UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Caracterização da proteína prefenato desidratase de *Mycobacterium tuberculosis* H37Rv

ANA LUIZA VIVAN

Tese submetida ao Programa de Pós Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para obtenção do título de Doutor em Ciências.

Orientador: Prof. Dr. Diógenes Santiago Santos

Co-orientador: Prof. Dr. Luiz Augusto Basso

Porto Alegre, julho de 2007.

O presente trabalho foi desenvolvido no Centro de Pesquisas em Biologia Molecular e Funcional – CPBMF localizado no TECNOPUC e pertencente ao Instituto de Biociências da Pontifícia Universidade Católica do Rio Grande do Sul – PUCRS.

Parte do trabalho foi desenvolvido no Laboratório de Biossistemas Moleculares na Universidade Paulista Júlio de Mesquita Filho – UNESP – Campus São José do Rio Preto – SP, sob orientação do Prof. Dr. Walter Filgueira de Azevedo Jr.

Este trabalho foi financiado pelo Projeto Millenium – Ministério da Ciência e Tecnologia – CNPq e a bolsa de doutorado foi concedida pela Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior – CAPES.

AGRADECIMENTOS

Gostaria de agradecer imensamente ao **Prof. Dr. Diógenes Santiago Santos**, orientador da presente tese, e ao **Prof. Dr. Luiz Augusto Basso**, pela oportunidade de realização deste trabalho no CPBMF – PUCRS.

Agradeço ao **Prof. Dr. Walter Filgueira de Azevedo Jr** pela oportunidade de desenvolver parte da tese em seu laboratório de Biosistemas Moleculares na Unesp – São José do Rio Preto – SP.

Ao Prof**. Dr. Júlio César Borges** - USP – São Carlos e ao **Prof. Dr. Carlos Ramos** – LNLS, pela importante contribuição nos estudos de ultracentrifugação analítica e dicroísmo circular.

Ao **Prof. Dr. Jose Ramon Beltran Abrego**, pelo auxílio na coleta de dados de SAXS e análise dos resultados.

A **CAPES** pela concessão da bolsa de doutorado.

Ao Projeto **Millennium MCT-CNPq** pelo apoio financeiro.

Ao Programa de Pós Graduação em **Genética e Biologia Molecular** pela oportunidade de realização do curso.

Ao **Laboratório Nacional de Luz Síncrotron**.

Aos amigos **Márcio Dias** e **Rafael Cáceres**, pelo apoio e participação no projeto e artigos publicados.

Aos amigos que fiz em São José do Rio Preto: **Marisa, Valmir, Pepeu, Márcia, Bruno, Janaína, Denise, Diego e Fernanda**.

A **Ana Paula** e todos os colegas do **CPBMF**, pela amizade e carinho.

Aos amigos do Laboratório de Bioquímica Estrutural da PUCRS (LaBioQuest): **Luís Fernando, Ivani, Guy, Paty, Raquel Dias, Dani e Raquel**.

As amigas do coração sempre presentes em todos os momentos: **Alessandra, Aline, Cynthia, Fabi, Ju Petek, Ju Cruz, Ju Mileski, Joana, Luciana, Lu Cardoso e Elma.**

Aos meus pais, **Delvina e Luiz** e minha irmã, **Márcia**, pela confiança e amor.

A todos muito OBRIGADA!

ÍNDICE

Lista de Abreviaturas

AIDS: Síndrome da Imunodeficiência Adquirida

AUC: *Analytical ultracentrifugation* ou ultracentrifugação analítica

CD: Dicroísmo circular

DOTS: *Directly Observed Therapy Short Course*

DMSO: dimetilsulfóxido

HIV: vírus da imunodeficiência humana

IPTG: isopropil β- *D*-tiogalactopiranosídeo

LB: Luria-Bertani

MDR-TB: *Multidrug resistant TB* ou tuberculose resistente a múltiplas drogas

OMS: Organização Mundial da Saúde

PAS: ácido para-aminosalicílico

PDT: prefenato desidratase

SAXS: *Small Angle x-ray scattering* ou Espalhamento de Raios X a baixo ângulo

TB: tuberculose

WHO: *World Health Organization*

XDR-TB: *Extensively drug resistant TB*

RESUMO

A tuberculose (TB) é uma das maiores causas de mortalidade em todo o mundo por um único agente infeccioso. Segundo a Organização Mundial da Saúde (OMS), todos os anos cerca de 8 milhões de pessoas são infectadas pelo *Mycobacterium tuberculosis*, o agente etiológico da TB, e aproximadamente 2 milhões morrem em todo mundo. Desde a década de 80, o aumento da prevalência da TB em países em desenvolvimento, o aumento e a proliferação de cepas resistentes a múltiplas drogas, a falta de recursos públicos adequados para o tratamento e a alta incidência de pacientes infectados pelo HIV, destacou a necessidade de desenvolver novos agentes antimicobacterianos. A via do ácido chiquímico ou chiquimato está presente em bactérias, algas, fungos, plantas e parasitas do filo apicomplexa. Esta via leva a biossíntese de corismato, um importante precursor metabólico de ácido fólico, menaquinonas, micobactinas e aminoácidos aromáticos. A primeira reação na biossíntese de fenilalanina é catalizada pela corismato mutase e involve a conversão de corismato a prefenato. A segunda reação é catalisada pela prefenato desidratase que realiza a descarboxilação e desidratação de prefenato a fenilpiruvato. As enzimas presentes na via do chiquimato e biossíntese de fenilalanina parecem ser essenciais ao *M. tuberculosis*, o que as torna promissoras como alvo para o desenvolvimento de novas drogas antimicobacterianas. O gene *phe*A que codifica a enzima prefenato desidratase (PDT) foi amplificado por PCR do DNA genômico, clonado em vetor pET23a(+) e superexpresso em células *E.coli* BL21(DE3). A proteína recombinante foi purificada por FPLC e a cristalização foi realizada pela técnica de difusão de vapor “hanging drop”. Os primeiros cristais apareceram sete dias após as gotas terem sido feitas. Cristais difrataram a 3.2 Á de resolução usando uma fonte de radiação síncrotron e pertencem ao grupo orthorrombico I222 ou $I2_12_12_1$. Análise da estrutura por substituição molecular foi realizada, contudo os resultados não foram satisfatórios. Assim, outras ferramentas foram utilizadas para inferir a estrutura da PDT. A técnica de modelagem molecular foi usada para inferir a estrutura tridimensional da PDT. Dicroísmo circular e ferramentas de bioinformática mostraram resultados similares quanto à estrutura secundária, sendo formada por 33% de hélices alfas e 18% de fitas betas. Dessa maneira, técnicas de espalhamento de raios X a baixo ângulo somado com ultracentrifugação analítica mostraram que o estado oligomérico da PDT é um homotetrâmero e com forma de um disco achatado. Dinâmica molecular demonstrou que o modelo predito é estável durante uma trajetória de 6ns. Dados experimentais e ferramentas de bioinformática corroboram os resultados aqui apresentados, dando evidências da estrutura tetramérica da PDT em solução. Além disso, este é o primeiro relato da caracterização estrutural da PDT de *M.tuberculosis*.

ABSTRACT

Tuberculosis (TB) is one of the major causes of deaths worldwide from an infectious agent. According to the World Health Organization (WHO), every year 8 million people are infected by *Mycobacterium tuberculosis*, the etiological agent of TB, and approximately 2 million people died in the world. Since 1980's the increasing of prevalence of TB in developing countries, the emergence and proliferation of multidrug-resistant and extensively drug resistant strains, the lack of adequate public resources for treatment and the high incidence of HIV-infected populations have highlighted the need for developing new antimycobacterial agents. The shikimate pathway is present in mycobacteria and absent in algae, plants, fungi and parasite of the phylum apicomplexa. This pathway leads to the biosynthesis of chorismate, an important metabolic precursor of folic acid, menaquinones, mycobactinas and aromatic amino acids. The first reaction in phenylalanine biosynthesis is catalysed by chorismate mutase and involves the conversion of chorismate to prephenate. The second reaction is catalysed by prephenate dehydratase and involves the decarboxilation and dehydratation of prephenate to phenylpiruvate. The enzymes of this pathway seem to be essential to M. tuberculosis and can thus be used as targets for the development of new antimycobacterial drugs. The *phe*A gene that encodes to prephenate dehydratase (PDT) was amplified by PCR from genomic DNA, cloned in pET23a(+) and overexpressed in *E. coli* BL21(DE3). The recombinant protein was purified by FPLC, and crystallization was performed by the hanging drop vapor diffusion method. Crystals appear seven days after drops were done. The crystal diffracted at 3.2 Å resolution using a synchrotron radiation source and belong to the orthorhombic group I222 or $I2_12_12_1$. Molecular replacement was used to solve the structure of PDT, but did not yield meaningful results. Other tools were used to infer the structural arrangement of PDT. Molecular modeling were used to infer the three dimensional structure of PDT. Circular dicroism and bioinformatics tools were in agreement in about secondary structure, showing 33% of α-helix e 18% of beta sheet. Small Angle X-Ray scattering and analytical centrifugation showed that the oligomeric state of this protein is a homotetramer and the PDT is a flat disk protein. Molecular dynamics showed that this model is stable at a trajectory of 6ns. Experimental and bioinformatics tools were in agreement and provide evidence for a tetrameric structure of PDT at solution. Therefore, this is the first report of a structural characterization of PDT from *M.tuberculosis*.

INTRODUÇÃO

O Problema: Tuberculose

A tuberculose (TB) é, depois da Síndrome da Imunodeficiência Adquirida (AIDS), a segunda maior causa global de mortes causada por um agente infeccioso (Frieden *et al*, 2003). Em 1993 a Organização Mundial da Saúde declarou a TB estado de emergência global, sendo a única doença a levar esta designação. Em 2005, foram estimados 8,8 milhões de novos casos de TB e aproxidamente dois milhões de óbitos (WHO, 2007). Estima-se que um terço da população mundial esteja infectado com o bacilo causador da doença, contudo somente 10% desses indivíduos desenvolverão a doença ativa (Donoghue *et al*, 2004). O desenvolvimento da doença é regulado pela integridade do sistema imune do hospedeiro, que pode permanecer com o bacilo latente no organismo ou desenvolver a doença ativa.

O ressurgimento da TB na década de 80 pode ser atribuído a vários fatores, como o aumento na resistência a drogas, o surgimento da AIDS, aumento na imigração de indivíduos de países com alta prevalência da doença, transmissão em prisões, hospitais, abrigos e a deterioração do sistema de saúde (Brennan, 1997; Fatkenheuer *et al*, 1999). Embora tenha ressurgido em países desenvolvidos, a TB sempre foi um problema de saúde pública presente em países em desenvolvimento como o Brasil (Ruffino-Neto, 2002). Estima-se que

95% dos casos ocorram em países em desenvolvimento, onde ocorrem 98% das mortes, ocupando o Brasil a 13ª posição entre 22 países onde a TB é bem disseminada (Ruffino-Neto, 2002).

O HIV (Vírus da Imunodeficiência Adquirida), agente etiológico da AIDS, é capaz de deteriorar o sistema imune, o que facilita a infecção ou a reativação da TB, sendo esta responsável pela maior mortalidade entre os pacientes soropositivos (Fatkenheuer *et al.*, 1999). Em 2000, cerca de dois milhões de pessoas morreram de TB e 13% destas estavam coinfectadas com o HIV (Frieden *et al.,* 2003). Assim, atualmente a AIDS representa o maior fator de risco para o desenvolvimento de TB, aumentando suas taxas de mortalidade e morbidade. De acordo com a Organização Mundial da Saúde, há aproximadamente 20 milhões de pessoas infectadas na África com HIV e pelo menos a metade está coinfectada com TB, sendo que a incidência no Brasil é de aproximadamente 200.000 coinfectados com TB e HIV.

