**UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS
DEPARTAMENTO DE FÍSICA E INFORMÁTICA**

**Modelagem Molecular da Enzima
gliceraldeído-3-fosfato desidrogenase
de *T. cruzi* e Análise de Potenciais
Inibidores Específicos.**

EZEQUIEL HORACIO PANEPUCCI

Dissertação apresentada ao Instituto de
Física de São Carlos da Universidade de
São Paulo para a obtenção do título de
Mestre em "Física Aplicada".

ORIENTADOR: PROF. DR. GLAUCIUS OLIVA

**SÃO CARLOS
1994**

MEMBROS DA COMISSÃO JULGADORA DA DISSERTAÇÃO DE MESTRADO DE ~~EZEQUIEL MONCÍO PINEPUCCI~~ APRESENTADA
AO INSTITUTO DE FÍSICA DE SÃO CARLOS, DA UNIVERSIDADE DE SÃO PAULO, EM 16/12/1994

COMISSÃO JULGADORA:

Prof. Dr. Glaucius Oliva

Prof. Dra. Rosemary Sanches

Prof. Dr. Amando Siuiti Ito

*Dedico esta dissertação à meus pais,
irmãos e, especialmente, à Gisele.*

AGRADECIMENTOS

À minha mãe Lucia e ao meu pai Horacio, pelo apoio, pela compreensão, pelas valorosas discussões que travamos e pelo carinho que me dedicaram e dedicam ainda hoje.

À minha esposa Gisele pelo amor, carinho, compreensão, companheirismo, pelas correções ortográficas e pelas calorosas discussões.

Aos meus irmãos, Roberto, Luciano e Rodrigo (em ordem decrescente de idade) pelo apoio e amizade.

À minha amiga Gabriela Castellano pelo companheirismo, pela força e pela amizade.

Ao Prof. Glaucius pela orientação e amizade.

Ao Prof. Richard pelas sugestões, discussões e amizade.

Ao amigo Eduardo Horjales pelas ótimas discussões e pela amizade.

À Profa. Yvonne pelo apoio e pelas discussões.

Aos Profs. Castellano e Julio pelas discussões cristalográficas e pela amizade.

À moçada do grupo, Cristininha, Teresa, Valma, Bianca, Beatriz, Ignez, Célia, Dulce, Paulão, Marcão, Rodrigo, João, Geraldo, Antônio, Jorge e Ednilson pela amizade, pelo companheirismo e pelas discussões.

À secretária Maria Helena pela ajuda inestimável e pela amizade.

Ao Gert Vriend pelas discussões computacionais.

Aos técnicos Augusto, Geraldo e Vanda pelo apoio e pela amizade.

Às bibliotecárias pela presteza e simpatia.

Aos funcionários do Instituto de Física de São Carlos pela eficiência.

Ao CNPq pelo suporte financeiro.

ÍNDICE

1. INTRODUÇÃO	5
1.1 DESENHO RACIONAL DE DROGAS	6
1.2 A DOENÇA DE CHAGAS	9
1.2.1 <i>Glicólise em Trypanosoma cruzi e uma Possível Proteína Alvo</i>	10
1.3 ESTRUTURA E FUNÇÃO DE PROTEÍNAS	12
1.3.1 <i>Função Das Proteínas</i>	12
1.3.2 <i>Estrutura de Proteínas</i>	13
1.3.3 <i>Estrutura e Função de Proteínas</i>	21
1.4 MÉTODOS EMPÍRICOS PARA A DETERMINAÇÃO DA ESTRUTURA MOLECULAR DE UMA PROTEÍNA	22
1.4.1 <i>Cristalografia de Proteínas</i>	22
1.4.2 <i>Ressonância Magnética Nuclear</i>	24
1.5 MÉTODOS SEMI-EMPÍRICOS DE DETERMINAÇÃO DE ESTRUTURA	25
1.5.1 <i>Previsão de estrutura secundária</i>	26
1.5.2 <i>Modelagem por homologia</i>	27
1.6 ESTRUTURA GERAL DAS ENZIMAS GAPDH	32
2. MODELAGEM DA GGAPDH	35
2.1 ALINHAMENTO	35
2.2 CONSTRUÇÃO E AVALIAÇÃO DO MODELO	37
2.3 SIMULAÇÕES COMPUTACIONAIS	45
2.3.1 <i>Dinâmica molecular</i>	45
2.3.2 <i>Minimização de energia</i>	45
2.4 COMPARAÇÃO DA GGAPDH DE <i>TRYPANOSOMA CRUZI</i> COM A GAPDH HUMANA	46
3. INIBIDORES ESPECÍFICOS PARA GGAPDH DE TRYPANOSOMA CRUZI	48
3.1 COMPOSTO 1: 2-METIL ADENOSINA	48
3.2 COMPOSTO 2: 8-TIENIL ADENOSINA	50
3.3 COMPOSTO 3: 2'-DESOXI-2'-(3-METOXI BENZAMIDO) ADENOSINA	52
3.4 COMPOSTO 4: 2'-DESOXI-2'-(2H-CHROMEN BENZAMIDO) ADENOSINA	55
3.5 COMPOSTO 5: 2'-DESOXI-2'-(BENZOIMIDAZOL) ADENOSINA	56
4. UM PROGRAMA PARA VISUALIZAÇÃO DE ESTRUTURAS DE PROTEÍNAS	58
4.1 INTRODUÇÃO	58
4.2 USANDO O PROGRAMA	59
5. CONCLUSÕES	63
6. APÊNDICES	66
6.1 ALINHAMENTO ENTRE 78 SEQUÊNCIAS DE GLICERALDEÍDO-3-FOSFATO DESIDROGENASES DE VÁRIAS ESPÉCIES	66
7. REFERÊNCIAS BIBLIOGRÁFICAS	71

LISTA DE FIGURAS

FIGURA 1: REAÇÃO CATALISADA PELA GLICERALDEÍDO-3-FOSFATO DESIDROGENASE.	13
FIGURA 2: A) FORMA GERAL PARA UM AMINOÁCIDO, <i>R</i> REPRESENTA A CADEIA LATERAL. B) LIGAÇÃO PEPTÍDICA ($\Omega=180^\circ\pm20$) E ÂNGULOS TORSIONAIS Φ E Ψ	14
FIGURA 3: NOMENCLATURA DOS ÁTOMOS NA CADEIA LATERAL DE UM AMINOÁCIDO.	15
FIGURA 4: HÉLICE A MOSTRANDO A DISPOSIÇÃO DAS LIGAÇÕES DE HIDROGÊNIO ENTRE OS ÁTOMOS DA CADEIA PRINCIPAL.	17
FIGURA 5: FOLHA B MOSTRANDO A ASSOCIAÇÃO NA FORMA PARALELA (A ESQUERDA) E ANTIPARALELA (A DIREITA).....	19
FIGURA 6: FLUXOGRAMA GENÉRICO PARA A CONSTRUÇÃO POR HOMOLOGIA DE UM MODELO 3-D DE PROTEÍNA.....	30
FIGURA 7: REPRESENTAÇÃO DOS DOIS DOMÍNIOS DE GAPDH:	33
FIGURA 8: TETRÂMERO DO MODELO DA gGAPDH DE <i>TRYPANOSOMA CRUZI</i> VISTO AO LONGO DO EIXO <i>Q</i> . (MODELO FINAL DO TRABALHO).....	34
FIGURA 9: MONÔMERO DA gGAPDH DE <i>TRYPANOSOMA CRUZI</i> . AS SETAS REPRESENTAM FITAS B E AS ESPIRAIS HÉLICES A. FIGURA CRIADA COM O PROGRAMA MOLSCRIPT ⁽¹⁾	34
FIGURA 10: ALINHAMENTO FINAL USADO NA CONSTRUÇÃO DO MODELO DA ENZIMA GLICERALDEÍDO-3-FOSFATO DESIDROGENASE DE <i>TRYPANOSOMA CRUZI</i>	38
FIGURA 11: DIAGRAMA DE RAMACHANDRAN SOBRE A ESTATÍSTICA DO PROGRAMA PROCHECK41	
FIGURA 12: CADEIA PRINCIPAL DO MODELO NA REGIÃO DA ALA 72 ANTES DE SER MODELADA (LINHA TRACEJADA) E DEPOIS (LINHA SÓLIDA).....	42
FIGURA 13: CADEIA PRINCIPAL DO MODELO (LINHA TRACEJADA) NA REGIÃO DE ARG 11: A) ANTES DA MODELAGEM E B) DEPOIS DA MODELAGEM	43
FIGURA 14: DISTRIBUIÇÃO DE χ_1 E χ_2 PARA O MODELO FINAL.	44
FIGURA 15: SUPERPOSIÇÃO DOS MONÔMEROS DE <i>T. CRUZI</i> E DA ENZIMA HUMANA.	47
FIGURA 16: 2-METIL ADENOSINA. CONSTANTE DE INIBIÇÃO $IC_{50} = 700$ mM.....	49
FIGURA 17: SÍTIO DE LIGAÇÃO DO INIBIDOR 2-METIL ADENOSINA. AS LINHAS FINAS REPRESENTAM A PROTEÍNA, AS GROSSAS O INIBIDOR E AS LINHAS TRACEJADAS REPRESENTAM AS INTERAÇÕES INIBIDOR/PROTEÍNA.....	50
FIGURA 18: SEGUNDO INIBIDOR: GRUPO TIENILA NA POSIÇÃO 8 DO GRUPO ADENINA. CONSTANTE DE INIBIÇÃO $IC_{50} = 500$ mM.....	51
FIGURA 19: CONTATOS ENTRE O GRUPO TIENILA E A ENZIMA DE <i>T. CRUZI</i> MOSTRANDO AS INTERAÇÕES COM LEU 112 E MET 38.	52
FIGURA 20: TERCEIRO INIBIDOR: 2'-DESOXI-2'-(3-METOXI BENZAMIDO) ADENOSINA.....	53
FIGURA 21: SUPERPOSIÇÃO DAS ENZIMAS HUMANA EM ROSA E DE <i>T. CRUZI</i> EM AMARELO NA REGIÃO QUE COMPÕE A FENDA HIDROFÓBICA.	54
FIGURA 22: GRÁFICO DOS FATORES DE AGITAÇÃO TÉRMICA DA REGIÃO DA FENDA HIDROFÓBICA. ESTÃO GRAFICADOS OS VALORES PARA AS ESTRUTURAS DE <i>T. BRUCEI</i> , <i>B. STEAROTHERMOPHILUS</i> E HUMANA.	55
FIGURA 23: 2'-DESOXI-2'-(2H-CHROMEN BENZAMIDO) ADENOSINA.....	56
FIGURA 24: COMPOSTO 4 INSERIDO NA FENDA HIDROFÓBICA DA ENZIMA DE <i>T. CRUZI</i> . AS LINHAS TRACEJADAS DENOTAM AS LIGAÇÕES DE HIDROGÊNIO INIBIDOR-ENZIMA.....	56
FIGURA 25: 2'-DESOXI-2'-(BENZOIMIDAZOL) ADENOSINA.....	57
FIGURA 26: A ENZIMA DE <i>T. CRUZI</i> E O INIBIDOR ESTÃO COM A SUPERFÍCIE DE VAN DER WAALS REPRESENTADA, MOSTRANDO A EFICIÊNCIA DO EMPACOTAMENTO.	57
FIGURA 27: JANELA DO PDBV MOSTRANDO AS JANELAS PRINCIPAL E DE RAMACHANDRAN E OS MENUS.	59
FIGURA 28: ESQUEMA GERAL PARA O TRABALHO DE DESENHO RACIONAL DE DROGAS.....	64

LISTA DE TABELAS

TABELA 1: PROGRAMAS GRÁFICOS PARA MODELAGEM MOLECULAR E BASES DE DADOS DISPONÍVEIS.....	6
TABELA 2: ETAPAS NO PROCESSO DE DESENVOLVIMENTO DE UMA NOVA DROGA.....	7
TABELA 3: CLASSIFICAÇÃO E NOMENCLATURA DOS AMINOÁCIDOS	14
TABELA 4: CLASSIFICAÇÃO DAS VOLTAS.	20
TABELA 5: ALINHAMENTO ENTRE DUAS SEQUÊNCIAS IMAGINÁRIAS.....	29
TABELA 6: RELAÇÃO DAS PROTEÍNAS HOMÓLOGAS À ENZIMA GGAPDH DE <i>TRYPANOSOMA</i> <i>CRUZI</i> COM ESTRUTURA RESOLVIDA.....	35
TABELA 7: DIFERENÇAS ENTRE GGAPDH DE <i>TRYPANOSOMA CRUZI</i> E GAPDH HUMANA.....	46

RESUMO

Foi construído o modelo tridimensional da enzima gliceraldeído-3-fosfato desidrogenase de *Trypanosoma cruzi*, o causador da doença de Chagas, usando técnicas computacionais de modelagem por homologia. O modelo foi comparado à enzima muscular homóloga de humanos quanto às diferenças no sítio de ligação do cofator NAD^+ .

Sobre o modelo da enzima de *T. cruzi* foram construídas moléculas análogas ao grupo adenosina do cofator NAD^+ como tentativa de se obter um inibidor seletivo à enzima do parasita e não efetivo quanto à enzima de humanos. Alguns dos compostos haviam sido ensaiados quanto á afinidade pela enzima análoga de *Trypanosoma brucei*.

Foi escrito um programa de visualização gráfica de modelos moleculares que permite a análise dos parâmetros esteroquímicos com rapidez e simplicidade.

ABSTRACT

The threedimensional model of the glycosomal enzyme glyceraldehyde-3-phosphate dehydrogenase from *Trypanosoma cruzi*, the causative agent of Chagas disease, was built using homology modelling computational techniques. The model was compared against the human muscle homologous enzyme with regard to the differences in the NAD^+ binding site.

Adenosine analogs were designed based on the model of the *T. cruzi* enzyme as an attempt to build selective inhibitors of the parasite enzyme with no activity against the human enzyme. Some of the compounds were previously tested for their affinity for the homologous enzyme from *Trypanosoma brucei*.

A graphical program was written that allows the representation and analysis of stereochemical parameters of molecular models with ease and speed.

1. INTRODUÇÃO

O avanço recente na tecnologia computacional, tanto em "hardware" quanto em "software", tem provido os cientistas com ferramentas quase imprescindíveis para o seu trabalho. A área de biologia molecular, que trata do estudo do funcionamento de sistemas biológicos a nível molecular, tem se beneficiado desta nova situação para desenvolver e planejar de forma mais eficiente seus experimentos e modelos teóricos.

