

Universidade de São Paulo

Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto
Faculdade de Medicina de Ribeirão Preto
Informática Biomédica

Predição da estrutura 3D da proteína Hsp65 e da
estrutura secundária de seu
RNA mensageiro usando análises *in silico*

Renata Ariza Marques Rossetti
renata_ariza@yahoo.com.br

Ribeirão Preto – SP
Dezembro 2008.

Universidade de São Paulo

Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto
Faculdade de Medicina de Ribeirão Preto
Informática Biomédica

Predição da estrutura 3D da proteína Hsp65 e da
estrutura secundária de seu
RNA mensageiro usando análises *in silico*

Trabalho de conclusão de curso
apresentado como parte dos requisitos
para a obtenção do título de bacharel
em Informática Biomédica

Aluna: Renata Ariza Marques Rossetti
Orientadora: Arlete A. M. Coelho-Castelo
Co-orientadora: Silvana Giuliatti

Ribeirão Preto – SP
Dezembro 2008.

Dedicatória

Não poderia dedicar este trabalho a outras pessoas que não meus pais. Graças a eles pude fazer uma ótima faculdade sem me preocupar com nada. Quando digo com nada, não penso apenas em dinheiro, mas sim em compreensão. Por mais brigas e discussões que tivemos ao longo desses cinco anos, sempre soube (bem no fundo!) que podia contar com eles para o que precisasse. Muitas vezes não queria escutar aquele tipo de conselho, mas depois de passada a situação percebia que não havia conselho melhor a ser dado.

Obrigada pelas infinitas vezes que vieram me buscar ou levar de Ribeirão, por todos os cuidados com meu pequeno apartamento, pela compreensão nas brigas e discussões, enfim, obrigado por tudo! Amo vocês!

Gostaria de dedicar também às pessoas que me ajudaram de forma direta na Monografia.

Primeiramente a minha orientadora, Arlete, muito obrigada pela oportunidade e por acreditar em mim, nada mais prova isso do que mandar um projeto pra Fapesp e conseguir a bolsa mesmo eu tendo alguns, vamos dizer, obstáculos no currículo.

Silvana, obrigado por tornar essa monografia possível, pois sem sua orientação não saberia nem como começar.

Carlos, não existe palavras para explicar como sou grata a você. Mesmo não sendo meu orientador, co-orientador, colaborador, me ajudou de todas as formas possíveis. Quando achei que nada mais tinha solução, você me orientava, respondia e-mail mesmo estando em outro país. Obrigada pela paciência.

Agradecimentos

Esperei cinco anos por esse momento, e esses cinco anos passaram como cinco meses graças a muitas pessoas que deixaram marcas. Marcas essas de coisas positivas e negativas, pois nada é perfeito, mas hoje quando me lembro dos momentos só coisas felizes me vêm a mente. Como era gostoso ir para a faculdade todo dia de manhã, mesmo não suportando mais ouvir falar de cálculo, Java, AED....

Não vou me esquecer nunca da casa da Dura lá no centro quando ela morava ainda com a Ju Sheila e a Tort's. Tinha época que passava mais tempo lá do que na minha própria casa. Ai Dura... quantas coisas não passamos juntas, né? As três mosqueteira, Dura, Ms e eu.... quantos bares, quantas andanças pela cidade a noite, principalmente se tivesse chovendo! Obrigada por tornarem minha faculdade tão legal e movimentada!

Mouse, Baiano e Dizi... como agüentei vocês tanto tempo!? Nunca era tão bom estudar como quando vocês iam em casa, por mais incrível que pareça a gente realmente estudava e graças a esses dias muitas DPs minhas foram evitadas. O Baiano nos abandonou, e tivemos no último ano um remember de como era bom... SO!

Os anos foram passando e as pessoas só foram se tornando mais especiais. Naty e Lixo, meus amiguinhos de São Paulo, metidos por isso. Não sei dizer o ponto que nos tornamos amigos, talvez uma aula chata em que ficamos jogando liga os pontos, uma ida do nada ao Intercomp, não sei, mas conforme o tempo ia passando a amizade entre nós ia só crescendo. Hoje é difícil acreditar que o Lixo virou meu namorado Rodrigo!

No último ano para minha feliz surpresa me aproximei mais de pessoas muito especiais: Jaq, Jujuba, Tang, Juliano, Raul, Monique que fico até um pouco triste por isso não ter acontecido antes, mas como disse a uma amiga: Amizade não tem hora, nem lugar para começar!

Há também pessoinhas que são impossíveis de esquecer, mas que não teve um momento pra se tornarem amigas, simplesmente aconteceu da forma mais natural possível: Pato, Rapel, Rato, Mentira e Hélio. Como

esquecer da caminhada diária pra faculdade, acordava cedo só para não receber bronca da Mentira por me atrasar!

Aliás, se hoje entrego essa Monografia, muito é graças ao Rapel, que me apresentou ao Júlio, e tudo teve início.

Não posso deixar de agradecer meus veteranos, que tornaram a faculdade mágica desde o primeiro dia de aula, ou melhor, desde o trote na matrícula: Boi, Pólis, Piffer, Pedro, Brody, “Campo Minado”.

Por último, gostaria de agradecer, ao pessoal do laboratório: Isabela, Thiago, Piá, Queijo, Lú, Ana Paula, Ana Flávia, Zé, Willian, Carol e Júlio. Último, não por importância, mas sim por ordem de chegada. Espero conviver muito ainda com vocês.

Um agradecimento especial a Fapesp, não só pelo apoio financeiro, mas também pelos conselhos dados através dos referis dos relatórios.

Resumo

Um dos alvos a ser estudado e esclarecido é o mecanismo da vacina de DNAhsp65 de *Mycobacterium leprae* (hsp65), que está sendo desenvolvido pelo Centro de Pesquisa em Tuberculose da Faculdade de Medicina de Ribeirão Preto – USP (CPT-FMRP) e o Núcleo de Pesquisa em Tuberculose. O Interesse nesse estudo decorre de publicações anteriores onde foi mostrado que essa vacina é capaz de gerar uma atividade profilática e terapêutica contra a infecção por *M. tuberculosis*, (LOWRIE, 1999). Além disso, essa vacina foi a primeira a ser usada em estudo clínico, podendo ser considerado um avanço pioneiro em terapia gênica em nosso País. Uma das características importantes do antígeno hsp65 é ser uma chaperona, o que pode implicar seu papel em mecanismos celulares, contribuindo, assim, para a eficácia da vacina. Análises *in silico* demonstram que a hsp65 apresenta quatro sítios possíveis de fosforilação o que pode gerar peptídeos contendo esses resíduos *in vivo*. Embora a importância de se obter um cristal da molécula seja imprescindível, a bioinformática dispõe de várias ferramentas que facilitam o designer de moléculas *in silico*, a partir de seqüências de aminoácidos. Assim, o objetivo desse projeto é determinar, através de programas apropriados, a estrutura tridimensional da hsp65 de *M. leprae* baseada na seqüência de aminoácidos da mesma. Outro objetivo foi a determinação da estrutura secundária do RNAm, importante para o avanço em terapia gênica usando moléculas de RNA. Tais dados são de extrema utilidade na determinação das seqüências mais imunogênicas tanto da proteína como do RNA. Os resultados obtidos favorecem o desenvolvimento de novas formas vacinais e, até mesmo, o uso de peptídeos ou RNA para o controle da resposta imune. Somado a isso, esse estudo dará uma visão global das moléculas em questão, facilitando o entendimento de suas funções biológicas.

Sumário

Lista de Figuras	1
Capítulo 1: Introdução	2
1.1. Tuberculose	2
1.2. Princípios Básicos de Biologia Celular	3
1.3. Vacinas de DNA e RNA.....	6
1.4. Estrutura das Proteínas	8
1.5. Proteína de Choque Térmico de 65KDa (hsp65)	10
1.6. Bioinformática.....	12
1.7. Bioinformática.....	13
Capítulo 2: Métodos	14
2.1. Modelagem do RNAm da HSP65 de <i>M.leprae</i>	14
2.2. Modelagem da proteína hsp65 de <i>M.leprae</i>	15
Capítulo 3: Resultados e Discussão.....	18
3.1. Modelagem do RNAm da HSP65 de <i>M.leprae</i>	18
3.2. Modelagem da proteína hsp65 de <i>M.leprae</i>	22
3.2.1. Escolha das proteínas moldes	22
3.2.2. Alinhamento da 1SJP com a 1WE3.....	26
3.2.3. Alinhamento da hsp65 de <i>M.leprae</i> com o modelo resultante do alinhamento múltiplo.....	28
3.2.4. Construção do modelo da hsp65 de <i>M.leprae</i>	30
3.2.5. Análise Estereoquímica	31
Capítulo 4: Conclusão Final	37
Referências Bibliográficas.....	38

Lista de Figuras

Figura 1: Tabela do código genético.....	04
Figura 2: Estrutura do aminoácido.....	08
Figura 3: Esquema da modelagem por homologia.....	14
Figura 4: Modelo dos RNA mensageiros de estudo.....	19
Figura 5: Parte específica do RNA mensageiro para análise.....	20
Figura 6: Trecho do script <u>build_profile.py</u>	21
Figura 7: Alinhamento entre hsp65 de <i>M.leprae</i> e <i>M.tuberculosis</i>	23
Figura 8: Alinhamento entre hsp65 de <i>M.leprae</i> e a 1WE3	24
Figura 9: Trecho do script <u>salign.py</u>	25
Figura 10: Alinhamento entre 1SJP e 1WE3.....	26
Figura 11: Trecho do script <u>salign2d.py</u>	27
Figura 12: Alinhamento múltiplo entre 1SJP, 1WE3 e hsp54 de <i>M.leprae</i> ..	28
Figura 13: Trecho do script <u>model_mult.py</u>	29
Figura 14: Modelo molecular da hsp65 de <i>M.leprae</i>	29
Figura 15: Gráfico de Ramachandran.....	31
Figura 16: Gráfico da cadeia principal	32
Figura 17: Gráfico gerado pelo VERIFY3D.....	34
Figura 18: Trecho do arquivo gerado pelo WHATIF	35

Capítulo 1: Introdução

1.1. Tuberculose

A tuberculose é uma patologia que acompanha o homem desde a milhares de anos, existindo relatos da doença na Grécia e Roma antigas. No entanto, foi somente na segunda metade do século XIX que se atribuiu à bactéria *Mycobacterium tuberculosis* a causa da infecção. A tuberculose afeta principalmente os pulmões, sendo transmitida pelo ar através da tosse de pessoas infectadas pela bactéria. Apesar de acompanhar a humanidade há muito tempo, a tuberculose ainda é motivo de grande preocupação, pois, segundo a Organização Mundial de Saúde (OMS), 1,5 milhões de pessoas morreram vítimas dessa doença em 2006. Soma-se também o fato que outras 200 mil morreram pela associação da tuberculose com a AIDS (Síndrome da Imunodeficiência Adquirida). Assim, o subdesenvolvimento em muitos países aliados a emergência do HIV, fazem da tuberculose um dos principais problemas de saúde pública.

Uma alternativa para o controle da doença surgiu por volta de 1921, pelo desenvolvimento de uma vacina atenuada. Essa vacina, utilizada até hoje, na maioria dos países é conhecida como BCG (*Bacille Calmette-Guérin*), e, considerada a primeira vacina do século XX. Ela foi obtida através da atenuação do *Mycobacterium bovis*, sendo aplicada em crianças com menos de cinco anos de idade. Um dos principais problemas com a vacina BCG é a não proteção de adultos (LUGOSI, 1992). Isso é um dos principais fatores que levam a busca de novas estratégias vacinais, incluindo as vacinas de DNA.

