UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS: BIOQUÍMICA

Chiquimato Desidrogenase de *Mycobacterium tuberculosis*: mecanismos cinético e químico da enzima recombinante

Isabel Osorio da Fonseca

Orientador: Prof. Dr. Luiz Augusto Basso

Tese apresentada ao Programa de Pós-Graduação em Ciências Biológicas: Bioquímica da Universidade Federal do Rio Grande do Sul como pré-requisito para obtenção do grau de Doutor.

Porto Alegre, agosto de 2006.

Esta tese foi julgada e aprovada para a obtenção do grau de doutor em Ciências Biológicas: Bioquímica pelo orientador e pela comissão examinadora no programa de Pós – Graduação em Ciências Biológicas: Bioquímica.

Orientador:

Prof. Dr. Luiz Augusto Basso
Faculdade de Biociências
Pontifícia Universidade Católica do Rio Grande do Sul

Comissão Examinadora:

Dr. Carlos Alberto Saraiva Gonçalves
Membro do Programa de Pós-Graduação em Ciências Biológicas: Bioquímica
Universidade Federal do Rio Grande do Sul

Dr. Arthur Germano Fett-Neto
Membro do Programa de Pós-Graduação em Biologia Celular e Molecular
Universidade Federal do Rio Grande do Sul

Dr. Walter Filgueira de Azevedo Jr.
Departamento de Ciências Fisiológicas da Faculdade de Biociências
Pontifícia Universidade Católica do Rio Grande do Sul

Este trabalho foi realizado no Centro de Pesquisas em Biologia Molecular e Funcional, Instituto de Pesquisas Biomédicas, TECNOPUC, Pontifícia Universidade Católica do Rio Grande do Sul, sob orientação do Prof. Dr. Luiz Augusto Basso.

Este trabalho é dedicado...

ao meu marido, Kiko, por todo o amor, compreensão, força, carinho e apoio incondicional.
ao meu filho, Jota, pela compreensão, carinho, paciência e apoio que me deram forças para vencer mais essa etapa.

A eles, meu MMMUUUUIIIIITTTTOOOO OBRIGADO por me amarem e estarem ao meu lado...

Agradecimentos

Ao meu orientador **Dr. Luiz Augusto Basso**, pelos ensinamentos, pela confiança... sinceramente... MUITO OBRIGADO!!!

Ao **Dr. Diógenes Santiago Santos**, pela oportunidade, pela confiança, pelos ensinamentos, pelo exemplo... MUITO OBRIGADO POR TUDO!!!

Ao **meu colega Rafael** pela enorme força para o desenho dos experimentos, análise dos resultados, por todos os ensinamentos... MMUUIIITTTTOOOOO OBRIGADO!!!

A colega **Cláudia** pelos ensinamentos em bioinformática... MUITO OBRIGADA!!!

Aos **meus pais** pela força, pelo exemplo, pelo amor, por acreditarem em mim... AMO MMUUIIITTTTOOOOO VOCÊS!!!

Aos meus colegas, **Cristopher**, **Clarissa**, **Jotinha**, **Bruna**, **Raquel**, **Zé**, **Fer, Isabelinha**... pela amizade, pelas contribuições, pelo apoio, pelas brincadeiras, pelas conversas... que tornaram ainda mais agradáveis meus dias... MUITO OBRIGADO!!!

À **Clotilde**, **Renilda** e **Gleci** pela dedicação zelando sempre a manutenção e funcionamento do laboratório... MUITO OBRIGADO!!!

Ao **Programa de Pós-Graduação em Ciências Biológicas: Bioquímica** da Universidade Federal do Rio Grande do Sul pela oportunidade que me foi dada.

À CAPES pelo apoio financeiro que me auxiliou nesse período.

Índice

Parte I 1

Resumo **2**

Abstract **3**

Lista de Abreviaturas **4**

Introdução **5**

Tuberculose **5**

Via do Chiquimato e Chiquimato desidrogenase **11**

Objetivo **19**

Parte II 21

Capítulo I: **22**

Functional shikimate dehydrogenase from *Mycobacterium tuberculosis* H37Rv: purification and characterization.

Capítulo II: **23**

Shikimate Dehydrogenase from *Mycobacterium tuberculosis* H37Rv: Kinetic and Chemical Mechanisms.

Parte III 24

Discussão

Capítulo I: **25**

Functional shikimate dehydrogenase from *Mycobacterium tuberculosis* H37Rv: purification and characterization.

Capítulo II: **35**

Shikimate Dehydrogenase from *Mycobacterium tuberculosis* H37Rv: Kinetic and Chemical Mechanisms.

Conclusão **50**

Referências **51**

Anexos

Anexo I **55**

1. AZEVEDO, WF; SILVA, RG; **FONSECA, IO**; RENARD, G; BASSO, LA; SANTOS, DS. (2003) **Docking and small angle X-ray scattering studies of purine nucleoside phosphorylase.** *Biochemical and Biophysical Research Communications* **309**, 928-933.

Anexo II **56**

2. POLETTO, SS; **FONSECA, IO**; CARVALHO, LP; BASSO, LA; SANTOS, DS. (2004) **Selection of an *Escherichia coli* host that expresses mutant forms of *Mycobacterium tuberculosis* 2-trans enoyl-ACP(CoA) reductase and 3-Ketoacyl-ACP (CoA) reductase enzymes.** *Protein Expression and Purification* **34**, 118-125.

Anexo III **57**

3. RIZZI, C; **FONSECA, I.O.**; GALLAS, M; WEBER, P; FRAZZON, J; SANTOS, DS; BASSO, LA. (2005) **DAHP synthase of *Mycobacterium tuberculosis* H37Rv: Cloning, overexpression, and purification of the functional enzyme.** *Protein Expression and Purification* **40**, 23-30.

Anexo IV **58**

4. PATENTE (2005) SANTOS, D. S.; BASSO, L.A.; RENARD, G.; **FONSECA, I.O.**; CHIES, J.M. **Método de Obtenção de Seqüências Nucleotídicas Quiméricas e Seqüência Nucleotídica Quimérica** - INPI Registro PI0506047-8.

Anexo V **59**

5. LEOPOLDINO, A.M.; CANDURI, F.; CABRAL, H.; JUNQUEIRA, M.; MARQUI, A.B.T.; APPONI, L.H.; **FONSECA, I.O.**; DOMONT, G.B.; SANTOS, D. S.; VALENTINI, S. (2006) **Expression, purification, and circular dichroism analysis of human CDK9.** *Protein Expression and Purification* **47**, 614-620.

PARTE I

Resumo

O aumento na prevalência da tuberculose (TB), a emergência de cepas resistentes a múltiplas drogas de *Mycobacterium tuberculosis*, o agente etiológico da TB, e o efeito devastador da co-infecção pelo HIV têm enfatizado a urgente necessidade no desenvolvimento de novos agentes antimicobacterianos. A análise completa da seqüência genômica do *M. tuberculosis* H37Rv mostrou a existência de genes envolvidos na via de biossíntese de aminoácidos aromáticos, a via do chiquimato, e evidências experimentais indicam que esta via é essencial para o *M. tuberculosis* e apresenta-se ausente no homem. Os produtos dos genes que são essenciais para o crescimento dos microrganismos fazem deles alvos atrativos para a ação de drogas, pois a inibição de sua função pode matar o bacilo. Previamente, nosso grupo relatou a clonagem do gene *aroE* de *M. tuberculosis* e a expressão do seu produto na forma solúvel, a enzima chiquimato desidrogenase (*Mtb*SD), que catalisa o quarto passo na via do chiquimato. Neste trabalho apresentamos, no primeiro artigo, intitulado "*Functional shikimate dehydrogenase from Mycobacterium tuberculosis H37Rv: purification and characterization*", a purificação da *Mtb*SD solúvel e ativa por cromatografia líquida, o seqüenciamento do N-terminal, a espectrometria de massas, a determinação do estado oligomérico por cromatografia em gel filtração, a estabilidade térmica, a determinação de parâmetros cinéticos aparentes em estado estacionário para as reações direta e reversa, a constante de equilíbrio aparente e energia de ativação para a reação química catalisada pela *Mtb*SD. No segundo artigo, intitulado "*Shikimate Dehydrogenase from Mycobacterium tuberculosis H37Rv: Kinetic and Chemical Mechanisms*", descrevemos os parâmetros de velocidade inicial na reação direta, os estudos de inibição pelos produtos, os efeitos isotópicos cinéticos primários do deutério, os efeitos isotópicos cinéticos do solvente, os efeitos isotópicos cinéticos múltiplos, *proton inventory*, o efeito de pH na química ácido/base para a ligação e catálise dos substratos, e a estrutura 3D da *Mtb*SD obtida *in silicon* pela modelagem por homologia. Esses resultados servem como base para os estudos cinéticos e estruturais que podem auxiliar no desenho racional de inibidores a serem testados como agentes antimicobacterianos.

Abstract

The increasing prevalence of tuberculosis (TB), the emergence of multidrug-resistant strains of *Mycobacterium tuberculosis*, the causative agent of TB, and the devastating effect of co-infection with HIV have highlighted the urgent need for the development of new antimycobacterial agents. Analysis of the complete genome sequence of *M. tuberculosis* H37Rv shows the presence of genes involved in the aromatic amino acid biosynthetic pathway, the shikimate pathway, and experimental evidence showed that this pathway is essential for *M. tuberculosis* and it is absent in humans. The gene products that are essential for the growth of the microorganisms make them attractive drug targets since inhibiting their function may kill the bacilli. Our group has previously reported the cloning of *M.* tuberculosis *aro*E gene and the expression of the product in the soluble form, the shikimate dehydrogenase (*Mtb*SD) enzyme, that catalysis the fourth reaction in the shikimate pathway. Currently, in the first manuscript "*Functional shikimate dehydrogenase from Mycobacterium tuberculosis H37Rv: purification and characterization*", we present the purification of soluble and active *Mtb*SD, N-terminal sequencing, mass spectrometry, assessment of the oligomeric state by gel filtration chromatography, thermal stability, determination of apparent steady-state kinetic parameters for both forward and reverse directions, apparent equilibrium constant, and energy of activation for the enzyme-catalyzed chemical reaction. In the second manuscript "*Shikimate Dehydrogenase from Mycobacterium tuberculosis H37Rv: Kinetic and Chemical Mechanisms*", we describe the kinetic mechanism, initial velocity patterns in the forward direction, product inhibition studies, primary deuterium kinetic isotope effects, solvent kinetic isotope effects, multiple isotope effects, proton inventory, pH rate profile and the *Mtb*SD 3D structure obtained in silicon by homology modeling. These results should be useful as a solid base for structural and kinetic studies, which can aid in the rational design of inhibitors to be tested as antimycobacterial agents.

Lista de Abreviaturas

DAHP	**3-desoxi-D-arabinoheptulosonato-7-fosfato**
DHS	**3-desidrochiquimato**
***Ecoli*QSD**	**quinato-chiquimato desidrogenase de *Escherichia coli***
***Ecoli*SD**	**chiquimato desidrogenase de *Escherichia coli***
EPSP	**5-enolpiruvilchiquimato-3-fosfato**
E4P	**eritrose-4-fosfato**
***Hinfl*SD**	**chiquimato desidrogenase de *Haemophilus influenzae***
IPTG	**β-D-tiogalactopiranosídeo**
***Mjan*SD**	**chiquimato desidrogenase de *Methanacoccus jannaschii***
***Mtb*SD**	**chiquimato desidrogenase de *Mycobacterium tuberculosis***
PEP	**Fosfoenolpiruvato**
***Psat*SD**	**chiquimato desidrogenase de *Pisum sativum***
SD	**chiquimato desidrogenase**
SHK	**chiquimato ou ácido chiquímico**
TB	**tuberculose**

Introdução

Tuberculose

A tuberculose (TB) é a doença infecto-contagiosa responsável pela maior parte da mortalidade humana causada por um único agente infeccioso, o *Mycobacterium tuberculosis* (ENARSON & MURRAY, 1996).

No século XIX, a TB já era uma doença devastadora, com taxa de transmissão e índice de morbidade altíssimos; período que corresponde ao começo da revolução industrial, que trouxe como conseqüência a formação de aglomerados urbanos com péssimas condições de higiene e habitação (DORMANDY, 2000).

A infecção ocorre por meio da inalação do bacilo, presente em partículas infectivas, expelidas por pacientes com tuberculose ativa através da tosse ou espirro. Elas apresentam dimensão suficiente para permanecerem suspensas no ar de ambientes fechados e assim alcançar os alvéolos pulmonares de outros indivíduos (DUNLAP e cols., 2000; GLICKMAN & JACOBS, 2001).

Após a infecção, a doença é regulada fundamentalmente pelo sistema imune do hospedeiro, que determina se o microrganismo é eliminado, condicionado a um estado de latência, ou se está propenso a desenvolver a doença ativa no seu hospedeiro.

Assim que a micobactéria chega aos bronquíolos e alvéolos, os macrófagos fagocitam o invasor. A partir daí, a infecção irá ou não se estabelecer dependendo da virulência bacteriana e da capacidade bactericida dos macrófagos do hospedeiro. Caso a micobactéria não seja eliminada pela primeira linha de defesa do organismo, a erosão da bactéria no interstício pulmonar estabelecerá uma inflamação local; logo, o dano tissular, a formação dos granulomas e a agregação destes formarão o tubérculo -

a característica nodular da lesão de TB (MILBURN, 2001). Aproximadamente 4 meses após a infecção, ocorre a eliminação da maior parte dos bacilos e cessa a infecção primária.

A tuberculose conhecida como tuberculose extra-pulmonar ou miliar, representa 15% dos pacientes com a infecção ativa. Ela é causada pelo crescimento bacteriano excessivo no granuloma, que atinge então a corrente sangüínea e disseminando-se para a pleura, linfonodos, fígado, baço, ossos e articulações, coração, cérebro, sistema genito-urinário, meninges, peritônio ou na pele.

Uma outra possibilidade é a coexistência pacífica entre invasor e hospedeiro humano, que recebe a denominação de infecção latente. Nesse caso se estabelece um reservatório bacteriano no indivíduo infectado, onde o metabolismo bacteriano está reduzido como conseqüência da ação do sistema imunológico do hospedeiro. No momento em que o sistema imune do hospedeiro apresentar falhas, como subnutrição, câncer, diabetes, quimioterapia, entre outros, pode ocorrer a reativação do bacilo, levando o indivíduo a desenvolver TB ativa (PARRISH e col., 1998).

O *M. tuberculosis* adapta-se a uma variedade de condições ambientais, dentro e fora do hospedeiro, sendo então considerado o patógeno de maior sucesso da história. Esta característica peculiar do bacilo faz com que a TB permaneça como uma das principais causas de morte no mundo (ENSERINK, 2001). A exata forma como o microrganismo permanece silencioso no corpo humano, em trégua com o sistema imune, tem sido um longo mistério (WICKELGREN, 2000).

A TB começou a ser esclarecida no meio científico em 1882, quando no IV Congresso Mundial de TB, o médico e bacteriologista alemão Robert Koch tornou

pública a identificação do agente etiológico da doença. Este marco da história científica tornou possível a procura de terapias eficazes no tratamento de pacientes com TB.

A descoberta da estreptomicina, em 1944, marcou a era de ouro no desenvolvimento de drogas para TB. Outras drogas, ainda hoje utilizadas para o tratamento da doença, foram sendo introduzidas: o ácido p-aminosalicílico (1946), isoniazida (1952), ciclosserina (1955), canamicina (1957), rifampicina (1965), etionamida (1966), etambutol (1968) e pirazinamida (1970) (DUNCAN, 2003). A administração de estreptomicina, isoniazida e ácido para-amino-salicílico contra a doença ativa reduziu consideravelmente a mortalidade por TB (BLOOM & MURRAY, 1992; DANIEL, 1997). Essas drogas e as vacinas de prevenção da infecção trouxeram um período de relativa tranqüilidade com respeito à doença.

Em 1993, a TB foi declarada “emergência de saúde global” pela Organização Mundial da Saúde (OMS). Estima-se que anualmente ocorram 8 milhões de novos casos e 3 milhões de mortes devido a esta doença (DUNLAP e cols., 2000). Em torno de um terço da população mundial está hoje infectada pelo *M. tuberculosis* e 10% deste número desenvolverá a doença ao longo da vida (ENSERINK, 2001). Além disso, a OMS e o Centro de Controle e Prevenção de Doenças dos Estados Unidos indicam a TB como responsável por 27% de todas as mortes evitáveis em todo o mundo (SEPKOWITZ e cols., 1995). Estima-se que até 2020, um bilhão de pessoas se tornarão infectadas, 200 milhões de pessoas desenvolverão a doença e 70 milhões morrerão devido à TB se a atual tendência permanecer inalterada (PASQUALOTO & FERREIRA, 2001).

A argumentação de que a tuberculose é uma doença reemergente é válida apenas para alguns países europeus e para os Estados Unidos; pois no Brasil, a TB não é um problema reemergente e sim um problema persistente (RUFFINO-NETTO, 2002).

Os fatores responsáveis pelo ressurgimento da TB, até então considerada vencida em países desenvolvidos são: (1) advento da Síndrome da Imunodeficiência Adquirida (AIDS); (2) deterioração dos programas de saúde pública visando o controle da TB; (3) aumento da transmissão do *M. tuberculosis* em hospitais e prisões; e (4) surgimento de cepas resistentes a múltiplas drogas (MDR, resistentes a pelo menos isoniazida e rifampicina) (ESPINAL, 2003).

A TB e a AIDS são referenciadas como co-epidemia, visto que nesta situação a probabilidade de desenvolvimento da doença aumenta até 30 vezes. A taxa de reativação de TB latente em pacientes HIV-positivos é de 2,3 a 13,3%, dependendo do nível de imunossupressão, contra 0,2% em pessoas imunocompetentes (KNIGGE e cols., 2000). Pacientes HIV-positivos infectados com *M. tuberculosis* progridem para TB ativa em uma taxa de 37% nos primeiros 6 meses, ao contrário de pacientes imunocompetentes cuja taxa é de 2 a 5% nos primeiros dois anos (SEPKOWITZ, 1995). Esse aumento considerável na probabilidade leva a uma aceleração da transmissão do agente etiológico (KNIGGE e cols., 2000) e, portanto, a co-infecção TB - HIV representa um problema de efeito devastador tanto para pacientes infectados como para a população em geral.

O aumento da transmissão do *M. tuberculosis* em hospitais e prisões ocorre devido aos aglomerados humanos com condições de higiene e habitação péssimas.

Um agravante deste quadro decorre da deterioração dos sistemas de saúde visando o controle da TB (FÄTKENHEUER e col., 1999).

A TB resistente a múltiplas drogas (MDR-TB – "*multidrug-resistant tuberculosis"*) está relacionada à resistência aos medicamentos para combater a doença, visto que ocorre a resistência à pelo menos rifampicina e isoniazida. Em 2000 foi estimado que 3,2% dos novos casos são de MDR-TB (Espinal, 2003). Segundo a OMS, ocorrem 300.000 novos casos de MDR-TB anualmente, onde 79% são de "super cepas", que se refere às cepas resistentes a pelo menos três das quatro principais drogas usadas no tratamento de TB (WHO, 2004).

Através de um estudo realizado em 35 países pela OMS e a União Internacional Contra a Tuberculose e Doenças Pulmonares (IUATLD – *"International Union Against Tuberculosis and Lung Disease"*) foi demonstrado que a doença consiste em um problema global, pois todas as regiões analisadas apresentaram linhagens de *M. tuberculosis* com resistência à pelo menos uma das drogas antituberculose, sendo normalmente a isoniazida ou estreptomicina (PABLOS-MENDEZ e col., 1998).

O MDR-TB surge sob pressão seletiva da quimioterapia inapropriada ou da não adesão ao tratamento (RILEY, 1993). Terapia inadequada compreende o uso de uma única droga, combinações inadequadas de drogas, má-absorção de drogas administradas, etc. Nesses casos o *M. tuberculosis* estará exposto a concentrações sub-letais de compostos bactericidas, quando as cepas mais resistentes na população bacteriana sobreviverão e superarão a cepa sensível (PETRINI & HOFFNER, 1999). No caso da co-epidemia de HIV-TB, o aparecimento das linhagens MDR-TB foi favorecido pelo maior índice de abandono do tratamento (BRENNAN, 1997).

Atualmente o tratamento recomendado pela OMS consiste na administração combinada dos antibióticos isoniazida, rifampicina, pirazinamida e estreptomicina (ou etambutol) durante 2 meses, seguida pela combinação de isoniazida e rifampicina por pelo menos mais 4 meses.

Mesmo com um tratamento bem estabelecido, a eficácia do tratamento contra o bacilo de Koch através da utilização de antibióticos tem sido dificultada, em parte, principalmente pela oposição do paciente em cumprir com o tratamento de 6 meses (YOUNG, 1998). A longa duração e os desagradáveis efeitos colaterais do tratamento levaram a OMS ao desenvolvimento de um programa conhecido como DOTS (*directly observed treatment short-course*) que visa acompanhar a adesão dos pacientes ao tratamento, monitorar seu progresso e observar a ingestão de cada dose da medicação (*NSB* Editorial comment, 2000).

O programa DOTS combina cinco elementos: compromisso político, serviços de microscopia, suprimento de medicação, sistemas de monitoramento e observação direta do tratamento (PASQUALOTO & FERREIRA, 2001). Entretanto, em alguns países super populosos o programa não consegue alcançar 100% dos indivíduos infectados, não chegando aos objetivos primeiramente propostos. Uma alternativa é a obtenção de drogas que simplifiquem o tratamento da TB facilitando a execução do programa DOTS (DUNCAN, 2003).

Desenvolvimento de Agentes anti-TB

Frente a essa exposição de fatos que envolvem a TB, faz-se necessária à introdução de novas drogas anti-TB, pois desde a metade da década de 70 não houve a descoberta, nem tampouco o desenvolvimento de novas drogas de primeira linha que

possam ser utilizadas no combate ao bacilo causador da TB (PETRINI & HOFFNER, 1999).

As novas drogas teriam como objetivo a efetividade no tratamento de MDR-TB e TB latente, além da redução no tempo do tratamento (O´BRIEN & NUNN, 2001). Isto representaria vantagem significativa em relação às drogas atualmente utilizadas. DUNCAN (2003) estima que o desenvolvimento completo de uma nova droga anti-TB custe de 100 a 800 milhões de dólares. Considerando-se que 95% dos novos casos de TB ocorram em países em desenvolvimento, poucos pacientes poderiam pagar os altos custos necessários para o lucro da indústria farmacêutica (O´BRIEN & NUNN, 2001). O provável retorno financeiro é baixo e, conseqüentemente, o interesse neste setor também é baixo (DUNCAN, 2003).

Mais de um século após a descoberta de Robert Koch, está disponível uma nova perspectiva para o desenvolvimento de novas drogas, em decorrência da seqüência completa do genoma do *M. tuberculosis*, marcando uma nova era na batalha contra esse patógeno (COLE e cols., 1998).

Utilizando a análise de homologia de seqüências, foi possível a identificação de genes estruturais envolvidos em rotas metabólicas. A identificação e validação de rotas essenciais ao microrganismo, ausentes no hospedeiro, são o primeiro passo no desenvolvimento de inibidores específicos de baixa toxicidade (DUNCAN, 2003). Alguns genes expressos na infecção latente foram descritos, e suas proteínas podem ser bons alvos para o desenvolvimento de drogas para o tratamento da TB latente (WICKELGREN, 2000).

Uma possibilidade para a obtenção de novas drogas é o desenho racional que pode ser alcançado quando existe informação detalhada sobre a interação entre proteína alvo - ligante, fornecida pela estrutura tridimensional do alvo (DUNCAN, 2003). Com a obtenção da estrutura 3D do alvo é possível fazer uma triagem em biblioteca virtual de compostos químicos que possam ser utilizados como drogas naturais ou que sirvam como base para o desenho racional de novas drogas sintéticas. As informações sobre os mecanismos cinético e químico também são importantes para o desenho racional de inibidores.

Uma alternativa atraente para obter novos inibidores específicos para alvos definidos é a busca desses compostos na biodiversidade brasileira, e para isso é necessária à obtenção de um alvo isolado e funcional. Utilizando a tecnologia de detecção através da Ressonância Plasmônica de Superfície (SPR), imobiliza-se o alvo e após faz-se a triagem de ligantes (possíveis inibidores) presentes em extratos de plantas oriundos da biodiversidade brasileira e/ou seus compostos purificados. Essa tecnologia monitora as interações bimoleculares entre compostos e o alvo de interesse em tempo real, detectando a ligação de compostos ao alvo pelo aumento de massa em relação ao estado inicial. Esse tipo de análise e metodologia é ideal para a descoberta de novas drogas por ter alta sensibilidade, e pelo fato de que os compostos químicos analisados não necessitam de marcação radioativa nem de fluorescência. Em posse dos possíveis inibidores, segue-se para a determinação das constantes de inibição (K_i) para a caracterização da natureza da inibição (BASSO e col., 2005).

Dentre as vias metabólicas identificadas no genoma do *M. tuberculosis* está a via do chiquimato (Figura 1) que representa um alvo validado para o desenvolvimento

de herbicidas e agentes antimicrobianos, pois estão presentes em algas, vegetais, bactérias, fungos (BENTLEY, 1990) e parasitas do filo Apicomplexa (ROBERTS e cols., 1998), mas ausente em vertebrados. Foi sugerido que esta via é essencial para o crescimento do *M. tuberculosis* mesmo na presença de suplementos exógenos (PARISH & STOKER, 2002). Com isso, inibidores específicos das enzimas presentes nessa via apresentam potencial como agentes terapêuticos no tratamento da TB.

A via do chiquimato é composta de uma seqüência de sete reações que vão desde a condensação da *D*-eritrose 4-fosfato (E4P) e do fosfoenolpiruvato (PEP) até a formação do corismato. O corismato leva à biossíntese de precursores de aminoácidos aromáticos, vitaminas E e K, ácido fólico (BENTLEY, 1990), naftoquinonas, menaquinonas e micobactinas (em micobatérias) (Figura 2) (RATLEDGE, 1982).

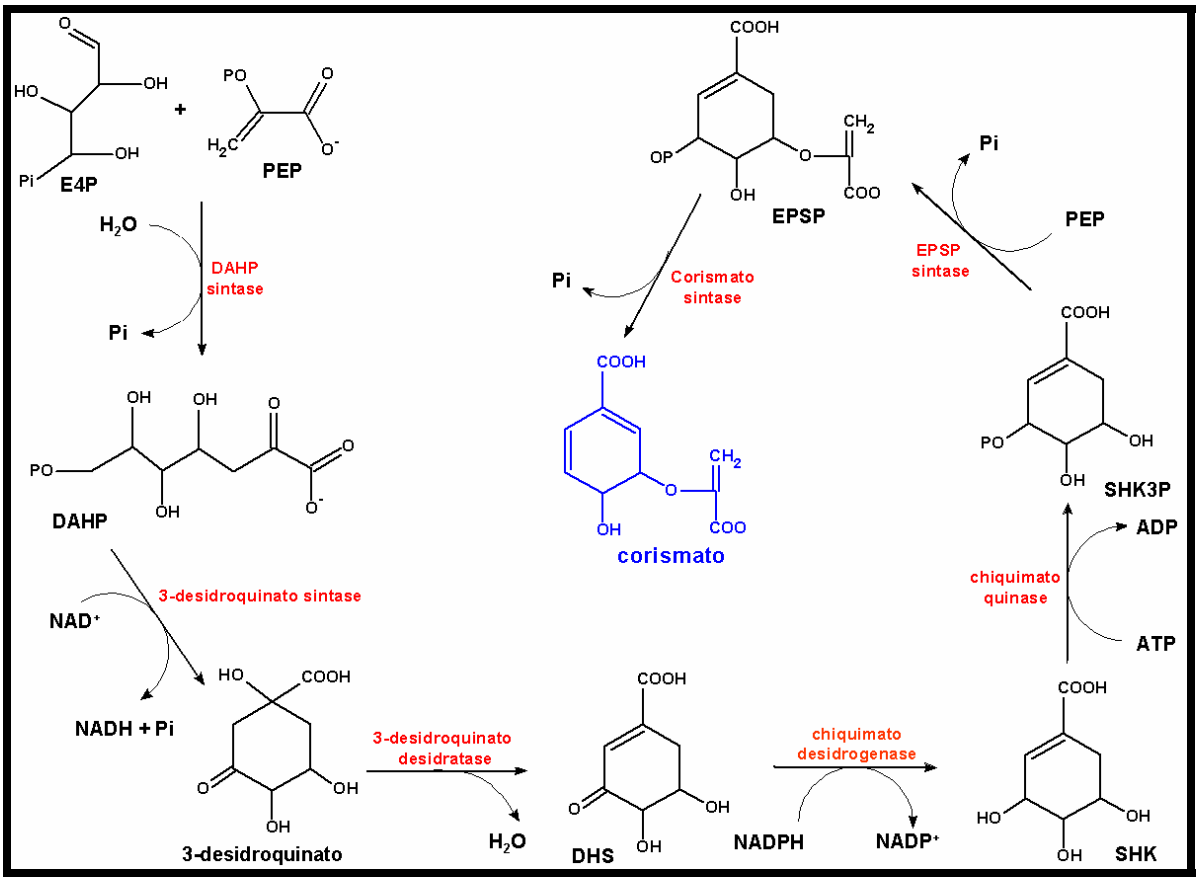


Figura 1: A Via do Chiquimato. Esta via é composta por sete etapas enzimáticas que culminam com a formação do corismato, o precursor chave para formação de compostos aromáticos essenciais aos microrganismos.

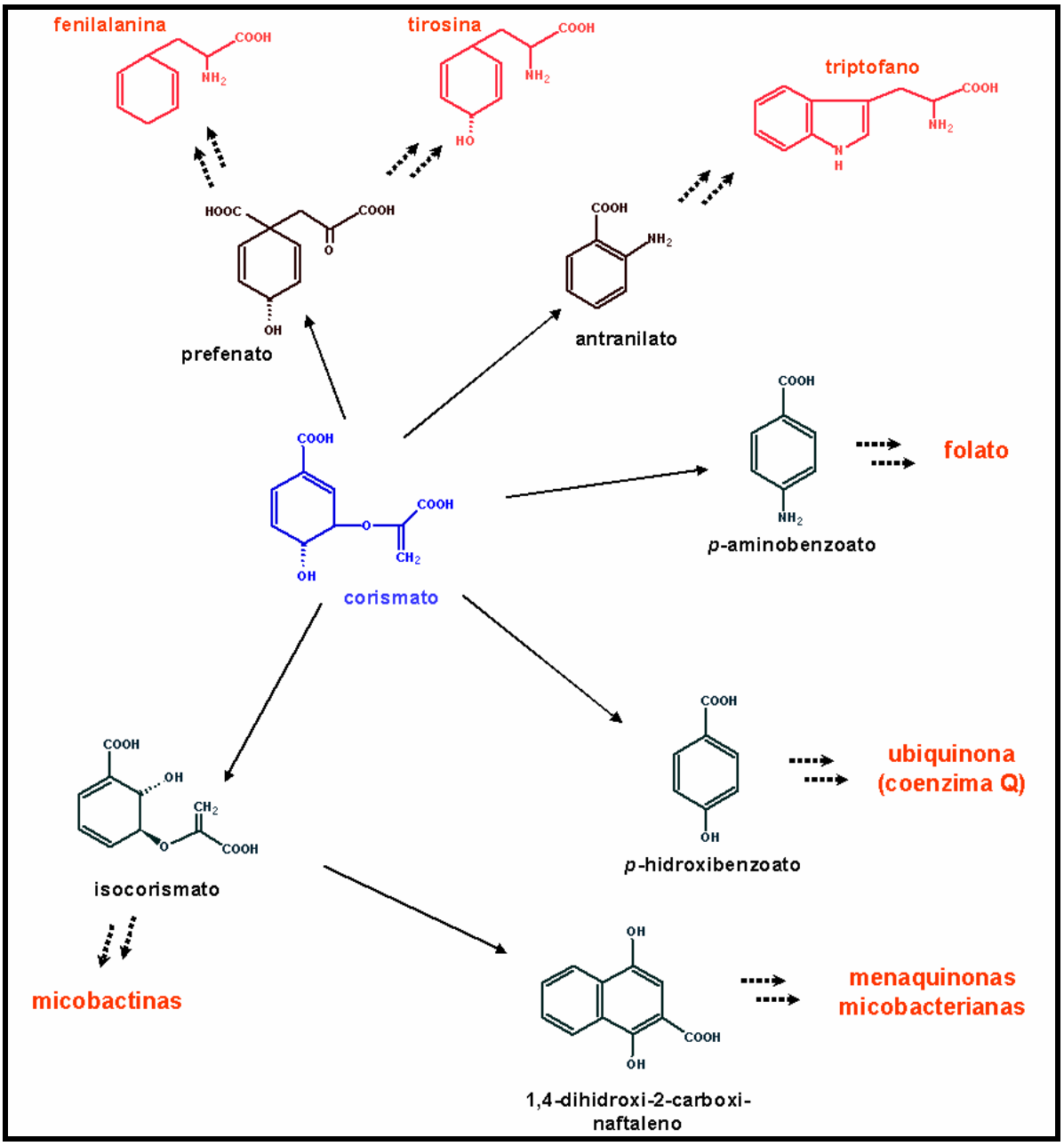


Figura 2: Importância da Via do Chiquimato. O produto da via do chiquimato, corismato e os compostos essenciais sintetizados a partir dele.

O trabalho aqui descrito mostra o estudo sobre os mecanismos cinético e químico feito com um alvo enzimático, a chiquimato desidrogenase de *M. tuberculosis* (*Mtb*SD), presente na quarta etapa de catálise da via do chiquimato (Figura 1). A *Mtb*SD catalisa a redução reversível do 3-desidrochiquimato (DHS) a chiquimato (SHK), utilizando especificamente o cofator NADPH (Figura 3) (BENTLEY, 1990; CHAUDHURI e cols., 1985).

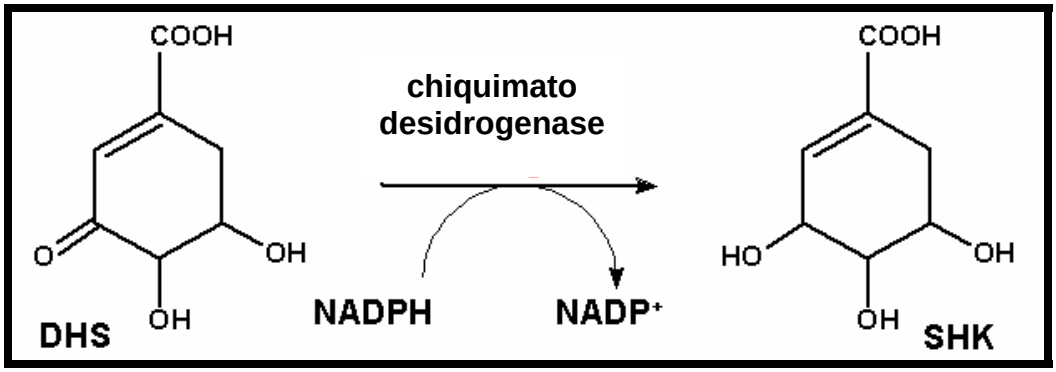


Figura 3: Reação enzimática catalisada pela enzima chiquimato desidrogenase. A redução do 3-desidrochiquimato dependente de NADPH leva à formação do D-chiquimato.

A proteína *Mtb*SD, codificada pelo gene *aro*E (Rv2552c), possui 269 aminoácidos e pertence à superfamília das oxiredutases dependentes de NAD(P)H, que atuam em rotas anabólicas e catabólicas (BENACH e cols., 2003).

Em 2002, a primeira descrição sobre a enzima *Mtb*SD foi publicada (MAGALHÃES e cols.). O artigo trata da clonagem do gene *aro*E da cepa *M. tuberculosis* H37Rv em vetor para expressão de proteínas heterólogas pET23a(+) (Novagen®) e expressão da enzima recombinante em *E. coli* BL21(DE3). O trabalho

relatou que o sistema pET23a(+)-*aro*E em *E. coli* BL21(DE3) é excelente para a expressão do alvo *Mtb*SD recombinante sem adição do indutor IPTG e a obtenção da *Mtb*SD recombinante solúvel e funcional foi feita por meio da utilização do procedimento de congelamento e descongelamento em ciclos repetidos da *E. coli* BL21(DE3) com pET23a(+)-*aro*E para o extravasamento do conteúdo celular (MAGALHÃES e cols.).

A expressão da *Mtb*SD também foi obtida em vetor de expressão pET28b (Novagen®). A purificação da *Mtb*SD foi feita em coluna de níquel por meio de ligação da cauda de histidina da proteína recombinante. Os valores de K_m aparente para os substratos ácido chiquímico e $NADP^+$ foram de 0,030 mM e 0,063 mM, respectivamente, e o k_{cat} foi de 399 s^{-1}. Esses valores foram determinados sem retirar da cauda de histina da *Mtb*SD recombinante, o que pode influenciar os dados obtidos (ZHANG e cols., 2005).

A estrutura da chiquimato desidrogenase codificada pelo gene *aro*E de *E. coli* (*Ecoli*SD) foi determinada com resolução de 1,5 Å. A enzima apresenta uma arquitetura com dois domínios α/β separados por um sulco grande que corresponde ao sítio catalítico da enzima (MICHEL e col., 2003). O sítio ativo da *Ecoli*SD foi identificado pelos resíduos Ser14, Ser16, Lys65, Asn86, Thr101, Asp102 e Gln244, conservados na família das chiquimato desidrogenase. A caracterização cinética aparente da proteína *Ecoli*SD, determinação de K_m e k_{cat}, foi feita utilizando a reação reversa que corresponde à catálise de oxidação do ácido chiquímico. A *Ecoli*SD trata-se de um monômero em solução (CHAUDURI & COGGINS, 1985).

Um estudo semelhante foi feito com a chiquimato desidrogenase codificada pelo gene *aro*E de *Methanococcus jannaschii* (*Mjan*SD). A estrutura foi determinada em 2,35 Å e a arquitetura descrita é semelhante à *Ecoli*SD. A proteína *Mjan*SD apresenta-se na forma dimérica em solução (PADYANA e BURLEY, 2003).

Objetivo

Considerando que a tuberculose é problema de saúde pública global, o quadro alarmante com o advento da AIDS, a inexistência de programas eficazes de saúde publica com relação à TB, o surgimento de cepas resistentes, o interesse na obtenção de novos antimicobacterianos, a via do chiquimato é um alvo promissor, e que inibidores específicos das enzimas presentes nessa via apresentam potencial como agentes terapêuticos no tratamento da TB, **"o presente trabalho visa elucidar os mecanismos cinético e químico da enzima chiquimato desidrogenase de *M. tuberculosis* (*Mtb*SD), que servirão como base para o alcance de um inibidor específico com possível utilização no tratamento da tuberculose."**

Para atingir tal objetivo foram feitas:

- ✓ **Caracterização da *Mtb*SD**

  Determinação dos parâmetros cinéticos aparentes em ambas as direções da reação, da constante de equilíbrio da reação, da estabilidade térmica da enzima recombinante e da energia de ativação para a reação química.

- ✓ **Determinação dos mecanismos cinético e químico para a *Mtb*SD**

  Com o intuito de investigar os mecanismos cinético e químico foram feitos experimentos de velocidade inicial em estado estacionário, de estudos de inibição pelos produtos, de efeitos isotópicos cinéticos primário, de efeitos isotópicos cinéticos de solvente, de efeitos isotópicos cinéticos múltiplos, de "*proton inventory*" e de efeito de pH na química ácido/base na reação.

- ✓ **Estudos estruturais da *Mtb*SD utilizando modelagem por homologia**

  Os estudos estruturais *in silicon* da *Mtb*SD são úteis na interpretação dos resultados obtidos nos experimentos cinéticos e possibilita a identificação dos possíveis resíduos envolvidos na ligação e catálise dos substratos. Para tal foi modelada a estrutura 3D para a *Mtb*SD em *silicon* utilizando como molde a chiquimato desidrogenase de *E. coli*.

PARTE II

Capítulo I

Fonseca, I. O., Magalhães, M. L. B., Oliveira, J. S., Silva, R. G., Mendes, M. A., Palma, M. S., Santos, D. S., Basso, L. A. (2006) "**Functional shikimate dehydrogenase from *Mycobacterium tuberculosis* H37Rv: purification and characterization**" *Protein Expr. Purif.* **46**, 429-437.

Arquivo nomeado:

Functional shikimate dehydrogenase from *Mycobacterium tuberculosis* H37Rv.pdf

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Protein Expression and Purification 46 (2006) 429–437
Protein Expression & Purification
www.elsevier.com/locate/yprep

Functional shikimate dehydrogenase from *Mycobacterium tuberculosis* H37Rv: Purification and characterization

Isabel O. Fonseca [a], Maria L.B. Magalhães [a], Jaim S. Oliveira [a], Rafael G. Silva [a], Maria A. Mendes [b], Mario S. Palma [b], Diógenes S. Santos [a,*], Luiz A. Basso [a,*]

[a] *Centro de Pequisas em Biologia Molecular e Funcional, TECNOPUC, Programa de Pós-Graduação em Ciências Biológicas: Bioquímica, Programa de Pós-Graduação em Biologia Celular e Molecular (PUCRS-UFRGS), Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre—RS 90619-900, Brazil*

[b] *Laboratório de Biologia Estrutural e Zooquímica, Centro de Estudos de Insetos Sociais, Departamento de Biologia, Instituto de Biociências, Universidade Estadual Paulista, Rio Claro—SP 13506-900, Brazil*

Received 18 July 2005, and in revised form 30 September 2005
Available online 27 October 2005

Abstract

Tuberculosis (TB) poses a major worldwide public health problem. The increasing prevalence of TB, the emergence of multi-drug-resistant strains of *Mycobacterium tuberculosis*, the causative agent of TB, and the devastating effect of co-infection with HIV have highlighted the urgent need for the development of new antimycobacterial agents. Analysis of the complete genome sequence of *M. tuberculosis* shows the presence of genes involved in the aromatic amino acid biosynthetic pathway. Experimental evidence that this pathway is essential for *M. tuberculosis* has been reported. The genes and pathways that are essential for the growth of the microorganisms make them attractive drug targets since inhibiting their function may kill the bacilli. We have previously cloned and expressed in the soluble form the fourth shikimate pathway enzyme of the *M. tuberculosis*, the *aroE*-encoded shikimate dehydrogenase (mtSD). Here, we present the purification of active recombinant *aroE*-encoded *M. tuberculosis* shikimate dehydrogenase (mtSD) to homogeneity, N-terminal sequencing, mass spectrometry, assessment of the oligomeric state by gel filtration chromatography, determination of apparent steady-state kinetic parameters for both the forward and reverse directions, apparent equilibrium constant, thermal stability, and energy of activation for the enzyme-catalyzed chemical reaction. These results pave the way for structural and kinetic studies, which should aid in the rational design of mtSD inhibitors to be tested as antimycobacterial agents.

Keywords: *Mycobacterium tuberculosis*; Shikimate pathway; Shikimate dehydrogenase; *aroE*; Enzyme kinetics; Protein purification; Drug target

The causative agent of tuberculosis (TB),[1] *Mycobacterium tuberculosis*, infects approximately 32% of the world's human population. TB remains the leading cause of mortality due to a bacterial pathogen. Currently, there are 8 million new cases and 2 million deaths annually from tuberculosis, and it is predicted that a total of 225 million new cases and 79 million deaths will occur between 1998 and 2030 [1]. Approximately 2 billion individuals are believed to harbor latent TB based on tuberculin skin test surveys [2], which represents a considerable reservoir of bacilli. Possible factors underlying the resurgence of TB worldwide include the HIV epidemic, increase in the homeless population, and decline in health care structures and national surveillance [3]. The pandemic of AIDS has had

* Corresponding authors. Fax: +55 51 33203629.
E-mail addresses: diogenes@pucrs.br (D.S. Santos), luiz.basso@pucrs.br (L.A. Basso).

[1] *Abbreviations used:* DHS, 3-dehydroshikimate; ESI-MS, electrospray ionization mass spectrometry; IPTG, isopropyl β-D-thiogalactopyranoside; LB, Luria–Bertani; MDR-TB, multi-drug-resistant tuberculosis; mtSD, *Mycobacterium tuberculosis* shikimate dehydrogenase; $NADP^+$, oxidized β-nicotinamide adenine dinucleotide phosphate; NADPH, reduced β-nicotinamide adenine dinucleotide phosphate; SHK, D-shikimate; SDS–PAGE, sodium dodecylsulfate–polyacrylamide gel electrophoresis; TB, tuberculosis.

1046-5928/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.pep.2005.10.004

a major impact on the TB problem, owing not only to increased reactivation of latent TB but also to acceleration of transmission in HIV/TB co-infected patients following the increase in the number of smear-positive infectious pulmonary TB cases [4]. Another contributing factor is the evolution of multi-drug TB (MDR-TB), defined as resistant to isoniazid and rifampicin, which are the most effective first-line drugs [5]. MDR-TB is more difficult and more expensive to treat, and more likely to be fatal [6]. According to the 2004 Global TB Control Report of the World Health Organization, there are 300,000 new cases per year of MDR-TB worldwide, and 79% of MDR-TB cases are now "super-strains," resistant to at least three of the four main drugs used to treat TB [7]. The factors that most influence the emergence of MDR-TB strains include inappropriate treatment regimens, and patient non-compliance in completing the prescribed courses of therapy due to the lengthy standard "short-course" treatment or when the side effects become unbearable [8]. Thus, there is a need for the development of new antimycobacterial agents to both treat *M. tuberculosis* strains resistant to existing drugs and shorten the duration of short-course treatment to improve patient compliance [9].

The shikimate pathway is an attractive target for the development of herbicides and antimicrobial agents because it is essential in algae, higher plants, bacteria, and fungi, but absent from mammals [10]. The mycobacterial shikimate pathway leads to the biosynthesis of chorismic acid, which is a precursor of aromatic amino acids, naphthoquinones, menaquinones, and mycobactins [11]. The salicylate-derived mycobactin siderophores have been shown to be essential for *M. tuberculosis* growth in macrophages [12]. In addition, the shikimate pathway has more recently been shown to be essential for the viability of *M. tuberculosis* [13]. Accordingly, the essentiality of mycobacterial shikimate pathway and its absence from human host indicate that any enzyme of this pathway represents a promising target for the development of non-toxic antimycobacterial agents.

Analysis of the complete genome sequence of *M. tuberculosis* shows the presence of seven *aro* genes predicted to be involved in the shikimate pathway [14]. Amongst them, the *aroE*-encoded shikimate dehydrogenase (SD) has been predicted by DNA sequence homology to be present in *M. tuberculosis* H37Rv strain. Shikimate dehydrogenase (EC 1.1.1.25) catalyzes the fourth reaction in the shikimate pathway. We have previously reported the cloning and expression of *M. tuberculosis* SD (mtSD) [15]. In addition, measurements of the NADPH-dependent reduction of 3-dehydroshikimate to shikimate catalyzed by mtSD confirmed the correct assignment to the structural gene encoding SD in *M. tuberculosis* [15]. Here, we report the purification to homogeneity of recombinant and functional mtSD. The purification protocol yielded approximately 11 mg of homogeneous recombinant mtSD from 14 L of *Escherichia coli* cell culture. We also present N-terminal amino acid sequencing and electrospray ionization mass spectrometry (ESI-MS) data that unambiguously demonstrate the identity and purity of homogenous recombinant mtSD protein. The estimated molecular mass of native homogeneous recombinant protein determined by gel filtration indicates that mtSD enzyme is a dimer in solution with a subunit molecular mass value of 27,207 Da determined by ESI-MS. The apparent kinetic parameters for mtSD were determined for all substrates in both forward and reverse reactions. The mtSD thermal stability was evaluated, and an estimate for the activation energy (E_a) was obtained from a linear plot of $\log k$ versus $1/T$ (K^{-1}). A comparison of polypeptide sequences of SDs from *M. tuberculosis*, *E. coli*, *Haemophilus influenzae*, and *Methanococcus jannaschii* allowed identification of amino acid residues that are likely to be involved in 3-dehydroshikimate/shikimate binding.

The results presented here will pave the way for structural and functional efforts currently underway in our laboratory. It is hoped that these studies will provide a framework on which to base the design of new agents with antitubercular activity and low toxicity to humans.

Materials and methods

Overexpression and release of mtSD

The recombinant plasmid pET23a(+)::*aroE* was transformed into *E. coli* BL21 (DE3) host cells by electroporation, and selected on LB agar plates containing 50 μg mL^{-1} carbenicillin. Single colonies were used to inoculate 14 L of LB medium containing 50 μg mL^{-1} carbenicillin with no addition of isopropyl β-D-thiogalactopyranoside (IPTG), and grown for 24 h at 37 °C at 180 rpm as described elsewhere [15]. Cells (49 g) were harvested by centrifugation at 14,900*g*, for 30 min at 4 °C, and stored at −20 °C. Cells expressing recombinant mtSD were placed into metal containers to allow fast temperature equilibrium to be reached, which is necessary for increased efficiency of cell disruption by the freeze–thaw method [15]. Cell paste was placed into a dry-ice/ethanol bath for 2 min and immediately transferred to an ice-water bath for no longer than 8 min; this cycle was repeated 10 times. The cells were dissolved in 196 mL of 50 mM Tris–HCl, pH 7.8 (buffer A). After incubating the mixture for 30 min on ice, cell debris was removed by centrifugation (48,000*g* for 1 h) and the supernatant containing soluble mtSD was collected.