Existem atualmente à disposição de qualquer cientista inúmeros bancos de dados com informações que vão desde sequência de aminoácidos de proteínas até as estruturas tridimensionais das mesmas. Estes bancos de dados são mantidos por programas e computadores dedicados e podem ser acessados interativamente através de uma rede de computadores que cobre todo o planeta usando um computador pessoal conectado à esta rede. Hoje em dia pode-se descobrir, em aproximadamente 10 minutos, se uma determinada proteína já teve sua estrutura resolvida e transferir as informações pertinentes da base de dados pesquisada para o seu próprio computador pessoal. Nestes mesmos 10 minutos pode-se ainda vasculhar um banco de dados de sequências de proteínas usando como sonda uma proteína recém sequenciada e obter como resultado uma lista das 200 sequências mais parecidas à sequência problema.

Por outro lado, a análise de dados experimentais e de modelos teóricos também ficou bastante facilitada com o avanço computacional. Computadores dedicados à representação gráfica tridimensional são usados para visualizar e alterar a estrutura tridimensional de uma biomolécula qualquer bem como submetê-la a simulações computacionais. Atualmente os programas de modelagem molecular permitem desde a simples visualização de uma estrutura determinada experimentalmente até a construção por métodos *ab-initio* de uma estrutura tridimensional de uma proteína fictícia. Na Tabela 1 estão listados alguns dos programas gráficos dedicados a este tipo de trabalho.

Em particular, os grupos ligados a pesquisa de novas drogas desenvolveram uma linha de pesquisa que faz uso desta tecnologia para guiar o desenho de novos fármacos de forma mais objetiva, chamou-se a esta linha de pesquisa de *desenho racional de drogas*. O desenho racional de drogas pode ser usado quando está disponível a estrutura tridimensional de alguma proteína (*proteína alvo*) relacionada à causa da doença, seja ela de um vírus, de uma bactéria ou uma proteína defectiva do próprio organismo portador da doença.

Esta dissertação tem por objetivo o estudo de inibidores para a enzima glicosomal gliceraldeído-3-fosfato desidrogenase de *Trypanosoma cruzi*, que é o causador da doença de Chagas, testados por experimentos bioquímicos na enzima homóloga de *Trypanosoma brucei*. Este estudo envolverá a construção do modelo tridimensional da proteína e dos inibidores bem como da construção do complexo do modelo da proteína com cada um dos inibidores. Também serão construídos modelos do complexo entre os inibidores e a estrutura cristalográfica da enzima homóloga humana. Os complexos serão avaliados quanto às interações inibidor-enzima de modo a comparar o modo de ligação destes inibidores na enzima de *T. cruzi* e na enzima homóloga humana.

Tabela 1: Programas gráficos para modelagem molecular e bases de dados disponíveis.

Programa	Base de Dados
<i>TOM</i> ⁽¹⁾ †	OWL ⁽²⁾
<i>WHATIF</i> ⁽³⁾ †	SwissProt ⁽⁴⁾
<i>O</i> ⁽⁵⁾ †	PIR ⁽⁶⁾
<i>PDBV</i> †	PDB ⁽⁷⁾
<i>INSIGHT</i> ⁽⁸⁾ ‡	GenBank ⁽⁹⁾
	NRL-3D ⁽¹⁰⁾

† Acadêmico; ‡ Comercial;

1.1 Desenho Racional De Drogas

O processo padrão para se encontrar uma nova droga consiste de um ciclo de tentativa e erro que começa com uma seleção de possíveis drogas de um banco de dados de moléculas sintéticas e naturais. Estas moléculas (aproximadamente 100.000) são testadas *in vitro* contra o causador da doença

e as que não passam pelo limite mínimo de atuação são descartadas. As que sobram são submetidas a testes pré-clínicos com cobaias e as que passam por estes testes são então submetidas a testes clínicos com humanos. A Tabela 2⁽¹¹⁾ resume as etapas envolvidas no desenvolvimento de uma nova droga com uma estimativa de tempo de conclusão de cada etapa.

Tabela 2: Etapas no processo de desenvolvimento de uma nova droga.

Etapas	Tempo (anos)
Descobrimiento do composto de partida	1-2
Otimização do composto de partida	1-2
Testes <i>in vitro</i> e <i>in vivo</i>	1-2
Testes toxicológicos	1-3
Testes de segurança para humanos	1
Testes de eficácia para humanos	1-2
Tempo total de desenvolvimento	6-12

O tempo total de desenvolvimento de uma nova droga bem como o custo envolvido neste processo tem sido significativamente reduzido com a crescente utilização da nova tecnologia de desenvolvimento racional de drogas baseado em estruturas. Este processo é usado quando alguma proteína (chamada de proteína alvo) relacionada à causa de determinada doença tem a sua estrutura tridimensional elucidada. Existem vários métodos que podem ser empregados na obtenção de compostos de partida, a partir dos quais são feitos refinamentos até que se obtenha a forma final do medicamento que poderá ser comercializado.

Um dos métodos parte de uma varredura inicial de uma base de dados de compostos de interesse farmacêutico usando como parâmetro de exclusão um teste de inibição da proteína alvo com cada um dos possíveis candidatos a droga. Este método tem a vantagem de lidar com compostos de partida para os quais, geralmente, já se conhece o comportamento *in vivo*, isto é, a toxicidade no organismo, biodisponibilidade, cinética de absorção e eliminação, etc. Uma indústria farmacêutica pode vasculhar uma base de 100.000 compostos em aproximadamente 3 meses com o auxílio de tecnologia robótica.

Outro método usa os cofatores e/ou substratos naturais de cada enzima que são computacionalmente modificados usando-se a estrutura da proteína alvo como guia⁽¹²⁾. A vantagem do uso de cofatores e/ou moléculas de substrato como compostos de partida é que estes são muito bem aceitos pelo organismo. O composto assim obtido deve ser testado para atividade inibitória contra a proteína alvo.

Um terceiro método parte da construção *de novo* de moléculas a partir da conformação da proteína alvo usando um banco de fragmentos de moléculas na construção de um composto de partida^(13,14). Estas moléculas devem, como no caso anterior, ser testadas *in vitro* para atividade de inibição contra a proteína alvo.

Uma vez encontrados os compostos de partida que apresentam uma constante de inibição razoável, estes devem passar para a etapa de otimização. Esta etapa envolve a determinação da estrutura do complexo proteína alvo-inibidor para que se possa estabelecer qual o modo de ligação do inibidor no complexo. Uma vez conhecido o modo de ligação do inibidor com a proteína alvo podem-se fazer modificações do inibidor inicialmente proposto de modo a aumentar a especificidade e a constante de ligação desta pequena molécula para com a proteína.

Um bom exemplo da eficiência do desenho racional de drogas foi apresentado por Colman e colaboradores⁽¹⁵⁾ num trabalho que envolveu a cooperação entre uma universidade e uma empresa farmacêutica. O trabalho reporta a obtenção de drogas capazes de inibir a replicação *in vitro* e *in vivo* dos vírus da gripe, influenza A e B. Estes vírus têm a capacidade de promover mutações nos resíduos expostos de suas enzimas antigênicas numa taxa muito elevada de modo que o sistema imune do hospedeiro não possa criar defesas com eficiência. A estrutura cristalográfica da enzima viral neuraminidase, crítica para a movimentação do vírus na mucosa do trato respiratório e para a sua evasão de células já infectadas, submetida às técnicas acima descritas, permitiu o desenvolvimento de dois inibidores que atuam em uma região do sítio ativo.

1.2 A Doença de Chagas

A doença de Chagas foi descoberta pelo cientista Carlos Chagas que por volta de 1909 encontrou formas epimastigotas de um flagelado no intestino de um inseto hematófago conhecido como barbeiro (*Triatoma infestans*, *Panstrongylus megistus* e *Triatoma sordida* entre outros) que infestava as moradias das áreas rurais de Minas Gerais. Em seguida, Chagas provou que estes flagelados eram infectantes para sagüis, cães e roedores. Mais tarde identificou o flagelado, que foi batizado de *Trypanosoma cruzi*, no sangue de uma criança que vivia em uma casa infestada pelo barbeiro providenciando assim o primeiro exemplo de uma doença descoberta primeiro no transmissor da doença e depois no ser humano. Foi constatado que mais de 190 espécies de mamíferos e invertebrados são portadores do *T. cruzi* mostrando que a reserva natural do tripanossomo é praticamente inesgotável sendo portanto inexequível qualquer tentativa de erradicação dos tripanossomos.

A infecção pelo *Trypanosoma cruzi* não ocorre durante a picada do barbeiro, mas sim quando o indivíduo picado coça a região atingida levando para dentro do orifício recém-aberto as fezes do barbeiro que contém os metatripanossomos fazendo com que estes alcancem a corrente sangüínea. Uma vez na corrente sangüínea, os tripanossomos invadem as células do hospedeiro e passam para a forma amastigota. A forma amastigota multiplica-se por divisão binária e pode permanecer na célula por longos períodos de tempo. Durante o ciclo de vida do *T. cruzi* as formas amastigotas transformam-se em tripomastigotas (formas providas de flagelo) e vão para a corrente sangüínea levando a infecção para outras células do organismo hospedeiro.⁽¹⁶⁾

A doença manifesta-se de forma não muito severa, entre duas a três semanas após a infecção, podendo levar a quadros de miocardite e meningoencefalite em crianças muito jovens, esta última sendo fatal em aproximadamente 50% dos casos. As piores complicações da doença de Chagas desenvolvem-se de 10 a 20 anos após a infecção, e são caracterizadas por cardiopatia crônica (27% dos infectados), complicações crônicas

digestivas (6%) e neurológicas (3%). Os pacientes com graves doenças crônicas vão se tornando cada vez mais debilitados até a morte, geralmente por falha do coração. A Organização Mundial da Saúde estima que entre 16 e 18 milhões de pessoas estão infectadas com o *Trypanosoma cruzi*, destes, aproximadamente 2 a 3 milhões já devem ter desenvolvido as complicações crônicas, enquanto que 3 milhões ou mais ainda estão no estágio de incubação e poderão desenvolver a doença no futuro. Ocorrem por consequência da doença de Chagas aproximadamente 45.000 mortes por ano (estimativa baseada em estudos longitudinais para o biênio 1991-1992).⁽¹⁷⁾

1.2.1 Glicólise em *Trypanosoma cruzi* e uma Possível Proteína Alvo

A forma intravertebral (sangüínea) do *Trypanosoma cruzi* possui um mitocôndrio inativo fazendo com que o flagelado seja completamente dependente da via glicolítica para as suas necessidades energéticas, isto é, todo o ATP sintetizado pelo *T. cruzi* é obtido a partir da conversão de glicose em piruvato. De acordo com Oppenheimer^(18,19), o *T. cruzi* tem uma organela chamada glicossomo, característica da família tripanossomatida, que contém nove enzimas envolvidas nas reações da via glicolítica. Esta compartimentalização faz com que a glicólise proceda a uma taxa superior à observada em qualquer outro organismo. Foi mostrado por ensaios *in vitro* e *in vivo*⁽²⁰⁾ que a inibição da glicólise causa o desaparecimento dos tripanossomos da corrente sangüínea do hospedeiro mamífero. Foi observado que a enzima glicossomal gliceraldeído-3-fosfato desidrogenase (gGAPDH) que participa da glicólise na etapa em que o gliceraldeído-3-fosfato é convertido em 1,3-bisfosfoglicerato tem também um papel importante no controle do fluxo de metabólitos na interface glicossomo-citossol⁽²¹⁾.

Visto que o bloqueio da via glicolítica no *Trypanosoma cruzi* causa a sua morte, investigou-se quais enzimas poderiam servir como proteína alvo para desenho racional de drogas. A enzima glicossomal gliceraldeído-3-fosfato desidrogenase mostrou-se adequada ao trabalho pelos seguintes motivos:

1. foram verificadas diferenças de afinidade pelo cofator entre as enzimas humana e de *T. cruzi*,
2. um estudo comparativo entre a sequência de *Trypanosoma cruzi* e as estruturas das enzimas humana e de *Bacillus stearothermophilus* evidenciou diferenças na região do sítio de ligação do NAD^+ ,
3. estavam disponíveis estruturas cristalográficas de enzimas homólogas (ver Tabela 6) e
4. a enzima encontra-se expressada em uma cepa de *Escherichia coli*.

A constatação de que a enzima GAPDH humana apresenta uma afinidade por NAD^+ 5 a 10 vezes maior que a afinidade da gGAPDH de *T. cruzi* pelo mesmo cofator implica diferenças na estrutura 3-D do sítio de ligação do cofator entre as duas proteínas.

O objetivo de um trabalho de desenho racional de drogas é a obtenção de um inibidor que seja o mais específico possível para a enzima alvo. Deste modo a existência de diferenças em regiões importantes para a enzima (como, por exemplo, a região de ligação do cofator NAD^+) é fundamental para um melhor desempenho do inibidor. No Apêndice 6.1 encontra-se o alinhamento completo das 78 sequências de GAPDH de diferentes espécies encontradas no Banco de Dados SwissProt. Neste alinhamento as colunas demarcadas com fundo cinza representam posições para as quais 75% das sequências apresentam o mesmo aminoácido. O alinhamento mostra que estas enzimas foram bastante conservadas durante o processo evolutivo, sendo este um ponto não muito favorável quando se está procurando uma proteína alvo para desenho de drogas. Entretanto, um estudo comparativo entre a estrutura cristalográfica da GAPDH humana e a sequência da enzima homóloga de *Trypanosoma brucei* evidenciou diferenças na região do sítio de ligação do NAD^+ confirmando a viabilidade desta enzima como alvo para um trabalho de desenho racional de drogas⁽²²⁾.

A disponibilidade de estruturas cristalográficas de enzimas homólogas à proteína alvo é imprescindível para a criação de um modelo tridimensional da macromolécula.

A disponibilidade da enzima em quantidade para a execução de testes de atividade e para a determinação da estrutura cristalográfica do complexo com os inibidores é imprescindível para que o ciclo do desenho racional de drogas possa ser completado.

1.3 Estrutura e Função de Proteínas

1.3.1 Função Das Proteínas

Proteínas são usadas por organismos vivos na formação de tecidos conjuntivos (colágeno e queratina), no transporte de pequenas moléculas através do organismo (hemoglobina e mioglobina) e na catálise de reações químicas (GAPDH, anidrase carbônica, rodopsina e tirosina quinase entre outras). Em vírus as proteínas existem para formar a cápsula viral e na forma de enzimas chave na infestação das células e posterior replicação. As proteínas que catalizam algum tipo de reação são chamadas de enzimas e formam a classe de proteínas que apresenta maior variabilidade funcional e conformacional.

