

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL – UFRGS**  
**INSTITUTO DE CIÊNCIAS BIOLÓGICAS - ICBS**  
**DEPARTAMENTO DE BIOQUÍMICA**  
**PROGRAMA DE PÓS GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS – BIOQUÍMICA**

**MODELAGEM MOLECULAR E ESTUDOS DE  
DOCKING DA ENZIMA CORISMATO SINTASE DE  
*MYCOBACTERIUM TUBERCULOSIS***

**Cláudia Lemelle Fernandes**

PORTO ALEGRE  
2006

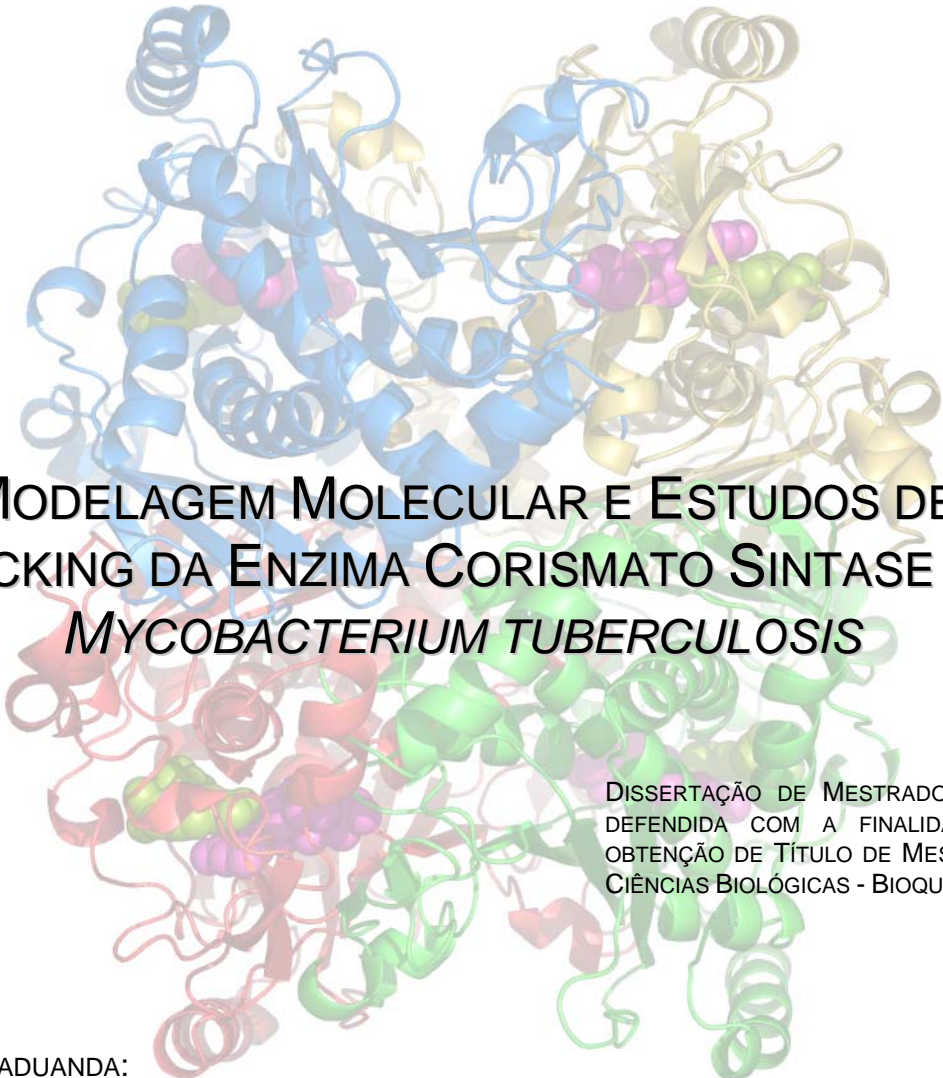


UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL – UFRGS

INSTITUTO DE CIÊNCIAS BIOLÓGICAS

DEPARTAMENTO DE BIOQUÍMICA

PROGRAMA DE PÓS GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS – BIOQUÍMICA



MODELAGEM MOLECULAR E ESTUDOS DE  
DOCKING DA ENZIMA CORISMATO SINTASE DE  
*MYCOBACTERIUM TUBERCULOSIS*

DISSERTAÇÃO DE Mestrado a ser  
defendida com a finalidade de  
obtenção de título de Mestre em  
Ciências Biológicas - Bioquímica

PÓS-GRADUANDA:

**Cláudia Lemelle Fernandes**

ORIENTADOR:

**Luiz Augusto Basso**

Porto Alegre, agosto de 2006.

## AGRADECIMENTOS

- Aos professores Luiz Augusto Basso e Diógenes Santiago Santos pela oportunidade de ter desenvolvido este trabalho.
- Ao professor Osmar Norberto de Souza pela sua importância na minha formação profissional e pela sua atenção dedicada nestes anos.
- Ao Laboratório de bioinformática, Modelagem e Simulação de Biosistemas (LABIO) da Faculdade de Informática da PUCRS, onde todo este trabalho foi desenvolvido.
- Aos meus colegas de laboratório e de maneira muito especial a Ardala Breda e Evelyn Koeche Schroeder que mais que colegas são amigas com quem sempre pude contar.
- Aos amigos que fiz no PPG de Ciências da Computação da PUCRS e também de forma especial à Mariana Luderitz Kolberg, Sandra Rosa e Pedro Velho.
- A meus pais, Valmir Dasso Fernandes e Nancy Maria Lemelle Fernandes pelo seu apoio incondicional.
- À Marília Burger mais que minha cunhada uma amiga com quem pude desabafar.
- Ao meu irmão Paulo Fernandes sempre uma inspiração para eu seguir no caminho da pesquisa pelo seu apoio e conselhos.
- A Capes, CNPq e FAPERGS que financiaram parte dos recursos envolvidos neste projeto.

NÃO VEMOS AS COISAS  
COMO ELAS SÃO,  
MAS COMO NÓS SOMOS

ANAIS NIN

## ÍNDICE

PARTE I	
RESUMO .....	01
ABSTRACT .....	02
LISTA DE ABREVIACÕES .....	03
1 INTRODUÇÃO .....	04
1.1 A TUBERCULOSE .....	04
1.2 A VIA DO ÁCIDO CHIQUÍMICO .....	07
1.2.1 A VIA DO ÁCIDO CHIQUÍMICO PASSO A PASSO .....	08
1.3 A COENZIMA FMN .....	10
1.4 A ENZIMA CORISMATO SINTASE .....	11
1.4.1 A ESTRUTURA DA CS .....	13
2 OBJETIVOS .....	15
PARTE II	
3 ARTIGO 1 .....	16
Fernandes, C.L.; Breda, A.; Santos, D.S.; Basso, L.A.; Norberto de Souza, O., <b>2006</b> . A Structural Model for Chorismate Synthase from <i>Mycobacterium tuberculosis</i> in Complex with Coenzyme and Substrate. <i>Computers in Biology and Medicine</i> , no prelo.	
4 ARTIGO 2 .....	27
Fernandes, C.L.; Santos, D.S.; Basso, L.A.; Norberto de Souza, O., <b>2005</b> . Structure Prediction and Docking Studies of Chorismate Synthase from <i>Mycobacterium tuberculosis</i> . <i>Lectures Notes in Computer Science</i> , <b>3594</b> , 118-127.	
PARTE III	
5 RESULTADOS E DISCUSSÃO .....	38
5.1 ESTUDO DA SEQUÊNCIA DE <i>MTB</i> .....	38
5.2 MODELAGEM MOLECULAR POR HOMOLOGIA .....	40
5.2.1 RESULTADOS DO PROCHECK .....	42

5.2.2 RESULTADOS DO VERIFY 3D.....	44
5.3 A ESTRUTURA QUATERNÁRIA.....	45
5.4 ESTUDOS DE <i>DOCKING</i> .....	46
5.4.1 INTERAÇÕES COM O FMN.....	48
5.4.2 INTERAÇÕES COM O EPSP.....	49
6 CONCLUSÃO E PERSPECTIVAS.....	51
7 REFERÊNCIAS.....	53
ANEXOS	
ANEXO 1: ALINHAMENTO MÚLTIPLO DA ENZIMA CORISMATO SINTASE.....	58
ANEXO 2: ARQUIVOS DE ENTRADA PARA AS MINIMIZAÇÕES DE ENERGIA.....	65

# PARTE I

## RESUMO

As enzimas da via do ácido chiquímico constituem um excelente alvo para o desenho de novos agentes antibacterianos. Esta rota é encontrada em bactérias, fungos, plantas e parasitas do *filo* apicomplexa, mas está ausente em mamíferos. A Corismato sintase (CS) catalisa o último passo desta rota, produto que é utilizado em outras reações biossintéticas, como biossíntese de aminoácidos aromáticos, folato, vitamina K e ubiquinona. Esta reação é a mais incomum de toda rota e é única na natureza. Ela converte 5-enolpiruvil-chiquimato-3-fosfato (EPSP) em corismato na presença de uma flavina mononucleotídeo reduzida (FMNH<sub>2</sub>) como coenzima. A predição da estrutura tridimensional (3D) da enzima CS de *Mycobacterium tuberculosis* (Mtb) foi feita pela técnica de modelagem comparativa por homologia, utilizando a estrutura cristalográfica de CS de *Streptococcus pneumoniae* (PDB ID: 1QX0) como molde ( $\approx$  42% de identidade), e o programa MODELLER6v2. Adicionalmente, com o intuito de entender as possíveis formas de ligação do substrato e da coenzima a enzima EPSP e flavina mononucleotídeo (FMN), respectivamente, foi feito um *docking* geométrico. A minimização de energia de todo o complexo mostrou, como esperado, que a maioria das interações ocorridas no molde são preservadas na estrutura de *Mtb*, exceto pela His11, Arg139 e Gln255. Entretanto, novas interações envolvendo Arg111, Gli113 e Ser317 também foram observadas. Este conhecimento poderá facilitar na busca por novos inibidores para esta enzima como fármacos alternativos no tratamento da tuberculose (Tb).



## ABSTRACT

The enzymes of the shikimate pathway constitute an excellent target for the design of new antibacterial agents. This pathway is found in bacteria, fungi, plants and apicomplexan parasites but is absent in mammals. Chorismate synthase (CS) catalyzes the last step of this pathway, the product of which is utilized in other enzymatic transformations like the biosynthesis of aromatic amino acids, folate, vitamin K and ubiquinone. This reaction is the most unusual of the entire pathway and is unique in nature. It converts EPSP to chorismate in the presence of a reduced FMN coenzyme. Structure prediction used the comparative protein structure modeling methodology. The three-dimensional (3D) structure prediction of the enzyme CS of *Mycobacterium tuberculosis* was performed using the crystal structure (PDB ID: 1QX0) of CS from *Streptococcus pneumoniae* as template ( $\approx 42\%$  identity), and the MODELLER6v2 package. Additionally, in order to understand the possible binding modes of substrate and coenzyme to the enzyme, EPSP and FMN, respectively, were geometrically docked to CS. Energy minimization of the whole complex showed, as expected, that most of the template interactions are preserved in the *Mtb* structure, except for His11, Arg139 and Gln255. However, novel interactions involving Arg111, Gly113 and Ser317 were also observed. This knowledge should facilitate the search for inhibitors of this enzyme as alternative agents to treat tuberculosis.

## LISTA DE ABREVIÇÕES

3D – Tridimensional

CS – Corismato sintase

E4P – Eritrose 4-fosfato

EPSP – 5-enolpiruvil-chiquimato-3-fosfato

FMN – Flavina mononucleotídeo oxidada

FMNH<sub>2</sub> - Flavina mononucleotídeo reduzida

MS – Ministério da saúde

*Mtb* - *Mycobacterium tuberculosis*

NADPH – Nicotinamida adenina dinucleotídeo-fosfato reduzida

PDB – Protein Data Bank

PEP – Fosfoenolpiruvato

OMS - Organização Mundial da Saúde

RMSD - Desvio médio quadrático

Tb – tuberculose

# 1 INTRODUÇÃO

## 1.1 A Tuberculose

Dentre as doenças infecto-contagiosas a Tb requer um estudo mais aprofundado. A Tb, através do seu agente etiológico, o *Mtb*, é a responsável pelo maior número de mortes por um único agente infeccioso no mundo (Garble, 1991).

Os números desta epidemia são alarmantes. Estima-se que um terço da população mundial esteja infectada com o *Mtb* e a expectativa é que 10% delas (ou aproximadamente 200 milhões de pessoas) desenvolvam a doença (Daffé, 1998). Segundo a Organização Mundial de Saúde (OMS), foram estimados 8,8 milhões de novos casos em 2003 (OMS, 2005).

No Brasil, estima-se que cerca de 50 milhões de pessoas estejam infectadas. Estes números colocam o Brasil junto com o Perú como os responsáveis por 50% de todos os casos das Américas. Conforme dados do Ministério da Saúde (MS), foram cerca de 78.000 novos casos registrados em 2002, mantendo uma média de aproximadamente 80.000 novos casos registrados por ano (MS, 2005). Entretanto, de acordo com a OMS, estima-se que a média anual no Brasil seja de 90-95 mil casos por ano entre novos e re-tratados (OMS, 2005).

A Tb é uma doença infecciosa causada por micobactérias do complexo *Mycobacterium tuberculosis*, das quais a principal é o *Mtb*. O *Mtb* é um bacilo pequeno (1,4 µm de comprimento por 0,3-0,6 µm de diâmetro), sem flagelos com parede celular rica em lipídeos, classificado pelos métodos colorimétricos como bacilo álcool ácido resistente (BAAR). É uma bactéria de crescimento lento, estritamente aeróbia, mas podendo sobreviver em ambiente de micro-aerofilia e que coloniza preferencialmente as cavidades das paredes pulmonares (Grosset, 1980),

porém a partir do pulmão a infecção pode se espalhar para outras partes do corpo do hospedeiro ([Russell, 2001](#)).

A infecção é normalmente transmitida de pessoa para pessoa por inalação de pequenas partículas de aerossóis expelidas por pacientes com a doença na forma pulmonar e que, se inaladas por um indivíduo sadio, se alojarão no espaço alveolar, onde iniciam a multiplicação, posteriormente sendo fagocitadas por macrófagos do hospedeiro. O acúmulo de células de defesa do organismo como monócitos, neutrófilos e macrófagos no local da infecção gera uma estrutura granulomatosa compacta, a qual é chamada de tubérculo ([Hiryanna & Ramakrishnan, 1986](#)).

A forma como o *Mtb* sobrevive no hospedeiro é um dos problemas no combate à doença. A infecção inicial é assintomática e controlada facilmente na maioria dos casos, porém o *Mtb* normalmente sobrevive dentro dos tecidos durante anos em estado de dormência. Eventualmente, devido a um momento de deficiência do sistema imunológico, a bactéria dormente é reativada causando o ressurgimento da doença, o que pode acontecer até muitas décadas após a infecção inicial ([Manabe & Bishai, 2000](#)). A existência destas formas dormentes dificulta o controle da doença, provavelmente devido à baixa atividade bactericida das drogas antituberculose contra estes bacilos em estado dormente ou semi-dormente. Por isso, o tratamento deve ser longo (de no mínimo 6 meses) e envolve a combinação de diversas drogas ([Coates, 1998](#)).

A prescrição mais comum para o tratamento da Tb utiliza a combinação das drogas isoniazida, rifampicina, pirazinamida e etambutol ou estreptomicina nos primeiros dois meses, com o intuito de exercer uma ação bactericida contra o bacilo ativo, e, posteriormente, isoniazida e rifampicina para os últimos quatro meses. Estas têm a função de esterilização contra bacilos dormentes ou semi-dormentes.

Este tratamento, apesar de longo, é importante para o controle da doença (Basso & Blanchard, 1998). Deve-se ressaltar, entretanto, a importância do tratamento. A morte por Tb em casos não tratados pode chegar a 40-60% dos casos, enquanto, com tratamento esse número cai para 10% com o uso correto de antibióticos.

Além da quimioterapia, a vacinação com BCG (*Bacillus Calmette-Guerin*) é usada como tratamento profilático contra infecções micobacterianas. Porém este tratamento tem resultados diversos, protegendo algumas populações e em outras não tendo a mesma eficácia (OMS, 2005). Os resultados da vacinação no Brasil relatam uma cobertura de 100% da população desde 1995 até 2002 (MS, 2005).

A partir da década de 40 até o início dos anos 80, o tratamento com estes quimioterápicos facilitou o controle e o declínio do número de casos da doença. Entretanto, o número de casos começou a crescer novamente a partir da segunda metade da década de 80. Alguns fatores inter-relacionados vêm contribuindo para o reaparecimento mundial da tuberculose. Entre os mais significativos estão a epidemia de AIDS, a diminuição do poder sócio-econômico da população, a aglomeração em grandes cidades, o declínio das condições de saúde pública e o abandono do tratamento (Basso *et al.*, 1998). Aliado a isso, cepas de *Mtb* resistentes às drogas antituberculose têm sido identificadas globalmente e de forma alarmante (Heym *et al.*, 1995; Cohn *et al.*, 1997). Exemplos do surgimento de cepas resistentes no Brasil também foram notificados (Telles *et al.*, 2005; Silva *et al.*, 2003).

No Brasil há ainda um outro aspecto preocupante do panorama da doença. Segundo a OMS, taxas superiores a 45% de pacientes previamente tratados apresentam multi-resistência adquirida, atestando a falência do atual esquema de tratamento da Tb em nosso país (OMS, 2005). As taxas de abandono do tratamento no país apresentam uma variação constante de 10% entre 1993 e 2002, enquanto

as taxas de cura apresentam um percentual entre 75 e 80% a partir de 1994 (MS, 2005).

Em 1998, Cole e colaboradores publicaram a seqüência completa do genoma do *Mtb* H37Rv. O genoma compreende 4.411.529 pares de bases e contém cerca de 4.000 genes, reanotados em 2002 (Cole *et al.*, 1998; Camus *et al.*, 2002). Com relação a prevalência de cada uma das bases, há uma preferência para produção de aminoácidos com códons ricos em C e G (Ala, Gli, Pro Arg, Trp) em comparação a aminoácidos com códons ricos em A e T (Asn, Lis, Phe, Tir) (Cole *et al.*, 1998). A enzima objeto do nosso estudo é a CS, codificada pelo gene *aroC*, que contém 1.201 pares de bases ou 401 aminoácidos (Bentley, 1990).

Com base neste panorama, o estudo da Tb e de seu agente é mais do que justificado. O conhecimento apropriado desta doença e seu do agente etiológico são fundamentais para os avanços em seu combate, e conseqüentemente para a melhoria das condições de saúde pública mundial. Isso possibilitará o desenvolvimento de novos fármacos para o combate ao *Mtb*, tanto na sua forma sensível como na forma resistente às drogas atuais.

## **1.2 A Via do Ácido Chiquímico**

Na biossíntese de aminoácidos aromático, a rota inicial, ou de caminho comum, é conhecida como via do ácido chiquímico. A partir de seu produto final, o corismato, a rota se divide para produção de fenilalanina, tirosina e triptofano além de outros metabólitos que também são oriundos do corismato, como a ubiquinona, menaquinona, ácidos micólicos e vitamina K (Figura 1) e que são igualmente importantes para a viabilidade do organismo (Du *et al.*, 2000; Pittard, 1996).

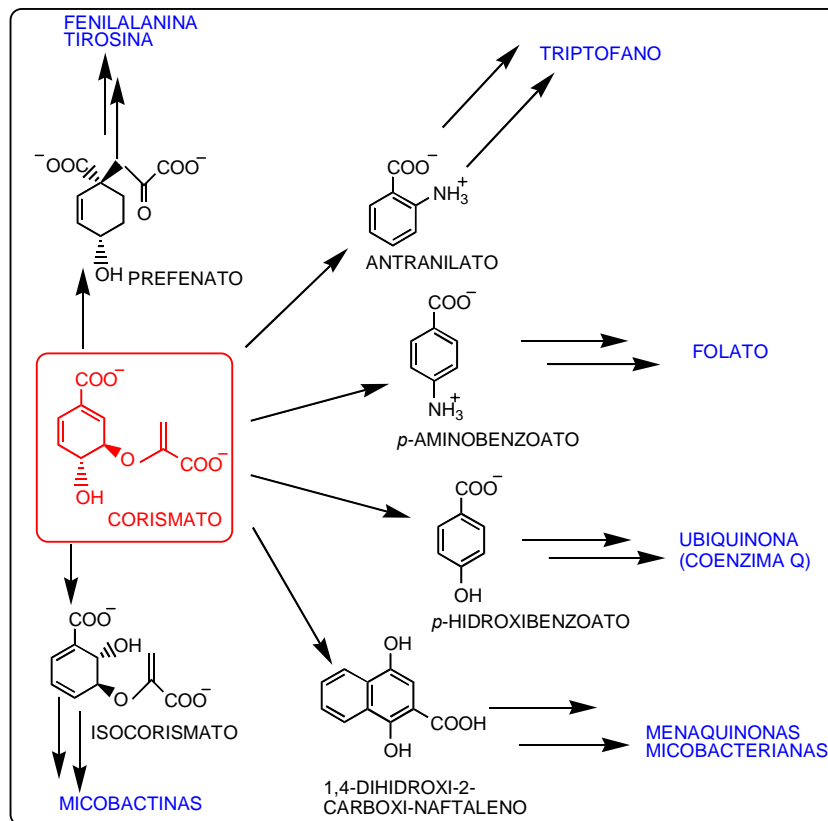


FIGURA 1: OS DESTINOS DO CORISMATO A PARTIR DA VIA DO ÁCIDO CHIQUÍMICO E OS PRODUTOS FINAIS DE CADA UMA DAS VIAS NO *MTB* (EM AZUL).

Esta via está presente somente em fungos, bactérias, protozoários do *filo* apicomplexa e em vegetais, sendo então uma via ideal como alvo para o estudo de novos fármacos e de pesticidas para a lavoura, como é o caso do glifosato (Roundup®), um inibidor que atua sobre a 5-enolpiruvil chiquimato-3-fosfato sintase (EPSP sintase), a sexta enzima da via do ácido chiquímico (Sikorski & Gruys, 1997).

### 1.2.1 A via do ácido chiquímico passo-a-passo

Os precursores desta via são a eritrose-4-fosfato (E4P) e o fosfoenolpiruvato (PEP). No estudo das vias metabólicas é usual se classificar uma via de acordo com seus precursores. Aqui temos dois tipos de precursores. O PEP é um intermediário glicolítico enquanto a E4P é um intermediário da via pentose-fosfato (Lehninger et

al., 1995). A via completa (Figura 2) possui sete reações, que serão descritas a seguir:

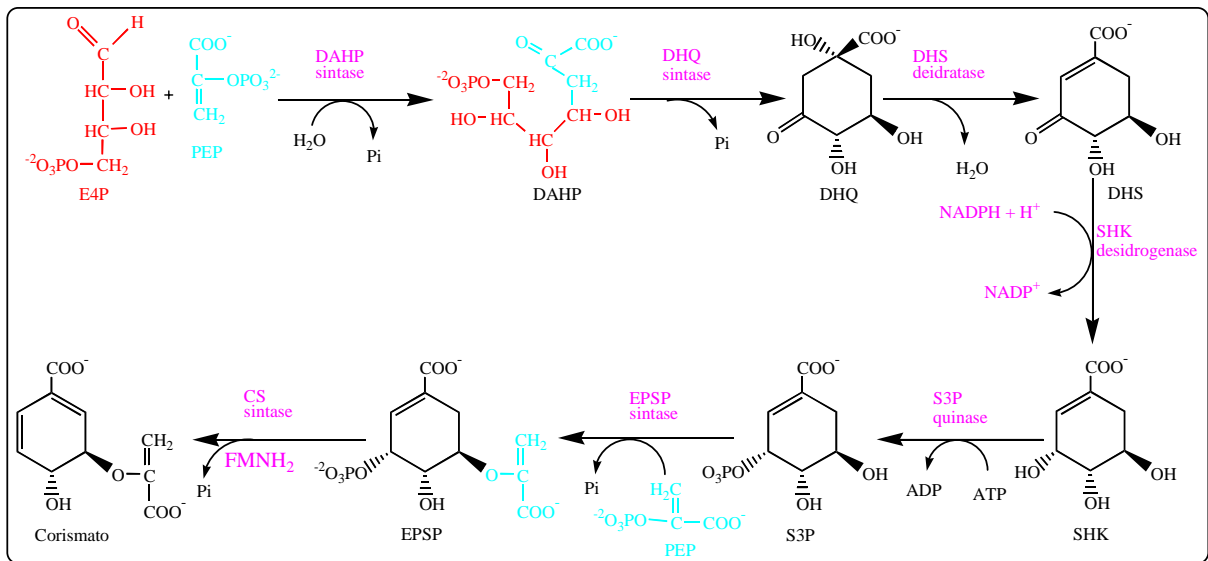


FIGURA 2: VIA DO ÁCIDO CHIQUÍMICO. VER DESCRIÇÃO DETALHADA NO TEXTO. ADAPTADO DE PITTARD, 1996.