Purification of recombinant mtSD

All steps of the purification protocol of recombinant mtSD were performed on ice or at 4 °C. The supernatant containing soluble mtSD was incubated with 1% (w/v) of streptomycin sulfate for 30 min and centrifuged at 48,000*g* for 30 min. The supernatant was dialyzed twice against buffer A, using a dialysis tubing with molecular weight exclusion limit of 6000–8000 Da. This sample was

clarified by centrifugation (48,000*g* for 30 min) and loaded on a Q-Sepharose fast flow (26 cm × 9.5 cm) column (Amersham Biosciences) pre-equilibrated with the same buffer. The column was washed with 5 column volumes of buffer A and the absorbed material was eluted with a linear gradient (0–100%) of 20 column volumes of 50 mM Tris–HCl, pH 7.8, 0.5 M NaCl (buffer B) at 1 mL min^{-1}. The fractions containing mtSD were pooled (55 mL) and ammonium sulfate was added to a final concentration of 1 M, and clarified by centrifugation (48,000*g* for 30 min). The supernatant was loaded on a Phenyl-Sepharose High Performance (Amersham Biosciences) column pre-equilibrated with 50 mM Tris–HCl, pH 7.8, 1 M $(NH_4)_2SO_4$ (buffer C). The column was washed with 5 column volumes of buffer C and the bound proteins were eluted with a 20-column volume linear gradient (0–100%) of buffer A at 1 mL min^{-1}. The mtSD-containing fractions were pooled (43 mL), concentrated to less than 4 mL using an Amicon ultrafiltration cell (MWCO 10,000 Da), and loaded on a Sephacryl S-200 (26 cm × 60 cm) (Amersham Biosciences) column pre-equilibrated with buffer A. The recombinant mtSD protein was eluted in a total volume of 29 mL at a flow rate of 0.5 mL min^{-1}, and loaded on an anion-exchange Mono-Q column (Amersham Biosciences) pre-equilibrated with buffer A. The column was washed with 5 column volumes of buffer A and the absorbed material was eluted with a 20-column volume linear gradient (0–100%) of buffer B at 1 mL min^{-1}. Elution profiles were followed at 280 and 215 nm. Homogeneous mtSD was eluted in a total volume of 12.5 mL and stored at −20 °C. Protein purification was monitored by SDS–PAGE [16], and the protein concentration was determined by the method of Bradford et al. [17] using the Bio-Rad protein assay kit (Bio-Rad) and bovine serum albumin as standard.

Enzyme activity assay of mtSD

The screening of fractions of the purification protocol containing shikimate dehydrogenase was performed by assaying enzyme activity in the reverse direction in 100 mM Tris–HCl, pH 9.0, at 25 °C. The 500 μL assay mixture contained 2 mM $NADP^+$ and 4 mM D-shikimate (SHK) [18], and the reaction was initiated with addition of 1 μL of the chromatographic fractions. Measurement of the $NADP^+$-dependent oxidation of SHK to form NADPH and 3-dehydroshikimate (DHS) catalyzed by mtSD was continuously monitored by the increase in absorbance at 340 nm ($\varepsilon_{NADPH} = 6.18 \times 10^3\ M^{-1}\ cm^{-1}$). One unit of enzyme activity (U) is defined as the amount of enzyme catalyzing the conversion of 1 μmol of $NADP^+$ per minute at 25 °C.

Determination of apparent kinetic parameters

Determination of the apparent steady-state kinetics parameters, V_{max} and K_m, for DHS and NADPH in the forward reaction, was carried out at varying concentration of one substrate (5, 10, 20, 30, 50, 100, and 200 μM) while the other was maintained at constant saturation level. The reverse reaction was performed in the same conditions of pH (100 mM Tris–HCl buffer, pH 7.0), temperature (25 °C) and range of concentrations of substrates. The reaction was initiated with addition of 6 pmol of homogeneous *M. tuberculosis* SD enzyme and was monitored for 1 min. The kinetic data were analyzed by double reciprocal plots.

Shikimate dehydrogenase activity measurements were based on decreasing concentration of NADPH upon DHS reduction or on increasing concentration of NADPH upon SHK oxidation. The reaction catalyzed by mtSD was continuously monitored by measuring change in absorbance at 340 nm ($\varepsilon_{NADPH} = 6.18 \times 10^3\ M^{-1}\ cm^{-1}$).

Determination of the energy of activation and thermo stability of recombinant mtSD

The energy of activation (E_a) was estimated for recombinant mtDS by the following Arrhenius equation: $k = Ae^{-E_a/RT}$, where k is the rate constant of the reaction at temperature T (in Kelvin), A is a pre-exponential factor (related to collision frequency and a steric factor); E_a is the activation energy; R is universal gas constant (8.3145 J K^{-1} mol^{-1}) and $e^{-E_a/RT}$ is the fraction of molecular collisions that have energy equal to or greater than the measured energy of activation (E_a) of the system at a particular temperature T. The E_a value was calculated from the slope of the linear plot $\log k$ versus $1/T$ (K^{-1}) fitting the data to the following equation: $\log k = (E_a/2.3RT) + \log A$. Measurements of mtSD enzyme activity were in triplicates and performed in 100 mM Tris–HCl buffer, pH 7.0, in the presence of saturating concentrations of SHK (200 μM) and $NADP^+$ (100 μM) at the following temperatures: 15, 20, 25, 30, and 37 °C.

For thermal stability determination, mtSD was incubated at 15, 25, 37, and 55 °C and the remaining enzyme activity was measured at different times of incubation up to 1 h, monitoring the mtSD reverse reaction in an assay mixture containing saturating concentrations of substrates (200 μM SHK and 100 μM $NADP^+$) in 100 mM Tris–HCl, pH 7.0, at 25 °C.

Mass spectrometry analysis

The homogeneity of recombinant protein preparation was assessed by mass spectrometry (MS), employing some adaptations made to the system described by Chassaigne and Lobinski [19]. Samples were analyzed on a triple quadrupole mass spectrometer, model QUATTRO II, equipped with a standard electrospray (ESI) probe (Micromass, Altrinchan), adjusted to ca. 250 μL min^{-1}. The source of temperature (80 °C) and needle voltage (3.6 kV) were maintained constant throughout the experimental data collection, applying a drying gas flow (nitrogen) of 200 L h^{-1} and a nebulizer gas flow of 20 L h^{-1}. The mass

spectrometer was calibrated with intact horse heart myoglobin and its typical cone-voltage induced fragments. The subunit molecular mass of *M. tuberculosis* SD was determined by ESI-MS, adjusting the mass spectrometer to give a peak width at half-height of 1 mass unit, and the cone sample to skimmer lens voltage controlling the ion transfer to mass analyzer was set to 38 V. About 50 pmol sample was injected into electrospray transport solvent. The ESI spectrum was obtained in the multi-channel acquisition mode, scanning from *m/z* 500 to 2000 at scan time of 7 s. The mass spectrometer is equipped with MassLynx and Transform softwares for data acquisition and spectra handling.

N-terminal amino acid sequencing

The N-terminal amino acid residues of purified recombinant mtSD were identified by automated Edman degradation sequencing using a PPSQ 21A gas-phase sequencer (Shimadzu).

Determination of native mtSD molecular mass

The molecular mass of native mtSD homogenous protein was estimated by gel-permeation chromatography on a Superdex 200 HR column (1.0 cm × 30 cm) (Amersham Biosciences). The column was eluted with 50 mM Tris–HCl containing 0.2 M NaCl, pH 7.8, at a flow rate of 0.4 mL min^{-1}. The eluate was monitored at 215 and 280 nm and the column was calibrated with the following protein standards (Amersham Biosciences): ribonuclease A (13,700 Da) from bovine pancreas, chymotrypsinogen (25,000 Da) from bovine pancreas, ovalbumin (43,000 Da) from hen egg, and albumin (67,000 Da) from bovine serum. Blue Dextran 2000 was used to determine the void volume (V_0). The K_{av} value was calculated for each protein using the equation $(V_e - V_0)/(V_t - V_0)$, where V_e is the elution volume for the protein and V_t is the total bed volume, and K_{av} was plotted against the logarithm of standard molecular weights.

Results and discussion

Expression in *E. coli* BL21 (DE3) of recombinant *M. tuberculosis* shikimate dehydrogenase (mtSD) and the method of disruption of transformed host cells were as described elsewhere [15]. Recombinant mtSD was purified as described under Materials and methods and the samples of each chromatographic step were analyzed by SDS–PAGE with Coomassie blue staining and assayed for enzyme activity in the reverse direction, following the increase in absorbance at 340 nm due to the $NADP^+$-dependent oxidation of D-shikimate to form NADPH and 3-dehydroshikimate. The recombinant protein was purified 8.5-fold (Table 1) to electrophoretic homogeneity (Fig. 1). The relative mobility of the polypeptide chain in SDS–PAGE indicates a homogeneous protein with a subunit molecular mass value of approximately 27 kDa (Fig. 1). Even though a large amount of cells were needed for recovery of homogeneous target protein in quantities necessary for kinetic and structural studies, the freeze–thaw method [20] was previously shown to be, amongst a number of experimental protocols tested to reduce insoluble protein production, the method of choice to obtain soluble mtSD in its active form [15]. Approximately 11 mg of homogeneous recombinant mtSD could be obtained from 49 g of *E. coli* BL21 (DE3) host cells following the purification protocol presented in Table 1, which required four chromatographic steps to obtain homogeneous mtSD. The recombinant protein eluted with approximately 50% of 50 mM Tris–HCl, pH 7.8, containing 0.5 M NaCl (buffer B), consistent with the theoretical value of 5.11 for the mtSD isoelectric point. Since the theoretical isoelectric point values for mtSD and *E. coli* SD (5.12) are quite similar, the anion-exchange chromatographic step using Q-Sepharose Fast Flow resin is unlikely to have separated these

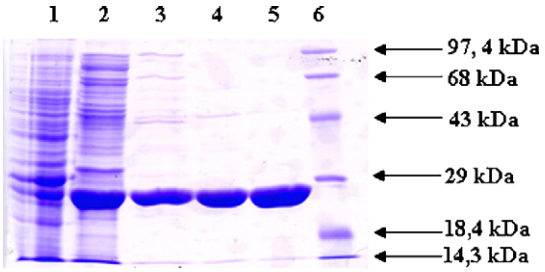


Fig. 1. SDS–PAGE (12.5%) analysis of pooled fractions from the various steps of the purification protocol of mtSD. Lane 1, crude extract (50 μg); lane 2, Q-Sepharose fast flow ion exchange (30 μg); lane 3, Phenyl-Sepharose hydrophobic interaction (10 μg); lane 4, Sephacryl S-200 gel permeation (10 μg); lane 5, Mono-Q ion exchange (17 μg); lane 6, MW marker *High Range* (Gibco).

Table 1
Purification of *M. tuberculosis* shikimate dehydrogenase from *E. coli* BL21 (DE3) [pET23a(+)::*aroE*] cells

Purification step	Total protein (mg)	Total activity (U)	Specific activity[a] (U mg^{-1})	Purification (fold)	Yield (%)
Crude extract	1040.4	475.3	0.46	1.0	100.0
Q-Sepharose	82.5	110.0	1.33	2.9	23.1
Phenyl-Sepharose	24.1	69.2	2.87	6.2	14.6
Sephacryl S-200	17.4	65.3	3.75	8.2	13.7
Mono-Q	10.8	42.1	3.90	8.5	8.9

The results presented are for a typical purification protocol from 49 g of *E. coli* host cells.
[a] U mL^{-1} $mg.^{-1}$

enzymes. Nevertheless, this chromatographic step resulted in 2.9-fold protein purification (Table 1) and removal of some noticeable contaminants with subunit molecular weight values larger than 29 kDa (Fig. 1). The recombinant mtSD protein in 50 mM Tris–HCl, pH 7.8, containing 1 M $(NH_4)_2SO_4$ adsorbed to a Phenyl-Sepharose High Performance column and was eluted with approximately 57% of 50 mM Tris–HCl, pH 7.8, buffer resulting in a 6.2-fold purification with removal of substantial amount of contaminants (Fig. 1). Gel filtration on Sephacryl S-200 resin was performed to desalt and further purify the recombinant protein. Elution of fractions containing mtSD from an anion-exchange Mono-Q column with approximately 50% of 50 mM Tris–HCl, pH 7.8, buffer containing 0.5 M NaCl resulted in homogeneous mtSD, with a protein yield of approximately 9% (Table 1). It is noteworthy that there was no need for elution with coenzyme from an affinity column as described for *E. coli* shikimate dehydrogenase [21]. Accordingly, the purification protocol here described represents an efficient and low-cost method to obtain homogeneous mtSD. The purification of *M. jannaschii* shikimate dehydrogenase by glutathione and Sepharose Q chromatography yielded, after proteolytic removal of the affinity tag, a protein with five N-terminal residues resulting from cloning artifact [22]. The strategy followed to obtain homogeneous *H. influenzae* shikimate dehydrogenase provided a protein with a C-terminal hexahistidine tag [23]. It has been shown that N- and C-terminal hexahistidine tags have a noticeable negative effect on protein solubility of recombinant proteins expressed in *E. coli* [24]. In addition, His-tagged proteins may have different structure [25] or biological activity [26] as compared to their native form. The mtSD cloning strategy we described elsewhere [15] and protein purification we presented here yielded homogeneous polypeptide chain with no extra amino acid residues.

The subunit molecular mass of active mtSD was determined to be 27,076 Da by electrospray ionization mass spectrometry (ESI-MS), indicating removal of the N-terminal methionine residue (predicted molecular mass: 27,207 Da). The ESI-MS result also revealed a peak at 54,150 Da, indicating that the enzyme could have a dimeric form. No peak could be detected at the expected molecular mass for *E. coli* SD (29,413 Da) and a degree of purity of 98% could be estimated by ESI-MS, thus providing evidence for the identity and purity of the recombinant protein. The first 11 N-terminal amino acid residues of mtSD were identified as SEGPKKAGVLG by the Edman degradation method. This result unambiguously identifies the homogeneous recombinant protein as mtSD and confirms removal of the N-terminal methionine. Modification at the N-termini is a common type of co-/post-translational alteration of proteins synthesized in prokaryotic cells. Methionine aminopeptidase-catalyzed cleavage of initiator methionine is usually directed by the penultimate amino acid residues with the smallest side chain radii of gyration (glycine, alanine, serine, threonine, proline, valine, and cysteine) [27]. Removal of N-terminal methionine from mtSD polypeptide chain conforms to this rule since serine is the penultimate amino acid residue.

The enzymatic activity of homogeneous recombinant mtSD purified was assayed in the reverse direction by continuously monitoring the increase in absorbance at 340 nm upon $NADP^+$-dependent oxidation of D-shikimate to form NADPH and 3-dehydroshikimate. The activity of mtSD was linearly dependent on sample volume added to the reaction mixture (Fig. 2), thereby showing that the initial velocity is proportional to total enzyme concentration and that true initial velocities are being measured. The *M. tuberculosis* SD was stable at −20 °C for at least 1 year.

A value of 58,367 Da for the molecular mass of homogeneous mtSD protein was estimated by gel filtration (data not shown). This result suggests that mtSD is a dimer in solution, in agreement with the ESI-MS results. Shikimate dehydrogenase from *M. jannaschii* has recently been shown to be a dimer in solution [22]. Whereas dehydrogenases usually form oligomers, shikimate dehydrogenase is present as a monomer in both *E. coli* [18] and *H. influenzae* [23].

The apparent kinetics parameters obtained are presented in Table 2. The plots fitted to a hyperbolic equation, indicating that the recombinant enzyme-catalyzed chemical

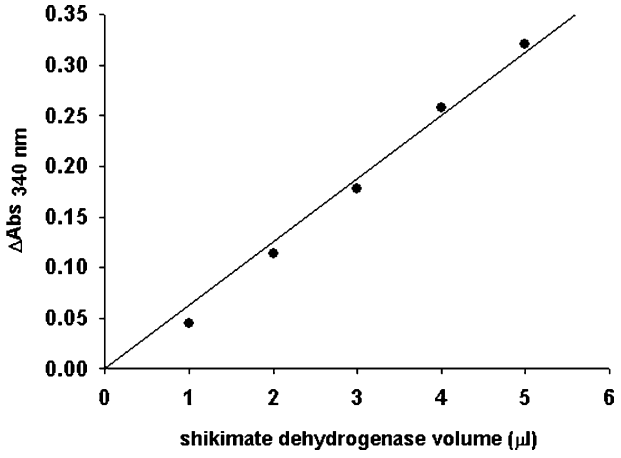


Fig. 2. Linear dependence of mtSD activity on homogeneous protein volume. The rates of mtSD enzyme activity were followed in the reverse reaction by continuously monitoring the increase of NADPH concentration at 340 nm. Reactions were started by addition of varying volumes of homogeneous mtSD protein solution.

Table 2
Apparent kinetic parameters for *M. tuberculosis* shikimate dehydrogenase[a]

Substrates	V_{max} (U mg^{-1})	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m (M^{-1} s^{-1})
DHS	108 ± 5	31 ± 2	49 ± 2	1.6 (±0.1) × 10^6
NADPH	100 ± 5	10 ± 1	45 ± 2	4.5 (±0.5) × 10^6
SHK	18 ± 1	50.18 ± 0.01	8.2 ± 0.5	1.63 (±0.01) × 10^5
$NADP^+$	12.9 ± 0.7	22 ± 2	5.9 ± 0.3	2.68 (±0.03) × 10^5

[a] All constants were measured in Tris–HCl 100 mM (pH 7.0) at 25 °C.

reaction obeys Michaelis–Menten kinetics for all substrates (Fig. 3). The K_m and V_{max} values for DHS were found to be, respectively, 31 μM and 108 U mg^{-1}; and for NADPH they were 10 μM and 100 U mg^{-1}. The k_{cat} for DHS at saturating NADPH was 49 s^{-1} yielding a k_{cat}/K_m of 1.6×10^6 M^{-1} s^{-1}. The K_m value for DHS is lower than the value for SD purified from *Pisum sativum* (340 μM); but the K_m value for NADPH is in agreement with the value of 4.3 μM determined at pH 7.4 [28]. For the reverse reaction, the K_m and V_{max} values for SHK were found to be, respectively, 50.18 μM and 18 U mg^{-1}; and for $NADP^+$ they were 22 μM and 12.9 U mg^{-1}. The k_{cat} for SHK at saturating $NADP^+$ was 8.2 s^{-1} yielding a k_{cat}/K_m of 1.63×10^5 M^{-1} s^{-1}. For the reverse reaction catalyzed by *P. sativum* SD, the K_m values were 10.3 μM for $NADP^+$ and 600 μM for SHK [28]. The K_m value for SHK is thus 10 times larger for *P. sativum* SD than for mtSD. A comparison of the SHK and $NADP^+$ K_m values for *E. coli* SD [29] with those obtained for mtSD shows that these two enzymes catalyze the reverse reaction with similar kinetic constants. A value of 19.6 for the apparent equilibrium constant (K_{eq}) under the experimental conditions given in the Materials and methods section was estimated by the Haldane equation using the apparent steady-state kinetic parameters for mtSD. An apparent equilibrium constant value of 10.3 for the reaction at pH 7.4 has been determined by finding a mixture of substrate concentrations that showed no measurable change in optical density at 340 nm [28]. However, it should be pointed out that the Haldane equation we used here is the kinetic relationship for rapid-equilibrium random Bi Bi system, and it was implicitly assumed that the dissociation constant value for a substrate binding second is not changed by the substrate binding first [30].

The temperature effects on recombinant homogeneous mtSD stability are presented in Fig. 5. Aliquots were removed for assay of residual enzyme activity after heating for 1, 5, 10, 15, 30, 45, and 60 min at temperature values of 15, 25, 37, and 55 °C. The homogeneous mtSD is very stable, maintaining the specific activity, measured at 25 °C, unchanged in the reverse reaction up to 37 °C for at least 1 h of incubation (Fig. 4). However, at 55 °C, there is a gradual loss of recombinant mtSD biological activity, with only 20% of the initial enzyme activity remaining after 1 h of incubation.

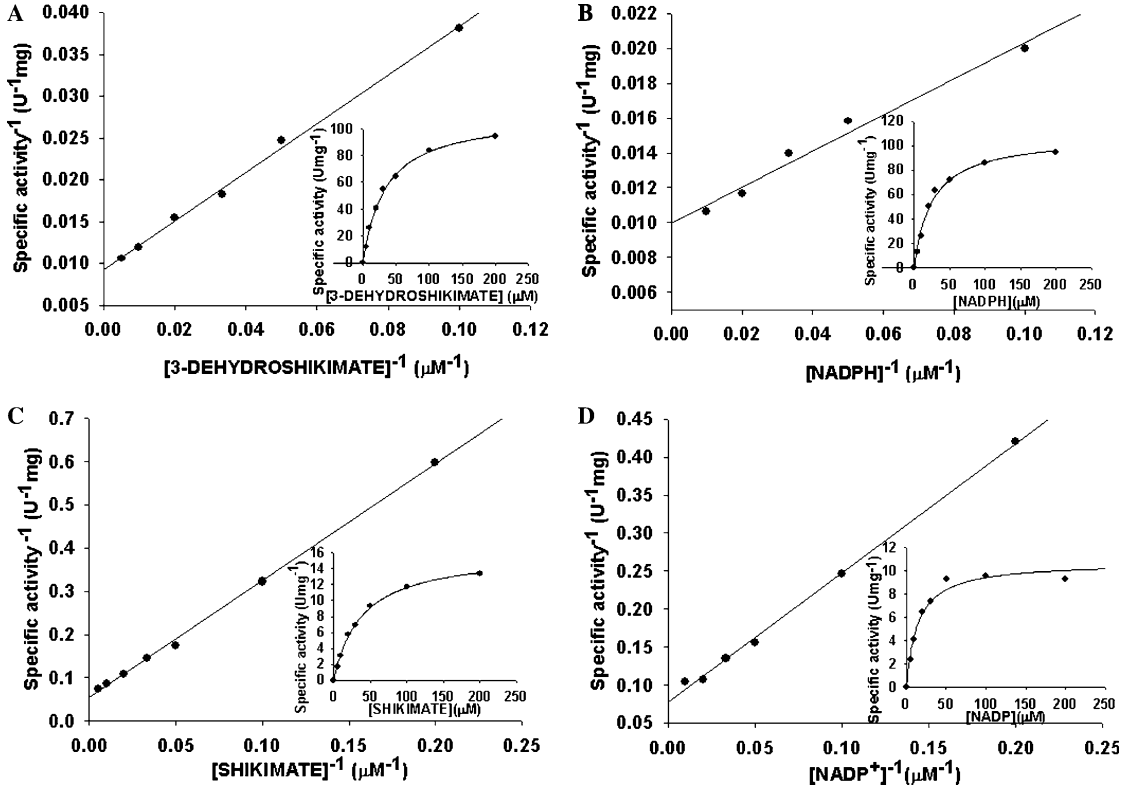


Fig. 3. Reciprocal plots with 3-dehydroshikimate (DHS), $NADP^+$, D-shikimate (SHK) or NADPH as variable substrate (5, 10, 20, 30, 50, 100, and 200 μM) while the concentration of the other substrate was maintained at constant saturation level in the forward and reverse reaction. The enzyme activity was assayed at 25 °C in 100 mM Tris–HCl buffer, pH 7.0. The reaction catalyzed by mtSD was continuously monitored by the absorbance at 340 nm ($\varepsilon_{NADPH} = 6.18 \times 10^3$ M^{-1} cm^{-1}). The kinetics data were analyzed by linear regression fit Michaelis–Menten kinetics. (A) 3-dehydroshikimate as variable substrate (5, 10, 20, 30, 50, 100, and 200 μM) while the NADPH was maintained at constant saturation level (200 μM) in the forward reaction. (B) NADPH as variable substrate (5, 10, 20, 30, 50, 100, and 200 μM) while 3-dehydroshikimate was maintained at constant saturation level (200 μM) in the forward reaction. (C) Shikimate as variable substrate (5, 10, 20, 30, 50, 100, and 200 μM) while $NADP^+$ was maintained at constant saturation level (100 μM) in the forward reaction. (D) $NADP^+$ as variable substrate (5, 10, 20, 30, 50, 100, and 200 μM) while shikimate was maintained at constant near-saturation level (100 μM) in the forward reaction.

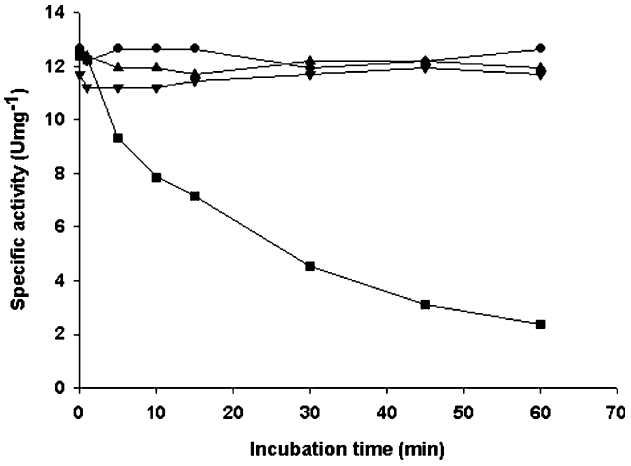


Fig. 4. Heat stability of mtSD activity as a function of time (minutes) of incubation at different temperatures. Incubation temperatures were as follows: 15 °C (●), 25 °C (▲), 37 °C (▼), and 55 °C (■). All reactions were carried out in 100 mM Tris–HCl, pH 7.0, assay mixture containing 200 μM SHK, 100 μM $NADP^+$, and 0.24 nM of homogeneous *M. tuberculosis* SD enzyme and the reaction was measured for 1 min at 25 °C.

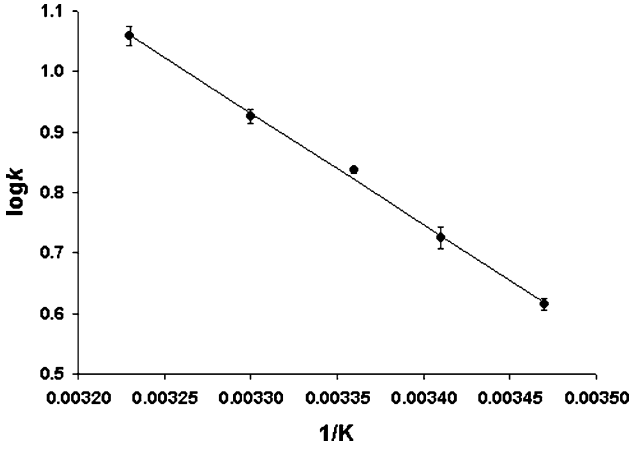


Fig. 5. The activation energy (E_a) estimated for homogeneous *M. tuberculosis* SD enzyme by the Arrhenius equation. All data were obtained in the reverse reaction in 100 mM Tris–HCl, pH 7.0, containing 200 μM SHK, 100 μM $NADP^+$, and 0.24 nM of homogeneous *M. tuberculosis* SD enzyme and the reaction was monitored for 1 min at the following temperatures: 288, 293, 298, 303, and 310 K.

From the linear plots of $\log k$ versus $1/T$ (K^{-1}), a value of 35.2 kJ mol^{-1} for E_a was obtained for mtSD (Fig. 5). It should be pointed out that the E_a value calculated from the Arrhenius plot is an apparent or "average value," and that the pre-exponential factor (A) was considered as temperature-independent in the temperature range used in our experiments. Keeping that in mind, 35.2 kJ mol^{-1} can be considered as the minimal amount of energy required to initiate the mtSD-catalyzed chemical reaction, since we measured the enzyme activity at saturating concentrations of substrates (200 μM SHK and 100 μM $NADP^+$). Interestingly, the Arrhenius plot is linear indicating no change in the rate-limiting step of the mtSD-catalyzed chemical reaction at different temperatures. In addition, there was no sudden drop in the Arrhenius plot at low $1/T$ (high T) values that could indicate protein denaturation, which is consistent with mtSD thermal stability up to 55 °C.

The three-dimensional structures of shikimate dehydrogenase from *E. coli* [29,31], *H. influenzae* [23,32], and *M. jannaschii* [22] have been determined. Sequence alignment of shikimate dehydrogenase from *M. tuberculosis* H37Rv strain, *E. coli*, *H. influenzae*, *M. jannaschii* was carried out using ClustalW [33,34]. The identity between the SD sequences from *M. tuberculosis* and *E. coli* is 24%. The substrate-binding site in the *E. coli* SD has been identified by the position of the nicotinamide ring of the cofactor and was delineated almost entirely by residues from the N-terminal domain [29]. This binding site is in a pocket where most of the residues absolutely conserved in the shikimate dehydrogenase (SDH) family are located, i.e., Ser14, Ser16, Lys65, Asn86, Thr101, Asp102, and Gln244 (*E. coli* SD numbering). These residues are conserved in the mtSD polypeptide sequence (Fig. 6) corresponding to the following residues: Ser18, Ser20, Lys69, Asn90, Thr104, Asp105, and Gln243 (*M. tuberculosis* SD numbering). However, the 3-dehydroshikimate-binding site has not unambiguously been shown for *E. coli* shikimate dehydrogenase since it was inferred from DTT and sulfate ions bound to the crystal and position of the nicotinamide ring of $NADP^+$ [29]. The sequence identity (24%) between mtSD and *M. jannaschii* SD is similar to the identity between the sequences of *M. tuberculosis* and *E. coli* SDs. The N-terminal domain (domain I) of *M. jannaschii* SD is responsible for 3-dehydroshikimate substrate binding [22]. The putative active site residues within domain I that are likely to be involved with substrate binding are invariant polar residues include Lys70, Asn91, and Asp106 (*M. jannaschii* SD numbering). Further examination of the active site cleft of *M. jannaschii* SD revealed another strictly conserved residue, Gln254. These residues correspond to Lys69, Asn90, Asp105, and Gln244 in *M. tuberculosis* SD. It has been proposed that these residues are likely to be involved in catalytic reduction of DHS to SHK catalyzed by *M. jannaschii* SD [22]. The amino acids Ser14, Ser16, Lys65, Asn85, Asp102, and Gln245 (*H. influenzae* SD numbering) have been suggested as some of the potential residues involved in 3-dehydroshikimate-binding site in *H. influenzae* SD [23]. All these residues are conserved in the mtSD polypeptide sequence (Fig. 6), and correspond to the same residues observed in the *E. coli* SD sequence. More recently, the crystal structure of a newly characterized shikimate dehydrogenase-like protein (HI0607) from *H. influenzae* has been determined [32], and the conserved residues Lys67, Asn88, Asp103, and Gln242 have been proposed to be involved in either catalysis or substrate binding. It should be pointed out that the three-dimensional structures of shikimate dehydrogenases determined were in complex with the cofactor and no three-dimensional structure in complex with 3-dehydroshikimate/shikimate molecules have been reported to date.

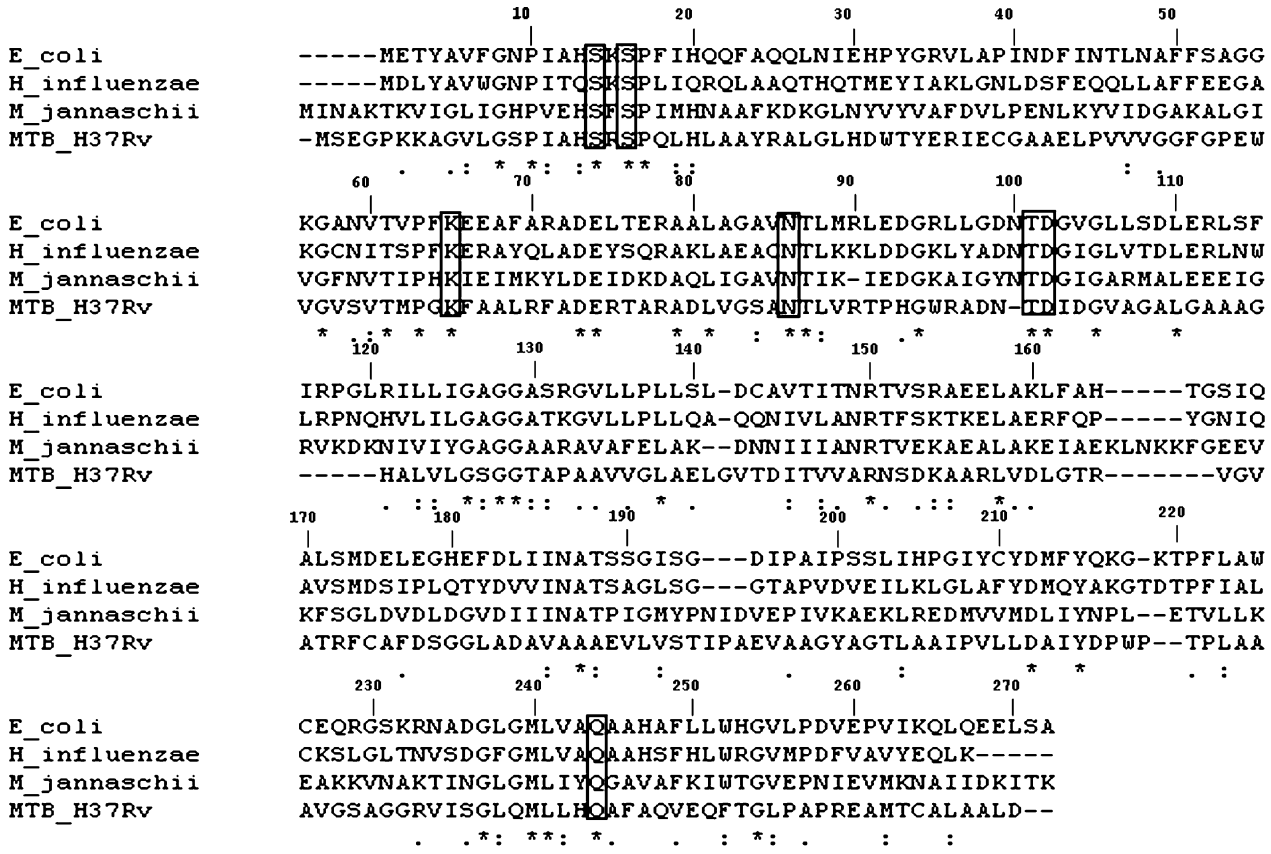


Fig. 6. Multiple sequence alignment of SD from *M. tuberculosis* using ClustalW. The boxes denote residues conserved in the SDH family involved in catalytic reduction of DHS to SHK. This sequence alignment was constructed using the following sequences from GenBank: *E. coli* K12 AroE (NP_417740, residues 1–172), *H. influenzae* AroE (YP_248346, residues 1–268), *M. jannaschii* AroE (NP_248077, residues 1–282), and *M. tuberculosis* AroE (CAB06186, residues 1–269). The symbols are as follows: * = identity, : = strong similarity, . = week similarity. These nomenclature is based on the properties of aminoacids using the substitution matrix BLOSUM [34].

The work presented here describes, to the best of our knowledge, the first purification protocol of recombinant *M. tuberculosis* SD to homogeneity and determination of its oligomeric state. In addition, here we present a detailed characterization of the homogeneous mtSD. It should be pointed out that expression of mtSD in soluble and active form proved to be laborious to achieve [15]. Protein production and crystallization must be optimized if structural genomics will ever reach its goal of solving the three-dimensional structure of the whole proteome encoded by a given genome [35]. Unfortunately, even when a genome can be sequenced, only up to 20% of the protein targets can produce soluble proteins under very basic experimental conditions [36]. Thus, expression of proteins in soluble form has been identified as an important bottleneck in efforts to determine biological activity and crystal structure of *M. tuberculosis* proteins [37].

The protocol for mtSD purification presented here should provide protein in quantities necessary for further enzymological studies and three-dimensional structure determination efforts. Determination of mtSD kinetic mechanism by steady-state and pre-steady-state kinetics, and isotope effects, site-directed mutagenesis, and chemical rescue will allow elucidation of its chemical and catalytic mechanism. The three-dimensional structures of shikimate dehydrogenases from *E. coli* [29,31], *H. influenzae* [23,32], and *M. jannaschii* [22] should facilitate screening of experimental conditions to obtain crystals of mtSD protein and may provide templates for mtSD structure determination by molecular replacement. All of the enzymes that make up the shikimate pathway are potential targets for the design of novel drugs directed against pathogenic bacteria. The structure of *M. tuberculosis* shikimate dehydrogenase and the understanding of the enzyme mode of action will be used as a platform for the design of effective inhibitors of this pathway aiming at the development of antitubercular agents.

Acknowledgments

Financial support for this work was provide by Millennium Initiative Program MCT-CNPq, Ministry of Health-Department of Science and Technology-UNESCO (Brazil) to D.S.S. and L.A.B. D.S.S. and L.A.B. also acknowledge grants awarded by PRONEX, CNPq/FAPERGS. D.S.S. (CNPq, 304051/1975-06), L.A.B. (CNPq, 520182/99-5), and M.S.P. (CNPq, 500079/90-0) are researchers awardees from the National Research Council (CNPq) of Brazil. We

thank Professor John W. Frost, Department of Chemistry of Michigan State University, for his generous gift of 3-dehydroshikimate substrate.

References

[1] World Health Organization, World Health Report on Infections Diseases: Removing Obstacles to Healthy Development. Atar, Switzerland. WHO/CDS/99.1, 1999.
[2] C. Dye, S. Scheele, P. Dolin, V. Pathania, M.C. Raviglione, Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country, JAMA 282 (1999) 677–686.
[3] B.R. Bloom, C.J.L. Murray, Tuberculosis: commentary on a reemergent killer, Science 257 (1998) 1055–1064.
[4] A. Knigge, H. Morr, A. Kilian, Tuberculosis and HIV/AIDS epidemics: opportunities for cross support, Deutsche Gesellschaft Für Technische Zusammenarbeit (GTZ) GmbH (2000) 1–7 (<www.gtz.de/>).
[5] L.A. Basso, J.S. Blanchard, Resistance to antitubercular drugs, Adv. Exp. Med. Biol. 456 (1998) 115–144.
[6] J.D. McKinney Jr., W.R. Jacobs, B.R. Bloom, Persisting problems in tuberculosis, in: R.M. Krause (Ed.), Emerging Infections, Academic Press, New York, 1998, pp. 51–146.
[7] World Health Organization, Anti-tuberculosis drug resistance in the world: Third Global Report (2004).
[8] K. Duncan, Progress in TB drug development and what is still needed, Tuberculosis 83 (2003) 201–207.
[9] R.J. O'Brien, P.P. Nunn, The need for new drugs against tuberculosis, Am. J. Respir. Crit. Care Med. 162 (2001) 1055–1058.
[10] R. Bentley, The shikimate pathway—a metabolic tree with many branches, Crit. Rev. Biochem. Mol. Biol. 25 (1990) 307–384.
[11] C. Ratledge, Nutrition, growth and metabolism, in: C. Ratledge, J.L. Stanford (Eds.), The Biology of the Mycobacteria, vol. 1, Academic Press, London, 1982, pp. 185–271.
[12] J.J. de Voos, K. Rutter, B.G. Schroder, H. Su, Y. Zhu, C.E. Barry III, The salicylate-derived mycobactin siderophores of *Mycobacterium tuberculosis* are essential for growth in macrophages, Proc. Natl. Acad. Sci. USA 97 (2000) 1252–1257.
[13] T. Parish, N.G. Stoker, The common aromatic amino acid biosynthesis pathway is essential in *Mycobacterium tuberculosis*, Microbiology 148 (2002) 3069–3077.
[14] S.T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S.V. Gordon, K. Eiglmeier, S. Gas, C.E. Barry Jr., F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, B.G. Barrell, Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence, Nature 393 (1998) 537–544.
[15] M.L. Magalhães, C.P. Pereira, L.A. Basso, D.S. Santos, Cloning and overexpression of functional shikimate dehydrogenase (EC 1.1.1.25) from *Mycobacterium tuberculosis* H37Rv, Protein Expr. Purif. 26 (2002) 59–64.
[16] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, Nature 227 (1970) 680–685.
[17] M.M. Bradford, R.A. McRorie, W.L. Williams, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, Anal. Biochem. 72 (1976) 248–254.
[18] S. Chaudhuri, J.R. Coggins, The purification of shikimate dehydrogenase from *Escherichia coli*, Biochem. J. 226 (1985) 217–223.
[19] H. Chassaigne, R. Lobinski, Characterization of horse kidney metallothionein isoforms by electrospray MS and reversed-phase HPLC–electrospray MS, Analyst 123 (1998) 2125–2130.
[20] B.H. Johnson, M.H. Hecht, Recombinant proteins can be isolated from *E. coli* cells by repeated cycles of freezing and thawing, Bio/Technology 12 (1994) 1357–1360.
[21] J. Maclean, S.A. Campbell, K. Pollock, S. Chackrewarthy, J.R. Coggins, A.J. Lapthorn, Crystallization and preliminary X-ray analysis of shikimate dehydrogenase from *Escherichia coli*, Acta Crystallogr. Sect. D. Biol. Crystallogr. 56 (2000) 512–515.
[22] A.K. Padyana, S.K. Burley, Crystal structure of shikimate 5-dehydrogenase (SDH) bound to NADP: insights into function and evolution, Structure 11 (2003) 1005–1013.
[23] S. Ye, F.v. Delft, A. Brooun, M.W. Knuth, R.V. Swanson, D.E. McRee, The crystal structure of shikimate dehydrogenase (AroE) reveals a unique NADPH binding mode, J. Bacteriol. 185 (2003) 4144–4151.
[24] E.A. Woestenenk, M. Hammarström, S. van dn Berg, T. Härd, H. Berglund, His tag effect on solubility of human proteins produced in *Escherichia coli*: a comparison between four expression vectors, J. Struct. Funct. Genomics 5 (2004) 217–229.
[25] A. Chant, C.M. Kraemer-Pecore, R. Watkin, G.G. Kneale, Attachment of a histidine tag to the minimal zinc finger protein of the *Aspergillus nidulans* gene regulatory protein AreA causes a conformational change at the DNA-binding site, Protein Expr. Purif. 39 (2005) 152–159.
[26] I. Fonda, M. Kenig, V. Gaberck-Porekar, P. Prostovaek, V. Menart, Attachment of histidine tags to recombinant tumor necrosis factor-alpha drastically changes its properties, Sci. World J. 2 (2002) 1312–1325.
[27] P.-H. Hirel, J.-M. Schmitter, P. Dessen, G. Fayat, S. Blanquet, Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid, Proc. Natl. Acad. Sci. USA 86 (1989) 8247–8251.
[28] D. Balinsky, A.W. Dennis, W.W. Cleland, Kinetic and isotope-exchange studies on shikimate dehydrogenase from *Pisum sativum*, Biochemistry 10 (1971) 1947–1952.
[29] G. Michel, A.W. Roszak, V. Sauvé, J. Maclean, A. Matte, J.R. Coggins, M. Cygler, A.J. Lapthorn, Structures of shikimate dehydrogenase AroE and its paralog YdiB, J. Biol. Chem. 278 (2003) 19463–19472.
[30] W.W. Cleland, An analysis of Haldane relationships, Methods Enzymol. 87 (1982) 366–369.
[31] J. Benach, I. Lee, W. Edstron, A.P. Kuzin, Y. Chiang, T.B. Acton, G.T. Montelione, J.F. Hunt, The 2.3 Å crystal structure of the shikimate 5-dehydrogenase orthologue YdiB from *Escherichia coli* suggests a novel catalytic environment for na NAD-dependent dehydrogenase, J. Biol. Chem. 278 (2003) 19176–19182.
[32] S. Singh, S. Korolev, O. Koroleva, T. Zarembinski, F. Collart, A. Joachimiak, D. Christendat, Crystal structure of a novel shikimate dehydrogenase from *Haemophilus influenzae*, J. Biol. Chem. 280 (2005) 17101–17108.
[33] BLAST: S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSIBLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402.
[34] CLUSTALW: J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673–4680.
[35] W.R.A. Taylor, A 'periodic table' for protein structures, Nature 415 (2002) 657–660.
[36] S.A. Lesley, P. Kuhn, A. Godzik, A.M. Deacon, I. Mathews, A. Kreusch, G. Spraggon, H.E. Klock, D. McMullan, T. Shin, J. Vincent, A. Robb, L.S. Brinen, M.D. Miller, T.M. McPhillips, M.A. Miller, D. Scheibe, J.M. Canaves, C. Guda, L. Jaroszewski, T.L. Selby, M.A. Elsliger, J. Wooley, S.S. Taylor, K.O. Hodgson, I.A. Wilson, P.G. Schultz, R.C. Stevens, Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline, Proc. Natl. Acad. Sci. USA 99 (2002) 11664–11669.
[37] R. Vicentelli, C. Bignon, A. Gruez, S. Canaan, G. Sulzenbacher, M. Tegoni, V. Campanacci, C. Cambillau, Medium-scale structural genomics: strategies for protein expression and crystallization, Acc. Chem. Res. 36 (2003) 165–172.

Capítulo II

Este artigo será submetido para a revista *ARCHIVES OF BIOCHEMISTRY AND BIOPHYSICS*

Fonseca I.O., Silva R.G., Fernandes C., Souza O.N., Basso L.A., Santos D.S. "**Shikimate Dehydrogenase from *Mycobacterium tuberculosis* H37Rv: Kinetic and Chemical Mechanisms.**

Kinetic and chemical mechanisms of shikimate dehydrogenase from *Mycobacterium tuberculosis*

Isabel O. Fonseca[‡π], Rafael G. Silva[‡π], Claudia Fernandes[†], Osmar Norberto de Souza[†], Luiz A. Basso[‡*], and Diógenes S. Santos[‡*]

[‡]Centro de Pequisas em Biologia Molecular e Funcional, Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), Porto Alegre - RS 90619-900, Brazil.

[π]Programa de Pós-Graduação em Ciências Biológicas: Bioquímica, Universidade Federal do Rio Grande do Sul (UFRGS).

[†]Laboratório de Bioinformática, Modelagem e Simulação de Biossistemas – LABIO, Faculdade de Informática, PUCRS, Porto Alegre - RS 90619-900, Brazil.

Short Title: *M. tuberculosis* shikimate dehydrogenase mode of action

Address correspondence to: Luiz A. Basso or Diógenes S. Santos, Centro de Pesquisas em Biologia Molecular e Funcional, Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), Porto Alegre - RS - 90619-900, Brazil. Phone/Fax: +55 51 33203629; E-mail: luiz.basso@pucrs.br or diogenes@pucrs.br

ABSTRACT

Mycobacterium tuberculosis shikimate dehydrogenase (*Mtb*SD) catalyzes the fourth reaction in the shikimate pathway, the NADPH-dependent reduction of 3-dehydroshikimate. To gather information on the kinetic mechanism, initial velocity patterns, product inhibition, and primary deuterium kinetic isotope effect studies were performed and the results suggested a steady-state ordered bi-bi kinetic mechanism. The magnitudes of both primary and solvent kinetic isotope effects indicated that the hydride transferred from NADPH and protons transferred from the solvent in the catalytic cycle are not significantly rate limiting in the overall reaction. Proton inventory analysis indicates that one proton gives rise to solvent isotope effects. Multiple isotope effect studies indicate that both hydride and proton transfers are concerted. The pH profiles revealed that acid/base chemistry takes place in catalysis and substrate binding. The *Mtb*SD 3D model was obtained *in silico* by homology modeling. Kinetic and chemical mechanisms for *Mtb*SD are proposed on the basis of experimental data.

Keywords: shikimate dehydrogenase, tuberculosis, drug target, 3-dehydroshikimate, shikimate, kinetic mechanism, chemical mechanism, *Mycobacterium tuberculosis*.

Introduction

Tuberculosis[1] (TB) is a pandemic, which even today, remains a major global health concern. Its causative agent, *Mycobacterium tuberculosis,* is one of the most prolific infectious agents affecting humans. A third of the world's population is thought to host *M. tuberculosis* and approximately 30 million people have died from the disease in the past decade [1]. The treatment of multidrug-resistant TB (MDR-TB), defined as resistant to at least isoniazid and rifampicin [2], requires the administration of second-line drugs that are more toxic and less effective, and are given for at least three times as long as, and 100 times as expensive as basic chemotherapy regimens [3]. More recently, a survey of the frequency and distribution of extensively drug-resistant (XDR) TB cases, which are defined as cases in persons with TB whose isolates were resistant to isoniazid and rifampicin and at least three of the six main classes of second-line drugs, showed that during 2000-2004, of 17,690 TB isolates, 20% were MDR and 2% were XDR, and that XDR-TB has a wide geographic distribution [4]. New antimycobacterial agents are thus needed to improve the treatment of MDR- and XDR-TB, and to provide more effective treatment of latent tuberculosis infection [5].

The availability of the complete genome sequence of *M. tuberculosis* allowed the identification of molecular targets against which new therapeutical agents may be developed [6]. Homologues to the seven enzymes of the shikimate pathway were identified in the genome sequence of *M. tuberculosis*. This pathway has been shown to be essential for the viability of *M. tuberculosis* [7]. Accordingly, the essentiality of the mycobacterial shikimate pathway and its absence from human host indicate that any enzyme of this pathway represents a promising target for the development of non-toxic antimycobacterial

agents. However, an understanding of the kinetic mechanism and availability of structural models are needed to help the rational inhibitor design to progress more rapidly [8].

Shikimate dehydrogenase (EC 1.1.1.25), the fourth enzyme in the shikimate biosynthesis pathway, catalyzes the NADPH-dependent reduction of 3-dehydroshikimate (DHS) to shikimate (SHK) (Fig. 1). We have previously reported cloning and expression of *M. tuberculosis* SD (*Mtb*SD) [9]. In addition, measurements of the $NADP^+$-dependent oxidation of shikimate to 3-dehydroshikimate catalyzed by recombinant *Mtb*SD confirmed the correct assignment to the structural gene encoding SD in *M. tuberculosis* [9]. More recently, we have reported purification to homogeneity of recombinant functional *Mtb*SD, *N*-terminal amino acid sequencing, electrospray ionization mass spectrometry analysis, chromatographic size exclusion, and enzyme activity measurements that unambiguously demonstrated the identity of recombinant *Mtb*SD, its oligomeric state, the apparent kinetic parameters for all substrates, the thermal stability, and the activation energy for the enzyme-catalyzed chemical reaction [10].

In the present work, initial velocity patterns in the forward direction, product inhibition studies, primary deuterium kinetic isotope effects, solvent kinetic isotope effects, proton inventory, multiple kinetic isotope effects, pH rate profiles, and homology modeling were performed to investigate the *Mtb*SD kinetic and chemical mechanisms. The data here presented provide a framework on which to base the rational design of inhibitors with possible antitubercular activity.