Estas enzimas associam-se ao substrato (molécula que sofrerá a reação química) e opcionalmente a algum cofator (molécula que participa da reação de alguma maneira) diminuindo a energia de ativação de uma determinada reação. Dentre os cofatores mais comuns estão o NAD^+ (nicotinamida adenosina difosfato), FAD (flavina adenosina difosfato), o ATP (adenosina trifosfato) e o GTP (guanosina trifosfato). Em particular, as gliceraldeído-3-fosfato desidrogenases são enzimas que catalizam a reação de oxidorredução, mostrada na Figura 1, que consiste na transformação de gliceraldeído-3-fosfato em 1,3-bisfosfoglicerato usando como cofator uma molécula de NAD^+ .

O funcionamento e as características de uma proteína são regidos por sua conformação tridimensional. Deste modo, para que seja possível a modificação das propriedades de determinada enzima ou a criação de uma

proteína com funções diferentes ou a elaboração de uma molécula pequena capaz de inibir uma determinada enzima deve-se, primeiro, entender a organização a nível atômico das estruturas proteicas.

Figura 1: Reação catalisada pela gliceraldeído-3-fosfato desidrogenase.

1.3.2 Estrutura de Proteínas

Proteínas são macromoléculas formadas por aminoácidos que ligam-se uns aos outros numa cadeia linear através da formação de ligações peptídicas. Estes aminoácidos são formados por uma parte constante (em **negrito** na Figura 2A) e uma parte variável (a cadeia lateral, simbolizada por *R* na Figura 2A). A sequência dos aminoácidos em uma proteína é chamada de **estrutura primária**. Esta estrutura primária se enovela de modo a formar a **estrutura terciária**, que é a conformação tridimensional de uma fita contínua de proteína. A conformação adotada pela associação de duas ou mais cadeias de proteína recebe o nome de **estrutura quaternária**. A parte constante é responsável pela formação das ligações peptídicas da cadeia principal.

Figura 2: A) Forma geral para um aminoácido, R representa a cadeia lateral. B) Ligação peptídica ($\Omega=180^\circ\pm 20$) e ângulos torsionais Φ e Ψ .

As cadeias laterais R são escolhidas de um elenco de 20 aminoácidos os quais são classificados por sua natureza química de acordo com a Tabela 3.

Tabela 3: Classificação e nomenclatura dos aminoácidos

Característica Química	Nome	Código Tres letras	Código Uma letra
alifáticos	GLICINA	GLY	G
	PROLINA	PRO	P
	ALANINA	ALA	A
	VALINA	VAL	V
	ISOLEUCINA	ILE	I
	LEUCINA	LEU	L
aromáticos	FENILALANINA	PHE	F
	TIROSINA	TYR	Y
	TRIPTOFANO	TRP	W
com enxofre	METIONINA	MET	M
	CISTEÍNA	CYS	C
alifáticos hidroxilados	SERINA	SER	S
	TREONINA	THR	T
ácidos	ASPARTATO	ASP	D
	ASPARAGINA	ASN	N
	GLUTAMATO	GLU	E
	GLUTAMINA	GLN	Q
básicos	LISINA	LYS	K
	ARGININA	ARG	R
	HISTIDINA	HIS	H

Os átomos da cadeia lateral são identificados através de uma nomenclatura padrão definida pela "IUPAC". A primeira letra do nome do átomo define o tipo do átomo, a segunda letra é do alfabeto grego e define a distância deste átomo com relação ao carbono alfa (CA) em termos do número de ligações covalentes e a terceira letra distingue átomos à mesma

distância de ligações do carbono alfa quando existe ramificação da cadeia lateral.

Figura 3: Nomenclatura dos átomos na cadeia lateral de um aminoácido.

A conformação das cadeias laterais é definida pelos ângulos diedros χ_n ($1 \leq n \leq 5$) exceto para os aminoácidos glicina e alanina que não possuem átomos suficientes para definir ângulos torsionais na cadeia lateral.

A conformação da cadeia principal de uma proteína é completamente caracterizada por três ângulos diedros: os ângulos ϕ (formado pelos átomos C_{i-1} , N_i , C^α_i e C_i), ψ (N_i , C^α_i , C_i e N_{i+1}) e Ω (C^α_i , C_i , N_{i+1} e C^α_{i+1}) mostrados na Figura 2B.

O ângulo Ω contém a ligação peptídica e é restrito a $(180 \pm 20)^\circ$ devido à conjugação dos elétrons π da ligação $C=O$ com o par de elétrons não ligado do nitrogênio. Os ângulos ϕ e ψ também são restritos, porém devido a impedimentos estéricos entre os átomos da cadeia lateral e os átomos da cadeia principal⁽²³⁾. As exceções são a prolina e a glicina, a prolina por ter o ângulo ϕ restrito em aproximadamente -64° devido à ligação covalente entre o $C\delta$ da cadeia lateral e o N da cadeia principal e a glicina por não apresentar restrições aos ângulos ϕ e ψ devido a ausência de cadeia lateral.

Verificou-se que o interior de proteínas hidrossolúveis era formado quase que exclusivamente por aminoácidos hidrofóbicos, isto levou à hipótese de que a principal força motriz que leva uma proteína hidrossolúvel a enovelar-se é a tendência de agrupar as cadeias laterais hidrofóbicas no interior da proteína deixando as cadeias hidrofílicas para fora. O problema do agrupamento de cadeias hidrofóbicas no interior da proteína é que a

cadeia principal também deve acomodar seus grupos polares (carbonila e amino) neste interior hidrofóbico. A natureza resolveu este problema com o uso de **estruturas secundárias**. A estrutura secundária aparece entre a estrutura primária e a estrutura terciária e é o elemento estrutural que forma o chamado **esqueleto** da proteína. Os elementos de estrutura secundária são trechos da estrutura primária que adotam uma conformação onde todos os aminoácidos têm aproximadamente os mesmos ângulos ϕ e ψ e quase todos os grupos carbonila e amino formam ligações de hidrogênio com outros grupos amino e carbonila da cadeia principal. Este padrão de ligações de hidrogênio resolve assim o problema de grupos carregados inseridos num ambiente hidrofóbico.

Atualmente as estruturas secundárias são classificadas como hélice α , hélice 3_{10} , hélice π , fita β e voltas. Estas estruturas secundárias são caracterizadas por ângulos ϕ e ψ e por padrões de ligação de hidrogênio.

As hélices α são responsáveis por aproximadamente metade dos elementos de estrutura secundária até agora determinados, sendo formadas através da repetição de ângulos $\phi \sim -60^\circ$ e $\psi \sim -50^\circ$ num determinado trecho da cadeia polipeptídica. Esta repetição propicia um arranjo ideal para a formação de ligações de hidrogênio entre os grupos carbonila e amino da cadeia principal. Mais precisamente, o oxigênio da carbonila do resíduo n faz uma ligação de hidrogênio com o amino do resíduo $n+4$. Dada a natureza periódica da hélice α pode-se definir o número de aminoácidos por volta (3,6 aa/volta), a altura de uma volta completa (5,4 Å) e o deslocamento por resíduo ao longo do eixo da hélice (1,5 Å).

Figura 4: Hélice α mostrando a disposição das ligações de hidrogênio entre os átomos da cadeia principal.

As hélices α são em sua maioria de mão-direita, isto é, olhando-se ao longo do eixo da hélice no sentido N-C terminal os aminoácidos “giram” no sentido horário. Hélices α de mão-esquerda são menos prováveis porque existem choques entre as cadeias laterais e os grupos carbonila da cadeia principal. As hélices α ocorrem geralmente em tamanhos que vão de 5 a 10 aminoácidos podendo chegar a 40 aminoácidos.

O arranjo periódico de aminoácidos numa hélice alfa faz com que todas as ligações de hidrogênio da cadeia principal estejam na direção do eixo da hélice alfa, como mostra a Figura 4, produzindo um dipolo na hélice alfa que gera uma carga parcial positiva na extremidade N-terminal e uma carga parcial negativa na extremidade C-terminal. A presença de carga e de grupos NH livres na extremidade N-terminal de uma hélice alfa propicia um ambiente ideal para a ligação de moléculas com carga negativa, como por exemplo, ATP e NAD^+ que contém grupos fosfato. Entretanto, raramente são encontrados ligantes positivos na extremidade C-terminal de uma hélice α , em contrapartida os aminoácidos com cadeia lateral positiva (LYS e ARG) “fecham” estas extremidades negativas.

As hélices 3_{10} não são tão comuns e nem tão extensas quanto as hélices alfa. As ligações de hidrogênio ocorrem entre o resíduo n e o resíduo $n+3$ com os ângulos $\phi \sim -49^\circ$ e $\psi \sim -26^\circ$. As ligações de hidrogênio nas hélices 3_{10} não são alinhadas como na hélice alfa e a geometria desta hélice provoca mais choques entre as cadeias laterais e a cadeia principal, desfavorecendo assim esta estrutura secundária.

A hélice π é ainda mais rara que a hélice 3_{10} , os ângulos que definem sua cadeia principal são $\phi \sim -57^\circ$ e $\psi \sim -70^\circ$ fazendo com que as ligações de hidrogênio ocorram entre o resíduo n e o resíduo $n+5$. Este tipo de estrutura possui uma cavidade no interior, ao longo do eixo da hélice, que diminui as interações de van der Waals.

A fita β é responsável pela formação das folhas β , a segunda estrutura secundária que ocorre com maior frequência entre as estruturas resolvidas de proteínas. A fita β é caracterizada pela repetição dos ângulos $\phi \sim -64^\circ$ e $\psi \sim +170^\circ$. Ao contrário das hélices, as fitas β não compõe por si só um elemento de estrutura secundária, elas devem se associar a outras fitas β através de ligações de hidrogênio para formar a folha β . Duas fitas β podem se associar de forma paralela ou antiparalela como mostra a Figura 5.

A associação na forma paralela ocorre quando as duas fitas estão orientadas com suas extremidades N-terminal fazendo ligação de hidrogênio uma com a outra. A associação antiparalela ocorre quando o N-terminal de uma fita faz ligações de hidrogênio com o C-terminal da outra fita.

Figura 5: Folha β mostrando a associação na forma paralela (a esquerda) e antiparalela (a direita).

As folhas β ocorrem ou somente com fitas paralelas, ou com fitas antiparalelas, ou com uma mistura de fitas paralelas e antiparalelas, sendo esta última modalidade mais incomum que as duas primeiras. Todas as folhas β têm uma torção de mão-direita sendo que as torções de mão-esquerda não foram encontradas até hoje.

As conexões entre os elementos de estrutura secundária são feitas por voltas ou por "loops". As voltas têm sempre quatro aminoácidos ($i, i+1, i+2$ e $i+3$) sendo que a distância entre $C\alpha^i$ e o $C\alpha^{i+3}$ deve ser menor que $7 \text{ \AA}$ e os resíduos ($i+1$ e $i+2$) não podem ter conformação de hélice α . De acordo com estas restrições, as voltas podem ser classificadas em termos ϕ e ψ como sendo dos tipos I, I', II e II'⁽²⁴⁾ (existem outras classificações mas estas quatro são as mais freqüentes). Algumas voltas têm preferência por certos aminoácidos como mostra a Tabela 4.

Tabela 4: Classificação das voltas.

Tipo	I	II	I'	II'	Resto
Ocorrência	41 %	26 %	6 %	6 %	27
Preferência	NENHUMA	G em i+2	G em i+2	G em i+1	NENHUMA

Os loops são trechos da sequência com comprimento de 5 a 40 aminoácidos que ficam geralmente expostos ao solvente e por isso têm uma quantidade maior de resíduos hidrofílicos. Os grupos carbonila e amino geralmente não fazem ligações de hidrogênio entre si e por isso ficam disponíveis para fazê-las com moléculas de água do solvente, íons ou outras moléculas presentes. Por estarem os loops expostos e por não terem ligações de hidrogênio regulares entre átomos da cadeia principal são geralmente mais flexíveis que os elementos de estrutura secundária. Foi verificado que as regiões de inserção e/ou deleção presentes em um alinhamento entre proteínas homólogas ocorrem quase que exclusivamente nas regiões de loops.

A variabilidade em tamanho e conformação dão aos loops também uma variedade de funções que vão desde a simples conexão de dois elementos de estrutura secundária até a ligação de proteínas inteiras (imunoglobulinas), de íons (calmodulina) e formação de sítios-ativos (enzimas).

Com o grande número de estruturas de proteínas resolvidas foi possível desenvolver uma classificação para os arranjos que determinados grupos de elementos de estrutura secundária adotam. Estes arranjos levam o nome de **domínios** e geralmente caracterizam uma função, por exemplo, as lactato desidrogenases e as gliceraldeído desidrogenases que são proteínas que ligam nucleotídeos têm no domínio de ligação um enovelamento de Rossmann, descrito na seção 1.6, que apresenta-se ideal para fazer a ligação destes nucleotídeos.

O enovelamento de Rossmann contém vários **motivos** do tipo $\beta\alpha\beta$, estes motivos ocorrem com frequência em vários tipos de domínios, como por exemplo nos barris $\alpha\beta$. Dois outros motivos encontrados em proteínas são o α -loop- α e o β -loop- β .

Uma proteína é geralmente formada por mais de um domínio, por exemplo, a piruvato quinase contém três domínios, um “sanduíche” de fitas β antiparalelas, um barril $\alpha\beta$ e uma folha de fitas paralelas formada por motivos $\beta\alpha\beta$.

Usando os elementos de estrutura secundária pode-se classificar as proteínas em três grupos, das proteínas que contém somente hélices α , das que contém somente fitas β ou das que contém hélices α e fitas β . Esta classificação mostrou-se útil, pois, associada a experimentos de dicroísmo circular, pode ajudar a prever a quantidade relativa de hélices α , fitas β e estruturas aleatórias (loops e voltas), fornecendo assim informações sobre a estrutura da proteína.