1. Condensação do PEP com a E4P formando um intermediário acíclico com sete carbonos, o 2-ceto-3-desoxi D-arabinoeptulose-7-fostato, pela ação da enzima 2-ceto-3-desoxi D-arabinoeptulose-7-fostato sintase (DAHP sintase), liberando fosfato inorgânico (Pi);
2. Ciclização do substrato em 3-desidroquinato, pela desidroquinato sintase (DHQ sintase), com a presença de  $\text{NAD}^+$  como coenzima;
3. Desidratação do substrato a 3-desidrochiquimato, pela 3-desidroquinato desidratase (DHS desidratase);
4. Redução do substrato pela chiquimato desidrogenase (SHK desidrogenase), com a coenzima NADPH se oxidando a  $\text{NADP}^+$  e formação do chiquimato;
5. Fosforilação do substrato através da chiquimato quinase (S3P quinase), com o gasto de uma molécula de ATP se convertendo a ADP;



6. Condensação de uma segunda molécula de PEP ao substrato, e liberação de outro fosfato inorgânico produzindo o EPSP pela ação da EPSP sintase;
7. Perda do grupamento fosfato do substrato, liberando fosfato inorgânico, e formando corismato, pela ação da corismato sintase (CS sintase) na presença de FMN na forma reduzida (FMNH<sub>2</sub>);

### 1.3 A coenzima FMN

A flavina mononucleotídeo (FMN) faz parte da família das flavinas. A característica comum dessa família é a presença do anel isoaloxazina (Figura 3). Dentre as flavinas estão a riboflavina (vitamina B2), com a presença apenas de uma ribose ligada ao anel isoaloxazina, o FMN (Figura 3), com a adição de um grupamento fosfato à ribose, e a flavina adenina dinucleotídeo (FAD) com uma adenina ligada à ribose por uma ligação fosfodiéster (Bornemann, 2002).

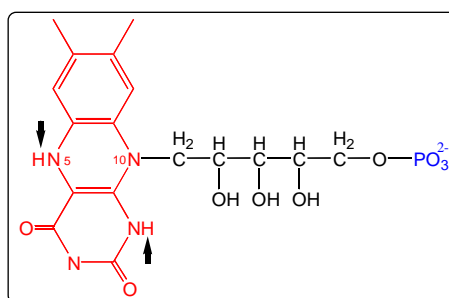


FIGURA 3: ESTRUTURA DO FMN NA FORMA REDUZIDA (FMNH<sub>2</sub>). EM VERMELHO O ANEL ISOALOXAZINA, EM PRETO A RIBOSE E EM AZUL O GRUPAMENTO FOSFATO, QUE DIFERENCIA O FMN DAS OUTRAS FLAVINAS. OS DOIS HIDROGÊNIOS RESPONSÁVEIS PELO ESTADO REDUZIDO ESTÃO ASSINALADOS PELAS SETAS.

A conformação do FMNH<sub>2</sub> induz uma modificação estrutural no anel isoaloxazina dobrando-o em um ângulo de 27,3° em relação ao eixo que passa pelos átomos N5 e N10, enquanto na forma oxidada o anel é sempre planar independente da protonação, (Figura 4) (Zheng & Ornstein, 1996). Esta modificação pode facilitar a entrada da coenzima no sítio ativo.

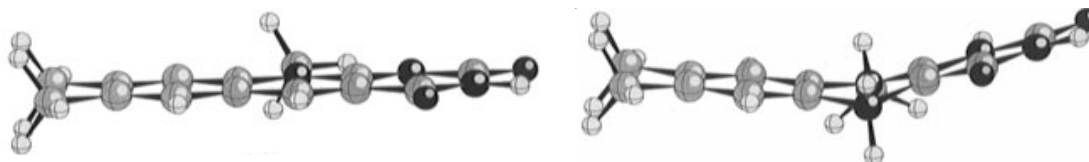


FIGURA 4: VISTA LATERAL DO ANEL ISOALOXAZINA. NA FORMA OXIDADA (ESQUERDA) E NA FORMA REDUZIDA (DIREITA). ADAPTADO DE ZHENG & ORNSTEIN, 1996.

#### 1.4 A enzima *Corismato Sintase (CS)*

A enzima CS ou O5-(1-carboxivinil)-3-fosfochiquimato fosfato liase (EC 4.6.1.4) cataliza a conversão do EPSP em corismato. Esta reação, do ponto de vista bioquímico, é excepcional pois tanto a enzima quanto a reação são únicas na natureza. Esta enzima requer a presença de um FMNH<sub>2</sub> como coenzima para ocorrência da reação (Bentley, 1990; Macheroux *et al.*, 1999).

A reação da CS acrescenta a segunda de três duplas ligações ao anel benzênico do substrato, e formalmente é descrita como uma reação concertada de eliminação *anti*-1,4 do grupo 3-fosfato e a perda do hidrogênio C-(6*proR*). Esta reação apresenta algumas particularidades, como o fato de não envolver nenhuma alteração no estado redox de substrato e produto e apresentar uma estereoquímica do tipo *anti* (eliminações do tipo 1-4 apresentam preferencialmente estereoquímica do tipo *syn*) (Macheroux *et al.*, 1999).

Como o mecanismo de reação ainda não foi completamente elucidado, há diversas alternativas propostas, mas a mais aceita sugere o envolvimento direto do FMN na reação de eliminação, com um mecanismo radicalar com uma transferência de um elétron do FMNH<sub>2</sub> para o substrato, iniciando a quebra da ligação C-O do grupamento fosfato, produzindo um derivado com um radical. Como não há mudança de estado redox na reação, uma transferência reversa do elétron doado é necessária para completar o ciclo catalítico (Figura 5) (Macheroux *et al.*, 1996;

Bornemann *et al.*, 1996). Este mecanismo é corroborado pela resolução da estrutura da CS de *Streptococcus pneumoniae*, onde o FMN, o EPSP e a enzima formam um complexo ternário, onde a porção do anel isoaloxazina do FMN está a uma distância de 3,3Å do EPSP (Macleán & Ali, 2003). A formação deste complexo induz uma mudança conformacional na enzima (Macheroux *et al.*, 1998).

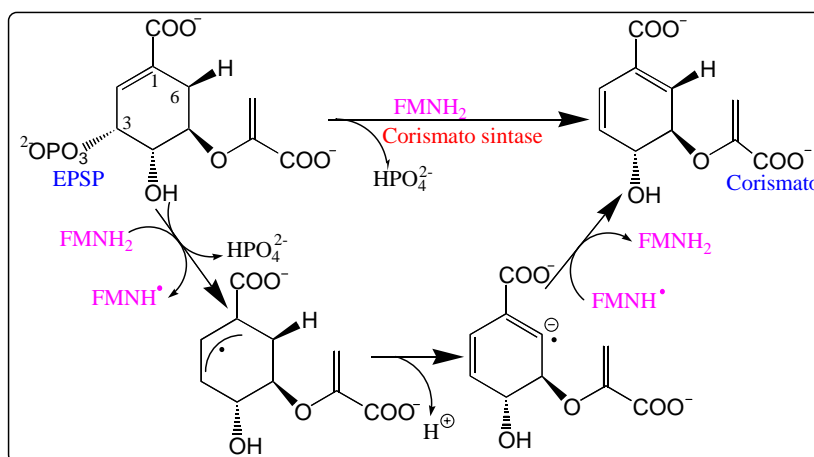


FIGURA 5: REAÇÃO CATALISADA PELA CORISMATO SINTASE E MECANISMO DE REAÇÃO MAIS ACEITO (ABAIXO), ONDE O  $FMNH_2$  DOA UM ELÉTRON PARA A QUEBRA DA LIGAÇÃO DO FOSFATO (3), SENDO ESTE REGENERADO POSTERIORMENTE PELA DOAÇÃO DE UM ELÉTRON PELO SUBSTRATO (6), PARA A COENZIMA.

Embora catalisando uma reação única na natureza em termos bioquímicos, a CS, do ponto de vista catalítico, é muito diferente nos diversos organismos estudados, apresentando diferentes propriedades, assim como diferentes tamanhos moleculares. Desta forma foi criada uma classificação para a CS de acordo com a forma com que a enzima obtém o  $FMNH_2$ .

Quando a CS realiza a reação a partir do  $FMNH_2$ , formado a partir de outra reação catalisada por alguma FMN redutase, ela é classificada como uma enzima monofuncional. As enzimas presentes em eubactérias, plantas e parasitas do filo apicomplexa são essencialmente monofuncionais (Macheroux *et al.*, 1999; Fitzpatrick *et al.*, 2001)

Nas bactérias gram-positivas como o *Bacillus subtilis* e *Staphylococcus aureus*, a CS também é monofuncional, dependendo da obtenção de FMN já reduzido para a ocorrência da reação. Entretanto sua estrutura quaternária se apresenta de forma diferente, pois forma um complexo heterotrimérico com uma enzima NADPH dependente, responsável pela redução do FMN para a ocorrência da reação da CS, e também com outra enzima da via do ácido chiquímico, a desidroquinato sintase, segunda enzima da via. Para a ocorrência da reação pela CS, além da enzima redutora do FMN e do NADPH, também é necessária a presença de um íon magnésio ( $Mg^{2+}$ ) (Hassan & Nester, 1978; Horsburgh *et al.*, 1996).

Já em fungos, a CS apresenta uma segunda atividade catalítica, a redução do FMN às custas de uma segunda coenzima, o NADPH. Este tipo específico de CS geralmente tem um peso molecular maior e é classificada como bifuncional. Possivelmente este peso molecular maior está relacionado com a presença deste domínio de ligação extra para o NADPH (Henstrand *et al.*, 1996). Estudos preliminares sobre o sítio de ligação do NADPH sugerem que este se encontra sobreposto ao sítio do FMN (Kitzing *et al.*, 2001).

#### **1.4.1 A Estrutura da CS**

Devido à sua natureza única, a primeira estrutura 3D de CS levou mais tempo para ser determinada do que a das outras enzimas da via do ácido chiquímico. A primeira estrutura 3D publicada data de 2003. Neste trabalho foi resolvida a estrutura 3D da CS de *S. pneumoniae* cristalizada junto com o substrato e a coenzima na forma oxidada. Foi identificada uma nova topologia característica desta proteína com um dobramento  $\beta$ - $\alpha$ - $\beta$  do monômero. A CS apresenta uma estrutura quaternária na forma de um homotetrâmero estabilizado por dois arranjos de fitas- $\beta$ , um

estabilizando a formação dos dímeros e o segundo estabilizando a interação entre os dois dímeros anteriormente formados (Maclean & Ali, 2003).

A partir deste mesmo trabalho, também foi estudada a interação do substrato e coenzima com os resíduos da proteína identificando os resíduos importantes para o entendimento da reação.

Em relação ao sítio de ligação do FMN, fica claro que este desempenha um papel de bloqueio para a saída do EPSP do sítio catalítico. Também foram identificados alguns resíduos que fazem interações com a coenzima, como os resíduos Lis311, Ala252 e Asn251, que fazem interações com a porção ribose do FMN (Figura 3). A porção isoaloxazina é mais hidrofóbica e faz poucas interações com a proteína como, por exemplo, com os resíduos conservados His110, Tre315 e Ser338 (Maclean & Ali, 2003).

O sítio de ligação do EPSP é bastante hidrofílico e básico, com 6 argininas e 2 histidinas concentradas na região de entrada do sítio. O EPSP interage com os resíduos Arg39, Arg45, Arg134, Arg48, Arg337, His10 e His110. (Maclean & Ali, 2003). Estas duas histidinas foram identificadas como sendo resíduos chave para a atividade da enzima, pois sua mutação acarreta uma redução de atividade para 10% (His 110) e 5% (His10) (Kitzing *et al.*, 2004).

Também foram publicados estruturas de CS monofuncionais de *Helicobacter pylori* (Ahn *et al.*, 2003), *Aquifex aeolicus* (Viola *et al.*, 2004) e recentemente a de *Mtb* (Dias *et al.*, 2006) além da bifuncional de *Saccharomyces cerevisiae* (Quevillon-Cheruel *et al.*, 2004).

Thomas e colaboradores (Thomas *et al.*, 2003) identificaram inibidores para a CS do grupo benzilideno-diidroxi-furanonas, porém estes estudos de inibição não apresentaram dados estruturais.

## 2 OBJETIVOS

Este trabalho tem por objetivo a modelagem molecular da enzima CS de *Mtb* e, posteriormente, a construção da estrutura quaternária da mesma para o entendimento dos eventos moleculares na interação com o substrato EPSP e com a coenzima da reação, o FMN, visando compreender quais resíduos são importantes para a interação intermolecular.

Para a realização deste trabalho foram propostas as seguintes etapas:

- ⇒ Estudo da seqüência de CS de *Mtb*, visando conhecer os possíveis pontos críticos na construção do modelo;
- ⇒ Procura de moldes para uso na técnica de Modelagem Molecular por Homologia;
- ⇒ Construção e avaliação de um modelo 3D da CS de *Mtb*;
- ⇒ *Docking* manual do substrato e coenzima no sítio catalítico da enzima modelada;
- ⇒ Uso de técnicas de minimização de energia para melhor representar o comportamento dos ligantes no sítio catalítico;
- ⇒ Avaliação e comparação das interações encontradas nos modelos de *docking* em relação ao molde cristalográfico da enzima de CS de *Streptococcus pneumoniae* (Macleay & Ali, 2003);

O conhecimento da forma de ligação da enzima-substrato-coenzima facilitará a procura por novos inibidores que ajam sobre esta enzima e que poderão ser candidatos a novos fármacos no combate à Tb.

## **PARTE II**

# ARTIGO 1

**A STRUCTURAL MODEL FOR CHORISMATE SYNTHASE FROM *MYCOBACTERIUM TUBERCULOSIS*  
IN COMPLEX WITH COENZYME AND SUBSTRATE.**

FERNANDES, C.L.; BREDAS, A.; SANTOS, D.S.; BASSO, L.A.; NORBERTO DE SOUZA, O.

***COMPUTERS IN BIOLOGY AND MEDICINE, 2006.***

*NO PRELO*





ELSEVIER

Computers in Biology and Medicine ■■■ (■■■■) ■■■–■■■

Computers in Biology  
and Medicine

www.intl.elsevierhealth.com/journals/cobm

## A structural model for chorismate synthase from *Mycobacterium tuberculosis* in complex with coenzyme and substrate

Cláudia Lemelle Fernandes<sup>a, b</sup>, Ardala Breda<sup>a, c</sup>, Diógenes Santiago Santos<sup>d, e</sup>, Luiz Augusto Basso<sup>d</sup>, Osmar Norberto de Souza<sup>a, c, d, e, \*</sup>

<sup>a</sup>Laboratório de Bioinformática, Modelagem e Simulação de Biosistemas - LABIO, Faculdade de Informática, PUCRS, Brazil

<sup>b</sup>Programa de Pós-Graduação em Ciências Biológicas - Bioquímica, UFRGS, Brazil

<sup>c</sup>Programa de Pós-Graduação em Biologia Celular e Molecular, PUCRS, Brazil

<sup>d</sup>Centro de Pesquisa em Biologia Molecular e Funcional, TECNOPUC, PUCRS, Brazil

<sup>e</sup>Instituto de Pesquisas Biomédicas, PUCRS, Porto Alegre, RS, Brazil

### Abstract

The enzymes of the shikimate pathway constitute an excellent target for the design of new antibacterial agents; chorismate synthase (CS) catalyzes the last step of this pathway. The prediction of *Mycobacterium tuberculosis* (MTB) CS three-dimensional structure and the geometric docking of the coenzyme FMN and the substrate EPSP were performed using the crystal structure of CS from *Streptococcus pneumoniae* as template. Energy minimization of the whole complex showed, as expected, that most of the template interactions are preserved in the MTB structure, except for HIS11, ARG139 and GLN255. However, novel interactions involving ARG111, GLY113 and SER317 were also observed. © 2006 Elsevier Ltd. All rights reserved.

**Keywords:** Shikimate pathway; Chorismate synthase; Molecular modeling; Energy minimization; Docking; *Mycobacterium tuberculosis*

### 1. Introduction

The shikimate pathway is the common way for the production of various products including folic acid, vitamin K, ubiquinone and the three aromatic amino acids, tryptophan, phenylalanine and tyrosine. In bacteria, fungi, plants and apicomplexan parasites, chorismate, the final product of the shikimate pathway, is the branch point in the biosynthesis for all these products that are essential for these species. The absence of the shikimate pathway in all other species makes it an attractive target for the development of new antibacterial agents [1,2].

Chorismate synthase (CS), the seventh and final step of the shikimate pathway, catalyzes the conversion of 5-enolpyruvylshikimate 3-phosphate (EPSP) to chorismate in the presence of a reduced flavin mononucleotide (FMN) as a coenzyme [3]. The reaction mechanism of the shikimate pathway

has been studied extensively and revealed that the reaction of CS is unique in nature. The reaction involves a 1, 4 elimination of phosphate and the loss of a proton of the C-6 hydrogen. This consists in the formation of the second out of three necessary double bonds to build an aromatic ring (Fig. 1). The enzyme activity requires a reduced FMN molecule which is not consumed during the reaction [4].

The function of the reduced FMN in catalysis was extensively studied. The most accepted mechanism suggests a direct role of reduced FMN in the elimination reaction. FMN transfers the electron transiently to phosphate and the substrate donates an electron to regenerate the FMN. This reaction does not involve an overall change in the redox state [3,5].

Recently, with the first high-resolution X-ray structure of CS from *Streptococcus pneumoniae* (SPN) with the substrate and the coenzyme in the oxidized form [6], the structure of CS from *Saccharomyces cerevisiae* [7], and the structure of CS from *Helicobacter pylori* with the coenzyme in the reduced form [8], studies on the binding mode of substrate and coenzyme in the active site has been started.

\* Corresponding author.

E-mail address: osmarns@inf.pucrs.br (Osmar Norberto de Souza).

How reduced FMN is obtained divides the CS into two classes, monofunctional and bifunctional. Bifunctional CS has an intrinsic ability to reduce flavin (specifically FMN) using NADPH. In monofunctional CS this catalytic activity is not present. The bifunctional enzyme is present in fungi and the monofunctional form in plants and bacteria [3].

The active site of SPN CS is very hydrophilic and extremely basic, with six arginine and two histidine residues. The two histidines in the active site, HIS10 and HIS110, are present in both classes of CS and across all known species, being HIS10 part of a characteristic CS signature sequence [6]. HIS110 is involved in FMN binding while HIS10 protonates the

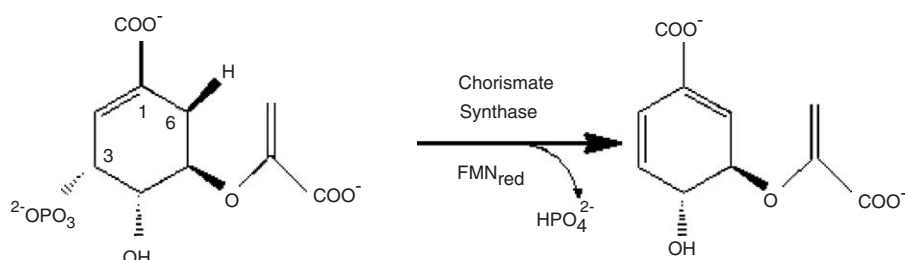


Fig. 1. Reaction catalyzed by CS. The elimination of the 3-phosphate and the loss of a proton in C-6 introduce a second double bond in the ring.

	10	20	30	40	50	60	70
TARGET	MLRWITAGESHGRALVAVVEGMVAGVHVTSADIADQLARRRLGYGRGARMTFERDAVTVLSGIRHGSTLG						
TEMPLATE	-MRYLTAGESHGPRLTAIIIEGIPAGLPLTAEDINEDLRRRQGGYGRGGRMKIENDQVVFTSGVRHGKTTG						
	:*:*:*****	*.*:*:*:	**:*:	*:*:	*:*:	*:*:	*:*:
	80	90	100	110	120	130	140
TARGET	GPIAIEIGNTEWPKWETVMAADPVDPAELADVARNAPLTRPRPGHADYAGMLKYGFDDARFVLERASARE						
TEMPLATE	APITMDVINKDHQKWLDDIMSAEDIEDRLKSKRK-----ITHPRPGHADLVGGIKYRFDDLRLNSLERSARE						
	.**::: *	**	*:*:*:	::	::	*:*****	*.* ** * * * * * * * * *
	150	160	170	180	190	200	210
TARGET	TAARVAAGTVARAFLRQALGVEVLSHVISIGASAPYEGPPRAEDLPAIDASPVRAYDKAAEADMIAQIE						
TEMPLATE	TMRVAVGAVAKRLLAELDMETIANHVVFVGGKEIDVPENLTVAEIKQRAAQSEVSIVNQEREQEIKDYID						
	*: ** * * * * * * * * *	*: *	*: *	*: *	*: *	*: *	*: *
	220	230	240	250	260	270	280
TARGET	AAKKDGDITLGGVVEAVALGLPVGLGSFTSGDHRLDSQLAAAVMGIQAIKVEIGDGFQTARRRGSRAHDE						
TEMPLATE	QIKRDGDTIGGVVETVVGVPVGLGSYVQWDRKLDARLAQAVVSINAFKGVFEGLGFEGYRKGQVMDE						
	*:*****:*****:*	*:*****:..	*:*****:..	*:*****:..	*:*****:..	*:*****:..	*:*****:..
	290	300	310	320	330	340	350
TARGET	MYPG-PDGVVVRSTNRAGGLEGGMTNGQPLRVRAAMKPISTVPRALATVDLATGDEAVAIHQ <del>RS</del> DVCAVPA						
TEMPLATE	ILWSKEDGYTRRTNNLGGFEGGMTNGQPIVVRGVMKPIPTLYKPLMSVDIETHEPYKATVERSDPTALPA						
	: . ** * * * *	**:******:	**:******:	**:******:	**:******:	**:******:	**:******:
	360	370	380	390	400		
TARGET	AGVVVETMVALVLARAALKFGGDSLAE <del>TQ</del> RNIAAYQRSVADREAPAA <del>RV</del> SG						
TEMPLATE	AGMVMEAVVATVLAQEILEK <del>FSS</del> DNLEELKEAVAKHRDYTKNY-----						
	**:*:*:*:* *	**:*:*:* *	**:*:*:* *	**:*:*:* *	**:*:*:* *	**:*:*:* *	**:*:*:* *
Identity (*): 172 is 42.79 %							
Strongly similar (:): 77 is 19.15 %							
Weakly similar (.): 33 is 8.21 %							
Different: 120 is 29.85 %							

Fig. 2. CLUSTAL W pairwise sequence alignment between the target (MTB CS) and template (IQXO). The amino acid residues of the CS signatures (G-[DES]-S-H-[GC]-x(2)-[LIVM]-[GTIV]-x-[LIVT]-[LIV]-[DEST]-[GH]-x-[PV], [GE]-x(2)-S-[AG]-R-x-[ST]-x(3)-[VT]-x(2)-[GA]-[STAVY]-[LIVMF], R-[SHF]-D-[PSV]-[CSAVT]-x(4)-[SGAIVM]-x-[IVGSTAPM]-[LIVM]-x-E-[STAHNCG]-[LIVMA]) are underlined.

leaving phosphate group of EPSP. (The numbering of histidine residues corresponds to the alignment shown in Fig. 2; HIS10 and HIS110 match HIS17 and HIS106, respectively, in the SPN CS original sequence [9].) Site-directed mutagenesis studies indicated that mutation of HIS110 to an alanine reduces CS activity to 10%; and, in changing HIS10 to an alanine only 5% of the original activity can be observed, demonstrating the importance of these histidines in the CS reaction. These results are the same for both bifunctional and monofunctional CS enzymes [9].