A transmissão da TB ocorre pelo ar quando indivíduos com a doença ativa falam, tossem ou espirram liberando partículas contendo o bacilo (Frieden *et al.*, 2003). Essas partículas são extremamente pequenas, com 1-5 μm de diâmetro, e por isso são capazes de alcançarem os alvéolos pulmonares e ali se replicarem (Dunlap *et al*, 2000). Indivíduos com TB latente, mas que não desenvolveram a doença, não são capazes de transmitir o microrganismo, sendo quatro os fatores determinantes na transmissão da TB: o número de

organismos que são liberados, a concentração de bacilos determinada pelo volume do espaço e ventilação, o tempo de exposição ao bacilo e o estado imune do indivíduo exposto (Dunlap *et al*., 2000).

O principal agente etiológico da TB humana foi descoberto por Robert Koch em 1882 sendo uma bactéria aeróbia, fracamente gram positiva conhecida como *Mycobacterium tuberculosis* (Bloom e Murray, 1992). Como uma das principais características do bacilo, a parede celular micobacteriana é formada de lipídeos de cadeia longa, conhecidos como ácidos micólicos e o peptidioglicano contém ácido N-glicolilmurâmico ao contrário das demais bactérias que apresentam ácido N-acetilmurâmico (Jarlier e Nikaido, 1994).

O seqüenciamento completo do genoma de *Mycobacterium tuberculosis,* linhagem padrão H37Rv, foi realizado por Cole e colaboradores em 1998 demonstrando que este patógeno apresenta um cromossomo circular com 4.411.529 pares de bases e 65,6% de bases G e C (Cole *et al*, 1998). Este estudo demonstrou que mesmo sendo o genoma de *M. tuberculosis* menor que o de outras bactérias, como *E. coli*, é um genoma muito versátil apresentando genes codificadores para vias anabólicas e catabólicas e síntese e degradação de aminoácidos, sendo capaz de expressar enzimas envolvidas na lipólise e lipogênese. Análise da sequência genômica demonstrou o metabolismo dinâmico do bacilo, apresentando rotas metabólicas como via da pentose fosfato, ciclo glioxalato e glicólise, sendo que comparações por homologia com

outras bactérias demonstrou a presença da via do ácido chiquímico e as vias de biossíntese de aminoácidos aromáticos. A identificação e validação de rotas essenciais ao microrganismo que sejam ausentes no hospedeiro são de grande importância para o desenvolvimento de inibidores específicos de baixa toxicidade (Duncan, 2003).

A via do ácido chiquímico está presente em bactérias, algas, plantas, fungos e parasitas do Filo Apicomplexa, estando ausente em mamíferos (Pittard, 1996; Roberts *et al*, 1998). Esta via realiza a conversão de fosfoenolpiruvato e eritrose-4-fosfato a corismato, um importante precursor metabólico na biossíntese de biomoléculas essenciais como ácido fólico, micobactinas, menaquinonas e aminoácidos aromáticos (Pittard, 1996). A essencialidade desta via pode ser demonstrada pela disrupção do gene *aro*K de *M. tuberculosis* (Parish e Stoker, 2002). As enzimas da via do chiquimato e biossíntese de aminoácidos aromáticos, devido a sua ausência no homem, tornam-se alvos atrativos para o desenvolvimento de drogas seletivas ao microrganismo e que, por sua vez, não apresentarão efeitos colaterais no hospedeiro.

MDR-TB

O tratamento atual consiste de duas fases: a primeira chamada de fase intensiva ou bactericida, com duração de dois meses e administração de pelo menos quatro diferentes drogas de primeira linha: isoniazida, rifampicina,

estreptomicina e pirazinamida; a segunda fase chamada de fase de esterilização com administração de isoniazida e rifampicina por pelo menos quatro meses. A utilização inadequada deste tratamento pode levar ao surgimento de cepas resistentes, sendo a prevalência ainda maior em pacientes previamente tratados com a quimioterapia do que novos casos (Pablos-Mendez *et al.*, 2002).

O surgimento de cepas resistentes é um importante fator no ressurgimento da tuberculose, aparecendo logo após a introdução dos primeiros antimicobacterianos (Pablos-Mendez *et al*, 2002). Após 4 a 6 semanas de tratamento, os sintomas da doença começam a desaparecer, o que leva os pacientes a abandonarem o tratamento. Entretanto, muitos pacientes voltam a desenvolver a doença e com isso, reiniciam a terapia, criando condições favoráveis à seleção de organismos resistentes (Bloom e Murray, 1992). Contudo, o surgimento de cepas resistentes não está diretamente envolvido com a utilização, por mais de 40 anos, de isoniazida e rifampicina (Espinal, 2003). O surgimento de cepas resistentes deve-se à utilização inadequada da terapia e da estratégia DOTS (*Directly Observed Therapy Short Course*) por pacientes com TB e onde o programa de controle da TB é fraco, com pouco apoio governamental.

Cepas resistentes a múltiplas drogas são conhecidas como MDR-TB (*multidrug resistant)* e são caracterizadas pela resistência a pelo menos duas das principais drogas utilizadas no tratamento de curta duração da TB:

isoniazida e rifampicina. O regime de tratamento para os casos de MDR-TB é conhecido como DOTS-PLUS e consiste na utilização de drogas de segunda linha que são menos eficientes e mais tóxicas, possuindo um custo maior do que drogas de primeira linha (Pablos-Mendez *et al*, 2002). A taxa de mortalidade de pacientes com MDR-TB foi estimada entre 40 e 60%, similar à taxa de pacientes com TB não tratados (Bloom e Murray, 1992).

Cepas XDR-TB (*Extensively Drug-Resistant tuberculosis*) foram identificadas como sendo resistentes a drogas de primeira linha, isoniazida e rifampicina e drogas de segunda linha como fluoroquinona, capreomicina, canamicina e amicacina. Casos de XDR-TB foram confirmados em diversos países, inclusive no Brasil. Em um estudo conduzido pelo Centro de Controle e Prevenção de Doenças (CDC) dos Estados Unidos, aproximadamente 20% dos casos novos identificados entre os anos de 2000-2004 eram MDR-TB e 2% eram XDR-TB (CDC, 2006).

Recentemente, uma amostra letal de *M.tuberculosis* XDR emergiu em um pequeno vilarejo na África do Sul, atingindo a quase totalidade da população. Os pacientes infectados com a amostra multiresistente não responderam ao tratamento com qualquer das drogas de primeira e segunda linhas disponíveis, resultando na morte de 52 pacientes de um grupo de 53 coinfectados com HIV, 16 dias após o diagnóstico (Gandhi *et al*, 2006).

Desenvolvimento de Novas Drogas

No início dos anos 90, a OMS introduziu a estratégia DOTS, de maneira a controlar a transmissão da TB e melhorar as condições de saúde. Esta estratégia consiste em serviços de microscopia, monitoramento, medicação, comprometimento político e observação direta do tratamento por agentes de saúde. Entre os anos de 1980 e 2005, 90 milhões de pacientes infectados com TB foram relatados à Organização Mundial da Saúde.

A introdução de tratamento quimioterápico contra TB começou nos anos 40 com a introdução de estreptomicina, ácido para-aminosalicílico (pouco utilizado) e isoniazida. A rifampicina foi introduzida no final dos anos 60 e a pirazinamida nos anos 70, contudo, esta última é utilizada com cautela devido a sua toxicidade (Petrini e Hoffner, 1999). A última droga foi introduzida há mais de 30 anos (O'Brien e Nunn, 2001).

O seqüenciamento do genoma de *M. tuberculosis* oferece uma nova alternativa para alvos potenciais para novas drogas. A eficiência de drogas contra TB ativa e latente, eficazes no tratamento de MDR-TB, menos tóxicas e um tratamento com menor duração são requisitos essenciais no desenvolvimento de drogas antimicobacterianas. A combinação da química moderna com *high-throughput screening* é uma poderosa estratégia na identificação de novos compostos (Young, 2003). Contudo, introduzir uma nova

droga em um programa de tratamento fraco e ineficaz apenas irá acelerar o desenvolvimento de resistência aos novos compostos (O'Brien e Nunn, 2001).

A enzima prefenato desidratase

Homólogos de enzimas da via do chiquimato (Figura 1) e biossíntese de aminoácidos aromáticos (Figura 2) foram identificados no genoma de *M. tuberculosis* H37Rv (Cole *et al*., 1998), incluindo o gene que codifica para a proteína prefenato desidratase (E.C. 4.2.1.51). Esta proteína é codificada pelo gene *pheA* (Rv3838) de 963 pares de bases e possui 386 aminoácidos, sendo a segunda enzima na via de biossíntese de fenilalanina realizando a descarboxilação e desidratação de prefenato a fenilpiruvato (Pittard, 1996).

Em *E. coli*, a prefenato desidratase está acoplada à corismato mutase, sendo uma enzima bifuncional denominada proteína-P (Davidson *et al*., 1972). A proteína P de *E. coli* possui dois sítios catalíticos acoplados independentemente e foi caracterizada como um homodímero com 43 kDa (Pittard, 1996). Em *Amycolatopsis methanolica* a prefenato desidratase foi caracterizada como homotetrâmero e subunidade com peso molecular de 34 kDa (Euverink *et al*., 1995).

