UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS

**Estrutura Cristalográfica da Enzima
Triosefosfato Isomerase de *Bacillus
stearothermophilus* e Desenho
Racional de Drogas Contra a Doença
do Sono**

Luís Fernando Delboni

OK.

Tese apresentada ao Instituto de
Física de São Carlos, USP, para a
obtenção do título de DOUTOR EM
CIÊNCIAS (FÍSICA APLICADA).

Orientador: Prof. Dr. Glaucius Oliva

Departamento de Física e Informática
1995

IFSC — SERVIÇO DE BIBLIOTECA E
INFORMAÇÃO

MEMBROS DA COMISSÃO JULGADORA DA TESE DE DOUTORADO DE LUIS
FERNANDO DELBONI APRESENTADA AO INSTITUTO DE FÍSICA DE SÃO CARLOS,
UNIVERSIDADE DE SÃO PAULO, EM 01/09/1995.

COMISSÃO JULGADORA:

Prof. Dr. Glaucius Oliva

Prof. Dr. Otaciro Rangel Nascimento

Profa. Dra. Márcia Carvalho de Abreu Fantini

Prof. Dr. Paulo Arruda

Prof. Dr. Marcel Tabak

Dedico esta tese à minha esposa Kátia,
aos meus filhos João Gabriel e Pedro
Henrique e em especial à meu pai
(em memória)

Agradecimentos

Acredito ser a parte mais difícil deste trabalho, o qual só foi possível com a colaboração de um número bastante elevado de pessoas.

Primeiramente, e especialmente, gostaria de agradecer ao Dr. Wim G. J. Hol por ter-me possibilitado a experiência nunca vivenciada anteriormente, oferecendo a oportunidade de fazer parte de seu grupo de pesquisa em dois países (Holanda e Estados Unidos). A experiência tanto profissional quanto pessoal são indescritíveis. Agradeço também à sua amizade durante os dois anos que estive participando de seu grupo de trabalho. O professor Wim não foi somente um orientador, mas principalmente um educador.

Agradeço ao grupo de pesquisa do Departamento de Química e Biofísica da Universidade Estadual de Groningen, especialmente ao técnico Tjaard Pjinning, pelo esforço em compreender-me e transmitir seus sólidos conhecimentos em bioquímica. Infelizmente não conseguiria reescrever os nomes dos muito outros alunos de graduação e doutorado que lá estavam, mas posso agradecer ao calor humano com o qual fui recebido.

Agradeço ao Prof. Dr. Rik Wierenga, por ter-me dado a oportunidade de conhecer o EMBL (European Molecular Biology Laboratory) em Heidelberg, e ter-me permitido entender os processos de cristalização da triosefosfato isomerase de *T. brucei*, através do seu técnico Johan Zeller, um pessoa com características especiais.

A todos os integrantes do grupo de Biologia Estrutural, na Universidade de Washington, em Seattle, onde passei a maior parte de meu treinamento no exterior. Agradeço aos técnicos Steve Safarty, Stuart e Isolda, pela ajuda constante no laboratório e nas coletas de dados. Aos colegas que como eu mudamos da Holanda para os Estados Unidos, que são Shekhar e Sharmila Mande (e seu filho Nikkel), Focco van der Akker, Jorg Hendle, Ingeborg Feil e Christophe Verlinde.

Aos integrantes do grupo de Seattle, Profa. Ellen Adams, Prof. Ron Stenkamp, Dr. Ethan Merritt, Dr. David Teller e ao Prof. Lile Jensen.

Ao Prof. Dr. Philip Petra pela amizade e ajuda inestimável nos processos bioquímicos da purificação de proteínas.

Aos pós-doutorandos Vivien, Hidong e Xiang que estavam ou vieram a se integrar ao grupo no decorrer de minha estadia.

Aos serviços da secretária Sue English.

Findo os agradecimentos no exterior agradeço ao Prof. Dr. Glaucius Oliva, do grupo de Cristalografia de Proteínas de São Carlos, por ter-me permitido participar durante meu trabalho de doutorado do grupo do Prof. Wim Hol.

À todos os integrantes do grupo de Cristalografia de Proteínas de São Carlos, em especial à técnica Wanda pela ajuda e amizade. Também à todos aqueles que já deixaram o grupo como àqueles que vieram a se integrar posteriormente, os quais tive a oportunidade de conhecer.

Agradeço aos meus pais e sogros, e especialmente à minha família (Kátia, João Gabriel e Pedro Henrique) pela inestimável dedicação que me prestaram durante tão árduo trabalho e pela compreensão de minhas ausências durante o período em que estive envolvido com o trabalho desta tese.

Às organizações financiadoras, CNPq e OMS por terem-me permitido realizar parte de meu doutorado no exterior.

Existe muitos outros mais a agradecer e então dirigo um agradecimento generalizado a todos que contribuíram de alguma forma para a realização deste trabalho

SUMÁRIO

Lista de Figuras	i
Lista de Tabelas	v
Lista de Abreviações	viii
Resumo	ix
Abstract	xi

Capítulo 1 - Introdução

1.1 Introdução.....	1.1
1.2 Desenho Racional de Drogas Baseado na Estrutura de Proteínas do <i>Trypanosoma brucei</i>	1.2
1.3 Estabilidade Térmica de Proteína.....	1.5

Capítulo 2 - Técnicas Utilizadas em Cristalografia de Proteínas

2.1 Introdução.....	2.1
2.2 Técnicas Utilizadas nesta Tese.....	2.2
2.2.1 Expressão de Proteínas em Bactérias.....	2.2
2.2.2 Purificação das Proteínas Expressas em Bactérias.....	2.4
2.2.3 Cristalização.....	2.5
2.2.4 Coleta de Dados de Difração de Raios-X.....	2.9
2.2.5 Determinação da Estrutura Cristalográfica.....	2.11
2.2.6 Modificação da Densidade Eletrônica: Promediação e Nivelamento do Solvente.....	2.15
2.2.7 Refinamento e Análise da Qualidade do Modelo.....	2.18
2.2.8 Cálculo de Mapas de Densidade Eletrônica e Análise Gráfica.....	2.22
2.2.9 Técnicas Usadas em Desenho Racional de Drogas.....	2.23

Capítulo 3 - Experimentos de Cristalização da Frutose-1,6-Bisfosfato Aldolase de *T. brucei* e Análise Estrutural Preliminar para o Desenvolvimento Racional de Drogas Contra a Doença do Sono

3.1 Introdução.....	3.1
---------------------	-----

3.2 Comparação entre as Sequências das Aldolases Humanas e de <i>T. brucei</i> Baseado na Estrutura Tridimensional da Aldolase de Músculo Humana.....	3.4
3.3 Parte Experimental.....	3.5
3.4 Expressão em Grandes Quantidades de Aldolase de <i>T. brucei</i> em <i>E. coli</i>	3.7
3.5 Experimentos de Cristalização.....	3.11
3.6 Trabalhos Futuros.....	3.13

Capítulo 4 - Experimentos de Difusão de Inibidores da Triosefosfato Isomerase de *T. brucei*. Em busca de Inibidores Seletivos para o Desenvolvimento Racional de Drogas Contra a Doença do Sono

4.1 Introdução.....	4.1
4.2 Expressão em Grandes Quantidades da TIM de <i>T. brucei</i> em <i>E. coli</i> ...	4.14
4.3 Experimentos de Cristalização.....	4.17
4.4 Estudos de Inibidores da TIM de <i>T. brucei</i>	4.19
4.4.1 Experimentos de Difusão do Composto $C_5N_2O_5PH_{13}$ na Forma Cristalina III.....	4.22
4.4.2 Experimentos de Difusão do Composto $C_5N_2O_5PH_{13}$ na Forma Cristalina II.....	4.23
4.4.3 Experimentos de Difusão do Composto $C_{13}N_2H_{14}PO_5$ na Forma Cristalina II.....	4.26
4.4.4 Experimentos de Difusão do Composto $C_4NH_{10}PO_5$ na Forma Cristalina II.....	4.29
4.4.5 Experimentos de Difusão do Composto $C_4NH_{10}PO_5$ na Forma Cristalina I.....	4.32
4.5 Conclusão.....	4.33

Capítulo 5 - Desenho Racional de Drogas contra a Doença do Sono: Procura por Novos Inibidores da TIM de *T. brucei*, na Conformação Aberta da Estrutura em Volta Flexível

5.1 Introdução.....	5.1
5.2 Descrição do Método de Desenho Racional de Drogas.....	5.5
5.3 Descrição do Programa DOCK.....	5.6
5.3.1 Descrição Molecular: Esferas.....	5.7

5.3.2 O Problema da Complementariedade: Gráfico Bifurcado.....	5.7
5.3.3 Determinação da Pontuação.....	5.9
5.3.3.1 Pontuação por Contato.....	5.10
5.3.3.2 Pontuação por Interação Eletrostática.....	5.11
5.4 Parâmetros Usados na Busca de Compostos a Serem Ajustados na	
TIM de <i>T. brucei</i>	5.12
5.4.1 Modelo da TIM de <i>T. brucei</i> Utilizado.....	5.13
5.4.2 Cálculo da Superfície Através dos Programas MS e	
SPHGEN.....	5.13
5.4.3 Cálculo da Malha para a Pontuação.....	5.14
5.4.3.1 Cálculo da Malha para Pontuação por Contato.....	5.14
5.4.3.2 Cálculo da Malha para Pontuação por Interação	
Eletrostática.....	5.15
5.4.4 Banco de Dados de Moléculas Pequenas.....	5.15
5.4.5 Os Parâmetros Utilizados para o Programa o DOCK.....	5.15
5.5 Resultados	5.17
5.5.1 Pontuação por Contato.....	5.17
5.5.2 Pontuação por Interação Eletrostática.....	5.20
5.5.3 Pontuação Combinada.....	5.21
5.6 Conclusão.....	5.23

Capítulo 6 - Determinação da Estrutura Cristalográfica da Triosefosfato Isomerase de *Bacillus stearothermophilus* em Complexo com 2-fosfoglicolato, um Inibidor Análogo ao Substrato, à 2,8 Å de Resolução

6.1 Introdução.....	6.1
6.2 Cristalização da TIM de <i>Bacillus stearothermophilus</i>	6.3
6.3 Coleta de Dados de Difração de Raios-X.....	6.4
6.4 Determinação da Estrutura.....	6.5
6.5 Promediação da Densidade Eletrônica e Nivelamento do Solvente.....	6.14
6.6 Qualidade da Estrutura.....	6.20
6.7 Características Estruturais da TIM de <i>B. stearothermophilus</i>	6.23
6.7.1 Comparação entre os Monômeros.....	6.23
6.7.2 Sítio Ativo e Ligação do Inibidor 2PG.....	6.27
6.7.3 Comparação da TIM de <i>B. stearothermophilus</i> com as	
outras TIM de Estruturas Conhecidas.....	6.30
6.8 Conclusão.....	6.43

Capítulo 7 - Considerações Finais

7.1 Desenho Racional De Drogas.....	7.1
7.2 Estabilidade Térmica de Proteínas.....	7.3

Referências.....	R1
-------------------------	-----------

Figura 1.1 -	O metabolismo da glicose e do glicerol no glicosomo. Abreviações para as proteínas são: HK, hexoquinase, PGI, fosfoglicose isomerase, PFK, 6-fosfofrutoquinase, ALDO, frutose-1,6-bisfosfato aldolase, TIM, triosefosfato isomerase, GAPDH, gliceraldeído-3-fosfato desidrogenase, PGK, 3-fosfoglicerato quinase, GDH, glicerol-3-fosfato desidrogenase, GK, glicerol quinase. As setas maiores indicam os compostos que são trocados com o citosol, P, indica fosfato inorgânico.....	1.3
Figura 2.1 -	Representação esquemática dos métodos de cristalização por difusão de vapor.....	2.8
Figura 2.2 -	Representação dos passos de montagem do cristal em tubos capilares para coleta de dados de difração que consiste em: (1) Introdução da solução no tubo capilar, (2) introdução do cristal através da utilização de uma pipeta, (3) cristal atinge o menisco inferior, (4) retirada da solução e (5) selagem do tubo capilar	2.11
Figura 2.3 -	Representação da função G para o caso de uma distribuição esférica (Equação 2.9).	2.13
Figura 2.4 -	Representação do processo de (a) nivelamento do Solvente e (b) promediação da Densidade eletrônica. Mais detalhes no texto.	2.18
Figura 3.1 -	Reação catalisada pela frutose-1,6-bisfosfato aldolase.....	3.1
Figura 3.2 -	Perfil de eluição da aldolase de <i>T. brucei</i> pela coluna "CM Sepharose Fast Flow" do fabricante Pharmacia.....	3.9
Figura 3.3 -	Perfil de eluição da aldolase de <i>T. brucei</i> pela coluna "Sephacryl S200 HR" do fabricante Pharmacia. Somente frações do pico foram mostradas neste gráfico.....	3.9
Figura 3.4 -	Eletroforese em resina feita no equipamento da Pharmacia PHAST com as amostras do pico que foi eluído da coluna Sephacryl S200 HR (da esquerda para a direita: frações 47, 48 e 49, padrão de peso molecular, frações 50, 51, 52, 53, 54, 55 e 56, padrão de peso molecular, frações 57, 58 e 59 e padrão de peso molecular). Como pode ser observado a proteína apresenta alto grau de pureza, sendo que a parte final do pico apresenta impurezas que foram removidas. Esta coluna apresentou resultados bastante promissores quanto à qualidade do material obtido.....	3.10
Figura 4.1 -	Reação catalisada pela TIM.....	4.1
Figura 4.2 -	Vista de uma subunidade da TIM em direção ao eixo do barril. Características da estrutura são mostradas na Figura.....	4.3

Figura 4.3 -	Esquema dos inibidores de TIM: (a) 2-fosfoglicolato (2PG) não observado em complexo com TIM de <i>T. brucei</i> , (b) 3-fosfopropionato (3PP), (c) 2-fosfoglicerato (2PGA), (d) glicerol-3-fosfato (G3P), (e) 3-fosfoglicerato (3PGA), (f) 2 (N-formil-n-hidroxi-amino) etil-fosfonato (IPP), (g) fosfoglicolo-hidroxamato (PGH) e (h) N-hidroxi-4-fosfono-butanamida.....	4.9
Figura 4.4 -	A molécula DG-11 modelada para inibir seletivamente a TIM de <i>T. brucei</i>	4.12
Figura 4.5 -	Coluna de troca iônica CM-Sepharose Fast Flow (Pharmacia). Passo de purificação da TIM de <i>T. brucei</i> expressa em <i>E. coli</i>	4.16
Figura 4.6 -	Coluna de filtração S200HR-Sephacryl (Pharmacia). Passo de purificação da TIM de <i>T. brucei</i> expressa em <i>E. coli</i>	4.17
Figura 4.7 -	Compostos sintetizados por (a) Dra. Willson, (b), (c) e (d) por Jean Lee.....	4.20
Figura 4.8 -	Compostos sintetizados pelo Dr. Witmans da Universidade de Groningen.....	4.21
Figura 4.9 -	Mapa de densidade eletrônica do cristal da forma cristalina II no qual se difundiu o composto $C_5N_2O_5PH_{13}$. Mapa do tipo $2F_0-F_c$, e contorno de 1,5 x rms do mapa.....	4.26
Figura 4.10 -	Mapa de densidade eletrônica do cristal da forma cristalina II no qual se difundiu o composto $C_{13}N_2H_{14}PO_5$. Mapa do tipo $2F_0-F_c$, e contorno de 1,5 x rms do mapa.....	4.29
Figura 5.1 -	Desenho esquemático do "ciclo de desenho racional de drogas".....	5.3
Figura 5.2 -	Distância interna de complementação. Distâncias internas do ligante e do receptor são comparadas. Se a distância interna não coincide em um dado vértice, a árvore de procura é interrompida neste vértice.....	5.8
Figura 5.3 -	Interação do composto 4PBH com resíduos do sítio ativo da TIM de <i>T. brucei</i> , como determinado por técnicas cristalográficas (Verlinde <i>et al.</i> , 1992b). Interações com moléculas de água foram removidas.....	5.18
Figura 5.4 -	Interações dos compostos (a) e443 e (b) Z978 com os resíduos do sítio ativo da TIM de <i>T. brucei</i> , obtidos através do programa DOCK. Maiores detalhes no texto.....	5.19
Figura 5.5 -	Interação do composto l355 com os resíduos do sítio ativo da TIM de <i>T. brucei</i> , obtido através da pontuação combinada. Maiores detalhes no texto.....	5.22

Figura 6.1a -	Seção da função de rotação cruzada. Pico com solução $\theta_1=217,57^\circ$, $\theta_2=75,00^\circ$, $\theta_3=60,90^\circ$, seção $\theta_2=75,00^\circ$. Níveis dos contornos com intervalo de 1σ . Valor mínimo 1σ	6.6
Figura 6.1b -	Seção da função de rotação cruzada. Pico com solução $\theta_1=160,79^\circ$, $\theta_2=52,50^\circ$, $\theta_3=75,53^\circ$, seção $\theta_2=52,50^\circ$. Níveis dos contornos com intervalo de 1σ . Valor mínimo 1σ	6.7
Figura 6.2a -	Coeficiente de correlação das soluções da função de rotação cruzada através do PC-refinamento do programa XPLOR. Coeficiente de correlação da lista de soluções da função de rotação cruzada. As duas soluções se sobressaem em relação às outras obtidas pelo programa.....	6.8
Figura 6.2b -	Melhora do coeficiente de correlação das soluções da função de rotação cruzada através do PC-refinamento do programa XPLOR. Refinamento do dímero. Pode ser observe o aparecimento de mais duas soluções com alto coeficiente de correlação.....	6.9
Figura 6.2c -	Melhora do coeficiente de correlação das soluções da função rotação cruzada através do PC-refinamento do programa XPLOR. Refinamento dos monômeros. Pode ser observe o aparecimento de mais uma solução e a convergência dos valores da correlação das novas soluções para valores próximos aos valores observados para as duas primeiras soluções. Ver discussão no texto.....	6.9
Figura 6.3a -	Seção da função translação da solução 1 da função de rotação cruzada (Tabela 6.3) para o maior pico, $14,1\sigma$ acima da media. Este pico está isolado nesta seção e o próximo pico mais alto encontra-se noutra seção com apenas $4,4\sigma$ acima do valor médio.....	6.10
Figura 6.3b -	Seção da função translação da solução 2 da função de rotação cruzada (Tabela 6.3) para o maior pico, $14,1\sigma$ acima da média. Este pico está isolado nesta seção e o próximo pico mais alto encontra-se noutra seção com apenas $4,4\sigma$ acima do valor médio.....	6.11
Figura 6.4a -	Densidade eletrônica do modelo (acima) antes da promediação e nivelamento do solvente e (abaixo) depois, mostrando comparativamente a melhora do mapa de densidade eletrônica. Contorno de $1,5\sigma$	6.16
Figura 6.4b -	Densidade eletrônica do modelo final.....	6.17
Figura 6.5 -	Densidade eletrônica do modelo da TIM de <i>B. stearothermophilus</i> , após a omissão dos resíduos da estrutura em volta flexível (acima) e densidade eletrônica do modelo final (abaixo). Contorno de $1,5\sigma$	6.19
Figura 6.6 -	Diagrama de Ramachandran das duas subunidades da TIM de <i>B. stearothermophilus</i> mostrando os resíduos que encontram combinações de ϕ , ψ fora das regiões permitidas.....	6.21

Figura 6.7 -	Densidade eletrônica do inibidor 2PG no sítio ativo, contorno de $1,5 \sigma$ (acima) e interações observadas entre o inibidor 2PG e os resíduos do sítio ativo da subunidade 1 da TIM de <i>B. stearothermophilus</i>	6.22
Figura 6.8 -	Estrutura do modelo da subunidade 1 da TIM de <i>B. stearothermophilus</i> feita através do programa MOLSCRIPT (Kraulis, 1991) mostrando em detalhes os resíduos do sítio ativo, o inibidor 2PG e a estrutura em volta flexível.....	6.24
Figura 6.9 -	Gráfico do desvio entre as duas subunidades 1 e 2 quando superpostas. Diferenças maiores são visíveis na região dos resíduos Asp 54, Asp 354 e C-terminal.....	6.25
Figura 6.10a-	Gráfico dos fatores de temperatura em função dos resíduos das cadeias principal e lateral da subunidade 1.....	6.26
Figura 6.10b-	Gráfico dos fatores de temperatura em função dos resíduos das cadeias principal e lateral da subunidade 2.....	6.26
Figura 6.11 -	Análise de:(a) 62 sequências de proteínas de <i>E. coli</i> disponíveis no banco de sequências do programa GCG mostrando a distribuição da razão Arg/(Arg+Lys), a qual apresenta o valor médio de 0,47; (b) 74 sequências de proteínas de <i>B. stearothermophilus</i> disponíveis no banco de sequências do programa GCG mostrando a distribuição da razão Arg/(Arg+Lys), a qual apresenta o valor médio de 0,44. A escolha das proteínas foi feita para se ter um máximo de proteínas homólogas dentro da análise.....	6.33
Figura 6.12 -	Distribuição da razão Arg/(Arg+Lys) nas 20 sequências de TIM conhecidas (todas as sequências presentes na Tabela 6.1 e duas outras não incluídas no banco de dados, <i>B. subtilis</i> e <i>B. megaterium</i>). Esta distribuição mostra valor médio de 0,32, sendo aparente a diferença relativa à todas as razões das outras sequências de TIM e a de <i>B. stearothermophilus</i> , que apresenta a razão de 0,478 (Tabela 6.9a).....	6.34
Figura 6.13 -	Distribuição das prolinas na estrutura da TIM de <i>B. stearothermophilus</i> . Nesta figura só estão representadas estruturas secundárias próximas às prolinas, que são predominantemente hélices.....	6.39
Figura 6.14 -	Par de mutação Gly/Lys e Asn/His.....	6.42

Tabela 1.1 -	Proteínas que apresentam estrutura em forma de barril TIM.	1.7
Tabela 3.1 -	Alinhamento das sequências foi feito através do programa CLUSTAL V (Higgins <i>et al.</i> , 1994 e 1992). Códigos são : T para <i>Trypanosoma brucei</i> , A, B e C para isoenzimas humana A, B e C respectivamente; ↓ indica o resíduo envolvido na base de Schiff. Estruturas secundárias: a - hélice α ; b - fita β . Negrito e * indicam resíduos conservados e corresponde à mutações funcionalmente conservadas. ↓ indica His 361 na sequência da isoenzima humana A discutida no texto.	3.2
Tabela 3.2 -	Distâncias em Å dos resíduos ao redor do sítio de ligação do fosfato que apresentam diferenças possíveis de serem exploradas para a seletividade de inibidores. DHAP significa resíduos próximos ao sítio de ligação do fosfato da dihidroxiacetona fosfato e GAP aos resíduos próximos ao sítio de ligação do gliceraldeído-3-fosfato.	3.5
Tabela 3.3 -	Condições de cristalização para a aldolase de <i>T. brucei</i> , utilizando proteína expressa em <i>E. coli</i> em grandes quantidades.	3.6
Tabela 3.4 -	Condições de cristalização que produziram os melhores cristais que foram expostos à radiação X no sincrotron SSRL em Stanford, na Califórnia, que apresentaram padrões de difração de até 3.0 Å de resolução.	3.11
Tabela 4.1 -	Distâncias em Å entre o átomo de fósforo do complexo G3P-TIM de <i>T. brucei</i> e átomos específicos do tripeptídeo Ala400-Tyr401-Tyr402.	4.11
Tabela 4.2 -	Coleta de dados de difração de um cristal nativo da forma cristalina III.	4.23
Tabela 4.3 -	Dados de difração coletados de um cristal da forma cristalina II na qual se difundiu o composto $C_5N_2O_5PH_{13}$.	4.24
Tabela 4.4 -	Estatísticas do refinamento da estrutura da forma cristalina II na qual se difundiu o composto $C_5N_2O_5PH_{13}$.	4.25
Tabela 4.5 -	Dados de difração coletados de um cristal da forma cristalina II na qual se difundiu o composto $C_{13}N_2H_{14}PO_5$.	4.27
Tabela 4.6 -	Estatísticas do refinamento da estrutura da forma cristalina II, na qual se difundiu o composto $C_{13}N_2H_{14}PO_5$, sintetizado por Jean Lee.	4.28
Tabela 4.7 -	Dados de difração da forma cristalina II na qual se difundiu o composto $C_4NH_{10}PO_5$.	4.30
Tabela 4.8 -	Estatísticas do refinamento da estrutura da forma cristalina II na qual se difundiu o composto $C_4NH_{10}PO_5$.	4.31