1.3.3 Estrutura e Função de Proteínas

A relação entre a estrutura tridimensional e a função de uma proteína fica evidente no caso da anemia falciforme. Nesta doença congênita ocorre uma mutação onde o GLU da posição 6 das cadeias β do tetramero $\alpha_2\beta_2$ da hemoglobina é trocado por uma VAL, esta mutação cria uma região hidrofóbica na proteína que se encaixa perfeitamente em um bolso hidrofóbico de outra hemoglobina provocando uma “polimerização” de hemoglobinas. A polimerização reduz drasticamente a solubilidade da hemoglobina desoxigenada levando a formação de precipitados fibrosos dentro das células vermelhas, o que causa a sua deformação provocando uma anemia crônica que pode levar o indivíduo à morte.

É interessante notar que esta doença ocorre apenas nos portadores homozigotos do gene que codifica a hemoglobina defeituosa, os portadores heterozigotos não apresentam qualquer sintoma. Uma pesquisa revelou que aproximadamente 40% dos negros africanos apresentam a condição de heterozigoto para o gene defeituoso, isto é explicado evolutivamente pela resistência que indivíduos heterozigotos apresentam contra a forma mais letal de malária naquela região.⁽²⁷⁾

Outro exemplo indicativo da relação estrutura-função em proteínas é encontrado na família das serino-proteases. As serino-proteases contém a

chamada tríade catalítica. Este conjunto de três aminoácidos essenciais à catálise da quebra da ligação peptídica nunca varia não importando a especificidade da reação. Das estruturas de serino-proteases resolvidas até hoje a enzima subtilisina não apresenta qualquer semelhança de sequência ou estrutura 3-D com as outras enzimas, não obstante, os aminoácidos importantes à ligação do substrato e à catálise estão praticamente na mesma conformação montando um sítio ativo igual ao das outras serino proteases. Este exemplo além de ressaltar a importância da estrutura para uma determinada função também fornece um exemplo de evolução convergente, isto é, a natureza modificou enzimas sem nenhuma relação prévia para que hoje em dia seus sítios ativos fossem praticamente idênticos.⁽²⁵⁾

1.4 Métodos empíricos para a determinação da estrutura molecular de uma proteína

Existem duas técnicas para se determinar a estrutura, a nível atômico, de uma proteína, a difração por raios-X ou nêutrons e a espectroscopia de ressonância magnética nuclear. Embora as técnicas de difração necessitem de quantidades grandes de proteína, num estado de altíssima pureza, para que possa ser obtido um cristal de proteína adequado aos experimentos de difração, ela não tem limites quanto ao tamanho da proteína. A técnica de espectroscopia de ressonância magnética nuclear (R.M.N), embora não necessite de cristais de proteína, tem seus limites no tamanho da proteína da qual se quer obter a estrutura tridimensional. Esta técnica tem sido usada na obtenção de estruturas de proteínas com até 25 KDa de massa molecular com qualidade comparável às estruturas cristalográficas de 2,0 Å de resolução⁽²⁶⁾.

1.4.1 Cristalografia de Proteínas

A maior parte das estruturas tridimensionais de proteína vem do resultado de experimentos de difração de raios-X. A técnica se baseia na absorção e emissão de raios-X pelos elétrons da proteína. Quando estes elétrons estão dispostos num arranjo periódico ordenado, como em um monocristal, a interferência entre os raios-X emitidos por cada elétron origina um padrão de difração. Este padrão de difração depende do grupo

espacial ao qual o cristal pertence e da molécula que está causando o espalhamento dos raios-X.

A experiência de difração de raios-X de um cristal qualquer resulta num conjunto de dados que contém os índices (h , k e l) das "reflexões", suas respectivas intensidades espalhadas (I) e um erro associado à medida de cada reflexão. Para se chegar a um mapa da densidade eletrônica do cristal, a partir do qual se constroa o modelo tridimensional da molécula, é necessário fazer uma transformada de Fourier usando as intensidades coletadas e suas respectivas fases. O problema surge porque as fases de cada reflexão não podem ser medidas experimentalmente. Para contornar este problema foram desenvolvidas as técnicas de substituição isomorfa múltipla e substituição molecular que permitem uma estimativa inicial das fases para cada reflexão hkl .

A técnica de substituição isomorfa múltipla requer a produção de cristais com a proteína nativa e cristais (com o mesmo grupo espacial e dimensões de cela unitária) com derivados de átomos pesados. Os cristais são coletados e com o uso do método de Patterson pode-se obter uma estimativa das fases iniciais.

A técnica de substituição molecular pode ser usada quando se sabe de antemão que a proteína da qual se quer resolver a estrutura é tridimensionalmente parecida com uma proteína que já tenha estrutura resolvida. Esta estrutura-sonda é então colocada na cela unitária do cristal, rotada e translacionada até que se encontre uma orientação adequada que forneça a estimativa inicial das fases para cada reflexão hkl .

O resultado final de uma determinação de estrutura de proteína usando a técnica de difração de raios-x é um arquivo contendo uma lista dos átomos da proteína com cinco valores experimentais, as coordenadas x , y e z , um fator de agitação térmica B e opcionalmente um fator de ocupação Occ .

Grosso modo, o fator de agitação térmica é uma medida da agitação do átomo dentro do cristal. Este valor pode ser útil para se determinar quais regiões de uma proteína são mais flexíveis e quais são mais rígidas. O fator

de ocupação é uma medida estatística da distribuição de conformações para um determinado átomo. É comum encontrar uma estrutura com duas conformações para um mesmo aminoácido (variando por exemplo no ângulo χ_1), a conformação mais freqüente leva um fator de ocupação maior enquanto que a menos freqüente leva um fator de ocupação menor. A soma dos fatores de ocupação para um mesmo átomo deve ser sempre igual a 1.

1.4.2 Ressonância Magnética Nuclear

A espectroscopia de ressonância magnética nuclear pode ser conduzida em substâncias que possuam átomos com momento nuclear magnético, geralmente estes átomos têm massa atômica ímpar (por exemplo ^1H , ^{11}B , ^{13}C , ^{15}N , ^{19}F , ^{31}P , etc). O próton pode ser considerado como uma unidade de carga que gira em torno de seu próprio eixo, como toda carga que gira este próton cria um pequeno campo magnético ao longo do seu eixo de rotação.

Se este núcleo for agora colocado sob a influência de uma campo magnético externo, ele irá precessionar em torno da direção do campo magnético de forma paralela ou antiparalela definindo os níveis de energia inferior e superior respectivamente. A energia entre estes dois estados é característica do tipo de núcleo e do campo magnético externo experimentado pelo mesmo. A aplicação de um pulso de rádio-freqüência apropriado a uma coleção destes núcleos induz a uma transição entre os dois estados que por sua vez gera uma magnetização transversal detectável experimentalmente. O valor da técnica de ressonância magnética nuclear vem do fato de que o campo magnético experimentado pelo núcleo é a soma do campo magnético externo e dos pequenos campos magnéticos gerados pelos átomos vizinhos ao núcleo que está sendo medido. O mesmo núcleo sob o mesmo campo magnético externo pode então gerar diferentes picos no espectro resultante devido aos diferentes vizinhos, esta diferença é chamada de deslocamento químico.

Existem essencialmente dois tipos de experimentos de R.M.N. bidimensionais que produzem informações bastante diferentes. No primeiro, chamado de espectroscopia correlacionada ("correlated spectroscopy" -

COSY), o experimento resulta num mapa bidimensional onde os picos definem conectividades entre prótons separados por não mais que três ligações covalentes. Estes mapas também fornecem informações sobre ângulos diedros. No segundo, chamado de espectroscopia bidimensional de melhoramento nuclear Overhauser ("2D nuclear Overhauser enhancement spectroscopy" - NOESY), o experimento fornece um mapa bidimensional onde os picos representam distâncias entre pares de prótons separados por menos de 5 Å. A intensidade de um pico fornece a distância entre os prótons que deram origem ao pico.

O resultado final para esta técnica é também um arquivo contendo uma lista dos átomos e de suas coordenadas. Neste caso entretanto não são definidos os fatores de agitação térmica e nem os fatores de ocupação. Frequentemente os arquivos de estruturas resolvidas por esta técnica apresentam não apenas uma conformação para a estrutura da proteína mas sim um conjunto de conformações que estão de acordo com os valores mensurados. A representação gráfica da superposição de todas as conformações definidas no arquivo permite, assim como no caso do fator de agitação térmica, encontrar quais regiões são mais flexíveis.

1.5 Métodos Semi-Empíricos de Determinação de Estrutura

Os métodos semi-empíricos estão baseados nos resultados anteriores de experimentos de determinação de estrutura, cinética enzimática, dicroísmo circular, mutagênese sítio-dirigida e outros. O conhecimento sobre estruturas de proteínas e seu funcionamento, adquirido com o passar dos anos e após muitos e variados experimentos, foi compilado numa série de regras que permite, hoje em dia, fazer a predição da estrutura tridimensional de uma proteína partindo apenas da sua sequência de aminoácidos.

A regra principal sobre a qual está apoiada a técnica de modelagem molecular pode ser resumida na frase:

"A estrutura tridimensional de uma proteína, no seu ambiente natural, é função de sua estrutura primária."

Este fato foi evidenciado após experimentos de desnaturação (destruição da estrutura terciária e secundária) e renaturação da ribonuclease bovina⁽²⁷⁾. Experimentos de desnaturação e renaturação de proteínas têm mostrado que tais moléculas, naturais ou artificiais, podem ser reversivelmente enoveladas e desenoveladas⁽²⁸⁾ sem a ajuda de outras enzimas. Estes experimentos implicam que em determinados ambientes, a sequência de aminoácidos é, por si só, suficiente para definir a estrutura tridimensional da proteína. Não obstante, existem proteínas que não são capazes de chegar à sua conformação final sem o auxílio de enzimas especiais que ajudam na fase final do seu enovelamento. Estas enzimas especiais são as *chaperoninas* que ligam proteínas globulares de até 90 KDa^(29,30).

Existem duas formas de se fazer a predição da estrutura terciária para uma sequência de aminoácidos, a primeira parte da informação compilada a partir das estruturas de proteína contidas no PDB resolvidas experimentalmente, a segunda usa as próprias estruturas experimentais como base na proposição do modelo para a estrutura terciária.

O resultado de análises bioquímicas e biofísicas, específicas para a proteína de interesse, podem fornecer indicações importantes para a construção do modelo da proteína. Por exemplo, experimentos de dicroísmo circular podem fornecer as proporções relativas entre os elementos de estrutura secundária, experimentos de mutagênese sítio-dirigida podem ajudar na localização de aminoácidos importantes à catálise e experimentos de cinética química indicam como está agindo um determinado inibidor proposto por desenho racional.

1.5.1 Previsão de estrutura secundária

No método de previsão de estrutura secundária as informações contidas no PDB são compiladas num conjunto de regras de modo a fornecer a probabilidade de ocorrência de cada aminoácido em cada uma das estruturas secundárias existentes (hélice alfa, fita beta ou volta). A proteína problema é

então analisada de acordo com as regras acima para prever qual a provável estrutura secundária associada a cada trecho da sequência problema.

Embora a amostragem estatística aumente dia-a-dia melhorando a confiabilidade do método, o processo de construção deste conjunto de regras é degenerativo, isto é, existe uma perda de informação quando os dados de uma estrutura terciária são compilados para informação de estrutura secundária. Por exemplo, duas fitas β com conformações tridimensionais diferentes são classificadas no conjunto de regras simplesmente como fitas β .

Com esta informação em mãos, tenta-se encontrar o melhor arranjo espacial entre os elementos de estrutura secundária que respeite a distribuição de aminoácidos para uma proteína de sua classe, a geometria comumente observada para distâncias e ângulos, contatos atômicos, etc. A estrutura da enzima interleucina-4 humana foi proposta a partir da predição de sua estrutura secundária com o auxílio de experimentos de dicroísmo circular.⁽³¹⁾

1.5.2 Modelagem por homologia

O método de modelagem por homologia não parte de um conjunto de regras elaborado a partir do PDB mas sim diretamente do PDB, ou seja, não existe a perda da informação original das estruturas no processo de construção do modelo.

O processo de construção de um modelo por homologia é baseado na hipótese de que a estrutura tridimensional é mais conservada que a estrutura primária. Isto equivale a dizer que dois trechos de proteína com sequência de aminoácidos semelhantes, mas não idênticos, devem ter uma conformação tridimensional semelhante. O alinhamento e a suposição de que sequências parecidas implicam em estruturas tridimensionais parecidas são os pontos de apoio da modelagem molecular por homologia.

Como as proteínas não trazem consigo a árvore genealógica de seus antepassados, deve existir uma maneira de se identificar as relações de homologia entre duas proteínas partindo apenas de suas sequências de aminoácidos. O parâmetro usado para saber se duas proteínas são homólogas

entre si é o grau de **identidade**, que é obtido a partir de uma comparação entre as sequências das proteínas de interesse.

A comparação entre as sequências de proteínas é feita com o programa MULTALIGN⁽³²⁾ que usa o método de Needleman e Wunsch⁽³³⁾ para encontrar partes equivalentes entre duas ou mais sequências de proteínas, o resultado deste estudo é chamado de **alinhamento**. O alinhamento pode ser entendido como uma tabela onde cada linha representa uma sequência e cada coluna, os aminoácidos topologicamente equivalentes. Ou seja, se todas as estruturas tridimensionais correspondentes às sequências alinhadas pelo programa fossem superpostas todos os aminoácidos de uma mesma coluna ocupariam a mesma posição espacial.

Se o alinhamento é feito apenas com sequências sem estrutura tridimensional disponível, o programa deve encontrar sozinho os pontos de inserção e deleção, podendo desta forma fornecer um alinhamento errôneo com inserções e/ou deleções nas regiões do esqueleto da proteína ou ainda desalinhando aminoácidos importantes para a função da enzima.

Quando estão disponíveis as estruturas tridimensionais de uma ou mais proteínas homólogas, o alinhamento é feito com base na superposição das estruturas⁽³⁴⁾, ou seja, primeiro encontram-se as posições espacialmente equivalentes (ou topologicamente equivalentes) e depois constroem-se o alinhamento. O alinhamento construído desta forma é chamado de alinhamento tridimensional. Este alinhamento tridimensional é mantido fixo enquanto que a sequência problema é alinhada contra o alinhamento tridimensional. Desta forma evita-se a criação de inserções e/ou deleções nas regiões do esqueleto da proteína.

A identidade entre duas proteínas é a razão entre o número de pares de aminoácidos idênticos, encontrado no alinhamento entre as duas sequências, pelo número total de aminoácidos alinhados. Por exemplo, na Tabela 5 o número de pares de aminoácidos idênticos é 7 e a quantidade de aminoácidos alinhados é 14, logo a identidade entre as sequências A e B é 50% ou 7/14.