The FMN coenzyme is deeply buried into the active site with EPSP blocking any possible exit. FMN makes one hydrogen bond to EPSP and a few polar interactions with the protein. On the other hand, EPSP makes several polar interactions and a few hydrophobic contacts with the protein [6].

*Mycobacterium tuberculosis* (MTB), the etiological agent of tuberculosis, is responsible for widespread human morbidity and mortality. The development of new effective chemotherapy should aid in the treatment and control of the disease. The World Health Organization (WHO) estimates that there were 8.8 million new cases of tuberculosis in 2002 and 4.4 million new cases in 2003. The global incidence rate of tuberculosis is growing at  $\sim 1.1\%$  per year and the number of cases at 2.4% per year. Through the years 1980–2004, a total of 81 million cases were reported to the WHO [10].

Sequencing of the MTB genome [11] has revealed a large number of individual enzymes potentially useful in drug design, including CS. Understanding the structure of MTB CS, together with its coenzyme and substrate binding modes, should facilitate the search for inhibitors of this enzyme as possible alternative agents to treat tuberculosis.

In this work, we present three-dimensional (3D) structural models for CS from MTB and evaluate their interactions with the substrate EPSP and the coenzyme FMN by geometric docking and energy minimization studies.

## 2. Materials and methods

The starting point of homology modeling is the identification of proteins in the protein data bank (PDB) [12] that are related to the target sequence and then select the templates. In this case, the structure prediction of CS from MTB was based on 3D structures for the homologous SPN CS protein (PDB ID: 1QXO), experimentally determined by X-ray diffraction at 2.0 Å resolution [6]. Blastp [13] was used to search for templates.

The next step is the multiple sequence alignment comparison. The objective of this alignment is to improve the sensitivity of the search and to find regions with high similarity. Possible templates and target sequences alignments were performed with CLUSTAL W [14] and required a small gap of four residues (insertions and/or deletions).

The program MODELLER6v2 [15] was used to build the protein models, using the standard protocol of the comparative protein structure modeling methodology [16]. The best model of each enzyme was evaluated and selected according to their stereochemical quality analyzed with PROCHECK [17]. Val-

idation of the models 3D profiles was performed with VERIFY 3D [18]. CS quaternary structure was built using other known CS structures [6–8,19], by structure fitting with Swiss-PdbViewer v3.7 [20].

The structures of EPSP and FMN were geometrically docked to the CS model of MTB using Swiss-PdbViewer v3.7 [20]. The topology of the CS-FMN-EPSP ternary complex, the FMN and EPSP charges and energy minimization, using the all-atom force field [21], were performed with the Leap, Antechamber and Sander modules, respectively, of AMBER 7 [22]. Hydrogen atom addition to the ligands and their optimization were performed with PRODRG2 server [23].

The ternary complex structure refinement (relaxation and optimization) was performed by two runs of energy minimization, each of 4000 steps, and a cutoff radius of 12 Å for the evaluation of non-bonded interactions, with decreasing force restraints. In the first run the coenzyme and substrate molecules were constrained to their initial positions (called herein constrained model) whilst the enzyme was allowed to move. In the second run the enzyme, coenzyme and substrate were allowed to move by slowly decreasing the restraining force constant (called herein unrestrained model). Both minimized models have the N-terminal amino acid residue missing. All calculations were carried out on a 2.4 GHz Pentium IV PC.

MTB CS structures (geometrically docked model, constrained and unrestrained models) were compared to the template by an evaluation of their root mean-square deviation (RMSD). Protein–ligand interactions were analyzed with LIGPLOT [24]. All protein figures were prepared with the PyMOL v0.98 graphics package [25].

## 3. Results and discussion

### 3.1. Homology modeling

CS is a protein of about 360–400 amino acid residues, except in apicomplexan parasites, where it has about 500 amino acids. It has a high degree of sequence conservation among species. The protein has three signature patterns (Fig. 2) from conserved regions rich in basic residues (mostly arginines) [26].

In the search for templates we found four candidates for modeling, CS of SPN (PDB ID: 1QXO) [6], *Aquifex aeolicus* (PDB ID: 1Q1L) [19], *S. cerevisiae* (PDB ID: 1R53) [7] and *H. pylori* (PDB ID: 1UMF) [8]. The CS from SPN was used as template to model the 3D structure of MTB CS. 1QXO turned to be a very attractive template for its structure contained not only the enzyme, but the coenzyme FMN in the oxidized form and the EPSP substrate. In addition, as in MTB, CS from SPN is monofunctional. Fig. 2 shows a pairwise alignment between the target and template sequences, whose identity is over 40%, well above the 30% limit usually required for comparative protein structure modeling [16].

Ten models of the enzyme were built. They were evaluated by PROCHECK [17] and VERIFY 3D [18] in order to choose the best one. Out of 335 non-glycine and non-proline residues, 318 or 94.9% are located in the most favored regions of the Ramachandran plot, while for the SPN CS template this

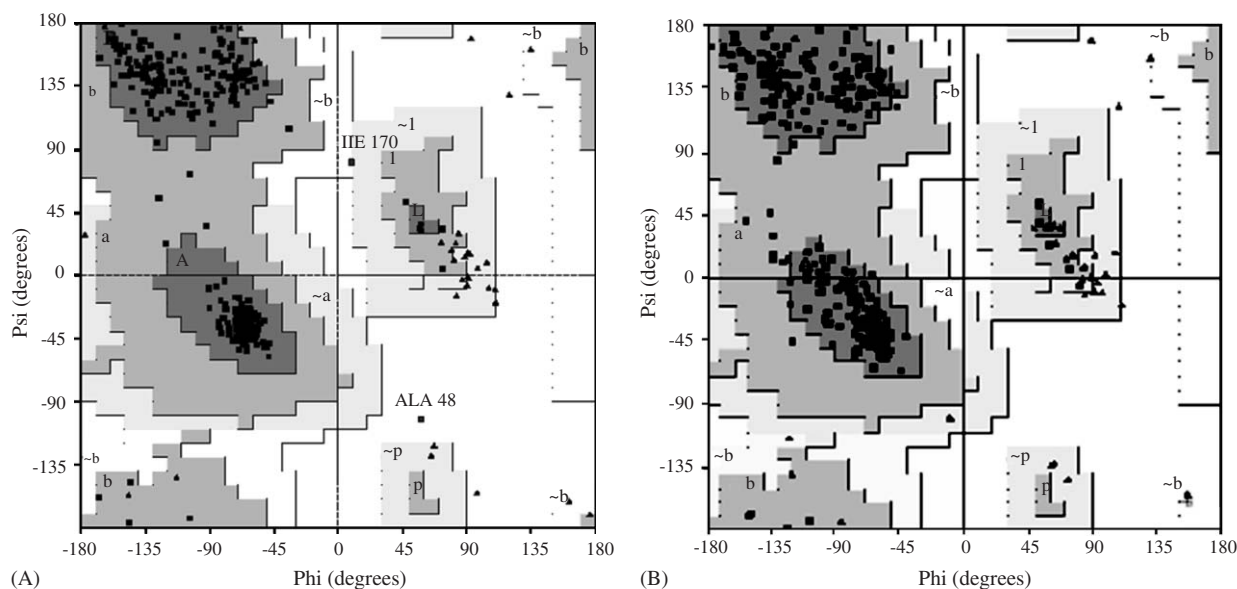


Fig. 3. Ramachandran plot for the best model of MTB CS (A) and for the SPN CS template (B).

Table 1  
Ramachandran plot statistics for the best model of CS MTB and for the template

Ramachandran plot statistics				
	Model		Template	
Residues in most favored regions [A,B,L]	318	94.9%	309	93.1%
Residues in additional allowed regions [a,b,l,p]	15	4.5%	23	6.9%
Residues in generously allowed regions [ $\sim$ a, $\sim$ b, $\sim$ l, $\sim$ p]	1	0.3%	0	0%
Residues in disallowed regions	1	0.3%	0	0%
Number of non-glycine and non-proline residues	335	100.0%	332	100%
Number of end-residues (excl. Gly and Pro)	1		2	
Number of glycine residues (shown as triangles)	43		38	
Number of proline residues	22		15	
Total number of residues	401		387	

value is 93.1% (Figs. 3A and B, and Table 1). The root mean square (RMS) of planar atoms from best-fit plane are less than 0.02 for rings and 0.01 otherwise. The PROCHECK results are given in Table 2. The best model of MTB CS is illustrated in Fig. 4. These results attest to the quality of the MTB CS model as good and appropriate for docking studies.

### 3.2. Docking studies

The coenzyme FMN and the EPSP substrate were geometrically docked into the active site of MTB CS. Based on the assumption that the binding is similar to SPN, we were able to geometrically place FMN and EPSP in their respective binding sites [6] and minimize the complex under two different conditions, thus producing the constrained and unrestrained models, as describe in the methods section. The backbone RMSD between the template, the geometrically docked model, and the

refined models (constrained and unrestrained) were 1.51, 1.57 and 1.58 Å, respectively. Both the refined models and the geometrically docked model were analyzed regarding their interactions with FMN and EPSP.

### 3.3. FMN interactions analysis

FMN interactions were calculated with LIGPLOT [24] for each of the models generated: geometrically docked, constrained and unrestrained, and compared to those observed in SPN CS (Table 3 and Fig. 5B).

**Geometrically docked model:** The interactions observed were similar to SPN, except one missing contact to GLN256 which in SPN corresponds to ASN251. The longer side chains of the non-conserved GLN256 and the conserved LYS315 are too close to FMN and thus need a conformational change to be properly accommodated. However, GLN256 makes an H bond with the FMN phosphate. LYS315 in MTB made an H bond with O3 in the aliphatic portion of FMN which was not present in SPN. All H bonds to FMN in SPN are also present in MTB, including the H bond with EPSP. Although this model presented some extra hydrophobic contacts, ILE255, ALA314 and ALA349, and one missing contact due to amino acid substitution, PRO314 in SPN corresponds to SER317 in MTB. Comparatively, MTB has more hydrophobic contacts (six) than the SPN (four).

**Constrained model:** There was no change in the hydrophobic contacts involving amino acids ALA137, MET313, ILE316, ALA345 and ALA348. Hydrogen bonding to ALA256 and HIS114 as well as to EPSP did not change. LYS314 overlap with FMN was removed while keeping its H bonding intact. THR318 which had only one H bond to FMN, now has two extra H bonds. A new H bond is observed with amino acid GLY113, which had not been seen even in SPN; this residue is not part of CS signature, and is in a coil region at the active

Table 2  
Quality of main-chain and side-chain parameters of modeled MTB CS. The model is verified at 2.0 Å resolution

Comparison values					
Stereochemical parameters	Number of data pts	Parameter value	Typical value	Band width	Number of bandwidths from mean
<i>Stereochemistry of main-chain</i>					
%-tage residues in A, B, L	335	94.9	83.8	10.0	1.1
Omega angle st. dev.	400	3.8	6.0	3.0	−0.7
Bad contacts 100 residues	3	0.7	4.2	10.0	−0.3
Zeta angle st. dev.	358	1.3	3.1	1.6	−1.1
H-bond energy st. dev.	238	0.7	0.8	0.2	−0.5
Overall G-factor	401	0.0	−0.4	0.3	1.2
<i>Stereochemistry of side-chain</i>					
Chi-1 gauche minus st. dev.	56	5.1	18.1	6.5	−2.0
Chi-1 trans st. dev.	89	9.1	19.0	5.3	−1.9
Chi-1 gauche plus st. dev.	121	6.2	17.5	4.9	−2.3
Chi-1 pooled st. dev.	266	7.0	18.2	4.8	−2.3
Chi-2 trans st. dev.	72	11.8	20.4	5.0	−1.7

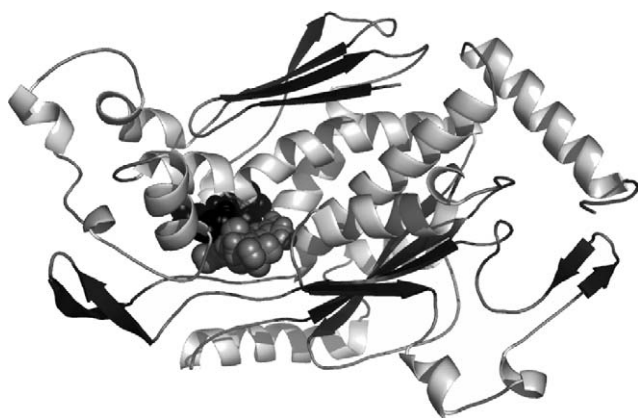


Fig. 4. Ribbon representation of MTB CS 3D structure. The structure contains 9  $\alpha$ -helices and 15  $\beta$ -strands in a  $\beta$ - $\alpha$ - $\beta$  secondary structure arrangement, an uncommon architecture. The  $\alpha$ -helices are shown in light gray and the  $\beta$  strands in black. Ligands are represented as spheres; EPSP in black and FMN in gray.

site. GLN255 no longer overlaps with FMN, but lost its only H bond. Indeed, now this residue makes only a hydrophobic contact with FMN.

*Unrestrained model:* The best model for CS–FMN interactions was the unrestrained, which reproduces very well the interactions observed at SPN CS. All H bonds previously observed are present in this model, except for changes at LYS314, with an extra H bond, and for a loss of one of the THR318 H bonds. The hydrophobic contact at ALA348 observed in the constrained model is lost, however it was also not observed in SPN (Fig. 5A).

### 3.4. EPSP interactions analysis

Substrate–protein interactions were calculated with LIGPLOT [24] for each of the models generated: geometrically docked, constrained and unrestrained, and compared to those observed in SPN CS, as done for FMN (Table 4 and Fig. 6B).

Table 3  
Protein–coenzyme (CS–FMN) interactions observed in each of the MTB models and SPN

Amino acids	SPN	Geometrically docked	Constrained	Unrestrained
EPSP	4001	403	402	402
GLY	—	—	113	113
HIS	110	115	114	114
ALA	133	138	137	137
ILE	—	255	—	—
GLN/ASN	ASN251 (2)	GLN256 <sup>a</sup>	GLN255	GLN255
ALA	252	257	256	256
MET	—	314	313	313
LYS	311	315 <sup>a</sup>	314	314 (2)
ILE	313	317	316	316
SER/PRO	PRO314	—	SER317	SER317
THR	315 (2)	319	318 (3)	318 (2)
ALA	342	346	345	345
ALA	—	349	348	—
Number of interactions	7/4	6/6 <sup>a</sup>	9/6	9/5

The first column indicates the ligand and amino acids involved in the interactions while the others indicate amino acid position. Residues GLN/ASN and SER/PRO are the only differences observed in the active site. Residues involved in H bond interactions and hydrophobic contacts are in italics and bold, respectively. Numbers in brackets correspond to the number of interactions made by the residue when different from 1.

<sup>a</sup>Residues overlapped with FMN atoms.

*Geometrically docked model:* The H bonds to EPSP in MTB CS are also very similar to those found in SPN. Three H bonds, involving amino acids HIS11, ARG49 and ARG139 were missing (HIS10, ARG48 and ARG134 in SPN), although ARG49 now makes a hydrophobic contact to EPSP. Furthermore, while ARG45 makes three H bonds to EPSP in SPN, in MTB ARG46 makes only one H bond. ARG40 changed from two to three H bonds, and ALA138 interaction remains the same. No hydrophobic contacts were observed in SPN while there are two in MTB, ARG49 (H bond interaction in SPN) and MET50.



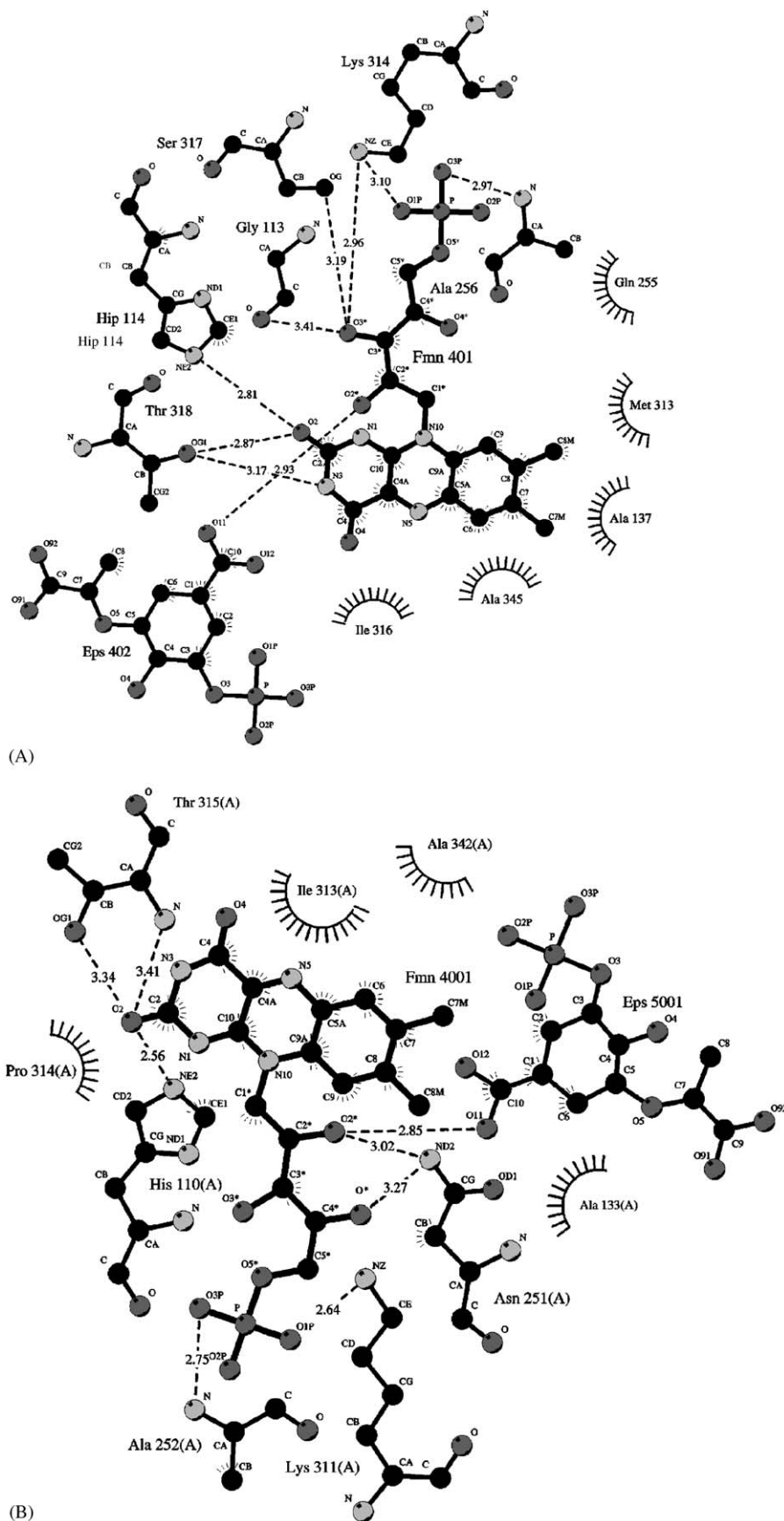


Fig. 5. LIGPLOT representation of FMN binding pattern to MTB CS (A) and for the SPN CS template (B). Dotted lines represent the hydrogen bonds and the half circles the hydrophobic contacts.

Table 4  
Protein–substrate (CS–EPSP) interactions observed in each of the MTB models and SPN

Amino acids	SPN	Geometrically docked	Constrained	Unrestrained
FMN	5001	402	401	401
HIS	10	—	—	—
ARG	39 (2)	40 (3)	39 (3)	39 (2)
ARG	45 (3)	46	45	45
ARG	48	<b>49</b>	—	—
MET	—	<b>50</b>	—	—
ARG	—	—	—	111
SER	—	—	—	<b>136</b>
ALA	133	138	137	137
ARG	134	—	—	—
Number of interactions	9/0	6/2	6/0	6/1

The first column indicates the ligand and amino acids involved in the interactions while the others indicate amino acid position. Residues involved in H bonds interactions and hydrophobic contacts are in italic and bold, respectively. Numbers in brackets correspond to the number of interactions made by the residue when different from 1.

*Constrained model:* All H bonds to amino acids ARG39, ARG45 and ALA137 were preserved. However, two hydrophobic contacts observed previously were missing.

*Unrestrained model:* The best model for CS–EPSP interactions was again the unrestrained; which reproduces very well the interactions observed in the SPN CS template. Differently from the constrained model, there are only two H bonds to ARG39, as observed in SPN CS. H bonds to ARG45 and ALA137 are maintained while a new H bond interaction is observed to the amino acid ARG111. This interaction was not seen in SPN CS; this residue is not part of the signature region as seen before for the FMN interaction to GLY113, which is also in a coil region at the active site. In addition, one new hydrophobic contact was observed at SER136, a coil residue in active site and part of the second signature region (Fig. 6).

Evaluating structural differences of FMN and EPSP of each model with respect to the crystal structure, it can be observed that energy minimization of FMN did not alter its structure significantly; the ring portion of the unrestrained model is only slightly different. The EPSP structure, however, shows conformational changes after energy minimization. Its phosphate portion moved 1.2 Å away from the crystallographic structure, which could explain the missing interactions to HIS11.

### 3.5. Quaternary structure

The quaternary structure of MTB CS (Fig. 7) was built using other known CS structures by structure fitting with Swiss-PdbViewer v3.7 [20]. As expected, the major subunit contacts that are involved in structure stabilization are very similar to those observed in SPN CS. The main feature of CS subunit interface stabilization is a  $\beta$  sandwich formed by adjacent antiparallel  $\beta$  sheets from two adjacent subunits

(Fig. 7I). Another key element that has correspondence in the MTB model is the interface formed by the extension of  $\beta$  sheets from each monomer, bridging them through hydrogen bonds [6] (Fig. 7II).

## 4. Conclusions

We have obtained a 3D structural model of MTB CS based on the crystal structure of an orthologous enzyme from SPN.

In addition, we modeled the interactions of the coenzyme FMN and the EPSP substrate with the enzyme using a simple, geometric docking approach. After docking studies we performed refinement with energy minimization using two different protocols. The minimized structures showed improvements when compared to the geometrically docked one; amino acids that were overlapping with FMN were then better accommodated in the active site, thus rescuing an interaction between SER317 (PRO314 in SPN CS) and FMN. Another residue exchange occurs at GLN255 (ASN251 in SPN CS). Although two hydrogen bonds are lost, the overall number of H bonds is greater than in the SPN CS.