Tabela 5.1 -	Lista dos compostos que apresentaram melhor resultado na pontuação combinada.....	5.22
Tabela 6.1 -	Alinhamento de sequências através do programa GCG (Genetic Computer Group, Inc., Copyright 1982-1992, Wisconsin, USA) de todas as sequências de TIM conhecidas. Em ordem: Copja, <i>coptis japonica</i> (Okada <i>et al.</i> , 1989); Maize, <i>zea mays</i> (maize) (Marchionni and Gilbert, 1986); Human, <i>homo sapiens</i> (Maquat <i>et al.</i> , 1985); Macmu, <i>macaca mulatta</i> (<i>rhesus macaque</i>) (Old and Mohrenweiser, 1988); Rabbit, <i>oryctolagus cuniculus</i> (rabbit) (Corran and Waley, 1975); Mouse, <i>mus musculus</i> (mouse) (Cheng <i>et al.</i> , 1990); Chick, <i>gallus gallus</i> (chicken) (Straus and Gilbert, 1985); Latch, <i>latimeria chalumnae</i> (latimeria) (Kolb <i>et al.</i> , 1974); Culta, <i>culex tarsalis</i> (mosquito) (Tittiger <i>et al.</i> , 1993); Drome, <i>drosophila melanogaster</i> (fruit fly) (Shaw-Lee <i>et al.</i> , 1991); Emeni, <i>emericella nidulans</i> (<i>aspergillus nidulans</i>) (Mcknight <i>et al.</i> , 1986); Schpo, <i>schizosaccharomyces pombe</i> (fission yeast) (Russell, 1985); Yeast, <i>saccharomyces cerevisiae</i> (baker's yeast) (Alber and Kawasaki, 1982); Trybb, <i>trypanosoma brucei brucei</i> (Swinkels <i>et al.</i> , 1986); Bacst, <i>Bacillus stearothermophilus</i> (Rentier-Delrue <i>et al.</i> , 1993); Ecoli, <i>escherichia coli</i> (Pichersky <i>et al.</i> , 1984); Corgl, <i>corynebacterium glutamicum</i> (Eikmanns, 1992); Morsp, <i>moraxella sp</i> (Rentier-Delrue <i>et al.</i> , 1993). A última linha de cada sequência de alinhamento mostra através de um asterístico os resíduos que são conservados em todas as sequências de TIM.....	6.2
Tabela 6.2 -	Cristalização, coleta e processamento dos dados de difração da TIM de <i>Bacillus stearothermophilus</i>	6.5
Tabela 6.3 -	Determinação da estrutura da TIM de <i>Bacillus stearothermophilus</i> por substituição molecular usando a estrutura da TIM de <i>T. brucei</i> como modelo inicial.....	6.8
Tabela 6.4 -	Refinamento do modelo da TIM de <i>B. stearothermophilus</i> por XPLOR....	6.13
Tabela 6.5 -	Promediação e nivelamento do solvente do modelo da TIM de <i>B. stearothermophilus</i>	6.15
Tabela 6.6 -	Estatística do refinamento da estrutura da TIM de <i>B. stearothermophilus</i>	6.20
Tabela 6.7 -	Relação de átomos da molécula do inibidor e da TIM de <i>B. stearothermophilus</i> que interagem e as respectivas distâncias em Å.....	6.28
Tabela 6.8 -	Estruturas de TIM de organismos mesofílicos, descritas na literatura.....	6.30

Tabela 6.9a - Razão Arg/(Arg+Lys) para as enzimas TIM de estrutura conhecida. Como pode ser visto, TIM de <i>B. stearothermophilus</i> apresenta a maior razão.....	6.30
Tabela 6.9b - Composição em porcentagem de aminoácidos por tipo nas seis sequências de TIM de estruturas tridimensionais conhecidas.....	6.31
Tabela 6.9c - Composição em porcentagem de aminoácidos nas seis sequências de TIM de estruturas conhecidas. Resultados relevantes estão em negrito.....	6.31
Tabela 6.10a- Interações das Argininas da subunidade 1 na estrutura da TIM de <i>B. stearothermophilus</i>	6.36
Tabela 6.10b- Interações das Argininas da subunidade 2 na estrutura da TIM de <i>B. stearothermophilus</i>	6.37

Lista de Abreviações

PEG -	Polietileno glicol
Fosfato -	Tampão constituído de fosfato de sódio e potássio, mistura esta que define o pH da solução
Phosphate 96 -	Soluções preparadas com tampão fosfato em várias unidades de pH assim como com vários compostos usados para promover a precipitação da proteína como PEG, sulfato de amônia e aditivos iônicos para testes de cristalização. Estas soluções são conhecidas como Matricial pois cobrem uma região vasta de condições que podem cristalizar a proteína, de uma forma racional e disposta em forma matricial para melhor interpretação dos resultados
Tris/HCl	Tampão tris-(hidroximetil)aminometano o qual foi ajustado com ácido HCl
TEA/HCl	Tampão trietanolamina o qual foi ajustado com HCl
MOPS	Ácido 3-[N-morfolino]propanosulfônico
MES	Ácido 2-[N-morfolino]etanosulfônico
fosfato/ citrato de sódio -	Tampão preparado pela mistura de tampão fosfato e citrato de sódio
NaCl	Cloreto de sódio
A.S.	Sulfato de amônia
DTT	Ditriteitol, composto utilizado para manter reduzidas as cisteínas
EDTA	ácido etileno-diamino-tetra acético, quelante de metais
SDS	Sódio dodecil sulfato
psi	Unidade de pressão, sendo dada em libras por polegada ao quadrado
rpm	Rotações por minuto
K_i , K_a , K_d e IC_{50}	Constante de inibição, constante de afinidade, constante de dissociação e metade do valor da atividade enzimática de uma enzima, obtido na presença de algum inibidor, respectivamente.

Resumo

O atual crescimento do número de estruturas tridimensionais de proteínas está produzindo um banco de dados de estruturas que é muito útil como ponto de partida para o desenvolvimento de novas moléculas relevantes do ponto de vista médico tais como drogas e vacinas. Neste trabalho, parte do projeto de desenvolvimento de novas drogas contra a doença do sono foi feita baseada nos estudos cristalográficos, de ligação de compostos e de modelagem molecular de duas enzimas glicolíticas presentes no glicossomo de *T. brucei*. Como o alvo para novas drogas é o ciclo da glicólise, o qual também é presente nos seres humanos, o novo ligante deve ser seletivo.

Frutose-1,6-bisfosfato aldolase (ALDO) é uma das proteínas alvo para o desenho racional de drogas. A análise estrutural e estudos cristalográficos da ALDO de *T. brucei* foram feitos, através do uso de proteína produzida em grandes quantidades por via recombinante. Vários possíveis sítios de seletividade foram encontrados em ALDO de *T. brucei* quando comparada às três sequências das isoenzimas humanas, tendo como base a estrutura tridimensional da isoenzima ALDO A. O braço flexível da região C-terminal é um forte alvo para compostos seletivos devido à não-conservação das sequências. Cristais foram crescidos e padrões de difração de raios-X foram obtidos até 3,0 Å, mas ainda os cristais são pequenos e mostram-se sensíveis à radiação.

Triosefosfato isomerase (TIM) é a outra proteína alvo analisada neste trabalho. A estrutura da TIM de *T. brucei* é conhecida a alta resolução. Cristais de proteína expressa em bactérias foram obtidos após extensivos experimentos. Foi identificado que usar proteína recentemente preparada é essencial para obter cristais que apresentem boa qualidade de difração. Depois de obtidos cristais de proteína expressa em grandes quantidades, foram feitos experimentos de difusão de novos inibidores, coleta de dados de difração e análises das estruturas dos possíveis complexos.

A conhecida estrutura em volta flexível da TIM fecha quando se liga o substrato ou um inibidor análogo ao substrato no sítio ativo. Desenho racional de drogas tem sido seguido usando como padrão a conformação "fechada" da estrutura em volta. No entanto, com a obtenção de um novo complexo, no qual a estrutura em volta adota a conformação "aberta", foi feita uma busca de novos compostos em bancos de dados, que não fossem derivados de fosfatos e fosfonatos, através do programa DOCK. Com este procedimento são propostos novos compostos para os quais serão utilizados no ciclo do desenho racional de drogas.

A estrutura da triosefosfato isomerase de *Bacillus stearothermophilus* que apresenta estabilidade térmica em complexo com o inibidor competitivo 2PG foi determinada por cristalografia de raios-X à resolução de 2,8 Å. A estrutura foi resolvida por substituição molecular

usando o programa XPLOD. Promediação da densidade eletrônica de ordem dois e nivelamento do solvente foram aplicados para melhorar a qualidade do mapa. Ambos os sítios ativos estavam ocupados pelo inibidor e a estrutura em volta flexível adota a conformação "fechada" em ambas as subunidades. O fator cristalográfico R é de 17,6% com boa geometria. *Bacillus stearothermophilus* é considerado um organismo de estabilidade térmica moderada. Encontram-se disponíveis cinco estruturas de triosefosfato isomerase de organismos mesofílicos. A estrutura obtida neste trabalho é a primeira TIM reportada que apresenta estabilidade térmica. Vários artigos têm listados fatores de estabilidade térmica, os quais foram analisados em todas as estruturas de TIM disponíveis. Neste trabalho foi conduzida uma comparação entre a estrutura termofílica e as mesofílicas disponíveis, da qual concluiu-se que a interação hidrofóbica na formação do dímero e o alto número de prolínas são os mais importantes fatores que contribuem para a estabilidade térmica da TIM de *B. stearothermophilus*. Também concluiu-se que a razão $\text{Arg}/(\text{Arg}+\text{Lys})$ é elevada em TIM de *B. stearothermophilus* mais devido ao baixo número de lisinas do que a um alto número de argininas. Analisou-se as argininas na TIM de *B. stearothermophilus* como também as interações das argininas e lisinas em todas as outras estruturas de TIM e não foi possível relacionar o valor aumentado da razão $\text{Arg}/(\text{Arg}+\text{Lys})$ com a estabilidade térmica, fator este indicado anteriormente como importante para a estabilidade térmica.

Abstract

The current growth in the number of known 3-dimensional proteins structures is producing a database of structures that is very useful as starting point for the development of new medically relevant molecules such as drugs and vaccines. In this work part of the project of developing new drugs against the sleeping sickness has been done based on crystallographic, ligand-binding and molecular modeling studies of two glycolytic glycosomal enzymes from *T. brucei*. As the target for new drugs is the glycolysis pathway, which is present also in the human being, the new ligand must be selective.

Fructose-1,6-bisphosphate aldolase (ALDO) is one of the target protein for the rational drug design. The structural analysis and crystallographic studies of ALDO from *T. brucei* have been done, from overexpressed protein. Several possible sites for selective inhibitor were found in *T. brucei* ALDO when compared with the three human isoenzyme sequences, based on the human ALDO A structure. The flexible C-terminal arm is a promising target for selective compounds, due to the non-conserved sequence. Crystals were grown, and X-ray diffraction was obtained up to 3.0 Å, but still the crystals are small and show radiation damage.

Triosephosphate isomerase (TIM) is the other target protein analyzed in this work. The structure of TIM from *T. brucei* is already known at high resolution. Crystals from overexpressed protein were obtained after extensive experiments. It was identified that the use of fresh protein is essential to grow good diffracting crystals. After obtaining crystals from overexpressed protein, it has been done soaking of new inhibitors, diffraction data collection and analysis of the structures of the possible complexes.

The well known flexible loop in TIM closes upon the binding of either the substrate or an inhibitor analogue to the substrate in the active site. Rational drug design has been followed based on the "closed" conformation of the flexible loop. Nevertheless, with a new complex structure, in which the flexible loop adopts the "open" conformation, it has been done a search for new compounds in the database, non-derived from phosphate and phosphonate, through the program DOCK. Based on this procedure it was proposed new lead compounds which will be used on the rational drug design cycle.

The structure of the thermostable triosephosphate isomerase from *Bacillus stearothermophilus* in complex with the competitive inhibitor 2PG was determined by X-ray crystallography to a resolution of 2.8 Å. The structure was solved by molecular replacement using XPLOR. Two fold averaging and solvent flattening were applied to improve the quality of the map. Both of the active sites were occupied by the inhibitor and the flexible loop adopts the

"close" conformation in both subunits. The crystallographic R-factor is 17.6% with good geometry. *Bacillus stearothermophilus* is considered a moderate thermophile organism. There are already five triosephosphate isomerase structures available from mesophilic organisms. The structure reported here is the first thermostable TIM reported. It has been conducted in this work a comparison between the thermophilic and the mesophilic protein structures available. Several reports had listed thermostable factors, which have been analyzed in all TIM structures available, from which it has been concluded that the hydrophobic interaction upon the dimer formation and the higher number of proline residues are the most important factors that contribute to the thermostability of *B. stearothermophilus* TIM. Also it has been concluded that the ratio Arg/(Arg+Lys) is increased in *B. stearothermophilus* TIM more due to lower number of lysine instead of a higher number of arginine residues. Arginines have been analyzed in *B. stearothermophilus* TIM as well as the interactions of arginines and lysines in all other TIM structures and it has not been possible to relate the increased ratio Arg/(Arg+Lys) with thermostability, which was previously reported as important in thermostability.

Capítulo 1

Introdução

1.1 Introdução

Segundo relatórios da Organização Mundial de Saúde ainda não existem drogas eficientes contra várias das doenças que afetam os países localizados na região tropical, como países africanos, Índia, China e países da América Latina. Várias doenças tropicais afetam milhões de pessoas nestas regiões, como malária, esquistossomose, filariose, lepra, tripanossomíase africana, doença de Chagas e leishmaniose. Há um grande esforço da Organização Mundial da Saúde em controlar ou mesmo erradicar estas doenças dos países tropicais, os quais são economicamente pobres. Estes fatores agravam em muito a situação pois muitas destas doenças poderiam ser controladas através da instrução e melhora nas condições sanitárias, o que claramente é um problema econômico e político destes países. Os meios científicos para obtenção de medicamentos eficientes contra estas doenças são, claramente, os mais difíceis por envolver um problema interdisciplinar bastante amplo. No entanto o conhecimento conquistado por tais pesquisas serão úteis para futuros desenvolvimentos de drogas que venham a ser eficientes para várias outras doenças, como vem sendo estudado em países desenvolvidos. Desta forma o trabalho em desenvolvimento de medicamentos, através do desenho racional, dentro do contexto brasileiro, tem dois objetivos básicos: o desenvolvimento científico deste campo de pesquisa assim como a solução de problemas sociais e econômicos ocasionados pelas doenças tropicais.

Atualmente, faz-se necessária uma droga mais eficiente contra as várias doenças citadas acima em vista dos dados levantados pela Organização Mundial de Saúde. A farmacologia tem apresentado um desenvolvimento bastante rápido, com o acúmulo de informações estruturais sobre proteínas e o melhor entendimento da bioquímica dos organismos causadores das doenças resultando em novas técnicas farmacológicas. As informações estruturais de proteínas, DNA e RNA tem possibilitado um amplo conhecimento do metabolismo dos organismos vivos, que permite também o entendimento do funcionamento básico de doenças assim como características particulares de organismos parasitas, que podem ser utilizadas como alvo para projetos de medicamentos.

O desenho racional de drogas é uma das técnicas que tem se desenvolvido através da cristalografia de raios X de proteínas (Capítulo 2), apoiada pela modelagem molecular em computadores, pela biologia molecular e bioquímica de proteínas e síntese química de compostos (Verlinde *et al*, 1994). A farmacologia tem apresentado desta forma revoluções devido ao desenvolvimento das técnicas de estudo dos fenômenos bioquímicos e com o desenvolvimento simultâneo dos computadores.

1.2 Desenho racional de drogas baseado na estrutura de proteínas do *Trypanosoma brucei*

Os protozoários hemoflagelados *Trypanosoma brucei gambiense* e *Trypanosoma brucei rhodesiense* (Molyneux *et. al.*, 1983) são os agentes causadores da doença do sono. Outras espécies como *Trypanosoma vivax* e *Trypanosoma congolense* são responsáveis pela tripanossomiase africana em gado, onde a doença é conhecida por "nagana". Estas doenças são transmitidas pela mosca "tse-tse" que se encontra numa área de 10 milhões de quilômetro quadrados é reconhecidamente uma limitação a produtividade desta região e um problema social extremamente sério no continente Africano.

O tripanossomo responsável pelas doenças nos seres humanos e nos animais tem um ciclo de vida que constitui em hospedeiro-mosca-hospedeiro, no qual o vetor é o inseto do gênero *Glossina*. Os diferentes estágios neste processo estão associados com diferentes características morfológicas e bioquímicas do parasita. *T. brucei* por exemplo pode ter morfologia intermediária entre dois extremos enquanto no hospedeiro mamífero: uma forma alongada presente particularmente logo após a infecção e uma forma mais compacta que desenvolve mais tarde e é a forma que é transmitida à mosca "tse-tse". A mosca adquire o parasita através da picada em animais infectados (homem, gado ou outros animais selvagens). O parasita eventualmente reside nas glândulas salivares e proventriculos do inseto na forma epimastigota e na forma metacíclica, a qual é a forma que infecta os mamíferos. O ciclo então é completo quando a mosca pica os mamíferos, infectando-os.

Anualmente 25 000 novos casos de tripanossomiase humana são reportados. Na maioria dos casos, a infecção é fatal se não tratada. Tratamento tem sido feito com drogas como triparsamida, suramina, melarsoprol ou pentamidina. Estas drogas têm sucesso moderado frequentemente apresentam efeitos colaterais (Gutteridge, 1985). Uma nova droga, α -difluoremetilornitina, é mais efetiva, mas requer altas quantidades para o completo tratamento sendo então proibida o uso sem cuidados médicos em hospitais.

As perspectivas para uma vacina contra o tripanossomo é bastante remota, desde que o tripanossomo possui um sistema sofisticado conhecido como variação antigênica (Borst, 1986) que consegue resistir aos ataques do sistema imunológico através da variação da glicoproteína antigênica da superfície (conhecida como VSG) entre um número de alternativas possíveis codificada pelo genoma do parasita. Por este motivo existe a necessidade de desenvolvimento de drogas mais eficientes contra o tripanossomo.

O primeiro projeto no mundo a abordar uma doença tropical utilizando técnicas de modelagem molecular baseada em estruturas tridimensionais de proteínas, desenvolvido no grupo

do Dr. Wim Hol, na Universidade de Washington, Seattle, WA, USA, no intuito de desenhar drogas racionalmente contra proteínas bioquimicamente importantes do tripanossoma *Trypanosoma brucei*. O grupo do Dr. Fred Opperdoes, do International Institute of Cellular and Molecular Pathology (ICP), Bruxelas, na Bélgica tem estudado o *T. brucei* para o entendimento dos mecanismos bioquímicos deste organismo. Opperdoes e Borst (Opperdoes e Borst, 1977) descobriram a existência de uma organela neste organismo, que mantém nove enzimas do ciclo da glicólise e o metabolismo do glicerol, chamada glicosomo. O ciclo metabólico catalisado pelas enzimas contidas no glicosomo pode ser visto na Figura 1.1. Esta organela é uma característica única de tripanossomatídeos, que pode ser explorada no desenho racional de drogas. Características particulares da constituição da membrana assim como mecanismo de importação das enzimas podem ser exploradas no intuito de desenvolver uma droga que venha a interferir com esses processos. No entanto o mecanismo de importação das enzimas não é claramente entendido até o momento.

Figura 1.1 - O metabolismo da glicose e do glicerol no glicosomo. Abreviações para as proteínas são: HK, hexoquinase, PGI, fosfoglicose isomerase, PFK, 6-fosfofrutoquinase, ALDO, frutose-1,6-bisfosfato aldolase, TIM, triosefosfato isomerase, GAPDH, gliceraldeído-3-fosfato desidrogenase, PGK, 3-fosfoglicerato quinase, GDH, glicerol-3-fosfato desidrogenase, GK, glicerol quinase. As setas maiores indicam os compostos que são trocados com o citosol. P_i indica fosfato inorgânico.

T. brucei apresenta uma taxa de consumo de glicose 50 vezes maior que a dos mamíferos (Wang, 1984) quando na forma encontrada na circulação sanguínea no hospedeiro humano. Esta

forma do tripanossomo, conhecida como tripomastigota, consome cerca de 85 nmol de glicose por minuto por miligrama de proteína. Desta taxa, e assumindo que a proteína constitui 10% (p/p) do protozoário (Misset *et al.*, 1986), pode ser estimado que o tripanossomo consome a quantidade de glicose igual ao seu peso em 7 horas.

Na forma encontrada na circulação sanguínea, a mitocôndria foi reduzida a um canal periférico sem um ciclo de Krebs funcional. Também o tripanossomo desta espécie não possui formas de estoque de energia metabólica, como carboidratos e moléculas com grupos fosfato de alta energia .

Isto aumenta a vulnerabilidade deste organismo para compostos que venham a afetar o ciclo da glicólise (Oppenheimer, 1983). Isto foi comprovado através de teste tanto *in vitro* (Clarkson e Brohn, 1976) como *in vivo* (Oppenheimer e Borst, 1977) da adição de uma mistura de ácido salicílico-hidroxâmico e glicerol

Devido à importância deste ciclo metabólico no tripanossoma, decidiu-se concentrar nas enzimas glicossomais como alvos potenciais para o desenho de drogas. Este ciclo metabólico oferece várias oportunidades para o desenho de novos compostos terapêuticos, seguindo um ciclo de desenho racional de drogas (Figura 5.1), que podem ser:

- (i) Explorar o fato de que as proteínas glicossomais têm que ser translocadas para dentro do glicossomo. Interferência com o processo de passagem através da membrana do glicossomo é uma das possibilidades já que não existe este tipo de organela no organismo humano. No entanto o mecanismo de importação das proteínas no glicossomo ainda não é claramente entendido;
- (ii) Explorar diferenças nos centros catalíticos das enzimas glicossomais e as enzimas humanas análogas. Com o conhecimento do mecanismo de catálise, da sequência de aminoácidos e da estrutura tridimensional pode ser possível conseguir uma inibição seletiva das proteínas do tripanossomo, deixando inafetada ou levemente afetada as enzimas humanas homólogas. Este é o enfoque utilizado nesta tese;
- (iii) Interferir na formação dos oligômeros (todas as enzimas glicolíticas exceto fosfoglicerato quinase são oligoméricas). Compostos que interfiram na formação dos oligômeros mas não afetem o processo de oligomerização nas enzimas humanas homólogas podem ser fortes candidatos para drogas;

Existem vários outros processos metabólicos ou características destes organismos que podem ser exploradas. O enfoque dado ao item (ii) foi devido à importância do ciclo da glicólise

e também à disponibilidade de conhecimento suficiente para a obtenção de proteína pura em quantidades adequadas para o desenvolvimento do projeto.

Duas proteínas foram estudadas neste trabalho: frutose-1,6-bisfosfato aldolase (ALDO, Capítulo 3) e triosefosfato isomerase (TIM, Capítulos 4 e 5) de *T. brucei*. A estrutura da TIM foi determinada a alta resolução por Wierenga *et al.* (Wierenga *et al.*, 1991), e vários complexos com inibidores que se ligam ao sítio ativo foram obtidos (veja Capítulo 4 desta tese). O projeto somente pôde ser continuado com a expressão das enzimas em grandes quantidades em bactérias, que proporcionou resolver problemas básicos quanto a cristalização da proteína. Vários compostos foram difundidos em cristais no intento de obter maiores informações a respeito dos modos de interação com a proteína no seu sítio ativo.

O desenho racional de drogas tem ganho muito com a organização de bancos de dados de compostos disponíveis comercialmente, assim como informações estruturais dos mesmos. A disponibilidade de programas também tem contribuído para este desenvolvimento. Estes programas podem executar uma busca nos bancos de dados e testar cada composto em sítios definidos pelo usuário, na proteína de interesse, com a intenção de encontrar algum composto que satisfaça as condições de interações com a proteína no referido sítio. Vários resultados têm sido obtidos e uma discussão bastante detalhada de um dos programas (DOCK) é feita no Capítulo 5. Neste Capítulo o programa DOCK é aplicado usando a TIM de *T. brucei* como proteína alvo e um banco de dados com aproximadamente 40.000 compostos foi usado para encontrar algumas moléculas que podem interagir favoravelmente com o sítio ativo.

ALDO de *T. brucei* foi a parte do projeto que apresentou maiores dificuldades. A cristalização desta proteína ainda oferece resistência quanto à qualidade dos cristais. O projeto avançou de forma apreciável após a expressão em grandes quantidades das enzimas em bactérias, depois da qual cristais melhores foram obtidos e difrataram, ainda que insuficiente para a determinação da estrutura.

1.3 Estabilidade térmica de proteína

Vários fatores têm sido reportados como importantes para a estabilidade térmica de proteínas como ligações de hidrogênio, pontes salinas e hidrofobicidade (Fontana, 1991, Creighton, 1991, Matthews, 1993). Os fatores de estabilidade térmica variam de proteína para proteína significando que talvez as proteínas atinjam a estabilidade térmica pelo acúmulo de

modificações pequenas que melhoram a estabilidade das subunidades assim como também dos oligômeros. Segundo os resultados até agora obtidos não existem evidências de fatores únicos de estabilidade térmica, significando que cada proteína é um caso particular e deve ser analisado como tal.

Várias estruturas de proteínas de organismos que apresentam estabilidade térmica foram determinadas e as características particulares discutidas. A estrutura da triosefosfato isomerase de *Bacillus stearothermophilus* é determinada nesta tese (Capítulo 6). A estrutura da triosefosfato isomerase foi primeiramente determinada por Barner *et. al.* (1975), a qual apresentou um motivo estrutural novo, conhecido atualmente como barril TIM. Esta estrutura pode ser vista na Figura 4.2 (Capítulo 4) e é formada por uma repetição de um motivo volta-fita β -volta-hélice α , tal que as oito fitas β formam um barril de fitas paralelas fechado no interior da proteína, circundado por oito hélices α . Este motivo estrutural tem sido observado em várias outras proteínas apresentando diferentes funções, entre elas a triosefosfato isomerase (que também da nome ao barril) e a aldolase, discutida na seção anterior (Brändén, 1991). Devido a observação em várias proteínas (Tabela 1.1) e o grande interesse no desenho de novos inibidores para o desenho de drogas este motivo estrutural foi e ainda encontra-se em estudo.

A estrutura da triosefosfato isomerase de vários organismos é conhecida, sendo apenas de organismos mesofilicos (não estáveis a alta temperatura). *Bacillus stearothermophilus* era conhecido como um organismo termofílico (estável a temperatura mais elevada) mas atualmente com o conhecimento de um grande número de bactérias que vivem em águas com temperaturas superiores a 100°C e a alta pressão, *B. stearothermophilus* é considerado um organismo de estabilidade térmica moderada. Mesmo assim a elucidação da estrutura de proteínas do *B. stearothermophilus* pode fornecer maiores informações a respeito dos processos utilizados por proteínas como fatores de estabilidade térmica.

