



Modelagem Molecular da Chiquimato Quinase de
"Mycobacterium leprae"

José Renato Panzarini Junior

Dissertação apresentada para obtenção do título de Mestre em Biofísica Molecular, área de concentração Biofísica Molecular do Departamento de Física do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista "Julio de Mesquita Filho" – UNESP.

Orientador: Prof. Dr. Walter Filgueira de Azevedo Júnior

São José do Rio Preto
Estado de São Paulo – Brasil
Agosto-2006

“A humanidade está sofrendo não porque tenha deixado de progredir, mas,
porque progrediu depressa demais.”

(Albert Einstein)

AGRADECIMENTOS

Ao Prof. Dr. Walter Filgueira de Azevedo Junior, pela orientação, companheirismo, pela oportunidade de completar uma etapa em minha vida, pela amizade, paciência, pelos conselhos e apoio durante o decorrer do trabalho,

A minha família, principalmente aos meus pais e avós que sempre me incentivaram e estiveram do meu lado dando todo o apoio necessário para superar qualquer barreira e com isso alcançar algo melhor em minha vida,

Aos colegas de sala Nelson José e Karisa Karla que tiveram muita paciência dando sugestões e discutindo sobre os problemas encontrados no dia a dia durante o desenvolvimento do meu trabalho,

A Helen que também sempre esteve presente participando e discutindo sobre os problemas do meu projeto,

A todos os outros colegas do Laboratório de Sistemas Biomoleculares pelo companheirismo e amizade durante todo esse tempo junto,

Aos professores participantes da banca pela participação e pelas sugestões para que possa finalizar o meu projeto,

A todos os professores e funcionários do departamento que estiveram presentes durante esses anos,

A Capes, CNPq e a FAPESP pelo apoio financeiro e a manutenção dos laboratórios,

SUMÁRIO

LISTA DE FIGURAS.....	4
LISTA DE TABELAS.....	5
LISTA DE ABREVIACOES.....	6
RESUMO.....	7
ABSTRACT.....	9
1. INTRODUO.....	11
1.1. Via Metablica do cido Chiqumico.....	13
1.2. Chiquimato Quinase.....	21
2. OBJETIVOS.....	24
3. METODOLOGIA E FERRAMENTAS.....	25
3.1. Modelagem Molecular por Homologia ou Comparativa.....	25
3.1.1. Identificao e Seleo de Protenas molde.....	27
3.1.2. Alinhamento.....	28
3.1.3. Construo do Modelo.....	28
3.1.4. Avaliao do Modelo.....	33
3.2. Mtodos de Avaliao.....	36
3.2.1. PROCHECK.....	36
3.2.2. Diagrama de Ramachandran.....	37
3.2.3. G-Factor.....	38
3.2.4. VERIFY-3D.....	39
3.2.5. WHATIF/WHATCHECK.....	40
3.2.6. RMSD (Desvio Mdio Quadrtico).....	41
4. RESULTADOS E DISCUSSO.....	44
4.1. Comparao da Seqncia Primria.....	44
4.2. Qualidade do Modelo.....	45
4.3. Descrio do Modelo.....	59
5. CONCLUSO.....	61

6. REFERÊNCIAS BIBLIOGRÁFICAS..... 63

LISTA DE FIGURAS

	Páginas
Figura 1: Distribuição geográfica da taxa de prevalência do <i>Mycobacterium leprae</i> no mundo.....	12
Figura 2 – Estruturas químicas do ácido chiquímico e do ácido corísmico.....	14
Figura 3 – A via metabólica do ácido chiquímico.....	16
Figura 4 – Biossíntese dos aminoácidos aromáticos a partir do corismato.....	17
Figura 5 – O corismato e os compostos essenciais sintetizados a partir do corismato.....	20
Figura 6 – Reação catalisada pela enzima SK.....	21
Figura 7 – Esquema dos passos da modelagem molecular comparativa.....	27
Figura 8 – Modelagem comparativa de estrutura de proteína por satisfação de restrição espacial.....	31
Figura 9 – Parte de uma cadeia polipeptídica indicando os graus de liberdade torcional de cada unidade peptídica e os ângulos de rotação ϕ e Ψ	37
Figura 10 – Diagrama de Ramachandran.....	37
Figura 11 – Alinhamento das seqüências primárias para da Chiquimato Quinase de <i>M. leprae</i> e a Chiquimato Quinase de <i>M. tuberculosis</i>	45
Figura 12 – Diagrama de Ramachandran dos modelos finais e moldes.....	47
Figura 13 – Gráfico do programa VERIFIC 3D dos modelos finais da Chiquimato Quinase de <i>M. leprae</i>	49 e 50
Figura 14 - Gráfico do programa VERIFIC 3D para os moldes.....	51 e 52
Figura 15 – Sobreposição dos C_{α} dos modelos.....	54
Figura 16 – Superfície molecular das estruturas modeladas.....	55
Figura 17 – Estruturas secundárias da Chiquimato Quinase de <i>M. leprae</i>	60 e 61

LISTA DE TABELAS

	Páginas
Tabela 1 – Tabela com as chiquimato quinases de estruturas depositadas no PDB.	23
Tabela 2 – Resultado das análises do Diagrama de Ramachandran dos moldes e dos modelos de CQ de <i>M. leprae</i> .	46
Tabela 3 – Resultado dos valores do G-factor dos moldes e dos modelos de CQ de <i>M. leprae</i> .	48
Tabela 4 – Resultado dos valores do VERIFY 3D dos moldes e dos modelos de CQ de <i>M. leprae</i> .	49
Tabela 5 – Resultado dos valores do RSMD da geometria ideal dos moldes e dos modelos de CQ de <i>M. leprae</i>	52
Tabela 6 – Resultado dos valores do RSMD da sobreposição ideal dos moldes e dos modelos de CQ de <i>M. leprae</i>	53
Tabela 7 – Ligações intermoleculares de hidrogênio da chiquimato do modelo da <i>mlSK</i> comparando os valores com as ligações da chiquimato com <i>mtSK</i> (1WE2).	56
Tabela 8 – Ligações intermoleculares de hidrogênio do ADP, do modelo da <i>mlSK</i> comparando os valores com as ligações da chiquimato com <i>mtSK</i> (1L4Y).	57
Tabela 9 – Ligações intermoleculares de hidrogênio do ADP, do modelo da <i>mlSK</i> comparando os valores com as ligações da chiquimato com <i>mtSK</i> (1WE2).	58
Tabela 10 – Ligações intermoleculares de hidrogênio do ADP, do modelo da <i>mlSK</i> comparando os valores com as ligações da chiquimato com <i>mtSK</i> (2DFT).	59

LISTA DE ABREVIACÕES

3D – Tridimensional

ATP – Adenosina Trifosfato

ADP – Adenosina Difosfato

CQ – Chiquimato Quinase

CS – Corismato sintase

DAHHP – 3-deoxi-D-arabino-ácido heptulosônico 7-fosfato

DAHPS – 3-deoxi-D-arabino-ácido heptulosônico 7-fosfato sintase

DHQP – Desidroquinato desidratase

DHQS – 3-Desidroquinato sintase

E4P – Eritrose-4-fosfato

EPSPS - 5-Enoilpiruvilchiquimato 3-fosfato sintase

Mg²⁺ - Íon de magnésio (II)

MI - *Mycobacterium leprae*

NADH – Nicotinamida adenina dinucleotídeo

NMP – Monofosfato de nucleosídeo

OMS – Organização Mundial de Saúde

PABA – Ácido *p*-aminobenzóico

PDF – Função de Densidade de Probabilidade

PEP – Fosfoenolpiruvato

Phe – Aminoácido Fenilalanina

RMN – Ressonância Magnética Nuclear

RMSD – Desvio médio quadrático

S3P – Chiquimato 3-fosfato

SB – Chiquimato-ligado

SD – Chiquimato-5-desidrogenase

SHI – Chiquimato quinase

Trp – Aminoácido Triptofano

Tyr – Aminoácido Tirosina

RESUMO

A lepra ou hanseníase é uma doença crônica infecciosa de pele e nervos periféricos causados pelo patógeno intracelular obrigatório do *Mycobacterium leprae* (*M. leprae*). A lepra afeta milhões de pessoas em todo mundo, principalmente os habitantes de países pobres em desenvolvimento e industrializados.

O seqüenciamento do genoma do *M. leprae* revelou a presença de genes da via metabólica do ácido chiquímico, que tem uma importância fundamental na produção de compostos aromáticos em plantas, bactérias e fungos. A via metabólica do ácido chiquímico consiste em sete enzimas que catalisam a conversão seqüencial de eritrose-4-fosfato (E4P) e fosfoenolpiruvato (PEP) em corismato, a qual está ausente em humanos, minimizando os efeitos tóxicos das drogas no organismo e aumentando a especificidade das novas drogas desenvolvidas.

A análise do modelo estrutural da chiquimato quinase de *M. leprae* (MISK) complexada com ácido chiquímico, adenosina difosfato (ADP) e magnésio (Mg^{+2}), a qual catalisa o quinto passo da via metabólica do ácido chiquímico obtidos por métodos de modelagem molecular comparativa foi validado e poderá ser utilizado para o desenho de drogas específicas contra o *M. leprae* e vários outros microrganismos que apresentam a mesma via.

Este trabalho aumenta a certeza de que a modelagem comparativa é uma ferramenta útil em bioinformática estrutural, uma vez que não se tem acesso às estruturas determinadas experimentalmente, podendo ser valiosa na anotação de

seqüências genômicas, contribuindo para a genômica estrutural e funcional, e simulações de *docking* proteína-ligante.

ABSTRACT

Leprosy or Hansen's disease is a chronic infection of the skin and peripheral nerves caused for obligate intracellular pathogen *Mycobacterium leprae*. The leprosy is affecting million persons in whole world, principally the inhabitants of poor countries in development and industrialized.

The genome sequencing of *M. leprae* has revealed the presence of genes of metabolic pathway of the shikimic acid, what is pivotal importance to production of aromatic compounds in plants, bacteria and fungi. The metabolic pathway of the shikimic acid consists of seven enzymes that catalyse the sequential conversion of erythrose-4-phosphate (E4P) and phosphoenolpyruvate (PEP) to chorismate, which is absent in humans, minimizing the toxic effects of the drugs in the organism and increasing the specificity of the new drugs developed.

The analysis of shikimate kinase structural model of *M. leprae* (MISK) complex with shikimic acid, adenosine diphosphate (ADP) and magnesium (Mg^{+2}), which catalysing the fifth step of shikimic acid metabolic pathway obtained by methods of molecular comparative modeling was validated and it will be able to draw specific drugs against *M. leprae* and several other microorganisms that presents the same pathway.

This work increase the conviction that comparative modeling is an useful tool in structural bioinformatics, once that have not access to the experimentally determined estructures, can be valuable in annotating genome sequence,

contributing to the structural and functional genomics, and protein-ligand docking simulations.

1. INTRODUÇÃO

A lepra ou hanseníase é uma doença crônica infecciosa de pele e nervos periféricos causada pelo patógeno intracelular obrigatório *Mycobacterium leprae*. Os danos para os nervos periféricos resultam no enfraquecimento motor e sensor com características de deformação e inabilidade (BRITTON & LOCKWOOD, 2004).

Em 1985, foram estimadas 12 milhões de pessoas com lepra no mundo, uma prevalência de 12 por 10.000. Em 2000, a Organização Mundial de Saúde (OMS) relatou que foram 597.000 casos registrados e 719.000 novos casos detectados durante 2002, resultando em uma prevalência global de pacientes leprosos registrados abaixo de 1 por 10.000. Quinze países ainda tem uma prevalência maior que 1 por 10.000, principalmente países da Ásia, África e América do Sul. Há uma concentração de 83% dos casos registrados somente em 6 países: Índia, Brasil, Burma, Indonésia, Madagascar e Nepal (BRITTON & LOCKWOOD, 2004). Com isso, fica clara a presença da bactéria em países pobres em desenvolvimento e industrializados.

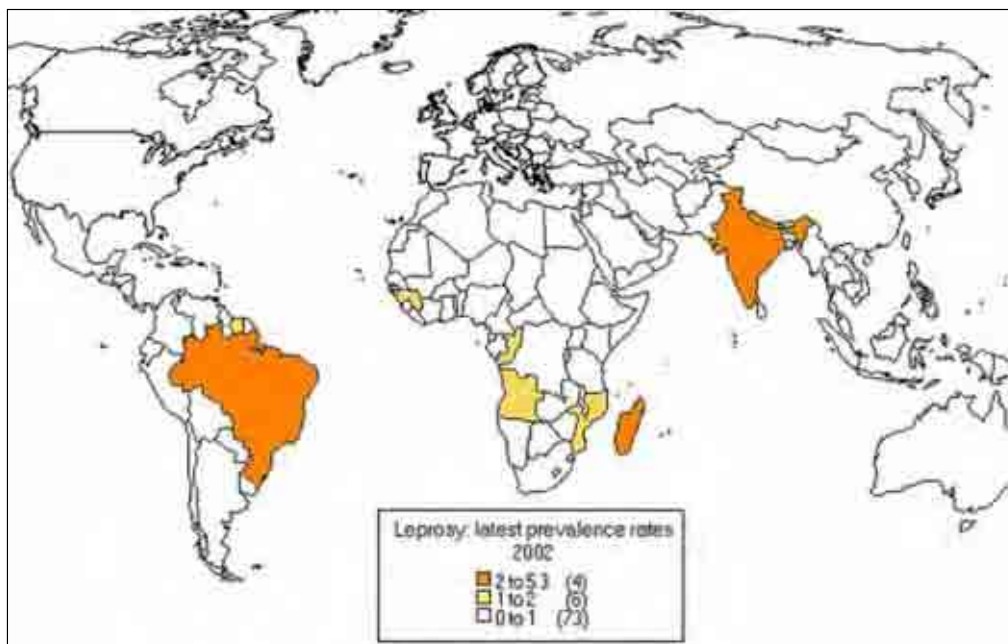


Figura 1: Distribuição geográfica da taxa de prevalência do *Mycobacterium leprae* no mundo em 2002(OMS).

A lepra vem infeccionando a população humana por milhares de anos e confundido cientistas desde a identificação do agente etiológico, o *M. leprae* (MONOT et. al., 2005).

O *M. leprae* foi descoberto em 1873 pelo cientista norueguês Gerhard Armauer Hansen, sendo morfológicamente muito semelhante ao bacilo causador da tuberculose (<http://www.brasilecola.com/doencas/hanseniose.htm>). O *M. leprae* é um organismo resistente, pode sobreviver fora do corpo durante até 45 dias. Em países onde a lepra é endêmica, tais como Etiópia e Indonésia, até 5% da população carregam o DNA do *M. leprae*, sem evidência da doença.

A transmissão do *M. leprae* se dá através da mucosa nasal de pacientes não tratados, provavelmente o *mycobacterium* sobrevive no meio ambiente antes de infectar o próximo hospedeiro (LOCKWOOD & SUNEETHA, 2005). Outra possibilidade é o contato direto com a pele através de feridas de doentes. No

entanto, é necessário um contato íntimo e prolongado para a contaminação, como a convivência de familiares na mesma residência. Daí a importância do exame dos familiares do doente de hanseníase. A maioria da população adulta é resistente à hanseníase, mas as crianças são mais susceptíveis, geralmente adquirindo a doença quando há um paciente contaminado na família. O período de incubação varia de 2 a 7 anos e entre os fatores predisponentes estão o baixo nível sócio-econômico, a desnutrição e a superpopulação doméstica (<http://www.dermatologia.net/neo/base/doencas/hanseniasse.htm>).

Como não se transmite com facilidade, não são registradas grandes epidemias. A hanseníase chegou ao Brasil na época do descobrimento, espalhou-se e permanece, até hoje, sendo considerada um dos mais sérios problemas da saúde pública no país. Ocupamos o 1º lugar da América Latina e 2º lugar do mundo em número de doentes, ficando atrás somente da Índia (BRITTON & LOCKWOOD, 2004).

1.1. Via Metabólica do Ácido Chiquímico

A via metabólica do ácido chiquímico é uma das rotas biossintéticas que leva à produção de alguns dos metabólitos primários e secundários importantes na natureza (GANEM, 1978; DE OLIVEIRA, 2001). O ácido chiquímico (Figura 2) foi primeiramente isolado por Eykmann, em 1885, a partir da fruta da planta oriental *Illicium religiosum*, sendo seu nome derivado do nome em japonês para esta planta, shikimi-no-ki (CAMPBELL *et al.*, 1993). Entretanto, sua estrutura permaneceu desconhecida por aproximadamente 50 anos. Somente na década de 30, graças aos esforços pioneiros de Fisher, Freudenberg e Karrer, a estrutura

completa e a estereoquímica absoluta deste ácido carboxílico foi determinada (GANEM, 1978).

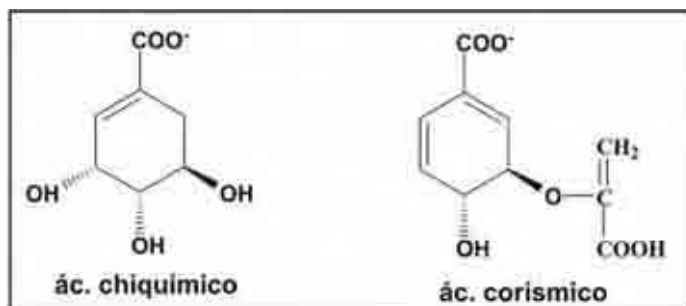


Figura 2 – Estruturas químicas do ácido chiquímico e do ácido corísmico.