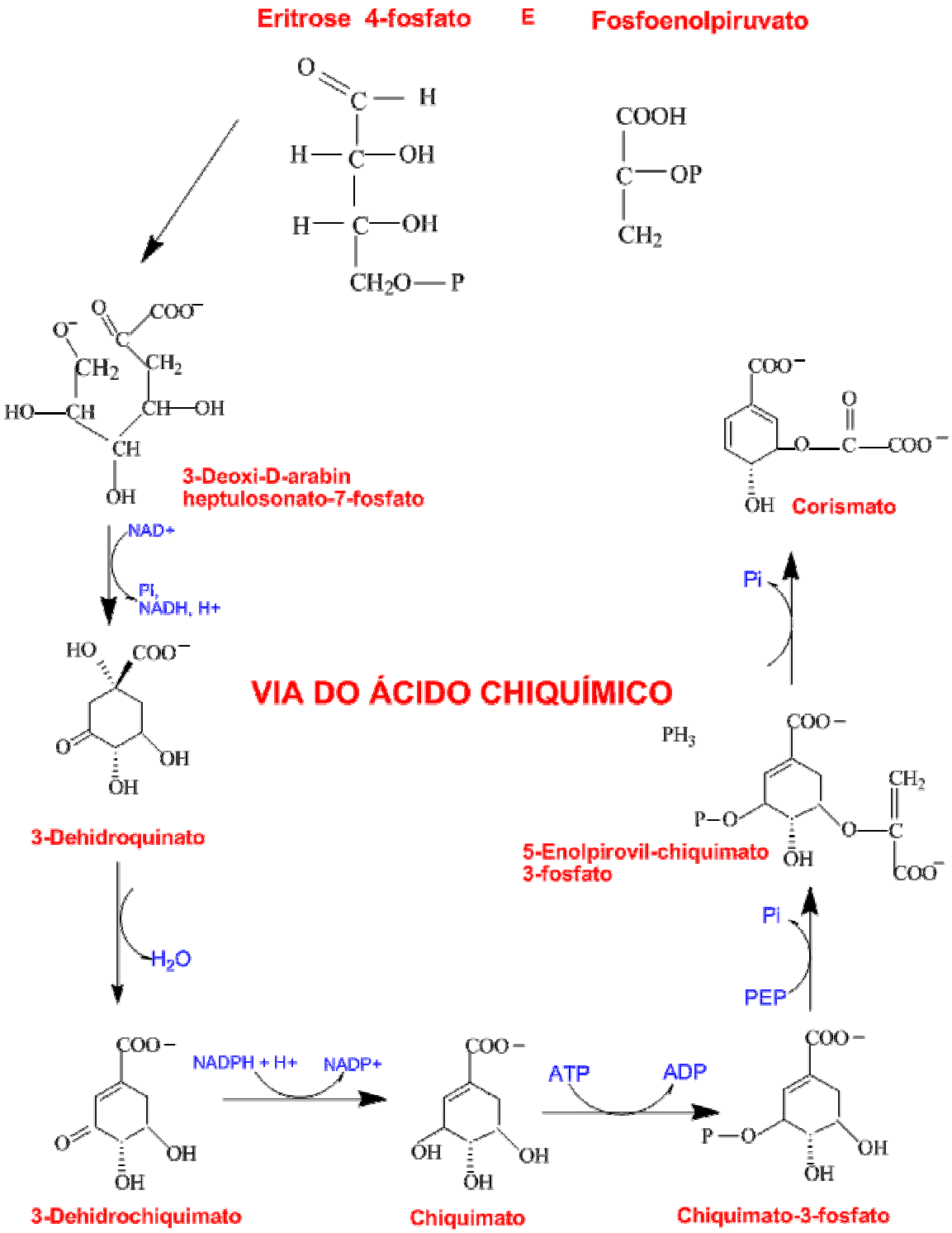

Figura 1: Representação esquemática da via do ácido chiquímico.

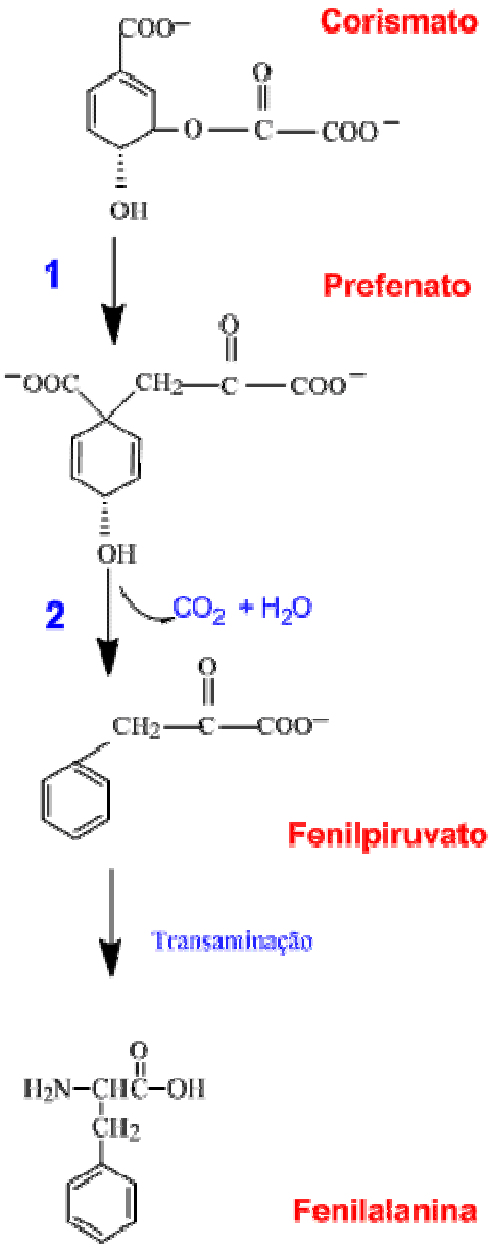

Figura 2: Representação da via de biossíntese de fenilalanina. 1: reação catalisada pela enzima prefenato desidratase. **2**: reação catalisada por uma transaminase.

Em *M. tuberculosis,* a prefenato desidratase foi anteriormente caracterizada como uma enzima monofuncional, regulada alostericamente pelos aminoácidos aromáticos (Prakash *et al*, 2005). Contudo, apesar desta enzima ter sido caracterizada bioquimicamente, nenhum trabalho elucidando sua estrutura tridimensional foi até então reportado.

Dessa maneira, a presente tese traz dois artigos desenvolvidos durante o doutorado, relatados nos CAPÍTULOS 1 e 2. O Capítulo 1 apresenta o artigo "**Crystallization and Preliminary X-ray diffraction analysis of prephenate dehydratase from *Mycobacterium tuberculosis* H37Rv**" que apresenta a clonagem, expressão, purificação e cristalização da enzima em estudo, enquanto o Capítulo 2 apresenta o artigo "**Structural studies of Prephenate Dehydratase from *Mycobacterium tuberculosis* H37Rv by SAXS, ultracentrifugation and computational analysis**" que apresenta a caracterização tridimensional da proteína através da bioinformática, ultracentrifugacão analítica e SAXS. Assim, este vem a ser o primeiro relato da estrutura tridimensional da prefenato desidratase de *M. tuberculosis* H37Rv.

OBJETIVOS

Objetivos Gerais

O objetivo geral deste trabalho foi a caracterização da proteína prefenato desidratase de *M. tuberculosis* H37Rv.

Objetivos Específicos

Os objetivos específicos deste trabalho foram purificação, cristalização e caracterização da estrutura tridimensional através de técnicas de baixa resolução como SAXS, ultracentrifugação e dicroísmo circular combinados com estudos de modelagem e dinâmica molecular.

CAPÍTULO 1

Artigo Publicado na Revista Acta Crystallographica Section F 62: 357 – 360.

"Crystallization and Preliminary X-ray diffraction analysis of prephenate dehydratase from *Mycobacterium tuberculosis* H37Rv"

Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

Crystallization and preliminary X-ray diffraction analysis of prephenate dehydratase from Mycobacterium tuberculosis H37Rv

Ana Luiza Vivan,[a,b] Márcio Vinícius Bertacini Dias,[c] Cristopher Z. Schneider,[d,b] Walter Filgueira de Azevedo Jr,[b] Luiz Augusto Basso[b]* and Diógenes Santiago Santos[b]*

[a]Programa de Pós Graduação em Genética e Biologia Molecular, Departamento de Genética, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil, [b]Centro de Pesquisa em Biologia Molecular e Funcional, Instituto de Pesquisas Biomédicas, Pontifícia Universidade Católica do Rio Grande do Sul–PUCRS, Av. Ipiranga, 6681 Prédio 92A–TECNOPUC–Partenon, CEP 90619-900, Porto Alegre, RS, Brazil, [c]Programa de Pós Graduação em Biofísica Molecular, Departamento de Física, IBILCE/UNESP, São José do Rio Preto, SP, 15054-000, Brazil, and [d]Programa de Pós Graduação em Biologia Celular e Molecular, Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

Correspondence e-mail: luiz.basso@pucrs.br, diogenes@pucrs.br

Received 20 December 2005
Accepted 20 February 2006

Tuberculosis remains the leading cause of mortality arising from a bacterial pathogen (Mycobacterium tuberculosis). There is an urgent need for the development of new antimycobacterial agents. The aromatic amino-acid pathway is essential for the survival of this pathogen and represents a target for structure-based drug design. Accordingly, the M. tuberculosis prephenate dehydratase has been cloned, expressed, purified and crystallized by the hanging-drop vapour-diffusion method using PEG 400 as a precipitant. The crystal belongs to the orthorhombic space group I222 or $I2_12_12_1$, with unit-cell parameters a = 98.26, b = 133.22, c = 225.01 Å, and contains four molecules in the asymmetric unit. A complete data set was collected to 3.2 Å resolution using a synchrotron-radiation source.

1. Introduction

According to the World Health Organization, one-third of the world population is asymptomatically infected with Mycobacterium tuberculosis and approximately eight million people will develop active tuberculosis (TB) and three million people will die every year (World Health Organization, 2005). The increasing prevalence of TB, the emergence of multi-drug-resistant strains (MDR-TB) as a consequence of inappropriate treatment regimens or patient non-compliance in completing the prescribed courses of therapy and the devastating effect of co-infection with HIV have highlighted the urgent need for the development of new antimycobacterial agents. The shikimate pathway is present in algae, higher plants, fungi, bacteria and parasites from the phylum Apicomplexa, but is absent in mammals (Pittard, 1996; Roberts et al., 1998). The essentiality of this pathway in M. tuberculosis has been demonstrated by disruption of the aroK gene (Parish & Stoker, 2002), which codes for the shikimate kinase that catalyzes the fifth step in chorismate biosynthesis. In mycobacteria, chorismate is the precursor of aromatic amino acids, naphthoquinones, menaquinones and mycobactins (Ratledge, 1982). Amino-acid auxotrophs of mycobacteria have displayed restricted growth in macrophages (Bange et al., 1996; Gordhan et al., 2002), thereby suggesting that amino-acid availability is limited in the vacuoles of the macrophages within which the bacillus resides (Russell, 2005). Accordingly, the aromatic amino-acid biosynthesis pathway is an attractive target for the development of antimycobacterial agents. Homologues of enzymes of the phenylalanine-biosynthesis pathway have been identified in the genome sequence of M. tuberculosis H37Rv strain (Cole et al., 1998), including the pheA-encoding prephenate dehydratase gene (Rv3838c). Prephenate dehydratase catalyzes the decarboxylation and dehydration of prephenate to form phenylpyruvate, which in turn is converted to phenylalanine by the tyrosine aminotransferase enzyme. In Amycolatopsis methanolica, the prephenate dehydratase was characterized as a homotetrameric protein (150 kDa) with a subunit molecular weight of 34 kDa (Euverink et al., 1995). In Escherichia coli, the pheA-encoded bifunctional P-protein possesses chorismate mutase and prephenate dehydratase activities. More recently, the M. tuberculosis pheA gene has been shown to encode an allosterically regulated monofunctional prephenate dehydratase that only catalyzes the dehydration and concomitant decarboxylation to form phenylpyruvate (Prakash et al., 2005). Here, we report an alternative cloning

strategy and the expression, purification to homogeneity and preliminary crystallographic analysis of recombinant prephenate dehydratase from M. tuberculosis. The results presented here will pave the way for the three-dimensional structure determination of wild-type M. tuberculosis prephenate dehydratase, which in turn should provide a framework on which to base the rational design of chemotherapeutic agents to treat TB. In addition, this crystal structure will provide insight into the allosteric activation of prephenate dehydratase by phenylalanine, tyrosine and tryptophan.