Vários fatores de estabilidade térmica são conhecidos. Hidrofobicidade é um dos fatores mais importante tanto em estabilidade quanto direcionar o enovelamento de proteínas. Termodinamicamente, a entropia aumenta com o enovelamento da proteína, onde resíduos hidrofóbicos são retirados do meio hidrofílico, formando um centro hidrofóbico. A análise do interior e das interações hidrofóbicas fornece um importante fator de comparação entre estruturas homólogas para o entendimento de aumento de estabilidade térmica. No caso de estruturas oligoméricas, a análise das interações entre cadeias pode fornecer informações sobre a estabilidade. Outros tipos de interações como ligações de hidrogênio, pontes salinas, pontes dissulfeto são fatores que contribuem para estabilidade de proteínas. Estruturas secundárias normalmente são formadas preferencialmente por certos resíduos que tornam-na mais estáveis.

Tabela 1.1 - Proteínas que apresentam estrutura em forma de barril TIM

Proteína	Designação	Referência
Triosefosfato isomerase	TIM	Banner <i>et. al.</i> , 1975
2-queto-3-deoxi-6fosfogluconato aldolase	KDPGA	Mavridis <i>et. al.</i> , 1982
Piruvato quinase	PK	Muirhead <i>et. al.</i> , 1986
ciclodextrina glicosil transferase	CGT	Klein e Schulz, 1991
flavocitocromo b2	FCB2	Xia e Mathews, 1990
ribulose-1,5-bisfosfato carboxilato N-(5'-fosforibosil)-antranilato isomerase	RUBISCO	Schneider <i>et. al.</i> , 1990
Adenosine desaminase	PRAI	Priestle <i>et. al.</i> , 1991
Mandelato racemase	ADA	Wilson <i>et. al.</i> , 1991
Apo-enolase	MR	Neidhart <i>et. al.</i> , 1990
D-xilose-isomerase	ENO*	Stec e Lebioda, 1990
Aldolase	XYI	Henrick <i>et. al.</i> , 1989
α -amilase	ALDO	Gamblin <i>et. al.</i> , 1991
Glicolato oxidase	AMY	Buisson <i>et. al.</i> , 1987
Trimetilamina desidrogenase	GAO	Lindqvist, 1986
subunidade α da triptofano sintetase	TMDH	Lim <i>et. al.</i> , 1986
indolo-3-glicerolfosfato sintetase	TRPS α	Hyde <i>et. al.</i> , 1988
Celobioidroglase II	IGPS	Priestle <i>et. al.</i> , 1991
Enzima muconato lactonizante	CBH	Rouvinen <i>et. al.</i> , 1991
Aldose redutase	MLE#	Goldman <i>et. al.</i> , 1987
	AR‡	Roudeau <i>et. al.</i> , 1992

* $\beta\beta\alpha\alpha-(\beta\alpha)_6$
7 fitas
‡ Falta uma hélice

Vários autores calcularam um fator de propensão de cada resíduo em estruturas secundárias como hélice α . Ala, Val e Ile são preferencialmente encontrados em hélices α sendo que Pro e Gly mais frequentemente encontrados em voltas. Pro terminam hélices devido a quebra na sequência de ligações de hidrogênio característica desta estrutura secundária, e por outro lado diminui o grau de liberdade, portanto introduzindo um grau de rigidez maior a estrutura, e

portanto fornecendo maior estabilidade térmica a estrutura. Cavidades normalmente decrescem a estabilidade de proteínas pois aumentam a superfície hidrofóbica exposta ao solvente. Vários outros fatores são observados como estabilizantes, muitos sendo particular de cada estrutura quando comparada com suas homólogas.

Foi neste contexto que a estrutura da TIM de *B. stearothermophilus* foi determinada neste trabalho e analisada comparativamente com as estruturas das outras TIM de organismos mesofílicos (Capítulo 6). Apesar de *Bacillus stearothermophilus* ser um organismo de estabilidade térmica moderada, foi possível, através da análise comparativa, observar fatores que são responsáveis pela estabilidade térmica aumentada desta proteína.

Capítulo 2

Técnicas Utilizadas em Cristalografia de Proteínas

2.1 Introdução

Em cristalografia de proteínas faz-se necessário o uso de várias técnicas. Para se entender esta necessidade deve-se notar que para obter a estrutura de uma proteína, precisa-se primeiramente obter uma proteína de interesse, que normalmente é um trabalho realizado por bioquímicos de proteínas ou biólogos. Estes em alguns casos fornecem a proteína em um grau de pureza adequado para experimentos de cristalização mas, algumas vezes, alguns passos adicionais de purificação devem ser realizados. Com o desenvolvimento da biologia molecular é possível introduzir o gene que expressa a proteína em bactérias ou outros sistemas, os quais então passarão a produzir a proteína de interesse em grandes quantidades.

Com proteína pura inicia-se o processo de cristalização, o qual é basicamente uma técnica bioquímica onde uma solução é preparada, na qual a solubilidade da proteína é decrescida lentamente, passando pelo ponto de precipitação da proteína, a qual em muitos casos produz monocristais com qualidade para coleta de dados de difração de raios-X.

Após ter dados coletados, dependendo se já existe uma proteína homóloga com estrutura conhecida ou não, utiliza-se as técnicas de substituição molecular ou substituição isomorfa múltipla, respectivamente, para determinar a estrutura. Estas duas técnicas têm bases cristalográficas bastante distintas e ultimamente, com o grande número de estruturas determinadas, a técnica da substituição molecular tem sido mais amplamente usada.

Tendo determinado a estrutura, inicia-se o processo de refinamento do modelo, monitorando a qualidade do mesmo através da comparação dos dados coletados com os dados calculados através do modelo. Existem vários pacotes de programas, os quais usam diversas técnicas para o refinamento de estruturas, assim como existem diversos pacotes de programas para cálculos de mapas de densidade eletrônica, os quais são utilizados para o aprimoramento dos modelos, assim como para a observação de modificações ocorridas com a ligação de um composto à proteína ou para a análise local de mutação sítio dirigida. Análises dos modelos são feitas visualmente em computadores com ótima qualidade gráfica, através de pacotes de programas que permitem a visualização da densidade eletrônica e também permitem ao usuário introduzir modificações nas posições dos resíduos, ângulos de torção, posições de átomos, etc.

Com um modelo da proteína refinado e convergido para a melhor solução, pode-se passar à fase de análise das características da proteína ou mesmo ao uso do modelo para uma busca de compostos que venham a se ligar algum sítio de interesse da proteína, o qual vai ser então posteriormente usado para novas análises cristalográficas. Todas estas técnicas são largamente usadas por cristalógrafos de proteínas. Desta forma faz-se necessário introduzir

sumariamente estas técnicas, principalmente as que foram usadas nesta tese.

2.2 Técnicas Utilizadas nesta Tese

Para facilitar a descrição das técnicas, far-se-á uma divisão em tópicos, os quais serão: Expressão de proteínas em bactérias, purificação, cristalização, coleta de dados de difração, determinação da estrutura cristalográfica, modificação da densidade eletrônica, refinamento e análise da qualidade do modelo, cálculo de mapas de densidade eletrônica e análise gráfica e finalmente técnicas usadas em desenho racional de drogas.

2.2.1 Expressão de Proteínas em Bactérias

A biologia molecular tem evoluído muito nos últimos 20 anos depois do trabalho pioneiro de Paul Berg, Herbert Boyer e Stanley Cohen que desenvolveram a tecnologia do DNA recombinante (Jackson *et al.*, 1972, Cohen *et al.*, 1973, Cohen, 1985) e a expressão de proteínas em bactéria, a qual tem permitido um grande avanço em outras áreas.

Esta expressão inicialmente começa com a identificação do fragmento de DNA que codifica a proteína de interesse. A técnica conhecida como PCR (polymerase chain reaction, ou seja, reação em cadeia da polimerase) é em alguns casos utilizada para amplificar o DNA, ou seja, produzir cópias em grandes quantidades de um padrão inicial de DNA. Desta forma pode-se então obter o fragmento de interesse em grandes quantidades.

Este fragmento de DNA é então covalentemente ligado a um **vetor** de DNA. Uma característica essencial de um vetor é que ele pode se replicar de forma autônoma em um hospedeiro apropriado. Plasmídeos (ocorrem naturalmente em círculos de DNA que atuam como um cromossomo acessório) e fago λ (um vírus) são escolhidos como vetores para clonagem em *E. coli*.

Bactérias mutantes que não degradam rapidamente DNA estranho são frequentemente usadas como células hospedeiras, as quais pegam moléculas de DNA naturalmente do meio. Como o processo não é muito eficiente, as células alvo podem ser infectadas com o DNA de interesse através do uso de partículas de vírus rearranjadas para conter a molécula de DNA recombinante,

como foi citado acima.

O passo final então é selecionar quais as células que contém o DNA recombinante de interesse. Isto pode ser feito através da introdução nos plasmídeos de vetores que conferem resistência a um antibiótico, o qual pode ser usado para eliminar as células não transformadas. Existem outros métodos de isolamento mas este foi o utilizado nesta tese.

Em maiores detalhes os dois casos de expressão de proteínas utilizados nesta tese, para triosefosfato isomerase (TIM) e frutose-1,6-bisfosfato aldolase (ALDO) , foram construídos usando os mesmos hospedeiros e vetores, os quais serão descritos.

O vetor usado para a expressão de ambas as proteínas foi o pET3A. Este vetor se encontra sob o controle de um promotor T7 (o qual dirige a transcrição através do promotor *lacUV5*), que é induzido por isopropil- β -D-tiogalactopiranosida (IPTG). Este vetor foi introduzido na *E. coli* (DE3)pLysS, uma linhagem que contém o gene que expressa T7 RNA polimerase. A presença de IPTG na cultura deste sistema induz a transcrição da T7 RNA polimerase que por sua vez transcreve o DNA de interesse no plasmídeo. Este sistema apresenta também gene para a expressão da T7 lisozima, uma proteína que liga à T7 RNA polimerase e inibe a transcrição anteriormente à indução (para maiores detalhes do sistema usado veja Studier *et al.*, 1990).

A expressão é feita em culturas de bactérias na presença do meio LB, o qual contém 1% de tripton, 0,5% de extrato de levedura e 0,5% de cloreto de sódio, que basicamente consiste em alimento para as bactérias e fonte de material necessário para a produção de proteínas. Antes da expressão propriamente dita, é feita uma cultura em meio sólido em placas contendo 1,6% de ágar em meio LB, previamente autoclavados. Após a autoclavagem o meio encontra-se na forma líquida, ao qual são adicionados os antibióticos necessários para selecionar somente as bactérias que possuem o plasmídeo de interesse. Esta solução é colocada em placas estéreis, e quando resfriada, solidifica-se. Sobre este meio, bactérias são espalhadas cuidadosamente, através de movimentos em zig-zag para que somente colônias simples de apenas uma linhagem possam se desenvolver. Este sistema é incubado a 37°C durante a noite.

Destas colônias são retiradas as bactérias para iniciar a expressão propriamente dita. Uma parte desta colônia é adicionada em meio LB com antibióticos para se iniciar uma pré-cultura. Esta cultura é incubada durante toda a noite sob agitação a 200 rpm e a 37°C. No próximo dia esta cultura é distribuída em um volume bem maior, normalmente em 1 litro de meio LB e incubada sob as mesmas condições. Após obter uma quantidade razoável de bactérias, a qual é medida por absorção óptica da solução, o indutor IPTG é adicionado para iniciar a expressão da proteína de interesse. Depois de expressa, inicia-se o processo de purificação.

2.2.2 Purificação das Proteínas Expressas em Bactérias

A proteína de interesse encontra-se dentro das bactérias e precisa, inicialmente, ser removida. Porém um cuidado deve ser seguido aqui, pois além da proteína de interesse, a bactéria produz proteínas importantes para seu funcionamento, que incluem, entre outras, as proteases. No processo de lise das células (quebra das paredes celulares para liberar as proteínas) as proteases também serão liberadas e encontrar-se-ão em presença de todas as outras proteínas inclusive a expressa em grande quantidade. Para evitar a degradação da proteína de interesse, adiciona-se inibidores de proteases como PMSF (floreto de fenil-metil-sulfonil) entre outros.

A lise (ou quebra das paredes celulares) pode ser feita de várias maneiras, como através do uso de sonicadores e através do uso de prensas. O método usado nesta tese foi a prensa, chamada de prensa francesa. As células que encotram-se na presença do meio LB são centrifugadas e o sobrenadante é retirado. As células que estão no precipitado são ressuspensas em tampão na presença de inibidores de protease. Esta solução é introduzida dentro de um pistão, o qual possui uma torneira com uma esfera plástica ou de metal ao redor da qual a solução irá sair. Antes de abrir a torneira o pistão é submetido a uma alta pressão. Quando a torneira é aberta, a solução passa ao redor da esfera sob alta pressão, fazendo com que as paredes celulares se rompam. Desta forma obtém-se a lise.

O material lisado agora contém pedaços de parede celular que podem ser precipitados por centrifugação e o sobrenadante conterà a proteína de interesse juntamente com várias outras proteínas da bactéria.

Existem vários métodos para a purificação de proteínas. Um primeiro passo é o fracionamento por sulfato de amônia. Diferentes proteínas possuem diferentes solubilidade em sulfato de amônia. Desta forma estas diferenças podem ser usadas como um passo de purificação, o qual é usado neste trabalho.

Depois deste primeiro passo de purificação inicia-se outros métodos, com colunas de cromatografia, as quais consistem de uma coluna contendo uma resina com características particulares que ligam determinadas proteínas com diferentes afinidades as quais posteriormente são retiradas através da introdução de um composto que venha a competir com os sítios de ligação, liberando as proteínas em ordem de menor para maior afinidade, separando assim as proteínas. Este é o processo usado na cromatografia de troca iônica. A coluna usada nesta tese foi a CM Sepharose Fast Flow, do fabricante Pharmacia. Esta resina tem as características acima mencionadas e é um trocador de cátions, ou seja, o material da coluna, no pH de uso, apresenta cargas negativas. Então toda a proteína que apresenta uma carga líquida positiva irá se ligar à

coluna, sendo que as proteínas que neste pH apresentam carga líquida negativa não irão se ligar à coluna. Este passo é extremamente importante na purificação de proteínas glicosomais de *T. brucei*, pelo fato destas proteínas apresentarem um ponto isoelétrico elevado (entre 9,0 e 11,0), acima da grande maioria das proteínas da bactéria, possibilitando então a remoção de um número apreciável de proteínas da bactéria neste passo da purificação.

O passo seguinte de purificação usa um método completamente diferente, ainda sendo cromatografia, mas agora de filtração, o qual simplesmente separa as proteínas por pesos moleculares, retardando as moléculas menores e deixando as maiores passarem mais livremente, devido a presença de poros na matriz da resina da coluna, o qual retarda proteínas de tamanhos diferentes. A resina utilizada neste passo foi a S200 do fabricante Pharmacia.

Para as duas proteínas deste trabalho somente estes três métodos de purificação foram suficientes para obtê-las num grau elevado de pureza. A análise da pureza da proteína foi feita através da técnica de eletroforese, a qual consiste em uma separação por peso molecular. Isto se faz adicionando SDS à solução, que distribui igualmente cargas negativas sobre a proteína, a qual é desnaturada (através da adição de β -mercaptoetanol e levando à fervura) para que a carga líquida seja proporcional ao peso molecular. Aplica-se a proteína em uma resina de poliacrilamida com densidade conhecida e submete-se a resina a um campo elétrico, o qual faz com que as proteínas se movam com velocidades proporcionais aos pesos moleculares, devido à resistência ao movimento imposta pela resina de poliacrilamida.

Antes de iniciar os trabalhos de cristalização, as proteínas são submetidas a testes de atividade enzimática. Enzimas catalisam reações e para verificar sua integridade realiza-se testes de sua atividade como catalisadores das reações. Desta forma os testes de atividade são realizados para verificar quão íntegras estão as proteínas. Cada proteína tem o seu protocolo de ensaio de atividade enzimática, o qual não é o intuito deste trabalho, desta forma não será descrito. Estes protocolos podem ser acompanhados em maiores detalhes para ALDO em Callens *et al.* (Callens *et al.*, 1991) e para TIM em Lambier *et al.* (Lambier *et al.*, 1987).

2.2.3 Cristalização

Para a utilização da técnica de difração de raios-X na determinação de estrutura de proteínas deve-se primeiramente obter cristais. A técnica de obtenção de cristais de proteínas não difere muito da obtenção de cristais de pequenas moléculas mas há mais variáveis envolvidas assim

como cuidados específicos que devem ser seguidos. O primeiro cristal de proteína reportado foi obtido por Bernal e Crowfoot, do qual obtiveram o primeiro padrão de difração da pepsina (Bernal e Crowfoot, 1934). Desde este primeiro resultado muitas macromoléculas de várias famílias têm sido cristalizadas (Gilliland, 1988) pelo uso principalmente de métodos empíricos e sem nenhum controle racional do mecanismo de crescimento. No início dos trabalhos em biologia estrutural o interesse se concentrava principalmente no desenvolvimento da técnica de raios-X para a determinação de estruturas mais do que na racionalização da cristalização. Como atualmente a cristalografia é uma técnica bem estabelecida, a produção de cristais de proteínas que difratam a alta resolução é o fator limitante. Com o desenvolvimento da biologia molecular (veja seção 2.2.1) e da engenharia de macromoléculas através de mutações sítio dirigidas, a necessidade de cristais de macromoléculas que difratem a alta resolução tem aumentado grandemente.

A cristalização de materiais biológicos é um processo multiparamétrico, envolvendo os três clássicos passos de nucleação, crescimento e terminação do crescimento. Fatores envolvidos no processo de cristalização são: supersaturação (a qual pode ser atingida através da utilização de diferentes compostos como sulfato de amônia, cloreto de sódio, polietilenoglicol, glicerol, etc), temperatura, pH, taxas de equilíbrio, força iônica e pureza dos compostos químicos usados, difusão e convecção, volume e geometria das amostras, partículas sólidas, paredes e interfaces, densidade e viscosidade, pressão, campos elétricos e magnéticos; parâmetros bioquímicos e biofísicos das macromoléculas como sensibilidade da conformação à parâmetros físicos, presença de ligantes, aditivos especiais, propriedades como oxidação e idade das amostras; parâmetros relacionados à pureza como contaminantes, heterogeneidade na sequência de aminoácidos, heterogeneidade conformacional, etc.

A grande diferença entre cristalização de moléculas pequenas e macromoléculas biológicas consiste basicamente na flexibilidade e versatilidade conformacional e consequentemente maior sensibilidade às condições externas das macromoléculas. Para crescer cristais de qualquer composto, as moléculas têm que ser levadas a um estado de supersaturação estado termodinamicamente instável, o qual pode ser desenvolvido em uma fase cristalina ou nos casos desfavoráveis, em fase amorfa, quando o sistema retorna ao equilíbrio. A supersaturação pode ser atingida pela lenta evaporação de um solvente ou variando alguns dos parâmetros mencionados acima. Devido à necessidade de uma força iônica e de um pH definidos para estabilidade e funcionalidade das macromoléculas, os cristais devem ser obtidos de soluções mais complexas ao invés de soluções puramente aquosas, normalmente usadas para o crescimento de cristais de pequenas moléculas. A cristalização se inicia por uma fase de nucleação (que é a formação do primeiro agregado ordenado) a qual é seguida por uma fase de crescimento. A

nucleação requer um grau maior de supersaturação do que o crescimento, e a taxa de cristalização aumenta quando a supersaturação aumenta. Desta forma nucleação e crescimento deveriam ser processos separados. A terminação do crescimento de cristais pode ter várias causas: defeitos no crescimento e envelhecimento da amostra são algumas delas.

Os cristais de macromoléculas são mantidos por interações como ligações de hidrogênio, pontes salinas, interações de van der Waals, interações entre dipolos, somente nas regiões de contato, pois por serem grandes e normalmente globulares, elas interagem através de uma pequena região da superfície, sendo que grande parte da superfície das macromoléculas interage com solvente que constitui normalmente cerca de 50% do cristal, mas existem casos de cristais com valores tão baixos quanto 28% de solvente como também cristais com valores tão altos quanto 80% de solvente. Um número bastante usado em cristalografia de proteínas que envolve a quantidade de solvente nos cristais foi introduzido por Matthews (Matthews, 1968). Este número indica a razão entre o volume da unidade assimétrica pelo peso molecular da proteína contida nela (V_m), sendo portanto dado em $\text{\AA}^3\text{Da}^{-1}$. Normalmente este valor encontra-se por volta de $2,7 \text{\AA}^3\text{Da}^{-1}$ mas existem casos de proteínas com valores variando de $1,7 \text{\AA}^3\text{Da}^{-1}$ até $6,15 \text{\AA}^3\text{Da}^{-1}$. Através do V_m é possível obter o conteúdo percentual em volume da proteína e do solvente no cristal. Estes valores são obtidos pelas relações:

onde $V_{\text{proteína}} \times 100$ é o volume percentual ocupado pela proteína no cristal e $V_{\text{solvente}} \times 100$ é o volume percentual ocupado pelo solvente. Desta forma, em geral encontra-se cristais com volumes de

$$V_{\text{proteína}} = \frac{1,23}{V_m} \quad e \quad V_{\text{solvente}} = 1 - \frac{1,23}{V_m} \quad (2.1)$$

solvente tão baixos quanto 28% assim como valores tão altos quanto 80%. Claro que alguns casos mais raros podem ser observados com valores fora destes limites mais frequentemente observados. Casos onde os valores de solvente são bastante elevados podem dificultar a coleta de dados mas por outro lado permite a melhora da densidade eletrônica através do nivelamento do solvente (seção 2.2.6).

A constituição dos cristais de macromoléculas biológicas torna-os frágeis ao manuseio, assim como sensíveis ao solvente. Mudanças nas características do solvente, como aumento ou diminuição de um composto presente na cristalização podem quebrar ou dissolver os cristais.

Com a crescente necessidade de cristais de macromoléculas que difratem à alta resolução, técnicas de cristalização mais racionais têm surgido. Diálise é um dos métodos que permite uma variação fácil dos diferentes parâmetros que influenciam na cristalização de

macromoléculas biológicas. Estas moléculas são separadas do solvente por uma membrana semipermeável a qual permite a passagem de pequenas moléculas como íons, aditivos, tampões, etc, mas não permite que as macromoléculas biológicas circulem através da membrana.

Outra técnica, e provavelmente a mais usada, é a difusão de vapor (a qual foi a predominante neste trabalho). A técnica constitui em misturar uma gota da solução contendo a macromolécula biológica e tampão, com uma gota do agente cristalizante, tampão e aditivos e então equilibrar contra um reservatório contendo a solução do agente cristalizante em uma concentração mais elevada do que a da gota. O equilíbrio vai ser atingido pela difusão de espécies voláteis na solução (como água e solventes orgânicos) até a pressão de vapor na gota se igualar à do reservatório. Três métodos são usados nesta técnica, "hanging drop" (gota pendurada), "sitting drop" (gota sentada) e "sandwich drop" (gota entre duas superfícies) que são apresentadas na Figura 2.1.

Figura 2.1 - Representação esquemática dos métodos de cristalização por difusão de vapor

Outra técnica conhecida é o banho, no qual o agente cristalizante e a macromolécula biológica são misturados tal que a supersaturação é atingida imediatamente.

Difusão em interfaces também é uma técnica usada, no caso difusão entre duas fases líquidas, uma delas sendo o agente cristalizante e a outra sendo a solução de macromolécula biológica.

Tendo-se de obtido cristais de macromoléculas biológicas, há varias circunstâncias nas quais é necessária a troca da solução na qual estes foram crescidos. Mudando a solução do cristal é possível ligar inibidores, ativadores, substratos, produtos e átomos pesados à macromolécula,

a qual pode ter suficiente grau de liberdade para sofrer algumas mudanças conformacionais em resposta a esses compostos.

Complexos de macromoléculas podem ser obtidos por co-cristalização, diretamente da solução, ou por difusão do ligante no cristal já formado, colocando-se o cristal em uma solução que contenha o ligante de interesse. Cada um dos métodos tem suas vantagens e desvantagens. Enquanto a difusão simplifica o problema cristalográfico, pois permite simplesmente analisar as diferenças causadas pela ligação do ligante a uma forma cristalina de estrutura já conhecida, a co-cristalização pode mudar a conformação da proteína a tal ponto de interferir na solubilidade da proteína em relação a conformação não ligada, que poderá não cristalizar ou cristalizar numa forma não isomorfa.

Neste trabalho a difusão é usada extensivamente para troca de tampões e introdução de inibidores em cristais previamente obtidos por difusão de vapor. Os cristais são transferidos da solução de cristalização para soluções diferentes nas quais o cristal se mostra estável. Isto é feito neste trabalho pois a forma cristalina obtida para TIM de *T. brucei* possuía uma quantidade muito elevada de sulfato que é um inibidor do sítio ativo. Após esta transferência, os cristais são retransferidos para soluções contendo os inibidores de interesse, que irão difundir através dos muitos canais existentes no cristal e possivelmente se ligar ao sítio do inibidor na proteína.

Maiores detalhes a respeito da técnica de cristalização pode ser obtida através do livro "Crystallization of Nucleic Acids and Proteins, A practical approach", editado por A. Ducruix e R. Giegé, The Practical Approach Series, Oxford University Press.

2.2.4 Coleta de Dados de Difração de Raios-X

Tendo-se obtido os cristais, e após estes serem crescidos a dimensões adequadas, ou mesmo após a difusão de compostos de interesse, pode-se montá-los para a coleta de dados de difração. Este é um passo que exige bastante habilidade manual. Não existe um método padrão e cada pessoa desenvolve sua própria técnica.

A montagem de cristais para coleta de dados de difração inicia com a escolha dos cristais. Cristais crescidos através da metodologia descrita na seção 2.2.3 podem ser melhorados em termos de tamanho a partir do processo de "seeding", que consiste em utilizar pequenos cristais como semente para um processo de aumento das suas dimensões. Este método vem sendo largamente utilizado em cristalografia de proteínas por possibilitar a utilização de condições de

cristalização já obtidas com sucesso e também por separar os processos de nucleação e crescimento dos cristais. Cristais de dimensões entre 0,2 a 0,5 mm são de tamanho razoável para coleta de dados em geradores convencionais. Caso não se obtenha cristais de tais dimensões (normalmente se não são das dimensões acima são menores) então faz-se necessário o uso de raios-X de maior energia, conseguidas em fontes de radiação síncrotron. Neste trabalho tanto as fontes convencionais, como as fontes síncrotron foram usadas. A escolha do cristal passa por uma análise visual da qualidade das faces, se existem trincas e mais detalhadamente por uma análise através de microscópio óptico equipado com polarizador e analisador.