Materials and Methods

Enzymatic Assay for MtbSD - All enzyme activity assays were carried out at 25°C in 100 mM potassium phosphate (pH 7.3) by monitoring the decrease in absorbance at 340 nm ($\varepsilon = 6220\ M^{-1}\ cm^{-1}$) accompanying the conversion of NADPH to $NADP^+$ in the presence of DHS and *Mtb*SD. The final enzyme concentration was 1.8 nM. To obtain initial velocity measurements, rate values were determined under experimental conditions in which less than 5% of substrate was consumed.

Initial Velocity and Product Inhibition - To determine the steady-state kinetic parameters and initial velocity patterns, *Mtb*SD activity was measured in the presence of varying concentrations of one substrate and several fixed-varied concentrations of the other. Product inhibition patterns were determined by measuring initial rates at five concentrations of one substrate, fixed non-saturating concentration of the co-substrate, and fixed-varying levels of product (either $NADP^+$ or SHK).

Kinetic Isotope Effects and Proton Inventory - The synthesis of [4*S*-^{2}H]NADPH was performed as described by Ottolina *et al.* (1989) [11]. Both NADPH and NADPD substrates were purified on an FPLC Mono-Q column (GE Healthcare) as previously described, and the fractions with absorbance ratios $A_{260nm}/A_{340nm} \leq 2.3$ were pooled [12]. All measurements were performed in duplicate. Primary deuterium kinetic isotope effects were determined by measuring initial rates in the presence of varying concentrations of one substrate and five fixed concentrations of the other, with either NADPH or NADPD. Solvent kinetic isotope effects were determined by measuring initial velocities using a

saturating level of one substrate and varying concentrations of the other in either H_2O or 90 atom % D_2O. Multiple kinetic isotope effects were obtained by determining the primary isotope effects using D_2O as solvent. The proton inventory was determined using saturating concentrations of DHS and NADPH at various mole fractions of D_2O. Each proton inventory individual initial rate datum is the average of triplicate measurements.

pH-Rate Profiles - To determinate the pH dependence of k_{cat} and k_{cat}/K_m, initial velocities were measured in the presence of varying concentrations of one substrate and a saturating level of the other at different pHs in a mixture of potassium phosphate and boric acid. The mixed buffers were used over the following pH values: 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, 9.0, 9.5, and 10.0. The *Mtb*SD enzyme had been incubated in buffers at all the pH values mentioned above and assayed under standard conditions to ensure enzyme stability at the tested pHs.

Data Analysis - Initial velocity kinetic data were fitted to appropriate equations by using the nonlinear regression function of SigmaPlot 2000 (SPSS, Inc.). Substrate hyperbolic saturation curves at a single concentration of the fixed substrate and varying concentrations of the other substrate were fitted to equation 1. Intersecting initial velocity patterns were fitted to equation 2, which describes a sequential mechanism. For equations 1 and 2, v is the measured reaction velocity, V is the maximal velocity, A and B are the concentrations of substrates (DHS and NADPH), K_a, and K_b are the corresponding Michaelis- Menten constants, and K_{ia} is the dissociation constant for substrate A.

$$v = VA / (K_a + A) \qquad \text{(Eq. 1)}$$

$$v = VAB / (K_aB + K_bA + K_{ia}K_b + AB) \qquad \text{(Eq. 2)}$$

Competitive and noncompetitive inhibition data were fitted to, respectively, equation 3 and 4. For equations 3 and 4, I represent the inhibitor concentration, and K_{is} and K_{ii} are the slope and intercept inhibition constants, respectively.

$$v = VA / [K_a (1 + I/K_{is}) + A] \qquad \text{(Eq. 3)}$$

$$v = VA / [K_a (1 + I/K_{is}) + A (1 + I/K_{ii})] \qquad \text{(Eq. 4)}$$

The kinetic isotope effect data were fitted to equations 5 or 6, which describe, respectively, effects on both V and V/K and on V only, where Fi represents the fraction of isotopic label, and $E_{V/K}$ is the isotope effect on V/K minus one and E_V is the isotope effect on V minus one.

$$v = VA / [K (1 + FiE_{V/K}) + A (1 + FiE_V)] \qquad \text{(Eq. 5)}$$

$$v = VA / [\mathrm{K} + A (1 + FiE_V)] \qquad \text{(Eq. 6)}$$

Data for pH profiles data that showed a decrease in $\log k_{cat}$ and $\log k_{cat}/K_m$ with a slope of -1 as the pH values increased were fitted to equation 7, where y is the apparent kinetic parameter, C is the pH-independent plateau value of y, H is the hydrogen ion concentration, and K_b is the apparent base dissociation constant for ionizing groups.

$$\log y = \log [C/(1 + K_b/H)] \qquad \text{(Eq. 7)}$$

Homology Modeling - Primarily, homology modeling is the identification and selection of template proteins from the Protein Data Bank (PDB) [13] that are related to the target sequence. The program Blastp [14] was used to search for templates. Multiple sequence alignment comparisons were carried out to improve the sensitivity of the search and to find regions with high similarity using ClustalW [15]. The program MODELLER6v2 [16] was used to build the protein models, using the standard protocol of the comparative protein structure modeling methodology [17]. The best model of each enzyme was evaluated and

selected according to their stereochemical quality analyzed with PROCHECK [18]. Validation of the 3D profiles of the models was performed with VERIFY 3D [19]. Here, the structure prediction of *Mtb*SD was based on 3D structure for the homologous SD from *Escherichia coli* (*E. coli*) (PDB ID: 1NYT), experimentally determined by X-ray diffraction at 1.5Å resolution [20]. Root-mean square deviation (RMSD) between template and model structures was calculated using SwissPdbViewer [21]. Drawing of protein structures were performed with the program SwissPdbViewer [21].

Results

Initial Velocity and Product Inhibition Patterns - The constants for steady-state kinetics of *Mtb*SD were determined under initial velocity experimental conditions. When DHS was varied at fixed-varied NADPH concentrations, the lines intersected on the *x*-axis and on the left of the *y*-axis (Fig. 2A); when NADPH was varied with fixed-varied DHS concentrations, the lines intersected below the *x*-axis and on the left of the *y*-axis (Fig. 2B). The data were fitted to equation 2, and yielded the following values: K_{DHS} = 44 ± 3 μM, K_{iaDHS}= 30 ± 9, K_{NADPH} = 34 ± 2 μM, k_{cat} = 78 ± 2 s^{-1}, k_{cat}/K_{DHS} = 1.8 x 10^6 M^{-1} s^{-1}, and k_{cat}/K_{NADPH} = 2.3 x 10^6 M^{-1} s^{-1}.

Product inhibition experiments were performed and the data fitted to equation 3 or 4. The results are given in Table 1. $NADP^+$ is a competitive inhibitor (C) and a noncompetitive inhibitor (NC) versus NADPH and DHS, respectively. For SHK product, the inhibition pattern is C and NC versus, respectively, DHS and NADPH.

Kinetic Isotope Effects and Proton Inventory - The primary deuterium kinetic isotope effects on *V/K* and *V* for both substrates are presented in Table 2. The primary kinetic isotope effect values of 1.8 for $^{D}V_{DHS}$ and 1.5 for $^{D}V_{NADPH}$ using [4*S*-^{2}H]NADPH indicate that the C_4-*proS* hydride (B side) is transferred to DHS in the oxy-reduction reaction catalyzed by *Mtb*SD.

The $^{D}(V/K_{app})_{DHS}$ values decreased to a limiting value of 1.0 ± 0.03 as the concentrations of either NADPH or NADPD increased (Fig. 3), whereas $^{D}(V/K_{app})_{NADPH}$ (1.4 ± 0.3) was independent of fixed-varied concentrations of DHS (Fig. 3 - inset).

Solvent kinetic deuterium isotope effects were determined at pH 7.3 (Fig. 4A and Fig. 4B), a pH region in which the kinetic parameters are independent of small changes in pH. Small solvent kinetic isotope effects on V were observed for DHS (1.5 ± 0.3) and NADPH (1.3 ± 0.2), whereas no solvent effects on V/K were observed (Table 2). The proton inventory, a relationship between V and the mole fraction of D_2O, was linear for *Mtb*SD (Fig. 4A - inset).

In order to distinguish between stepwise and concerted mechanisms, multiple isotope effects were evaluated measuring the primary isotope effects for both substrates in either D_2O (Fig. 5A and 5B) or H_2O (Fig. 5A and 5B - insets), and the results are summarized in Table 2. The $^{D}V/K_{NADPH}$ value in H_2O (1.4 ± 0.3) was within standard error of the $^{D}V/K_{NADPH}$ value in D_2O (1.5 ± 0.1), whereas $^{D}V_{NADPH}$ in D_2O (2.5) increased as compared to the value obtained in H_2O (1.5). The magnitude of the $^{D}V_{DHS}$ value in D_2O (2.6) also increased, and, most important, running the reaction in D_2O allowed the observation of a small, but significant, $^{D}V/K_{DHS}$, (1.3 ± 0.1) in contrast to the unity obtained in H_2O.

pH Rate Profiles - To probe the role of acid/base chemistry in the mechanism of *Mtb*SD, the pH dependence of k_{cat} and k_{cat}/K_m for DHS and NADPH was determined over the pH range of $5.5 \leq pH \leq 10.0$. The pH profiles of k_{cat} for both substrates are very similar and decrease at high pHs with a slope of -1 (Fig. 6A), demonstrating that deprotonation of a single residue with apparent pK_a value of 8.9 ± 0.1 abolishes *Mtb*SD catalytic activity. The k_{cat}/K_m profiles for DHS and NADPH show a decrease at high pH values with a slope of -1

(Fig. 6B and Fig. 6C), suggesting that deprotonation of a single ionizable group with pK_a value of 9.1 ± 0.1 diminishes DHS and NADPH binding.

Homology Modeling -In the search for templates to perform the homology modeling we found three candidate structures, namely, *E. coli* SD (PDB ID: 1NYT) [20], *E. coli* YdiB (PDB ID: 109B) [20] and *Methanococcus jannaschii* SD (PDB ID: 1NVT) [22]. The Blastp results listed *M. jannaschii* SD (PDB ID: 1NVT) as the best template for *Mtb*SD. However, the structure coordinates were not available at the time of this modeling work. The second best result was YdiB (PDB ID: 109B), although its function is different from that of *Mtb*SD. YdiB was characterized as a dual specificity quinate/shikimate dehydrogenase that utilizes either NAD^+ or $NADP^+$ as cofactor. Multiple sequence alignment, with the insertion of two gap regions into the *Mtb*SD sequence, illustrates this difference (Fig. 7). Segments of the target sequence which have no equivalent in the template are the most difficult regions to model. Hence, the *E. coli* SD (PDB ID: 1NYT), solved experimentally by X-ray diffraction at 1.5Å resolution [20], was used as a template to model the 3D structure of *Mtb*SD. They both have the same function and a sequence identity of approximately 25%, the limit usually allowed for comparative protein structure modeling [14].

Ten models of the enzyme were built, and evaluated by PROCHECK [18] and VERIFY 3D [19] to choose the best one. *Mtb*SD contains 269 amino acid residues. Out of 220 non-glycine and non-proline residues, 199 or 90.5% were located in the most favored regions of the Ramachandran plot. The best model of *Mtb*SD is illustrated in Figure 8. The backbone RMSD between template and model structures is 1.93 Å. Considering the

sequence divergence between these ortholog sequences, approximately 25% sequence identity, this RMSD value is expected and the modeling results, overall, demonstrate that the *Mtb*SD model obtained is highly satisfactory and can thus be used to infer structure-activity relationships for this enzyme (Fig. 8).

Discussion

Kinetic Mechanism of MtbSD - The families of intersecting double-reciprocal plots observed in the initial velocity for the forward direction for both substrates are consistent with a sequential mechanism and thus both substrates must attach to the enzyme forming a ternary complex before any product is released. Accordingly, a ping-pong mechanism could be discarded. In addition, a rapid equilibrium ordered mechanism was ruled out because the families of double reciprocal plots obtained intersect on the left of the *y*-axis (Fig. 2A and Fig. 2B).

The steady-state kinetic constant values given here are larger than the apparent kinetic parameters we have previously reported [10], whose data were collected in Tris-HCl 100mM, pH 7.0. We verified a significant reduction in the catalytic rate for this enzyme when the assay mix was buffered using Tris-HCl instead of potassium phosphate (data not show). In addition, here we report true steady-state kinetic parameters instead of apparent kinetic parameters.

Determination of steady-state kinetic parameters for *Pisum sativum* SD [23] in 100 mM potassium phosphate, pH 7.4, yielded values of 340 ± 40 μM for K_{DHS}, which is approximately 8-fold larger than the value for *Mtb*SD (44 μM), and 4.3 ± 0.5 μM for

K_{NADPH}, which is approximately 8-fold smaller than the value here reported for *Mtb*SD (34 μM).

The first step in determining the mechanism of action of an enzyme is to establish its kinetic mechanism, the order of substrate binding and release of products. This is most commonly achieved by product inhibition studies in which we observe the patterns obtained when substrate concentrations are varied at fixed levels of other substrate(s) in the presence of products, yielding information not available from initial velocity studies.

Product inhibition data for *Mtb*SD (Table 1) were similar to those for *P. sativum* SD. For the reaction of *P. sativum* SD [23], NADPH was found to give linear competitive inhibition with $NADP^+$ as variable substrate, and NADPH was a noncompetitive inhibitor *versus* SHK at low $NADP^+$ concentrations. The results for SHK inhibition versus DHS indicated slightly noncompetitive inhibition, thereby suggesting that the mechanism is ordered [23]. Product inhibition data presented here suggest two possible kinetic mechanisms for *Mtb*SD, steady-state ordered and rapid equilibrium random with two dead-end ternary complexes (*Mtb*SD-NADPH-SHK and *Mtb*SD-$NADP^+$-DHS). An application of primary deuterium isotope effects is to distinguish among possible kinetic mechanisms. Accordingly, to differentiate between steady-state ordered and rapid equilibrium random, analysis of primary kinetic isotope effects was carried out.

Measurements of the steady-state primary deuterium kinetic isotope effects on apparent *V/K* for one substrate were carried out at various cosubstrate concentrations (Fig. 3). Based on the mechanistic deductions from isotope effects for multireactant enzymes developed by Cook and Cleland [24], the primary isotope effect on $^D(V/K_{app})_B$ is independent of the concentration of A and $^D(V/K_{app})_A$ decreases as the concentration of B

increases and reaches a limiting value of 1.0 at infinite concentration of B for a steady-state ordered mechanism. A and B refer to, respectively, the first and second substrates to bind to the enzyme. The kinetic isotope effects on *V/K* results (Fig. 3) are in agreement with a steady-state ordered bi-bi mechanism with DHS binding first followed by NADPH binding to *Mtb*SD active site (Fig. 9). In addition, the DHS substrate is sticky, hence it reacts to give SHK as fast as, or faster than, it dissociates from the enzyme [24]. Based on initial velocity patterns, product inhibition, analysis of Haldane relationships and isotope-exchange studies, an ordered kinetic mechanism with NADPH binding first has been proposed for *P. sativum* SD [23]. Initial velocity and dead-end inhibition studies showed that the kinetic mechanism is ordered bi-bi for mouse class II alcohol dehydrogenase with coenzyme binding first [25]. The NAD-dependent mouse class II alcohol dehydrogenase is structurally homologous to *E. coli* SD dinucleotide-binding domain [20].

Stereospecificity and Rate-limiting Steps - The primary deuterium kinetic isotope effects indicate that the C_4-*proS* hydride (B side) is transferred to DHS in the oxy-reduction reaction catalyzed by *Mtb*SD. In contrast, *E. coli* SD has been shown to transfer the C_4-*proR* hydride (A-side) of NADPH [26], in agreement with the crystal structure of the enzyme [20]. Isotope effects on *V* report on events following formation of the ternary complex capable of undergoing catalysis, which include the chemical steps, possible enzyme conformational changes, and product release [27]. Isotope effects on *V/K* report on steps in the reaction mechanism from the binding of the isotopically labeled substrate to the first irreversible step, usually the release of the first product. The apparent classical limit for primary deuterium kinetic isotope effects on the maximal velocity is around 8 if carbon-hydrogen bond cleavage is the rate-determining step, even though values as small as 2 have been used, in a less rigorous practice, as evidence for rate-limiting steps [27]. The primary

kinetic isotope effect values of 1.8 for ${}^{D}V_{DHS}$ and 1.5 for ${}^{D}V_{NADPH}$ using [4*S*-^{2}H]NADPH as reductant indicate that the hydride transfer is only partly rate limiting for *Mtb*SD enzyme catalysis. Isotope-insensitive steps such as enzyme isomerization and/or product release are probably contributing to the rate-limiting steps of the *Mtb*SD reaction. The ${}^{D}(V/K_{app})_{NADPH}$ and ${}^{D}(V/K_{app})_{DHS}$ values indicate that substrate binding steps make a small contribution to the rate-limiting steps.

Steady-state solvent isotope effects were evaluated to assess the contribution of solvent proton transfer to a step in the enzymatic mechanism. The effect on *V* arises from solvent-exchangeable protons being transferred during catalysis. The values for solvent isotope effect on *V* for DHS (1.5) and NADPH (1.3) indicate that solvent-exchangeable protons being transferred during catalysis is only partly rate limiting in the overall reaction. No solvent isotope effect on *V*/*K* was observed for DHS and NADPH.

The number of protons transferred during the oxy-reduction reaction catalyzed by *Mtb*SD was determined by the proton inventory technique. Measurements of *V* in different isotopic solvent mixtures (*V* relative versus mol fraction of D_2O) showed a linear relationship (Fig. 4A – inset), suggesting that a single proton is transferred in the step that exhibits the solvent isotope effect [28]. A similar result was observed for the *M. tuberculosis* NADPH-dependent mycothione reductase [29].

Chemical Mechanism of MtbSD - Double isotope effect studies are able to distinguish whether two different isotopic substitutions affect the same or different chemical steps. Assuming that the primary deuterium kinetic isotope effect is expressed only on the hydride transfer reaction (Fig. 5A and 5B – insets) and solvent isotope effects affect only the alkoxide protonation, double isotope effects can distinguish whether these

isotopic substitutions affect the same or different chemical steps (Fig. 5A and 5B). Theory predicts that if protonation and hydride transfer occur in the same transition state, the primary isotope effects will be larger or unchanged with D_2O as compared to H_2O. On the other hand, if hydride transfer and protonation occur in distinct steps, the primary isotope effects will be smaller with D_2O as solvent, as proton transfer will become more rate limiting [30, 31]. The increased values for the primary isotope effects on *V/K* measured in D_2O ($^{D}V/K_{DHS(D_2O)} = 1.3$ and $^{D}V/K_{NADPH(D_2O)} = 1.5$) as compared to values measured in H_2O ($^{D}V/K_{DHS} = 1.0$ and $^{D}V/K_{NADPH} = 1.4$) are in agreement with both hydride and proton transfer taking place in the same step (concerted mechanism), thus implying a single transition state. It should be pointed out that use of D_2O allowed the observation of a significant value for $^{D}V/K_{DHS}$ (1.3), even though this substrate is sticky and reaction proceeds through a steady-state ordered kinetic mechanism.

The β-elimination catalyzed by bovine liver crotonase is concerted [32], as well as are the reactions catalyzed by pig liver acyl-CoA dehydrogenase [33] and isocitrate dehydrogenase [34]. Concerted reactions are a common strategy utilized by enzymes to avoid unstable intermediates, such as some enolate anions [35, 36].

The role of acid/base chemistry in the mechanism of *Mtb*SD was determined using the pH dependence of k_{cat} and k_{cat}/K_m for DHS and NADPH in pH range of $5.5 \leq pH \leq 10.0$ (Fig. 6). In this experiment the k_{cat} for both substrates are very similar and decrease at high pHs with a slope of -1 (Fig. 6A), indicating that deprotonation of a single residue with apparent pK_a value of 8.9 ± 0.1 abolishes *Mtb*SD catalytic activity. The k_{cat}/K_m profiles for DHS and NADPH show that deprotonation of a single ionizable group with pK_a value of 9.1 ± 0.1 abolishes DHS and NADPH binding (Fig. 6B and 6C). It is likely that the

ionization behavior of the same group is being observed in the pH profiles for k_{cat} and k_{cat}/K_m, and the difference in the *pK* values may reflect perturbation of the *pK* value upon substrate(s) binding to *Mtb*SD. The *pK* values for k_{cat} and k_{cat}/K_m lie in the normal *pK* range for ε-amino of lysine, thiol of cysteines and phenolic hydroxyl of tyrosine amino acids. The pH-rate profiles of *H. influenzae* SD for the reverse reaction indicated that a group with *pK* value of ~ 8.1 needs to be deprotonated for activity, and site-directed mutagenesis results were consistent with Lys-67 playing a key role in catalysis [37].

Based on the double isotope effects and pH-rate profiles, we propose a chemical mechanism for *Mtb*SD (Fig. 10), in which hydride transfer and solvent proton transfer are concerted, and an amino acid residue with pK_a value of 8.9 is involved in catalysis.

Analysis of the structure of *E. coli* SD shows that DHS is positioned for stereospecific reduction to SHK and the side chain of Lys65 has been proposed to hydrogen bond to the C-4 hydroxyl group of DHS/SHK (*E.coli* SD numbering) [20]. This amino acid residue has been proposed to be the acid/base catalytic group that donates a proton to the carbonyl of DHS during reduction and that removes a proton during oxidation of SHK [20]. In agreement, previous studies of *P. sativum* SD showed that substrate-like inhibitors of the enzyme require a C-4 hydroxyl, whereas either a C-5 hydroxyl or carboxylate group is needed for strong binding [38]. Analogs of DHS substrate that lack the C-4 and C-5 hydroxyls were used to demonstrate the role of both C-5 hydroxyl and C-4 hydroxyl on the substrate specificity for the *E. coli* SD, and the results showed that the C-4 hydroxyl has a very significant effect on the specificity of the substrate [39]. It has been suggested that the C-4 hydroxyl group hydrogen bonds to a charged group in the *E. coli* SD active site based on an estimation of the binding energy given by k_{cat}/K_m [39]. However, it should be kept in

mind that participation of a lysine side chain residue in the chemical reaction catalyzed by *E. coli* SD is based on a molecular model, since the crystal structure is for *E. coli* SD-$NADP^+$ binary complex, and *E. coli* SD transfers the C_4-*proR* hydride (A-side) of NADPH. Based on pH rate profiles of *P. sativum* SD, it has been postulated that a group with pK_a value of 9.4, possibly or an ε-amino group of lysine, binds the carboxylate ion of the substrate, while a group of pK_a 8.6, possibly a sulphydryl residue, interacts with the C-4 hydroxyl group of DHS/SHK [40]. The crystal structure of *M. jannaschii* SD has been determined and the residues involved in DHS binding and its catalytic reduction were identified, amongst them Lys70 (*M. jannaschii* SD numbering) [22]. The side chain of Asp102 of *E. coli* SD has also been suggested to form a hydrogen bond to C-4 hydroxyl group of the substrate [20]. In addition, analysis of the crystal structure of *H. influenzae* SD and site-directed mutagenesis studies have shown that Asp103 and Lys67 (*H. influenzae* SD numbering) may function as a catalytic pair involved in acid/base catalysis for this enzyme [37]. However, the pH-rate profiles for *Mtb*SD did not show any residue whose protonation would abolish enzyme activity at low pH values. However, it could be argued that the lowest pH value (5.0) of the pH-rate profiles presented here would not allow detection of carboxylic groups of aspartate side chains since its *pK* value is approximately 4. Notwithstanding, the *pK* for the carboxylate group of DHQ/SHK is approximately 4.1 and would be a difficult task to discriminate between substrate and/or amino acid ionizing groups. We have previously shown by analysis of multiple sequence alignment of *Mtb*SD, *E. coli* SD, *H. influenzae* SD-like, and *M. jannaschii* SD that Lys69 and Asp105 (*M. tuberculosis* numbering) are conserved [10]. The pH-rate profiles described here show participation of a group whose deprotonation abolishes binding (pK_a = 9.1) and catalytic

activity (pK_a = 8.9), consistent with participation of Lys69. At any rate, site directed mutagenesis of these residues and measurements of the steady-state kinetic parameters for *Mtb*SD enzyme reaction are currently underway to evaluate the role, if any, of Lys69 and Asp105 may play in binding and/or catalysis.

MtbSD Structure - The *Mtb*SD structure belongs to the homologous superfamily denominated NAD(P)-binding Rossmann-fold domain [41]. The SDH family provides a new example of a protein family displaying the dinucleotide binding-fold, without significant sequence homology with other Rossmann-fold families [20]. The structure has an elongated shape with two domains (Fig. 8). Each domain has an α/β architecture. The domains are linked by an α-helix and a turn that keep them together and form a deep groove into which the NADP cofactor binds. The NADP-binding domain, in the C-terminus, adopts a nearly canonical Rossmann fold, a six-stranded parallel β-sheet with six α-helices, three on each side of the β-sheet. The N-terminus consists of a six-stranded β-sheet and six α-helices, however the arrangement is irregular, since the near central β5 strand is in an antiparallel orientation with respect to the other strands [20].

In the analysis of the *E. coli* SD structure utilized as template, the substrate-binding site was identified by the position of the nicotinamide ring of the cofactor [20]. The following residues conserved in the SD family were identified: Ser14, Ser16, Lys65, Asn86, Thr101, Asp102, and Gln244 [20]. Both carboxyl oxygens from DHS form two hydrogen bonds to the protein using probably the conserved serine residues at positions 14 and 16 [20]. These residues were identified in the *Mtb*SD model (Ser18, Ser20, Lys69, Asn90, Thr104, Asp105, and Gln243; *Mtb*SD numbering), and they are present in

equivalent positions as showed in Fig. 11. The conserved residues Ser14, Ser16 and Tyr215 have been suggested to be involved in substrate carboxylate binding for *E. coli* SD [20]. These residues were found in the *Mtb*SD model (Fig. 11) to be in equivalent positions and appear to play a role in binding of the carboxylate group of DHS/SKH.

In the multiple sequence alignment between the target (*Mtb*SD) and templates (*E. coli* YdiB and *E. coli* SD) we identified the sequence Gly124-Ser125-Gly126-Gly127-Thr128-Ala129 as *Mtb*SD sequence pattern (G [A,s,g] G G [A,t] [A,S,g]), corresponding to the diphosphate-binding loop, which is conserved in the entire SDH family [20] (Fig. 7). Interaction between the diphosphate-binding loop and the active site in the *E. coli* SD occurs through Gly129 and Ala130 residues, which correspond to, respectively, Thr128 and Ala129 in *Mtb*SD (Fig. 7). For the binding of the nicotinamide in *E. coli* SD it was verified that the amide group N-7 of the nicotinamide ring is hydrogen-bonded to the carbonyl group of Met213 and of the invariant Gly237 [20]. In the *Mtb*SD model, the Gly236 (the correspondent position for Gly237) is hydrogen-bonded to the amide group N-7. On the other hand, the *Mtb*SD model presents Ala213 residue in equivalent position of the *E. coli* Met213. Notwithstanding, an equivalent bond with the amide group N-7 of nicotinamide ring is maintained with *Mtb*SD Ala213. The *E. coli* SD Arg150 and Arg154 play a crucial role in adenine phosphate binding as they form an "electrostatic clamp" that sandwiches the phosphate substituent of $NADP^+$ [20]. In the *Mtb*SD model it was found that Arg149 and Lys153 residues occupy equivalent positions, and it is thus tempting to suggest that these residues play a role in adenine phosphate binding in *Mtb*SD.

It has been suggested that NAD^+ binding in *E. coli* YdiB is favored by the substitution of residues Thr151 and Arg154 of *E. coli* SD (*E. coli* SD numbering) by Asp158 and Phe160 (*E. coli* YdiB numbering), respectively [20]. The hydrophobic residue

Phe160 creates a neutral environment, which is less discriminating than the basic binding pocket with Arg154 from the SD structure. The Asn157 and Lys160 residues of $NADP^+$-dependent *Mtb*SD, which correspond to Thr151 and Arg154 of $NADP^+$-dependent *E. coli* SD [20], occupy equivalent positions and appear to play a similar role in creating a more discriminating environment in favor of $NADP^+$ substrate.

The work here presented is, to the best of our knowledge, the first detailed report on the steady-state velocity patterns, product inhibition, primary deuterium kinetic isotope effects, solvent kinetic isotope effects, proton inventory, double isotope effects, pH-rate profiles, and molecular modeling of *Mtb*SD. Initial velocity patterns in the forward reaction, product inhibition studies, and primary deuterium kinetic isotope effects allowed us to propose a steady-state ordered bi-bi kinetic mechanism for catalysis, with DHS binding first followed by NADPH binding to the *Mtb*SD enzyme catalytic site. The primary deuterium kinetic isotope effects indicated that the C_4-*proS* hydride is transferred from the NADPH co-substrate in a step that is only partly rate limiting. Solvent kinetic isotope effects demonstrated that proton transfer from the solvent is only partly rate limiting. Proton inventory results indicated that a single proton is transferred in the solvent-sensitive step. Double isotope effects showed that transfer of hydride and proton occur in concert. The pH-rate profiles revealed that a charged group, probably the ε-amino group of Lys69 plays an important role in catalysis and substrate binding. The homology 3D model for *Mtb*SD identified the probable residues in the substrate binding site, amongst them Lys69, in agreement with pH studies, and Asp105. These results allowed us to propose a kinetic and chemical mechanism for *Mtb*SD. The enzyme kinetics and modeling studies provide a framework on which to base the design of enzyme inhibitors with potential use as

antitubercular agents. Site-directed mutagenesis, equilibrium binding spectrofluorimetry, and pre-steady state experiments are currently underway to improve our understanding of the mechanism of action of *Mtb*SD.

Acknowledgements

Financial support for this work was provided by Millennium Initiative Program MCT-CNPq, Ministry of Health-Department of Science and Technology (DECIT)-UNESCO (Brazil) to D.S.S. and L.A.B. D.S.S. and L.A.B. also acknowledge grants awarded by CNPq, FINEP, and PRONEX/FAPERGS/CNPq. D.S.S. (CNPq, 304051/1975-06) and L.A.B. (CNPq, 520182/99-5) are researchers awardees from the National Research Council of Brazil. We thank Professor John W. Frost, Department of Chemistry of Michigan State University, for his generous gift of 3-dehydroshikimate substrate.

References

1. C. Dye, S. Scheele, P. Dolin, V. Pathania, M. C. Raviglione, JAMA 282 (1999) 677-686.
2. L. A. Basso, J. S. Blanchard, Adv. Exp. Med. Biol. 456 (1998) 115-144.
3. A. Plabos-Méndez, D.K. Gowda, T.R. Frieden, World Health Organ. 80 (2002) 489-495.
4. CDC (Centers for Disease Control and Prevention), Morb. Mortal. Wkly. Rep. 55 (2006) 301-305.
5. R. J. O'Brien, P. P. Nunn, Am. J. Respir. Crit. Care Med. 162 (2001) 1055-1958.
6. S. T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry 3rd, F. Tekaia, K. Badcock, D. Basham, D. Brown,

T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M. A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead, B. G. Barrell, Nature 393 (1998) 537-544.

7. T. Parish, N. G. Stoker, Microbiology 148 (2002) 3069-3077.
8. J.R. Coggins, C. Abell, L.B. Evans, M. Frederickson, D.A. Robinson, A.W. Roszak, A.P. Lapthorn, Biochem. Soc. Trans. 31 (2003) 548-552.
9. M. L.Magalhães, C. P. Pereira, L. A. Basso, D. S. Santos, Protein Expr. Purif. 26 (2002) 59-64.
10. I. O. Fonseca, M. L. B Magalhães, J. S. Oliveira, R. G. Silva, M. A.Mendes, M. S. Palma, D. S. Santos, L. A. Basso, Protein Expr. Purif. 46 (2006) 429-437.
11. G. Ottolina, S. Riva, G. Carrea, B. Danieli, A. F. Buckmann, Biochem. Biophys. Acta 998, (1989) 173-178.
12. G. A.Orr, J. S. Blanchard, Anal. Biochem. 142, (1984) 232-234.
13. H.M. Berman, J. Westbrook, Z. Feng, , G. Gilliland, T.N. Bhat, H. Weissig, I.N.Shindyalov, P.E. Bourne, Nucleic Acids Res. 28, (2000) 235-242.
14. S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Nucleic Acids Res. 25 (1997) 3389-3402.
15. J.D. Thompson, D.G. Higgins, T.J. Gibson, Nucleic Acids Res. 22 (1994) 4673-4680
16. A Šali, T.L. Blundell, J. Mol. Biol. 234 (1993) 779-815.
17. M.A. Martí-Renom, A.C. Stuart, A. Fiser, R. Sánchez, F. Melo, A. Šali, Annu. Rev. Biophys. Biomol. Struct. 29 (2000) 291-325.

18. R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, J. Appl. Cryst. 26 (1993) 283-291.

19. R. Lüthy, J.U. Bowie, D. Eisenberg, Nature 356 (1992) 83-85.

20. G. Michel, A. W. Roszak, V. Sauvé, J. Maclean, A. Matte, J. R. Coggins, M. Cygler, A. J. Lapthorn, J. Biol. Chem. 278 (2003) 19463-19472.

21. N. Guex, M. C. Peitsch, Electrophoresis 18 (1997) 2714-2723.

22. A.K Padyana, S.K. Burley, Structure 11 (2003) 1005-1013.

23. D. Balinsky, A. W. Dennis, W. W. Cleland, Biochemistry 10 (1971) 1947-1952.

24. P. F. Cook, W. W. Cleland, Biochemistry 20 (1981) 1790-1796.

25. P. Strömberg, S. Svensson, K. B. Berst, B. V. Plapp, J. O. Höög, Biochemistry 43 (2004) 1323-1328.

26. P. Dansette, R. Azerad, Biochimie 56 (1974) 751-755.

27. D. B. Northrop, Biochemistry 14 (1975) 2644-2651.

28. D. M. Quinn, L. D. Sutton, in: P. F. Cook (Ed.), Enzyme Mechanism from Solvent Isotope Effects, CRC Press, Florida, 1991, pp 73-126.

29. M. P. Patel, J. S. Blanchard, Biochemistry 40 (2001) 5119-5126.

30. J. D. Hermes, C. A. Roeske, M. H. O'Leary, W. W. Cleland, Biochemistry 21 (1984) 5106-5114.

31. J. G. Belasco, J. Albery, J. R. Knowles, J. Am. Chem. Soc. 105 (1983) 2475-2477.

32. Gerlt, J. A. Bioorganic Chemistry: Peptides and Proteins, Oxford University Press, New York, 1998.

33. B. J. Bahnson, V. E. Anderson, Biochemistry 30 (1991) 5894-5906.

34. B. Pohl, T. Raichle, S. Ghisla, Eur. J. Biochem. 160 (1986) 109-115.

35. P. F. Cook, W. W. Cleland, Biochemistry 20 (1981) 1797-1805.

36. A. Thibblin, W. P. Jencks, J. Am. Chem. Soc. 101 (1979) 4963-4973.

37. S. Singh, S. Korolev, O. Koroleva, T. Zarembinski, F. Collart, A. Joachimiak, D. Christendat, J. Biol. Chem. 280 (2005) 17101-17108.

38. D. Balinsky, D. D. Davies, Biochem. J. 80 (1961) 296-300.

39. T. D. Bugg, C. Abell, J. R. Coggins, Tetrahedron Lett. 29 (1988) 6779-6782.

40. A. W. Dennis, D. Balinsky, Int. J. Biochem. 3 (1972) 93-102.

41. A. G. Murzin, S. E. Brenner, T. Hubbard, C. Chothia. J. Mol. Biol. 247 (1995) 536-540.

Figure Legends

Fig. 1. Shikimate dehydrogenase-catalyzed reaction.

Fig. 2. Initial velocity patterns for *Mtb*SD with both substrates DHS (***A***) and NADPH (***B***) as variable substrate. Each curve represents varied-fixed levels of the cosubstrate. Both DHS and NADPH concentrations varied from 5 to 200 μM. One unit of enzyme activity (U) is defined as the amount of enzyme catalyzing the conversion of 1 μmol of $NADP^+$ per minute at 25 °C.

Fig. 3. Dependence of the apparent ${}^{D}V/K_{DHS}$ values on the concentration of the cosubstrate. The data were fitted to an equation describing a hyperbolic decay, which yielded a limiting

value of 1.0 ± 0.03 for the $^{D}V/K_{DHS}$. The inset shows that $^{D}V/K_{NADPH}$ values do not depend on the concentration of the cosubstrate.

Fig. 4. Solvent isotope effects for *Mtb*SD. (**A**) DHS as varied substrate, with saturating concentration of the cosubstrate. Reaction mix contained either 0 (●) or 90 (▼) atom % D_2O. Inset represents the proton inventory on *Mtb*SD, with both substrates at saturating concentrations. (**B**) NADPH as varied substrate, with saturating concentration of the cosubstrate. Reaction mix contained either 0 (●) or 90 (▼) atom % D_2O.

Fig. 5. Multiple isotope effects for *Mtb*SD. (**A**) NADPD as varied substrate, with saturating concentration of DHS. Reaction mix contained either 0 (•) or 90 (▼) atom % D_2O. Inset represents the primary isotope effect, NADPH (•) or NADPD (▼) as varied substrate with saturating concentration of the DHS. (**B**) DHS as varied substrate, with saturating concentration of NADPD. Reaction mix contained either 0 (•) or 90 (▼) atom % D_2O. Inset represents the primary isotope effect, DHS as varied substrate with saturating concentration of NADPH (•) or NADPD (▼).

Fig. 6**.** Dependence of *Mtb*SD kinetics parameters on pH. (**A**) pH dependence of log k_{cat}, (**B**) pH dependence of log k_{cat}/K_{DHS}, and (**C**) pH dependence of log k_{cat}/K_{NADPH}. Experimental data were fitted to equation 7.

Fig. 7. ClustalW multiple sequence alignment between the target (*Mtb*SD) and templates (*E. coli* YdiB and *E. coli* SD). YdiB is more similar to *Mtb*SD than to *E. coli* SD and optimal alignment between the three sequences requires the insertion of two gap regions in

the C-terminus of *Mtb*SD and *E. coli* SD sequences (underlined in YdiB). α-helices and β-strands are represented as **H** and **S**, respectively. This sequence alignment was created using the following sequences from GeneBank^{TM}: *M. tuberculosis* SD (CAB06186, residues 1–269), *E. coli* K12 YdiB (NP 416207, residues 1-288), and *E. coli* K12 SD (NP_417740, residues 1-272).

Fig. 8. Stereo ribbon representation of the *Mtb*SD structure. The *Mtb*SD structure is characterized by a nearly canonical Rossmann fold with a six-stranded parallel β sheet (C-terminal domain on top), and a α/β N-terminal domain (bottom) with an antiparallel β-strand (β5). The α-helices are shown in light gray and the β-strands in black.

Fig. 9. Proposed kinetic mechanism for *M. tuberculosis* shikimate dehydrogenase.

Fig. 10. Proposed chemical mechanism for *M. tuberculosis* shikimate dehydrogenase-catalyzed reaction. **R** represents ribose, adenosine diphosphate, and 2'-phosphate moieties of NADPH.

Fig. 11. The conserved residues in the nicotinamine ring pocket current in the structures of both *E. coli* SD (black) and in the *Mtb*SD (gray). The tyrosine and serine residues involved in hydrogen-bond with the DHS carboxylate group are in equivalent positions in both *E. coli* SD (Tyr215, Ser14, and Ser16; black) and in the *Mtb*SD (Tyr215, Ser18, and Ser20; gray).

Table1: Product Inhibition Patterns for *M. tuberculosis* Shikimate Dehydrogenase [a]

Varied substrate	Product inhibition	Inhibition type [b]	K_{is} (μM) [c]	K_{ii} (μM) [d]
DHS	SHK	C	62.9 ± 10.5	
NADPH	SHK	NC	256.3 ± 35.4	142.5 ± 74.0
DHS	$NADP^+$	NC	30.4 ± 1.3	28.3 ± 7.6
NADPH	$NADP^+$	C	8.3 ± 1.3	

[a] At 25°C and 100 mM potassium phosphate, pH 7.3. [b] C = competitive, NC-MT = noncompetitive mixed-type. [c] K_{is} is the slope inhibition constant. [d] K_{ii} is the intercept inhibition constant.

Table 2: Kinetic Isotope Effects for *M. tuberculosis* Shikimate Dehydrogenase

Parameter	Isotope Effect*
$^{D}V/K_{DHS}$	1.0 ± 0.03
$^{D}V_{DHS}$	1.8 ± 0.09
$^{D}V/K_{NADPH}$	1.4 ± 0.3
$^{D}V_{NADPH}$	1.5 ± 0.1
$^{D_2O}V/K_{DHS}$	1.0 ± 0.01
$^{D_2O}V_{DHS}$	1.5 ± 0.3
$^{D_2O}V/K_{NADPH}$	1.0 ± 0.03
$^{D_2O}V_{NADPH}$	1.3 ± 0.2
$^{D}V/K_{DHS(D_2O)}$	1.3 ± 0.1
$^{D}V_{DHS(D_2O)}$	2.6 ± 0.01
$^{D}V/K_{NADPH(D_2O)}$	1.5 ± 0.1
$^{D}V_{NADPH(D_2O)}$	2.5 ± 0.1

* value ± standard error obtained from fitting the data to the appropriated equation

[1]**Abbreviations used:** DHS, 3-dehydroshikimate; MDR-TB, multidrug-resistant tuberculosis; *Mtb*SD, *Mycobacterium tuberculosis* shikimate dehydrogenase; $NADP^+$, oxidized β-nicotinamide adenine dinucleotide phosphate; NADPH, reduced β-nicotinamide adenine dinucleotide phosphate; NADPD, deuterated β-nicotinamide adenine dinucleotide phosphate; SD, shikimate dehydrogenase; SHK, D-shikimate; TB, tuberculosis; XDR-TB, extensively drug-resistant tuberculosis.

Figure 1

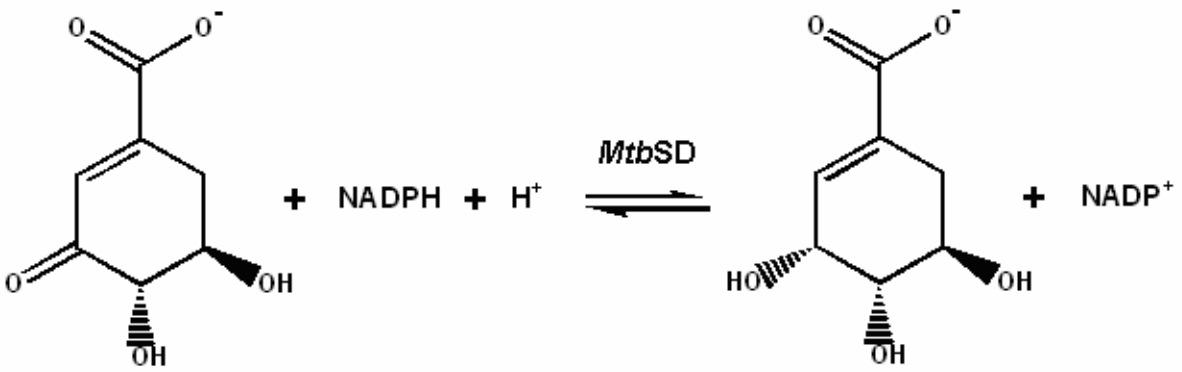


Figure 2A

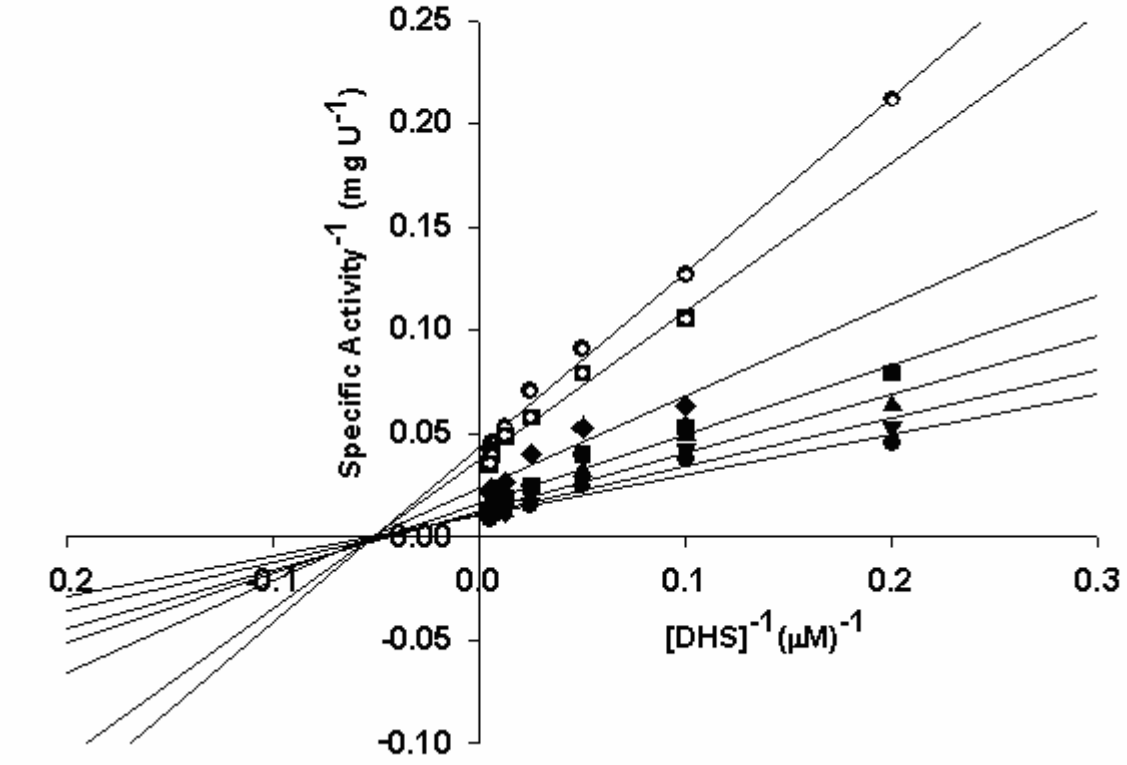

0.25
0.20
0.15
0.10
0.05
0.00
-0.05
-0.10
Specific Activity-1 (mg U-1)
0.2
0.1
0.0
0.1
0.2
0.3
[DHS]-1 (µM)-1

Figure 2B

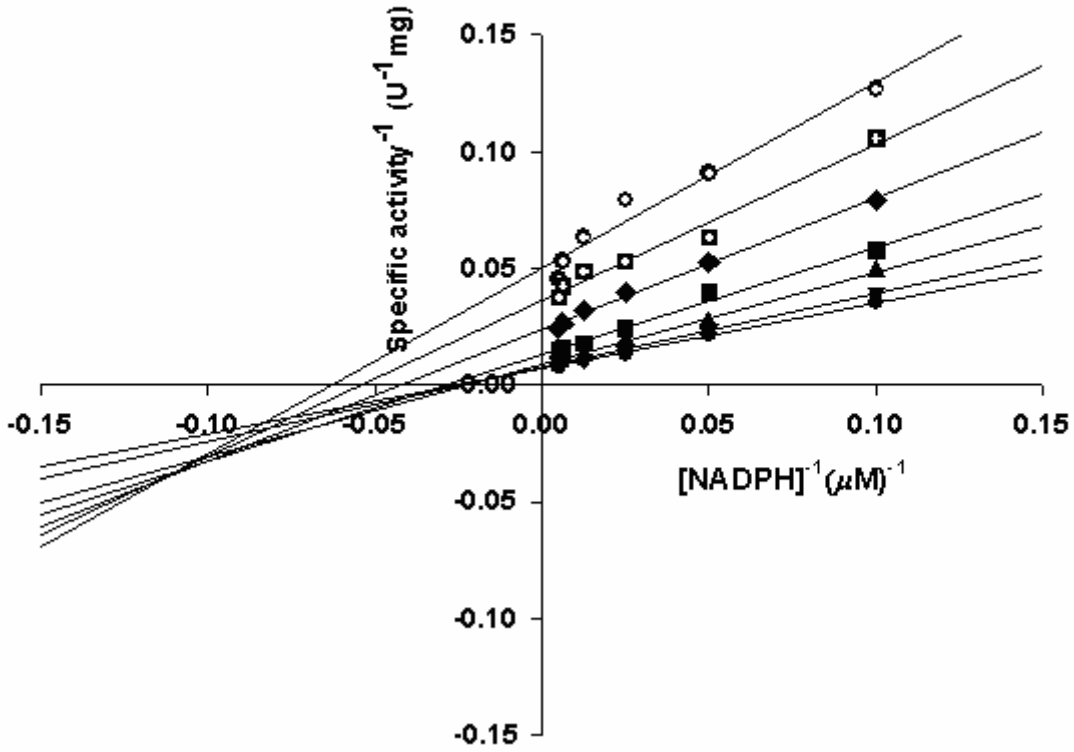

Specific activity^{-1} (U^{-1}mg)
0.15
0.10
0.05
0.00
-0.05
-0.10
-0.15
-0.15
-0.10
-0.05
0.00
0.05
0.10
0.15
[NADPH]$^{-1}$ (μM)$^{-1}$

