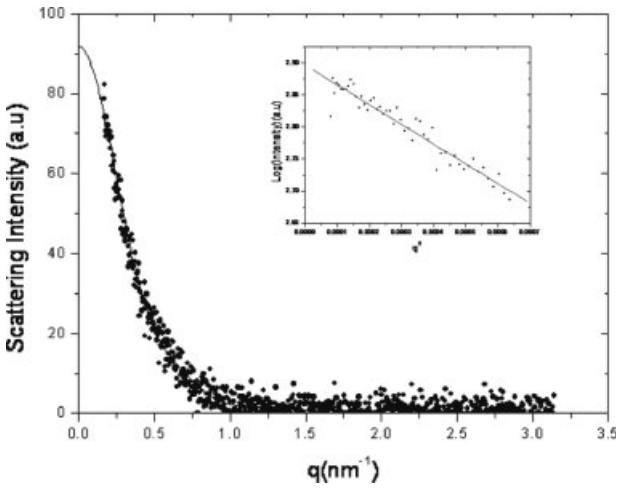

Figure 3
Experimental SAXS curve of MtbPDT. The MtbPDT SAXS experiments were carried out with a protein concentration extrapolated to 0 mg/mL in 50 mM boric acid pH 9.5 at 25°C. Circles indicate the raw data. Inset the Guinier plot of MtbPDT.

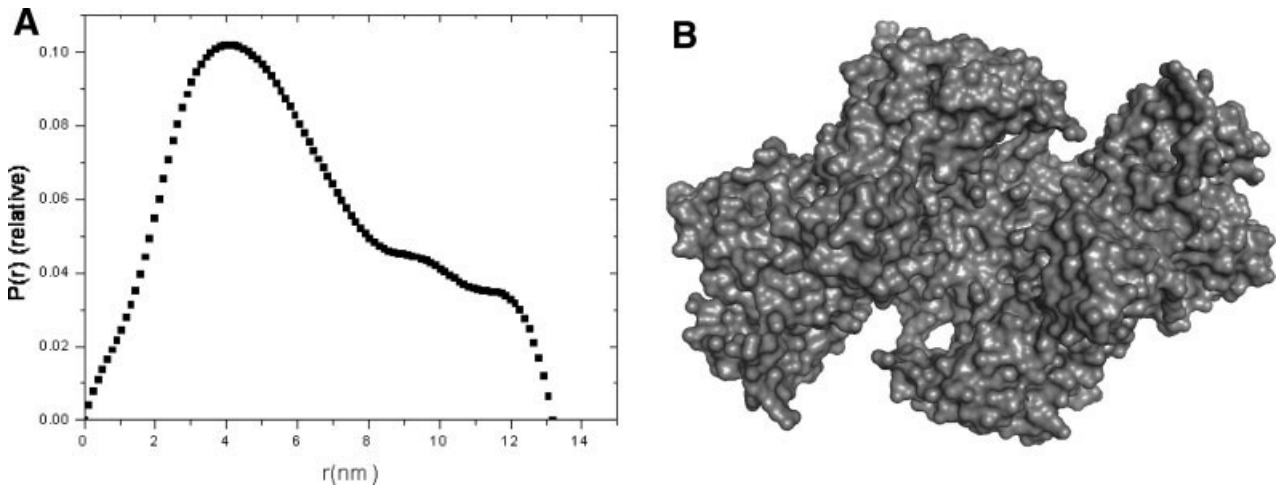

Figure 4

*(**A**) MtbPDT distance pair distribution function (pr) in boric acid buffer, calculated from the SAXS profile by indirect Fourier transformation using the GNOM program.[23] This profile is typical for a flat disk protein. (**B**) Representation of tetramer structure of MtbPDT using the program Pymol.*

not shown). Therefore, the *Mtb*PDT behaved as an asymmetric tetramer.

