

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
FACULDADE DE BIOCÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR**

KELEN BEIESTORF ROCHA

***VIRTUAL SCREENING E DINÂMICA MOLECULAR
PARA IDENTIFICAÇÃO DE INIBIDORES
DA ENZIMA CORISMATO SINTASE
DE *MYCOBACTERIUM TUBERCULOSIS****

Porto Alegre, RS

2011

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
FACULDADE DE BIOCÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR**

***VIRTUAL SCREENING* E DINÂMICA MOLECULAR
PARA IDENTIFICAÇÃO DE INIBIDORES
DA ENZIMA CORISMATO SINTASE
DE *MYCOBACTERIUM TUBERCULOSIS***

Dissertação apresentada ao Programa
de Pós-Graduação em Biologia
Celular e Molecular como requisito
para obtenção do grau de Mestre.

Autora

Kelen Beiestorf Rocha

Orientador

Prof. Dr. Walter Filgueira de Azevedo Júnior

Porto Alegre, RS

Abril, 2011

Dedico este trabalho aos meus pais,
exemplos de amor e vida.

AGRADECIMENTOS

Agradeço a todos aqueles que direta ou indiretamente contribuíram para este trabalho.

Em especial aos meu pais, pelo incentivo e amor incondicional e por terem sido sempre, meu maior exemplo de caráter, luta e vida.

Ao meu namorado pelo tempo de ausência, pelo carinho e o apoio em todos os momentos.

Ao meu orientador, pelo aprendizágio e pelo meu treinamento científico.

E aos meu colegas de laboratório pela companhia, amizade e disponibilidade.

SUMÁRIO

DEDICATÓRIA.....	i
AGRADECIMENTOS	ii
SUMÁRIO.....	iii
LISTA DE SIGLAS E ABREVIATURAS	iv
LISTA DE FIGURAS	vi
LISTA DE TABELAS	vii
LISTA DE GRÁFICOS	viii
RESUMO.....	ix
ABSTRACT.....	x
.....
1. INTRODUÇÃO	1
1.1 Tuberculose	1
1.2 Via do ácido chiquímico	6
1.3 Corismato sintase	9
1.3.1 Cofator e substrato	11
1.4 Desenvolvimento de drogas e a abordagem computacional	13
1.4.1 Modelagem molecular por homologia	13
1.4.2 <i>Docking</i> molecular	14
1.4.3 Dinâmica molecular	18
2. OBJETIVOS	20
2.1 Objetivo geral.....	20
2.2 Objetivos específicos	20
3. ARTIGO CIENTÍFICO	21
4. RESULTADOS E DISCUSSÃO	64
4.1 Construção do complexo ternário <i>MtCS:FMN:EPSP</i>	64
4.1.1 Modelagem por homologia	65
4.1.2 <i>Docking</i> geométrico	69
4.1.2.1 Análise os sítios ativos do FMN e do EPSP	70
4.2 <i>Redocking</i> da estrutura cristalográfica.....	75
4.3 <i>Virtual screening</i>	77
4.3.1 Ligante A.....	78
4.3.2 Ligante B.....	81
4.3.3 Ligante C	83
4.4 Dinâmica molecular	87
4.4.1 Flexibilidade e estabilidade das estruturas apo	89
4.4.2 Estabilidade dos ligantes no sítio ativo e principais resíduos responsáveis por esta estabilização	95
5. CONCLUSÕES E PERSPECTIVAS	101
6. REFERÊNCIAS	103
7. ANEXOS	109

LISTA DE SIGLAS E ABREVIATURAS

- TB** - Tuberculose
Mt - *Mycobacterium tuberculosis*
BCG - *Bacillus Calmette-Guérin*
OMS - Organização Mundial da Saúde
DOTS - *Directly Observed Treatment Short Course* (Tratamento Diretamente Supervisionado de Curta Duração)
AIDS - *Acquired Immunodeficiency Syndrome* (Síndrome da Imunodeficiência Adquirida)
HIV - *Human Immunodeficiency Virus* (Vírus da Imunodeficiência Humana)
TB-MDR - TB-Multi-Resistente a Drogas
TB-XDR - TB-Extensivamente Resistente a Drogas
SRI - Síndrome da Reconstituição Imune
CS - Corismato Sintase
VS - *Virtual Screening* (Triagem Virtual)
DM - Dinâmica Molecular
DAHPS - 3-Deoxi-D-Arabino-Heptulose 7-Fosfato Sintase
DHQS - 3-Desidroquinato Sintase
DHQD - 3-Desidroquinato Desidratase
SD - Chiquimato Desidrogenase
SQ - Chiquimato Quinase
EPSPS - 5-Enolpiruvilchiquimato-3-Fosfato Sintase
cDNA - *Complementary - Deoxyribonucleic Acid* (Complementares - Ácido Desoxirribonucleico)
E4P - Eritrose-4-Fosfato
PEP - Fosfoenolpiruvato
PABA - Ácido p-Aminobenzóico
EPSP - 5-Enolpiruvilchiquimato-3-Fosfato
FMNH₂ ou FMN_{red} - Flavina Mononucleotídeo Reduzida
FMN - Flavina Mononucleotídeo Oxidada
FAD - Flavina Adenina Dinucleotídeo
NADPH - Nicotinamida Adenina Dinucleotídeo Fosfato Reduzida
PDB - *Protein Data Bank* (Banco de Dados de Proteínas)
3D - Tridimensional
NMR - Ressonância Magnética Nuclear
RMSD - *Root Mean Square Deviation* (Desvio Médio Quadrático)
MtPNP - Purina Nucleosídeo Fosforilase de *Mt*
ADMETox - Absorção, Distribuição, Metabolismo, Excreção e Toxicidade
GNU - *General Public License* (Licença Pública Geral)
SPDBV - Swiss PDB Viewer

NCBI - *National Center for Biotechnology Information* (Centro Nacional para Informações de Biotecnologia)

SAVES - *Structural Analysis and Verification Server* (Servidor de análise e verificação estrutural)

NIH MBI Laboratory for Structural Genomics and Proteomics - *National Institute Health - Molecular Biology Institute - Laboratory for Structural Genomics and Proteomics* (Instituto Nacional de Saúde – Instituto de Biologia Molecular - Laboratório de Genômica e Proteômica Estrutural)

SpCS - Corismato sintase de *Streptococcus pneumoniae*

IUPAC - *International Union of Pure and Applied Chemistry* (União Internacional de Química Pura e Aplicada)

SPC/E - *Extended Simple Point Charge*

PME - *Particle-Mesh Ewald*

RMFS - *Root Mean-Square Fluctuation* (Flutuação dos Desvio Quadrático Médio)

LISTA DE FIGURAS

Figura 1 – Distribuição geográfica mundial de novos casos de tuberculose estimadas pela OMS no ano de 2007	5
Figura 2 – Distribuição geográfica no Brasil de notificação de TB - casos novos e recidivas, 2007	5
Figura 3 – Via metabólica do ácido chiquímico.....	8
Figura 4 – Reação enzimática mediada pela corismato sintase.....	9
Figura 5 – Representação da estrutura secundária do monômero da <i>MtCS</i>	10
Figura 6 – Representação da estrutura química do EPSP e do FMN.....	11
Figura 7 – Representação de fitas da estrutura 3D de <i>MtCS</i> na região modelada sobreposta a estrutura cristalográfica que serviu como molde.....	66
Figura 8 – Gráficos de Ramachadran da estrutura modelada e do molde.....	68
Figura 9 – Representação do EPSP e do resíduo ARG 341 antes e após minimização de energia.....	70
Figura 10 – Representação de fitas da estrutura 3D de <i>MtCS</i> dando ênfase ao sítio ativo da enzima	71
Figura 11 – Ligplot demonstrando as ligações de hidrogênio e os contatos hidrofóbicos feitos pelo EPSP com a estrutura de <i>SpCS</i> e <i>MtCS</i>	73
Figura 12 – Estrutura da enzima <i>MtCS</i> -FMN-EPSP e do espaço 3D que delimita as simulações <i>docking</i> no sítio catalítico do EPSP e representação do resultado de <i>redocking</i> – sobreposição da estrutura cristalográfica do EPSP	76
Figura 13 – Ligplot - demonstrando as ligações de hidrogênio e os contatos hidrofóbicos feitos pelo ligante A com a enzima <i>MtCS</i>	80
Figura 14 – Ligplot - demonstrando as ligações de hidrogênio e os contatos hidrofóbicos feitos pelo ligante B com a enzima <i>MtCS</i>	82
Figura 15 – Ligplot - demonstrando as ligações de hidrogênio e os contatos hidrofóbicos feitos pelo ligante C com a enzima <i>MtCS</i>	84
Figura 16 – Representação da estrutura química e da mesma estrutura docada e sobreposta ao EPSP do ligante A, ligante B e ligante C.....	85
Figura 17 – Representação em <i>ribbon</i> das estruturas iniciais e finais sobrepostas da DM	94

LISTA DE TABELAS

Tabela 1 – Informações sobre as estruturas de CS depositadas no PDB.....	12
Tabela 2 – Resultado da predição da estrutura secundária dos resíduos modelados gerados pelo <i>web server</i> APSSP2.	67
Tabela 3 – Resultado dos gráficos estatísticos de Ramachadran do modelo e do molde.....	68
Tabela 4 – Resíduos localizados até 4 Å nos sítios ativos da FMN e EPSP das estruturas cristalográficas de <i>SpCS</i> (1QXO) com 388 aminoácidos e <i>MtCS</i> (modelo final - modelada e minimizada) com 392 aminoácidos.....	74
Tabela 5 – Tabela demonstrando com quais resíduos da enzima cada molécula docada faz interação, que tipo de interação e entre quais átomos.....	86
Tabela 6 – Informações físico-químicas da estruturas analisadas.....	86

LISTA DE GRÁFICOS

Gráfico 1 – Representação gráfica do RMSD de todos os C α das estruturas do modelo de partida em função do tempo – estruturas apo.....	90
Gráfico 2 – Representação gráfica do raio de giro das estruturas do modelo de partida em função do tempo – estruturas apo	91
Gráfico 3 – Representação gráfica do RMSF das estruturas do modelo de partida em relação aos resíduos – estruturas apo.....	93
Gráfico 4 – Representação gráfica do RMSD de todos os C α das estruturas do modelo de partida em função do tempo – complexos com ligantes	95
Gráfico 5 – Representação gráfica do raio de giro das estruturas do modelo de partida em função do tempo – complexos com ligantes.....	96
Gráfico 6 – Gráfico mostrando os principais resíduos da proteína que fizeram ligação de hidrogênio com os ligantes em função do número de vezes em que estes resíduos apareceram nos <i>snapshots</i> extraídos a cada 1 ns (do 1° ao 18°ns) das trajetórias das simulações de DM.....	97
Gráfico 7 – Representação gráfica do RMSF das estruturas do modelo de partida em relação aos resíduos – de todos os complexos	98
Gráfico 8 – Representação gráfica do RMSF das estruturas do modelo de partida em relação aos resíduos – de cada complexo	99

RESUMO

O aumento da incidência de cepas resistentes de *Mycobacterium tuberculosis*, aliados a co-infecção com o vírus da imunodeficiência humana e a ausência de novas terapias anti-tuberculose nos últimos anos, evidenciam a necessidade da descoberta de novos agentes terapêuticos para o tratamento da tuberculose.

Considerando, a partir estudos anteriores, que a via metabólica do ácido chiquímico do *Mycobacterium tuberculosis* é essencial para sobrevivência do organismo, a predição de inibidores para corismato sintase, sétima enzima desta rota, abre a possibilidade de desenvolvimento de novas quimioterapias anti-micobacterianas.

Neste trabalho, através de técnicas computacionais de modelagem molecular, *virtual screening* e dinâmica molecular, foi possível elucidar aspectos moleculares importantes da interação da corismato sintase com seu substrato e cofator, bem como, propor três potenciais inibidores para esta enzima, contribuindo como início de uma pesquisa sobre compostos com potencial farmacológico anti-tuberculose.

Palavras-chaves: corismato sintase – *virtual screening* – dinâmica molecular - *Mycobacterium tuberculosis* - via do ácido chiquímico - tuberculose

ABSTRACT

The increasing incidence of resistant strains of *Mycobacterium tuberculosis*, combined with co-infection of the human immunodeficiency virus and the absence of new anti-tuberculosis therapy in recent years highlight the need for discovery of new therapeutic agents for the treatment of tuberculosis.

Seeing from previous studies that shikimate pathway of *Mycobacterium tuberculosis* is essential for survival of the organism, the prediction of inhibitors for chorismate synthase, the seventh enzyme of this route, open up the possibility of development of new anti-mycobacterial chemotherapy.

In this work, using computational techniques of molecular modeling, virtual screening, and molecular dynamics, was possible to elucidate the molecular interaction of chorismate synthase with its substrate and cofactor, and propose three potential inhibitors for this enzyme, contributing to early research on compounds with potential anti-tuberculosis.

Key words: chorismate synthase - virtual screening – molecular dynamics -
Mycobacterium tuberculosis - shikimate pathway – tuberculosis

1. INTRODUÇÃO

1.1 Tuberculose

O aumento e a propagação de cepas resistentes a antibióticos é um dos mais sérios problemas de saúde pública mundial da última década (Amer *et al.*, 2008). Neste contexto insere-se a tuberculose (TB), uma doença infecto-contagiosa, causada principalmente pelo bacilo *Mycobacterium tuberculosis* (Mt).

A incidência de tuberculose teve um rápido declínio no início do século XX nos países emergentes, principalmente devido a melhorias em saneamento e habitação. Esta tendência foi acelerada inicialmente pela introdução da vacina da cepa atenuada do *Mycobacterium bovis* - Bacilo Calmette-Guérin (BCG) (Parish & Stoker, 2002; Ducati *et al.*, 2006) e pela descoberta das propriedades antibacterianas e antituberculares da estreptomicina (1944), isoniazida e pirazinamida (1952) e a introdução de etambutol e rifampicina (Basso *et al.*, 2005; Ducati *et al.*, 2006).

Entretanto, nas últimas décadas, o ressurgimento da doença está acontecendo e de forma mundial. Em 1993, a gravidade da situação levou a Organização Mundial de Saúde (OMS) a declarar a tuberculose como uma emergência global na tentativa de aumentar a consciência pública e política para este problema (Parish & Stoker, 2002; WHO, 2009). Além disso, propôs também estratégias para otimizar a resposta e a aderência ao tratamento, como a implantação do programa de Tratamento Diretamente Supervisionado de Curta Duração - DOTS (*Directly Observed Treatment Short Course*) e a quimioterapia anti-TB padronizada (WHO, 2009).

Os últimos relatórios divulgados em 2009, relataram 9,27 milhões novos casos (Figura 1) e, aproximadamente, 2,7 milhões de mortes por tuberculose entre os anos

de 2007 e 2008, posicionando a doença, em relação à incidência e mortalidade, atrás apenas da AIDS (*Acquired Immunodeficiency Syndrome*), entre as patologias infecciosas (WHO, 2009).

No Brasil, que ocupa o 13º lugar entre os 22 países com maior índice de tuberculose no mundo, o empenho do governo na promoção de serviços sociais tem ampliado a visibilidade da TB como um problema de saúde pública. A expansão dos serviços de DOTS tem progredido e medidas de controle da patologia têm priorizado 315 de um total de 5.565 municípios que representam 70% dos casos de tuberculose do país (Figura 2). Concomitantemente, iniciativas especiais de controle da doença em grupos vulneráveis, como populações indígenas e de prisioneiros, tem sido executadas em colaboração com as organizações não governamentais. Apesar dos progressos realizados na contenção/erradicação da tuberculose, as taxas de detecção de casos e o sucesso do tratamento no país ainda estão abaixo das metas globais (WHO, 2009).

Entre os fatores responsáveis pela consolidação deste cenário, dois principais se destacam. O primeiro é a ocorrência da tuberculose em pacientes infectados com HIV (*Human Immunodeficiency Virus*), que possuem maior susceptibilidade de desenvolver a doença devido a um sistema imunológico comprometido. Segundo a OMS, cerca de 1,37 milhões ou 14,8% do número total de casos de incidência de TB em todo o mundo no ano de 2007, eram HIV-positivos. O número de óbitos neste grupo foi de 456 mil (equivalente a 26% das mortes por tuberculose em indivíduos HIV-positivos e HIV-negativos, e 23% dos 2 milhões de mortes relacionadas com o HIV) (WHO, 2009).

O outro fator é o surgimento de cepas resistentes aos antimicrobianos utilizados no tratamento, como a isoniazida e a rifampicina, em consequência,

principalmente, das terapias inadequadas e do uso indiscriminado destes antibióticos (Basso *et al.*, 2005; Da Silveira *et al.*, 2005). De acordo com a OMS, em 2007, foram aproximadamente 0,5 milhões de casos de TB multi-resistente a drogas (TB-MDR) em todo o mundo, sendo que até ao final de 2008, 55 países e territórios já haviam relatado pelo menos um caso de tuberculose extensivamente resistente aos medicamentos (TB-XDR) (WHO, 2009).

Esse cenário assustador é diretamente relacionada ao tratamento da doença. A terapêutica atualmente recomendada pela OMS, combina a administração de isoniazida, rifampicina, pirazinamida e etambutol (ou estreptomicina), as chamadas drogas de primeira linha (Da Silveira *et al.*, 2005; Raman *et al.*, 2008) durante os primeiros 2 meses, e um adicional de quatro meses com uma combinação de isoniazida e rifampicina (Basso *et al.*, 2005; Ducati *et al.*, 2006). No entanto, o longo período de tratamento e os efeitos colaterais indesejáveis, acabam levando a interrupção prévia da quimioterapia. Além disso, a incompatibilidade entre as quimioterapias dos tratamentos TB/HIV, como as interações medicamentosas entre rifampicina e os inibidores da protease, o risco da Síndrome da Reconstituição Imune (SRI) (Raman *et al.*, 2008) e as implicações relacionadas à incapacidade destas drogas atuarem sobre a forma latente da bactéria (Basso *et al.*, 2005), agravam ainda mais a situação.

Além destes problemas acerca das medicações já existentes para tuberculose, a indústria farmacêutica, concentrada principalmente em países desenvolvidos, reluta em investir na pesquisa de doenças negligenciadas e consideradas de países pobres, por causa do retorno financeiro que não é garantido (Dias *et al.*, 2007a).

Em face a estes dados, bem como o fato de nenhuma nova droga foi desenvolvida nas últimas décadas para essa patologia, o desenvolvimento de

antibacterianos direcionados para alvos em potencial, não somente é um processo contínuo, mas também neste momento, uma urgente necessidade.

Neste trabalho enfocamos a corismato sintase (CS) de *Mt*, uma enzima que participa da via do ácido chiquímico e se constitui um importante alvo molecular. Aqui são discutidos os aspectos estruturais e moleculares desta proteína, bem como de possíveis compostos inibidores identificados e analisados por técnicas computacionais como o *virtual screening* (VS) ou triagem virtual de ligantes e a simulação de dinâmica molecular (DM).

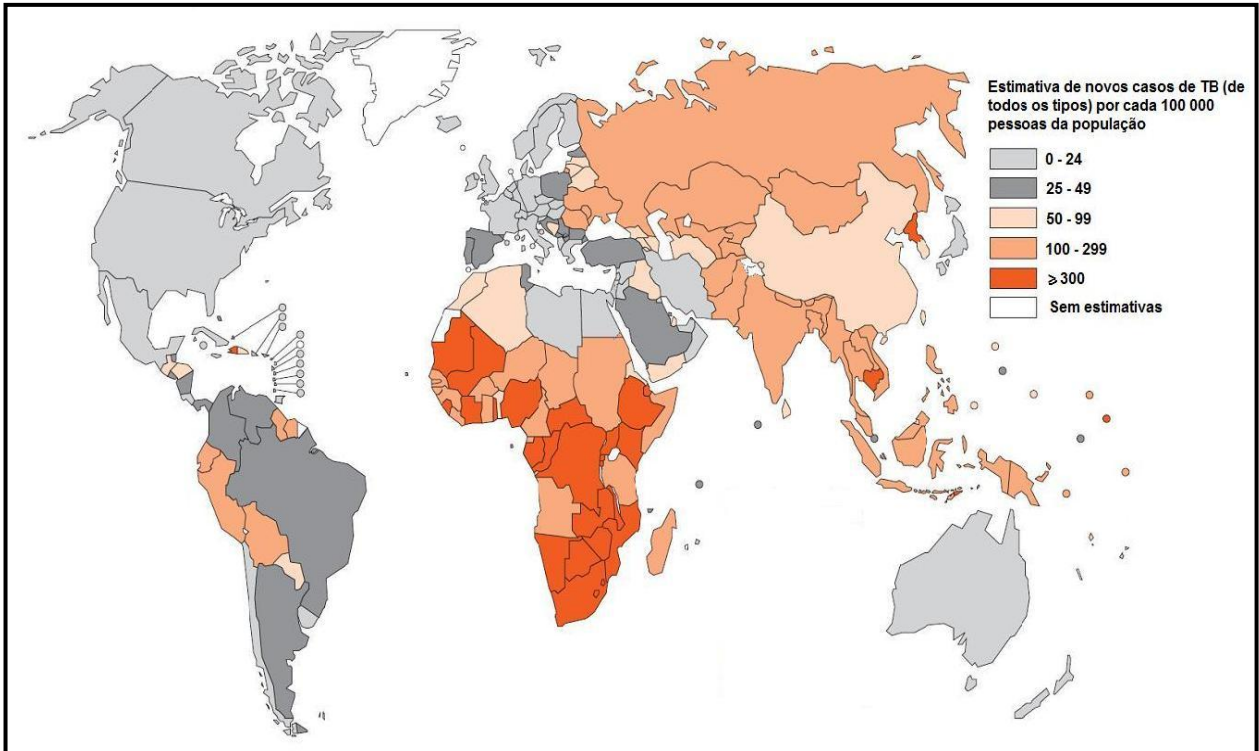


FIGURA 1. Distribuição geográfica mundial de novos casos de tuberculose estimadas pela OMS no ano de 2007. Destaque para os países da África e leste da Europa (WHO, 2009).

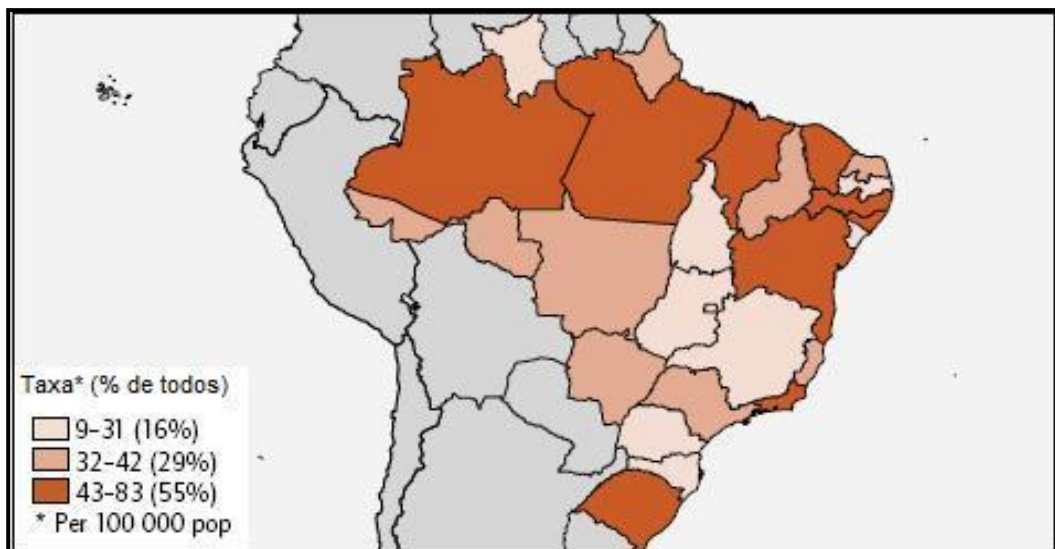


FIGURA 2. Distribuição geográfica no Brasil (em porcentagem) de notificação de TB - casos novos e recidivas, 2007 (WHO, 2009).

1.2 Via do ácido chiquímico

A caracterização estrutural de enzimas que compõem os processos metabólicos de patógenos é fundamental para o desenvolvimento de fármacos baseado na estrutura (Da Silveira *et al.*, 2005), principalmente se estas enzimas fazem parte apenas de rotas metabólicas de patógenos e sejam ausentes em hospedeiros, o que as confere a propriedade de toxicidade seletiva. Desta forma, uma via sintética que tem despertado um grande interesse como alvo para desenho de agentes antimicrobianos e herbicidas, é a via do ácido chiquímico ou chiquimato (Alibhai & Stallings, 2001; Schönbrunn *et al.*, 2001; Park *et al.*, 2004).

Esta, por sua vez, é uma rota em comum para a produção de vários metabólitos incluindo ácido fólico, vitamina K, ubiquinona, fenilalanina, triptofano e tirosina em bactérias, fungos, algas, plantas e parasitas do filo Apicomplexa. Entretanto, ela encontra-se ausente em mamíferos, que dependem da dieta para obtenção destes compostos (Da Silveira *et al.*, 2005; Dias *et al.*, 2006).

A via de biossíntese do ácido chiquímico foi identificada em *Mycobacterium tuberculosis* H37Rv (Roberts *et al.*, 1998) após o sequenciamento do genoma do bacilo (Cole *et al.*, 1998) e do reconhecimento dos genes que codificam as sete enzimas que a compõem: *aroG* (3-deoxi-D-arabino-heptulosonato 7-fosfato sintase ou DAHPS), *aroB* (3-desidroquinato sintase ou DHQS), *aroD* (3-desidroquinato desidratase ou DHQD), *aroE* (chiquimato desidrogenase ou SD), *aroK* (chiquimato quinase ou SQ), *aroA* (5-enolpiruvilchiquimato-3-fosfato sintase ou EPSPS) e *aroF* (corismato sintase) (Parish & Stoker, 2002).

Algumas destas enzimas de *Mt*, como a DAHPS, a CS, a EPSPS e a SQ, tiveram seus cDNAs (*Complementary-Deoxyribonucleic Acid*) clonados e foram

expressas e purificadas com as respectivas atividades (Oliveira *et al.*, 2001; 2003; Rizzi *et al.*, 2005). Outras também já foram cristalizadas e tiveram suas estruturas determinadas (De Azevedo *et al.*, 2002; Pereira *et al.*, 2003; 2004; Dias *et al.*, 2004, 2006, 2007b; Oliveira *et al.*, 2006).

Composta por sete passos enzimáticos, a rota metabólica do ácido chiquímico catalisa seqüencialmente a conversão de dois produtos do metabolismo de carboidratos, a eritrose-4-fosfato (E4P) e o fosfoenolpiruvato (PEP), em corismato ou ácido corísmico (Figura 3).

Este produto final, como o seu próprio nome em grego diz (corismato = separação, divisão, divórcio) (Dias *et al.*, 2007a; De Azevedo *et al.*, 2009) é o principal ponto de ramificação e precursor da síntese e vários metabólitos, tais como o PABA (ácido p-aminobenzóico), o ácido p-hidroxibenzóico (precursor da coenzima Q), as micobactinas e os aminoácidos aromáticos fenilalanina (Phe), triptofano (Trp) e tirosina (Tyr) (Dias *et al.*, 2006; Fernades *et al.*, 2007).

Com o intuito de demonstrar experimentalmente a importância desta via para o bacilo, foi realizada o *knockout* do gene *aroK*, que codifica a enzima chiquimato quinase. Os resultados obtidos apresentaram fortes evidências que a via do ácido chiquímico é essencial para a viabilidade do *Mt* (Parish & Stoker, 2002), tornando, desta forma, as enzimas que a compõem, alvos atrativos para o desenvolvimento de drogas anti-tuberculose.

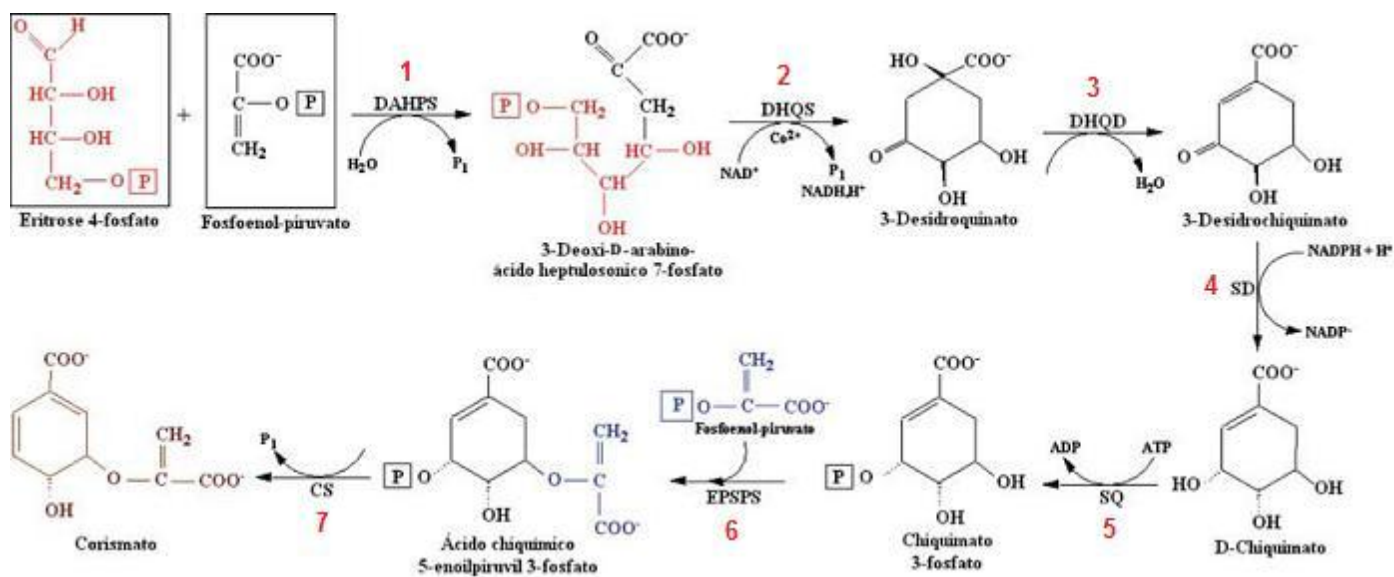


FIGURA 3. Via metabólica do ácido chiquímico. Numeradas em vermelho, as sete enzimas que catalizam as respectivas etapas (Da Silveira *et al.*, 2005).

1.3 Corismato sintase

A corismato sintase (EC 4.2.3.5) catalisa o último passo da rota, convertendo o substrato 5-enolpiruvilchiquimato-3-fosfato (EPSP) em corismato na presença de um cofator, uma flavina mononucleotídeo reduzida (FMN_{red} ou FMN_{red}) que não é consumida durante a reação (Kitzing *et al.*, 2004; Fernandes *et al.*, 2007). Este produto é um precursor não-aromático comum para a biossíntese de uma classe importante de metabólitos aromáticos e representa o maior ponto de bifurcação de outras vias do patógeno (Amer *et al.*, 2008).

A reação enzimática da CS (Figura 4) é descrita como uma reação de eliminação *anti*-1,4 do grupo 3-fosfato e perda do hidrogênio C(6*proR*) a partir do EPSP (Bornemann *et al.*, 2000). Esta reação catalítica é a única conhecida com tal transformação em sistemas biológicos e exatamente por isso, intensamente estudada (Macheroux *et al.*, 1996; Bornemann *et al.*, 1996, 2000; Osborne *et al.*, 2000; Kitzing *et al.*, 2004; Rauch *et al.*, 2007) fazendo com que esta seja uma enzima de natureza singular (Dias *et al.*, 2006).

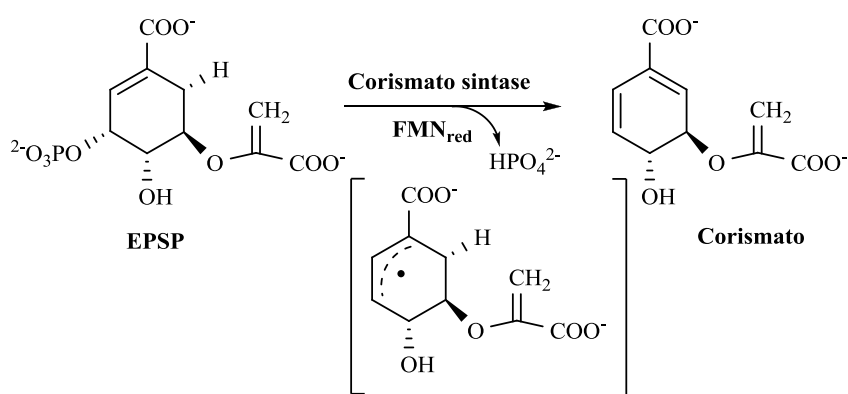


FIGURA 4. Reação enzimática mediada pela corismato sintase.

A estrutura da enzima pertence a família α - β e apresenta uma topologia dominante do tipo β - α - β *sandwich*, formando um homotetrâmero composto por um dímero de dímeros. Cada monômero (Figura 5) é formado por quatro hélices- α na região central, $\alpha 1$, $\alpha 5$, $\alpha 11$ e $\alpha 8$, e duas folhas compostas por 4 fitas- β antiparalelas cada, $\beta 1$, $\beta 2$, $\beta 7$ e $\beta 4$, e $\beta 8$, $\beta 9$, $\beta 15$ e $\beta 10$, respectivamente. Os parâmetros da cela unitária foram estabelecidos como $a = b = 129,7$, $c = 156,8$ Å e o peso molecular de 41.792 Da (Dias *et al.*, 2006).

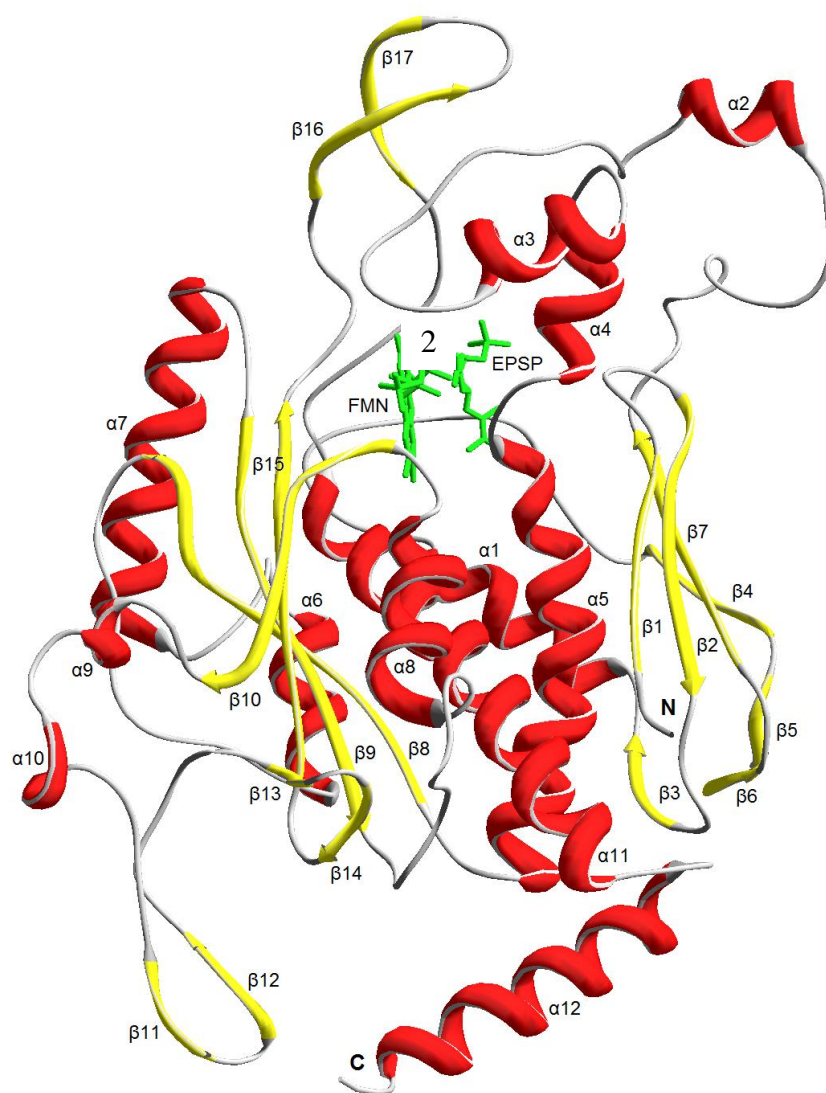


FIGURA 5. Representação da estrutura secundária do monômero da *MICS* (Vermelho: hélices- α ; amarelo: fitas- β ; cinza: voltas e alça).