Figure 3

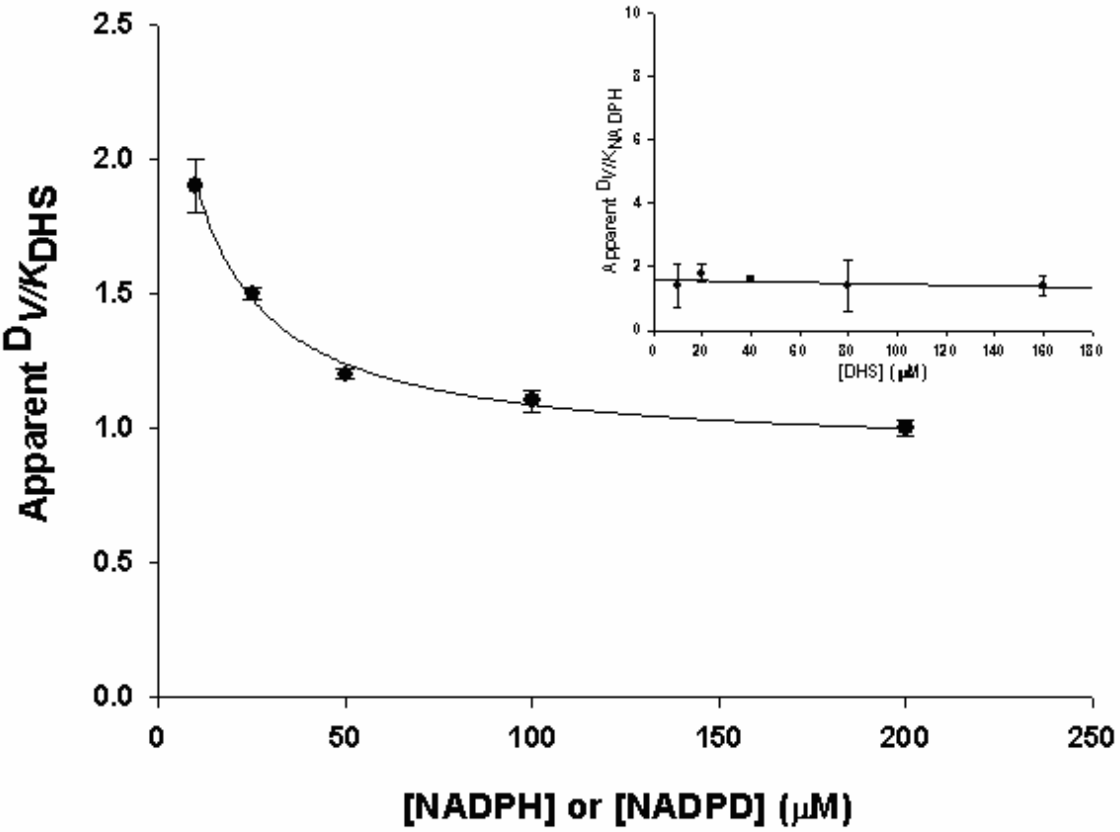

Apparent $^{D}V/K_{DHS}$
2.5
2.0
1.5
1.0
0.5
0.0
0
50
100
150
200
250
[NADPH] or [NADPD] (μM)
Apparent $^{D}V/K_{NADPH}$
10
8
6
4
2
0
0
20
40
60
80
100
120
140
160
180
[DHS] (μM)

Figure 4A

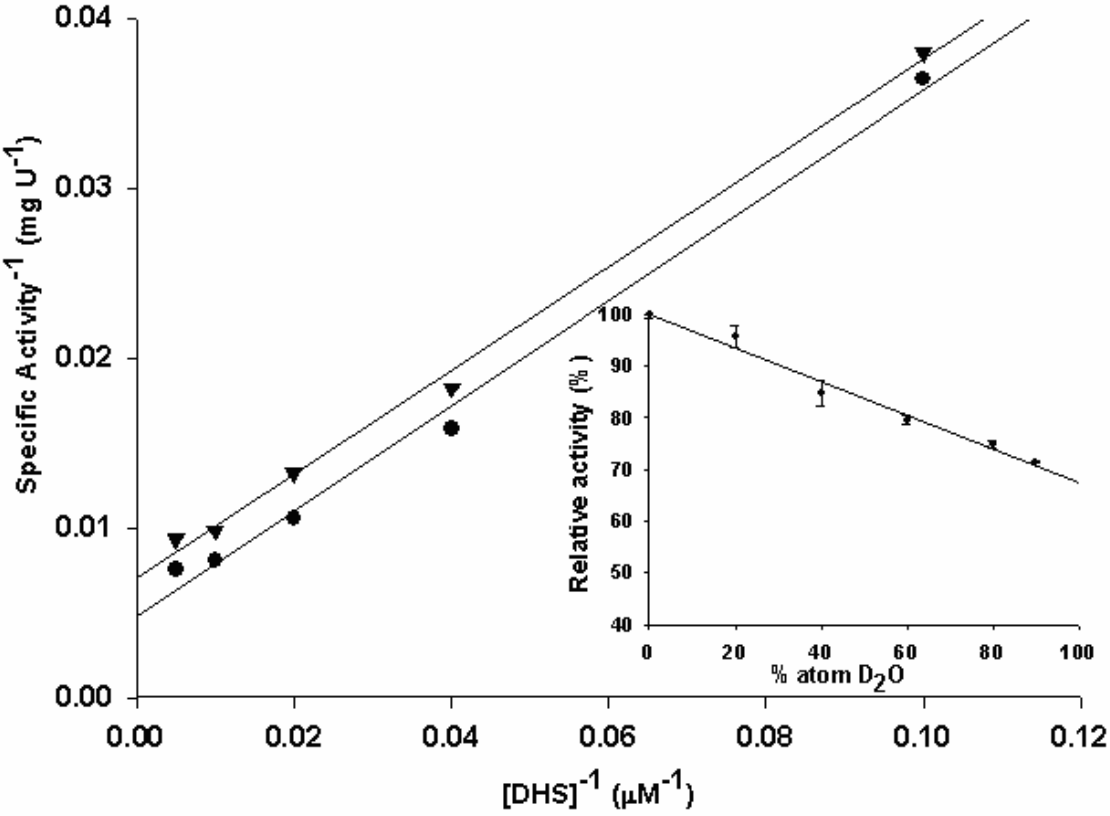

Specific Activity^{-1} (mg U^{-1})
0.04
0.03
0.02
0.01
0.00
0.00
0.02
0.04
0.06
0.08
0.10
0.12
[DHS]$^{-1}$ (µM^{-1})
Relative activity (%)
100
90
80
70
60
50
40
0
20
40
60
80
100
% atom D_2O

Figure 4B

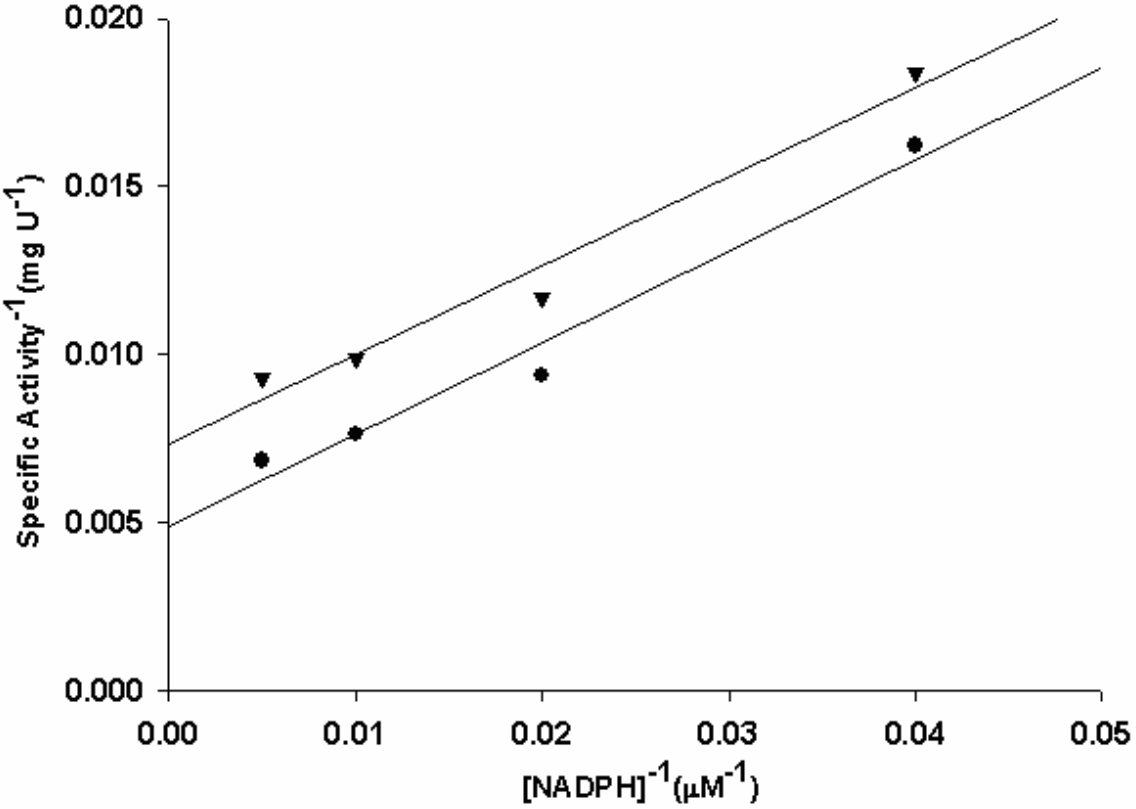

0.020
0.015
0.010
0.005
0.000
Specific Activity-1(mg U-1)
0.00
0.01
0.02
0.03
0.04
0.05
[NADPH]-1(µM-1)

Figure 5A

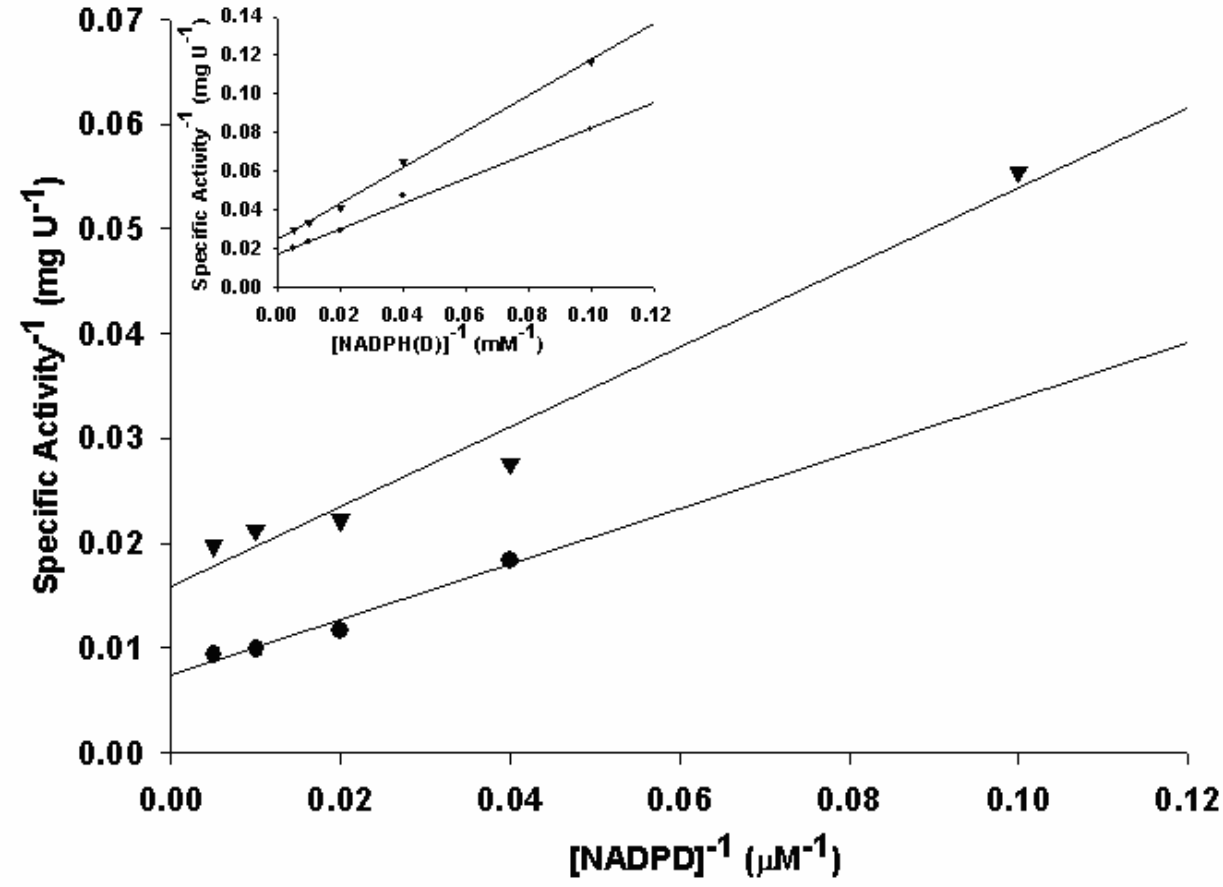

Specific Activity^{-1} (mg U^{-1})
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.00
0.02
0.04
0.06
0.08
0.10
0.12
[NADPD]$^{-1}$ (μM^{-1})
Specific Activity^{-1} (mg U^{-1})
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.00
0.02
0.04
0.06
0.08
0.10
0.12
[NADPH(D)]$^{-1}$ (mM^{-1})

Figure 5B

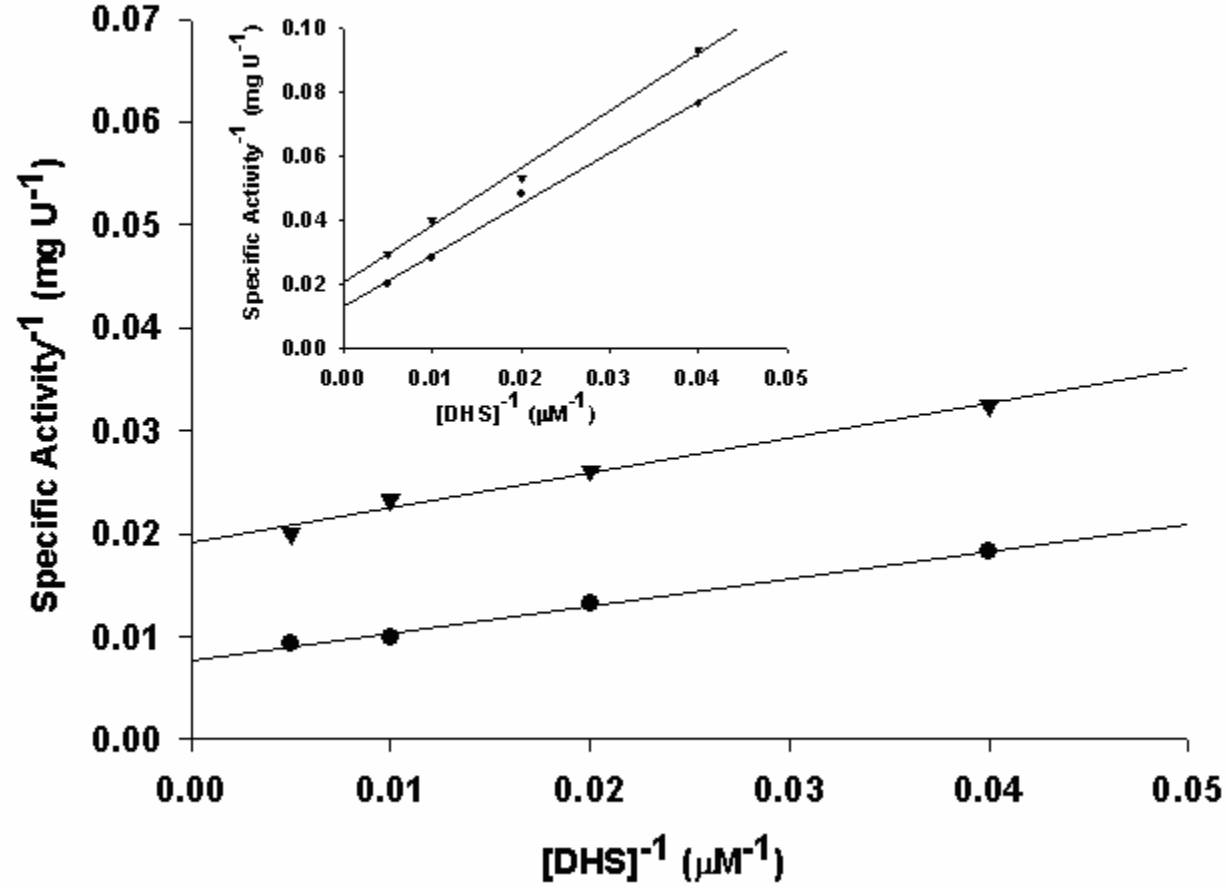

Specific Activity^{-1} (mg U^{-1})
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0.00
0.01
0.02
0.03
0.04
0.05
[DHS]$^{-1}$ (μM^{-1})
Specific Activity^{-1} (mg U^{-1})
0.10
0.08
0.06
0.04
0.02
0.00
0.00
0.01
0.02
0.03
0.04
0.05
[DHS]$^{-1}$ (μM^{-1})

Figure 6

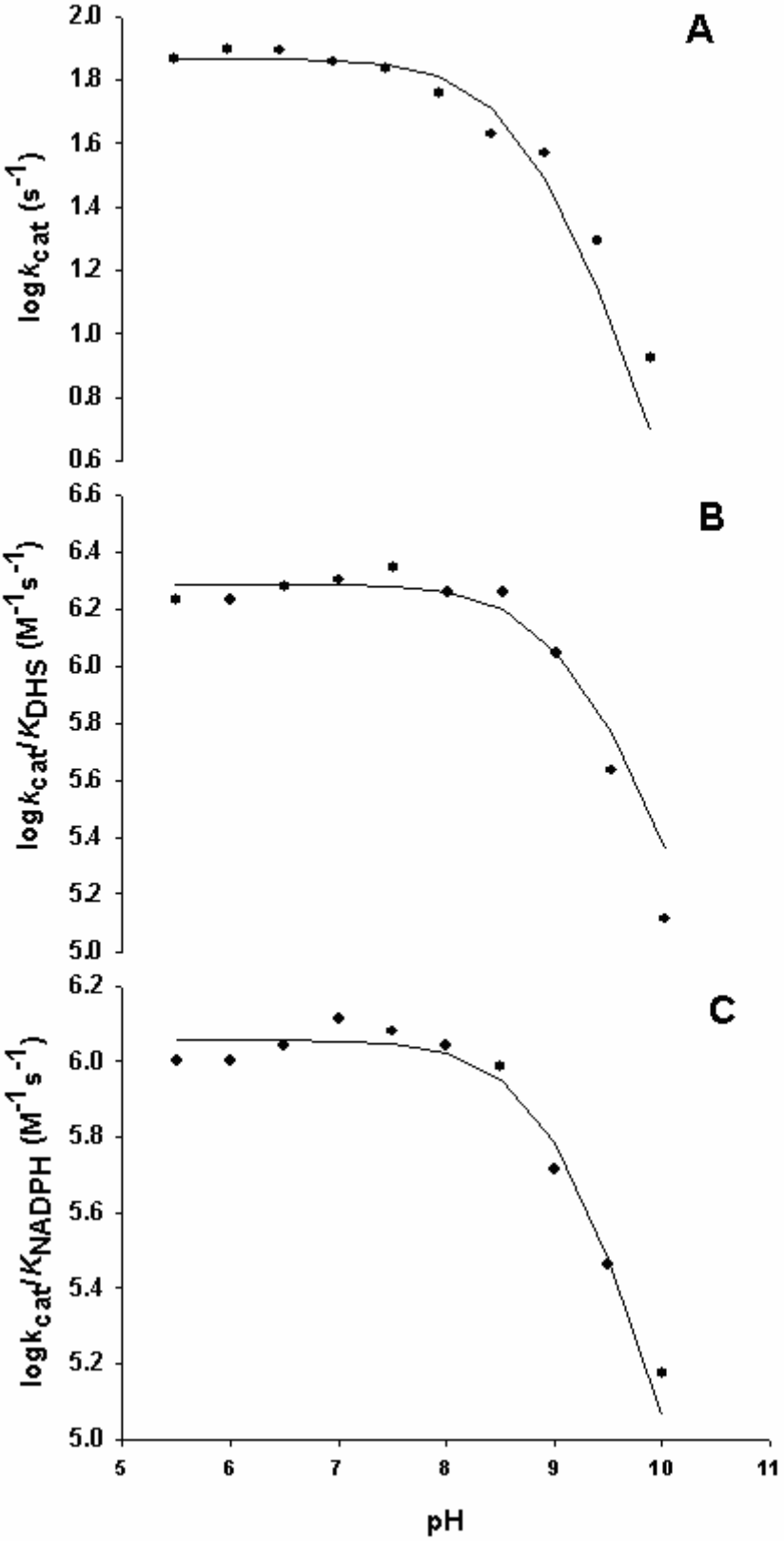


Figure 7

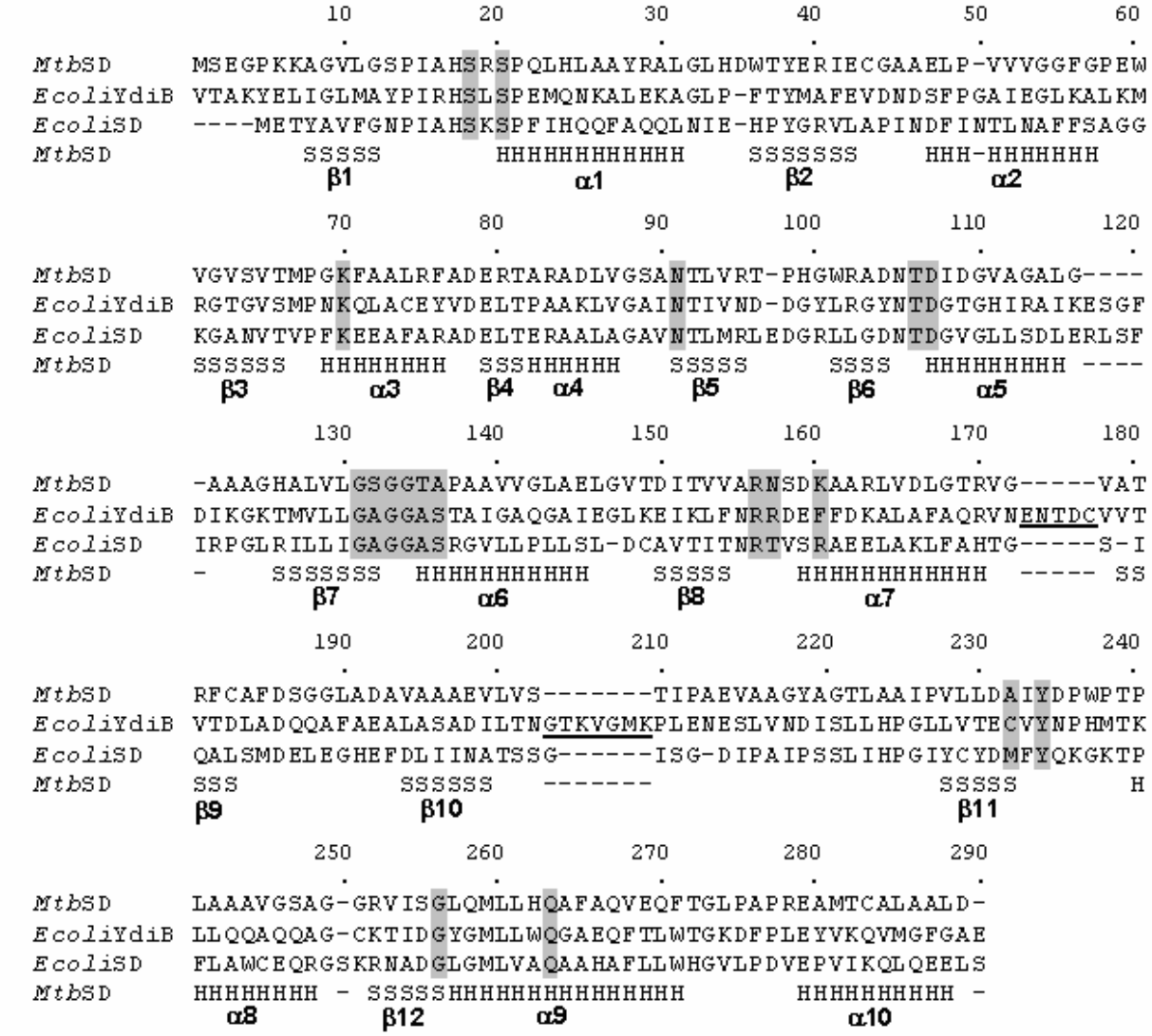

10 20 30 40 50 60
MtbSD MSEGPKKAGVLGSPIAHSRSPQLHLAAYRALGLHDWTYERIECGAAELP-VVVGGFGPEW
EcoliYdiB VTAKYELIGLMAYPIRHSLSPEMQNKALEKAGLP-FTYMAFEVDNDSFPGAIEGLKALKM
EcoliSD ----METYAVFGNPIAHSKSPFIHQQFAQQLNIE-HPYGRVLAPINDFINTLNAFFSAGG
MtbSD SSSSS HHHHHHHHHHHH SSSSSSS HHH-HHHHHHH
β1 α1 β2 α2
70 80 90 100 110 120
MtbSD VGVSVTMPGKFAALRFADERTARADLVGSANTLVRT-PHGWRADNTDIDGVAGALG----
EcoliYdiB RGTGVSMPNKQLACEYVDELTPAAKLVGAINTIVND-DGYLRGYNTDGTGHIRAIKESGF
EcoliSD KGANVTVPFKEEAFARADELTERAALAGAVNTLMRLEDGRLLGDNTDGVGLLSDLERLSF
MtbSD SSSSSS HHHHHHHH SSSHHHHHH SSSSS SSSS HHHHHHHHH ----
β3 α3 β4 α4 β5 β6 α5
130 140 150 160 170 180
MtbSD -AAAGHALVLGSGGTAPAAVVGLAELGVTDITVVARNSDKAARLVDLGTRVG-----VAT
EcoliYdiB DIKGKTMVLLGAGGASTAIGAQGAIEGLKEIKLFNRRDEFFDKALAFAQRVNENTDCVVT
EcoliSD IRPGLRILLIGAGGASRGVLLPLLSL-DCAVTITNRTVSRAEELAKLFAHTG-----S-I
MtbSD - SSSSSSS HHHHHHHHHHH SSSSS HHHHHHHHHHHH ----- SS
β7 α6 β8 α7
190 200 210 220 230 240
MtbSD RFCAFDSGGLADAVAAAEVLVS-------TIPAEVAAGYAGTLAAIPVLLDAIYDPWPTP
EcoliYdiB VTDLADQQAFAEALASADILTNGTKVGMKPLENESLVNDISLLHPGLLVTECVYNPHMTK
EcoliSD QALSMDELEGHEFDLIINATSSG------ISG-DIPAIPSSLIHPGIYCYDMFYQKGKTP
MtbSD SSS SSSSSS ------- SSSSS H
β9 β10 β11
250 260 270 280 290
MtbSD LAAAVGSAG-GRVISGLQMLLHQAFAQVEQFTGLPAPREAMTCALAALD-
EcoliYdiB LLQQAQQAG-CKTIDGYGMLLWQGAEQFTLWTGKDFPLEYVKQVMGFGAE
EcoliSD FLAWCEQRGSKRNADGLGMLVAQAAHAFLLWHGVLPDVEPVIKQLQEELS
MtbSD HHHHHHHH - SSSSSHHHHHHHHHHHHHH HHHHHHHHH -
α8 β12 α9 α10

Figure 8

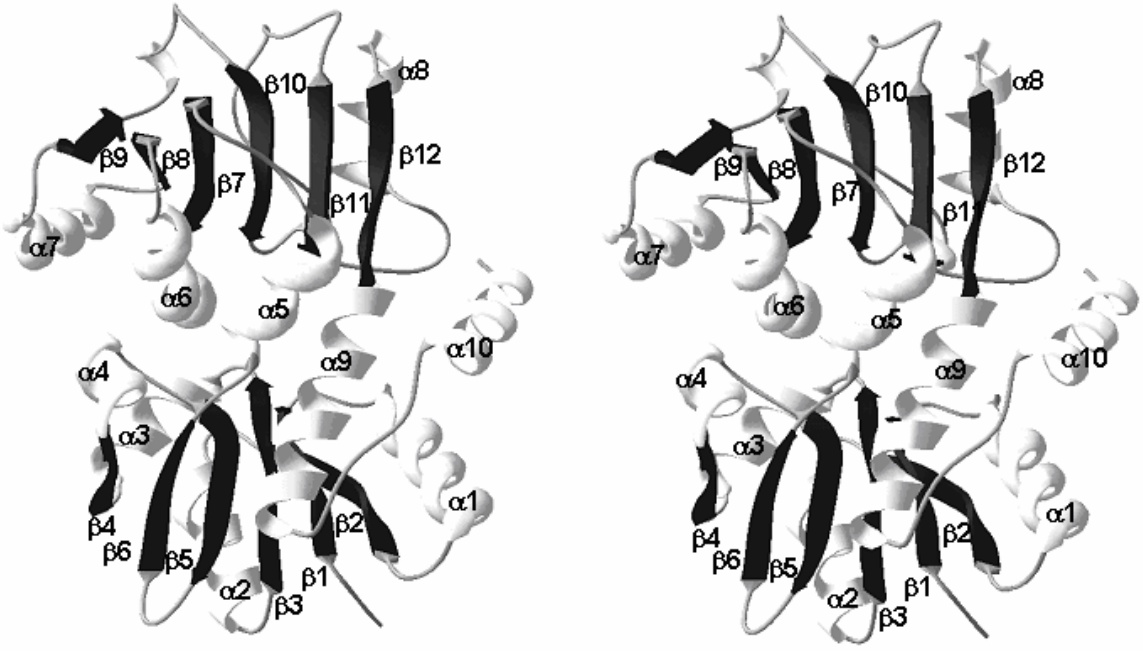

β10
α8
β9
β8
β12
β7
β11
α7
α6
α5
α10
α9
α4
α3
α1
β4
β2
β6
β5
β1
α2
β3
β10
α8
β9
β8
β12
β7
β1
α7
α6
α5
α10
α9
α4
α3
α1
β4
β2
β6
β5
β1
α2
β3

Figure 9

Figure 10

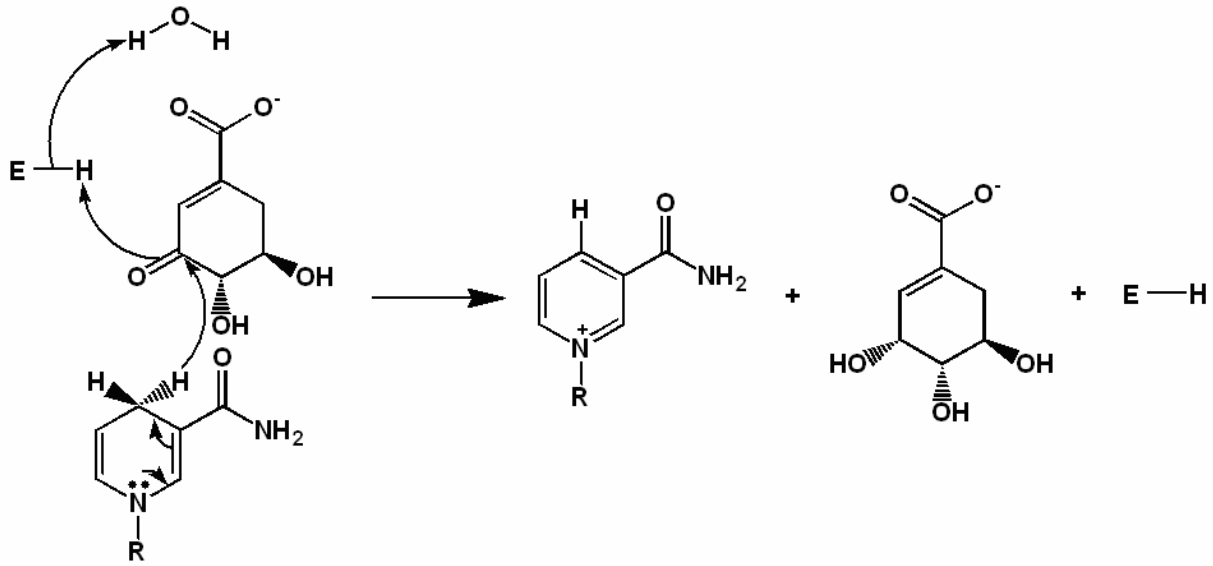
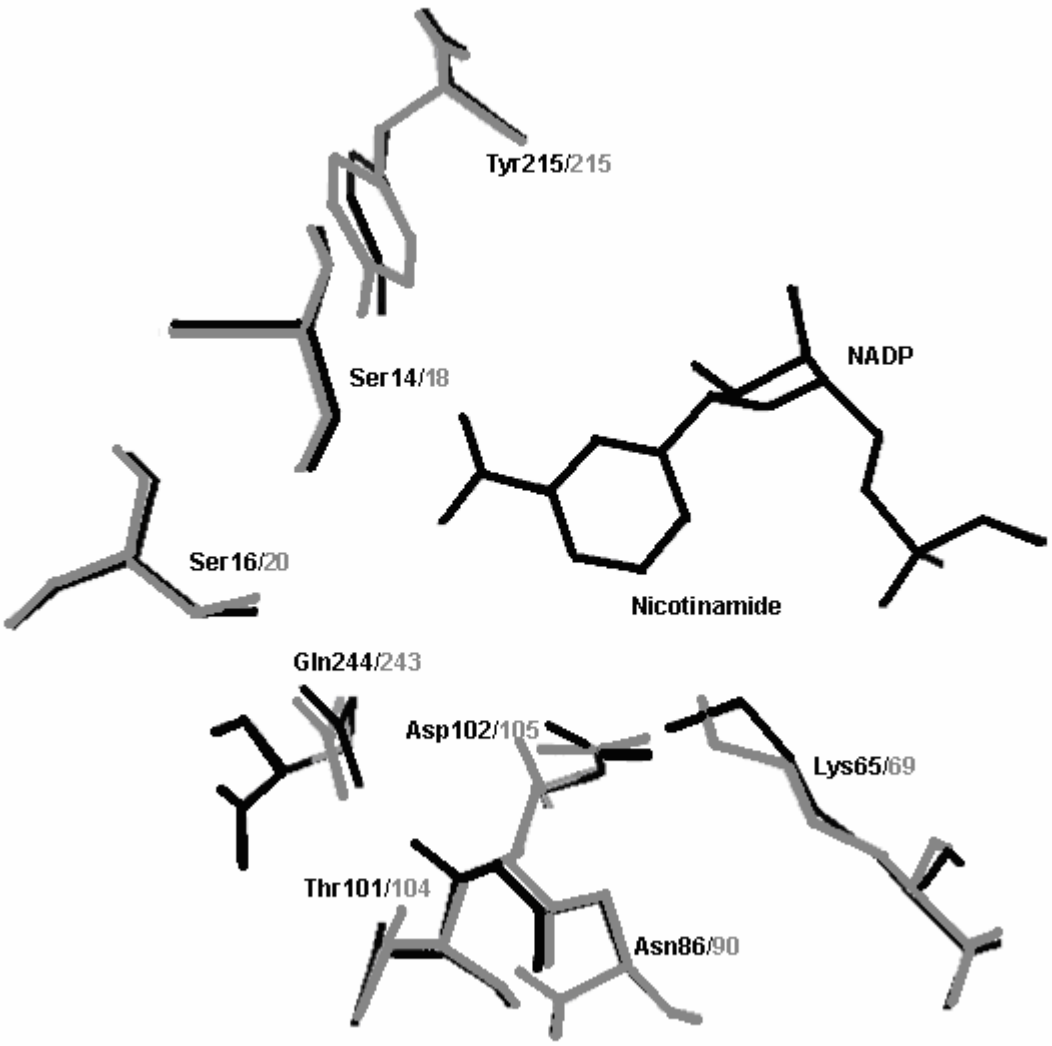


Figure 11


Tyr215/215
Ser14/18
NADP
Ser16/20
Nicotinamide
Gln244/243
Asp102/105
Lys65/69
Thr101/104
Asn86/90

PARTE III

Resultados e Discussão

ARTIGO 1: "*Functional shikimate dehydrogenase from Mycobacterium tuberculosis H37Rv: Purification and characterization*"

Obs.: As **figuras** e **tabelas em negrito**, mencionadas na discussão a seguir, estão identificadas de acordo com o artigo a que se referem.

Purificação e Análise da enzima recombinante MtbSD

No artigo intitulado "*Functional shikimate dehydrogenase from Mycobacterium tuberculosis H37Rv: Purification and characterization*", o protocolo de purificação para a enzima *Mtb*SD utiliza quatro etapas cromatográficas, onde a proteína recombinante *Mtb*SD foi purificada 8,5 vezes. Foram necessárias 49g de células *E. coli* BL21(DE3) com vetor de expressão pET23a(+) contendo o gene *aro*E de *M. tuberculosis* para obtenção de ~ 10 mg de *Mtb*SD pura. Esse protocolo de purificação e a análise da proteína recombinante pura foram feitos na dissertação de mestrado da Maria de Lourdes Borba Magalhães pelo Programa de Pós-Graduação em Biologia Molecular e Celular da Universidade Federal do Rio Grande do Sul.

Neste protocolo a primeira etapa cromatográfica, utilizando-se da *Q-Sepharose Fast Flow*, resultou na maior perda de todo o processo. Apesar disso, essa etapa foi interessante por possibilitar a redução da quantidade das proteínas totais em 90% e concentrar a proteína de interesse. Na etapa subseqüente utilizou-se uma coluna de interação hidrofóbica, a *Phenyl Sepharose High Performance*, que resultou em uma taxa de purificação de quatro vezes e um rendimento de 60% em relação à etapa anterior. A terceira

etapa foi feita em gel filtração, através da matriz *Sephacryl High Resolution S-200* e por fim, a coluna de troca aniônica, *Mono Q-HR*, foi necessária para o refinamento da purificação da *Mtb*SD (**Tabela 1 e Figura1**).

A proteína purificada foi ensaiada monitorando a absorbância (340 nm) da redução do cofator $NADP^+$ a NADPH por meio da oxidação do *D*-chiquimato (SHK) que leva a formação do 3-desidrochiquimato (DHS) (Figura 3). É possível verificar que a atividade enzimática da proteína purificada foi linearmente dependente do volume de proteína homogênea à mistura de reação, mostrando que a velocidade inicial é proporcional à concentração total de enzima (**Figura 2**).

A análise do produto da purificação, por espectrometria de massas em ESI-MS, mostra dois picos (Figura 4). Um pico em 27.207 Da que corresponde à massa molecular da seqüência primária prevista para a *Mtb*SD, com a perda da metionina N-terminal; e o segundo pico em 54.150 Da que corresponde à presença da forma dimérica da *Mtb*SD. O grau de pureza obtido da proteína pura foi de 98%. A ausência de um pico em 29.413 Da no espectro de massas confirma que a *Mtb*SD foi separada da enzima *Ecoli*SD (CHAUDURI & COGGINS, 1985), hospedeiro utilizado na expressão da proteína alvo.

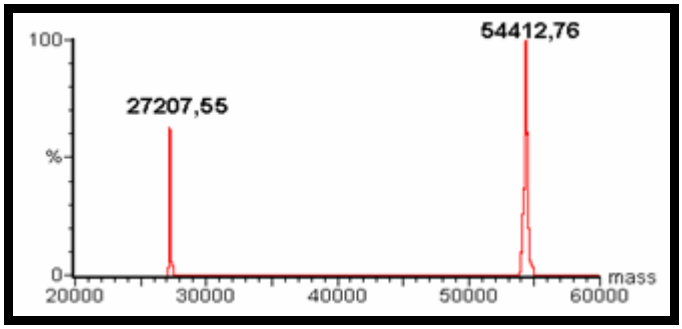


Figura 4: Espectro de massas da *Mtb*SD. O espectro mostra a massa molecular da enzima na forma monomérica e dimérica. O eixo de valores x representa a massa molecular e o eixo de valores y a intensidade do pico.

A comprovação da natureza dimérica da enzima em solução foi feita por meio da cromatografia líquida em gel filtração utilizando-se a coluna *Superdex-200* e uma curva de calibração (Figura 5). As SDs de *E. coli* (MICHEL e cols., 2003) e *H. influenzae* (YE e col., 2003) apresentam-se na forma monomérica em solução, em contrapartida a *Mjan*SD é dimérica (PADYANA & BURLEY, 2003).

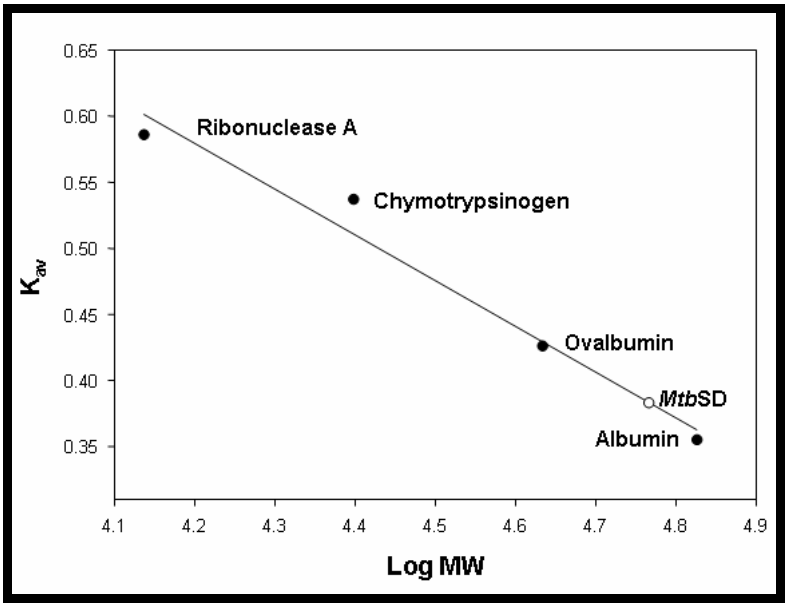


FIGURE 5: Determinação da massa molecular da *Mtb*SD em solução. A amostra foi aplicada na coluna *Superdex 200* HR 10/30 e a corrida foi feita em um fluxo de 0.4 mL min^{-1} em FPLC. Os padrões de massa molecular foram Ribonuclease A (13.700 Da), Chymotrypsinogen (25.000 Da), Ovalbumin (43.000 Da) e Albumin (67.000 Da).

A seqüência dos resíduos de aminoácidos no N-terminal da proteína purificada, SEGPKKAGVLG, foi determinada pelo método de química degradativa de Edman, confirmando assim que a proteína pura trata-se da *Mtb*SD e que realmente ocorre a remoção da metionina N-terminal, um tipo comum de modificação pós-traducional de proteínas sintetizadas em sistemas procariotos.

Posteriormente este protocolo de purificação foi otimizado, reduzindo-se a purificação para três etapas cromatográficas. Em torno de 5 g de células

foram necessárias para obtenção de 130,3 mg de proteínas totais, cujo rendimento de 13,5 % resultou em 1,5 mg de *Mtb*SD pura. Essa preparação protéica foi utilizada para todos os ensaios de caracterização e de cinética.

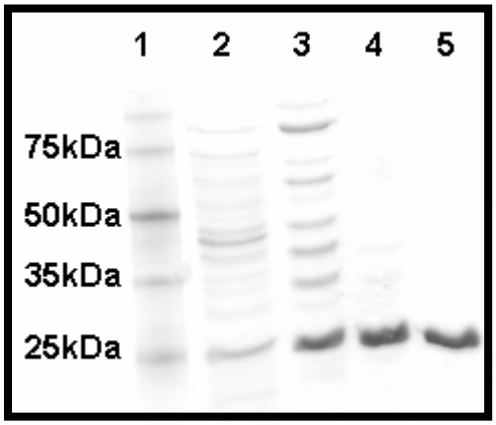


Figura 6: Análise por SDS-PAGE das frações oriundas das etapas cromatográficas da purificação da *Mtb*SD. Canaleta 1: marcador de peso molecular *Perfect Marker* (GIBCO); **canaleta 2:** extrato bruto (40 μg de proteínas); **canaleta 3**: *pool* das frações da *Phenyl Sepharose FF* (40 μg de proteínas), **canaleta 4**: *pool* das frações da *Sephacryl-S200* (20 μg de proteína), **canaleta 5**: *pool* das frações da *Mono-Q* (10 μg de proteína).

Tabela1: Purificação da enzima chiquimato desidrogenase de *M. tuberculosis* H37Rv.

Etapa de Purificação	Proteínas totais (mg)	Unidades totais (U)	Atividade específica (U/mg)	Taxa de purificação	Rendimento (%)
Extrato bruto	130,3	156,5	1,2	1	100
Phenyl Sepharose FF	25,1	113,0	4,5	3,8	72,2
Sephacryl S200	8,9	81,9	9,2	7,7	52,3
MonoQ	1,5	21,2	14,1	11,8	13,5

Analisando a tabela acima de purificação, a otimização da purificação ocorreu pela mudança da primeira etapa cromatográfica, onde a coluna de troca iônica, a *Q-Sepharose Fast Flow* foi substituida pela coluna de interação hidrofóbica *Phenyl Sepharose Fast Flow*, visto que a coluna cromatográfica de interação hidrofóbica foi bem sucedida no primeiro protocolo de purificação

proposto e a coluna *Q-Sepharose Fast Flow* resultou na maior perda do processo anterior. Este novo protocolo de purificação reduz uma etapa na purificação da *Mtb*SD recombinante, melhorando os valores de rendimento e taxa de purificação.

Os protocolos de purificação estabelecidos aqui, não usam a adição de seqüências extras para o alcance da proteína purificada; como é o caso da *Hinflu*SD que utiliza uma cauda de histidinas (YE e cols., 2003). Já foi demonstrado que esse tipo de artefato pode exercer um efeito negativo em posteriores estudos da proteína recombinante, alterando à nível de atividade biológica ou até mesmo alterações estruturais (WOESTENENK e cols., 2004). Na purificação da enzima *Mjan*SD utilizaram duas etapas cromatográficas e após mais uma etapa de remoção proteolítica da cauda (PADYANA & BURLEY, 2003).

Caracterização e Determinação dos Parâmetros Cinéticos Aparentes

O artigo em questão também descreve os parâmetros cinéticos aparentes no sentido direto e reverso da reação catalisada pela enzima *Mtb*SD recombinante. É interessante ressaltar que esses resultados para a reação direta foram os primeiros descritos na literatura para a família das SDs de procariotos.

Os dados foram coletados e analisados em equação hiperbólica, indicando que a reação catalisada pela *Mtb*SD segue a cinética de Michaelis-Menten para todos os substratos (**Figura 3**). Os valores de K_m e V_{max} encontrados para 3-desidrochiquimato (DHS) foram de 31 μM e 108 U mg^{-1}, respectivamente; e para o cofator NADPH foram de 10 μM e 100 U mg^{-1} (**Tabela 2**). O valor de K_m para o substrato DHS é bastante inferior ao valor

encontrado para a SD purificada de *Pisum sativum* (*Psat*SD), cujo valor descrito é de 340 μM (BALINSKY e cols., 1971). Entretanto, o valor de K_m encontrado para o cofator NADPH é próximo ao determinado para a enzima *Psat*SD em condições semelhantes de pH (4,3 μM). Na reação reversa, os valores aparentes de K_m e V_{max} para os substratos SHK e $NADP^+$ foram de 50,18 μM e 18 U mg^{-1}; e de 22 μM e 12,9 U mg^{-1}, respectivamente (**Tabela 2**). O valor de K_m para o SHK é 10 vezes maior para *Psat*SD em relação ao de *Mtb*SD. No entanto, os valores de K_m para os substratos SHK e $NADP^+$ das enzimas *Mtb*SD e *Ecoli*SD são bastante próximos, indicando que as duas enzimas catalisam a reação reversa com constantes cinéticas semelhantes.

Os valores de V_{max} para os substratos DHS e SHK rederam k_{cat} e k_{cat}/K_m para esses substratos de 49 s^{-1} e 1,6 x 10^6 M^{-1} s^{-1}, e 8,2 s^{-1} e 1,63 x 10^5 M^{-1} s^{-1}, respectivamente (**Tabela 2**).

A constante de equilíbrio aparente (K_{eq}) também foi estimada utilizando os parâmetros cinéticos aparentes em estado estacionário. Por meio da equação de Haldane o valor de 19,6 foi alcançado para K_{eq}. A K_{eq} determinada em pH 7.4 para *Psat*SD foi de 10,3. É importante ressaltar que a metodologia utilizada para essa determinação considera o mecanismo cinético como sendo aleatório em rápido equilíbrio. Nesse caso é assumido que a constante de dissociação para o segundo substrato não é alterado pela ligação do primeiro substrato.

O efeito da temperatura na estabilidade da proteína *Mtb*SD foi verificado. Foram incubadas amostras da proteína recombinante em temperaturas diferentes (15, 25, 37 e 55°C) e alíquotas foram retiradas em tempos diferentes de incubação para realização da atividade enzimática (25°C). A *Mtb*SD se

mostra uma proteína bastante estável nas temperaturas de 15, 25 e 37°C durante 1 hora de incubação (**Figura 4**). Entretanto, a atividade enzimática foi reduzida gradativamente em 1 hora de incubação a 55°C, chegando a 20% da atividade inicial no último ponto analisado. A estabilidade da *Mtb*SD também é constatada a temperatura de -20°C, onde a proteína mantém sua atividade enzimática por pelo menos 1 ano.

Pela análise do gráfico log*k* verso 1/*T* (K^{-1}), chegou-se a um valor de 35,2 kJ mol^{-1} de energia de ativação (E_a) para *Mtb*SD (**Figura 5**). É importante ressaltar que a E_a calculada pelo gráfico de Arrhenius é um valor aparente e que o fator pré-exponencial (A) foi considerado como independente da temperatura no intervalo de temperatura utilizado nos experimentos. Logo, esse valor para a E_a pode ser considerado a quantidade mínima de energia necessária para iniciar a reação química catalisada pela *Mtb*SD em concentrações de saturação de SHK e $NADP^+$ testadas. Outra observação feita foi que o gráfico se apresenta linear, indicando que a etapa limitante de velocidade não altera quando a temperatura é mudada. É importante ressaltar que a veracidade dos dados sobre a E_a da *Mtb*SD é constatada pela inexistência de queda repentina do gráfico que indicaria uma possível desnaturação da enzima, consistente com os dados de estabilidade da proteína.