Figure 2(B) gives the equilibrium sedimentation data at three different speeds and a protein concentration of 500 μg/mL (see Methods for more details). The experimental data were fitted to a self-association model and the M_W of *Mtb*PDT fixed as a tetramer of 134.5 kDa. The analyses of the residuals plot suggest that the model of *Mtb*PDT as a tetramer fitted well with the experimental equilibrium sedimentation data [Fig. 2(B), lower panels], thereby indicating that *Mtb*PDT is a tetramer in solution. When the M_W was allowed to float, the fitting of experimental data resulted in a M_W of about 133 ± 2 kDa (Table I).

The HydroPro software[27] was employed to estimate the hydrodynamic properties of either *Mtb*PDT model built by homology modeling or the low resolution *ab initio* model generated from the SAXS data (described

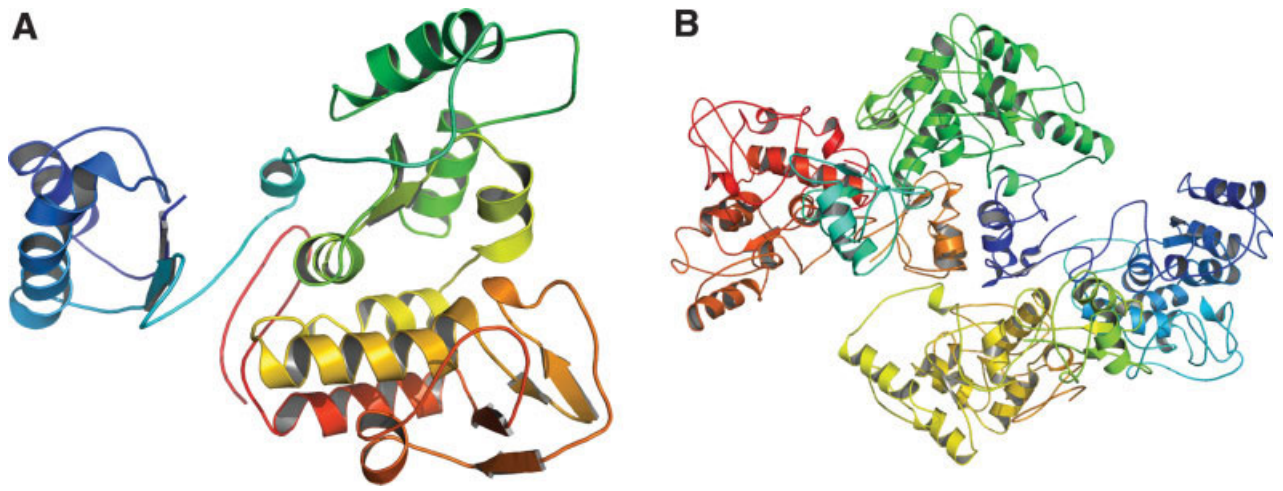

Figure 5

*Ribbon diagram of molecular model of MtbPDT. (**A**) View of the monomer generated using PARMODEL. (**B**) The MtbPDT tetramer was obtained with molecular docking.*