1.3.1 Cofator e substrato

O cofator FMN (flavina mononucleotídeo ou riboflavina-5'-fosfato mononucleotídeo) (Figura 6-b) pertence a uma família de flavoproteínas que possuem uma variada e rica funcionalidade na natureza. Uma das principais características destas proteínas é a presença de um sistema de anéis tricíclico isoaloxazina que é o responsável por esta diversidade química (Bornemann, 2000). Entre os principais representantes estão a riboflavina (vitamina B2), que apresenta apenas uma ribose ligada ao anel isoaloxazina, a flavina adenina dinucleotídeo (FAD), que possui uma adenina ligada uma ribose por uma ligação fosfodiéster e o FMN, que tem um grupamento fosfato ligado uma ribose (Bornemann, 2000).

Exatamente por ser a principal forma de riboflavina encontrada nas células e tecidos humanos, este cofator não se caracteriza por ser um bom alvo, pois não possui toxicidade seletiva, ou seja, um inibidor para enzima *MtCS* análogo ao FMN poderia também interferir em outros processos biológicos do hospedeiro, no caso, o ser humano. Partindo deste princípio, o alvo deste trabalho é o EPSP, o substrato da reação catalizada pela enzima *MtCS*. A molécula do EPSP pode ser dividida em três regiões: carboxila, enol-piruvil e do grupo fosfato, como observado na figura 6-a.

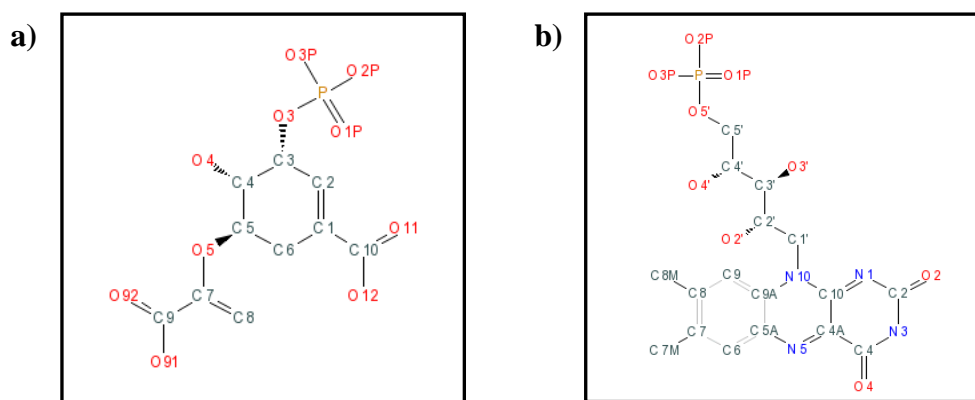


FIGURA 6. Representação da estrutura química - a) EPSP e b) FMN.

Outra característica interessante da enzima *MtCS* é em relação à redução do cofator FMN que se distingue em duas classes entre os microrganismos que possuem a via do chiquimato (Ely *et al.*, 2008). A CS de leveduras tem a capacidade de utilizar uma β -nicotinamida adenina dinucleotídeo fosfato (NADPH) para a redução do FMN oxidado, havendo desta forma, uma atividade catalítica adicional e, portanto, sendo chamada de bifuncional, enquanto todas as outras são chamadas de monofuncionais (Dias *et al.*, 2006).

As informações sobre as coordenadas atômicas das estruturas cristalográficas da CS, estão acessíveis publicamente através PDB (*Protein Data Bank*) (Berman *et al.*, 2002). O PDB é um banco de dados internacional com informações e coordenadas de estruturas tridimensionais (3D) de complexos de moléculas biológicas, determinadas através de métodos experimentais como cristalografia com difração de raios X e Ressonância Magnética Nuclear (NMR) (Berman *et al.*, 2002).

Atualmente existem onze estruturas depositadas, tanto nas formas apo (não-complexadas), quanto complexadas com outras moléculas, com diferentes resoluções e de variados organismos, como pode ser observado na tabela 1.

TABELA 1. Informações sobre as estruturas de CS depositadas no PDB.

PDB	Complexo	Organismo	Resolução (Å)
2QHF ^a	CS complexada com NCA (Nicotinamida)	<i>Mycobacterium tuberculosis</i>	1.65
2O11 ^b	CS Apo	<i>Mycobacterium tuberculosis</i>	1.65
2O12 ^c	CS complexada com FMN	<i>Mycobacterium tuberculosis</i>	1.72
2G85 ^d	CS Apo	<i>Mycobacterium tuberculosis</i>	2.22
1ZTB ^e	CS Apo	<i>Mycobacterium tuberculosis</i>	2.65
1UMO ^f	CS complexada com FMN	<i>Helicobacter pylori</i>	1.95
1UMF ^f	CS Apo	<i>Helicobacter pylori</i>	2.25
1SQ1 ^g	CS Apo	<i>Campylobacter jejuni</i>	2.80
1QXO ^h	CS complexada com FMN e EPS	<i>Streptococcus pneumoniae</i>	2.00
1R52 ⁱ	CS Apo	<i>Saccharomyces cerevisiae</i>	2.89
1R53 ⁱ	CS Apo	<i>Saccharomyces cerevisiae</i>	2.20
1Q1L ^j	CS Apo	<i>Aquifex aeolicus</i>	2.05

Nota: As seguintes referências foram usadas para compilar estes dados - ^a[35]; ^b[36]; ^c[37]; ^d[38]; ^e[12]; ^f[39]; ^g[40]; ^h[41]; ⁱ[42]; ^j[43].

1.4 Desenvolvimento de drogas e a abordagem computacional

A compreensão dos mecanismos de interação molecular receptor-ligante e a racionalização de como o fármaco em potencial poderá interagir neste processo, além de ser um dos principais desafios da biologia molecular, é um dos aspectos centrais para o sucesso e planejamento de novos fármacos dentro da área de desenho racional de drogas. Além disso, o conhecimento detalhado da estrutura da proteína-alvo torna-se fundamental, bem como, a via metabólica na qual ela participa. Por esta razão, ferramentas de bioinformática para construção, simulação e análise de estruturas 3D tornam-se fundamentais.

1.4.1 – Modelagem molecular por homologia

A modelagem por homologia comparativa é usualmente o método de escolha quando existe uma clara relação de homologia entre a seqüência da proteína alvo e pelo menos uma estrutura 3D conhecida. Esta técnica computacional é baseada na hipótese que estruturas terciárias de duas proteínas possam ser similares se suas seqüências primárias sejam relacionadas (Canduri *et al.*, 2003). Portanto, se semelhanças entre duas proteínas são detectáveis à nível de seqüência, a analogia estrutural pode ser presumida. Dentre os programas utilizado, um dos mais conhecidos é o MODELLER 9v7 (Sali & Blundell, 1993; Fiser *et al.*, 2000; Marti-Renom *et al.*, 2000, Eswar *et al.*, 2006).

O programa MODELLER 9v7 utiliza um método de modelagem por satisfação de restrições espaciais, utilizando distâncias geométricas ou técnicas de otimização.

Partes específicas do modelo da proteína, como *loops*, também podem ser refinados.

O processo total de predição do modelo possui quatro etapas: 1) determinação do *fold* ou seja, da estrutura 3D relacionada com a seqüência primária da proteína a ser modelada (alvo) e que será utilizada como molde; 2) alinhamento do alvo com o molde; 3) construção dos modelos e 4) avaliação e seleção das estruturas (Eswar *et al*, 2006).

O resultado final é um modelo 3D da seqüência alvo contendo todos os átomos das cadeias principal e lateral sem hidrogênio (Canduri *et al*, 2003).

1.4.2 – Docagem molecular

Uma metodologia que se mostra eficiente para identificação de novos inibidores é a abordagem computacional de simulação do encaixe de um ligante no sítio ativo de uma enzima. Esta metodologia é chamada de docagem molecular (Shoichet, 2004).

Esta simulação pode ser realizada utilizando-se um algoritmo capaz de prever o modo ideal de ligação de uma molécula pequena (ligante) ao sítio ativo de uma macromolécula alvo. No docagem molecular, a estrutura do alvo molecular (as coordenadas atômicas da enzima) permanece fixa e diversas posições de encaixe possíveis para o ligante são identificadas computacionalmente. A predição do ligante mais adequado pode ser feita a partir da aplicação de funções escores empíricas, que analisam a interação do complexo (Shoichet, 2004). Essa abordagem computacional permite a triagem *in silico* de grandes bibliotecas de compostos, avaliando afinidade e a especificidade a partir de propriedades estruturais e

químicas, como tamanho, geometria, distribuição de cargas, polaridade e potencial de interações hidrofóbicas e ligações de hidrogênio. Assim, o objetivo da triagem de bancos de possíveis ligantes é identificar compostos que se ligam mais fortemente a uma proteína alvo em relação ao seu substrato natural. Ao fazer isso, a reação bioquímica que a proteína alvo catalisa pode ser alterada ou impedida (inibição), agindo assim, como potenciais inibidores e conseqüentemente, possíveis fármacos.

Mesmo ainda com algumas limitações computacionais, atualmente o acesso via *internet* a banco de dados de estruturas cristalográficas, como PDB (Berman *et al.*, 2002), e de vastas bibliotecas de pequenas moléculas comercialmente acessíveis, como o ZINC *database* (Irwin & Shoichet, 2005) fornecem um grande número de possíveis novos inibidores, os quais podem ser comprados ou sintetizados, e então testados *in vitro*, agilizando desta forma a pesquisa farmacêutica (Shoichet, 2004).

Esta abordagem, denominada de VS, é executada através de *softwares* de simulação de docagem molecular, como o *MolDock* 4.1.0 (Thomsen & Christensen, 2006).

MolDock ou *Molegro Virtual Dock* 4.1.0 (Thomsen & Christensen, 2006) é uma implementação de uma variação do algoritmo evolucionário, guiado por evolução diferencial, uma técnica de otimização iterativa inspirada levemente na evolução darwiniana, porém, mais simplificada. Este algoritmo utiliza uma abordagem alternativa para selecionar e modificar candidatos à solução (“indivíduos”). A principal inovação, é a idéia de criar descendentes (“filhos”) a partir de uma diferença ponderada dos “pais”.

Na primeira etapa do algoritmo do *MolDock* todos os indivíduos são inicialmente posicionados aleatoriamente por meio de operadores de variação, como mutação e

recombinação, dentro do espaço conformacional delimitado (sítio ativo) e avaliados segundo um critério de ajuste (função escore). Na próxima etapa, para cada indivíduo da população anterior, um descendente é gerado, adicionando-se uma diferença ponderada das soluções “pais”, que são aleatoriamente selecionados a partir da população. Posteriormente, se o descendente for mais apto (segundo o critério de ajuste), substitui o “pai”, caso contrário o “pai” continua e passa para próxima geração. Este procedimento é repetido várias vezes, até que o critério de parada seja satisfeito (número máximo de iterações atingido ou se a variância da população está abaixo de um valor de corte, 0,01, por exemplo).

O algoritmo de evolução diferencial tem mostrado sucesso na previsão da interação proteína-ligante devido ao fato da população inicial, que é aleatoriamente gerada, ser grande, permitindo uma exploração ampla das possibilidades de complexos proteína-ligante. Esta iteração restringe o número de soluções que, em casos favoráveis, leva à solução certa, ou bem próxima da posição cristalográfica.

A função escore (energia) usada para avaliar a interação proteína-ligante e como critério de seleção dos indivíduos da população usa a seguinte equação: **Escore=E inter+E intra**, onde **E inter** leva em conta a energia intermolecular e **E intra** a energia do ligante (Thomsen & Christensen, 2006). A conformação final docada é selecionados de acordo com seu escore de energia, sendo que a hipótese geral é que os escores mais baixos de energia representam as melhores interações proteína-ligante.

Além da função escore intrínseca do programa *MolDock*, outro critério que pode ser utilizado para estimar a afinidade de ligação de um complexo é através do coeficiente de dissociação, ou pK_d , calculado através de uma função polinomial com o programa POLSCORE (De Azevedo & Dias, 2008; Dias *et al.*, 2008). Trabalhos

recentes que aplicaram a abordagem de uma análise computacional adicional para reclassificar os resultados de docagem obtidos com *MolDock*, para a enzima *MtPNP* (Purina Nucleosídeo Fosforilase de *Mt*), apresentaram informações mais confiáveis, com a eliminação de falsos-positivos, confirmadas por análise experimental (Ducati *et al.*, 2010).

Em resumo, a docagem molecular pode ser formulada como um problema de otimização, onde a tarefa é encontrar o modo de fixação do ligando com a menor energia de ligação.

Além disso, outros parâmetro podem ser analisados para avaliar a qualidade de um possível fármaco, como as regras de Lipinski (Lipinski *et al.*, 2001) e ADMETox (Miteva *et al.*, 2006).

As regras de Lipinski são baseadas nas propriedades físico-químicas de compostos que passaram na fase 1 de testes clínicos que envolvem determinação de aspectos relacionados à toxicidade e farmacocinética. Os critério analisam o número de doadores e aceptadores de hidrogênio, a massa molecular e coeficiente de partição ($\log P$) ou lipofilicidade da molécula.

Já os critérios de avaliação ADMETox (Absorção, Distribuição, Metabolismo, Excreção e Toxicidade) podem ser empregados em pequenas moléculas através de *web server*, como o *FAF-drugs* (Miteva *et al.*, 2006), e possuem além dos parâmetros das regras de Lipinski, outros que avaliam propriedades como área de superfície polar, número de ligações rígidas e flexíveis, número e tamanho de anéis, número de átomos de carbono e não-carbono, carga total, entre outros.

1.4.3 – Dinâmica molecular

Outro método de análise computacional é a simulação por DM, a qual pode fornecer a descrição detalhada dos movimentos individuais de partículas em função do tempo. Esta simulação permite a elucidação de problemas freqüentes que requerem uma compreensão das interações que acontecem a nível molecular, podendo ser usadas para responder questões específicas, sobre as propriedades de um sistema biológico modelo, de maneira mais simples que nos sistemas experimentais disponíveis.

Nesta abordagem, as simulações são realizadas em água (solvatada), utilizando as estruturas cristalográficas ou complexos binários obtidos a partir das simulações de docagem, como modelos iniciais. A proteína complexada com o ligante passa por um estágio inicial de minimização de energia, posterior termalização (aquecimento) e segue para a fase de produção onde serão coletados os dados. A análise destas simulações permite observar as modificações estruturais ocorridas devido à interação do complexo proteína-ligante, bem como prover a descrição detalhada dos movimentos individuais das partículas em função do tempo. Um dos pacote de programas mais utilizados para DM é o GROMACS 4.0.5 (Van der Spoel *et al.*, 2005; Hess *et al.*, 2008), disponível como *freeware* sob licença GNU (*General Public License*).

Devido ao custo computacional relativamente alto, a dinâmica molecular é utilizada apenas para análise dos melhores complexos proteína-ligante obtidos das simulações de docagem.

Em resumo, a abordagem *in silico* da interação proteína-ligante, através de VS e DM, utiliza métodos dedutíveis, fazendo a integração entre informações

disponíveis de biologia celular, molecular e farmacológica com a tecnologia de *softwares* de análise de sistemas biomoleculares, tornando possível a simulação de aspectos dinâmicos da estrutura 3D de macromoléculas biológicas.

2. OBJETIVOS

2.21 Objetivo geral:

Identificar potenciais inibidores para corismato sintase de *Mycobacterium tuberculosis*.

2.3 Objetivos específicos:

- Estudar os mecanismos moleculares de interação entre substrato, cofator e a enzima CS.

- Identificar, a partir da triagem *in silico* (VS) de duas bibliotecas de compostos químicos, a SIGMA-Aldrich (2006 – versão 5) e a ACROS Organics (2006 – versão 4) (Irwin & Shoichet, 2005), totalizando 19.615 moléculas disponíveis comercialmente para testes *in vitro*, pequenas moléculas que apresentam maior especificidade e afinidade, e que sejam potenciais inibidores da enzima CS.

- Analisar as interações e a estabilidade das estruturas dos complexos da enzima com os diferentes ligantes identificados no VS.

3. ARTIGO CIENTÍFICO

Analysis of the molecular basis for binding of *Mycobacterium tuberculosis* chorismate synthase toward its natural substrate and possible inhibitors: A docking and molecular dynamics study

Kelen B. Rocha^{1,2} and Walter Filgueira de Azevedo Jr.^{1,2*}

¹Faculdade de Biociências, Laboratório de Bioquímica Estrutural, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre – RS, Brazil;

²Programa de Pós-Graduação em Biologia Celular e Molecular.

(*) Corresponding author: Walter Filgueira de Azevedo Jr.
Phone/ FAX: +55-51-33204529 E-mail: walter@azevedolab.net

Abstract

The increased incidence of multi- and extensively-drug resistant *Mycobacterium tuberculosis* strains, combined with co-infection human immunodeficiency virus and the absence of new anti-tuberculosis chemotherapy in the last years show the need of finding new and more effective drugs to treat tuberculosis. For this, the rational development of new therapeutic agents, which is based on detailed knowledge of three-dimensional structure of the target protein, understanding its interaction with the ligand and the rationalization of how the drug can potentially interact in this process, it is fundamental. In this context, is inserted the chorismate synthase, an enzyme with unique characteristics and peculiarities of reaction in nature, and that catalyzes the conversion of the last step of the shikimate pathway. This metabolic route is essential for survival of mycobacteria, but is absent in humans, who make chorismate synthase, an interesting potential target for a specific chemotherapeutic attack. In this work, we report docking studies simulation and for the first time, the dynamic simulation for this enzyme in complex with natural substrates, with a inhibitor the *Escherichia coli* chorismate synthase and with two ligands obtained through virtual screening. The results presented concerning possible inhibitors proposed are optimistic about the characteristics observed binding and geometric complementarity, as well as their chemical and pharmacological conditions, common to drugs already known and tested. This information can be used both to determine structural priorities in rational drug design, how to refine and improve approaches and protocols for computational screening.

Keywords: Molecular dynamics; Chorismate synthase; *Mycobacterium tuberculosis*; Virtual screening.

Introduction

The incidence of tuberculosis (TB) had a rapid decline in the early XX century in the developing countries, mainly due to improvements in sanitation and housing. This trend was accelerated initially by the introduction of the vaccine strain attenuated *Mycobacterium bovis* bacille Calmette-Guérin (BCG) [1,2] and by the discovery of the antibacterial and antitubercular properties of the streptomycin (1944), isoniazid and pyrazinamide (1952) and the introduction of ethambutol and rifampicin [2,3].

Notwithstanding, in recent decades, the resurgence of the disease is occurring worldwide. In 1993, gravity of the situation led the World Health Organization (WHO) declared TB a global emergency, in an attempt to increase public and political awareness to this problem [4]. The latest reports released in 2009, announce 9.27 million new cases, and of these, approximately 4.1 million or 44% were new smear-positive, placing tuberculosis as the world's second commonest cause of death from infectious disease [4].

In face of these scenario, as well as the fact that no new drug was identified in the past years for this pathology, the development of antibiotics targeted to potential targets, not only is an ongoing process, but now an urgent necessity. In order to better understand the molecular biology of bacteria, computational tools are being applied on potential targets for identification and developed of new lead compounds.

In this work we focused on chorismate synthase (CS; EC 4.6.1.4) (Fig. 1) from *Mycobacterium tuberculosis* (*Mt*), an important molecular target of shikimate pathway. This route catalyzes the sequential conversion of two products of the metabolism of carbohydrates, erythrose-4-phosphate and phosphoenolpyruvate in chorismate, a precursor common to the biosynthesis of important metabolites (Fig. 2) [5] in bacteria, fungi, algae, plants and parasites of the *Apicomplexa* phylum. However, it is absent in mammals, which depend on diet to

obtain these compounds [6,7], feature that makes this route preference in studies of drug development.

The CS is the seventh enzyme of this pathway and catalyzes the converting 5-enolpyruvylshikimate 3-phosphate (EPSP) to chorismate in the presence of a mandatory reduced flavin mononucleotide (FMN_{red}) as cofactor [5] (Fig. 3). This enzymatic reaction involves an usual *trans*-1,4 elimination of the C-3 phosphate and C-6 *proR* hydrogen from EPSP, indicating that the reaction may be nonconcerted [8]. This catalysis reaction is only known in biological systems, making the CS unique in nature [9].

We report docking studies simulation in complex with three different ligands in order to determine the structural basis for specificity of these ligands against *MtCS*. The molecular dynamics (MD) simulation were performed in order to evaluate stability of *MtCS* model. The structural knowledge and understanding of the mechanisms of binding to the active site of CS may contribute to the rational development of selective inhibitors for the enzyme and thus, potential anti-tuberculosis drugs.

Methodology

Computational methods for building three dimensional structure of the MtCS

The starting point for the work was the construction and validation of the quaternary structure *MtCS*:FMN:EPSP, used for virtual screening (VS) and DM. The search and identification of the crystal structure used as template was performed in the Protein Data Bank (PDB) [10]. Among the twelve structures of CS of the different organisms deposited with various resolutions and shapes five are from *Mycobacterium tuberculosis*, and only the PDB access code 2O12 [11] complexed with the FMN substrate, thus becoming the selected.

However, the chosen structure had two problems. When the crystal structure was solved, six residues, from 47 to 52, and 9 residues from the C-terminal did not present electronic density, and were omitted in the final model. The other problem was regarding the location of the active site of coenzyme EPSP, our target for molecular docking. To solve these problems, first a homology modeling was built and soon after, a geometric docking or overlapping structures, followed by an energy minimization (EM) of the ternary complex.

Molecular modeling

The final *MtCS* model was built without a short region of 15 residues present in the C-terminal. This flexible C-terminal region is far from the active site and it is missing in all CS structures solved so far, therefore omitted from the model. Six residues that were missing in the *MtCS* crystallographic structure were included in the *MtCS* model, they are Gly47, Ala48, Arg49, Met50, Thr51 and Phe52. In order to check the correct sequence of *MtCS* this information was retrieved from the National Center for Biotechnology Information (NCBI) GenBank database (NCBI: NP_217056.1). All waters and ligands were removed from the template (PDB access code 2O12 [11]).

The software MODELLER 9v7 [12,13,14,15] was used to construct models of the enzyme, using standard methodology for comparative modeling of proteins. We generated 1000 models. The best model was selected based on MODELLER objective function. The overall stereochemical quality of the final models for enzyme was assessed by the program PROCHECK [16]. The protein model was used for the identification of active site in the manual docking geometry and for docking simulations. After the modeling the 3D structure of CS, the next step was the location of the binding sites of the substrate EPSP and the FMN cofactor from the manual docking and construction of the ternary complex *MtCS:FMN:EPSP*.

This process was accomplished by the overlap of α -carbon of the model with the crystallographic structure: 2O12 [11], for the location of the binding site of FMN and 1QXO [17], for the location of the binding site of EPSP, using the Swiss-PDBViewer v4.0.1 program (SPDBV) [18]. The choice of structure 1QXO (*Streptococcus pneumoniae* - SpCS) [17] occurred based on the identity above 40% of sequences MtCS and SpCS, and on fact be the only crystal structure of CS complexed with the substrate EPSP.

The final model MtCS:FMN:EPSP was subjected to an EM (Steepest Descent) using the GROMACS 4.0.5 package [19] to correct any potential steric clashes and unfavorable contacts between the residues of the monomer of the protein and ligands (the protocol used for EM is described below in the “*Molecular Dynamics simulations*”). This technique aims at finding a set of coordinates that minimize the potential energy of the system of interest.

Redocking

Docking simulations were performed using MolDock 4.1.0 [20]. The molecular docking algorithm MolDock is based on a new heuristic search algorithm that combines differential evolution with a cavity prediction algorithm. The docking scoring function is an extension of the piecewise linear potencial (PLP) including new hydrogen bonding and electrostatic terms. To further improve docking accuracy, a re-ranking scoring function is introduced, which identifies the most promising docking solution from the solutions obtained by the docking algorithm [20].

Because there is no crystal structure of MtCS complexed with the substrate EPSP, for the simulation, we have used the atomic coordinates of the monomer D of 1QXO. The option of docking only with the monomer form, is justified by the fact that the binding site of EPSP only presents intermolecular interaction with the monomer. Moreover, according to Dias and

colleagues [9], the monomers B, C and D of *SpCS*, are at greatest structural similarity to that of *MtCS*.

The used protocol restricted the torsion angles of the enzyme and the cofactor, but allowed flexibility to the substrate. The center of the radius of the sphere that defines the docking simulations was 15 Å, on the catalytic site of the EPSP (Fig. 4). All solvent molecules and the co-crystallized ligands were removed from the structure. The results were evaluated by the function re-ranck Score of the program.

Virtual Screening

In the VS simulation we used the libraries of compounds from SIGMA and ACROS Organics [21] with 19.615 molecules, made available through the ZINC database [21]. The protocol established and validated in redocking, with RMSD (Root Mean Square Displacement) of ≈ 0.3 Å (in the three simulations) was employed in VS (Fig. 5).

Different orientations of the ligands were searched and ranked based on their energy scores. After, those with lowest energy were filtered by using the criterion of satisfaction of the rules of Lipinski [22] and toxicity – ADMETox (absorption, distribution, metabolism, excretion, and toxicity) [23], thus showing promising features of toxicity and pharmacokinetics.

The selected molecules were then analyzed in the SEA (Similarity Ensemble Approach) [24], a server that correlates the chemical properties of a compound with a database with activity determined molecules. In addition, the calculation pK_d of the ten best results was also done through the a polynomial empirical scoring function - POLSCORE [25,26] to estimate the ligand-binding affinity and applied to rank docking results.

Molecular Dynamics simulations

After the identification and analysis of candidate inhibitors of CS, the next step is to verify the stability of these compounds in the active site of the enzyme by MD technique. Simulations of structures complexed with substrate and cofactor (system 1), ligand A (system 2), ligand B (system 3), and ligand C (system 4) were held separately and only in monomeric form enzyme (monomer D).

In addition, as complementary study to verify enzyme stability in terms of their quaternary structure and the role of ligands in this stabilization, three DM were carried out, all apo - monomer (system 5 - D monomer), dimeric (system 6 - chains C and D) and tetrameric (system 7 - chains A, B, C and D).

All DM simulations were performed in periodic boundary conditions with GROMACS 4.0.5 [19] package using the Gromos 96.1 (53A6) force fields [27]. The molecular topology file and the force field parameters were generated with the Dundee PRODRG server 2.5 (beta) program [28] except for the charges of the cofactor, substrate and ligands, which were submitted to single-point *ab initio* calculations at HF/6-311G (d, p) (B3LYP) level in order to obtain the Mülliken charges through the program Gaussian 03 [29]. This is because the Gromos 96.1 force fields are well parameterized for proteins, but the parameters for small molecules are still limited for simulations of more complicated biological systems.

Before MD simulations, the systems were solvated with explicit solvation model SPCE (Extended Simple Point Charge) [30], incorporated in a cubic box system, with the minimum distance between the protein surface and the box face were 10 Å length in all directions. To neutralize the negative charge density of the systems, sodium ions were added. The systems were then subjected to EM by steepest descent algorithm [31]. After the systems were gradually heated from 50 to 298.15 K in five steps (50-100, 100-150, 150-200, 200-250 and

250 to 298.15 K), and the speed of each step integration, re-scaled according to the Maxwell-Boltzmann distribution in order to adjust the kinetic energy of the system until the final temperature. This period is called thermalization.

Soon after, the systems were subjected to a short MD, with position restrains for a period of 20 ps. Finally, a longer dynamic, with a trajectory of 18 ns for the monomeric structures and 10 ns for dimeric and tetrameric structures were accomplished, this time without restrains. All simulations were performed under standard temperature (298.15 K) and pressure (1 bar) using a time of coupling constant temperature of 0.1 ps and a time of pressure coupling constant of 1.0 ps. All bond lengths including hydrogen atoms were constrained by the LINCS algorithm [32]. Electrostatic interactions were calculated using the particle-mesh Ewald algorithm (PME) [33] with interpolation order of 4 and a spacing of 1.2 Å. The cutoff radius for the van der Waals interactions was 10 Å. The simulations were performed with a time steps of 2 fs in the integration of the equations of motion, and coordinates were saved every 1 ps. All analyses were performed on the ensemble of system configurations extracted at 0.5 ps time intervals from the simulation.

The convergences of the different simulations were analyzed in terms of the RMSD from the initial models structures and root mean-square fluctuation (RMSF). The radius of gyration (Rg), which demonstrates the changes at the level of compaction of the enzyme, was also calculated. The intermolecular hydrogen bonds were assessed by the Ligplot v4.0 [34], which is a program that implements the distance equation and identifies the pairs of atoms of the residues that make hydrogen bonding between protein and ligand and the manipulation of structures, was performed with SPDBV [18].

All docking simulations were conducted on an iMac (Intel Core 2 Duo 2.66 GHz, 2 GB DDR3 SDRAM 1066 MHz), while the MD simulations and results analyses were performed on an Intel Xenon ® T3500 DuoQuad Core 3.00 GHz.

Result and discussion

Quality of the models

The overall stereochemical quality of the modeled structure showed better results than the crystal structure used as a template, which was also analyzed as control. The Ramachandran plot which show the backbone ϕ and ψ dihedral angles, showed 314 or 96% of model residues in the most favorable regions, against 303 or 94.7% of the crystallographic structure 2012 (Table 1). Moreover, an analysis was performed on the main chain and side chain, with confidence intervals as the program parameters, and results showed that within the average or better. In addition, superposition of the model against crystallographic structure 2012 indicates RMSD of 0.15 Å.

To predict the secondary structure of the modeled region (residues 47-52), it was used web server APSSP2 [35] where we obtained results from the relationship between secondary structure of the amino acid and probability of correct prediction. All six residues presented as the most likely secondary coil (data no showing), corroborating the modeled structure (Fig. 6).

Docking geometry and active site

There is no crystallographic structure available for *MtCS* complexed with cofactor and substrate, thus the atomic coordinates of crystallography structure of the template *SpCS* (access code 1QXO) were used as basic models for construction of the ternary complex *MtCS:FMN:EPSP*. Subsequently, the quaternary structure were generated by overlapping modeled monomer with each of the chains of the tetrameric structure 1QXO. The

displacement of each monomer, about the crystallographic structure used, obtained by RMSD, was $\approx 1.8 \text{ \AA}$.

After the docking geometry, we can observe that the major orientations of active site residues of the EPSP, as Arg341, possibly be doing steric clash with the substrate, exactly because the crystal structure used as a template (2O12) was crystallized without the EPSP. This problem was corrected during the EM which had exactly this purpose.

In the final result of the modeled structure, the Arg49 and Arg341 are those which appear more displaced in relation to crystal structure 1QXO which served as reference for the location of the active site of EPSP in the structure of *MtCS*. This perhaps is one explanation for the difference between the number of hydrogen bonds of this residues with the phosphate portion of EPSP from *SpCS* (Arg337), which are only two, while in *MtCS*, the Arg341 makes four bonds. Moreover, according to previous studies [9], Arg341 is also part of the loop 29 of *MtCS* (between residues Ile338 and Ala346) which possesses a key role in opening/closing of the site of cofactor FMN.

The Arg48 and Arg45, which correspond the Arg46 and Arg49 in *MtCS*, are part of the loop 5 (*MtCS*) and are important residues of the active site of *SpCS*, making 3 and 2 hydrogen bonds with the EPSP, respectively. But in *MtCS*, only Arg49 appears and doing hydrophobic contact.

The residues Ala138 (equivalent Ala133 in *SpCS*), Arg40 (equivalent Arg39 in *SpCS*) and the cofactor FMN, appear doing the same bonds with the EPSP, only with slightly different lengths. Just the Arg139 (equivalent Arg134 in *SpCS*) appears doing hydrophobic contact with the EPSP in *MtCS*, instead of hydrogen bonding, as seen in *SpCS*. This may be due to the small displacement of this residue regarding the structure of *SpCS*.

An important feature seen in the geometric docking is relative adjacency of active sites of FMN cofactor and the substrate EPSP. Evidence from experiments [36] with *Escherichia*

coli CS (*EcCS*) in relation to the K_d of oxidized FMN in the presence or absence of EPSP, validate the theory that the binding in the sites are coordinated - the first entry of FMN in the deeper part of the site, followed by binding of EPSP, which blocks any possible exit of FMN from the active site and which has more contact with the solvent. These factors are manifested in the decrease in the K_d of FMN, suggesting a greater than 1000-fold decrease in the off rate of FMN in the presence of EPSP [17].

Another important feature also observed is the presence in the site geometrically docked of two main residues, His11 and His115, considered essential for catalysis of the reaction and conserved in the primary sequences of CS of various organisms, both of which appear in the same positions corresponding residues in *SpCS*, His10 and His110.

The change of three amino acids Ser131 → Ala136; Asn251 → Gln256 and Pro314 → Ser318, corroborating previous studies [37] were also observed.

In plots generated by Ligplot [34] program we can see all these observations and the similarities between the active sites of *SpCS* (Fig. 7-a) and *MtCS* (Fig. 7-b). The main residues of the active sites of the FMN and EPSP of the structure modeled and minimized of *Mt* and crystallographic of *Sp* are shown in table 2.

Virtual screening

As final result of the VS simulation, two molecules were selected under the ZINC041147896 code (called *ligand A*) and ZINC00005085 (called *ligand B*). Moreover, in order to correlate our results with experimental data, we also did molecular docking of the structure (6R)-6-fluoro-5-enolpyruvylshikimate 3-phosphate (called *ligand C*), an inhibitor of *Escherichia coli* CS [38].

Ligand A

The ligand A, obtained by screening the library of compounds Acros, has the IUPAC (International Union of Pure and Applied Chemistry) name of [(3*S*,3*aR*,6*S*,6*aR*)-6-[5-(4-imidazol-1-ylphenoxy)tetrazol-1-yl]-2,3,3*a*,5,6,6*a*-hexahydrofuro[3,2-*b*]furan-3yl](furan-2ylmethyl azanium).

When compared, the energy of binding of ligand A and B, -214.744 and -180.995, respectively, showed up better than the energy of binding of the natural substrate EPSP, which was \approx -159.954.