The major difference in protein–substrate, CS–EPSP, interactions was the loss of HIS11 H bond, an interaction not seen in any of the models. This could be due to the fact that HIS11 is involved in neutralizing the departing phosphate group from ESPPS by proton donation one step after FMN binding [6]. The FMN used in our studies is in the oxidized form, different from the redox state necessary for the reaction to take place. This may be the reason why the interaction was not observed here. Another possibility is that the phosphate portion to which HIS11 H bonds is very different from its crystallographic conformation (1.2 Å).

Another important interaction that has been lost, that of ARG139 to the pyruvyl portion, could also be a consequence of conformational changes to ESPPS after refinement. The pyruvyl portion where this interaction takes place moved during minimization (1.7 and 1.4 Å for the constrained and unrestrained models), thus placing its atoms far from the active site ARG139.

Despite the differences reported here, MTB CS interactions with its coenzyme and substrate, best revealed by the unrestrained model, reproduces well the interaction patterns observed in the SPN CS template. This should be expected, to a certain degree, for CS have very similar physico-chemical properties (3) and, additionally, in the study reported here, both enzymes are monofunctional. Hence, we do not expect the activity or the catalytic cycle of MTB CS to be very different from that of SPN CS.

Future work will consider molecular dynamics and docking simulations with the reduced FMN to further investigate MTB CS–FMN–EPSP interactions so as to improve our understanding about these interactions.

Understanding the structure of MTB CS, together with its coenzyme and substrate binding modes, should provide ligand, as well as receptor-based pharmacophoric models to be used in high throughput virtual screening of small-molecule public libraries in order to accelerate the search for inhibitors of this enzyme as alternative agents to treat tuberculosis.

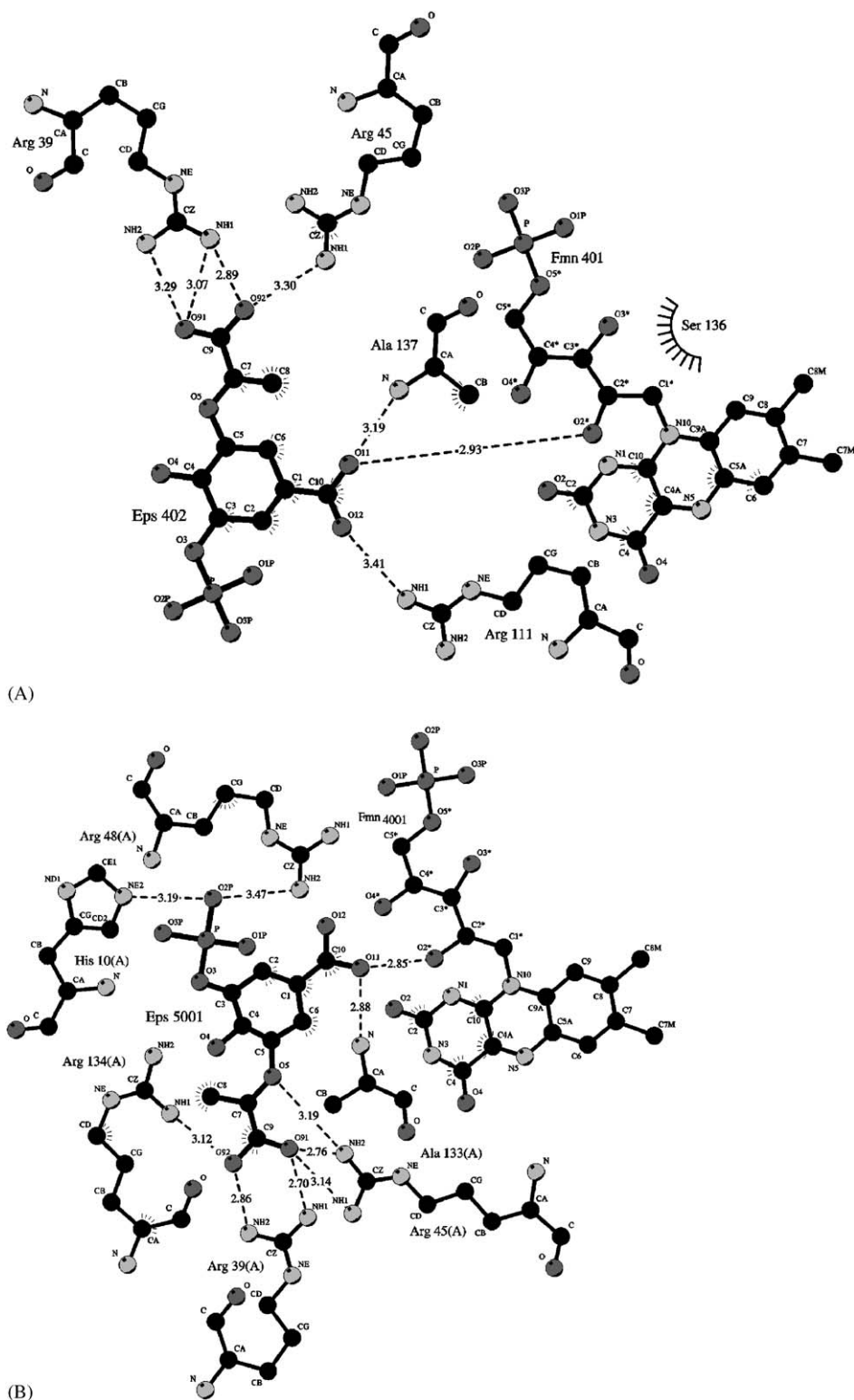


Fig. 6. LIGPLOT representation of EPSP (Eps) binding pattern to MTB CS (A) and for the SPN CS template (B). Dotted lines represent the hydrogen bonds and the half circles the hydrophobic contacts.

## 5. Summary

The enzymes of the shikimate pathway constitute an excellent target for the design of new antibacterial agents. This path-

way is found in bacteria, fungi, plants and apicomplexan parasites but is absent in mammals. CS catalyzes the last step of this pathway, the product of which is utilized in other enzymatic transformations like the biosynthesis of aromatic amino





Fig. 7. The tetrameric conformation of MTB CS quaternary structure represented as ribbons. The arrow is pointing towards the FMN coenzyme and the EPSP substrate, both represented as spheres, in the active site of each subunit. I: the  $\beta$  sandwich interactions. II: the  $\beta$  sheet that bridges two subunits, fundamental to holding them together.

acids, folate, vitamin K and ubiquinone. This reaction is the most unusual of the entire pathway and is unique in nature. It converts EPSP to chorismate in the presence of a reduced FMN coenzyme. Structure prediction used the comparative protein structure modeling methodology. The 3D structure prediction of the enzyme was performed using the crystal structure (PDB ID: 1QX0) of CS from SPN as template ( $\approx 42\%$  identity), and the MODELLER6v2 package. Additionally, in order to understand the possible binding modes of coenzyme and substrate to the enzyme, FMN and EPSP, respectively, were geometrically docked to CS. Three different MTB CS models were produced and analyzed. The starting homology model indicated that the quality of the MTB CS structure is good enough and appropriate for docking studies. Despite its quality this initial model had some overlapping atoms between CS (GLN256 and LYS315) and FMN, hence further energy minimizations were performed to allow ligand and enzyme to relax in order to obtain a binding pattern as close as possible to that of SPN CS. Therefore, two minimized models were generated under different restraining (constrained and unrestrained) conditions, both showing structural improvements, such as amino acids rearrangements, that removed the overlapping. The unrestrained model turned to be best one, since it not only reproduced in MTB CS the binding pattern of FMN and EPSP observed in the template SPN CS, but also revealed novel interactions in the MTB ternary complex. This knowledge should facilitate the search for inhibitors of this enzyme as alternative agents to treat tuberculosis.

## Acknowledgments

This project was supported by grants from CAPES and FAPERGS to O.N.S. and FINEP and Millennium Institute (CNPq/MCT) to D.S.S. and L.A.B.; C.L.F. and A.B., we are supported by MSc scholarships from CAPES. We also thank the referees for their valuable suggestions.

## References

- [1] R. Bentley, The shikimate pathway—a metabolic tree with many branches, *Crit. Rev. Biochem. Mol. Biol.* 25 (5) (1990) 307–384.
- [2] F. Roberts, C.W. Roberts, J.J. Johnson, D.E. Kyle, T. Krell, J. Coggins, G.H. Coombs, W.K. Milhous, S. Tzipori, D.J.P. Ferguson, D. Chakrabarti, R. McLeod, Evidence for the shikimate pathway in apicomplexan parasites, *Nature* 393 (1998) 801–805.
- [3] P. Macheroux, J. Schmid, N. Amrhein, A. Schaller, A unique reaction in a common pathway: mechanism and function of chorismate synthase in the shikimate pathway, *Planta* 207 (1999) 325–334.
- [4] P. Macheroux, S. Bornemann, S. Ghisla, R.N.F. Thorneley, Studies with flavin analogs provide evidence that a protonated reduced FMN is the substrate-induced transient intermediate in the reaction of *Escherichia coli* chorismate synthase, *J. Biol. Chem.* 271 (42) (1996) 25850–25858.
- [5] S. Bornemann, D.J. Lowe, R.N.F. Thorneley, The transient kinetics of *Escherichia coli* chorismate synthase: substrate consumption, product formation, phosphate dissociation, and characterization of a flavin intermediate, *Biochemistry* 35 (30) (1996) 9907–9916.
- [6] J. MacClean, S. Ali, The structure of chorismate synthase reveals a novel flavin binding site fundamental to a unique chemical reaction, *Structure* 11 (2003) 1499–1511.
- [7] S. Quevillon-Cheruel, N. Leulliot, P. Meyer, M. Graille, M. Bremang, K. Blondeau, I. Sorel, A. Poupon, J. Janin, H. van Tilbeurgh, Crystal structure of the bifunctional chorismate synthase from *Saccharomyces cerevisiae*, *J. Biol. Chem.* 279 (1) (2004) 619–625.
- [8] H.J. Ahn, H.-J. Yoon, B.I. Lee, S.W. Suh, Crystal structure of chorismate synthase: a novel FMN-binding protein fold and functional insights, *J. Mol. Biol.* 336 (2004) 903–915.
- [9] K. Kitzing, S. Auweter, N. Amrhein, P. Macheroux, Mechanism of chorismate synthase. Role of the two invariant histidine residues in the active site, *J. Biol. Chem.* 10 (2004) 9451–9461.
- [10] World Health Organization, USA, Global Tuberculosis Control Report 2004, (<http://www.who.int/tb/publications/2005/en/index.html>).
- [11] S.T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S.V. Gordon, K. Eiglmeier, S. Gas, C.E. Barry, 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, Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence, *Nature* 393 (1998) 537–544.
- [12] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, *Nucleic Acids Res.* 28 (2000) 235–242.
- [13] 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.
- [14] 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 (22) (1994) 4673–4680.
- [15] A. Šali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, *J. Mol. Biol.* 234 (3) (1993) 779–815.

- [16] M.A. Martí-Renom, A.C. Stuart, A. Fiser, R. Sánchez, F. Melo, A. Šali, Comparative protein structure modeling of genes and genomes, *Annu. Rev. Biophys. Biomol. Struct.* 29 (2000) 291–325.
- [17] 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.
- [18] R. Lüthy, J.U. Bowie, D. Eisenberg, Assessment of protein models with three-dimensional profiles, *Nature* 356 (1992) 83–85.
- [19] C.M. Viola, V. Saridakis, D. Christendat, Crystal structure of chorismate synthase from *aquifex aeolicus* reveals a novel beta alpha beta sandwich topology, *Proteins* 54 (1) (2004) 166–169.
- [20] N. Guex, M.C. Peitsch, SWISS-MODEL and The Swiss-PdbViewer: an environment for comparative protein modeling, *Electrophoresis* 18 (1997) 2714–2723.
- [21] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz Jr., D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, A second generation force field for the simulation of proteins. Nucleic acids, and organic molecules, *J. Am. Chem. Soc.* 117 (1995) 5179–5197.
- [22] D.A. Case, D.A. Pearlman, J.W. Caldwell, T.E. Cheatham III, J. Wang, W.S. Ross, C.L. Simmerling, T.A. Darden, K.M. Merz, R.V. Stanton, A.L. Cheng, J.J. Vincent, M. Crowley, V. Tsui, H. Gohlke, R.J. Radmer, Y. Duan, J. Pitera, I. Massova, G.L. Seibel, U.C. Singh, P.K. Weiner, P.A. Kollman, AMBER 7, University of California, San Francisco, 2002.
- [23] A.W. Schuettelkopf, D.M.F. van Aalten, PRODRG—a tool for high-throughput crystallography of protein–ligand complexes, *Acta Crystallogr. D60* (2004) 1355–1363.
- [24] A.C. Wallace, R.A. Laskowski, J.M. Thornton, LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions, *Protein Eng. Des. Sel.* 8 (1995) 127–134.
- [25] W.L. DeLano, The PyMOL Molecular Graphics System, DeLano Scientific, San Carlos CA USA, 2002 (<http://www.pymol.org>).
- [26] N. Hulo, C.J.A. Sigrist, V. Le Saux, P.S. Langendijk-Genevaux, L. Bordoli, A. Gattiker, E. De Castro, P. Bucher, A. Bairoch, Recent improvements to the PROSITE database, *Nucleic Acids Res.* 32 (2004) D134–D137.

## **ARTIGO 2**

**STRUCTURE PREDICTION AND DOCKING STUDIES OF CHORISMATE SYNTHASE FROM  
*MYCOBACTERIUM TUBERCULOSIS.***

FERNANDES, C.L.; SANTOS, D.S.; BASSO, L.A.; NORBERTO DE SOUZA, O.

***LECTURES NOTES IN COMPUTER SCIENCE, 3594:118-127, 2005.***

# Structure Prediction and Docking Studies of Chorismate Synthase from *Mycobacterium Tuberculosis*

Cláudia Lemelle Fernandes<sup>1</sup>, Diógenes Santiago Santos<sup>2,3</sup>,  
Luiz Augusto Basso<sup>2,4</sup>, and Osmar Norberto de Souza<sup>1,2,3</sup>

<sup>1</sup> Laboratório de Bioinformática,  
Modelagem e Simulação de Biosistemas - LABIO  
Pontifícia Universidade Católica do Rio Grande do Sul

<sup>2</sup> Centro de Pesquisa em Biologia Molecular e Funcional, TECNOPUC, PUCRS

<sup>3</sup> Instituto de Pesquisas Biomédicas, PUCRS

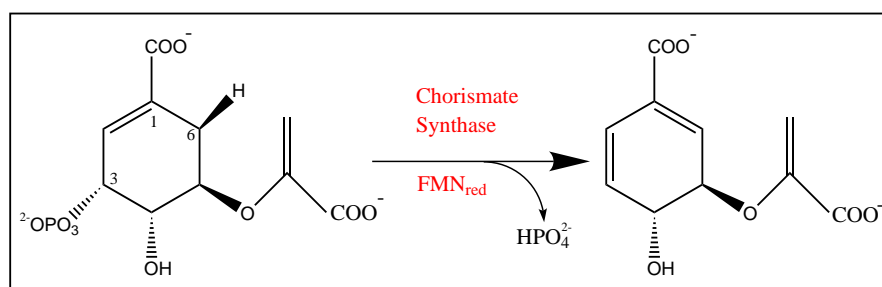
<sup>4</sup> Dep. de Biologia Molecular e Biotecnologia, UFRGS,  
Porto Alegre, RS, Brasil  
osmarns@inf.pucrs.br

**Abstract.** The enzymes of the shikimate pathway constitute an excellent target for the design of new antibacterial agents. This pathway is found in bacteria, fungi, plants and apicomplexan parasites but is absent in mammals. Chorismate Synthase (CS) catalyzes the last step of this pathway, the product of which is utilized in other enzymatic transformations like the biosynthesis of aromatic amino acids, folate, vitamin K and ubiquinone. This reaction is the most unusual of the entire pathway and is unique in nature. It converts EPSP to chorismate in the presence of a reduced FMN cofactor. Structure prediction used the comparative protein structure modeling methodology. The three-dimensional (3D) structure prediction of the enzyme was performed using the crystal structure (PDB ID: 1QX0) of CS from *Streptococcus pneumoniae* as template ( $\approx 42\%$  identity), and the MODELLER6v2 package. Additionally, in order to understand the possible binding modes of substrate and cofactor to the enzyme EPSP and FMN, respectively, were geometrically docked to CS. FMN binding to CS of *M. tuberculosis* (MTB) is similar to that of the *S. pneumoniae* template despite the change of Asn251 in *S. pneumoniae* to Gln256 in MTB. The longer side chain of Gln256 is overlapping with the FMN cofactor and a small conformational change is needed in order to properly accommodate this interaction. EPSP binding mode is also very similar to that of the template with three hydrogen bonds missing. This could be due to artifacts from the simple geometric docking we performed. Refinement with energy-based docking algorithms should relax the enzyme and substrates, thus promoting the expected interactions between them. Understanding the structure of MTB CS together with its cofactor and substrate binding modes should facilitate the search for inhibitors of this enzyme as alternative agents to treat tuberculosis.

## 1 Introduction

The shikimate pathway is the common way for the production of various products including folic acid, vitamin K, ubiquinone and the three aromatic amino acids. In bacteria, fungi, plants and apicomplexan parasites, chorismate, the final product of the shikimate pathway, is the branch point in the biosynthesis for all these products that are essential for these species. The absence of the shikimate pathway in all other species makes it an attractive target for the development of new antibacterial agents [3, 16].

Chorismate Synthase, the seventh and final step of the shikimate pathway, catalyses the conversion of 5-enolpyruvylshikimate 3-phosphate (EPSP) to chorismate in the presence of a reduced flavin mononucleotide (FMN) as a cofactor [12]. The reaction mechanism of the shikimate pathway has been studied extensively and revealed that the reaction of CS is unique in nature. The reaction involves a 1,4 elimination of phosphate and the loss of a proton of the C-6 hydrogen. This consists in the formation of the second out of three necessary double bonds to build an aromatic ring (Fig. 1). The enzyme activity requires a reduced FMN molecule which is not consumed during the reaction [3, 11].



**Fig. 1.** Reaction catalyzed by Chorismate Synthase. The elimination of the 3-phosphate and the loss of a proton in C-6 introducing a second double bound in the ring

The function of the reduced FMN in catalysis was extensively studied. The most accepted mechanism suggests a direct role of reduced FMN in the elimination reaction. FMN transfer the electron transiently to phosphate and the substrate donates an electron to regenerate the FMN. This reaction does not involve an overall change in the redox state [5, 12].

Recently, with the first high-resolution X-ray structure of CS from *S. pneumoniae* with the substrate and the cofactor in the oxidized form, [13] the structure of CS from *Saccharomyces cerevisiae* [15], and the structure of CS from *Helicobacter pylori* with the cofactor in the reduced form [1], studies on the binding mode of substrate and cofactor in the active site has started.

How reduced FMN is obtained divides the CS into two classes, monofunctional and bifunctional. Bifunctional CS has an intrinsic ability to reduce flavin (specifically FMN) using NADPH. In monofunctional CS this catalytic activity is

not present. The bifunctional enzyme is present in fungi and the monofunctional form in plants and bacteria [12].

The active site of *S. pneumoniae* CS is very hydrophobic and extremely basic, with six arginine and two histidine residues. The FMN cofactor is deeply buried into the active site with EPSP blocking any possible exit. FMN makes one hydrogen bond with EPSP and a few polar interactions with the protein. On the other hand, EPSP makes several polar interactions and a few hydrophobic contacts with the protein [13]. The two histidines are present in both classes of CS and mutation of these residues to two alanines reduces the activity of both bifunctional and monofunctional enzymes to 5% [8].

*M. tuberculosis*, the etiological agent of tuberculosis, is responsible for widespread human morbidity and mortality. The development of new effective chemotherapy should aid in the treatment and control of the disease [21].

Sequencing of the MTB genome has revealed a large number of individual enzymes potentially useful in drug design [6], including CS. Understanding the structure of MTB CS, together with its cofactor and substrate binding modes, should facilitate the search for inhibitors of this enzyme as possible alternative agents to treat tuberculosis.

In this work we present 3D structural models for CS from MTB and evaluate their interactions with the substrate EPSP and the cofactor FMN by docking simulations.

## 2 Materials and Methods

The starting point of homology modelling is the identification of proteins in the Protein Data Bank (PDB) [4] that are related to the target sequence and then select the templates. In this case, the structure prediction of CS from MTB was based on 3D structures for the homologous *S. pneumoniae* CS protein (PDB ID: 1QXO) [13] found using Blastp [2].

The next step is the multiple sequence alignment comparisons. The objective of this alignment is to improve the sensitivity of the search and to find the regions with high similarity. Possible templates and target sequences alignments were performed with ClustalW [18] and required a small gap, of four residues (insertions and/or deletions).

The program MODELLER6v2 [19] was used to build the protein models, using the standard protocol for comparative protein structure modelling methodology [14]. The best model of each enzyme was evaluated and selected according to their stereochemical quality analyzed with PROCHECK [9]. Validation of the models 3D profiles was performed with VERIFY 3D [10].

The structures of EPSP and FMN were geometrically and manually docked to the CS model of MTB using SwisPdbViewer [7].

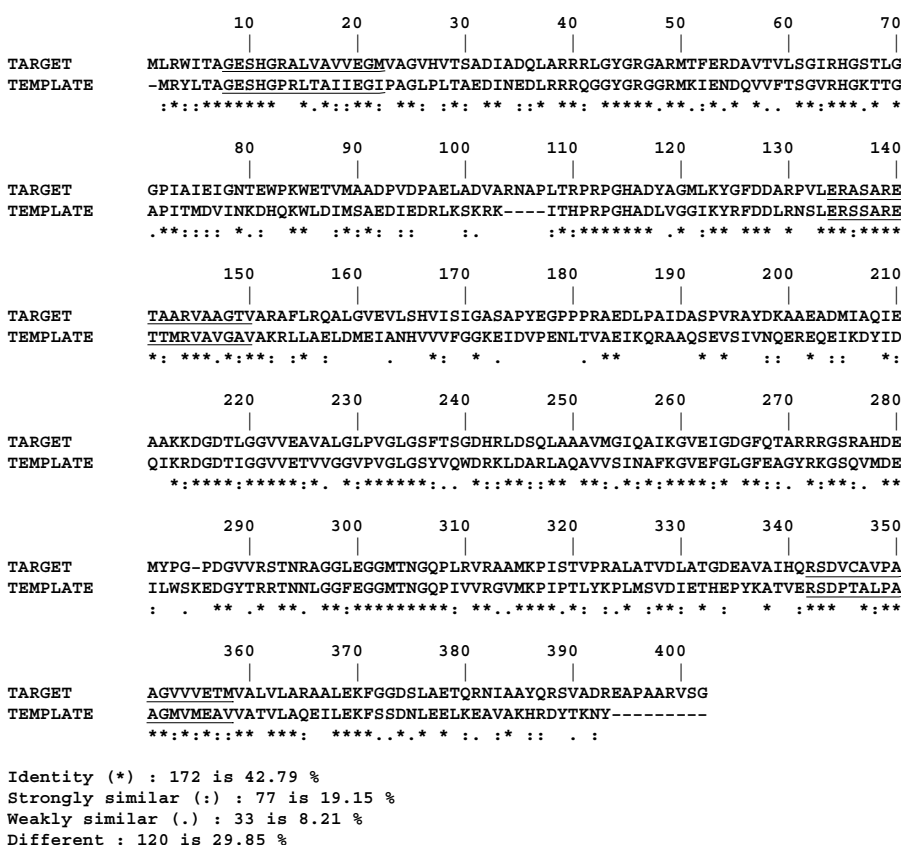
The schematic diagrams of protein-ligand interactions of the best docking results were performed with LIGPLOT [20]. All figures were prepared with SwisPdbViewer [7].