1.2. Princípios Básicos de Biologia Celular

A informação genética direciona de algum modo à síntese de proteínas. As proteínas são os principais constituintes das células e determinam não somente a estrutura da célula, mas também suas funções. Portanto, as informações genéticas presentes no Ácido Desoxirribonucléico (DNA) determinam as seqüências de aminoácidos das proteínas. Os processos responsáveis pela concretização desse fluxo de informações são a transcrição (mecanismo para a cópia de DNA para RNA) e tradução (processo de produção da proteína através do RNA).

Tanto o DNA como o Ácido Ribonucléico (RNA), são polímeros lineares compostos de quatro tipos distintos de nucleotídeos unidos por ligações fosfodiéster. Nucleotídeos são formados por um grupo fosfato, uma pentose (molécula de açúcar com cinco carbonos) e uma base nitrogenada, podendo ser Adenina (A), Citosina (C), Guanina (G), Timina (T) ou Uracila (U).

As diferenças fundamentais entre eles são: o açúcar do DNA é a desoxirribose, enquanto o do RNA é a ribose; o DNA tem como bases nitrogenadas A, G, C e T, enquanto o RNA tem a U no lugar da T; e estruturalmente, o DNA ocorre nas células como uma hélice de fita dupla, enquanto o RNA é fita simples. É devido a fita simples do RNA que ele pode se dobrar em diversas formas, assim como uma cadeia de polipeptídicos pode dobrar-se para produzir a conformação final da proteína.

O RNA mensageiro (RNAm) é obtido através de um processo conhecido como transcrição. Neste processo, a informação contida no Ácido Desoxirribonucléico (DNA), será transmitida para o RNAm, responsável por levar a informação genética contida no núcleo da célula (onde o DNA reside) para o citosol.

A transcrição é realizada apenas em um dos filamentos do DNA, sempre no sentido 5' para o 3'. A RNA polimerase é a enzima responsável pela transcrição. Ela se liga a seqüências específicas de DNA chamadas de promotores para iniciar a síntese de RNA. Estas seqüências antecedem a região que codifica as proteínas, e contêm seqüências de DNA curtas e conservadas. A RNA polimerase

move-se ao longo do DNA e sintetiza seqüencialmente a cadeia de RNA. O DNA é desenrolado adiante da polimerase em movimento, e a hélice é reformada atrás dela. A RNA polimerase reconhece o finalizador que faz com que outros ribonucleotídios não sejam incorporados.

Diversos tipos de RNA são produzidos nas células. O RNA mensageiro (RNAm) é responsável por dirigir a síntese protéica. Para produzir o RNAm nas células eucarióticas, o comprimento inteiro do gene, incluindo os íntros e éxons, é transcrito em uma longa molécula de RNA (transcrito primário). Antes que o RNAm saia do núcleo todas as seqüências de íntros são removidas e o éxons ligados uns aos outros. O resultado é uma molécula muito menor, a qual contém uma seqüência codificante não interrompida. Tal processo é chamado de splicing.

A tradução consiste na conversão da informação de RNA para proteína, ou seja, converte a informação para outra linguagem, de nucleotídeo para aminoácido. Para isso há um sistema de criptografia na natureza que associa 64 possíveis combinações de três nucleotídeos (códon) em apenas 20 aminoácidos, mostradas na Figura 1.

		2ª base			
		U	C	A	G
1ª base	U	UUU (Phe/F) Fenilalanina	UCU (Ser/S) Serina	UAU (Tyr/Y) Tirosina	UGU (Cys/C) Cisteína
		UUC (Phe/F) Fenilalanina	UCC (Ser/S) Serina	UAC (Tyr/Y) Tirosina	UGC (Cys/C) Cisteína
		UUA (Leu/L) Leucina	UCA (Ser/S) Serina	UAA "Ocre" (Stop)	UGA "Opala" (Stop)
		UUG (Leu/L) Leucina	UCG (Ser/S) Serina	UAG "Âmbar" (Stop)	UGG (Trp/W) Triptofano
	C	CUU (Leu/L) Leucina	CCU (Pro/P) Prolina	CAU (His/H) Histidina	CGU (Arg/R) Arginina
		CUC (Leu/L) Leucina	CCC (Pro/P) Prolina	CAC (His/H) Histidina	CGC (Arg/R) Arginina
		CUA (Leu/L) Leucina	CCA (Pro/P) Prolina	CAA (Gln/Q) Glutamina	CGA (Arg/R) Arginina
		CUG (Leu/L) Leucina	CCG (Pro/P) Prolina	CAG (Gln/Q) Glutamina	CGG (Arg/R) Arginina
	A	AUU (Ile/I) Isoleucina	ACU (Thr/T) Treonina	AAU (Asn/N) Asparagina	AGU (Ser/S) Serina
		AUC (Ile/I) Isoleucina	ACC (Thr/T) Treonina	AAC (Asn/N) Asparagina	AGC (Ser/S) Serina
		AUA (Ile/I) Isoleucina	ACA (Thr/T) Treonina	AAA (Lys/K) Lisina	AGA (Arg/R) Arginina
		AUG (Met/M) Metionina, Start	ACG (Thr/T) Treonina	AAG (Lys/K) Lisina	AGG (Arg/R) Arginina
	G	GUU (Val/V) Valina	GCU (Ala/A) Alanina	GAU (Asp/D) Ácido aspártico	GGU (Gly/G) Glicina
		GUC (Val/V) Valina	GCC (Ala/A) Alanina	GAC (Asp/D) Ácido aspártico	GGC (Gly/G) Glicina
		GUA (Val/V) Valina	GCA (Ala/A) Alanina	GAA (Glu/E) Ácido glutâmico	GGA (Gly/G) Glicina
		GUG (Val/V) Valina	GCG (Ala/A) Alanina	GAG (Glu/E) Ácido glutâmico	GGG (Gly/G) Glicina

Figura 1: Tabela para conversão de nucleotídeo para aminoácido.

A tradução é um mecanismo mais complexo que a transcrição. Assim apenas um breve resumo será apresentado. Há três etapas.

A primeira etapa é a iniciação. A subunidade menor do ribossomo liga-se a extremidade 5' do RNAm, que desliza ao longo do RNAm até encontrar o códon de iniciação. O RNA transportador (tRNA) que estará transportando o aminoácido correspondente, liga-se ao códon por complementaridade.

A segunda fase é o alongamento. Ocorrem ciclos repetidos de transporte de aminoácidos, havendo a formação de ligações peptídicas entre os aminoácidos que vão sendo traduzidos. O ribossomo vai avançando ao longo do RNAm de três em três bases, sempre da extremidade 5' para a 3'. Os tRNA que já se ligaram inicialmente, vão se desprendendo do RNAm sucessivamente.

A última etapa é a finalização. O ribossomo encontra o códon de finalização terminando o alongamento. O último tRNA coloca o aminoácido correspondente, as unidades dos ribossomos se separam e a cadeia polipeptídica é liberada (ALBERTS, 2005)(TURNER, 2005).

1.3. Vacinas de DNA e RNA

Como a vacina de BCG só é eficaz na fase infantil, a busca por novas alternativas têm sido uma constante. Assim com o desenvolvimento da estratégia de vacinas de DNA, que são capazes de estimular ambas as subpopulações de linfócitos TCD4 e TCD8 para um padrão TH1, houve uma nova expectativa para o desenvolvimento e controle de doenças infecciosas que necessitam desse perfil de ativação do sistema imune, como a tuberculose.

O grande mérito dessas vacinas é induzir células CD8 antígeno específicas e já carregar um adjuvante mediado pelos CpGs procarióticos, presente na construção da vacina (GURUNATHAN, 2000). Os meios de inoculação do plasmídeo podem ser vários, sendo o mais comum pelo método intramuscular, mas existem também as formas de inoculação pela via oral, pela via nasal (em forma de aerossol) e a forma intradérmica, esta última por meios de bombardeamento de micropartículas de ouro cobertas pelo DNA (LIMA, 2003).

Atualmente é demonstrado que a proteína mais eficaz para o desencadeamento da resposta imune contra a tuberculose, por meio de vacina de DNA, é a proteína de choque térmico de 65 KDa (Hsp65), sendo obtida da *Mycobacterium leprae* (SILVA, 1996). Desse modo, várias abordagens têm sido utilizadas usando essa construção plasmidial como modelo. Assim, foi demonstrado que após a imunização intramuscular o DNA plasmidial pode ser capturado e expresso por células B o que sugere que essas células também possam participar como apresentadoras de antígenos nesses modelos (COELHO-CASTELO, 2003).

Outro aspecto determinado foi com relação à biodistribuição da vacina e sua incorporação ao genoma da célula eucariótica. Os resultados obtidos demonstraram que a vacina não se incorpora ao genoma do hospedeiro, mas há uma ampla distribuição dependendo da dose vacinal utilizada (COELHO-CASTELO, 2006). Esses resultados são importantes com relação à biossegurança dessas construções em uso humano. Porém, um aspecto ainda pouco explorado

diz respeito à participação do antígeno hsp65 em modulação celular, desde que o mesmo é uma chaperonina e apresenta atividade enzimática de endopeptidase.

O uso do RNAm traz algumas vantagens quando utilizado para a construção de vacinas. O antígeno é produzido assim que o RNAm penetra no citosol da célula hospedeira, diferentemente do que ocorre na vacinação com DNA plasmidial, onde o DNA precisa vencer a barreira da membrana nuclear para ser primeiramente transcrito e depois traduzido no citosol. Dessa forma, com o uso do RNAm, o antígeno a ser produzido não enfrentaria tantos obstáculos, podendo ativar o sistema imune mais rapidamente. O uso de RNAmhsp65 veiculado pela rota intranasal, mostrou proteção contra a tuberculose experimental, reforçando a importância dessas moléculas em estratégia profiláticas e terapêuticas (LORENZI, 2008).

O entendimento da interação entre moléculas de DNA e RNA com receptores presentes em células do sistema imune, pode facilitar o desenvolvimento de formulações vacinais mais eficazes e com redução na dose utilizada. Assim, a determinação ou predição da estruturas dessas moléculas é de extrema importância para decifrar os mecanismos envolvidos na proteção.

1.4. Estrutura das Proteínas

Proteínas são longas cadeias de aminoácidos, cada qual unido a seu vizinho através de ligações peptídicas. Cada proteína tem uma seqüência de aminoácidos que lhe é característica.

Aminoácidos são moléculas formadas por átomos de carbono, hidrogênio, oxigênio e nitrogênio, sendo divididos em quatro partes: grupo amina (H_3N^+), grupo carboxílico (COO^-), carbono alpha e um radical característico de cada aminoácido (R).

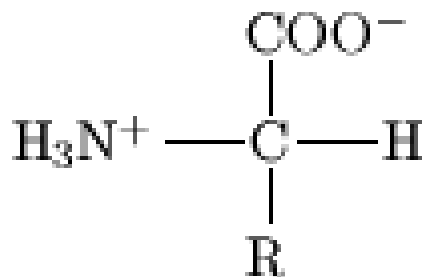


Figura 2: Estrutura geral de um aminoácido.

Ligação peptídica é uma ligação química que ocorre entre dois aminoácidos quando o grupo carboxila de uma molécula reage com o grupo amina de outra molécula, liberando uma molécula de água. Reações desse tipo são classificadas de síntese por condensação.

As proteínas possuem quatro tipos de estruturas. A estrutura primária é a seqüência repetida de átomos de carbono alpha e ligações peptídicas fornecendo o arcabouço do polipeptídio, enquanto as diferentes cadeias laterais dos aminoácidos conferem funcionalidade à proteína.

A natureza altamente polar dos grupos carboxila e amina das ligações peptídicas dão a ligação entre o carbono e o nitrogênio um caráter parcial de ligação dupla. Isto torna a ligação peptídica uma unidade rígida e planar, mas há uma rotação livre entre as ligações peptídicas adjacentes. Esta polaridade também favorece a formação de pontes de hidrogênio entre as unidades peptídicas apropriadamente espaçadas e orientadas. Assim, as cadeias

polipeptídicas são capazes de se dobrar em um número regular de estruturas que são mantidas juntas por estas pontes de hidrogênio, são as chamadas estruturas secundárias. Os principais tipos são alpha-hélice e folha-beta.