Although the ligand A have showed a binding energy lower than the ligand B, which implies a greater affinity with the enzyme, the number of hydrogen bonds was smaller, only two, however in two important residues of the active site (Fig. 8).

The hydrogen bonds of the ligand A with the O2' of FMN, also appears in the interact with the ESPS in *MtCS* and *SpCS*, being the only binding between the cofactor and substrate, appearing to be important for the stabilization of a possible competitive inhibitor. The bond lengths of O11 of EPSP with O2' of FMN is 3.20 Å and 2.85 Å in *MtCS* and *SpCS*, respectively. While in the ligand A, the length is 2.89 Å and binding of the atom O2' of FMN is the N3. The second hydrogen bond of the ligand A is with the residue His115, between atoms N2 and NE2, respectively, with the distance of 2.68 Å. Previously published studies of site-directed mutagenesis [39] demonstrated the importance of His115 in catalytic reaction of CS, when the mutation from His110 (*SpCS*) to an alanine reduced the enzyme activity by 10%.

Regarding the hydrophobic interactions, eight residues of the enzyme makes contact with the ligand A: Met50, Phe52, Leu133, Gln340, Arg139, His11, Arg49 and Arg341, being that the last four also appear in EPSP interactions.

When compared with the ligand B, eight out of the ten residues that make interactions with the enzyme appears in common, but not all the same type. The Arg139, for example, that in ligand A makes only hydrophobic contact, in ligand B makes a hydrogen bond of 3.13 Å. In relation to the ligand C, only six residues appear in the active site of both, however, the hydrogen bond of both with the FMN is in the same atom O3, with only a difference in distance of 2.89 Å (ligand A) to 2.97 Å (ligand C).

The physicochemical properties of the ligand showed up satisfactory and obeying the rules of Lipinski and toxicity. The molecular weight of this compound is the higher among others rated, even than the substrate natural (Table 3) but within the limit of 500 Da, consonant with six rings. However, its surface polar area is closer to the area of the EPSP than in relation to the ligand B. The pK_d , (POLSCORE [25,26]) used as an additional tool for evaluating the binding affinity of the compound to the enzyme, showed a result slightly smaller than the ligand B, but higher than the C ligand and the substrate EPSP (Table 4). This means that presenting a higher binding affinity and therefore, a characteristic quite satisfactory for a possible competitive inhibitor.

Ligand B

The ligand B, is named by IUPAC methyl 4-(2-benzylbenzoyl)-2,5-dimethyl-1H-pyrrole-3-carboxylate and by SIGMA of F131.

This compound has several studies related to several pharmacological properties, for example, as agent for pancreatic disorders, HIV protease inhibitor, inhibitor of the synthesis of leukotriene, antagonist/agonist of benzodiazepine, cardiotoxic, among others [24]. Moreover, this compound also showed structural relationship with chemical compounds with

antituberculosis activity, with E-value of $3,68 \cdot 10^{-3}$ and Tanimoto coefficient (TC) of 0.33 [24].

The ligand B when compared to A, it makes more hydrogen bonds (five in total), however has fewer hydrophobic contacts, being one with the FMN. However, all protein residues that interact with the ligand B also interact with ligand A, such as Arg49, Phe52, Leu133, Gln340 and Arg341 (Fig. 9). In the ligand A, Arg139 appears making just hydrophobic contact, on the other hand, in the ligand B this residue makes a hydrogen bond, as well as His115.

Compared with the ligand C, the ligand B has six residues making interactions in common, plus the cofactor. Of these, five also are part of the active site of EPSP.

Regarding the physicochemical properties of this compound, two features stand out and are complementary. The polar surface area, which is nearly half of the other analyzed and partition coefficient, which is positive. We can relate these two features by being directly reversed. Compounds with higher partition coefficient, are more hydrophobic and has higher affinity for the organic phase, and therefore tend to more easily overcome the hydrophobic biomembranes, such as phospholipid bilayer structure of cells, that coming to be a feature satisfactory for a drug with performance in this niche.

Ligand C

Osborne and colleagues demonstrated experimentally the inhibition of *EcCS* with an analogue compound of the EPSP, the (6R)-6-fluoro-5-enolpyruvylshikimate 3-phosphate. When incubated with the enzyme, the ligand C showed an $IC_{50} = 0.5 \mu M$, but when the enzyme was pre-incubated, the concentration of the compound needed to inhibit 50% of the enzyme activity increased to $250 \mu M$ [38].

Based on this data, as well as the high degree of sequence conservation CS enzyme between species [37] this inhibitor was subjected to docking and MD against structure of *MtCS*, in order to correlate information with ligands screened in VS and to assess whether this compound also has inhibitory potential in *MtCS*.

Ligand C, among the docked molecules, is the one which shows great similarity to the residues of the enzyme that makes the EPSP interaction, probably for to be substrate analogue.

Among these, Arg40, Arg49 and Arg341, which in the ligands appear only making hydrophobic contact, arise making hydrogen bonds with the molecule. Another similarity is the Ala138 and Arg139 which form hydrogen bond with EPSP, ligand C and the ligand B. However, two other residues that were not part of any active site appear making interaction with this molecule, Ser141 and Thr137 (Fig. 10).

The physicochemical properties of the ligand C and substrate, exactly for the structural similarity, are very alike. Nevertheless, contrary to expectations, the binding energy of the ligand C, compared with the EPSP, was higher and lower pK_a . This may be due to the weights of each property of the algorithms of the programs, but not totally rule out this molecule as a possible competitive inhibitor.

Study of the stability and of the active site residues by DM

Stability of the quaternary structure

To add information to work, the study the stability of the enzyme in different compositions of quaternary structure was performed by three independent MD simulations,

with the monomeric forms (system 5 - D monomer), dimeric (System 6 - chains C and D) and tetrameric (System 7 - chains A, B, C and D) of the non-complexed enzyme.

To obtain an estimate of the quality of convergence of the trajectories of MD, the main chain RMSD of models in the initial structures was calculated, as demonstrated in figure 11. The tetrameric structure (Fig. 11 - green line) is the more stable ($2.2 \pm 0.3 \text{ \AA}$), followed by the dimer (Fig. 11 - red line), which increased slightly at the beginning of simulation, stabilizing after 5000 ps or 5 ns ($3.3 \pm 0.8 \text{ \AA}$). The monomeric form (Fig. 11 - black line) showed greater variation ($3.9 \pm 0.8 \text{ \AA}$), and at the end of the simulation of 10 ns, the structure had not yet stabilized. This important finding suggests that the enzyme *MtCS* is stable only in the form of dimer or trimer and/or complexed. This information corroborating with the experiments of Dias and workers [9], through gel filtration column to determination of the quaternary structure *MtCS* in solution. In this study, they observed that in solution the enzyme CS tends to be in balance in the dimeric and tetrameric forms, but that the increase of enzyme concentration and or the binding of a molecule leads to conformational changes of the dimers, resulting in tetramerization of the enzyme.

Another important property, which describes the overall spread of the molecule and is defined as the root mean square distance of the collection of atoms from their common centre of gravity, is the radius of gyration. The second graph (Fig. 12) shows that the size of the radius of the dimeric structure (Fig. 12 - red line) and tetrameric (Fig. 12 - green line) are larger, hence why these structures have two and four times the waste, respectively. As expected, the monomeric system (Fig. 12 - black line) showed the greatest variation, but keeping the protein folding of their original design.

Flexibility of the structures was described by RMSF, which reflects the movement of residues of the enzyme, as shown in figure 13 and 14. The largest fluctuations were observed

in the loops and C-terminal regions, with lower values of RMSF are the central regions of the structure.

In the system 5, corresponding to the monomer (Fig. 13-a and 14-a) can be defined four regions with greater variation, located between residues 84-104 (R1), 110-132 (R2), 261-296 (R3) and 318-340 (R4). The areas corresponding to regions 2, 3 and 4 are located near the interface of monomers C and D, which could explain the greater variation or movement in these areas monomeric system.

In graph corresponding to the dimer (Fig. 13-b), is also possible to distinguish four regions with the greatest fluctuation, between residues 77-106 (R1), 171-191 (R2), 277-294 (R3) and 313-341 (R4). However, this change in position of residues along the trajectory simulation is lower than in the monomeric system.

In the tetrameric structure, the fluctuation of the waste appears to be more homogeneous and five regions with the highest oscillation can be observed between residues 41-55 (R1), 80-108 (R2), 170-193 (R3), 280-291 (R4) and 318-349 (R5) (Fig. 13-c and 14-c).

In all three systems, it is possible to observe common regions with more fluctuation. The residues located between α -helix 3 and β -strand 7, corresponding to region 1 in the monomeric and dimeric structures, and 2 in the tetrameric structure, are located in a peripheral part of the protein and no contact with other monomers, being this a possible explanation for a greater fluctuation. Three other common areas can also be observed: R2 (dimer) to R3 (tetramer), R3 and R4 (monomer and dimer) with R4 and R5 (tetramer), respectively.

The *MtCS* dimers are characterized by a large antiparallel eight-stranded β -sheet constituted of two sheets (one sheet in each monomer). This large sheet is the main feature of the dimer interface, and the two β -strands 10 form the most important region of stabilization of the dimer, together with the α -helix 8 that interacts with the other α -helix 8 of the adjacent monomer. In the simulation of the monomer, we can observe the intense movement in this

region, more specifically between from residues 261 to 264 (R3), which are part of the β -strand 10, but that in the dimeric and tetrameric systems was seen with less intensity.

Stability of the ligands in the active site and key residues responsible for this stabilization

To analyze the stability of the ligands complexed with *MtCS*, the RMSDs of α -carbons of the structures during the simulation trajectories (18 ns) were measured (Fig. 15). It was possible to observe an increase in the deviation in all systems in the first part of the dynamic and stabilization at around 10 ns, except for the apo structure (blue line). This leads us to believe that the binding of a small molecule to the enzyme has a key role in its stabilization, as mentioned earlier and that the compounds proposed here as potential inhibitors, are stable in the active site of the enzyme.

The average values of RMSD of the systems *MtCS*-FMN complexed with EPSP and ligand C appear equal to $3.6 \pm 0.6 \text{ \AA}$. This result is probably related to chemical and structural similarity of the two compounds and also suggests a similar behavior of the protein. The system apo presented the highest mean RMSD, $4.7 \pm 1.1 \text{ \AA}$, followed by the ligand (4.3 \pm 0.4 \AA) and B (4.1 \pm 0.9 \AA), respectively.

Complementing the stability aspect, in figure 16, which shows the variation of the radius of gyration during the simulations, we can observe from the 15 ns, the convergence of all systems, and the difference between the averages of only $\approx 0.4 \text{ \AA}$. This represents that even with increased temperature, the protein folding, both with different ligands such as apo, remains stable.

To analyze the active site residues that make interactions with the EPSP and the ligands A, B and C, were analyzed the snapshots extracted from the simulations of the trajectories of

all complexed systems every nanosecond (from the first to eighteenth) using the program Ligplot [34]. The results, shown in figure 17, highlight the main residues that interacted at some point in the trajectory with the ligands A, B and C (combined - in pink) in comparison with the natural substrate of the enzyme EPSP (purple).

Regarding the RMSF of the structures, we can see, unlike apo systems shown in figure 18-e, a greater similarity between the regions and the size of the fluctuations in all analyzed complexes, indicating that, when attached to a molecule, the tendency of movement the active site appears to be very similar. Regions 1 and 2, corresponding to residues 75-138 and 261-301, can be observed in all complexes (fig. 18) as the largest areas of floating of the residues. At the peak of these regions are residues 90, 96, 94, 100, 123 (R1) and 265, 269, 270, 274, 283, and 286 (R2).

However, the system complexed with ligands (Fig. 18-b) differs from the others. A third area of greatest fluctuation appears between residues 331-346.

We conclude from these data that residues Arg49, Ser137, Ala138, Arg139, and Arg341, have a key role in stabilizing the binding of a molecule in the catalytic site of the substrate EPSP from *MtCS*. Moreover, the results of MD simulations indicated good stability of the ligands A, B and C within the active site of the enzyme, suggesting that the molecules proposed are potential inhibitors for *MtCS* and promising for anti-mycobacterial chemotherapy.

Conclusion

This paper presents for the first time, modeling, docking studies, and MD based on a crystal structure of *MtCS*, which gives our results, a higher reliability. Previous works, which were used as a base, modeling and analysis presented *MtCS* only from the structure of *SpCS* [37]. Studies of MD, which provide more realistic simulation system for being a solution, also are unprecedented for this enzyme, explaining, supplementing and enriching the structural data of *MtCS* and its interaction with the cofactor and substrate for the study of candidates for inhibitors examined here, as well as for future work on this issue.

The results of modeling the geometric docking and EM of the structure built, demonstrated a good quality of the proposed model and its validation for their use in molecular docking simulations.

The results concerning possible inhibitors, called ligand A, B, and C, are promising about the characteristics observed of the binding, geometric complementarity and stability in the enzyme active site, as well as their chemical and pharmacological conditions, common to drugs already known and tested.

Moreover, the results obtained from simulations of MD were quite interesting in relation to apo enzyme stability in different oligomeric states, suggesting that the main factors responsible for this stability is directly related to the quaternary structure and the binding of a molecule.

However, despite the bioinformatics is a powerful tool in drug development by making the integration of available information on cell biology, molecular pharmacology and technology with software for analyzing biomolecular systems, in vitro tests are needed, as well as other stages of prediction of new drugs.

Acknowledgments

This work was supported by grants from CNPq and Instituto Nacional de Ciência e Tecnologia do Conselho Nacional de Desenvolvimento Científico e Tecnológico-Ministério de Ciência e Tecnologia (INCT-Tuberculose, CNPq-MCT, Brazil). KBR would like to thank CNPq for the fellowship. WFA is senior researcher for CNPq (Brazil).

References

- [1] Parish T, Stoker NG. (2002) The common aromatic amino acid biosynthesis pathway is essential in *Mycobacterium tuberculosis*. *Microbiology* 148:3069-3077.
- [2] Ducati RG, Ruffino-Netto A, Basso LA, Santos DS (2006) The resumption of consumption: a review on tuberculosis. *Mem Inst Oswaldo Cruz* 101:697-714
- [3] Basso LA, da Silva LHP, Fett-Neto AG, de Azevedo Jr WF, Moreira IS, Palma MS, Calixto JB, Filho SA, dos Santos R R, Soares M, Santos DS (2005) The use of biodiversity as source of new chemical entities against defined molecular targets for treatment of malaria, tuberculosis, and T-cell mediated diseases – a review. *Mem Inst Oswaldo Cruz* 100:475-506
- [4] World Health Organization. Global Tuberculosis Control: Surveillance, planning, financing. WHO Report 2009. WHO, Geneva, Switzerland. 2009
- [5] Ely F, Nunes JES, Schroeder EK, Frazzon J, Palma MS, Santos DS, Basso LA (2008) The *Mycobacterium tuberculosis* Rv2540c DNA sequence encodes a bifunctional chorismate synthase. *BMC Biochem* 9:13
- [6] da Silveira, NJF, Uchoa HB, Pereira JH, Canduri F, Basso LA, Palma MS, Santos DS, de Azevedo Jr WF (2005) Molecular models of protein targets from *Mycobacterium tuberculosis*. *J Mol Model* 11:160-166
- [7] Dias MVB, Ely F, Palma MS, de Azevedo Jr WF, Basso LA, Santos DS (2007) Chorismate synthase: an attractive target for drug development against orphan diseases. *Curr Drug Targets* 8:437-444
- [8] Bornemann S, Theoclitou ME, Brune M, Webb MR, Thorneley RNF, Abell C (2000) A secondary beta deuterium kinetic isotope effect in the chorismate synthase reaction. *Bioorg Chem* 28:191-204
- [9] Dias MVB, Borges JC, Ely F, Pereira JH, Canduri F, Ramos CH, Frazzon J, Palma MS, Basso LA, Santos DS, de Azevedo Jr WF (2006) Structure of chorismate synthase from *Mycobacterium tuberculosis*. *Struct Biol* 154:130-143
- [10] Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook JD, Zardecki C (2002) The Protein Data Bank. *Acta Crystallogr D Biol Crystallogr* 58:899–907

- [11] Bruning M, Bourenkov GP, Strizhov NI, Bartunik HD (2006) Complexes of chorismate synthase from *Mycobacterium tuberculosis* (PDB: 2O12) - To be published
- [12] Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234:779–815
- [13] Fiser A, Do RK, Sali A (2000) Modeling of loops in protein structures. *Protein Sci* 9:1753-1773
- [14] Marti-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, Sali A (2000) Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* 29:291-325
- [15] Eswar N, Marti-Renom MA, Webb B, Madhusudhan MS, Eramian D, Shen M, Pieper U, Sali A (2006) Comparative protein structure modeling with MODELLER. *Current Protocols in Bioinformatics*, John Wiley & Sons, Inc., Supplement 15, 5.6.1-5.6.30
- [16] Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 8:477-486
- [17] Maclean J, Ali S (2003) The structure of chorismate synthase reveals a novel flavin binding site fundamental to a unique chemical reaction. *Structure* 11:1499-1511
- [18] Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18:2714–2723
- [19] van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ (2005) GROMACS: fast, flexible and free. *J Comput Chem* 26:1701–1718
- [20] Thomsen R, Christensen MHJ (2006) MolDock: a new technique for high-accuracy molecular docking. *Med Chem* 49:3315-3321
- [21] Irwin JJ, Shoichet BK (2005) ZINC - a free database of commercially available compounds for virtual screening. *J Chem Inf Model* 45(1):177-182
- [22] Lipinski CA, Lombardo F, Dominy W, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 46:3-26
- [23] Miteva MA, Violas S, Montes M, Gomez D, Tuffery P, Villoutreix BO (2006) FAF-Drugs : free ADME/tox filtering of compound collections. *Nucleic Acids Res* 34:W738-44
- [24] Keiser MJ, Roth BL, Armbruster BN, Ernsberger P, Irwin JJ, Shoichet BK (2007) Relating protein pharmacology by ligand chemistry. *Nat Biotech* 25:197-206
- [25] de Azevedo Jr WF, Dias R (2008) Evaluation of ligand-binding affinity using polynomial empirical scoring functions. *Bioorg Med Chem* 16:9378-9383

- [26] Dias R, Timmers LFSM, Caceres RA, de Azevedo Jr WF (2008) Evaluation of molecular docking using polynomial empirical scoring functions. *Curr Drug Targets* 9:1062–1070
- [27] Oostenbrink C, Soares TA, van der Vegt NF, van Gunsteren WF (2005) Validation of the 53A6 GROMOS force field. *Eur Biophys J* 34:273–284
- [28] Schuttelkopf AW, van Aalten DM (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr* 60: 1355-1363.
- [29] Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, *et al.* (2003) Density functional calculations of adsorbates on graphene sheets in uniform electric fields were performed using the generalized-gradient approximation and the Perdue, Burke, Ernzerhof parameterization of exchange and correlation in the Gaussian 03 localized-basis package Gaussian 03, revision B01. Gaussian, Pittsburgh, PA
- [30] Berendsen HJC, Grigera JR, Straatsma TP (1987) The missing term in effective pair potentials. *J Phys Chem* 91:6269–6271
- [31] Arfken G (1985) The Method of Steepest Descents. §7.4. In: *Mathematical Methods for Physicists*, 3rd ed. Orlando, FL Academic Press, pp 428–436.
- [32] Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: a linear constraint solver for molecular simulations. *J Comput Chem* 18:1463–1472.
- [33] Darden T, York D, Pedersen L (1993) Particle Mesh Ewald - an $n \cdot \log(n)$ method for ewald sums in large systems. *J Chem Phys* 98:10089–10092.
- [34] Wallace AC, Laskowski RA, Thornton JM (1995) LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. *Prot Eng* 8:127-134
- [35] Raghava GPS (2000) Protein secondary structure prediction using nearest neighbor and neural network approach. *CASP* 4:75-76
- [36] Macheroux P, Schönbrunn E, Svergun DI, Volkov VV, Koch MH, Bornemann S, Thorneley RNF (1998) Evidence for a major structural change in *Escherichia coli* chorismate synthase induced by flavin and substrate binding. *Biochem J* 335: 319-327
- [37] Fernandes CL, Breda A, Santos DS, Basso LA, Souza ON (2007) A structural model for chorismate synthase from *Mycobacterium tuberculosis* in complex with coenzyme and substrate. *Comput Biol Med* 37:149-158
- [38] Osborne A, Thorneley RNF, Abell C, Bornemann S (2000) Studies with substrate and cofactor analogues provide evidence for a radical mechanism in the chorismate synthase reaction. *J Biol Chem* 275:35825-35830
- [39] Kitzing K, Auweter S, Amrhein N, Macheroux P (2004) Mechanism of chorismate synthase. *J Biol Chem* 279:9451-9461

List of abbreviations

TB: tuberculosis

BCG: bacille Calmette-Guérin

WHO: World Health Organization

CS: chorismate synthase

Mt: *Mycobacterium tuberculosis*

EPSP: 5-enolpyruvylshikimate-3-phosphate

FMN_{red}: reduced flavin mononucleotide

MD: molecular dynamics

VS: Virtual Screening

PDB: Protein Data Bank

EM: energy minimization

3D: three-dimensional

NCBI: National Center for Biotechnology Information

SPDBV: Swiss-PDBViewer

SpCS: CS from *Streptococcus pneumonia*

PLP: piecewise linear potencial

RMSD: Root Mean Square Displacement

ADMET_{tox}: absorption, distribution, metabolism, excretion, and toxicity

SEA: Similarity Ensemble Approach

POLSCORE: polynomial empirical scoring function

SPCE: Extended Simple Point Charge

PME: particle-mesh Ewald

RMSF: root mean-square fluctuation

R_g: radius of gyration

EcCS: CS from *Escherichia coli*

IUPAC: International Union of Pure and Applied Chemistry

Legends of figures

Fig. 1 Chorismate synthase from *Mycobacterium tuberculosis* (*MtCS*) monomer. The structure is presented as ribbon diagram and flavin mononucleotide (FMN) and 5-enolpyruvylshikimate 3-phosphate (EPSP) are presented as stick

Fig. 2 The shikimate pathway catalyses the conversion of phosphoenolpyruvate and erythrose-4-phosphate to chorismate, a key precursor of mycobacterial biosynthesis of folates, ubiquinones, and aromatic amino acids

Fig. 3 Reaction catalyzed by chorismate synthase from *Mycobacterium tuberculosis* (*MtCS*)

Fig. 4 *MtCS*-EPSP-FMN. Docking sphere used in the redocking and virtual screening simulations. This sphere has a radius of 15 Å and it is centered at the active site of ESPS - generated by MolDock program

Fig. 5 Result of redocking. Superposition of the crystallographic structure of the EPSP (colored by CPK) and pose (green) - generated by MolDock program

Fig. 6 Representation of primary and secondary structure 3D *MtCS* in the modeled region superimposed on the crystal structure that served as a template. Arrows highlight the discontinuity of the crystal structure and absence of 6 amino acids (green: template - PDB: 2012; red: structure of the final model; gray: the modeled region representing its secondary structure; blue: modeled region representing its primary structure)

Fig. 7 Ligplot - representation of EPSP binding to *SpCS* (a) and for the *MtCS* (b). Dotted lines represent the hydrogen bonds and the half circles the hydrophobic contacts

Fig. 8 Ligplot - representation of ligand A binding to *MtCS*. Dotted lines represent the hydrogen bonds and the half circles the hydrophobic contacts

Fig. 9 Ligplot - representation of ligand B binding to *MtCS*. Dotted lines represent the hydrogen bonds and the half circles the hydrophobic contacts

Fig. 10 Ligplot - representation of ligand C binding to *MtCS*. Dotted lines represent the hydrogen bonds and the half circles the hydrophobic contacts

Fig. 11 Graphical representation of root-mean-square deviation (RMSD) of all $C\alpha$ as a function of time. Monomer (black line), dimer (red line) and tetramer (green line)

Fig. 12 Graphical representation of radius of gyration. Monomer (black line), dimer (red line) and tetramer (green line)

Fig. 13 Graphical representation of root-mean-square fluctuations (RMSF) of all $C\alpha$ from starting structure of models as a function of time. a) monomer (black line); b) dimer (red line); c) tetramer (green line); d) all systems together. In gray, the regions of residues highlighted in the text and in figure 14

Fig. 14 Ribbon representation in the initial structures (green) and final (blue) superimposed (MD -10 ns). Arrows indicate the region and the movement exercised during simulation. a)

monomer; b) dimer; c) tetramer; N: N-terminal; C: C-terminal (some regions described in text and graphics are not represented in the figures)

Fig. 15 Graphical representation of root-mean-square deviation (RMSD) of all C α as a function of time. Apo (blue line), *MtCS*-FMN-EPSP (black line), *MtCS*-FMN-ligand A (yellow line), *MtCS*-FMN-ligand B (red line) and *MtCS*-FMN-ligand C (green line)

Fig. 16 Graphical representation of radius of gyration. Apo (blue line), *MtCS*-FMN-EPSP (black line), *MtCS*-FMN-ligand A (yellow line), *MtCS*-FMN-ligand B (red line) and *MtCS*-FMN-ligand C (green line)

Fig. 17 Graphic showing the main residues of the protein that made the hydrogen bond relative to the number of times they appeared in the snapshots taken every 1 ns (from 1-18 ns) of the trajectories of MD simulations (pink: sum of ligands A, B and C; lilac: EPSP)

Fig. 18 Graphical representation of root-mean-square fluctuations (RMSF) of all C α from starting structure of models as a function of time. *MtCS*-FMN-EPSP (black line), *MtCS*-FMN-ligand A (blue line), *MtCS*-FMN-ligand B (red line) and *MtCS*-FMN-ligand C (green line)

Legends of the tables

Table 1 Ramachandran plot statistics for the best model of *MtCS* and for the template

Table 2 Residues located up to 4 Å of the active sites of FMN and EPSP in crystallographic structure of *SpCS* (1QXO) and *MtCS* (final model - modeled and minimized); (h) - α -helix; residues in bold also appear in Ligplot

Table 3 Hydrogen bonds (HB) and van der Waals contacts (VW) of *MtCS* with ligands A, B, and C (Ligplot)

Table 4 Data of the structures analyzed. PSA: polar surface area; LogP: partition coefficient; AHB: acceptor H-bonds; DHB: donor H-bonds; p*K*_d: dissociation constant; MolDock Score: binding energy MolDock program

Figure 1

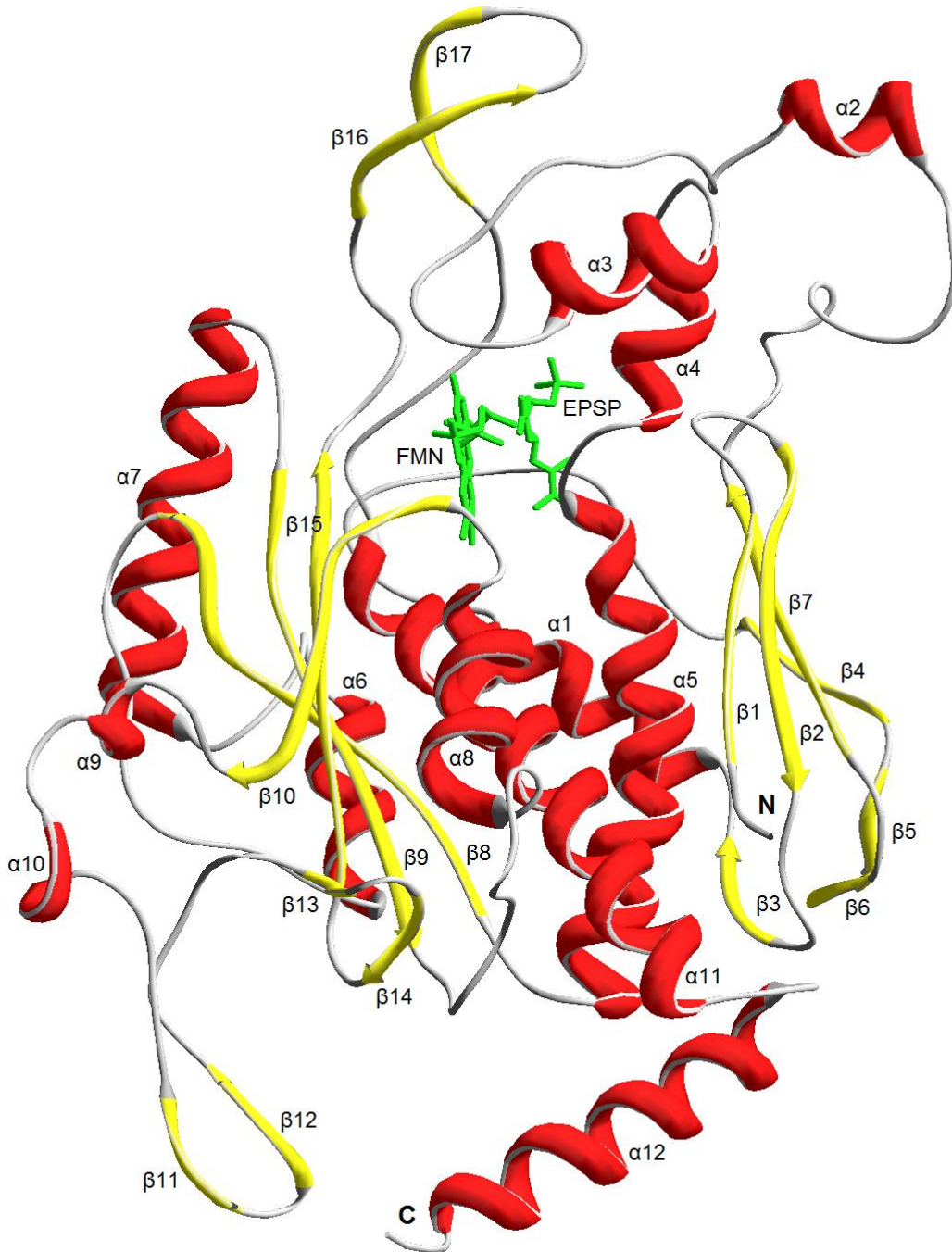


Figure 2

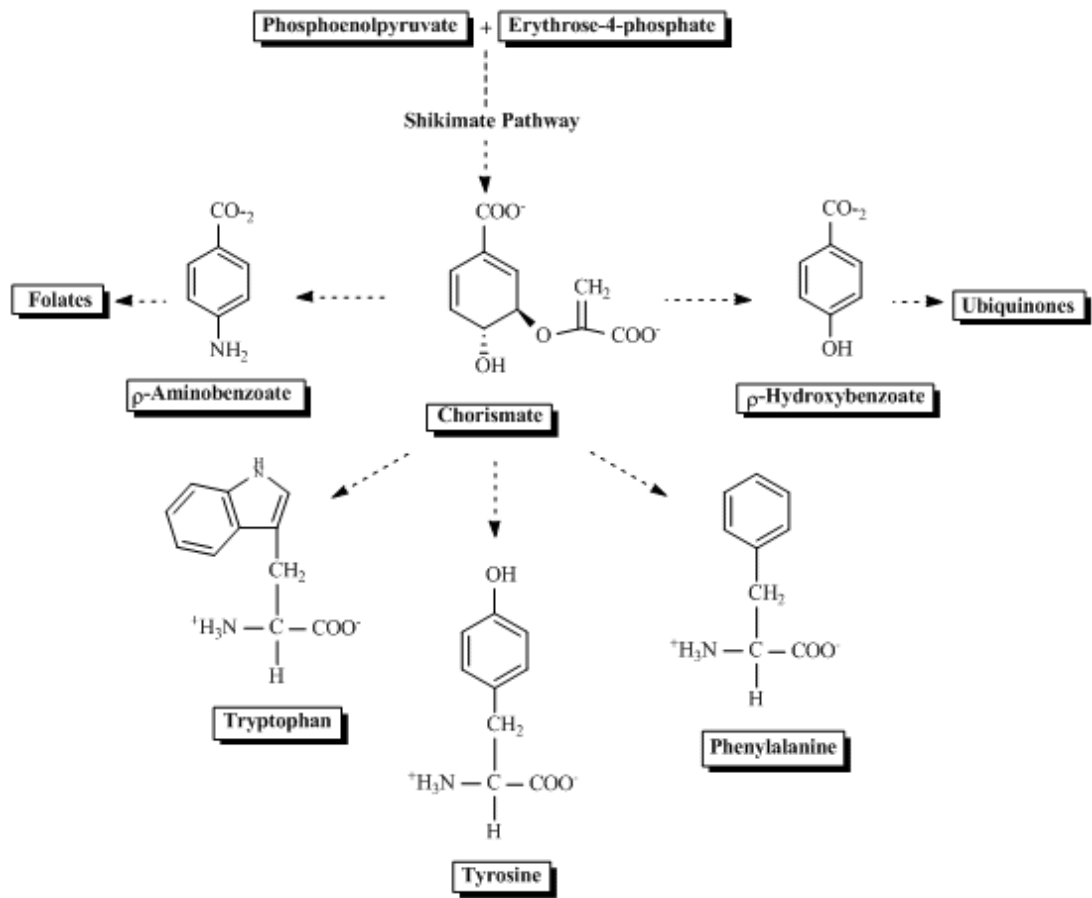


Figure 3

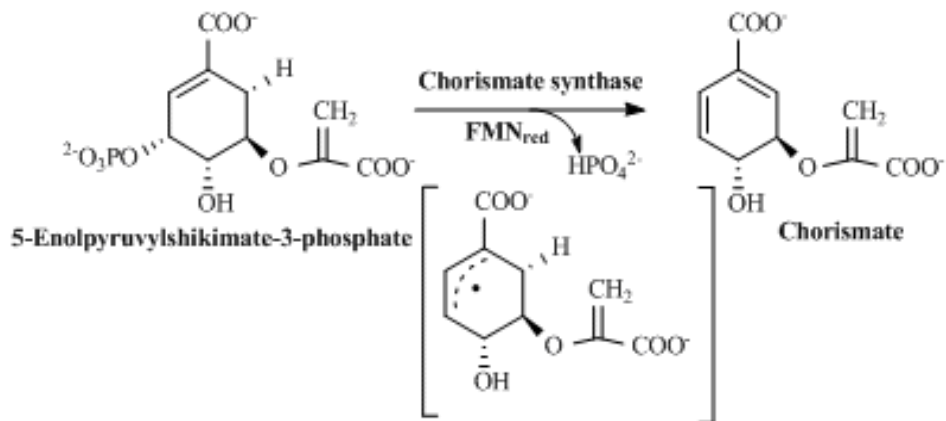


Figure 4

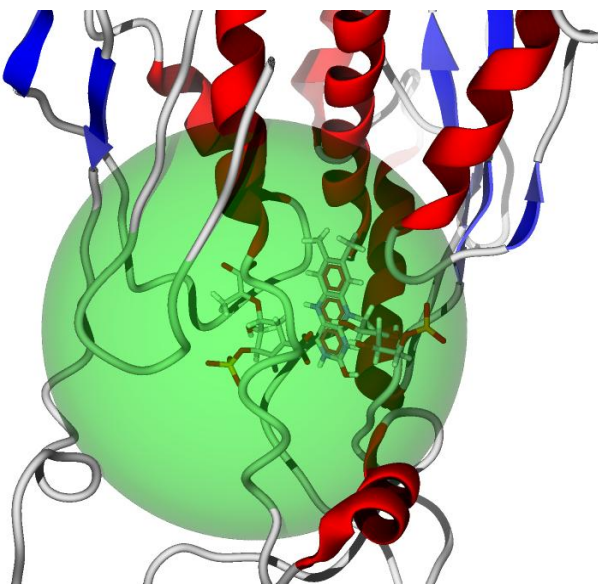


Figure 5

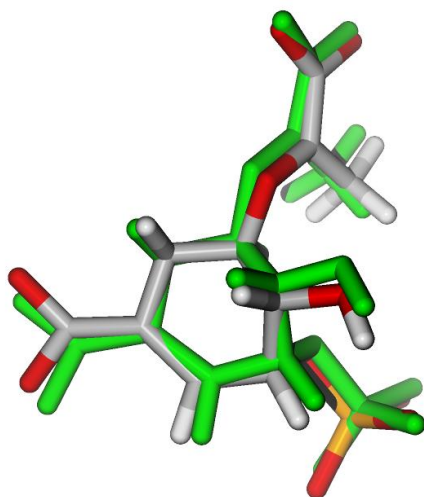


Figure 6

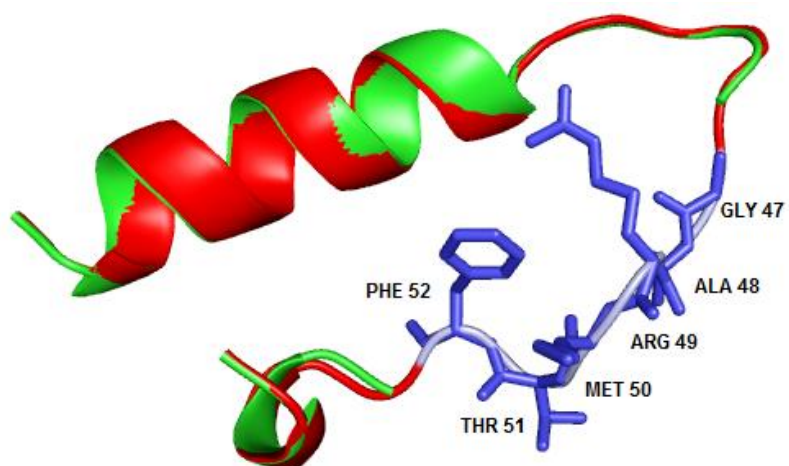
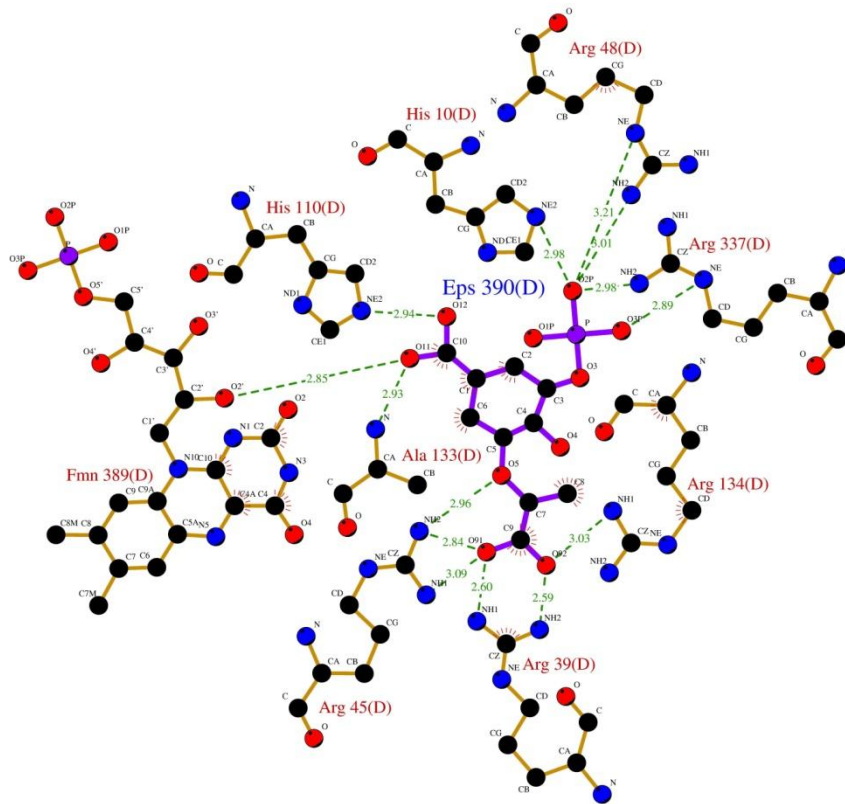