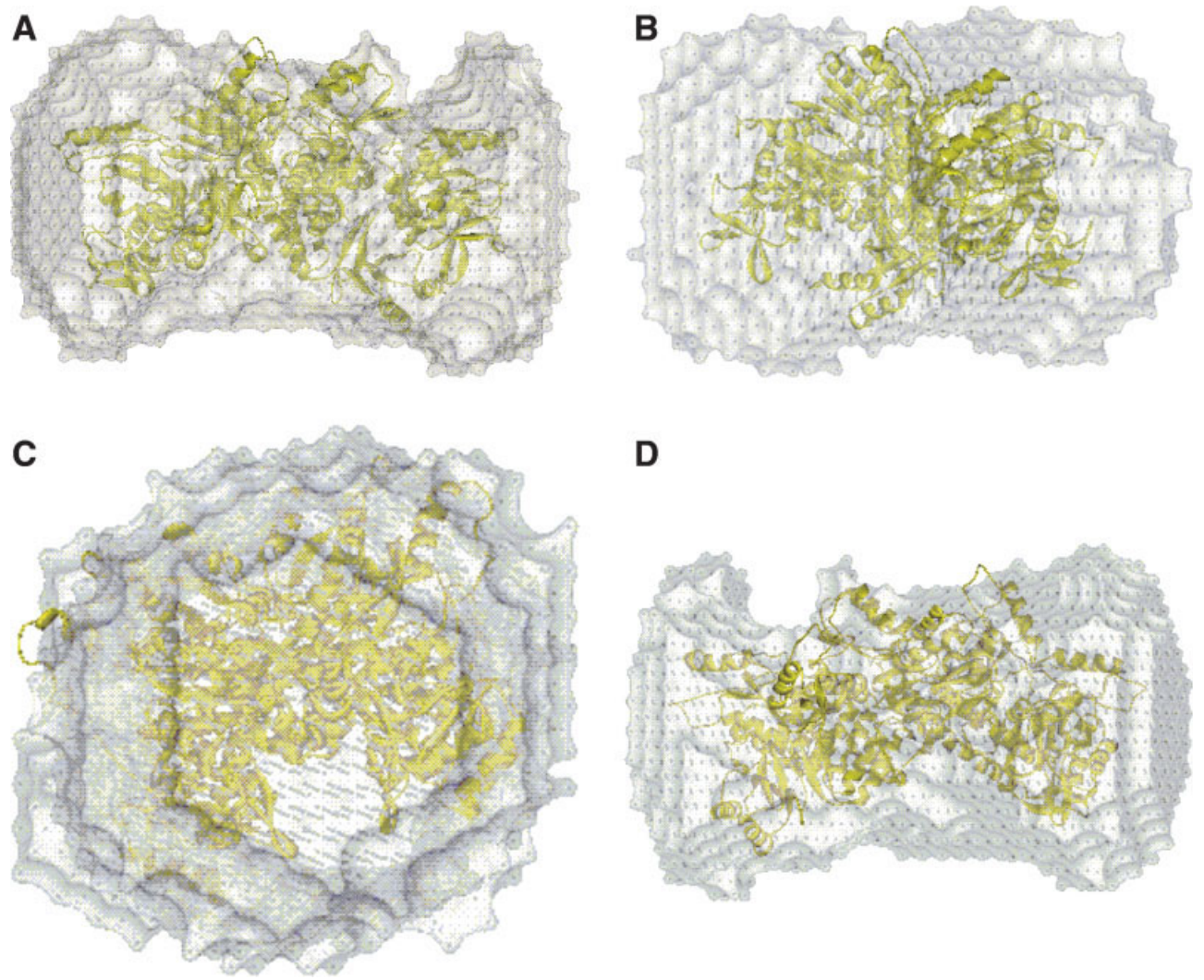

Figure 6

*Low resolution beads model of tetrameric MtbPDT. Models constructed using the ab initio modeling program DAMMIN to analyze SAXS data. (**A**) Figure showing MtbPDT at 0°. (**B**) Figure showing MtbPDT at 60°. (**C**) Figure showing MtbPDT at 90°. (**D**) Figure showing MtbPDT at 180°. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]*

later). Both models were evaluated as tetramers and the results are given in Table I. The D_{max} observed for the models in HydroPro software were around 145 Å long and were in good agreement with the experimental data from SAXS, 130 Å, and with the value estimated from the hydrodynamic data, considering that they are low-resolution models. The value of predicted $s_{20,w}$ estimated for the *ab initio* model was very close to the experimental data (5.9 S vs. 6.22 S). However, for the *Mtb*PDT homology model it was larger (6.93 S vs. 6.22 S). The analyses of the *Rg*, f/f_0, and *Rs* also suggested that the *Mtb*PDT homology model was more discrepant than the PDT *ab initio* model. These differences may be explained by the flexibility of the protein in solution that would increase the protein volume and its interaction area with the solvent. Taken together, the hydrodynamic analysis suggested that the models developed are fair models of *Mtb*PDT protein as a tetramer and showed that it had an asymmetric shape.