Estudos subsequentes de Bernhard Davis (1950) e de outros pesquisadores como David Sprinson e Frank Gibson nos 25 anos seguintes (GIBSON, 1999) revelaram uma rota bem definida de oito metabólitos intermediários que conduziam, a partir da glicose, primeiro para o ácido chiquímico, e depois ao ácido corísmico (Figura 2). Estas substâncias constituíam uma rota biossintética comum que seria re-direcionada para a formação de vários metabólitos aromáticos mono e policíclicos em plantas e microrganismos por diferentes caminhos (GANEM, 1978).

A via metabólica do ácido chiquímico conecta o metabolismo de carboidratos a biossíntese de compostos aromáticos, numa seqüência de sete etapas metabólicas. A etapa inicial é a condensação do PEP com E4P. Na segunda etapa o 3-deoxi-D-arabino-ácido heptulosônico 7-fosfato (DAHP) perde seu grupamento fosforila e cicliza-se para 3-desidroquinato. A desidratação fornece então 3-desidrochiquimato, que é reduzido usando como co-fator nicotinamida adenina dinucleotídeo (NADH) a D-chiquimato. Na quinta etapa ocorre uma fosforilação específica do grupo 3-hidroxil da D-chiquimato usando

adenosina trifosfato (ATP), como co-fator. Na sexta etapa uma segunda molécula de PEP condensa-se com chiquimato 3-fosfato (S3P) para produzir 5-enoilpiruvil 3-fosfato. Na etapa final o 5-enoilpiruvil 3-fosfato perde seu grupamento fosforila, produzindo corismato (Figura 3), o precursor dos aminoácidos aromáticos (Phe, Trp e Tyr) (Figuras 4) e de muitos metabólitos secundários aromáticos. Todas as vias intermediárias também podem ser consideradas como compostos de conexão que podem servir como substrato para outras vias metabólicas (HERRMANN & WEAVER, 1999).

A via metabólica do ácido chiquímico é composta pelas enzimas: 3-deoxi-D-arabino-ácido heptulosônico 7-fosfato Sintase (DAHPS), 3-Desidroquinato Sintase (DHQS), 3-Desidroquinato Desidratase (DHQD), Chiquimato-5-Desidrogenase (SD), Chiquimato Quinase (SK), 5-Enoilpiruvilchiquimato 3-fosfato Sintase (EPSPS) e Corismato Sintase (CS).

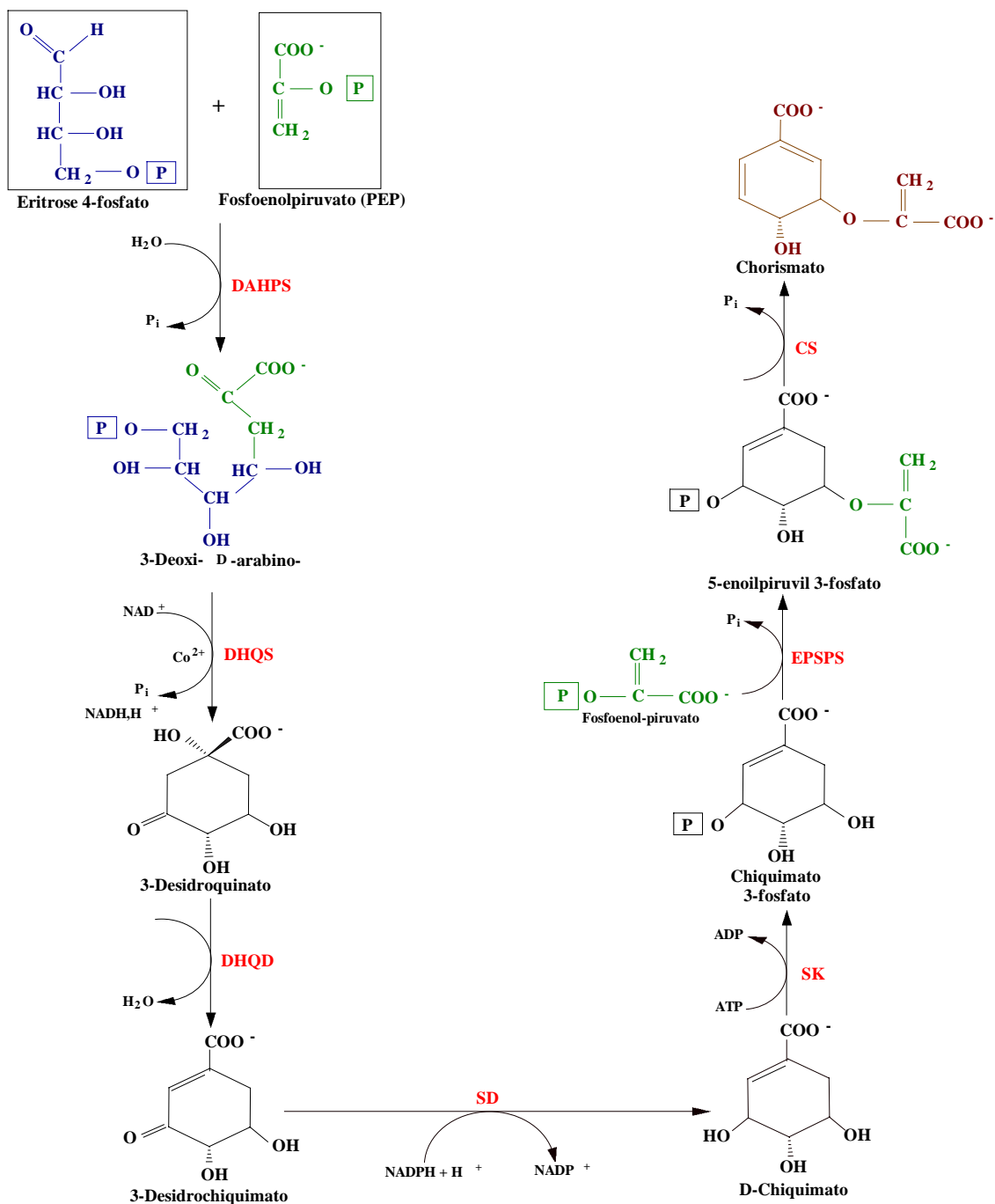


Figura 3 – A via metabólica do ácido chiquímico. As sete enzimas estão indicadas à direita em vermelho.

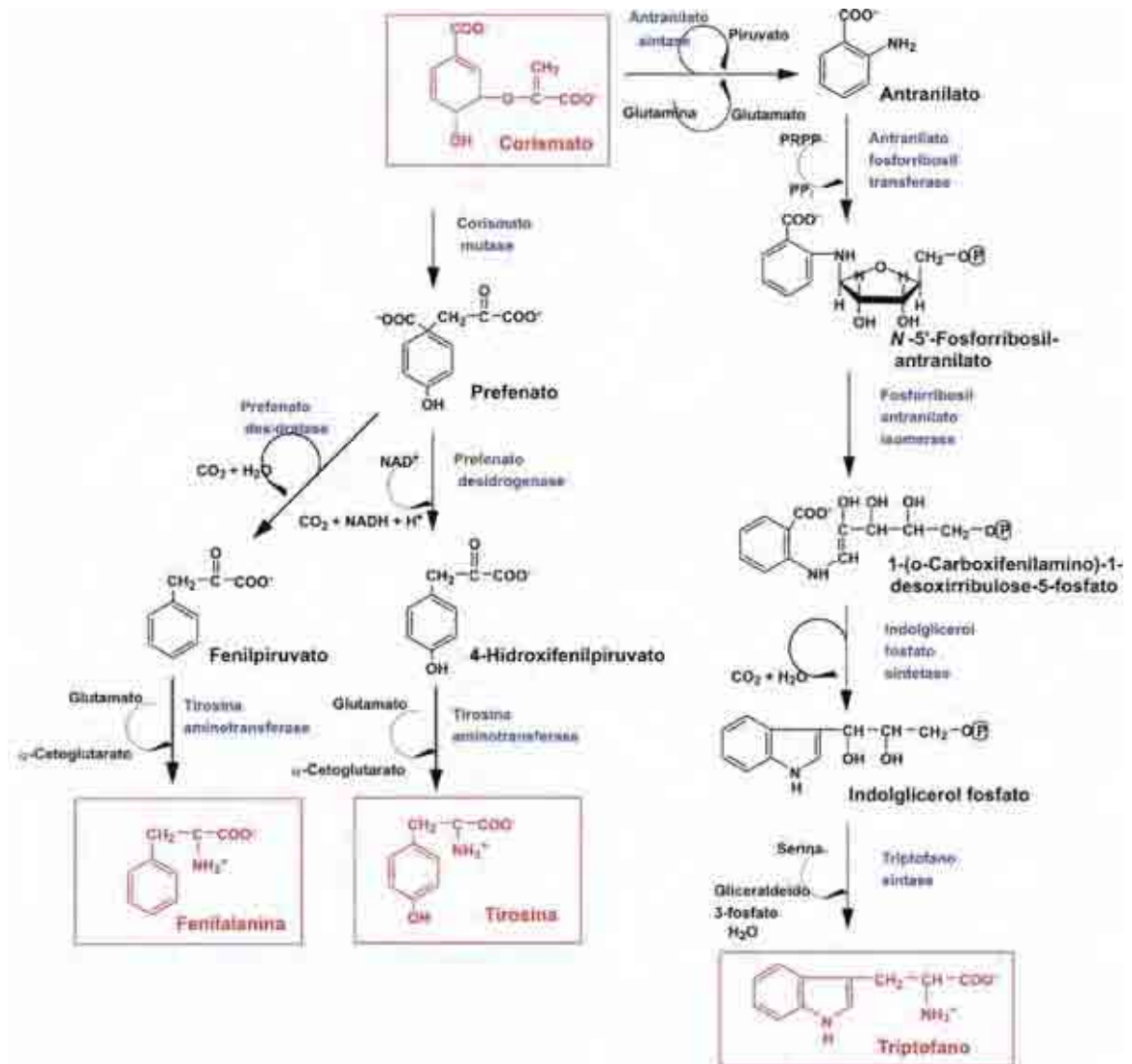


Figura 4 – Biossíntese dos aminoácidos aromáticos a partir do corismato. O corismato e os aminoácidos aromáticos estão indicados por quadros, e as enzimas estão indicadas em azul. (DE OLIVEIRA, 2001).

A via metabólica do ácido chiquímico é essencial para algas, plantas, bactérias e fungos, mas é ausente em animais, pois esses adquirem os aminoácidos aromáticos através da sua alimentação (COGGINS *et al.*, 2003). Por causa disso as enzimas pertencentes a esta via representam importantes alvos para o desenvolvimento de herbicidas (KISHORE & SHAH, 1988), antimicrobianos potencialmente não tóxicos (DAVIES *et al.*, 1994) e drogas antiparasitas (ROBERTS *et al.*, 1998) (KRELL *et al.*, 1998; PEREIRA *et al.*, 2004a; PEREIRA *et al.*, 2004b).

Em microrganismos, a via metabólica do ácido chiquímico é regulada pela inibição por retroalimentação e pela repressão da primeira enzima: a DAHPS nas vias intermediárias ou para produtos finais. Em plantas, nenhum inibidor fisiológico pela retroalimentação foi identificado. Esta diferença entre microrganismos e plantas é refletida na variação das estruturas primárias das DAHPS (HERRMANN & WEAVER, 1999). A organização e a estrutura molecular das enzimas da via metabólica do ácido chiquímico variam consideravelmente entre grupos de microrganismos.

As bactérias têm sete polipeptídios individuais, cada um possui uma única atividade enzimática, e são codificados por genes separados. As plantas têm um arranjo molecular similar às bactérias, com a exceção do DHQD e SD, que possuem domínios separados em um polipeptídio bifuncional (CAMPBELL *et al.*, 2004; HERRMANN & WEAVER, 1999). Em contraste, todos os fungos examinados têm as enzimas DAHPS e CS como polipeptídios monofuncionais e o restante das enzimas como um polipeptídio pentafuncional denominado

AROM. O polipeptídeo AROM tem os domínios análogos às enzimas bacterianas: DHQS, EPSPS, SK, DHQD e SD (CAMPBELL *et al.*, 2004). As enzimas DHQS e EPSPS formam o domínio N-terminal e SK, DHQD e SD formam o domínio C-terminal da proteína. A proteína de AROM está relacionada à regulação da transcrição eucariótica (CARPENTER *et al.*, 1998; HERRMANN & WEAVER, 1999).

Um aspecto evolucionário muito interessante do metabolismo aromático em fungos é uma duplicação do gene *arom* que resulta nos genes que codificam proteínas regulatórias para a degradação do quinato (HERRMANN & WEAVER, 1999). Os complexos AROM estudados foram *Neurospora crassa*, *Aspergillus nidulans* (CARPENTER *et al.*, 1998), *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, e *Pneumocystis carinii* (HERRMANN & WEAVER, 1999).

O corismato é o último produto comum da via metabólica do ácido chiquímico e precursor chave para a biossíntese de PABA, ácido *p*-hidroxibenzóico, micobactinas e dos aminoácidos aromáticos essenciais (Figuras 4 e 5) (ROBERTS *et al.*, 2002; DOSSELAERE & VANDERLEYDEN, 2001).

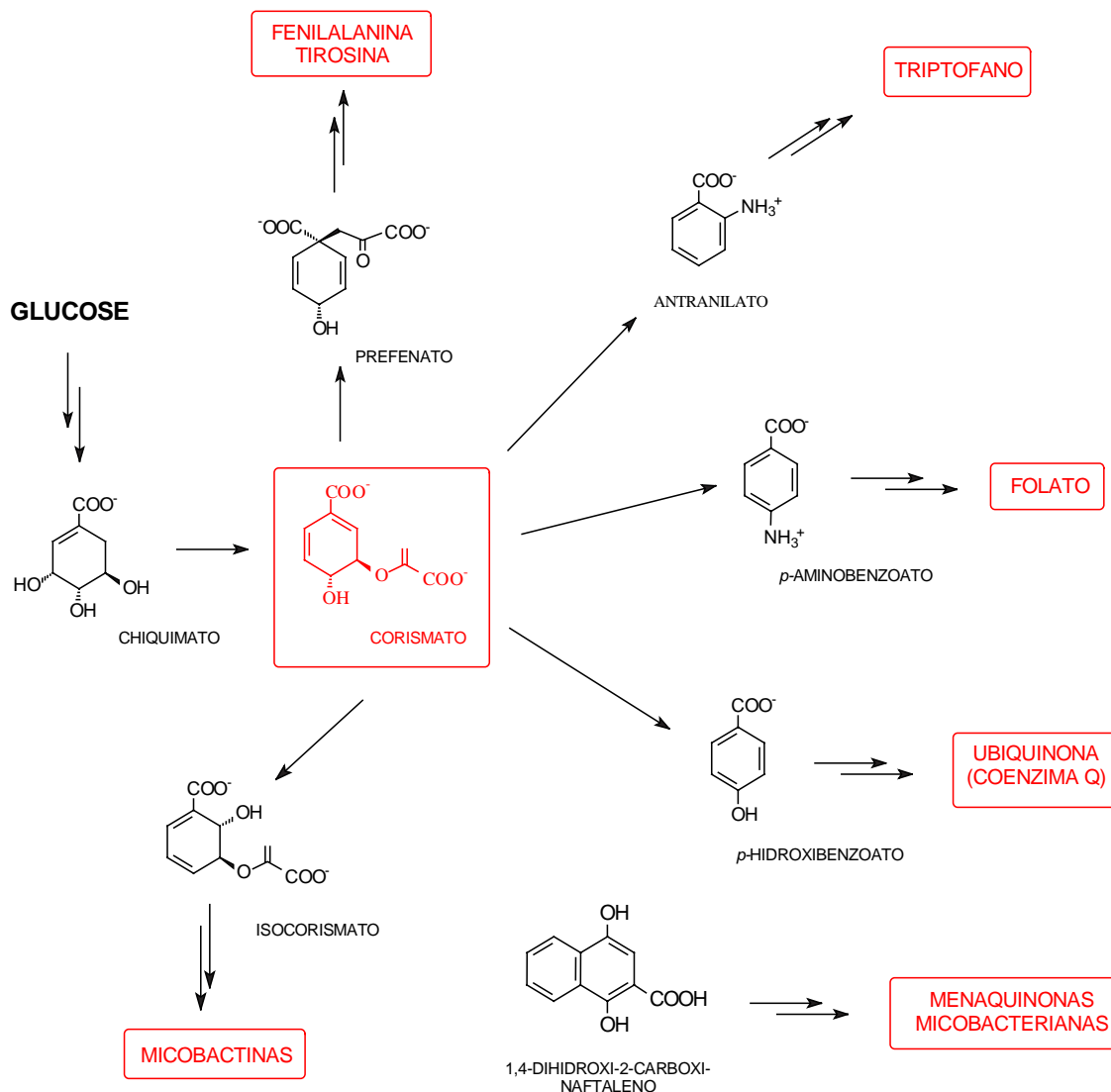


Figura 5 – O corismato e os compostos essenciais sintetizados a partir do corismato estão destacados por quadros em vermelho (DE OLIVEIRA, 2001).

As bactérias gastam mais de 90 % da sua energia metabólica total na biossíntese de proteína. Conseqüentemente, a via metabólica do ácido chiquímico em bactérias serve quase exclusivamente para sintetizar aminoácidos aromáticos (HERRMANN, 1995a). O bloqueio de qualquer uma das sete etapas da via metabólica do ácido chiquímico por meio da inibição de uma de suas enzimas

específicas poderia, em princípio, levar a bactéria à morte pela não disponibilização de metabólitos essenciais para a sua sobrevivência (HERRMANN, 1995a). Qualquer microrganismo poderá, em princípio, ser combatido por inibidores das enzimas da via metabólica do ácido chiquímico que impossibilitarão a produção do corismato.