Figure 7

a)



b)

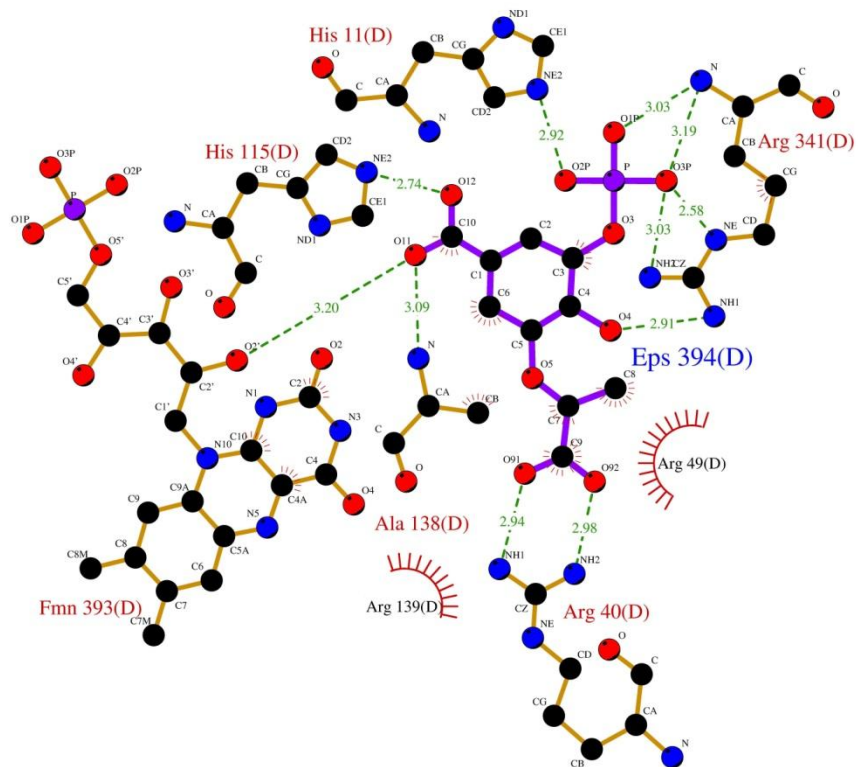


Figure 8

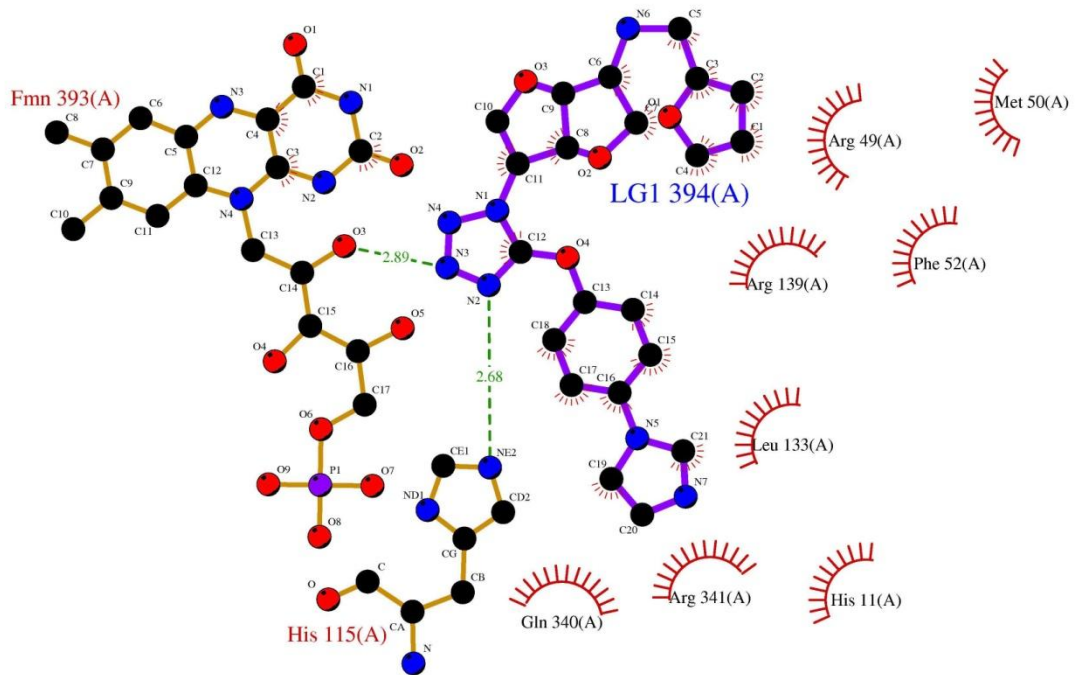


Figure 9

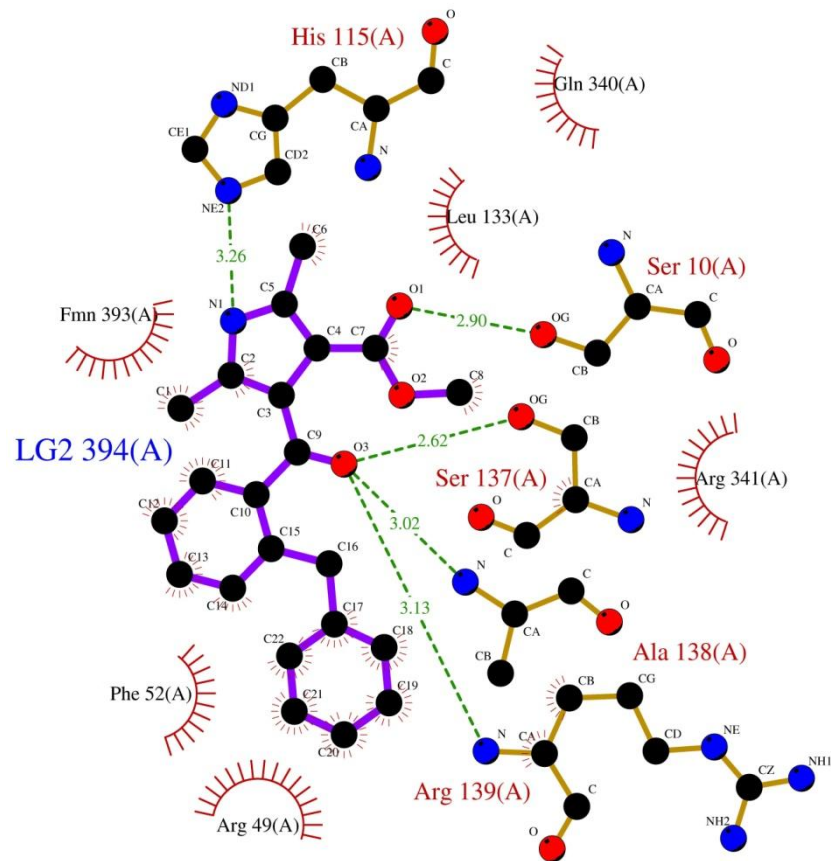


Figure 10

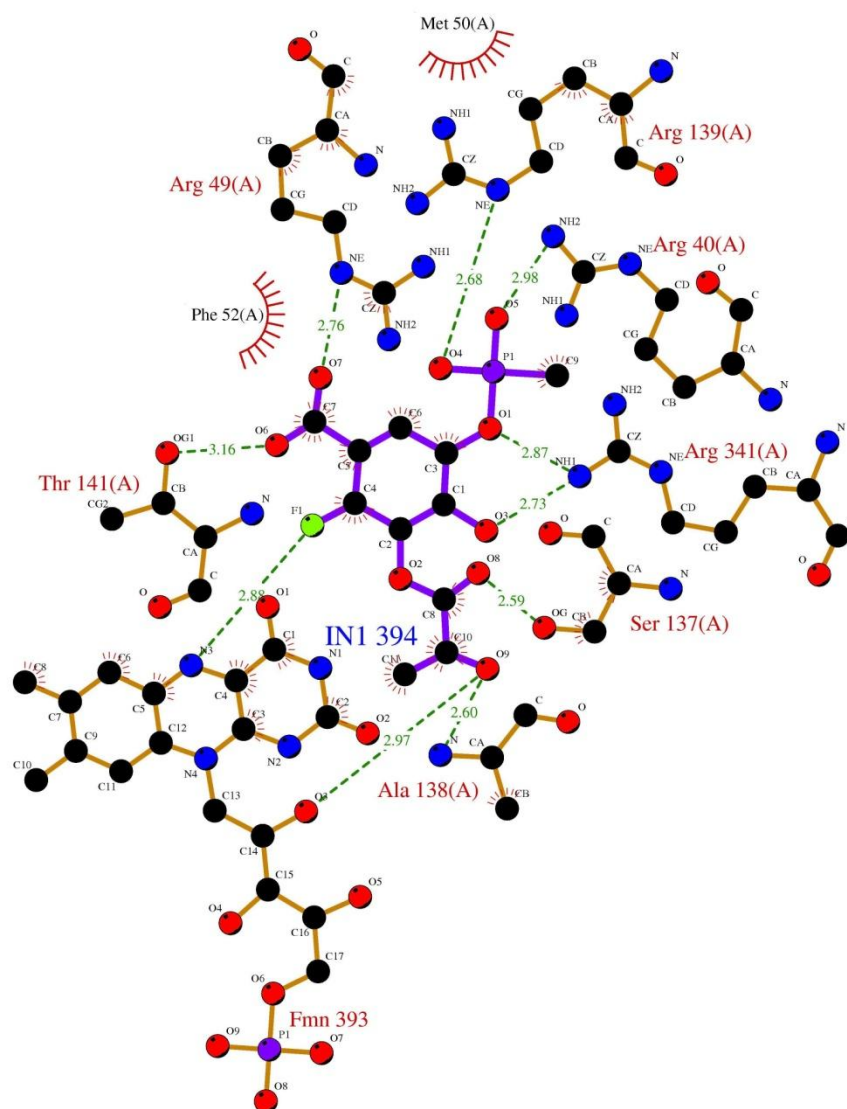


Figure 11

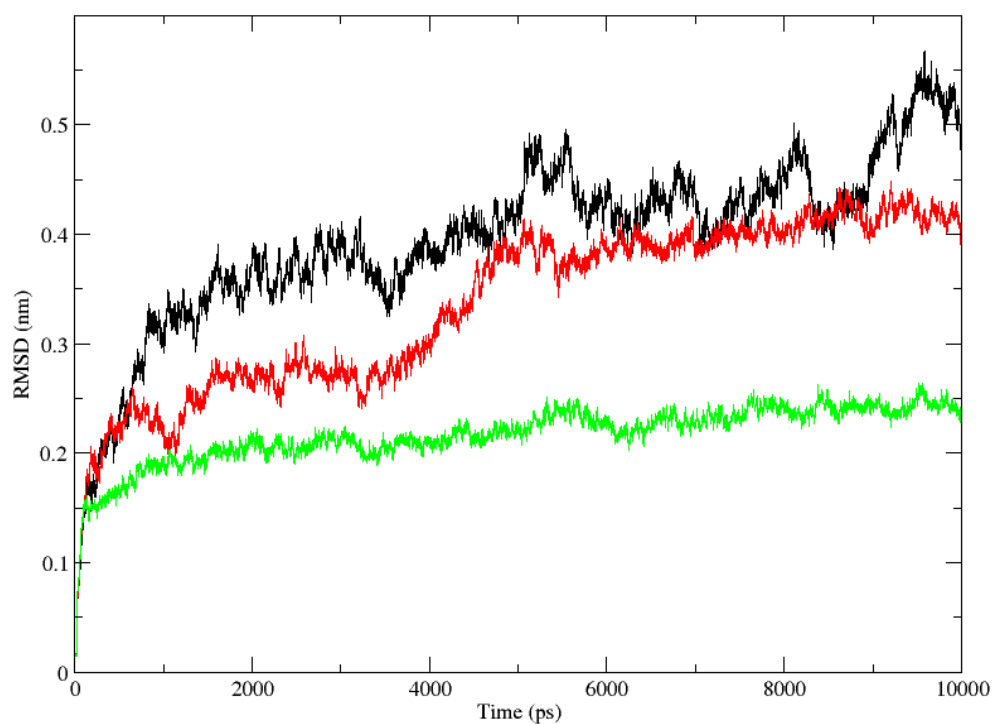


Figure 12

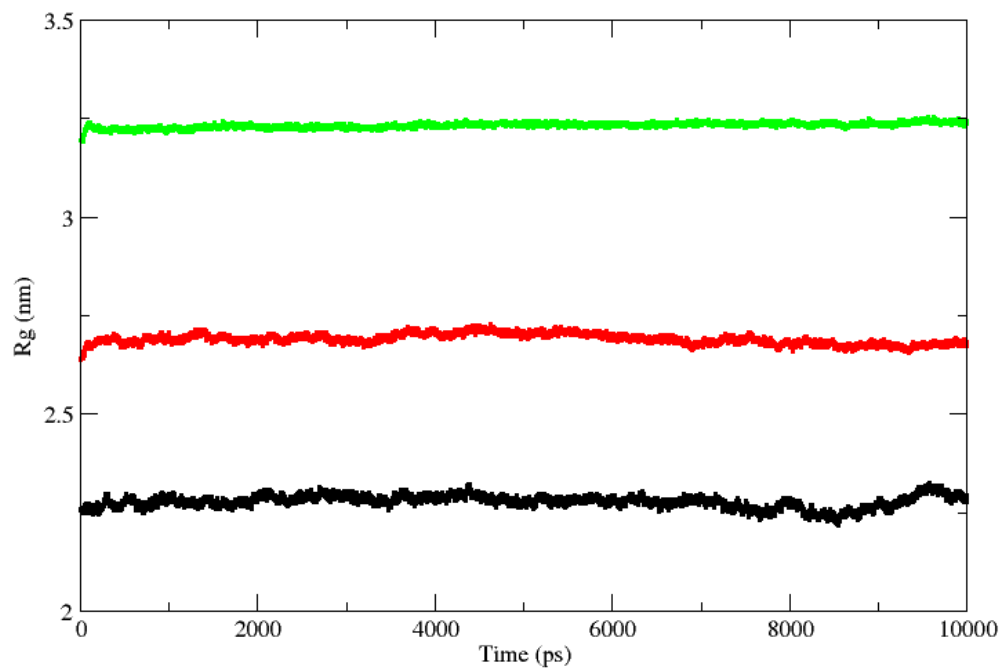
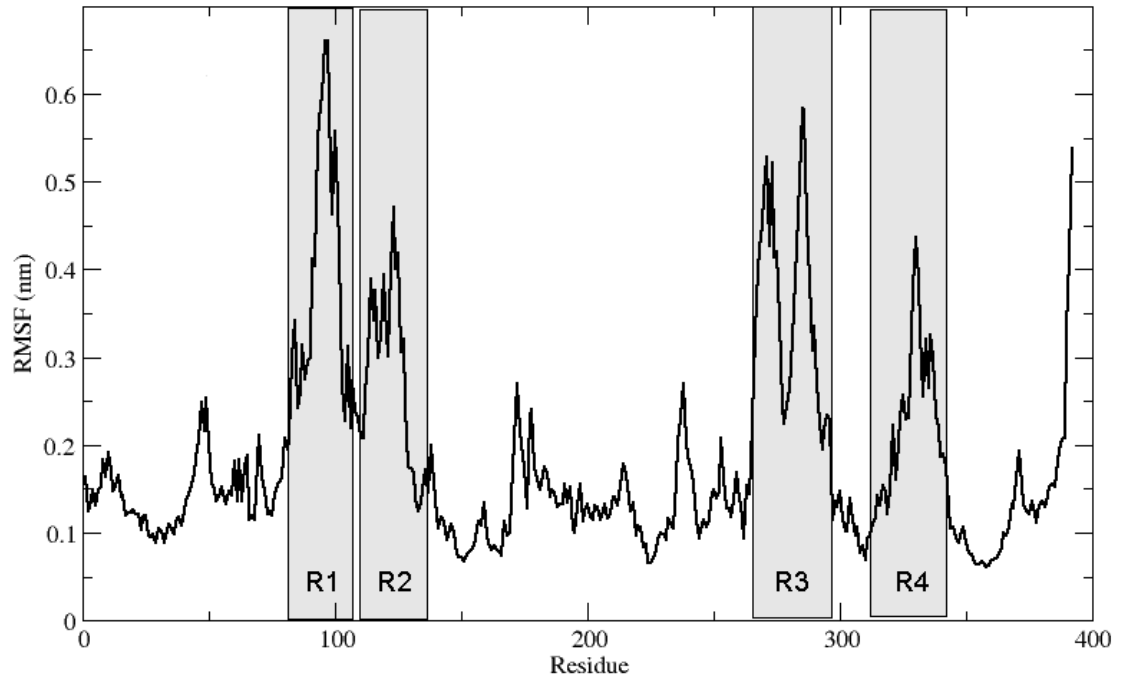
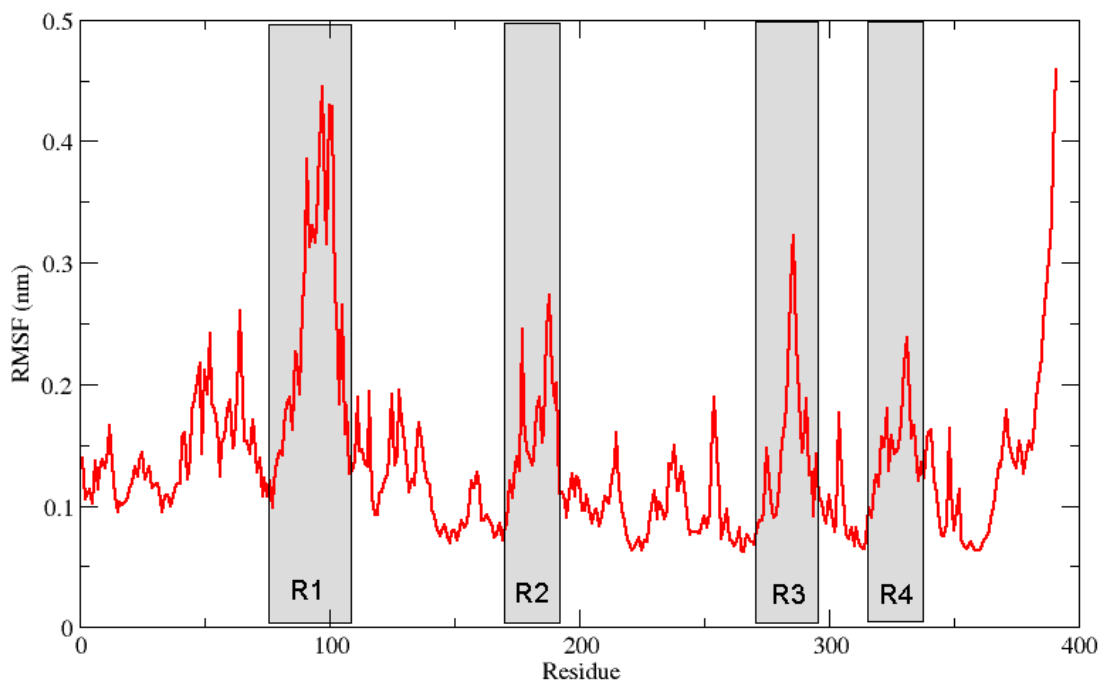


Figure 13

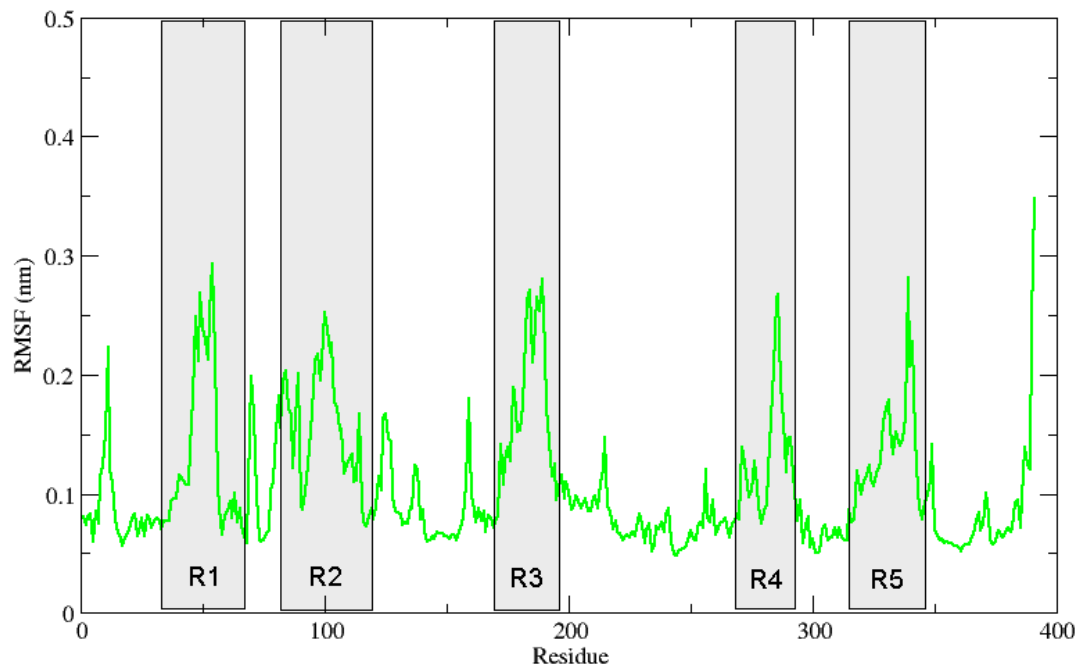
a)



b)



c)



d)