As diferentes partes da estrutura secundária e regiões de conexão se dobram em uma bem definida estrutura terciária. A tendência geral da estrutura da molécula é se dobrar com aminoácidos hidrofílicos na superfície e os hidrofóbicos no interior. A estrutura é estabilizada por interações não covalentes e às vezes por pontes dissulfídicas. A natureza da estrutura terciária é inerente à estrutura primária e, nas condições corretas, a seqüência primaria se dobrará espontaneamente na estrutura terciária de menor energia. Para este dobramento pode-se ter ajuda de proteínas especiais, chamadas chaperonas.

Muitas proteínas são compostas de duas ou mais cadeias polipeptídicas, chamadas de unidades. O conjunto dessas unidades forma a estrutura quaternária da proteína. Tais unidades podem ser idênticas ou diferentes. É a estrutura quaternária que permite que moléculas protéicas muito grandes sejam feitas (TURNER, 2005).

A forma tridimensional das proteínas é inteiramente determinada por sua seqüência de aminoácidos. Cada seqüência de aminoácidos resulta em uma única conformação, sendo esta marginalmente estável. Essa conformação irá determinar as exatas propriedades químicas da proteína, o que lhe permite exercer na célula a sua exata função. As proteínas são construídas de forma que a mudança de poucos aminoácidos pode, às vezes, comprometer a estrutura da proteína, e dessa forma causar uma perda de função.

Uma proteína com várias conformações diferentes e propriedades variáveis não seria biologicamente útil, sendo eliminadas por seleção natural no curso do processo evolutivo (ALBERTS, 2005).

1.5. Proteína de Choque Térmico de 65KDa (hsp65)

As proteínas de choque térmico (hsps) são altamente conservadas entre as espécies e desempenham um importante papel no dobramento e translocação de proteínas e também na montagem de complexos protéicos. As hsps estão presentes nos diversos compartimentos celulares e exercem funções vitais na homeostase celular, atuando como chaperonas, transportando peptídeos, participando do dobramento e desdobramento de proteínas e prevenindo a agregação protéica (LINDQUIST, 1986). A maioria das hsps é expressa constitutivamente, mas sua síntese é aumentada nas células de procariotos e eucariotos, quando estas são submetidas a algum tipo de estresse térmico ou químico (LINDQUIST, 1998).

A síntese de hsp é aumentada para proteger células eucarióticas e procarióticas de possíveis danos durante períodos de estresse causados por infecções, inflamações ou eventos similares. As hsps constitutivas representam até 5% do conteúdo protéico de uma célula normal. Elas encontram-se entre as proteínas mais conservadas filogeneticamente e são agrupadas em famílias conforme seus pesos moleculares e a homologia de suas seqüências sendo caracterizadas como: família hsp110, hsp90, hsp70, hsp60 e a família de baixo peso molecular como a hsp18 e hsp14 (LINDQUIST, 1998).

Elas estão entre os antígenos imunodominantes encontrados nas micobactérias, principalmente as proteínas de 70, 65, 18 e 14 KDa. Em camundongos infectados com tuberculose, de 10 a 20% de todas as células T respondem especificamente para o antígeno hsp65 (KAUFMANN, 1987). A proteína hsp65 de *M.leprae* tem recebido maior atenção devido ao amplo envolvimento com células e mediadores do sistema imune nas infecções micobacterianas (SILVA, 1994).

As funções imunológicas das hsps começaram a ser relatadas nos anos 80 quando observou-se que preparações homogêneas de certas hsps isoladas de células tumorais induziam resposta contra câncer, enquanto preparações correspondentes de tecidos normais não o faziam (SRIVASTAVA, 1998). A imunogenicidade das hsps derivadas de tumor resultava da associação das

mesmas com peptídeos gerados pela degradação de proteínas expressas por células das quais as hsps eram purificadas. Dessa forma, as hsps derivadas de células tumorais estavam não-covalentemente associadas a peptídeos derivados de antígenos tumorais, do mesmo modo que as hsps provenientes de tecidos não tumorais apresentavam-se complexadas com peptídeos normais e por isso não eram capazes de induzir resposta antitumoral específica (SRIVASTAVA, 2002)(SANTOS, 2008).

1.6. Bioinformática

A bioinformática é a combinação de conhecimentos estatísticos, biológicos e informáticos, que tem por objetivo representar, armazenar e distribuir dados, além de desenvolver e utilizar ferramentas para revelar os conhecimentos que podem ser obtidos através desses dados.

O começo da bioinformática se deu nos anos 60, quando as primeiras seqüências de proteínas foram coletadas por Margaret Dayhoff e colaboradores, construindo as matrizes de substituição PAM, utilizadas para o alinhamento de seqüências protéicas. Nos anos 70 surgem os programas para alinhamento de seqüências. O número de seqüências biológicas cresce exponencialmente passando a serem depositadas em bancos de dados, principalmente as seqüências de DNA. Na década de 80 dá-se início as buscas por similaridades entre as seqüências, sendo o primeiro programa desenvolvido, o FASTA (MOUNT, 2001). Atualmente com a crescente quantidade de informações geradas pelo projeto genoma, assim como muitos outros, a bioinformática torna-se uma importante área para explorar todos os conhecimentos provenientes de tais informações.

Na área protéica, o estudo das estruturas tridimensionais já vem sendo realizado há mais de três décadas. O conhecimento da estrutura permite predizer a localização de sítios catalíticos e de interação, assim como auxiliam na identificação de alvos para os estudos de metagênese sítio-dirigida. A modelagem protéica pode ser realizada de forma computacional, obtendo assim mais rapidamente resultados dificilmente obtidos pela cristalografia.

A área relacionada aos estudos do RNA ainda é bem recente, mas já se dispõe de mecanismos rápidos e eficientes para a determinação da estrutura secundária do RNAm.

1.7. Bioinformática

O objetivo geral deste projeto é utilizar ferramentas de bioinformática para modelagem molecular por homologia da chaperonina hsp65 de *M. leprae*.

Como objetivo específico tem-se:

- a determinação da estrutura secundária do RNA mensageiro correspondente.

Capítulo 2: Métodos

2.1. Modelagem do RNAm da HSP65 de *M.leprae*

As maiorias dos RNAs são fita simples, mas podem apresentar estrutura secundária através do emparelhamento entre pares de bases. Tal emparelhamento ocorre principalmente entre as bases G e C, A e U, e ocasionalmente, entre G e U (MACHADO-LIMA, 2008). O objetivo principal dessa formação é proteger algumas bases de moléculas de água. Algumas bases não formam pares originando laços (*loops*) ou bulbos (*bulges*).

Dada uma seqüência de RNA, o número de possíveis estruturas secundárias cresce exponencialmente conforme o tamanho da seqüência. Assim, para realizar a predição da estrutura secundária há diversos programas que utilizam diferentes métodos. No estudo em questão foi utilizado o *MFOLD* (ZUKER, 2003), que tem como método o modelo de mínima energia livre.

O algoritmo de mínima energia livre tem por objetivo obter a estrutura mais estável da molécula de RNA, analisando-se o resultado da energia da molécula depois de interações par a par, entre os vizinhos mais próximos. Quanto mais negativa for a energia de uma estrutura, mais provável é a dessa formação da estrutura secundária.

Alguns parâmetros para a formação de uma estrutura secundária estável são: a maior quantidade do par CG, que gera uma maior estabilidade; o número de bases não pareadas, que resultará numa menor estabilidade; o número de bases na região de grampos (*hairpins*), já que *loops* com mais de 10 bases e menos de 5 requerem mais energia; e o número de bases numa região fita, já que longas fitas resultam em mais ligações.

2.2. Modelagem da proteína hsp65 de *M.leprae*

A estrutura terciária de uma proteína determina a função que esta irá exercer. Sendo assim, a modelagem computacional pode ser empregada na predição de estruturas tridimensionais de proteínas.

Quando uma proteína apresenta um alto grau de similaridade com outra proteína que já possui uma estrutura tridimensional conhecida pode-se usar a modelagem por homologia (GIBAS e JAMBECK, 2001). Tal abordagem se baseia no conceito de evolução molecular, ou seja, a semelhança nas estruturas primárias das proteínas resulta em similaridade estrutural entre elas (MOUNT, 2001).

Para realizar uma modelagem por homologia o primeiro passo é a identificação e a seleção de proteínas homólogas através de buscas em banco de dados. Em seguida, deve-se alinhar a proteína de estudo com a proteína modelo, sendo essa fase parte fundamental do processo de modelagem (GIBAS e JAMBECK, 2001). A partir do alinhamento constrói-se o modelo da proteína em questão. Tal modelo deverá ser validado, através de softwares de validação estereoquímica. Na Figura 3 pode-se observar um esquema para melhor compreensão das etapas acima mencionadas.

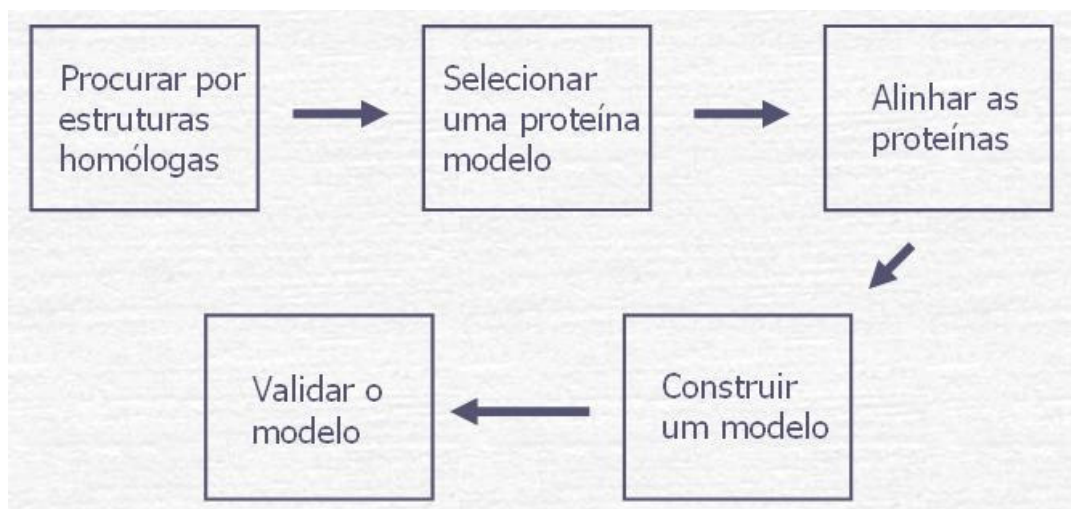


Figura 3: Esquema que demonstra os passos a serem seguidos por uma modelagem por homologia.

O software escolhido para realizar a modelagem da proteína de estudo foi o *Modeller* (ESWAR, 2000), tal escolha se deve ao fato de muitos estudos comprovarem sua eficácia. O *Modeller* faz uso do Python como linguagem de programação, não incluindo interface gráfica. É escrito em Fortran 77, sendo distribuído gratuitamente para instituições acadêmicas ou sem fins lucrativos.

Para iniciar análises da proteína, o DNA atualmente utilizado pelo laboratório na vacina de DNAhsp65 contra tuberculose foi seqüenciado. Com os dados do seqüenciamento, um alinhamento local através do software BLAST (Basic Local Alignment Sequence Tool) (ALTSCHUL, 1990) foi realizado encontrando como resultado a seqüência de código M14341, correspondente a proteína de choque térmico de 65 kDa (hsp65) do *M.leprae*, como era o esperado. O seqüenciamento foi realizado visando garantir as análises futuras, excluindo qualquer possibilidade de mutação garantindo a determinação da real seqüência da proteína a ser modelada.