Análise da Seqüência Proteica da MtbSD

Através de um alinhamento múltiplo de seqüências, a seqüência de resíduos de aminoácidos da *Mtb*SD foi comparada com outras três seqüências de SDs, a de *E. coli* (MICHEL e col., 2003), a de *H. influenzae* (YE e col., 2003)

e *M. jannaschii* (PADYANA & BURLEY, 2003), cujas estruturas tridimensionais já foram determinadas por cristalografia (**Figura 6**).

A maior identidade encontrada foi de 24% entre as seqüências de *Mtb*SD e *Ecoli*SD, e entre *Mtb*SD e *Mjan*SD. Apesar desse valor baixo, considerado o limite para realizar estudos de modelagem molecular, vários resíduos conservados da família das SDs foram encontrados nas seqüências analisadas.

O sítio de ligação do substrato na *Ecoli*SD, localizado no domínio N-terminal, foi identificado pelo posicionamento do anel nicotinamida do cofator NADPH na estrutura resolvida. Este sítio está em um sulco encontrada na estrutura da *Ecoli*SD, onde a maior parte dos resíduos altamente conservados da família das SDs estão presentes, como por exemplo, Ser14, Ser16, Lys65, Asn86, Thr101, Asp102 e Gln244 (numeração pela seqüência de *Ecoli*SD) (MICHEL e col., 2003). Esses resíduos estão conservados na seqüência polipeptídica de *Mtb*SD, correspondendo a Ser18, Ser20, Lys69, Asn90, Thr104, Asp105 e Gln243 (numeração pela seqüência de *Mtb*SD).

O domínio N-terminal (domínio I) da estrutura da *Mjan*SD é responsável pela ligação do substrato DHS e os resíduos presentes neste sítio, aparentemente envolvidos nessa ligação, Lys70, Asn91 e Asp106 (numeração pela seqüência de *Mjan*SD), são resíduos polares invariáveis em SD. Outra revelação feita pela estrutura da *Mjan*SD, foi o resíduo Gln254 (numeração pela seqüência de *Mjan*SD), que também aparece fazendo parte da estrutura deste sítio e está conservado na família das SDs. Foi proposto que esses resíduos estão envolvidos com a redução catalítica de DHS a SHK pela *Mjan*SD (PADYANA & BURLEY, 2003). Pelo alinhamento estes resíduos na

seqüência da *Mtb*SD correspondem a Lys69, Asn90, Asp105 e Gln244 (numeração pela seqüência *Mtb*SD).

Pelo estudo estrutural da *Hinfl*SD, foi sugerido que os aminoácidos Ser14, Ser16, Lys65, Asn85, Asp102 e Gln245 (numeração pela seqüência de *Hinfl*SD) estão envolvidos no sítio de ligação do DHS (YE e cols., 2003). Pelo alinhamento múltiplo esses resíduos correspondem aos resíduos mencionados anteriormente. Na enzima SD-*like* de *H. influenzae* foi proposto que os resíduos conservados Lys67, Asn88, Asp103 e Gln242 estão envolvidos tanto na catálise quanto na ligação do substrato (SINGH e cols., 2005). Todos os resíduos mencionados nos trabalhos da *Hinflu*SD e da SD-*like* de *H. influenzae* estão conservados na seqüência de *Mtb*SD e estão presentes nas demais seqüências analisadas.

Este primeiro trabalho descreve o primeiro protocolo de purificação para a enzima *Mtb*SD, que é capaz de obter quantidade suficiente de enzima pura para fazer experimentos de cinética e de cristalografia. Além disso, apresenta uma caracterização detalhada informando o estado oligomérico e as constantes cinéticas aparentes para ambas reações, direta e reversa, entre outras. As constantes cinéticas da reação direta foram as primeiras descritas na literatura para as SDs de procariotos. Como perspectiva tem-se a determinação do mecanismo cinético em estado estacionário e pré-estacionário, os efeitos isotópicos, a mutagênese sítio direcionada e "salvamento químico" que permitirão a elucidação dos mecanismos cinético e químico. A estrutura e o conhecimento do modo de ação da *Mtb*SD serão utilizados como "conhecimento base" para o desenho de inibidores efetivos da

via do chiquimato, pela inibição específica da SD. Esses inibidores serão potenciais agentes anti-TB.

ARTIGO 2: *"Shikimate Dehydrogenase from Mycobacterium tuberculosis H37Rv: Kinetic and Chemical Mechanisms"*

Obs.: As **figuras** e **tabelas em negrito**, mencionadas na discussão a seguir, estão identificadas de acordo com o artigo a que se referem.

Parâmetros de Velocidade Inicial e Inibição pelos Produtos

O primeiro passo para esclarecer o mecanismo de ação de uma enzima é estabelecer seu mecanismo cinético, determinando assim a ordem de ligação dos substratos ao sítio ativo da enzima e posteriormente de liberação dos produtos. O inicio dessa caracterização foi feita pela determinação das velocidades iniciais e posteriormente plotando esses valores no gráfico de Lineweaver-Burk para visualização do perfil das famílias de retas alcançadas.

Para determinar os parâmetros cinéticos em estado estacionário e os padrões de velocidade inicial, a atividade da *Mtb*SD foi determinada variando a concentração de um dos substratos em algumas concentrações fixas do outro substrato e vice-versa.

Esses dados foram plotados no gráfico de Lineweaver-Burk, e as famílias de retas observadas na reação direta para os dois substratos são consistentes com o mecanismo seqüencial e como conseqüência ocorrendo a formação de um complexo ternário no decorrer da reação. Com isso foi descartada a possibilidade do mecanismo ping-pong para a *Mtb*SD, que apresenta no plote de Lineweaver-Burk uma família de retas paralelas. Nos ensaios onde DHS foi variado em sete concentrações fixas de NADPH (10, 25, 50, 100 e 200 μM), as retas se interceptaram sobre o eixo *x* e a esquerda do eixo *y* (**Figura 2A**). Quando o substrato NADPH foi variado, seguindo as

mesmas condições de ensaio para DHS, as retas se interceptaram abaixo do eixo *x* e a esquerda do eixo *y* (**Figura 2B**). Com estes gráficos foi descartado também o mecanismo ordenado em equilíbrio rápido, pois nesse mecanismo se espera uma família de retas que se interceptam no eixo *y*.

Os dados coletados foram ajustados na equação geral para mecanismo seqüencial $v = VAB / (K_aB + K_bA + K_{ia}K_b + AB)$, onde *v* é a velocidade medida da reação, *V* é a velocidade máxima, *A* e *B* são as concentrações dos substratos, K_a e K_b são as constantes de Michaelis-Menten dos substratos *A* e *B* e K_{ia} é a constante de dissociação para o substrato *A*. Os valores para as constantes de Michaelis-Menten dos substratos DHS e NADPH é de 44 ± 2,8 μM e 34 ± 2,2 μM, respectivamente. A constante catalítica (k_{cat}) foi de 77,8 ± 2,3 s^{-1}, alcançando então as constantes de especificidade de 1,8 x 10^6 M^{-1} s^{-1} (k_{cat}/K_{DHS}) e de 2,3 x 10^6 M^{-1} s^{-1} (k_{cat}/K_{NADPH}). Esses valores são maiores que os parâmetros cinéticos aparentes apresentados no artigo "*Functional shikimate dehydrogenase from Mycobacterium tuberculosis H37Rv: Purification and characterization*", onde os dados foram coletados em tampão Tris HCl 100 mM pH 9,0. Foi verificado um aumento significativo na velocidade catalítica da *Mtb*SD quando no ensaio de reação utilizou-se tampão fosfato de potássio 100 mM pH 7,3 ao invés de tampão Tris HCl 100 mM pH 9,0. Os parâmetros cinéticos em estado estacionário para a *Psat*SD, em tampão fosfato de potássio 100 mM pH 7,3, são: K_{DHS} 340 ± 40 μM e K_{NADPH} 4.3 ± 0.5 μM (BALINSKY e col., 1971). Observando esses valores, se constata que K_{DHS} para *Psat*SD é 8 vezes maior e que K_{NADPH} é 8 vezes menor em relação aos valores determinados para *Mtb*SD.

O estudo de inibição pelos produtos foi o segundo passo para a determinação do mecanismo cinético da *Mtb*SD. Os dados coletados foram ajustados nas seguintes equações: $v = VA / [K_a (1 + I/K_{is}) + A]$ ou $v = VA / [K_a (1 + I/K_{is}) + A (1 + I/K_{ii})]$, onde v é a velocidade medida da reação, V é a velocidade máxima, A é a concentração do substrato, K_a é a constante de Michaelis-Menten do substrato A, K_{is} é a constante de inibição angular, K_{ii} é a constante de inibição linear e I é a concentração do inibidor.

O substrato $NADP^+$ mostrou-se um inibidor competitivo para NADPH e não competitivo para DHS. O produto SHK é um inibidor competitivo e não competitivo para os substratos DHS e NADPH, respectivamente (**Tabela 1**). Esse perfil de inibição pelos produtos é compatível com os mecanismos cinéticos ordenado em estado-estacionário e aleatório em rápido equilíbrio. Dados similares para a inibição pelos produtos foram descritos para a *Psat*SD (BALINSKY e col., 1971).

Esses dados de inibição pelos produtos indicam que no decorrer da reação dois possíveis complexos não produtivos podem ser formados, o complexo *Mtb*SD-NADPH-SHK e o complexo *Mtb*SD-$NADP^+$-DHS, caso o mecanismo cinético tratar-se de aleatório em rápido equilíbrio.

Efeitos Cinéticos Isotópicos Primário do Deutério

A diferenciação entre os dois possíveis mecanismos para a *Mtb*SD foi feita através de experimentos de efeitos cinéticos isotópicos primário. Para isso foram determinados os parâmetros cinéticos em velocidade inicial variando um dos substratos em cinco concentrações fixas do outro substrato, utilizando NADPH ou β-nicotinamida adenina dinucleotídeo fosfato deuterado ([4*S*-

^{2}H]NADPH /NADPD). Os efeitos isotópicos cinéticos primário foram observados sobre o *V/K* e o *V* em relação à reação não marcada.

O efeito isotópico sobre *V/K* descreve o efeito a partir da ligação do substrato marcado até a primeira etapa irreversível da reação, que geralmente está representada pela liberação do primeiro produto. O efeito isotópico sobre *V* exibe o efeito nos eventos seguintes após a formação do complexo ternário produtivo, incluindo as etapas químicas, possíveis isomerizações da enzima até a liberação dos produtos (NORTHROP, 1975).

Os valores coletados foram ajustados nas seguintes equações: $v = VA / [K (1 + FiE_{V/K}) + A (1 + FiE_V)$ ou $v = VA / [K + A (1 + FiE_V)]$, as quais descrevem o efeito sobre o *V* e *V/K*, e apenas sobre *V*, respectivamente. Os termos *Fi* representa a fração isotópica marcada, *Evk* o efeito isotópico sobre *V/K* menos 1 e *Ev* o efeito isotópico sobre *V* menos 1 (**Tabela 2**).

O valor do efeito isotópico cinético primário de 1,8 para $^{D}V_{DHS}$ e de 1,5 para $^{D}V_{NADPH}$ utilizando [4*S*-^{2}H]NADPH como redutor, indica que a etapa de transferência do hidreto, que compreende parte da etapa química é parcialmente limitante para a velocidade catalítica da *Mtb*SD. Isso sugere que outras etapas insensíveis ao efeito isotópico, tais como isomerização da enzima e/ou liberação dos produtos, estão envolvidas com as etapas limitantes de velocidade da reação catalisada pela *Mtb*SD. Os valores baixos para $^{D}V/K_{NADPH}$ e $^{D}V/K_{DHS}$ indicam que a etapa de ligação dos substratos até a formação do complexo irreversível tem uma pequena contribuição na velocidade de catalise pela *Mtb*SD.

A magnitude dos efeitos isotópicos cinéticos utilizando o [4*S*-^{2}H]NADPH indica que o hidrogênio *proS* do C_4 (lado B) do cofator é transferido na forma

de hidreto para o substrato DHS na reação de catálise pela *Mtb*SD. Estudo similar feito com a *Ecoli*SD demonstrou que o hidreto transferido na mesma reação de oxiredução é o *proR* do C_4 (lado A) do NADPH (DANSETTE & AZERAD, 1974), dado que confere com o estudo da estrutura 3D da proteína *Ecoli*SD (MICHEL e col., 2003).

A dedução de mecanismo cinético utilizando os efeitos isotópicos para enzimas com multireagentes foi desenvolvida por Cook & Cleland (1981), utilizando os valores dos efeitos isotópicos cinéticos aparentes sobre *V/K* para um substrato em várias concentrações do cosubstrato.

Os valores do $^{D}(V/K_{app})_{NADPH}$ (1,4 ± 0,3) são independentes da concentração de DHS, entretanto os valores de $^{D}(V/K_{app})_{DHS}$ diminuem até o valor de 1,0 ± 0,03 quando a concentração de NADPH ou NADPD é aumentada (**Figura 3**). Segundo as deduções mecanísticas de Cook & Cleland quando o efeito isotópico primário $^{D}(V/K_{app})_{B}$ é independente da concentração de A e o efeito isotópico primário $^{D}(V/K_{app})_{A}$ diminui com o aumento da concentração de B, chegando a 1,0 em concentração infinita de B, o mecanismo cinético da reação em estudo é ordenado em estado estacionário. A e B refere-se ao primeiro e ao segundo substrato, respectivamente, que se liga a enzima. Frente a esses dados chega-se a conclusão de que o mecanismo cinético para a *Mtb*SD é o mecanismo ordenado em estado estacionário, e que a ordem de ligação ao sítio catalítico da enzima é DHS primeiro, seguida da ligação do cofator NADPH (**Figura 10**).

Utilizando os dados dos parâmetros de velocidade inicial, de inibição pelos produtos, da análise da relação de Haldane e dos estudos *isotope-exchange*, o mecanismo cinético ordenado com NADPH ligando primeiro foi

proposto para a *Psat*SD (BALINSKY e col., 1971). A enzima álcool desidrogenase NAD-dependente (classe II) de camundongo, estruturalmente homóloga ao domínio de ligação do cofator da *Ecoli*SD (MICHEL e col., 2003), apresenta o mecanismo seqüencial ordenado bi-bi com o cofator NADPH ligando antes do outro substrato (Strömberg e col., 2004).

Efeito Isotópico Cinético do Solvente e "Proton Inventory".

Para analisar a contribuição da etapa de transferência de próton pelo solvente na velocidade da reação enzimática e para determinar o número de prótons que são transferidos pelo solvente foram feitos experimentos de efeito isotópico cinético do solvente e "*próton inventory*", respectivamente.

O efeito isotópico cinético do solvente foi determinado medindo as velocidades iniciais usando concentrações saturantes de um substrato variando a concentração do outro em H_2O e em 90% de D_2O (**Figura 4A e 4B**). Os valores coletados foram ajustados nas equações utilizadas para determinação dos efeitos isotópicos primário. Os efeitos cinéticos isotópicos do solvente observados sobre o *V/K* e o *V* estão mostrados na **Tabela 2**.

Os valores dos efeitos isotópicos do solvente sobre *V* (^{D_2O}V) foram pequenos, atingindo para DHS a magnitude de 1,5 ± 0,3 e para NADPH de 1,3 ± 0,2. O efeito sobre *V/K* ($^{D_2O}V/K$) não foi observado. Esses dados sugerem que a transferência de próton pelo solvente não se trata da etapa limitante de velocidade, e sim um contribuinte parcial para a velocidade da reação catalisada pela *Mtb*SD.

"Proton inventory" foi determinado utilizando a relação entre a velocidade inicial em concentrações saturantes de NADPH e DHS e várias concentrações de D_2O (20, 40, 60, 80 e 90%). Esta relação foi linear, sugerindo que ocorra a

transferência de um único próton pelo solvente na reação de oxiredução catalisada pela *Mtb*SD (QUINN & SUTTON, 1991). Um resultado similar foi observado para a enzima micotiona redutase dependente de NADPH de *M. tuberculosis* (PATEL & BLANCHARD, 2001) (**Figura inserida 4A**)

Efeitos Isotópicos Cinéticos Múltiplos

A fim de caracterizar o mecanismo químico da reação catalisada pela *Mtb*SD, foram feitos os efeitos isotópicos cinéticos múltiplos através da determinação dos efeitos isotópicos cinéticos do solvente utilizando NADPD como substrato variado. Deste modo distingui-se entre o mecanismo químico concertado, onde a transferência do próton e do hidreto ocorrem na mesma etapa e o mecanismo químico *stepwise*, onde a transferência do próton e do hidreto ocorrem em etapas distintas.

A teoria diz que, se caso a protonação e a transferência de hidreto ocorrerem na mesma etapa de transição, o efeito isotópico primário será maior ou igual com o uso de D_2O como solvente quando comparado com a H_2O. Por outro lado, se caso a transferência de hidreto e a protonação ocorrerem em etapas distintas, o efeito isotópico primário será menor com o uso de D_2O como solvente, visto que a transferência de próton se tornará limitante na velocidade da reação (HERMES e col., 1984 e BELASCO e col., 1983).

Os resultados dos efeitos isotópicos cinéticos múltiplos estão mostrados na **tabela 2**. O aumento nos valores de todos os efeitos isotópicos primários medidos em D_2O indicam o mecanismo químico concertado para a *Mtb*SD (**Figura 5A e 5B**). Um dado importante é o efeito isotópico sobre $^{D}V/K_{DHS}$, que não era observado pelo efeito isotópico cinético primário e passou para 1,3 ± 0,1 quando foi determinado em D_2O como solvente.

O mecanismo concertado é uma estratégia comum utilizada pelas enzimas para evitar a formação de intermediários instáveis (COOK & CLELAND, 1981 e THIBBLIN & JENCKS, 1979). A β-eliminação catalisada pela crotonase é do tipo concertado (GERLT, 1998), bem como as reações catalisadas pela acil-CoA desidrogenase (BAHNSON & ANDERSON, 1991) e a isocitrato desidrogenase (POHL e col., 1986) de fígado suíno.

Estudos de pH

O papel da química ácido/base no mecanismo da *Mtb*SD foi determinado através das constantes k_{cat} e k_{cat}/K_m em diferentes pHs (5,5 – 10,0). Previamente foi verificada a estabilidade da enzima *Mtb*SD em cada um dos pHs testados, para a certificação da veracidade dos dados coletados. Em cada pH testado, as velocidades iniciais foram medidas variando a concentração de um dos substratos mantendo o outro em concentração saturante. Os dados coletados foram ajustados na equação $\log y = \log [C/(1 + K_b/H)]$ onde y é o parâmetro cinético aparente, C é valor de y no platô independente do pH, H é a concentração de H^+ e K_b é a constante de dissociação aparente para os grupos ionizados.

Os efeitos do pH na k_{cat} para os substratos DHS e NADPH são similares, diminuindo os valores em pHs elevados com uma inclinação de -1 (**Figura 6A**). Isso demonstra que a deprotonação de um único resíduo com pK_a aparente de 8,9 $\pm$ 0,1 anula a atividade catalítica da *Mtb*SD.

A constante k_{cat}/K_m para ambos os substratos sofreu um decréscimo em altos pHs com uma inclinação de -1, sugerindo que um único grupo deprotonado com valor de pK_a aparente igual a 9,1 $\pm$ 0,1, reduz a ligação entre substrato-enzima (**Figura 6B e 6C**).

Provavelmente é a ionização do mesmo grupo que esta sendo observado pelo efeito do aumento de pH sobre as constantes analisadas. O valor de pK_a para as constantes k_{cat} e k_{cat}/K_m encontra-se na faixa normal de pK_a para os grupos ε-amino da lisina, tiol da cisteína e hidroxila fenólica da tirosina. Frente aos resultados de efeitos isotópicos cinéticos múltiplos e estudos de pH foi proposto um mecanismo químico para a *Mtb*SD mostrado na **figura 11**.

A análise da estrutura da *Ecoli*SD propõe que a cadeia lateral do resíduo Lys65 estabelece uma ponte de hidrogênio ao grupo hidroxila do C-4 do DHS/SHK (MICHEL e col., 2003). Os autores propuseram ainda que tal resíduo seja o grupo catalítico ácido/base que doa um próton para o grupo carbonil do substrato DHS durante a sua redução (MICHEL e col., 2003). Compostos análogos ao substrato DHS que não continham as hidroxilas no C-4 e C-5 foram utilizadas para demonstrar o papel dessas hidroxilas na especificidade da enzima *Ecoli*SD pelo substrato. Os resultados mostraram que a hidroxila do C-4 tem um efeito bastante significativo na especificidade da enzima pelo substrato (BUGG e col., 1983). A avaliação feita na energia de ligação através do k_{cat}/K_m sugere que a hidroxila no C-4 estabelece uma ponte de hidrogênio com um grupo carregado no sítio ativo da *Ecoli*SD (BUGG e col., 1983). É interessante ressaltar que a participação da cadeia lateral da lisina na reação química catalisada pela *Ecoli*SD está baseada em um modelo molecular, e a estrutura determinada para *Ecoli*SD por cristalografia é o complexo binário enzima-$NADP^+$ e que a transferência do hidreto nessa reação é dada pelo hidrogênio *proR* do C-4 do cofator. A cadeia lateral do resíduo Asp102 da

*Ecoli*SD se propõe que também esteja envolvida com a hidroxila no C-4 do DHS por meio de uma ponte de hidrogênio (MICHEL e col., 2003).

Estudos prévios da *Psat*SD mostraram que inibidores do tipo substrato-*like* necessitam da hidroxila no C-4, ao passo que tanto uma hidroxila quanto um grupo carboxila no C-5 é necessário para uma ligação forte entre inibidor-enzima (BALINSKY & DAVIES, 1961). Os efeitos de pH sobre a velocidade da *Psat*SD sugerem que uma cisteína ou um grupo ε-amino está protonado para interagir com o grupo hidroxila no C-4 do DHS/SHK (DENNIS & BALINSKY, 1972).

A estrutura da *Mjan*SD foi determinada e os resíduos envolvidos na ligação e na sua redução catalítica do DHS foram identificados, dentre eles está a Lys 70 (PADYANA & BURLEY, 2003). A *Hinfl*SD tem sua estrutura resolvida pela análise cristalográfica e estudos de mutagênese sítio direcionada mostrou que o par Asp103 - Lys67 pode estar envolvido no mecanismo de catálise desta enzima (SINGH e col., 2005).

Em nossos estudos de efeitos de pH na reação catalisada pela *Mtb*SD não foi identificado qualquer resíduo cuja protonação anule a atividade enzimática em pH baixo. Entretanto, os efeitos dos pHs abaixo de 5,5 não foram testados e com isso não foi permitida a detecção do grupo carboxila presente na cadeia lateral do aspartato, visto que o valor de pK_a para esse grupo é de aproximadamente 4. Os testes em pHs inferiores a 5,5 foram evitados pelo fato de que o *pK* do grupo carboxila do DHS/SHK é ~ 4,1 e isso dificultaria a interpretação dos resultados.

Anteriormente foi feita uma análise da seqüência de *Mtb*SD através do alinhamento múltiplo de seqüências utilizando: a *Ecoli*SD, a *Hinfl*SD-like e a

*Mjan*SD, cujas estruturas já foram determinadas (FONSECA e col., 2006). Dentre os resíduos identificados como conservados para a família das SDs foram identificados os resíduos Lys69 e Asp105 da *Mtb*SD que correspondem aos resíduos discutidos nas demais estruturas (FONSECA e col., 2006). Os dados descritos aqui no estudo de pH para *Mtb*SD mostrou a participação de um grupo cuja deprotonação desfavorece a ligação da *Mtb*SD ao substrato (pK_a = 9.1) e reduz a atividade catalítica (pK_a = 8.9), que é consistente com a participação da Lys69 identificada na seqüência primária da enzima alvo.

Modelagem por Homologia

O primeiro passo para execução da técnica de modelagem por homologia é buscar um molde para os estudos. Foram encontrados três candidatos: a *Ecoli*SD de (PDB ID: 1NYT) (MICHEL e col., 2003), a quinato/chiquimato desidrogenase (QSD) de *E. coli* (PDB ID: 109B) (MICHEL e col., 2003) and a *Mjan*SD (PDB ID: 1NVT) (PADYANA & BURLEY, 2003). O resultado do "*BLASTp*" identificou a *Mjan*SD (PDB ID: 1NVT) como melhor modelo para a *Mtb*SD. Entretanto, as coordenadas da estrutura não estavam disponíveis para o estudo. O segundo candidato pelos dados obtidos no "*Blastp*" foi a QSD de *E. coli* (PDB ID: 109B), entretanto sua função é diferente da exercida pela *Mtb*SD e além disso a QSD tem a capacidade de utilizar como cofator tanto o NADH quanto o NADPH. Outro ponto desfavorável em relação ao molde QSD é a inserção de dois *gaps* na seqüência da *Mtb*SD observados no alinhamento múltiplo de seqüências (**Figura 7**). Esses segmentos que não apresentam equivalência no molde são regiões difíceis de resolver na estrutura a ser modelada.

Frente a essas justificativas, a *Ecoli*SD (PDB ID: 1NYT) foi eleita como molde para a *Mtb*SD, pois apresentam a mesma função biológica, a sua estrutura foi resolvida em 1,5Å (MICHEL e col., 2003) e a identidade entre as duas seqüências é de 25%, limite permitido para os estudos de modelagem estrutural de proteínas sejam feitos (ALTSCHUL e col., 1997).

Partindo desse molde, dez modelos foram construídos para a *Mtb*SD. Todos os modelos foram avaliados pelos programas PROCHECK (LASKOWSKI e col. 1993) (**Tabela 4**) e VERIFY 3D (LUTHY e col., 1992) para eleger o melhor modelo. Dos 269 resíduos de aminoácidos da *Mtb*SD na melhor estrutura modelada, 90,5% dos resíduos estão localizados na região mais favorável no gráfico de Ramachandran (**Figura 8 e Tabela 3**). O RMSD da cadeia principal entre as estruturas do molde e da modelada é de 1,93 Å. Considerando as divergências entre as duas seqüências utilizadas no estudo, o valor para RMSD é plausível e os resultados gerais da estrutura modelada mostram que o molde eleito é bastante satisfatório e pode ser utilizado para analisar a relação estrutural e funcional da *Mtb*SD.

A estrutura da *Mtb*SD pertence a superfamília denominada domínio "*Rossmann-fold*" de ligação a NAD(P) (MURZIN e col., 1995). A estrutura apresenta uma forma alongada com dois domínios, que apresentam uma arquitetura α/β (**Figura 9**). Esses domínios estão ligados por uma α-hélice e uma volta que os mantém juntos e formando um sulco onde o cofator NADPH liga-se. O domínio de ligação do cofator NADPH, localizado na porção C-terminal, apresenta uma estrutura do tipo "*Rossmann-fold*", que compreende em seis fitas β paralelas e seis α hélices, onde cada três α hélices estão dispostas sobre cada lado do leque central de fitas β. A porção N-terminal é

bastante semelhante ao domínio C-terminal descrito, entretanto ocorre um arranjo irregular que compreende a orientação antiparalela da fita β-5 em relação às demais fitas (MICHEL e col., 2003).

Na análise da estrutura da *Ecoli*SD, utilizada como molde, foi identificado o sítio de ligação dos substratos pela posição do anel nicotinamida do cofator, e neste sulco os resíduos de aminoácidos, Ser14, Ser16, Lys65, Asn86, Thr101, Asp102 e Gln244 foram identificados. Ambos os oxigênios carboxil do substrato DHS formam pontes de hidrogênio com o sítio ativo da proteína, provavelmente pelos resíduos conservados de serina nas posições 14 e 16 (MICHEL e col., 2003). Tais resíduos foram identificados no modelo para *Mtb*SD, Ser18, Ser20, Lys70, Asn91, Thr106, Asp107 e Gln263, em posições estruturalmente equivalentes (**Figura 12**). Na estrutura da *Ecoli*SD foi identificado o resíduo Tyr215 que também está envolvido na ligação do substrato DHS no sítio ativo da enzima através de ponte de hidrogênio. No modelo da *Mtb*SD o resíduo equivalente é Tyr215 (**Figura 12**).

No alinhamento múltiplo de seqüências entre o alvo (*Mtb*SD) e os moldes (*Ecoli*QSD and *Ecoli*SD) foi identificada a seqüência Gly124-Ser125-Gly126-Gly127-Thr128-Ala129 como a seqüência padrão da *Mtb*SD (G [A,s,g] G G [A,t] [A,S,g]), correspondendo ao *loop* de ligação ao difosfato, que está conservado na família das SDs (MICHEL e col., 2003) (**Figura 7**).

Para a ligação da nicotinamida na *Ecoli*SD foi verificado que o N-7 do grupo amida do anel nicotinamida é ligado por pontes de hidrogênio ao grupo carbonil da Met213 e da Gly237 (MICHEL e col., 2003). A Gly236 (posição correspondente para a Gly237) é conservada no modelo da *Mtb*SD, estabelecendo o mesmo tipo de interação com o cofator; em contrapartida na

posição equivalente a Met213 encontra-se a Ala213 que mantém a mesma natureza de ligação com o N-7 do anel (**Figura 7**).

As argininas 150 e 154 na *Ecoli*SD têm um papel crucial na ligação do sítio ativo ao fosfato da adenine pela formação de um "grampo eletrostático" (*electrostatic clamp*) que envolve o fosfato (MICHEL e col., 2003). No modelo da *Mtb*SD são encontrados os resíduos correspondentes Arg149 e Lys153. Provavelmente no modelo em estudo estes resíduos têm o mesmo papel, visto que aparecem pontes de hidrogênio entre eles e o fosfato da adenina (**Figura 7**).

O resíduo de Ser189 na *Ecoli*SD é ligado ao oxigênio do anel da ribose fosfato do cofator por ponte de hidrogênio. Analisando o modelo *Mtb*SD essa interação é dada pelo resíduo de Val189. A interação entre o grupo fosfato da ribose e o sítio ativo está sendo feito por ponte de hidrogênio com Asn149 e Asn150 na *Ecoli* SD e na *Mtb*SD, respectivamente (**Figura 7**).

A interação entre a porção difosfato do cofator e o sítio ativo da *Ecoli*SD ocorre através da Gly129 e da Ala130, que corresponde a Thr128 e a Ala129 no modelo em estudo, respectivamente.

Michel e col. (2003) mostraram que a ligação do NAD^+ na *Ecoli*QSD é favorecida pela substituição dos resíduos Thr151 e Arg154, presentes na *Ecoli*SD, pelos resíduos Asp158 e Phe160, respectivamente. A explicação para tal, é o ambiente hidrofóbico criado pelo resíduo apolar Phe160, que reduz a especificidade pelo NADPH e permitindo assim, que tanto NADH quanto NADPH sejam utilizados como cofatores pela *Ecoli*QSD. Em posição equivalente aos resíduos Thr151 e Arg154 na estrutura da *Ecoli*SD estão os resíduos Asn157 e Lys160 na estrutura modelada da *Mtb*SD. Tanto no molde

quanto na estrutura modelada, esses resíduos estabelecem ponte de hidrogênio e é essa interação que resulta em uma orientação similar das cadeias laterais e como conseqüência um ambiente equivalente para a ligação do NADPH (**Figura 13**).

O trabalho apresentado é a primeira descrição detalhada sobre os mecanismos cinético e químico para a *Mtb*SD, propondo um resíduo envolvido na catálise e na etapa de ligação dos substratos.

Como perspectivas próximas há os estudos de mutagênese sítio direcionada para confirmar a importância do resíduo Lys69 na ligação dos substratos e na velocidade de catálise pela *Mtb*SD e os estudos de fluorescência de proteína que fornecerão as constantes de associação e dissociação em equilíbrio para ambos substratos. A etapa subseqüente será a cinética em estado pré-estacionário que possibilitará a determinação de constantes de velocidade de primeira e segunda ordens o que possibilitará um melhor detalhando do mecanismo cinético.

Estes estudos juntamente com os estudos estruturais do sítio ativo da enzima fornecerão a base sólida para o desenho racional de inibidores para a *Mtb*SD, que poderão representar possíveis agentes quimioterápicos para tratamento da tuberculose.

Conclusões

- A proteína recombinante purificada é a *Mtb*SD na forma dimérica;
- A reação catalisada é reversível, com uma *K*eq de 19,6, uma E*a* = 35,2 kJ mol-1 e segue a cinética de Michaelis-Menten;
- O mecanismo cinético é ordenado em estado estacionário, com DHS ligando primeiro à enzima;
- As transferências de hidreto e próton contribuem modestamente para determinação da velocidade da reação e ocorrem de forma concertada;
- Há catálise ácido/base na reação e a protonação de um grupo é essencial para a catálise e a ligação dos substratos;
- Todos os resíduos sugeridos importantes para catálise na SD de *E. coli* estão conservados e em posições semelhantes no sítio ativo da *Mtb*SD.

Referências

[1]. ALTSCHUL, S.F., MADDEN, T.L., SCHÄFFER, A.A., ZHANG, J., ZHANG, Z., MILLER, W., LIPMAN, D.J. (1997) *Nucleic Acids Res*. **25**, 3389-3402

[2]. BALINSKY, D., DAVIES, D. D. (1961) *Biochem. J*. **80**, 296-300

[3]. BALINSKY, D., DENNIS, A.W., CLELAND, W.W. (1971) *Biochemistry* **10** 1947–1952.

[4]. BASSO, L.A., SILVA, L.H.P., FETT-NETO, A.G., AZEVEDO, W.F., MOREIRA, I.S., PALMA, M.S., CALIXTO, J.B., ASTOLFI, S., SANTOS, R.R., SOARES, M.B.P., SANTOS, D.S. (2005) *Mem Inst Oswaldo Cruz* **100,** 475-506.

[5]. BELASCO, J. G., ALBERY, J., AND KNOWLES, J. R. (1983) *J. Am. Chem. Soc.* **105**, 2475-2477.

[6]. BUGG, T. D., ABELL, C., COGGINS, J. R. (1988) *Tetrahedron Lett*. **29**, 6779-6782

[7]. BENACH, J., LEE, I., EDSTRON, W., KUZIN, A. P., CHIANG, P., ACTON, T. B., MONTELIONE, G. T. AND HUNT J. F. (2003) *The Journal of Biological Chemistry* **278**, 21, 19176-19182.

[8]. BENTLEY, R. (1990) *Critical Reviews in Biochemistry and Molecular Biology* **25**, 307-384.

[9]. BLOOM, B.R. & MURRAY, C.J.L. (1992) *Science* **257**, 1055-1064.

[10]. BRENNAN, P.J. (1997) *FEMS Immunology and Medical Microbiology* **18**, 263-269.

[11]. CHAUDHURI, S. AND COGGINS J. R. (1985) *Biochemical Journal* **226**, 217-223.

[12]. COLE, S.T., BROSCH, R., PARKHILL, J. *et al*. (1998) *Nature* **393**, 537–544.

[13]. COOK, P. F., CLELAND, W. W. (1981) *Biochemistry* **20**, 1790-1796

[14]. DANSETTE, P., AZERAD, R. (1974) *Biochimie* **56**, 751-755

[15]. DANIEL, T.M. Captain of death: the story of tuberculosis. *1th ed. New York. University of Rochester Press*, 1997.

[16]. DENNIS, A. W., BALINSKY, D. (1972) *Int. J. Biochem*. **9**, 93-102

[17]. DORMANDY, T. (2002) The white death: a history of tuberculosis, *New York University Press.*

[18]. DUNCAN, K. (2003) *Tuberculosis* **83**, 201-207.

[19]. DUNLAP, N. E., BASS, J., FUJIWARA, P., HOPEWELL, P., HORSBURGH JR., C. R., SLAFINGER, M. AND SIMONE, P. M. (2000) *American Journal of Respiratory and Critical Care Medicine* **161**, 1376-1395.

[20]. ENARSON, D.A. & MURRAY, J.F. (1996) Global epidemiology of tuberculosis. In: Tuberculosis (Rom, W.M. & Garay, S., Eds.), pp. 57-75. Little, Brown and Co., Boston, MA.

[21]. ENSERINK, M. (2001) *Science* **293**, 5528, 234-236.

[22]. ESPINAL M. A. (2003) *Tuberculosis* **83**, 44-51.

[23]. FÄTKENHEUER, G., TAELMAN, H., LEPAGE, P., SCHWENK, A. & WENZEL, R. (1999) *Diagn. Microbiol. Infect. Dis.* **34**, 139-146.

[24]. FONSECA, I. O., MAGALHÃES, M. L. B., OLIVEIRA, J. S., SILVA, R. G., MENDES, M. A., PALMA, M. S., SANTOS, D. S., BASSO, L. A. (2006) *Protein Expr. Purif.* **46**, 429-437

[25]. GLICKMAN, M.S. & JACOBS, W.R. Jr. (2001) *Cell* **104**, 477-485.

[26]. HERMES, J. D., ROESKE, C. A., O'LEARY, M. H., AND CLELAND, W. W. (1984) *Biochemistry* **21**, 5106-5114.

[27]. KNIGGE, A., MORR, H. AND KILLIAN, A. (2000) Tuberculosis and HIV/AIDS. Epidemics: Opportunities for cross support. *Disease Control (Tuberculosis) and Health Promotion AIDS Control in Developing Countries*. Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ) GmbH.

[28]. LASKOWSKI, R.A., MACARTHUR, M.W., MOSS, D.S., THORNTON, J.M. (1993) *J. Appl. Cryst*. **26**, 283-291

[29]. LÜTHY, R., BOWIE, J.U., EISENBERG, D. (1992) *Nature* **356**, 83-85

[30]. MURZIN A. G., BRENNER S. E., HUBBARD T., CHOTHIA C. (1995) *J. Mol. Biol.* **247**, 536-540

[31]. MAGALHÃES, M. L. B., PEREIRA, C. P., BASSO, L. A. AND SANTOS D. S. (2002) *Prot. Exp. and Purif.* **26**, 59-64.

[32]. MICHEL, G., ROSZAK, A.W., SAUVE, V., MACLEAN, J. MATTE, A., COGGINS, J.R., CYGLER, M., LAPTHORN, A.J. (2003) *J. Biol. Chem*. **278**, 19463-19472.

[33]. MILBURN, H. J. (2001) *Current Opinion in Pulmonary Disease* **7**, 133-141.

[34]. NORTHROP, D. B. (1975) *Biochemistry* **14**, 2644-2651

[35]. *NSB* Editorial comment. Taming tuberculosis-again. *Nature Structural Biology* **7**, 87-88 (2000).

[36]. O´BRIEN, R. J. AND NUNN P. P. (2001) *American Journal of Respiratory and Critical Care Medicine* **162**, 1055-1058.

[37]. PADYANA, A.K., BURLEY, S.K. (2003) *Structure* **11**, 1005-1013.

[38]. PABLOS-MENDEZ, A., RAVIGLIONE, M.C. *et al.* (1998) *The New England Journal of Medicine* **338**, 1641-1649.

[39]. PARISH, T. AND STOKER, N. G. (2002) *Microbiology* **148**, 3069-3077.

[40]. PARRISH, N.M., DICK, J.D. & BISHAI, W.R. (1998) *Trends in Microbiology* **6**, 107-112.

[41]. PASQUALOTO, K.F.M. & FERREIRA, E.I. (2001) *Current Drug Targets* **2**, 427-437.

[42]. PATEL, M. P., AND BLANCHARD, J. S. (2001) *Biochemistry* **40**, 5119-5126.

[43]. PETRINI, B. & HOFFNER, S. (1999) *International Journal of Antimicrobial Agents* **13**, 93-97.

[44]. QUINN, D. M., AND SUTTON, L. D. (1991) *Enzyme Mechanism from Solvent Isotope Effects* (Cook, P. F., Ed.) 1st ed., pp 73-126, CRC Press, Florida.

[45]. RATLEDGE, C. (1982) Nutrition, Growth and Metabolism in “The Biology of the Mycobacteria” (Ratledge, C. and Stanford, J. L. Ed.) vol. 1, pp. 185-271, Academic Press, London.

[46]. RILEY, L.W. (1993) *Clinical Infectious Diseases* **17**, 442-446.

[47]. ROBERTS, F., ROBERTS, C. W., JOHNSON, J. J., KYLE, D. E., KRELL, T., COGGINS, J. R., COOMBS, G. H., MILHOUS, W. K. TZIPORI, S., FERGUSON, D. J. P., CHAKRABARTI, D. AND MCLEOD, R. (1998) *Nature* **393,** 801-805.

[48]. RUFFINO-NETTO, A. (2002) *Rev. Soc. Bras. Med. Trop.* **35**, 51-58.

[49]. STRÖMBERG, P., SVENSSON, S., BERST, K. B., PLAPP, B. V., HÖÖG, J. O. (2004) *Biochemistry* **43**, 1323-1328

[50]. SINGH, S., KOROLEV, S., KOROLEVA, O., ZAREMBINSKI, T., COLLART, F., JOACHIMIAK, A., CHRISTENDAT, D. (2005) *J. Biol. Chem*. **280**, 17101-17108

[51]. SEPKOWITZ, K. A., RAFFALI, J., RILEY, L., KIEHN, T. E. AND ARMSTRONG, D. (1995) *Clinical Microbiology Reviews* 180-199.

[52]. ZHANG, X., ZHANG, S., HAO, F., LAI, X., YU, H., HUANG, Y. AND WANG, H. (2005) *J Biochem Mol Biol.* **38**, 624-31.

[53]. Ye, S., Delft, F.V., Brooun, A., Knuth, M.W., Swanson, R.V., McRee, D.E. (2003) *J. Bacteriol*. **185,** 4144–4151.

[54]. YOUNG, D.B. (1998) *Nature* **393**, 515-516.

[55]. WICKELGREN, I. (2000) *Science* **288**, 5470, 1314-1315.

[56]. WOESTENENK, E.A., HAMMARSTRO, M., BERG, S., HÄRD, T., BERGLUND, H. J. (2004) *Struct. Funct. Genomics* **5,** 217–229.

ANEXO I

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochemical and Biophysical Research Communications 309 (2003) 923–928
BBRC
www.elsevier.com/locate/ybbrc

Docking and small angle X-ray scattering studies of purine nucleoside phosphorylase

Walter Filgueira de Azevedo Jr.,[a,b,*] Giovanni César dos Santos,[a] Denis Marangoni dos Santos,[a] Johnny Rizzieri Olivieri,[a,b] Fernanda Canduri,[a,b] Rafael Guimarães Silva,[c] Luiz Augusto Basso,[c] Gaby Renard,[c] Isabel Osório da Fonseca,[c] Maria Anita Mendes,[b,d] Mário Sérgio Palma,[b,d] and Diógenes Santiago Santos[c]

[a] *Departamento de Física, UNESP, São José do Rio Preto, SP 15054-000, Brazil*
[b] *Center for Applied Toxinology, Instituto Butantan, São Paulo, SP 05503-900, Brazil*
[c] *Rede Brasileira de Pesquisas em Tuberculose, Departamento de Biologia Molecular e Biotecnologia, UFRGS, Porto Alegre, RS 91501-970, Brazil*
[d] *Laboratory of Structural Biology and Zoochemistry, Department of Biology, Institute of Biosciences, UNESP, Rio Claro, SP 13506-900, Brazil*

Received 19 August 2003

Abstract

Docking simulations have been used to assess protein complexes with some success. Small angle X-ray scattering (SAXS) is a well-established technique to investigate protein spatial configuration. This work describes the integration of geometric docking with SAXS to investigate the quaternary structure of recombinant human purine nucleoside phosphorylase (PNP). This enzyme catalyzes the reversible phosphorolysis of *N*-ribosidic bonds of purine nucleosides and deoxynucleosides. A genetic deficiency due to mutations in the gene encoding for PNP causes gradual decrease in T-cell immunity. Inappropriate activation of T-cells has been implicated in several clinically relevant human conditions such as transplant rejection, rheumatoid arthritis, lupus, and T-cell lymphomas. PNP is therefore a target for inhibitor development aiming at T-cell immune response modulation and has been submitted to extensive structure-based drug design. The present analysis confirms the trimeric structure observed in the crystal. The potential application of the present procedure to other systems is discussed.

Keywords: Geometric docking; SAXS; Purine nucleoside phosphorylase; Bioinformatics

Recent developments in the algorithm for protein docking allowed the prediction of the conformation of quaternary structures of several proteins. Among all available algorithms for docking of biological macromolecules the geometric docking has been proved to generate reasonable models of several macromolecular assemblies [1,2]. Small angle X-ray scattering (SAXS) technique provides information on the structural characteristics of macromolecules in solution at a superatomic scale. One of the procedures to obtain structural information from SAXS results is based on the comparison between structure functions of proposed models with different configurations of monomers or subdomains with those determined from experiments. Even though this technique cannot guarantee the uniqueness of the model, it is widely used and was demonstrated to yield useful information on the structure, on structural variations, and on the quaternary structure of a number of macromolecules of biological interest [3]. The Guinier analysis of SAXS intensity provides a structural parameter, the radius of gyration of the macromolecule in solution, which is independent of any a priori model. The SAXS method also yields information on the spatial configuration of the macromolecular subdomains but ignores internal structural details and dynamics features such as, vibration, rotation or internal conformational changes [4].

Purine nucleoside phosphorylase (PNP) catalyzes the reversible phosphorolysis of the ribonucleosides and

* Corresponding author. Fax: +55-17-221-2247.
E-mail address: walterfa@df.ibilce.unesp.br (W.F. de Azevedo Jr.).

0006-291X/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2003.08.093

2′-deoxyribonucleosides of guanine, hypoxanthine, and a number of related nucleoside compounds [5], except adenosine. Human PNP is an attractive target for drug design and it has been submitted to extensive structure-based design. PNP inhibitors could be used in the following applications: (1) treatment of T-cell leukemia; (2) suppression of the host versus graft response in organ transplantation recipients; (3) treatment of secondary or xanthine gout by restricting purine catabolites to the more soluble nucleosides; and (4) in combination with nucleosides to prevent their degradation by PNP metabolism [6]. More recently, the structure of human PNP has been solved using cryocrystallographic techniques at 2.3 Å resolution, which allowed a redefinition of the residues involved in the substrate binding sites [7,8]. The crystallographic structure is a trimer, however, there is a report of dimeric structure for the human enzyme [9], which may change the subunit interface. Since the active site is located near the interface of two subunits, changing the putative interactions between enzyme and inhibitors should have a bearing on structure-based inhibitor design.

Here we report the combination of geometric docking simulations and SAXS studies to assess the human PNP quaternary structure in solution. The general procedure, here described, may be used to study the spatial configuration of the macromolecular subdomains of other proteins in solution.

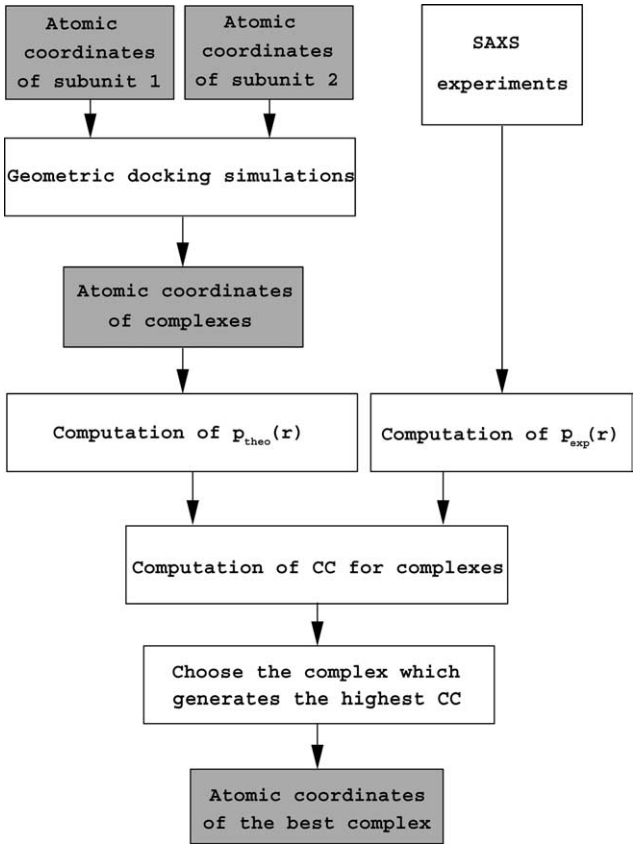


Fig. 1. Flowchart describing the overall strategy to assess protein complex conformations.

Materials and methods

Integration of geometric docking simulation and SAXS experiments. In order to assess the quaternary structure of PNP a scheme was used that involved both geometric docking simulations and SAXS experiments. A flowchart describing the overall strategy is shown in Fig. 1. This procedure was used to generate the dimeric models for PNP and the trimeric structure was built using the crystallographic symmetry. In order to speed up the geometric docking simulations, a parallel version of the program GRAMM [10] was used to generate the dimeric models for human PNP. Each step of the procedure is described in the following sections.

Geometric docking simulations. In order to obtain the dimeric structure of human PNP, docking simulation was performed using the geometric recognition algorithm, which was developed to identify molecular surface complementarity. The monomeric structure of human PNP (PDB access code: 1M73) [7] was docked against its own structure. It generated a total of 100 dimers. The geometric recognition algorithm is based on a geometrical approach and involves an automated procedure including: (i) a digital representation of the molecules by three-dimensional discrete functions; (ii) the calculation of a correlation function that assesses the degree of molecular surface overlap and penetration upon relative shifts of the molecules in three dimensions; and (iii) a scan of the relative orientations of the molecules [10]. The procedure is equivalent to a six-dimensional search but considerably faster by design, and the computation time is only moderately dependent on molecular size. This procedure has been applied to assess protein–protein and protein–ligand interactions. The geometric recognition algorithm was implemented in the program GRAMM [10]. All geometric docking simulations were performed on a Beowulf cluster, with 16 nodes (B16/AMD Athlon 1800+; BioComp, São José do Rio Preto, SP, Brazil).

SAXS studies. X-ray scattering data were collected at room temperature using Cu Kα X-rays radiation generated by a Rigaku RU300 rotating anode generator operated at 50 kV and 90 mA and collimated with a block slit system [11]. The scattering intensity was measured using a linear position sensitive detector (CBPF-Brazil).