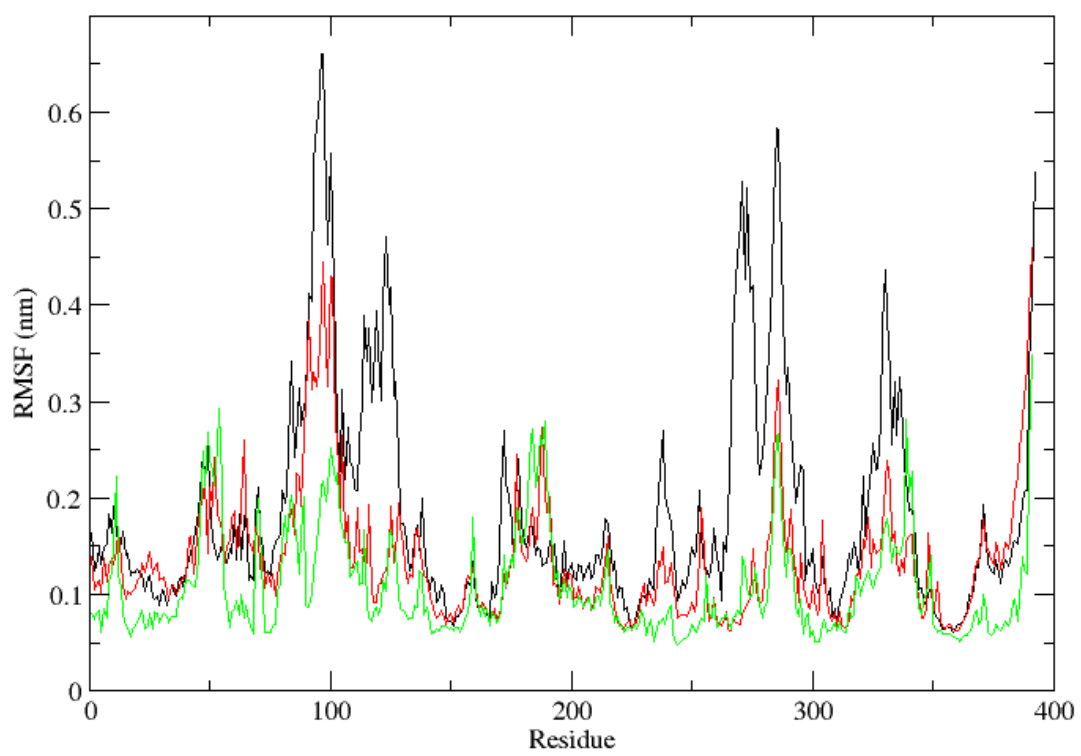


Figure 14

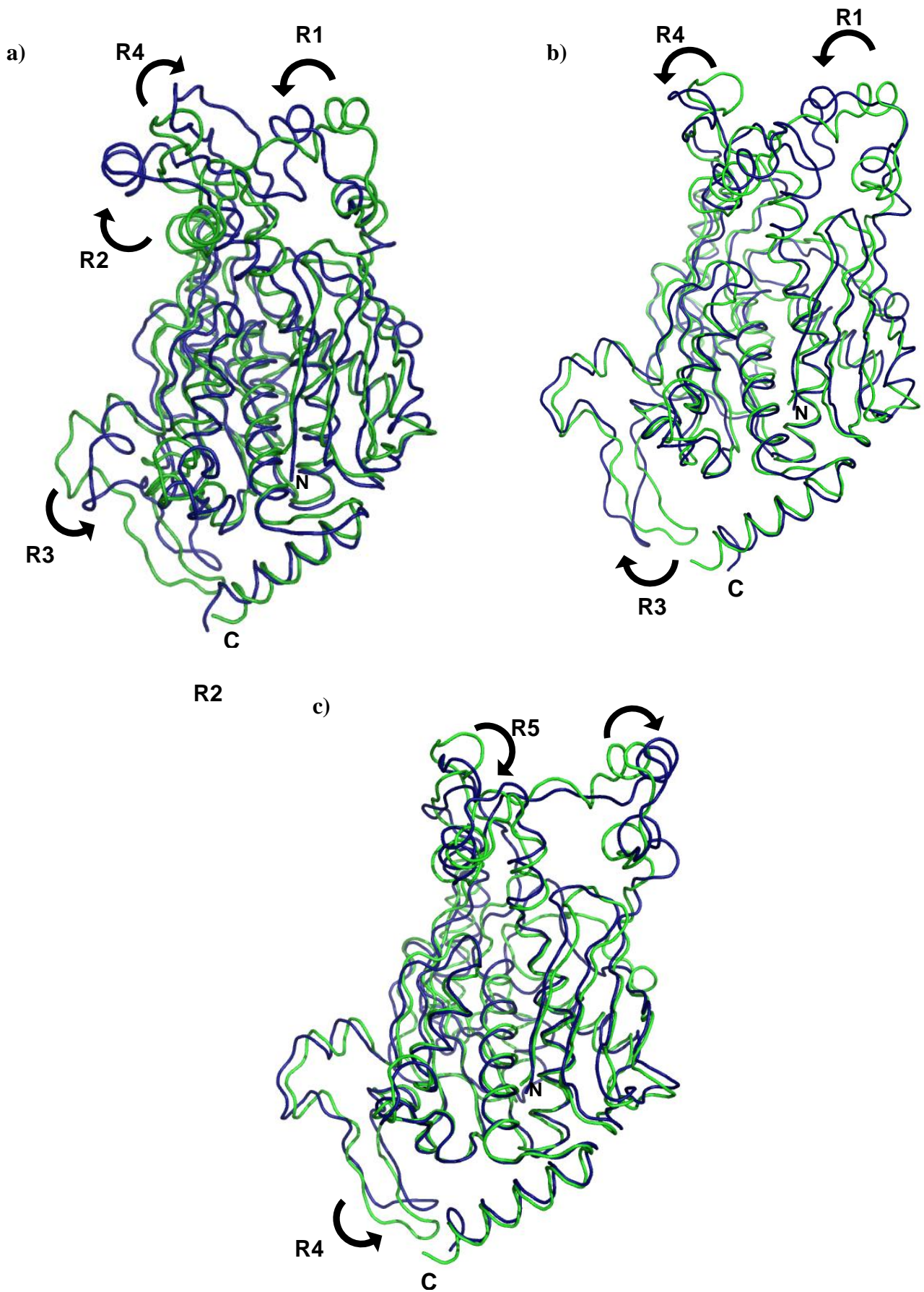


Figure 15

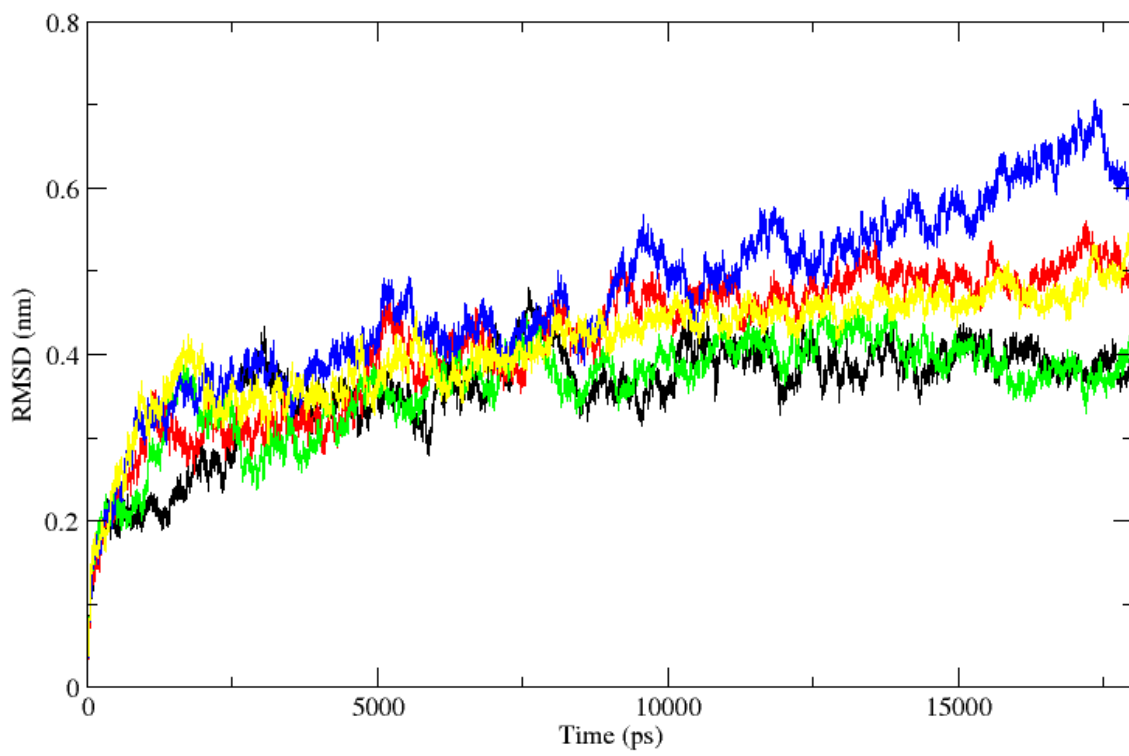


Figure 16

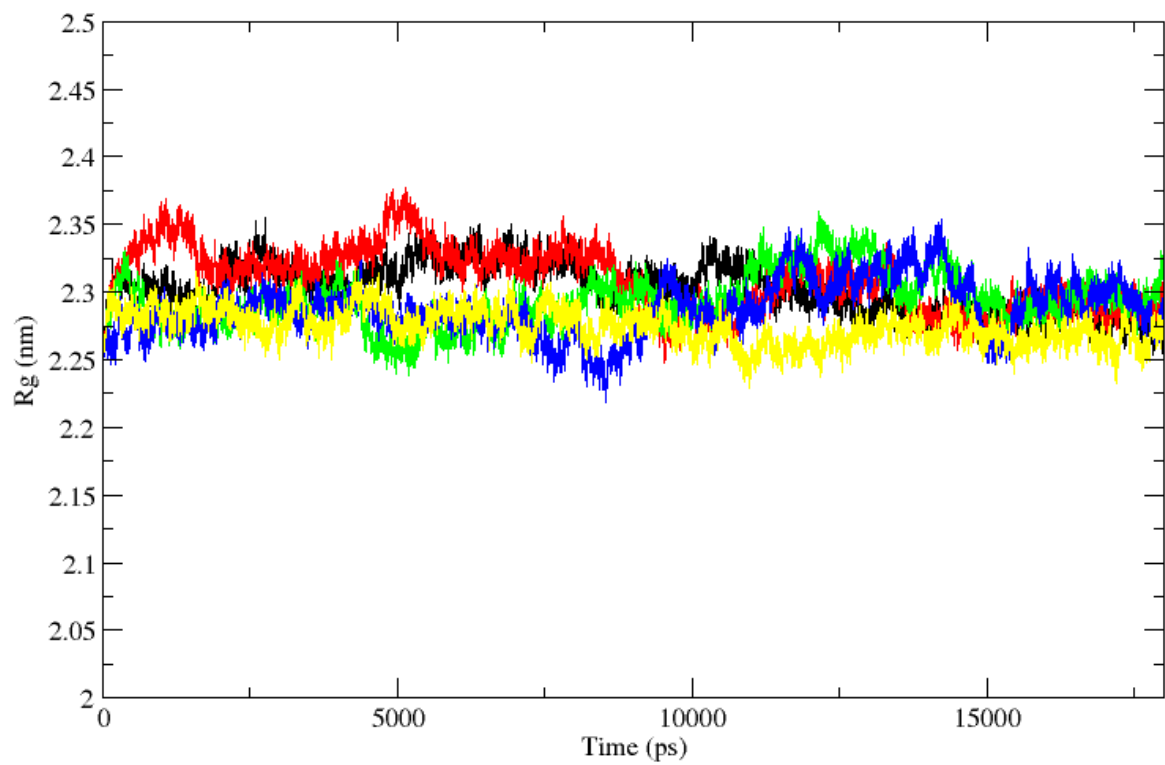


Figure 17

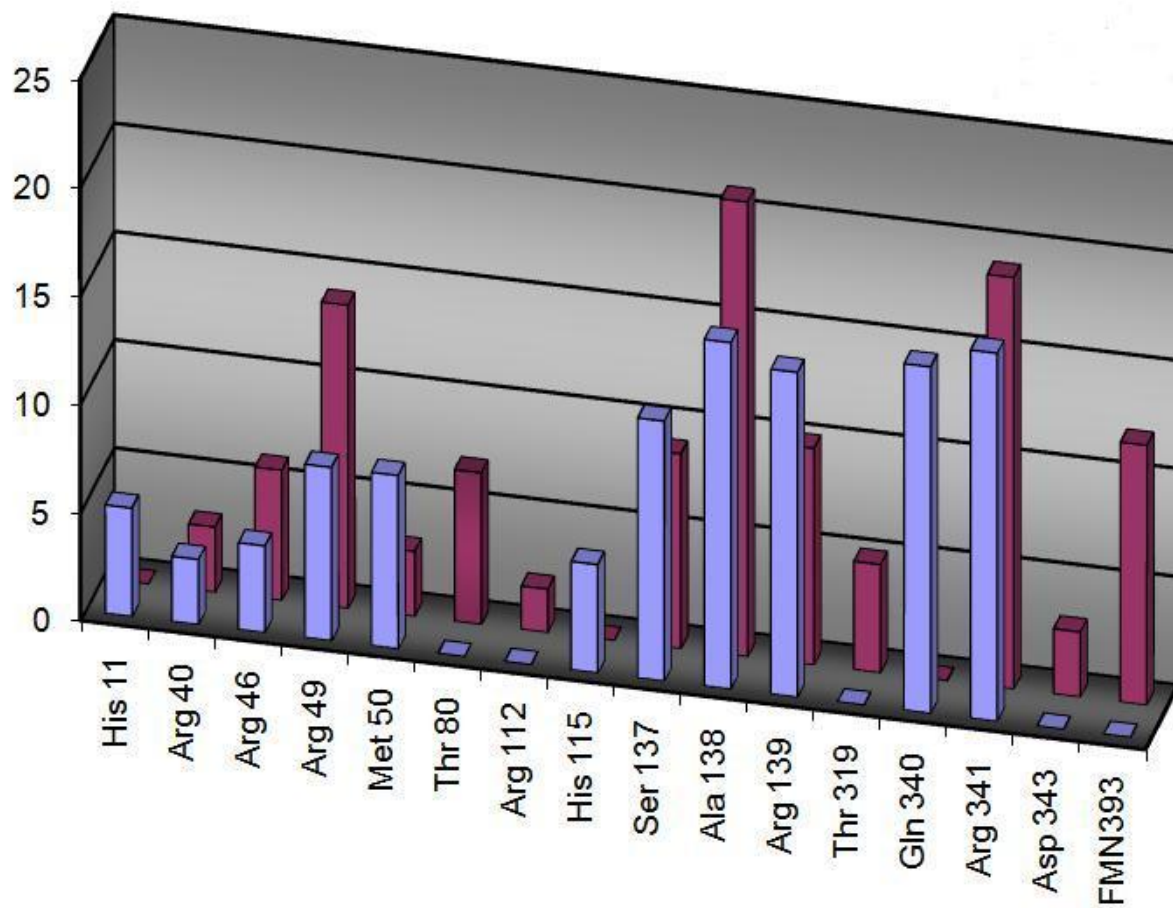
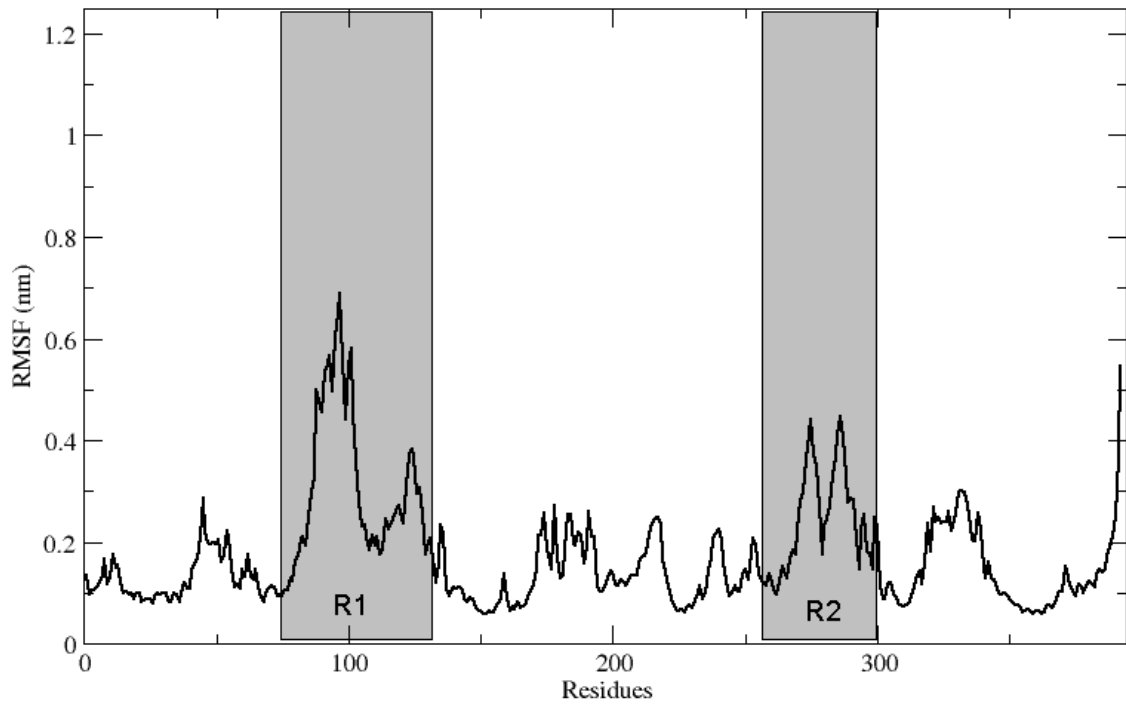
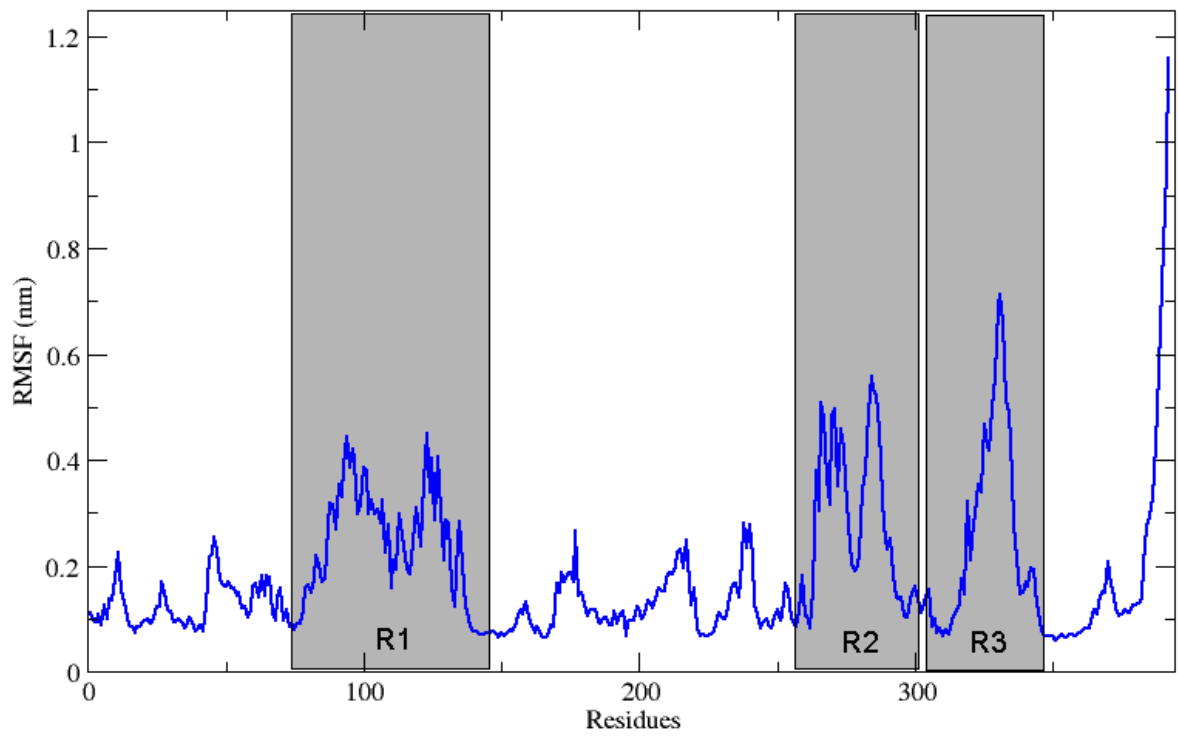


Figure 18

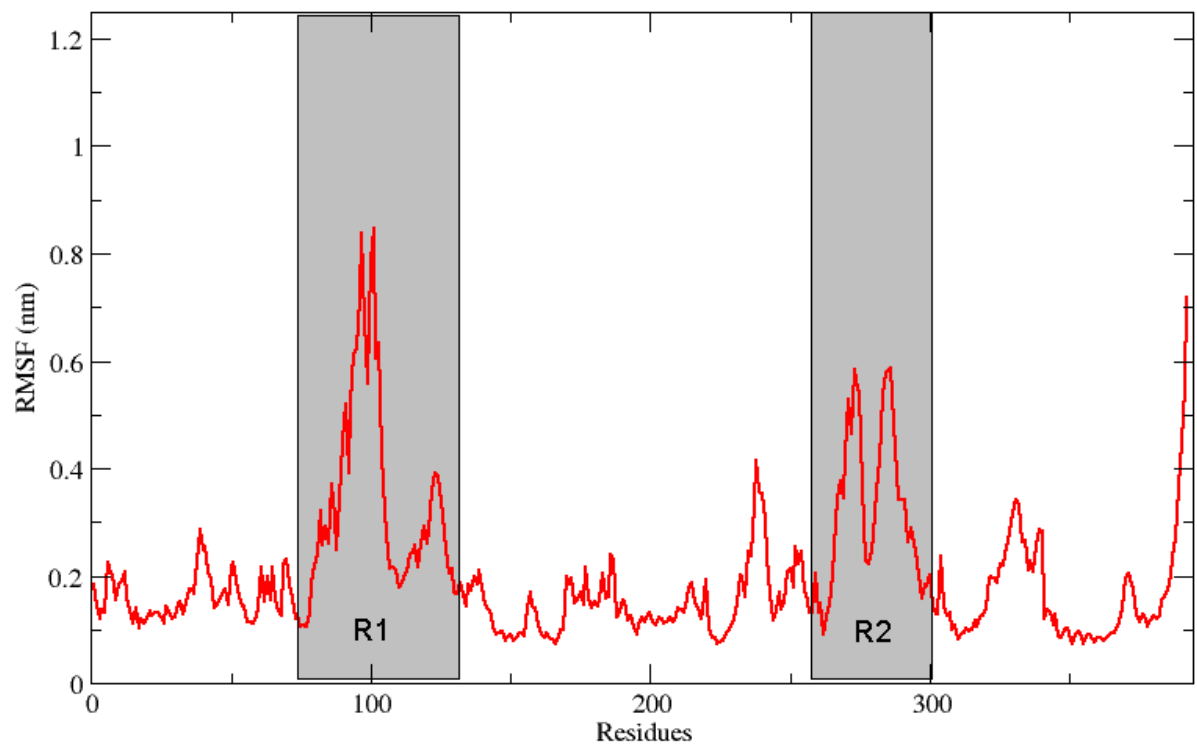
a)



b)



c)



d)

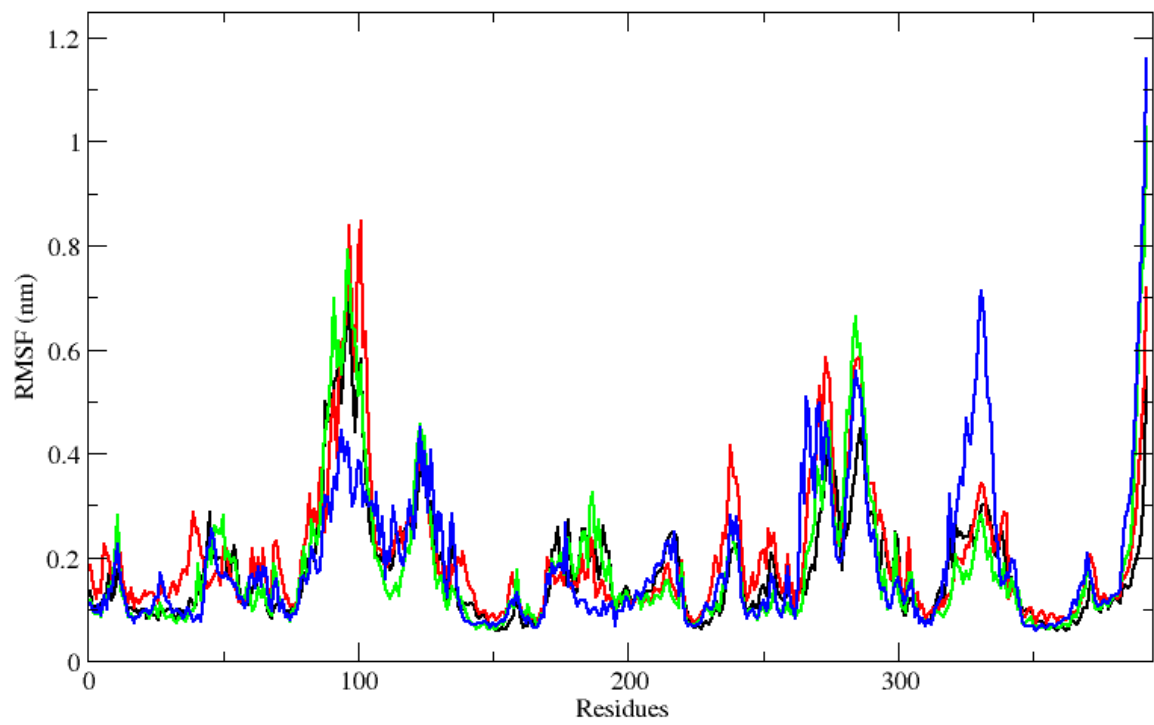


Table 1

	Model		Template	
Residues in most favoured regions [A,B,L]	314	96%	303	94.7%
Residues in additional allowed regions [a,b,l,p]	12	3.7%	16	5.0%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%	0	0.0%
Residues in disallowed regions	1	0.3%	1	0.3%
Number of non-glycine and non-proline residues	327	100%	320	100%
Number of end-residues (excl. Gly and Pro)	2		4	
Number of glycine residues	42		41	
Number of proline residues	21		21	
Total number of residues	392		386	

Table 2

PDB: 1QXO	Model of <i>MtCS</i>
HIS 10	HIS 11
(h) ARG 39	(h) ARG 40
ARG 45	ARG 46
ARG 48	ARG 49
MET 49 (with selenium)	MET 50 (does not appear in 4 Å)
GLY 109	GLY 114
HIS 110	HIS 115
SER 131	ALA 136
SER 132	SER 137
ALA 133	ALA 138
(h) ARG 134	(h) ARG 139
(h) THR 137	
ILE 250 (does not appear in 4 Å)	ILE 255
ASN 251	GLN 256
(s) ALA 252	(s) ALA 257
(s) MET 310 (with selenium)	(s) MET 314 (does not appear in 4 Å)
(s) LYS 311	(s) LYS 315
ILE 313	ILE 317
PRO 314	SER 318
THR 315	THR 319
GLU 336 (does not appear in 4 Å)	GLN 340
ARG 337	ARG 341
SER 338	SER 342
ASP 339	ASP 343
ALA 342	ALA 346
ALA 345	ALA 349 (does not appear in 4 Å)

Table 3

Ligand A			Ligand B			Ligand C		
Residue	Interaction	Atoms	Residue	Interaction	Atoms	Residue	Interaction	Atoms
HIS 11	VW		SER 10	HB	OG-O1	ARG 40	HB	NH2-O5
ARG 49	VW		ARG 49	VW		ARG 49	HB	NE-O7
MET 50	VW		PHE 52	VW		MET 50	VW	
PHE 52	VW		HIS 115	HB	NE2_N1	PHE 52	VW	
HIS 115	HB	NE2-N2	LEU 133	VW		SER 137	HB	OG-O8
LEU 133	VW		SER 137	HB	OG-O3	ALA 138	HB	N-O9
ARG 139	VW		ALA 138	HB	N-O3	ARG 139	HB	NE-O4
GLN 340	VW		ARG 139	HB	N-O3	THR 141	HB	OG1-O6
ARG 341	VW		GLN 340	VW		ARG 341	HB	NH1-O3
FMN (393)	HB	O3-N3	ARG 341	VW		ARG 341	HB	NH1-O1
			FMN (393)	VW		FMN (393)	HB	O3-O9
						FMN (393)	HB	N3-F1

Table 4

Compound	Chemical formula	Molecular weight (g/mol)	PSA (Å ²)	LogP	AHB	DHB	pK _a	MolDock Score
Ligand A	C ₂₁ H ₂₂ N ₇ O ₄ ⁺	436,4	119	-0.88	8	1	7,29675	-214,744
Ligand B	C ₂₂ H ₂₁ NO ₃	347,4	59.2	4.6	3	1	7,6304	-180,995
Ligand C	C ₁₀ H ₁₂ FO ₁₀ P	342,1	171	-1.9	11	5	5,67831	-150,823
EPSP	C ₁₀ H ₁₃ O ₁₀ P	324,0	171	-1.8	10	5	6,04056	-160,066

4. RESULTADOS E DISCUSSÃO

4.1 Construção do complexo ternário *Mt*CS:FMN:EPSP

O ponto de partida do trabalho foi a construção e validação da estrutura quaternária *Mt*CS:FMN:EPSP que seria utilizada para o VS e para a DM.

A busca e identificação da estrutura cristalográfica utilizada como molde foi realizada no PDB (Berman *et al.*, 2002). Dentre as doze estruturas de CS de diferentes organismo depositadas com diversas resoluções e conformações (Tabela 1), cinco correspondem ao *Mt*, sendo apenas o PDB código de acesso: 2O12 (Bruning *et al.*, em publicação), complexada com o cofator FMN, tornando-se assim, o selecionado. Porém, a estrutura escolhida apresentava dois problemas. O primeiro era a ausência de 15 resíduos de aminoácido - do 47 ao 52 e mais 9 resíduos da porção C-terminal, que provavelmente não apresentaram densidade eletrônica durante a resolução da estrutura cristalográfica e foram omitidos do modelo final. O outro problema era em relação a localização do sítio ativo do substrato EPSP, nosso alvo de docagem molecular por apresentar toxicidade seletiva, como já descrito anteriormente. Para a resolução destes problemas, primeiramente foi feito uma modelagem por homologia e logo após, docagem geométrica ou sobreposição de estruturas, seguido de uma minimização de energia do complexo ternário *Mt*CS:FMN:EPSP.

4.1.1 Modelagem por homologia

Como o objetivo da modelagem era a reconstrução da estrutura 3D da *MtCS* (PDB: 2O12/molde) contendo os resíduos GLY 47, ALA 48, ARG 49, MET 50, THR 51 e PHE 52, a própria seqüência primária da enzima, com 401 aminoácidos foi utilizada (alvo) (NCBI: NP_217056.1). O molde foi utilizado sem as moléculas de água da estrutura cristalográfica e em sua conformação apo (ou seja, sem ligantes), que por sua vez, apresenta 386 resíduos de aminoácidos, sendo que dos 15 resíduos ausentes, apenas os 6 citados acima, foram modelados (Figura 7). Por ser uma região bastante flexível, com um número expressivo de aminoácidos faltantes (9) e com peso mínimo no objetivo final deste trabalho, a região C-terminal não foi modelada.

O programa MODELLER 9v7 (Sali & Blundell, 1993; Fiser *et al.*, 2000; Marti-Renom *et al.*, 2000, Eswar *et al.*, 2006) foi utilizado para a construção dos modelos da enzima, utilizando o protocolo padrão de metodologia de modelagem comparativa de proteínas (Anexo 1).

Foram gerados 1000 modelos. O melhor resultado selecionado, ranquiado pela função energia do programa, foi novamente submetido a modelagem, mas apenas para o refinamento do *loop*, onde mais 50 modelos foram gerados.

A estrutura final obtida, foi avaliada através do pacote de análise PROCHECK (Laskowski *et al.*, 1993), que avalia a qualidade estereoquímica do modelo pelo *web server* SAVES (*Structural Analysis and Verification Server*) do NIH MBI *Laboratory for Structural Genomics and Proteomics*.

A qualidade estereoquímica da estrutura modelada apresentou resultados melhores do que da estrutura cristalográfica utilizada como molde, que também foi

analisada como controle. O gráfico de Ramachadran (Figura 8) onde são medidos os ângulos de rotação *phi* e *psi*, apresentou 314 ou 96% dos resíduos do modelo em regiões favoráveis, contra 303 ou 94,7% (Tabela 3). Além disso, foram feitas análises da cadeia principal e lateral, com intervalos de confiança conforme os parâmetros do programa, e que apresentaram resultados dentro da média ou melhores.

A verificação do deslocamento do modelo em relação ao molde também foi feita, obtendo-se um RMSD de 0,15 Å.

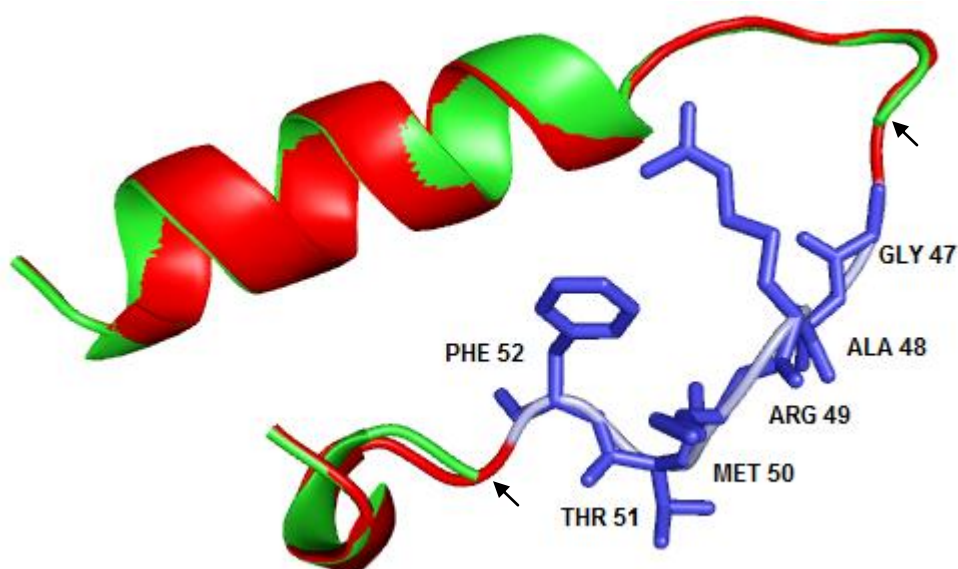


FIGURA 7. Representação da estrutura primária e secundária 3D de *MtCS* na região modelada sobreposta a estrutura cristalográfica que serviu como molde. Setas destacando a descontinuidade da estrutura cristalográfica e ausência dos 6 aminoácidos. (Verde: molde (PDB: 2012); vermelho: estrutura do modelo final; cinza: região modelada representando sua estrutura secundária; azul: região modelada representando sua estrutura primária).

Para predição da estrutura secundária dos resíduos modelados, foi utilizado o *web server* APSSP2 (Raghava, 2000) onde foram obtidos resultados de relação entre a estrutura secundária do aminoácido e a probabilidade de predição correta. Todos os 6 resíduos apresentaram como forma de estrutura secundária mais provável *coil* (Tabela 2), corroborando com a estrutura modelada.

TABELA 2. Resultado da predição da estrutura secundária dos resíduos modelados gerados pelo *web server* APSSP2.

Resíduo de aminoácidos	Estrutura secundária	Índice de Predição (até 1)
GLY 47	C (<i>Coil</i>)	1.0
ALA 48	C (<i>Coil</i>)	0.9
ARG 49	C (<i>Coil</i>)	0.9
MET 50	C (<i>Coil</i>)	1.0
THR 51	C (<i>Coil</i>)	0.8
PHE 52	C (<i>Coil</i>)	0.6

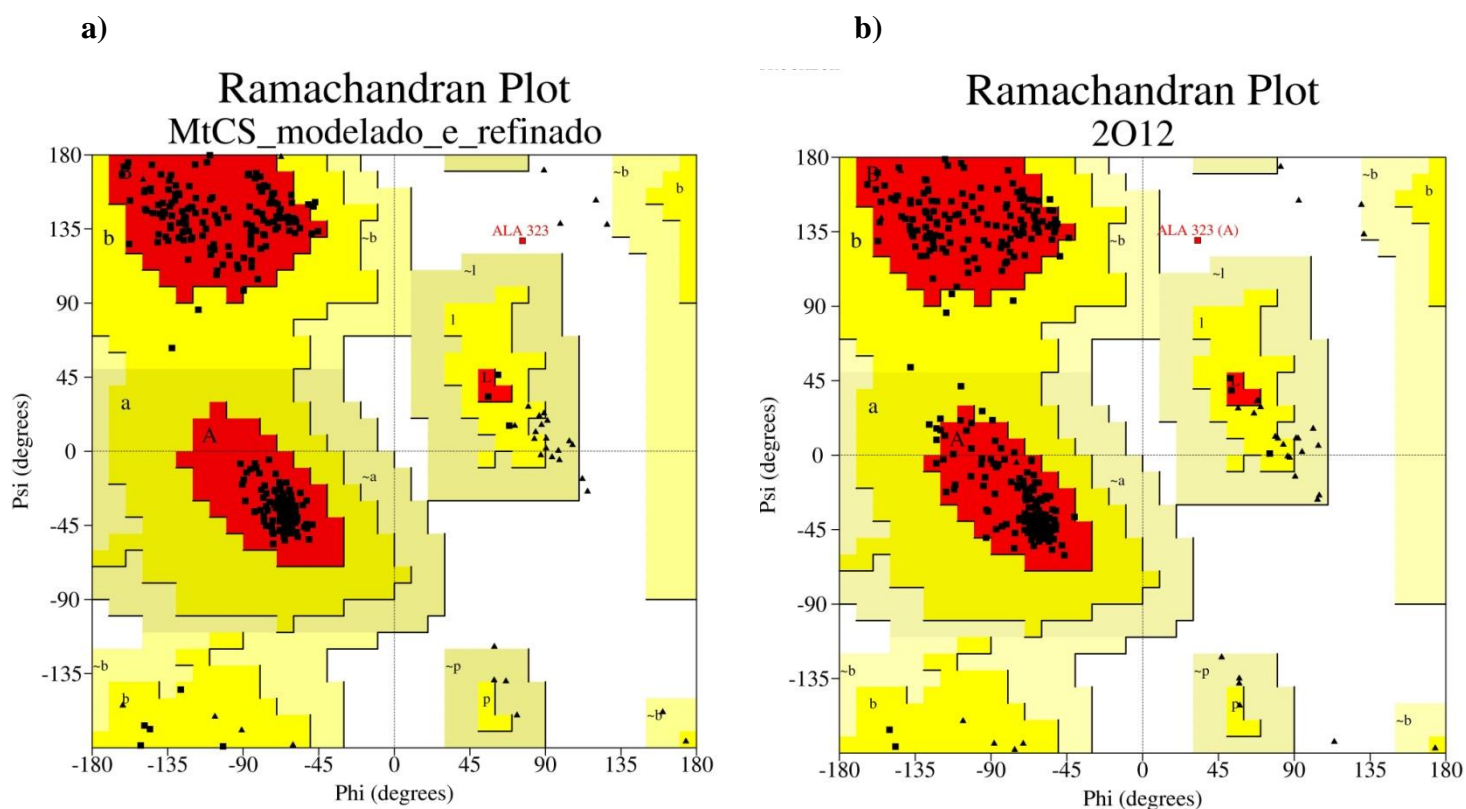


FIGURA 8. Gráficos de Ramachadran – a) modelado e b) molde (estrutura cristalográfica PDB: 2012).

TABELA 3. Resultado dos gráficos estatísticos de Ramachadran do modelo e do molde.

	Modelo		Molde	
Resíduos nas regiões mais favoráveis [A,B,L]	314	96%	303	94.7%
Resíduos nas regiões adicionalmente favoráveis [a,b,l,p]	12	3.7%	16	5.0%
Resíduos nas regiões generosamente permitidas [~a,~b,~l,~p]	0	0.0%	0	0.0%
Resíduos nas regiões desfavoráveis	1	0.3%	1	0.3%
Número de resíduos não-glicina e não-prolina	327	100%	320	100%
Número de resíduos terminais (exceto glicina e prolina)	2		4	
Número de resíduos de glicina (triângulos)	42		41	
Número de resíduos de prolina	21		21	
Número total de resíduos	392		386	

4.1.2 Docagem geométrica:

Após a modelagem da estrutura 3D da CS, o próximo passo foi a localização dos sítios ativos do substrato EPSP e do cofator FMN a partir da docagem geométrica manual e a construção do complexo ternário *MtCS:FMN:EPSP*.

Este processo foi realizado mediante a sobreposição dos carbonos α da estrutura modelada com as cristalográficas: 2O12 (para a localização do sítio ativo da FMN) e 1QXO (para a localização do sítio ativo do EPSP), através do programa SPDBV (Guex & Peitsch, 1997). A escolha da estrutura 1QXO se deu baseada na identidade acima de 40% das seqüências de *MtCS* e *SpCS* (CS de *Streptococcus pneumoniae*) e no fato de ser a única estrutura cristalográfica de CS complexada com o substrato EPSP.

O modelo final gerado *MtCS:FMN:EPSP* foi submetido a uma minimização de energia ou otimização de geometria através do pacote de programas GROMACS 4.0.5 (Van der Spoel *et al.*, 2005; Hess *et al.*, 2008) para corrigir possíveis choques estereoquímicos e contatos desfavoráveis entre os resíduos do monômero da proteína e ligantes.

Posteriormente, as estruturas quartenárias foram geradas, sobrepondo a unidade modelada com cada uma das cadeias da estrutura tetramérica 1QXO. O deslocamento de cada monômero, em relação a estrutura cristalográfica utilizada, obtidos pelo RMSD, foi de $\approx 1,8 \text{ \AA}$.

4.1.2.1 Análise dos sítios ativos do FMN e do EPSP

Após a docagem geométrica, pode-se observar que orientações de importante resíduos do sítio ativo do EPSP, como a ARG 341, possivelmente estariam fazendo choque estereoquímico com o substrato, exatamente pelo fato da estrutura cristalográfica utilizada como molde (2O12) ter sido cristalizada sem o EPSP. Este problema foi corrigido durante a fase de minimização de energia, que tinha exatamente este propósito (Figura 9).

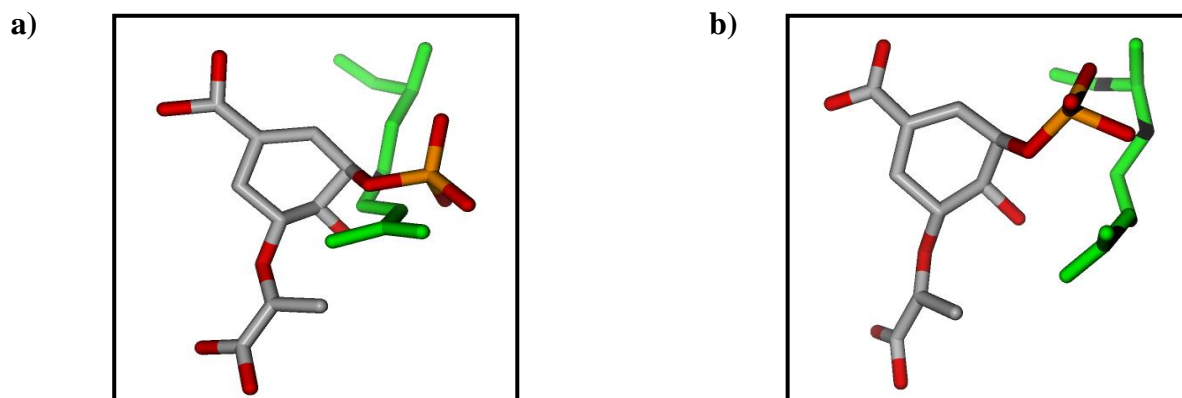


FIGURA 9. EPSP (colorido por CPK) e o resíduo ARG 341 (verde) - a) antes da minimização de energia demonstrando o possível choque estereoquímico; b) após a minimização de energia.