### 3 Results

#### 3.1 Homology Modelling

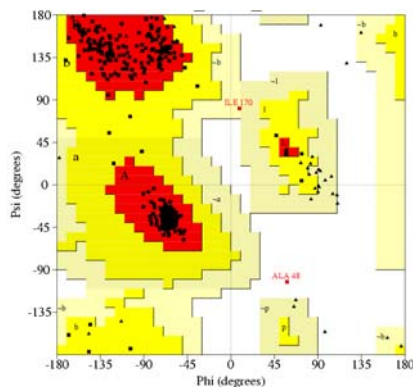
Corismate Synthase is a protein of about 360 to 400 amino acid residues, except in apicomplexan parasites, where it has about 500 amino acids. It has a high degree of sequence conservation among species. The protein has three signature patterns (Fig. 2) from conserved regions rich in basic residues (mostly arginines) [17].

In the search for templates, CS from *Aquifex aeolicus* (PDB ID: 1Q1L) had the best score. However, some residues were missing from the crystal structure which then had to be discarded. Therefore, the second highest score structure,



**Fig. 2.** ClustalW pairwise sequence alignment between the target (*M. tuberculosis* CS) and template (1QXO). The amino acid residues of the CS signatures (G-[DES]-S-H-[GC]-x(2)-[LIVM]-[GTIV]-x-[LIVT]-[LIV]-[DEST]-[GH]-x-[PV], [GE]-x(2)-S-[AG]-R-x-[ST]-x(3)-[VT]-x(2)-[GA]-[STAVY]-[LIVMF], R-[SHF]-D-[PSV]-[CSAVT]-x(4)-[SGAIVM]-x-[IVGSTAPM]-[LIVM]-x-E-[STAHNCG]-[LIVMA]) are underlined





Plot Results		
Residues in most favoured regions (red)	318	94.9%
Residues in additional allowed regions (dark yellow)	15	4.5%
Residues in generously allowed regions (light yellow)	1	0.3%
Residues in disallowed regions (white)	1	0.3%
<hr/>		
Number of non-glycine and non-proline residues	335	100.0%
Number of end-residues (excl. Gly and Pro)	1	
Number of glycine residues (shown as triangles)	43	
Number of proline residues	22	
<hr/>		
Total number of residues	401	

**Fig. 3.** Ramachandran plot for the best model of MTB CS

CS from *S. pneumoniae* (PDB ID: 1QX0), was used as template to model the 3D structure of MTB CS. 1QX0 turned to be a very attractive template for its structure contained not only the enzyme, but the cofactor FMN in the oxidized form and the EPSP substrate. In addition, as in MTB, CS from *S. pneumoniae* is monofunctional.

Fig. 2 shows a pairwise alignment between the target and template sequences, whose identity is over 40%, well above the 30% limit usually required for comparative protein structure modelling [14].

Ten models of the enzyme were built. They were evaluated by PROCHECK [9] and VERIFY 3D [10] in order to choose the best one. Out of 335 non-glycine and non-proline residues, 318 or 94,9%, were located in the most favored regions of the Ramachandran plot (Fig. 3). The root mean-square (RMS) of planar atoms from best-fit plane are less than 0.02 for rings and 0.01 otherwise. The PROCHECK results are given in Table 1. The best model of MTB CS is illustrated in Fig. 4.

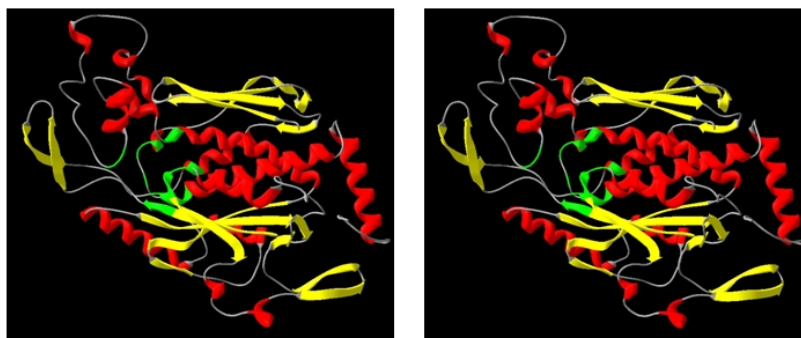
### 3.2 Docking Studies

The cofactor FMN and the EPSP substrate were geometrically and manually docked into the active site of MTB CS. Based on the assumption that the binding is similar to *S. pneumoniae*, we were able to manually place FMN and EPSP in their respective binding sites [13].



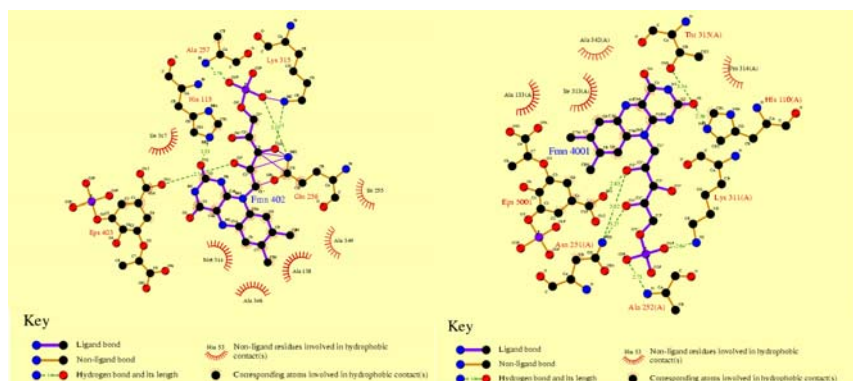
**Table 1.** Quality of main-chain and side-chain parameters of modelled MTB CS. The model is verified at 2 Å resolution

Comparison Values					
Stereochemical parameters	Number of data pts	Parameter value	Typical value	Band width	Number of bandwidths from mean
<b>Stereochemistry of main-chain</b>					
%-tage residues in A, B, L	335	94.9	83.8	10.0	1.1
Omega angle st dev	400	3.8	6.0	3.0	-0.7
Bad contacts 100 residues	3	0.7	4.2	10.0	-0.3
Zeta angle st dev	358	1.3	3.1	1.6	-1.1
H-bond energy st dev	238	0.7	0.8	0.2	-0.5
Overall G-factor	401	0.0	-0.4	0.3	1.2
<b>Stereochemistry of side-chain</b>					
Chi-1 gauche minus st dev	56	5.1	18.1	6.5	-2.0
Chi-1 trans st dev	89	9.1	19.0	5.3	-1.9
Chi-1 gauche plus st dev	121	6.2	17.5	4.9	-2.3
Chi-1 pooled st dev	266	7.0	18.2	4.8	-2.3
Chi-2 trans st dev	72	11.8	20.4	5.0	-1.7

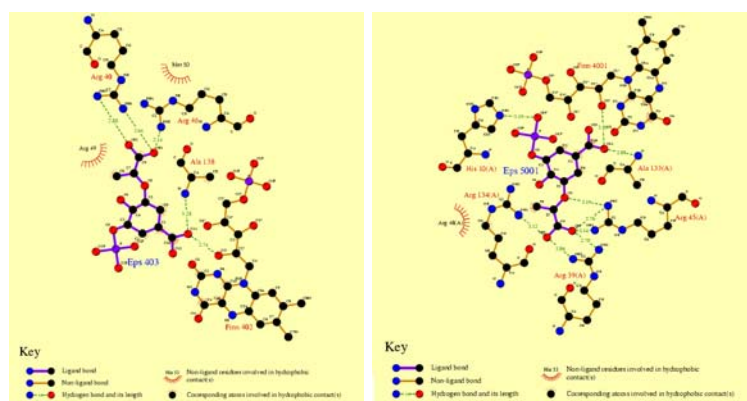
**Fig. 4.** Stereoview of *M. tuberculosis* CS three-dimensional structure (ribbon representation) looking across the active site. The structure contains 9  $\alpha$ -helices (red) and 15  $\beta$ -strands (yellow). The secondary structures making up the active site are colored green

Thus, we have similar interactions except one missing contact to Gln256 which in *S. pneumoniae* corresponds to Asn251. The longer side chains of the non-conserved Gln256 and the conserved Lys315 are too close to FMN and thus need a conformational change to be properly accommodated.

All H bonds to FMN in *S. pneumoniae*, with exception of that involving Thr315, are present in MTB, including the H bond with EPSP. The H bonds to FMN involving Asn251 in *S. pneumoniae* were not reproduced in the MTB model for its structural equivalent in MTB, Gln256, is far too close to the FMN molecule. However, Gln256 makes a H bond with the FMN phosphate. Lys315 in



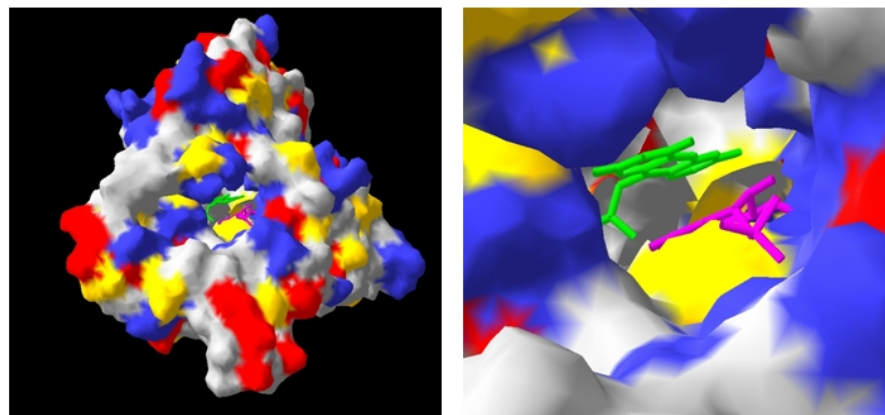
**Fig. 5.** Ligplot representation of the FMN binding pattern to CS from MTB (Left, Fmn 402) and *S. pneumoniae* (Right, Fmn 4001)



**Fig. 6.** Ligplot representation of the binding pattern of EPSP to MTB CS (Left, Eps 403) and *S. pneumoniae* (Right, Eps 5001)

*M. tuberculosis* made a H bond with O3 in the aliphatic portion of FMN which was not present in *S. pneumoniae*. MTB has more hydrophobic contacts (six) than the *S. pneumoniae* (four) (Fig. 5).

The H bonds to EPSP in MTB CS are also very similar to those found in *S. pneumoniae*. Two H bonds, involving amino acids His11 and Arg139, were missing. They correspond to His10 and Arg134, respectively, in the *S. pneumoniae*. Furthermore, while Arg45 makes two H bonds to EPS in *S. pneumoniae*, in MTB, its equivalent, Arg46, makes only one H bond. The hydrophobic contacts were very similar, only one in the *S. pneumoniae* (Arg48) and two in MTB, Arg49 and Met50 (Fig. 6).



**Fig. 7.** Molecular surface representation of MTB CS colored by amino acid type. Apolar, gray; polar, yellow; acidic or negatively charged, red; and basic or positively charged, blue. (Left) The enzyme subunit looking into the active site. (Right) A close-up view of the active site entrance. The binding site entrance is composed mainly of arginines (blue). The cofactor FMN (green) and substrate EPSP (pink) are bound deep inside

The molecular surface of MTB CS (Fig. 7) shows that the active site entrance is very hydrophilic with many basic amino acids that are involved in EPSP binding.

#### 4 Discussion and Conclusions

We have obtained the 3D structure of MTB CS based on the crystal structure of an orthologous enzyme from *S. pneumoniae*. In addition, we modelled the interactions of the cofactor FMN and the EPSP substrate with the enzyme using a simple, geometric, docking approach.

The quality of the MTB CS model is good and appropriate for docking studies.

The geometric docking we used is adequate for an initial study only. As observed, the side chains of Gln256 and Lys315 need to undergo some conformational changes to better accommodate the FMN cofactor in its binding site. Nonetheless, our docking analysis showed that the binding mode of EPSP and FMN is similar in both *S. pneumoniae* and MTB CS, as we should expect based on sequence homology.

The hydrophilic amino acids making up the binding site of EPSP are conserved, but the interactions between enzyme and substrate need some additional studies. His11 and Arg139 are important for enzyme activity, but the H bonds their side chains make to the EPSP substrate are missing in MTB. Further docking refinements with energy-based docking algorithms should relax

the enzyme and substrates, hence promoting the expected interactions between them.

Understanding the structure of *M. tuberculosis* Chorismate Synthase, together with its cofactor and substrate binding modes, should provide a working model to be used in high throughput virtual screening of small-molecule public libraries so as to accelerate the search for inhibitors of this enzyme as alternative agents to treat tuberculosis.

## Acknowledgments

The authors wish to thank Afonso Sales for valuable technical support.

We are grateful to the referees for careful reading of our manuscript and valuable suggestions.

This project was supported by grants from CAPES and FAPERGS to O.N.S., and FINEP and Millennium Institute, CNPq/MCT, to D.S.S. and L.A.B.

C.L.F. is supported by a MSc scholarship from CAPES.

## References

1. Hyung Jun Ahn, Hye-Jin Yoon, Byung II Lee, and Se Won Suh. Crystal Structure of Chorismate Synthase: A novel FMN-binding Protein Fold and Functional Insights. *Journal of Molecular Biology*, 336:903–915, 2004.
2. Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 25:3389–3402, 1997.
3. Ronald Bentley. The Shikimate Pathway - A metabolic Tree with Many Branches. *Critical Reviews in Biochemistry and Molecular Biology*, 25(5):307–384, 1990.
4. H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne. The protein data bank. *Nucleic Acids Research*, 28:235–242, 2000.
5. Stephen Bornemann, David J. Lowe, and Roger N. F. Thorneley. The Transient Kinetics of *Escherichia coli* Chorismate Synthase: Substrate Consumption, Product Formation, Phosphate Dissociation, and Characterization of a Flavin Intermediate. *Biochemistry*, 35(30):9907–9916, 1996.
6. S. T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eglmeier, S. Gas, C. E. Barry, 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, and B. G. Barrell. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*, 393:537–544, 1998.
7. Nicolas Guex and Manuel C. Peitsch. SWISS-MODEL and The Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis*, 18:2714–2723, 1997.

8. Karina Kitzing, Sigrid Auweter, Nikolaus Amrhein, and Peter Macheroux. Mechanism of Chorismate Synthase: Role of the Two Invariant Histidine Residues in the Active Site - *In Press*. *Journal of Biological Chemistry*, December, 10 2003.
9. Roman A. Laskowski, Malcolm W. MacArthur, David S. Moss, and Janet M. Thornton. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Appl. Cryst.*, 26:283–291, 1993.
10. Roland Lüthy, James U. Bowie, and David Eisenberg. Assessment of protein models with three-dimensional profiles. *Nature*, 356:83–85, 1992.
11. Peter Macheroux, Stephen Bornemann, Sandro Ghisla, and Roger N. F. Thorneley. Studies with Flavin Analogs Provide Evidence That a Protonated Reduced FMN is the Substrate-induced Transient Intermediate in the Reaction of *Escherichia coli* Chorismate Synthase. *The Journal of Biological Chemistry*, 271(42):25850–25858, 1996.
12. Peter Macheroux, Jürg Schmid, Nikolaus Amrhein, and Andreas Schaller. A Unique reaction in a Common Pathway: Mechanism and Function of Chorismate Synthase in the Shikimate Pathway. *Planta*, 207:325–334, 1999.
13. John MacClean and Sohail Ali. The Structure of Chorismate Synthase Reveals a Novel Flavin Binding Site Fundamental to a Unique Chemical reaction. *Structure*, 11:1499–1511, 2003.
14. Marc A. Martí-Renom, Ashley C. Stuart, András Fiser, Roberto Sánchez, Francisco Melo, and Andrej Šali. Comparative Protein Structure Modeling of Genes and Genomes. *Annual Reviews Biophys. Biomol. Structure*, 29:291–325, 2000.
15. Sophie Quevillon-Cheruel, Nicolas Leulliot, Philippe Meyer, Marc Graille, Michael Bremang, Karine Blondeau, Isabelle Sorel, Anne Poupon, and Joël Janin nad Herman van Tilbeurgh. Crystal Structure of the Bifunctional Chorismate Synthase from *Saccharomyces cerevisiae*. *The Journal of Biological Chemistry*, 279(1):619–625, 2004.
16. Fiona Roberts, Craig W. Roberts, Jennifer J. Johnson, Dennis E. Kyle, Tino Krell, John Coggins, Graham H. Coombs, Willbur K. Milhous, Saul Tzipori, David J. P. Ferguson, Debopam Chakrabarti, and Rima McLeod. Evidence for the Shikimate pathway in Apicomplexan Parasites. *Nature*, 393:801–805, 1998.
17. C.J.A. Sigrist, L. Cerutti, N. Hulo, A.Gattiker, L. Falquet, M. Pagni, A. Bairoch, and P.Bucher. PROSITE: a documented database using patterns and profiles as motif descriptors. *Bioinform.*, 3:265–274, 2002.
18. Julie D. Thompson, Desmond G. Higgins, and Toby 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 Research*, 22(22):4673–4680, 1994.
19. Andrej Šali and Tom L. Blundell. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *Journal of Molecular Biology*, 234(3):779–815, 1993.
20. Andrew C. Wallace, Roman A. Laskowski, and Janet M. Thornton. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Engineering Design and Seletion*, 8:127–134, 1995.
21. World Health Organization, USA. Global Tuberculosis Control Report 2004. <http://www.who.int/tb/en>, 2004.

## **PARTE III**

## 5 Resultados e Discussão

### 5.1 Estudo da sequência de CS de *Mtb*

A proteína CS de diferentes organismos conhecidos tem entre 360 e 400 aminoácidos, exceto em parasitas do filo apicomplexa e em vegetais superiores.

A análise da CS de *Mtb*, via Scan prosite (Sigrist *et al.*, 2002) que identifica os resíduos funcionais presentes na sequência, localizou três regiões de assinatura da proteína altamente conservadas entre as diferentes espécies.

- Primeira região: GESHGRALVAVVEGM

Assinatura: G-[DES]-S-H-[GC]-x(2)-[LIVM]-[GTIV]-x-[LIVT]-[LIV]-[DEST]-[GH]-x-[PV];

- Segunda região: ERASARETAARVAAGTV

Assinatura: [GE]-x(2)-S-[AG]-R-x-[ST]-x(3)-[VT]-x(2)-[GA]-[STAVY]-[LIVMF];

- Terceira região: RSDVCAVPAAGVVETM

Assinatura: R-[SHF]-D-[PSV]-[CSAVT]-x(4)-[SGAIVM]-x-[IVGSTAPM]-[LIVM]-x-E-[STAHNCG]-[LIVMA].

Dentre os resíduos encontrados nas assinaturas encontram-se vários correspondentes aos citados anteriormente nos estudos da estrutura cristalográfica (Maclean & Ali, 2003), e também a His10, citada como importante para a funcionalidade da enzima de *S. pneumoniae* (Kitzing *et al.*, 2004). Este resíduo (destacado em vermelho) encontra-se na primeira região de assinatura, e é um dos resíduos conservados entre todas as espécies.

Na análise da composição de aminoácidos da enzima, realizada com o programa Protparam (Gasteiger *et al.*, 2005), observou-se o alto conteúdo de resíduos arginina

(36) e a predominância de resíduos carregados positivamente (Arg + His + Lis) em relação aos resíduos carregados negativamente (Asp + Glu): 52 a 48, respectivamente, o que era esperado visto que o substrato EPSP é uma molécula com carga efetiva negativa.

TABELA1: COMPOSIÇÃO DE AMINOÁCIDOS (AA) DA PROTEÍNA CS DE *MTB*.

AA	total	%	AA	total	%	AA	total	%	AA	total	%
Ala	70	17,5%	Gln	10	2,5%	Leu	28	7,0%	Ser	19	4,7%
Arg	36	9,0%	Glu	23	5,7%	Lis	8	2,0%	Tre	20	5,0%
Asn	5	1,2%	Gli	43	10,7%	Met	11	2,7%	Trp	3	0,7%
Asp	25	6,2%	His	8	2,0%	Phe	6	1,5%	Tir	7	1,7%
Cis	1	0,2%	Ile	17	4,2%	Pro	22	5,5%	Val	39	9,7%

A partir de uma busca no BLASTp (Altschul *et al.*, 1997) por seqüências similares à seqüência alvo, o programa ClustalX (Thompson *et al.*, 1997) foi utilizado para o alinhamento múltiplo da CS de *Mtb* com as demais espécies de CS, visando a observação da conservação dos resíduos entre estas enzimas, assim como a avaliação da semelhança da CS de *Mtb* com as enzimas reconhecidamente mono ou bifuncionais (Anexo 1). Observou-se que o que melhor se conserva entre todas as seqüências é realmente a região das assinaturas. Ainda a partir deste alinhamento foi realizado um cladograma, onde se observou que a CS de *Mtb* se localiza evolutivamente mais próxima às proteínas monofuncionais (Figura 7).



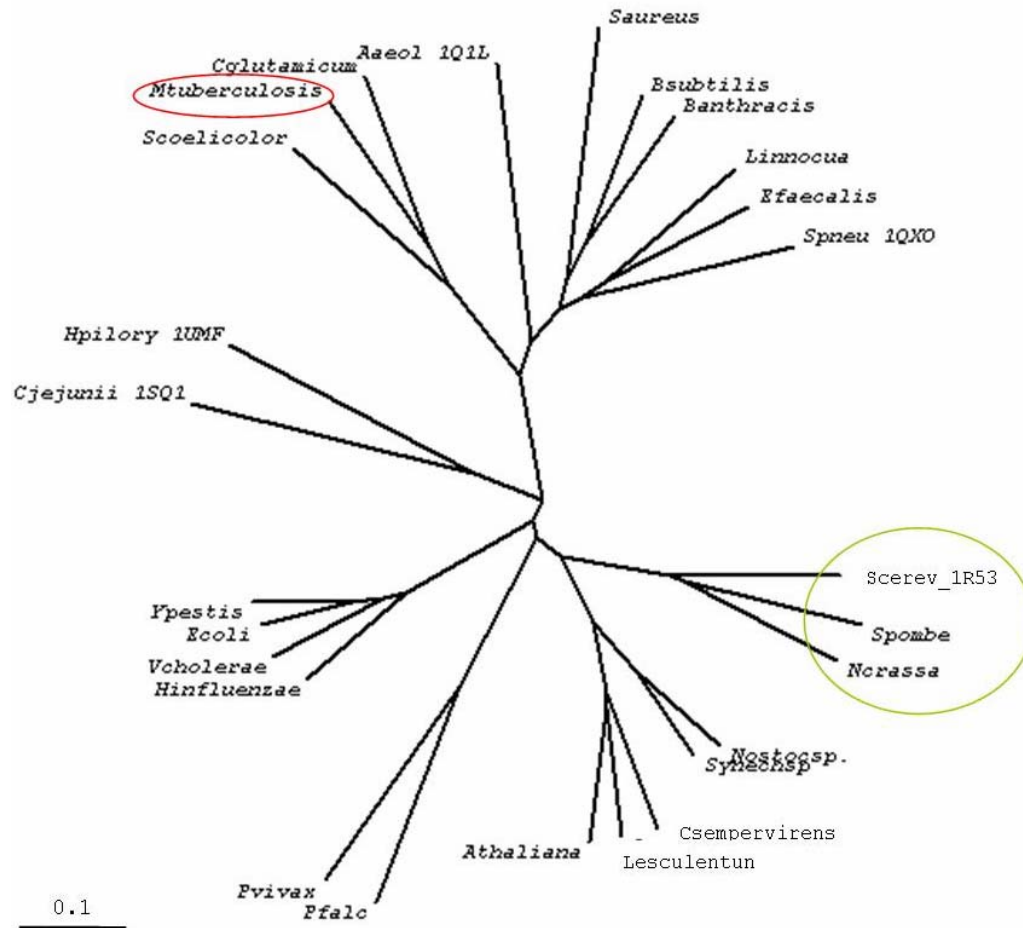


FIGURA 7: CLADOGRAMA GERADO A PARTIR DO ALINHAMENTO MÚLTIPLO DE CS DE VÁRIAS ESPÉCIES (*Mtb* NO CÍRCULO VERMELHO)(ANEXO 1). AQUELAS ESPÉCIES QUE POSSUEM ESTRUTURA 3D ESTÃO INDICADAS PELO COM O CÓDIGO PDB DE SUA ESTRUTURA APÓS O NOME. AS ÚNICAS ESPÉCIES RECONHECIDAMENTE BIFUNCIONAIS ESTÃO INDICADAS PELO CÍRCULO VERDE.