Foi realizado um BLAST contra o banco de dados do Swiss-prot, para obter a seqüência a ser modelada. A seqüência deverá corresponder a parte codificadora do gene, ou seja, a parte que é o resultado da tradução do RNAm. Obteve-se como resultado a seqüência de código P09239 correspondente a chaperonina de 65 Kda do *M.leprae*.

Já era de conhecimento do laboratório a alta similaridade compartilhada entre a hsp65 de *M.leprae* e a hsp65 de *M.tuberculosis*. Visando conhecer um pouco mais a similaridade entre essas duas proteínas realizou-se o estudo de seus domínios. Os domínios formam unidades estruturais e funcionais semi-independentes dentro de uma única cadeia polipeptídica. Os domínios podem ter uma função específica, e quando tal função é necessária em muitas proteínas diferentes, é encontrada a mesma estrutura de domínio.

A procura dos domínios das proteínas foi realizada através dos programas CDSearch (MARCHLER-BAUER, 2004) e Pfam (FINN, 2006). Ambos os programas encontraram o mesmo domínio para as duas proteínas, o Cpn60_TCP1. Enquanto o Pfam encontrou do resíduo 22 até o 522, o CDSearch encontrou do 22 até o 519. Este domínio está relacionado à família das

chaperoninas, a qual inclui as hsp60 e hsp65. Os resultados apresentados confirmaram as funções das proteínas estudadas, tanto de proteína ligadora de ATP, quanto de proteína ligadora de proteína, características clássicas de chaperoninas.

Capítulo 3: Resultados e Discussão

3.1. Modelagem do RNAm da HSP65 de *M.leprae*.

A estrutura secundária do RNAm do gene HSP65 de *M.leprae* foi determinada através do software MFOLD, como mencionado anteriormente. Modelou-se a seqüência completa do gene HSP65 *M.leprae*, assim como a seqüência aberta de leitura (ORF), correspondente aos primeiros 1500 pb. As duas seqüências foram analisadas, pois ambas são utilizadas em protocolos vacinais atualmente utilizados no laboratório.

Os resultados obtidos no modelamento estrutural demonstraram que há uma diferença considerável entre os dois RNAs modelados, tanto na estrutura quanto no valor da energia livre (Figura 4), sendo que o pre-RNA-HSP65 libera uma quantidade maior de energia quando comparado com o mRNA-HSP65 (dG: - 1511 Kcal/mol e dG: - 682 Kcal/mol, respectivamente). O comprimento da seqüência dos RNAm influencia diretamente na estabilidade, verifica-se assim que o mRNA-HSP65 possui uma estabilidade maior do que o pre-RNA-HSP65, já que este possui 3000 pb.

Essa avaliação é importante, pois nessa região podem existir estruturas moleculares como *hairpins* e *loops* que podem ter influência crucial no controle de expressão gênica pós-transcricional (KOZAK, 1986), (HELKE, 1993), (BABENDURE, 2006). Essa influência depende do tamanho, da estabilidade e também da posição dessas estruturas em relação ao sítio de iniciação da tradução. Como pode ser visualizado na Figura 5, a estrutura tem um dG -17,20 e apenas um *hairpin* que está distante 18 bases do sítio de iniciação da tradução. Esse dado pode ser considerado interessante, pois Babendure e colaboradores (2006) comentam e comprovam que valores de dG maiores que -30 têm relação com uma diminuição da expressão protéica no modelo de GFP.

A descoberta de que RNA interagem com células eucarióticas, ativando células do sistema imune, e modulando a resposta vieram primeiramente de

estudos virais. Atualmente, sabe-se que o próprio RNAm eucarioto interage com receptores do tipo Toll 7 e 8, levando a ativação celular, com produção de mediadores inflamatórios ou citocinas que podem tanto controlar uma infecção como causar maiores danos.

O uso de RNA total já é bem documentado no controle de tumores, assim, sabe-se que a ligação do RNA em células alvo leva produção de citocinas do padrão Th1, como IFN-gama e IL-12, envolvidas na resistência contra tumores. Tais citocinas também estão envolvidas na resistência contra reações inflamatórias crônicas, como a tuberculose.

A

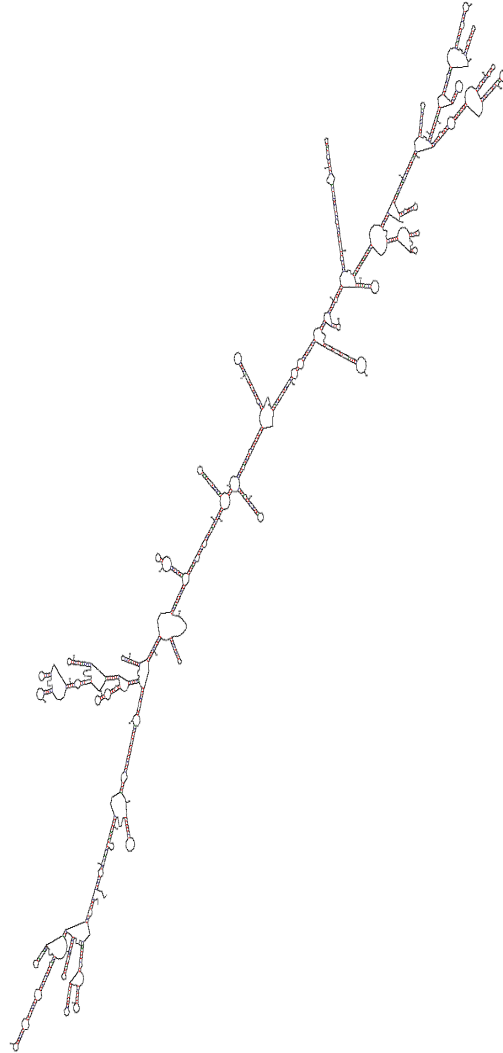
Output of sir_graph
by D. Stewart and M. Zuker



dG = -1511.19 [initially -1604.30] hsp_RNA_first

B

Output of sir_graph
by D. Stewart and M. Zuker



dG = -682.13 [initially -731.40] RNA_seq541aa_30/01

Figura 4: Análise da estrutura molecular dos RNAs mensageiros codificando a proteína Hsp65 de Mycobacterium leprae (A: pre-RNA-HSP65 e B: mRNA-HSP65) . A análise da estrutura molecular dos RNAs mensageiros foi realizada utilizando-se o programa Mfold. dG: energia liberada em Kcal/mol.

3.2. Modelagem da proteína hsp65 de *M.leprae*

3.2.1. Escolha das proteínas moldes

Através do script `buil.profile.py` foi obtido as proteínas que poderiam servir de molde para a modelagem da proteína de estudo. As proteínas encontradas são mostradas na Figura 6 no primeiro campo em destaque (código do PDB). Como se pode notar algumas delas possuem uma baixa porcentagem de similaridade, o que resultaria em um molde ruim para a modelagem. Geralmente, trabalha-se com uma taxa de similaridade superior a 60%, englobando ainda seis possíveis proteínas moldes.

```
Dynamically allocated memory at amaxprofile [B,KiB,MiB]: 7681574 7501.537 7.326
> 1a6dA 1 202 15850 503 541 27.86 0.0 2 456 9 51 5 503
> 1a6dB 1 203 15850 502 541 25.51 0.0 3 449 21 51 13 502
> 1kp8A 1 587 75550 525 541 59.16 0.0 4 522 2 52 2 525
> 1dk7A 1 2268 22700 146 541 61.38 0.0 5 145 189 33 1 145
> 1kid 1 3561 30100 193 541 62.69 0.0 6 193 182 37 1 193
> 1iokA 1 4727 65700 482 541 58.63 0.0 7 478 2 52 2 482
> 1la1A 1 5798 30750 192 541 63.54 0.0 8 192 186 37 1 192
> 1q3qA 1 7268 16700 518 541 29.94 0.0 9 455 22 52 27 517
> 1s1pA 1 7930 102200 447 541 95.53 0.0 10 447 61 51 1 447
> 1srvA 1 8019 24300 145 541 68.06 0.0 11 144 190 33 1 144
> 1we3A 1 9005 83550 527 541 64.71 0.0 12 523 2 52 1 527
```

Figura 6: Trecho do script `buil_profile.py`, que mostra em destaque o código PDB das possíveis proteínas a serem usadas como molde, o score de similaridade com proteína de estudo e o tamanho da proteína, respectivamente.

Confirmando o conhecimento do laboratório sobre a alta similaridade entre a hsp65 de *M.tuberculosis* e *M.leprae* – 1SJP (código no PDB), optou-se por realizar a modelagem com essa proteína.

Realizado o alinhamento entre as duas proteínas verificou-se que tanto a parte N-terminal, como a C-terminal da hsp65 de *M.leprae* não conseguiu obter um bom alinhamento. Tal fato se deve a falta de dados da proteína molde, assim o *Modeller* não encontraria os parâmetros necessários para um bom enovelamento.

A Figura 7 corresponde a um alinhamento entre as proteínas realizado pelo ClustalW (D.HIGGINS, 1994), podendo-se comprovar a alta similaridade entre as duas proteínas. Aparece em destaque nas seqüências das proteínas o domínio Cpn60_TCP1. O ClustalW utiliza a Matrix Gonnet 250 pela sua sensibilidade. As penalidades para os *gaps* são principalmente os seguintes: para a abertura do

gap, 10.0; por estender um *gap*, 0.2 e pela separação entre *gaps*, 4.0. O *score* obtido no alinhamento apresentado foi de 93, comprovando a alta similaridade entre as proteínas envolvidas. Como pode ser observado, as partes iniciais e finais não obtiveram um bom alinhamento.

```

CLUSTAL W (1.83) multiple sequence alignment

hsp65      MAKTIAYDEEARRGLERGLNSLADAVKVTLLGPKGRNVVLEKKWCAPTIITNDGVSIKKEIE 60
lsjp      -----WCAPTIITNDGVSIKKEIE 18
                *****

hsp65      LEDPYEKIGAEIVKREVAKRITDDVAGDCTTTATVLAQALVKEGLRNVAAGANPLGLKRCIE 120
lsjp      LEDPYEKIGAEIVKREVAKRITDDVAGDCTTTATVLAQALVREGLRNVAAGANPLGLKRCIE 78
                *****

hsp65      KAVDRVTETLLKDAKEVETREQIAATAAISAGDQSIGDLIAEAMDKVGNECVITVRESMT 180
lsjp      KAVDRVTETLLKCAKEVETREQIAATAAISAGDQSIGDLIAEAMDKVGNECVITVRESMT 138
                ***.*****.*****

hsp65      FGLQLELTGCMRFDKCYISCYFVTD AERQEA VLEEPYILLWSSKRVSTVKDLLP LLEKVIQ 240
lsjp      FGLQLELTGCMRFDKCYISCYFVTD PERQEA VLEDPYILLWSSKRVSTVKDLLP LLEKVIQ 198
                *****

hsp65      ACKSLILIAEDVEGEALSTLVVNKIRCTFRKSVAVKAPGFCDRRKAMLDMAILTCAQVIS 300
lsjp      ACKPLILIAEDVEGEALSTLVVNKIRCTFRKSVAVKAPGFCDRRKAMLDMAILTGGQVIS 258
                ***.*****.*****

hsp65      EEVGLTLENADLSLLCKARKVVMTRDETTIVEGAGD TDAIACRVAQIRTEIENS DSDYDR 360
lsjp      EEVGLTLENADLSLLCKARKVVMTRDETTIVEGAGD TDAIACRVAQIRQEIENS DSDYDR 318
                *****

hsp65      EKLQERLAKLAGCVAVIKACAAT EVELKERKHRIEDAVRNAKA AVEECIVAGCGVTL LQA 420
lsjp      EKLQERLAKLAGCVAVIKACAAT EVELKERKHRIEDAVRNAKA AVEECIVAGCGVTL LQA 378
                *****

hsp65      APALDKLKLTCDEATGANIVKVALEAPLQKIAFN SCMEPGVVAEKVRNLSVCHGLNAA TC 480
lsjp      APTLDELKLEGEATGANIVKVALEAPLQKIAFN SGLPEGVVAEKVRNLPAGHGLNAA TC 438
                **.*.*** *****.*****.*****

hsp65      EYEDLLKAGVADPVKVT RSALQNAAS IAGLFLT ERTAAP ASDPTGCMGGM DTEAVVADPK 540
lsjp      VYEDLLAAGVADPVKVT RSALQNAAS IAGLFLT TEAVVADKPEKEKASVPC -CGDMGGM D 497
                *****

hsp65      F----- 541
lsjp      FHHHHH 504
                *

```

Figura 7: Alinhamento da hsp65 de *M.leprae* e da hsp65 de *M.tuberculosis*, usando o programa ClustaW.