1.2. Chiquimato Quinase

A chiquimato quinase é a quinta enzima da via metabólica do ácido chiquímico (Figura 6), catalisa a fosforilação específica do grupo 3-hidroxil do ácido chiquímico usando Adenosina Trifosfato (ATP) como um co-substrato (KRELL *et al.*, 1998; PEREIRA *et al.*, 2004a; PEREIRA *et al.*, 2004b).

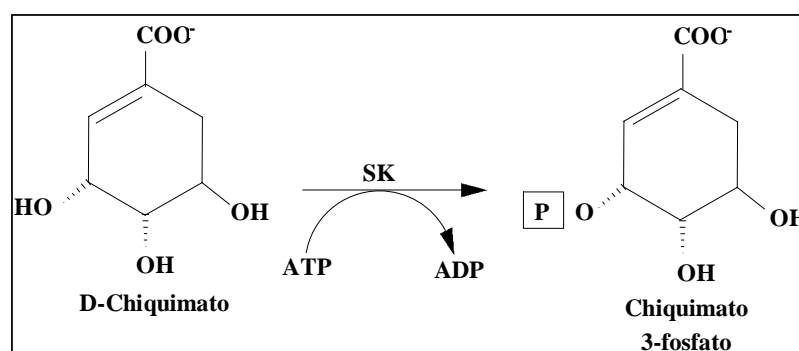


Figura 6 – Reação catalisada pela enzima SK.

A SK pertence à família de quinases do monofosfato de nucleosídeo (NMP). Uma característica das quinases de NMP é que sofrem uma grande mudança conformacional durante a catálise (VONRHEIN *et al.*, 1995). Existem duas regiões flexíveis na estrutura que são o sítio de ligação SB e o domínio LID. Na SK, o domínio de ligação do chiquimato (SB) corresponde ao domínio de NMPB de quinases NMP (PEREIRA *et al.*, 2004b). As quinases de NMP (YSAI

& YAN, 1999) são compostas de três domínios: o NÚCLEO (*CORE*), a TAMPA (*LID*), e o domínio de ligação do NMP (NMPB).

O domínio *CORE* contém um laço de ligação do fosfato (*P-loop*) altamente conservado, o domínio *LID* fecha o sítio de ATP e carrega resíduos funcionalmente essenciais, e o domínio NMPB é responsável pelo reconhecimento e ligação específica de NMP (GU *et al.*, 2002).

Os três motivos funcionais de ligação do nucleotídeo das enzimas são reconhecidos para SK de *M. tuberculosis* (PEREIRA *et al.*, 2004a,b; GU *et al.*, 2002), incluindo o motivo Walker A, um motivo Walker B e um laço de ligação da adenina. O motivo Walker A é localizado entre a primeira folha β (β_1) e a primeira hélice α (α_1), contendo a sequência conservada GXXXXGKT/S (WALKER *et al.*, 1982), onde X representa um aminoácido qualquer. Este motivo dá forma ao laço de ligação do fosfato (*P-loop*). Além do motivo Walker A, uma segunda sequência conservada ZZDXXG chamada de motivo Walker B (WALKER *et al.*, 1996) é observada, onde Z representa um resíduo hidrofóbico (PEREIRA *et al.*, 2004a,b), onde se acomoda o íon Mg^{2+} . Este domínio é encontrado em todas as enzimas SK conhecidas e também em diversas proteínas que ligam ATP e GTP (KRELL *et al.*, 1998). Todas estas proteínas que contêm o laço de ligação do fosfato (*P-loop*) apresentam uma estrutura em seu domínio central formada de quatro fitas β paralelas e por dois pares de hélices α -flanqueadas (MILNER-WHITE *et al.*, 1991).

Foi colocada abaixo uma tabela com algumas informações sobre as estruturas de chiquimato quinase já conhecidas que foram depositadas no PDB.

Tabela 1 – Tabela com as estruturas de chiquimato quinases depositadas no PDB.

Organismo	Código de Acesso PDB	Resolução (Å)	Referências
<i>Mycobacterium tuberculosis</i>	1WE2	2,30	PEREIRA <i>et al.</i> , 2004a.
<i>Erwinia chrysanthemi</i>	2SHK	2,60	KRELL <i>et al.</i> , 1997.
<i>Mycobacterium tuberculosis</i>	1U8A	2,15	DHALIWAL <i>et al.</i> , 2004.
<i>Mycobacterium tuberculosis</i>	1L4Y	2,00	GU <i>et al.</i> , 2002.
<i>Mycobacterium tuberculosis</i>	1L4U	1,80	GU <i>et al.</i> , 2002.
<i>Erwinia chrysanthemi</i>	1E6C	1,80	KRELL <i>et al.</i> , 2001.
<i>Erwinia chrysanthemi</i>	1SHK	1,90	KRELL <i>et al.</i> , 1997.
<i>Escherichia coli</i>	1KAG	2,05	ROMANOWSKI <i>et al.</i> , 2002.
<i>Campylobacter jejuni</i>	1VIA	1,57	BADGER <i>et al.</i> , 2005.
<i>Mycobacterium tuberculosis</i>	2DFT	2,80	DIAS <i>et al.</i> , (submeter).

2. OBJETIVOS

O principal objetivo deste trabalho foi obter o modelo tridimensional da chiquimato quinase (CQ) de *M. leprae* e analisar o comportamento da região que forma o LID na CQ, através da aplicação de métodos de modelagem molecular por homologia ou comparativa, tendo vínculo com o projeto Genoma Estrutural que tem por sua vez o objetivo de fornecer a estrutura tridimensional de proteínas, que foram identificadas nos projetos de seqüenciamento de genomas, sendo elas determinadas por técnicas experimentais tais como cristalografia de raios X, ressonância magnética nuclear ou por modelagem molecular.

Levando em consideração este objetivo geral, a realização deste trabalho consiste em modelarmos a chiquimato quinase de *M. leprae* em três complexos diferentes (ADP:MG:CL, ADP:MG:SHI:CL e ADP:MG), que ainda não tem nenhuma estrutura cristalográfica. Tendo a finalidade de propor a compreensão da interação proteína/ligante e fornecer um modelo para desenho de inibidores. Qualquer microrganismo poderá, em princípio, ser combatido por inibidores dessa enzima da via metabólica do ácido chiquímico, que impossibilitarão a produção do ácido corísmico. A utilização de uma das enzimas da via metabólica do ácido chiquímico como alvos para o desenvolvimento de inibidores pode ser plenamente justificada e validada pela inibição específica da 5-Enoilpiruvilchiquimato 3-Fosfato Sintase (EPSPS) pelo glifosato (N-(fosfonometil) glicina), o qual é um herbicida efetivo e específico.

3. METODOLOGIA E FERRAMENTAS

3.1. Modelagem Molecular por Homologia ou Comparativa

O conhecimento da estrutura tridimensional de uma proteína constitui uma informação valiosa para a determinação de sua função, pois com isso pode-se permitir a identificação de domínios, sítios catalíticos, sítios de regulação alostérica dentre outras funções. Apesar de todas as técnicas inovadoras, sobretudo nas áreas de cristalografia de raios X e de ressonância magnética nuclear (RMN), muitos problemas persistem, pois não é fácil a obtenção de amostras em quantidades suficientes para os ensaios necessários e os cristais obtidos nem sempre tem uma boa qualidade para a obtenção de dados de difração de raios X. Além disso, em certas classes de proteínas, como por exemplo, as proteínas de membrana celular, a determinação estrutural é uma desafio, pois essas proteínas raramente cristalizam e dificilmente podem ser tratadas de modo satisfatório por RMN (FILHO SANTOS e DE ALENCASTRO, 2003).

Relacionamentos evolucionários permitem o uso de métodos computacionais para prever a estrutura tridimensional de uma proteína baseada na sua similaridade com estruturas de proteínas conhecidas (ARCURI *et al.*, 2004; PEREIRA *et al.*, 2003; DA SILVEIRA *et al.*, 2004). Entre todas as aproximações teóricas atuais, a modelagem molecular comparativa ou por homologia de estruturas de proteína é o único método computacional que tem o potencial de construir um modelo 3D para uma proteína de estrutura desconhecida (o alvo) baseada em uma ou mais proteínas relacionadas de estruturas conhecidas (os moldes) com precisão e confiança comparáveis a

estruturas experimentais de baixa resolução a partir da seqüência de aminoácidos (TRAMONTANO *et al.*, 2001; BLUNDELL *et al.*, 1987; GREER, 1981; JOHNSON *et al.*, 1994; SALI & BLUNDELL, 1993; SALI, 1995; SANCHEZ & SALI, 1997; MARTI-RENOM *et al.*, 2000; FISER *et al.*, 2002; SANCHEZ & SALI, 2000).

A técnica de modelagem, em geral, é cercada por diversas dificuldades, tais como alinhamentos problemáticos que geram os erros na modelagem comparativa, modelagem de Alças (*loops*) existência de múltiplos conformacionais, a modelagem de sítios de ligação e modelagem das cadeias laterais (MUNIZ, 2003).

O principal fator que influencia a obtenção de uma boa modelagem molecular está correlacionado com a identidade da seqüência em que o modelo está sendo baseado, levando em consideração como corte uma identidade da seqüência menor que 30%, pois abaixo dessa identidade aumentam rapidamente os erros de uma modelagem molecular. Os modelos comparativos têm sido usados em estudos de mecanismos catalíticos de enzimas, desenho e melhoramento de ligantes, *docking* de macromoléculas, *screening* virtuais e *docking* de pequenos ligantes, etc (FISER *et al.*, 2002).

Os métodos atuais de modelagem molecular consistem basicamente em quatro passos: identificação e seleção de proteínas moldes; alinhamento das seqüências primárias; construção das coordenadas atômicas do modelo e validação. A seguir é mostrado um esquema geral do processo de modelagem de proteínas por homologia (Figura 7) (SALI e OVERINGTON, 1994).

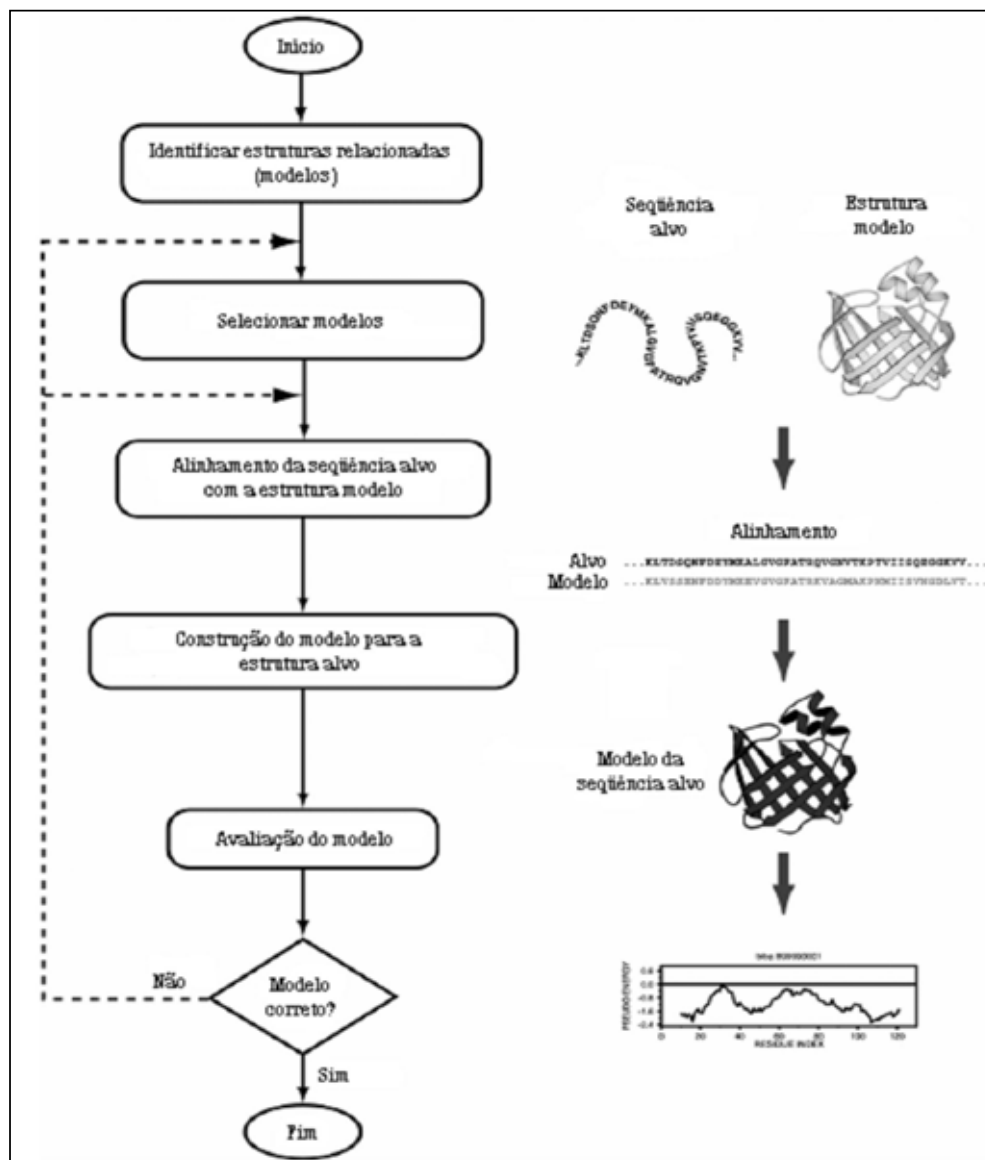


Figura 7 – Esquema dos passos da modelagem molecular comparativa (SALI & BLUNDELL, 1993).

A modelagem molecular, portanto, consiste em vários passos para a obtenção do modelo:

3.1.1. Identificação e seleção de proteínas molde: essa etapa consiste em identificar uma ou mais seqüências de proteínas de estrutura tridimensional conhecida, para servir de molde para a determinação da estrutura tridimensional

da seqüência a ser modelada (proteína alvo). A busca por este molde pode ser feita no PDB (BERNSTEIN et al., 1977; BERMAN et al., 2000) (<http://www.rcsb.org.pdb>) desde que seja conhecida a família da proteína alvo, ou podem ser utilizadas outras ferramentas de busca como o BLAST (ALTSCHUL et al., 1997) ou FASTA (PEARSON, 1990).

3.1.2. Alinhamento: depois de definido o molde, pode-se executar a etapa seguinte que é o alinhamento da seqüência alvo com a seqüência molde. O objetivo do alinhamento é alinhar resíduos estruturalmente equivalentes levando em conta características estruturais comuns, tais como, elementos de estrutura secundária e resíduos catalíticos (FILHO SANTOS & DE ALENCASTRO, 2003). Neste processo, podem ocorrer os chamados “gaps” que são espaços vazios e são representados por linhas tracejadas, correspondem principalmente a regiões de alças. Se tivermos selecionado mais de uma seqüência molde, teremos que alinhar a seqüência alvo com todas as seqüências dos moldes selecionadas, obtendo-se dessa forma um alinhamento múltiplo. Existem vários programas para realizar o alinhamento das seqüências, tais como, BLAST (ALTSCHUL et al., 1997), o FASTA (PEARSON, 1990), o CLUSTAL (THOMPSON, et al., 1994) e o MULTALIN (CORPET, 1998), todos se encontram disponíveis na internet.

3.1.3. Construção do modelo: existem vários métodos para executar-se a construção do modelo tridimensional da proteína alvo. Um deles leva em consideração a modelagem através da montagem de corpo rígido, construindo o modelo através de algumas regiões centrais, alças e cadeias laterais, que são

obtidas ao se dissecar as estruturas relacionadas. Um outro método, utiliza segmentos modelados emparelhando-os e confiando nas posições aproximadas dos átomos conservados nos modelos e calculando as coordenadas dos outros átomos. Já um terceiro método, faz a modelagem satisfazendo as restrições espaciais, usando a distância geométrica ou técnicas de otimização, para satisfazer às restrições espaciais, obtidas do alinhamento da sequência designada com o modelo estrutural. As precisões dos métodos são relativamente semelhantes quando adequadamente usados (FISER e SALI, 2000).

Neste trabalho de modelagem da Chiquimato Quinase de *M. leprae*, foi utilizada uma aproximação automatizada para modelagem comparativa de estrutura de proteína baseada na satisfação da restrição espacial (Figura 8) (SALI & BLUNDELL, 1993; FISER, 2000) que é executada pelo programa MODELLER (<http://salilab.org/modeller>) (SALI & BLUNDELL, 1993; FISER *et al.*, 2002), através do servidor de *web* PARMODEL (UCHÔA *et al.*, 2004)(<http://www.biocristalografia.df.ibilce.unesp.br/bmsystools>).

A modelagem comparativa por satisfação de restrição espacial consiste de três etapas basicamente para construção do modelo (alvo). Na primeira etapa da construção do modelo, as restrições espaciais da sequência alvo são calculadas, a partir do segundo alinhamento gerado. No geral, as restrições são expressas como funções de densidade de probabilidade condicional (pdf) $\mathbf{P}(\mathbf{f}|\mathbf{I})$ para a característica espacial restrita \mathbf{f} , dado diversas variáveis \mathbf{I} que foram encontradas para as características a serem preditas. Há três tipos de restrições, dependem da

origem e da natureza da informação **I** (SALI & BLUNDELL, 1993; FISER *et al.*, 2002).