## 5.2 Modelagem Molecular por Homologia

Conforme os resultados encontrados no BLASTp (Altschul *et al.*, 1997), foram encontradas 5 estruturas de CS de outros organismos que poderiam ser utilizadas como moldes para a modelagem da CS de *Mtb*. Por ainda não ter sido publicado o artigo da estrutura de *Campylobacter jejuni* (1SQ1) ela não foi considerada como um candidato a molde.

TABELA 2: RESULTADOS DO BLASTP PARA SEQUÊNCIAS COM ESTRUTURA 3D.

<i>PDB ID</i>	<i>Espécie</i>	<i>Score(Bits)</i>	<i>E value</i>
1QXO	<i>Streptococcus pneumoniae</i>	256	7e-69
1Q1L	<i>Aquifex aeolicus</i>	253	6e-68
1R53	<i>Saccharomyces cerevisiae</i>	139	7e-34
1UMF	<i>Helicobacter pylori</i>	119	1e-27
1SQ1	<i>Campylobacter jejuni</i>	111	2e-25

A partir destes resultados foram feitos alinhamentos par-a-par com cada uma das seqüências selecionadas, para avaliar qual seria o melhor candidato a molde para a modelagem. Este alinhamento avalia a quantidade de resíduos idênticos entre a seqüência do molde e a do modelo (identidades), bem como os resíduos similares, além de possíveis regiões de baixa identidade, que poderiam causar problemas no processo de modelagem.

TABELA 3: RESULTADOS DOS ALINHAMENTOS PAR-A-PAR DA SEQÜÊNCIA DE CS DE *MTB* COM OS CANDIDATOS A MOLDES.

<i>Espécies</i>	<i>Identidade %</i>	<i>Similaridade forte %</i>	<i>Similaridade fraca %</i>	<i>Diferentes %</i>	<i>Gaps</i>
1QX0	42,79	19,15	8,21	29,85	5
1Q1L	41,73	17,27	10,55	30,46	16
1R53	31,18	17,27	11,99	39,57	40
1UMF	29,54	19,85	7,26	43,34	45

A estrutura de CS de *S. pneumoniae* (1QX0) foi, então, escolhida como molde para a construção do modelo 3D de CS de *Mtb*. Esta estrutura possui uma maior identidade em comparação às outras, e também apresenta a menor quantidade de

gaps (regiões de inserções ou deleções) em relação à seqüência de *Mtb*, o que é muito importante na construção de um modelo pela técnica de modelagem molecular por homologia (Martí-Renon *et al.*, 2000). A estrutura 1QX0 tem resolução de 2,0 Å e, além de ser o melhor candidato a modelo, ainda apresenta a vantagem de ter sido cristalizada com o EPSP e o FMN, o que facilitará o estudo posterior do *docking* destes no sítio ativo, uma vez que a cavidade do sítio será modelada já aberta para receber os ligantes.

Foram produzidos 10 modelos com o protocolo padrão do programa MODELLER6v2 (Šali & Blundell, 1993), que posteriormente foram avaliados com os programas PROCHECK (Laskowski *et al.*, 1993), para avaliar a estereoquímica do modelo, e Verify 3D (Luthy *et al.*, 1992), que produz uma validação teórica da estrutura.

### **5.2.1 Resultados do PROCHECK**

PROCHECK é um pacote de análise para validação estereoquímica da estrutura 3D. A primeira análise feita pelo programa é o gráfico de Ramachandran (Figura 7), onde são medidos os ângulos de rotação  $\phi$  e  $\psi$ . Conforme estudos em 118 estruturas cristalográficas com resolução de pelo menos 2 Å (Engh & Huber, 1991), sabe-se que estes ângulos tendem a ficar em regiões específicas deste gráfico, classificadas como regiões favoráveis (vermelho), adicionalmente favoráveis (amarelo), generosamente permitidas (amarelo claro) e desfavoráveis (branco). Sabe-se que para um modelo ter boa qualidade ele deve ter pelo menos 90% dos resíduos na região favorável (Laskowski *et al.*, 1993). Adicionalmente à análise do modelo é feita também a análise do molde como controle.

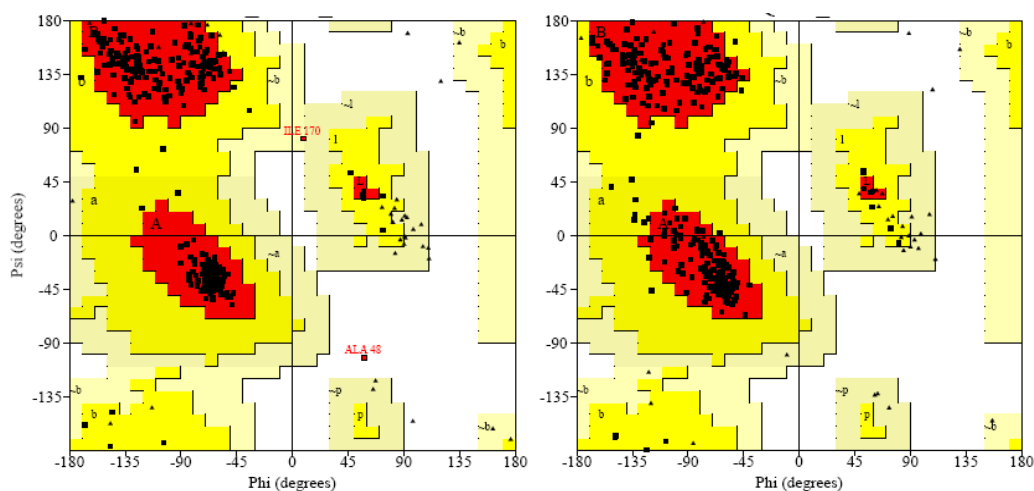


FIGURA 7: GRÁFICOS DE RAMACHANDRAN DO MODELO (ESQUERDA) E DO MOLDE (DIREITA).

TABELA 4: RESULTADOS DO GRÁFICO DE RAMACHANDRAN DO MODELO (ESQUERDA) E MOLDE (DIREITA).

Estatísticas do gráfico de Ramachandran				
	Modelo		Molde	
Resíduos nas regiões mais favoráveis [A,B,L]	318	94,90%	309	93,10%
Resíduos nas regiões adicionalmente favoráveis [a,b,l,p]	15	4,50%	23	6,90%
Resíduos nas regiões generosamente permitidas [~a,~b,~l,~p]	1	0,30%	0	0%
Resíduos nas regiões desfavoráveis	1	0,30%	0	0%
No. de resíduos não-glicina e não-prolina	335	100.0%	332	100%
No. de resíduos terminais (exceto. Gly e Pro)	1		2	
No. de resíduos de glicina (triângulos)	43		38	
No. de resíduos de prolina	22		15	
No. total de resíduos	401		387	

Além desta análise, foram realizadas análises para a cadeia principal e lateral, separadamente (Parte II, Artigo 1, Tabela 2). Estas análises foram expressas em gráficos, com um intervalo de confiança conforme os parâmetros do programa. Caso

estas análises ficassem abaixo do intervalo de confiança, o modelo necessitaria de revisão. Todos os parâmetros analisados do modelo apresentaram-se dentro da média ou melhores.

### 5.2.2 Resultados do Verify 3D

O programa Verify 3D faz uma análise teórica da estrutura 3D de cada aminoácido (perfil 3D) dentro da estrutura da proteína. Esta análise é feita comparando a estrutura do aminoácido a um banco de conformações possíveis para este aminoácido (perfil 1D). Para a comparação é computado um escore, indicando a viabilidade da conformação encontrada (escore 3D-1D). Quanto maior este escore melhor será a estrutura (Figura 8).

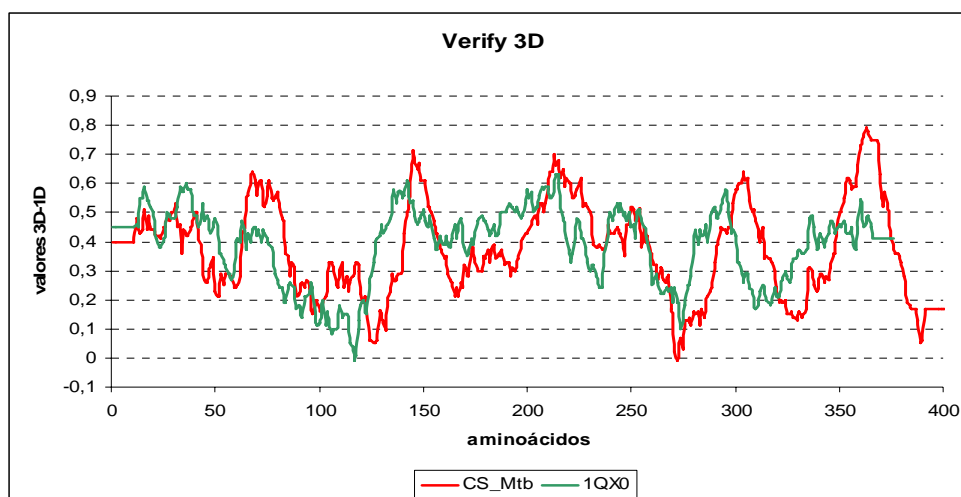


FIGURA 7: GRÁFICO DO VERIFY 3D DO MODELO DA ESTRUTURA DE CS DE *MTB* (VERMELHO) E DO MOLDE 1QX0 (VERDE), INDICANDO A QUALIDADE SEMELHANTE ENTRE AMBOS.

A partir destes resultados foi escolhido um entre os dez modelos construídos para a continuação dos estudos da estrutura quaternária da proteína, e também os estudos de *docking* do substrato e da coenzima.

A estrutura modelada da CS de *Mtb* possui topologia semelhante à do molde. Sua topologia é do tipo  $\beta$ - $\alpha$ - $\beta$ , uma arquitetura típica de CS, apresentando 9  $\alpha$ -hélices e 15 fitas- $\beta$ .



FIGURA 8: REPRESENTAÇÃO DE FITAS DA ESTRUTURA 3D DA CS DE *Mtb* MODELADA A PARTIR DA 1QX0. COLORIDA DE ACORDO COM A ESTRUTURA SECUNDÁRIA:  $\alpha$ -HÉLICES (VERMELHO) E FITAS- $\beta$  (AMARELO) VOLTAS E ALÇAS (CINZA).

### 5.3 Estrutura quaternária

A construção da estrutura quaternária da CS de *Mtb* foi feita a partir da estrutura molde, com a sobreposição das unidades modeladas com cada uma das cadeias do molde. A partir deste modelo foi realizada uma minimização de energia com o programa AMBER7 (Cornel *et al.*, 1995; Case *et al.*, 2002) para corrigir os possíveis contatos desfavoráveis entre os resíduos dos monômeros da enzima. Após a minimização de energia, foi possível observar os arranjos de fitas- $\beta$  na interface das subunidades, que estabilizam a formação do tetrâmero (Figura 9), semelhante ao observado no molde.

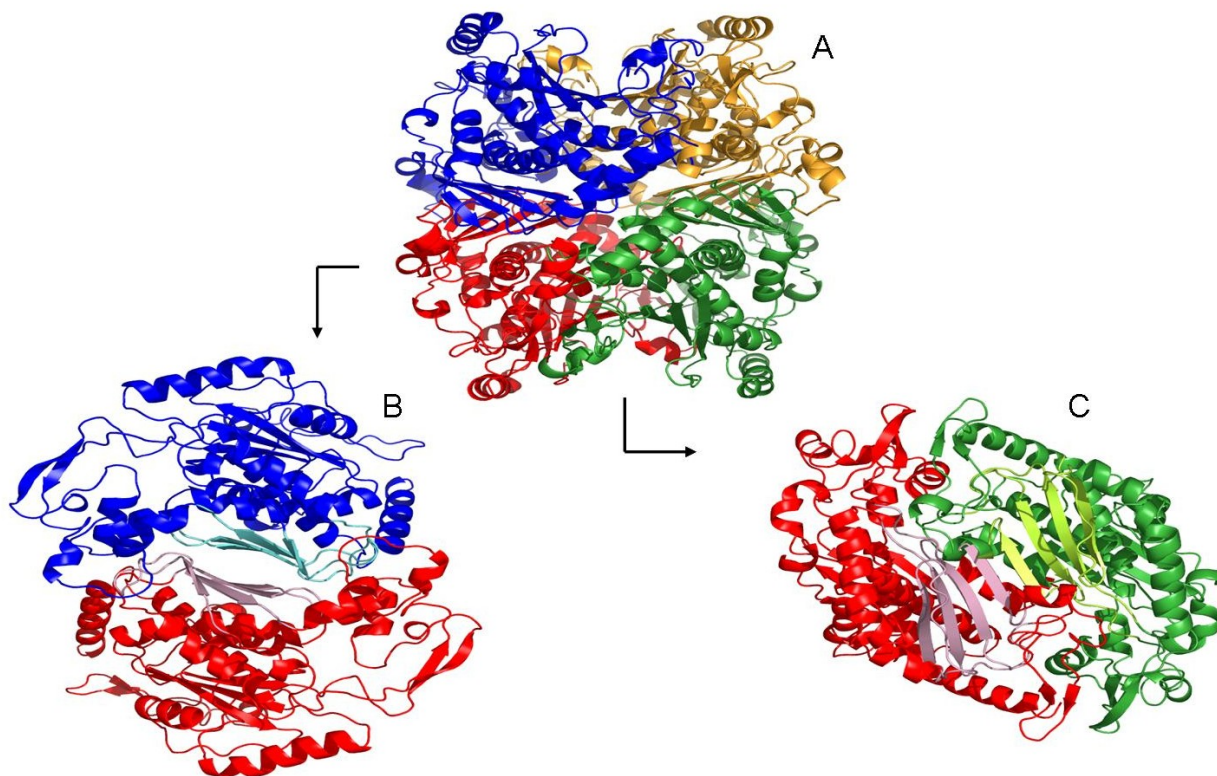


FIGURA 9: (A) TETRÂMERO DA CS DE *MTB*. AS INTERAÇÕES QUE ESTABILIZAM O TETRÂMERO, ISTO É, OS DOIS ARRANJOS DE FITAS- $\beta$  ESTÃO REPRESENTADOS SEPARADAMENTE NOS DÍMEROS EM CORES MAIS CLARAS. (B) VISTA LATERAL EM RELAÇÃO À (A) E (C) VISTA INFERIOR EM RELAÇÃO À (A).

#### 5.4 Estudos de Docking

O primeiro passo no estudo das interações entre os ligantes e a enzima é a colocação da coenzima e do substrato no local do sítio ativo. Para tanto foi feito um *docking* geométrico manual, baseado na localização já conhecida do sítio ativo da estrutura molde (Macleán & Ali, 2003). A cavidade do sítio está localizada entre um dos arranjos de fitas e duas grandes alças da enzima, que devem ter a função de abertura do sítio para a entrada dos ligantes. O FMN e o EPSP ocupam sítios de ligação adjacentes na enzima. Conforme estudos anteriores, o FMN seria o primeiro a entrar no sítio ativo (Macheroux *et al.*, 1996), seguido pelo EPSP. No *docking* manual, se observa que o sítio ativo é uma cavidade realmente estreita, com uma grande quantidade de



resíduos carregados positivamente na parte externa. Assim como observado no molde, a enzima de *Mtb* também apresenta seis argininas e 2 histidinas nesta região (Figura 10).

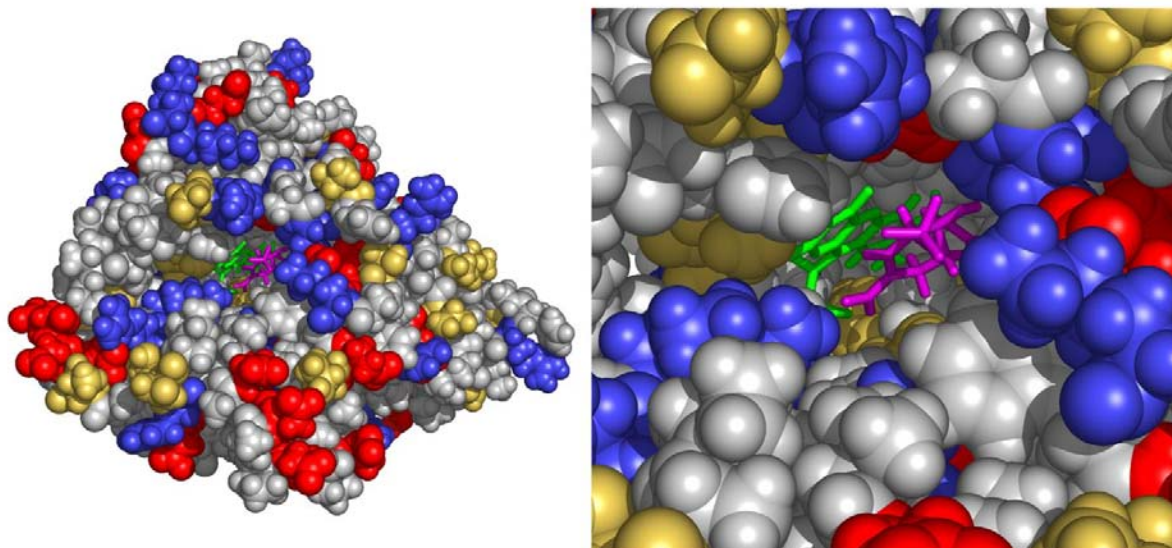


FIGURA 10: REPRESENTAÇÃO UTILIZANDO AS ESFERAS ATÔMICAS DA CS DE *Mtb*, COLORIDAS DE ACORDO COM A NATUREZA DOS AMINOÁCIDOS: APOLAR (CINZA), POLAR (AMARELO), POSITIVO (AZUL) E NEGATIVO (VERMELHO). À DIREITA, VISTA APROXIMADA DO SÍTIO ATIVO COM O EPSP (ROSA) E O FMN (VERDE) NO SÍTIO DE LIGAÇÃO. NOTA-SE A PRESENÇA DE VÁRIOS RESÍDUOS POSITIVOS NESTA REGIÃO.

Em uma primeira análise do *docking* geométrico observou-se que as interações que ocorrem entre os ligantes e a enzima CS de *Mtb* são bem semelhantes às ocorridas na enzima de *S. pneumoniae*. As exceções ficam por conta do resíduo Gln256 (que no molde é a Asn251, um resíduo menor e que faz 2 pontes de hidrogênio com a porção ribose do FMN). No *Mtb*, a Gln256, mais volumosa, aproxima-se muito do FMN. Há ainda outro resíduo que, embora esteja conservado, também se encontra muito próximo ao ligante: a Lis315, próxima ao grupamento fosfato do FMN (Figura 5 do Artigo 2). Nas interações com o EPSP não há nenhum caso onde os resíduos da enzima e os ligantes estejam muito próximos, porém algumas interações entre o EPSP e alguns resíduos não se conservam (Figura 6 do Artigo 2). Como o modelo de *docking* geométrico



manual não se mostrou suficiente para entender as interações entre a enzima e os ligantes, foram propostos dois protocolos de minimização de energia, para melhor acomodar os ligantes dentro do sítio ativo. No primeiro, a enzima foi submetida a minimização de energia mas os ligantes foram mantidos rígidos (modelo com restrição). No segundo, os ligantes também foram minimizados, sendo permitido a estes também se acomodar na cavidade (modelo sem restrição). Em ambos modelos minimizados, o primeiro aminoácido N-terminal foi retirado.

Para analisar o deslocamento dos modelos durante a minimização foi medido o RMSD (desvio médio quadrático) da cadeia principal do modelo original (utilizado no *docking* manual) e os modelos minimizados com e sem restrição, em relação à estrutura do molde 1QX0. Os valores de RMSD foram 1,51 Å (original), 1,57 Å (com restrição) e 1,58 Å (sem restrição), indicando as oscilações ocorridas após a minimização.

#### **5.4.1 Interações com o FMN**

As interações do FMN com os três modelos e o molde encontram-se descritas na Tabela 3 e na Figura 5 do Artigo 1.

Dentre os resíduos e a molécula de EPSP envolvidos nas interações com o FMN, tanto o EPSP como a His110 e a Ala252 têm suas pontes de hidrogênio conservadas nos três modelos. Quanto às interações hidrofóbicas, três são completamente conservadas: Ala133, Ile313 e Ala342 (números correspondentes à CS de *S. pneumoniae*).

Há apenas dois resíduos diferentes entre os sítios ativos da CS de *S. pneumoniae* e *Mtb*: a já mencionada substituição de Asn251 para Gln256 e a substituição da Pro314 para Ser318. No *S. pneumoniae*, a Asn251 realiza duas pontes

de hidrogênio com a porção ribose do FMN. No modelo original do *Mtb*, a Gln256 está muito próxima ao FMN, mas nos modelos minimizados este resíduo passou a fazer contatos hidrofóbicos com o FMN. Já a Pro314 da CS de *S. pneumoniae* realiza um contato hidrofóbico com o FMN enquanto nos modelos minimizados do *Mtb*, o resíduo Ser317 equivalente realiza uma ponte de hidrogênio com a porção ribose.

O resíduo Tre315, embora sempre interaja com a porção isoaloxazina do FMN, apresenta uma variação no número de pontes de hidrogênio formadas nos diversos modelos. A interação hidrofóbica do FMN com a Ile255 da CS de *Mtb* só foi observada no modelo original, provavelmente devido à limitação do processo de *docking* manual. A interação hidrofóbica da Ala349 só apareceu em dois dos modelos (original e com restrição), não sendo detectada no modelo de minimizado sem restrições.

Foram observadas duas interações novas entre a CS de *Mtb* e o FMN, em relação a *S. pneumoniae*: uma ponte de hidrogênio com a Gli113 (numeração dos modelos minimizados) que apareceu nos dois modelos minimizados, e um contato hidrofóbico com a Met314 (numeração do modelo original) que ocorreu nos três modelos de *Mtb*.

As observações acima indicam que, provavelmente, estes resíduos tenham uma participação importante na associação da CS de *Mtb* com a molécula de FMN.

#### **5.4.2 Interações com o EPSP**

Os resultados analisados a seguir estão representados na Tabela 4 e na Figura 6 do Artigo 1.

Há 9 resíduos que fazem interação com o EPSP além do FMN, analisado anteriormente. Destes, apenas a Ala133 faz uma ponte de hidrogênio com a carbonila do EPSP tanto no molde como nos modelos (numeração conforme o molde). Há três

pontes de hidrogênio que se formavam no molde e não foram observadas em nenhum dos modelos: a His10 e Arg48 que interagem com a porção fosfato e a Arg134 que interage com a porção piruvil.

No modelo original foram observados dois contatos hidrofóbicos desfavoráveis que posteriormente foram corrigidos com a minimização de energia, a Arg49 e Arg50. Ainda sobre as argininas que interagem com o EPSP, a Arg39 e Arg45 fazem interação em todos os modelos e no molde, porém estas interações têm algumas pequenas diferenças, tanto no número quanto na posição das pontes de hidrogênio formadas com o EPSP. Todas as interações são com a porção piruvil do substrato.

Quanto às interações que se formaram apenas no *Mtb*, temos a Arg111, fazendo uma ponte de hidrogênio com a carbonila do EPSP no modelo sem restrição e a Ser136 fazendo uma interação hidrofóbica também com este modelo.