2. Materials and methods

2.1. Cloning, protein expression and purification

The pheA gene (Rv3838c) encoding prephenate dehydratase from

(PCR) from genomic DNA. The forward (5′-TGCATATGGTGC-GTATCGCTTACCTCGGTCC-3′) and reverse (5′-ACAAGCTTT-CATGCTTGCGCCCCCTGGTCG-3′) synthetic oligonucleotide primers were based on the amino-terminal coding and carboxy-terminal non-coding strands of the pheA gene (Cole et al., 1998) containing 5′ NdeI and 3′ HindIII restriction sites (shown in bold), respectively. The PCR product was cloned into pET-23a(+) expression vector (Novagen) and the recombinant plasmid was sequenced to confirm the identity of the cloned DNA fragment and to ensure that no mutations had been introduced by the PCR amplification step. To overexpress M. tuberculosis prephenate dehydratase protein, E. coli BL21(DE3) electrocompetent cells were transformed with the recombinant plasmid. Transformed cells were grown in Luria–Bertani (LB) medium containing 50 mg ml^{-1} carbenicillin at 310 K to an OD_{600} value of 0.5 and were then grown for an additional 3, 6, 9, 12 or 24 h in the presence or absence of 1 mM isopropyl -d-thiogalacto- sic (IPTG). Cells were harvested by centrifugation (17 300g for 15 min at 277 K), resuspended in 50 mM Tris–HCl pH 7.8 and lysed by sonication. All subsequent steps were performed on ice or at 277 K. The cell lysate was centrifuged at 48 000g for 1 h at 277 K. The supernatant was incubated with 1%(w/v) streptomycin sulfate for 30 min and centrifuged (48 000g for 30 min at 277 K). A saturated solution of ammonium sulfate was added to the supernatant to a final concentration of 30% and the mixture was centrifuged at 48 000g for 30 min at 277 K. The resultant pellet was resuspended in 50 mM Tris-HCl pH 7.8 and clarified by centrifugation. The supernatant was loaded onto a Q-Sepharose Fast Flow (2.6 8.2 cm) anion-exchange column (GE Healthcare) and fractionated using a 0.0–0.5 M NaCl linear gradient. The fractions were pooled and ammonium sulfate was added to a final concentration of 0.6 M; the mixture was then loaded onto a HiLoad 16/10 Phenyl Sepharose HP hydrophobic interaction column (GE Healthcare). The active fractions were loaded onto a Mono Q HR 16/10 anion-exchange column (GE Healthcare) and eluted using a 0.0–0.5 M NaCl linear gradient.

2.2. Enzyme assay

Prephenate dehydratase activity assays were carried out as described elsewhere (Gething et al., 1976). In summary, 400 ml reaction mixture containing 0.5 mM barium prephenate, 1.0 mM EDTA, 0.01% bovine serum albumin and 50 mM DTT in 20 mM Tris–HCl pH 8.2 was pre-incubated for 5 min at 310 K. Recombinant prephenate dehydratase was added to this mixture and incubated for 5 min at 310 K. The reaction was stopped by the addition of 800 ml 1.5 M NaOH and phenylpyruvate formation was measured at 320 nm. Appropriate controls were performed in the absence of recombinant enzyme solution. The molar extinction coefficient value of 17 500 M^{-1} cm^{-1} (Cotton & Gibson, 1965) was used to calculate the phenylpyruvate concentration.

2.3. Determination of protein concentration

Protein concentrations were determined by the method of Bradford et al. (1976) using the Bio-Rad Laboratories protein-assay kit and bovine serum albumin as standard.

2.4. Crystallization

Crystallization trials were initially performed by the hanging-drop vapour-diffusion method at 292 K. Hampton Crystal Screen and Crystal Screen 2 kits (Hampton Research) were used to determine the initial crystallization conditions. Hanging drops were prepared by

in 50 mM Tris–HCl pH 7.8 and 1 ml reservoir solution. Crystals were obtained with a reservoir solution containing 0.1 M HEPES pH 7.5, 28%(v/v) PEG 400, 0.2 M calcium chloride.

2.5. Data collection

The data set for recombinant M. tuberculosis prephenate dehydratase was collected at a wavelength of 1.438 Å using a synchrotron-radiation source (Station MX1, LNLS, Campinas) and a MAR CCD detector. The crystal was flash-frozen at 100 K in liquid nitrogen. The oscillation range used was 0.8 , the crystal-to-detector distance was 150 mm and the exposure time was 90 s. The crystal diffracted to 3.2 Å resolution. All data were processed and scaled using the programs MOSFLM and SCALA from the CCP4 program suite (Collaborative Computational Project, Number 4, 1994).

3. Results and discussion

The probable 963 bp pheA gene was successfully PCR amplified from M. tuberculosis H37Rv genomic DNA. Under the experimental conditions tested, the presence of 10% dimethyl sulfoxide (DMSO) proved to be essential for PCR amplification. DMSO is a co-solvent that improves the denaturation of GC-rich DNA, consistent with the 65.6% GC content of M. tuberculosis genomic DNA (Cole et al., 1998). The PCR fragment was cloned into pET-23a(+) expression vector between NdeI and HindIII restriction sites. Nucleotide-sequence analysis confirmed the identity of the cloned gene and demonstrated that no mutations had been introduced by the PCR amplification step. The cloning strategy previously published for M. tuberculosis recombinant prephenate dehydratase resulted in a mutant protein containing 5′-Leu-Glu-(His) -3′ extra C-terminal amino-acid residues (Prakash et al., 2005). The cloning strategy reported here results in M. tuberculosis recombinant prephenate dehydratase with no extra amino-acid residues (wild type). Recombinant plasmid was introduced into into E. coli BL21 (DE3) host cells and the best experimental protocol for prephenate dehydratase protein expression was determined to be 9 h of cell growth in the absence of IPTG induction after a value of 0.5 for OD_{600} had been reached. Analysis by SDS–PAGE showed a significant amount of a 33 kDa protein, in agreement with the predicted molecular weight for this protein. In pET expression vectors, the target genes are cloned under control of the bacteriophage T7 late promoter and the expression hosts contain a chromosome copy of T7 RNA polymerase gene under lacUV5 control, the expression of which should be induced by IPTG (isopropyl -d-thiogalactoside) to ensure tight control of recombinant gene basal expression (Studier & Moffatt, 1986; Dubendorff & Studier, 1991). In agreement with the results

Table 1
Summary of data-collection statistics for M. tuberculosis prephenate dehydratase.

Values in parentheses refer to the highest resolution shell.

X-ray wavelength (Å)	1.438
Temperature (K)	100
Resolution range (Å)	65.94–3.20 (3.37–3.20)
Total/unique reflections	71611/23215
Space group	I222 or $I2_12_12_1$
Matthews coefficient (Å^3 Da^{-1})	2.7
Unit-cell parameters	
a (Å)	98.26
b (Å)	133.22
c (Å)	225.01
= = ( )	90
Mosaicity ( )	0.43
Data completeness (%)	94.4 (97.2)
Average I/ (I)	5.7 (1.5)
Multiplicity	3.1 (3.0)
R_{merge}† (%)	0.120 (0.434)

† $R_{merge} = \sum_h \sum_i |I(h) - \langle I(h) \rangle| / \sum_h \sum_i I(h)$ where I(h) is the intensity of reflection h, $\sum_h$ is the sum over all reflections and reflection h.

presented here, leaky expression has been shown to occur in the pET system (Kelley et al., 1995; Basso et al., 2001; Oliveira et al., 2001; Magalhães et al., 2002; Silva et al., 2003). It has been proposed that leaky protein expression is a property of the lac-controlled system as cells approach stationary phase in complex medium and that cyclic AMP, acetate and low pH are required to achieve high-level expression in the absence of IPTG induction, which may be part of a

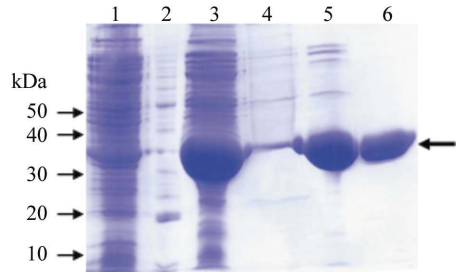

Figure 1
SDS–PAGE analysis of recombinant M. tuberculosis prephenate dehydratase protein purification. Lane 1, crude extract; lane 2, BenchMark Ladder protein standards (Invitrogen; 10–220 kDa); lane 3, ammonium sulfate precipation; lane 4, Q-Sepharose Fast Flow anion exchange; lane 5, phenyl Sepharose HP hydrophobic interaction; lane 6, Mono Q HR anion exchange.

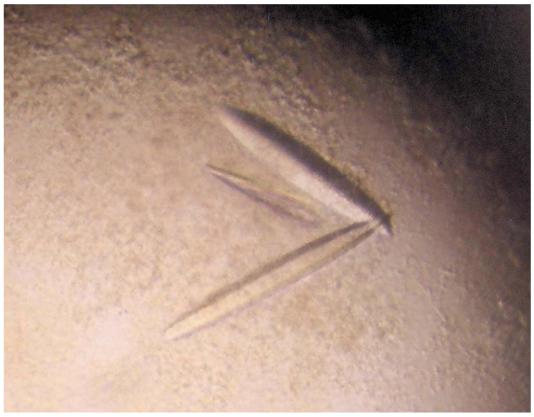

Figure 2
Crystal of recombinant M. tuberculosis prephenate dehydratase with approximate dimensions of 0.5 0.01 0.02 mm.

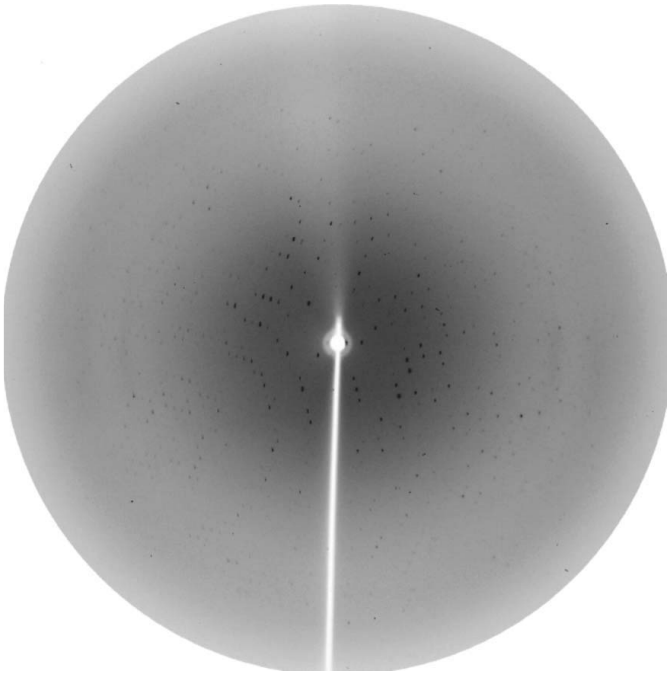

Figure 3
X-ray diffraction pattern of recombinant M. tuberculosis prephenate dehydratase.

general cellular response to nutrient limitation (Grossman et al., 1998).

The protocol described in x2 resulted in a 2367-fold protein purification, yielding approximately 70 mg of homogeneous recombinant M. tuberculosis prephenate dehydratase (Fig. 1) from 3 l growth medium (approximately 15 g cell paste) with a specific activity value of 1.42 U mg^{-1}. This value for specific activity is lower than that previously reported (Prakash et al., 2005) probably because the prephenate substrate used in the assay described here is not saturating. Small crystals of prephenate dehydratase appeared after 7 d (Fig. 2). The crystals diffracted to 3.2 Å resolution and belong to the orthorhombic space group I222 or $I2_12_12_1$ (Fig. 3). Assuming the asymmetric unit content to be four monomers of molecular weight 33 600 Da, the V_M value is 2.7 Å^3 Da^{-1} (Matthews, 1968). Assuming a value of 0.74 cm^3 g^{-1} for the protein partial specific volume, the calculated solvent content in the crystal is 54.7%. Unit-cell parameters and other statistics are given in Table 1. Attempts were made to solve the three-dimensional structure by molecular replacement using the program AMoRe (Navaza, 1994) with phenylalanine hydroxylase as the search model (Kobe et al., 1999; PDB code 2phm), since it showed the highest degree of similarity (60%) using the MULTALIGN program. However, no meaningful result was obtained, which could be a consequence of the low identity (20%) between the search model and the structure under study. Heavy-atom screening is currently under way in order to try to solve the structure by multiple isomorphous replacement. The work described here will pave the way for M. tuberculosis prephenate dehydrogenase structure determination, which in turn should provide a three-dimensional framework on which to base the design of antimycobacterial agents.