Small angle X-ray scattering results

Figure 3 shows the SAXS profile of *Mtb*PDT at a protein concentration extrapolated to 0 mg mL^{-1}.

A Guinier plot of the scattering data shows a linear correlation between the intensity and the *q* values and

the slope of this plot provides the radius of gyration (R_g) of the protein, which was $R_g = 45.77 \pm 1.81$ Å.

The pair distance distribution function, $p(r)$ shown in the Figure 4 was calculated from the data in Figure 3. The maximum diameter D_{max} for the *Mtb*PDT was estimated to be 130 Å from this analysis. The $p(r)$ function shows a peak of 40 Å with a shoulder at higher r value. This profile is typical for a flat disk protein.[21]

To gain further insight into the structure of *Mtb*PDT, the construction of 20 models by *ab initio* based on the SAXS data using the program DAMMIN[38] was carried out. These models were analyzed by DAMAVER program to obtain the most probable average model, and the superposition of the SAXS curves and the predicted model generated by molecular model by SUPCOMB 20.[23,26] The shape determination of a protein by *ab initio* from solution scattering has been validated by a posteriori comparison with high resolution crystallography structures.[21]

Considering the fact that no close homologue of *Mtb*PDT has been found, this model has proposed a tetrameric structure for this protein.

Molecular modeling

The value of RMSD for the superposition of the predicted model and the template was 1.34 Å. These values are due to the low identity between the template and predicted model. Analysis of the structural quality of the model indicates that the generated model is appropriated for structural studies. The predicted model for monomeric *Mtb*PDT is shown in Figure 5(A), and tetrameric *Mtb*PDT shown in Figure 5(B) was obtained from docking the dimeric model against itself. A low resolution beads model of tetrameric *Mtb*PDT is given in Figure 6.

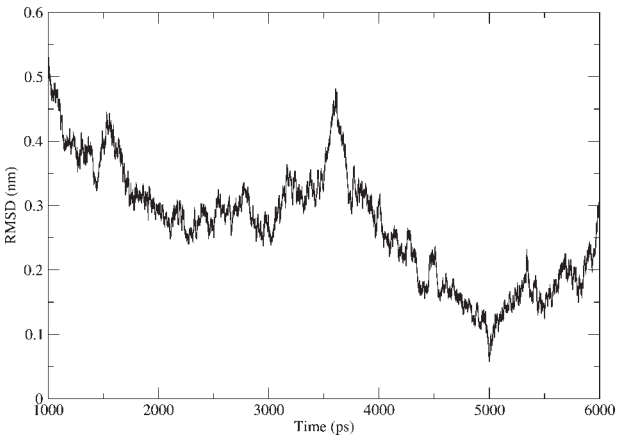

Figure 7

Root mean square deviation (RMSD) of C-alpha from MtbPDT for 6 ns. The trajectory collection was initiated after 1 ns of dynamics to ensure complete system equilibration.

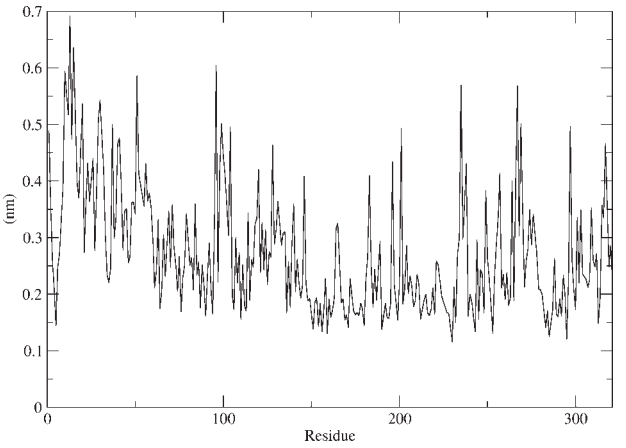

Figure 8

Root mean square fluctuation (RMSF) of 321 residues of MtbPDT for last 5 ns of molecular dynamics simulation.