No resultado final da estrutura modelada, a ARG 341, bem como a ARG 49, são os que aparecem mais deslocados em relação a estrutura cristalográfica 1QXO, que serviu de referência para a localização do sítio ativo do EPSP na estrutura de CS de *Mt*. Esta talvez seja umas das explicações para a diferença entre o número de ligações de hidrogênio deste resíduos com a porção fosfato do EPSP de *SpCS* (ARG 337), que são apenas duas, enquanto em *MtCS*, a ARG 341 faz quatro ligações. Além disso, de acordo com estudos anteriores, a ARG 341 também faz parte do *loop* 29 de *MtCS* (entre os resíduos ILE 338 e ALA 346) que possui um papel fundamental na abertura/fechamento do sítio do cofator FMN (Dias *et al.*, 2006).

Já as ARG 45 e ARG 48, que em *MtCS* correspondem as ARG 46 e ARG 49, fazem parte do *loop* 5 (*MtCS*) (Figura 10) e são importantes resíduos do sítio ativo em *SpCS*, fazendo 3 e 2 ligações de hidrogênio com o EPSP, respectivamente. Porém em *MtCS*, apenas a ARG 49 aparece e fazendo somente contato hidrofóbico.

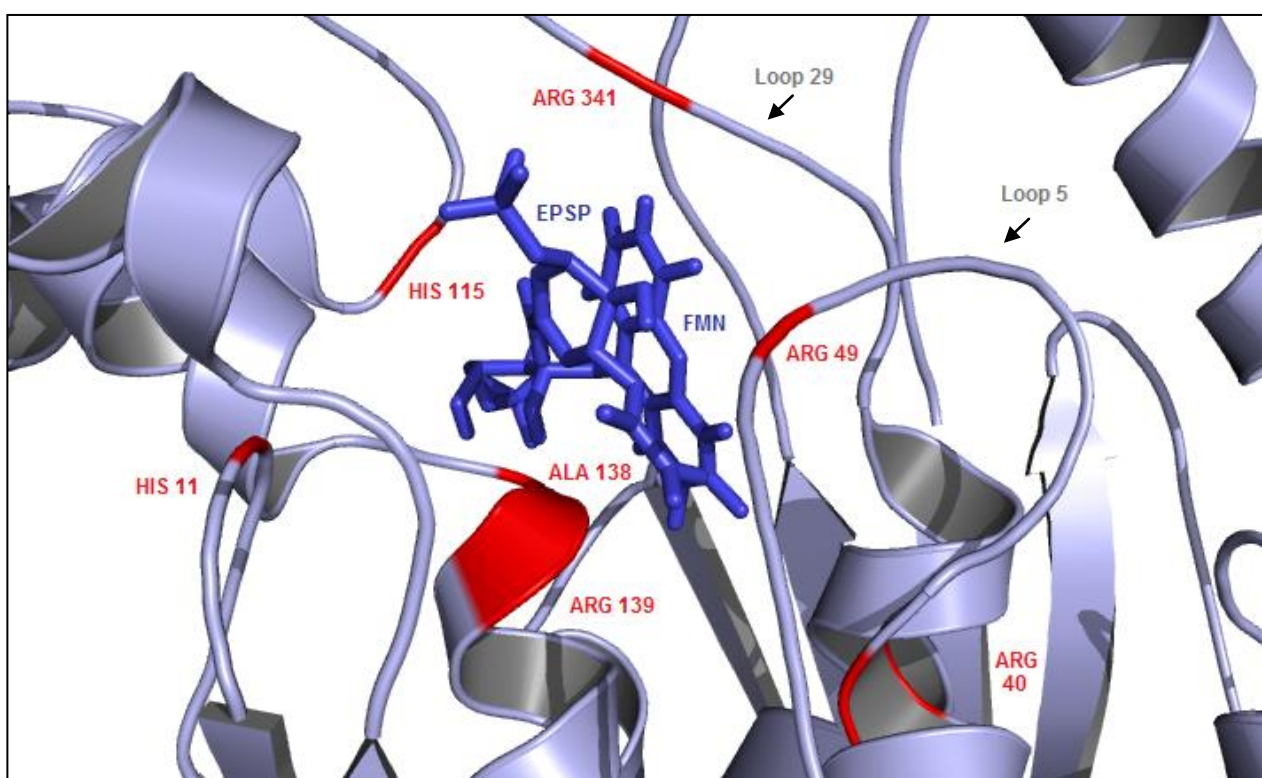


FIGURA 10. Representação da estrutura secundária em 3D de *MtCS* (cinza) dando ênfase ao sítio ativo da enzima. Setas destacando os dois principais *loops* envolvidos. (Azul: cofator FMN e substrato EPSP; vermelho: localização na estrutura secundária dos principais resíduos que fazem parte do sítio ativo).

Os resíduos ALA 138 (ALA 133 em *SpCS*) e ARG 40 (ARG 39 em *SpCS*), bem como o cofator FMN, aparecem fazendo as mesmas ligações com o EPSP, apenas com comprimentos sutilmente diferentes. Somente a ARG 139 (ARG 134 em *SpCS*) aparece fazendo contato hidrofóbico com o EPSP em *MtCS*, ao invés de ligação de hidrogênio, como visto em *SpCS*. Isto pode ser devido ao pequeno deslocamento deste resíduo em relação a estrutura de *SpCS*.

Uma importante particularidade vista na docagem geométrica, é em relação adjacência do sítios ativos do cofator FMN e do substrato EPSP (Figura 10). As evidências de experimentos (Macheroux *et al.*, 1998) com CS de *E. Coli* em relação ao K_d do FMN oxidado na presença ou não do EPSP, validam a tese de que as ligações nos sítios são coordenadas – primeiro a entrada do FMN, na parte mais aprofundada do sítio, seguida pela ligação do substrato, que além de impedir a saída do cofator, possui mais contato com o solvente (Maclean & Ali, 2003).

Outra característica relevante também observada, é a presença, no sítio docado geometricamente, dos dois principais resíduos, HIS 11 e HIS 115, considerados essenciais para catálise da reação e conservados nas seqüências primárias de CS de diversos organismos, sendo que ambos aparecem nas mesmas posições dos resíduos correspondentes em *SpCS*, HIS 10 e HIS 110.

A alteração de três aminoácidos SER 131 → ALA 136; ASN 251 → GLN 256 e PRO 314 → SER 318, corroborando com estudos anteriores (Fernandes *et al.*, 2007), também foram observadas.

No Ligplot v.4.0 (Wallace *et al.*, 1995), que é um programa que implementa a equação de distância e identifica os pares dos átomos dos resíduos que fazem ligação de hidrogênio entre a proteína e os ligantes, podemos visualizar todas essas observações, bem como as semelhanças entre os sítios ativos de *MtCS* e *SpCS* (Figura 11).

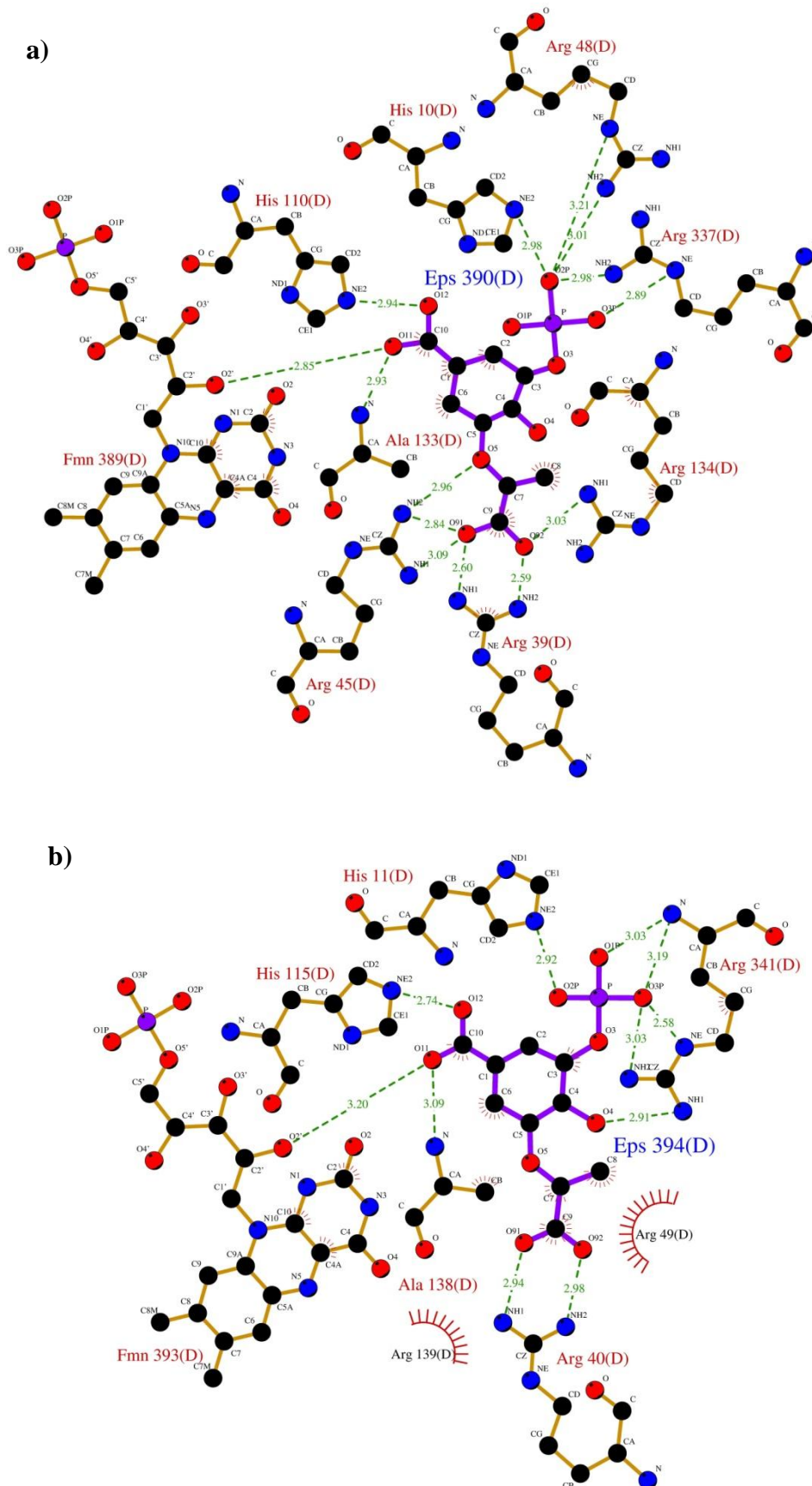


FIGURA 11 Ligplot - demonstrando as ligações de hidrogênio e os contatos hidrofóbicos feitos pelo EPSP com a estrutura de *SpCS* (a) e *MtCS* (b). (Representação: átomos de oxigênio (esferas vermelhas); átomos de carbono (esferas pretas); átomos de nitrogênio (esferas azuis); ligações de hidrogênio (linhas tracejadas em verde); ligações entre os átomos do ligante (linhas em lilás); ligações entre os átomos do FMN (linhas em caramelo); contato hidrofóbicos (semicírculos em vermelho))

Os principais resíduos dos sítios ativos do cofator FMN e do substrato EPSP da estrutura modelada e minimizada de *Mt* e da cristalográfica de *Sp* estão demonstrados na tabela 4.

TABELA 4. Resíduos localizados até 4 Å nos sítios ativos da FMN e EPSP das estruturas cristalográficas de *SpCS* (1QXO) com 388 aminoácidos e *MtCS* (modelo final - modelada e minimizada) com 392 aminoácidos; (h) – resíduos que fazem parte de hélices- α ; resíduos destacados em negrito aparecem também no Ligplot.

PDB: 1QXO	Modelo de <i>MtCS</i>
HIS 10	HIS 11
(h) ARG 39	(h) ARG 40
ARG 45	ARG 46
ARG 48	ARG 49
MET 49 (com selênio)	MET 50 (não aparece a 4 Å)
GLY 109	GLY 114
HIS 110	HIS 115
SER 131	ALA 136
SER 132	SER 137
ALA 133	ALA 138
(h) ARG 134	(h) ARG 139
(h) THR 137	
ILE 250 (não aparece em 4 Å)	ILE 255
ASN 251	GLN 256
(s) ALA 252	(s) ALA 257
(s) MET 310 (com selênio)	(s) MET 314 (não aparece a 4 Å)
(s) LYS 311	(s) LYS 315
ILE 313	ILE 317
PRO 314	SER 318
THR 315	THR 319
GLU 336 (não aparece em 4 Å)	GLN 340
ARG 337	ARG 341
SER 338	SER 342
ASP 339	ASP 343
ALA 342	ALA 346
ALA 345	ALA 349 (não aparece em 4 Å)

4.2 Redocagem da estrutura cristalográfica

A simulação de redocagem visa recuperar, a partir da simulação computacional, a posição original de um ligante presente numa estrutura cristalográfica de um complexo binário proteína-ligante. Partindo da hipótese que a posição do ligante no complexo é a de menor energia, os algoritmos que simulam a busca, devem apresentar potencial de recuperar a posição cristalográfica deste ligante. Estas posições resultantes da simulação de docagem são chamadas de *poses*. Espera-se que numa simulação que obteve sucesso, a *pose* de menor energia seja aquela mais próxima da posição cristalográfica do ligante.

Este método é utilizado como critério de avaliação de acurácia preditiva do programa e são utilizados os valores de RMSD, das coordenadas atômicas cristalográficas comparadas as poses obtidas no redocagem. Valores de RMSD menores que 0,5 Å indicam que a simulação de redocagem teve sucesso, e que o protocolo pode ser usado para o VS (Thomsen & Christensen, 2006).

Como já mencionado e justificado anteriormente, por não haver uma estrutura cristalográfica de CS de *Mt* complexada com o substrato EPSP, para a simulação de redocagem no programa *MolDock* 4.1.0 (Thomsen & Christensen, 2006) foram utilizada as coordenadas 3D da estrutura do monômero D do 1QXO de *SpCS*, sem as águas cristalográficas e sem selênio. A opção de docagem apenas com a forma de monômero, justifica-se pelo fato de que o sítio catalítico do EPSP somente apresenta interação com resíduos do próprio monômero em que se situa, sendo as distâncias entre as unidades do tetrâmero de mais 27 Å. Além disso, segundo Dias e colaboradores (Dias *et al.*, 2006), os monômeros B, C e D de *SpCS*, são os que apresentam maior similaridade com *MtCS*.

O protocolo utilizado restringiu as torções da enzima e do cofator, localizado no sítio adjacente, mas permitiu as do substrato. O centro do raio da esfera que delimita as simulações de docagem foi de 15 Å, sobre o sítio catalítico do EPSP (Figura 12-a). Os demais parâmetros do protocolo utilizado encontram-se no anexo C. Os resultados foram avaliados pela função *re-ranck Score* do programa. O RMSD encontrado foi de $\approx 0,3$ Å na três simulações de redocagem (Figura 12-b), validando o programa e o protocolo, sendo utilizado para o VS.

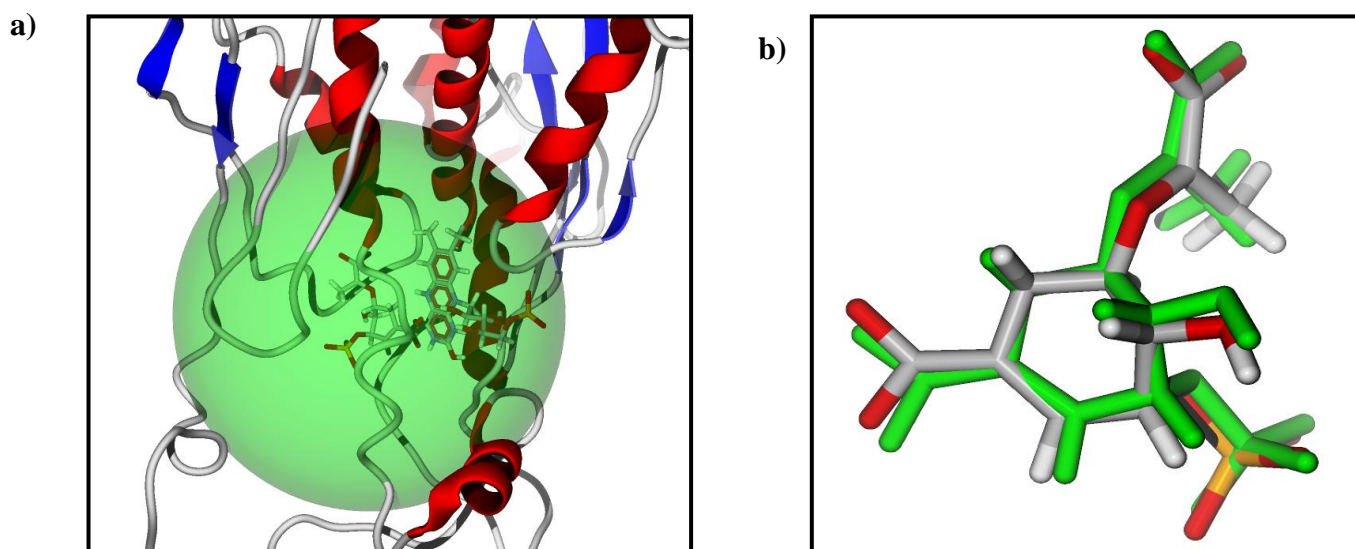


FIGURA 12. a) Estrutura da enzima *MICS-FMN-EPSP* (colorida de acordo com a estrutura secundária: hélices- α em vermelho, fitas- β em azul; voltas e alças em cinza e substrato e cofator coloridos por CPK) e do espaço 3D (esfera verde) que delimita as simulações de docagem no sítio catalítico do EPSP - gerado pelo programa *MolDock*; b) Representação do resultado de redocagem – sobreposição da estrutura cristalográfica do EPSP (colorida por CPK) e da *pose* (verde).

4.3 Virtual screening

Na simulação de VS foram utilizadas as bibliotecas de compostos SIGMA-Aldrich e a ACROS Organics (Irwin & Shoichet, 2005), totalizando 19.615 moléculas, disponibilizadas através do banco de dados ZINC (Irwin & Shoichet, 2005). O protocolo, já estabelecido e validado na redocagem, foi executado no programa *MolDock* 4.1.0 (Thomsen & Christensen, 2006).

Os resultados foram ranquiados e pré-selecionados através da função intrínseca do programa *re-ranck Score* (Thomsen & Christensen, 2006). Após, os que apresentaram melhor energia foram filtrados através da utilização dos critério de satisfação das regras de Lipinski (Lipinski *et al.*, 2001) e de toxicidade – ADMETox (Miteva *et al.*, 2006), ou seja, apresentavam características de toxicidade e farmacocinética aceitáveis.

As moléculas selecionadas foram analisadas então no SEA (*Similarity Ensemble Approach*) (Keiser *et al.*, 2007), um servidor que correlacionam as propriedades químicas de um composto com um banco de dados de moléculas com atividade determinada. Além disso, o cálculo do pK_d dos dez melhores resultados também foi feito, através do programa POLSCORE (De Azevedo & Dias, 2008; Dias *et al.*, 2008), para estimar o coeficiente de dissociação do complexo.

Como resultado final, duas moléculas foram selecionadas sob o código de ZINC04147896 (denominada de **ligante A**) e a ZINC00005085 (denominada de **ligante B**). Além disso, com o objetivo de correlacionar nossos resultados com dados experimentais, foi feito também, docagem molecular da estrutura (6R)-6-fluoro-5-enoilpiruvilchiquimato 3-fosfato (denominado de **ligante C**), um inibidor de CS de *Escherichia coli* (Osborne *et al.*, 2000).

4.3.1 Ligante A

O ligante A, obtido da triagem da biblioteca de compostos Acros, possui a denominação pela IUPAC (*International Union of Pure and Applied Chemistry*) de *[(3S,3aR,6S,6aR)-6-[5-(4-imidazol-1-ylphenoxy)tetrazol-1-yl]-2,3,3a,5,6,6a-hexahydrofuro[3,2-b]furan-3-yl]-(furan-2ylmethyl) azanium*.

Quando comparadas, as energias de ligação dos ligante A e B, de - 214,744 e -180,995 respectivamente, apresentaram-se melhor do que a energia de ligação do substrato natural EPSP, que foi de \approx -159,954.

Apesar do ligante A ter apresentado uma energia de ligação menor do que o ligante B, o que se pressupõe uma maior afinidade com a enzima, o número de ligações de hidrogênio, como observado na figura 13, foi menor, apenas duas. Entretanto, ambas ligações em resíduos importantes do sítio ativo.

A ligação de hidrogênio do ligante A com o O2' do FMN, também aparece na interação com o ESPS em *MtCS* e *SpCS*, sendo a única ligação entre o cofator e o substrato, parecendo ser importante para a estabilização de um possível inibidor competitivo. Em *MtCS* e *SpCS*, o comprimento da ligação do O11 do EPSP com o O2' do FMN é de 3,20 Å e 2,85 Å respectivamente. Enquanto no ligante A, o comprimento é de 2,89 Å, e a ligação do átomo O2' do FMN é com o N3.

A segunda ligação de hidrogênio do ligante A é com o resíduo HIS 115, entre os átomos N2 e NE2, respectivamente, com a distância de 2,68 Å. Estudos anteriores de mutagênese sítio-direcionadas (Kitzing *et al.*, 2004), demonstraram a importância da HIS 115 na reação de catálise da CS, quando a mutação da HIS 110 (*SpCS*) para uma alanina, reduziu a atividade da enzima em 10%.

Em relação as interações hidrofóbicas, oito resíduos da enzima fazem contato com o ligante A: MET 50, PHE 52, LEU 133, GLN 340, ARG 139, HIS 11, ARG 49 e ARG 341, sendo que os quatro últimos, também aparecem nas interações com EPSP.

Quando comparado com o ligante B, oito dos dez resíduos que fazem interação com a enzima aparecem em comum, sendo que nem todas do mesmo tipo. A ARG 139, por exemplo, que no ligante A faz apenas contato hidrofóbico, no ligante B faz uma ligação de hidrogênio de 3,13 Å. Já em relação ao ligante C, apenas seis resíduos aparecem no sítio ativo dos dois, entretanto a ligação de hidrogênio de ambos com o FMN é no mesmo átomo, o O3, apenas com uma diferença de distância de 2,89 Å (ligante A) para 2,97 Å (ligante C).

As propriedades físico-químicas do ligante A apresentam-se satisfatórias, obedecendo as regras de Lipinski e de toxicidade. O peso molecular deste composto, entre os demais avaliados, é o que se apresenta maior, até mesmo do que o do substrato natural (Tabela 5), mas dentro do limite de 500 Da, condizentes com seus seis anéis. Contudo, sua área de superfície polar está mais próxima da área do EPSP, do que em relação ao ligante B. O pK_d , utilizado como mais uma ferramenta de avaliação da afinidade da ligação do composto à enzima, também foi estimado, apresentando um resultado ligeiramente menor do que o ligante B, entretanto maior do que o ligante C e o substrato EPSP, ou seja, apresentando uma maior afinidade de ligação e conseqüentemente, uma característica bastante satisfatória para um possível inibidor competitivo (Tabela 6).

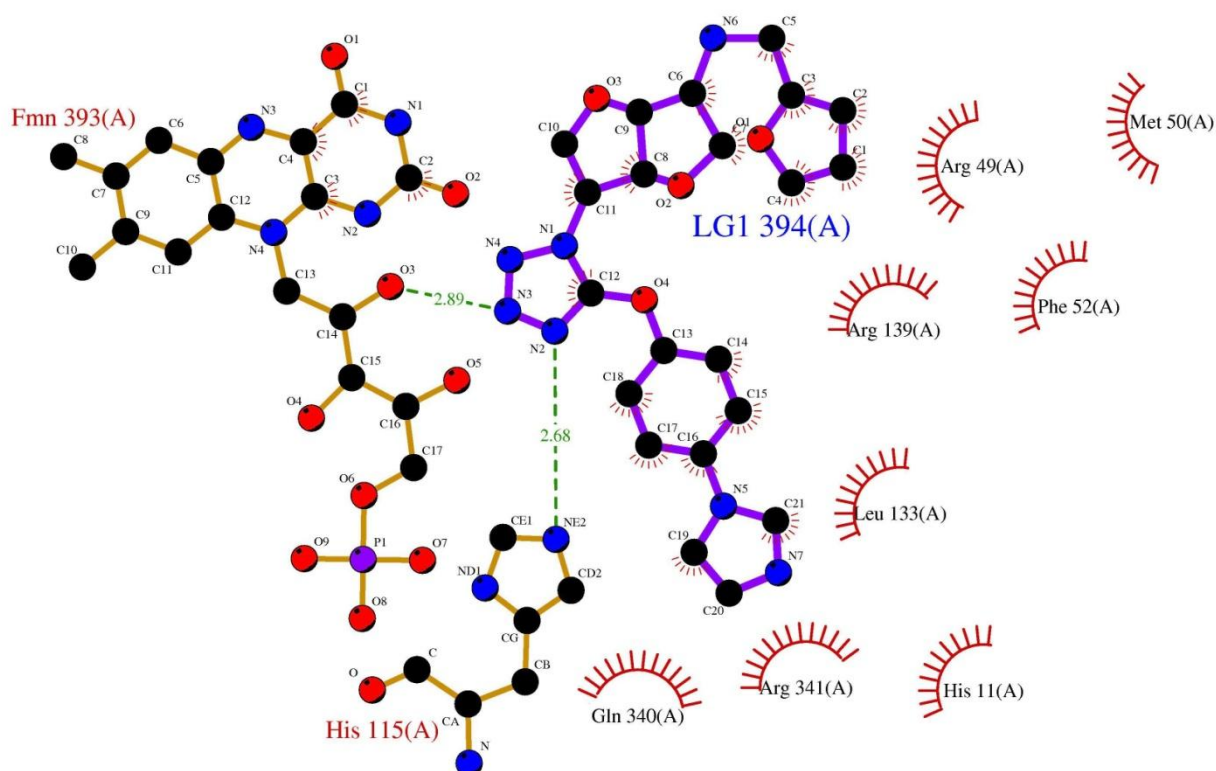


FIGURA 13. Ligplot - demonstrando as ligações de hidrogênio e os contatos hidrofóbicos feitos pelo ligante A com a enzima *MtCS*.

(Representação: átomos de oxigênio (esferas vermelhas); átomos de carbono (esferas pretas); átomos de nitrogênio (esferas azuis); ligações de hidrogênio (linhas tracejadas em verde); ligações entre os átomos do ligante (linhas em lilás); ligações entre os átomos do FMN (linhas em caramelo); contato hidrofóbicos (semi-círculos em vermelho)).

4.3.2 Ligante B

O ligante B, recebe a denominação pela IUPAC de *methyl 4-(2-benzylbenzoyl)-2,5-dimethyl-1H-pyrrole-3-carboxylate* e pela SIGMA de F131.

Este composto possui vários estudos relacionados a diversas propriedades farmacológicas, como agente para desordens pancreáticas, inibidor de HIV protease, inibidor da síntese de leucotrienos, antagonista/agonista de benzodiazepínicos, cardiotônico, entre outros (Keiser *et al.*, 2007). Mas além destes e mais importante para o trabalho, também apresentou relação estrutural química com compostos com atividade anti-tuberculose, com *E-value* de 3, 68.10^{-3} e coeficiente de Tanimoto (TC) de 0.33.

O ligante B, em comparação ao A, faz mais ligações de hidrogênio, cinco no total, em compensação, possui menos contatos hidrofóbicos, sendo que um deles com o FMN. Entretanto, todos os resíduos que fazem interação com o ligante B, também fazem com o A, como a ARG 49, PHE 52, LEU 133, GLN 340 e ARG 341 (Figura 14). No ligante A, a ARG 139 aparece fazendo apenas contato hidrofóbico, já no ligante B, este resíduo faz uma ligação de hidrogênio, assim como a HIS 115.

Quando comparado com o inibidor, o ligante B apresenta seis resíduos fazendo interações em comum, mais o cofator. Destes, cinco também fazem parte do sítio do EPSP.

Em relação as propriedades físico-químicas deste composto, duas particularidades se destacam e são complementares. A área de superfície polar, que é praticamente a metade dos demais analisados e o coeficiente de partição, que é positivo. Podemos relacionar estas duas características por serem diretamente inversas. Compostos que apresentam maior coeficiente de partição, são mais

hidrofóbicos e tem maior afinidade pela fase orgânica, e por isso, tendem a ultrapassar com maior facilidade as biomembranas hidrofóbicas, como bicamadas lipídicas das células, que vêm a ser uma característica satisfatória para um fármaco com atuação nesse nicho.

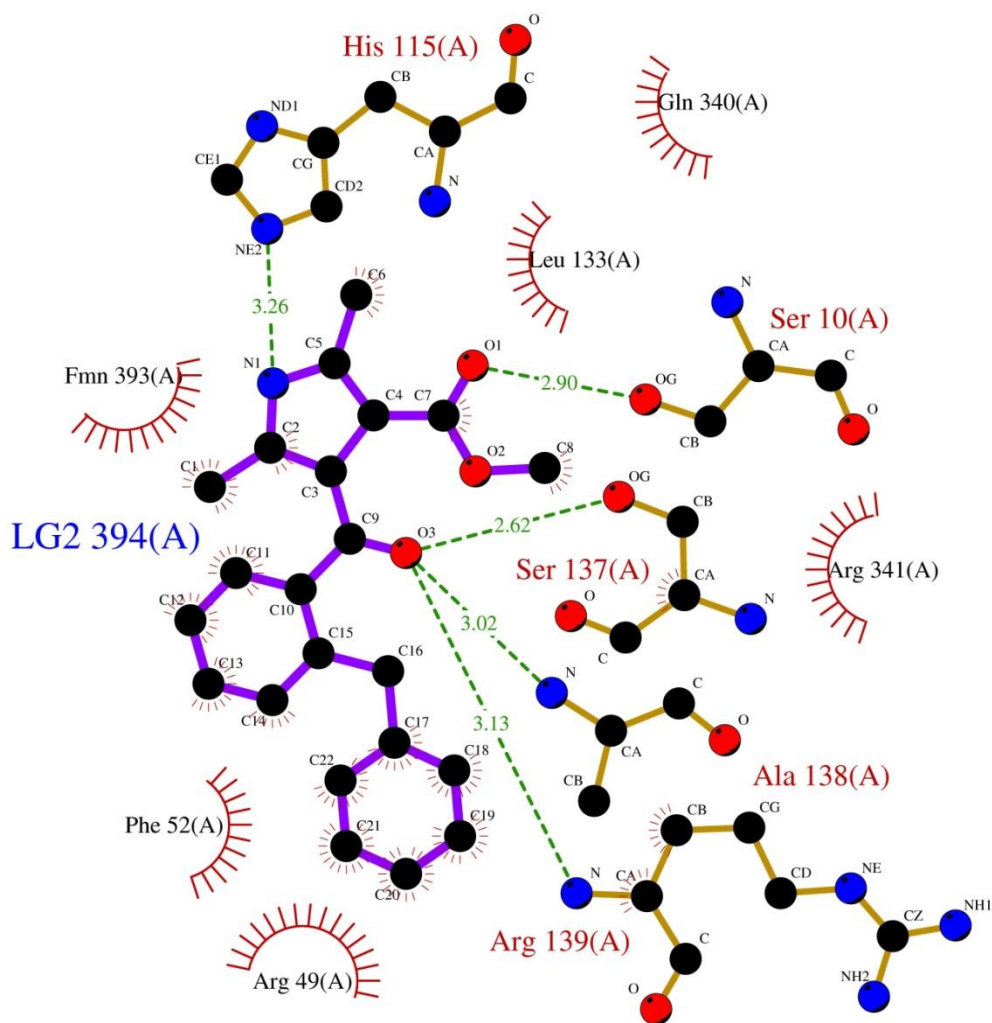


FIGURA 14. Ligplot - demonstrando as ligações de hidrogênio e os contatos hidrofóbicos feitos pelo ligante B com a enzima *MtCS*. (Representação: átomos de oxigênio (esferas vermelhas); átomos de carbono (esferas pretas); átomos de nitrogênio (esferas azuis); ligações de hidrogênio (linhas tracejadas em verde); ligações entre os átomos do ligante (linhas em lilás); ligações entre os átomos do FMN (linhas em caramelo); contato hidrofóbicos (semi-círculos em vermelho)).

4.3.3 Ligante C

Osborne e colaboradores, demonstraram experimentalmente, a inibição da enzima CS de *Escherichia coli* com um composto análogo do EPSP, o (6R)-6-fluoro-5-enoilpiruvilchiquimato 3-fosfato. Quando incubado com a enzima, o ligante C apresentou um $IC_{50}=0,5 \mu M$, porém quando a enzima não foi previamente incubada, a concentração deste composto necessária para inibir 50% da atividade da enzima, aumentou para $250 \mu M$ (Osborne *et al.*, 2000).

Baseado neste dados, bem como no alto grau de conservação da seqüência da enzima CS entre as espécies (Fernades *et al.*, 2007), este inibidor foi submetido a docagem e dinâmica molecular contra uma estrutura de *MtCS*, com o objetivo de correlacionar informações com os ligantes triados no VS, bem como avaliar se este composto também tem potencial inibitório em *Mt*.

O ligante C, por ser análogo do EPSP, é entre as moléculas docadas, o que apresenta maior similaridade em relação ao resíduos da enzima com que o substrato faz interação. Entre estes, a ARG 49 e a ARG 341, que nos ligantes aparecem apenas fazendo contato hidrofóbico, e a ARG 40, surgem fazendo ligações de hidrogênio com a molécula. Outra similaridade é os resíduos ALA 138 e ARG 139 que fazem ligação de hidrogênio tanto com EPSP, quanto com o ligante C e o ligante B. Porém, dois outros resíduos, que não faziam parte de nenhum outro sítio ativo aparecem fazendo interação com esta molécula, a THR 141 e a SER 137 (Figura 15).

As características físico-químicas do ligante C e do substrato, exatamente pela semelhança estrutural, são bastante parecidas. Porém, ao contrário do esperado, a energia de ligação do ligante C, comparada com a do EPSP, foi maior e o pK_d

menor. Este fato pode ser devido aos pesos de cada propriedade dos algoritmos dos programas, mas não descarta totalmente esta molécula como um possível inibidor competitivo.

A representação da estrutura química e docada de cada ligante pode ser observada na figura 16.