Financial support for this work was provided by Millennium Initiative Program MCT–National Research Council of Brazil (CNPq), Ministry of Health–Secretary of Health Policy and PRONEX/FAPERGS/CNPq (Brazil) to DSS and LAB. WFA thanks

FAPESP (SMOLBNet, Proc. 01/07532-0 and 03/12472-2). DSS (304051/1975-06), LAB (520182/99-5) and WFA (300851/98-7) are research career awardees from CNPq. ALV is a scholarship holder from CAPES, Brazil. We thank the LNLS technical staff for assistance during data collection.

References

Bange, F. C., Brown, A. M. & Jacobs, W. R. Jr (1996). Infect. Immun. **63**, 1794–1799.
Basso, L. A., Santos, D. S., Shi, W., Furneaux, R. H., Tyler, P. C., Schramm, V. L. & Blanchard, J. S. (2001). Biochemistry, **40**, 8196–8203.
Bradford, M. M., McRorie, R. A. & Williams, W. L. (1976). Anal. Biochem. **72**, 248–254.
Cole, S. T. et al. (1998). Nature (London), **393**, 537–544.
Collaborative Computational Project, Number 4 (1994). Acta Cryst. D**50**, 760–763.
Cotton, R. G. H. & Gibson, F. (1965). Biochim. Biophys. Acta, **100**, 75–88.
Dubendorff, J. W. & Studier, F. W. (1991). J. Mol. Biol. **219**, 45–59.
Euverink, G. J. W., Wolters, D. J. & Dijkhuizen, L. (1995). Biochem. J. **308**, 313–320.
Gething, M. J. H., Davidson, B. E. & Dopheide, T. A. A. (1976). Eur. J. Biochem. **71**, 317–325.
Gordhan, B. G., Smith, D. A., Alderton, H., McAdam, R. A., Bancroft, G. J. & Mizrahi, V. (2002). Infect. Immun. **70**, 3080–3084.
Grossman, T. H., Kawasaki, E. S., Punreddy, S. R. & Osburne, M. S. (1998). Gene, **209**, 95–103.
Kelley, K. C., Huestis, K. J., Austen, D. A., Sanderson, C. T., Donoghue, M. A., Stickel, S. K., Kawasaki, E. & Osburne, M. (1995). Gene, **156**, 33–36.
Kobe, B., Jennings, I. G., House, C. M., Michell, B. J., Goodwill, K. E., Santarsiero, B. D., Stevens, R. C., Cotton, R. G. H. & Kemp, B. E. (1999). Nature Struct. Biol. **6**, 442–448.
Magalhães, M. L. B., Pereira, C. P., Basso, L. A. & Santos, D. S. (2002). Protein Expr. Purif. **26**, 59–64.
Matthews, B. W. (1968). J. Mol. Biol. **33**, 491–497.
Navaza, J. (1994). Acta Cryst. A**50**, 157–163.
Oliveira, J. S., Pinto, C. A., Basso, L. A. & Santos, D. S. (2001). Protein Expr. Purif. **22**, 430–435.
Parish, T. & Stoker, N. G. (2002). Microbiology, **148**, 3069–3077.
Pittard, A. J. (1996). Escherichia coli and Salmonella: Cellular and Molecular Biology, edited by F. C. Neidhardt, pp. 458–484. Washington: ASM Press.
Prakash, P., Pathak, N. & Hasnain, S. E. (2005). J. Biol. Chem. **280**, 20666–20671.
Ratledge, C. (1982). The Biology of the Mycobacteria, edited by C. Ratledge & J. L. Stanford, pp. 185–271. London: Academic Press.
Roberts, F., Roberts, C. W., Johnson, J. J., Kylell, D. E., Krell, T., Coggins, J. R., Coombs, G. H., Milhous, W. K., Tziporis, S., Ferguson, D. J. P., Chakrabarti, D. & McLeod, R. (1998). Nature (London), **393**, 801–805.
Russell, D. G. (2005). Tuberculosis and the Tubercle Bacillus, edited by S. T. Cole, K. D. Eisenach, D. N. McMurray & W. R. Jacobs Jr, pp. 427–435. Washington DC: ASM Press.
Silva, R. G., Carvalho, L. P. S., Oliveira, J. S., Pinto, C. A., Mendes, M. A., Palma, M. S., Basso, L. A. & Santos, D. S. (2003). Protein Expr. Purif. **27**, 158–164.
Studier, F. W. & Moffatt, B. A. (1986). J. Mol. Biol. **189**, 113–130.
World Health Organization (2005). World Health Organization Report: Global Tuberculosis Control – Surveillance, Planning, Financing. Geneva, Switzerland: World Health Organization.

CAPÍTULO 2

Artigo Submetido à Revista Proteins: Structure, Function and Bioinformatics.

"Structural studies of Prephenate Dehydratase from *Mycobacterium tuberculosis* H37Rv by SAXS, ultracentrifugation and computational analysis"

Structural Studies of Prephenate Dehydratase from *Mycobacterium tuberculosis* H37Rv by SAXS, ultracentrifugation and computational analysis

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

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 – Programa de Pós Graduação em Biologia Celular e Molecular, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre – RS, Brazil.

3 – Departamento de Física, IBILCE/UNESP, São Jose do Rio Preto – SP, Brazil.

4 - Laboratório Nacional de Luz Síncrotron; Campinas, SP, Brazil.

5 – Faculdade de Biociências, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre – RS, Brazil.

6 - Centro de Pesquisas em Biologia Molecular e Funcional, Instituto de Pesquisas Biomédicas, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre – RS, Brazil.

7 - Departamento de Química e Física Molecular, Instituto de Química de São Carlos, Universidade de São Paulo, São Carlos, SP, Brazil.

Short title: Mycobacterial prephenate dehydratase

Keywords: Molecular Modeling; Small-angle X-ray scattering; Molecular Dynamics; Analytical Ultracentrifugation; oligomeric state; Bioinformatics; three-dimensional structure; Circular Dichroism

* To whom correspondence may be addressed: Diogenes S. Santos or Walter F. de Azevedo
Centro de Pesquisas em Biologia Molecular e Funcional – PUCRS, Av. Ipiranga, 6681 Prédio 92A, Porto Alegre 90619-900, RS, Brazil
Phone/FAX: +55 51 33203629
Email: diogenes@pucrs.br or walter.junior@pucrs.br

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, catalysed by chorismate mutase. The second reaction is catalysed 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 (*Mtb*PDT) in solution. Small angle X-ray scattering and ultracentrifugation analysis showed that the protein oligomeric state is a tetramer and *Mtb*PDT is a flat disk protein. Bioinformatics tools were used to infer the structure of *Mtb*PDT. A molecular model for *Mtb*PDT is presented and molecular dynamics simulations indicate that *Mtb*PDT is stable.

Experimental and molecular modeling results were in agreement and provide evidence for a tetrameric state of *Mtb*PDT in solution.

Abbreviations used: TB, tuberculosis; MDR-TB, multi-drug resistant tuberculosis; SAXS, Small Angle X-Ray Scattering; PDT, prephenate dehydratase; RMSD, root mean square deviation; PDB, protein data bank; AUC: Analytical ultracentrifugation; CD, circular dichroism; MD, Molecular dynamics; MM, molecular mass; ns, nanoseconds; *D*, Diffusion coefficient; $D^0_{20,w}$, standard diffusion coefficient at 0 mg/mL of protein; DLS, dynamic light scattering; *Dmax,* maximum distance; f/f_0, frictional ratio; *p(r)*, distance distribution function; *s*, sedimentation coefficient; $s_{20,w}$, standard sedimentation coefficient; $s^0_{20,w}$, standard sedimentation coefficient at 0 mg/mL of protein; *Rg,* radius of gyration; Rs, Stokes Radius; [Θ], residual molar ellipticity.

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 8 million people developed TB and 3 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 non-compliance). Another contributing factor is the evolution of multi-drug 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 above, 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 50mM Tris-HCl (pH 7.8), containing 50mM NaCl and 1mM 2-mercaptoethanol. The *D* is related to the frictional coefficient (*f*) by the Einstein equation:

$$D = \frac{RT}{N_A f} \quad \text{Equation 1}$$

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:

$$f = 6\pi\eta Rs \quad \text{Equation 2}$$

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:

$$R_0 = \left(\frac{3 M_W V_{bar}}{4 \pi N_A} \right)^{1/3} \qquad \text{Equation 3}$$

Where V_{bar} is the partial specific volume. The ratio between f to f_0 supply the so called frictional ratio (f/f_0) that indicates the shape asymmetry of a hydrated protein, when compared to a globular protein of same M_W, being is an important parameter of protein structure studies [15,18]. Through the f_0, one can obtain the maximum diffusion coefficient (D_{sphere}), applying the equation 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 applying the Stokes and Svedberg equations as described above [15]. The Sednterp also estimated the buffer viscosity (η = 1.0167 x10^{-2} poise) and density (ρ = 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 1,000 µg/mL in 50mM Tris-HCl (pH 7.8) containing 50mM NaCl and 1mM 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 solves the Lamm equation in order to discriminate diffusion from the

sedimentation boundary spreading [16, 17]. The *D* estimated by the DLS experiments was used as prior knowledge for calculation of sedimentation coefficient distributions c(s). The SedFit software simulated *D* values for fitting the experimental sedimentation velocity data. The apparent sedimentation coefficients (*s*) were found as the maximum of Gaussian fitting 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 [18]. Changes in the $s^0_{20,w}$ induced by pH, ionic strength, ligands, temperature are due to conformational changes [18]. The Sednterp software estimated the standard sedimentation coefficients ($s_{20,w}$) at each protein concentration from the *s* [15,18], 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.

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

The SedFit (Version 9.4) was also employed to estimate the M_W by the maximum of Gaussian curves in the c(M) plot. This software also extracted a weigth-avarege shape factor f/f_0 which fitted well the experimental data [16,17,18]. The f/f_0 was also estimated by the ratio $s_{sphere}/s^0_{20,w}$ [18].

The sedimentation equilibrium experiments were carried out at 20°C, at speeds of 6,000, 8,000, 10,000 and 12,000 rpm with the AN-60Ti rotor, and the scan data acquisition was at 234 nm. *Mtb*PDT in 50mM Tris-HCl (pH 7.8) containing 50mM NaCl and 1mM 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]:

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

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 and/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 -1,000 µg/mL of *Mtb*PDT in 50mM Tris-HCl (pH 7.8) containing 50mM NaCl and 1mM 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:

$$I(h) = I(0)e\,[-h^2R_g^2/3] \qquad \text{Equation 6}$$

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

$$p(r) = 1/2\pi^2 \int_0^\infty (rh)I(h) \sin (rh)\, dh \qquad \text{Equation 7}$$

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 $p(r)_{exp}$ 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, equations 8 and 9, where Q is invariant [24].

$$V = 2\pi^2 I(0)/Q \qquad \text{Equation 8}$$

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

The SAXS measurements were performed at 25ºC using *Mtb*PDT in solution at concentrations of 9mg/ml, 6mg/ml and 3mg/ml in 50mM Tris-HCl pH 7.2 and 50mM 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 minutes. 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 $Å^{-1}$ < q < 0.450 $Å^{-1}$. The distance between detector and specimen was 831mm.