The SAXS measurements were performed within an angular range defined by $0.02\,\text{Å}^{-1} < h < 0.450\,\text{Å}^{-1}$ where $h = (4\pi\sin)/\lambda$, 2θ being the angle between the incident and the scattered X-ray beam and λ the X-ray wavelength. The contributions to the scattering intensity from the solvent, capillary, and air were subtracted from the total intensity.

Recombinant human PNP was expressed and purified as previously described [12]. The SAXS measurements were carried out using human PNP solution, which was concentrated to 12 mg ml^{-1} against 10 mM potassium phosphate buffer (pH 7.1). The counting time was 12 h. The extrapolated experimental SAXS intensity function was desmeared to suppress the influence from the slit collimation system yielding the corrected intensities, $I(h)$.

A structural parameter related to the overall size of the macromolecule, the radius of gyration R_g, was determined by using the Guinier equation [13]

$$I(h) = I(0)\exp\left[-\frac{h^2R_g^2}{3}\right]. \quad (1)$$

Eq. (1) applies to macromolecules in the limits of a dilute solution and small h values. More detailed information of the molecular structure can be obtained from the distance distribution function $p(r)$, which is related to the SAXS desmeared (free from geometrical collimation effects) intensity $I(h)$ by

$$p(r) = \frac{1}{2\pi^2}\int_0^\infty I(h)(hr)\sin(hr)\,\mathrm{d}h. \quad (2)$$

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 radius of gyration of macromolecules in solution is usually determined by applying Eq. (1). The distance distribution function, $p_{\exp}(r)$, has been determined by indirect Fourier transformation using the ITP program [11]. This program was also used to determine the intensity $I(h)$, free from smearing collimation effects. The theoretical function $p_{\text{theo}}(r)$ was calculated using the program MULTIBODY [11], modified in order to make molecular model building easier [4]. The program MULTIBODY calculates the resulting function $p(r)$ of the complete set of atomic coordinates of each structural model for the macromolecule. In the present study we calculated the $p_{\text{theo}}(r)$ for monomer, for the dimers, generated by geometric docking, and for the trimer, obtained by application of crystallographic rotations.

Correlation between geometric docking simulations and SAXS experiments. To assess the correlation between theoretical and experimental distance distribution function $p_{\exp}(r)$ and $p_{\text{theo}}(r)$, respectively, we have calculated the linear correlation coefficient (CC), which is defined as follows [14]:

$$\mathrm{CC} = \frac{\sum_i^n \left[\left(|p_{\exp,i}(r)|^2 - \overline{|p_{\exp}(r)|^2}\right) \times \left(|p_{\text{theo,i}}(r)|^2 - \overline{|p_{\text{theo}}(r)|^2}\right)\right]}{\left[\sum_i^n \left(|p_{\exp,i}(r)|^2 - \overline{|p_{\exp}(r)|^2}\right)^2 \sum_i^n \left(|p_{\text{theo,i}}(r)|^2 - \overline{|p_{\text{theo}}(r)|^2}\right)^2\right]^{\frac{1}{2}}}, \quad (3)$$

where $\overline{|p_{\exp}(r)|^2}$ is the mean of the $|p_{\exp,i}(r)|^2$, $\overline{|p_{\text{theo}}(r)|^2}$ is the mean of $|p_{\text{theo,i}}(r)|^2$, and sums are made over all available $p(r)$. When a correlation is known to be significant, CC is one conventional way of summarizing its strength. The complex, which generates the highest correlation coefficient, is considered the right macromolecular conformation. In the present work, we calculated the CC for all dimeric models obtained from the geometric docking simulations and for the monomeric and trimeric structures.

Results and discussion

Guinier plot ($\log I(h)$ versus h^2) of the desmeared scattering function is displayed in Fig. 2. The slope of linear portion of this plot was determined to obtain the radius of gyration of human PNP. The R_g value was 29.8 Å.

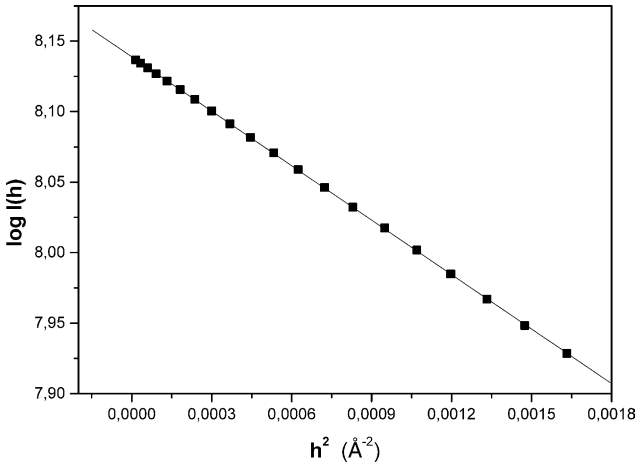


Fig. 2. Guinier plot of the SAXS intensities, $I(h)$, for human PNP. The straight line was obtained by least-squares fitting in the region $h^2 < 1.8.10^{-3}$ Å^2.

PNPs from most mammalian and some of the bacterial sources appear to be trimeric although dimeric quaternary structures have been proposed for the human enzyme [9]. Analysis of the crystallographic structures of human PNP indicates a trimeric structure (PDB access codes: 1ULA, 1ULB, 1M73, and 1PWY) [6–8,15]. However, in a number of instances the quaternary structure observed in the crystalline state is not conserved in solution [3]. Furthermore, in the case of human PNP the low pH used in the crystallization condition [16] may generate differences in the spatial configuration of the macromolecular subdomains observed in the crystal when compared to the structure in physiological pH. In addition, since the active site of the PNP is located near the interface of two subunits within the trimer, the precise information about the biological unit in solution is of capital importance to guide the structure-based design of inhibitors because its target is a structure as close as possible to the structure found in the physiological conditions, where the drug will act. Up to now all structure-based designs of PNP inhibitors have used the low-resolution structures of human PNP (PDB access codes: 1ULA and 1ULB) and consider the trimer as the target for molecular modeling studies.

Three families of theoretical models, based on the high-resolution crystallographic structure (PDB access code: 1M73) [7], were used to determine the theoretical distance distribution function, $p_{\text{theo}}(r)$, using the program MULTIBODY and then compared with the experimental function $p(r)$ determined using the ITP program from SAXS data. Figs. 3A–C show structural models and the experimental distribution function against the theoretical distribution function for the monomer, dimer, and trimer, of human PNP, respectively.

The atomic coordinates for monomeric structure were obtained from the asymmetric unit content of the crystallographic structure of human PNP solved at 2.3 Å resolution [7]. Previous statistical analysis of low-resolution docking indicated that gross structural features of protein–protein interactions could be identified for a significant percentage of protein complexes [1]. Therefore, the low-resolution protocol of the GRAMM program [10] was used to generate the dimeric models. A total of 100 models for the dimeric structure were built, only the complex, which generated the highest correlation coefficient between theoretical and experimental distance distribution function is shown in Fig. 3B. The trimeric structure was built applying two successive rotations of 120° along z-axis on the atomic coordinates of the monomer. The radii of gyration for the structural models are 18.7, 26.5, and 28.7 Å, for the monomer, dimer, and trimer, respectively.

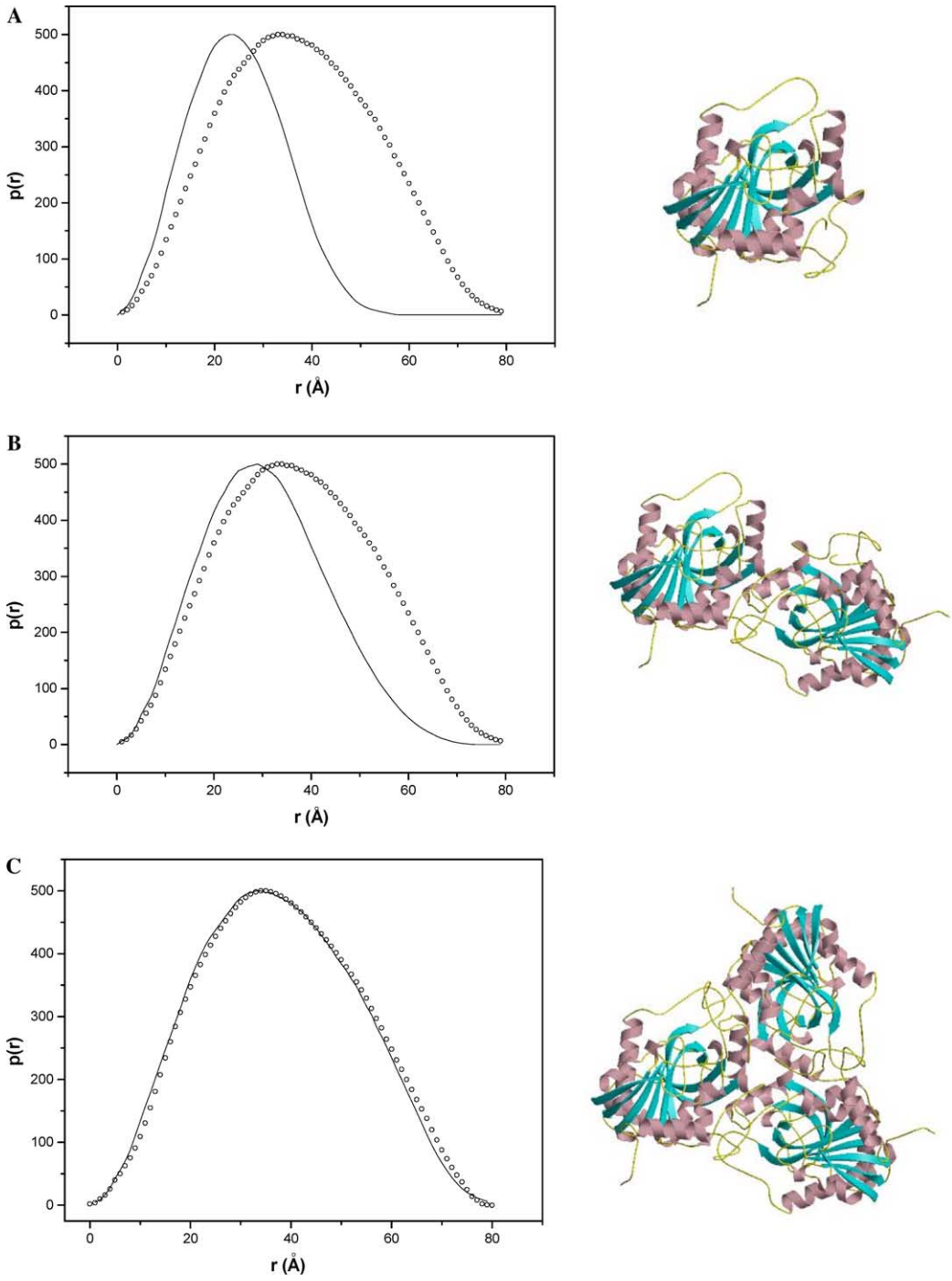


Fig. 3. Proposed structural models for human PNP and the corresponding theoretical (continuous line) and experimental (dotted line) distance distribution functions, $p(r)$, for: (A) monomer, (B) dimer, and (C) trimer. The model figures were generated by MOLSCRIPT [27] and Raster3D [28].

The value CC lies between −1 and 1. It has a value of 1, when the data points lie on a perfect straight line with positive slope. If the data points lie on a perfect straight line with negative slope, then CC has the value −1 [14]. The correlation coefficient between theoretical and experimental distance distribution functions ranges from 0.591 to 0.995, and the highest correlation coefficient was obtained for the trimeric structure, which also presented the radius of gyration closer to the experimental radius of gyration.

The contact area at interface between each subunit in the PNP trimer is 1124 $Å^2$, which indicates that the subunits are strongly bound to each other. Fig. 4 shows the electrostatic potential surface at subunit interface of the trimeric structure generated using GRASP [17]. Analysis of the electrostatic potential surface at the

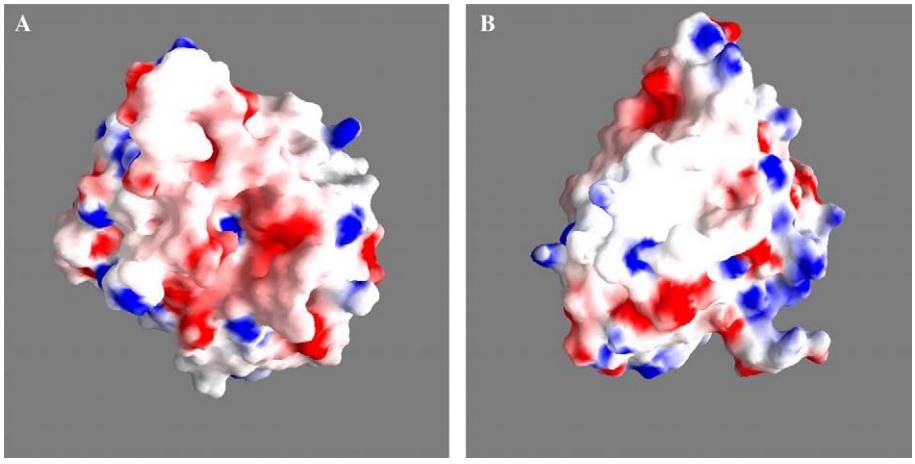


Fig. 4. Electrostatic potential surface at subunit interface of human PNP, calculated with GRASP [17] and shown from −10 to +10 kT. Uncharged regions are white.

subunit interface indicates good shape complementarity and some charge complementarity; however, most of the contacts are hydrophobic and involve residues Tyr88, Phe141, Phe159, Phe200, and Leu209.

The trimeric PNP structure has been extensively used for structure-based studies of PNP inhibitors [5,6,18–25]. However, the quaternary structure of human PNP in solution and in physiological pH has not been previously investigated using low-resolution methods, such as SAXS. The present analysis of the SAXS experiments integrated with geometric docking simulation strongly indicates that human PNP is a trimer in solution, the agreement found between the experimental and theoretical $p(r)$ functions for the trimer suggests that structure in solution adopts approximately the same conformation identified in the high-resolution crystallographic structure (PDB access code: 1M73) [7]. The radius of gyration determined for the trimeric structure is slightly smaller than that determined from the Guinier plot ($\log I(h)$ versus h^2) of the desmeared scattering function. The possible reasons for this discrepancy may be the cryogenic conditions used to solve the high-resolution structure of human PNP and the absence of solvents in the theoretical model.

The integration of a high-efficient algorithm for geometric docking with SAXS experiments allowed the investigation of the possible quaternary structures not observed in the crystalline state, such as the putative PNP dimeric structure [9]. The procedure adopted to analyze the interaction between PNP subunits can be used for other protein complexes. The main applications of the present methodology are: (1) analysis of interactions between biological macromolecules using structural models obtained from crystallography or NMR, (2) validation of structural models obtained from molecular modeling [26] of complexes of biological macromolecules, and (3) analysis of complexes of biological macromolecules in conditions closer to the biological environment.

Geometric docking simulations may be omitted from the strategy if the atomic coordinates for the complexes are available. We are applying the procedure, here described, to assess the quaternary structure of a number of protein complexes, such as hemoglobins, PNPs, and crotoxin.

Acknowledgments

This work was supported by grants from FAPESP (SMOLBNet, Proc. Num. 01/07532-0), CNPq, CAPES and Instituto do Millenium (CNPq-MCT). W.F.A. (CNPq, 300851/98-7), M.S.P. (CNPq, 500079/90-0), and L.A.B. (CNPq, 520182/99-5) are researchers for the Brazilian Council for Scientific and Technological Development.

References

[1] A. Tovchigrechko, C.A. Wells, I.A. Vakser, Docking of protein models, Protein Sci. 11 (2002) 1888–1896.
[2] N. Jing, C. Marchand, J. Liu, R. Mitra, M.E. Hogan, Y. Pommier, Mechanism of inhibition of HIV-1 integrase by G-tetrad-forming oligonucleotides in vitro, J. Biol. Chem. 275 (2000) 21460–21467.
[3] D.I. Svergun, C. Barberato, M.H.J. Koch, L. Fetler, P. Vachette, Large differences are observed between the crystal and solution quaternary structures of allosteric aspartate transcarbamylase in the R state, Proteins Struct. Funct. Genet. 27 (1997) 110–117.
[4] J.R. Olivieri, A.F. Craievich, The subdomain structure of human serum albumin in solution under different pH conditions studied by small angle X-ray scattering, Eur. Biophys. J. 24 (1995) 77–84.
[5] J.A. Montgomery, Purine nucleoside phosphorylase: a target for drug design, Med. Res. Rev. 13 (1993) 209–228.
[6] S.E. Ealick, Y.S. Babu, C.E. Bugg, M.D. Erion, W.C. Guida, J.A. Montgomery, J.A. Secrist III, Application of crystallographic and modeling methods in the design of purine nucleoside phosphorylase inhibitors, Proc. Natl. Acad. Sci. USA 91 (1991) 11540–11544.

[7] W.F. de Azevedo Jr., F. Canduri, D.M. Santos, R.G. Silva, J.S. Oliveira, L.P.S. Carvalho, L.A. Basso, M.A. Mendes, M.S. Palma, D.S. Santos, Crystal structure of human purine nucleoside phosphorylase at 2.3 Å resolution, Biochem. Biophys. Res. Commun. 308 (2003) 545–552.
[8] D.M. Santos, F. Canduri, J.H. Pereira, M.V.B. Dias, R.G. Silva, M.A. Mendes, M.S. Palma, L.A. Basso, W.F. de Azevedo, D.S. Santos, Crystal structure of human purine nucleoside phosphorylase complexed with acyclovir, Biochem. Biophys. Res. Commun. 308 (2003) 553–559.
[9] A.S. Lewis, B.A. Lowy, Human erythrocytes purine nucleoside phosphorylase: molecular weight and physical properties, J. Biol. Chem. 254 (1979) 9927–9932.
[10] E. Katchalski-Katzir, I. Shariv, M. Eisenstein, A.A. Friesem, C. Aflalo, I.A. Vakser, Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques, Proc. Natl. Acad. Sci. USA 89 (1992) 2195–2199.
[11] O. Glatter, in: O. Glatter, O. Kratky (Eds.), Small Angle X-Ray Scattering, Academic Press, London, 1982.
[12] R.G. Silva, L.P. Carvalho, J.S. Oliveira, C.A. Pinto, M.A. Mendes, M.S. Palma, L.A. Basso, D.S. Santos, Cloning, overexpression, and purification of functional human purine nucleoside phosphorylase, Protein Expr. Purif. 27 (2003) 158–164.
[13] A. Guinier, G. Fournet, Small-Angle Scattering of X-rays, Wiley, New York, 1955.
[14] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipes in FORTRAN. The Art of Scientific Computing, second ed., Cambridge University Press, New York, 1992.
[15] S.E. Ealick, S.A. Rule, D.C. Carter, T.J. Greenhough, V. Babu, W.J. Cook, J. Habash, J.R. Helliwell, J.D. Stoeckler, R.E. Parks Jr., F. Chen, C.E. Bugg, Three-dimensional structure of human erythrocytic purine nucleoside phosphorylase at 3.2 Å resolution, J. Biol. Chem. 265 (1990) 1812–1820.
[16] W.J. Cook, S.E. Ealick, C.E. Bugg, J.D. Stoeckler, R.E. Parks Jr., Crystallization and preliminary X-ray investigation of human erythrocytic purine nucleoside phosphorylase, J. Biol. Chem. 256 (1981) 4079–4080.
[17] A. Nicholls, K. Sharp, B. Honig, Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons, Proteins Struct. Funct. Genet. 11 (1991) 281–296.
[18] P.W.K. Woo, C.R. Kostlan, J.C. Sircar, M.K. Dong, R.B. Gilbertsen, Inhibitors of human purine nucleoside phosphorylase. Synthesis and biological activities of 8-amino-3-benzylhypoxanthine and related analogues, J. Med. Chem. 35 (1992) 1451–1457.
[19] J.-W. Chern, H.-Y. Lee, C.-S. Chen, Nucleosides. 5. Synthesis of guanine and formycin B derivatives as potential inhibitors of purine nucleoside phosphorylase, J. Med. Chem. 36 (1993) 1024–1031.
[20] M.D. Erion, S. Niwas, J.D. Rose, S. Ananthan, M. Allen, J.A. Secrist III, Y.S. Babu, C.E. Bugg, W.C. Guida, S.E. Ealick, J.A. Montgomery, Structure-based design of inhibitors of purine nucleoside phosphorylase. 3. 9-Arylmethyl derivatives of 9-deazaguanine substituted on the methylene group, J. Med. Chem. 36 (1993) 3771–3783.
[21] J.A. Secrist III, S. Niwas, J.D. Rose, Y.S. Babu, C.E. Bugg, M.D. Erion, W.C. Guida, S.E. Ealick, J.A. Montgomery, Structure-based design of inhibitors of purine nucleoside phosphorylase. 2. 9-Alicyclic and 9-heteroalicyclic derivatives of 9-deazaguanine, J. Med. Chem. 36 (1993) 1847–1854.
[22] W.C. Guida, R.D. Elliot, H.J. Thomas, J.A. Secrist III, Y.S. Babu, C.E. Bugg, M.D. Erion, S.E. Ealick, J.A. Montgomery, Structure-based design of inhibitors of purine nucleoside phosphorylase. 4. A study of phosphate mimics, J. Med. Chem. 37 (1994) 1109–1114.
[23] S. Niwas, P. Chand, V.P. Pathak, J.A. Montgomery, Structure-based design of inhibitors of purine nucleoside phosphorylase. 5. 9-Deazahypoxanthines, J. Med. Chem. 37 (1994) 2477–2480.
[24] P.E. Morris, A.J. Elliott, S.P. Walton, C.H. Williams, J.A. Montgomery, Synthesis and biological activity of a novel class of purine nucleoside phosphorylase inhibitors, Nucleosides Nucleotides Nucleic Acids 19 (2000) 379–404.
[25] A. Andricopulo, R.A. Yunes, Structure–activity relationships for a collection of structurally diverse inhibitors of purine nucleoside phosphorylase, Chem. Pharm. Bull. 49 (2001) 10–17.
[26] A. Sali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, J. Mol. Biol. 234 (1993) 779–815.
[27] P.J. Kraulis, MOLSCRIPT: a program to produce both detailed and schematic plots of proteins, J. Appl. Cryst. 24 (1991) 946–950.
[28] E.A. Merritt, D.J. Bacon, Raster3D: photorealistic molecular graphics, Methods Enzimol. 277 (1997) 505–524.

ANEXO II

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Protein Expression and Purification 34 (2004) 118–125
Protein Expression & Purification
www.elsevier.com/locate/yprep

Selection of an *Escherichia coli* host that expresses mutant forms of *Mycobacterium tuberculosis* 2-*trans* enoyl-ACP(CoA) reductase and 3-ketoacyl-ACP(CoA) reductase enzymes

Simone S. Poletto,[a] Isabel O. da Fonseca,[a] Luiz P.S. de Carvalho,[a] Luiz A. Basso,[a,*] and Diógenes S. Santos[b,*]

[a] *Departamento de Biologia Molecular e Biotecnologia, Universidade Federal do Rio Grande do Sul, Avenida Bento Gonçalves, 9500, Porto Alegre, RS 91501-970, Brazil*

[b] *Faculdade de Farmácia Pontifícia, Universidade Católica do Rio Grande do Sul, Porto Alegre, RS 90619-900, Brazil*

Received 14 August 2003, and in revised form 17 October 2003

Abstract

Tuberculosis (TB) still remains a worldwide health concern. Efforts to understand the complex biology of *Mycobacterium tuberculosis*, the causative agent of TB, are important for new antitubercular drug development. Despite the completion of the genome sequence and the development of new genetic tools to manipulate this organism, the availability of sufficient amounts of mycobacterial proteins still remains an essential and laborious step to study the biochemical features of this pathogen. The T7-RNA polymerase-based pET system has been largely employed to express mycobacterial proteins in *Escherichia coli*, but it presents some limitations. To overcome problems with unstable expression of an *M. tuberculosis inh*A-encoded enoyl reductase mutant protein and lack of expression of two *mab*A-encoded ketoacyl reductase mutants, a sub-population of *E. coli* BL21(DE3) host cells was selected from a small-opaque colony. This empirically selected host, named BL21(DE3)NH, allowed stable expression of these mutant proteins. Although the mechanism that led the BL21(DE3)NH host to express the recombinant mutant proteins remains unknown, the persistent phenotype points to a stable genetic switch. This genetic alteration resulted in a tight control of the highly processive T7 RNA polymerase. Moreover, the absolute requirement for IPTG to obtain protein expression in the BL21(DE3)NH host cells suggests that no inherent defect in the transcriptional activity of the T7 promoter is present. Empirical host selection requires no further genetic manipulation of recombinant plasmids and may represent a means of obtaining tailor-made *E. coli* strains that overcome toxic effects associated with heterologous protein expression.

Keywords: Protein expression; *Escherichia coli*; Host selection; Mycobacteria

Tuberculosis (TB), caused by *Mycobacterium tuberculosis*, remains one of the deadliest diseases in the world. It is estimated that 8.2 million new TB cases occurred worldwide in the year 2000, 1.8 million deaths occurred in the same year, and more than 95% of those were in developing countries [1]. Possible factors underlying the resurgence of TB include the HIV epidemic, increase in the homeless population, and decline in health care structures and national surveillance [2]. Another contributing factor is the evolution of multidrug-resistant strains of *M. tuberculosis* (MDR-TB), defined as resistant to at least isoniazid and rifampicin, which are the most effective first-line drugs [3]. Treatment of MDR-TB requires the administration of second-line drugs (amikacin, kanamycin, capreomycin, cycloserine, *para*-aminosalicylic acid, ethionamide, and fluoroquinolones) that are more toxic and less effective and are given for at least three times as long and at 100 times the cost of basic short-course chemotherapy regimens [4]. Hence, new antimycobacterial agents are needed to improve the treatment of MDR-TB, to shorten the treatment course to improve patient compliance, and to

* Corresponding author. Fax: +55-51-33167309.
E-mail addresses: labasso@dna.cbiot.ufrgs.br (L.A. Basso), diogenes@pucrs.br (D.S. Santos).

1046-5928/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.pep.2003.10.009

provide for more effective treatment of latent tuberculosis infection [5].

The study of biology, pharmacology, and host–pathogen interactions is important for the development of new drugs as well as vaccines against *M. tuberculosis*. The completion of the genome sequence of *M. tuberculosis* [6] and availability of genetic tools to manipulate this organism [7] have accelerated these studies. Despite these advances, the availability of sufficient amounts of proteins of *M. tuberculosis* still remains an essential and laborious step to study the biochemical features of this pathogen. This fact is due to (i) the slow-growth of the bacillus whose generation time is 24 h; (ii) since aerosols containing bacteria are the way of spread of the disease, a P3 facility is required to grow the bacillus culture; (iii) and the highly lipophilic envelope of mycobacteria that represents an obstacle to protein extraction and purification. On the other hand, the improvements in the recombinant DNA technology, the advances in molecular genetic studies in mycobacteria and, as previously pointed out, the availability of the complete genome sequence have facilitated the cloning and heterologous expression of mycobacterial proteins. *Escherichia coli*[1] and *Mycobacterium smegmatis* (a fast-growing bacterium that belongs to the same genus of *M. tuberculosis*) are the first choices for hosts [8]. *E. coli*, in particular, remains one of the most attractive systems for heterologous protein expression due to its ability to grow rapidly and at high density on inexpensive substrates, its well-characterized genetics, and the availability of an increasingly large number of cloning vectors and mutant host strains [9]. The pET system (Novagen) developed by Studier and Moffat [10] has been one of the most widely used systems for expression of mycobacterial proteins in *E. coli* [11–15]. In this system, target genes are positioned downstream of the bacteriophage T7 late promoter on medium copy number plasmids. The highly processive T7 RNA polymerase gene (DE3 lysogen) of production hosts is placed under control of the IPTG-inducible *lac*UV5 promoter. However, expression of membrane proteins and proteins toxic to the host cells [16], formation of inclusion bodies [17] or complete absence of expression are amongst the difficulties encountered with mycobacterial proteins.

The *inh*A-encoded 2-*trans* enoyl-ACP (CoA) reductase (ENR) and *mab*A-encoded 3-ketoacyl-ACP (CoA) reductase (KAR) are enzymatic components of the Type II fatty acid synthase (FAS-II) of mycobacteria and are thus involved in mycolic acid biosynthesis [18]. The mycobacterial NADH-dependent ENR is the target for isoniazid [11,19], the most prescribed drug to treat tuberculosis, making it an excellent target for TB drug development. Another reductive step of the mycobacterial FAS-II system is catalyzed by the NADPH-dependent KAR enzyme [20]. The FAS-II system enzymes have been shown to be essential for mycobacterial growth and configured into high throughput screening for identification of new drugs against TB [21]. The amino acid residues tyrosine-158 and lysine-165 have been implicated in the catalytic mechanism of *M. tuberculosis* ENR [11,22,23]. Moreover, the side chain of Y158 of *M. tuberculosis* ENR has been shown to undergo a rotation upon binding of triclosan, an enzyme inhibitor, to form a hydrogen bond with the hydroxyl group of this inhibitor [18]. The mycobacterial KAR and ENR enzymes are members of the short-chain dehydrogenase/reductase (SDR) family and thereby display the amino acid signature of this family, S–$(X)_{12}$–Y–$(X)_3$–K [24]. As part of our efforts to develop novel antitubercular drugs based on rational design, we are studying the catalytic mechanism of *M. tuberculosis* ENR and KAR enzymes, identifying amino acid residues involved in either catalysis or substrate binding, and trying to solve their three-dimensional structures. Accordingly, seven mutant proteins were produced for the putative catalytic amino acids (ENR: Y158F, K165A, K165Q; KAR: S140T, S140A, Y153F, and K157A), cloned into the pET system, and transformed into *E. coli* BL21(DE3) host strain. However, the Y158F ENR mutant failed to show stable protein expression in this system, and no expression could be obtained for the Y153F and K157A KAR mutant enzymes.

Miroux and Walker [16] have selected mutant hosts derived from *E. coli* BL21(DE3) that allowed protein expression at higher levels than the parent host. The mutant host C41(DE3) has been employed to express *E. coli* inner membrane enzyme acyl acyl carrier protein synthase [25] and *M. tuberculosis* acyl carrier protein (AcpM) and malonyl-CoA:AcpM transacylase [15]. Building on the strategy adopted by Miroux and Walker [16], we selected a sub-population of *E. coli* BL21(DE3), named BL21(DE3)NH, that was able to express theY158F ENR, Y153F KAR, and K157A KAR mutant proteins, which could not be expressed in the *E. coli* BL21(DE3) parent strain. Here, we report site-directed mutagenesis to obtain theY158F ENR, Y153F KAR, and K157A KAR mutants, and selection of a host cell that was able to express these mutant proteins in soluble form. Scanning electron microscopy has been performed to evaluate changes in BL21(DE3)NH morphology as compared to the BL21(DE3) parent strain.

[1] *Abbreviations used: E. coli*, *Escherichia coli*; ENR, enoyl reductase; InhA, enoyl-ACP (CoA) reductase from *Mycobacterium tuberculosis*; KAR, 3-ketoacyl-ACP (CoA) reductase; LB, Luria–Bertani medium; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; IPTG, isopropyl-β-D-thiogalactopyranoside; MabA, 3-ketoacyl-ACP (CoA) reductase from *Mycobacterium tuberculosis*; PCR, polymerase chain reaction; bp, base pairs; WT, wild type; EtdBr, ethidium bromide; SEM, scanning electron microscopy; OsO_4, osmium tetraoxide; expression plasmid names consist of the name of the parent plasmid vector followed by the recombinant DNA insert they encode.

Materials and methods

Materials

The pET-23d(+)::*inh*A recombinant plasmid was a gift from Dr. John S. Blanchard at Albert Einstein College of Medicine (AECOM, Bronx, NY) and pET-3d::*mab*A recombinant plasmid was a gift from Dr. William R. Jacobs Jr. at AECOM (Bronx, NY). pET-23d(+), pET-23a(+), BL21(DE3) *E. coli* strain, and Perfect Protein Marker were purchased from Novagen; Luria–Bertani medium (LB), oligonucleotides, T4 DNA ligase, IPTG, DH10B cells, and DNA and protein molecular weight standards were from Gibco-BRL; carbenicillin was from Sigma; *Bam*HI, *Hin*dIII, *Nco*I, and *Nde*I restriction enzymes were from Boehringer–Mannheim and New England Biolabs; *Pfu* DNA polymerase was from Stratagene; Agarose was from SEA KEM GTG; Zero Blunt PCR Cloning Kit pCR-Blunt and Top10 One Shot were from Invitrogen; Quantum Prep Plasmid Miniprep kit, QIAEX II Agarose Gel Extraction kit, and QIAprep Spin Miniprep kit were from Qiagen; Thermo Sequenase radiolabeled terminator cycle sequencing kit was from Amersham Life Science; and SDS–polyacrylamide gels and Tris–glycine running buffer were from Bio-Rad.

Cloning of wild-type mabA gene into pET23a(+) vector

The synthetic oligonucleotide primers used in this study are listed in Table 1. The primers mabA5, containing a *Nde*I restriction site (bold), and mabA3, containing a *Hin*dIII restriction site (bold), were used to amplify the *mab*A gene from the pET-3d::*mab*A recombinant plasmid using *Pfu* DNA polymerase at standard conditions. The PCR product was purified from agarose gel using QIAEX II Kit and inserted into a pCR-blunt vector. *E. coli* Top 10 One-Shot competent cells were transformed with the recombinant plasmid.

Table 1
Primers used for amplification and mutagenesis of *inh*A and *mab*A genes from *M. tuberculosis* H37Rv

Primers	Restriction sites	Sequence
P1	*Nco*I	5′-att gaa **cca tgg** cag gac tgc tgc acg-3′
JSB02	*Bam*HI	5′-g**cg gat cc**g cta gag caa ttg ggt gtg cgc-3′
Y158a	—	5′-atg ccg gcc ttc aac tgg atg-3′
Y158b	—	5′-cat cca gtt gaa ggc cgg ca t-3′
mabA5	*Nde*I	5′-att **cat atg** act gcc aca gcc act gaa gg-3′
mabA3	*Hin*dIII	5′-t **aag ctt** tca gtg gcc cat acc cat gcc-3′
Y153F5	—	5′-cag gcc aac ttc gca gcc tcc-3′
Y153F3	—	5′-gga ggc tgc gaa gtt ggc ctg-3′
K157A5	—	5′-gca gcc tcc gcg gcc gga gtg-3′
K157A3	—	5′-cac tcc ggc cgc gga ggc tgc-3′

The sequences underlined correspond to degenerate codons and the sequences in bold correspond to restriction sites.

Plasmid DNA was extracted using Quantum Prep Plasmid Miniprep Kit (Qiagen). The insert was removed by *Nde*I and *Hin*dIII double digestion and ligated into a pET23a(+) digested with the same restriction enzymes. The integrity of the *mab*A gene was confirmed by DNA sequencing.

Site-directed mutagenesis of inhA and mabA genes

Site-directed mutations were introduced in the *inh*A and *mab*A genes using the method described by Ho et al. [26]. A list of the primers used for all amplification steps is given in Table 1. It should be noted that only the primers to obtain theY158F ENR, Y153F KAR, and K157A KAR proteins by site-directed mutagenesis are described here, since these were the mutants that showed unstable or no protein expression. The experimental protocol described here is exemplified by Y158F *inh*A mutant gene production. First, two independent PCRs using *Pfu* DNA polymerase were performed using primers P1 and Y158b—reaction 1, and JSB02 and Y158a—reaction 2, and the plasmid pET23d(+)::*inh*A as template. The PCR program was as follows: one step of 98 °C for 3 min, 35 cycles at 98 °C for 1 min, 65 °C for 1 min, and 72 °C for 2 min, followed by a final extension step at 72 °C for 2 min, using a DNA Thermal Cycler (PTC200, MJ, USA). The PCR products of these reactions were purified by electrophoresis on 1% agarose gel and DNA was extracted using the QIAEX II Kit. A second PCR was performed using the products of the reactions 1 and 2 as templates and the primers P1 and JSB02. The overlap PCR conditions were identical to those used in the first round of amplification. The PCR-amplified band was purified and cloned as described in the previous section, yielding pET23d(+)::*inh*AY158F recombinant plasmid. The primers used to make the *mab*A mutants are listed in Table 1 and followed the same protocol just described, yielding pET23a(+)::*mab*AY153F and pET23a(+)::*mab*AK157A recombinant plasmids. The sequence of the DNA inserts in all recombinant plasmids was confirmed by the dideoxynucleotide chain termination method using the Thermo Sequenase radiolabeled terminator cycle sequencing kit (Amersham Biosciences).

Protein expression of WT and ENR mutants in BL21(DE3)

To assess the expression of these constructs, BL21(DE3) electrocompetent cells were transformed with the following recombinant plasmids: pET23-d(+)::*inh*A and pET23d(+)::*inh*AY158F and grown in LB medium containing carbenicillin ($50\,\mu g\,ml^{-1}$) at various times after induction with IPTG (1 mM) at 37 °C. Control experiments were performed under the same experimental conditions, except that *E. coli*

BL21(DE3) host cells harbored the expression vector lacking the target gene. Cells were harvested (14,000*g* for 10 min) at 4, 8, 12, 16, 20, and 24 h after induction, lysed by sonication, and cell debris was removed by centrifugation (20,000*g* for 20 min). The resulting soluble protein content was analyzed by 12% SDS–PAGE.

Stable host production

Escherichia coli BL21(DE3) electrocompetent cells were transformed with pET23d(+)::*inh*AY158F, plated on LB solid medium containing carbenicillin (50 μg ml^{-1}), and grown overnight at 37 °C, yielding subpopulations of large-translucent and small-opaque colonies. These two types of colonies were examined for their ability to express Y158F ENR protein in liquid culture, using the protocol of IPTG induction as previously described for wild-type ENR. The degree of protein expression, as monitored by SDS–PAGE, was dependent on colony morphology, with smaller colonies expressing the desired protein, whereas larger colonies did not. Since sub-culturing may eventually result in loss of phenotype, *E. coli* BL21(DE3) host cells were transformed with pET23d(+)::*inh*AY158F recombinant plasmid and plated on LB–carbenicillin overnight at 37 °C. A single small colony was used to inoculate 5 ml LB–carbenicillin medium, grown to midlog phase (0.4–0.6 OD_{600}) at 37 °C, and induced with 1 mM IPTG for 3 h. At midlog phase of growth, an aliquot was withdrawn to prepare a glycerol culture stock. The remaining of cells were harvested by centrifugation at 14,000*g* for 10 min. Pelleted material was solubilized in 400 μl of 20 mM Pipes buffer, pH 7.3, disrupted by sonication, and the resulting crude extract was centrifuged at 14,000*g* for 10 min. Expression of the recombinant protein in soluble form was confirmed by 12% SDS–PAGE. The glycerol stock of this recombinant protein expressing host was cured of pET-23d(+)::*inh*AY158F by growth in LB liquid medium in the absence of carbenicillin for 15 days. Each day, a portion of the culture (5 μl) was used to inoculate 5 ml of LB. After 15 days, small-colony derived BL21(DE3) cells lacking the recombinant plasmid arose, which was verified both by its inability to grow on LB solid medium containing carbenicillin and no recovery of recombinant plasmid from the cells. To confirm the ability of the new host, henceforth named BL21(DE3)NH (NH standing for New Host), to confer stable overexpression of Y158F ENR protein, BL21(DE3)NH electrocompetent cells were transformed with either pET-23d(+) plasmid as control or pET-23d(+)::*inh*AY158F recombinant plasmid, plated on LB-agar plate containing carbenicillin (50 μg ml^{-1}), and grown overnight at 37 °C. Seven single colonies were separately used to inoculate 5 ml LB–carbenicillin liquid medium, grown to an OD_{600} of 0.4–0.6, induced with 1 mM IPTG, and grown for 3 h at 37 °C after induction. The cells were harvested by centrifugation at 14,000*g* for 10 min, pelleted material was solubilized in 400 μl of 20 mM Pipes buffer, pH 7.3, and disrupted by sonication. Soluble proteins were analyzed by SDS–PAGE.

Protein expression of mycobacterial KAR mutants in BL21(DE3)NH host cells

To address the ability of this new host to express proteins that, in our hands, have failed to do so in the BL21(DE3) parent host, BL21(DE3), and BL21(DE3) NH electrocompetent cells were transformed with the following plasmids: pET-23a(+), pET-23a(+)::*mab*AY 153F and pET-23a(+)::*mab*AK157A, plated on LB–carbenicillin medium, and incubated overnight at 37 °C. Single colonies were used to inoculate 5 ml LB medium containing carbenicillin (50 μg ml^{-1}) and 1 mM of IPTG was added to the liquid cultures as they reached an OD_{600} between 0.4 and 0.6. The cultures were grown at 37 °C for 8 h as determined for the wild-type KAR protein in BL21(DE3) (data not shown). Cells were harvested by centrifugation (14,000*g* for 10 min) and pellets were solubilized in 400 μl of 50 mM Tris–HCl, pH 7.8, buffer, disrupted by sonication, and clarified by centrifugation (14,000*g* for 10 min). The soluble fraction of the cellular extracts was analyzed by SDS–PAGE.

Scanning electronic microscopy

The procedure for scanning electronic microscopy of *E. coli* was carried out as described elsewhere [27,28]. Slides of *E. coli* BL21(DE3) and BL21(DE3)NH were obtained by cutting single colonies (1 mm^2 and 0.5 mm of thickness) from the LB agar plates, followed by a 2-h immersion in Karmovisk solution at room temperature to chemically stabilize the sample and an additional fixation with 1% OsO_4. Samples were then washed 10 times with cacodylate buffer solution. For dehydration, samples were washed with 30–100% ethanol and, subsequently, with aqueous carbonic acid. The specimens were coated with gold by the sputtering process to obtain a good secondary electron production for image formation. Micrographs were taken using Zeiss DMS 940A and operated in electron emission mode at 20 kV accelerating voltage. Samples were observed in 3000× amplification at 12 mm distance.

Results and discussion

Site-directed mutagenesis of both *inh*A and *mab*A genes was performed by the overlap extension method [26] and mutations were confirmed by DNA sequencing. The *inh*A mutant was cloned into pET-23d(+) and *mab*A mutants were inserted into pET-23a(+) expression plasmid.

Overexpression of mycobacterial K165A, K165Q, and Y158F ENR mutant proteins is shown in Fig. 1. However, mycobacterial Y158F ENR mutant protein overexpression was not stable, that is, not all colonies picked up at random expressed the recombinant protein. The appearance of two types of colonies with distinct morphologies (large and small) of BL21(DE3) transformed with pET-23d(+)::*inhA*Y158F plasmid was observed on LB–carbenicillin plates incubated overnight at 37 °C. The small colonies, which have diameters approximately 30% smaller than those of large colonies, were opaque, while the large colonies were translucent. In addition, the small-opaque colonies expressed the desired proteins while the large-translucent colonies did not. Variations in colony morphology of *E. coli* hosts have been associated with detection of recombinant colonies [29] and expression of the recombinant gene [30]. Moreover, empirical selection has been employed by Miroux and Walker [16] to tailor expression hosts to overcome toxic effects associated with protein overexpression. Building on Miroux and Walker strategy, a sub-population of BL21(DE3) cells having a small-opaque phenotype that expressed the Y158F ENR mutant protein was selected, isolated, and cured. The selected host cell, named BL21(DE3)NH, transformed with pET-23d(+)::*inh*AY158F and plated on LB–carbenicillin solid medium, yielded a population of small colonies only. To evaluate whether protein expression of the recombinant protein requires IPTG induction, seven isolated colonies were picked up at random, suspended in 200 µl LB–carbenicillin medium and 50 µl of the resulting suspension used to inoculate 5 ml LB–carbenicillin medium, and the experimental protocol described in Material and methods was followed. Fig. 2A shows a representative result for a single colony, suggesting that expression of Y158 ENR in the BL21(DE3)NH host can be achieved only in the presence of the inducer. Fig. 2B shows that all seven colonies express the recombinant Y158F ENR mutant protein in soluble form upon IPTG induction.

A protocol had been designed to optimize protein expression of mycobacterial wild-type KAR, involving addition of 1 mM IPTG to an LB–carbenicillin liquid culture ($0.4 < OD_{600\,nm} < 0.6$) of BL21(DE3) and further growth for 8 h at 37 °C. The mycobacterial Y153F and K157A KAR mutants, however, did not show any expression under these experimental conditions (Fig. 3A). To test the ability of the empirically selected host cells to express these mutants, BL21(DE3)NH cells were transformed with pET23a(+)::*mab*AY153F and pET23a(+)::*mab*AK157A recombinant plasmids, and pET23a(+) plasmid (control). LB–carbenicillin liquid medium was inoculated with single colonies, and cell growth and induction was carried out as for the

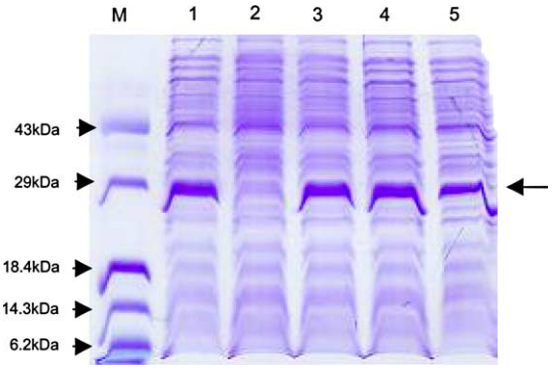


Fig. 1. SDS–PAGE (12%) of the soluble fraction of the cell extracts expressing *M. tuberculosis* ENR proteins upon IPTG induction in *E. coli* BL21(DE3) host cells. Lane M, Protein molecular weight standards (Gibco-BRL: insulin α and β chains, 3 kDa; bovine trypsin inhibitor, 6.2 kDa; lysozyme, 14.3 kDa; β-lactoglobulin, 18.4 kDa; carbonic anhydrase, 29 kDa; and ovalbumin, 43 kDa); lane 1, wild-type ENR; lane 2, pET-23d(+); lane 3, K165A ENR; lane 4, K165Q ENR; and lane 5, Y158F ENR. The arrow points the band corresponding to *M. tuberculosis* ENR proteins (~29 kDa).

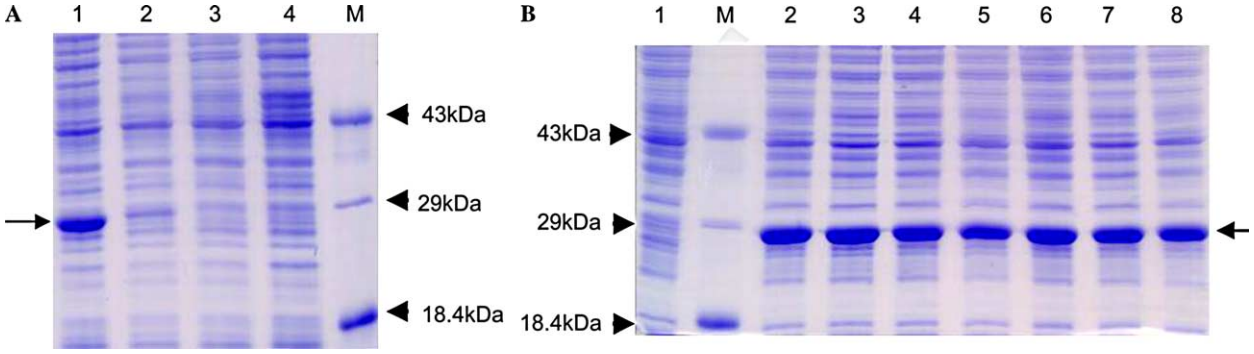


Fig. 2. SDS–PAGE (12%) of the soluble fraction of the cell extracts of BL21(DE3)NH transformed with either pET-23d(+) (control) or pET-23d(+)::*inh*AY158F recombinant plasmid. (A) Lane 1, 1 mM IPTG; lane 2, no IPTG; lane 3, control with 1 mM IPTG; lane 4, control in the absence of IPTG; and lane M, protein molecular weight standards (same as Fig. 1). (B) Lane 1, control with 1 mM IPTG; lane M, protein molecular weight standards (same as Fig. 1); and lanes 2–8, protein expression of seven colonies of BL21(DE3)NH transformed with pET-23d(+)::*inh*AY158F plasmid upon 1 mM IPTG induction. The arrows point the band corresponding to mycobacterial ENR (~29 kDa).