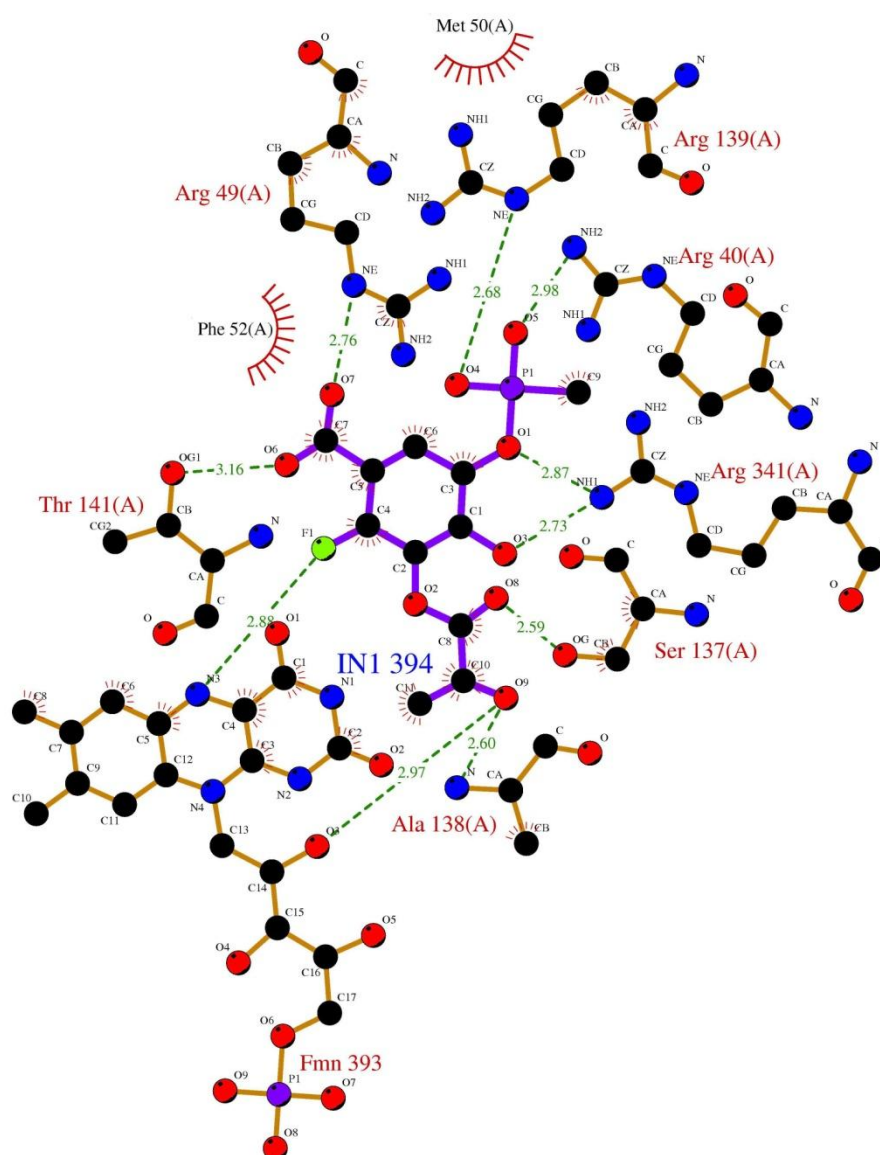


FIGURA 15. Ligplot - demonstrando as ligações de hidrogênio e os contatos hidrofóbicos feitos pelo ligante C com a enzima *MfCS*.

(Representação: átomos de oxigênio (esferas vermelhas); átomos de carbono (esferas pretas); átomos de nitrogênio (esferas azuis); ligações de hidrogênio (linhas tracejadas em verde); ligações entre os átomos do ligante (linhas em lilás); ligações entre os átomos do FMN (linhas em caramelo); contato hidrofóbicos (semi-círculos em vermelho)).

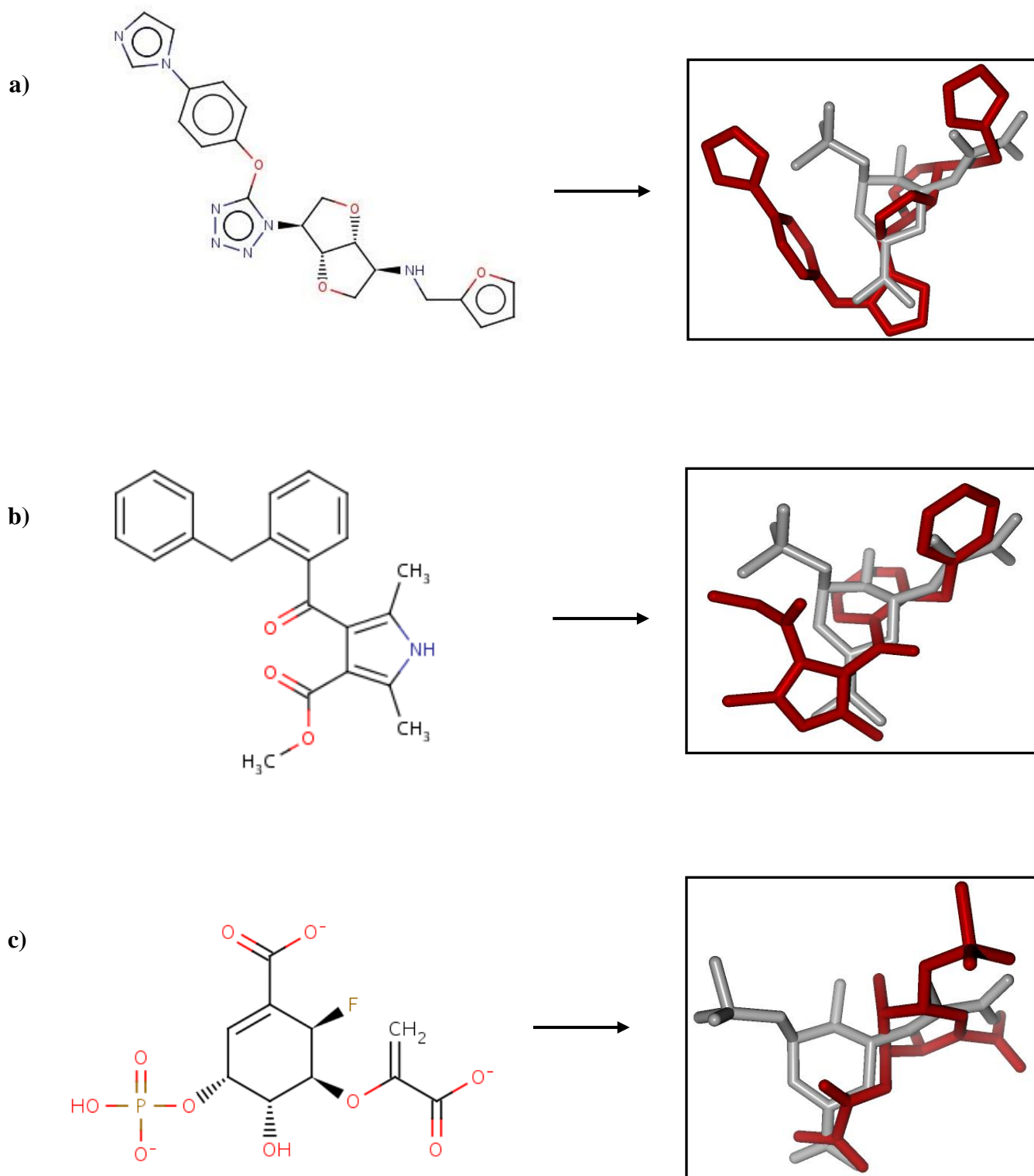


FIGURA 16. Representação da estrutura química (esquerda) e da mesma estrutura docada e sobreposta ao EPSP (direita – em cinza o EPSP e em vermelho o composto; geradas pelo programa *MolDock*). a) Ligante A; b) Ligante B; c) Ligante C.

TABELA 5. Tabela demonstrando com quais resíduos da enzima cada molécula docada faz interação, que tipo de interação e entre quais átomos. HB= *Hydrogen Bonds* (Ligações de hidrogênio); VW= interações de van der Waals (para localizar átomos, ver as figuras do Ligplot).

Ligante A			Ligante B			Ligante C		
Resíduo	Interação	Átomos	Resíduo	Interação	Átomos	Resíduo	Interação	Átomos
HIS 11	VW		SER 10	HB	OG-O1	ARG 40	HB	NH2-O5
ARG 49	VW		ARG 49	VW		ARG 49	HB	NE-O7
MET 50	VW		PHE 52	VW		MET 50	VW	
PHE 52	VW		HIS 115	HB	NE2_N1	PHE 52	VW	
HIS 115	HB	NE2-N2	LEU 133	VW		SER 137	HB	OG-O8
LEU 133	VW		SER 137	HB	OG-O3	ALA 138	HB	N-O9
ARG 139	VW		ALA 138	HB	N-O3	ARG 139	HB	NE-O4
GLN 340	VW		ARG 139	HB	N-O3	THR 141	HB	OG1-O6
ARG 341	VW		GLN 340	VW		ARG 341	HB	NH1-O3
FMN (393)	HB	O3-N3	ARG 341	VW		ARG 341	HB	NH1-O1
			FMN (393)	VW		FMN (393)	HB	O3-O9
						FMN (393)	HB	N3-F1

TABELA 6. Informações físico-químicas da estruturas analisadas. (PSA (*Polar Surface Area*) – Área de Superfície Polar; *LogP* – log de P (coeficiente de partição); AHB (*Acceptor H-Bonds*) – Número de aceitadores de ligação de hidrogênio; DHB (*Donor H-Bonds*) – Número de doadores de ligação de hidrogênio; *pK_d* – log de K_d (coeficiente de dissociação); *MolDock Score* - energia de ligação do programa *MolDock*).

Composto	Fórmula química	Peso molecular (g/mol)	PSA (Å ²)	LogP	AHB	DHB	pK _d	MolDock Score
Ligante A	C ₂₁ H ₂₂ N ₇ O ₄ ⁺	436,4	119	-0.88	8	1	7,29675	-214,744
Ligante B	C ₂₂ H ₂₁ NO ₃	347,4	59.2	4.6	3	1	7,6304	-180,995
Ligante C	C ₁₀ H ₁₂ FO ₁₀ P	342,1	171	-1.9	11	5	5,67831	-150,823
EPSP	C ₁₀ H ₁₃ O ₁₀ P	324,0	171	-1.8	10	5	6,04056	-160,066

4.4 Dinâmica molecular

Após a identificação e análise dos candidatos a inibidores de CS, o passo seguinte foi a verificação da estabilidade destes compostos no sítio ativo da enzima, através da técnica de DM. As simulações das estruturas complexadas com substrato e cofator (sistema 1), ligante A e cofator (sistema 2), ligante B e cofator (sistema 3), e ligante C e cofator (sistema 4), foram realizadas separadamente e apenas na forma monomérica (monômero D) da enzima. Isto se deve ao fato relevante de que os sítios ativos do EPSP encontram-se situados separadamente em cada monômero, sem a interação de outra subunidade, com mais de 27 Å de distância um do outro.

Além disso, de forma complementar ao estudo, para averiguar a estabilidade da enzima em relação a sua estrutura quaternária e o papel dos ligantes nesta estabilização, mais três dinâmicas, sem ligantes, foram simuladas.

Todas as simulações foram realizadas em condições periódicas de contorno usando o pacote GROMACS 4.0.5 (van der Spoel *et al.*, 2005) e o campo de força 96.1 Gromos (53A6) (Oostenbrink *et al.*, 2005). O arquivo de topologia molecular e os parâmetros de campo de força, foram gerados por PRODRG 2.5 *server* (beta) (Schuttelkopf & van Aalten, 2004) exceto para as cargas do cofator, substrato e ligantes, que foram submetidos a cálculos *single-point ab initio* em nível HF/6-311G (d,p) (B3LYP) para a obtenção das cargas através do programa Gaussian 03 (Frisch *et al.*, 2003).

Antes de simulações de DM, os sistemas foram solvatadas com o modelo de solvatação explícita SPC/E (Berendsen *et al.*, 1981), incorporados em um sistema de caixa cúbica, com a distância mínima entre a superfície da proteína e a interface da

caixa de 10 Å. Para neutralizar a densidade de carga negativa dos sistemas, íons de sódio foram adicionados. Os sistemas foram então submetidos a uma minimização de energia pelo algoritmo *steepest descent* (Arfken, 1985). Após, os sistemas foram gradualmente aquecidos de 50 para 298,15 K, em cinco passos (50-100, 100-150, 150-200, 200-250 e 250-298,15 K), sendo que a velocidade de cada passo de integração, re-escalada de acordo com a distribuição de Maxwell-Boltzmann, a fim de ajustar a energia cinética do sistema até a temperatura final. Este período é chamado de termalização.

Logo após, os sistemas foram submetidos a uma MD curta, com restrições de posição, por um período de 20 ps. Finalmente, uma dinâmica mais longa, com uma trajetória de 18 ns para as estruturas monoméricas e 10 ns para as estruturas diméricas e tetraméricas, foram realizadas, desta vez, sem nenhuma restrição. Todas as simulações foram realizados sob condições normais de temperatura (298,15 K) e pressão (1 bar), utilizando um tempo de acomodamento de temperatura constante de 0,1 ps e um tempo de acomodamento de pressão constante de 1,0 ps. Todos os comprimentos de ligação, incluindo átomos de hidrogênio, foram restringidos pelo algoritmo LINCS (Hess *et al.*, 1997). As interações eletrostáticas foram calculados usando algoritmo *Particle-mesh Ewald* (PME) (Darden *et al.*, 1993), com ordem de interpolação de 4 e um espaçamento de 0,12 nm. O raio de corte para as interações de van der Waals foi de 1,0 Å. As simulações foram realizadas com um intervalo de tempo (passos) de 2 fs, e as coordenadas foram salvas a cada 1 ps.

A flexibilidade e as mudanças conformacionais da proteína, observadas durante a simulação, foram descritas através do RMSD e as flutuações da estrutura média da proteína, expressas em termos de flutuações dos desvios quadráticos médios

(RMFS). O raio de giro, que demonstra as mudanças em nível de compactação da enzima também foi calculado.

4.4.1 Flexibilidade e estabilidade das estruturas apo

Para agregar informações ao trabalho, o estudo da estabilidade da enzima, em diferentes composições de estrutura quaternária, foi realizado através de três simulações de DM independentemente, com as formas monomérica (sistema 5 - monômero D), dimérica (sistema 6 - cadeias C e D) e tetramérica (sistema 7 - cadeias A, B, C e D) da enzima não-complexada.

Para a obtenção de uma estimativa da qualidade da convergência das trajetórias da DM, o RMSD das cadeias principais dos modelos das estruturas iniciais foi calculado, como demonstrado no gráfico 1.

Como é possível observar, a estrutura tetramérica (Gráfico 1- linha verde) é a que apresenta maior estabilidade (variação de $2.2 \pm 0.3 \text{ \AA}$), seguido pelo dímero (Gráfico 1- linha vermelha), que teve um pequeno aumento do desvio no começo da simulação, estabilizando após os 5000 ps ou 5 ns (variação de $3.3 \pm 0.8 \text{ \AA}$). Esta diferença é parcialmente causada pelas limitações do campo de força, mas também porque os átomos na simulação estão se movendo e vibrando em torno de uma estrutura de equilíbrio.

Já a forma monomérica (Gráfico 1- linha preta) demonstrou maior variação ($3.9 \pm 0.8 \text{ \AA}$), sendo que ao final da simulação de 10 ns, a estrutura ainda não havia estabilizado totalmente. Este importante dado sugere que a enzima *MtCS* somente encontra-se estável na forma de dímero ou trímero e/ou complexada, corroborando com os experimentos de Dias e colaboradores (Dias *et al.*, 2006) com coluna de gel

filtração para determinação da estrutura quaternária da *MtCS* em solução. Neste estudo, eles observaram que em solução, a enzima CS tende estar em equilíbrio nas formas diméricas e tetraméricas, mas que o aumento da concentração da enzima e/ou a ligação de alguma molécula levam a mudanças conformacionais dos dímeros, resultando numa tetramerização da enzima.

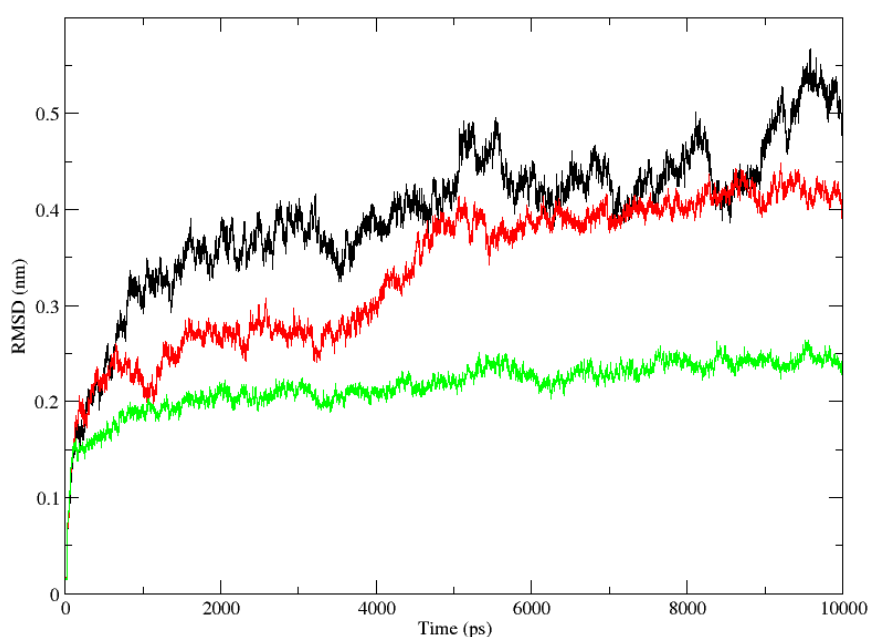


GRÁFICO 1. Representação gráfica do RMSD (nm) de todos os C α das estruturas do modelo de partida em função do tempo (ps) (Monomérica – linha preta; Dimérica – linha vermelha; Tetramérica – linha verde).

Outra propriedade importante, que informa sobre o estado de empacotamento da estrutura ao redor do centro de massa da proteína, é o raio de giro. No gráfico 2, podemos observar que o tamanho do raio da estrutura dimérica (linha vermelha) e tetramérica (linha verde) são maiores, conseqüentemente por estas estruturas terem o dobro e o quádruplo de resíduos, respectivamente. Como esperado, o sistema monomérico (linha preta) foi o que apresentou maior variação, mas mantendo o empacotamento da sua estrutura inicial.

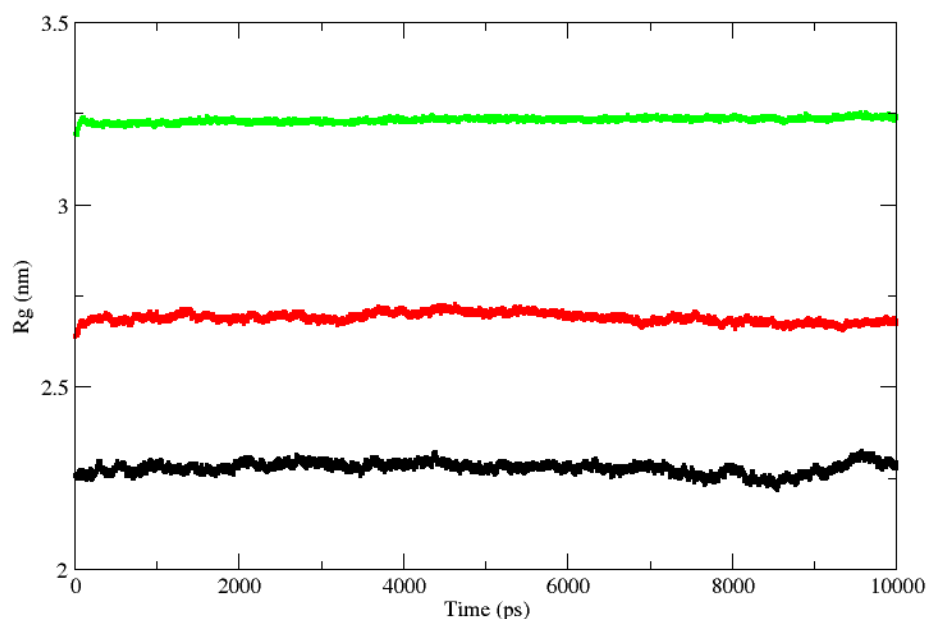


GRÁFICO 2. Representação gráfica do raio de giro (nm) das estruturas do modelo de partida em função do tempo (Monomérica – linha preta; Dimérica – linha vermelha; Tetramérica – linha verde).

Já a flexibilidade das estruturas foi descrita através do RMSF, que reflete a movimentação dos resíduos da enzima, como observado no gráfico 3 e visualizado na figura 17. As maiores flutuações foram observadas nas regiões dos *loops* e nas regiões C-terminais, sendo os valores mais baixos do RMSF correspondem às regiões centrais da estrutura.

No sistema correspondente ao monômero (Gráfico 3-a e figura 17-a), podem ser definidas quatro regiões com maior variação, localizada entre os resíduos 84-104 (R1), 110-132 (R2), 261-296 (R3) e 318-340 (R4). As áreas correspondentes as regiões 2, 3 e 4 estão localizada próximo a interface dos monômeros D e C, o que justificaria a maior variação ou movimentação destas áreas no sistema monomérico.

No gráfico 3-b (Figura 17-b), correspondente ao dímero, é possível também, distinguir 4 regiões com maior flutuação, entre os resíduos 77-106 (R1), 171-191 (R2), 277-294 (R3) e 313-341 (R4). Entretanto, esta variação na posição dos

resíduos ao longo da trajetória de simulação é menor do que no sistema monomérico.

Já na estrutura tetramérica, a flutuação dos resíduos parece ser mais homogênea e cinco regiões com maior oscilação podem ser observadas, entre os resíduos 41-55 (R1), 80-108 (R2), 170-193 (R3), 280-291 (R4) e 318-349 (R5) (Gráfico 3-c e figura 17-c).

Nos três sistemas, é possível observar regiões comuns com maior flutuação. Os resíduos localizado entre fita- β 7 e a hélice- α 3, que correspondem a região 1 nas estruturas monoméricas e diméricas, e 2 na estrutura tetramérica, estão situados em uma parte mais periférica da proteína e sem contato com outros monômeros, sendo esta uma possível explicação para maior flutuação. Outras três regiões em comuns também podem ser observadas: R2 (dímero) com R3 (tetrâmero); R3 e R4 (monômero e dímero) com R4 e R5 (tetrâmero), respectivamente.

Os dímeros da *MtCS* são caracterizados por uma grande folha compostas por 8 fitas- β antiparalelas (uma folha de cada monômero). Esta grande folha é a principal característica da interface dos dímeros, sendo que as duas fitas- β 10 formam a região mais importante de estabilização do dímero, juntamente com a hélice- α 8 que interage com a outra hélice- α 8 do monômero adjacente. Na simulação de dinâmica do monômero, podemos observar a movimentação mais intensa desta região, mais especificamente do resíduo 261 ao 264 (R3), que fazem parte da fitas- β 10, mas que nos sistemas dimérico e tetramérico foi vista com menor intensidade.

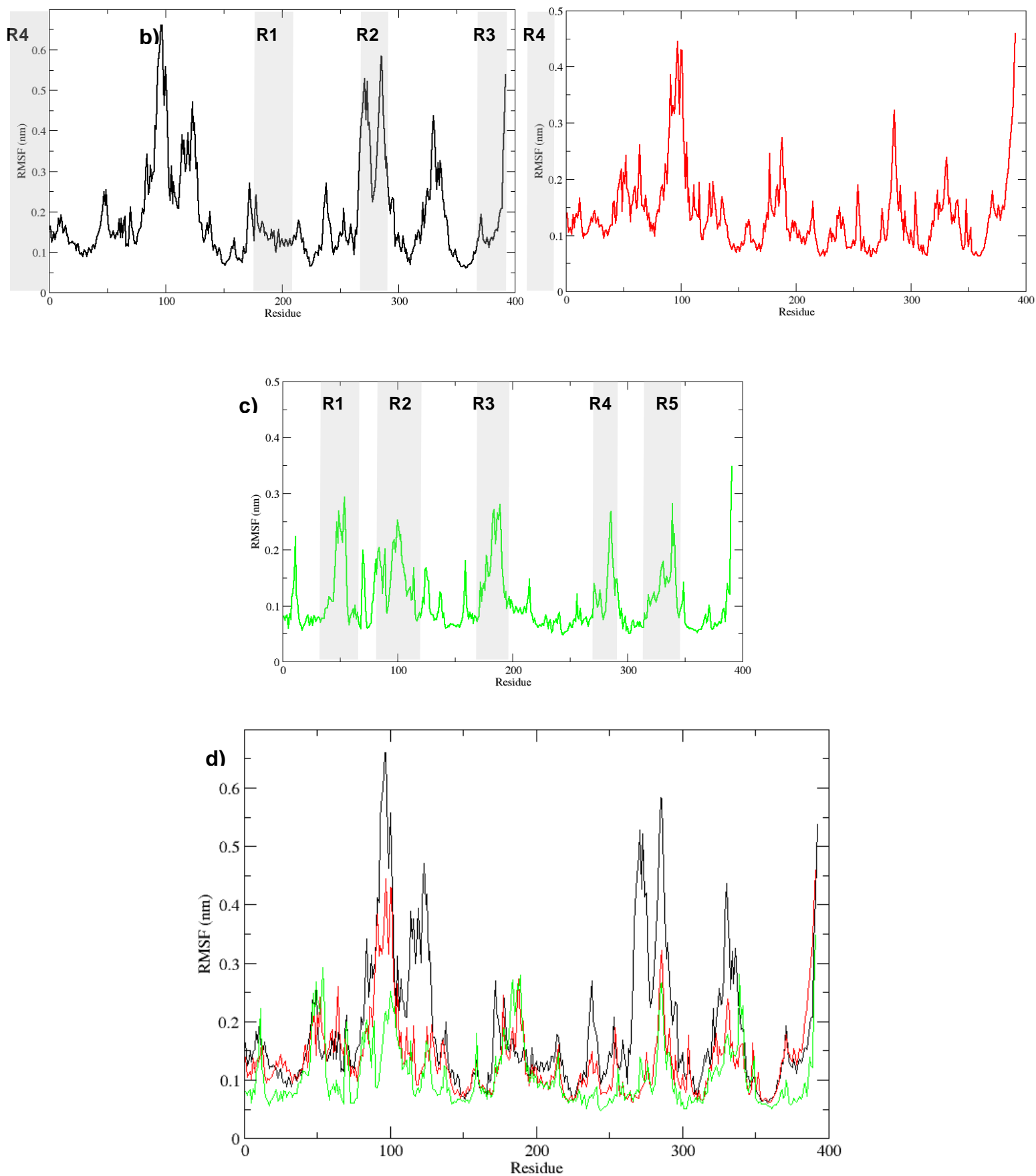


GRÁFICO 3. Representação gráfica do RMSF (nm) das estruturas do modelo de partida em relação aos resíduos. a) Monomérica – linha preta; b) Dimérica – linha vermelha; c) Tetramérica – linha verde; d) todos sistemas juntos. Em cinza, as regiões dos resíduos destacadas no texto e na figura 17.

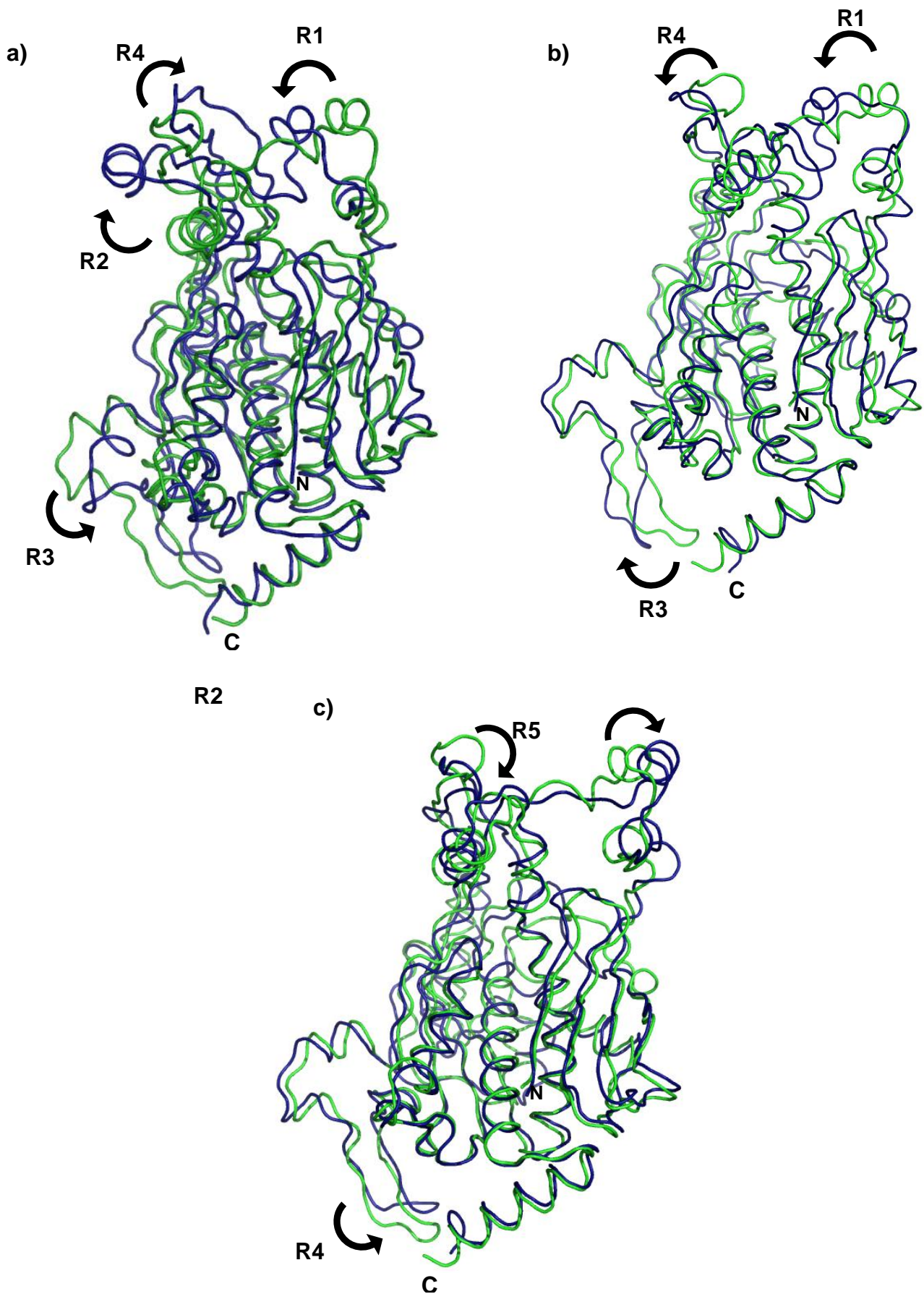


FIGURA 17. Representação em *ribbon* das estruturas iniciais (verde) e finais (azul) sobrepostas da DM (em 10 ns). Setas indicando a região e o movimento exercido durante a simulação. a) Monomérica; b) Dimérica; c) Tetramérica N: região N-terminal da enzima; C: região C-terminal da enzima (algumas regiões descritas no texto e nos gráficos não estão representadas nas figuras).

4.4.2 Estabilidade dos ligantes no sítio ativo e principais resíduos responsáveis por esta estabilização

Para analisar a estabilidade dos ligantes quando complexados com a *MtCS*, os RMSDs dos carbonos- α das estruturas, durante as trajetórias de 18 ns de simulação, foram mensurados (Gráfico 4).

Foi possível observar um aumento do desvio em todos os sistemas na primeira parte da dinâmica e a estabilização em torno dos 10 ns, com exceção da estrutura apo (azul). Isso leva-nos a crer que, a ligação de uma pequena molécula à enzima possui um papel fundamental em sua estabilização, como mencionado anteriormente e que os compostos aqui propostos como possíveis inibidores, permanecem estáveis no sítio ativo da enzima.

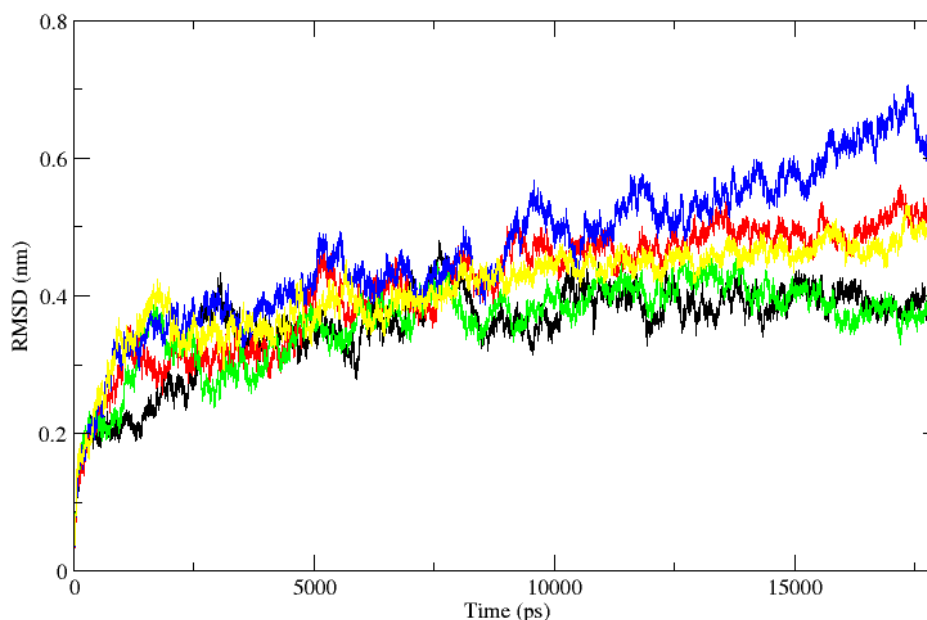


GRÁFICO 4. Representação gráfica do RMSD (nm) de todos os C α das estruturas do modelo de partida em função do tempo (ps) (Apo – linha azul; *MtCS*-FMN-EPSP – linha preta; *MtCS*-FMN-ligante A – linha amarela; *MtCS*-FMN-ligante B – linha vermelha; *MtCS*-FMN-ligante C – linha verde).

Os valores médios de RMSD dos sistemas *MtCS*-FMN complexados com EPSP e com ligante C aparecem iguais, $3.6 \pm 0.6 \text{ \AA}$. Este resultado provavelmente está relacionado a semelhança química e estrutural dos dois compostos e também sugere um comportamento similar da proteína. O sistema apo foi o que apresentou a maior média de RMSD, $4.7 \pm 1.1 \text{ \AA}$ (em 18ns), seguido pelo ligante A ($4.3 \pm 0.4 \text{ \AA}$) e B ($4.1 \pm 0.9 \text{ \AA}$), respectivamente.

Complementando o aspecto estabilidade, no gráfico 5, que mostra a variação do raio de giro durante as simulações, podemos observar, a partir do 15° ns, a convergência de todos os sistemas, tendo a diferença entre as médias de apenas $\approx 0.4 \text{ \AA}$. Isto representa que, mesmo com aumento de temperatura o enovelamento da enzima, tanto com diferentes ligantes, como apo, permanece estável.

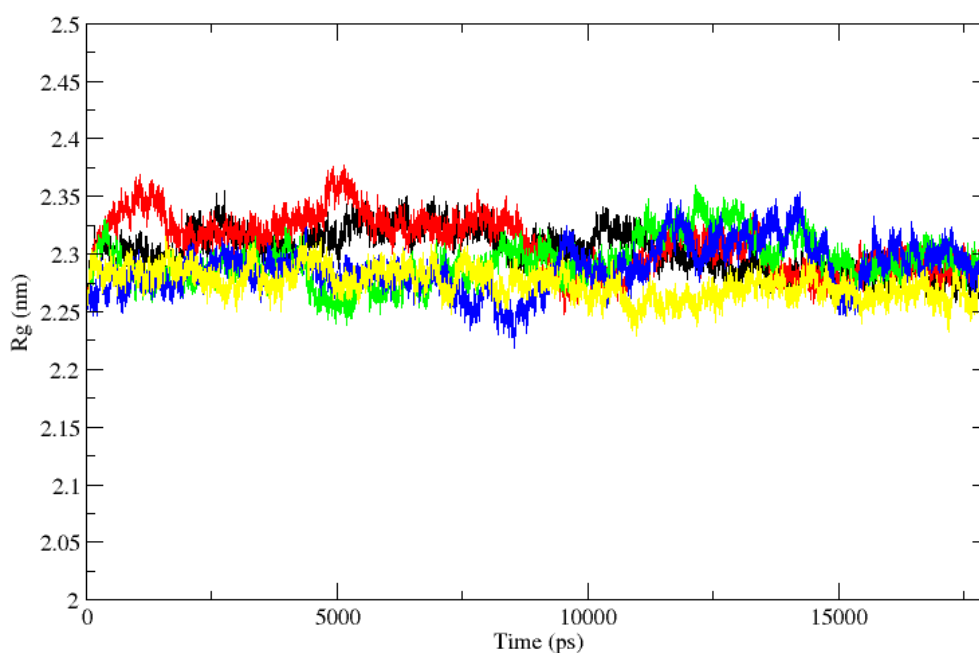


GRÁFICO 5. Representação gráfica do raio de giro (nm) das estruturas do modelo de partida em função do tempo (Apo – linha azul; *MtCS*-FMN-EPSP – linha preta; *MtCS*-FMN-ligante A – linha amarela; *MtCS*-FMN-ligante B – linha vermelha; *MtCS*-FMN-ligante C – linha verde).