Monocristais assim analisados extinguem a luz bem definidamente quando girados em intervalos de 90° . Escolhido o melhor cristal, prepara-se os capilares usados para montagem. Estes capilares possuem uma parede de vidro bem fina para evitar absorção dos raios-X. Algumas vezes é importante lavar o capilar com uma solução a base de silicone para evitar que alguns cristais crescidos em solventes muito viscosos venham a se movimentar durante a coleta de dados.

Os capilares são então vedados em uma das extremidades e com o auxílio de uma seringa o líquido mãe (que muitas vezes não é o mesmo utilizado na cristalização, sendo geralmente 5% mais concentrado em termos do agente precipitante) é introduzido no capilar. O cristal escolhido então é transferido, através do uso de uma pipeta, da solução em que se encontra para o capilar. Espera-se o cristal afundar no solvente até o menisco inferior (Figura 2.2). Com o auxílio da mesma seringa usada anteriormente retira-se a maior parte do solvente. O uso de papel de filtro pode ser recomendado para secar o excesso de solvente ao redor do cristal. Coloca-se um pouco de solvente próximo ao cristal para manter a pressão de vapor dentro do capilar, e então evitar que o cristal perca água ou solventes voláteis e venha a se quebrar. Fecha-se a outra extremidade com cera ou mesmo com óleo de alta pressão de vapor.

Após a montagem, o capilar é colocado numa cabeça goniométrica, que permite três graus de liberdade angulares, e três graus de liberdade translacionais. Esta cabeça goniométrica é colocada entre o colimador do gerador de raios-X, e o detector de área. Inicia-se o processo de coleta de dados. Detalhes sobre a coleta de dados não é o objetivo desta tese e podem ser acompanhados nos manuais dos diversos detectores de área. Basicamente os cristais são alinhados através de algumas fotos e depois os dados são coletados, sendo que o intervalo angular é baseado no grupo espacial. Passos importantes no processamento são integração das intensidades difratadas e posteriormente o cálculo da média entre reflexões equivalentes, que vai fornecer o R_{sym} ou R_{merge} , valor que mostra a qualidade do padrão de difração do cristal. Baixos R_{sym} , evidenciam bons cristais e bons dados quando o número de reflexões coletadas encontra-se em torno de quatro vezes ou mais o das reflexões únicas.

Figura 2.2 - Representação dos passos de montagem do cristal em tubos capilares para coleta de dados de difração que consiste em: (1) Introdução da solução no tubo capilar. (2) introdução do cristal através da utilização de uma pipeta. (3) cristal atinge o menisco inferior. (4) retirada da solução e (5) selagem do tubo capilar.

2.2.5 Determinação da Estrutura Cristalográfica

Várias técnicas encontram-se disponíveis para a determinação de estruturas de proteínas dependendo de cada caso. Substituição Molecular, Substituição Isomorfa Múltipla, Dispersão Anômala Múltipla, Microscopia Eletrônica, entre outras, podem ser combinadas para a determinação de estruturas. Não existe um procedimento padrão quando o caso é especial. O pesquisador deve usar as técnicas adequadas e tentar todos os métodos possíveis caso não consiga resultados satisfatórios e, é claro, usar o bom senso.

Substituição molecular é empregada quando a proteína em estudo tem uma proteína homóloga com estrutura tridimensional já conhecida. Proteínas homólogas geralmente apresentam estruturas tridimensionais bastante similares. A substituição molecular resume-se a um problema com três estágios que são: (i) determinação da orientação relativa (rotação) de estruturas não conhecidas porém com estruturas homólogas, no mesmo ou em diferentes cristais, conhecida como *função rotação*; (ii) determinação da posição da estrutura conhecida no sistema cristalino da estrutura a ser determinada, conhecida com *função translação*, e (iii) determinação das fases usando o conhecimento das operações não cristalográficas obtidas nos itens (i) e (ii).

O item (i), para o caso em questão da substituição molecular como método de determinação de estrutura, consiste em encontrar a orientação de uma estrutura molecular conhecida numa cela unitária do cristal de interesse que apresente o melhor acordo possível entre os dados de difração calculados pelo modelo e os dados de difração observados, ou seja, a molécula conhecida é usada para procurar na cela não conhecida um padrão similar.

Este processo pode ser representado por equações. Para isto inicialmente deve-se introduzir a função de Patterson:

$$P(uvw) = \int_{v.c.u.} \rho(x,y,z) \rho(x+u,y+v,z+w) dV \quad (2.2)$$

onde v.c.u é o volume da cela unitária, e ρ é a densidade eletrônica.

Como a densidade eletrônica, ρ , é a transformada de Fourier dos fatores de estrutura F, então a função de Patterson pode ser escrita como:

$$P(\bar{u}) = \frac{1}{V^2} \sum_{\bar{h}} \sum_{\bar{h}'} F_{\bar{h}} F_{\bar{h}'} \exp(-2\pi i \bar{h}' \cdot \bar{u}) \int_{v.c.u.} \exp[-2\pi i (\bar{h} + \bar{h}') \cdot \bar{x}] dV \quad (2.3)$$

A integração é diferente de zero somente quando $\bar{h} = -\bar{h}'$, quando então o valor da integral é igual ao volume da cela unitária V. Logo a função de Patterson é:

$$P(\bar{u}) = \frac{1}{V} \sum_{\bar{h}} |F_{\bar{h}}|^2 \cos(2\pi \bar{h} \cdot \bar{u}) \quad (2.4)$$

Que simplifica para:

$$P(uvw) = \frac{2}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=0}^{\infty} F(hkl)^2 \cos 2\pi(hu + kv + lw) \quad (2.5)$$

onde aqui foi usado a lei de Friedel, que estabelece a igualdade $F_{\bar{h}} = F_{\bar{h}}$

A função rotação foi introduzida por Rossmann e Blow (1962) e é representada por:

$$R = \int_U P_2(X_2) P_1(X_1) dX_1 \quad (2.6)$$

onde a função de Patterson P_1 é mantida fixa e a função P_2 é rotacionada dentro do volume U. A função de Patterson neste caso, para o raio de integração, está representando os vetores intramoleculares (vetores entre átomos da mesma molécula). Esta função R então vai apresentar

um máximo quando os dois conjuntos de vetores tiverem a mesma orientação.

Substituindo a equação (2.4) na equação (2.6) obtém-se:

$$R = \frac{1}{V^2} \sum_{\bar{p}} \sum_{\bar{h}} |\mathbf{F}_{\bar{p}}| |\mathbf{F}_{\bar{h}}| \int_{U_h} \cos [2 \pi (\bar{h} + \bar{h}') \cdot \mathbf{X}_h] d\mathbf{X}_h \quad (2.7)$$

A integral na equação (2.7) é uma função de interferência cujo valor pode ser escrito como $(U/V) |G_{\bar{h}, \bar{h}'}|$, para uma integração sobre um volume simétrico sobre a origem de Patterson. Desta forma a função rotação pode ser escrita como:

$$R = (U/V^3) \sum_{\bar{p}} \sum_{\bar{h}} |\mathbf{F}_{\bar{p}}|^2 |\mathbf{F}_{\bar{h}}|^2 G_{\bar{h}, \bar{h}'} \quad (2.8)$$

Para uma partícula esférica, a função G de interferência apresenta a forma:

$$G = \frac{3 [\sin (2 \pi x) - 2 \pi x \cos (2 \pi x)]}{(2 \pi x)^3} \quad (2.9)$$

a qual apresenta o valor máximo de 1 na origem e não excede a 0,086 fora do intervalo $-0,725 < x < 0,725$, sendo portando todos os termos fora deste intervalo praticamente nulos (Figura 2.3).

A função de rotação normalmente é calculada transformando o sistema cristalográfico em coordenadas cartesianas e então convertidos para um sistema de ângulos, que pode ser o sistema de Euler $(\theta_1, \theta_2, \theta_3)$ ou o sistema de Lattman $(\theta_1 = \theta_1 + \theta_3, \theta_2 = \theta_1 - \theta_3, \theta_3)$. Uma lista de todas as unidades assimétricas, em termos dos ângulos de Euler ou de Lattman, pode ser acompanhado em Narasinga *et al.* (1980).

Figura 2.3 - Representação da função G para o caso de uma distribuição esférica(Equação 2.9).

Uma vez determinada a orientação, pode-se então posicionar a molécula no espaço, relativa aos elementos de simetria cristalográficos, ou seja, tendo-se a orientação da molécula com um modelo conhecido, é necessário definir sua posição na cela, relativa à origem selecionada, que é basicamente determinada pela função translação, expressa por (Rossmann *et al.*, 1964):

$$T(\bar{\Delta}) = \int_{|u| < 2r} P(\bar{u} + \bar{\Delta}) P([C]\bar{u} - \bar{\Delta}) d\bar{u} \quad (2.10)$$

onde Δ é o vetor de translação, que é independente da origem. Rossman *et al.* (1964) expressaram a função acima como:

$$T(\bar{\Delta}) = \left(\frac{u}{V^3} \right) \sum_{\bar{h}} \sum_{\bar{p}} |F_{\bar{h}}|^2 |F_{\bar{p}}|^2 G_{\bar{h}\bar{p}} \cos \{ 2 \pi (\bar{h} - \bar{p}) \cdot \bar{\Delta} \} \quad (2.11)$$

onde a soma é conduzida de tal forma a remover os vetores intramoleculares ($|u| < 2r$). Como na função rotação, a função translação depende criticamente da escolha do raio do conjunto de vetores e o resultado pode ser analisado, para cada caso, pelo teste de diferentes raios. Para uma revisão mais completa veja Rossmann (1990).

Nesta tese, esta técnica foi utilizada para a determinação da estrutura da TIM de *Bacillus stearothermophilus* (Capítulo 6 desta tese) através do programa XPLOR (Brünger, 1990a).

Além da substituição molecular, mais recentemente foi implementado um método mais eficiente de selecionar as soluções listadas pelo programa, método este conhecido como refinamento de correlação de Patterson (refinamento PC). Este procedimento consiste no refinamento de um parâmetro p (Brünger, 1990b) da lista de soluções da substituição molecular. O refinamento é conduzido contra o coeficiente de correlação negativo pois os algoritmos de minimização normalmente localizam mínimos ao invés de máximos. Os parâmetros p podem ser grupos de átomos representados por um corpo rígido, coordenadas atômicas, fatores de temperatura ou de ocupação. Desta forma, o refinamento PC pode ser a minimização através do refinamento por corpo rígido de todas as soluções listadas. A minimização do coeficiente de correlação negativo -PC é equivalente a minimizar os erros das fases (Hauptman, 1982). Maior detalhes deste processo implementado no programa XPLOR pode ser acompanhado em Brünger (1990b).

O passo final consiste então na função translação utilizando a orientação que produziu o melhor coeficiente de correlação, na função de rotação. Um caso detalhado da aplicação desta técnica pode ser acompanhada em Brünger (1990b).

Como as outras técnicas não foram utilizadas neste trabalho detalhes, podem ser

acompanhados de forma mais generalizada em Blundell e Johnson (Blundell e Johnson, 1976), e mais recentemente em *Methods in Enzymology* (editado por Harold W. Wychoff e C. H. W. Hirs) volumes 114 e 115 (Diffraction Methods for Biological Macromolecules, Parte A e B, Academic Press, Inc., 1985) e Jan Drenth através de um livro bastante recente que engloba os princípios da cristalografia de proteínas (Principles of Protein X-ray Crystallography, Springer-Verlag New York, Inc., 1994). Outra fonte mais básica segundo uma visão bioquímica da cristalografia de proteínas, onde as técnicas não são discutidas em detalhes pode ser obtida do livro de Gale Rhodes (Crystallography Made Crystal Clear. A guide for users of Macromolecular Models, Academic Press, Inc., 1993).

2.2.6 Modificação da Densidade Eletrônica: Promediação e Nivelamento do Solvente

Depois de determinada a estrutura da macromolécula, a análise do mapa de densidade eletrônica pode ser feita para a melhora do modelo. Em alguns casos o mapa de densidade eletrônica não apresenta qualidade adequada para a interpretação dos resíduos, cadeias laterais e em alguns casos até mesmo o traço da cadeia principal. Nestes casos, quando existe mais de uma molécula por unidade assimétrica, pode-se aplicar a promediação da densidade eletrônica para melhorar a qualidade do mapa. Se o cristal apresenta alta quantidade de solvente o nivelamento do solvente pode permitir uma melhor definição do envelope da proteína, assim como melhorar a qualidade do mapa. Aqui serão descritas as bases destas duas técnicas as quais foram usadas no Capítulo 6 desta tese.

Para se aplicar a técnica de modificação da densidade eletrônica por promediação e nivelamento do solvente primeiramente deve-se identificar que no cristal em questão existe simetria não-cristalográfica, ou seja, que além da simetria do grupo espacial, existem, na unidade assimétrica, cópias idênticas da mesma moléculas, relacionadas por operações de simetria não-cristalográficas, por exemplo no caso fortuíto da cristalização de mais de uma molécula na unidade assimétrica, da presença de oligômero (dímero, trímero, etc) ou no caso de se obter duas formas cristalinas da mesma proteína. A simetria não-cristalográfica pode ser própria (constituída de apenas uma rotação) ou imprópria (constituída de uma rotação mais uma translação).

A técnica historicamente iniciou com a identificação de subunidades idênticas na unidade assimétrica por Rossmann e Blow (1962). A identificação de subunidades idênticas foi equacionada através de restrições devido a simetria não-cristalográfica no espaço recíproco

(gerando as equações da substituição molecular - MRE) por Main e Rossmann (1966). O solvente foi então considerado como uma restrição a mais por Main (Main, 1967). Crowther reformulou as equações da substituição molecular (MRE) em uma forma linear (Crowther, 1967, 1969) e finalmente Bricogne obteve a equivalência entre as equações do espaço recíproco e da densidade eletrônica no espaço real (Bricogne, 1974).

O método é sumariamente apresentado abaixo para o caso do nivelamento do solvente. O fator de estrutura de uma molécula na unidade assimétrica (F) pode ser escrito como:

$$F_p^- = \int_U \rho(\bar{x}_n) \exp(2\pi i \bar{p} \cdot \bar{x}_n) d\bar{x}_n + F_p^s \quad (2.12)$$

onde ρ é a densidade eletrônica no espaço real. O primeiro termo é integrado sobre a molécula e o segundo termo é resultado da integração sobre o solvente. Como a densidade eletrônica é a transformada de Fourier dos fatores de estrutura então:

$$\rho(\bar{x}_n) = \sum_{\bar{h}} F_{\bar{h}} \exp(2\pi i \bar{p} \cdot \bar{x}_n) \quad (2.13)$$

Substituindo a equação (2.13) na equação (2.12) obtém-se:

$$F_p^- = \frac{1}{V} \sum_{\bar{h}} F_{\bar{h}} \int_U \exp(2\pi i \bar{h} - \bar{p}) \cdot \bar{x} d\bar{x}_n + F_p^s \quad (2.14)$$

que facilmente pode ser reescrita através da função de interferência descrita na seção 2.2.5 (Figura 2.3) de Rossmann e Blow (1962):

$$F_p^- = \frac{U}{V} \sum_{\bar{h}} F_{\bar{h}} G_{\bar{h}-\bar{p}} + F_p^s \quad (2.15)$$

que posteriormente foi implementada no espaço real por Wang (1985).

Para o caso da promediação, a densidade é promediada em todas os pontos equivalentes por simetria não-cristalográfica como pode ser visto pela equação (2.16):

$$\rho_{avg}(\bar{x}) = \frac{1}{N} \sum_{n=1}^N \rho(\bar{x}_n) = \frac{1}{NV} \sum_{n=1}^N \sum_{\bar{h}} F_{\bar{h}} \exp(2\pi i \bar{p} \cdot \bar{x}_n) \quad (2.16)$$

onde o somatório é feita sobre todas as moléculas equivalentes existentes na unidade assimétrica.

Substituindo a equação (2.16) na equação (2.12) obtém-se:

$$F_p^- = \frac{1}{NV} \sum_{\bar{h}} F_{\bar{h}}^- \sum_{n=1}^N \int_U \exp (2 \pi i \bar{h} \cdot \bar{x}_n) \exp [-2 \pi i (-\bar{p} \cdot \bar{x})] d\bar{x} + F_p^-^s \quad (2.17)$$

Como:

$$\bar{x}_n = [R_n^-] \bar{x} + \bar{t}_n \quad (2.18)$$

Então a equação (2.17) fica:

$$F_p^- = \frac{1}{NV} \sum_{\bar{h}} F_{\bar{h}}^- \sum_{n=1}^N \exp (2 \pi i \bar{h} \cdot \bar{t}_n) \int_U \exp (2 \pi i \{ \bar{h} [R_n^-] - \bar{p} \} \cdot \bar{x}) d\bar{x} + F_p^-^s \quad (2.19)$$

E finalmente com a substituição da função de interferência de Rossmann e Blow (1962) (seção 2.2.5, Figura 2.3):

$$F_p^- = \frac{U}{NV} \sum_{\bar{h}} F_{\bar{h}}^- \sum_{n=1}^N \exp (2 \pi i \bar{h} \cdot \bar{t}_n) G_{\bar{h}[R_n^-] - \bar{p}} + F_p^-^s \quad (2.20)$$

Então o fator de estrutura é recalculado por um somatório sobre todos os fatores de estrutura anteriores ponderado pela função G de Rossmann e Blow (1962). A função de interferência G apresenta um pico pronunciado na origem e é atenuada radicalmente além do valor 0,725 (veja seção 2.2.5, Figura 2.3) fazendo com que apenas os fatores de estrutura dentro deste limite contribuam significativamente aos valores recalculados. Desta forma o nivelamento do solvente tem seu efeito aumentado em estruturas que apresentam uma grande quantidade de solvente, pois para uma mesma molécula em um cristal com mais solvente, observa-se uma unidade assimétrica maior e portanto, no espaço recíproco, uma rede mais contraída, resultando então na participação de um número maior de fatores de estrutura no cálculo do novo fator de estrutura (equação 2.15). A promediação apresenta resultados mais apreciáveis para casos de muitas moléculas equivalentes na unidade assimétrica. Isto se observa pelo somatório na equação (2.20). O fator de estrutura obtido, far-se-á da contribuição dos que se encontram envolvidos pela função G (equação 2.20) e também pela promediação dos fatores de estrutura que se encontram relacionados por simetria não-cristalográfica.

Na Figura 2.4a pode ser visto a representação do espaço recíproco através de uma rede bidimensional. O círculo maior representa o limite de resolução, e o círculo menor representa a função G de Rossmann e Blow (1962). No exemplo da Figura 2.4a, a cela unitária aumentou de duas vezes, então a cela no espaço recíproco diminuiu de duas vezes. Logo a função G de

Rossmann e Blow (1962) envolve mais pontos da rede que irão contribuir para o fator de estrutura recalculado. Na Figura 2.4b representa-se a promediação, com uma simetria não-cristalográfica de ordem 3, a qual é representada por $G_{h[R_n]-p}$, onde R_n representa a matriz da operação de simetria não-cristalográfica.

Figura 2.4 - Representação do processo de (a) nivelamento do solvente e (b) promediação eletrônica. Mais detalhes no texto

2.2.7 Refinamento e Análise da Qualidade do Modelo

Existem, em princípio, muitas maneiras nas quais modelos de proteínas podem ser

refinados. O problema inicial aqui é tratar o modelo obtido por substituição molecular como modelo inicial. Desta forma deve-se fazer os ajustes do modelo, observadas as características da densidade eletrônica obtida.

No caso de substituição molecular as fases são originadas da posição dos átomos existentes no modelo usado inicialmente como padrão. Informações previamente introduzidas, como restrições da natureza dos aminoácidos, permitem reduzir os parâmetros a serem refinados. Também a densidade eletrônica pode ser aprimorada com o uso da modificação da densidade eletrônica por promediação assim como por nivelamento do solvente (seção 2.2.6).

O programa XPLOR (Brünger, 1990a) apresenta várias opções para o refinamento, incluindo refinamento de corpo rígido, minimização de energia, dinâmica molecular de alta temperatura e resfriamento lento, refinamento de fatores de temperatura, refinamento com omissão de átomos, entre outros. O processo a ser seguido depende do usuário e das condições iniciais do mapa de densidade eletrônica e do modelo inicial.

O refinamento de corpo rígido é normalmente usado no início para o caso da determinação da estrutura ter sido obtida por substituição molecular. Neste caso refina-se somente três coordenadas e três ângulos, ou seja, o modelo é refinado como um todo, permitindo apenas rotação e translação. Caso o modelo apresente mais de uma subunidade, pode-se refinar as subunidades independentemente, permitindo um refinamento na orientação de cada subunidade, que na verdade é um refinamento da matriz de simetria não- cristalográfica. Estes passos são importantes no caso do uso da modificação da densidade eletrônica por promediação (seção 2.2.6).

Minimização de energia pressupõe que a energia estereoquímica do modelo seja dada por:

$$E_{chem} = w_b \sum E_{dist. lig.} + w_c \sum E_{ang. lig.} + w_d \sum E_{ang. tor.} + w_e \sum E_{const. nl} + w_f \sum E_{gr. plan.} + w_g \sum E_{vol. quiral} \quad (2.21)$$

onde $E_{dist. lig.}$ é a energia da distância de ligação, $E_{ang. lig.}$ é a energia do ângulo de ligação, $E_{ang. tor.}$ é a energia do ângulo torsional, $E_{const. nl}$ é a energia de contatos entre átomos não-ligados, $E_{gr. plan.}$ é a energia de grupos planares e $E_{vol. quiral}$ é a energia de quiralidade do carbono α . O deslocamento da posição de equilíbrio tende a aumentar todos os termos da energia estereoquímica. As constantes w_b a w_g são usadas para manter a proporcionalidade ou mesmo o peso de cada termo durante o processo de minimização. A energia é minimizada pelo deslocamento dos átomos no sistema de coordenada cartesiana em direção à posição de menor energia. XPLOR (Brünger, 1990a) possui arquivos contendo as constantes e as distâncias de ligações e ângulos para cada

aminoácido, assim como as constantes de torsão.

Em geral a conformação do modelo usado para a determinação da estrutura, para o caso da substituição molecular, não coincide com o modelo da proteína de interesse. Diferenças normalmente ocorrem na superfície e em estruturas em volta (loops) que são regiões mais flexíveis da proteína. Estas diferenças conformacionais apresentam barreiras energéticas muito elevadas para serem vencidas através da minimização de energia. Desta forma a dinâmica molecular de alta temperatura e resfriamento lento pode ser utilizada para resolver estas diferenças conformacionais no refinamento.

Neste caso a função de energia a ser minimizada é uma função híbrida:

$$E_{pot} = E_{chem} + w_a E_{X-ray} \quad (2.22)$$

onde E_{chem} corresponde a função de energia apresentada na equação (2.21), e E_{X-ray} consiste do resíduo cristalográfico. Em geral, E_{X-ray} pode consistir de qualquer função que tende a zero quando o modelo reproduz os dados, como por exemplo, E_{X-ray} pode ser $(1-c)$ onde c é o coeficiente de correlação linear entre os dados observados e calculados (entre $|F_{obs}(\bar{h})|$ e $|F_{calc}(\bar{h})|$ (Brünger, 1990a). O peso w_a relaciona E_{chem} à E_{X-ray} . Como mencionado a simples minimização da equação (2.22) pode não atingir o mínimo global e ficar aprisionada num mínimo local. Para ultrapassar os mínimos locais precisa-se permitir que o sistema ganhe energia para superar as barreiras. Este processo foi aplicado pela primeira vez em cristalografia por Brünger *et al.* (Brünger *et al.*, 1987), que implementaram o processo conhecido como "simulated annealing" (Kirkpatrick *et al.*, 1983) com um algoritmo de dinâmica molecular (Verlet, 1967). Este método vai ser referido neste trabalho como refinamento por dinâmica molecular. Vários autores usaram este método com sucesso (uma lista de exemplos pode ser obtida em Brünger *et al.*, 1990).

O processo deste tipo de refinamento consiste em elevar a temperatura do sistema, que vai aumentar a energia cinética dos átomos como consequência, a qual pode ser suficiente para ultrapassar barreiras conformacionais locais. Após esta elevação a temperatura é decrescida em passos até chegar à temperatura mínima, depois da qual o modelo passa por um refinamento com minimização de energia mantendo a temperatura constante.

Brünger *et al.* (1990) testaram vários protocolos deste tipo de refinamento mostrando que o decréscimo lento da temperatura (em passos de 25 K fs⁻¹), utilizando um acoplamento fraco da temperatura (Berendsen *et al.*, 1984) como processo de controle da temperatura apresentou o melhor resultado. Este método é conhecido como "slow-cooling protocol" para o refinamento em dinâmica molecular, o qual vai ser referido nesta tese como dinâmica molecular de resfriamento lento. Este foi um dos passos do refinamento da TEM de *Bacillus*

stearothermophilus (Capítulo 6 desta tese).

Para aprimorar o modelo adiciona-se aos fatores de estrutura um termo dependente da temperatura pois as expressões para o espalhamento consideram o átomo em repouso e o movimento térmico interfere no padrão de difração. Debye mostrou que o movimento térmico decresce a intensidade dos raios-X espalhados por um fator $\exp[-B(\sin^2\theta/\lambda^2)]$ onde $B=8\pi^2\overline{u}/3=8\pi^2\overline{u_x}$ e onde $\overline{u}$ é o deslocamento médio dos átomos ao longo dos planos de reflexão e $\overline{u_x}$ é sua componente x. Em quase todas as estruturas cristalinas, B é determinado empiricamente e para proteínas varia entre 12 e 20 Å² para átomos contidos na cadeia principal, sendo maiores para átomos nas cadeias laterais, principalmente para aqueles situados na superfície da proteína. Estes valores correspondem a um deslocamento entre 0,15 Å e 0,5 Å respectivamente. O refinamento de B leva em consideração também fatores como desordem estática, erro nos fatores de escala de medida, absorção e incorreções nas curvas de espalhamento atômico. Então os valores do deslocamento médio podem ser considerados como aproximados, apesar de nas proteínas a contribuição maior ser devido ao movimento dos átomos. Normalmente em proteínas, devido ao grande número de parâmetros a ser refinado, refina-se um fator de temperatura isotrópico.