Primeiramente, as restrições são obtidas para aqueles resíduos do alvo que são alinhados com os resíduos do molde. Esta homologia derivada das restrições limitando as distâncias entre os átomos da cadeia principal e cadeia lateral, bem como os ângulos diedros da cadeia principal (Φ , Ψ e Ω) e os ângulos diedros da cadeia lateral (χ_i), onde estas restrições foram obtidas de uma análise estatística de 105 famílias alinhadas que incluem 416 proteínas definidas estruturalmente (SALI & OVERINGTON, 1994). Por exemplo, uma restrição numa certa distância entre dois $C\alpha$ - $C\alpha$ equivalentes em duas estruturas de proteína relacionadas são bem descritas por uma soma ponderada de duas funções Gaussianas que correspondem às duas distâncias do molde, respectivamente (SALI & BLUNDELL, 1993; FISER *et al.*, 2002).

A segunda classe de restrições reflete as preferências estatísticas extraídas das estruturas de proteínas conhecidas no geral e são relacionadas aos potenciais estatísticos médios de força (FISER *et al.*, 2002). As restrições dependem somente dos tipos restritos de átomos ou resíduos e não da estrutura do molde. Estas restrições são aplicadas à cadeia principal e aos ângulos diedros da cadeia lateral dos resíduos alvos que não são alinhados com os resíduos do molde e para as distâncias entre todos os pares de átomos não ligados. São usados porque foram encontrados para resultar em modelos mais exatos do que os termos correspondentes do campo de força mecânica molecular (FISER *et al.*, 2000; SALI & BLUNDELL, 1993; FISER *et al.*, 2002).

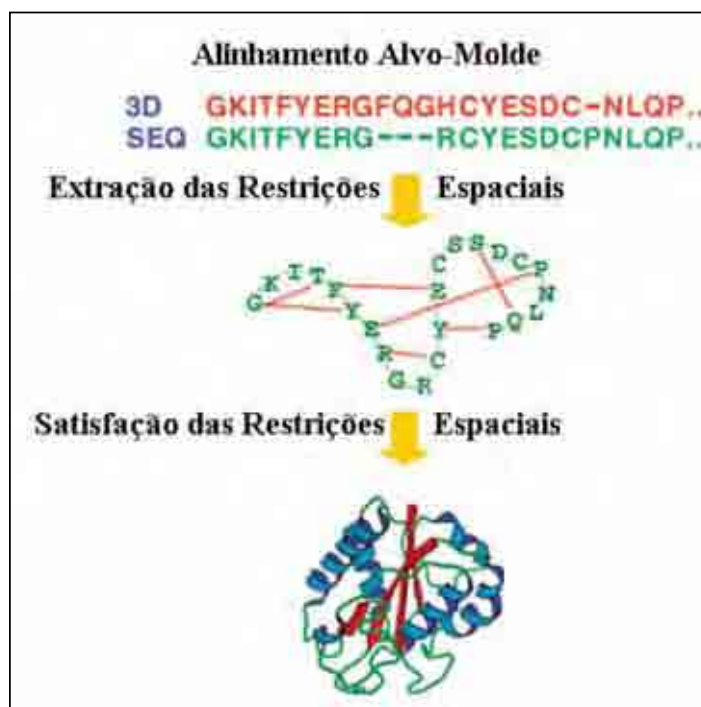


Figura 8 – Modelagem comparativa de estrutura de proteína por satisfação de restrição espacial (FISER *et al.*, 2002). Primeiro, a estrutura 3D conhecida usada como modelo é alinhada com a seqüência (alvo) a ser modelada. Segundo características espaciais são transferidas do modelo para o alvo. Terceiro e último passo, o modelo 3D é obtido por satisfação de todas as restrições que são possíveis.

O terceiro tipo de restrição é obtido a partir do campo de força mecânica molecular do CHARMM-22 (MACKERELL *et al.*, 1998) e incluem restrições de ligações químicas. Estas restrições mecânicas moleculares reforçam estereoquimicamente o modelo correto. Depois que todas as pdfs são calculadas, seus logaritmos são somados para obter uma função objetiva que dependa do modelo e dê sua probabilidade. Finalmente, o modelo que contém todos os átomos excluindo o hidrogênio é calculado otimizando a função objetiva no espaço cartesiano. A otimização é realizada pelo método da função do alvo variável que emprega gradientes conjugados e a dinâmica molecular com *simulated annealing* (SALI & BLUNDELL, 1993; FISER *et al.*, 2002).

A saída é um modelo 3D para a seqüência alvo contendo todas as cadeias principais e cadeias laterais dos átomos sem o hidrogênio. Além da construção do modelo, o MODELLER (SALI & BLUNDELL, 1993) executa tarefas auxiliares adicionais como: incluir um alinhamento de duas seqüências de proteína ou de seus perfis, alinhamento múltiplo de seqüências e/ou de estruturas de proteína, cálculo de árvores filogenéticas, e a modelagem de alças nas estruturas de proteína (JACOBSON & SALI, 2004). O programa é usado com uma linguagem de *script* e não têm gráficos. É escrito na linguagem de programação Fortran90 e funciona em computadores com o sistema operacional Unix (SALI & BLUNDELL, 1993).

PARMODEL (UCHÔA *et al.*, 2004) é um *pipeline* e um servidor *web* para modelagem comparativa automatizada de estruturas de proteínas 3D. O PARMODEL tem como objetivo executar a modelagem, avaliação, visualização e otimização dos modelos de proteínas, conseqüentemente é subdividido em quatro módulos: Modelagem Parmodel, Avaliação Parmodel, Visualização Parmodel, e Otimização Parmodel. O PARMODEL automatiza e integra os principais programas usados na modelagem comparativa como: MODELLER (SALI & BLUNDELL, 1993), PROCHECK (LASKOWSKI *et al.*, 1993), Raster3D (BACON & ANDERSON, 1988; MERRITT, 1994), MOLSCRIPT (KRAULIS, 1991) e GROMACS (VAN GUNSTEREN *et al.*, 1996; BERENDSEN *et al.*, 1995; LINDAHL *et al.*, 2001).

PARMODEL executa a paralelização do MODELLER em um *cluster Beowulf* de 16 PCs rodando em Linux. O usuário pode construir vários modelos

para uma mesma proteína (até 1000) e pode aumentar a possibilidade de obter um bom modelo, porque as mudanças nas restrições espaciais iniciais gerarão diferentes conformações na estrutura final.

O módulo de modelagem do PARMODEL requer algumas informações e arquivos na interface de entrada de dados para começar a modelagem dos modelos, como: (i) a seqüência de aminoácido da proteína (alvo), (ii) até oito estruturas de proteínas (moldes), (iii) o número de modelos 3D (alvo) a ser modelado, (iv) escolher o método experimental dos modelos, (v) escolher o número de ligantes do modelo (alvo) ou se não vai ser modelado com ligantes, (vi) escolher que programas serão usados para avaliar os modelos: PROCHECK (LASKOWSKI *et al.*, 1993) e 3DAnalysis (LUTHY *et al.*, 1992; BOWIE *et al.*, 1991) e (vii) escolher como os modelos finais serão classificados, pela função objetiva do MODELLER ou pela qualidade estereoquímica. Depois de entrar com as informações é só submeter os dados de entrada e aguardar o retorno do melhor modelo gerado. O módulo de modelagem do PARMODEL retorna os resultados da modelagem por e-mail, que contém um arquivo comprimido com o melhor modelo da estrutura de proteína (alvo), o resultado do alinhamento, uma lista com a função objetiva do MODELLER ou da qualidade estereoquímica do modelo para todos os modelos construídos e os resultados da análise para o melhor modelo.

3.1.4. Avaliação do Modelo:

A validação do modelo é uma etapa essencial que pode ser executada em diferentes níveis de organização estrutural. O modelo pode ser avaliado como um

todo ou por regiões. Como a qualidade do modelo depende muito da estrutura molde utilizada, esta última deve ter preferencialmente uma resolução acima de 2,0Å e com um fator-R inferior a 20%, pois com isso o modelo terá qualidades proporcionais a da estrutura 3D utilizada como molde (SANCHEZ AND SALI, 1998). Primeiramente, precisamos notar se existem grandes diferenças conformacionais que não podem ser explicadas entre os elementos de estrutura secundária das estruturas molde e da estrutura modelada (FILHO SANTOS e DE ALENCASTRO, 2003).

A qualidade estereoquímica do modelo é de fundamental importância, sendo o PROCHECK (LASKOWSKI, et al., 1993) o programa mais utilizado na avaliação dos parâmetros estereoquímicos. Ele avalia os comprimentos de ligação, os ângulos planos, a planaridade dos anéis de cadeias laterais, a quiralidade, as conformações das cadeias laterais, a planaridade das ligações peptídicas, os ângulos torcionais da cadeia principal e das cadeias laterais, o impedimento estérico entre pares de átomos não-ligados e a qualidade do gráfico de Ramachandran (RAMAKRISHNAN e RAMACHANDRAN, 1965). O gráfico de Ramachandran é útil porque ele define os resíduos que estão nas regiões mais favoráveis e nas regiões desfavoráveis e orienta a avaliação da qualidade dos modelos teóricos ou experimentais de proteínas (RAMAKRISHNAN e RAMACHANDRAN, 1965). Também foram usados os seguintes programas: WHATCHECK/WHATIF (HOOFT *et al.*, 1996; VRIEND, 1990), VERIFY-3D (BOWIE *et al.*, 1991; LUTHY *et al.*, 1992; KABSCH & SANDER, 1983), programa X-PLOR (SCHWIETERS *et al.*, 2003; BRÜNGER, 1992), o LSQKAB

do pacote CCP4 (Collaborative Computational Project N° 4., 1994) para medir a distância entre os átomos de $C\alpha$ - $C\alpha$ sobrepostos (KABASH & SANDER, 1983), o PROCHECK (LASKOWSKI *et al.*, 1993) com o G-factor. Na avaliação da interação entre enzima-ligante foram verificadas as ligações intermoleculares de hidrogênio.

Para obter a análise do programa VERIFY-3D (BOWIE *et al.*, 1991; LUTHY *et al.*, 1992; KABSCH & SANDER, 1983), e o diagrama de Ramachandran foi utilizado o módulo de avaliação do programa PARMODEL (UCHÔA *et al.*, 2004). O módulo de avaliação PARMODEL permite que o usuário avalie os modelos gerados por modelagem, difração de raios X e RMN. O módulo requer como entrada apenas um arquivo das coordenadas atômicas no formato “pdb” da proteína que será avaliada, os resultados são mostrados na mesma página da web. As saídas geradas podem ser salvas em arquivos com extensão “pdf” para o PROCHECK. Para o VERIFY-3D os valores de referência são mostrados em um gráfico.

3.2. Métodos de Avaliação

3.2.1. PROCHECK

A análise da geometria global da estrutura ou de cada resíduo individualmente é feita pelo PROCHECK (LASKOWSKI *et al.*, 1993), utilizando para isso parâmetros estereoquímicos derivados de estruturas de alta resolução ou bem refinadas (MORRIS *et al.*, 1992) que constituem sua base de dados para verificação da qualidade da estrutura 3D da proteína. Os parâmetros estereoquímicos usados são aqueles descritos com detalhe em MORRIS *et al.* (1992). As verificações também fazem uso do comprimento e ângulo de ligação ideal, derivados de uma análise detalhada (ENGH & HUBER, 1991) de estruturas de pequenas moléculas em Cambridge Structural Database (CSD) (ALLEN *et al.*, 1979). Estes parâmetros utilizados como informações estereoquímicas checados pelo PROCHECK são: ligações covalentes, planaridade de grupos planares (aromáticos, ligações peptídicas, etc.) ângulos diedros, quiralidade, interações não covalentes, ligações de hidrogênio da cadeia principal e pontes de dissulfeto.

O PROCHECK requer como entrada um arquivo das coordenadas atômicas da estrutura da proteína a ser avaliada no formato “pdb” e produz representações coloridas facilmente interpretadas, descrevendo a estrutura de uma proteína, e também comparando duas estruturas de proteínas relacionadas, juntamente com uma lista detalhada de resíduo por resíduo. Dentre as várias análises fornecidas pelo PROCHECK foram utilizadas neste trabalho: o Diagrama de Ramachandran e o G-factor, que são descritos a seguir.

3.2.2. Diagrama de Ramachandran

As saídas do PROCHECK compreendem um número de gráficos com uma detalhada listagem resíduo a resíduo. Dentre esses gráficos está o gráfico de Ramachandran que exibe os ângulos de torção ϕ e Ψ para todos os resíduos da estrutura.

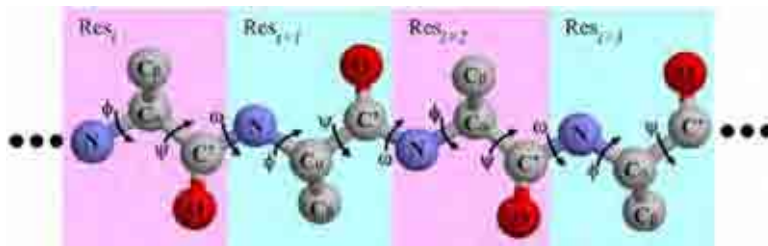


Figura 9 – Parte de uma cadeia polipeptídica indicando os graus de liberdade torcional de cada unidade peptídica e os ângulos de rotação ϕ e Ψ .

Os resíduos de glicina são identificados por triângulos. A coloração do gráfico permite delimitar as regiões (MORRIS et al., 1992). Regiões em vermelho correspondem as combinações mais favoráveis dos valores de phi-psi (figura 10). O ideal é que pelo menos 90% dos resíduos estejam nestas regiões mais favoráveis. A porcentagem dos resíduos nestas regiões é um dos melhores guias para avaliar a qualidade estereoquímica (LASKOWSKI et al., 1993).

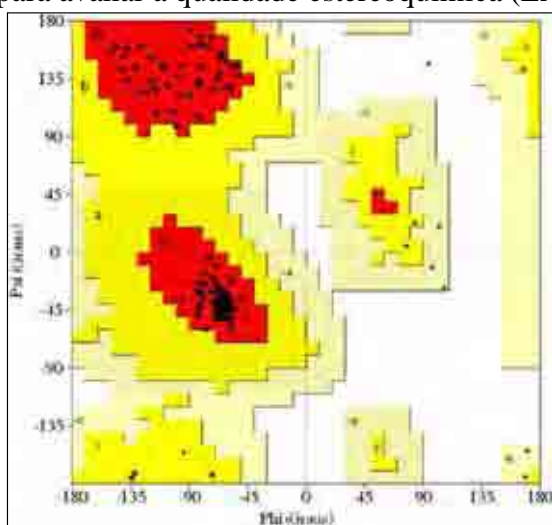


Figura 10 – Diagrama de Ramachandran. A região permitida em vermelho, a região adicionalmente permitida em amarelo escuro, a região generosamente permitida em amarelo claro e a não permitida em branco. E as regiões onde estão as hélices- α (A e L) e folhas- β (B).

Os ângulos ϕ (phi) e Ψ (psi), são utilizados para a designação de rotações em torno das ligações C-N e C $_{\alpha}$ -C, respectivamente. A conformação do esqueleto de uma proteína pode ser descrita pela especificação dos valores de ϕ e Ψ para cada resíduo. Em proteínas, freqüentemente ocorrem dois tipos de estrutura secundária, que são as estruturas de hélices α e de folhas β pregueadas, mantidas por ligações de hidrogênio. As hélices α e as folhas β não são as únicas estruturas secundárias possíveis mas, elas são, sem dúvida, as mais importantes (RAMAKRISHNAN e RAMACHANDRAN, 1965).

3.2.3. G-factor

O G-Factor fornece uma medida de uma propriedade estereoquímica, se é dada como normal ou não, sendo que o valor ideal do G-factor deve ser acima de -0,5; no intervalo de -0,5 a -1,0 existem resíduos na proteína que deverão ser investigados e, abaixo de -1,0, os modelos apresentam qualidade ruim e deverão ser rejeitados (MORRIS *et al.*, 1986). Os valores gerais são obtidos a partir de uma média de todos os G-Fatores diferentes para cada resíduo da estrutura (LASKOWSKI *et al.*, 1993). Valores de Fator G são computados para as seguintes propriedades (LASKOWSKI *et al.*, 1993):

Ângulos de Torção: combinações de phi-psi; combinações de chi1-chi2; torções de chi1 para aqueles resíduos que não possuem um chi2; Ângulos de Torção combinados chi3 e chi4 e Ângulos de Torção Omega.

Geometria Covalente: calcula comprimentos de ligações entre cadeias principais e Ângulos de ligações entre cadeias principais.

3.2.4.VERIFY-3D

O VERIFY-3D (BOWIE *et al.*, 1991; LUTHY *et al.*, 1992; KABSCH & SANDER, 1983) mede a compatibilidade entre a seqüência de aminoácidos de uma proteína e o modelo da sua estrutura tridimensional, usando um perfil 3D. Isto é feito reduzindo a estrutura tridimensional a uma representação unidimensional simplificada chamada cadeia de ambiente, que então é comparada com a seqüência de aminoácidos unidimensional. O VERIFY-3D também pode ser usado para verificar a validade de uma estrutura de proteína obtida através de modelagem molecular medindo a compatibilidade daquela estrutura com a própria seqüência da proteína. Apesar da aplicação, o método implica três operações básicas:

- Redução da estrutura tridimensional a uma cadeia unidimensional de ambientes de resíduos. Esses ambientes são categorizados segundo a área de resíduo que está na parte interna da proteína, a fração da área da cadeia lateral que está exposta às moléculas polares e a estrutura secundária local.
- Geração de uma matriz de comparação dependente da posição conhecida com perfil 3D. É calculada através da cadeia de ambientes de resíduo, penalidade de “gaps” e por uma matriz de classificação pré-calculada. A matriz de classificação é determinada através da probabilidade de encontrar cada um dos 20 aminoácidos em cada uma das

classes de ambiente como observado em um banco de dados de estruturas conhecidas e seqüências relacionadas.