Tabela 5: Alinhamento entre duas sequências imaginárias

Sequência A	A	G	I	I	K	E	K	M	E	P	W	I	D	
Sequência B	A	G	V	I	R			M	N	P	W	L	V	D

Dada a definição para o grau de identidade, podemos definir os pontos limitantes de um trabalho de modelagem por homologia em função da identidade entre a sequência da proteína problema e a sequência da proteína de base. As diferenças entre as estruturas tridimensionais da proteína problema e da proteína de base aumentam com o decréscimo da identidade sequencial, em consequência a confiabilidade do modelo diminui. Também foi verificado que a qualidade da estrutura de base influi na qualidade da estrutura modelada, ou seja, usar uma estrutura de base com resolução de 1,5 Å é melhor que usar uma estrutura resolvida a 3,5 Å.

Os desvios entre o modelo contruído por homologia e a estrutura real determinada por raios-X varia ao longo da sequência da proteína. De um modo geral, as maiores diferenças ocorrem nos loops da superfície da proteína, enquanto que os trechos de estrutura secundária e o esqueleto hidrofóbico da proteína apresentam os menores desvios.

O método para se criar um modelo por homologia varia conforme a identidade entre a sequência problema e a estrutura de base. De um modo geral o esquema apresentado na Figura 6 mostra os passos necessários para se obter um modelo. O primeiro passo na construção do modelo consiste na obtenção do esqueleto da proteína. A partir do alinhamento e da análise das estruturas de base o esqueleto pode ser obtido avaliando-se quais são as regiões estruturalmente conservadas, geralmente estas regiões são compostas pelos elementos de estrutura secundária que compõem o esqueleto da proteína e por alguns loops importantes à estrutura.

A segunda etapa consiste da modelagem dos loops. Se existir algum loop homólogo em alguma das proteínas de base o procedimento para a incorporação deste loop consiste primeiramente na superposição local dos trechos anteriores e posteriores ao loop que se deseja incorporar, e

finalmente na cópia das coordenadas deste loop incorporando-o ao arquivo do modelo.

Os loops que não forem encontrados em nenhuma das estruturas de base devem ser obtidos a partir do banco de dados PDB usando o método desenvolvido por Jones e Thirup⁽³⁵⁾ que se baseia na conformação da cadeia principal do trecho anterior e posterior ao loop que se deseja construir.

Uma vez modelados o esqueleto da proteína e os loops, o próximo passo consiste na colocação das cadeias laterais de cada aminoácido. A substituição de cadeias laterais pode ser feita por diferentes métodos dos quais os mais comuns são:

1. cálculo da posição dos novos átomos gerando uma cadeia lateral com ângulos $\chi = 180^\circ$ (rotinas REPLACE e REFI do programa TOM),
2. aproveitamento dos átomos da cadeia lateral original como um guia no cálculo da posição dos átomos da nova cadeia lateral, diminuindo assim problemas de choque com o restante da proteína (rotina MUTATE do programa WHATIF),
3. procura de rotâmeros⁽³⁶⁾ específicos para a cadeia principal local (rotina DGMUT do programa WHATIF).

Qualquer mudança usando um dos métodos acima descritos deve ser seguida de uma análise para a verificação de choques da nova cadeia lateral com o restante da proteína. Caso seja necessário, a cadeia lateral é ajustada manualmente (opção TORS do programa TOM e do programa WHATIF).

A terceira opção tem se revelado mais confiável que as duas primeiras visto que o desvio quadrático médio entre a cadeia lateral modelada e a cadeia lateral determinada cristalograficamente é menor que os desvios obtidos quando são usados os métodos 1 e 2. Neste método a nova cadeia lateral é pesquisada no banco de estruturas de proteína levando-se em conta dois critérios, a conformação da cadeia principal que se estende desde dois aminoácidos anteriores até dois posteriores ao resíduo que será trocado (num total de 5 resíduos) e o tipo do aminoácido que deverá ser colocado no lugar.^(37,38)

Uma vez terminada a substituição dos resíduos é feita uma minimização de energia de toda a estrutura para se fazer um "ajuste fino" da geometria e das interações não ligadas.

Feito isso o modelo da proteína está pronto. Em casos de baixa homologia quando a construção do alinhamento não é totalmente confiável, é recomendável submeter o modelo a diversos programas que checam a realidade e a confiabilidade da estrutura proposta. Um destes programas é o PROCHECK⁽⁵⁰⁾ que compara os parâmetros da estrutura (desvio padrão do comprimento de ligações, ângulos de ligação, quiralidade, ângulos torsionais,

etc) a parâmetros equivalentes de estruturas de alta resolução. O módulo QUALITY⁽³⁹⁾ do programa WHATIF realiza um exame do ambiente de cada aminoácido da estrutura e compara este ambiente com aqueles encontrados nas estruturas de alta resolução do PDB. Os índices de qualidade para cada aminoácido são graficados numa escala em que valores abaixo de determinado limite estão garantidamente (pelo autor do programa) errados. Também um valor médio para todo o modelo é fornecido que, se abaixo de um certo limite, indica que a proteína está erroneamente modelada.

1.6 Estrutura geral das enzimas GAPDH

As gliceraldeído-3-fosfato desidrogenases são constituídas de um homotetrâmero com cada subunidade pesando entre 35 e 40 KDa. A Figura 8:

Tetrâmero do modelo da gGAPDH de *Trypanosoma cruzi* visto ao longo do eixo Q. (Modelo final do trabalho)

Figura 9 mostra o enovelamento do monômero da GAPDH com a molécula de NAD⁺ ligada. As quatro subunidades estão agrupadas segundo 3 eixos de simetria molecular de ordem 2 como mostra a **Error! Reference source not found.** criando três tipos de interface entre as subunidades. Seguindo a nomenclatura descrita para a estrutura da GAPDH de *Bacillus stearothermophilus*⁽⁴³⁾ as subunidades recebem os códigos **O**, **P**, **Q** e **R** e os eixos que relacionam estas subunidades são chamados de P, Q e R.

Cada monômero é formado por dois domínios⁽⁴⁰⁾, o domínio de ligação do NAD⁺ e o domínio catalítico. O domínio de ligação do NAD⁺ compreende dois trechos da cadeia polipeptídica, do resíduo 1 ao 165 e do 335 ao 359 e possui uma estrutura terciária semelhante ao enovelamento de Rossmann. A topologia do domínio de ligação do NAD⁺ está mostrada na Figura 7:I. Este domínio é composto de onze fitas β ($\beta 1$, $\beta 2$, ... e $\beta 11$) das quais nove formam uma folha β e cinco hélices alfa ($\alpha 1$, $\alpha 2$ e $\alpha 5$ de um lado da folha e $\alpha 3$ e $\alpha 4$ do outro lado).

Figura 7: Representação dos dois domínios de GAPDH:
I) domínio de ligação do NAD⁺;
II) domínio catalítico.

O domínio catalítico (Figura 7:II) se estende do resíduo 166 ao 334 e é formado por uma folha β com sete fitas ($\beta 1'$, $\beta 2'$, ... e $\beta 7'$) e quatro hélices alfa ($\alpha 1'$, $\alpha 2'$, $\alpha 3'$ e $\alpha 4'$ todas de um mesmo lado da folha).

Das três interfaces, a relacionada pelo eixo $\underline{P}$ (Q/R e P/O) é a que promove a maior parte das interações entre dois monômeros. O eixo $\underline{P}$ promove a formação de um “sanduíche” entre duas folhas β de dois domínios catalíticos como pode ser visto na **Error! Reference source not found.**

A interface formada pelo eixo $\underline{R}$ (O/R e P/Q) envolve a formação de contatos nas imediações do sítio de ligação do NAD⁺.

O eixo $\underline{Q}$ (P/R e Q/O) promove apenas alguns contatos em suas interfaces, sendo que nenhum deles corresponde a regiões de interesse para o presente trabalho.

Figura 8: Tetrâmero do modelo da gGAPDH de *Trypanosoma cruzi* visto ao longo do eixo Q. (Modelo final do trabalho)

Figura 9: Monômero da gGAPDH de *Trypanosoma cruzi*. As setas representam fitas β e as espirais hélices α . Figura criada

2. MODELAGEM DA gGAPDH

2.1 Alinhamento

O primeiro passo na modelagem por homologia da enzima gGAPDH de *Trypanosoma cruzi* é a obtenção de um alinhamento entre as sequências com estrutura resolvida e a sequência da enzima problema, para isto foi necessário recuperar as estruturas resolvidas do banco de dados PDB. A busca no banco de dados PDB resultou numa lista de seis enzimas homólogas à enzima de *T. cruzi* com estrutura resolvida mostradas na Tabela 6. Das seis estruturas encontradas, as quatro primeiras foram resolvidas com o cofator NAD⁺ e as duas últimas foram resolvidas sem o cofator.

Tabela 6: Relação das proteínas homólogas à enzima gGAPDH de *Trypanosoma cruzi* com estrutura resolvida.

Código da Enzima (PDB)	Origem da Enzima (espécie)	Número de aminoácidos	Identidade com <i>T.cruzi</i>	Resolução (Å)	Fator R
1GGA ⁽⁴²⁾	<i>Trypanosoma brucei</i>	359	324/359	3,2	0,176
1GD1 ⁽⁴³⁾	<i>Bacillus stearo-thermophilus</i>	334	175/334	1,8	0,177
1GPD ⁽⁴⁴⁾	<i>Homarus americanus</i>	333	199/333	2,9	-----
HUMANA ⁽⁴⁵⁾	<i>Homo sapiens</i>	334	176/334	2,3	0,33
2GD1 ⁽⁴⁶⁾	<i>Bacillus stearothermophilus</i>	334	175/334	2,5	0,177
4GPD ⁽⁴⁷⁾	<i>Homarus americanus</i>	333	199/333	2,8	0,218

O alinhamento da sequência de *Trypanosoma cruzi* contra as sequências das enzimas homólogas encontradas no banco de dados PDB foi feito em três etapas.

A primeira etapa consistiu numa superposição aproximada das estruturas de base. Estas superposições foram feitas mantendo-se a estrutura de *Trypanosoma brucei* imóvel enquanto a segunda estrutura era rotacionada e transladada de modo a superpor-se à primeira estrutura. Esta superposição inicial foi feita com a rotina MOTIVS⁽⁴⁸⁾ do programa WHATIF que identifica trechos estruturalmente equivalentes entre duas estruturas.

Na segunda etapa a superposição inicial foi analisada e as regiões que faziam parte do esqueleto da proteína (as regiões estruturalmente conservadas) foram selecionadas para a superposição. Estas rotinas de superposição usam o método descrito por Kabsch⁽⁴⁹⁾ que encontra a melhor matriz de rotação para colocar um conjunto de vetores sobre um segundo conjunto de vetores através da minimização da soma dos quadrados das diferenças entre pares equivalentes de coordenadas. Esta superposição é usualmente feita com os átomos C α pertencentes aos trechos de estrutura selecionados.

Com as estruturas superpostas fez-se uma análise detalhada da superposição entre a estrutura de *T. brucei* com a estrutura de *B. stearothermophilus* e depois da superposição com a estrutura da enzima humana. Desta análise determinou-se os pares de aminoácidos topologicamente equivalentes entre as estruturas e contruiu-se então o alinhamento tridimensional.

O alinhamento tridimensional foi então fixado nas regiões de estrutura secundária e contra este alinhamento foi alinhada a sequência da gGAPDH de *Trypanosoma cruzi*. O alinhamento final que foi usado para a construção do modelo está mostrado na Figura 10.

O alinhamento consiste de seis blocos de quatro linhas cada donde a primeira linha refere-se à sequência de *T. cruzi*, a segunda de *T. brucei*, a terceira refere-se à sequência da enzima humana e a quarta à enzima de *B. stearothermophilus*. Neste alinhamento as regiões demarcadas com uma onda triangular definem as regiões de hélices alfa e as regiões demarcadas com uma caixa cinza definem os trechos de fita β . As posições do alinhamento demarcadas com um asterisco denotam as diferenças entre a sequência da gGAPDH de *T. cruzi* e de *T. brucei*.

Com o alinhamento em mãos foi construída a Tabela 6 que resume as relações entre as várias enzimas homólogas de estrutura resolvida com a enzima de *Trypanosoma cruzi*. Na primeira coluna está o código PDB da estrutura (exceto para a estrutura humana obtida via uma comunicação

peçoal), na segunda coluna está a espécie da qual foi extraída a enzima, na terceira coluna está o número de aminoácidos da enzima, na quarta coluna está a identidade com relação à enzima de *T. cruzi* expressa como a razão aminoácidos-idênticos/aminoácidos-alinhados, na quinta coluna está a resolução da estrutura e na sexta coluna o fator R para a estrutura.

Os valores das colunas cinco e seis definem a qualidade da estrutura, ou seja, baseando-se nestes valores pode-se optar por um determinado trecho de uma estrutura em detrimento de outra que apresenta menor confiabilidade.

2.2 Construção e Avaliação do Modelo

Toda a numeração de aminoácidos será feita de acordo com o alinhamento da Figura 10 que é a mesma para a estrutura de *Trypanosoma brucei*.

Avaliando o alinhamento da Figura 10 optou-se por escolher a estrutura de baixa resolução de *Trypanosoma brucei* como base para o modelo da enzima de *Trypanosoma cruzi*. Esta escolha foi feita por dois motivos, primeiro, a estrutura de *T. Brucei*, embora de baixa resolução, não apresentava inserções e/ou deleções com relação à sequência de *T. cruzi*, segundo estava disponível a estrutura de alta resolução de *Bacillus stearothermophilus* que mantém uma alta identidade com a enzima de *T. cruzi*, deste modo a estrutura de alta resolução de *B. stearothermophilus* foi usada como guia para tomar decisões em regiões incertas.

Figura 10: Alinhamento final usado na construção do modelo da enzima gliceraldeído-3-fosfato desidrogenase de *Trypanosoma cruzi*.

Embora o alinhamento da Figura 10 mostra-se uma identidade de 90 % entre a sequência de *T. brucei* e de *T. cruzi* sugerindo que a cadeia principal de 1GGA pudesse ser adotada de forma integral como ponto de partida para a construção do modelo, constatou-se após uma análise das estruturas de base na estação gráfica e usando o programa PROCHECK a necessidade de correções em quatro pontos da cadeia principal da estrutura 1GGA que apresentava ângulos ϕ/ψ não permitidos.