As interações da enzima com a porção fosfato do EPSP não foram detectadas nos modelos da CS de *Mtb*. A porção fosfato destes modelos apresenta-se voltada para fora da enzima. O deslocamento desta porção do substrato, em relação a sua posição cristalográfica, foi de 1,2 Å, o que pode justificar a perda destas interações com resíduos tão importantes da enzima.

## 6. CONCLUSÃO E PERSPECTIVAS

A análise da estrutura de CS de *Mtb*, gerada a partir deste estudo de Modelagem Molecular por Homologia, demonstrou a boa qualidade do modelo proposto para esta enzima e a possibilidade de sua utilização nos estudos de *docking*.

A minimização de energia do complexo ternário, formado pela CS de *Mtb*, o FMN e o EPSP, mostrou melhorias nas estruturas propostas quando comparadas à estrutura do complexo ternário gerada via *docking* geométrico manual. Entretanto, a minimização de energia sem restrições de movimentos de enzima e ligantes levou ao modelo que melhor descreve as interações intermoleculares.

De acordo com os estudos realizados e com o cladograma apresentado, pode-se concluir que, assim como a CS de *S. pneumoniae* (molde), a CS de *Mtb* é monofuncional, e, portanto, deve apresentar um padrão de interação com seus ligantes semelhante ao molde. Estudos adicionais de *docking* molecular, utilizando o programa AutoDock 3.1 (Morris *et al.*, 1998), poderão representar uma nova alternativa para o estudo destas interações.

Simultaneamente ao nosso estudo de modelagem molecular, a estrutura cristalográfica da CS de *Mtb* foi determinada com uma resolução de 2,65 Å (Dias *et al.*, 2006; PDB ID: 1ZTB). A avaliação comparativa demonstrou que ambas estruturas 3D (determinada por cristalografia e modelada por homologia) apresentam um dobramento similar. A sobreposição da estrutura cristalográfica e do nosso modelo apresentou um RMSD de 2,11 Å (para a cadeia principal), quando medido em relação ao modelo original de CS, e um RMSD de 2,14 Å, quando medido em relação ao modelo de CS minimizado sem restrições.

Como a estrutura da 1ZTB foi cristalizada sem a presença de ligantes, ela apresenta-se em uma conformação mais fechada em relação ao modelo proposto. A presença de FMN e EPSP no modelo proposto por homologia induz uma mudança conformacional na enzima CS para acomodação destes ligantes, conforme descrito ([Macheroux et al., 1998](#)). Os RMSDs observados descrevem esta pequena diferença. Assim, podemos considerar que a construção da estrutura da CS de *Mtb* via Modelagem Molecular por Homologia foi bem sucedida como representação da estrutura 3D desta enzima.

A utilização do modelo de CS de *Mtb* proposto via Modelagem Molecular por Homologia, assim como dos estudos preliminares de sua associação com os ligantes FMN e EPSP nos permitem um melhor entendimento das interações do complexo enzima-substrato-coenzima ([Fernandes et al., 2005](#); [Fernandes et al., 2006](#)), e deverá, portanto, facilitar a busca por novos inibidores desta enzima e o desenvolvimento de novos fármacos para o tratamento da Tb.

## 6 REFERÊNCIAS:

- Ahn, H.J.; Yoon, H.J.; Lee, B.I.; Suh, S.W. *J. Mol. Biol.*, 336:903-915, 2004.
- Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. *Nucleic Acids Res.*, 25:3389–3402, 1997.
- Basso, L.A. & Blanchard, J.S. *Adv. Exp. Med. Biol.*, 456:115-144, 1998.
- Basso, L.A.; Zheng, R.; Musser, J.M.; Jacobs Jr., W.R.; Blanchard, J.S. *J. Infect.Dis.*, 178:769-775, 1998.
- Bentley, R. *Crit. Rev. Biochem. Mol. Biol.*, 25(5):307–384, 1990.
- Bornemann, S.; Lowe, D.J.; Thorneley, R.N.F. *Biochemistry*, 35(30):9907-9916, 1996.
- Bornemann, S. *Nat. Prod. Rep.*, 19(6):761-772, 2002.
- Camus, J.C.; Pryor, M.J.; Médigue, C.; Cole, S.T. *Microbiology*, 148:263-270, 2002.
- Case, D.A.; Pearlman, D.A.; Caldwell, J.W.; Cheatham III, T.E.; Wang, J.; Ross, W.S.; Simmerling, C.L.; Darden, T.A.; Merz, K.M.; Stanton, R.V.; Cheng, A.L.; Vincent, J.J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R.J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G.L.; Singh, U.C.; Weiner, P.K.; Kollman, P.A. University of California, San Francisco, 2002.
- Coates, A.R.M. *FEMS Microbiology Letters*, 158:139-45, 1998.
- Cohn, D.L.; Bustreo, F.; Raviglione, M.C. *Clinical Infectious Disease*, 24:S121-30, 1997.

- Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry III, C.E.; Tekaia, F.; Badcock, K.; Basham, D.; Brown, D.; Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.; Murphy, L.; Oliver, K.; Osborne, J.; Quail, M.A.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, J.; Squares, R.; Squares, S.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G. *Nature*, 393:537-544, 1998.
- Cornell, W.D.; Cieplak, P.; Bayly, C.I.; Gould, I.R.; Merz Jr., K.M.; Ferguson, D.M.; Spellmeyer, D.C.; Fox, T.; Caldwell, J.W.; Kollman, P.A. *J. Am. Chem. Soc.*, 117:5179-5197, 1995.
- Daffé, M. *Microbiology*, 144:577-87, 1998.
- Dias, M.V.; Borges, J.C.; Ely, F.; Pereira, J.H.; Canduri, F.; Ramos, C.H.; Frazzon, J.; Palma, M.S.; Basso, L.A.; Santos, D.S.; de Azevedo W.F.Jr. *J. Struct. Biol.*, 154(2):130-143, 2006.
- Du, W.; Wallis, N.G.; Mazzulla, M.J.; Chalker, A.F.; Zhang, L.; Liu, W.S.; Kallender, H.; Payne, D. *European Journal of Biochemistry*, 267:222-27, 2000.
- Engh, R. A. & Huber, R. *Acta Cryst.*, A47:392-400, 1991.
- Fernandes, C.L.; Santos, D.S.; Basso, L.A.; Norberto de Souza, O. *Lectures Notes in Computer Science*, 3594:118-127, 2005.
- Fernandes, C.L.; Breda, A.; Santos, D.S.; Basso, L.A.; Norberto de Souza, O. *Computers in Biology and Medicine*, 2006, no prelo.
- Fitzpatrick, T.; Ricken, S.; Lanzer, M. Amrhein, N.; Macheroux, P.; Kappes, B. *Mol. Microbiol.*, 40(1):65-75, 2001.
- Garble, T. *Mol. Gen. Genet.*, 228: 385-92, 1991.

- Gasteiger, E.; Hooland, C.; Gattiker, A.; Duvaud, S. Wilkins, M.R.; Appel, R.D.; Bairoch, A. Ed. John M. Walker, Humana Press, 571-607, 2005.
- Grosset, J. *Clin. Chest. Med.*, 1:231-241, 1980.
- Hasan, N. & Nester, E.W. *J. Biol. Chem.*, 253(14):4993-4998, 1978.
- Horsburgh, M.J.; Foster, T.J.; Barth, P.T.; Coggins, J.R. *Microbiology*, 142:2943-2950, 1996.
- Henstrand J.M.; Schaller, A.; Braun, M.; Amrhein, N.; Schmid, J. *Mol. Microbiol.*, 22(5):859-866, 1996.
- Heym, B.; Alzari, P.M.; Honore, N.; Cole, S.T. *Mol. Microbiol.*, 15:235-245, 1995.
- Hiriyan, K.T. & Ramakrishnan, T. *Arch. Microbiol.*, 144:105-109, 1986.
- Hulo, N.; Sigrist, C.J.A.; Le Saux, V.; Langendijk-Genevaux, P.S.; Bordoli, L.; Gattiker, A.; De Castro, E.; Bucher, P.; Bairoch, A. *Nucleic Acids Res.*, 32:D134-D137, 2004.
- Kitzing, K.; Macheroux, P.; Amrhein, N. *J. Biol. Chem.*, 276(46):42658-42666, 2001.
- Kitzing, K.; Auweter, S.; Amrhein, N.; Macheroux, P. *J. Biol. Chem.*, 10:9451-9461, 2004.
- Laskowski, R.A.; MacArthur, M.W.; Moss, D.S.; Thornton, J.M. *J. Appl. Cryst.*, 26:283-291, 1993.
- Lehninger, A.L.; Nelson D. L.; Cox, M. M. 2<sup>a</sup> Edição. São Paulo; Sarvier, 1995.
- Lüthy, R.; Bowie, J.U.; Eisenberg, D. *Nature*, 356:83–85, 1992.



- Macheroux, P.; Bornemann, S.; Ghisla, S.; Thorneley, R.N.F. *J. Biol. Chem.*, 271(42):25850–25858, 1996.
- Macheroux, P.; Schonbrunn, E.; svergun, D.I.; Volkov, V.V.; Koch, M.H.J.; Bornemann, S.; Thorneley, R.N.F. *Biochem J.*, 335:319-327, 1998.
- Macheroux, P.; Schmid, J.; Amrhein, N.; Schaller, A. *Planta*, 207:325–334, 1999.
- Maclean, J. & Ali, S. *Structure*, 11:1499–1511, 2003.
- Manabe, Y.C. & Bishai, W.R. *Nat. Med.*, 6:1327-1329, 2000.
- Martí-Renom, M.A.; Stuart, A.C.; Fiser, A.; Sánchez, R.; Melo, F.; Šali, A. *Annu. Rev. Biophys. Biomol. Struct.*, 29:291-325, 2000.
- Morris, G. M.; Goodsell, D. S.; Halliday, R.S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Computational Chemistry*, 19: 1639-1662. 1998
- MS - Ministério da Saúde – Portal da saúde - Tópicos de saúde, Tuberculose: [http://portal.saude.gov.br/portal/svs/visualizar\\_texto.cfm?idtxt=23493](http://portal.saude.gov.br/portal/svs/visualizar_texto.cfm?idtxt=23493). Julho de 2006.
- OMS - World Health Organization -. WHO | Tuberculosis. <http://www.who.int/tb/en/>. Julho de 2006.
- Pittard, A.J. 2ª edição, volume 1. ASM Press, Washington DC. Cap 28:458-84, 1996.
- Quevillon-Cheruel, S.; Leulliot, N.; Meyer, P.; Graille, M.; Bremang, M.; Blondeau, K.; Sorel, I.; Poupon, A.; Janin, J.; van Tilbeurgh, H. *J. Biol. Chem.*, 279(1):619-625, 2004.
- Russell, D.G. *Nat. Rev. Mol. Cell Biol.*, 2:569-77, 2001.

- Šali, A. & Blundell, T.L. *J. Mol. Biol.* 234(3):779-815, 1993.
- Sigrist, C.J.; Cerutti, L.; Hulo, N.; Gattiker, A.; Fakquet, L.; Pagni, M.; Bairoch, A.; Bucher, P. *Brief Bioinform.*, 3930:265-274, 2002.
- Silva, M.S.; Senna, S.G.; Ribeiro, M.O.; Valim, A.R.; Telles, M.A.; Kritski, A.; Morlock, G.P.; Cooksey, R.C.; Zaha, A.; Rossetti, M.L. *J. Clin. Mricobiol.*, 41(9):4471-4474, 2003.
- Sikorski, J.A. & Gruys, K.J. *Accounts of Chemical Research*, 30:2-8, 1997.
- Telles, M.A.; Ferrazoli, L.; Waldman, E.A.; Giampaglia, C.M.; Martins, M.C.; Ueki, S.Y.; Chimara, E.; Silva, C.A.; Waldman, C.C.; Heyn, I.; Hirono. I.U.; Riley, L.W. *Int. J. Tuberc. Lung Dis.*, 9(9):970-976, 2005.
- Thomas, M.G.; Lawson, C.; Allanson, N.M.; Leslie, B.W.; Bottomley, J.R.; McBride, A.; Olusanya, O.A. *Bioorg. Med. Chem. Lett.*, 13(3):423-426, 2003.
- Thompson, J.D.; Ginson, T.J.; Plewniak, F.; Jeanmougin, F.; Higgins, D.G. *Nucleic Acids Res.*, 25(24):4876-4882, 1997.
- Viola, C.M.; Saridakis, V.; Christendat, D. *Proteins*, 54(1):166–169, 2004.
- Zheng, Y.J. & Ornstein, R.L. *J. Biomol. Struct. Dyn.* 14(2):231-233, 1996.

## **ANEXOS**

## Anexo 1

Alinhamento múltiplo produzido com o programa ClustalX com uma matriz da série BLOSUM para as seqüências de:

- *Saccharomyces cerevisiae* (Scerev\_1R53) Número de acesso: [gi|40889609|pdb|1R53|A](#)
- *Schizosaccharomyces pombe* (Spombe) Número de acesso: [gi|63054453|ref|NP\\_588359.2|](#)
- *Neurospora crassa* (Ncrassa) Número de acesso: [gi|9367484|emb|CAB97473.1|](#)
- *Nostoc* sp. (Nostocsp) Número de acesso: [gi|17130142|dbj|BAB72754.1|](#).
- *Synechocystis* sp. (Synechsp) Número de acesso: [gi|16330007|ref|NP\\_440735.1|](#)
- *Bacillus subtilis* (Bsubtilis) Número de acesso: [gi|143806|gb|AAA20859.1|](#)
- *Bacillus anthracis* (Banthraxis). Número de acesso: [gi|47503399|gb|AAT32075.1|](#)
- *Listeria innocua* (Linnocua) Número de acesso: [gi|16414543|emb|CAC97272.1|](#)
- *Staphylococcus aureus* (Saureus) Número de acesso: [gi|49483655|ref|YP\\_040879.1|](#)
- *Enterococcus faecalis* (Efaecalis) Número de acesso: [gi|29343589|gb|AAO81351.1|](#)
- *Escherichia coli* (Ecoli) Número de acesso: [gi|16130264|ref|NP\\_416832.1|](#)
- *Yersinia pestis* (Ypestis) Número de acesso: [gi|45437065|gb|AAS62617.1|](#)
- *Vibrio cholerae* (Vcholerae) Número de acesso: [gi|75824680|ref|ZP\\_00754128.1|](#)
- *Haemophilus influenzae* (Hinfluenzae) Número de acesso: [gi|1573154|gb|AAC21865.1|](#)
- *Mycobacterium tuberculosis* (Mtuberculosis) Número de acesso: [gi|15609677|ref|NP\\_217056.1|](#)

- *Corynebacterium glutamicum* (Cglutamicum) Número de acesso: [gi|41325843|emb|CAF21632.1|](#)
- *Streptomyces coelicolor* (Scoelicolor) Número de acesso: [gi|8249955|emb|CAB93376.1|](#)
- *Streptococcus pneumoniae* (Spneu\_1QXO) Número de acesso: [gi|40889471|pdb|1QXO|D](#)
- *Campylobacter jejuni* (Cjejunii\_1SQ1) Número de acesso: [gi|47169186|pdb|1SQ1|A](#)
- *Helicobacter pylori* (Hpilory\_1UMF) Número de acesso: [gi|49259375|pdb|1UMF|D](#)
- *Aquifex aeolicus* (Aaeol\_1Q1L) Número de acesso: [gi|37927759|pdb|1Q1L|D](#)
- *Plasmodium falciparum* (Pfalci) Número de acesso: [gi|2258447|gb|AAB63293.1|](#)
- *Plasmodium vivax* (Pvivax) Número de acesso: [gi|18030075|gb|AAL56611.1|](#)
- *Lycopersicon esculentum* (Lesculentum) Número de acesso: [gi|410482|emb|CAA79859.1|](#)
- *Corydalis sempervirens* (Csempervirens) Número de acesso: [gi|18256|emb|CAA43034.1|](#)
- *Arabidopsis thaliana* (Athaliana) Número de acesso: [gi|53749180|gb|AAU90075.1|](#)

# CLUSTAL X (1.81) MULTIPLE SEQUENCE ALIGNMENT

File: C:Corismatoalig\_asta\_final5.ps  
Page 1 of 5

Date: Wed Jul 26 11:30:24 2006

```

:: : **** : :: . . . : : : * : . * : * : : *
Spombe -----MSFSGTLFKVITYGESHCKSVGCVIVEGCPGPNLIESDVVOLTTRRRPGQSNLITPRNEKDQVIOQSGTEFGVTLGTFPIGMLVLNQDQKPH 91
Ncrassa -----MSTFGHYFRVITYGESHCKSVGCVIVDGVPPGMELTEDDIQPMTRRRPGQSAITTPRDEKDRVIOQSGTEFGVTLGTFPIGMLVMNEDQRFK 91
Scerev_1R53 -----MSTFGKLFrvITYGESHCKSVGCVIVDGVPPGMSLTDADIQPLTRRRPGQSKLSTPRDEKDRVETQSGTEFGKTLGTFPIAMMIKNEQDRPH 91
Nostocsp. -----MGNTFGHLFRITTFGESHGGGVGVVLDGCPPLLEISPEEIQLDLRRRPGQSKITTPRKEADTCEILSGVYEGKTLGTFPIISILVRNKDTRPQ 92
Synechsp -----MGNTFGSLFRITTFGESHGGGVGVVLDGCPPLLEISPEEIQLDLRRRPGQSKITTPRKEADQCEILSGVYEGKTLGTFPIALLVRNKDARSQ 92
Lesculentum MASFVPTKQFVGASSSDIGSSRLVSLQLPSKFS--SSNFHLPSR-PSQLKRLEIQAASTFGNYFRVITTFGESHGGGVGCVIIDGCPPLRPLSESDMVELDRRRPGQSRITTPRKEADTCKISSGADGLTTSPIKVEVFNTRDGRN 146
Athaliana -----MASSSLTSKSLGSKLGSLSLPSLRLRSLSPAVOISLR-TQTRKNFQIQATGSSYGFHFRVITTFGESHGGGVGCVIIDGCPPLRPLSESDMVELDRRRPGQSRITTPRKEADTCKISSGADGLTTSPIHVFVFNTRDGRN 142
Csempervirens MASSLSTKPFLLSGRRRSITDGSQWSYFQTSDLRQLSNQSVQISVRRQIAPLKLVVQASGSSFGKVFQVITTFGESHGGGVGCVIIDGCPPLRPLSESDMVELDRRRPGQSRITTPRKEADTCKIYSVADGFTTSPHISVFNTRDGRN 150
Linnocua -----MRYLTAGESHGPGLTIIIEGLPAGMPLLAEDVNLKELRRQGGHGRGARMRIEKDQVQITAGIRHGKTLGAPVAMFVENKDWKHW 84
Efaecalis -----MRFITAGESHGPGLTIIIEGLPAGLPLSSEINRELARRGGYGRGGRMKIEKDQVQITAGIRHGKTLGAPVTLIVENKDWKHW 84
Spneu_1QX0 -----MRYLTAGESHGPGLTIIIEGLPAGLPLSSEINRELARRGGYGRGGRMKIENDQVQITAGIRHGKTLGAPITMDVINKDWKHW 84
Bsubtilis -----MRYLTAGESHGPGLTIIIEGVPAGLYITTEEDINFELARRQGGHGRGGRMKIEKDQAKIMSQVRRHARTLGSPIALVVENKDWKHW 84
Banthracis -----MRYLTAGESHGPGLTIIIEGVPAGLYITTEEDINFELARRQGGHGRGGRMKIEMDVEIVSGVRRHGMLGSPITLIVKNDFFKHW 84
Saureus -----MRYLTAGESHGPGLTIIIEGVPAGLYITTEEDINFELARRQGGYGRGGRMKIEKDVEIVSGVRRNGYTLGSPITMVVITDDFFTHW 84
Aaeol_1Q1L -----GSHXSLRVLRLFTAGESHKGGLTALLEGIPANLPLSEEEINHELRRRQGGYGRGGRMKIEKDTAEILSGVRFKTLGSPIALFIRNRDWNEN 92
Mtuberculosis -----MLRWITAGESHGRALVAVVEGMVAGVHVTSADIADQLARRRLGYGRGARMFERRDAVTVLSGIRHGKTLGSPIALEIGNTEWPKW 85
Cglutamicum -----MLGMLRWITAGESHQALIAIVHEMPAGVPVTKDEVSYQLARRRLGYGRGARMKFEQDALFTLGTIRHGKTLGSPISIMIGNTEWPKW 88
Scoelicolor -----MSRLRWITAGESHPALVATLEGLPAGVPIITTEMVADHLARRRLGYGRGARMKFERDEVTVLGGVRRHGLTMGSPVAVMVGNTWPKW 87
Cjejunii_1SQ1 -----XNTFTRLKFTSFESHGVAVGCIIIDGXPAVKFDEEFLQNELDKRKGK-SKFATPRKESDKAQVLSGVFEGYTTGHPPIAIVVFNENAHSK 90
Hpilory_1UMF -----MNTLGRFLRLTFGESHGDVIGVLDGMPGKIDYALLENEMKRRQGGRNVEITPRKEDDKVEITSGVFEFDFSTGTFPIGFLIHNQRASK 91
Ecoli -----MAGNTIGQLFRVITTFGESHGLALGCVIVDGVPPGIPITLTDLQHDLLDRRRPGTSRYTTQRRPEPQVKILSGVFEVTTGTSIGLLIENTDORSQ 93
Ypestis -----MMAGNSIQGFRVITTFGESHGIALGCVIIDGVPPIPIITLTDLQHDLLDRRRPGTSRYTTQRRPELQVRILSGVFEVTTGTSIGLMIENTDORSQ 94
Vcholerae -----MAGNSIQGFRVITTFGESHGIALGCVIIDGVPPIPIITLTDLQHDLLDRRRPGTSRYTTQRRPEDEVKILSGVFEKTTGTSIGLLIENTDORSK 93
Hinfluenzae -----MAGNTIGQLFRVITTFGESHGIALGCVIIDGVPPIPIITLTDLQHDLLDRRRPGTSRYTTQRRPEDEVKILSGVFEKTTGTSIGLMIENTDORSQ 93
Pfalca -----MSTYGTLLKVTTFGESHGKALGCVIIDGFLSNIEINFDLIQKDLRRRPNQSKLTSNRNEPKDLVILSGFDENKTLGTFITFLIYNEDIKKE 91
Pvivax -----MSTYGTLLKVTTFGESHGKALGCVIIDGFLSNIEINFDLIQKDLRRRPNQSKLTSNRNEPKDLVILSGFDENKTLGTFITFLIYNEDIKKE 91
ruler 1.....10.....20.....30.....40.....50.....60.....70.....80.....90.....100.....110.....120.....130.....140.....150