Em alguns casos a conformação do modelo usado em substituição molecular difere consideravelmente do modelo da proteína que se deseja refinar. Esta conformação no espaço real introduz uma modificação que se encontra espalhada em todo o espaço recíproco (pois o fator de estrutura é uma transformada de Fourier da densidade eletrônica), e muitas vezes não se consegue obter a densidade eletrônica na posição correta. A remoção destes átomos e um passo de minimização de energia (e em alguns casos uma dinâmica molecular de resfriamento lento) remove a influência do posicionamento incorreto, fornecendo densidade eletrônica em posições diferentes daquelas observadas anteriormente. Este foi o caso para a estrutura em volta flexível no refinamento da estrutura da TEM de *Bacillus stearothermophilus*, a qual se encontrava na conformação conhecida como aberta e depois do refinamento por omissão foi observada densidade eletrônica na conformação conhecida como fechada, com diferença de até 8 Å na posição dos átomos localizados no centro da estrutura em volta (Capítulo 6 desta tese).

Após a obtenção do modelo final, análises devem ser feitas para verificar a qualidade geométrica dos comprimentos das ligações, ângulos entre átomos ligados, ângulos de torção e ângulos impróprios. Esta análise pode ser feita por vários programas. XPLOR (Brünger, 1990a) permite fazer tais análises, mas análises mais completas podem ser feitas por programas como PROCHECK (Programs to Check the Stereochemical Quality of Protein Structures, Laskowski,

MacArthur, Moss and Thornton, Oxford Molecular Ltd., Morris *et al.*, 1992) entre outros.

Estas análises observam quanto estes valores desviam de valores médios. Também o diagrama de Ramachandran mostra onde teoricamente os ângulos diédricos Ψ e ϕ dos peptídeos devem ser observados para cada tipo de estrutura secundária nas proteínas e o programa PROCHECK (Morris *et al.*, 1992) apresenta regiões mais flexíveis onde combinações destes ângulos foram observados em um número razoável de proteínas. Alguns valores proibidos podem ser observados caso exista uma justificação para tais.

O pacote de programa XPLOR (Brünger, 1990a) permite também tratamento de dados obtidos por NMR (Ressonância Magnética Nuclear) de proteínas em solução para determinação de estruturas. Este pacote tem sido amplamente utilizado no refinamento de estruturas e tem fornecido bons modelos com boa qualidade geométrica. Este programa foi amplamente utilizado neste trabalho nos Capítulos 5 e 6 desta tese.

2.2.8 Cálculo de Mapas de Densidade Eletrônica e Análise Gráfica

Com fases iniciais obtidas por substituição molecular ou por substituição isomorfa múltipla pode-se calcular um mapa de densidade eletrônica através de diversos pacotes de programas. O pacote mais amplamente utilizado por cristalógrafos de proteínas é o CCP4 (CCP4, 1994. The SERC (UK) Collaborative Computational Project Number 4, distribuído pelo Laboratório de Daresbury, Warrington, Inglaterra) que permite ampla opção para manipulação de dados, desde processamento até refinamento e análises finais de modelos. Este pacote vem sendo continuamente revisado apresentando versões recentes que envolvem ampla abordagem de técnicas cristalográficas. Um pacote que apresenta características similares foi desenvolvido pelo grupo de Cristalografia da Universidade Estadual de Groningen, chamado de BIOMOL (pacote de programs do Laboratório de Físico-Química da Universidade Estadual de Groningen, Groningen, Holanda). Este pacote foi o utilizado neste trabalho devido à predominante utilização no grupo do Dr. Wim Hol, no qual grande parte deste trabalho foi desenvolvido. Este programa foi utilizado predominantemente para os cálculos de densidade eletrônica, promediação de densidade eletrônica e nivelamento do solvente. Os cálculos seguiam a ordem abaixo:

- (i) Obtenção das amplitudes e fases através do modelo, que foram obtidas pelo programa XPLOR (Brünger, 1990a) ou através da inversão do mapa obtido do modelo de substituição molecular, através dos programas EDXZY e

- FSINV do pacote BIOMOL;
- (ii) Escalonamento das amplitudes através do programa KBRANI do pacote BIOMOL;
 - (iii) Obtenção das figuras de méritos das fases e coeficientes dos mapas que podiam ser $m(F_o - F_c)$ ou $m(2F_o - F_c)$, através do programa SIGMAA (Read, 1986) do pacote BIOMOL. Ambos os mapas foram utilizados sendo que o primeiro conhecido como mapa diferença indica regiões onde deve existir algum grupo no caso de densidade positiva ou desaparecer algum grupo no caso de densidade negativa. Este tipo de mapa foi predominantemente usado durante a localização e refinamento das estruturas em volta que não possuíam densidade eletrônica definida durante o processo de refinamento.
 - (iv) Finalmente, cálculo do mapa de densidade eletrônica pelos programas FSFOR e SHFMFF do pacote BIOMOL, os quais calculam os mapas em uma cela unitária e ao redor de uma molécula respectivamente.

Os mapas assim obtidos são analisados através de programas gráficos como O (Jones *et al.*, 1991), o qual possibilita a análise do modelo, sua manipulação (como inserção de resíduos ou átomos, deleção, manipulação em geral do modelo), e visualização do mapa de densidade eletrônica. O programa O também oferece possibilidades de refinamentos locais depois de cada manipulação do modelo que permitem ajustar as modificações aos parâmetros geométricos dos resíduos. Estes dois pacotes são utilizados constantemente durante o refinamento da estrutura pois permitem a visualização do progresso do refinamento do modelo. À medida que o refinamento vai se desenvolvendo, programas são utilizados para avaliar qualitativamente o modelo, como desvios das distâncias de ligação padrão de cada par de átomo, desvios dos ângulos de torsão, desvio dos ângulos diédricos, contatos de van der Waals entre outras características que serão apresentadas durante a discussão de cada modelo estudado neste trabalho.

2.2.9 Técnicas Usadas em Desenho Racional de Drogas.

Este é um campo que vem se desenvolvendo grandemente em todos os países, o qual consiste na procura racional de novos compostos que tenham as características de serem utilizados como drogas e que portanto venham a curar doenças que afetam milhões de pessoas.

Cristalografia de proteínas, por estudar moléculas extremamente importantes para os seres vivos, tem permitido além do entendimento dos processos metabólicos dos seres vivos também a possibilidade de solucionar problemas causados por deficiências genéticas assim como combate a organismos invasores causadores de doenças nos seres humanos e em animais economicamente importantes para o homem.

A concentração neste trabalho é no desenvolvimento de compostos que venham a inibir seletivamente proteínas glicolíticas do *T. brucei*, agente causador da doença do sono. A descrição neste capítulo de todo o processo de desenho racional de drogas seria redundante pois uma descrição detalhada de todo o processo e programas utilizados foi feita no Capítulo 5 desta tese.

Capítulo 3

Experimentos de Cristalização da Frutose-1,6-Bisfosfato Aldolase de *T. brucei* e Análise Estrutural Preliminar para o Desenvolvimento Racional de Drogas Contra a Doença do Sono

3.1 Introdução

Frutose-1,6-bisfosfato aldolase do *Trypanosoma brucei* é uma das enzimas glicolíticas incluída num projeto amplo de desenvolvimento racional de drogas contra a doença do sono. Este projeto é parte de uma colaboração com Dr. Fred Oppendoes e Dr. Paul Michels do International Institute of Cellular and Molecular Pathology (ICP), em Bruxelas, na Bélgica, os quais são os pesquisadores responsáveis pela obtenção da proteína em grandes quantidades, através da clonagem e expressão em *E. coli*, além da caracterização e medidas da atividade enzimática e constante de inibição para novos compostos tidos como potenciais inibidores.

A aldolase catalisa a quebra reversível da frutose-1,6-bisfosfato em gliceraldeído-3-fosfato e dihidroxiacetona fosfato (Figura 3.1) (Horecker *et al.*, 1972). Em função de seu mecanismo de catálise, duas classes de aldolase foram identificadas: classe I e II. Na classe I o mecanismo de reação envolve a formação de uma base de Schiff entre o substrato e uma lisina do sítio ativo, a qual estabiliza a carga negativa do grupo carbônio da dihidroxiacetona fosfato. Enzimas da classe II não formam um intermediário ligado covalentemente à enzima, mas a estabilização do carbônio é conseguida por um íon bivalente, Zn^{2+} . Aldolase da classe II é encontrada apenas nos organismos procarióticos e eucarióticos mais primitivos como levedura, fungos e algas, enquanto que aldolase da classe I tem sido encontrada em representantes de todos os reinos filogenéticos:

Figura 3.1 - Reação catalisada pela frutose-1,6-bisfosfato aldolase da classe I

bactérias, protistas e metazoos.

Aldolases da classe I são enzimas homotetraméricas com peso molecular de 40 KDa por subunidade. Estruturas tridimensionais foram obtidas das aldolases de músculo humano(classe I) (Gamblin *et al.*, 1990, 1991), músculo de coelho (Sygush *et al.*, 1987), fígado de coelho (White *et al.*, 1991), *Drosophila melanogaster* (Hester *et al.*, 1991) e

Tabela 3.1 - Alinhamento das sequências foi feito através do programa CLUSTAL V (Higgins *et al.*, 1994 e 1992). Códigos são : T para *Trypanosoma brucei*, A, B e C para isoenzimas humana A, B e C respectivamente; † indica o resíduo envolvido na base de Schiff. Estruturas secundárias: a - hélice α ; b - fita β . Negrito e * indicam resíduos conservados e . corresponde a mutações funcionalmente conservadas. ‡ indica His 361 na sequência da isoenzima humana A discutida no texto.

		63
	aaaaaaaaaaaaa bbbbbbbb aaaaaaaa a	
T	SKRVEVLLTQLPAYNRLKTPYE ELI ETAKKMTAP GKGLLA DE STG SCSKRFAGIGLSNTAE	
A	PYQ-----YPALTPEQKK ELSDIA HRIVAP GKGI LA DE ST GSIA KRLQSIGTENTEE	
B	AHR-----FPALTQEQKK ELSEIA QSIVANG GKGI LA DE SVGTMGNRLQRIKVENTEE	
C	PHS-----YPALSAEQKK ELSDIAL RIVAP GKGI LA DE SVGSMARLSQIGVENTEE	
.	. * ** . * . * ***.***** * . . . * **	
	126	
	aaaaaaaaaaaaa bbbbbbb aaaaaa aaaaaa bbbbbbbb	
T	HRRQYRALMLECEG-FEQYIS GVILHDE TVYQAKTGETFPQYLRRRGVVP GIKTDC GLEPLV	
A	NRRFYRQLLLLTADDRVNPCIG GVILF HETLYQKADGRFPQVIKSKGGVVG IKVDKGV VPLA	
B	NRRQFREILFSVDSSINQSIG GVILF HETLYQKDSQKLFERNILKEKGIV VGIKLDQ GGAPLA	
C	NRRLYRQVLFSADDRVKKCIG GVIFF HETLYQKDDNGVPFVRTIQDKGIV VGIKVDKGV VPLA	
.	.*. * * * * * *	
	189	
	aaaaaaaaaaaaa bbbbbbbbbb aaaaaaaaaaaaaaaaaaaaa	
T	EGAKGEQMTAGLDGYIKRAKKYYAMGCR FKWR NVYKIQNGTVSEAVVRFNAETLARYA IL LSQ	
A	-GTNGETTTQGLDGLSERCAQYKKDGAD FAKWR CVLKIGEHTPSALAIMENANVLARYA SI CQ	
B	-GTNKETTTIQGLDGLSERCAQYKKDGVDF GKWR AVLRIADQCPSSLAIQENANALARYA SI CQ	
C	-GTDGETTTQGLDGLSERCAQYKKDGAD FAKWR CVLKISERTPSALAILENANVLARYA SI CQ	
.	* . * **** * . * * * * * * . * . * . * . * . * . *	
	252	
	aa bbbbbbb aaaaaaaaaaaaaaaaaaaaaa bbbbbbb	
T	LCGLVPIVEPEVMIDGTHDIETCQRVSQHVWSEVVSALHRHGVVWEGCLLKPNMVVPGAESGL	
A	QNGIVPIVEPEIILPDGDHDLKRCQYVTEKVLAAVYKALSDHHIYLEGTL LLKPNMV VTGHACTQ	
B	QNGLVPIVEPEVIPDGDHDLHCQYVTEKVLAAVYKALNDHHVYLEGTL LLKPNMV TGHACTK	
C	QNGIVPIVEPEIILPDGDHDLKRCQYVTEKVLAAVYKALSDHHVYLEGTL LLKPNMV TPGHACPI	
.	* * . * . * . * . * . * . * . * . * . *	
	315	
	aaaaaaaaaaaaa bbbbbbbbbb aaaaaaaaaaaaaa bbbbbbb a	
T	KGHAEQVAEYTVKTLARVIPALPGVT FLSGGL SEVMASEYLNAMNNCPLPRPWKL TF SYARA	
A	KFSHEEIAMATVTALRRTVPAVVTG ITFLSGG QSEEEASINLNAINKCPL LPWALT TF SY GRA	
B	KYTPEQVAMATVTALHRTVPAAVPGIC FLSGGM SEEDATLNLNAINLCPL PKPWALS TF SY GRA	
C	KYTPEEIAMATVTALRRTVPAVPGVT FLSGGS SEEEASFNLNAINCCPL PRPWALT TF SY GRA	
.	* . . * ** . * * . * . * . * . * . * . * . * . *	
	374	
	aaaaaaaaaaaaa aaaaaaaaaaaaaaaaaaaaaa	
T	LQSSAIKRWGGKESGVEAGRRAFMHRAK NSLA QLGKYNRADDDKD--SQSLYVAGNTY	
A	LQASALKAWGGKKENLKAAQEEYVKRALANSLACQ GKY TPSGQAGAAASESLFVSNHAY	
B	LQASALAAWGGKAANKEATQEA FMKRAM ANCQAAGQYVHTGSSGAASTQSLFTACYTY	
C	LQASALNAWRGQRDNAGAATEEFIKRAEVNGLAAQ GKYE GSGEDGGAAQSLYIANHAY	
.	.*. * . * . * . * . * . * . * . * . *	

fosfogluconato aldolase de bactéria (Lebioda *et al.*, 1982). A arquitetura molecular das aldolases da classe I é um barril de 8 fitas α/β , similar ao da triosefosfato isomerase, primeiramente observado por Banner *et al.* (Banner *et al.*, 1975) e subsequentemente em várias outras enzimas (Farber e Petsko, 1990; maiores detalhes no Capítulo e Figura 4.2). Diferentemente das outras enzimas com estrutura em barril α/β que apresentam o sítio ativo próximo à superfície, a aldolase

apresenta o sítio ativo localizado no centro do barril, contendo não apenas resíduos hidrofóbicos mas também vários resíduos carregados.

Mamíferos apresentam três diferentes isoenzimas da classe I, distribuídas especificamente em diferentes tecidos. Elas foram classificadas como A (tecido embrionário e muscular), B (fígado e rim) e C (no cérebro). O alinhamento das sequências das três isoenzimas humana e da aldolase glicossomal do *T. brucei* pode ser vista na Tabela 3.1.

A aldolase de *T. brucei* apresenta uma homologia de 47,2, 45,5, 46,3% com isoenzimas humana A, B e C respectivamente. A homologia entre as isoenzimas humana são: A e B, 69,4%; B e C, 70,2%; A e C, 81,8%.

Vários resíduos são conhecidos como formadores dos sítios de ligação dos grupos fosfato (o fosfato do gliceraldeído-3-fosfato e da dihidroxiacetona fosfato, produtos da reação de catálise), como Lys 146, Arg 148, Lys 41, Arg 42 e Arg 303 (numeração da estrutura da isoenzima humana A). Resíduo Lys 229 forma a base de Schiff com o substrato.

Como nas outras estruturas em barril α/β , o sítio ativo é acessível pelo lado C-terminal do barril. O acesso no entanto é modulado pela região C-terminal, na forma de um braço com um cotovelo flexível, o qual cobre o sítio ativo. Enquanto o sítio ativo é conservado evolutivamente, o braço C-terminal aparece com uma alta variação. Comparação de sequências e estudos de mutagênese sugerem que a região C-terminal contribui para a especificidade da enzima em relação ao substrato (Sygush *et al.*, 1987, Takahashi *et al.*, 1989). Mais recentemente Littlechild e Watson (1993) propuseram uma explicação estrutural para a participação da Tyr C-terminal (Tyr 363 na sequência humana) a qual é ancorada no sítio ativo pela interação com uma His (His 361 na sequência humana) do braço C-terminal e com o fosfato do carbono 1 do substrato (Gamblin *et al.*, 1991). Esta His, nas isoenzimas A e C, está trocada por uma Tyr na isoenzima B e por uma Asn na aldolase de *T. brucei* (Tabela 3.1). Resultados de análises enzimáticas e especificidade de substratos mostram que tanto a frutose-1,6-bisfosfato (FBF) quanto a frutose-1-fosfato (F1F) são substratos mas com diferentes especificidades. Enquanto as isoenzimas humana A e C apresentam uma especificidade 50 e 10 vezes maior para a FBF relativa a F1F, respectivamente, a isoenzima B não apresenta especificidade (razão igual a 1). Desta forma não somente a Tyr C-terminal é importante na catálise como também seu posicionamento estrutural no sítio ativo que é mantido através da interação de resíduos do braço flexível.

Desenho de compostos que venham a ser usados como inibidores específicos de aldolase de *T. brucei*, deve levar em consideração a existência de três isoenzimas no organismo humano e a seletividade deve ser conseguida através de um composto que explore esta diferença. Possíveis sítios de divergência entre as estruturas são os resíduos do braço flexível, já que não existe

conservação do braço como um todo, mas somente alguns resíduos enzimaticamente ativos. A dificuldade atual é que como consequência de sua flexibilidade, em todas as estruturas resolvidas o braço flexível apresenta uma densidade eletrônica muito pobre, sendo difícil a utilização dos modelos atuais das aldolases para desenhos de inibidores que venham a utilizar o sítio de ligação do braço flexível.

Uma primeira estrutura resolvida na presença de um dos vários inibidores análogos ao substrato conhecidos possibilitaria ao mesmo tempo fornecer informações de como o substrato se liga ao sítio ativo, possivelmente fornecendo então dados para o entendimento do mecanismo enzimático, assim como também de como explorar as diferenças entre as estruturas para uma modelagem de um inibidor que apresente seletividade. Desta forma mais do que apenas a estrutura da aldolase de *T. brucei*, a estrutura de complexos de todas as isoenzimas humanas seria de muita importância para o sucesso do desenho racional de drogas.

3.2 Comparação entre as Sequências das Aldolases Humanas e de *T. brucei* Baseada na Estrutura Tridimensional da Aldolase de Músculo Humano.

Como no momento somente a estrutura tridimensional de uma das isoformas humanas encontra-se disponível, um primeiro trabalho pôde ser feito baseado na comparação das sequências e a posterior análise da estrutura tridimensional disponível para a possibilidade de se encontrar um sítio de seletividade que seja comum quando comparadas à sequência da aldolase de *T. brucei* (Delboni e cols., 1991).

Neste trabalho definiu-se uma esfera ao redor dos resíduos Lys 107, Lys 146 e Lys 229, todas envolvidas no sítio ativo, através de interações com o fosfato do substrato ou através da formação da base de Schiff. Foram incluídos todos os resíduos ao redor do átomo N ζ de cada uma destas lisinas dentro de um raio de 18 Å. Todos os resíduos encontrados dentro desta esfera foram analisados relativamente ao alinhamento das sequências mostrado na Tabela 3.1.

Baseado na estrutura da aldolase humana (Gamblin *et al.*, 1990 e 1991) foram encontrados vários possíveis sítios de seletividade, sendo que alguns poderão ser futuramente explorados com a disponibilidade da estrutura da aldolase de *T. brucei*.

Na Tabela 3.2 pode ser visto o resultado de tal comparação. Como pode ser visto, existem várias regiões que se mostram promissoras para serem utilizadas como sítios que poderão apresentar seletividade relativa à todas as isoenzimas humanas. Infelizmente nem todas as

diferenças são acessíveis ao solvente o que torna o uso destas diferenças bastante restrito. Por outro lado, estas diferenças, apesar de não acessíveis, podem apresentar efeitos indiretos como uma conformação estrutural diferente, diferença esta que pode ser explorada no desenho de inibidores seletivos. Mas este resultado somente poderá ser analisado quando as estruturas tridimensionais das aldolases de *T. brucei* e das isoenzimas humanas B e C forem elucidadas. Fica bastante claro que a disponibilidade da estrutura da aldolase de *T. brucei* é crucial para a continuidade de um projeto de desenho racional de inibidores específicos da enzima do tripanossomo, bem como a obtenção de mais informações de como o substrato se liga ao sítio ativo.

Tabela 3.2 - Distâncias em Å dos resíduos ao redor do sítio de ligação do fosfato que apresentam diferenças possíveis de serem exploradas para a seletividade de inibidores. DHAP corresponde aos resíduos próximos ao sítio de ligação do fosfato da dihidroxiacetona fosfato e GAP aos resíduos próximos ao sítio de ligação do gliceraldeído-3-fosfato.

a. DHAP					
d (Å)	<i>T. brucei</i>	Human A	Human B	Human C	
9	His 88	Phe 79	Phe	Phe	
~15	Asp 89	His 80	His	His	
~11	Thr 118*	Val 108*	Leu	Val	
~18	Cys 120	Lys 110	Gln	Lys	
~16	Glu 123	Val 113	Ala	Val	
~11	Gln 133	Thr 122	Thr	Thr	
~15	Ala 136	Gln 125	Gln	Gln	
~10	Leu 285	Gln 274	Met	Gln	

b. GAP					
d (Å)	<i>T. brucei</i>	Human A	Human B	Human C	
~6	Cys 49*	Ile 39*	Thr	Ser	
~10	Phe 53*	Leu 43*	Leu	Leu	
~11	Ala 54	Gln 44	Gln	Ser	
~10	Gly 55	Ser 45	Arg	Gln	
~7	Ser 318	Ala 307	Ala	Ala	

* Estes resíduos não são, na isoenzima humana, acessíveis ao solvente, não sendo provavelmente diretamente úteis para desenho de inibidores. Indiretamente, portanto, eles podem afetar a estrutura da enzima do parasita próximo aos sítios de ligação dos fosfatos. O quanto estes possíveis efeitos indiretos podem ser explorados para o desenho de inibidores seletivos depende da determinação da estrutura da enzima do parasita.

3.3 Parte experimental

O projeto de determinação da estrutura tridimensional da aldolase de *T. brucei* começou

já há muitos anos, primeiramente com 17 mg de proteína cedidas pelo grupo do Dr. Fred Oppendoes ao grupo do Dr. Wim Hol, em Groningen, Holanda. Dr. Rik Wierenga, então em seu programa de doutorado, obteve apenas alguns pequenos cristais os quais não difrataram nem mesmo sob radiação síncrotron.

Devido ao sucesso obtido com a determinação da estrutura da TIM de *T. brucei*, este projeto foi mantido latente. Este trabalho foi re-iniciado com a expressão da enzima no laboratório do Dr. Fred Oppendoes com a participação do Dr. Paul Michels, que forneceram 70 mg de proteína para os ensaios de cristalização.

Tabela 3.3 - Condições de cristalização para a aldolase de *T. brucei* utilizando proteína expressa em *E. coli* em grandes quantidades¹.

1. Amostra I (preparada em Bruxelas pelo Dr. F. Oppendoes do ICP) (experimentos de cristalização pelos métodos da gota pendurada e da gota sentada, Capítulo 2, seção 2.2.3)
 - Experimentos de cristalização anteriores: PEG 8000 + Fosfato
 - Resultados: Agulhas
 - Vários aditivos foram usados mas eles não afetaram a qualidade dos cristais
 - Método de cristalização Matricial "Phosphate 96":
 1. Solução de proteína: 2 mg/ml em 10 mM Tris/HCl pH 7.5, 100 mM NaCl
 2. Solução do poço: 0.1 M tampão fosfato/citrato de sódio pH 5.0-7.0; 2 M A. S. ou 2.0-2.5 M NaCl.
 3. Gota: Razão de proteína para solução do poço. 2:1-6:1 - Resultados: Muitos cristais pequenos de forma cúbica e hexagonal (60 µm em cada direção)
 - Tentativas:
 - a. "Seeding" (micro e macro, Capítulo 2, seção 2.2.4) não melhorou os resultados
 - b. Cristalização em resina não produziu nenhum cristal
 - c. 4 °C não melhorou os resultados
 - d. Co-cristalização com inibidores não melhorou os resultados
2. Amostra II (expressão em Seattle)
 - Condições: Mesmas que as anteriores.
 - Resultados: Mesmos que os anteriores.

Com a disponibilidade de proteína, milhares de condições de cristalização foram preparadas pelo autor desta tese durante um ano, com obtenção de cristais que apresentaram até 1 mm em uma direção, mas dimensões muito reduzidas nas duas outras direções (ao redor de 0,01 mm). Condições de cristalização nas quais estes cristais apareceram podem ser vistas na Tabela 3.3.

¹ Para abreviações e métodos apresentados nesta seção e em toda a tese vide Lista de Abreviações e/ou Capítulo 2

Vários aditivos e variações das condições de cristalização foram tentados sem resultados satisfatórios que mostrassem um progresso na melhora das outras duas dimensões.