- Alinhamento de uma seqüência com o perfil 3D. O resultado do escore de alinhamento é a medida da compatibilidade da seqüência com a estrutura descrita pelo perfil 3D.

Para verificação da estrutura é necessário compensar os efeitos de comprimento e também comparar com o escore obtido de estruturas conhecidas.

Lüthy *et al.* (1992) determinaram empiricamente o índice global esperado de $S_{\text{calc}} = \exp(-0,83 + 1,008 \times \ln(L))$ de compatibilidade entre a seqüência e a estrutura 3D onde L é o comprimento da seqüência e S_{calc} representando, então, a soma das probabilidades individuais dos resíduos. Lüthy *et al.* (1992) também sugere um limite de $0,45 \times S_{\text{calc}}$ para a confiabilidade da compatibilidade seqüência/estrutura. Valores menores que $0,45 \times S_{\text{calc}}$ indicam uma estrutura incorreta, valores em torno de $0,45 \times S_{\text{calc}}$ podem estar corretos, mas possuem qualidade questionável e valores acima de $0,45 \times S_{\text{calc}}$ indicam uma estrutura correta. O VERIFY-3D requer como entrada um arquivo de coordenadas atômicas no formato pdb.

3.2.5. WHATIF/WHATCHECK

O WHATIF (VRIEND, 1990) pode calcular quase todas as propriedades de proteínas, desde a área de superfície acessível ao solvente até valores de pKa para contato com a dinâmica molecular com GROMOS (VAN GUNSTEREN *et al.*, 1987).

O WHATCHECK (HOOFT *et al.*, 1996) fornece acesso a um subconjunto de verificações de qualidade estrutural WHATIF. WHATCHECK dá informações sobre a formação de regiões centrais hidrofóbicas, a acessibilidade de resíduos e átomos com as moléculas do solvente (água), a distribuição espacial de grupos iônicos, a distribuição das distâncias atômicas e das ligações de hidrogênio da cadeia principal. Ou seja, retrata a estereoquímica, comprimentos de ligações, ângulos e diedros, entre outras quantidades na forma de um relatório gerado no formato pdf.

3.2.6. RMSD (Desvio Médio Quadrático)

O RMSD pode ser calculado de duas maneiras: (i) uma função de todos os átomos de uma proteína (ENGH & HUBER, 1991) e (ii) uma função de algum subconjunto dos átomos, como a cadeia principal da proteína ou posições de C_{α} apenas (KABASH & SANDER, 1983). O parâmetro mais comum que expressa a diferença entre duas estruturas protéicas é o RMSD, ou desvio médio da raiz quadrada, em posições atômicas entre as duas estruturas.

Quando o RMSD é calculado como uma função de todos os átomos de uma proteína, podemos adotar a definição de que o RMSD serve para analisar a medida de quanto à posição de cada átomo de uma estrutura varia com relação a um vetor aceito, e é denominado como RMSD da geometria ideal (KEARSLEY, 1989). O RMSD da geometria ideal utiliza como parâmetros o comprimento e ângulo de ligação ideal, derivados de uma análise detalhada (ENGH & HUBER, 1991) de estruturas de pequenas moléculas em Cambridge Structural Database

(CSD) (ALLEN *et al.*, 1979). O programa utilizado para executar os cálculos de comprimento e ângulos de ligação foi o programa X-PLOR (SCHWIETERS *et al.*, 2003; BRÜNGER, 1992).

Quando o RMSD é calculado como uma função de algum subconjunto de átomos é comum a utilização de um subconjunto de átomos da proteína usada como template para quando duas estruturas protéicas são comparadas; para o caso quando as proteínas não são idênticas entre si em seqüência e, temos que, os únicos átomos entre comparação de posição (um-a-um) que podem ser efetuados são os átomos da cadeia principal (C_{α} - C_{α}). Neste contexto abordado a orientação de uma estrutura molecular se torna importante, uma vez que as estruturas de proteínas são geralmente descritas em coordenadas atômicas cartesianas, e surgem como uma orientação incorporada em relação ao espaço. O RMSD é uma função da distância entre os átomos em uma estrutura e os mesmos átomos em outra estrutura. Portanto, se uma molécula começar em uma posição diferente do sistema de coordenadas de referência da outra molécula, o RMSD entre as duas proteínas será calculado independentemente. Para computar RMSDs significativos, as duas estruturas em consideração devem primeiro ser superpostas, desde que possível. A sobreposição das estruturas de proteínas começa geralmente com uma comparação de seqüências. A comparação de seqüências define as relações um-a-um entre os pares de átomos onde o RMSD é computado. As relações átomo-a-átomo, para fins de comparação de estruturas, podem ocorrer na verdade entre resíduos que não estão na mesma posição relativa na seqüência de aminoácidos. As inserções e deleções da seqüência

podem forçar duas seqüências a ficarem sem registro entre si, enquanto a arquitetura central das duas estruturas permanece similar.

Uma vez definida as relações átomo-a-átomo entre duas estruturas, com um programa de sobreposição como o LSQKAB do pacote CCP4 (Collaborative Computational Project N° 4., 1994) alcançamos uma sobreposição ótima entre as duas estruturas, isto é, a sobreposição com o menor RMSD possível. A sobreposição de um par de átomos pode perfeitamente deixar outro par de átomos à parte. Os algoritmos de sobreposição otimizam a orientação e a posição espacial de duas moléculas (translação e rotação) entre si. Uma vez efetuadas as superposições ótimas de todos os pares de estruturas, os valores de RMSD que são computados como resultado podem ser comparados entre si, já que as estruturas foram removidas para a mesma estrutura de referência antes de fazer os cálculos de RMSD (GIBAS & JAMBECK, 2001).

Para visualização dos elementos de estrutura secundária das proteínas modeladas foi usado o programa MOLMOL (KORADI *et al.*, 1996) e para a análise de interações proteína-substrato foi usado o programa XtalView (MCREE, 1999).

Essas foram às ferramentas utilizadas para auxiliar nas análises quantitativas e qualitativas da proteína em estudo, possibilitando a avaliação e a validação dos modelos gerados através do método de modelagem molecular comparativa.

4. RESULTADOS E DISCUSSÃO

Foram construídos modelos de CQ de *M. leprae* em complexo com ADP:Mg:Cl, ADP:Mg:SHI:Cl e ADP:Mg, para cada modelo foi utilizado um molde diferente (1L4Y, 1WE2 e 2DFT) sendo todos os 03 moldes do mesmo microorganismo *Mycobacterium tuberculosis*. Foi gerado um total de 1000 modelos de CQ de *M. leprae* para cada molde, sendo escolhido o modelo de melhor qualidade estereoquímica utilizando o PROCHECK (LASKOWSKI *et al.*, 1993) e a função objetiva do MODELLER (SALI & BLUNDELL, 1993).

4.1. Comparação da Seqüência Primária

A porcentagem de identidade da seqüência primária da CQ de *M. leprae* contra as seqüências primárias dos moldes indica que essas estruturas 3D utilizadas são bons modelos para usarmos como molde na modelagem da CQ de *M. leprae* (alvo), pois apresentam identidade maior que 30%. O valor da identidade entre as seqüências é maior que 77%, o que justifica a escolha dos moldes. Como as três seqüências primárias dos moldes utilizados são do mesmo microorganismo, o alinhamento será o mesmo, então servirá como referência para os três modelos gerados (Figura 11).

O alinhamento entre as seqüências primárias da CQ de *M. leprae* (alvo) e a CQ de *M. tuberculosis* (molde – 1WE2) é mostrado na figura 11. Esse alinhamento foi executado pelo MODELLER o qual é baseado no algoritmo de programação dinâmica, proposto por Needleman e Wunsch para alinhamento de seqüências (NEEDLEMAN & WUNSCH, 1970). Na figura de alinhamento, os

espaços (gaps) são representados por linhas tracejadas, os resíduos idênticos são representados por um * e os resíduos não idênticos são representados por espaços vazios.

```

          10      20      30      40      50      60
1WE2  -APKAVLVGLPGSGKSTIGRRLLAKALGVLLDVAIEQRTGRSIADIFATDGEQEFRI
lmSK  MAPKAVLVGLPGAGKSTIGRRLLSKALGVSLDAAIEKQTGRSIADIFAIDGEEEFRI
      ***** ***** ***** ***** *** ***** ***** *** *****

          70      80      90      100     110     120
1WE2  EEDVVRAALADHDGVLSLGGAVTSPGVRAALAGHTVVYLEISAAEGVRRRTGGNGVRPLL
lmSK  EEGVVRAALVEHDGVVSLGGAVTSPGVCAALAGHIVIYLEINAEEMRRACGSTVRPLL
      ** ***** ***** ***** * ***** * * ** * *****

          130     140     150     160
1WE2  AGPDRAEKYRALMAKRAPLYRRVATMRVDTNRRNPGAVVRHILSRL-
lmSK  AGPDRAEKFQDLMARRVPLYRRVATIRVDTNCHNLGAVVRYIMARLQ
      ***** ** * ***** ***** * ***** * **

```

Figura 11 – Resultado do alinhamento das seqüências primárias da CQ de *M. leprae* (mlSK) e a CQ de *M. tuberculosis* (molde - 1WE2).

4.2. Qualidade do Modelo

Avaliar a qualidade do modelo é fundamental. A qualidade do modelo construído será proporcional à qualidade da estrutura 3D utilizada como molde (SANCHEZ *et al.*, 1998), por isso deve-se avaliar o modelo como um todo e por região. Os modelos (alvos) foram construídos baseados nas coordenadas atômicas das estruturas resolvidas por cristalografia; 1WE2 (CQ:ADP:SHI:Cl:Mg) com resolução de 2,3 Å (PEREIRA *et al.*, 2004a; PEREIRA *et al.*, 2004b); 1L4Y(CQ:ADP:Cl:Mg) com resolução de 2,0 Å (Gu *et al.*, 2002) e 2DFT (CQ:ADP:Mg) com resolução 2.80 Å (DIAS *et al.*, submetido). As coordenadas atômicas de todas as moléculas de água foram retiradas dos moldes.

Após a etapa inicial de escolha do melhor modelo, análises mais criteriosas e detalhadas foram efetuadas com a utilização do VERIFY 3D,

WHATCHECK, RMSD da geometria ideal, RMSD da sobreposição $C\alpha - C\alpha$ e PROCHECK, a fim de verificar a qualidade do modelo construído. Os resultados obtidos do modelo final foram comparados com o valor da estrutura 3D (molde).

O programa PROCHECK foi utilizado para analisar a geometria global do modelo através do diagrama de Ramachandran (Tabela 2 e Figura 12 A-F). A análise dos diagramas de Ramachandran dos modelos da CQ de *M. leprae* (alvo) mostrou que mais de 97% dos resíduos estão sobre a região mais favorável e para a CQ de *M. tuberculosis* (moldes) mostrou que mais de 88% dos resíduos estão sobre a região mais favorável.

Tabela 2 – Resultado das análises do Diagrama de Ramachandran dos moldes e dos modelos de CQ de *M. leprae*
* Entre os parênteses estão os valores obtidos para as estruturas 3D usadas como moldes.

Modelos	Região do Diagrama de Ramachandran (%)			
	Mais Favorável	Adicionalmente Permitida	Generosamente Permitida	Não Permitida
CQ:ADP:CL:MG	99,3 (90,5)	0,0 (8,8)	0,7 (0,7)	0,0 (0,0)
CQ:ADP:SHI:CL:MG	97,2 (93,4)	2,1 (5,9)	0,7 (0,7)	0,0 (0,0)
CQ:ADP:MG	97,2 (88,4)	2,8 (10,9)	0,0 (0,8)	0,0 (0,0)

Nas figuras 12 A-F a região em vermelho é a mais favorável, a região amarela escuro é a adicionalmente permitida, a região amarela claro é a generosamente permitida e a região em branco é a região não permitida. Os resíduos de glicina são representados por triângulos.

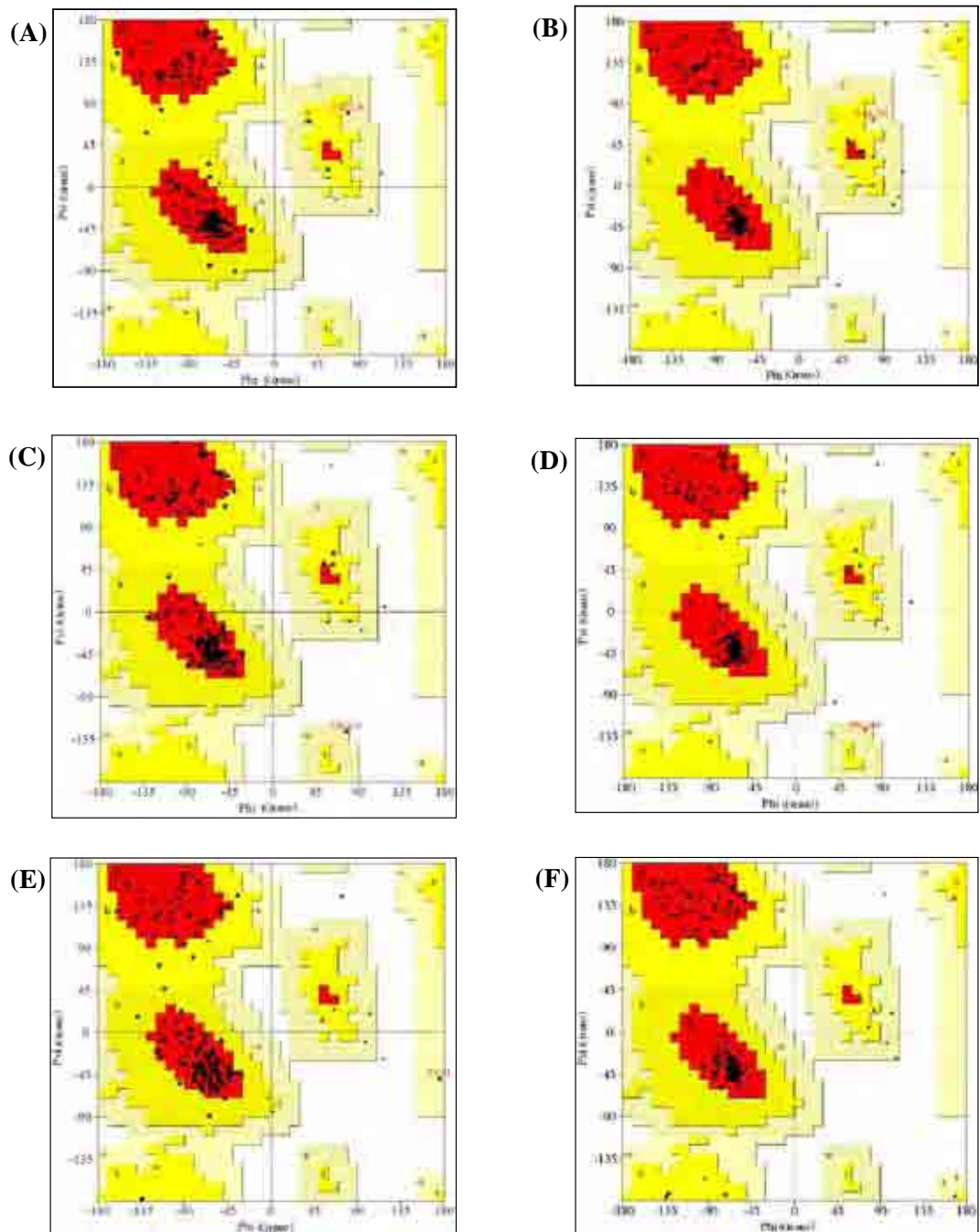


Figura 12 – Diagrama de Ramachandran: (A) 1L4Y; (B) Modelo gerado pelo 1L4Y; (C) 1WE2; (D) Modelo gerado pelo 1WE2; (E) 2DFT e (F) Modelo gerado pelo 2DFT.

Como podemos observar os modelos apresentaram uma qualidade estereoquímica ligeiramente melhor do que a dos moldes utilizados.

Através do programa WHATCHECK foi constatado que os ambientes químicos de resíduos individuais se enquadram nos padrões frequentemente encontrados em estruturas protéicas. Segundo os valores de referência do programa para modelagem molecular, o modelo final pode ser considerado de boa qualidade.

O G-factor global é uma medida da normalidade da estrutura como um todo. Representa uma média para o G-factor de cada resíduo na estrutura. O G-factor médio dos ângulos de torção e geometria covalente representa uma medida do desvio de uma dada propriedade estereoquímica. O resultado dos valores do G-factor, para os ângulos de torção, geometria covalente e para o G-factor global estão colocados na tabela 3.

Tabela 3 – Resultado dos valores do G-factor dos moldes e dos modelos de CQ de *M. leprae*
*Entre parênteses estão os valores obtidos para as estruturas 3D usadas como moldes.