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 (Rg), Stokes radius (Rs), the maximum distance (*Dmax*) 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 MtbPDT. 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].

In order 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 6 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, 49.63Å, 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 dynamic 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 K to 310 K in 5 steps (50 K to 100 K, 100 K to 150 K, 150 K to 200 K, 200 K to 250 K and 250 K to 310), 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 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 X 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 (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.

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 2A shows the sedimentation coefficient distribution c(s) of the *Mtb*PDT at different protein concentration, where the maximum of Gaussian curves gave the *s*. The results suggest that the protein behaved as a monodisperse system without aggregates or other species with different oligomeric states. In order to avoid interferences caused by the solution viscosity and density [18], we determined the $s^0_{20,w}$ (Figure 2A, *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 x 10^{-7} cm^2/seg, which was similar to the experimental *D*

obtained from DLS data (Table I). The $D^0_{20,w}$ estimated for *Mtb*PDT was smaller than for it as either monomer (D_{sphere}= 10 x 10^{-7} cm^2/seg) or dimer (D_{sphere}= 7.9 x 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). This software also fitted the f/f_0 of the protein. 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 (Equation 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 (Equation 1 and 2). The ratios $s_{sphere}/s^0_{20,w}$ and $D_{sphere}/D^0_{20,w}$ showed similar values (data not shown). Therefore, the *Mtb*PDT behaved as an asymmetric tetramer.

Figure 2B 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 the experimental equilibrium

sedimentation data (Figure 2B, 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 below). Both models were evaluated as tetramers and the results are given in Table I. The Dmax observed for the models were around 145 Å long and were in good agreement with the experimental data from SAXS and with the value estimated from the hydrodynamic data. The value of predicted $s_{20,w}$ estimated for the *ab initio* model was very close to the experimental data (5.9 S *versus* 6.22 S). However, for the *Mtb*PDT homology model it was larger (6.93 S *versus* 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 ± 1,81 Å. The Guinier plot could be seen inside the Figure 3.

The pair distance distribution function, p(r) shown in the Figure 4 was calculated from the data in Figure 3. The maximum diameter Dmax 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].

In order 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 [39] 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

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 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 5A, and tetrameric *Mtb*PDT shown in Figure 5B was obtained from docking the dimeric model against itself. A low resolution beads model of tetrameric *Mtb*PDT is given in Figure 6.

Molecular dynamics

In order to verify the stability of *Mtb*PDT, MD was performed for 6 ns and trajectory was collected after 1ns to ensure complete system equilibration. The root mean square deviation (RMSD) of C-alpha (C□□ 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 3.5ns a decreasing value for RMSD of C□ was observed indicating equilibration of at least 2.5 ns of simulation□□□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 [40]. 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 mM TrisHCl, 100 mM NaCl and 1mM DTT pH 7.8 as running buffer. Our result showed that *Mtb*PDT as a pentameric enzyme (data not shown), and 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 6ns trajectory.

In future works crystallography X-Ray diffraction might be used to solve three-dimensional structure of prephenate dehydratase.

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

This work was supported by grants from CAPES and Millennium Institutes (CNPq-MCT) to DSS, LAB and WFA. 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 Epidemioly of tuberculosis. The 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. Bulletin of the World Health Organization. 2002; 80 (6): 489-500.

4. Pittard AJ, Biosyntheses of the Aromatic Amino Acids. In. Neidhardt FC, Ingraham JL, Low KB, Magasanik B, Schaechter M, Umbarger HE, editor. *Escherichia coli* and *Salmonella:* Cellular and Molecular Biology*,* vol. 1, Washington: American Society for Microbiology. 1987; 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. 393 (1998) 801-805.

6. Parish T, Stoker NG. The common aromatic amino acid biosynthesis pathway is essencial in *Mycobaterium tuberculosis.* Microbiology. 2002; 148: 3069-3077.

7. Ratledge C. The Biology of the Mycobacteria, Vol. 1, edited by Ratledge C & Stanford JL. 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. The Journal of Biological Chemistry. 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. Biochemical Journal. 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. The Journal of Biological Chemistry. 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 Jr WF, Basso LA, Santos DS. Crystallization and preliminary X-ray diffraction analysis of prephenate dehydratase of *Mycobacterium tuberculosis* H37Rv. Acta Crystallogr F, 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. Schuck P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys. J. 2000; 78: 1606-1619.

17. 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.

18. Laue T. Biophysical studies by ultracentrifugation . Curr. Opin. Struct. Biol. 2001; 11: 579-583.

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. Current Opin Structural Biology. 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. Review of Scientific Instruments. 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, In: Glatter O & Kratky O, editor. Small-angle X-ray Scattering, London: Academic Press; 1982; p 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 Jr JC, Candur F, de Azevedo Jr WF, 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. Cryst. 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. Assesment 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 Research. 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 Comp. 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. Chen R, Li L, Weng Z. PROTEINS: Structure, Function, and Genetics 2003; 52: 80-87.

39. Svergun DI, Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys J. 1999; 76: 2879–2886.

40. 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 (11): 6248-6253.

Figure 1: *Mtb*PDT presents high α-helix content. Residual molar ellipticity [Θ] of *Mtb*PDT 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^{o}C and in a 0.1 mm pathlength cell.

Figure 2: AUC experiments show *Mtb*PDT as a tetramer. A) Sedimentation velocity experiments were carried out at 20^{o}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 *Mtb*PDT 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^{o}C and in 8,000, 10,000 and 12,000 rpm.

Figure 3: Experimental SAXS curve of *Mtb*PDT. The *Mtb*PDT SAXS experiments were carried out with a protein concentration extrapolated to 0 mg/mL in 50mM Boric Acid pH 9.5 at 25º C. Circles indicate the raw data. *Inset* the Guinier plot of *Mtb*PDT.

Figure 4: (A) *Mtb*PDT 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 *Mtb*PDT using the program Pymol.

Figure 5: Ribbon diagram of molecular model of *Mtb*PDT. **(A)** View of monomer generated using PARMODEL. **(B)** The *Mtb*PDT tetramer was obtained with molecular docking.

Figure 6: Low resolution beads model of tetrameric *Mtb*PDT. Models constructed using the *ab initio* modeling program DAMMIN to analyze SAXS data. **(A)** Figure showing *Mtb*PDT at 0º. **(B)** Figure showing *Mtb*PDT at 60º **(C)** Figure showing *Mtb*PDT at 90º **(D)** Figure showing *Mtb*PDT at 180º.

Figure 7: Root mean square deviation (RMSD) of C-alpha from *Mtb*PDT for 6ns. The trajectory collection was initiated after 1ns of dynamics to ensure complete system equilibration.

Figure 8: Root mean square deviation (RMSD) of 321 residues from *Mtb*PDT**.**

Figure 1

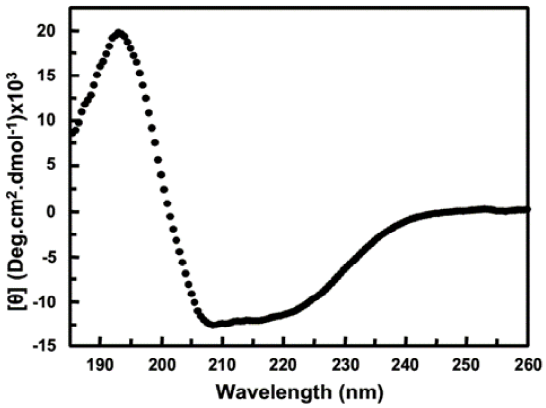
20
15
10
5
0
-5
-10
-15
[θ] (Deg.cm².dmol⁻¹)x10³
190
200
210
220
230
240
250
260
Wavelength (nm)

Figure 2A

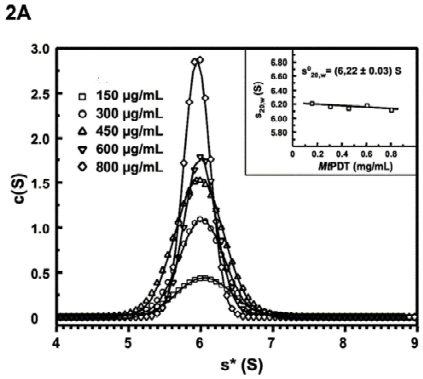

Figure 2B

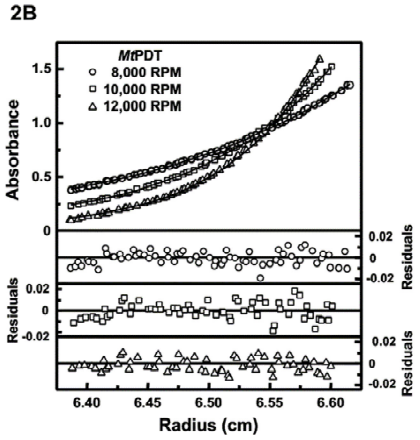

Figure 3

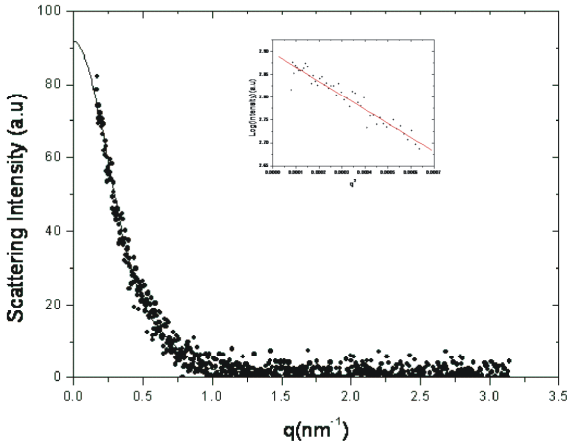
Scattering Intensity (a.u)
q(nm-1)
0
20
40
60
80
100
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5