Assim como em TIM de *T. brucei* (vide Capítulo 4, neste trabalho) o problema da idade da proteína foi então levantado como possível explicação para a qualidade dos cristais, assim como o seu grau de pureza. Desta forma a cristalização deveria seguir imediatamente a expressão da proteína em grande quantidade e purificação, processos estes que deveriam sofrer modificações para melhorar a qualidade da solução de proteína obtida, para fornecer resultados mais satisfatórios.

Desta forma, assim como em TIM, a expressão de aldolase de *T. brucei* em grande quantidade foi implementada no grupo do Dr. Wim Hol na Universidade de Washington, Seattle, USA. Este trabalho inicialmente seguiu o protocolo do Dr. Paul Michels, no qual foram introduzidas posteriormente modificações que otimizaram o rendimento final e principalmente a qualidade da pureza da enzima.

3.4 Expressão em Grandes Quantidades da aldolase de *T. brucei* em *E. coli*

O gene da aldolase foi amplificado pelo método PCR (Reação em Cadeia da Polimerase) e foi clonado próximo ao sítio de ligação do ribossomo do vetor de expressão pET3A que permite indução com IPTG. Esta construção foi então introduzida em *E. coli* BL21(DE3)pLysS, uma linhagem que contém o gene da polimerase T7 RNA no seu cromossomo (Studier *et al.*, 1990), sob o controle do promotor lacUV5. As bactérias assim construídas foram gentilmente cedidas pelo Dr. Paul Michels, do ICP, Bruxelas (Chevalier *et al.*, submetido). Uma pré-cultura de bactérias então foi crescida em meio sólido (placas de ágar em meio LB - 1,6% de ágar em meio LB que consiste de 1% triptona, 0,5% extrato de levedura e 0,5% de cloreto de sódio, sendo autoclavadas antes do uso e adicionado 0,01% de ampicilina no momento do uso), a 37°C durante toda a noite, para isolar uma única linhagem. Após este primeiro procedimento as bactérias foram então inoculadas em 5 ml do meio LB mais 100 µg/ml de ampicilina e 25 µg/ml de cloranfenicol , e mantidas a 37°C sob agitação de 200 rpm durante a noite. O volume de 1 ml da pré-cultura foi então inoculado em 1 litro de meio LB com 100 µg/ml de ampicilina e 25 µg/ml de cloranfenicol, distribuídos em 4 frascos tipo erlenmeyer para manter uma alta razão de ar para propiciar alto nível de oxigenação. Esta cultura foi incubada durante 6 horas a 37°C sob agitação de 200 rpm.

Quando a absorção óptica atingiu o nível de 0,7 a 0,8, a indução foi feita com 0,4 mM de IPTG (concentração final de IPTG na solução), e a cultura foi incubada por mais 2 horas a 37 °C sob agitação de 200 rpm (expressão mantida durante toda a noite mostrou pessimos resultados). Bactérias depois de induzidas sob as condições mencionadas foram então centrifugadas a 5000 rpm, durante 10 minutos no rotor GSA da Sorvall. O precipitado então foi resuspenso em 40 ml de tampão A (100 mM TEA/HCl pH 7,6, 0,1 mM DTT, 0,2 mM EDTA e 0,1 mM PMSF). As paredes celulares foram quebradas através de duas passagens da solução de bactérias pela prensa francesa a 13 000 psi. Um dos passos existentes no protocolo do grupo do Dr. Paul Michels, o tratamento do produto da prensa francesa com endonuclease (750 unidades de Benzon nuclease - marca registrada de deoxy-ribonuclease) durante 30 minutos a 37°C e o passo seguinte de adicionar 20 mg de sulfato de protamina em agitação leve por 30 minutos à temperatura ambiente foi removido, depois de verificado o seu baixo rendimento. Este passo visava a quebra dos ácidos nucléicos em oligonucleotídeos que então seriam precipitados pelo sulfato de protamina, diminuindo assim a viscosidade final. Provavelmente este precipitando também a proteína ou mesmo o tratamento prolongado a uma temperatura poderia contribuir para a degradação proteolítica da enzima, apesar da presença de inibidores. Como a solução obtida depois da prensa francesa não apresentou viscosidade muito elevada, a remoção destes passos não afetou os passos seguintes e melhorou o rendimento final da purificação. Apesar deste passo ter sido usado no primeiro experimento de expressão, a quantidade de oligonucleotídeos no volume morto da primeira coluna de purificação foi expressivo, mostrando a ineficiência do processo.

O produto da prensa francesa foi centrifugado durante 30 minutos a 12 000 rpm, a 4°C no rotor SS34 da Sorvall. O sobrenadante, que contém a proteína, então foi submetido a uma primeira purificação por fracionamento com 65% de saturação de sulfato de amônia, adicionados lentamente em grãos finos e lentamente agitado em gelo e posteriormente mantendo a solução em repouso por 10 minutos. A solução foi então centrifugada durante 30 minutos a 12 000 rpm, a 4°C no rotor SS34 da Sorvall. Segundo o protocolo do Dr. Paul Michels, o precipitado deveria ser resuspenso no tampão B (25 mM TEA/HCl pH 8,0, 0,1 mM DTT, 0,2 mM EDTA, 1 mM azida sódica, 0,1 mM PMSF) de baixa força iônica para que a solução final tivesse força iônica próxima à do tampão C (25 mM TEA/HCl pH 8,0, 0,1 mM DTT, 0,2 mM EDTA, 1 mM azida sódica e 50 mM NaCl), tampão este usado para equilibrar a coluna de troca iônica do próximo passo de purificação.

Este processo mostrou-se trabalhoso e de difícil controle, já que medidas constantes da força iônica da solução (através da medida da condutividade) deveriam ser feitas e

também frequentes passos de concentrações e diluições para o ajuste final. Este passo foi então substituído por uma diálise durante a noite contra o tampão C. Uma pequena quantidade de precipitado apareceu no saco de diálise, inexpressiva frente à quantidade de proteína presente em solução.

Figura 3.2 - Perfil de eluição da aldolase de *T. brucei* pela coluna "CM Sepharose Fast Flow" do fabricante Pharmacia.

Figura 3.3 - Perfil de eluição da aldolase de *T. brucei* pela coluna "Sephacryl S200 HR" do fabricante Pharmacia. Somente frações do pico foram mostradas neste gráfico.

A solução final de 48 ml, depois da diálise foi centrifugada a 10 000 rpm, a 4°C, durante 30 minutos no rotor SS34 da Sorvall e aplicada na coluna "CM sepharose fast flow" da Pharmacia, dimensões de 1,5 cm de diâmetro por 29 cm de comprimento (51 ml), equilibrada e empacotada com tampão C. A proteína foi então eluída através do gradiente 0,05-1,5 M de cloreto de sódio (125 ml de tampão C + 125 ml de tampão C' - 25 mM

TEA/HCl pH 8,0, 0,1 mM DTT, 0,2 mM EDTA, 1 mM azida sódica e 1,5 M NaCl), num fluxo de 1,1 ml/min, coletadas em frações de 3,3 ml (3 min/fração). O resultado da coluna pode ser visto na Figura 3.2

Frações de 17 à 22 foram colocadas em diálise contra o tampão D (10 mM TEA/HCl pH 7,5, 0,1 mM DTT, 0,2 mM EDTA, 1 mM azida sódica e 500 mM NaCl) e depois concentradas até um volume final de 1,5 ml. A solução foi então aplicada na coluna S200 HR Sephacryl (Pharmacia), de dimensões 1,5 cm de diâmetro por 125 cm de comprimento, num total de 221 ml, empacotada e equilibrada com o tampão D.

A proteína foi então eluída num fluxo de 0,7 ml/min em frações de 3,5 ml (5 min/fração). O resultado pode ser visto na Figura 3.3. A medida de atividade foi feita para algumas frações e mostrou altíssima atividade, aproximadamente 3 vezes mais ativa que as amostras recebidas do Dr. Paul Michels (como mostrado por Dr. Michels, Chevalier *et al.*, submetido). A proteína fornecida inicialmente apresentava 20 U/mg, enquanto que o experimento de expressão apresentado nesta seção mostrou uma atividade de 50 U/mg).

Figura 3.4 - Eletroforese em resina feita no equipamento da Pharmacia PHAST com as amostras do pico que foi eluído da coluna Sephacryl S200 HR (da esquerda para a direita: frações 47, 48 e 49, padrão de peso molecular, frações 50, 51, 52, 53, 54, 55 e 56, padrão de peso molecular, frações 57, 58 e 59 e padrão de peso molecular). Como pode ser observado a proteína apresenta alto grau de pureza, sendo que a parte final do pico apresenta impurezas que foram removidas. Esta coluna apresentou resultados bastante promissores quanto à qualidade do material obtido.

A proteína assim obtida apresentou padrões de pureza excelentes em eletroforese de resina de poliácridamida em SDS (Figura 3.4).

3.5 Experimentos de Cristalização

Os experimentos de cristalização foram iniciados com a amostra fornecida pelo Dr. Paul Michels como descrito acima e os resultados não foram promissores. O experimento de expressão em Seattle foi bem sucedido, apesar da primeira expressão produzir apenas 8 mg de proteína devido à problemas nas colunas tanto de troca iônica como de filtração.

Melhoras no protocolo de expressão e purificação possibilitaram uma melhora na qualidade da proteína obtida (vide seção anterior) como também um aumento no rendimento (34 mg de proteína pura). A primeira amostra de proteína obtida, apesar dos problemas durante a purificação, mostraram resultados positivos com cristais sendo obtidos em condições diferentes das inicialmente observadas, apesar das mesmas terem sido testadas anteriormente, sendo que a morfologia mudou drasticamente. Cristais em forma hexagonal com dimensões que não ultrapassavam os 0,05 mm surgiram depois de um tempo de crescimento de aproximadamente 1 mês (Tabela 3.4). Em uma viagem para o síncrotron de CHESS, na Universidade de Cornell os cristais foram expostos a radiação X não mostrando nenhum padrão de difração. Com a proteína obtida através do protocolo modificado e apresentado na seção anterior, novos experimentos de cristalização foram feitos.

Tabela 3.4 - Condições de cristalização que produziram os melhores cristais que foram expostos no síncrotron SSRL em Stanford, na Califórnia, que apresentaram padrões de difração de até 3.0 Å de resolução.

Condições de cristalização da aldolase de <i>T. brucei</i> (Experimento de expressão em Seattle)	
Método:	Gota pendurada ou gota sentada
Solução de proteína:	3.4 mg/ml em 10 mM Tris/HCl pH 7.0-7.5, 50 mM NaCl
Solução do poço:	1.5-2.0 M NaCl, 0.1 M fosfato e citrate de sódio, pH 5.0-6.0
Combinação na gota:	2-4 volumes de proteína + 1 volume da solução do poço
Cristais:	Ao redor de 0.05 mm, cúbicos e hexagonais.
Co-cristalizações*:	Ribose-5-fosfato (5.3 mM), 6-fosfogluconato (4.2 mM), glucose-1-fosfato (8.2 mM), fosfoenolpiruvato (2.6 mM), levamisole (0.0875 mM), albendazole (39 mM)

* Todas os experimentos de co-cristalização foram feitos com excesso de inibidor. Os valores entre parênteses são os fornecidos como constante de inibição ou IC₅₀ dos compostos. Não foi possível verificar se os resultados de co-cristalização realmente produziram cristais complexados. Várias condições produziram cristais mas ainda pequenos para uma coleta de dados completa para a determinação da estrutura.

Experimentos de cristalização foram conduzidos imediatamente após a obtenção da

proteína pelo técnico de laboratório Steve Safarty do grupo do Dr. Wim Hol. Várias condições produziram cristais com boa morfologia mas ainda muito pequenos em dimensões, não ultrapassando os 0,05 mm nas dimensões máximas (Tabela 3.4) o que impedia qualquer tentativa de coleta nos geradores de radiação X da Universidade de Washington. Na próxima exposição dos cristais a radiação síncrotron, no Laboratório de Radiação Síncrotron de Stanford (SSRL), na Califórnia, obteve-se padrões de difração de até 3,0 Å de resolução. No entanto, os cristais não apresentaram estabilidade durante a exposição aos raios-X o que não possibilitou a obtenção de dados suficiente para a determinação da estrutura.

Os dados obtidos foram apenas suficiente para uma análise inicial que possibilitou a determinação do sistema cristalino e parâmetros de rede. Os cristais expostos pertencem ao sistema cristalino cúbico e têm parâmetros de rede de 212,7 Å. Não houve possibilidade de determinação do grupo espacial sem ambiguidade que pode ser P23 ou I23.

Segundo estes resultados uma análise pode ser feita quanto à quantidade de solvente. Supondo que os cristais pertençam ao grupo espacial P32 (grupo espacial número 195, sistema cristalino cúbico), que apresenta 12 operações de simetria, a presença de um tetrâmero na unidade assimétrica forneceria um número de Matthews de 5,02 Å³/Da, que é um valor muito elevado, sendo correspondente a 75,5% de solvente. Dois tetrâmeros por unidade assimétrica teria $V_m = 2,51 \text{ Å}^3/\text{Da}$, tendo portanto 51% de solvente. A última suposição é mais razoável para este grupo espacial.

Para o caso do grupo espacial I23 (grupo espacial número 197, sistema cristalino cúbico), a unidade assimétrica corresponde a metade da unidade assimétrica do grupo espacial anterior devido a operação de corpo centrado, apresentando 24 operações de simetria. Portanto um tetrâmero por unidade assimétrica fornece $V_m = 2,51 \text{ Å}^3/\text{Da}$, valor este razoável. Devido a este grupo ser de maior simetria que o anterior, espera-se que este seja o grupo espacial deste cristal. No entanto o fato de o braço flexível na região C-terminal ainda estar parcialmente desordenado poderia ter motivado um empacotamento de menor simetria, podendo portanto ser o grupo espacial anterior, P32. Claro que uma coleta de dados com um número maior de cristais nos fornecer informações suficientes para identificar o grupo espacial mas esta discussão está estendida aqui. Esforços se concentram principalmente em resolver a fragilidade do cristal aos raios-X.

3.6 Trabalhos Futuros

Os trabalhos desenvolvidos até o momento com a proteína aldolase foram extensivo e objetivando inicialmente a obtenção de cristais. Todo o trabalho foi concentrado na determinação da estrutura e portanto direcionado na obtenção de cristais que difratassem a alta resolução. Para o objetivo de desenho racional de drogas há a necessidade de obter um modelo com resolução superior a 2,5 Å. Portanto não apenas foi procurado resolver o problema de cristalização, mas obter condições reprodutíveis e cristais com boa qualidade de difração de raios X. A continuidade deste trabalho é totalmente dependente dos cristais pois os testes de novos inibidores modelados deverão ser feitos através de difusão. Co-cristalização só será realizada caso os cristais não resistam a difusão.

Portanto o trabalho desenvolvido neste projeto foi concentrado na cristalização e o objetivo principal (que não foi atingido durante o desenvolvimento deste projeto) seria a determinação da estrutura da aldolase de *T. brucei*. Fica claro do exposto nos capítulos anteriores que para um projeto de desenho racional de drogas, o domínio das condições de cristalização são cruciais (como poderá ser visto no Capítulo 4).

O trabalho desenvolvido pelo autor desta tese neste projeto foi basicamente a implementação do sistema de expressão através de clones fornecidos pelos colaboradores Dr. Paul Michels e Dr. Fred Oppendoes, melhora do protocolo de purificação para obter proteína em grau de pureza suficiente para cristalização e obtenção de informações iniciais de parâmetros de rede e sistema cristalino dos cristais obtidos.

As expectativas de continuidade deste projeto são agora bastante promissoras já que cristais foram obtidos e apresentaram um padrão de difração razoável. Novos experimentos variando as condições de cristalização assim como a possibilidade de coletas de dados a baixas temperaturas podem fornecer dados de difração suficientes para o início da determinação da estrutura pelo método da substituição molecular.

Posteriormente, a determinação da estrutura de complexos com inibidores, poderá permitir maiores informações a respeito do modo de ligação no sítio ativo assim como da posição do braço flexível sobre o sítio ativo. Estas informações irão permitir uma análise mais sólida da estrutura para possíveis desenhos de compostos de partida para o início do ciclo visando a obtenção de inibidores mais potentes e seletivos.

Agradecimentos devem aqui ser feitos à Dra. Ingeborg Feil que teve participação decisiva na expressão da aldolase e do Dr. Philips Petra, profundo conhecedor das técnicas de purificação que teve participação imprescindível neste projeto.

Todos os passos seguidos neste projeto podem ser generalizados para outros sistemas que tenham como objetivo o desenho racional de drogas. Neste mesmo enfoque, colaboração com o grupo do Dr. Paul Michels já teve início, o qual forneceu bactérias clonadas com o gene para expressão da enzima gliceraldeído-3-fosfato desidrogenase (GAPDH) de *T. cruzi*, agente causador da doença de Chagas, para o grupo de Cristalografia de Proteínas do Instituto de Física de São Carlos, USP, projeto este que permitirá a continuidade dos conhecimentos adquiridos durante o treinamento no grupo do Dr. Wim Hol.

Capítulo 4

**Experimentos de Difusão de Inibidores da
Triosefosfato Isomerase de *T. brucei*. Em busca de
Inibidores Seletivos para o Desenvolvimento
Racional de Drogas Contra a Doença do Sono**

4.1. Introdução

Triosefosfato isomerase de *T. brucei* é uma das proteínas envolvidas num projeto mais amplo de desenho racional de drogas para doenças tropicais como já mencionado nos Capítulos 1 e 3 e como será mencionado no Capítulo 5. Este projeto tem estreita colaboração com Dr. Fred Oppendoes e Dr. Paul Michels do ICP, Bruxelas, na Bélgica, Dr. Rik Wierenga do EMBL, Heidelberg, na Alemanha, Dr. Michael Kahn da Universidade de Washington e Dr. J. Perrie da Univerisdade de Paul Sabatier, Toulouse, França.

Figura 4.1 - Reação catalisada pela TIM.

Triosefosfato isomerase (TIM) é uma enzima glicolítica bem conhecida, que catalisa a interconversão do gliceraldeído-3-fosfato em dihidroxiacetona fosfato (Figura 4.1). TIM é uma enzima homodimérica com cerca de 250 resíduos em cada subunidade, tendo um peso molecular aproximado de 55 KDa. Alguns detalhes do mecanismo enzimático são apresentados no Capítulo 6. Evidências do mecanismo enzimático mostram que a TIM é uma enzima extremamente eficiente (Knowles e Alberly, 1977). Na espécie humana a deficiência desta enzima gera problemas sérios de saúde e a morte no estágio inicial de vida (Maquat *et al.*, 1985). O tripanossomo, o qual é mais dependente da glicólise do que os seres humanos, dependerá até mesmo mais do funcionamento apropriado desta enzima. Por esta razão inibidores da TIM de *T. brucei* que não afetem ou afetem parcialmente a enzima homóloga humana, são de grande interesse para o tratamento da doença do sono.

A estrutura tridimensional de várias enzimas TIM foram elucidadas pela técnica de cristalografia de raios X (galinha, Banner *et al.*, 1975, Alber *et al.*, 1981; levedura, Lolis e Petsko, 1990, Lolis *et al.*, 1990; *E. coli*, Noble *et al.*, 1993, *T. brucei*, Wierenga *et al.*, 1991b; humana, Mande *et al.*, 1994 e *Bacillus stearothermophilus*, Delboni *et al.*, em preparação -

Capítulo 6 deste trabalho) a qual evidenciou um enovelamento bastante usado pela natureza em várias proteínas que apresentam diferentes mecanismos enzimáticos: 8 fitas β formando um barril central envolvidas por 8 hélices α , ligadas por estruturas em volta. Uma vista esquemática pode ser observada na Figura 4.2, a qual apresenta características da TIM elucidadas pelos autores das estruturas determinadas até o momento que são:

- Três resíduos no sítio ativo que interagem diretamente com o substrato: Lys 13, His 95 e Glu 167 (numeração da TIM de *T. brucei*);
- His 95 encontra-se situada na parte N-terminal de uma curta "hélice do sítio ativo" compreendendo os resíduos 95 à 105;
- O grupo fosfato do substrato interage com peptídeos da parte N-terminal da "hélice que liga o grupo fosfato" (Lolis *et al.*, 1990);
- Na região N-terminal da fita β_6 , que compreende os resíduos 167-180, forma-se a chamada "estrutura em volta flexível" que cobre o sítio ativo durante a catálise (Alber *et al.*, 1981; Joseph *et al.*, 1990). Esta estrutura em volta pode ocupar as conformações "aberta" (na ausência de inibidores ou do substrato) ou "fechada" (na presença de inibidores ou substrato), conformações estas que serão extensivamente discutidas abaixo nos diferentes complexos já reportados.
- Contatos entre subunidades são feitos pelas hélices α_2 e α_3 , mais a estrutura em volta entre a fita β_1 e a hélice α_1 e mais a estrutura em volta entre a fita β_2 e a hélice α_2 , enquanto os resíduos 74-77, entre a fita β_3 e hélice α_3 da chamada "estrutura em volta de interface", atingem o sítio ativo da outra subunidade (Wierenga *et al.*, 1987).

A comparação de todas as estruturas de TIM (Capítulo 6 desta tese) mostra a grande similaridade das estruturas, apesar de algumas apresentarem similaridades na sequência inferiores a 50%. No contexto do projeto de desenho racional de drogas contra a doença do sono, a enzima TIM encontra-se em fase mais avançada pois já se conhecem as estruturas nas duas espécies envolvidas no problema da doença do sono: humana e *T. brucei*. Neste caso a seletividade do inibidor pode ser buscada com mais facilidade com a disponibilidade das duas estruturas, e da diferença nas resoluções das estruturas (TIM humana foi refinada somente até 2,8 Å de resolução enquanto que TIM de *T. brucei* foi refinada até 1,83 Å de resolução). A comparação das duas estruturas foi feita por Mande *et al.* (Mande *et al.*, 1994) a qual confirmou as conclusões já obtidas por Wierenga *et al.* (Wierenga *et al.*, 1987): resíduos num raio de 10 Å

Portanto TIM parece ser um alvo que apresenta dificuldades para a obtenção de um

possível resistência às drogas.

O estudo de complexos da enzima com os candidatos à inibidores pode ser feito de duas formas diferentes: por difusão, colocando-se o cristal em soluções que contenham o inibidor e por co-cristalização da enzima na presença do inibidor. Ambos os casos apresentam dificuldades. No primeiro a forma cristalina pode apresentar um empacotamento que esteja bloqueando o sítio ativo, íons podem estar ocupando o sítio ativo ou mesmo mudanças conformacionais da proteína podem destruir a rede cristalina e desta forma o cristal. Já no segundo caso na maioria das vezes a adição do inibidor causa mudanças conformacionais na proteína que então não permitem a cristalização da mesma nas condições obtidas anteriormente, o que leva então à busca de novas condições de cristalização, basicamente início de um novo projeto, que atualmente é possível devido à quantidade de proteína que se obtém da expressão em grande quantidade em bactéria. Ainda assim pode ser um problema limitado pelas quantidades de inibidores fornecidas pelos colaboradores já que a co-cristalização exige uma quantidade muito maior de inibidor do que a requerida em experimentos de difusão. A estabilidade dos inibidores também é importante para o caso da co-cristalização, que exige um tempo maior para obtenção dos cristais, ou seja, o tempo requerido na cristalização, tempo este controlado para o caso da difusão.

Como os primeiros cristais da TIM de *T. brucei* foram obtidos em alta concentração de sulfato de amônia (2,4 M - forma cristalina I) e como é sabido que o K_i do íon sulfato é de 4 mM (Lambeir *et al.*, 1987) é, desta forma, impossível a sua utilização, estando os cristais na solução de cristalização, para experimentos de difusão. Esta forma cristalina mostra o íon sulfato presente no sítio ativo da subunidade 2, sendo que neste sítio a estrutura em volta flexível era "quase-fechada". Na subunidade 1 a estrutura em volta flexível encontra-se aberta, cristalinos e a conformação das cadeias laterais não fornece condições para a ligação de um íon sulfato. Na posição do sulfato na subunidade 1 foi refinada uma molécula de água (Wierenga *et al.*, 1991b).

Uma segunda forma cristina (forma cristalina II) foi obtida em solução PEG+sulfato de amônia [18% PEG 6000 (p/v) com 50 mM de sulfato de amônia na solução de proteína], a qual não mostra a presença do íon sulfato em nenhum dos sítios de ligação. O experimento de difusão desta forma cristalina em 400 mM de sulfato de amônia mostrou a presença do íon em ambos os sítios com as estruturas em voltas flexíveis na conformação aberta (Noble, 1992). Esta forma cristalina, diferentemente da forma cristalina I, apresenta dois monômeros por unidade assimétrica ao invés de um dímero, mas existe semelhanças entre as duas subunidades da forma cristalina I e esta forma cristalina. A estrutura em volta flexível da subunidade 1 da forma cristalina II é semelhante à estrutura em volta flexível da subunidade 1 da forma cristalina I, a qual também se

encontra envolvida em contatos cristalinos. A estrutura em volta da subunidade 2 da forma cristalina II, que é semelhante à estrutura em volta da subunidade 2 da forma cristalina I, encontra-se livre de contatos cristalinos. Desta forma apesar de serem semelhantes devido à ausência de interações da estrutura em volta flexível da subunidade 2 de ambas as formas cristalinas, estas duas subunidades apresentam diferentes conformações pois na forma cristalina I a presença do íon sulfato faz com que a estrutura em volta apresente a conformação "quase-fechada" enquanto que na forma cristalina II esta estrutura em volta apresenta-se na conformação aberta. O íon da subunidade 1, na forma cristalina II após o experimento de difusão, possui baixa ocupação (Noble, 1992) e poderia, da mesma forma que na subunidade 1 da forma cristalina I, ser uma molécula de água.