Para analisar quais e com que frequência os resíduos do sítio ativo fazem interação com os respectivos ligantes, foram analisados os *snapshots*, extraídos das

trajetórias das simulações de todos os sistemas complexados a cada nanosegundo, do primeiro ao décimo oitavo, utilizando o programa Ligplot v.4.0 (Wallace *et al.*, 1995). Os resultados, mostrado no gráfico 6, evidenciam os principais resíduos que interagiram, em algum momento da trajetória com os ligantes A, B e C (somados - em rosa) em comparação com o substrato natural da enzima, o EPSP (lilás).

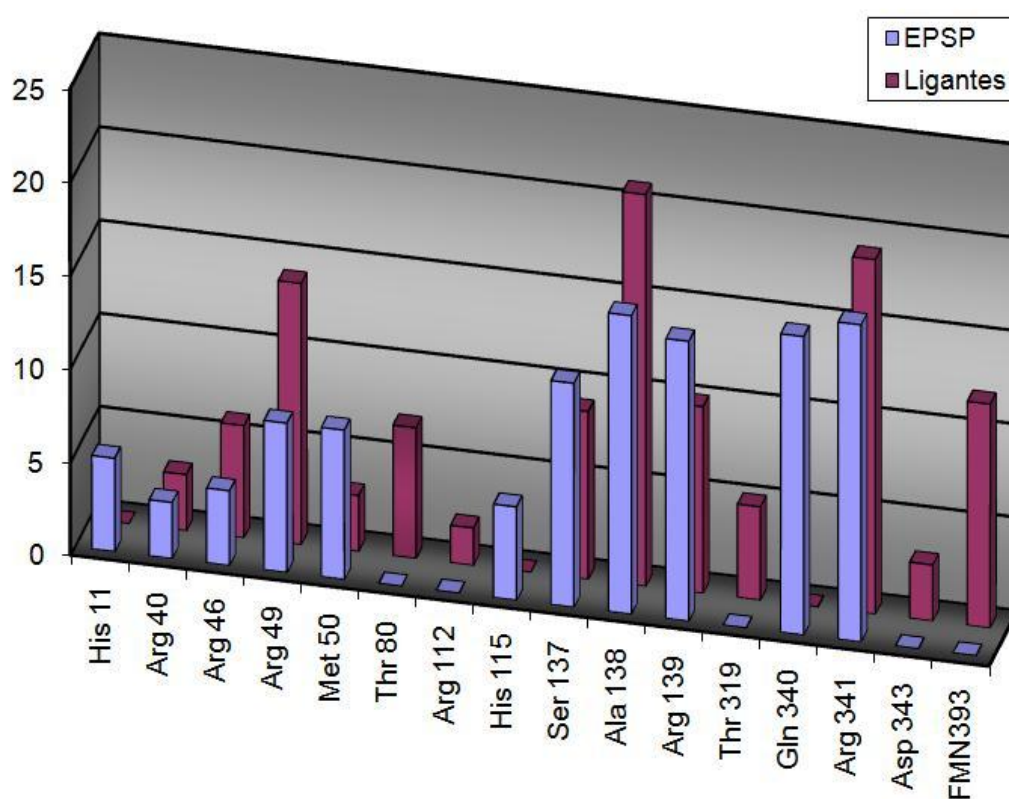


GRÁFICO 6. Gráfico mostrando os principais resíduos da proteína que fizeram ligação de hidrogênio em função do número de vezes em que estes apareceram nos *snapshots* extraídos a cada 1 ns (do 1° ao 18°ns) das trajetórias das simulações de DM (Soma dos ligantes A, B e C - em rosa); EPSP – lilás).

Em relação ao RMSF das estruturas, podemos observar, diferentemente dos sistemas apo demonstrados no gráfico 3, uma maior similaridade entre as regiões e o tamanho das flutuações em todos os complexos analisados, indicando que, quando ligada a uma molécula, a tendência de movimentação dos resíduos do sítio

ativo parece ser bastante parecida. As regiões 1 e 2, correspondentes aos resíduos 75-138 e 261-301, podem ser observadas em todos os complexos (Gráficos 7 e 8) como as maiores zonas de flutuação de resíduos. No pico destas regiões estão os resíduos 90, 96, 94, 100, 123 (R1) e 265, 269, 270, 274, 283 e 286 (R2).

Entretanto, o sistema complexado com o ligante A (Gráfico 7 e 8-b) difere-se dos demais. Uma terceira região de maior flutuação aparece entre os resíduos 331-346, tendo como pico o resíduo 330.

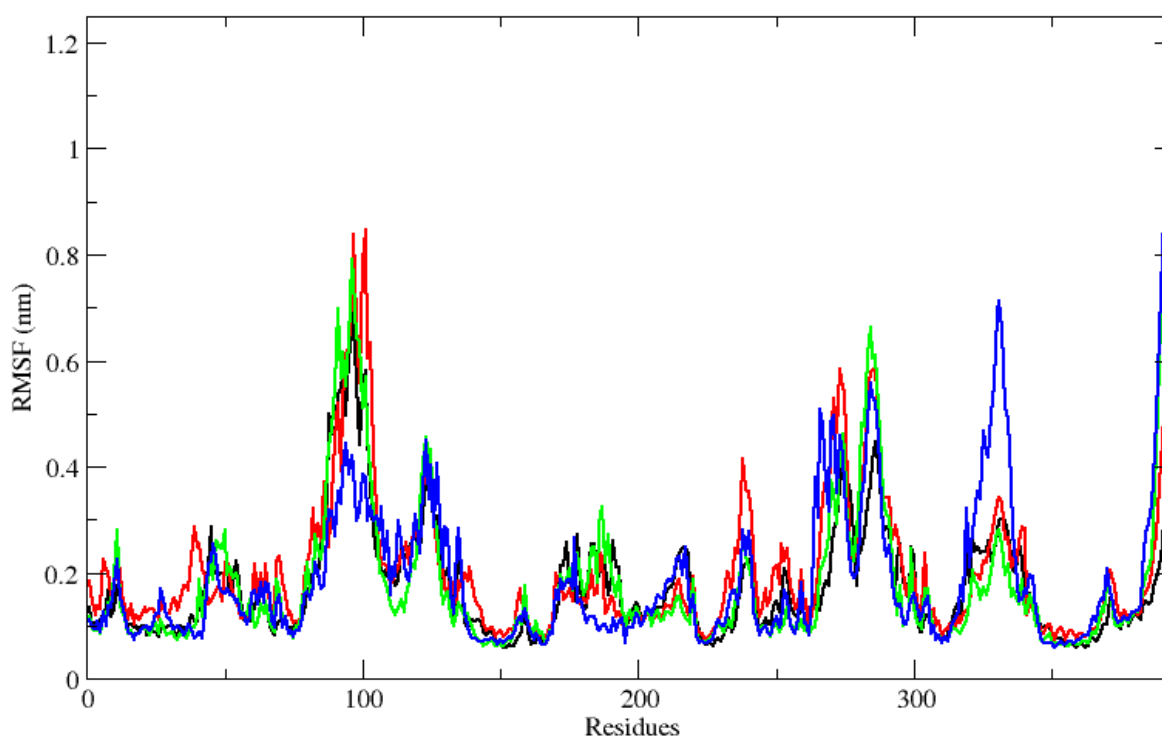


GRÁFICO 7. Representação gráfica do RMSF (nm) das estruturas do modelo de partida em relação aos resíduos. (*MiCS-FMN-EPSP* – linha preta; *MiCS-FMN-ligante A* – linha azul; *MiCS-FMN-ligante B* – linha vermelha; *MiCS-FMN-ligante C* – linha verde).

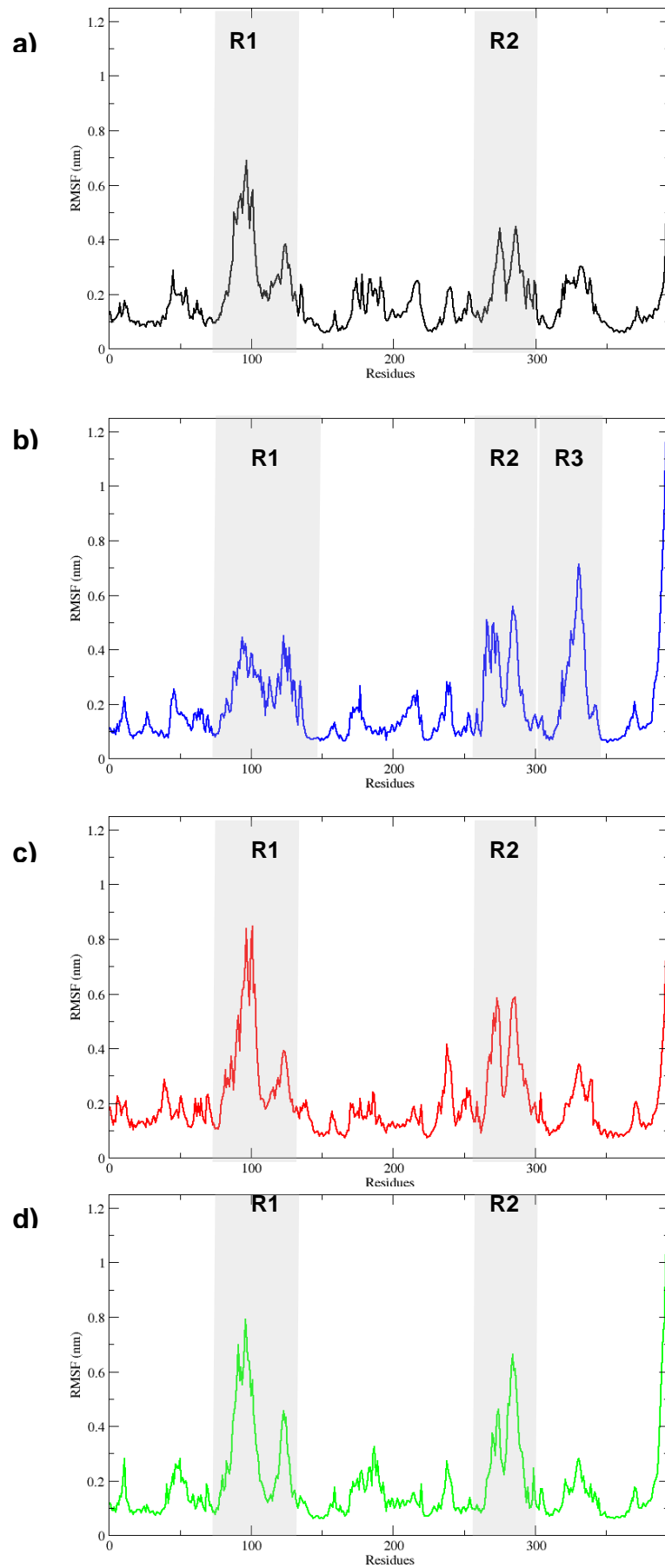


GRÁFICO 8. Representação gráfica do RMSF (nm) das estruturas do modelo de partida em relação aos resíduos. **a)** *MtCS-FMN-EPSP* – linha preta; **b)** *MtCS-FMN-ligante A* – linha azul; **c)** *MtCS-FMN-ligante B* – linha vermelha; **d)** *MtCS-FMN-ligante C* – linha verde).

Podemos concluir, a partir destes dados que os resíduos Arg 49, Ser 137, Ala 138, Arg 139 e Arg 341, possuem um papel chave na estabilização da ligação de um molécula ao sítio catalítico do substrato EPSP na *MtCS*. Além disso, os resultados das simulações de DM indicaram estabilidade dos ligantes A, B e C no interior do sítio ativo da enzima, sugerindo que as moléculas propostas são inibidores em potencial para *MtCS* e promissores para a quimioterapia anti-micobacteriana.

6. CONCLUSÕES E PERSPECTIVAS:

Este trabalho apresenta, pela primeira vez, modelagem e estudos de docagem e dinâmica molecular baseados em uma estrutura cristalográfica de *MtCS*, o que confere aos nossos resultados, uma maior confiabilidade. Trabalhos anteriores, que foram utilizados como base, apresentaram modelagem e análises de *MtCS* apenas a partir da estrutura de *SpCS* (Fernades *et al.*, 2007). Estudos de dinâmica molecular, que fornecem informações mais realistas do sistema por ser uma simulação em solução, também são inéditos para esta enzima, elucidando, complementando e enriquecendo dados sobre a estrutura de *MtCS* e sua interação com o cofator e substrato, para o estudo dos candidatos a inibidores aqui analisados, bem como para trabalhos futuros acerca deste tema.

A análise dos resultados da modelagem, da docagem geométrica e da minimização de energia da estrutura construída, demonstraram uma boa qualidade do modelo proposto e a sua validação para sua utilização nas simulações de docagem molecular.

Os resultados apresentados, relativos aos possíveis inibidores propostos, denominados de ligante A, B e C, mostram-se otimistas em relação as características observadas de ligação, complementariedade geométrica e estabilidade no sítio catalítico da enzima, bem como suas propriedades químicas e farmacológicas favoráveis, comuns a drogas já conhecidas e testadas existentes.

Além disso, os resultados obtidos a partir das simulações da dinâmica molecular, mostraram-se bastante interessantes em relação a estabilidade da enzima apo em diferentes estados oligoméricos, sugerindo que os principais fatores

responsáveis por esta estabilidade estão diretamente relacionada com a estrutura quaternária e com a ligação de alguma molécula.

Estas informações podem ser utilizadas tanto para determinar prioridades estruturais no desenho racional de drogas, como para refinar e otimizar abordagens e protocolos para triagem computacional.

Entretanto, apesar da bioinformática ser uma poderosa ferramenta no desenvolvimento de drogas, exatamente por fazer a integração de informações disponíveis de biologia celular, molecular e farmacológica com a tecnologia de *softwares* de análise de sistemas biomoleculares, teste *in vitro* são necessários, bem como as demais etapas da predição de novos fármacos.

7. REFERÊNCIAS:

1. Amer FA, El-Behedy EM, Mohtady HA. New targets for antibacterial agents. *Biotechnology and Molecular Biology Reviews*. 2008;3(3):046-057.
2. Parish T, Stoker NG. The common aromatic amino acid biosynthesis pathway is essential in *Mycobacterium tuberculosis*. *Microbiology*. 2002;148:3069-3077.
3. Ducati RG, Ruffino-Netto A, Basso LA, Santos DS. The resumption of consumption: a review on tuberculosis. *Mem Inst Oswaldo Cruz*. 2006;101(7):697-714.
4. Basso LA, Da Silva LHP, Fett-Neto AG, De Azevedo WFJr, Moreira IS, Palma MS, Calixto JB, Astolfi Filho S, Dos Santos RR, Soares MBP, Santos DS. The use of biodiversity as source of new chemical entities against defined molecular targets for treatment of malaria, tuberculosis, and T-cell mediated diseases - a review. *Mem Inst Oswaldo Cruz*. 2005; 100(6):475-506.
5. Global tuberculosis control: surveillance, planning, financing. WHO Report 2009. Geneva, World Health Organization (WHO/HTM/TB/2009.411).
6. Da Silveira NJF, Uchoa HB, Pereira JH, Canduri F, Basso LA, Palma MS, Santos DS, De Azevedo WFJr. Molecular models of protein targets from *Mycobacterium tuberculosis*. *J Mol Model*. 2005;11(2):160-166.
7. Raman K, Yeturu K, Chandra N. TargetTB: a target identification pipeline for *Mycobacterium tuberculosis* through an interactome, reactome and genome-scale structural analysis. *BMC Systems Biology*. 2008; 2:109.
8. Dias MVB, Ely F, Palma, MS, De Azevedo WFJr, Basso LA, Santos DS. Chorismate synthase: an attractive target for drug development against orphan diseases. *Curr Drug Targets*. 2007a; 8(3): 437-444.
9. Alibhai MF, Stallings WC. Closing down on glyphosate inhibition - with a new structure for drug discovery. *Proc Natl Acad Sci USA*. 2001;98:2944-2946.
10. Schonbrunn E, Eschenburg S, Shuttleworth WA, Schloss JV, Amrhein N, Evans JN, *et al*. Interaction of the herbicide glyphosate with its target enzyme 5-enolpyruvylshikimate 3-phosphate synthase in atomic detail. *Proc Natl Acad Sci USA*. 2001;98:1376-1380.
11. Park H, Hilsenbeck JL, Kim HJ, Shuttleworth WA, *et al*. Structural studies of *Streptococcus pneumoniae* EPSP synthase in unliganded state, tetrahedral intermediate-bound state and S3P-GLP-bound state. *Mol Microbiol*. 2004;51:963-971.

- 12.** Dias MVB, Borges JC, Ely F, Pereira JH, Canduri F, Ramos CH, Frazzon J, Palma MS, Basso LA, Santos DS, De Azevedo WFJr. Structure of chorismate synthase from *Mycobacterium tuberculosis*. *J Struct Biol.* 2006;154(2):130-143.
- 13.** Roberts F, Roberts CW, Johnson JJ, Kyle DE, Krell T, Coggins JR, Coombs GH, Milhous WK, Tzipori S, Ferguson DJ, *et al.* Evidence for the shikimate pathway in apicomplexan parasites. *Nature.* 1998;393:801-805.
- 14.** Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, *et al.* Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature.* 1998;393:537-544.
- 15.** Oliveira JS, Pinto CA, Basso LA Santos DS. Cloning and overexpression in soluble form of functional shikimate kinase and 5-enolpyruvylshikimate synthase enzymes from *Mycobacterium tuberculosis*. *Protein Expr. Purif.* 2001;22:430-435.
- 16.** Oliveira JS, Mendes MA, Palma MS, Basso LA, Santos DS. One-step purification of 5-enolpyruvylshikimate-3-phosphate synthase enzyme from *Mycobacterium tuberculosis*. *Protein Expr Purif.* 2003;28:287-292.
- 17.** Rizzi C, Frazzon F, Ely F, Weber PG, Fonseca IO, Gallas M, Oliveira JS, Mendes MA, Souza BM, Palma MS, Santos DS, Basso LA. DAHP synthase from *Mycobacterium tuberculosis* H37Rv: cloning, expression, and purification of functional enzyme. *Protein Expr Purif.* 2005;40(1):23-30.
- 18.** De Azevedo WFJr, Canduri F, De Oliveira JS, Basso LA, Palma MS, Pereira JH, Santos DS. Molecular model of shikimate kinase from *Mycobacterium tuberculosis*. *Biochem Biophys Res Commun.* 2002;295:142-148.
- 19.** Pereira JH, Canduri F, De Oliveira JS, Da Silveira NJ, Basso LA, Palma MS, De Azevedo WFJr, Santos DS. Structural bioinformatics study of EPSP synthase from *Mycobacterium tuberculosis*. *Biochem Biophys Res Commun.* 2003;312:608-614.
- 20.** Pereira JH, Oliveira JS, Canduri F, Dias MVB, Palma MS, Basso LA, Santos DS, De Azevedo WFJr. Structure of shikimate kinase from *Mycobacterium tuberculosis* reveals the binding of shikimic acid. *Acta Crystallogr Sect D Biol Crystallogr.* 2004; 60(Pt 12 Pt 2): 2310-2319.
- 21.** Dias MVB, Ely F, Canduri F, Pereira JH, Frazzon J, Basso LA, Palma MS, De Azevedo WFJr, Santos DS. Crystallization and preliminary X-ray diffraction analysis of chorismate synthase *Mycobacterium tuberculosis*. *Acta Crystallogr Sect D Biol Crystallogr.* 2004; 60(11): 2003-2005.
- 22.** Dias MVB, Faim LM, Vasconcelos IB, Oliveira JS, Basso LA, Santos DS, De Azevedo WFJr. Effects of magnesium and chloride ions and shikimate on the structure of shikimate kinase from *Mycobacterium tuberculosis*. *Acta Crystallograph Sect F Struct Biol Cryst Commun.* 2007b;63(1):1-6.
- 23.** Oliveira JS, Pereira JH, Canduri F, Rodrigues NC, de Souza ON, De Azevedo WFJr, Basso LA, Santos DS. Crystallographic and pre-steady-state kinetics studies

on binding of NADH to wild-type and isoniazid-resistant enoyl-ACP(CoA) reductase enzymes from *Mycobacterium tuberculosis*. J Mol Biol. 2006; 359(3):646-666.

24. De Azevedo WF Jr, Caceres RA, Pauli I, Timmers LFSM, Barcellos GB, Rocha KB, Soares MBP. Protein-drug interaction studies for development of drugs against *Plasmodium falciparum*. Curr Drug Targets. 2009;10(3):271-278.

25. Fernandes CL, Breda A, Santos DS, Basso LA, Souza ON. A structural model for chorismate synthase from *Mycobacterium tuberculosis* in complex with coenzyme and substrate. Computers in Biology and Medicine. 2007;37(2):149-158.

26. Kitzing K, Auweter S, Amrhein N, Macheroux P. Mechanism of chorismate synthase. J Biol Chem. 2004;279:9451-9461.

27. Bornemann S, Theoclitou ME, Brune M, Webb MR, Thorneley RNF, Abell C. A secondary beta deuterium kinetic isotope effect in the chorismate synthase reaction. Bioorg Chem. 2000;28:191-204.

28. Macheroux P, Petersen J, Bornemann S, Lowe DJ, Thorneley RNF. Binding of the oxidized, reduced, and radical flavin species to chorismate synthase. An investigation by spectrophotometry, fluorimetry, and electron paramagnetic resonance and electron nuclear double resonance spectroscopy. Biochemistry. 1996; 35(5):1643-1652.

29. Bornemann S, Lowe DJ, Thorneley RNF. The transient kinetics of *Escherichia coli* chorismate synthase: substrate consumption, product formation, phosphate dissociation, and characterization of a flavin intermediate. Biochemistry. 1996;35(30): 9907-9916.

30. Osborne A, Thorneley RNF, Abell C, Bornemann S. Studies with substrate and cofactor analogues provide evidence for a radical mechanism in the chorismate synthase reaction. J Biol Chem. 2000;275:35825-35830.

31. Rauch G, Ehammer H, Bornemann S, Macheroux, P. Mutagenic analysis of an invariant aspartate residue in chorismate synthase supports its role as an active site base. Biochemistry. 2007;46(12):3768-3774.

32. Ely F, Nunes JES, Schroeder EK, Frazzon J, Palma MS, Santos DS, Basso LA. The *Mycobacterium tuberculosis* Rv2540c DNA sequence encodes a bifunctional chorismate synthase. BMC Biochemistry. 2008;9:13.

33. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook JD, Zardecki C. The protein data bank. Acta Crystallogr D Biol Crystallogr. 2002;58:899-907.

34. Bruning M, Bourenkov GP, Strizhov NI, Bartunik HD. *Mycobacterium tuberculosis* chorismate synthase in complex with NCA. 2007 (PDB: 2QHF) - To be published.

35. Bruning M, Bourenkov GP, Strizhov NI, Bartunik HD. Complexes of chorismate

synthase from *Mycobacterium tuberculosis*. 2006 (PDB: 2O11) - To be published.

36. Bruning M, Bourenkov GP, Strizhov NI, Bartunik HD. Complexes of chorismate synthase from *Mycobacterium tuberculosis*. 2006 (PDB: 2O12) - To be published.

37. Dias MVB, Dos Santos BB, Ely F, Basso LA, Santos DS, De Azevedo WFJr. Crystal structure of chorismate synthase from *Mycobacterium tuberculosis* at 2.22 angstrom of resolution. 2006 (PDB: 2G85) - To be published.

38. Ahn HJ, Yoon HJ, Lee B, Suh SW. Crystal structure of chorismate synthase: a novel FMN-binding protein fold and functional insights. *J Mol Biol.* 2004;336:903-915 (PDB: 1UM0 e 1UMF).

39. Forouhar F, Lee I, Vorobiev SM, Xiao R, Acton TB, Montelione GT, Tong L, Hunt JF. Crystal structure of the chorismate synthase from *Campylobacter jejuni*. Northeast Structural Genomics Target BR19. 2004 (PDB: 1SQ1) - To be published.

40. Maclean J, Ali S. The structure of chorismate synthase reveals a novel flavin binding site fundamental to a unique chemical reaction. *Structure.* 2003;11:1499-1511(PDB: 1QXO).

41. Quevillon-Cheruel S, Leulliot N, Meyer P, Graille M, Bremang M, Blondeau K, Sorel I, Poupon A, Janin J, Van Tilbeurgh H. Crystal structure of the bifunctional chorismate synthase from *Saccharomyces cerevisiae*. *J Biol Chem.* 2004;279:619-625 (PDB: 1R52 e 1R53).

42. Viola CM, Saridakis V, Christendat D. Crystal structure of chorismate synthase from *Aquifex aeolicus* reveals a novel beta alpha beta sandwich topology. *Proteins: Struct Funct Genet.* 2004;54:166-169 (PDB: 1Q1L).

43. Canduri F, Silveira NJF, Camera JCJr, De Azevedo WFJr. Structural bioinformatics study of cyclin-dependent kinases complexed with Inhibitors. *Eclética Quím.* 2003;28:45-53.

44. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol.* 1993;234:779-815.

45. Fiser A, Do RK, Sali A. Modeling of loops in protein structures. *Protein Science.* 2000;9:1753-1773.

46. Marti-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct.* 2000;29:291-325.

47. Eswar N, Marti-Renom MA, Webb B, Madhusudhan MS, Eramian D, Shen M, Pieper U, Sali A. Comparative protein structure modeling with MODELLER. *Current Protocols in Bioinformatics*, John Wiley & Sons, Inc. 2006;Supplement 15:5.6.1-5.6.30.

48. Shoichet, B.K. Virtual screening of chemical libraries. *Nature*. 2004;16(432):862-865.
49. Irwin JJ, Shoichet BK. ZINC - a free database of commercially available compounds for virtual screening. *J Chem Inf Model*. 2005;45(1):177-182.
50. Thomsen R, Christensen MHJ. MolDock: a new technique for high-accuracy molecular docking. *Med Chem*. 2006;49:3315-3321.
51. De Azevedo WFJr, Dias R. Evaluation of ligand-binding affinity using polynomial empirical scoring functions. *Bioorg Med Chem*. 2008;16:9378-9383.
52. Dias R, Timmers LFSM, Caceres RA, De Azevedo WFJr. Evaluation of molecular docking using polynomial empirical scoring functions. *Curr Drug Targets*. 2008;9:1062-1070.
53. Ducati RG, Basso LA, Santos, DS, De Azevedo WFJr. Crystallographic and docking studies of purine nucleoside phosphorylase from *Mycobacterium tuberculosis*. *Bioorg Med Chem*. 2010; 18:4769-4774.
54. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*. 2001;46:3-26.
55. Miteva MA, Violas S, Montes M, Gomez D, Tuffery P, Villoutreix BO. FAF-Drugs : free ADME/tox filtering of compound collections. *Nucleic Acids Res*. 2006;34:W738-744.
56. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJ. GROMACS: fast, flexible and free. *J Comput Chem*. 2005;26:1701-1718.
57. Hess E, Kutzner C, Van Der Spoel D, Lindahl B. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput*. 2008;4(3):435-447.
58. Guex N, Peitsch MC. SWISS_MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*. 1997;18:2714-2723.
59. Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. *Protein Eng*. 1995;8:127-134.
60. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA, Jr, Vreven T, Kudin KN, Burant JC, Millam IM, Iyengar SS, *et al*. Gaussian 03 - Revision C.02. Pittsburgh. Gaussian Inc.; 2003.
61. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst*. 1993;26:283-291.
62. Raghava GPS. Protein secondary structure prediction using nearest neighbor and neural network approach. *CASP4*. 2000;75-76.

- 63.** Macheroux P, Schönbrunn E, Svergun DI, Volkov VV, Koch MH, Bornemann S, Thorneley RNF. Evidence for a major structural change in *Escherichia coli* chorismate synthase induced by flavin and substrate binding. *Biochemical Journal*. 1998;335 (Pt 2):319-327.
- 64.** Keiser MJ, Roth BL, Armbruster BN, Ernsberger P, Irwin JJ, Shoichet BK. Relating protein pharmacology by ligand chemistry. *Nat Biotech*. 2007;25(2):197-206.
- 65.** Oostenbrink C, Soares TA, Van Der Vegt NF, Van Gunsteren WF. Validation of the 53A6 GROMOS force field. *Eur Biophys J*. 2005;34:273-284.
- 66.** Schüttelkopf AW, Van Aalten DM. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr*. 2004;60:1355-1363.
- 67.** Berendsen HJC, Postma JPM, Van Gunsteren WF, Hermans J. Interaction models for water in relation to protein hydration. In: Pullman B (ed) *Intermolecular forces*. The Netherlands. Reidel D Publishing Company Dordrecht; 1981:331-342.
- 68.** Arfken G The Method of Steepest Descents. In: *Mathematical Methods for Physicists*. Orlando. FL Academic Press; 1985:428-436.
- 69.** Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: a linear constraint solver for molecular simulations. *J Comput Chem*. 1997;18:1463-1472.
- 70.** Darden T, York D, Pedersen L. Particle Mesh Ewald - an $n \cdot \log(n)$ method for ewald sums in large systems. *J Chem Phys*. 1993;98:10089-10092.

8. ANEXOS:

ANEXO A: Comprovante de submissão do artigo

De: em.jmm.0.203cc8.d60b6162@editorialmanager.com em nome de Journal of Molecular Modeling

Enviada: sex 7/1/2011 16:21

Para: Walter Filgueira de A Junior

Assunto: Submission Confirmation

Dear Mr Walter de Azevedo Jr.,

Thank you for submitting your manuscript entitled: "Analysis of the molecular basis for binding of Mycobacterium tuberculosis chorismate synthase toward its natural substrate and possible inhibitors: A docking and molecular dynamics study. to the Journal of Molecular Modeling.

The manuscript number will be sent to you after your submission has been assigned to the Editor.

If your manuscript is accepted for publication in Journal of Molecular Modeling, you may elect to submit it to the Open Choice program. For information about the Open Choice program, please access the following URL:
<http://www.springer.com/openchoice>

If you have any questions, please do not hesitate to contact us.

Kind regards

ANEXO B: Alinhamento das seqüências da estrutura 3D da MtCS (PDB: 2O12/molde) contendo os resíduos GLY 47, ALA 48, ARG 49, MET 50, THR 51 e PHE 52 e a própria seqüência primária da enzima, com 401 aminoácidos (alvo).

a) alinhamento feito pelo programa MODELLER; b) ajuste manual do alinhamento feito pelo programa MODELLER.

a)

```

aln.pos          10          20          30          40          50          60
2O12A.pdb MLRWITAGESHGRALVAVVEGMVAGVHVTSADIADQLARRRLGYGR-----RDAVTVLSGIRHGS
MtCS       MLRWITAGESHGRALVAVVEGMVAGVHVTSADIADQLARRRLGYGRGARMTFERDAVTVLSGIRHGS
_consrvd   *****

```

b)

```

aln.pos          10          20          30          40          50          60
2O12A.pdb MLRWITAGESHGRALVAVVEGMVAGVHVTSADIADQLARRRLGYGR-----RDAVTVLSGIRHGS
MtCS       MLRWITAGESHGRALVAVVEGMVAGVHVTSADIADQLARRRLGYGRGARMTFERDAVTVLSGIRHGS
_consrvd   *****

aln.pos          70          80          90          100         110         120         130
2O12A.pdb TLGGPIAIEIGNTEWPKWETVMAADPVDPAELADVARNAPLTRPRPGHADYAGMLKYGFDDARPVLE
MtCS       TLGGPIAIEIGNTEWPKWETVMAADPVDPAELADVARNAPLTRPRPGHADYAGMLKYGFDDARPVLE
_consrvd   *****

aln.pos          140         150         160         170         180         190         200
2O12A.pdb RASARETAARVAAGTVARAFRLRQALGVEVLSHVISIGASAPYEGPPPRAEDLPAIDASPVRAYDKAA
MtCS       RASARETAARVAAGTVARAFRLRQALGVEVLSHVISIGASAPYEGPPPRAEDLPAIDASPVRAYDKAA
_consrvd   *****

aln.pos          210         220         230         240         250         260
2O12A.pdb EADMIAQIEAAKKGDTLGGVVEAVALGLPVGLGSFTSGDHRLDSQLAAAVMGIQAIKVEIGDGFQ
MtCS       EADMIAQIEAAKKGDTLGGVVEAVALGLPVGLGSFTSGDHRLDSQLAAAVMGIQAIKVEIGDGFQ
_consrvd   *****

aln.po 270         280         290         300         310         320         330
2O12A.pdb TARRRGSRAHDEMYPGPDGVVVRSTNRAGGLEGGMTNGQPLRVRAAMKPISTVPRALATVDLATGDEA
MtCS       TARRRGSRAHDEMYPGPDGVVVRSTNRAGGLEGGMTNGQPLRVRAAMKPISTVPRALATVDLATGDEA
_consrvd   *****

aln.pos          340         350         360         370         380         390
2O12A.pdb VAIHQRSQDVCAVPAAGVVVETMVALVLARAALKFGGDSLAEQQRNIAAYQRSVADR
MtCS       VAIHQRSQDVCAVPAAGVVVETMVALVLARAALKFGGDSLAEQQRNIAAYQRSVADR
_consrvd   *****

```

ANEXO C: Parâmetros do programa MolDock: os valores usados como parâmetros de *input* para as simulações de docagem no *MolDock*:

Preparation (arquivo de entrada proteína/ligantes)

Assign all below: Custom

Assign bonds: if missing

Assign bonds orders and hybridization: if missing

Create explicit hydrogens: if missing

Assign charges (calculated by MVD): always

Detect flexible torsions in ligands: always

Assign Tripos atom types: if missing

Scoring fuction

Score: Moldock Score Grid

Grid resolution (Å): 0.30

Binding site

Origin – Center:

X: 6.22

Y: 25.30

Z: -36,01

Radius: 15

Search algorithm

Algorithm: MolDock Optimizer

Number of runs: 10

Constrain poses to cavity

After docking: Optimize H-Bonds

Parameter settings

Population size: 50

Max iterations: 2000

Scaling factor: 0.5

Crossover rate: 0.9

Offspring scheme: Scheme 1

Termination scheme: Variance-based

Poses Clustering

Return multiple poses for each run

Max number of poses returned: 5

Cluster similar poses. RMSD threshold: 1.00

Ignore similar poses (for multiple runs only). RMSD threshold: 1.00