Figure 4A

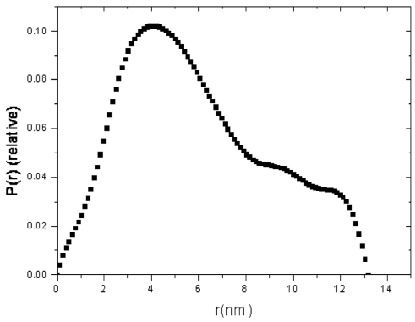

Figure 4B

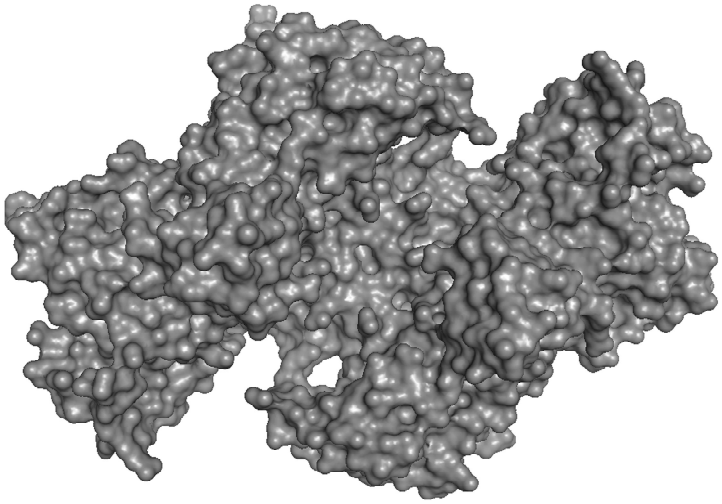

Figure 5A

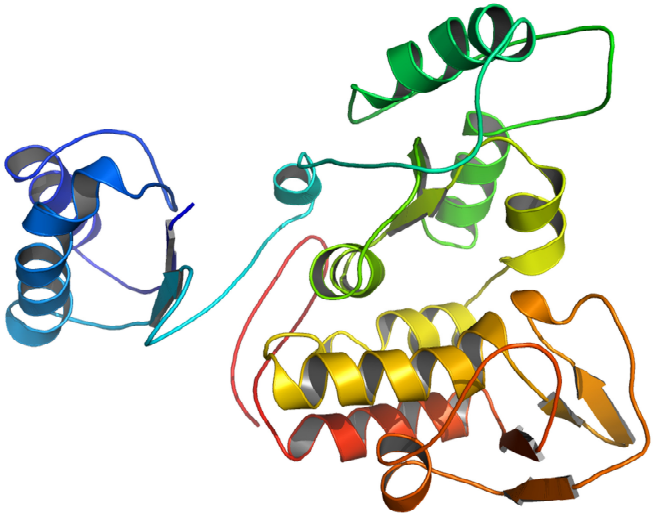

Figure 5B

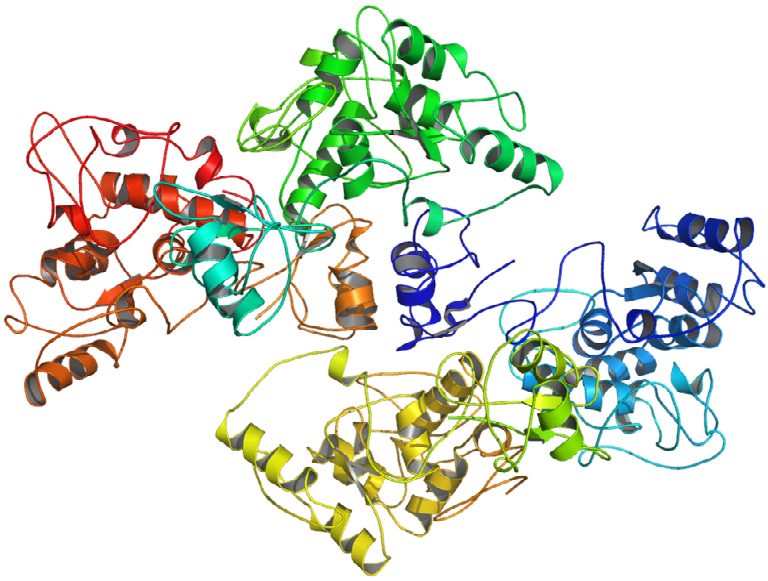

Figure 6 A

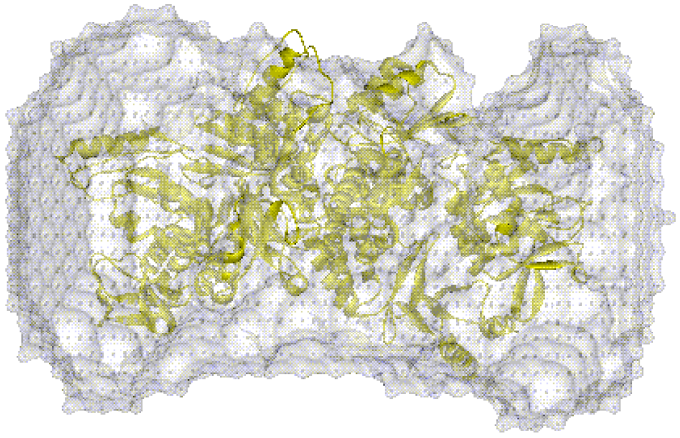

Figure 6 B

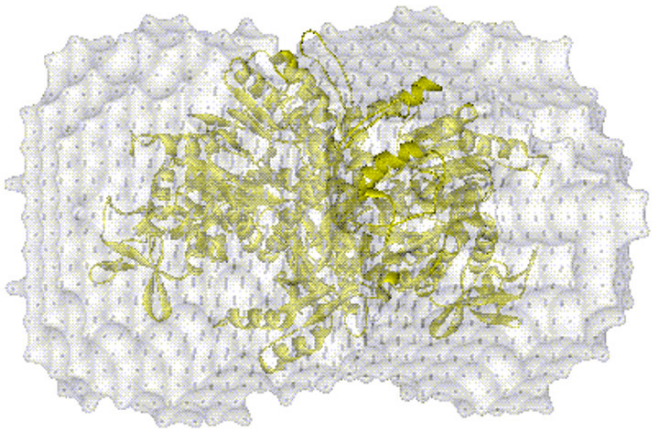

Figure 6 C

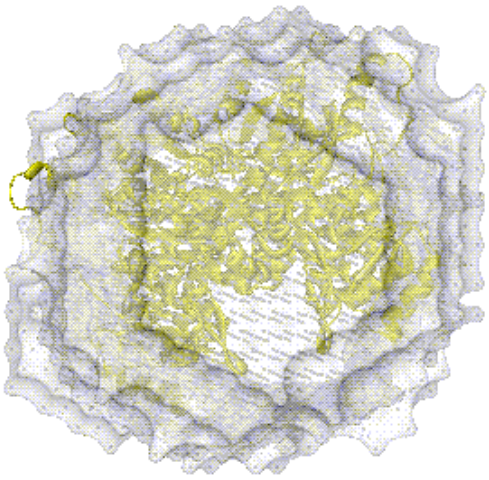

Figure 6 D

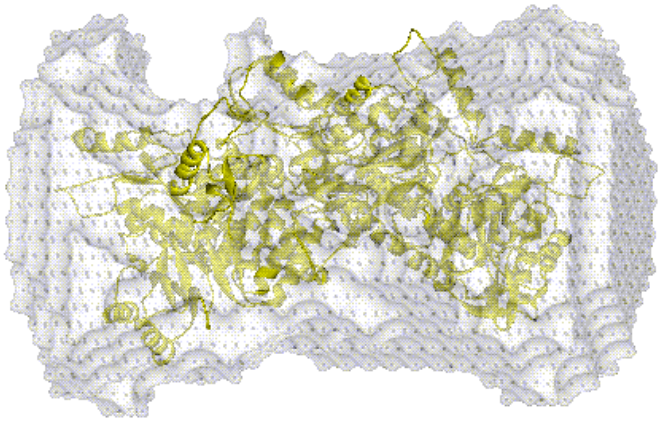

Figure 7

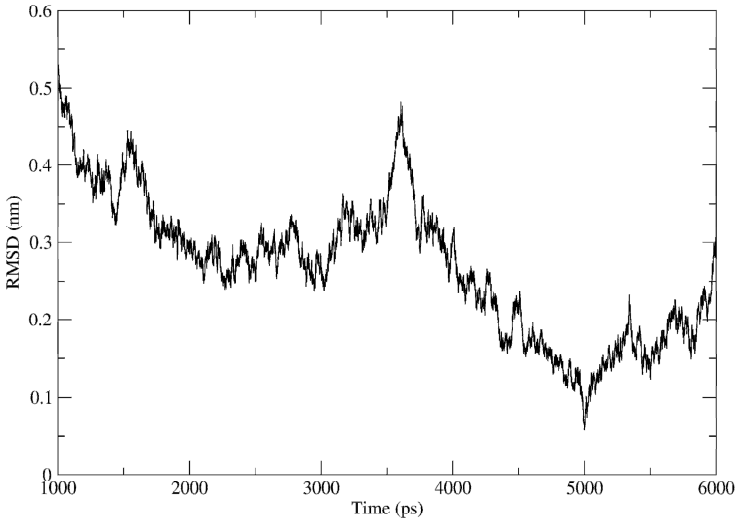

Figure 8

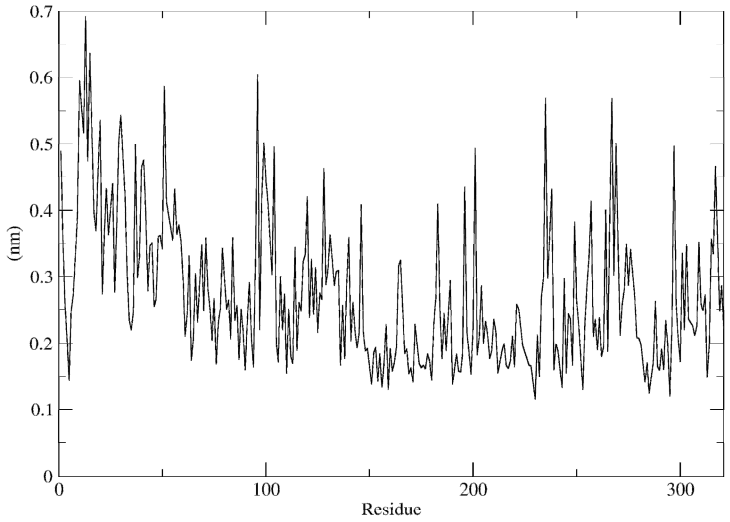

Table I
Structural and hydrodynamic parameters derived from SAXS, DLS and estimated to the PDT models by HydroPro Software.

*Mtb*PDT hydrodynamic properties	Predicted¤ (Tetramer)	DLS	AUC	HydroPro§ (using models)		SAXS
				Homology	*Ab initio*	
M_W (kDa)	134.5	125	128 ± 4[#] 133 ± 2[&] 130 ± 2[&#]	-	-	-
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[¥]	46.5	54.6	-
f/f_0	1.0	1.41	1.43 ± 0.03[#] 1.46 ± 0.01[¥]	1.33	1.56	-
Dmax (Å)	-	-	142[¥]	146	147	130
Rg (Å)	-	-	-	38.2 ± 0.1	42.3 ± 0.1	45.8

¤ Values predicted for PDT as a tetrameric globular particle (predicted by Sednterp software)

[#] calculated from SedFit c(M) fitting of the sedimentation velocity data;

[&] from equilibrium sedimentation data;

[&#] from *s/D* ratio (Equation 4);

§ 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);

[¥] data from extrapolation of the Sednterp software assuming the protein as a hydrated tetramer with an oblate shape.

DISCUSSÃO

Neste estudo foi utilizada uma série de ferramentas de biologia molecular e bioinformática para a caracterização da proteína prefenato desidratase (PDT) de *Mycobacterium tuberculosis* H37Rv. O gene *phe*A foi amplificado por PCR na presença de 10% de DMSO e clonado em vetor de expressão pET23a(+). A proteína recombinante foi superexpressa em células de *E.coli* BL21(DE3) sem a presença de IPTG, um importante indutor de expressão de proteínas que se encontram sob controle do promotor *Lac*UV5. A expressão não induzida de genes controlados pelo promotor Lac ocorre em células que se encontram em fase estacionária em meio complexo, dependendo da presença de cAMP para expressão na ausência de IPTG (Grossman *et al*., 1998). A proteína foi purificada até a homogeneidade por cromatografia líquida, com rendimento de aproximadamente 70mg a partir de 3 litros de meio de cultura LB.