A forma cristalina I não pode ser usada para experimentos de difusão com outros compostos se não for transferida para uma solução livre do íon sulfato. Mostrou-se que os cristais assim crescidos poderiam ser transferidos para uma solução de 44% (p/v) PEG 6000 sem perda apreciável do poder de difração (Schreuder *et al.*, 1988). Desta forma este cristal pode ser utilizado no processo de experimentos de difusão dos cristais em soluções com inibidores. A forma cristalina II também pode ser usada desde que transferida para uma solução livre de sulfato, resultado este observado por Noble (1992) com um dos complexos que serão descritos adiante. Desta forma, as duas formas cristalinas podem ser usadas no estudo de complexos da TIM de *T. brucei* com inibidores através de experimentos de difusão.

Vários complexos puderam então ser estudados através deste procedimento. Mesmo tendo um cristal livre de inibidores, ainda vários outros problemas podem ser importantes para o estudo de complexos. A acessibilidade do sítio ativo é de extrema importância e este é um dos problemas encontrados em ambas as formas cristalinas descritas acima. O sítio ativo da subunidade 1 encontra-se fixo na conformação aberta por interações com uma subunidade vizinha. Por outro lado o sítio ativo da subunidade 2 encontra-se livre de interações com subunidades vizinhas, o qual pode então ser utilizado para o estudo de complexos. Também para o estudo de inibidores que venham possivelmente a ser usados como drogas, constantes de inibição menor que μM são recomendadas, mas não necessárias para o início do ciclo de desenvolvimento da droga.

Para o caso da forma cristalina I, o inibidor análogo ao substrato, 2-fosfoglicolato, que possui um K_i de $27 \mu\text{M}$, deteriorou o cristal durante o experimento de difusão (Noble, 1992, Delboni, resultado não publicado) mostrando que esta forma cristalina não suporta os experimentos de difusão com inibidores com baixo K_i , os quais possivelmente produzem mudanças conformacionais no sítio ativo da subunidade 1, destruindo assim a rede cristalina. Inibidores com afinidades mais baixas, com K_i acima de $500 \mu\text{M}$, podem ser estudados através

da forma cristalina I.

Um inibidor com mais baixo K_i foi estudado com a forma cristalina II (fosfoglicolato-hidroxamato que possui K_i de $4 \mu\text{M}$).

Segundo os resultados apresentados, o projeto encontra-se provido de duas formas cristalinas que permitem estudar complexos com inibidores incluídos em um amplo intervalo de constante de inibição.

Apesar de se ter um inibidor que permita a difusão e ter o cristal transferido para um líquido mãe, livre de íons, que são inibidores competitivos do sítio ativo, o resultado pode apresentar-se negativo em vários casos. Assim foi o ocorrido com o composto ácido-1-aminonaftaleno-4,6,8-trissulfônico (ANTS) que foi fornecido pelo Dr. Nickel, Bonn, Alemanha, e que apresentou um K_i de $300 \mu\text{M}$ (medidas feitas pelos Drs. A. M. Lambeir e Fred Oppendoes, resultado não publicado). Depois da difusão dados foram coletados até a resolução de $2,8 \text{ \AA}$ no detector de área FAST e a estrutura foi determinada usando-se o modelo da TIM da forma cristalina I. Refinamento por dinâmica molecular convergiu rapidamente para um fator R de 13,9% com boa geometria e 30 moléculas de água (Wierenga *et al.*, 1991a). Nenhum inibidor foi observado na densidade eletrônica do sítio ativo. Por outro lado vários outros compostos apresentaram resultados positivos os quais serão brevemente descritos aqui.

O complexo entre 2-fosfoglicolato (2PG) e TIM foi largamente estudado por vários autores em várias TIM (Lolis e Petsko, 1990, complexo com a TIM de levedura, Mande *et al.*, 1994, complexo com a TIM humana e Delboni *et al.*, 1994, complexo com a TIM de *Stearothermophilus*. Uma discussão a respeito das características destes complexos encontra-se no Capítulo 6 desta Tese). No entanto este complexo não foi obtido para a TIM de *T. brucei* (Noble, 1992). Em todos os casos acima, os complexos foram obtidos por co-cristalização, mostrando que muito provavelmente um complexo entre a TIM de *T. brucei* e 2PG só será possível por co-cristalização. Como mencionado a constante de inibição deste composto é de $27 \mu\text{M}$ e a difusão na forma cristalina I deteriorou o cristal (Noble, 1992, Delboni, resultado não publicado). Difusão na forma cristalina II não foi testada pois acredita-se que este complexo não introduzirá novas informações dentro do número de complexos já obtidos.

Compostos promissores, como inibidores que poderiam ser desenvolvidos em drogas seletivas, deveriam se ligar ao sítio de ligação do fosfato e se estender para o sítio seletivo à mais de $12 \text{ \AA}$ do sítio do fosfato, ao invés de se estender para dentro do sítio ativo como nos complexos já estudados.

Uma das tentativas de se obter um composto que se ligue à parte "não-fosfato", fora do sítio e em direção ao sítio de seletividade foi sugerida por Callens e Oppendoes: ribose-5-fosfato

(R5P) que possui um K_i de aproximadamente 5 mM (Dra. Mia Callens e Dr. Fred Oppendoes, resultado não publicado). Da mesma forma os cristais foram transferidos para um líquido mãe livre de sulfato, 44 % (p/v) PEG 6000 (Schreuder *et al.*, 1988) contendo 15 mM de ribose-5-fosfato. Dados foram coletados num detector de área FASTO à 2,4 Å de resolução. Refinamento da estrutura foi feito e um fator R de 15% foi obtido com boa geometria e 156 moléculas de água e somente um íon mono hidrogênio fosfato foi encontrado na densidade eletrônica no sítio de ligação do fosfato (Verlinde *et al.*, 1991). O íon fosfato encontra-se 1,0 Å mais próximo do sítio ativo que o íon sulfato e a estrutura em volta flexível encontra-se aproximadamente 0,5 Å mais fechada que no caso do íon sulfato. As moléculas de água também parecem ocupar posições diferentes, formando um arranjo diferente. Observa-se também que o resíduo catalítico Glu 467 adota uma conformação diferente. A diferença em tamanho dos dois íons é menor que 0,1 Å e então parece que a presença de um hidrogênio deve ser o responsável pelos resultados observados (Verlinde *et al.* 1991).

Fosfatos e fosfonatos apresentam diferenças em relação à sua afinidade pela TIM. Como pode ser observado, 2-fosfoglicolato (2PG) possui um K_i de 27 μ M para a TIM de *T. brucei* (Lambeir *et al.*, 1987), é um composto análogo ao estado de transição (Figura 4.3) o qual foi estudado por Wolfenden e co-autores (Wolfenden, 1969, Collins, 1974). Sugeriu-se então, baseado na estrutura do 2-fosfoglicolato, que o 3-fosfonopropionato (3-PP) (Figura 4.3) poderia ser um inibidor em potencial (Dr Alan Horn, resultado não publicado). Ensaio enzimático mostrou constante de inibição muito pior que a do 2PG (ao redor de 10^3 vezes pior, Dr. Fred Oppendoes, resultado não publicado). Novamente os cristais foram transferidos para um líquido mãe livre de sulfato e em 44 % (p/v) PEG 6000. Estes cristais foram colocados em um banho no qual se adicionou 100 mM de 3-PP. Dados de difração foram coletados até resolução máxima de 2,8 Å e o modelo atômico refinado até um fator de concordância $R=12,5\%$, com 38 moléculas de água e 9 átomos do ligante (Noble *et al.*, 1991b). A molécula de 3-PP liga-se ao sítio ativo com todos os átomos de oxigênio envolvidos em ligações de hidrogênio, uma das quais com a cadeia lateral do resíduo Glu 467. Um aspecto importante desta estrutura é que a posição do fosfato do 3-PP está à 0,2 Å do fosfato do mono hidrogênio fosfato discutido acima. Também a estrutura em volta flexível encontra-se completamente fechada como no caso do mono hidrogênio fosfato e não "quase fechada" como no caso do íon sulfato. Portanto a razão para a baixa afinidade da TIM de *T. brucei* à molécula 3-PP não é devido à estrutura em volta flexível estar mais aberta, nem devido à posição não usual do grupo fosfato. Também não é devido à falta de interação dos oxigênios pois a troca de um oxigênio do fosfato por um grupo CH_2 não resulta em ligação de hidrogênio não satisfeita. Portanto, razões para esta diferença em

afinidade podem ser:

1. ângulos de torção menos favoráveis em 3-PP ligado à TIM quando comparado a 2-fosfoglicolato (2PG);
2. Diferenças nas interações com moléculas de água dos dois inibidores;
3. Diferenças no pK_a dos grupos fosfato e fosfonato (veja também Belasco *et al.*, 1978).

Nenhuma discussão além deste ponto pode ser estendida pela ausência da estrutura do complexo da molécula de 2PG e TIM de *T. brucei* e também pela baixa resolução da estrutura do complexo de TIM de *T. brucei* com 3-PP.

Outros inibidores estudados foram glicerol-3-fosfato (G3P) e 3-fosfoglicerato (3PGA) (Figura 4.3) complexados com TIM de *T. brucei*. Ambos os compostos têm um átomo a mais que a molécula 3-PP. Em adição, 3PGA e G3P têm um grupo hidroxil adicional na posição 2 quando comparados ao 3-PP. 3PGA e G3P diferem meramente pela troca do carboxilato do 3PGA para um grupo hidroximetil em G3P. Isto remove uma carga perto do Glu 467 do complexo com G3P quando comparado com o complexo com 3PGA.

As estruturas tridimensionais dos complexos com 3PGA e G3P foram determinadas de forma análoga à citada acima para o 3PP. O fator R do complexo 3PGA depois do refinamento foi de 14% para uma resolução de 2,2 Å e do complexo G3P foi de 13,7 % com essencialmente a mesma resolução (Noble *et al.*, 1991b).

Todas as estruturas dos complexos são bastante similares apesar da diferença entre os três ligantes. A posição do fosfato dos ligantes 3PGA, G3P, 3PP e HPO_4^{2-} está dentro de 0,2 Å um em relação ao outro, ou seja, parece haver um sítio bem definido para fosfatos e fosfonatos. Este "sítio do fósforo" parece ser independente da natureza precisa do inibidor ligado ao sítio ativo.

No complexo TIM-3PP a cadeia lateral do Glu 467 está envolvida numa ligação de hidrogênio com o grupo carboxilato da molécula 3PP como já descrito. Esta conformação do Glu 467 é chamada de "swung-in"¹. Apesar da molécula 3PGA apresentar o mesmo grupo carboxilato, a conformação do resíduo Glu 467 é a chamada "swung-out". Esta conformação difere da "swung-in" por uma rotação de 120° sobre a ligação $C^\alpha-C^\beta$, 40° sobre $C^\beta-C^\gamma$ e 60° sobre $C^\gamma-C^\delta$. Em vista das similaridades era esperado que a conformação do resíduo Glu 467 fosse a mesma para todos os complexos mas não é verdade para o caso do complexo TIM-3PGA que adota a conformação "swung-out" (maiores detalhes a respeito destas conformações podem

¹ A tradução deste termo, específico para a conformação de um resíduo na estrutura da TIM, não foi feita por ser encontrado apenas em referências específicas da estrutura da enzima TIM.

ser vistos no Capítulo 6 desta tese).

Figura 4.3 - Esquema dos inibidores de TIM: (a) 2-fosfoglicolato (2PG) não observado em complexo com TIM de *T. brucei*. (b) 3-fosfopropionato (3PP). (c) 2-fosfoglicerato (2PGA). (d) glicerol-3-fosfato (G3P). (e) 3-fosfoglicerato (3PGA). (f) 2 (N-formil-n-hidroxi-amino) etil-fosfonato (IPP). (g) fosfoglicolato-hidroxi-amato (PGH) e (h) N-hidroxi-4-fosfono-butanamida.

Outro complexo dentro do projeto de desenho racional de drogas foi do 2-fosfoglicerato (2PGA - Figura 4.3) com TIM de *T. brucei*. Este composto apresenta uma constante de inibição de 6,9 mM (Lambeir *et al.*, 1987). Este inibidor é similar ao já discutido acima, 3PP, já que os dois tem um grupo carboxílico separados por três ligações covalentes do átomo de fósforo (Figura 4.3). 2PGA é uma molécula maior que 3PP já que tem um grupo metoxi adicional. Dados de difração foram coletados até 2,4 Å de resolução e o refinamento da estrutura apresentou um fator de concordância R final de 14,9% com 130 moléculas de água e 11 átomos do inibidor (Noble *et al.*, 1991a).

Comparação do complexo da TIM com 2PGA com todos os complexos já apresentados acima mostram um desvio de 0,27 Å (quando comparado com G3P) e até 0,33 Å (comparado

com sulfato, 3PGA e 3PP) na posição do grupo fosfato. 2PGA ajusta-se bem apesar de ser maior que 3PP pois o Glu 467 adota aqui a conformação "swung out" que fornece mais espaço para ser ocupado.

Dois outros complexos foram obtidos por Noble (Noble, 1992): fosfoglicolohidroxamato (PGH) e 2-(N-formil-hidroxiamino-etilfosfonato) (IPP) (Figura 4.3). PGH possui uma constante de inibição de $4\text{ }\mu\text{M}$ (Collins, 1974). A difusão com $16\text{ }\mu\text{M}$ deste composto na forma cristalina II resultou na ruptura do cristal. A estabilidade do cristal para difusão deste inibidor foi obtida com a transferência do cristal para uma solução contendo 44% PEG 6000 (p/v). Nestas condições o cristal sobreviveu a um experimento de difusão com até $40\text{ }\mu\text{M}$ de PGH, o qual possibilitou obter dados de difração até $2,5\text{ }\text{\AA}$ de resolução. O refinamento convergiu para um fator de concordância R de 14,7% com 61 moléculas de água e 10 átomos do PGH. A estrutura revelou apenas uma molécula ligada à subunidade 2 desta forma cristalina, que se encontra de contatos cristalinos, na qual a estrutura em volta flexível adotou a conformação fechada. O resíduo Glu 467 adota a conformação "swung in" em acordo com os outros inibidores similares apresentados anteriormente.

IPP foi difundido com sucesso na forma cristalina I, previamente transferida para uma solução contendo 44% (p/v) PEG 6000 e 100 mM de sulfato de amônia. O IC_{50} (concentração na qual a atividade enzimática decresce para 50% do valor observado na ausência de inibidor) do IPP é de 0,5 mM (Dra. Mía Callens, resultado não publicado). A difusão foi obtida em uma solução contendo 44% PEG 6000 (p/v) com 2,5 mM de IPP ao invés de 100 mM de sulfato de amônia. Deste cristal foi possível coletar dados de difração até $2,5\text{ }\text{\AA}$ de resolução e a estrutura foi refinada até um fator de concordância R de 15,3% com 106 moléculas de água e 10 átomos do IPP (Noble, 1992). Este complexo liga-se também à subunidade 2 de forma bastante similar aos complexos já discutidos, com a estrutura em volta flexível na conformação fechada e o Glu 467 na conformação "swung in". Detalhes das interações destes inibidores com a proteína podem ser vistos nas publicações referidas.

Há ainda mais um complexo da TIM de *T. brucei* com N-hidroxi-4-fosfono-butanamida (4PBH), o qual tem um carbono a mais que os inibidores descritos acima (Figura 4.3) e K_i de 0,33 mM. Este complexo teve a estrutura determinada de forma similar (Verlinde *et al.*, 1995) usando a forma cristalina I transferida para 50% (p/v) PEG 6000 com 4,3 mM de 4PBH. O interessante deste resultado é que este inibidor liga-se à conformação aberta abrindo mais uma possibilidade para procura de potenciais inibidores para o projeto de desenho racional de drogas como será discutido no Capítulo 5 deste trabalho.

Os resultados apresentados acima são apenas os primeiros passos de um longo caminho

em direção ao desenho seletivo de inibidores da TIM do tripanossomo. Muitos dos complexos acima não apresentam a importância que se pretende em desenho de novas drogas. Nenhum dos inibidores acima apresenta extensões para a região seletiva, sendo apenas variantes do mesmo modo de ligação ao sítio ativo. Apenas o último complexo TIM-4PBH (Verlinde *et al.*, 1992b) apresenta um novo estado conformacional do sítio ativo que pode ser explorado para desenhos de novas drogas (Capítulo 5 desta tese). Como mostrado na Tabela 4.1 as regiões hidrofóbicas que apresentam diferenças relativas entre TIM de *T. brucei* e humana, encontram-se à distâncias ao redor de 12-19 Å do sítio de ligação do grupo fosfato. Neste caso pode-se observar que devido à presença na espécie humana de apenas uma isoenzima, este caso parece ser mais simples que no caso descrito da frutose-1,6-bisfosfato aldolase que apresenta três diferentes isoenzimas o que torna o problema da seletividade muito mais complexo (Capítulo 3 desta tese).

Tabela 4.1 - Distâncias em Å entre o átomo de fósforo do complexo G3P-TIM de *T. brucei* e átomos específicos do tripeptídeo Ala400-Tyr401-Tyr402.

Átomo	Distância (Å)
C α (Ala 400)	15.6
C β (Ala 400)	14.4
C α (Tyr 401)	16.7
OH (Tyr 401)	12.1
C α (Tyr 402)	19.0

Vários enfoques podem ser dados a este problema. No caso particular da TIM de *T. brucei* não somente a conformação "fechada" da estrutura em volta flexível que foi tratada até o momento como também a conformação "aberta" observada em um dos inibidores (TIM-4PBH), servem como base de um projeto para procura de inibidores seletivos (a conformação "aberta" da estrutura em volta flexível é tratada no próximo capítulo desta tese).

Existem, por outro lado, vários programas que permitem uma procura sistemática de inibidores analisando-se todos os compostos disponíveis no mercado e com estruturas conhecidas pela avaliação do seu ajuste ao "sítio-alvo" da enzima. Diferentes programas tratam o problema de maneiras diferentes, sendo que alguns simplesmente procuram satisfazer as ligações de hidrogênio e as interações hidrofóbicas, ponderando então o resultado final segundo às quantidades e qualidades destas interações. Este é o caso do programa LUDI (Böhm, 1992).

Existe outro programa que procura encaixar os compostos no sítio ativo simplesmente através de um ajuste espacial e separadamente calculando um índice de qualidade que depende do grau de complementariedade estérica e das interações eletrostáticas. Este é o caso do programa DOCK (Shoichet *et al.*, 1992) que é discutido e aplicado no Capítulo 5 deste trabalho. Também existe o procedimento mais rudimentar da análise gráfica do sítio de ligação para fornecer sugestões de um possível ligante que venha a preencher o espaço entre o sítio de ligação do fosfato e interagir com os sítios seletivos entre a TIM das diferentes espécies. Neste último processo utiliza-se do senso crítico do pesquisador que em alguns casos se sobressai aos resultados dos programas. No entanto, muitos dos resultados propostos são compostos de difícil síntese química o que torna os resultados de pouco uso. Através de uma análise detalhada do caminho entre o sítio de ligação do fosfato e do sítio seletivo, Dr. C. L M. J. Verlinde propôs uma molécula, a qual é apresentada na Figura 4.4 (Verlinde *et al.*, 1992a). Esta molécula apresenta quatro fragmentos cada um dos quais têm diferentes funções relativamente ao sítio de ligação da enzima e encontram-se envolvidos em diferentes interações com a enzima, a saber:

Figura 4.4 - A molécula DG-11 modelada para inibir seletivamente a TIM de *T. brucei*.

- (a) A âncora a qual bloquearia a entrada do sítio ativo;
- (b) O "preenchedor da cavidade NN", o qual se estende na cavidade bem definida pelos resíduos Asn 311 e Asn 315 (N311-N315);
- (c) O fragmento de "interface" localizado na interface das duas subunidades, o qual serve de ligação entre os dois fragmentos mencionados nos itens (a) e (b) e o próximo mencionado no item (d);

- (d) O fragmento "seletivo" o qual se ajusta na área hidrofóbica compreendida pelos resíduos Ala 400 e Tyr 401, que deveria prover o inibidor com seletividade contra a proteína humana.

Vários fatores, como problemas com a modelagem, moléculas de água que não foram consideradas, erros experimentais assim como adaptabilidade da proteína fazem com que este composto muito dificilmente venha a ser testado em um único passo. Ao invés disto, um tratamento de "fragmentos ligados" pode ser mais realístico, ou seja, diferentes fragmentos funcionais poderiam ser estudados separadamente. Quando terminado o estudo cristalográfico dos vários fragmentos em complexo com a proteína, então o próximo passo seria a ligação destes fragmentos formando uma molécula maior, caso seja possível e viável através da síntese química. Esta molécula deveria em princípio ligar-se melhor que os fragmentos separados em particular se ela pudesse ser feita mais rígida.

Claramente existe um longo caminho a ser seguido e parte do mesmo teve início com os trabalhos realizados nesta tese. Quando o autor desta tese iniciou sua participação neste projeto, este encontrava-se paralisado por aproximadamente dois anos por problemas na cristalização da TIM de *T. brucei* apesar de cristais já terem sido obtidos de proteína expressa em *E. coli*, porém sem repetibilidade. Através de uma análise sistemática concluiu-se que o tempo decorrido entre a preparação da enzima e sua utilização em experimentos de cristalização era um fator crucial na qualidade dos cristais. Quanto mais antiga a proteína, piores os cristais e o tempo médio era bastante curto, sendo cerca de um a dois meses. Cristais que cresciam até 1 mm passavam por um decréscimo e perda de uma dimensão, transformando-se em placas e finalmente em precipitado amorfo.

Com a conclusão observada acima uma única solução foi possível: a cristalização imediatamente após a expressão em bactérias. Para isto toda a expressão deveria ser feita na Universidade de Washington, em Seattle, USA e não no ICP, em Bruxelas, na Bélgica pois o atraso no envio da proteína tornaria os resultados insatisfatórios como havia acontecido anteriormente. Com a colaboração de vários grupos encarregados da síntese, pôde-se retomar o projeto onde havia sido interrompido. Nas próximas seções será descrito todo o trabalho de expressão, cristalização nos diferentes grupos espaciais e finalmente difusão de inibidores e determinação da estrutura tridimensional dos possíveis complexos, realizados pelo autor desta tese.

4.2 Expressão em Grandes Quantidades da TIM de *T. brucei* em *E. coli*

Observou-se claramente uma correlação marcante entre o tempo da preparação da amostra e a qualidade dos cristais obtidos. Se todos os experimentos de cristalização não fossem feitos no prazo de aproximadamente um mês, a proteína tornava-se inadequada para utilização. Desta forma como esta proteína já se encontrava em um sistema de expressão, o qual foi gentilmente cedido pelo Dr. Rik Wierenga, EMBL, Heidelberg, Alemanha, decidiu-se então por iniciar a expressão em grandes quantidades no laboratório do Dr. Wim Hol, na Universidade de Washington, em Seattle.

No primeiro experimento de expressão os problemas surgiram na purificação, no qual buscou-se otimizar as etapas do processo de purificação em FPLC (Fasto Performance Liquid Chromatography, Pharmacia). Vários problemas surgiram como dificuldade em ligar a proteína à coluna de troca iônica e perda de atividade já que a purificação foi feita à temperatura ambiente. Desta forma manteve-se as etapas de purificação o mais próximo ao protocolo fornecido pelo Dr. Wierenga.

Uma pré-cultura das bactérias contendo o gene para expressão da TIM do *T. brucei* (plasmídeo pET3A TIM(12) BL21(DE3)) foi crescida em meio sólido (placas de ágar em meio LB - 1,6% de ágar em meio LB que consiste de 1% triptona, 0,5% extrato de levedura e 0,5% de cloreto de sódio, sendo autoclavadas antes do uso e adicionado 0,01% de ampicilina no momento do uso), a 37°C durante toda a noite, para isolar uma única linhagem. Após este primeiro procedimento a bactéria foi então inoculada em 2 ml do meio LB mais 0,01% de ampicilina, mantida a 37°C sob agitação a 200 rpm durante toda a noite. A pré-cultura foi então inoculada em 200 ml de meio LB com 0,01% de ampicilina, em um frasco tipo erlenmeyer para manter uma alta razão de ar, proporcionando assim alto nível de oxigenação. Esta cultura foi incubada durante 6 horas a 37°C sob agitação a 200 rpm. A esta solução foram adicionadas 800 ml do meio LB com 0,01% de ampicilina e incubado à 37°C até $OD_{600}=0,6$. Como no caso da aldolase (Capítulo 3) a solução foi distribuída em 5 frascos tipo erlenmeyer tendo cada um 200 ml da solução para manter a alta razão de oxigenação.

Neste estágio fez-se a indução através da adição de 96 mg de IPTG e incubou-se três horas e toda a noite, à 37°C agitando à 200 rpm. Neste sistema não existe o problema de manter a indução durante toda a noite. Ao contrário o rendimento é extraordinário quando se mantém a indução durante toda a noite o que não acontece com outros sistemas como no caso da expressão da aldolase do *T. brucei*, já mencionado no capítulo anterior e no caso da

gliceraldeído-3-fosfato desidrogenase (GAPDH) de *L. mexicana* (Dr. Hidong Kim, comunicação pessoal) e *T. cruzi* (A. P. U. Araujo e D. H. F. Souza, comunicação pessoal).

Após a indução, as células foram centrifugadas durante 8 minutos a 8000 rpm, a 4°C no rotor GSA da Sorvall. As células foram ressuspensas em 40 ml de tampão A (25 mM MES/NaOH pH 6,5, 1 mM DTT, EDTA e 0,2 mM PMSF) e lisadas através de três passagens pela prensa francesa a uma pressão de 16 000 psi. Assim como na expressão da aldolase (Capítulo 3), o tratamento com endonuclease e sulfato de protamina não foram incluídos, o que não afetou o rendimento final. Desta forma o lisado foi centrifugado durante 30 minutos a 15 000 rpm, a 4°C no rotor SS34 da Sorvall. Devido a ausência de sal no tampão A, TIM apresenta baixa solubilidade e encontra-se no precipitado. Este precipitado foi ressuspensão em 40 ml do tampão B (25 mM MES/NaOH pH 6,5, 1 mM DTT, EDTA, 0,2 mM PMSF e 200 mM NaCl), o qual resolubilizou a enzima TIM, devido a presença de 200 mM de cloreto de sódio. Novamente a solução foi centrifugada durante 30 minutos, a 15 000 rpm, a 4°C no rotor SS34 da Sorvall.