Analysis of Ramachandran plot for the *Mtb*PDT model indicated that the model presents 89.6% of the residues in most favored regions and has no residues in disallowed regions. The accuracy of the protein model can be assessed by its 3D profile, which generated a 3D-1D plot with a score for each amino acid. The 3D profile score (S) for the compatibility of the sequence with the model is the sum, over all residue positions, of the 3D-1D scores for the amino acid sequence of the protein.[31] The 3D profile score (S) was 35.52, the 3D Profile Ideal Score (IS) was 146.59, and the quality of the model (S/IS) was 0.24. Although these values are low, the Ramachandran plot have shown that no residue were found in disallowed region and 89.6% of the residues are in the most favored regions, what shows that this model is good enough to be used in molecular dynamics.

Molecular dynamics

To verify the stability of *Mtb*PDT, MD was performed for 6 ns and trajectory was collected after 1 ns to ensure complete system equilibration. The root mean square deviation (RMSD) of C-alpha (Cα over the model of *Mtb*PDT trajectory was fitted to the starting structures and is plotted in Figure 7 revealing an average RMSD of 0.27 ± 0.1 nm. After 3.5 ns a decreasing value for RMSD of Cα was observed indicating equilibration of at least 2.5 ns of simulation. The flexibilities of each atom residue of MtbPDT could be seen at Figure 8. Folding secondary and tertiary structures were maintained during the simulation (data not shown). Thus, turns and loops contributed to raising the average value of RMSD and the latter does not represent the stability of *Mtb*PDT model running in solution.

CONCLUSIONS

In this study, we have used independent tools to infer new insights on the structure of *Mtb*PDT. Experimentally, we investigated the *Mtb*PDT oligomeric state in solution using small angle X-ray scattering and analytical ultracentrifugation. Both techniques strongly indicate that *Mtb*PDT acts as a homotetramer. AUC and HydroPro software also provided information about the hydrodynamics properties of *Mtb*PDT and have shown that the predicted model is good for *Mtb*PDT. Although the different techniques have shown some differences between the predicted model and experimental determined model for *Mtb*PDT, these could be due to the flexibility of the protein. Circular dichroism method and bioinformatics tools were in agreement showing a similar result about *Mtb*PDT secondary structure. SAXS also showed that *Mtb*PDT is a flat disk protein, with an asymmetric shape and also a tetramer.

The oligomeric structure of PDT from *M. tuberculosis* is similar to *A. methanolica* PDT[10] that is homotetrameric, and different of P-protein of *E.coli* that was characterized as a dimer.[39] These differences could be due to chorismate mutase function in P-protein. Size exclusion chromatography was used to determine the oligomeric state of MtbPDT. The experiment was performed on a Superdex 200 HR 10/30 using 10 m*M* TrisHCl, 100 m*M* NaCl, and 1 m*M* DTT pH 7.8 as running buffer. Our results corroborates with AUC experiments that shows the enzyme as a tetramer. This difference could be attributed to the asymmetric form of the protein. A previous work has shown PDT from *M. tuberculosis* as a dimer,[11] but this discrepancy may be due to the difference during cloning and purification process or to the protein concentration used in gel filtration chromatography.

Altogether, we have presented a theoretical investigation using bioinformatics tools to infer the three-dimensional structure of *Mtb*PDT. This analysis corroborates the experimental data. Our dynamics simulations have demonstrated that this model is stable during 6 ns trajectory.