Modelos	G-factor		
	Ângulos Torção	Geometria Covalente	Global
CQ:ADP:CL:MG	0,06 (0,3)	-0,52 (0,57)	-0,16 (0,41)
CQ:ADP:SHI:CL:MG	0,0 (0,08)	-0,24 (0,23)	-0,08 (0,15)
CQ:ADP:MG	0,02 (-0,72)	-0,25 (-1,04)	-0,08 (-0,81)

Para verificar a confiabilidade na compatibilidade da seqüência/estrutura foi utilizado o programa VERIFY 3D. O valor de Escore Total, o valor de Escore Ideal e o Escore S_{calc} final são mostrados na tabela 4. Os valores obtidos indicam que a estrutura do modelo final tem compatibilidade entre a seqüência primária do modelo e a estrutura 3D construída, pois o valor gerado para o modelo final ficou acima do limite de 0,45.

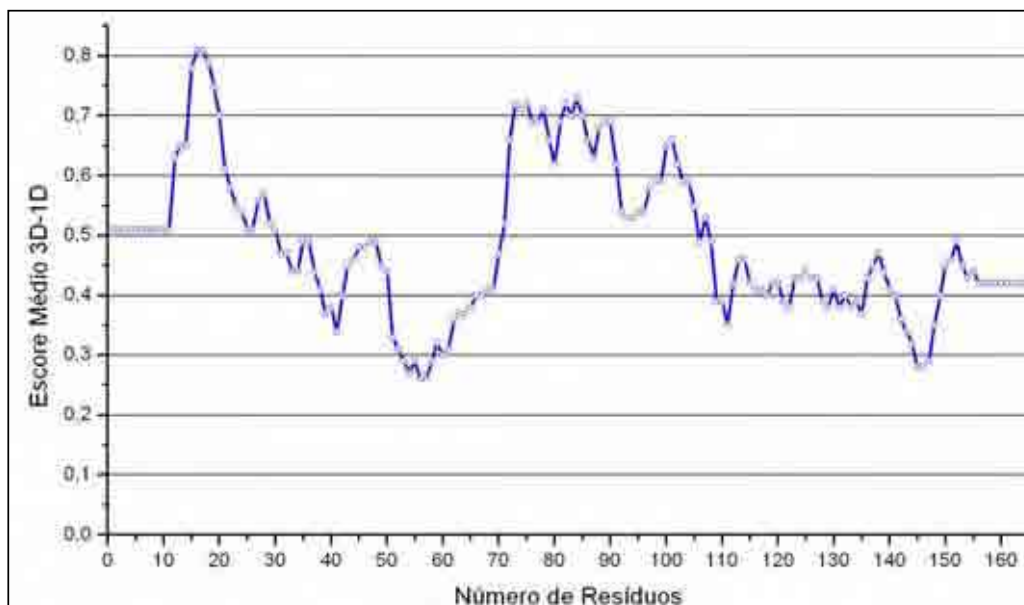
Tabela 4 – Resultado dos valores do VERIFY 3D dos moldes e dos modelos de CQ de *M. leprae*

*Entre parênteses estão os valores obtidos para as estruturas 3D usadas como moldes.

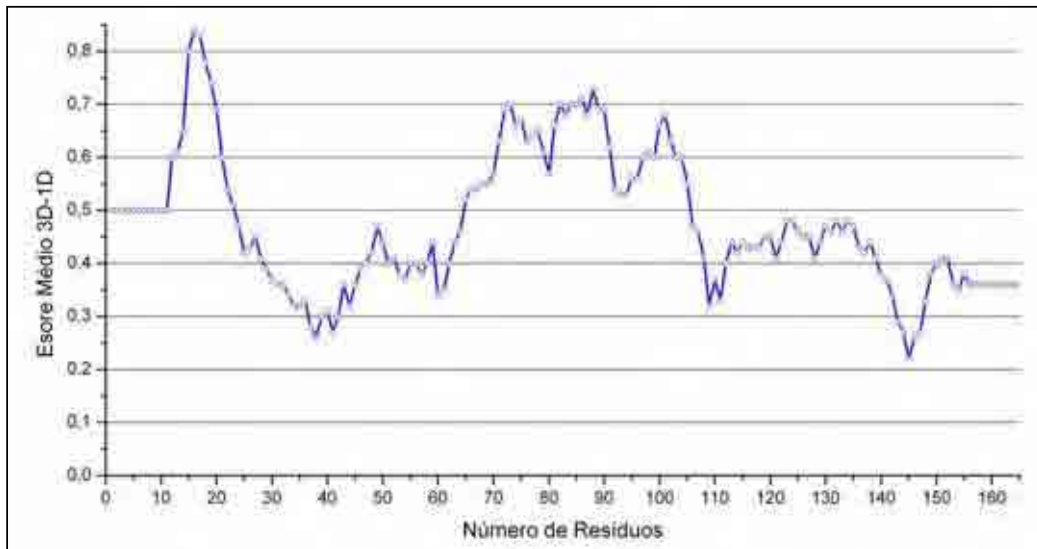
Modelos	VERIFY 3D		
	Escore Total	Escore Ideal	Escore S_{calc}
CQ:ADP:CL:MG	77,97(82,02)	75,41 (74,95)	1,03S (1,09S)
CQ:ADP:SHI:CL:MG	76,99 (82,79)	75,41 (74,95)	1,02S (1,10S)
CQ:ADP:MG	74,85 (81,09)	75,41 (70,37)	0,99S (1,15S)

Foi verificado através do gráfico de X versus Y, obtido pelo programa VERIF 3D, que nenhum resíduo se encontra abaixo da linha zero, que seria uma região onde os aminoácidos se encontrariam em uma conformação desfavorável.

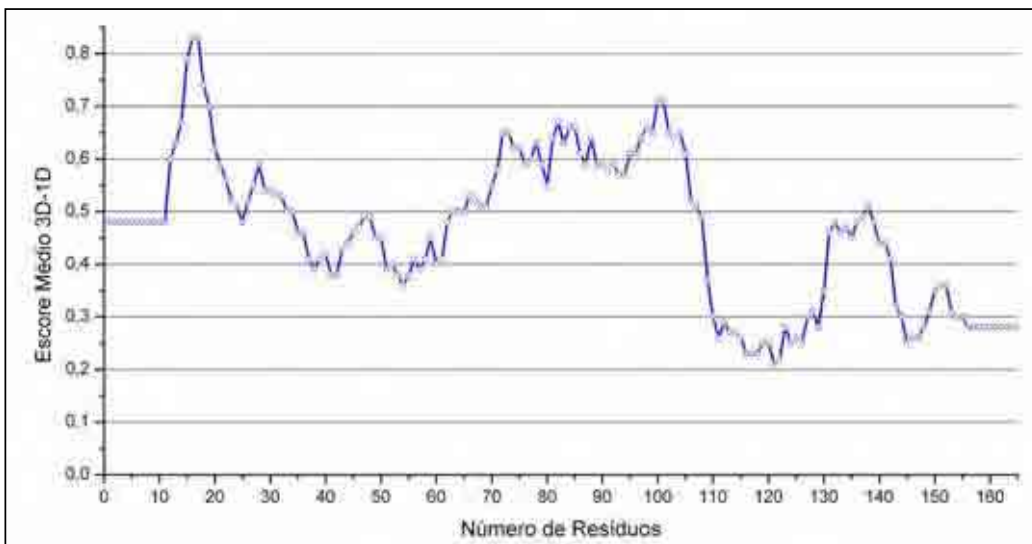
Esse resultado é mostrado para os modelos nas figuras 13 A-C.



(A)



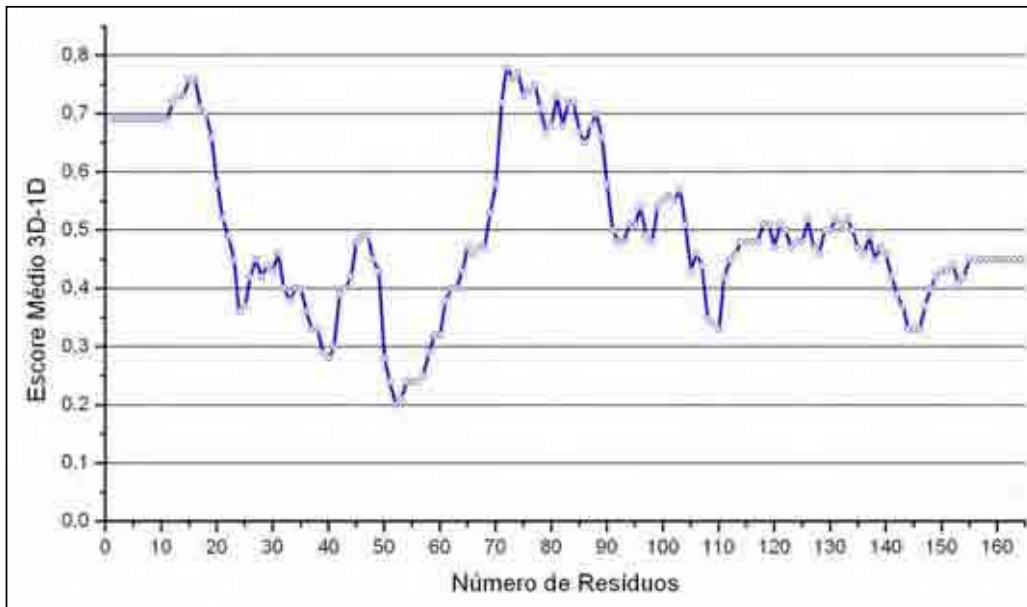
(B)



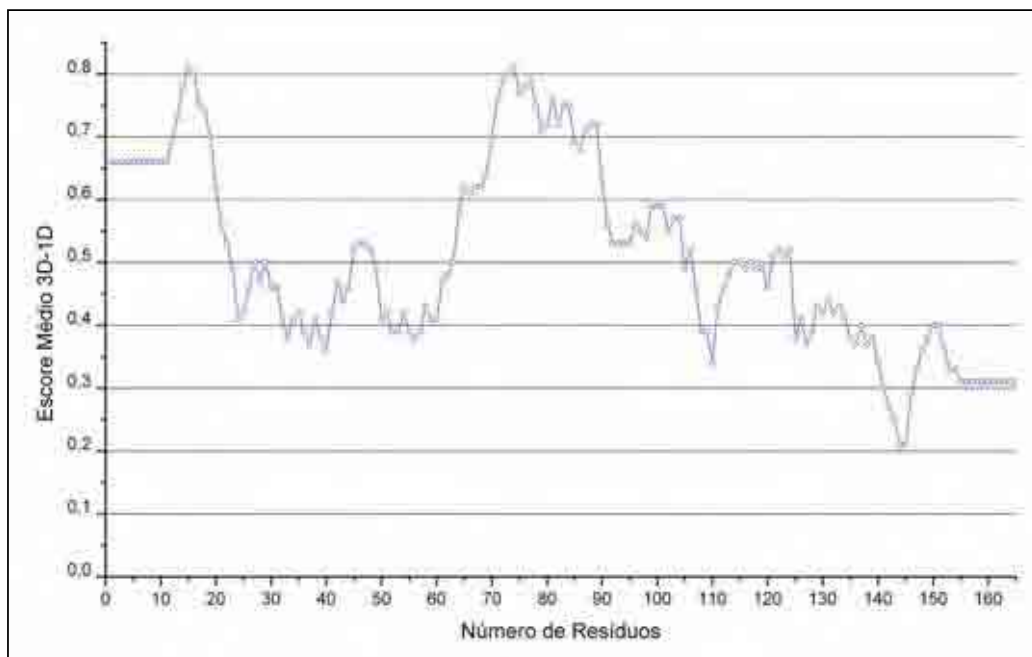
(C)

Figura 13 – Gráfico do programa VERIFY 3D dos modelos finais da CQ de *M. leprae*: (A) Modelo gerado pelo 1L4Y; (B) Modelo gerado pelo 1WE2 e (C) Modelo gerado pelo 2DFT.

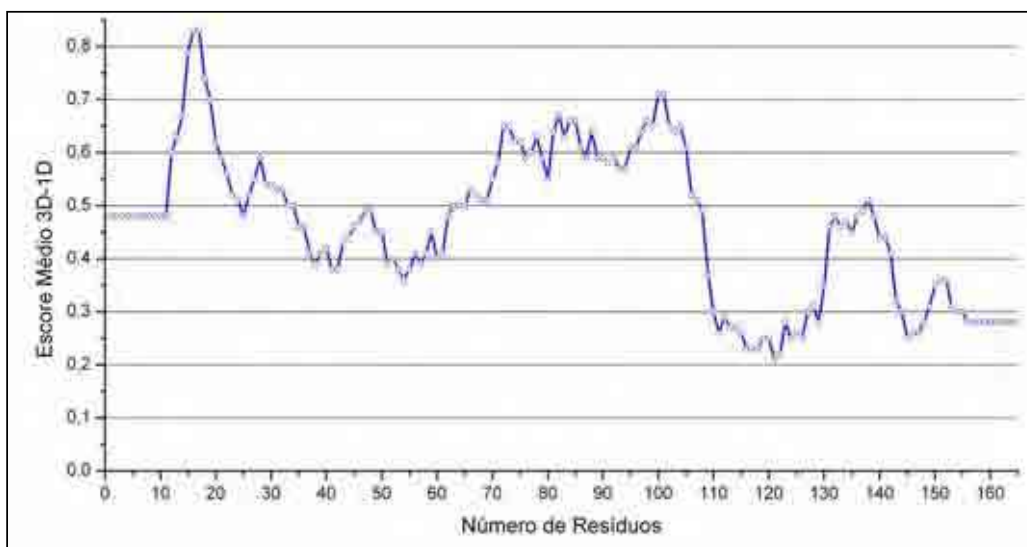
As figuras 14 A-C mostram os resultados encontrados para as estruturas cristalográficas utilizadas como molde.



(A)



(B)



(C)

Figura 14 – Gráfico do programa VERIFY 3D dos moldes: (A) 1L4Y; (B) 1WE2 e (C) 2DFT.

O valor do RMSD da geometria ideal (Tabela 5) para o comprimento e ângulo de ligação foi calculado através do programa X-PLOR (SCHWIETERS *et al.*, 2003; BRÜNGER, 1992), que gera um arquivo com os valores do RMSD para as ligações e ângulos diedros e impróprios bem como contatos não ligados, os valores ideais para o comprimento e ângulo de ligação devem estar abaixo 0,1 Å e 10,0 graus respectivamente, para que a estrutura da proteína como um todo esteja aceitável.

Tabela 5 – Resultado dos valores do RSMD da geometria ideal dos moldes e dos modelos de CQ de *M. leprae*
*Entre parênteses estão os valores obtidos para as estruturas 3D usadas como moldes.

Modelos	RMSD Geometria Ideal	
	Comprimento de Ligação (Å)	Ângulo Ligação (°)
CQ:ADP:CL:MG	0,023 (0,005)	4,409 (1,10)
CQ:ADP:SHI:CL:MG	0,020 (0,017)	3,893 (1,90)
CQ:ADP:MG	0,020	3,849

No RMSD da sobreposição foi considerada somente a sobreposição C_α – C_α com o auxílio do programa LSQKAB do pacote CCP4 (Collaborative Computational Project N° 4., 1994). Com a sobreposição é possível verificar se há possíveis alterações no posicionamento das cadeias laterais dos resíduos ou se há alguma alteração na conformação da estrutura da proteína, além de determinar o RMSD da sobreposição (Tabela 6) . A sobreposição é feita entre o modelo e o molde utilizando as coordenadas atômicas de cada um.

Tabela 6 – Resultado dos valores do RMSD da sobreposição ideal dos moldes e dos modelos de CQ de *M. leprae*

Modelos	RMSD sobreposição (Å)
CQ:ADP:Cl:Mg	0,244
CQ:ADP:SHI:Cl:Mg	0,055
CQ:ADP:Mg	0,083

O LSQKAB gera também um arquivo com as coordenadas atômicas da sobreposição. Este arquivo de coordenadas atômicas foi utilizado como entrada no programa XtalView (McREE, 1999) para visualizar e analisar estas estruturas (modelos gerados) com base em sua sobreposição (Figura 15).

Como citado anteriormente, CQ de *M. leprae* pertence à família das nucleosídeo monofosfato (NMP) quinases, que são compostas por 3 domínios: CORE, LID e domínio de ligação NMP. As quinases apresentam grande movimento durante a catálise para defender o sítio ativo da água evitando a hidrólise do ATP (PEREIRA, 2005). A sobreposição das estruturas geradas

(Figura 16) mostra as mudanças conformacionais nos domínios LID e SB devido à ligação do chiquimato (substrato) e ADP.

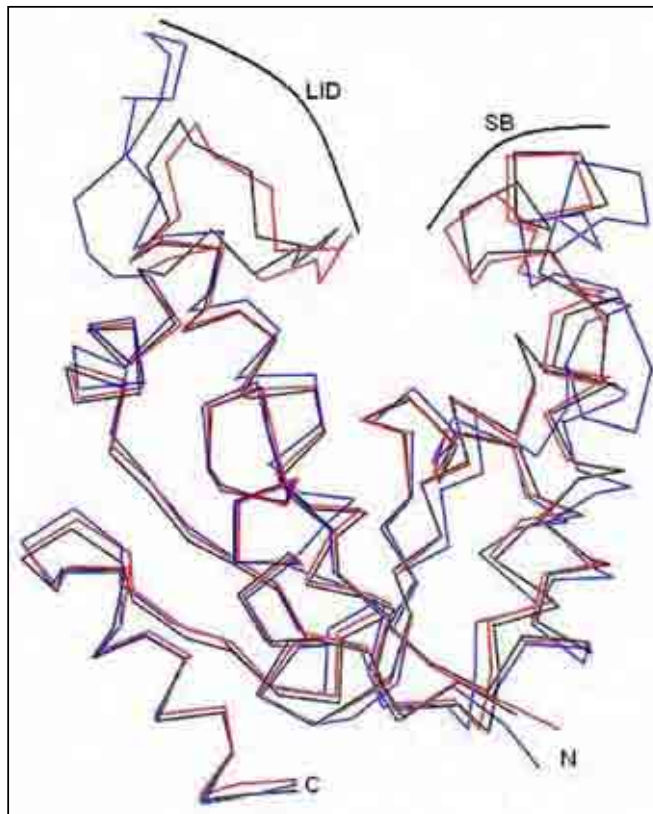


Figura 15 – Sobreposição dos modelos de CQ *M. leprae* (ADP:Cl:Mg) em preto, (ADP:SHI:Cl:Mg) em vermelho e (ADP:Mg) em azul.