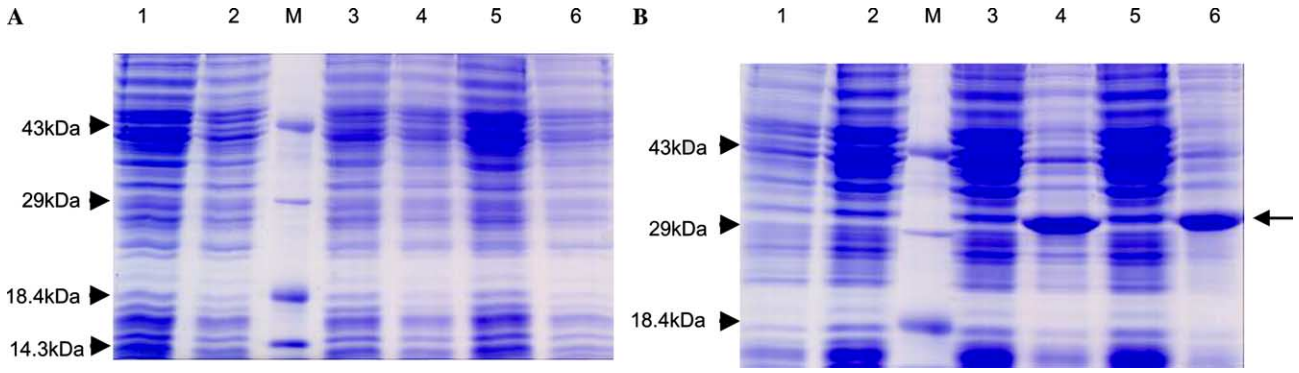


Fig. 3. SDS–PAGE (12%) of the soluble fraction of cell extracts of either BL21(DE3) or BL21(DE3)NH transformed with either pET-23a(+) (control), pET-23a(+)::*mab*AK157A, or pET-23a(+)::*mab*AY153F recombinant plasmids. (A) Plasmids transformed into BL21(DE3). Lane 1, control in the absence of IPTG; lane 2, control with 1 mM IPTG; lane M, protein molecular weight standards (same as Fig. 1); lane 3, K157A KAR in the absence of IPTG; lane 4, K157A KAR with 1 mM IPTG; lane 5, Y153F KAR in the absence of IPTG; and lane 6, Y153F KAR with 1 mM IPTG. (B) Plasmids transformed into BL21(DE3)NH. Lane 1, control in the absence of IPTG; lane 2, control with 1 mM IPTG; lane M, protein molecular weight standards (same as Fig. 1); lane 3, K157A KAR in the absence of IPTG; lane 4, K157A KAR with 1 mM IPTG; lane 5, Y153F KAR in the absence of IPTG; and lane 6, Y153F KAR with 1 mM IPTG. The arrow points to the band corresponding to the expected molecular weight of *M. tuberculosis* KAR proteins (25.6 kDa).

wild-type KAR enzyme. Cells were harvested and the resulting pellets were suspended in Tris–HCl 50 mM, pH 7.8, disrupted by sonication and the soluble fractions were analyzed by SDS–PAGE. High level of protein expression was observed in BL21(DE3)NH host cells only upon IPTG induction (Fig. 3B).

The intrinsic toxicity to *E. coli* BL21(DE3) host cells of pET vectors harboring no insert in the presence of IPTG [16] has not been observed for the BL21(DE3)NH host cells (Figs. 2 and 3B). Moreover, the BL21(DE3)NH host phenotype was stable, since it continued to give rise to small colonies on LB–carbenicillin agar plates upon transformation with the recombinant plasmids tested.

Scanning electron microscopy (SEM) results show that the *E. coli* BL21(DE3)NH cells have a distinct morphology of long bacillary shape (Fig. 4A). The elongation is evident and contrasts with the *E. coli* BL21(DE3) morphology (Fig. 4B). The BL21(DE3)NH host cell is a Gram-negative bacterium. Moreover, the biochemical properties of BL21(DE3)NH host cells were determined by standard methods for *E. coli* identification: urease production, indole production, carbohydrate fermentation, β-galactosidase activity, citrate utilization, and growth on MacConckey agar medium. All these tests confirm that BL21(DE3)NH host cell is a strain of *E. coli*.

The mechanisms of translucent to opaque phenotype, variations in colony morphology (large and small), and alteration to the long bacillary shape of BL21(DE3)NH cells remain unknown. It has been shown that opacity is correlated with a robust transcription of the recombinant gene whereas translation of recombinant RNA is not required [30]. Interestingly, uncoupling of transcription from translation has been implicated in toxicity of recombinant protein expression leading to host cells producing large amounts of the recombinant mRNA while target proteins are maintained at rather low levels [16]. In addition, expression of gene products that do not contribute to the metabolism of bacterium, under the growth conditions used in a particular experiment, has been shown to result in cumulative breakdown of rRNAs and ensuing loss of ribosomes and protein synthetic capacity [31]. Although the mechanism that led the empirically selected host to tolerate the recombinant mutant protein expression is unknown, the BL21(DE3)NH cells derived from small-opaque colonies transformed with recombinant plasmids resulted in small-opaque transformants, suggesting that the persistence of the phenotype points to a stable genetic switch.

Leaky expression has been shown to occur in the pET system [13,32,33]. The BL21(DE3)NH transformants showed an absolute requirement for IPTG induction to achieve expression of the recombinant proteins tested (Figs. 2 and 3B). Accordingly, the genetic alteration of the empirically selected host reported here appears to have resulted in a tight control of the highly processive T7 RNA polymerase that is under control of the IPTG-inducible *lac*UV5 promoter. Moreover, the IPTG absolute requirement for expression of the mycobacterial ENR and KAR proteins in the BL21(DE3)NH host cell suggests that no inherent defect in the transcriptional activity of the T7 promoter is present.

It has been suggested that the mechanism responsible for the differences in colony morphology is bacterial growth rate [29]. Although we have also found this to be generally true, any causal relationship between opacity and growth rate seems to be unwarranted because it is somewhat difficult to establish a quantitative relationship between them based on a qualitative observation (opacity).

Protein production and crystallization must be optimized if structural genomics will ever reach its goal of solving the three-dimensional structure of the whole

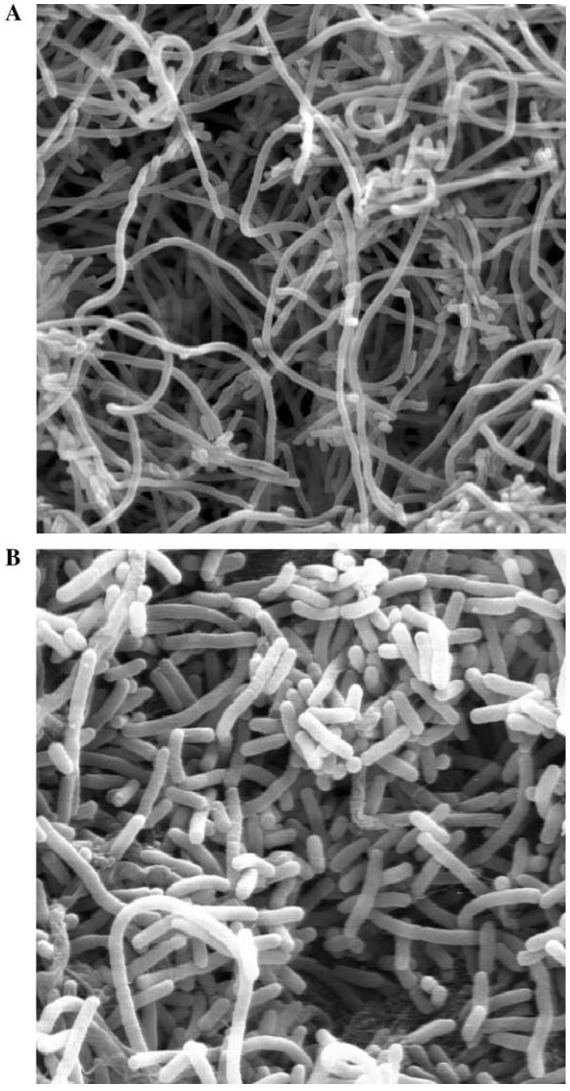


Fig. 4. Scanning electron microscopy of *E. coli* hosts observed at 3000× amplification. (A) *E. coli* BL21(DE3)NH and (B) *E. coli* BL21(DE3).

proteome encoded by a given genome [34]. Unfortunately, even when a genome can be sequenced, only up to 20% of the protein targets can produce soluble proteins under very basic experimental conditions [35]. Thus, expression of proteins in soluble form has been identified as an important bottleneck in efforts to determine biological activity and crystal structure of *M. tuberculosis* proteins [17]. A number of strategies for optimizing the yields of heterologous protein expression in *E. coli* have been described [36,37], including the use of different types of promoters, placement of transcription terminators upstream of the promoter that drives expression of the gene of interest, reduction of potential secondary-structure formation at the 5′ end of the transcript, use of translational enhancer, positioning of specific sequences in the 5′ untranslated or 3′ unstranslated regions of labile heterologous mRNAs to prolong their half-lives, altering rare codons in the target gene or coexpressing genes that encode rare tRNAs, minimizing proteolysis by mutations to avoid the *E. coli* "N-end rule" proteolytic pathway, and altering fermentation conditions such as nutrient composition, temperature, and pH. Incidentally, although wild-type KAR has nine rare codons (2× CCC–proline, 5× GGA–glycine, and 2× ATA–isoleucine) and wild-type ENR has six rare codons (2× CCC–proline, 3× GGA–glycine, and 1× AGG–arginine), they have no adverse effect on the synthesis and yield of these recombinant proteins, which corroborates the conclusion arrived by others [37] that the mere presence of rare codons in a gene does not necessarily dictate poor translation of that gene.

The *E. coli* host reported here may represent an additional tool to obtain mycobacterial soluble proteins using T7-RNA polymerase-based systems as has been the case for the C41(DE3) and C43(DE3) hosts [15,16,25] now marketed by Avidis S.A. (France). Empirical host selection requires no further genetic manipulation of recombinant plasmids and may yield strains that overcome toxic effects associated with the overexpression of heterologous proteins. As pointed out by Miroux and Walker [16], even though the host selection described here is empirical, it has the advantage that it encompasses the entire complexity of the biology of the expression system and provides a means of modifying it. The expression of *M. tuberculosis* ENR and KAR mutants will provide protein in quantities necessary for determination of the mechanisms of action of these enzymes by steady-state and pre-steady-state kinetics as well as for crystallization trials aiming at X-ray data collection. Enzymological and structural studies of *M. tuberculosis* ENR and KAR mutants should help in the design of enzyme inhibitors to be tested as new antitubercular agents.

Acknowledgments

Financial support for this work was provided by Millennium Initiative Program MCT-CNPq, Ministry of Health-Secretary of Health Policy (Brazil) to D.S.S. and L.A.B. D.S.S. and L.A.B. also acknowledge grants awarded by PADCT, CNPq, and FINEP. L.A.B. (CNPq, 520182/99-5) is a researcher for the Brazilian Council for Scientific and Technological Development. We also thank Dr. Luiz Antônio Suita de Castro and Embrapa Clima Temperado for the preparation of micrographs and laboratory facilities.

References

[1] M.C. Raviglione, The TB epidemic from 1992 to 2002, Tuberculosis 83 (2003) 4–14.

[2] B.R. Bloom, C.J.L. Murray, Tuberculosis: commentary on a reemergent killer, Science 257 (1992) 1055–1064.
[3] L.A. Basso, J.S. Blanchard, Resistance to antitubercular drugs, Adv. Exp. Med. Biol. 456 (1998) 115–144.
[4] A. Plabos-Méndez, D.K. Gowda, T.R. Frieden, Controlling multidrug-resistant tuberculosis and access to expensive drugs: a rational framework, Bull. World Health Org. 80 (2002) 489–495.
[5] R.J. O'Brien, P.P. Nunn, The need for new drugs against tuberculosis, Am. J. Respir. Crit. Care Med. 162 (2001) 1055–1958.
[6] S.T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S.V. Gordon, K. Eiglmeier, S. Gas, C.E. Barry III, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M.A. Quail, M.A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, S. Squares, R. Squares, J.E. Sulston, K. Taylor, S. Whitehead, B.G. Barrell, B.G. Deciphering, Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence, Nature 393 (1998) 537–544.
[7] M.S. Glickman, W.R. Jacobs Jr., Microbial pathogenesis of *Mycobacterium tuberculosis*: dawn of a discipline, Cell 104 (2001) 477–485.
[8] T.M. Shinnick, C.H. King, F.D. Quinn, Molecular biology, virulence, and pathogenicity of mycobacteria, Am. J. Med. Sci. 309 (1995) 92–98.
[9] F. Baneyx, Recombinant protein expression in *Escherichia coli*, Curr. Opin. Biotechnol. 10 (1999) 411–421.
[10] F.W. Studier, B.A. Moffat, Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes, J. Mol. Biol. 189 (1986) 113–130.
[11] A. Quémard, J.C. Sacchettini, A. Dessen, C. Vilchèze, R. Bittman, W.R. Jacobs Jr., J.S. Blanchard, Enzymatic characterization of the target for isoniazid in *Mycobacterium tuberculosis*, Biochemistry 34 (1995) 8235–8241.
[12] L.A. Basso, D.S. Santos, W. Shi, R.H. Furneaux, P.C. Tyler, V.L. Schramm, J.S. Blanchard, Purine nucleoside phosphorylase from *Mycobacterium tuberculosis* analysis of inhibition by a transition-state analogue and dissection by parts, Biochemistry 28 (2001) 8196–8203.
[13] J.S. Oliveira, C.A. Pinto, L.A. Basso, D.S. Santos, Cloning and overexpression in soluble form of functional shikimate kinase and 5-enolpyruvylshikimate 3-phosphate synthase enzymes from *Mycobacterium tuberculosis*, Protein Express. Purif. 3 (2001) 430–435.
[14] W. Shi, L.A. Basso, D.S. Santos, P.C. Tyler, R.H. Furneaux, J.S. Blanchard, S.C. Almo, V.L. Schramm, Structures of purine nucleoside phosphorylase from *Mycobacterium tuberculosis* in complex with immucillin-H and its pieces, Biochemistry 28 (2001) 8204–8215.
[15] L. Kremer, K.M. Nampoothiri, S. Lesjean, L.G. Dover, S. Graham, J. Betts, P.J. Brennan, D.E. Minnikin, C. Locht, G.S. Besra, Biochemical characterization of acyl carrier protein (AcpM) and malonyl-CoA:AcpM transacylase (mtFabD), two major components of *Mycobacterium tuberculosis* fatty acid synthase, J. Biol. Chem. 276 (2001) 27967–27974.
[16] B. Miroux, J.E. Walker, Over-production of proteins in *Escherichia coli*: mutant host that allow synthesis of some membrane proteins and globular proteins at high levels, J. Mol. Biol. 260 (1996) 289–298.
[17] R. Vicentelli, C. Bignon, A. Gruez, S. Canaan, G. Sulzenbacher, M. Tegoni, V. Campanacci, C. Cambillau, Medium-scale structural genomics: strategies for protein expression and crystallization, Acc. Chem. Res. 36 (2003) 165–172.
[18] E.K. Schroeder, O.N. de Souza, D.S. Santos, J.S. Blanchard, L.A. Basso, Drugs that inhibit mycolic acid biosynthesis in *Mycobacterium tuberculosis*, Curr. Pharm. Biotechnol. 3 (2002) 197–225.
[19] A. Banerjee, E. Dubnau, A. Quémard, V. Balasubramanian, K.S. Um, T. Wilson, D. Collins, G. Lisle, W.R. Jacobs Jr., *inhA*, a gene encoding a target for isoniazid and ethionamide in *Mycobacterium tuberculosis*, Science 263 (1994) 227–230.
[20] A. Banerjee, M. Sugantino, J.C. Sacchettini, W.R. Jacobs Jr., The *mab*A gene from the *inh*A operon of *Mycobacterium tuberculosis* encodes a 3-ketoacyl reductase that fails to confer isoniazid resistance, Microbiology 144 (1998) 2697–2704.
[21] P.J. Brennan, Structure, function, and biogenesis of the cell wall of *Mycobacterium tuberculosis*, Tuberculosis 83 (2003) 91–97.
[22] D.A. Rozwarski, C. Vilchèze, M. Sugantino, R. Bittman, J.C. Sacchettini, Crystal structure of the *Mycobacterium tuberculosis* enoyl-ACP reductase, InhA, in complex with NAD^+ and a C16 fatty acyl substrate, J. Biol. Chem. 274 (1999) 15582–15589.
[23] S. Parikh, D.P. Moynihan, G. Xiao, P.J. Tonge, Roles of tyrosine 158 and lysine 165 in the catalytic mechanism of InhA, the enoyl-ACP reductase from *Mycobacterium tuberculosis*, Biochemistry 38 (1999) 13623–13634.
[24] B. Persson, Y. Kallberg, U. Oppermann, H. Jörnvall, Coenzyme-based functional assignments of short-chain dehydrogenases/reductases (SDRs) family, Chem. Biol. Interact. 143-144 (2003) 271–278.
[25] J. Shanklin, Overexpression and purification of the *Escherichia coli* inner membrane enzyme acyl acyl carrier protein synthase in an active form, Protein Express. Purif. 18 (2000) 355–360.
[26] S.N. Ho, H.D. Hunt, R.M. Horton, J.K. Pullen, L.R. Pease, Site directed mutagenesis by overlap extension using the polymerase chain reaction, Gene 77 (1989) 51–59.
[27] M.A. Hayat, Basic Electron Microscopy Techniques, Nostrand Press, New York, 1972.
[28] M.A. Hayat, Basic Techniques for Transmission Electron Microscopy, Academic Press, Orlando, 1986.
[29] R.C. Austin, D. Singh, P.C.Y. Liaw, H.J. Craig, Visual detection method for identifying recombinant bacterial colonies, Bio.Techniques 18 (1995) 380–384.
[30] S. Barik, Relationship between opacity of transformed *E. coli* colonies and over-expression of the recombinant transcript, BioTechniques 22 (1997) 112–118.
[31] H. Dong, L. Nilsson, C.G. Kurland, Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction, J. Bacteriol. 177 (1995) 1497–1504.
[32] T.H. Grossman, E.S. Kawasaki, S.P. Punreddy, M.S. Osburne, Spontaneous cAMP-dependent derepression of gene expression in stationary phase plays a role in recombinant expression instability, Gene 209 (1998) 95–103.
[33] K.C. Kelley, K.J. Huestis, D.A. Austen, C.T. Sanderson, M.A. Donoghue, S.K. Stickel, E.S. Kawasaki, M.S. Osburne, Regulation of sCD4-183 gene expression from phage-T7-based vectors in *Escherichia coli*, Gene 156 (1995) 33–36.
[34] W.R.A. Taylor, A 'periodic table' for protein structures, Nature 415 (2002) 657–660.
[35] S.A. Lesley, P. Kuhn, A. Godzik, A.M. Deacon, I. Mathews, A. Kreusch, G. Spraggon, H.E. Klock, D. McMullan, T. Shin, J. Vincent, A. Robb, L.S. Brinen, M.D. Miller, T.M. McPhillips, M.A. Miller, D. Scheibe, J.M. Canaves, C. Guda, L. Jaroszewski, T.L. Selby, M.A. Elsliger, J. Wooley, S.S. Taylor, K.O. Hodgson, I.A. Wilson, P.G. Schultz, R.C. Stevens, Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline, Proc. Natl. Acad. Sci. USA 99 (2002) 11664–11669.
[36] G. Hannig, S.C. Makrides, Strategies for optimizing heterologous protein expression in *E. coli*, Trends Biotechnol. 16 (1998) 54–60.
[37] S.C. Makrides, Strategies for achieving high-level expression of genes in *Escherichia coli*, Microbiol. Rev. 60 (1996) 512–538.

ANEXO III

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Protein Expression and Purification 40 (2005) 23–30
Protein Expression & Purification
www.elsevier.com/locate/yprep

DAHP synthase from *Mycobacterium tuberculosis* H37Rv: cloning, expression, and purification of functional enzyme

Caroline Rizzi[a], Jeverson Frazzon[b], Fernanda Ely[a], Patrícia G. Weber[a], Isabel O. da Fonseca[a], Michelle Gallas[a], Jaim S. Oliveira[a], Maria A. Mendes[c], Bibiana M. de Souza[c], Mário S. Palma[c], Diógenes S. Santos[d,*], Luiz A. Basso[a,*]

[a] *Departamento de Biologia Molecular e Biotecnologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS 91501-970, Brazil*
[b] *Departamento de Ciência dos Alimentos, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS 91501-970, Brazil*
[c] *Departamento de Biologia/CEIS, Universidade do Estado de São Paulo, Rio Claro, SP 13506-900, Brazil*
[d] *Centro de Pesquisa e Desenvolvimento em Biologia Molecular e Funcional, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, RS 90619-900, Brazil*

Received 13 April 2004, and in revised form 18 June 2004
Available online 8 December 2004

Abstract

Tuberculosis (TB), caused by *Mycobacterium tuberculosis*, remains the leading cause of mortality due to a bacterial pathogen. According to the 2004 Global TB Control Report of the World Health Organization, there are 300,000 new cases per year of multidrug resistant strains (MDR-TB), defined as resistant to isoniazid and rifampicin, and 79% of MDR-TB cases are now "super strains," resistant to at least three of the four main drugs used to treat TB. Thus there is a need for the development of effective new agents to treat TB. The shikimate pathway is an attractive target for the development of antimycobacterial agents because it has been shown to be essential for the viability of *M. tuberculosis*, but absent from mammals. The *M. tuberculosis aroG*-encoded 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase (mtDAHPS) catalyzes the first committed step in this pathway. Here we describe the PCR amplification, cloning, and sequencing of *aroG* structural gene from *M. tuberculosis* H37Rv. The expression of recombinant mtDAHPS protein in the soluble form was obtained in *Escherichia coli* Rosetta-gami (DE3) host cells without IPTG induction. An approximately threefold purification protocol yielded homogeneous enzyme with a specific activity value of 0.47 U mg^{-1} under the experimental conditions used. Gel filtration chromatography results demonstrate that recombinant mtDAHPS is a pentamer in solution. The availability of homogeneous mtDAHPS will allow structural and kinetics studies to be performed aiming at antitubercular agents development.

Keywords: *Mycobacterium tuberculosis*; Shikimate pathway; DAHP synthase; Protein expression

Tuberculosis (TB)[1] remains the leading cause of mortality due to a bacterial pathogen, *Mycobacterium tuberculosis*. The interruption of centuries of decline in case rates of TB occurred, in most cases, in the late 1980s and involved the USA and some European countries due to increased poverty in urban settings and the immigration

* Corresponding authors. Fax: +55 51 3166234.
E-mail addresses: diogenes@pucrs.br (D.S. Santos), labasso@dna.cbiot.ufrgs.br (L.A. Basso).

[1] *Abbreviations used:* TB, tuberculosis; MDR-TB, multidrug-resistant; PEP, phosphoenolpyruvate; E4P, D-erythrose-4-phosphate; DAHPS, 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase; DMSO, dimethyl sulfoxide; LB, Luria–Bertani; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; IPTG, isopropyl β-D-thiogalactoside; β-ME, β-mercaptoethanol; DAHPS(Phe), phenylalanine-regulated DAHPS; DAHPS(Try), tyrosine-regulated DAHPS; DAHPS(Trp), tryptophan-regulated DAHPS.

1046-5928/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.pep.2004.06.040

from TB high-burden countries [1]. Thus, no sustainable control of TB epidemics can be reached in any country without properly addressing the global epidemic. It is estimated that 8.2 million new TB cases occurred worldwide in the year 2000, with approximately 1.8 million deaths in the same year, and more than 95% of those were in developing countries [2]. Approximately, 2 billion individuals are believed to harbor latent TB based on tuberculin skin test surveys [3], which represents a considerable reservoir of bacilli. Possible factors underlying the resurgence of TB worldwide include the HIV epidemic, increase in the homeless population, and decline in health care structures and national surveillance [4]. Another contributing factor is the evolution of multi-drug resistant strains (MDR-TB), defined as resistant to isoniazid and rifampicin, which are the most effective first-line drugs [5]. According to the 2004 Global TB Control Report of the World Health Organization, there are 300,000 new cases per year of MDR-TB worldwide, and 79% of MDR-TB cases are now "super strains," resistant to at least three of the four main drugs used to treat TB [6]. The factors that most influence the emergence of drug-resistant strains include inappropriate treatment regimens, and patient noncompliance in completing the prescribed courses of therapy due to the lengthy standard "short-course" treatment or when the side effects become unbearable [7]. Hence, faster acting and effective new drugs to better combat TB, including MDR-TB, are needed.

The shikimate pathway is an attractive target for the development of herbicides and antimicrobial agents because it is essential in algae, higher plants, bacteria, and fungi, but absent from mammals [8]. In mycobacteria, the shikimate pathway leads to the biosynthesis of chorismic acid, which is a precursor for the synthesis of aromatic amino acids, naphthoquinones, menaquinones, and mycobactins [9]. The salicylate-derived mycobactin siderophores have been shown to be essential for *M. tuberculosis* growth in macrophages [10]. More recently, the shikimate pathway has been shown by disruption of *aroK* gene, which codes for the shikimate kinase enzyme, to be essential for the viability of *M. tuberculosis* [11]. The absence from the human host and essentiality of mycobacterial shikimate pathway indicate that any of its enzymes are promising targets for the development of potentially non-toxic antimycobacterial agents.

Homologues to enzymes in the shikimate pathway have been identified in the genome sequence of *M. tuberculosis* [12]. The first committed step in the shikimate pathway is catalyzed by 3-deoxy-D-*arabino*-heptulosonate 7-phosphate (DAHP) synthase (DAHPS; EC 4.1.2.15). DAHPS catalyzes the stereospecific condensation of phosphoenolpyruvate (PEP) and D-erythrose 4-phosphate (E4P), forming DAHP and inorganic phosphate [13]. Based on phylogenetic analysis, DAHPS has been divided into two classes, class I and class II [14]. *Escherichia coli* expresses three DAHPS isoenzymes that are representative of class II and require divalent metal for activity [15], which play a role in catalysis and/or structural integrity [16]. Each isoenzyme is specifically inhibited by one of the three aromatic amino acids [8]. DAHPS(Phe), a homotetramer encoded by the *aroG* gene, is feedback inhibited by phenylalanine; the *aroH*-encoded DAHPS(Trp) and *aroF*-encoded DAHPS(Tyr) are homodimers feedback inhibited by, respectively, tryptophan and tyrosine. In *M. tuberculosis* genome, however, only the *aroG* (Rv2178c) encoded DAHPS isoenzyme (mtDAHPS) has been proposed to be present by sequence homology.

To determine the mechanism of action of mtDAHPS by steady-state and pre-steady-state kinetics as well as for X-ray crystal structure determination aiming at the rational design of antimycobacterial agents, expression of *aroG* encoded mtDAHPS in functional form and in large quantity are needed. Accordingly, we here describe the PCR amplification, cloning, sequencing, expression in *E. coli* Rosetta-gami (DE3) cells, purification to homogeneity, oligomeric state determination, and assay of mtDAHP enzyme activity. Measurements of enzyme activity confirm the correct assignment to the structural gene encoding mtDAHPS in *M. tuberculosis*. The availability of mtDAHPS will allow enzyme kinetics and structural studies to be undertaken to provide a framework on which to base the design of new agents with antitubercular activity with, hopefully, low toxicity.

Materials and methods

PCR amplification and cloning of M. tuberculosis aroG gene

The design of synthetic oligonucleotide primers used for PCR amplification of *aroG* gene (5′-gga**catatg**aactgg accgtcgacatac-3′ and 5′-c**ggatcc**tcagtcccgcagcatctccgc-3′) was based on the complete genome sequence of *M. tuberculosis* H37Rv [12]. These primers were complementary to, respectively, the amino-terminal coding and carboxy-terminal noncoding strands of *aroG* gene containing 5′*Nde*I and 3′*Bam*HI restrictions sites, which are in bold. This pair of primers was used to amplify the *M. tuberculosis aroG* gene (1389 bp) from genomic DNA using standard PCR conditions and the enzyme *Pfu* DNA polymerase (Stratagene), which is a thermostable polymerase that exhibits low error rate, thus lowering the likelihood of introducing unwanted mutations. PCR amplification required the presence of 10% of dimethyl sulfoxide (DMSO) in the reaction mixture. The PCR product was purified by electrophoresis on low melting agarose, digested with *Nde*I and *Bam*HI (Boehringer–Mannheim), and cloned into pET-23a(+) (Novagen)

expression vector, which had previously been digested with the same restriction enzymes. To both confirm the identity of the cloned gene and ensure that no mutations were introduced by the PCR amplification step, the DNA sequence of the amplified *M. tuberculosis aroG* structural gene was determined by dideoxy-chain termination method [17], using the Thermo Sequenase radiolabeled terminator cycle sequencing kit (Amersham Biosciences).

Expression of mtDAHPS

The recombinant plasmid pET-23a(+)::*aroG* was transformed into electrocompetent *E. coli* Rosetta-gami (DE3) cells (Novagen), and selected on LB agar plates containing 50 μg mL^{-1} carbenicillin, 15 μg mL^{-1} kanamycin, 34 μg mL^{-1} chloramphenicol, and 12.5 μg mL^{-1} tetracycline. Single colonies were used to inoculate 500 mL LB medium, containing the same antibiotics and concentrations of LB solid medium, and grown at 37 °C and 180 rpm for 24 h, without addition of isopropyl β-D-thiogalactopyranoside (IPTG). Cells were harvested by centrifugation at 48,000*g* for 20 min at 4 °C, and stored at −20 °C. For protein expression analysis, 10 mg of stored cells was resuspended in 500 μL Buffer A (50 mM Tris–HCl, pH 7.8), disrupted by sonification, and cell debris was removed by centrifugation. Both soluble and insoluble fractions were analyzed by SDS–PAGE 12% [18]. Control experiments were performed under the same experimental conditions except that *E. coli* host cells were transformed with the expression vector lacking the target gene.

Purification of recombinant mtDAHPS

Approximately, 36 g of cells was collected by centrifugation (48,000*g* for 20 min) from 6 L of LB medium. All subsequent steps were performed on ice or at 4 °C. Frozen cells (36 g) were thawed and resuspended in Buffer A (4 mL of buffer per gram of cell paste) containing 1 mM β-mercaptoethanol (β-ME) (Sigma) and 0.2 mg mL^{-1} of lysozyme, and the mixture was stirred for 30 min. Cells were disrupted by sonication, and cell debris was removed by centrifugation (48,000*g* for 30 min). The supernatant was incubated with 1% w/v of streptomycin sulfate for 15 min, and centrifuged (48,000*g* for 30 min). Solid ammonium sulfate was added to the supernatant fraction to a concentration of 25% saturation, incubated for 30 min, and centrifuged as above. The resultant pellet was resuspended in 70 mL Buffer A containing 1 mM β-ME and dialyzed against three changes of 2 L of the same buffer using a dialysis tubing with molecular weight cut-off of 12,000–4000 Da. The sample was clarified by centrifugation and loaded on a Q-Sepharose Fast Flow (2.6 cm × 8.2 cm) anion exchange column (Amersham Biosciences) previously equilibrated with Buffer A and fractionated using a 600 mL 0.0–0.6 M NaCl linear gradient. The fractions containing mtDAHPS (0.35–0.38 M NaCl) were pooled, concentrated to 8.0 mL using an Amicon ultrafiltration cell (MW 30,000 Da), and loaded on a Sephacryl S-200 HR (2.6 cm × 60 cm) gel filtration column (Amersham Biosciences) at 0.5 mL min^{-1}. The protein was eluted with Buffer B (Buffer A containing 200 mM NaCl) at the same flowrate. The active fractions were loaded on a Mono Q HR 16/10 anion exchange column (Amersham Biosciences) equilibrated with Buffer A and eluted with 400 mL linear 0.0–0.6 M NaCl gradient. The active fractions, which exhibited a single band on SDS–PAGE, were pooled, quickly frozen in liquid nitrogen, and stored at −80 °C.

Determination of protein concentration

Protein concentrations were determined using the Bio-Rad Laboratories protein assay kit (Bradford method) [19] and bovine serum albumin as standard.

Determination of mtDAHPS molecular mass

The molecular mass of native mtDAHPS homogeneous protein was determined by gel filtration chromatography using a Sephacryl S-200 (HR 10/30) (Amersham Biosciences) equilibrated with Buffer B at a flowrate of 0.4 mL min^{-1}. Protein molecular mass standards were from Gel Filtration LMW and HMW Calibration Kit from Amersham Biosciences. Protein elution was monitored at 280 nm.

mtDAHPS enzyme assay

Enzyme activity of recombinant mtDAPHS protein was assayed in the forward direction by a continuous spectrophotometric method described by Shoner and Hermann [20], monitoring the decrease in phosphoenolpyruvate (PEP) concentration at 232 nm ($\varepsilon = 2.8 \times 10^3$ M^{-1} cm^{-1}) on a Multi-Spec 1501 photodiode array spectrophotometer (Shimadzu). All reactions were carried out at 25 °C and initiated with addition of enzyme to a reaction mixture containing: 50 mM Tris–HCl, pH 7.0, 400 μM E4P (Sigma), 1 mM β-ME, and 200 μM PEP (Acrós Organics) in a total volume of 500 μL. One unit of enzyme activity (U) is defined as the amount of enzyme catalyzing the conversion of 1 μmol PEP/min at 25 °C.

N-terminal amino acid sequencing

The N-terminal amino acid residues of homogeneous recombinant mtDAHPS were identified by automated Edman degradation sequencing using a PPSQ 21A gas-phase sequencer (Shimadzu).

Mass spectrometry analysis

The homogeneity of recombinant protein preparation was assessed by mass spectrometry (MS), employing some adaptations made to the system described by Chassaigne and Lobinski [21]. Samples were analyzed on a triple quadrupole mass spectrometer, model QUATTRO II, equipped with standard electrospray (ESI) probe (Micromass, Altrinchan), adjusted to ca. 250 μL min^{-1}. The source temperature (80 °C) and needle voltage (3.6 kV) were maintained constant throughout the experimental data collection, applying a drying gas flow (nitrogen) of 200 Lh^{-1} and a nebulizer gas flow of 20 Lh^{-1}. The mass spectrometer was calibrated with intact horse heart myoglobin and its typical cone-voltage induced fragments. The subunit molecular mass of recombinant protein mtDAHPS was determined by ESI-MS, adjusting the mass spectrometer to give a peak with at half-height of 1 mass unit, and the cone sample to skimmer lens voltage controlling the ion transfer to mass analyzer was set to 38 V. About 50 pmol (10 μL) of each sample was injected into the electrospray transport solvent. The ESI spectrum was obtained in the multi-channel acquisition mode, scanning from 500 to 1800 *m/z* at a scan time of 7 s. The mass spectrometer is equipped with MassLynx and Transform software for data acquisition and spectra handling.

Results and discussion

The PCR amplification of *aroG* structural gene from *M. tuberculosis* H37Rv genomic DNA required the presence of 10% DMSO in the reaction mixture (data not shown). DMSO is a cosolvent that improves GC-rich DNA denaturation and helps to overcome the difficulties of polymerase extension through secondary structures, altering the structural conformation of DNA templates [22]. This result is consistent with the 65.6% G + C content of *M. tuberculosis* H37Rv genome [12]. PCR fragment was inserted into pET23a(+) expression vector [23] between *Nde*I and *Bam*HI restriction sites. DNA sequencing of the entire *aroG* structural gene by the dideoxy chain termination method both confirmed the identify of the cloned PCR product and showed that no mutations were introduced by the DNA amplification step.

Recombinant plasmids were introduced into *E. coli* BL21(DE3) host cells by electroporation. Unfortunately, recombinant mtDAHPS remained in the insoluble fraction (Fig. 1). Since one of the goals of the present work was to confirm the correct assignment to the structural gene encoding mtDAHPS, efforts were made to express recombinant *M. tuberculosis* DAHPS in its soluble, active form avoiding unfolding and refolding protocols because they cannot guarantee that they will yield large

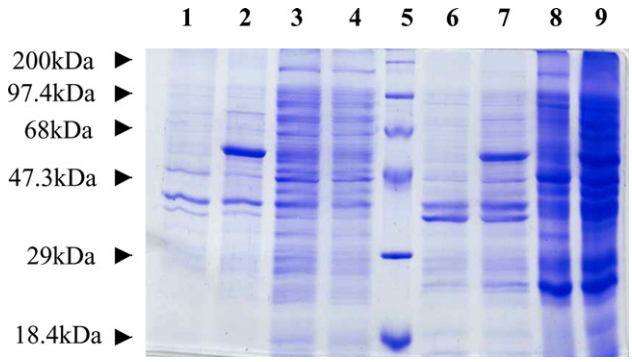


Fig. 1. SDS–PAGE (12%) of the soluble and insoluble fractions of the cell extracts of either BL21(DE3) or Rosetta-gami (DE3) host cells transformed with either pET-23a(+) (control) or pET-23a(+)::*aroG*. Expression conditions were 24 h at 37 °C without IPTG addition. Lane 1: insoluble fraction of BL21 (DE3) transformed with pET-23(+); lane 2: insoluble fraction of BL21(DE3) transformed with pET-23(+)::*aroG*; lane 3: soluble fraction of BL21(DE3) transformed with pET-23(+); lane 4: soluble fraction of BL21(DE3) transformed with pET-23(+)::*aroG*; lane 5: MW marker "High range" (Gibco-BRL); lane 6: insoluble fraction of Rosetta-gami (DE3) transformed with pET-23(+); lane 7: insoluble fraction of Rosetta-gami (DE3) transformed with pET-23(+)::*aroG*; lane 8: soluble fraction of Rosetta-gami (DE3) transformed with pET-23(+); and lane 9: soluble fraction of Rosetta-gami (DE3) transformed with pET-23(+)::*aroG*. Molecular mass of mtDAHPS is approximately 50.6 kDa.

amounts of biologically active product [24]. In addition, a number of protocols were tested to obtain mtDAHPS in the soluble fraction to no avail, including buffer additives (urea, deoxycholic acid, Triton X-100, and high NaCl concentrations) and reduced cultivation temperature (20, 25, and 30 °C). In practice, it is usually worthwhile to test several different vector/host combinations to obtain the best possible yield of protein in its desired form. Accordingly, a number of commercially available strains of *E. coli* host cells were tested in an attempt to produce mtDAHPS in the soluble fraction. Analysis of the relationship between codon preference and expression level led to the classification of *E. coli* genes into three main classes [25]. Class II genes, which correspond to genes highly and continuously expressed during exponential growth that is likely to resemble the tRNA population available for recombinant protein expression, have a number of avoided codons with frequencies of less than 6%. Insufficient tRNA pools can lead to premature translational termination, translation frameshifting or amino acid misincorporation that might result in expression of nonproperly folded recombinant protein [26]. Rare codons near the N-terminus of a coding sequence can have a severe effect on heterologous expression in *E. coli* [27]. Four rare codons for heterologous gene expression in *E. coli* are present near the N-terminus of *M. tuberculosis aroG* structural gene (1 × AUA for isoleucine, 3 × CCC for proline). To test whether these rare codons may have any effect on *aroG* expression, *E. coli* Rosetta (DE3) strain harboring tRNA genes for AGG, AGA, AUA, CUA, CCC, and

GGA rare codons on a chloramphenicol-resistant plasmid [28] was transformed with pET-23a(+)::*aroG* recombinant plasmid. Disappointingly, recombinant mtDAHPS remained in the insoluble fraction (data not shown), thereby discarding any effect of the mycobacterial *aroG* rare codons on recombinant protein expression. Although *E. coli* DAHPS has no disulfide bridges despite possessing three cysteine residues, there is no experimental evidence for the absence of disulfide bridges in mtDAHPS, which possesses five cysteine residues (Cys 87, Cys 231, Cys 365, Cys 420, and Cys 440), and a less reducing cytoplasmatic environment could improve mtDAHPS solubility. The Origami *E. coli* host strains (Novagen) have mutations in both the thioredoxin reductase (*trxB*) and glutathione reductase (*gor*) genes, which greatly enhances disulfide bond formation in the cytoplasm [29,30]. Unfortunately, none of the protocols tested yielded soluble mtDAHPS. The Rosetta-gami (DE3) *E. coli* host strain (Novagen) combines the features of Rosetta and Origami strains. SDS–PAGE analysis showed that expression of recombinant mtDAHPS protein in its soluble form with the expected molecular mass (~51 kDa) could be achieved using the Rosetta-gami (DE3) cells grown at 37 °C for 24 h with no IPTG induction (Fig. 1). The underlying reason for this result is unclear; however, it underscores the need for optimization of vector/host combinations to achieve soluble recombinant protein expression before attempting any unfolding/refolding protocols.

It should be pointed out that a screening of experimental conditions was carried out to obtain high yield of recombinant protein expression, including temperature of growth, culture aeration, medium type, hours of growth after IPTG induction, and hours of growth in the absence of IPTG. The best results were obtained from Rosetta-gami (DE3) *E. coli* cells grown for 24 h at 37 °C in LB medium without IPTG induction as described above. In the pET vector system (Novagen), target genes are positioned downstream of the bacteriophage T7 late promoter. Typically, production hosts contain a prophage (λDE3) encoding the highly processive T7 RNA polymerase under control of the IPTG-inducible *lac*UV5 promoter that would ensure tight control of recombinant gene basal expression [31,32]. In agreement with the results presented here, leaky expression has been shown to occur in the pET system [33–37]. 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 general cellular response to nutrient limitation [38].

Enzyme activity measurements demonstrated that there was a 92-fold increase in specific activity for mtDAHPS when Rosetta-gami (DE3) *E. coli* harboring either pET-23a(+)::*aroG* or pET-23a(+) crude extracts were compared (Table 1), indicating that mtDAHPS was expressed in its soluble and functional form. The purification protocol of recombinant mtDAHPS, protein determination, enzyme assay, and SDS–PAGE analysis were as described in Materials and methods. Recombinant mtDAHPS enzyme was purified approximately 3-fold (Table 2) to electrophoretic homogeneity (Fig. 2). The purification protocol yielded approximately 5 mg of homogeneous protein. A significant loss in protein yield occurred in the 25% ammonium sulfate precipitation step (Table 2) because some mtDAHPS remained in the supernatant. Protein precipitations by higher ammonium sulfate concentrations were also carried out, yielding larger amounts of recombinant mtDAHPS in the pellet (data not shown). However, a number of contaminants co-precipitated with mtDAHPS. In particular, a contaminant that co-eluted in subsequent chromatographic steps when larger than 25% ammonium sulfate concentrations were used. Accordingly, the 25% ammonium sulfate precipitation step was deemed more appropriate for the purification protocol because a significant amount of contaminants remained in the supernatant, while mtDAHPS with a lower protein-contaminating background remained in the pellet thus making the

Table 1
Measurements of recombinant DAHPS enzyme activity

Cell extract[a]	Specific activity[b] (SA, U mg^{-1})	AS cloned/SA control
Control	0.0018	1
DAHPS	0.1651	92

[a] Crude cell extract in 50 mM Tris–HCl, pH 7.0.
[b] U mL^{-1}/mg mL^{-1}.

Table 2
Purification of *M. tuberculosis* 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase expressed in *E. coli* Rosetta-gami(DE3) transformed with pET-23a(+)::*aroG*[a]

Purification step	Total protein (mg)	Total enzyme activity (U)	Specific activity[b] (U mg^{-1})	Purification fold	Yield (%)
Crude extract	2442.38	403.23	0.17	1.00	100
Ammonium sulfate	136.78	44.11	0.32	1.95	11
Q-Sepharose Fast Flow	21.42	7.00	0.33	1.98	2
Sephacryl S-200 HR	17.83	4.48	0.25	1.52	1
Mono Q HR 16/10	4.77	2.24	0.47	2.84	0.6

[a] Typical purification protocol starting from 36 g wet weight cells obtained from 6 L of culture.
[b] U mL^{-1}/mg mL^{-1}.

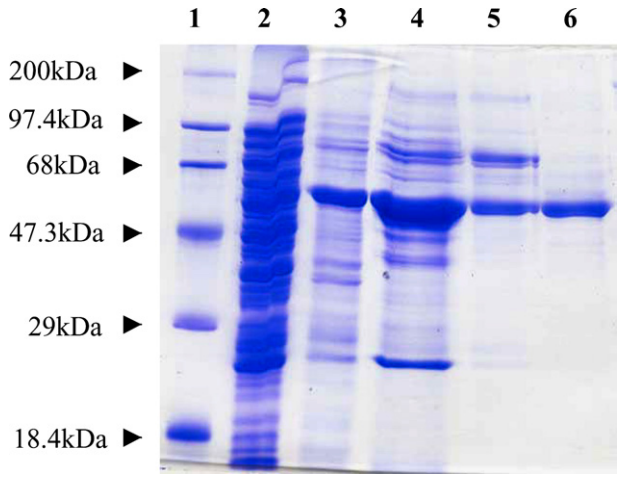


Fig. 2. SDS–PAGE analysis of pooled fractions from the purification steps of mtDAHPS. Lane 1, MW marker "High range" (Gibco-BRL); lane 2, crude extract; lane 3, ammonium sulfate precipitation; lane 4, Q-Sepharose Fast Flow ion exchange; lane 5, S-200 gel filtration; and lane 6, Mono Q ion exchange.

subsequent purification steps less demanding. The samples of each purification step were assayed for DAHPS enzyme activity and compared to control experiments to demonstrate that the observed values are actual velocities of mtDAHPS activity. Enzyme activity of homogeneous mtDAHPS was linearly dependent on sample volume added to the reaction mixture (Fig. 3), thereby showing that the initial velocity is proportional to total enzyme concentration and that true initial velocities are being measured. *E. coli* DAHPS(Phe) has been shown to be sensitive to oxidation, leading to inactivation [39]. Moreover, higher mtDAHPS enzyme activity values could be observed in the presence of β-ME (data not shown). Accordingly, β-ME was used in all steps of the mtDAHPS purification protocol. The mtDAHPS

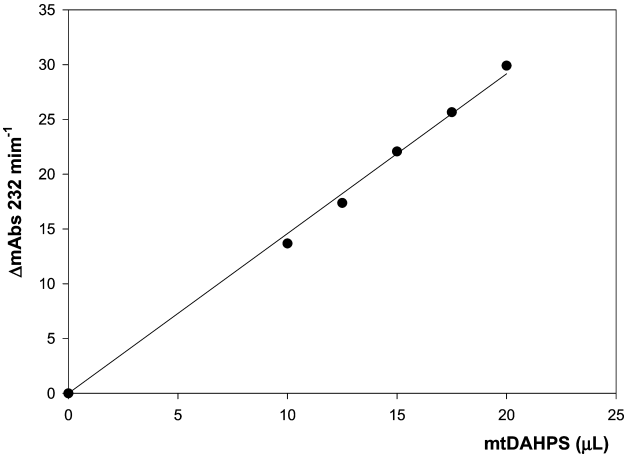


Fig. 3. Linear dependence of mtDAHPS activity on homogeneous protein volume. The rates of enzyme activity were performed in the forward direction by continuously monitoring the decrease of phosphoenolpyruvic acid at 232 nm.

specific activity has been found to be stable for at least two months when stored at −80 °C.

The subunit molecular mass of active mtDAHPS was determined to be 50510.38 Da by electrospray ionization mass spectrometry (ESI-MS), consistent with the post-translational removal of the N-terminal methionine residue from the full length gene product (predicted mass: 50641.51 Da). The ESI-MS result also revealed no peak at the expected mass for the three isoforms of *E. coli* DAHPS enzymes (38804.03, 38735.18, and 38009.53 Da), thus providing evidence for both the identity and purity of the recombinant protein. The first 11 N-terminal amino acid residues of the recombinant protein were identified as NWTVDIPIDQL by the Edman degradation chemistry protocol. This result unambiguously identifies the homogeneous recombinant protein as mtDAHPS and confirms removal of the N-terminal methionine residue from it. A common type of co-/post-translational modification of proteins synthesized in prokaryotic cells is modification at their N-termini. Methionine aminopeptidase catalyzed cleavage of initiator methionine is usually directed by the penultimate amino acid residues with the smallest side chain radii of gyration (glycine, alanine, serine, threonine, proline, valine, and cysteine) [40]. The N-terminal methionine was removed from the *E. coli* expressed recombinant mtDAHPS enzyme, consistent with the finding that some middle-sized penultimate amino acid residues (Asn, Asp, Leu, and Ile) undergo N-terminal processing [41].

A value of 253 ± 25 kDa was determined for the molecular mass of native mtDAHPS homogeneous protein by analytical gel filtration chromatography (data not shown), suggesting that mtDAHPS is a pentamer in solution. Whereas *E. coli* DAHPS(Phe) is a tetramer [42] and *E. coli* DAHPS(Trp) is a dimer [43]. Interestingly, more recently, recombinant DAHPS from *Pyrococcus furiosus* has been shown to be a dimer in solution and not to be inhibited by phenylalanine, tyrosine, or tryptophan [44].

DAHPS in most, not all, microorganisms is the target for pathway regulation by negative feedback inhibition, which controls carbon flow into the shikimate pathway. The most intensively investigated microorganism DAHPS has been the *E. coli* enzyme, which possesses three isoenzymes, each specifically regulated by one of three aromatic amino acid end products, either Phe, Tyr, or Trp [45]. The three isoforms have a common requirement for a metal cofactor, which can be similarly satisfied by a range of divalent metal ions [46]. DAHPS enzymes from a number of microorganisms have been studied, such as *Corynebacterium glutamicum* [47], *Thermotoga maritima* [48], *Bacillus subtilis* [49], *Saccharomyces cerevisiae* [50], and *P. furiosus* [44]. However, to the best of our knowledge, this is the first report on cloning, expression, and purification of functional DAHPS from *M. tuberculosis*.

Homogeneous mtDAHPS protein will provide protein in quantities necessary for studies on the enzyme mechanism of action by steady-state and pre-steady-state kinetics, its metal requirement, if any, and feedback inhibition by Phe, Tyr, and Trp. The expression and purification of mtDAHPS reported here will also provide protein for crystallization trials aiming at three-dimensional structure determination by X-ray diffraction. The three-dimensional structures of four forms of the Phe-regulated isoenzyme of *E. coli* DAHPS have been solved by X-ray crystallography [42,51–53], which should facilitate screening of experimental conditions to obtain crystals of mtDAHPS in complex with its substrates, possible metal cofactor and feedback inhibitors, if any. Expression of functional proteins in soluble form has been identified as an important bottleneck in efforts to determine biological activity and crystal structure of *M. tuberculosis* proteins [54]. We hope that the results reported here will contribute to efforts towards the structure determination of potential targets in *M. tuberculosis*. The enzymological and structural studies on mtDAHPS should help in the design of enzyme inhibitors to be tested as antimycobacterial agents.

Acknowledgments

Financial support for this work was provided by Millennium Initiative Program MCT-CNPq, Ministry of Health-Department of Science and Technology-UNESCO (Brazil) to D.S.S. and L.A.B. D.S.S. and L.A.B. also acknowledge grants awarded by PADCT, CNPq, and FINEP. L.A.B. (CNPq, 520182/99-5), D.S.S. (CNPq, 304051/1975-06), and M.S.P. (CNPq, 300337/2003-50) are researchers awardees from the National Council for Scientific and Technological Development of Brazil. M.A.M. is a post-doctoral fellow from FAPESP (01/05060-4).