Os trechos que necessitavam de correções compreendiam os aminoácidos 2, 10-11, 70-72 e 357. Os aminoácidos 2 e 357 foram modelados usando a rotina TORS do programa TOM.

Os aminoácidos 10 ao 11 foram modelados com a estrutura 1GD1 de alta resolução através de uma superposição local dos trechos polipeptídicos anterior e posterior à região em questão.

Os aminoácidos 70 ao 72 não apresentavam loops equivalentes nas estruturas de base e portanto foram modelados com a rotina DGLOOP do programa TOM.

Após a correção da cadeia principal foi feita a mudança das 35 cadeias laterais que diferiam da estrutura de base (1GGA) para os aminoácidos correspondentes da enzima de *T. cruzi*. Esta etapa foi realizada usando-se a rotina MUTATE do programa WHATIF. A rotina MUTATE calcula as posições dos átomos da nova cadeia lateral usando os átomos da cadeia lateral original como guia.

A terceira etapa consistiu do ajuste manual de cada cadeia lateral trocada com a rotina TORS do programa TOM. O objetivo deste ajuste é acomodar a nova cadeia lateral dentro do modelo, removendo na medida do possível todos os choques que tenham resultado da troca dos aminoácidos.

A quarta etapa consistiu da colocação do cofator NAD^+ no modelo do monômero. O modelo do cofator foi obtido da estrutura de *Trypanosoma brucei*.

Feito isto o modelo do monômero estava pronto e passou-se à etapa seguinte que envolveu a criação do tetrâmero para ajustar as cadeias laterais envolvidas nas interfaces monômero-monômero. Para construir o tetrâmero foram feitas quatro cópias do modelo do monômero e cada uma delas foi superposta a uma das cadeias **O**, **P**, **Q** e **R** da estrutura de *Trypanosoma brucei*.

O modelo foi avaliado quanto a qualidade estereoquímica usando o programa PROCHECK⁽⁵⁰⁾. Este programa avalia a distribuição dos ângulos ϕ/ψ de acordo com uma estatística feita sobre um conjunto de estruturas de alta resolução extraídas do banco de dados PDB. Os ângulos são graficados num diagrama equivalente ao de Ramachandran onde as regiões de maior probabilidade para os ângulos ϕ/ψ são sombreadas. Neste diagrama os resíduos de glicina são representados por triângulos e todos os outros são representados por quadrados.

A distribuição dos ângulos ϕ/ψ para o modelo final no diagrama de Ramachandran (Figura 11) mostrou que dentro dos limites impostos à técnica de modelagem e à resolução da estrutura de base (3,2 Å) existem apenas alguns aminoácidos fora das regiões permitidas.

Figura 11: Diagrama de Ramachandran sobre a estatística do programa PROCHECK

O aminoácido ILE 2 ($\phi \sim -45$, $\psi \sim -60$; 32 % de conservação no alinhamento do Apêndice 6.1) originalmente em posição não permitida ($\phi \sim 45$, $\psi \sim -80$) foi modelado para uma região aceitável de modo a manter as ligações de hidrogênio e fazer com que a isoleucina originalmente exposta ao solvente fosse colocada para dentro de um bolso hidrofóbico.

Figura 12: Cadeia principal do modelo na região da ALA 72 antes de ser modelada (linha tracejada) e depois (linha sólida).

O aminoácido C-terminal ALA 357 ($\phi \sim 45$, $\psi \sim 140$; 8 % de conservação no alinhamento da 6.1) foi modelado para uma região permitida.

O aminoácido ALA 72 ($\phi \sim 67$, $\psi \sim -80$) está localizado em uma inserção (resíduos 67 ao 74 da Figura 10, exclusivo às enzimas de *Trypanosoma cruzi* e *Trypanosoma brucei*) que compõe um loop exposto e flexível. A ligação peptídica entre os resíduos 71 e 72 foi modelada de modo a corrigir o ângulo ϕ da ALA 72. A Figura 12 mostra a cadeia principal desta região antes (linha tracejada) e depois (linha sólida) da modelagem. Este loop foi modelado com a rotina DGLOOP do programa TOM.

O aminoácido VAL 254 ($\phi \sim 94$, $\psi \sim 142$; 88 % de conservação no alinhamento da 6.1) é o último resíduo de um loop de cinco aminoácidos (conservação média para os cinco resíduos é de 82,5 % de acordo com o alinhamento da 6.1) que conecta as fitas beta $\beta 3'$ e $\beta 4'$. Estas fitas β estão no segundo domínio e fazem parte dos contatos intermonômeros. A mesma conformação de ϕ e ψ é observada na estrutura de alta resolução de *Bacillus stearothermophilus* e portanto não foi feita nenhuma modelagem para este resíduo.

A região da cadeia principal do resíduo ARG 11 (95 % de conservação no alinhamento da 6.1) que estava originalmente em uma região não permitida ($\phi \sim 50$ e $\psi \sim 60$) foi modelada e apresenta agora ângulos permitidos. Esta modificação também colocou o amino deste resíduo em posição favorável para fazer ligações de hidrogênio com um grupo fosfato do cofator o que não acontecia na estrutura de *Trypanosoma brucei*. Esta ligação de hidrogênio ocorre em todas as outras estruturas resolvidas com o cofator. A Figura 13 mostra a conformação deste trecho de peptídeo antes (A) e depois (B) da modelagem ressaltando a ligação de hidrogênio com o cofator NAD^+ . A linha sólida representa a estrutura de *Bacillus stearothermophilus* e a linha tracejada representa o modelo.

O resíduo treonina 282 (6 % de conservação no alinhamento da 6.1) está calizado em uma região de muitos contatos e não pôde ser modelado. A posição correspondente no alinhamento da 6.1 apresenta o resíduo glicina em 90 % das sequências.

A distribuição dos ângulos χ_1 e χ_2 do modelo não mudou da estrutura

Figura 13: Cadeia principal do modelo (linha tracejada) na região de ARG 11: A) antes da modelagem e B) depois da modelagem

de base para a estrutura do modelo visto que foram feitas apenas algumas mudanças de cadeias laterais.

O diagrama da Figura 14 mostra a distribuição destes ângulos para o modelo final. As cruzes tracejadas que aparecem na Figura 14 representam combinações de χ_1 e χ_2 para as quais os choques entre a cadeia lateral e a cadeia principal são mínimos. Deve-se esperar portanto que os resíduos apresentem seus ângulos χ_1 e χ_2 ao redor destas posições.

A distribuição dos ângulos ω , que definem a planaridade das ligações peptídicas, não apresenta desvios significativos conferindo um desvio médio quadrático de $5,0^\circ$ para toda a molécula. Nenhum carbono alfa apresentou quiralidade errada.

Figura 14: Distribuição de χ_1 e χ_2 para o modelo final.

A quantidade de choques entre átomos não ligados está dentro da faixa estatisticamente aceitável para estruturas resolvidas a até 2,0 Å de resolução segundo o programa PROCHECK.

2.3 Simulações Computacionais

O módulo DISCOVER do programa INSIGHTII utilizado para realizar as minimizações de energia e dinâmica molecular contém restrições quanto ao número de átomos do sistema a ser calculado, de modo que as simulações de energia foram realizadas utilizando-se apenas o monômero da gGAPDH e não o tetrâmero como seria ideal.

O sistema simulado consiste de uma molécula do monômero de gGAPDH e uma camada de moléculas de água de 3 Å ao redor da molécula. Uma quantidade maior de moléculas de água não pôde ser adicionada devido às restrições do programa.

2.3.1 Dinâmica molecular

O tetrâmero da gGAPDH de *T. cruzi* contém vários contatos intermonômeros que são parte essencial para a estabilização da estrutura quaternária e das estruturas secundárias de interface da estrutura. Um experimento de dinâmica molecular com o monômero conduzido em condições amenas levaram ao desenovelamento das estruturas secundárias localizadas nas regiões intermonômeros mostrando que seria necessário adicionar restrições especiais para estas regiões. Como já era sabido que um dos pontos de interesse localizava-se numa região intermonômeros, e que o programa era restrito e não permitia a condução de uma dinâmica molecular com o tetrâmero, decidiu-se por abandonar este tipo de simulação.

2.3.2 Minimização de energia

A minimização de energia foi conduzida usando-se a expressão para o campo de forças CVFF⁽⁵¹⁾ (Campo de Forças de Valência Consistente) que contém parâmetros gerais (para qualquer molécula orgânica pequena) e específicos para aminoácidos. Na primeira etapa a minimização foi controlada pela técnica de "steepest descent" (derivada mais íngreme) que remove contatos interatômicos, distâncias e ângulos grosseiros.

Na última etapa a minimização passou a ser controlada pela técnica de "conjugate gradient" (gradiente conjugado) que conduz a minimização por caminhos mais coerentes levando a estrutura a um ponto de menor energia.

2.4 Comparação da gGAPDH de *Trypanosoma cruzi* com a GAPDH humana

As maiores diferenças entre as sequências da gGAPDH de *Trypanosoma cruzi* e da GAPDH humana localizadas no alinhamento da Figura 10 são as inserções 1 (23-25) e 2 (67-74) que formam loops expostos ao solvente e que não fazem contatos intermoleculares.

As diferenças mais importantes ocorrem na região de ligação do NAD^+ . As interações do grupo adenosina com a gGAPDH de *Trypanosoma cruzi* são diferentes das que ocorrem com a enzima humana. Na Tabela 7 estão listadas as diferenças entre as estruturas da enzima humana e o modelo da estrutura do *T. cruzi*. A quarta coluna é usada para identificar os sítios explorados na construção de cada inibidor, estes sítios estão denotados no alinhamento da Figura 10 por setas numeradas.

Tabela 7: Diferenças entre gGAPDH de *Trypanosoma cruzi* e GAPDH humana.

Posição†	<i>T.cruzi</i>	Humana	Sítio
36	VAL	ASN	1
38	MET	PRO	3
112	LEU	VAL	2
205	VAL	PRO	3

† De acordo com o alinhamento da Figura 10

Para se fazer a avaliação das diferenças entre as enzimas de *T. cruzi* e a enzima humana foi necessário fazer uma superposição entre as estruturas usando as coordenadas dos grupos NAD^+ da enzima humana e as coordenadas do NAD^+ da estrutura de *Trypanosoma cruzi*.

Dos aminoácidos que fazem contato com o grupo adenina, o resíduo 36 apresenta uma mutação não conservativa da enzima humana (ASN; 95 % conservada de acordo com o alinhamento da seção 6.1) para a enzima de *Trypanosoma cruzi* (VAL). Esta mutação muda a característica do bolso que

interage com a posição 2 do anel da purina criando na enzima de *T. cruzi* um bolso hidrofóbico e na enzima humana um bolso hidrofílico.

A segunda região de interesse na elaboração de um inibidor seletivo encontra-se na posição 112 do alinhamento da Figura 10. A estrutura da enzima humana apresenta uma valina nesta posição enquanto que a estrutura de *Trypanosoma cruzi* apresenta uma leucina. Embora a diferença não seja na característica química da cadeia lateral (ambos são hidrofóbicos), a leucina apresenta uma cadeia lateral maior que a valina fazendo com que a região próxima ao NAD^+ fique mais fechada.

Uma terceira região que pode ser explorada na tentativa de se obter um inibidor compreende a interface criada pelo eixo P. A enzima humana apresenta esta interface mais fechada que a região correspondente na enzima de *Trypanosoma cruzi*.

Figura 15: Superposição dos monômeros de *T. cruzi* e da enzima humana.

3. INIBIDORES ESPECÍFICOS PARA gGAPDH DE TRYPANOSOMA CRUZI

Baseados na estrutura da GAPDH de *T.brucei*, Hol e colaboradores (comunicação pessoal) tem proposto alguns compostos de partida, planejados sobre a estrutura do grupo adenosina da molécula de NAD^+ . Neste trabalho buscamos analisar estes compostos quanto à sua potencial utilização na inibição da enzima de *T. Cruzi*.

A estrutura 3-D dos inibidores foi criada com o programa INSIGHTII a partir das coordenadas do cofator NAD^+ da estrutura de *Trypanosoma brucei*. Primeiro foram removidos os átomos dos grupos nicotinamida, ribose (da nicotinamida) e fosfato. Para cada um dos inibidores foram adicionados os grupos correspondentes seguido de uma regularização de geometria. Os inibidores foram então colocados dentro do modelo da enzima de *Trypanosoma cruzi* através da superposição dos átomos dos grupos ribose e adenina.

Para poder acomodar cada um dos inibidores da melhor forma possível foram necessários ajustes dos ângulos torsionais dos inibidores e de algumas cadeias laterais envolvidas na ligação do NAD^+ .

3.1 Composto 1: 2-metil adenosina

A primeira diferença a ser explorada entre a estrutura da GAPDH humana e a gGAPDH de *Trypanosoma cruzi* ocorre na posição 36 do alinhamento da Figura 10. O aminoácido ASN da enzima humana aparece com uma conservação de 92 % de acordo com o alinhamento da seção 6.1.

Nesta posição a enzima humana apresenta uma ASN que é substituída na enzima de *T. cruzi* por uma VAL. A presença de uma VAL nesta posição cria um bolso hidrofóbico na enzima de *T. cruzi* que não existe na enzima humana. As consequências da mudança da característica química de uma cadeia lateral já foram vistas para o caso da anemia falciforme na seção 1.3.3.

Este bolso contém resíduos que participam da ligação do grupo adenina do NAD^+ . Neste caso, a adição de um grupo metila na posição 2 do anel do

grupo adenina (Figura 16) promoveria uma maior interação do inibidor com a gGAPDH de *T. cruzi* que com a GAPDH humana através de um aumento das interações de van der Waals.

Figura 16: 2-metil adenosina. Constante de inibição $IC_{50} = 700 \mu M$.

A colocação do grupo metila na posição 2 sem o deslocamento da adenosina gera choques entre átomos os quais são removidos fazendo-se um rearranjo do inibidor no sítio de ligação. Este rearranjo faz com que as interações originais adenosina/enzima fiquem enfraquecidas nas duas proteínas. A Figura 17 mostra o inibidor no sítio de ligação. Deve-se notar que as ligações de hidrogênio entre o grupo ribose e a cadeia lateral do resíduo ASP 37 (linhas tracejadas na Figura 17) são mantidas embora levemente distorcidas.