In future works crystallography X-Ray diffraction might be used to solve three-dimensional structure of prephenate dehydratase, showing more accurate data on the structure of the active site.

In summary, this is the first report of *Mtb*PDT three-dimensional structure. Since PDT has no human homologue, the results here described will be useful to guide structure-based design of *Mtb*PDT inhibitors to treat tuberculosis.

ACKNOWLEDGMENTS

WFA, CHIR, LAB, and DSS are research career awardees from the National Research Council of Brazil (CNPq). JCB thanks FAPESP (04/08966-2) and RC thanks Hewlett-Packard of Brazil for fellowships. The authors would like to acknowledge LNLS technical staff for assistance at SAXS beam line.

REFERENCES

1. World Health Organization. World health organization report: global tuberculosis control—surveillance, planning, financing. Geneva, Switzerland, 2005.
2. Dye C. Global epidemiology of tuberculosis. Lancet 2006;367:938–939.
3. Pablos-Mendez A, Gowda DK, Frieden T. Controlling multidrug-resistant tuberculosis and access to expensive drugs: a rational framework. Bull World Health Organ 2002;80:489–500.
4. Pittard AJ. Biosyntheses of the aromatic amino acids. In: Neidhardt FC, Ingraham JL, Low KB, Magasanik B, Schaechter M, Umbarger HE, editors. *Escherichia coli* and *Salmonella*: cellular and molecular biology, Vol. 1. Washington: American Society for Microbiology; 1987. pp 458–479.
5. Roberts F, Roberts CW, Johnson JJ, Kylell DE, Krell T, Coggins JR, Coombs GH, Milhous WK, Triphuri S, Ferguson DJP, Chakrabarti D, McLeod R. Evidence for the shikimate pathway in apicomplexan parasites. Nature 1998;393:801–805.
6. Parish T, Stoker NG. The common aromatic amino acid biosynthesis pathway is essential in *Mycobaterium tuberculosis*. Microbiology 2002;148:3069–3077.
7. Ratledge C, Standord JL, editors. The biology of the Mycobacteria, Vol. 1. London: Academic Press; 1982.
8. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares R, Squares S, Sulston JE, Taylor K, Whitehead S, Barrell BG. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature 1998;393:537–544.
9. Davidson BE, Blackburn EH, Dopheide TAA. Chorismate mutase-prephenate dehydratase from *Escherichia coli* K-12. J Biol Chem 1972;247:4441–4446.
10. Euverink GJW, Wolkers DJ, Dijkhuizen L. Prephenate dehydratase of the actinomycete *Amycolatopsis methanolica*: purification and characterization of wild-type and deregulated mutant proteins. Biochem J 1995;308:313–320.
11. Prakash P, Aruna B, Sardesai AA, Hasnain SE. pheA (Rv3838c) of *Mycobacterium tuberculosis* encodes an allosteric regulated monofunctional prephenate dehydratase that requires both catalytic and regulatory domains for optimum activity. J Biol Chem 2005;280:20666–20671.
12. Gething MJH, Davidson BE, Dopheide TAA. Chorismate mutase/prephenate dehydratase from *Escherichia coli* K12: the effect of NaCl and its use in a new purification involving affinity chromatography on sepharosyl-phenylalanine. Eur J Biochem 1976;71:317–325.
13. Vivan AL, Dias MVB, Schneider CZ, de Azevedo WF, Jr, Basso LA, Santos DS. Crystallization and preliminary X-ray diffraction analysis of prephenate dehydratase of *Mycobacterium tuberculosis H37Rv*. Acta Crystallogr 2006;62:357–360.
14. Pace CN, Vajdos F, Fee L, Grimsley G, Gray T. How to measure and predict the molar absorption coefficient of a protein. Protein Sci 1995;4:2411–2423.
15. Lebowitz J, Lewis MS, Schuck P. Modern analytical ultracentrifugation in protein science: a tutorial review. Protein Sci 2002;11:2067–2079.
16. Laue T. Biophysical studies by ultracentrifugation. Curr Opin Struct Biol 2001;11:579–583.