Para que tais partes conseguissem encontrar uma boa base para a modelagem, optou-se por realizar a modelagem utilizando um alinhamento múltiplo entre duas proteínas. Continuou-se a utilizar a proteína 1SJP devido a sua alta similaridade com a proteína de estudo e para a melhor modelagem da parte N-terminal utilizou-se a Estrutura Cristalográfica do Complexo Chaperonina Cpn60/Cpn10/(ADP)7 da *Termus Thermophilus* – 1WE3 (código no PDB).

Descobriu através de buscas no bando de dados do PDB que a cadeia A da proteína 1WE3 é similar a hsp65 de *M.leprae* e contém a parte N-terminal para ser utilizada como parâmetro para o *Modeller*. Tal abordagem serviu também para verificar se os não enovelamentos das partes iniciais e finais da proteína era uma característica de micobactérias ou não. Já que, os dados obtidos mostravam que a

cauda amino e carboxi terminal poderia ser uma característica de hsp imaturas de procariotos.

O alinhamento obtido pelo ClustalW entre a proteína de estudo e a 1WE3 pode ser observado na Figura 8. Utilizaram-se os mesmos parâmetros que no alinhamento anterior, e o *score* obtido foi de 63, o que ainda demonstra uma boa similaridade entre as proteínas. Novamente o domínio Cpn60_TCP1 aparece em destaque.

```

CLUSTAL W (1.83) multiple sequence alignment

hsp65      MAKTIAYDEEARRGLERGLNSLADAVKVTLGPKGRNVVLEKKWGAPTIITNDGVSTAKEIE 60
1we3       MAKILVFDAAARRALERGVNAVAVKVTLGPKGRNVVLEKKFGSPITITKDGVTVAKEVE 60
***      .:.** ***,***:..*:*****:*****:*.***:***:***:*

hsp65      LEDPYEKIGAELVKEVAKKTDDVAGDGTTTATVLAQALVKEGLRNVAAAGANPLGLKRGIE 120
1we3       LEDHLENIQAQLLKEVASKTNDVAGDGTTTATVLAQAIVREGLRNVAAAGANPLALKRGIE 120
***      *:***:*.***:*.***:*****:*****:*.***:*****:*****

hsp65      KAVDKVTE TLLKDAKEVETKEQIAATAAISAGDQSIGDLIAEAMDKVGNEGVITVEESNT 180
1we3       KAVEAAVEKIKALAIIPVEDRKAIEEVATISANDPEVGKLIADAMEKVGKEGIIITVEESKS 180
***      ..*.: * ** :. * .*:*** * .:*.***:*.***:*.***:*****:..

hsp65      FGLQLELTEGMRFDKGYISGYFVTD AERQEA VLEEPYILLVSSKVSIVKDLLPLEKVIQ 240
1we3       LETELKFVEGYQPDFKGYISP YFVTINPETMEAVLEDAFILIVEKRVSNVRELLP ILEQVAQ 240
:      :*:..** :***** ***:..* *****:..**:*..***:*.***:***:* *

hsp65      AGKSLIIAEDVEGEALSTLVVWKIRGTFKSVAVKAPGFGDRRKAMLQDMAILTGAQVIS 300
1we3       TGRPLIIAEDVEGEALATLVVWKLRTLSVAAVKAPGFGDRRKEMLKDIAAVTGGTVIS 300
:*.*****:*****:***:..***** **:* * :*. ***

hsp65      EEVGLTLENTDLSLLGKARKVVMTKDETTIVEGAGD TDAIAGRVAQIRTEIENSDDYDR 360
1we3       EELGFLENATLSMLGRAERVRITKDETTIVGKGGKEDIEARINGIKKELETIDSEYAR 360
**:*:***: **:*:*.:* :***** * ..: * .*: * .:*.***:***:*

hsp65      EKLQERLAKLAGGVAVIKAGAAATEVELKERKHRIEDAVRNAKAAVEEGIVAGGGVTLQ 420
1we3       EKLQERLAKLAGGVAVIRVGAATETELKEKHRFEDALNATRAAVEEGIVPGGGVTLR 420
*****:*****:***:***:..*****:*****:***

hsp65      APALDKL--KLTGDEATGANIVKVALEAPLKQIAFN SGMPEGVVAEKVRNLSVGL--HGLN 476
1we3       ISAVEELIKKLEGDEATGAKIVRRAL EEPARQIAENAGYEGSVIVQQILAE TKNPRYGFN 480
.:*:.* ** *****:***: *** * :*** *:* * .*:..: : . :*:.*

hsp65      AATGEYEDLLKAGVADPVKVTRSALQNAASIAGLFLTTEAVVADKPEKTAAPASDPTGGM 536
1we3       AATGEFVDMVEAGIVDPKAVTRSA LQNAASIGALILTEAVVAEKPEKES--TPASAGA 538
*****: *::**:.** *****:*****:***:..*****:***: : ..*

hsp65      GGMDF 541
1we3      GDMDF 543

*.*.*

```

Figura 8: Alinhamento da hsp65 de *M.leprae* e da 1WE3, usando o programa ClustaW.

3.2.2. Alinhamento da 1SJP com a 1WE3

O alinhamento foi realizado pelo *Modeller* através do comando `salign()`, inserido no script `salign.py`, demonstrado na Figura 9. Será realizado o alinhamento entre essas duas proteínas para que futuramente possa ser realizado o alinhamento com a hsp65 de *M.leprae*.

O script tem como saída o alinhamento em dois formatos, no PIR, que será utilizado mais tarde por outro script de entrada, e no PAP, para melhor visualização. Além disso, gera uma proteína consenso, resultante da junção das duas proteínas utilizadas para fazer o alinhamento múltiplo.

```
aln = alignment(env)
for (code, chain) in (('1sjp', 'A'), ('1we3', 'A')):
    mdl = model(env, file=code, model_segment=('FIRST:'+chain, 'LAST:'+chain))
    aln.append_model(mdl, atom_files=code, align_codes=code+chain)

for (weights, write_fit, whole) in (((1., 0., 0., 0., 1., 0.), False, True),
                                     ((1., 0.5, 1., 1., 1., 0.), False, True),
                                     ((1., 1., 1., 1., 1., 0.), True, False)):
    aln.salign(rms_cutoff=3.5, normalize_pp_scores=False,
               rr_file='${LIB}/as1.sim.mat', overhang=30,
               gap_penalties_1d=(-450, -50),
               gap_penalties_3d=(0, 3), gap_gap_score=0, gap_residue_score=0,
               dendrogram_file='fm00495.tree',
               alignment_type='tree', # If 'progressive', the tree is not
                                     # computed and all structures will be
                                     # aligned sequentially to the first
               feature_weights=weights, # For a multiple sequence alignment only
                                     # the first feature needs to be non-zero
               improve_alignment=True, fit=True, write_fit=write_fit,
               write_whole_pdb=whole, output='ALIGNMENT QUALITY')
```

Figura 9: Trecho do script `salign.py`, que mostra em destaque a parte que utiliza o comando `salign()`.

A Figura 10 apresenta um trecho do arquivo PAP gerado pelo script `salign.py`. Pode-se notar que as seqüências escolhidas se complementam, resultando numa seqüência consenso que contém uma maior homologia com a proteína de estudo.

```

|_aln.pos      10      20      30      40      50      60
1sjpA      -----LEDPYEKIGAE-----
1we3A      AKILVFD--EAARRALERGVNAVAVKVTLGPRGRNVVLEKKFGSPTITKDGVTVAKEVELEDHLEN
_consrvd

|_aln.p      70      80      90      100     110     120     130
1sjpA      LVKEVAKKT-----TTATVLAQALVREGLRNVAAGANPLGLKRGIEKAVEKVTETLLKGAK
1we3A      IGAQLLKEVASKTNDVAGDGTTTATVLAQAIVREGLKNVAAGANPLALKRGIEKAVEAAVEKIKALAI
_consrvd      *                ***** * * * * * * * * * * * * * * * * * * * * * *

|_aln.pos     140     150     160     170     180     190     200
1sjpA      EVETKEQIAATAAISAGDQSIGDLIAEAMDKVGNEGVIITVEESNTFG-LQLELTEGMRFDKGYISGYF
1we3A      PVEDRKAIEEVATISANDPEVGKLIADAMEKVGKEGIIITVEESKSLETTELKFFVEGYQ-----
_consrvd      ** * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

|_aln.pos     210     220     230     240     250     260     270
1sjpA      VTDPERQEAVLEDP-----
1we3A      -----FDKGYISPYFVTNPETMEAVLEDAFILIVEKKVSNVRELLPILEQVAQTGKPLLIIEA
_consrvd

|_aln.pos     280     290     300     310     320     330     340
1sjpA      LVSSKVSTVKDLLPLLEKVI-----AGKPLLIIEAEDVEGEALSTLVVNKIRGTFK
1we3A      -----DVEGEALATLVVNLKRGTL SVA AVKAPGFGD-----
_consrvd      * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

|_aln.pos     350     360     370     380     390     400
1sjpA      SVAVKAPGFGDRRKAMLDMAILTG----GQVISEEVGLTLENADLSLLG-----
1we3A      -----RRKEMLKDIAAVTGGTVISEELGFK-----LENATLSMLGRAERVIT
_consrvd      * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

|_aln.p      410     420     430     440     450     460     470
1sjpA      --KARKVVTKDETTIVEGAGDTDAIAGRVAQIRQEIENSDDYDREKLQER-----
1we3A      KDETT-----IVGGKGGKEDI EARIN-----GIKKELETTDSEYAREKLQER
_consrvd      * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

|_aln.pos     480     490     500     510     520     530     540
1sjpA      LAKLAGGVAVIKAGAATEVELKERKHRIEDAVRNAKAAVEEGIVAGGGVTL LQAAPTLDL--KLEGD
1we3A      LAKLAGGVAVIRVGAATELTELKHKHRFEDALNATRAAVEEGIVPGGGVTL LRAI SAVEELIKKLEGD
_consrvd      ***** * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

|_aln.pos     550     560     570     580     590     600     610
1sjpA      EATGANIVKVALEAPLKQIAFNSGLEPGVVAEKVRNLP--AGHGLNAQTGVYEDLLAAGVADPVKVTR
1we3A      EATGAKIVRRALEEPARQIAENAGYEGSVIVQQILAETKNPRYGFNAATGEFVDMVEAGIVDPAKVTR
_consrvd      ***** * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

|_aln.pos     620     630
1sjpA      SALQNAASIAGLFLTTE-----
1we3A      SALQNAASIGALILTTEAVVAEKPEK
_consrvd      ***** * * * * *

```

Figura 10: Alinhamento da 1SJP e da 1WE3, através do script `salign.py`

3.2.3. Alinhamento da hsp65 de M.leprae com o modelo resultante do alinhamento múltiplo

O primeiro passo para a construção do modelo é o alinhamento entre a proteína de estudo e a proteína molde. Neste caso a proteína de estudo será alinhada com a proteína consenso obtida na fase anterior.

O alinhamento foi novamente através do comando `salign()`, inserido no script `align2d.py`, demonstrado na Figura 11.