Neste caso explora-se a diferença da característica química de um aminoácido importante para a ligação do cofator à enzima. Este inibidor apresenta uma constante de inibição $IC_{50} = 700 \mu M$

Figura 17: Sítio de ligação do inibidor 2-metil adenosina. As linhas finas representam a proteína, as grossas o inibidor e as linhas tracejadas representam as interações inibidor/proteína.

3.2 Composto 2: 8-tienil adenosina

O segundo ponto a ser explorado ocorre na posição 112 do alinhamento da Figura 10. De acordo com o alinhamento da seção 6.1 a região correspondente à posição 112 (110-114) da Figura 10 apresenta uma conservação média de 88 % para os aminoácidos THR, GLY, VAL, PHE e THR da enzima humana sendo que a valina da posição 112 apresenta uma conservação de 74 %. A enzima de *T. cruzi* apresenta os aminoácidos THR, GLY, LEU, PHE e THR neste trecho, diferindo portanto em apenas um dos aminoácidos.

Apesar das duas enzimas apresentarem aminoácidos hidrofóbicos nesta posição, a enzima de *T. cruzi* contém uma cadeia lateral de leucina maior que a cadeia lateral da valina na enzima humana. A colocação de um anel hidrofóbico (tienila, Figura 18) na posição 8 do grupo adenina cria mais um contato hidrofóbico entre este inibidor e a cadeia lateral da LEU na enzima de *T. cruzi*. O mesmo contato não é observado na enzima humana porque a cadeia lateral da valina não alcança o grupo tienil.

Figura 18: Segundo inibidor: grupo tienila na posição 8 do grupo adenina. Constante de inibição $IC_{50} = 500 \mu M$.

A Figura 19 mostra o contato com a LEU 112. Note-se que o grupo tienila além de fazer interações com a leucina 112 também faz contatos hidrofóbicos com a cadeia lateral da metionina 38. A metionina 38 é responsável por varias interações entre o anel adenina e a enzima. Na proteína humana a metionina 38 é substituída por uma prolina que não faz contatos com o grupo tienila.

A colocação deste grupo na posição 8 não gera distúrbios nas interações originais entre o grupo adenosina e a proteína aumentando assim a eficiência do inibidor que pode ser verificada na constante de inibição $IC_{50} = 500 \mu M$ para a enzima homóloga de *Trypanosoma brucei*.

Figura 19: Contatos entre o grupo tienila e a enzima de *T. cruzi* mostrando as interações com LEU 112 e MET 38.

3.3 Composto 3: 2'-desoxi-2'-(3-metoxi benzamido) adenosina

O terceiro ponto a ser explorado consiste de uma região inter-monômeros. Esta região é composta dos trechos que vão de 38 a 40 (do monômero O) e de 205 a 207 (do monômero R). A região de 38 a 40 apresenta uma conservação média de 70 % para os aminoácidos PRO, PHE e ILE da enzima humana. Na enzima de *T. cruzi* os aminoácidos correspondentes são MET, ASN e THR. A região de 205 a 207 apresenta conservação média de 57 % para os aminoácidos PRO, SER e ALA da enzima humana. A região equivalente na enzima de *T. cruzi* apresenta os resíduos VAL, SER e VAL.

Figura 20: Terceiro inibidor: 2' desoxi-2'(3-metoxi benzamido) adenosina

Esta região compõe uma fenda hidrofóbica em todas as GAPDHs com estrutura resolvida até hoje. Na enzima de *T. cruzi* esta fenda apresenta-se de forma mais pronunciada permitindo a acomodação de um grupo hidrofóbico. Esta mesma fenda na enzima humana é praticamente inexistente devido a uma diferença no curso do loop definido pelos aminoácidos 38 ao 40 conforme pode ser visto na Figura 21.

A enzima humana apresenta uma prolina na posição 39 que impede o recuo do loop não permitindo a acomodação de um grupo hidrofóbico nesta fenda. O composto 2'-desoxi-2'-(3-metoxi) benzamido adenosina preenche a fenda hidrofóbica em *T. cruzi* de forma que o grupo benzamido fica entre a cadeia lateral da valina 205R e a cadeia lateral da metionina 38O fazendo extensas interações de van der Waals com estes resíduos.

Ao mesmo tempo são mantidas as ligações de hidrogênio, originalmente entre os oxigênios 2'e 3' do grupo ribose da adenosina do cofator e o resíduo de ácido aspártico 37 sem maiores alterações dos ângulos χ deste aminoácido. O oxigênio do grupo metoxi faz uma ligação de hidrogênio com a cadeia lateral do resíduo asparagina 39 enquanto que o metil aponta para um bolso hidrofóbico.

Figura 21: Superposição das enzimas humana em rosa e de *T. cruzi* em amarelo na região que compõe a fenda hidrofóbica.

As interações entre o composto 3 e a enzima podem ainda ser vistas na Figura 27 da seção 4.1.

A Figura 22 mostra uma comparação dos fatores de agitação térmica da cadeia principal da enzima humana e das enzimas de base *B. stearothermophilus* e *T. brucei* na região da fenda hidrofóbica. A região marcada com o fundo cinza delimita os resíduos que formam a fenda. Pode-se notar que na região demarcada, a enzima de *T. brucei* apresenta valores maiores que as outras duas enzimas. As regiões ao redor das caixas cinzas apresentam valores comparáveis. A alta nos fatores de temperatura para enzima do tripanossomo pode ser explicada estruturalmente pela glicina 204 que corresponde a uma serina na enzima humana e na enzima do bacilo.

Os valores de agitação térmica mais elevados na enzima de *T. brucei* na região do sítio de ligação do cofator, e também o fato do sítio de ligação do cofator se apresentar estruturalmente mais acessível por causa do recuo do loop 38-41, podem explicar a menor afinidade desta enzima pelo NAD^+ . A

menor afinidade desta enzima pelo cofator deve ser uma das causas do bom desempenho da via glicolítica no *T. brucei* e no *T. cruzi*.

Figura 22: Gráfico dos fatores de agitação térmica da região da fenda hidrofóbica. Estão graficados os valores para as estruturas de *T. brucei*, *B. stearothermophilus* e humana.

3.4 Composto 4: 2'-desoxi-2'-(2H-chromen benzamido) adenosina

O composto 4 explora a mesma fenda hidrofóbica que o composto 3. Como pode ser visto na Figura 24, os oxigênios são responsáveis por um aumento das ligações de hidrogênio entre o inibidor e a enzima. Especificamente, os oxigênios 1 e 2 fazem ligações de hidrogênio com o ácido aspártico da posição 37 enquanto que o oxigênio 3 uma ligação de hidrogênio com a cadeia lateral da asparagina 39.

As interações de van der Waals entre o inibidor e os resíduos MET 38O e VAL 205R são intensificadas pela presença de um anel hidrofóbico maior.

Figura 23: 2'-desoxi-2'-(2H-chromen benzamido) adenosina.

Figura 24: Composto 4 inserido na fenda hidrofóbica da enzima de *T. cruzi*. As linhas tracejadas denotam as ligações de hidrogênio inibidor-enzima.

3.5 Composto 5: 2'-desoxi-2'-(benzoimidazol) adenosina

O composto 5 (Figura 25) também atua na fenda hidrofóbica como os compostos 3 e 4. Neste caso o nitrogênio N do anel substituinte faz ligações de hidrogênio com o resíduo de ácido aspártico 37 e o nitrogênio NH do mesmo anel faz uma ligação de hidrogênio com a cadeia lateral da asparagina 39. O anel hidrofóbico deste composto apesar de conter menos átomos para

fazer ligações de hidrogênio com a enzima, realiza as mesmas com uma geometria melhor que o composto 4. As interações de van der Waals podem ser vistas na Figura 26 na forma dos contatos entre as superfícies pontilhadas.

Figura 25: 2'-desoxi-2'-(benzoimidazol) adenosina

Figura 26: A enzima de *T. cruzi* e o inibidor estão com a superfície de van der Waals representada, mostrando a eficiência do empacotamento.

4. UM PROGRAMA PARA VISUALIZAÇÃO DE ESTRUTURAS DE PROTEÍNAS

4.1 Introdução

A falta de espaço em disco rígido e a inabilidade dos programas gráficos, disponíveis até então, de mostrar uma estrutura rapidamente sem a necessidade de criar arquivos temporários enormes levou ao desenvolvimento deste programa.

Este programa gráfico, específico para computadores da Silicon Graphics™, permite a visualização e mensuração de uma estrutura de proteína descrita em um arquivo no formato PDB. Podem ser feitas rotações, translações, zoom, seleção de átomos, medidas de distância, ângulos diedros, anotações, etc.

O programa foi escrito em linguagem C numa plataforma UNIX e está atualmente com aproximadamente 6.000 linhas de código. Neste programa os dados de cada átomo da proteína são lidos e armazenados numa estrutura que compreende os seguintes campos: nome do átomo (char[5]), tipo de átomo (int), nome do resíduo (char[4]), tipo do resíduo (int), número interno do resíduo (char[5]), identificador de cadeia (char), identificador de confômero (char), coordenadas xyz (float[3]), fator de ocupação (float) e fator de agitação térmica (float).

As funções definidas no programa são agrupadas quanto à sua utilização. Existem funções dedicadas à entrada e saída de dados, funções para busca de informação da estrutura de dados, funções de construção de objetos gráficos, funções de cálculo e funções de interação máquina/usuário.

O programa faz uso da biblioteca gráfica ("Graphics Library") da workstation permitindo maior rapidez gráfica interativa. O programa utiliza todo o "hardware" específico ("geometry engines") para manipulação do espaço tridimensional. Atualmente o equipamento permite a realização de "depthcueing", "antialiasing" e "z-buffer" que são técnicas para o melhoramento da percepção tridimensional do objeto gráfico. O depthcueing permite que uma determinada cor perca brilho à medida em que é deslocada

para longe do usuário. O z-buffer é uma memória da coordenada z de cada pixel que foi desenhado na tela. Antes de desenhar qualquer pixel na tela o computador checa a coordenada z deste pixel e se a coordenada z do novo pixel estiver “atrás” da coordenada z antiga então o pixel não é desenhado. Em outras palavras, este equipamento é responsável pela remoção de superfícies/linhas escondidas. O antialiasing é uma técnica usada para suavizar o serrilhado das linhas que surge devido ao tamanho dos pixels.

A Figura 27 mostra a janela principal do programa e uma janela auxiliar onde está representado o diagrama de Ramachandran. Na janela principal pode-se ver o composto 3 no sítio de ligação do NAD^+ e as interações do inibidor (linhas grossas) com a enzima (linhas finas).

4.2 Usando o programa

As rotações sobre X e Y são realizadas através da movimentação da seta (“mouse”) quando o botão do meio é mantido pressionado. A translação no plano XY é obtida quando todos os botões são pressionados

Figura 27: Janela do PDBV mostrando as janelas principal e de Ramachandran e os menus.

simultaneamente ao movimento do mouse. O zoom pode ser obtido movimentando-se o mouse com os dois botões da esquerda pressionados.

Para selecionar-se um átomo, deve-se levar a ponta da seta sobre algum átomo e apertar/soltar (clique) o botão da esquerda. Feito isto aparecerá o nome do aminoácido e número do aminoácido ao lado do átomo se o átomo for um C α , caso contrário aparecerá o nome do átomo e número do aminoácido selecionado. A seleção de um resíduo quando solicitada pelo programa é feita pela escolha de qualquer átomo do aminoácido.

O programa mantém uma memória dos últimos 4 átomos selecionados com o nomes PICK1, PICK2, PICK3 e PICK4, sendo que o último átomo selecionado é o PICK1. Sobre estes 4 átomos são realizadas todas as operações do programa.

Após a leitura do arquivo PDB o programa cria automaticamente quatro objetos gráficos pré-definidos, F1:traço dos carbonos α , F2:cadeia principal, F3:tudo o conteúdo do arquivo PDB e F7: traço dos carbonos α suporte. Estes objetos podem ser ligados e desligados através das teclas de função correspondentes. Estes objetos podem ainda ser alterados durante a sessão de trabalho com funções específicas aos objetos. Isto propicia ao usuário um meio rápido de estudar a estrutura sem perder tempo criando os objetos manualmente como ocorre com outros programas do gênero.

Além dos objetos gráficos pré-definidos, existem ainda mais três objetos especiais que podem ser facilmente criados, apagados ou alterados durante a sessão de trabalho, são eles: F4:esfera de resíduos, F5:cadeias laterais e F6:resíduos.

O objeto F1 pode ser redefinido da seguinte maneira, apaga-se o seu conteúdo (menu: [C-alfas]-> [Erase all]), seleciona-se o resíduo inicial (PICK1 ou PICK2) e final (PICK2 ou PICK1) do trecho a ser adicionado ao objeto e adiciona-se o objeto (menu:[C-alfas]->[Add]). O mesmo pode ser feito com os objetos F2 (menu:[backbone]) e F3 (menu:[zones]).

O objeto F7 não pode ser alterado nunca (pode apenas ser desligado) pois é usado como suporte para a seleção de aminoácidos.

O objeto F4 (esfera de resíduos) compreende um conjunto de tipos de átomos (S) que estão a determinada distância (R) de um átomo da estrutura (PICK1). O conjunto de átomos S pode ser: 1° qualquer átomo à distância R de PICK1 (menu:[Sphere]->[atoms]), 2° todo resíduo com pelo menos um átomo à distância R de PICK1 (menu:[Sphere]->[residues]) ou 3° a cadeia lateral de todo resíduo com pelo menos um átomo à distância R de PICK1 (menu:[Sphere]->[sidechains]).

O programa permite ainda graficar todos os ângulos ϕ/ψ da estrutura (menu:[Ramachandran]) em uma janela à parte podendo-se escolher como fundo para o diagrama entre estatística de PROCHECK ou as regiões teoricamente permitidas. Um aminoácido pode ser escolhido e ter seus ângulos graficados ou o par de ângulos pode ser escolhido do diagrama para se saber a qual resíduo pertence. Dentro da janela do diagrama de Ramachandran existe um menu diferenciado que permite customizar a aparência do gráfico. É ainda possível conectar pontos sucessivamente selecionados por linhas, o que pode ajudar a classificar os tipos de voltas.

O programa é ainda capaz de criar linhas, retângulos e elipses de diferentes tipos de linha, espessura e cor. Textos aleatórios podem ser inseridos em qualquer lugar da tela usando vários tipos de letra com acentuação.

Para facilitar a localização são permitidas 10 memórias de localização (ctrl ou shift + teclas 0 a 9 no teclado numérico) que permitem voltar ao mesmo ponto depois de várias rotações.