17. Schuck P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys J 2000;78:1606–1619.
18. Schuck P, Perugini MA, Gonzales NR, Howlett GJ, Schubert D. Size-distribution analysis of proteins by analytical ultracentrifugation: strategies and application to model systems. Biophys J 2000; 82:1096–1111.
19. Johnson ML, Correia JJ, Yphantis DA, Halvorson HR. Analysis of data from the analytical ultracentrifuge by nonlinear least-squares techniques. Biophys J 1981;36:575–588.
20. Böhm G, Muhr R, Jaenicke R. Quantitative analysis of protein far UV circular dichroism spectra by neural networks. Protein Eng 1992;5:191–195.
21. Svergun DI, Koch MHJ. Advances in structure analysis using small-angle scattering in solution. Curr Opin Struct Biol 2002;12:654–660.
22. Cavalcanti PP, Torriani IL, Plivelic TS, Oliveira CLP, Kellerman G, Neuenschwander R. Two new sealed sample cells for small angle X-ray scattering from macromolecules in solution and complex fluids using synchrotron radiation. Rev Sci Instrum 2004;75:4541–4546.
23. Svergun DI. Determination of the regularization parameter in indirect transform methods using perpetual criteria. J Appl Crystallogr 1992;25:495–503.
24. Porod G. General theory. In: Glatter O, Kratky O, editors. Small-angle X-ray scattering. London: Academic Press; 1982. pp 17–51.
25. Volkov VV, Svergun DI. Uniqueness of ab initio shape determination in small-angle scattering. J Appl Crystallogr 2003;36:860–864.
26. Kozin MB, Svergun DI. Automated matching of high- and low-resolution structural models. J Appl Crystallogr 2001;34:33–41.
27. García de la Torre J, Huertas ML, Carrasco B. Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. Biophys J 2000;78:719–730.
28. Uchoa HB, Jorge GE, da Silveira NJF, Camera JC, Jr, Candur F, de Azevedo WF, Jr. Parmodel: a web server for automated comparative modeling of proteins. BBRC 2004;325:1481–1486.
29. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–291.
30. Bowie JU, Luthy R, Einsenberg D. A method to identify protein sequences that fold into known three-dimensional structure. Science 1991;253:164–170.
31. Luthy R, Bowie JU, Einsenberg D. Assessment of protein models with three-dimensional profiles. Nature 1992;356:83–85.
32. Collaborative Computational Project N°4. The CCP4 suite: program for protein crystallography. Acta Crystallogr D 1994;50:760–763.
33. Tovchigrechko A, Vakser IA. GRAMM-X public web server for protein-protein docking. Nucleic Acids Res 2006;34:310–314.
34. Van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: fast. flexible and free. J Comput Chem 2005;26:1701–1718.
35. Chowdhuri S, Tan ML, Ichiye T. Dynamical properties of the soft sticky dipole-quadrupole-octupole water model: a molecular dynamics study. J Chem Phys 2006;125:144513.
36. Darden T, York D, Pedersen LA. Particle mesh Ewald: an N-log (N) method for Ewald sums in large systems. J Chem Phys 1993;98: 10089–10092.
37. Norberto de Souza O, Ornstein RL. Molecular dynamics simulations of a protein-protein dimer: particle-mesh Ewald electrostatic model yields far superior results to standard cutoff model. J Biomol Struct Dyn 1999;16:1205–1218.
38. Svergun DI. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys J 1999;76:2879–2886.
39. Zhang S, Pohnert G, Kongsaeree P, Wilson DB, Clardy J, Ganem B. Chorismate mutase-prephenate dehydratase from *Escherichia coli*: study of catalytic and regulatory domains using genetically engineered proteins. J Biol Chem 1998;273:6248–6253.