O script tem como saída o alinhamento em dois formatos, no PIR, que será utilizado mais tarde por outro script de entrada, e no PAP, para melhor visualização.

```
aln = alignment(env)
for (code, chain) in (('1s1p', 'A'), ('1we3', 'A')):
    mdl = model(env, file=code, model_segment=('FIRST:'+chain, 'LAST:'+chain))
    aln.append_model(mdl, atom_files=code, align_codes=code+chain)
for (weights, write_fit, whole) in (((1., 0., 0., 0., 1., 0.), False, True),
                                     ((1., 0.5, 1., 1., 1., 0.), False, True),
                                     ((1., 1., 1., 1., 1., 0.), True, False)):
    aln.salign(rms_cutoff=3.5, normalize_pp_scores=False,
              rr_file='${LIB}/as1.sim.mat', overhang=30,
              gap_penalties_1d=(-450, -50),
              gap_penalties_3d=(0, 3), gap_gap_score=0, gap_residue_score=0,
              dendrogram_file='fm00495.tree',
              alignment_type='tree', # If 'progresive', the tree is not
                                     # computed and all structures will be
                                     # aligned sequentially to the first
              feature_weights=weights, # For a multiple sequence alignment only
                                     # the first feature needs to be non-zero
              improve_alignment=True, fit=True, write_fit=write_fit,
              write_whole_pdb=whole, output='ALIGNMENT QUALITY')
```

Figura 11: Trecho do script `salign2d.py`, que mostra em destaque a parte que utiliza o comando `align2d()`.

A Figura 12 apresenta um trecho do arquivo PAP gerado pelo script `salign2d.py`. Pode-se notar que a parte N-terminal (em destaque na Figura) encontrou identidade com a proteína 1WE3, dessa forma o *Modeller* encontrará parâmetros para um bom enovelamento.

```

aln.pos      10      20      30      40      50      60
1sjpa      -----LEDPYEKIGAE-----
1we3A      AKILVFD--EAARRALERGVNAVAVKVTLGPRGRNVVLEKKFGSPTITKDGVTVAKEVELE
hsp65Comp1eta  MAKTI-AYDEEARRGLERGLNSLADAVKVTLGPKGRNVVLEKKWGAPTITNDGVSIACEIELE
_consrvd

aln.pos      70      80      90      100     110     120
1sjpa      -----LVKEVAKKT-----TTATVLAQALVREGLRNVAAGANPLGLKRGIEKAVEK
1we3A      DHLENIGAQLLKEVASKTNDVAGDGTATVLAQAIIVREGLKNVAAGANPLALKRGIEKAVEA
hsp65Comp1eta  DPYEKIGAEVLVKEVAKKTDVAGDGTATVLAQALVKEGLRNVAAGANPLGLKRGIEKAVDK
_consrvd      *          ***** * * * * *

aln.pos      130     140     150     160     170     180
1sjpa      VTETLLKGAKEVETKEQIAATAAISAGDQSIGDLIAEAMDKVGNEGVIITVEESNTFG-LQLEL
1we3A      AVEKIKALAI PVDRKAIIEEVATISANDPEVGKLIADAMEKVGKEGIITVEESKSLETCLKFV
hsp65Comp1eta  VTETLLKDAKEVETKEQIAATAAISAGDQSIGDLIAEAMDKVGNEGVIITVEESNTFG-LQLEL
_consrvd      *          * * * * * * * * * * * * * * *

aln.pos      190     200     210     220     230     240     250
1sjpa      TEGMRFDKGYISGYFVTDPERQEAILEDP-----
1we3A      EGYQ-----FDKGYISPYFVTNPETMEAVLEDAFILIVEKKVSNVRE
hsp65Comp1eta  TEGMRFDKGYISGYFVTDAERQEAILEEPPY-----ILLVSSKVS-----
_consrvd

aln.pos      260     270     280     290     300     310
1sjpa      -----YILLVSSKVSTVKDLLPLEKLVIG-----AGKPL
1we3A      LLPILEQVAQTGKPLLIIEA-----DVEGEALATLVVNKLRGTL SVAAVKA
hsp65Comp1eta  -----TV---KDLLPLLE-----KVIQ-----AGKSL
_consrvd      *          *

aln.pos      320     330     340     350     360     370
1sjpa      LIIAEDVEGEALSTLVVNKIRGTFKSVAVKAPGFGDRRKAMLQDMAITG----GQVISEEVG
1we3A      PGFGD-----RRKEMLKDIAAVTGGT VI SEELG
hsp65Comp1eta  LIIAEDVEGEALSTLVVNKIRGTFKSVAVKAPGFGDRRKAMLQDMAITG----AQVISEEVG
_consrvd      *          * * * * *

aln.pos      380     390     400     410     420     430     440
1sjpa      LTLENADLSLLG-----KARKVVVTKDETTIVEGAGD TDAIAGRVAQI
1we3A      FK-----LENATLSMLGRAERVIRITKDETT-----IVGGKGGKEDI EARI
hsp65Comp1eta  LT-----LENTDLSLLGKARKVVMTKDETTIV-----EGAGD TDAIAGRVAQI
_consrvd      *          * *

aln.pos      450     460     470     480     490     500
1sjpa      RQEIENS DSDYDREKLQER-----LAKLAGGVAVIKAGAATEVELKERK HRI
1we3A      N-----GIKKELETTDSEYAREKLQERLAKLAGGVAVIRVGAATETELKEK HRF
hsp65Comp1eta  RTEIENS DSDYDREKLQER-----LAKLAGGVAVIKAGAATEVELKERK HRI
_consrvd      * * * * * * * * * * * * * * *

aln.pos      510     520     530     540     550     560
1sjpa      EDAVRNAKAAVEEGIVAGGGV TLLQAAPTLDL--KLEGDEATGANIVKVALEAPLKQ IAFNS
1we3A      EDALNATRAAVEEGIVPGGGV TLLRAI SAVEELIKKLEGDEATGAKIVRRAL EEPARQIAENA
hsp65Comp1eta  EDAVRNAKAAVEEGIVAGGGV TLLQAAPALDKL--KLTGDEATGANIVKVALEAPLKQ IAFNS
_consrvd      * * * * * * * * * * * * * * * * * * * * * * * * * * *

aln.pos      570     580     590     600     610     620     630
1sjpa      GLEPGVVAEKVRNLP--AGHGLNAQTGVYEDLLAAGVADPVKVTR SALSALQNAASIAGLFLTTE-
1we3A      GYEGSVIVQQILAETKNPRYGFNAATGEFVDMVEAGIVDPKAVTR SALSALQNAASIGALILTTEA
hsp65Comp1eta  GMEPGVVAEKVRNLS--VGHGLNAATGEYEDLLKAGVADPVKVTR SALSALQNAASIAGLFLTTEA
_consrvd      * * * * * * * * * * * * * * * * * * * * * * * * * * *

aln.pos      640     650
1sjpa      -----
1we3A      VVAEKPEK-----
hsp65Comp1eta  VVADKPEKTAAPASDPTGGMGGMDF
_consrvd

```

Figura 12: Alinhamento múltiplo entre a 1SJP, 1WE3 e a hsp65 de *M.leprae*, através do script `salgn2d.py`. Em destaque, a parte N-terminal da proteína.

3.2.4. Construção do modelo da hsp65 de *M.leprae*

Finalmente gerou-se um modelo da hsp65 de *M.leprae*, baseando-se no alinhamento múltiplo entre a 1SJP e a 1WE3. O modelo é gerado através do script `model-mult.py`, que usou-se a classe `automodel`.

Um trecho do script é mostrado na Figura 13. Foram gerados dez modelos, todos no formato PDB.

```
env = environ()
a = automodel(env, alignfile='hsp65Completa-mult.ali',
              knowns=('1sjpA', '1we3A'), sequence='hsp65Completa')
a.starting_model = 1
a.ending_model = 10
a.make()
```

Figura 13: Trecho do script `model-mult.py`, que mostra em destaque a parte que utiliza a classe `automodel`.

Adotou-se como a melhor modelagem o modelo com menor valor da função do *Modeller*, apresentada na segunda linha do arquivo PDB. O modelo obtido pode ser visualizado na Figura 14.

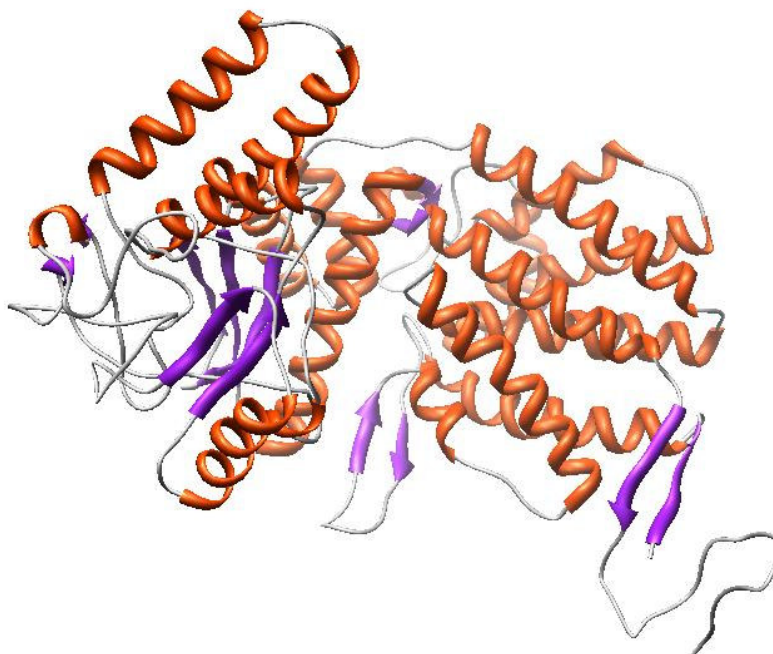


Figura 14: Modelo molecular da proteína hsp65 de *M.leprae* obtido pelo software *Modeller*, baseada no alinhamento múltiplo da 1SJP e da 1WE3.

3.2.5. Análise Estereoquímica

O nível de qualidade de um modelo gerado por homologia estrutural depende de um grande número de propriedades, como: exatidão estereoquímica, qualidade do empacotamento e confiabilidade do enovelamento. Os valores obtidos a partir dos modelos são comparados com os valores da proteína molde, descobrindo, assim, irregularidades estereoquímicas que revelam uma estrutura inadequada (SILVA, 2008).

A análise estereoquímica foi realizada utilizando-se três softwares: PROCHECK (LAKOWSKI, 1993), WHATIF (VRIEND, 1993) e VERIFY3D (BOWIE, 1991). Esses três softwares englobam as três principais propriedades a serem analisadas para a obtenção de um bom modelo.

O PROCHECK é responsável pela exatidão estereoquímica, avaliando parâmetros como: ângulos torcionais da cadeia principal, ângulos torcionais das cadeias laterais, maus contatos (ou impedimentos estéricos), energias das ligações de hidrogênio, planaridade das ligações peptídicas, entre outros. Há um valor calculado que representa uma qualidade estereoquímica média relativa aos parâmetros avaliados, chamado Fator G.

O programa gera vários gráficos, avaliando por completo a qualidade estereoquímica do modelo em comparação com estruturas bem definidas no mesmo nível de resolução. O principal gráfico gerado é o de Ramachandran que apresenta uma correlação entre os ângulos torcionais da cadeia principal (ϕ e ψ) para cada resíduo.

O gráfico de Ramachandran pode ser observado na Figura 15. A estrutura deve apresentar acima dos 90% de seus resíduos nas regiões favoráveis (região vermelha), ou seja, nas regiões A, B e L. Devem-se desconsiderar os resíduos de glicina (apresentados na figura como triângulos), já que eles não possuem cadeia lateral, os de prolina, pois nestes o Carbono alpha está ligado à cadeia lateral e os resíduos das extremidades (C-terminal e N-terminal), que apresentam padrões estereoquímicos diferentes dos outros resíduos (LAKOWSKI, 1993)(SILVA, 2008).