O programa permite ainda a gravação e posterior leitura de um arquivo de texto (alterável pelo usuário) que contém tudo o que foi feito durante uma sessão de trabalho.

O arquivo gravado contém uma seção em que se podem criar 10 objetos diferentes (teclas 0 a 9 no teclado numérico). Cada objeto destes pode conter até 50 instruções para criar objetos gráficos de proteína. Cada instrução é composta pelos seguintes campos: resíduo inicial, resíduo final, tipo (zone,

c-alfa, backbone), cor (por átomo, ou uma cor qualquer), tipo de linha (solida, tracejada ou pontilhada) e espessura de linha.

Este programa está em constante desenvolvimento e suas funções e características são determinadas pela necessidade de todos os componentes do grupo de pesquisa.

5. CONCLUSÕES

Foi construído o modelo 3-D da enzima gliceraldeído-3-fosfato desidrogenase de *Trypanosoma cruzi*.

A análise deste modelo quando comparada à estrutura da enzima homóloga humana indica diferenças estruturais significativas que podem ser exploradas no desenho de drogas de maior afinidade e especificidade pela enzima do parasita.

Os compostos descritos foram desenhados na tentativa de explorar as diferenças estruturais entre as duas enzimas. Destaca-se o composto 3 que foi sintetizado (Prof. J. Perrier, Université Paul Sabatier, Toulouse, França e Prof. P. Herdewijn, Universidade de Leuven, Bélgica) e testado quanto a afinidade e constante de inibição pela enzima homóloga de *Trypanosoma brucei*, apresentando uma afinidade 45 vezes maior pela enzima do parasita que pela enzima humana⁽⁵²⁾.

Este resultado é muito animador quanto à possibilidade de utilização deste composto como molécula de partida no desenvolvimento de uma droga para a doença de Chagas.

Como resultado deste trabalho de simulação computacional temos várias propostas para compostos com potencial ação inibitória. No entanto, o processo de planejamento e produção de drogas dificilmente é linear, em verdade o processo se desenvolve num ciclo (como o apresentado na Figura 28) em que são necessários vários refinamentos até a obtenção da forma final da droga.

Como pode ser visto na Figura 28, não obstante foram criados o modelo da enzima e desenhados os potenciais inibidores, testados os inibidores quanto a sua afinidade e atividade inibitória contra o parasita, se faz necessário fechar o ciclo com a determinação experimental da estrutura do complexo desta enzima com os inibidores propostos para se saber exatamente se o modo de ligação real concorda com aquele proposto na modelagem. O conhecimento exato do modo de ligação permite continuar o ciclo na tentativa de se otimizar os compostos já propostos.

Figura 28: Esquema geral para o trabalho de desenho racional de drogas.

O trabalho de determinação da estrutura da enzima do *Trypanosoma cruzi* já está em andamento no Grupo de Cristalografia de Proteínas do Instituto de Física de São Carlos. A enzima gGAPDH de *T. Cruzi* foi clonada⁽⁵³⁾ e a expressão em grande quantidade foi feita neste laboratório. A enzima ativa foi purificada e recentemente cristalizada (A.P.U.Araújo e D.H.F. Souza, trabalho de doutoramento).

Destacamos como aspectos importantes deste trabalho o completo domínio de todas as etapas envolvidas na nova tecnologia de desenho racional de drogas baseado em estrutura. É importante ressaltar que o Grupo de Cristalografia e Modelagem Molecular de Proteínas do Instituto de Física de São Carlos já está estabelecendo colaborações com outros grupos de pesquisa em doença de Chagas no Brasil, com o objetivo de estudar outros potenciais alvos proteicos para o desenho de inibidores.

Ressalta-se também o desenvolvimento de uma ferramenta de software gráfico, o programa PDBV descrito no capítulo 4 o qual tem sido usado extensivamente pelos outros membros do grupo em seus projetos de pesquisa.

Como perspectivas futuras para a continuação deste trabalho vemos a possibilidade de otimização dos presentes compostos, em particular foi verificado que a adição de um grupo etila na posição 4 do anel aromático do grupamento benzamido no composto 3, 4 ou 5 aumentaria as interações de van der Waals inibidor/enzima.

6. APÊNDICES

6.1 Alinhamento entre 78 sequências de gliceraldeído-3-fosfato desidrogenases de várias espécies. Desenhado pelo programa ALSCRIPT⁽⁵⁴⁾

[illegible]

Table with 10 columns: Species, 110, 120, 130, 140, 150, 160, 170, 180, 190. The table lists various plant species and their corresponding values across the columns. The species names are listed in the first column, and the values are listed in the subsequent columns. The values are often represented by letters and numbers, indicating specific data points for each species.

	310	320	330	340	350	360	370	380	390								
<i>M. formicicum</i>	KEG	GLST	AEFMEYAKELGRS	RNDLFR	GVW	WESLNI	V	DGEL	YYMQAIHQES	DVVPENVD	AI	RAMLE	BMEDNPS	KS	IEKTNK	KAMGIL	L
<i>M. bryantii</i>	EGGL	TS	AGFMEYAKDLGRS	RNDLFR	GVW	WESLNI	V	DGEL	YYMQAIHQES	DVVPENVD	AI	RAMLE	BMEDNPS	KS	IEKTNK	KAMGIL	L
<i>M. ferredus</i>	EDGL	TS	AEI MEYAKDLGRS	RNDLFR	GVW	WESLNI	V	DGEL	YYMQAIHQES	DVVPENVD	AI	RAMLE	BMEDNPS	KS	IEKTNK	KAMGIL	L
<i>P. woznesi</i>	EKG	PTST	AQILFAPARDLGRS	RNDLFR	GVW	WESLNI	V	DGEL	YYMQAIHQES	DVVPENVD	AI	RAMLE	BMEDNPS	KS	IEKTNK	KAMGIL	L
<i>T. h. huacel</i>	KNI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>T. cerui</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>T. maritima</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>T. aquaticus</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>C. glutamicum</i>	GE	T	LA	YS	EF	FL	VS	CD	FN	DS	H	S	T	FD	DA	GAG	I
<i>B. steatother.</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>B. nagnidans</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>B. subtilis</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>B. megaterium</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>S. commune</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>P. chrysosporium</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>A. hispanus</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>U. maydis</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>A. hispanus</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>C. glaucopetoides</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>T. kanningii</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>C. jamaicensis</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>C. lunata</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>C. terosotrophus</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>P. anserina</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>E. nidulans</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>T. kanningii</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>M. musculus</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>C. glaucus</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>R. norvegicus</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>Saccharo</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>H. sapiens</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>G. gallus</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>C. japonica</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>H. sapiens</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>D. melanogaster</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>D. melanogaster</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>D. hydei</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>H. americanus</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>C. elegans</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>C. elegans</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>C. elegans</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>C. elegans</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>C. briggsae</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>C. briggsae</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>S. marsonii</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>C. crispus</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>S. salba</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>A. thaliana</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>A. thaliana</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>D. carophyllus</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>M. liliiflora</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>M. crystallinum</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>A. majus</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>R. acer</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>H. vulgare</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>H. vulgare</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>Z. mays</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>P. hybida</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>P. sativum</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>P. crispum</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>P. aviculare</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>P. patens</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>P. infestans</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>S. cerevisiae</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>S. cerevisiae</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>S. cerevisiae</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>K. lactis</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>Z. mays</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>T. h. huacel</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>L. mexicanus</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>E. coli</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>A. variabilis</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>S. soleracea</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>A. variabilis</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>R. sphaeroides</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>E. coli</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>A. variabilis</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN
<i>E. coli</i>	KGI	L	GYTDRRLV	ADPINDR	RS	NI	YDS	KAT	L	GN	L	PN	RR	PK	KL	VS	WYDN

7. REFERÊNCIAS BIBLIOGRÁFICAS

- ¹ Referência do programa TOM
- ² Bleasby A.J., Wootton J.C., *Prot. Engng.* **3**(3) (1990) 153-159.
- ³ Vriend G., *J. Mol. Graph.* **8** (1990) 52-56.
- ⁴ Bairoch A., Boeckman B., *Nucleic Acids Res.* **19** (1991 - Apr 25) 2247-2249.
- ⁵ Jones T.A., Zou J-Y, Cowan S.W. e Kjeldgaard M., *Acta Cryst.* **A47** (1991) 110-119.
- ⁶ Sidman K.E., George D.G., Barker W.C., Hunt L.T., *Nucleic Acids Res* **16**(5) (1988 Mar 11) 1869-1871.
- ⁷ Bernstein F.C., Koetzle T.F., Williams G.J.B., Meyer E.F.Jr, Brice M.D., Rodgers J.R., Kennard O., Shimanouchi T., Tasumi M., *J. Mol. Biol.* **112** (1977) 535-542.
- ⁸ Dayringer H.E., Tramontano A., Fletterick R.J., *J. Mol. Graph.* **4** (1986) 82-87.
- ⁹ Benson D., Lipman D.J., Ostell J., *Nucleic Acids Res.* **21**(13) (1993 - Jul 1) 2963-2965.
- ¹⁰ Pattabiraman N., Namboodiri K., Lowrey A., Gaber B.P., *Protein Seq. Data Anal.* **3**(5) (1990 - Oct) 387-405
- ¹¹ Kuntz I.D., *Science* **257** (1992) 1078-1082.
- ¹² Böhm H.J., *J. Comput. Aided Mol. Des.* **6** (1991) 61-78.
- ¹³ Goodford P.J., *J. Med. Chem.* **28** (1985) 849-857.
- ¹⁴ Shoichet B.K., Bodian D.L., Kuntz I.D., *J. Comp. Chem.* **13**:3 (1992) 380-397.
- ¹⁵ Itzstein M. *Et al*, *Nature* **363** (1993) 418-423
- ¹⁶ Antônio R. L. Teixeira, Doença de chagas e outras doenças por tripanossomos. Brasília, Editora Universidade de Brasília, 1987.
- ¹⁷ Alvaro Moncayo, Chagas Disease, em: Annual Report of Tropical Diseases Research, 1993-1992
- ¹⁸ Oppendoes F.R., Borst P., *FEBS Letters* **80** (1977) 360-364.
- ¹⁹ Oppendoes F.R., *Ann. Rev. Microbiol.* **41** (1987) 127-151.
- ²⁰ Fairlamb A.H., Oppendoes F.R., Borst P., *Nature* **265** (1977) 270-271.
- ²¹ Clarkson A.B., Brohn F.H., *Science* **194** (1976) 204-206.
- ²² Stein S., Oppendoes F.R., *Eur. J. Biochem.* (1990).
- ²³ Ramachandran G.N., Sassiexharan V., *Adv. Prot. Chem.* **28** (1968) 283-437.
- ²⁴ Wilmot C.M., Thornton J.M., *J. Mol. Biol.* **203** (1988) 221-232.
- ²⁵ Wells J.A., *et al*, *Cold Spring Harbor Symp. Quant. Biol.* **52** (1987) 647-652.
- ²⁶ Clore G.M., Gronenborn A.M., *Protein Science*, **3** (1994) 372-390.
- ²⁷ Stryer L., *Biochemistry*, pg.32, 3rd edition, W.H.Freeman and Company, New York (1988)
- ²⁸ Schulz G.E., Schirmer R.H., *Principles of Protein Structure* (1979), Springer-Velag
- ²⁹ Martin J., Hartl F.U., *Structure* **1** (1993) 161-164.
- ³⁰ Ellis R.J., *Nature* **366** (1993) 213-214.
- ³¹ Ring C.S., Cohen F.E., *The FASEB Journal* **7** (Junho de 1993) 783-790.
- ³² Barton G.J., Sternberg M.J.E., *J. Mol. Biol.* **198** (1987) 327-337.
- ³³ Needleman S.B., Wunsch C.D., *J. Mol. Biol.* **48** (1970) 443-453.
- ³⁴ Barton G.J. e Sternberg M.J.E., *Prot. Eng.* **1** (1987) 89-94.
- ³⁵ Jones T.A., Thirup S., *EMBO J.* **5** (1986) 819-822.
- ³⁶ Vriend G., Sander C., Stouten P.F.W., *Prot.Engin.* **7** (1994) 23-29.
- ³⁷ Dunbrack R.L., Karplus M., *J. Mol. Biol.* **230** (1993) 543-574.
- ³⁸ Vriend G., Sander C., Stouten F.W., *Prot. Engin.* **7**:1 (1994) 23-29.
- ³⁹ Vriend G. e Sander C., *J. Appl. Cryst.* **26** (1993) 47-60.
- ⁴⁰ Rossmann M.G., Liljas A., *J. Mol. Biol.* **85** (1974) 177-181.

- ⁴¹ Kraulis P.J., *J. Appl. Cryst.* **24** (1991) 946-950.
- ⁴² Hol W.G.J. *et al*, *Proc. Natl. Acad. Sci.* **90** (1993) 2355-2359.
- ⁴³ Skarzynski T., Moody P.C.E., Wonacott A.J., *J. Mol. Biol.* **193** (1987) 171-187.
- ⁴⁴ Moras D., Olsen K.W., Sabesan M.N., Buehner M., Ford G.C., Rossmann M.G., *J. Biol. Chem.* **250:23** (1975) 9137-9162.
- ⁴⁵ Comunicação pessoal do Dr.H.C.Watson.
- ⁴⁶ Skarzynski T., Wonacott A.J., *J.Mol.Biol.* **203** (1988) 1097.
- ⁴⁷ Murthy M.R.N., Garavito R.M., Johnson J.E., Rossmann M.G., *J.Mol.Biol.*, **138** (1980) 859.
- ⁴⁸ Vriend G., Sander C., *Proteins* **11** (1991) 52-58.
- ⁴⁹ Kabsch W., *Acta Cryst.* **A32** (1976) 922-923.
- ⁵⁰ Laskowski R.A., *J. Appl. Cryst.* **26** (1993) 283-291.
- ⁵¹ Dauber-Osguthorpe P., Roberts V.A., Osguthorpe D.J., Wolff J., Genest M., Hagler A.T., *Proteins: Structure Funtion and Genetics*, **4** (1988) 31-47.
- ⁵² Verlinde C.L.M.J. *et al.*, e Hol W.G.J., *Structure*, **2** (1994) 577-587.
- ⁵³ Hannaert V. , Opperdoes F.R. e Michels P.A.M., *Protein Expression and Purification* (1994), submetido à publicação.
- ⁵⁴ Barton G.J., *Protein Engineering* **6:1** (1993) 37-40.