A sobreposição dos C_{α} dos modelos de CQ *M. leprae* (CQ:ADP:Mg) e CQ *M. leprae*(CQ:ADP:Cl:Mg) mostraram também que os domínios LID e SB sofreram um notável deslocamento em direção um ao outro na estrutura CQ *M. leprae* (CQ:ADP:SHI:Cl:Mg), como pode-se observar na figura 15.

As mudanças conformacionais descritas acima podem ser demonstradas pela redução da superfície molecular da CQ *M. leprae* (CQ:ADP:Mg) e CQ *M. leprae*(CQ:ADP:Cl:Mg) quando comparada com CQ *M. leprae* (CQ:ADP:SHI:Cl:Mg), figura 16 A-C.

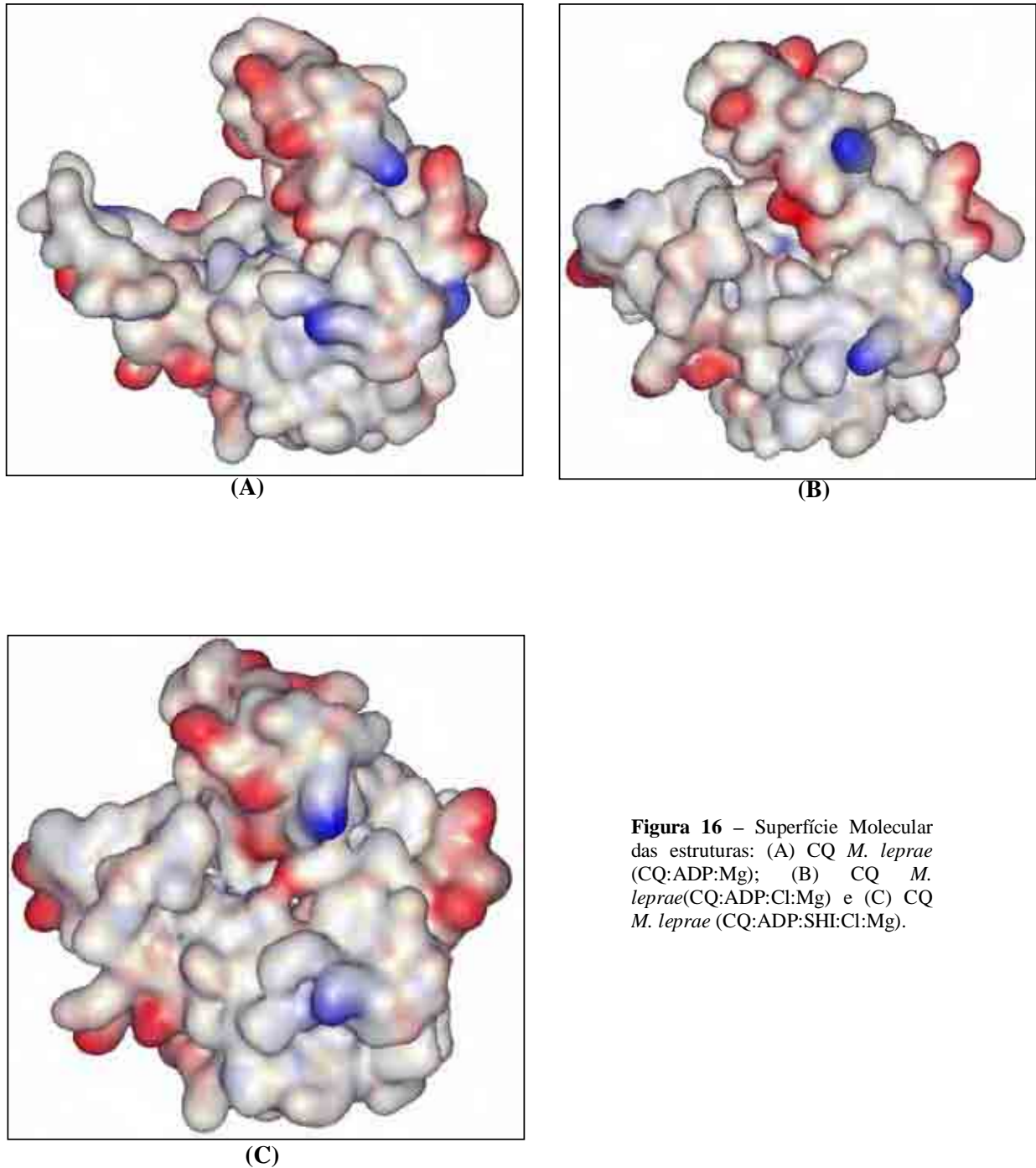


Figura 16 – Superfície Molecular das estruturas: (A) CQ *M. leprae* (CQ:ADP:Mg); (B) CQ *M. leprae*(CQ:ADP:Cl:Mg) e (C) CQ *M. leprae* (CQ:ADP:SHI:Cl:Mg).

As interações não covalentes como ligações intermoleculares de hidrogênio, interações iônicas (eletrostáticas) e interações hidrofóbicas e de van

der Waals são mais fracas que as ligações covalentes. As estruturas da maioria das moléculas biológicas são determinadas por influência coletiva de muitas interações não covalentes individuais. O efeito acumulativo dessas interações fracas se torna significativo, ajudando a estabilizar a estrutura e a definir os níveis de especificidade e afinidade entre proteína-ligante. As ligações intermoleculares de hidrogênio constituem uma das forças de estabilização mais importantes na estrutura de proteínas. As ligações de hidrogênio são incomuns entre as interações eletrostáticas porque são extremamente direcionais; elas se enfraquecem se o ângulo descrito pelos três átomos envolvidos for grande ou pequena. Por isso foi verificada a quantidade de ligações intermoleculares de hidrogênio nos moldes de CQ *M. leprae* com seus respectivos complexos e comparando-as com as informações dos moldes, nas tabelas 7, 8, 9 e 10.

Tabela 7 – Ligações de hidrogênio intermoleculares do SHI do modelo da *mlSK* comparando os valores com as ligações do SHI da *mtSK*.

SHI	<i>mlSK</i>	Distância (Å)	<i>MtSK</i> (1WE2)	Distância (Å)
O1	Gly80 N	3,07	Gly80 N	3,10
O1	Asp34 OD1	2,79	Asp34 OD1	2,82
O2	Asp34 OD1	2,80	Asp34 OD1	2,66
O1	Asp34 OD2	2,66	Asp34 OD2	2,57
O2	Asp34 OD2	2,78	Asp34 OD2	2,85
O4	Gly81 N	3,19	Gly81 N	3,22
O4	Arg136 NH2	2,75	Arg136 NH2	2,68
O5	Arg136 NH2	2,70	Arg136 NH2	2,34
O5	Arg58 NH2	2,71	Arg58 NH2	2,67
O5	Gly81 N	3,47	Gly81 N	3,49

Tabela 8 – Ligações de hidrogênio intermoleculares do ADP da *m*SK comparando os valores com as ligações do ADP da *mt*SK.

ADP	<i>m</i> SK		Distância (Å)	<i>Mt</i> SK(1L4Y)		Distância (Å)
O2B	Gly12	N	3,14	Ser 16	N	2,93
O2A	Gly12	N	2,73	Ser 16	N	3,13
O2A	Ala13	N	3,12	Thr 17	N	2,76
O5	Thr17	OG1	3,10	-----	-----	-----
N6	His153	O	2,98	Arg 153	O	2,96
O2A	Gly14	N	2,90	Thr 17	OG1	2,84
O1B	Gly14	N	3,15	Lys 15	NZ	2,81
O1A	Arg117	NH2	2,60	-----	-----	-----
O3B	Arg117	NH1	3,23	Gly 12	N	2,79
O4	Arg20	NH2	3,24	-----	-----	-----
O4	Thr17	OG1	2,91	-----	-----	-----
O1B	Lys15	N	2,41	Lys 15	N	2,89
O2B	-----	-----	-----	Ser 16	OG	3,01
O1B	-----	-----	-----	Ser 13	N	3,30
O1B	-----	-----	-----	Gly 14	N	3,09
O3B	-----	-----	-----	Arg 117	NH1	2,48
N7	Arg 110	NH1	3,21	-----	-----	-----

Tabela 9 – Ligações de hidrogênio intermoleculares do ADP *m*SK comparando os valores com as ligações do ADP da *mt*SK.

ADP	<i>m</i> SK	Distância (Å)	ADP	<i>Mt</i> SK(1WE2)	Distância (Å)
O2B	Ser16 N	3,11	O1B	Leu10 O	3,49
O2B	Ser16 OG	3,09	O3B	Gly12 N	3,15
O3B	Arg117 NH1	2,65	O1B	Ser13 N	3,08
O2A	Thr17 N	2,78	O1B	Ser13 OG	2,66
N6	His153 O	2,95	O3A	Gly14 N	2,75
O3B	Gly12 N	2,83	O3A	Lys15 N	3,35
O1B	Gly14 N	3,14	O2B	Lys15 NZ	2,80
O2A	Ser16 N	3,25	O3B	Lys15 NZ	2,54
O2A	Thr17 OG1	2,84	O1B	Lys15 NZ	2,82
O1B	Lys15 NZ	2,97	O2A	Ser16 OG	2,65
O1B	Lys15 N	2,74	O2A	Ser16 N	3,40
-----	-----	-----	O1A	Ser16 OG	2,98
-----	-----	-----	O2A	Thr17 OG1	3,02
-----	-----	-----	O2A	Thr17 N	3,08
-----	-----	-----	N1	Arg110 NH1	3,05
-----	-----	-----	O4	Arg110 NE	3,41
-----	-----	-----	N3	Arg110 NH1	2,88
-----	-----	-----	N7	Arg110 NH1	3,29
-----	-----	-----	N6	Arg152 O	2,52
-----	-----	-----	N1	Arg152 O	3,23
-----	-----	-----	N1	Arg153 O	3,37
-----	-----	-----	N6	Arg153 O	2,66
-----	-----	-----	N3	Arg153 NE	2,80
-----	-----	-----	N3	Arg153 NH2	2,68
-----	-----	-----	N1	Arg153 N	3,20

Tabela 10 – Ligações de hidrogênio intermoleculares do ADP da *m*SK comparando os valores com as ligações do ADP da *mt*SK.

ADP	<i>m</i> SK		Distância (Å)	<i>Mt</i> SK (2DFT)		Distância (Å)
O1B	Ser 16	N	2,59	Ser 16	N	2,58
O3B	Leu 10	O	3,10	Leu 10	O	3,13
N1	Arg 110	NH2	3,29	-----	-----	-----
O2A	Thr 17	N	2,74	Thr 17	N	2,74
N6	His 153	O	3,25	-----	-----	-----
O3B	Gly 12	N	2,75	Gly 12	N	2,76
O3A	Gly 14	N	3,14	Gly 14	N	3,15
O3B	Ala 13	N	3,26	Ser 13	N	3,23
O2A	Thr 17	OG1	2,66	Thr 17	OG1	2,54
O3B	Lys 15	NZ	2,91	Lys 15	NZ	2,94
O1B	Lys 15	N	2,63	Lys 15	N	2,55

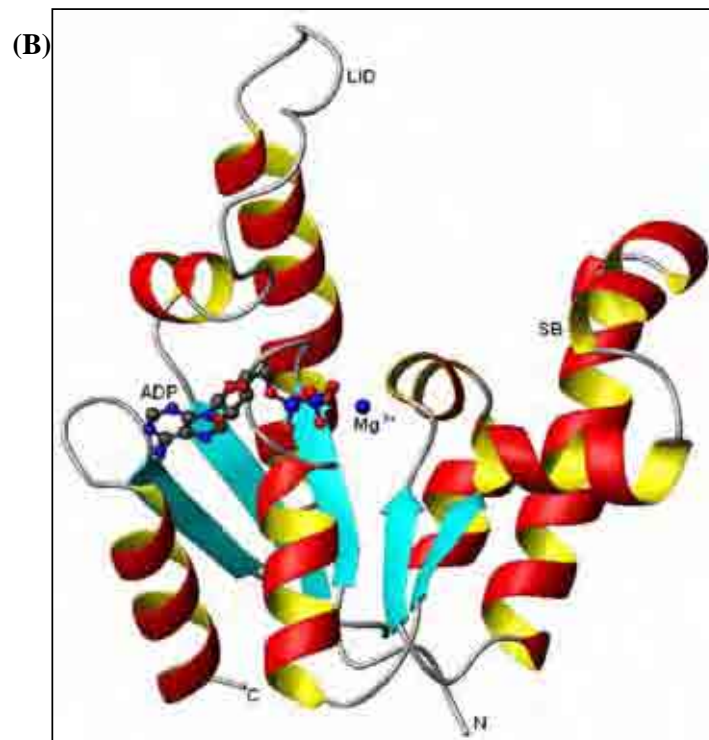
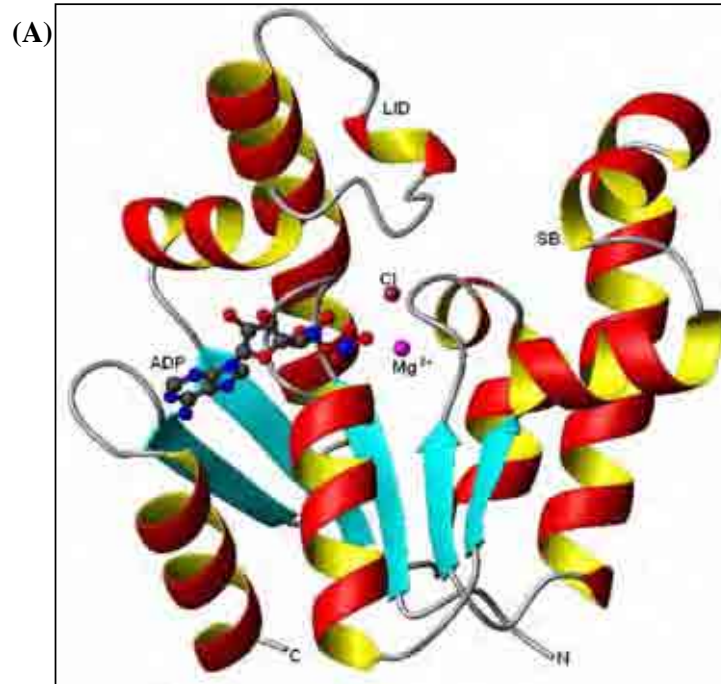
4.3. Descrição do modelo

A figura 17 A-C, mostra os modelos da CQ de *M. leprae* (*m*SK) que foram gerados através do programa MOLMOL (Koradi *et al.*, 1996).

A enzima *m*SK é uma proteína da classe α/β que consiste em um núcleo de cinco folhas- β cercada de hélices- α . As cinco fitas centrais estão em paralelo na ordem 23145 e esta folha- β é flanqueada por hélices- α ($\alpha 1$ e $\alpha 8$ em um lado, $\alpha 4$, $\alpha 5$ e $\alpha 7$ no outro) (PEREIRA *et al.*, 2004a). Esta ordem estabelecida pela folha- β 23145 classifica a enzima *m*SK como sendo pertencente à mesma família estrutural quinases NMP.

Existem duas regiões flexíveis na estrutura que são responsáveis pelo movimento: o sítio de ligação o de chiquimato (SB) e o LID que estão indicados na figura 17 A-C (PEREIRA *et al.*, 2004b). O domínio do SB é

formado pela fita $\beta 2$, hélices $\alpha 1$ e $\alpha 2$ e a região do N-Terminal da hélice $\alpha 4$ (PEREIRA *et al.*, 2004a,b). O sítio de Walker A localizado entre a fita $\beta 1$ e $\alpha 1$ forma um laço de fosfato (*P-loop*) (KRELL *et al.*, 1998).



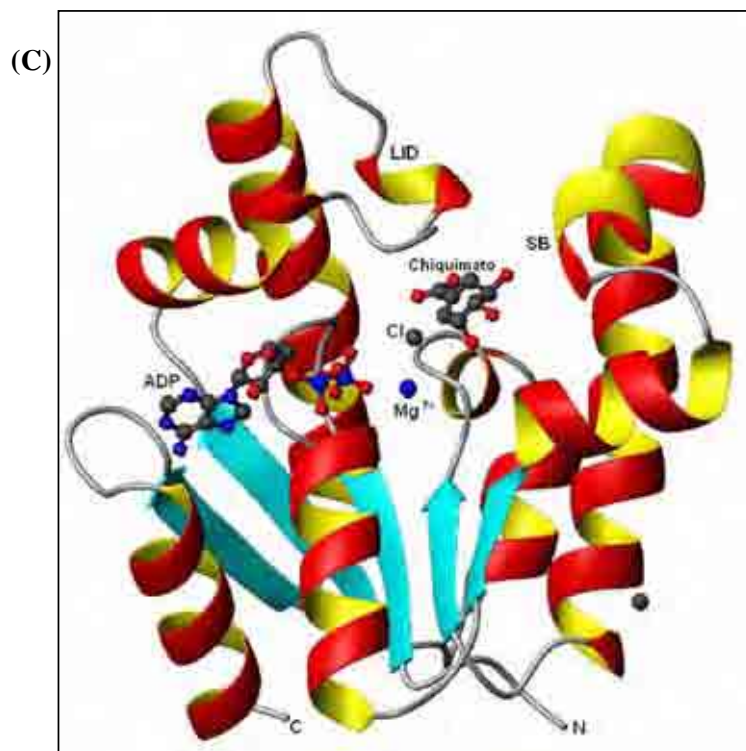


Figura 17 – Estrutura secundária da CQ de *M. leprae* em complexo (A) CQ:ADP:Cl:Mg; (B) CQ:ADP:Mg e (C) CQ:ADP:SHI:Cl:Mg.