References

[1] M.C. Raviglione, The TB epidemic from 1992 to 2002, Tuberculosis 83 (2003) 4–14.
[2] E.L Corbett., C.J. Watt, N. Walker, D. Maher, B.G. Williams, M.C. Raviglione, C. Dye, The growing burden of tuberculosis: global trends and interactions with the HIV epidemic, Arch. Intern. Med. 1639 (2003) 1009–1021.
[3] C. Dye, S. Scheele, P. Dolin, V. Pathania, M.C. Raviglione, Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country, JAMA 282 (1999) 677–686.
[4] B.R. Bloom, C.J.L. Murray, Tuberculosis: commentary on a reemergent killer, Science 257 (1998) 1055–1064.
[5] L.A. Basso, J.S. Blanchard, Resistance to antitubercular drugs, Adv. Exp. Med. Biol. 456 (1998) 115–144.
[6] World Health Organization, Anti-tuberculosis drug resistance in the world: Third Global Report, 2004.
[7] K. Duncan, Progress in TB drug development and what is still needed, Tuberculosis 83 (2003) 201–207.
[8] R. Bentley, The shikimate pathway—a metabolic tree with many branches, Crit. Rev. Biochem. Mol. Biol. 25 (1990) 307–384.
[9] C. Ratledge, Nutrition, growth and metabolism, in: C. Ratledge, J.L. Stanford (Eds.), The Biology of the Mycobacteria, vol. 1, Academic Press, London, 1982, pp. 185–271.
[10] J.J. de Voos, K. Rutter, B.G. Schroder, H. Su, Y. Zhu, C.E. Barry III, The salicylate-derived mycobactin siderophores of *Mycobacterium tuberculosis* are essential for growth in macrophages, Proc. Natl. Acad. Sci. USA 97 (2000) 1252–1257.
[11] T. Parish, N.G. Stoker, The common aromatic amino acid biosynthesis pathway is essential in *Mycobacterium tuberculosis*, Microbiology 148 (2002) 3069–3077.
[12] S.T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S.V. Gordon, K. Eiglmeier, S. Gas, C.E. Barry III, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, B.G. Barrell, Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence, Nature 393 (1998) 537–544.
[13] P.R. Srinivasan, D.B. Sprinson, 2-Keto-3-deoxy-D-*arabo*-heptonic acid 7-phosphate synthase, J. Biol. Chem. 234 (1959) 716–722.
[14] M.R. Birck, R.W. Woodard, *Aquifex aeolicus* 3-deoxy-D-*manno*-2-octulosonic acid 8-phosphate synthase: a new class of KDO 8-P synthase?, J. Mol. Evol. 52 (2001) 205–214.
[15] C.M. Stephens, R. Bauerle, Analysis of the metal requirement of 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase from *Escherichia coli*, J. Biol. Chem. 266 (1991) 20810–20817.
[16] P.A. Jordan, S. Bohle, C. Ramilo, J.N.S. Evans, New insights into the metal center of 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase, Biochemistry 40 (2001) 8387–8396.
[17] F. Sanger, S. Nicklen, R. Coulson, DNA sequencing with chain-terminating inhibitors, Pro. Natl. Acad. Sci. USA 74 (1977) 5436–5467.
[18] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, Nature 227 (1970) 680–685.
[19] M.M. Bradford, R.A. McRorie, W.L. Williams, A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding, Anal. Biochem. 72 (1976) 248–254.
[20] R. Shoner, K.M. Hermann, 3-Deoxy-D-*arabino*-heptulosonate 7-phosphate synthase, J. Biol. Chem. 252 (1976) 5440–5447.
[21] H. Chassaigne, R. Lobinski, Characterization of horse kidney metallothionein isoforms by electrospray MS and reversed-phase HPLC-electrospray MS, Analyst 123 (1998) 2125–2130.
[22] D. Pomp, J.F. Medrano, Organic solvents as facilitators of polymerase chain reaction, Biotechniques 10 (1991) 58–59.
[23] F.W. Studier, A.H. Rosenberg, J.J. Dunn, J.W. Dubendorff, Use of T7 RNA polymerase to direct expression of cloned genes, Methods Enzymol. 185 (1990) 60–89.
[24] C.H. Schein, Production of soluble recombinant proteins in bacteria, Bio/technology 7 (1989) 1141–1149.
[25] A. Hénaut, A. Danchin, Analysis and predictions from *Escherichia coli* sequences, or *E. coli* in silico, in: F.C. Neidhardt (Ed.), *Eschechichia coli* and *Salmonella*: Cellular and Molecular Biology, ASM Press, Washington, DC, 1996, pp. 2047–2066.
[26] C. Kurland, J. Gallant, Error of heterologous protein expression, Curr. Opin. Biotechnol. 7 (1996) 489–493.
[27] Y. Nakamura, T. Gojobori, T. Ikemura, Codon usage tabulated from international DNA sequence databases: status for the year 2000, Nucleic Acids Res. 28 (2000) 292.
[28] R. Novy, D. Drott, K. Yaeger, R. Mierendorf, Overcoming the codons bias of *Escherichia coli* for enhanced protein expression, Innovations 12 (2001) 1–3.
[29] W.A. Prinz, F. Aslund, A. Holmgren, J. Beckwith, The role of the thioredoxin and glutaredoxin pathways in redox-facilitate attempts at crystallizing and elucidating the protein disulfide bonds in the *Escherichia coli* cytoplasm, J. Biol. Chem. 272 (1997) 15661–15667.

[30] F. Baneyx, Recombinant protein expression in *Escherichia coli*, Curr. Opin. Biotechnol. 10 (1999) 411–421.
[31] F.W. Studier, B.A. Moffatt, Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes, J. Mol. Biol. 189 (1986) 113–130.
[32] J.W. Dubendorff, F.W. Studier, Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor, J. Mol. Biol. 219 (1991) 45–59.
[33] L.A. Basso, D.S. Santos, W. Shi, R.H. Furneaux, P.C. Tyler, V.L. Schramm, J.S. Blanchard, Purine nucleoside phosphorylase from *Mycobacterium tuberculosis*. Analysis of inhibition by a transition-state analogue and dissection by parts, Biochemistry 40 (2001) 8196–8203.
[34] J.S. Oliveira, C.A. Pinto, L.A. Basso, D.S. Santos, Cloning and overexpression in soluble form of functional shikimate kinase and 5-enoylpyruvylshikimate 3-phosphate synthase enzymes from *Mycobacterium tuberculosis*, Protein Expr. Purif. 3 (2001) 430–435.
[35] M.L.B. Magalhães, C.P. Pereira, L.A. Basso, D.S. Santos, Cloning and expression of functional shikimate dehydrogenase (EC 1.1.1.25) from *Mycobacterium tuberculosis* H37Rv, Protein Expr. Purif. 26 (2002) 59–64.
[36] R.G. Silva, L.P.S. Carvalho, J.S. Oliveira, C.A. Pinto, M.A. Mendes, M.S. Palma, L.A. Basso, D.S. Santos, Cloning, overexpression and purification of functional human purine nucleoside phosphorylase, Protein Expr. Purif. 27 (2003) 158–164.
[37] K.C. Kelley, K.J. Huestis, D.A. Austen, C.T. Sanderson, M.A. Donoghue, S.K. Stikel, E.S. Kawasaki, M.S. Osburne, Regulation of *sCD4-183* gene expression from phage T7 based vectors in *Escherichia coli*, Gene 156 (1995) 33–36.
[38] T.H. Grossman, E.S. Kawaski, S.R. Punreddy, M.S. Osburne, Spontaneous cAMP-dependent derepression of gene expression in stationary phase plays a role in recombinant expression instability, Gene 209 (1998) 95–103.
[39] R. Bauerle, O.K. Park, Metal-catalyzed oxidation of phenylalanine-sentitive 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase from *Escherichia coli*: inactivation and destabilization by oxidation of active-site cysteines, J. Bacteriol. 181 (1999) 1636–1642.
[40] W.T. Lowther, B.W. Matthews, Structure and function of the methionine aminopeptidases, Biochim. Biophys. Acta 1477 (2000) 157–167.
[41] P.-H. Hirel, J.-M. Schmitter, P. Dessen, G. Fayat, S. Blanquet, Extent of N-terminal methionine excision from *Escherichia coli* proteins is governed by the side-chain length of the penultimate amino acid, Proc. Natl. Acad. Sci. USA 86 (1989) 8247–8251.
[42] I.A. Shumilin, R.H. Kretsinger, R. Bauerle, Crystal structure of phenylalanine-regulated 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase from *Escherichia coli*, Struct. Folding Des. 7 (1999) 865–875.
[43] J.M. Ray, R. Bauerle, Purification and properties of tryptophan-sentitive 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase from *Escherichia coli*, J. Bacteriol. 173 (1991) 1894–1901.
[44] L.R. Schofield, M.L. Patchett, E.J. Parker, Expression, purification, and characterization of 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase from *Pyrococcus furiosus*, Protein Expr. Purif. 34 (2004) 17–27.
[45] K.M. Herrmann, The shikimate pathway as an entry to aromatic secondary metabolism, Plant Physiol. 107 (1995) 7–12.
[46] C.M. Stephens, R. Bauerle, Analysis of the metal requirement of 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase from *Escherichia coli*, J. Biol. Chem. 266 (1991) 20810–20817.
[47] H.F. Liao, L.L. Lin, H.R. Chin, W.H. Hsu, Serine 187 is a crucial residue for allosteric regulation of *Corynebacterium glutamicum* 3-deoxy-D-*arabino*-heptulosonate-7-phosphate, FEMS Microbiol. Lett. 184 (2001) 59–64.
[48] J. Wu, D.L. Howe, R.W. Woodard, *Thermotoga maritima* 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase: the ancestral eubacterial DAHP synthase?, J. Biol. Chem. 278 (2003) 27525–27531.
[49] R.A. Jensen, E.W. Nester, Regulatory enzymes of aromatic amino acid biosynthesis in *Bacillus subtilis*. I. Purification and properties of 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthetase, J. Biol. Chem. 241 (1966) 3365–3372.
[50] G. Paravicini, T. Schmidheini, G. Braus, Purification and properties of 3-deoxy-D-*arabino*-heptulosonate 7-phosphate synthase (phenylalanine-inhibitable) of *Saccharomyces cerevisiae*, Eur. J. Biochem. 186 (1989) 361–366.
[51] T. Wagner, I.A. Shumilin, R. Bauerle, R.H. Kretsinger, Structure of 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase from *Escherichia coli*: comparison of the Mn^{2+} 2-phosphoglycolate and the Pb^{2+} 2-phosphoenolpyruvate complexes and implications for catalysis, J. Mol. Biol. 301 (2000) 389–399.
[52] I.A. Shumilin, C. Zhao, R. Bauerle, R.H. Kretsinger, Allosteric inhibition of 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase alters the coordination of both substrates, J. Mol. Biol. 320 (2002) 1147–1156.
[53] I.A. Shumilin, R. Bauerle, R.H. Kretsinger, The high-resolution structure of 3-deoxy-D-*arabino*-heptulosonate-7-phosphate synthase reveals a twist in plane of bound phosphoenolpyruvate, Biochemistry 42 (2003) 3766–3776.
[54] R. Vicentelli, C. Bignon, A. Gruez, S. Canaan, G. Sulzenbacher, M. Tegoni, V. Campanacci, C. Cambillau, Medium-scale structural genomics: strategies for protein expression and crystallization, Acc. Chem. Res. 36 (2003) 165–172.

ANEXO IV

INPI INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL
PROTOCOLO GERAL
12/12/2005 016050005891 16:19 DERS
0000220505472390

Protocolo | Número (21)

DEPÓSITO
Pedido de Patente ou de Certificado de Adição

PI0506047-8

depósito / /

(espaço reservado para a etiqueta com o número e data de depósito)

Ao Instituto Nacional da Propriedade Industrial:

O requerente solicita a concessão de uma patente na natureza e nas condições abaixo indicadas:

1. Depositante (71)**:**

1.1 Nome: **UNIÃO BRASILEIRA DE EDUCAÇÃO E ASSISTÊNCIA**

1.2

1.2 Qualificação: **Soc.Civ.s.fim lu** 1.3 CGC/CPF: **88630413000109**

1.4 Endereço completo: **AV IPIRANGA, 6681,B/ PARTENON - POA/ RS, CEP 90001-970**

1.5 Telefone:

FAX:

☒ continua em folha anexa

2. Natureza:

☒ 2.1 Invenção ☐ 2.1.1. Certificado de Adição ☐ 2.2 Modelo de Utilidade

Escreva , obrigatoriamente e por extenso, a Natureza desejada: **INVENÇÃO**

3. Título da Invenção, do Modelo de Utilidade ou do Certificado de Adição (54):

"MÉTODO DE OBTENÇÃO DE SEQÜÊNCIAS NUCLEOTÍDICAS QUIMÉRICAS E SEQÜÊNCIA NUCLEOTÍDICA QUIMÉRICA"

☒ continua em folha anexa

4. Pedido de Divisão do pedido nº. , de

5. Prioridade Interna - O depositante reivindica a seguinte prioridade:

Nº de depósito Data de Depósito **(66)**

6. Prioridade - o depositante reivindica a(s) seguinte(s) prioridade(s):

País ou organização de origem	Número do depósito	Data do depósito

☐ continua em folha anexa

7. Inventor (72):

☐ Assinale aqui se o(s) mesmo(s) requer(em) a não divulgação de seu(s) nome(s) (art. 6º § 4º da LPI e item 1.1 do Ato Normativo nº 127/97)

7.1 Nome: **DIÓGENES SANTIAGO SANTOS**

7.2 Qualificação: **BRAS,DESQUITADO,**PROFESSOR UNIVERSITÁRIO, CPF 187292588-04

7.3 Endereço: **R FELIPE DE OLIVEIRA 734, B.PETRÓPOLIS-PORTO ALEGRE/RS**

ANEXO V

Available online at www.sciencedirect.com
SCIENCE DIRECT
Protein Expression and Purification 47 (2006) 614–620
Protein Expression & Purification
www.elsevier.com/locate/yprep

Expression, purification, and circular dichroism analysis of human CDK9

Andreia Machado Leopoldino [a], Fernanda Canduri [b], Hamilton Cabral [c], Magno Junqueira [d], Alessandra Bernadete Trovó de Marqui [a], Luciano H. Apponi [e], Isabel Osório da Fonseca [f], Gilberto Barbosa Domont [d], Diógenes S. Santos [f], Sandro Valentini [e], Gustavo Orlando Bonilla-Rodriguez [c], Marcelo Andrés Fossey [c], Walter Filgueira de Azevedo Jr. [g], Eloiza Helena Tajara [a,*]

[a] *Faculdade de Medicina de São José do Rio Preto, SJRP, SP, Brazil*
[b] *Universidade Federal do Mato Grosso do Sul, Campo Grande, MS, Brazil*
[c] *Universidade Estadual Paulista, SJRP, SP, Brazil*
[d] *Rede Proteômica-Rio, Universidade Federal do Rio de Janeiro, RJ, Brazil*
[e] *Faculdade de Ciências Farmacêuticas, UNESP, Araraquara, SP, Brazil*
[f] *Centro de Pesquisas em Biologia Molecular e Funcional/PUCRS, Porto Alegre, RS, Brazil*
[g] *Faculdade de Biociências-PUCRS, Porto Alegre, RS, Brazil*

Received 3 November 2005, and in revised form 2 February 2006
Available online 10 March 2006

Abstract

The human cyclin-dependent kinase 9 (CDK9) protein was expressed in *E. coli* BL21 using the pET23a vector at 30 °C. Several milligrams of protein were purified from soluble fraction using ionic exchange and ATP-affinity chromatography. The structural quality of recombinant CDK9 and the estimation of its secondary structure were obtained by circular dichroism. Structural models of CDK9 presented 26% of helices in agreement with the spectra by circular dichroism analysis. This is the first report on human CDK9 expression in *Escherichia coli* and structure analysis and provides the first step for the development of CDK9 inhibitors.

Keywords: CDK9; Cancer; AIDS; Structure; Dichroism analysis; Molecular modeling; Expression

The cyclin-dependent kinase 9 (CDK9)[1] is a cdc2-related kinase protein ubiquitously expressed in most cells [1] and involved in many physiological processes, including cell differentiation and apoptosis [2,3]. In contrast to other CDKs with cell cycle regulatory functions, this serine–threonine kinase exhibits protein levels unchanged in human cells entering and progressing through the cell cycle [4]. Originally named PITALRE for its PSTAIRE-like sequence [5,6], CDK9 is involved in transcriptional events [7] and is regulated by the regulatory subunit, cyclin T1, cyclin T2 or cyclin K [8,9]. Recently, a novel isoform of CDK9 of 55 kDa [10], which originates from an alternative upstream promoter was identified.

CDK9/cyclin T belongs to the multiprotein complex P-TEFb, an elongation factor for RNA polymerase II-directed transcription, and is responsible for the phosphorylation of the C-terminal domain of the largest subunit of RNA polymerase II [11]. It has been found that CDK9/cyclin T1 is also required for the viral transactivator called Tat to stimulate the processivity of RNA pol II in HIV,

* Corresponding author. Fax: +55 17 3201 5790.
E-mail address: tajara@famerp.br (E.H. Tajara).

[1] *Abbreviations used:* CDK9, cyclin-dependent kinase 9; LB, Luria–Bertani medium; MALDI-TOF, matrix-assisted desorption ionization time-of-flight; CD, circular dichroism.

1046-5928/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.pep.2006.02.012

suggesting a possible involvement of this kinase in AIDS (for references, see [12]). Ammosova et al. [13] observed dephosphorylation of CDK9 prior to its association with HIV-1 transcription initiation complex, which might be important for regulating HIV-1 transcription.

Recent data have indicated that Cdk9 is a critical determinant of cardiac hypertrophy, in vitro and in vivo. Sano and Schneider [14] observed that trophic signals for increasing cardiac mass activated Cdk9, suggesting a role in heart disease.

The CDK inhibitor Seliciclib (CYC202, R-roscovitine) competes for the ATP binding site on the kinase. It has greatest activity against CDK2/cyclin E, CDK7/cyclin H, and CDK9/cyclin T. Seliciclib induces apoptosis in tumor cell lines, reduces tumor growth in xenografts in nude mice and is currently in phase II clinical trials for lung cancer and B cell malignancies [15].

The functional aspects of human CDK9 have been studied by many groups but no conformational data is known. Since the CDK9 protein is a potential therapeutic target in AIDS, heart disease, and cancer, data on its 3D structure becomes an essential step in the design of new and specific inhibitors which could improve the treatment of patients.

In this work, we expressed and purified recombinant human CDK9 in *Escherichia coli* (BL21) suitable for future structural analysis. The secondary structural data obtained using circular dichroism was in agreement with the conformation obtained by homology modeling.

Materials and methods

PCR amplification and cloning of human cdk9 cDNA

For RT-PCR, the RNA total was isolated from human tissue using Trizol reagent (Invitrogen). To obtain the full-length *cdk9* cDNA, RT-PCR was carried out using the oligonucleotides primers forward Cg*CATATg*gCAAAG CAGTACGACTCGGTG and reverse gC*ggATCC***TCA**g AAgACgCgCTCAAACTCC designed based on the cDNA sequence from GenBank Accession No. NM_001261. The 5′ *Nde*I and 3′ *Bam*HI restriction sites are shown in italic. The start and stop codons are shown in bold and were designed to be complementary to 24 and 22 bases of the 5′ and 3′ ends of the CDK9 cDNA, respectively. The *cdk9* cDNA (1119 bp) was amplified using standard PCR conditions with *Pfu* DNA polymerase (Stratagene). To introduce the 5′ *Nde*I and 3′ *Bam*HI sites the amplified fragments were digested with *Nde*I and *Bam*HI, purified using the Qiagen Gel Purification, and inserted into a pET23a(+) expression vector (Novagen), previously digested with the same enzymes. The correct ORF was confirmed by restriction analysis and sequencing.

Expression of recombinant human CDK9

E. coli BL21(DE3)pLys cells were transformed with the pET23a(+)::cdk9 and recombinant colonies were selected on LB agar plates containing carbenicillin (50 μg mL^{-1}) and chloramphenicol (50 μg mL^{-1}). Following transformation, several recombinant colonies were used to inoculate 5 mL LB medium containing carbenicillin (50 μg mL^{-1}) and chloramphenicol (50 μg mL^{-1}). To determine the best condition of human CDK9 protein expression, *E. coli* BL21(DE3)pLys cells harboring pET23a(+)::cdk9 were grown in the presence of IPTG (final concentration of 1, 0.5, and 0.1 mM) at four different temperatures (37, 30, 28, and 20 °C) at 180 rpm after OD_{600} measurements (Abs 0.4–0.6) [16,17]. One sample was removed at each hour after induction. The cells were harvested by centrifugation, stored at −80 °C or resuspended in hypotonic lysis buffer (10 mM Tris–HCl, pH 7.4, 25 mM NaCl, and 1 mM EDTA), and disrupted by sonication (three short bursts, about 10 s each, allowing the bacterial suspension to cool on ice between each burst). The proteins and cell debris were separated by centrifugation at 20,800*g* for 30 min at 4 °C. The soluble and insoluble fractions were analyzed by SDS–PAGE.

Purification of recombinant human CDK9

Preculture (20 mL) of one single colony *E. coli* (DE3)pLys containing the recombinant plasmid was diluted in 500 mL of Luria–Bertani medium (LB) supplemented with appropriate antibiotics [carbenicillin (50 μg mL^{-1}) and chloramphenicol (50 μg mL^{-1})]. The culture was conducted at 30 °C at 180 rpm in a shaking incubator until the cells reached mid-log growth (OD_{600} measurements of 0.4–0.6). At this point, the expression of the target protein was induced by adding IPTG (0.1 mM) and continued incubation at 30 °C for 3 h. The cells were harvested by centrifugation and resuspended in 30 mL lysis buffer (10 mM Tris–HCl, pH 7.4, 25 mM NaCl, 1 mM EDTA, and 1 mM PMSF). After sonication (three short bursts, about 30 s each, allowing the bacterial suspension to cool on ice between each burst), the lysate was clarified by centrifugation for 1 h at 20,800*g* and 4 °C. The supernatant was loaded over a DEAE Sepharose column (2.5 × 9 cm) pre-equilibrated with lysis buffer. The absorbed proteins were eluted with a linear gradient (25–500 mM NaCl, elution volume 60 mL) by the use of a peristaltic pump at 60 mL/h. After, the fractions were visualized in silver-stained SDS gels, the selected CDK9 fractions were pooled and dialyzed against a buffer containing 10 mM Hepes, pH 7.4, 25 mM NaCl, and 1 mM EDTA. The dialyzed CDK9 were loaded onto an ATP affinity column (Sigma; 1 × 2.5 cm; 2 mL) pre-equilibrated with buffer A [10 mM Hepes, pH 7.4, 25 mM NaCl, 1 mM EDTA, 10% glycerol (v/v), and 0.5 mM dithiothreitol]. After washing, bound proteins were eluted with a 50 mL linear salt gradient (25–500 mM NaCl in buffer A). Fractions containing CDK9 were pooled and concentrated (up to approximately 6 mg/mL) and dialyzed using an Amicon ultrafiltration cell (MWC 10,000 Da) against 10 mM Hepes, pH 7.4, and 1 mM EDTA. The protein content was analyzed by SDS–PAGE and visualized using the Silver Staining kit, Protein (GE HealthCare).

Protein determination

The protein concentration was carried out as proposed by Bradford [18] using bovine serum albumin as standard, and after purification by direct absorbance readings at 280 nm.

Bidimensional electrophoresis

Bidimensional electrophoresis (2DE) [19] was performed with the IPGphor IEF system and Ruby electrophoresis unit (GE HealthCare). After sonication, the protein sample (100 μg) was diluted in buffer containing 8 M urea, 2% Chaps (3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate), 0.3% DTT (dithiothreitol), 0.5% IPG buffer, pH 3–10, bromophenol blue trace and loaded into immobilized pH gradient (IPG) gel strips. Isoelectric focusing was run at 20 °C with 8000 V for until 16,500 Vh. The IPG strips were subsequently either stored at −80 °C or equilibrated for 15 min each at room temperature in SDS equilibration buffer (50 mM Tris–HCl, pH 8.8, 6 M urea, 2% w/v SDS, 30% glycerol, and few grains of bromophenol blue) with 1% w/v DTT. The strips were again incubated in SDS equilibration buffer with 2.5% w/v iodoacetamide instead of DTT. The strips were sealed in 12.5% SDS–PAGE and electrophoresis was conducted at 30 mA/gel for 5 h. The gels were fixed overnight in 50% ethanol and 10% acetic acid, washed in destain solution (50% ethanol and 5% acetic acid) for 3 min and stained in 0.05% Coomassie blue solution in 40% methanol and 10% acetic acid for 90 min. The gels were soaked in destain solution for 15, 45, 120, and 120 min, respectively, and preserved in 5% acetic acid. The gels were incubated in 10% methanol for 48 h and storage in 50% glycerol until analysis. The 2D images were scanned by using the ImageScanner (GE HealthCare) and analyzed by using IMAGEMASTER 2D PLATINUM software (GE HealthCare). The intensity of each spot was first processed by background subtraction and then normalized between gels as a proportion of the total intensity protein from the gel. 2DE was chosen to CDK9 validation because it separated the proteins by charge and molecular weight (two biochemical characteristics). Therefore, it is more efficient for proteomics analysis and mass spectrometry than SDS–PAGE.

MS and protein identification

The matrix-assisted desorption ionization time-of-flight (MALDI-TOF) mass spectrometry (MS) was used to confirm the identity of the recombinant CDK9. The protein spot of interest identified by 2DE absent in control, was cut from the gel for subsequent tryptic digestion. For MALDI-TOF MS analysis, extracted peptides were resuspended with 0.1% trifluoroacetic acid in 50% acetonitrile, mixed with 1 μl of matrix (10 mg/ml α-cyano-4-hydroxycinnamic acid) and 1 μl of this mixture was pipetted in the MALDI plate and allowed to dry. Peptide maps were obtained using a Voyager–DE PRO MALDI-TOF mass spectrometer (Applied Biosystem) in positive ion reflectron mode. Peptide fragment mass values were searched against the database by using the MS-Fit program (http://prospector.ucsf.edu) with the minimum parameters: Swiss-Prot database, match with three peptides, mass tolerance of 50 ppm, and 15% coverage percentage.

Circular dichroism

Circular dichroism (CD) spectra in the UV–visible were recorded on a Jasco J710 dichrograph operating at room temperature, range 250–190 nm, interfaced with a PC. The spectrum was analyzed through the self-consistent method published by Sreerama and Wood [20], and an interactive graphic program for calculation the secondary structures by Deléage and Geourjon [21]. Measurements in the far UV were carried out on 25 μm cuvettes and the proteins were dissolved in 10 mM Hepes, pH 7.4, and 1 mM EDTA.

Molecular modeling

Homology modeling is usually the method of choice when there is a clear relationship of homology between the sequence of a target protein and at least one known structure. Model building of the CDK9 were carried out using the program Parmodel [22], which is a web server for automated modeling and protein structural assessment. Parmodel runs a parallelized version of MODELLER [23]. MODELLER is an implementation of an automated approach to comparative modeling by satisfaction of spatial restraints. The modeling procedure begins with alignment of the sequence to be modeled (target) with the sequence of related known three-dimensional structures (templates). This alignment is usually the input to the program. The output is a three-dimensional model for the target sequence containing all mainchain and sidechain non-hydrogen atoms [23].

Two models were built, the first in the absence of ligands, and the second in the presence of ATP. The models of CDK9 were based on the atomic coordinates of CDK2 apoenzyme (PDB code: 1HCL) [24] and CDK2:ATP (PDB code: 1HCK) [25].

Several slightly different models can be calculated by varying the initial structure. A total of 1000 models for each modeling were generated by the program Parmodel [22], the final models were selected based on stereochemical quality. All optimization process was performed on a Beowulf cluster with 16 nodes (BioComp, AMD Athlon XP 2100+).

Results and discussion

CDK9 is a serine–threonine kinase involved in transcription and responsible for the activation of RNA polymerase II. Due to its role in these processes, it is a potential therapeutic target in human disease. Therefore, data on its 3D

structure are essential to the development of new and specific inhibitors.

Studying protein 3D structure represents a challenge in protein biochemistry. In vitro studies such as crystallization are reliant on the solubilization, and on appropriate quantity and quality of proteins. Here, we present data that led to the successful expression and purification of human CDK9.

Overproduction of human CDK9

CDK9 protein has been produced using pET vector system and *E. coli* BL21(DE)pLys (Fig. 1, lane 4). The expression of CDK9 at 37 °C produced insoluble proteins that were resuspended in denaturing buffer containing 8 M urea. However, no CD spectrum was detected when an aliquot of dialyzed CDK9 (refolding) was analyzed by circular dichroism at a protein concentration of 13 mg/mL. Others studies [16,17] reported similar results indicating that refolding is incomplete. The incomplete refolding of CDK9 may be due to exposition to urea for a long period (up to 7 days) before diluting out the denaturant.

An optimization of the expression in different temperatures was used to yield soluble proteins. The best condition was by culturing the selected colonies with the plasmid pET23a(+)::cdk9 in LB medium at 30 °C and 0.1 mM IPTG (Fig. 1). The expression at 37 °C resulted in high amounts of precipitated proteins; at lower temperature (28 and 20 °C), the expression resulted in low amounts of protein (Fig. 2). In respect to IPTG, we found that cultures with lower IPTG concentrations produced appropriate amounts of soluble protein. Almost 40 mg mL^{-1} of total protein was synthesized and the soluble proteins were recovered after centrifugation (Table 1).

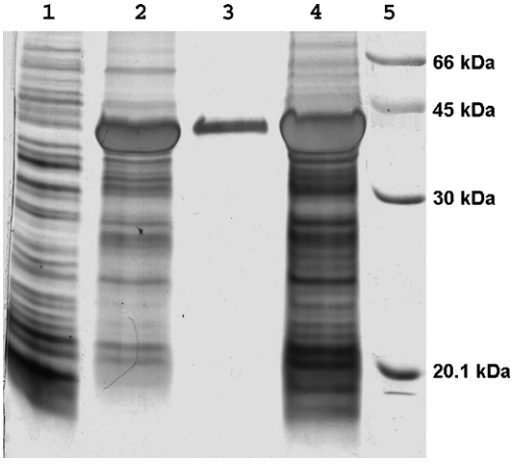


Fig. 1. Purification of the recombinant human CDK9 protein by ion exchange and ATP-affinity chromatography. Lane: 1, crude extract, no CDK9 expression (control); 2, ion exchange chromatography of total soluble protein; 3, ATP-affinity chromatography of human CDK9; 4, crude extract with recombinant human CDK9; 5, LMW protein marker (GE HealthCare). Silver staining. The resulting CDK9 is almost 99% pure.

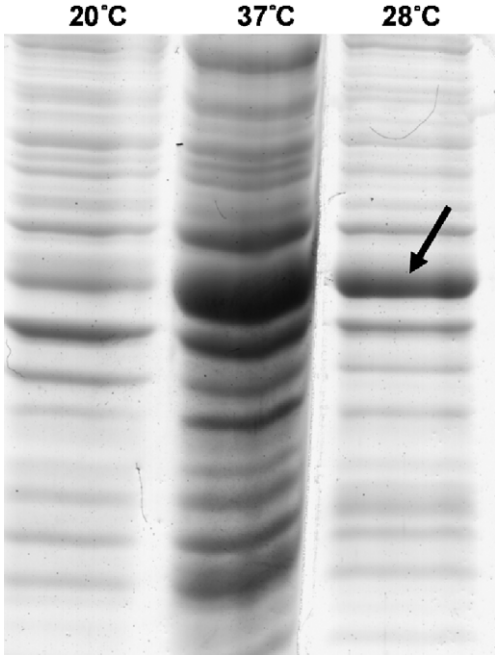


Fig. 2. SDS–PAGE of the crude extract containing recombinant human CDK9. Lanes: 1–3, crude extract from expression at 20, 37, and 28 °C, respectively. Coomassie blue staining.

By using direct absorbance readings at 280 nm, we estimated that, on average, about 10 mg of pure CDK9 was purified from a 500 mL culture of transformed bacteria.

Validation and purification of human CDK9

The proteins before and after induction by IPTG were analyzed by 2DE for validation. A spot with p*I* and molecular weight close to the theoretical values calculated for CDK9 (p*I* 8.97 and 42 kDa) was present in 2DE gel after induction but absent in control. Protein spot was excised, in-gel digested with trypsin, and identified by MALDI mass spectrometry. The processed peptide mass fingerprint data analyzed by the MS-Fit search identified the human CDK9 as the first hit with coverage of 82% (data not shown).

Recombinant protein was purified by ionic exchange (DEAE Sepharose) and ATP-affinity chromatography (Table 1; Fig. 1, lane 3). The resins and elution conditions were chosen based on our experience acquired during CDK2 purification [27]. The CDK2 purification protocol was optimized for CDK9. The fractions of purified CDK9 were pooled and concentrated for future analysis and

Table 1
The optimal purification

	Approximate yield (%)	Approximate purity[a] (%)
Cell lysate	50–60	20
DEAE column	70	50
ATP-affinity column	90	100

[a] The purities were estimated from silver-stained SDS gels and the quantification was estimated using direct absorbance readings at 280 nm.

maintained at 4 °C in the presence of proteases inhibitors to avoid degradation.

Structural analysis of CDK9

Circular dichroism is an important technique to estimate secondary structure content, to validate protein quality, and for monitoring conformational changes of proteins due to drug binding. The CDK9 deconvoluted CD spectrum showed that it contains 30–34% helix, 33–36% beta, and 21–29% of coil (Fig. 3). Such CD data permitted to obtain information about the CDK9 structures of the regions not modeled. The results suggest that human recombinant CDK9 maintains secondary structures and is viable to functional and structural studies.

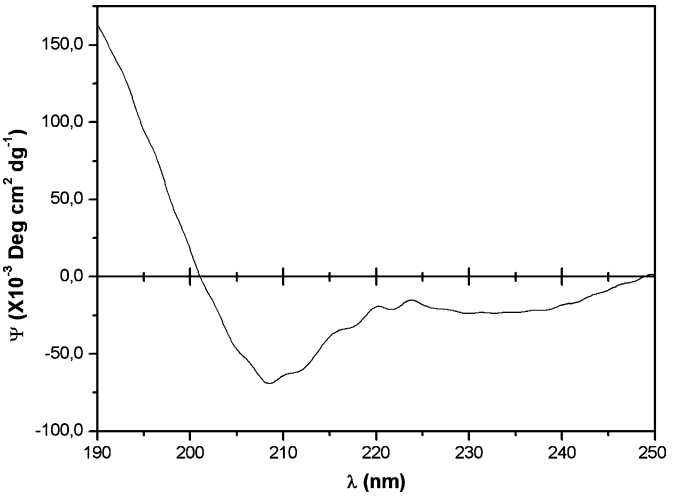


Fig. 3. Circular dichroism spectrum of human CDK9. The protein conformation consists of 30–34% helices, 33–36% beta-strands, and 21–29% coils.

Fig. 4 shows the sequence alignment of CDK9 and CDK2. The alignment presents 38% of identity, considering the region from residues 17 to 333. The overall stereochemical quality of the final CDK9 apoenzyme and CDK9:ATP models and average G-factor were calculated by the PROCHECK software [26]. Analysis of the Ramachandran plots indicates that 89.5% of the residues are in most favored regions, and 8.4% of the residues in additional allowed regions for the CDK9 apoenzyme whereas 87% of the residues are in most favored regions, and 9.9% in additional allowed regions for CDK9:ATP model. The G-factor values were 0.14 and 0.38 for CDK9 apoenzyme and CDK9:ATP models, respectively. Ideally, scores should be above −0.5.

The models for CDK9 are folded into the typical bilobal structure, with the smaller N-terminal lobe consisting predominantly of β-sheet structure and the larger C-terminal lobe consisting primarily of α-helices. The N-terminal lobe of CDK has a sheet of five antiparallel β-strands (β1–β5) and a single large helix (α1). The C-terminal lobe contains a pseudo-4-helical bundle (α2, 3, 4, 6), a small β-ribbon (β6–β8), and two additional helices (α5, 7). The ATP molecule is found in the cleft between the two lobes. The core (the β-sheet and the helical bundle) of the CDK9 structure is very similar to that of CDK2 [25,27,28], as shown in Figs. 5A–C.

The differences between the structures of CDK2 and CDK9 are in the N-terminal (residues 1–17) and C-terminal regions (residues 333–372); the CDK9 sequence is

```
                10        20        30        40        50        60
cdk2      ME---------------NFQKVEKIGEGTYGVVYKARNKLTGEVVALKKIR----TEGVP
cdk9      MAKQYDSVECPFCDEVSKYEKLAKIGQGTFGEVFKARHRKTGQKVALKKVLMENEKEGFP
          *                   *  *** ** * * ***   **  *****       ** *
                70        80        90        100       110       120
cdk2      STAIREISLLKELNHPNIVKLLDVIHT--------ENKLYLVFEFLHQDLKKFMDASALT
cdk9      ITALREIKILQLLKHENVVNLIEICRTKASPYNRCKGSIYLVFDFCEHDLAGLL-SNVLV
           ** ***  *  * * * * *     *            **** *   **        *
                130       140       150       160       170       180
cdk2      GIPLPLIKSYLFQLLQGLAFCHSHRVLHRDLKPQNLLINTEGAIKLADFGLARAFGVPVR
cdk9      KFTLSEIKRVMQMLLNGLYYIHRNKILHRDMKAANVLITRDGVLKLADFGLARAFSLAKN
             *  **     ** **   *    **** *  * **   *  ***********
                190       200       210       220       230       240
cdk2      T----YTHEVVTLWYRAPEILLGCKYYSTAVDIWSLGCIFAEMVTRRALFPGDSEIDQLF
cdk9      SQPNRYTNRVVTLWYRPPELLLGERDYGPPIDLWGAGCIMAEMWTRSPIMQGNTEQHQLA
               **  ******* ** ***   *    * *  *** *** **     *  *  **
                250       260       270       280       290       300
cdk2      RIFRTLGTPDEVVWPGVTSMPDY-KPSFPKWARQDFSKVVPP--LDEDGRSLLSQMLHYD
cdk9      LISQLCGSITPEVWPNVDNYELYEKLELVKGQKRKVKDRLKAYVRDPYALDLIDKLLVLD
           *    *     *** *     * *   *               *     *    *  *
                310       320       330       340       350       360
cdk2      PNKRISAKAALAHPFF------QDVTKPVP-HL--------------------------
cdk9      PAQRIDSDDALNHDFFWSDPMPSDLKGMLSTHLTSMFEYLAPPRRKGSQITQQSTNQSRN
          *  **    ** * **       *       **
                370
cdk2      ----------RL-
cdk9      PATTNQTEFERVF
          *
```

Fig. 4. Sequence alignment of CDK2 with CDK9 (38% identity). The alignment was performed with the program MULTIALIGN [30].

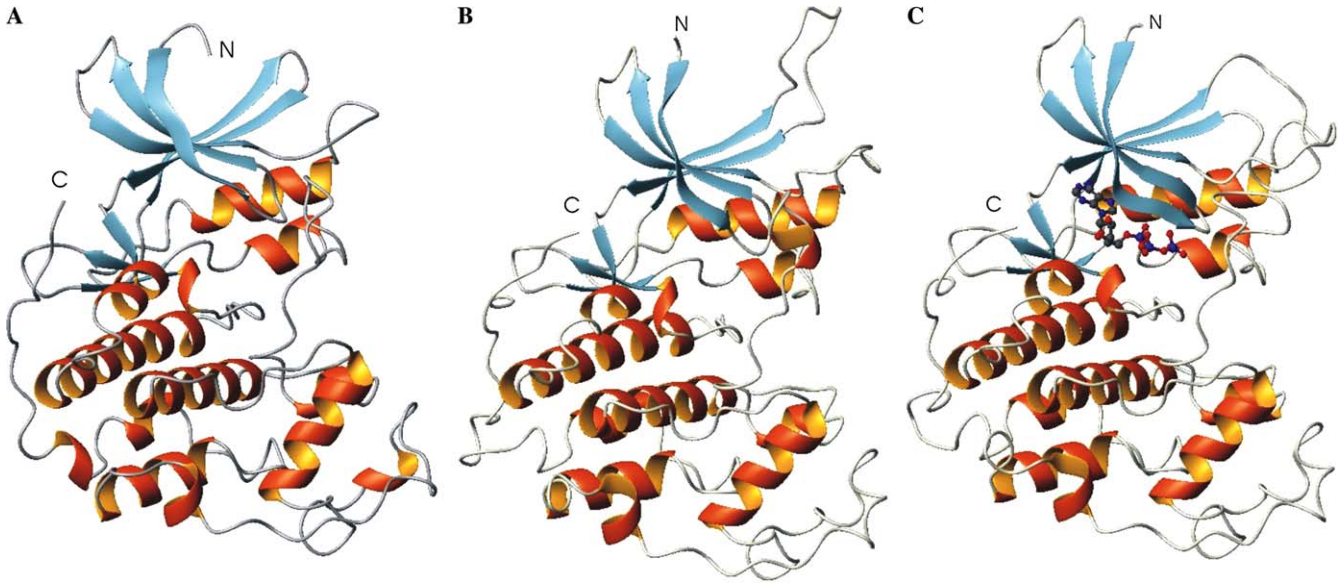


Fig. 5. Ribbon diagram of (A) CDK2 structure; (B) CDK9 model; and (C) CDK9:ATP model. Figure was generated by MolMol [31].

longer than CDK2. Therefore, the N- and C-terminal were not modeled, due to the absence of models. These regions can provide specificity to CDK9, such as the differences in the loops 52–55, 88–95, 182–185, and 317-322, which can be observed in the alignment of their primary sequences (Fig. 4). Considering the modeled region, CDK9 shows 26% of helices. Such result is consistent with CD data of the entire CDK9 protein.

Pinhero et al. [29] established a protocol for the expression and purification of the complex His_6-CDK9/CycT1 using recombinant baculovirus in *Spodoptera frugiperda* insect cells. On an average, from 1 L culture of infected Sf9 cells, these authors could purify about 1.2 mg of pure His_6-CDK9/CycT1. Compared to this study, our protocol in *E. coli* presents some advantages in respect to yield of purified protein, fusion tags, expression system, and operating costs/complexity.

In summary, ours results show that recombinant human CDK9 can be cloned and expressed in *E. coli* and purified from soluble fraction in two chromatographic steps. Mass spectrometry identification validated the expression of CDK9. Moreover, expression at a lower temperature using 0.1 mM IPTG resulted in a secondary structured recombinant protein as shown by CD analysis validating its use in future structural analysis. This is the first report on human CDK9 expression in *E. coli* and structure analysis and provides the first step for the development of new inhibitors for the treatment of AIDS, heart disease, and cancer.

Acknowledgments

This work was supported by grants from FAPESP (Proc.01/07532-0, 02/04383-7, and 02/09388-7), FAPERJ, CNPq, CAPES.

References

[1] X. Grana, A. De Luca, N. Sang, Y. Fu, P.P. Claudio, J. Rosenblatt, D.O. Morgan, A. Giordano, PITALRE, a nuclear CDC2-related protein kinase that phosphorylates the retinoblastoma protein in vitro, Proc. Natl. Acad. Sci. USA 91 (1994) 3834–3838.
[2] G. De Falco, Giordano A, CDK9: from basal transcription to cancer and AIDS, Cancer Biol. Ther. 1 (2002) 342–347.
[3] X. Yang, M.O. Gold, D.N. Tang, D.E. Lewis, E. Aguilar-Cordova, A.P. Rice, C.H. Herrmann, TAK, an HIV Tat-associated kinase, is a member of the cyclin-dependent family of protein kinases and is induced by activation of peripheral blood lymphocytes and differentiation of promonocytic cell lines, Proc. Natl. Acad. Sci. USA 94 (1997) 12331–12336.
[4] J. Garriga, S. Bhattacharya, J. Calbo, R.M. Marshall, M. Truongcao, D.S. Haines, X. Grana, CDK9 is constitutively expressed throughout the cell cycle, and its steady-state expression is independent of SKP2, Mol. Cell. Biol. 23 (2003) 5165–5173.
[5] J. Pines, The cell cycle kinases, Semin. Cancer Biol. 5 (1994) 305–313.
[6] P.D. Jeffrey, A.A. Russo, K. Polyak, E. Gibbs, J. Hurwitz, J. Massagué, N.P. Pavletich, Mechanism of CDK activation revealed by the structure of a cyclinA-CDK2 complex, Nature 376 (1995) 313–320.
[7] S.H. Chao, D.H. Price, Flavopiridol inactivates P-TEFb and blocks most RNA polymerase II transcription in vivo, J. Biol. Chem. 276 (2001) 31793–31799.
[8] J. Peng, N.F. Marshall, D.H. Price, Identification of a cyclin subunit required for the function of *Drosophila* P-TEFb, J. Biol. Chem. 273 (1998) 13855–13860.
[9] T.J. Fu, J. Peng, G. Lee, D.H. Price, O. Flores, Cyclin K functions as a CDK9 regulatory subunit and participates in RNA polymerase II transcription, J. Biol. Chem. 274 (1999) 34527–34530.
[10] S.M. Shore, S.A. Byers, W. Maury, D.H. Price, Identification of a novel isoform of Cdk9, Gene 307 (2003) 175–182.
[11] Y. Zhu, T. Pe'ery, J. Peng, Y. Ramanathan, N. Marshall, T. Marshall, B. Amendt, M.B. Mathews, D.H. Price, Transcription elongation factor P-TEFb is required for HIV-1 tat transactivation in vitro, Genes Dev. 11 (1997) 2622–2632.
[12] J. Garriga, X. Grana, Cellular control of gene expression by T-type cyclin/CDK9 complexes, Gene 337 (2004) 15–23.
[13] T. Ammosova, K. Washington, Z. Debebe, J. Brady, S. Nekhai, Dephosphorylation of CDK9 by protein phosphatase 2A and protein

phosphatase-1 in Tat-activated HIV-1 transcription, Retrovirology 2 (2005) 47.
[14] M. Sano, M.D. Schneider, Cyclins that don't cycle—cyclin T/cyclin-dependent kinase-9 determines cardiac muscle cell size, Cell Cycle 2 (2003) 99–104.
[15] D.E. MacCallum, J. Melville, S. Frame, K. Watt, S. Anderson, A. Gianella-Borradori, D.P. Lane, S.R. Green, Seliciclib (CYC202, R-Roscovitine) induces cell death in multiple myeloma cells by inhibition of RNA polymerase II-dependent transcription and down-regulation of Mcl-1, Cancer Res. 65 (2005) 5399–5407.
[16] J.J. Chalmers, E. Kim, J.N. Telford, E.Y. Wong, W.C. Tacon, M.L. Shuler, D.B. Wilson, Effects of temperature on *Escherichia coli* overproducing beta-lactamase or human epidermal growth factor, Appl. Environ. Microbiol. 56 (1990) 104–111.
[17] A. Lombardi, M. Sperandei, C. Cantale, P. Giacomini, P. Galeffi, Functional expression of a single-chain antibody specific for the HER2 human oncogene in a bacterial reducing environment, Protein Expr. Purif. 44 (2005) 10–15.
[18] M.M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, Anal. Biochem. 72 (1976) 248–254.
[19] A. Gorg, W. Postel, A. Domscheit, S. Gunther, Two-dimensional electrophoresis with immobilized pH gradients of leaf proteins from barley (*Hordeum vulgare*): method, reproducibility and genetic aspects, Electrophoresis 9 (1988) 681–692.
[20] N. Sreerama, R.W. Woody, Protein secondary structure from circular dichroism spectroscopy. Combining variable selection principle and cluster analysis with neural network, ridge regression and self-consistent methods, J. Mol. Biol. 242 (1994) 497–507.
[21] G. Deleage, C. Geourjon, An interactive graphic program for calculating the secondary structure content of proteins from circular dichroism spectrum, Comput. Appl. Biosci. 9 (1993) 197–199.
[22] H.B. Uchoa, G.E. Jorge, N.J. da Silveira, J.C. Câmera, F. Canduri, W.F. De Azevedo, Parmodel: a web server for automated comparative modeling of proteins, Biochem. Biophys. Res. Commun. 325 (2004) 1481–1486.
[23] A. Sali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, J. Mol. Biol. 234 (1993) 779–815.
[24] U. Schulze-Gahmen, J. Brandsen, H.D. Jones, D.O. Morgan, L. Meijer, J. Vesely, S.-H. Kim, Multiple modes of ligand recognition: crystal structures of cyclin-dependent protein kinase2 in complex with ATP and two inhibitors, olomoucine and isopentenyladenine, Proteins 22 (1995) 378–391.
[25] U. Schulze-Gahmen, H.L. De Bondt, S.-H. Kim, High-resolution crystal structures of human cyclin-dependent kinase 2 with and without ATP: bound waters and natural ligand as guides for inhibitor design, J. Med. Chem. 39 (1996) 4540–4546.
[26] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, J. Appl. Cryst. 26 (1993) 283–291.
[27] W.F.JR. De Azevedo, S. Leclerc, L. Meijer, L. Havlicek, M. Strnad, S.-H. Kim, Inhibition of cyclin-dependent kinases by purine analogues: crystal structure of human CDK2 complexed with roscovitine, Eur. J. Biochem. 243 (1997) 518–526.
[28] F. Canduri, H.B. Uchoa, W.F. De Azevedo Jr., Molecular models of cyclin-dependent kinase 1 complexed with inhibitors, Biochem. Biophys. Res. Commun. 324 (2004) 661–666.
[29] R. Pinhero, P. Liaw, K. Yankulov, A uniform procedure for the purification of CDK7/CycH/MAT1, CDK8/CycC and CDK9/CycT1, Biol. Proced. Online 6 (2004) 163–172.
[30] F. Corpet, Multiple sequence alignment with hierarchical clustering, Nucleic Acids Res. 16 (1988) 10881–10890.
[31] R. Koradi, M. Billeter, K. Wüthrich, MOLMOL: a program for display and analysis of macromolecular structures, J. Mol. Graph. 14 (1996) 51–55.