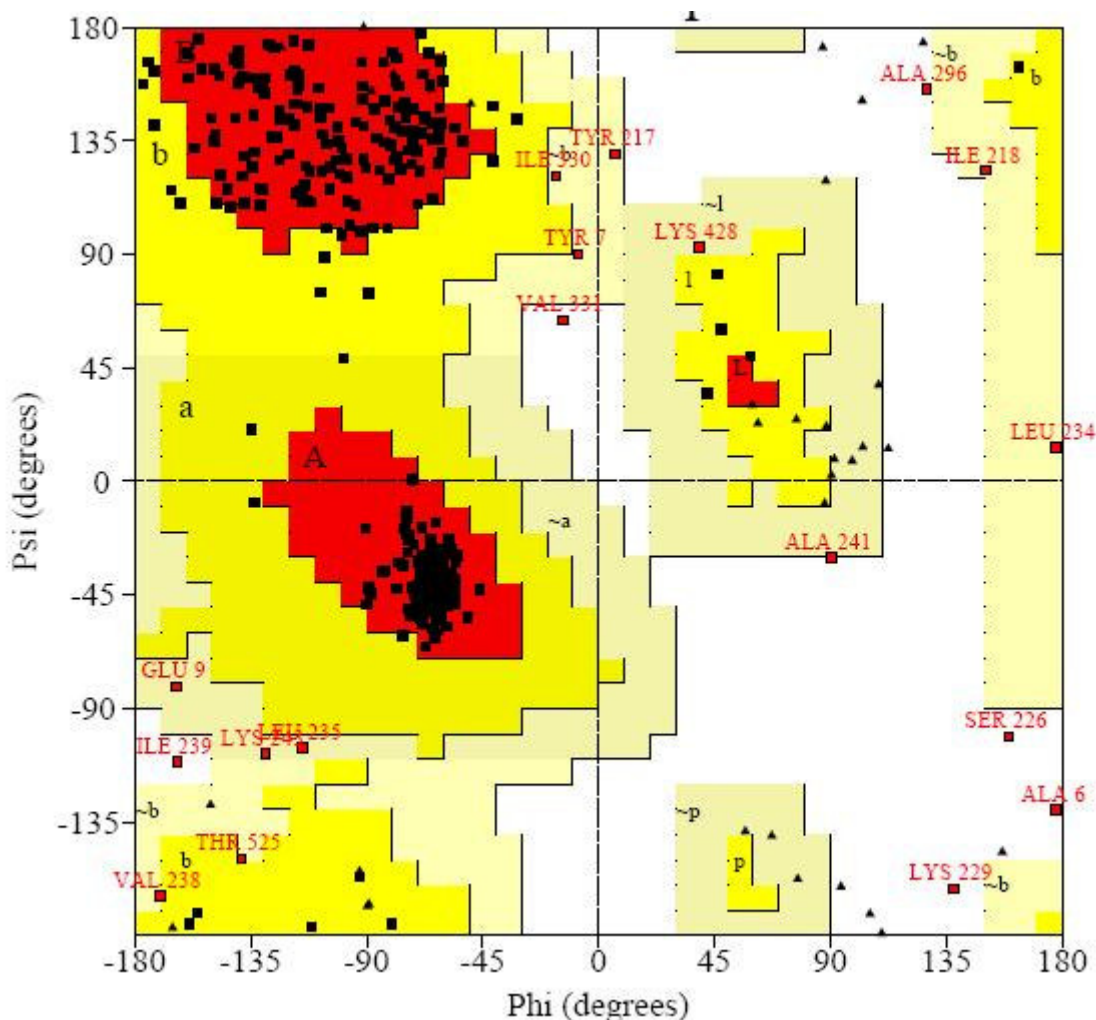


Figura 15: Gráfico de Ramachandran do modelo.

A porcentagem de resíduos encontrados nas regiões favoráveis foi de 90,7%, nas regiões adicionalmente permitidas (regiões a, b, l, p, ou seja, amarelo escuro) de 5,5%, nas regiões generosamente permitidas (regiões ~a, ~b, ~l, ~p, ou seja, amarelo claro) foi de 2,3% e nas regiões desfavoráveis (região em branco) foi de 1,5%. Os resíduos que se encontram em regiões desfavoráveis são um reflexo das estruturas usadas como molde, e alguns se encontram nas extremidades da proteína. Sendo assim, o gráfico revela uma boa validação do modelo.

Outro gráfico gerado pelo PROCHECK de extrema utilidade na avaliação de um modelo é o gráfico de parâmetros da cadeia principal, Figura 16. As figuras são compostas por bandas escuras em que a linha interior representa a média dos

valores em função da resolução, e as linhas das extremidades representam o desvio em relação à média. Os valores do modelo são marcados por quadrados. Os parâmetros avaliados são: a. Avaliação do gráfico de Ramachandran, b. Planaridade da ligação peptídica, c. Maus contatos atômicos, d. Distorção do carbono alpha, e. Energia de ligação de hidrogênios e f. Qualidade estereoquímica total (Fator G). O eixo vertical de todos os gráficos é a resolução da proteína em Angstroms, já o eixo horizontal é: a. porcentagem de resíduos em regiões favoráveis no gráfico de Ramachandran, b. ângulo de desvio da ligação peptídica, c. maus contatos atômicos a cada 100 resíduos, d. ângulo de desvio do carbono alpha, e. valor da energia em Kcal/mol e f. valor do fator G.

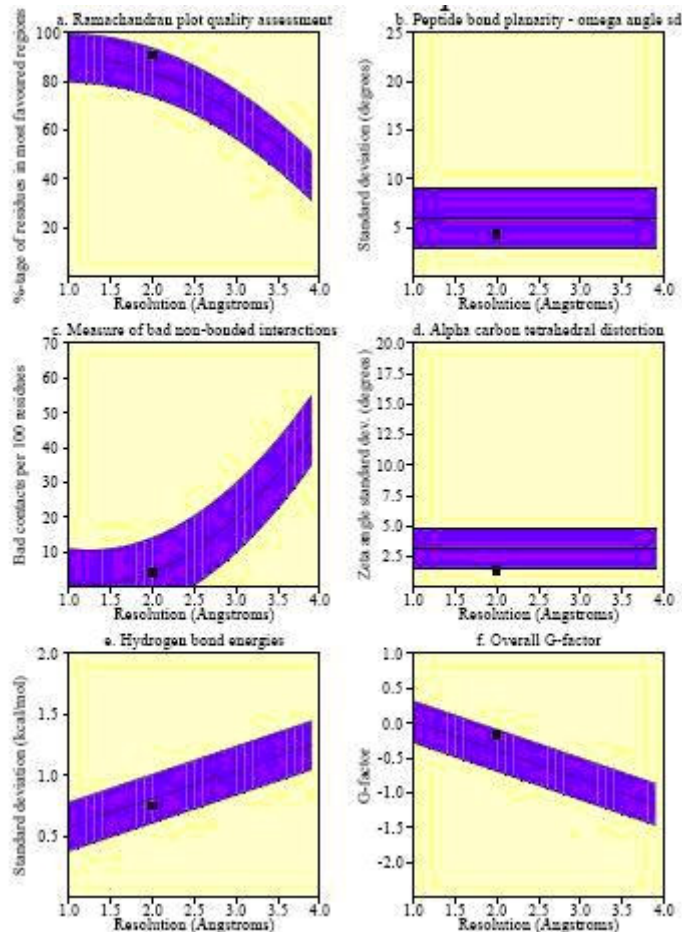


Figura 16: Gráficos dos parâmetros da cadeia principal do modelo.

Pode-se observar que todos os valores estão dentro das faixas permitidas, até mesmo a distorção do carbono alpha (gráfico d), apresentando um valor quase fora do limiar.

A análise realizada pelo VERIFY3D avalia a confiabilidade do enovelamento, determinando os ambientes químicos de cada resíduo do modelo e atribui *scores* com referência a uma matriz construída a partir de uma análise estatística envolvendo as estruturas protéicas do PDB (BOWIE, 1991)(SILVA, 2008). Essa matriz contém três propriedades que cada resíduo apresenta dentro de cada um dos dezoito ambientes químicos definidos. A posição de cada resíduo no modelo 3D é caracterizada pelo seu ambiente químico, e as preferências estatísticas para cada um dos aminoácidos são determinadas para cada um dos ambientes. Dessa forma o software avalia a compatibilidade da estrutura do modelo protéico com a sua seqüência, através do perfil 3D.

Os valores dos dez primeiros resíduos de cada extremidade são desconsiderados, e por este motivo se encontram no mesmo patamar de escore (BOWIE, 1991)(SILVA, 2008).

No gráfico gerado percebe-se que os *scores* de cada resíduo apresentam-se dentro da faixa considerada aceitável (entre 0,0 e 0,62), apenas quatro resíduos encontraram-se fora dessa faixa, mas englobando-se na faixa ainda aceitável (entre 0,0 e -0,03). Dessa forma, pode-se afirmar que o modelo proposto está de acordo com os parâmetros estereoquímicos descritos anteriormente.

Como o gráfico é muito extenso devido ao grande tamanho da proteína, a Figura 17 contém uma parte do começo da proteína, parte esta que apresentou grande problema de enovelamento.

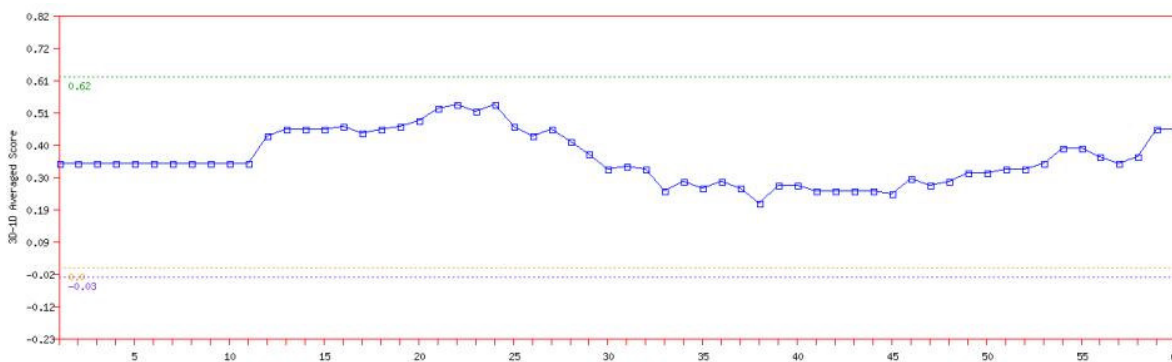


Figura 17: Gráfico *score* x resíduos, gerado pelo programa Verify 3D.

Completando a análise estereoquímica utilizou-se o WHATIF, responsável por avaliar a qualidade do empacotamento. Essa avaliação é realizada através do cálculo do chamado “índice da qualidade de contato”. Este índice é a medida entre a distribuição dos átomos ao redor de uma cadeia lateral de um aminoácido e as distribuições equivalentes observadas em proteínas com estruturas resolvidas. Foi utilizado o módulo *Coarse Packing Quality Control* do software WHATIF.

Um índice menor do que -5,0 para um resíduo significa contatos atômicos ruins ou incomuns, mas não necessariamente, uma estrutura incorreta, sendo necessário uma melhor analisada (VRIEND, 1993)(SILVA, 2008).

O valor médio do índice da qualidade de contato para o modelo protéico pode ser interpretado da seguinte forma:

- maior que -0,5: ótimo modelo;
- em -0,5: valor médio para um bom modelo;
- entre -1,0 e -0,5: ainda um bom modelo;
- em -1,5: ainda um bom modelo, mas com pequenos erros;
- em -2,0: modelo considerado pobre;
- em -3,0: modelo ruim;

O valor médio do modelo obtido foi de -0,880, sendo assim, o modelo obtido é considerado bom. Há resíduos com índices ruins, mas muitos deles são reflexos da baixa qualidade das proteínas modelos seguidas (1SJP e 1WE3). Como pode ser observado na Figura 18, os resíduos 42 e 58 apresentaram um índice menor que -5, mas ao analisar os índices dos resíduos correspondentes na proteína

molde observou-se que o índice também era ruim. A Figura 18 contém os índices apenas dos primeiros resíduos da proteína, já que o arquivo de saída era muito extenso.