Os testes de cristalização foram realizados no Laboratório de Biossistemas Moleculares, na Universidade Paulista Júlio de Mesquita Filho (Unesp) em São José do Rio Preto – SP, sob supervisão do Prof. Dr. Walter Filgueira de Azevedo Jr. A PDT foi cristalizada em solução contendo 100mM HEPES pH 7.5, 200mM cloreto de cálcio e 28% PEG 400. Os primeiros cristais apareceram sete dias após a confecção das primeiras gotas. Os dados cristalográficos foram coletados no Laboratório Nacional de Luz Síncrotron em Campinas - SP, o cristal de PDT difratou a 3.2 Å de resolução e a análise do padrão de difração de raios X demonstrou que a PDT pertence ao grupo espacial ortorrômbico I222 ou $I2_12_12_1$.

Para determinar o número de moléculas na unidade assimétrica foi utilizado o coeficiente de Mattews (V_m), que apresentou valor de 2,7 $Å^3Da^{-1}$, demonstrando que a unidade assimétrica da PDT apresenta quatro monômeros. Para resolução da estrutura tridimensional da proteína, foi utilizado o método de substituição molecular que tem como princípio a utilização de um modelo de estrutura já resolvida. Para isso a estrutura da fenilalanina hidroxilase (código PDB: 2PHM) foi utilizada como modelo (Kobe *et al.*, 1999) usando o programa AMORE (Navaza, 1994). Apesar da alta similaridade (60%), a fenilalanina hidroxilase apresentava apenas 20% de identidade, e com isso os resultados da substituição molecular não foram satisfatórios para a resolução da estrutura tridimensional da PDT. Quando não há um modelo de estrutura resolvida disponível para resolução através de substituição molecular, outras estratégias podem contribuir para a determinação da estrutura, como a técnica de substituição isomórfica múltipla (MIR - *Multiple Isomorphus Replacement*). Esta técnica consiste na utilização de substâncias contendo metais pesados adicionados ao cristal já formado, processo conhecido como *soaking*. Dessa maneira, o metal pesado irá formar um complexo com o cristal protéico e as informações originadas no padrão de difração serão obtidos analisando as trocas na intensidade das reflexões como resultado do processo de derivatização do cristal contendo o metal pesado (Ducruix e Giegé, 1999). Cristais de PDT foram submetidos ao *soaking* com diferentes metais pesados, contudo, alguns cristais não suportaram as mudanças causadas pela introdução dos metais e quebraram, enquanto outros, quando submetidos à radiação não apresentaram padrão de difração e ou resolução suficientes para que a fase pudesse ser encontrada e a estrutura da PDT fosse resolvida.

A técnica de espalhamento de raios X a baixo ângulo (SAXS - *Small Angle X-Ray Scattering*) permite inferir a estrutura da proteína em solução em baixa resolução, indicando a principal conformação que a partícula espalhada apresenta (Svergun e Koch, 2002). A PDT foi submetida à análise da forma e estado oligomérico por SAXS no Laboratório Nacional de Luz Síncrotron, sendo que os resultados demonstraram um tetrâmero com forma de disco achatado. Estes dados corroboram os dados de ultracentrifugação analítica (AUC), que também demonstrou ser a PDT um tetrâmero. Estudos feitos por dicroísmo circular mostraram resultados similares aos obtidos pela modelagem molecular sobre a estrutura secundária da proteína, apresentado aproximadamente 33% de hélices alfa e 18% de fitas betas.

Ferramentas computacionais foram utilizadas para inferir o modelo tridimensional da PDT, como a modelagem molecular que é capaz de predizer a estrutura de uma proteína através da sua seqüência de aminoácidos e homologia com uma estrutura já conhecida. Para isso, foi utilizado o software PARMODEL e como modelo a estrutura da fenilalanina hidroxilase (código PDB: 2PHM). Para verificação da qualidade estereoquímica do modelo predito utiliza-se o mapa de Ramachandran, que nada mais é que a representação visual da tendência de conformação das cadeias principais dos aminoácidos. Dessa forma, este gráfico fornece o ângulo de torsão psi (ψ) da cadeia principal pelo ângulo de torsão phi (φ) para cada resíduo de aminoácido na proteína. A análise do gráfico de Ramachandran para o modelo da PDT apresentou quase 90% dos seus

aminoácidos em regiões permitidas e nenhum aminoácido em região não permitida.

Dados experimentais realizados por AUC e pelo *software* HydroPro demostraram que o modelo predito é possível para a PDT. Simulações de dinâmica molecular confirmaram a estabilidade do modelo durante uma trajetória de 6ns.

A estrutura homotetramérica proposta para a PDT de *M.tuberculosis* é similar à descrita para a PDT de *A. methanolica* (Euverink *et al.*, 1995) e diferente da proteína-P de *E.coli*, descrita como um dímero (Zhang *et al.*, 1998). Essa discrepância com a proteína-P pode ser atribuída à bifuncionalidade da enzima, que apresenta função corismato-mutase acoplada à função prefenato desidratase. Anteriormente, um trabalho apresentou a PDT de *M.tuberculosis* como um dímero, mas a diferença com os resultados aqui apresentados deve-se as diferenças na clonagem e expressão da proteína e as técnicas utilizadas (Prakash *et al.*, 2005).

Este trabalho é o primeiro a caracterizar a estrutura tridimensional da PDT de *M. tuberculosis.* Como não há no ser humano homólogo para esta enzima, a caracterização da mesma aqui descrita torna-se importante no desenho racional de inibidores baseados em estruturas para o tratamento da TB.

BIBLIOGRAFIA

Bloom BR e Murray CJL (1992) Tuberculosis: Commentary on a Reemergent Killer. Science 257: 1055 – 1064.

Brennan PJ (1997) Tuberculosis in the context of emerging and reemerging diseases. FEMS Immunology and Medical Microbiology 18: 263 – 269.

Center for Disease Control and Prevention. (2006) Emergence of *Mycobacterium tuberculosis* with Extensive Resistance to Second-Line Drugs - Worldwide, 2000–2004. Morbidity and Mortality Weekly Report - MMWR. Vol 55, N 11: 301 – 328.

Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE, et al. (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature 393: 537 - 544.

Davidson BE, Blackburn EH e Dopheide TAA (1972) Chorismate Mutase-Prephenate dehydratase from *Escherichia coli* K-12. The Journal of Biological Chemistry 247: 4441-4446.

Donoghue HD, Spigelman M, Greenblatt CL, Lev-Maor G, Bar-Gal GK, Matheson C, Vernon K, Nerlich AG and Zink AR (2004) Tuberculosis: from prehistory to Robert Koch, as revealed by ancient DNA. The Lancet 4: 584 – 592.

Ducati RG, Ruffino-Neto A, Basso LA, Santos DS (2006) The Resumption of Consumption – A review on tuberculosis. Mem Inst Oswaldo Cruz 101(7): 697 – 714.

Ducruix A and Giegé R (1999) Crystallization of Nucleic Acids and Proteins: a pratical approach. 2nd edition. Oxford University Press, Oxford, 435 pp.

Duncan K (2003) Progress in TB drug development and what is still needed. Tuberculosis 83, 201-207.

Dunlap NE, Bass J, Fujiwara P, Hopewell P, Horsburg CR, Salfinger M, Simone, PM (2000) Diagnostic Standards and Classification of Tuberculosis in Adults and Children. American Journal of Respiratory and Critical Care Medicine. Vol 161, pp. 1376-1395.

Espinal MA (2003) The global situation of MDR-TB. Tuberculosis 83, 44-51.

Euverink GJW, Wolkers DJ, Dijkhuizen L (1995) Prephenate Dehydratase of the Actinomycete *Amycolatopsis methanolica*: Purification and Characterization of wild-type and deregulated mutant proteins. Biochemical Journal. 308: 313-320.

Fatkenheuer G, Taelman H, Lepage P, Schwenk A, Wenzel R (1999) The Return of Tuberculosis. Diag Microbiol Infec Dis 34: 139 – 146.

Frieden TR, Sterling TR, Munsiff SS, Watt CJ, Dye C. (2003) Tuberculosis. The Lancet 362: 887 – 899.

Gandhi NR, Moll A, Sturm W, Pawinski R, Govender T, Lallo U, Zeller K, Andrews J, Friedland G (2006) Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. Lancet 368: 1575 – 1580.

Grossman TH, Kawasaki ES, Punreddy SR, Osburne MS (1998) Spontaneous cAMP-dependent derepression of gene expression in stationary phase plays a role in recombinant expression instability. Gene 209, pp. 95-103.

Jarlier V and Nikaido H (1994) Mycobacterial cell wall: Structure and role in natural resistance to antibiotics. FEMS Microbiology Letters 123:11-18.

Kobe B, Jennings IG, House CM, Michell BJ, Goodwill KE, Santarsiero BD, Stevens RC, Cotton RGH, Kemp BE (1999). Structural basis of autoregulation of phenylalanine hydroxylase. Nature Structure Biology 6: 442–448.

Navaza J (1994). AMoRe: an automated package for molecular replacement. Acta Crystallographica A 50: 157–163.

O'brien RJ and Nunn PP (2001) The Need for New Drugs against Tuberculosis. Am J Respir Crit Care Med Vol 162: 1055-1058.

Pablos-Mendez A, Gowda DK, Frieden T. (2002) Controlling multidrug-resistant tuberculosis and access to expensive drugs: a rational framework. Bulletin of the World Health Organization 80 (6): 489-500.

Parish T and Stoker N (2002) The common aromatic amino acid biosynthesis pathway is essencial in *Mycobacterium tuberculosis*. Microbiology 148: 3069-3077.

Petrini B and Hoffner S (1999) Drug Resistant and multidrug-resistant tubercle bacilli. International Journal of Antimicrobial Agents 13: 93-97.

Pittard AJ (1996). Escherichia coli and Salmonella: Cellular and Molecular Biology, edited by F. C. Neidhardt, Washington: ASM Press 458–484 pp.

Prakash P, Aruna B, Sardesai AA, Hasnain SE (2005) pheA (Rv3838c) of *Mycobacterium tuberculosis* encodes an allosteric regulated monofunctional

prephenate dehydratase that requires both catalytic and regulatory domains for optimum activity. The Journal of Biological Chemistry 280: 20666-20671.

Roberts F, Roberts CW, Johnso JJ, Kylell DE, Krell T, Coggins JR, Coombs GH, Milhous WK, Triphuri S et al. (1998) Evidence for the shikimate pathway in apicomplexan parasites. Nature, Vol. 393, 25.

Ruffino-Neto A (2002) Tuberculose: a calamidade negligenciada. Revista da Sociedade Brasileira de Medicina Tropical 35(1): 51-58.

Svergun DI and Koch MHJ (2002) Advances in structure analysis using small-angle scattering in solution. Current Opinion Structural Biology 12: 654-660.

Young DB (2003) Ten years of research progress and what's to come. Tuberculosis 83: 77-81.

Zhang S, Pohnert G, Kongsaeree P, Wilson DB, Clardy J, Ganem B (1998) Chorismate mutase-Prephenate dehydratase from *Escherichia coli:* Study of catalytic and regulatory domains using genetically engineered proteins. Journal Biological Chemistry. 273 (11): 6248-6253.