5. CONCLUSÃO

Os resíduos identificados na ligação do chiquimato e ADP diretamente ou indiretamente são conservados dos moldes para os modelos gerados. As informações aqui apresentadas podem vir a contribuir para um melhor entendimento de como o sítio ativo da CQ se fecha diante do substrato (Chiquimato) e do co-fator (ADP) e com os íons (Mg^{2+} e Cl).

Na construção de modelos de estruturas tridimensionais por técnicas de modelagem molecular comparativa, a homologia entre as seqüências de aminoácidos é uma condição necessária para que um modelo de boa qualidade

possa ser obtido. Homologia estrutural revela a presença de ancestrais comuns, mas com divergências evolutivas por meio de mutações de seqüências. Essas premissas corroboram a utilização de programas que utilizam a semelhança estrutural para a modelagem molecular.

A precisão do modelo gerado determina sua utilidade. Como a precisão da modelagem comparativa de proteínas está relacionada à porcentagem de identidade no qual o modelo é baseado e na qualidade estereoquímica, foi verificado que os modelos finais apresentam uma precisão alta com base nos valores obtidos nas análises feitas até agora. Com isso, podemos afirmar com base nos resultados finais dos modelos da *MISK*, que suas estruturas têm qualidade suficiente para serem utilizadas no desenvolvimento de inibidores que forcem a estrutura em uma conformação fechada que seja incapaz de catalisar a transferência do fosfato para o chiquimato.

6. REFERÊNCIAS

ALLEN, F. H. et al; The Cambridge Crystallographic Data Centre: computer-based search, retrieval, analysis and display of information. **Acta Crystallogr.** v. B35, p. 2331-2339, 1979.

ALTSCHUL, S. F. et al; Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic Acids Res.** v. 25, p. 3389-3402, 1997.

ARCURI, H. A. et al; Molecular models for shikimate pathway enzymes of *Xylella fastidiosa*. **Biochem. Biophys. Res. Commun.** v. 320, n. 3, p. 979-991, 2004.

BACON, D. J.; ANDERSON, W. F. A fast algorithm for rendering space-filling molecule pictures. **J. Mol. Graphics** v. 6, p. 219–220, 1988.

BADGER, J. et al; Structural analysis of a set of proteins resulting from a bacterial genomics project. **Proteins.** v. 60, p. 787-796, 2005.

BAIROCH, A.; APWEILER, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. **Nucleic Acids Res.**, v. 28, p. 45-48, 2000.

BERENDSEN, H. J. C. et al; GROMACS: a message-passing parallel molecular dynamics implementation. **Comp. Phys. Commun.** v. 91, p. 43-56, 1995.

BERMAN, H. M. et al; The protein data bank. **Nucleic. Acids Res.** v. 28, p. 235–242, 2000.

BERNSTEIN, F. C. et al; The protein data bank: a computer-based archival file for macromolecular structures. **J. Mol. Biol.** v. 112, p. 535-542, 1977.

BLUNDELL, T.L. et al; Knowledge-based prediction of protein structures and the design of novel molecules. **Nature** v. 326, p. 347-352, 1987.

BOWIE, J. U. et al; A method to identify protein sequences that fold into a known three- dimensional structure. **Science** v. 253, p. 164-170, 1991.

BRITTON, W. J.; LOCKWOOD, D. H. J. Leprosy. **Lancet.** v. 363, p. 1209-1219, 2004.

BRÜNGER, A. T. **X-PLOR Version 3.1**: a system for crystallography and NMR, New Haven; Yale University Press, 1992.

CAMPBELL, M. M. et al; The biosynthesis and synthesis of shikimic acid, chirismic acid, and related compounds. **Synth. Stuttg.** v. 2, p. 179-193, 1993.

CAMPBELL, S. A. et al; A complete shikimate pathway in *Toxoplasma gondii* an accident eukaryotic innovation. **Int. J. Parasitol.** v. 34, p. 5-13, 2004.

CARPENTER, E. P. et al; Structure of dehydroquinate synthase reveals an active site capable of multistep catalysis. **Nature**, v. 394, n. 6690, p. 299-302, 1998.

COGGINS, J. R. et al; Experiences with the shikimate-pathway enzymes as targets for rational drug design. **Biochem. Soc. Trans.** v. 31, p. 548-552, 2003.

Collaborative Computational Project n° 4., The CCP4 suite: programs for proteins crystallography. **Acta Crystallogr.** v. D50, p. 760-763, 1994.

CORPET, F. Multiple sequence alignment with hierarchical clustering. **Nucl. Acids Res.** v. 16, n. 22, p. 10881-10890, 1988.

DA SILVEIRA, N. J. et al; Structural bioinformatics study of PNP from *Schistosoma mansoni*. **Biochem. Biophys. Res. Commun.** v. 322, p. 100-104, 2004.

DAVIES, G. M. et al; (6S)-6-fluoroshikimic acid, an antibacterial agent acting on the aromatic biosynthetic pathway. **Antimicrob. Agents Chemother.** v. 38, p. 403-406, 1994.

OLIVEIRA, J. S. **A chiquimato quinase (E. C. 2.7.1.71) e a 5-enolpiruvilchiquimato-3-fosfato sintase (E. C. 2.5.1.19) de *Mycobacterium tuberculosis*: alvos potenciais para o desenvolvimento de novas drogas anti-tuberculose.**, em 07/05/2001.

DHALIWAL, B. et al; Crystallographic studies of shikimate binding and induced conformational changes in *Mycobacterium tuberculosis* shikimate kinase. **Febs Lett.** v. 574, p. 49-54, 2004.

DIAS, M. V. B. et al; Effects of the magnesium and chloride ions and shikimate in the structure of shikimate kinase from *Mycobacterium tuberculosis*. **Arch. Biochem. Biophysics.** (No prelo).

DOSSELAERE, F.; VANDERLEYDEN, J. A metabolic node in action: chorismate-utilizing enzymes in microorganisms. **Crit. Rev. Microb.** v. 27, n. 2, p. 75-131, 2001.

ENGH, R. A.; HUBER, R. Accurate bond and angle parameters for X-ray protein structure refinement. **Acta Crystallogr.** v. A47, p. 392-400, 1991.

- FILHO, O. A. S.; ALENCASTRO, R. B. Modelagem de proteínas por homologia. **Quim. Nova** v. 26, n. 2, p. 253-259, 2003.
- FISER, A. et al; Modeling of loops in protein structures, **Protein Sci.** v. 9, p. 1753–1773, 2000.
- FISER, A. et al; Evolution and physics in comparative protein structure modeling. **Acc. Chem. Res.**, v. 35, p. 413-421, 2002.
- FISER, A.; SALI, A. Comparative protein structure modeling. **In: protein structural.** p. 167-206, 2003.
- FISER, A.; SALI, A. Comparative protein structure modeling with Modeller: a practical approach. **In Protein structural**, p. 1-73, 2000.
- GANEM, B. From glucose to aromatics: recent developments in natural products of the shikimic acid pathway. **Tetrahedron**, v. 34, p. 3353-3383, 1978.
- GASTEIGER, E. et al; SWISS-PROT: connecting biomolecular knowledge via a protein database. **Curr. Issues Mol. Biol.**, v. 3, n. 3, p. 47-55, 2001.
- GIBAS, C.; JAMBECK, P. **Desenvolvendo bioinformática**, Rio de Janeiro, Ed. Campus, p. 179-180, 2001.
- GIBSON, F. The elusive branch-point compound of aromatic amino acid biosynthesis. **Trends Biochem. Sci.**, v. 24, p. 36-38, 1999.
- GREER, J. Comparative model-building of the mammalian serine proteases. **J. Mol. Biol.**, v. 153, p. 1027-1042, 1981.
- GU, Y. et al; Cristal structure of shikimate kinase from *Micobacterium tuberculosis* reveals the dynamic role of the LID domain in catalysis. **J. Mol. Biol.**; v. 319, p. 779-789, 2002.
- HERRMANN, K. M. The shikimate pathway: early steps in the biosynthesis of aromatic compounds. **Plant Cell** v. 7, p. 907-919, 1995a.
- HERRMANN, K. M. The shikimate pathway as an entry to aromatic secondary metabolism. **Plant Physiol.** v. 107, p. 7-12, 1995b.
- HERRMANN, K. M.; WEAVER, L. M. The shikimate pathway. **Annu. Rev. Plant Mol. Biol.**, v. 50, n. 3, p. 473-503, 1999.
- HOOFT, R. W. et al; Errors in protein structures. **Nature** v. 381, p. 272, 1996.

JACOBSON, M.; SALI, A. Comparative protein structure modeling and its applications to drug discovery. **Annu. Rep. Med. Chem.**, v. 39, p. 259-276, 2004.

JOHNSON, M. S. et al; Knowledge-based protein modelling. **CRC Crit. Rev. Biochem. Mol. Biol.**, v. 219, p. 1-68, 1994.

KABSCH, W.; SANDER, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. **Biopolymers**, v. 22, n. 12, p. 2577-2637, 1983.

KEARSLEY, S. K. On the orthogonal transformation used for structural comparisons. **Acta Crystallogr.**, v. 45A, p. 208-210, 1989.

KISHORE, G. M.; SHAH, D. M. Amino acid biosynthesis inhibitors as herbicides. **Annu. Rev. Biochem.**, v. 57, p. 627-663, 1988.

KRAULIS, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. **J. Appl. Crystallogr.**; v. 24, p. 946-950, 1991.

KRELL, T. et al; Biochemical and X-ray crystallographic studies on shikimate kinase: the important structural role of the P-loop lysine. **Protein Sci.**, v. 10 p. 1137-1149, 2001.

KRELL, T. et al; Crystallization and preliminary X-ray crystallographic analysis of shikimate kinase from *Erwinia Chrysanthemi*. *Acta Crystallogr.*, v. 53, p. 612, 1997.

KRELL, T. et al; The three-dimensional structure of shikimate kinase. **J. Mol. Biol.**, v. 278, p. 983-997, 1998.

KORADI, R. et al; MOLMOL: a program for display and analysis of macromolecular structures. **J. Mol. Graphics**, v. 14: p. 51-55, 1996.

LASKOWSKI, R. A. et al; PROCHECK - a program to check the stereochemical quality of protein structures, **J. Appl. Cryst.**, v. 26, p. 283-291, 1993.

LINDAHL, E. et al; GROMACS 3.0: a package for molecular simulation and trajectory analysis. **J. Mol. Mod.**, v. 7, p. 306-317, 2001.

LOCKWOOD, D. H. J.; SUNEETHA, S. Leprosy: too complex a disease for a simple elimination paradigm. **Bull. World Health Organ.**, v. 83, n. 3, p. 230-235, 2005.

LUTHY, R. et al; Assessment of protein models with three-dimensional profiles. **Nature**, v. 356, p. 383-385, 1992.

- MACKERELL JR., A. D. et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. **J. Phys. Chem. B.**, v. 102, p. 3586-3616, 1998.
- MARTI-RENO, M. A. et al; Comparative protein structure modeling of genes and genomes. **Annu. Rev. Biophys. Biomol. Struct.**, v. 29, p. 291-325, 2000.
- McREE, D. E. XtalView/Xfit : a versatile program for manipulating atomic coordinates and electron density. **J. Struct. Biol.**, v. 125, p. 156-165, 1999.
- MERRITT, E. A. Raster3D version 2.0. A program for photorealistic molecular graphics. **Acta Crystallogr.**, v. 50, p. 869–873, 1994.
- MILNER-WHITE, J. E. et al; Evidence for an ancestral core structure in nucleotide-binding proteins with the type A-motif. **J. Mol. Biol.**, v. 221, p. 751-754, 1991.
- MONOT, M. et al; On the origin of leprosy. **Science**. v. 308, p. 1040-1042, 2005.
- MORRIS, A. L. et al; Stereochemical quality of protein structure coordinates. **J. Appl. Cryst.**, v. 12, p. 345-364, 1986.
- MORRIS, A. L. et al; Stereochemical quality of protein structure coordinates. **Proteins**, v. 12, p. 345-364, 1992.
- MUNIZ, J. R. C. **Aplicação da bioinformática no estudo dos genes e enzimas envolvidos na síntese da goma fastidiana produzida pela *Xylella fastidiosa***. São Carlos, 2003. 135 p. Tese (Mestrado) – Instituto de Física de São Carlos, Universidade de São Paulo.
- NEEDLEMAN, S.; WUNSCH, C. A general method applicable to the search for similarities in the amino acid sequence of two proteins. **J. Mol. Biol.**, v. 48, p. 443-453, 1970.
- PEARSON, W. R. Rapid and sensitive sequence comparison with FASTP and FASTA. **Methods Enzymol.**, v. 183, p. 63-98, 1990.
- PEREIRA, J. H. et al; Interaction of shikimic acid with shikimate kinase. **Biochem. Biophys. Res. Commun.**, v. 325, p. 10-17, 2004b.
- PEREIRA, J. H. et al; Structural bioinformatics study of EPSP synthase from *Mycobacterium tuberculosis*. **Biochem. Biophys. Res. Commun.**, v. 312, p. 608-614, 2003.
- PEREIRA, J. H. et al; Structure of shikimate kinase from *Mycobacterium tuberculosis* reveals the binding of shikimic acid. **Acta Crystallogr.**, v. D60, p. 2310–2319, 2004a.

ROBERTS, C. W. et al; The shikimate pathway and its branches in apicomplexan parasites. **J. Infect. Dis.**, v. 185, n. 1, p. S25-36, 2002.

ROBERTS, F. et al; Evidence for the shikimate pathway in apicomplexan parasites. **Nature**, v. 393, n. 6687, p. 801-805, 1998.

ROMANOWSKI, M. J.; BURLEY, S. K. Crystal structure of the Escherichia coli shikimate kinase I (AroK) that confers sensitivity to mecillinam. **Proteins**, v. 47, p. 558-562, 2002.

SALI, A; BLUNDELL, T. L. Comparative protein modelling by satisfaction of spatial restraints. **J. Mol. Biol.**, v. 234, p. 779-815, 1993.

SALI, A.; OVERINGTON, J. P. Derivation of rules for comparative protein modeling from a database of protein structure alignments. **Protein Sci.**, v. 3, p. 1582-1596, 1994.

SALI, A. et al; Evaluation of comparative protein modeling by MODELLER. **Proteins**, v. 23, p. 318-326, 1995.

SANCHEZ, R.; SALI, A. Advances in comparative protein-structure modelling. **Curr. Opin. Struct. Biol.**, v. 7, p. 206-214, 1997.

SANCHEZ, R.; SALI, A. Comparative protein structure modeling. Introduction and practical examples with modeller. **Methods Mol. Biol.**, v. 143, p. 97-129, 2000.

SANCHEZ, R.; SALI, A. Large-scale protein structure modeling of the *Saccharomyces cerevisiae* genome. **Proc. Natl. Acad. Sci.**, v. 95, p. 13597-13602, 1998.

SCHWIETERS, C. D. et al; The Xplor-NIH NMR molecular structure determination package. **J. Magn. Res.**, v. 160, p. 66-74, 2003.

THOMPSON, J. D. et al; CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. **Nucleic Acids Res.**, v. 22, p. 4673-4680, 1994.

TRAMONTANO, A. et al; Analysis and assessment of comparative modeling predictions in CASP4. **Proteins**, v. 45, p. 22-38, 2001.

UCHÔA, H. B. et al; Parmodel: a web server for automated comparative modeling of proteins. **Biochem. Biophys. Res. Commun.**, v. 325, p. 1481-1486, 2004.

VAN GUNSTEREN, W. F. et al; **Biomolecular simulations: the GROMOS96 manual and user guide.** VdF Hochschulverlag, Zurich, ETHZ, 1996.

VONRHEIN, C. et al; Movie of the structural changes during a catalytic cycle of nucleoside monophosphate kinases. **Structure**, v. 3, n. 5, p. 483-490, 1995.

VRIEND, G. WHAT IF: a molecular modeling and drug design program. **J. Mol. Graph.**, v. 8, p. 52-56, 1990.

WALKER G. E. et al; Evidence for a novel class of microbial 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase in *Streptomyces coelicolor* A3(2), *Streptomyces rimosus* and *Neurospora crassa*. **Microbiology**, v. 142, p. 1973-1982, 1996.

WALKER, J. E. et al; Distantly related sequences in the α - and β - subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. **EMBO J.**, v. 1, p. 945-951, 1982.

YSAI, M. D.; YAN, H. Nucleoside monophosphate kinases: structure, mechanism, and substrate specificity. **Adv. Enzymol. Relat. Areas Mol. Biol.**, v. 73, p. 103-134, 1999.