1	MET	(1)	:	-3.064	38	VAL	(38)	:	0.448	75	GLU	(75)	:	0.107
2	ALA	(2)	:	0.092	39	LEU	(39)	:	-0.177	76	VAL	(76)	:	0.389
3	LYS	(3)	:	-1.867	40	GLU	(40)	:	-4.045	77	ALA	(77)	:	0.100
4	THR	(4)	:	-3.446	41	LYS	(41)	:	-2.534	78	LYS	(78)	:	0.830
5	ILE	(5)	:	-0.528	42	LYS	(42)	:	-6.246	79	LYS	(79)	:	1.214
6	ALA	(6)	:	0.448	43	TRP	(43)	:	-5.119	80	THR	(80)	:	1.013
7	TYR	(7)	:	-7.340	44	GLY	(44)	:	-3.461	81	ASP	(81)	:	-1.357
8	ASP	(8)	:	-2.355	45	ALA	(45)	:	-2.052	82	ASP	(82)	:	-2.566
9	GLU	(9)	:	1.167	46	PRO	(46)	:	0.736	83	VAL	(83)	:	-3.318
10	GLU	(10)	:	-1.036	47	THR	(47)	:	0.859	84	ALA	(84)	:	-0.428
11	ALA	(11)	:	-0.925	48	ILE	(48)	:	-0.303	85	GLY	(85)	:	-2.767
12	ARG	(12)	:	-1.318	49	THR	(49)	:	-0.536	86	ASP	(86)	:	-2.375
13	ARG	(13)	:	1.233	50	ASN	(50)	:	-3.115	87	GLY	(87)	:	1.907
14	GLY	(14)	:	-0.298	51	ASP	(51)	:	2.643	88	THR	(88)	:	0.202
15	LEU	(15)	:	-0.552	52	GLY	(52)	:	-1.281	89	THR	(89)	:	0.043
16	GLU	(16)	:	-1.537	53	VAL	(53)	:	-0.031	90	THR	(90)	:	1.052
17	ARG	(17)	:	2.771	54	SER	(54)	:	-0.113	91	ALA	(91)	:	1.375
18	GLY	(18)	:	0.383	55	ILE	(55)	:	0.270	92	THR	(92)	:	0.722
19	LEU	(19)	:	-0.130	56	ALA	(56)	:	-0.328	93	VAL	(93)	:	1.291
20	ASN	(20)	:	0.096	57	LYS	(57)	:	-1.778	94	LEU	(94)	:	1.684
21	SER	(21)	:	-0.706	58	GLU	(58)	:	-4.944	95	ALA	(95)	:	1.405
22	LEU	(22)	:	0.633	59	ILE	(59)	:	-0.569	96	GLN	(96)	:	1.767
23	ALA	(23)	:	0.932	60	GLU	(60)	:	-3.524	97	ALA	(97)	:	1.328
24	ASP	(24)	:	1.071	61	LEU	(61)	:	-1.660	98	LEU	(98)	:	-0.119
25	ALA	(25)	:	-1.465	62	GLU	(62)	:	-4.367	99	VAL	(99)	:	0.127
26	VAL	(26)	:	-1.315	63	ASP	(63)	:	2.559	100	LYS	(100)	:	-0.736
27	LYS	(27)	:	-1.384	64	PRO	(64)	:	2.487	101	GLU	(101)	:	1.032
28	VAL	(28)	:	0.477	65	TYR	(65)	:	-0.210	102	GLY	(102)	:	0.501
29	THR	(29)	:	-2.033	66	GLU	(66)	:	-0.925	103	LEU	(103)	:	2.480
30	LEU	(30)	:	-1.693	67	LYS	(67)	:	-0.955	104	ARG	(104)	:	2.621
31	GLY	(31)	:	0.161	68	ILE	(68)	:	1.157	105	ASN	(105)	:	-0.856
32	PRO	(32)	:	-3.148	69	GLY	(69)	:	0.968	106	VAL	(106)	:	-1.313
33	LYS	(33)	:	-4.269	70	ALA	(70)	:	1.233	107	ALA	(107)	:	-3.424
34	GLY	(34)	:	-1.556	71	GLU	(71)	:	0.426					
35	ARG	(35)	:	-3.626	72	LEU	(72)	:	0.967					
36	ASN	(36)	:	-1.779	73	VAL	(73)	:	-0.149					
37	VAL	(37)	:	2.780	74	LYS	(74)	:	1.155					

Figura 18: Trecho do arquivo gerado pelo *Coarse Packing Quality Control*, do WHATIF, em destaque os resíduos 42 e 58.

Capítulo 4: Conclusão Final

Os objetivos do projeto foram alcançados, tanto na modelagem da estrutura secundária do RNAm, quanto na modelagem da estrutura terciária da proteína

O estudo estabeleceu a estrutura do RNAm, o que irá favorecer a digestão de determinadas partes desse RNA, permitindo o uso apenas do ligante, otimizando estratégias de terapia gênica.

Embora a validação da estrutura terciária da proteína aponte alguns resíduos de baixa qualificação, a estrutura apresentou um bom enovelamento. Tal estrutura possibilita a futura determinação das seqüências mais imunogênicas, favorecendo assim o desenvolvimento de novas formas vacinais.

Desta forma é possível concluir que com a modelagem das estruturas em questão houve um avanço no entendimento dessas moléculas, contribuindo para futuras pesquisas.

Referências Bibliográficas

ALBERTS, B. Fundamentos da Biologia Celular, Editora Universitária, 2005.

ALTSCHUL, S.F.; GISH, W.; MILLER, W. MYERS, E.W.; LIPMAN, D.J. Basic local alignment search tool. J. Mol. Biol., 215:403-410, 1990.

MACHADO-LIMA, A.; DEL PORTILLO, H.A.; DURHAM. Computational methods in noncoding RNA research. J. Math. Bio., 56: 15-49, 2008.

BABENDURE, J.R., BABENDURE, J.L. Control of mammalian translation by mRNA structure near caps. RNA, 12: 851-861, 2006.

BAUER-MARCHLER, A., BRYANT, S.H. CD-Search: protein domain annotations on the fly. Nucleic Acids Research, 32:W327-W331, 2004.

BOWIE, J.U., LUTHY, R. EISENBERG, D. A method to identify protein sequence that fold into known three-dimensional structure. Science, 253:164-170, 1991.

COELHO-CASTELO, A.A.M. B-Lymphocytes in bone marrow or lymph nodes can take up plasmid DNA after intramuscular delivery. Hum. Gene Ther., 14: 1279-1285, 2003.

COELHO-CASTELO, A.A.M. Tissue distribution of a plasmid DNA encoding Hsp65 gene is dependent on the dose administered through intramuscular delivery. Genet. Vaccines Ther., v.4, 2006.

ESWAR, N., MARTI-RENOM M.A., WEBB, B., MADHUSUDHAN, M.S., ERAMIAN, D., SHEN, M., PIEPER, U., SALI, A. Comparative protein structure modeling with MODELLER. Current protocols in Bioinformatics, John Wiley & Sons, Inc., supplement 15, 5.6.1-5.6.30, 2000.

HELKE, A., GEINSEN, R.M. An unstructured mRNA region and a 5' hairpin represent important elements of the E.coli translation initiation signal determined by using the bacteriophage T7 gene 1 translation start site. *Nucleic Acids Research*, 21: 5705-57011, 1993.

HIGGINS, D. THOMPSON, J., GIBSON, T., THOMPSON, J.D., HIGGINS, D.G., GIBSON, T.J. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22: 4673-4680, 1994.

FINN, R.D., MISTRY, J, SCHUSTER-BOCKLER, B., GRIFFITHS-JONES, S. HOLLICH, V., LASSMANN, T., MOXON, S., MARSHALL, M., KHANNA, A., DURBIN, R., EDDY, S.R., SONNHAMMER, E.L.L., BATEMAN, A. Pfam: clans, web tools and services. *Nucleic Acids Research*, 34: D247-D251, 2006.

GIBAS, C.; JAMBECK, P. *Desenvolvendo bioinformática*. Editora O'Reilly & Associates, 2001.

GURUNATHAN, S; KLINMAN, D.M; SEDER, R.A. DNA vaccines: immunology, application, and optimization. *Annu. Rev. Immunol.*, 18: 927-74, 2000.

KAUFMANN, S.H.E.; VATH, U.; THOLE, J.E.R. Enumeration of T cells reactive with Mycobacterium tuberculosis organisms specific for the recombination protein 64 kDa. *Eur. J. Immunol.*, v.18, p.1835-8, 1987.

LASKAWSKI, R.A., MACARTHUR, M.W., MOSS, D.S., THORNTON, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, 26: 283-291, 1993.

LIMA, K.M. Efficacy of DNA-hsp65 vaccination for tuberculosis varies with method of DNA introduction in vivo. *Vaccine*, 22: 49-56, 2003.

LINDQUIST, S. The heat-shock response. *Annu. Rev. Biochem.*, 55: 1151-1191, 1986.

LINDQUIST, J.A.; JENSEN, O.N.; MANN, M.; HAMMERLING, G.J. ER-60, a chaperone with thiol-dependent reductase activity involved in MHC class I assembly. *EMBO J.*, 17(8): 2186-2195, 1998.

LORENZI, J.C.C. RNA mensageiro codificador da proteína hsp65 de *Mycobacterium leprae* como uma nova estratégia vacinal no combate a tuberculose experimental. Dissertação de mestrado em Imunologia Básica e Aplicada. Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, 2008.

LUGOSI, L. Theoretical and methodological aspects of BCG vaccine from discovery of Calmette and Guérin to molecular biology. A review. *Tubercule and Lung Disease*, 73: 251-261, 1992.

KOZAK, M. Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc Natl. Acad Sci USA*, 83: 2850-2854, 1986.

MOUNT, D.W. *Bioinformatics: Sequence and Genome Analysis*. Editora Cold Spring Harbor Laboratory, 2001.

SANTOS, G.R. Efeito terapêutico da vacina DNA-Hsp65 em melanoma B16F10. Dissertação de doutorado em Imunologia básica e aplicada. Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, 2008.

SILVA, C.L.; LOWWRIE, D.B. A single micobacterial protein (hsp65) expressed by a transgenic antigen-presenting cell vaccinates mice against tuberculosis. *Immunol.*, 82: 244-248, 1994.

SILVA, C.L. Characterization of the T cells that confer a high degree of protective immunity against tuberculosis in mice after vaccination with tumour cells expressing mycobacterial hsp65. *Infect. Immun.*, 64: 2400-2407, 1996.

SILVA, V.B. Estudos de modelagem molecular e relação estrutura atividade de oncoproteína hnRNP K e ligantes. Dissertação de mestrado em Física Biológica. Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, 2008.

SRIVASTAVA, P.K.; MENORET, A.; BASU, S.; BINDER, R.J.; MCQUADE, K.L. Heat shock proteins come of age: primitive functions acquire new roles in na adaptaive world. *Immunity*, 8: 657-665, 1998

SRIVASTAVA, P,K. Roles of heat-shock proteins in innate and adaptive immunity. *Nat. Rev. Immunol.* 2: 185-194, 2002.

TURNER, P.C.; MCLENNAN, A.G. *Biologia Molecular*. Editora A.D.Bates & M.R.H. White, 2004.

VRIEND,G.; SANDER,C. Quality control of protein models: Directional atomic contact analysis. *J. Appl. Cryst.*, 26: 47-60, 1993.

ZUKER, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 31(13): 3406-15, 2003.