```



# CLUSTAL X (1.81) MULTIPLE SEQUENCE ALIGNMENT

File: C:Corismatoalig\_asta\_final5.ps  
Page 2 of 5

Date: Wed Jul 26 11:30:24 2006

	* * * *      **                      * * * * :      *                      :	
Spombe	DYS--DMDNYP-----RPSHADYTYLEKYGVKASSGGG-RSSARETIGRVAAGAIAEKYLLEAYGVEIVAFVSSVGKIAIPLHETASSA-----	172
Ncrassa	DYGNKTMITYP-----RPSHADWTYLEKYGVKASSGGG-RSSARETIGRVAAGAIAEKYLKLAYGVEIVAFVSSVG-----SEH-----	164
Scerev_1R53	DYS--DMDKFP-----RPSHADFTYSEKYGIKASSGGG-RASARETIGRVASGAIAEKFLAONSNVEIVAFVTOIG-----	159
Nostocsp.	DYD--EMAQKY-----RPSHADATYDAKYGIRNWQGGG-RSSARETIGRVAAGAIAKKILROVANVEVIGYVKRIK-----	160
Synechsp	DYN--EMAVKY-----RPSHADATYEAKYGIRNWQGGG-RSSARETIGRVAAGAIAKKILAFNGVEIVAVYKSIQ-----	160
Lesculentum	DYS--EMSLAY-----RPSHADATYDFKYGVRVSVQGGG-RSSARETIGRVAAGAVAKKILKLYSGAEVLAYVSOVH-----	214
Athaliana	DYS--EMSVAY-----RPSHADATYDMKYGVRVSVQGGG-RSSARETIGRVAAGALAKKILKQFAGTEILAYVSOVH-----	210
Csempervirens	DYS--EMAKAY-----RPSHADATYDFKYGVRVSVQGGG-RSSARETIGRVAAGALAKKILKAVAGTEVLAYVSOAH-----	218
Linnocua	ETVMSIEPVPEKNEK--SRR-----VSRPRPGHADLVGGMKYGHNDMRNVLERSSARETTVRVAAGAVAKKLLHELQ-IEVAGHVLEIGG-----	166
Efaecalis	TSVMSVEPVPEKQKK--IRR-----VSKPRPGHADLVGGMKYQHDDLNRNVLERSSARETTMRVAIGAVAKKLLAELD-IOVAGHVAVLGG-----	166
Spneu_1QXO	LDIMSARDIEDRLKS--KRK-----TTHPRPGHADLVGGIKYRFDDLNRNLSLERSSARETTMRVAVGAVAKRLLAELD-MEIANHVVVFGG-----	166
Bsubtilis	TKIMGAAPITEDIEKEMKRO-----ISRPRPGHADLNGAIKYNHRDMRNVLERSSARETTVRVAAGAVAKKILSELG-IKVAGHVLOIGA-----	168
Banthracis	TKVMGAAPISEKESKEMKRT-----ITKPRPGHADLNGAIKYGHRDIRNVLERSSARETTVRVAAGAVAKKILKELG-VEIAGHVLEIGG-----	168
Saureus	RKIMGAAPISEEARENMKRT-----ITKPRPGHADLVGGMKYNHRDLNRNVLERSSARETAARVAVGALCKVLLQLD-IDIYSRVVEIGG-----	168
Aaeol_1Q1L	KEKMAIEGEPSPSVVP-----FTRPRPGHADLVGGIKYQNRDLRNILERASARETAARVAVGAVCKKFLSEFG-IKIGSFVVSIGQ-----	172
Mtuberculosis	ETVMAADPVDPAHLADVARN-----APLTRPRPGHADYAGMLKYGFDDARVLERASARETAARVAAGTVAARAFLOALGVEVLSHVISIGA-----	172
Cglutamicum	TTIMSSDALDMEDPDNVAAMSSGRGAKLTRPRPGHADYAGMLKYGFDDARNVLERSSARETAARVAATVARSFLRETLGVEVLSHVISIGA-----	180
Scoelicolor	EQVMAADPVDPEVLAGLARN-----APLTRPRPGHADLACMKYGFDEARFILERASARETAARVALGAVARSYLKETAGVEIVSHVVELAS-----	174
Cjejunii_1SQ1	DYDNLKD-----LFRPAHADFTYFYKYGIRDHRGGG-RSSARESVARVAGGAVAAXLLREFD-ICVQSGVFGVGT-----	158
Hpilory_1UMF	DYDNIKN-----LFRPSHADFTYFHKYGIRDYFRGGG-RSSARESAIRVAAGAFAMLLREIG-IVCESGIIEIG-----	158
Ecoli	DYSAIKD-----VFRPGHADYTYEQKYGLRDYRGGG-RSSARETAMRVAAGAIKKYLAEKFGIEIRGCLTOMG-----	161
Ypestis	DYSAIKD-----VFRPGHADYTYEQKYGVDRDYRGGG-RSSARETAMRVAAGAIKKYLAQKFGVQVRGYLAOMG-----	162
Vcholerae	DYSDIKD-----KFRPGHADYTYHQYGVDRDYRGGG-RSSARETAMRVAAGAIKKYLAQEFGEIVRAYLSOMG-----	161
Hinfluenzae	DYGDIKD-----RFRPGHADFTYQOKYGIRDYRGGG-RSSARETAMRVAAGAIKKYLAHEHFGIEVRGFLSOIGN-----	162
Pfalca	DYNSFIN-----IPRPGHGDYTYFMKYHVKNKSSSS-RFSGRETAIRVAAGACIEQWLHTFYNCTIVCYVHSVGNIKLPEQVSKELFNKNPPSRDLVDVSYGTVRYNKKKIFMDCFNRIYDMNASMLKIDEYN	218
Pvivirus	NYSSFID-----IPRPGHGDYTYFMKYHVKNKSSSS-RFSGRETAIRVAAGACIEQWLHTFYNCTIVCYVHSVGNIKLPEQVSKDFE-RNPPSRDLVDVDAHGAVKYHCGRGTFTDFFGNVYNANGFELRGGEAA	217
ruler	.....160.....170.....180.....190.....200.....210.....220.....230.....240.....250.....260.....270.....280.....290.....300	



# CLUSTAL X (1.81) MULTIPLE SEQUENCE ALIGNMENT

File: C:Corismatoalig\_asta\_final5.ps  
Page 3 of 5

Date: Wed Jul 26 11:30:24 2006

	:	:::	*:	.	:::	..	.:	..	*:	*	*																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
Spombe	-----	ILDPEDD	FES	PI	ITAEYLKFLNKITRE	VDKTT	VRCPHAA	IAAKMAERIT	FRDNDHS	IGG	VT	CVIRNVPTG	-----	LGEPC	FDKLEAKLAHAMMS	IPATK	SF	IGSGREG	CKVAGS	SKHNDLFY	RNA	295																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																															
Ncrassa	-----	LFPPTAEH	PS	STN	EFLKLVNSITRETVDS	FLVRC	PDAA	BANKRMED	LITFRDNDHS	IGG	TV	TCVIRNVPSG	-----	LGEPA	FDKLEAMLAHAML	SIPATK	KG	FVFGSG	GGCEV	PGSIHND	PFVSAE	288																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																															
Scerev_1R53	-----	IKMNRDS	SFD	EFQ	LLNTITREKVDMSG	PIR	CPDAS	VAGLMVKEI	EYRGNKDS	IGG	VV	TCVIRNLP	-----	LGEPC	FDKLEAMLAHAML	SIPATK	KG	FVFGSG	GGCEV	PGSIHND	PFVSAE	279																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																															
Nostocsp.	-----	DLEG	---	VVD	NTVTLDQVESNIVR	CPD	GELADR	MIELIEQ	TGRQGD	IGG	VV	ECVARNVPKG	-----	LGEPV	FDKLEADIAKAVMS	LPA	SK	GFVFGSG	FAG	LLTGF	EHND	EYIDE	270																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
Synechsp	-----	DIEA	---	VVD	NTVTLDQVESNIVR	CPD	ECAEK	MIERID	QVLRQKDS	IGG	VV	ECVARNVPKG	-----	LGEPV	FDKLEADIAKAVMS	LPA	SK	GFVFGSG	FAG	LLTGF	EHND	EYIDE	270																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
Lesculentum	-----	QVVL	PED	LID	HQNVTL	EQI	ESNIVR	CPD	PEYAEK	MIAA	IDA	VRVGRD	-----	LGTP	VFDKLEAE	LAKACMS	LPA	T	KG	FVFGSG	FAG	FMTG	SEHN	DEFY	270																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																												
Athaliana	-----	HVVL	PEL	VD	HENLTL	EQI	ENNIVR	CPN	PEYAEK	MIAA	IDA	VRVGRD	-----	LGTP	VFDKLEAE	LAKACMS	LPA	T	KG	FVFGSG	FAG	FMTG	SEHN	DEFY	270																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																												
Csempervirens	-----	KVVL	PEL	VD	HETLSL	EQI	ESNIVR	CPD	SEYAEK	MIAA	IDA	VRVGRD	-----	LGSP	VFDKLEAE	LAKACMS	LPA	T	KG	FVFGSG	FAG	FMTG	SEHN	DEFY	270																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																												
Linnocua	-----	TRAN	LKR	DY	AVSEIQE	---	TSEAS	PVR	CLDEVA	AEMMQ	IDD	AKNGD	-----	LG	SYVQ	WDRK	LDA	KIAR	AIVS	INAF	KG	AE	FV	GG	FAG	280																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
Efaecalis	-----	IEA	TIP	EN	LTI	REIQE	---	RSE	QSAVR	VLD	PS	VEK	MKEL	IQ	KK	NGD	T	IGG	VV	EV	L	V	GG	V	PAG	280																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
Spneu_1QX0	-----	KE	IDV	P	EN	LTV	AEIQE	---	RAA	QSE	V	SIVN	QERE	Q	EI	KD	Y	D	Q	I	K	R	G	D	T	IGG	280																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										
Bsubtilis	-----	V	TE	S	P	R	CY	D	E	E	A	G	K	M	M	A	A	I	D	E	A	K	A	N	G	D	S	IGG	282																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																								
Banthracis	-----	V	K	A	K	H	S	N	L	S	I	E	E	I	Q	I	---	I	T	E	N	S	P	V	R	C	L	D	K	T	V	E	---	J	E	M	D	A	I	D	N	A	K	S	G	D	S	IGG	282																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
Saureus	-----	I	K	D	F	Y	D	S	---	E	T	F	K	A	---	N	L	R	N	D	V	R	V	I	D	S	I	A	Q	A	M	R	D	K	I	D	E	A	K	N	E	G	S	IGG	280																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																								
Aaeol_1Q1L	-----	K	E	V	E	L	K	D	K	S	V	F	A	N	P	E	K	L	S	Y	H	E	K	A	D	S	E	L	R	I	P	F	P	E	K	D	E	F	F	T	Y	I	D	E	V	K	E	G	S	L	G	G	F	E	V	F	A	L	N	V	P	P	G	---	L	G	S	H	I	Q	W	D	R	R	I	D	G	R	I	A	Q	A	X	S	I	Q	A	I	K	G	V	E	I	G	L	F	E	A	A	R	R	F	G	S	O	V	H	E	I	G	W	S	E	---	L	G	S	F	T	S	G	D	H	R	L	D	S	Q	L	A	A	V	M	G	I	Q	A	I	K	G	V	E	I	G	D	F	Q	T	A	R	R	R	G	S	R	A	H	D	E	M	P	G	P	---	L	G	S	H	I	S	G	E	D	R	L	D	A	Q	I	A	A	L	M	G	I	Q	A	I	K	G	V	E	I	G	D	F	E	A	R	R	R	G	S	E	A	H	D	E	V	L	D	D	---	L	G	S	H	V	W	D	R	K	L	D	A	R	L	A	G	A	L	M	G	I	Q	A	I	K	G	V	E	I	G	D	F	E	L	A	R	V	P	G	S	R	A	H	D	E	I	V	G	T	P	---	F	V	S	N	L	K	E	E	E	F	D	F	E	F	A	K	---	K	S	E	I	F	C	L	D	P	K	L	E	S	D	F	K	N	E	I	L	N	A	R	N	S	K	D	S	V	G	A	A	V	T	T	V	S	G	X	---	L	I	G	L	G	E	V	L	Y	D	K	L	D	S	K	L	A	H	A	L	X	G	I	N	A	V	K	A	V	E	I	G	E	G	I	N	A	S	K	X	R	G	S	C	N	N	D	A	L	K	D	G	K	---	K	S	E	I	F	A	L	D	E	E	Q	E	E	A	Q	K	T	A	I	Q	N	A	I	K	N	H	S	I	G	G	V	A	L	I	R	A	R	S	I	K	I	N	Q	K	L	P	I	G	L	G	L	Y	A	K	L	D	A	K	I	A	E	A	M	M	G	L	N	G	V	K	A	V	E	I	G	K	G	V	E	S	L	L	K	G	S	E	V	N	D	L	M	D	Q	K	G	---	Q	N	P	F	C	P	D	P	D	K	I	D	A	L	D	E	L	M	R	A	L	K	K	E	G	D	S	I	G	A	K	T	V	V	A	S	G	V	P	A	G	---	L	G	E	P	V	F	D	R	L	D	A	I	A	H	A	L	M	S	I	N	A	V	K	G	V	E	I	G	D	F	V	V	A	L	R	G	S	C	N	R	D	E	I	K	D	G	---	D	V	S	C	D	L	D	W	D	L	V	E	---	Q	N	P	F	C	P	D	A	S	K	L	E	P	L	D	A	L	M	R	E	L	K	K	A	G	D	S	I	G	A	K	T	V	V	A	E	N	V	P	V	G	---	L	G	E	P	V	F	D	R	L	D	A	I	A	H	A	L	M	S	I	N	A	V	K	G	V	E	I	G	D	F	V	V	R	Q	K	G	S	O	H	R	D	P	L	P	Q	G	---	N	N	D	F	C	P	D	V	K	V	A	A	F	D	E	L	I	R	E	L	K	K	E	G	D	S	I	G	A	K	T	V	V	A	T	G	V	P	V	G	---	L	G	E	P	V	F	D	R	L	D	A	I	A	H	A	L	M	S	I	N	A	V	K	G	V	E	I	G	D	F	V	V	R	Q	K	G	S	O	H	R	D	P	L	P	Q	G	---	I	K	L	A	P	Q	K	V	G	Q	I	D	W	E	K	V	N	---	S	N	P	F	C	P	D	S	A	V	E	K	F	D	E	L	I	R	E	L	K	K	E	G	D	S	I	G	A	K	L	T	V	I	A	E	N	V	P	V	G	---	L	G	E	P	V	F	D	R	L	D	A	I	A	H	A	L	M	S	I	N	A	V	K	G	V	E	I	G	D	F	V	V	R	Q	K	G	S	O	H	R	D	P	L	P	Q	G	---	K	N	T	L	T	I	P	S	I	D	N	T	Y	I	N	V	K	T	N	E	C	N	I	N	Q	V	D	N	N	H	N	N	I	N	D	K	D	N	T	F	N	N	S	E	K	S	D	E	W	I	Y	L	Q	T	R	C	P	H	P	T	A	V	Q	I	C	S	Y	L	K	L	K	N	G	D	S	V	G	G	I	A	T	C	I	Q	N	P	P	I	G	---	I	G	E	P	I	F	D	K	L	E	A	E	L	A	K	M	I	L	S	I	P	P	V	K	G	I	E	F	G	S	G	F	N	G	Y	M	F	G	S	M	H	D	I	F	I	P	V	E	---	A	D	G	T	P	I	D	G	N	P	T	D	D	P	L	A	N	A	R	E	R	L	S	D	P	G	E	C	T	L	L	Q	R	C	P	H	L	A	V	K	I	C	S	Y	L	K	L	K	A	G	D	S	IGG	V	A	T	C	I	A	H	N	V	P	V	G	---	I	G	E	P	I	F	E	K	M	E	A	E	L	G	K	I	L	S	I	P	A	V	K	G	I	E	F	G	S	G	F	D	G	Y	M	L	G	S	D	H	N	D	L	F	G	T	L	D	---	.....	310	.....	320	.....	330	.....	340	.....	350	.....	360	.....	370	.....	380	.....	390	.....	400	.....	410	.....	420	.....	430	.....	440	.....	450







# CLUSTAL X (1.81) MULTIPLE SEQUENCE ALIGNMENT

File: C:Corismatoalig\_asta\_final5.ps

Date: Wed Jul 26 11:30:24 2006

Page 5 of 5

```

      *           : .           :
Spombe  AKGRHDP CVVPR AIP IVEAMAALVVM DAVMIQQSRIASRNLLPNAQ----- 395
Ncrassa AKGRHDF SVVPR AVF IVEAMAALVIM DAVLAQQARHTAKSLLPPLKQINS GKDTV GNGVSENVG ESDLAQ 432
Scerev_1R53 AKGRHDP AVTPRAIP IVEAMTALVLADALLIQKARDFSRSVVH----- 376
Nostocsp. AKGRHDP CVLPR AVPMVEAMVALVLC D HLLRH HGQCKVL----- 362
Synechsp. AKGRHDP CVLPR AVPMVEAMAALVLC D HLLRFQAQCKTL----- 362
Lesculentum ARGRHDP CVVPR AVPMVEAMVALVLDQLMAQV SQMMFFINPELQEP LQSSPESAEVTL----- 440
Athaliana ARGRHDP CVVPR AVPMVEAMVALVLDQLMAQV AQCHLFFINPELQEP LQIEQPQNAIATL----- 436
Csempervirens ARGRHDP CVVPR AVPMVEAMVALVLLDQLMLQHAQGNLFSINPALQEP LSETVSSAAASLQGV----- 447
Linnocua  SVERSDSCAVPAASVVAEAVVAWEVAVVLEKFDGDRFDTLKKHV EERHNLTK E F----- 388
Efaecalis  SVERSDS TAVPAASVVC EAVVATEVAKAMLEKFDSDSFEQMKEAVKRYRLYTQNF----- 388
Spneu_1QX0 TVERSDP T ALPAAGMVEAVVATVLAQEILEKFS DNLEELKEAVAKHRDYTKNY----- 388
Bsubtilis  SIERSDSCAVPAASVVAEALSLGKLP SLNNSD----- 368
Banthracis SIERSDSCAVPAAGVVAESVVAWE LAHALVEQFGKDRMELIQQNI TQH NKYAK E F----- 390
Saureus    TIERSDSCAVPAASIVCEHVVAFEIAKALLEFQSNHIEQLQQQIADRRLQNLV E F----- 388
Aaeol_1Q1L GKERTDIVAVPAASVVG EAXLAIVLADALLEKLGGDFXEEVKKRFEDV VNHVKS F----- 401
Mtuberculosis IHQRSDVCAVPAAGVVVE T MVALVLARAAL EKF GGD SLAETQRNLAAYQRSVADREAPAA RVS G----- 401
Cglutamicum IHQRSDVCAVPAAGVVAEAMVTLVLARAVLQKFGD SLSETKSNIDT LKNIEERMKFEGLDGA----- 410
Scoelicolor HHQRSDVSAVPAAGIVAEAMVALVLADAVA EKF GGD SVTE RRRNVRSLDNLDIR----- 394
Cjejunii_1SQ1 LKGRHDP CVGVRG SVVASXVRLVLADCLLLNASANLNNLKNAYGLKLEHHHHHH----- 370
Hpilory_1UMF LKGRHDP CIAIRGSVVCESLLALVLADMVLLNLT SKIEYLKTIYNEN----- 365
Ecoli      TKGRHDP CVGIRAVPIAEAMLAIVLMDHLLRQRAQNDVKTIDIPRW----- 361
Ypestis    TRGRHDP CVGIRAVPIAEAMMAIVLMDHLLRQRAQC GDVASDVPRW----- 362
Vcholerae  TKGRHDP CVGIRAVPIAEAMLAIVLMDHLLRHRGQNGVVITTPKI----- 361
Hinfluenzae TKGRHDP CVGIRAVPIAEAMVAIVLLD HLLRFKAQCK----- 357
Pfalco    VQGRHDS CILPRLPPIIEASSMVIGDLILRQISKYGD KKLPTLFRNM----- 527
Pvivirus   VTGMHDC CILPRLPPIIEAS SIVIGDMILRQVAKYGHSKLPSVGLSRSTRE----- 547
ruler     .....610.....620.....630.....640.....650.....660.....670.

```



## Anexo 2:

Arquivos de entrada para as minimizações de energia:

⇒ Modelo sem restrição.

### mini\_01.in

```
# Restrained Minimization.NTF=1, NTC=1. NTR=1.
&cntrl  imin=1,
        ntp=50,
        ntf=1,dielc=1.0, igb=1, scnb=2.0, scee=1.2,
        ibelly=0, ntr=1, cut=12.0,
        ntc=1, tol=0.00001, ntb=0
        maxcyc=500, ncyc=100, ntmin=1, dx0=0.01, dxm=0.5, drms=0.0001,
&end
Restraining Protein
1000.000
RES 1 400
END
Restraining FMN EPS
1000.000
RES 401 402
END
END
```

### mini\_02.in

```
# Restrained Minimization.NTF=1, NTC=1.NTR=1.
&cntrl  imin=1,
        ntp=50,
        ntf=1,dielc=1.0, igb=1, scnb=2.0, scee=1.2,
        ibelly=0, ntr=1, cut=12.0,
        ntc=1, tol=0.00001, ntb=0
        maxcyc=500, ncyc=100, ntmin=1, dx0=0.01, dxm=0.5, drms=0.0001,
&end
Restraining Protein
500.000
RES 1 400
END
Restraining FMN EPS
1000.000
RES 401 402
END
END
```

### mini\_03.in

```
# Restrained Minimization.NTF=1, NTC=1.NTR=1.
&cntrl  imin=1,
        ntp=50,
        ntf=1,dielc=1.0, igb=1, scnb=2.0, scee=1.2,
        ibelly=0, ntr=1, cut=12.0,
```

```
        ntc=1, tol=0.00001,ntb=0
        maxcyc=500, ncyc=100, ntmin=1, dx0=0.01, dxm=0.5, drms=0.0001,
&end
Restraining Protein
250.000
RES 1 400
END
Restraining FMN EPS
1000.000
RES 401 402
END
END
```

#### **mini\_04.in**

```
# Restrained Minimization.NTF=1, NTC=1.NTR=1.
&cntrl  imin=1,
        ntp=50,
        ntf=1,dielc=1.0, igb=1, scnb=2.0, scee=1.2,
        ibelly=0, ntr=1, cut=12.0,
        ntc=1, tol=0.00001,ntb=0
        maxcyc=500, ncyc=100, ntmin=1, dx0=0.01, dxm=0.5, drms=0.0001,
&end
Restraining Protein
100.000
RES 1 400
END
Restraining FMN EPS
1000.000
RES 401 402
END
END
```

#### **mini\_05.in**

```
# Restrained Minimization.NTF=1, NTC=1.NTR=1.
&cntrl  imin=1,
        ntp=50,
        ntf=1,dielc=1.0, igb=1, scnb=2.0, scee=1.2,
        ibelly=0, ntr=1, cut=12.0,
        ntc=1, tol=0.00001,ntb=0
        maxcyc=500, ncyc=100, ntmin=1, dx0=0.01, dxm=0.5, drms=0.0001,
&end
Restraining Protein
50.000
RES 1 400
END
Restraining FMN EPS
1000.000
RES 401 402
END
END
```

#### **mini\_06.in**

```

# Restrained Minimization.NTF=1, NTC=1.NTR=1.
&cntrl  imin=1,
        ntp=50,
        ntf=1,dielc=1.0, igb=1, scnb=2.0, scee=1.2,
        ibelly=0, ntr=1, cut=12.0,
        ntc=1, tol=0.00001,ntb=0
        maxcyc=500, ncyc=100, ntmin=1, dx0=0.01, dxm=0.5, drms=0.0001,
&end
Restraining Protein
10.000
RES 1 400
END
Restraining FMN EPS
1000.000
RES 401 402
END
END

```

### **mini\_07.in**

```

# Unrestrained Minimization.NTF=1, NTC=1.
&cntrl  imin=1,
        ntp=50,
        ntf=1,dielc=1.0, igb=1, scnb=2.0, scee=1.2,
        ibelly=0, ntr=1, cut=12.0,
        ntc=1, tol=0.00001, ntb=0
        maxcyc=1000, ncyc=100, ntmin=1, dx0=0.01, dxm=0.5, drms=0.0001,
&end
Restraining FMN EPS
1000.000
RES 401 402
END
END

```

⇒ Modelo com restrição: foram usados os mesmos parâmetros para a minimização de energia da proteína nos 7 estágios, porém a minimização dos ligantes FMN e EPS foi mantida em 10000.000 para todos os estágios.

```

Restraining FMN EPS
10000.000

```