O primeiro passo de fracionamento por sulfato de amônia foi feito levando a solução a 45% de saturação (27,7 g/100 ml) mantida em gelo por 10 minutos. A solução foi então centrifugada durante 30 minutos, a 15 000 rpm, a 4°C no rotor SS34 da Sorvall. O precipitado foi descartado e o sobrenadante foi usado para o próximo passo do fracionamento por sulfato de amônia. A solução foi levada até 65% de saturação em sulfato de amônia (adição de 13,2 g/100 ml em uma solução à 45% de saturação em sulfato de amônia produz uma solução final de 65% de saturação em sulfato de amônia) e mantida em gelo por 10 minutos. A solução final foi então centrifugada durante 30 minutos, a 15 000 rpm, a 4°C, no rotor SS34 da Sorvall. O precipitado continha a enzima TIM de *T. brucei*, verificado por ensaio enzimático e eletroforese em resina de poliacrilamida em presença de SDS. O precipitado foi ressuspensão no tampão C (100 mM TEA/HCl pH 8,0, 1 mM DTT, EDTA, azida sódica e 50 mM NaCl) e dialisado durante toda a noite contra o tampão C para certificar que a força iônica final era a do tampão C. Esta foi a diferença introduzida também neste protocolo assim como na purificação da aldolase. A segurança observada na força iônica através da diálise supera a rapidez do método de troca de tampão através de sucessivas concentrações. Observou-se a presença de material precipitado no saco de diálise mas a quantidade não foi significativa.

A solução assim obtida foi concentrada até um volume de 1,5 ml e aplicada na coluna CM sepharose fasto flow (Pharmacia) empacotada e equilibrada com tampão C, dimensões de 1,5 cm de diâmetro por 15 cm de comprimento, totalizando 26 ml. A proteína foi eluída em um

gradiente de 50 ml de tampão C e 50 ml de tampão C' (100 mM TEA/HCl pH 8,0, 1 mM DTT, EDTA, azida sódica e 150 mM NaCl), fluxo de 1,25 ml/min, frações de 1 ml/tubo. O resultado pode ser visto na Figura 4.5.

As frações 52 a 67 que mostraram o melhor padrão de pureza em eletroforese foram unidas e dialisadas contra o tampão da próxima coluna de filtração, tampão D (100 mM TEA/HCl pH 7,5, 1 mM DTT, EDTA, azida sódica e 500 mM NaCl) e depois concentradas até 3 ml. Pode ser observado que no padrão de eluição aparecem dois picos com atividade de TIM (Figura 4.5). Estes dois picos podem sugerir duas isoenzimas ou mesmo duas enzimas TIM diferentes. Devido à grande quantidade obtida no segundo pico e por ser este resultado de uma expressão induzida fica claro que o segundo pico é a TIM de *T. brucei*.

Figura 4.5 - Coluna de troca iônica CM-Sephacryl Fast Flow (Pharmacia). Passo de purificação da TIM expressa em *E. coli*.

A amostra então foi aplicada na coluna Sephacryl S200HR (Pharmacia), empacotada e equilibrada com tampão D, com dimensões 1,5 cm de diâmetro por 125 cm de comprimento, num total de 221 ml. A proteína foi então eluída num fluxo de 0,7 ml/min em frações de 1 ml (1,4 min/fração).

O resultado pode ser visto na Figura 4.6. O ensaio enzimático foi feito apenas para algumas frações e mostrou alta atividade, ao redor de 13 000 U/mg (cristais de boa qualidade têm aparecido de proteína com atividade enzimática entre 4000 e 13 000 U/mg).

O rendimento do experimento de expressão foi de 70 mg de proteína de alto grau de pureza. Esta proteína foi utilizada para os experimentos de cristalização. A proteína não usada

nos experimentos de cristalização foi estocada em 80% de sulfato de amônia, à 4°C e posteriormente utilizada em medidas de constante de inibição dos compostos sintetizados por Jean Lee, discutidos nas seções seguintes.

Figura 4.6 - Coluna de filtração S200HR-Sephacryl (Pharmacia). Passo de purificação da TIM de *T. brucei* expressa em *E. coli*.

4.3 Experimentos Cristalização

Condições de cristalização para TIM de *T. brucei* são conhecidas para três diferentes grupos espaciais, com três diferentes empacotamentos cristalinos. As condições de cristalização são:

1. Forma cristalina I, grupo espacial P2₁2₁2₁ - A estrutura desta forma cristalina foi refinada até 1,83 Å por Wierenga (Wierenga *et al.*, 1991b), cristalizada em experimentos de gota pendurada.

Solução de proteína 4,5 mg/ml em 0,9 M sulfato de amônia, 25 mM Tris/HCl pH 7,8, 1 mM EDTA, azida sódica, DTT.

Solução do poço: 2,2-2,4 M sulfato de amônia, 0,2 M MOPS pH 7,0-8,0, 1 mM EDTA, azida sódica e DTT.

Tempo: Cristais aparecem após duas semanas; crescimento se completa após um mês.

Cristais apresentam boa morfologia tridimensional e dimensões excelentes (alguns chegam a ter mais de 1 mm). Cristais apresentam um dímero por unidade assimétrica, tendo a estrutura em volta flexível da subunidade 1 na conformação aberta devido ao contato cristalino com outra subunidade relacionada por operação de simetria, e a estrutura em volta flexível da subunidade 2 na conformação fechada com um íon sulfato ocupando o sítio ativo. Esta é a melhor forma cristalina para o estudo de complexos apesar da necessidade de troca do líquido mãe para uma solução livre de sulfato, que é um inibidor competitivo do sítio ativo.

2. Forma cristalina II, grupo espacial C2, com 2 subunidades por unidade assimétrica. A estrutura desta forma cristalina foi resolvida por Noble (1992) e usada para o estudo de um dos complexos com TIM de *T. brucei* discutidos acima.

Solução de proteína: 5 mg/ml em 50 mM sulfato de amônia, 25 mM Tris/HCl pH 7,8, 1 mM EDTA, azida sódica, DTT, 25 mM NaCl.

Solução do poço: PEG 6000 16-24% p/v, 0,2 M Tris/HCl pH 8,0-9,0, 1 mM EDTA, azida sódica, DTT.

Tempo: Cristais aparecem após uma semana sem aparente melhora depois disto.

Cristais apresentam boa morfologia com dimensões que chegam a 1 mm. O empacotamento cristalino apresenta duas subunidades de dímeros diferentes. A região da estrutura em volta flexível apresenta uma conformação bastante semelhante às duas subunidades da forma cristalina descrita no item 1. Uma apresenta a estrutura em volta flexível na conformação aberta devido aos contatos cristalinos e a outra apresenta a estrutura em volta flexível também na conformação aberta, sem nenhum íon sulfato no sítio ativo e livre de contatos cristalinos. Esta forma cristalina foi usada em dois experimento de difusão em 400 mM de sulfato de amônio e 40 μ M de PGH (fosfoglicolato-hidroxamato) com prévia transferência para uma solução de 44% (p/v) de PEG 6000 para melhor estabilidade como já discutido anteriormente (seção 4.1).

3. Forma cristalina III, grupo espacial C2, 1 monômero por unidade assimétrica. A e desta forma cristalina foi determinada por Noble (1992).

Solução de proteína: 5 mg/ml em 25 mM Tris/HCl pH 7,6, 1 mM EDTA, azida sódica, DTT, 50 mM NaCl.

Solução do poço (1): PEG 6000 16-22% p/v, 0,2 M Tris/HCl pH 8,0-9,0, 1 mM EDTA, azida sódica, DTT.

Solução do poço (2): PEG 6000 16-22% p/v, 0,2 M Tris/HCl pH 8,0-9,0, 1 mM EDTA, azida sódica, DTT, **5% DMSO**.

Tempo: Cristais aparecem no máximo após dois dias.

Em ambas as condições descritas acima foi possível obter os cristais, os quais estão distribuídos em várias morfologias, sendo facilmente encontrados bons cristais com mais de 1 mm em tamanho. Nesta forma cristalina a subunidade na unidade assimétrica encontra-se com a estrutura em volta flexível na conformação aberta mantida através de contatos cristalinos. Esta forma cristalina parece não ser a ideal para os estudos de complexos já que a única conformação não permite mudanças conformacionais na estrutura em volta flexível. Talvez estudos de complexos que se liguem à conformação aberta poderiam ser testados mas ainda assim inibidores que se liguem a conformação aberta poderiam motivar mudanças conformacionais na estrutura que poderiam levar à destruição da rede cristalina e por sua vez do cristal.

4.4 Estudos de Inibidores da TIM de *T. brucei*

Como apresentado acima o problema da obtenção dos cristais de TIM de *T. brucei* à partir de proteína recombinante foi por nós solucionada. Este problema tem sido o maior empecilho neste projeto nos últimos 3 anos. O grupo do Dr. Wim Hol não vinha obtendo mais cristais devido ao problema de envelhecimento da proteína o qual foi solucionado com a expressão e utilização imediata da proteína, trabalho este descrito na seção 4.3 acima. Centenas de cristais foram obtidos da expressão que produziu 70 mg de proteína, os quais foram utilizados o mais rápido possível. Mesmo preparando o maior número de condições de cristalização não foi possível a utilização de toda a proteína, a qual foi utilizada nos experimentos de ensaios enzimáticos e medidas de constante de inibição pela estudante de doutorado Jean Lee do grupo do Dr. Michael Kahn da Universidade de Washington, Seattle, WA, USA (trabalho em andamento).

Tendo os cristais, os experimentos de difusão de inibidores potenciais foram iniciados imediatamente com os compostos disponíveis, os quais eram: Composto $C_5N_2O_5PH_{13}$ sintetizado por Dra. M. Willson, do grupo do Dr. J. Pierre, da Universidade de Toulouse, na França. Este composto como pode ser visto na Figura 4.7 é parte do composto D-11 modelado pelo Dr. C. Verlinde (Figura 4.4). Desta forma o composto sintetizado baseou-se no resultado da modelagem do Dr. Verlinde, seguindo portanto o ciclo de modelagem de compostos com

Figura 4.7 - Compostos sintetizados por: (a) Dra. Willson. (b), (c) e (d) por Jean Lee.

potencialidade para obtenção de compostos seletivos. O grupo do Dr. Fred Oppendoes determinou em duas medidas da constante de inibição o valor médio de 14 mM (comunicação pessoal). Este então foi o primeiro composto a ser utilizado por dois motivos: primeiramente por ter sido fornecido há muito tempo e devido ao fato da falta de cristais os experimentos não foram executados anteriormente e segundo por fazer parte do projeto de desenho racional de drogas. O grupo do Dr. Michael Kahn motivado pela síntese de compostos que pudessem inibir a TIM de *T. brucei* fez seus próprios modelos, usando como base o modelo da TIM de *T. brucei* na conformação fechada da estrutura em volta flexível. Este projeto paralelo gerou três compostos, dois deles derivado do triptofano, com constante de inibição bastante pobre, superior a 10 mM para o composto 69 (Figura 4.7 - medidas do IC₅₀ feitas em colaboração com Jean Lee). Estes compostos não seguiram uma metodologia na modelagem. Basicamente os autores deste trabalho pela simples análise visual do sítio ativo propuseram derivados de triptofano pela simplicidade da síntese e também por mostrar-se promissor em preencher o sítio ativo. Este tipo de modelagem normalmente não é o mais indicado para o sucesso pois não explora todas as possibilidades de análise do composto. Cada composto proposto tem que ser analisado em termos de energia de interação, interações via ligações de hidrogênio ou pontes salinas e interações hidrofóbicas. Este procedimento detalhado será o assunto do Capítulo 5 desta tese.

Figura 4.8 - Compostos sintetizados pelo Dr. Witmans da Universidade de Groningen

Finalmente também utilizou-se compostos recebidos pelo grupo do Dr. Cornelis J. Witmans, do Centro Universitário para a Farmácia, da Universidade de Groningen, Holanda, que

sintetizou fosfonatos ao invés de fosfatos. São compostos muito mais estáveis mas apresentam afinidades muito baixas para TIM de *T. brucei* como já apresentado na seção 4.1. Estes compostos podem ser vistos na Figura 4.8.

Estes compostos foram testados mas sabia-se que os resultados obtidos não introduziriam maiores informações das já obtidas nos estudos anteriores. Segundo o ciclo do desenho racional de drogas, o composto mais promissor era o sintetizado pela Dra. Willson, que seguia resultados de modelagem do Dr. Verlinde mas devido a disponibilidade dos compostos, os experimentos foram conduzidos com aqueles fornecidos pelos três grupos colaboradores. Apesar de solucionados os problemas com a cristalização, ainda assim o projeto apresenta suas dificuldades com o manuseio dos melhores cristais durante os experimentos de difusão.

4.4.1 Experimentos de Difusão do Composto $C_5N_2O_5PH_{13}$ na Forma Cristalina III.

Este composto foi sintetizado pela Dra. Willson, o qual pode ser visto na Figura 4.7a. Apesar da forma cristalina III apresentar apenas um monômero por unidade assimétrica e ter a estrutura em volta flexível na conformação aberta, inicialmente por ter sido a primeira forma cristalina a ser obtida em nossos experimentos de cristalização e ser uma forma cristalina livre do íon sulfato, esta foi então a forma cristalina usada. Uma coleta de dados da estrutura nativa foi então realizada e os resultados estão sumarizados na Tabela 4.2.

Verificando que os cristais difratam razoavelmente bem foi feita a difusão do composto $C_5N_2O_5PH_{13}$ (Figura 4.7), que consistiu em:

- Cristais foram transferidos para uma solução de 200 μ l contendo PEG 6000 22% p/v, 0,2 M Tris/HCl pH 8,0, 1 mM EDTA, azida sódica, DTT, 80 mM do composto da Dra. M. Willson por 18 horas.
- Os cristais do item anterior foram transferidos para uma solução de 200 μ l contendo PEG 6000 22% p/v, 0,2 M Tris/HCl pH 8,0, 1 mM EDTA, azida sódica, DTT, 200 mM do composto da Dra. M. Willson por 18 horas.

Três cristais foram montados e expostos aos raios-X. As imagens de difração foram coletadas mas não puderam ser integradas, pois havia clara característica de múltiplas difração duplicados em várias das imagens. Visualmente os cristais não sofreram modificações durante a difusão do composto $C_5N_2O_5PH_{13}$. Uma explicação razoável pode ser que a TIM de *T. brucei*

Tabela 4.2 - Coleta de dados de difração de um cristal nativo da forma cristalina III

Parâmetros de rede:	a=95.54 Å β=116.12°	b=48.25 Å	c=68.24 Å
Grupo espacial:	C2 (uma subunidade por unidade assimétrica.)		
Número de cristais expostos:	2		
Estabilidade aos raios X	2 dias		
Reflexões coletadas	18102		
Reflexões únicas	10616		
Completeza	86.4%		
Na região entre 2.5-2.3 Å	76.5 %		
R _{sym}	3.3%		

tenha sofrido mudanças conformacionais que afetaram a rede cristalina sem deteriorar os cristais. Os cristais submetidos à difusão difrataram até uma resolução razoavelmente próxima à obtida na coleta de dados do cristal nativo. Este resultado indica que esta forma cristalina não é adequada à experimentos de difusão. Como o composto testado apresenta uma constante de inibição relativamente baixa, apesar de algumas mudanças conformacionais na proteína, o cristal ainda foi preservado. Caso o composto apresentasse uma constante de inibição muito melhor (em torno de μM) muito provavelmente o cristal não suportaria as mudanças conformacionais e seria destruído por completo.

4.4.2 Experimentos de Difusão do Composto C₅N₂O₅PH₁₃ na Forma Cristalina II.

Como o experimento de difusão na forma cristalina III não mostrou resultados satisfatórios (seção 4.4.1) e com a subsequente obtenção de diferentes formas cristalinas com a preparação de proteína recombinante, foi então decidido repetir o experimento de difusão com o mesmo composto, C₅N₂O₅PH₁₃, mas com a nova forma cristalina.

Tabela 4.3 - Dados de difração coletados de um cristal da forma cristalina II na qual se difundiu o composto $C_3N_2O_5PH_{13}$.

Parâmetros de rede (em Å)	a=93,95 b=48,31 c=130,76 $\beta=99,81^\circ$
Grupo espacial	C2 (2 monômeros por unidade assimétrica)
Reflexões coletadas	46068 (até 2,4 Å)
Reflexões únicas	16676 (até 2,4 Å)
Completeza	80,3% (8,0-2,5 Å)
última camada de resolução	54,8% (2,6-2,5 Å)
R_{sym}	3,85%
R_{σ}	7,82% (2,6-2,5 Å)

O procedimento de difusão foi bastante semelhante ao descrito na seção anterior com uma diferença nos passos iniciais, já que a forma cristalina II foi crescida na presença de sulf. amônia. O cristal foi inicialmente transferido para uma solução com uma concentração de sulfato de amônia reduzida a metade (25 mM) no mesmo tampão no qual o cristal foi crescido em 44% PEG (p/v). Este cristal então foi transferido para soluções contendo o inibidor, inicialmente a 80 mM e no final a 200 mM. As trocas de solução foram feitas duas vezes para cada concentração, sendo que o volume de cada banho foi sempre em torno de 250 μ l. O cristal foi montado em capilares e os dados foram coletados no detector de área XENTRONICS Siemens/Nicolet X1000, usando radiação do gerador de anodo rotatório RU200 da Universidade de Washington, Seattle, USA. A estatística dos dados coletados pode ser vista na Tabela 4.3

Inicialmente a estrutura da TIM de *T. brucei* refinada por Noble (1992) no grupo espacial C2 foi utilizada para um ciclo de refinamento de corpo rígido, minimização de energia e refinamento dos fatores de temperatura. Neste primeiro estágio as duas estruturas em volta flexíveis foram omitidas. Na observação do mapa de densidade eletrônica ficou claro que a estrutura em volta flexível na subunidade 1 adota a conformação aberta o que permitiu sua inclusão com leves ajustes. Novo ciclo de refinamento permitiu a observação da estrutura em volta flexível da subunidade 2 também na conformação aberta. Devido à fraca densidade ainda neste estágio resolveu-se submeter o modelo a um ciclo de refinamento por dinâmica molecular com resfriamento lento, seguido por minimização de energia e refinamento dos fatores de temperatura (através do programa XPLOR, Brünger, 1990a) que resultou em um fator de concordância R final de 21,0%, no intervalo de resolução de 8,0 a 2,5 Å, com 15743 reflexões,

apresentando boa geometria (Tabela 4.4).

Tabela 4.4 - Estatísticas do refinamento da estrutura da forma cristalina II na qual se difundiu o composto $C_5N_2O_5PH_{13}$.

Fator R	21.0%
Resolução	8.0-2.5 Å
Reflexões	15743
Átomos de proteína incluídos	3700
desvio quadrático médio	
de ligações covalentes	0.012 Å
dos ângulos entre átomos ligados	1.65°
dos ângulos diédricos	23.51°
dos ângulos impróprios	1.46°

Os mapas de densidade eletrônica foram calculados através do pacote de programas BIOMOL (BIOSIN, Universidade de Groningen), com fatores de estrutura e fases calculados pelo programa XPLOR (Brünger, 1990a), escalonados pelo programa KBRANI (programa do pacote BIOMOL), fases ponderadas pelo programa SIGMAA (Read, 1986) e mapa calculado pelo programa FSFOR (programa do pacote BIOMOL). A análise do mapa foi feita através do programa O (Jones *et al.*, 1991).

No mapa de Fourier diferença no modelo final, constatou-se que não havia nenhuma densidade eletrônica na região onde se esperaria observar a ligação do composto (Figura 4.9).

Vários níveis de contorno foram utilizados na análise dos mapas, assim como mapas do tipo $2F_o - F_c$ e $F_o - F_c$. O grupo fosfato deveria apresentar uma densidade eletrônica apreciável e seria de fácil observação (como exemplo vide Capítulo 6 desta tese). Como pode ser visto na Figura 4.9 há duas regiões de densidade eletrônica esféricas próximas ao sítio de ligação do fosfato. Não há nenhuma densidade eletrônica significativa que se estenda destas regiões esféricas que pudesse se assemelhar ao composto $C_5N_2O_5PH_{13}$. Estas densidades eletrônicas esféricas que apresentam níveis de $4,0 \sigma$ podem ser moléculas de água bem localizadas ou um grupo fosfato com um fator de ocupação menor que 1. Como as duas opções foram já analisadas anteriormente não prosseguimos com esta estrutura.

Figura 4.9 - Mapa de densidade eletrônica do cristal da forma cristalina II no qual se difundiu o composto $C_5N_2O_3PH_{13}$. Mapa do tipo $2F_o - F_c$, e contorno de $1,5 \times rms$ do mapa.

4.4.3 Experimentos de Difusão do Composto $C_{13}N_2H_{14}PO_5$ na Forma Cristalina II

Como pode ser observado na Figura 4.7, foram sintetizados três compostos, dois deles derivados do triptofano, como resultado de uma modelagem manual baseada na análise da estrutura da TIM de *T. brucei* na conformação fechada. Esta modelagem não seguiu um procedimento mais detalhado de cálculos de energia e de interações hidrofóbicas entre o composto e a proteína. A idéia básica partiu da possibilidade de ser um inibidor com fácil síntese química. Os compostos sintetizados não apresentavam alto grau de pureza, o que para este estágio do trabalho não era importante. Por outro lado impurezas poderiam dificultar experimentos de co-cristalização.

Cristais da forma cristalina II foram transferidos como descrito na seção 4.4.2 para uma solução contendo 44% de PEG 6000 (p/v). O composto selecionado para a difusão (composto (b) da Figura 4.7) não é solúvel a esta concentração de PEG. Gradativamente testes foram feitos com cristais, transferindo-os para soluções contendo quantidades decrescentes em PEG. A

concentração mínima na qual o cristal ainda é estável foi de 25% de PEG (p/v) com a presença de 25 mM de sulfato de amônia. Esta foi então a condição inicial para a difusão. Ainda nesta concentração de PEG o composto não pôde ser solubilizado em concentrações maiores que 100 mM.

Testes preliminares de atividade enzimática da TIM de *T. brucei* e medidas da constante de inibição e IC₅₀ mostraram valores de no mínimo 10 mM (Delboni e Lee, dados não publicados). Desta forma 100 mM seria o suficiente para se obter um complexo.

Tabela 4.5 - Dados de difração coletados de um cristal da forma cristalina II no qual se difundiu o composto C₁₃N₂H₁₄PO₅.

Parâmetros de rede (em Å)	a=93.63 β=100.28°	b=47.99	c=130.01
Grupo espacial	C2 (2 monômeros por unidade assimétrica)		
Reflexões coletadas	20678 (até 2.4 Å)		
Reflexões únicas	14260 (até 2.4 Å)		
Completeza	73.3% (10.0-2.6 Å)		
última camada	53.5% (2.7-2.6 Å)		
R _{sym}	5.08%		
..R _{sym}	11.25% (2.7-2.6 Å)		

Devido à instabilidade do composto, adotou-se um procedimento de difusão por 6 horas, feito com transferências seguidas, a cada 90 minutos, para soluções novas, livres de sulfato de amônia. Quatro trocas foram feitas para certificar que a quantidade de sulfato de amônia havia decrescido a níveis insignificantes. O cristal foi montado em capilar e os dados foram coletados no detector de área XENTRONICS Siemens/Nicolet X1000, usando radiação de um gerador de anodo rotatório RU200 da Universidade de Washington, Seattle, USA. A estatística dos dados coletados pode ser vista na Tabela 4.5.

O refinamento seguiu o mesmo procedimento descrito na seção anterior, partindo-se do modelo refinado por Noble (1992). Iniciou-se por um refinamento de corpo rígido, seguido de refinamento posicional com minimização de energia e finalmente refinamento do fator de temperatura. Ao longo deste procedimento as estruturas em volta flexíveis foram omitidas. O mapa de densidade eletrônica foi calculado da mesma forma que a descrita na seção 4.4.1. A

estrutura em volta flexível da subunidade 1 foi facilmente identificada na conformação aberta e então adicionada. A estrutura em volta flexível da subunidade 2 não apresentou uma densidade eletrônica bem definida, provavelmente encontrando-se desordenada, devido provavelmente à fraca ocupação do composto no sítio ativo. Tentativas de observação da densidade eletrônica do composto no sítio ativo foram feitas através do cálculo de vários mapas com níveis variados e mapas $2F_o - F_c$ e $F_o - F_c$, sem resultados apreciáveis. O modelo foi refinado mantendo a estrutura em volta flexível omitida até um fator de concordância R final de 19,3%, com 12 961 reflexões no intervalo de 10,0 a 2,6 Å apresentando boa geometria (Tabela 4.6). A densidade eletrônica ao redor do sítio ativo pode ser vista na Figura 4.10.

Tabela 4.6 - Estatísticas do refinamento da estrutura da forma cristalina II na qual se difundiu o composto $C_{13}N_2H_{14}PO_5$.

Fator R	19,3%
Resolução	10.0-2.6 Å
Reflexões	12961
Átomos de proteína	3700
desvio da ligação covalente	0.012 Å
desvio dos ângulos dos átomos ligados	1.71°
desvio dos ângulos diédricos	23.88°
desvio dos ângulos impróprios	1.48°

Apesar de todas as dificuldades apresentadas neste experimento um resultado positivo foi observado: a desordem da estrutura em volta flexível da subunidade 2, a qual pode ser devida à ocupação parcial e talvez desordenada do inibidor, mostrando que este composto muito provavelmente deve se ligar ao sítio ativo. No entanto este composto apresenta grau de solubilidade muito limitada.

Provavelmente re-sintetizar o composto em forma de sal para obter um composto maior grau de solubilidade deve solucionar este problema, o que esta sendo buscado pelos químicos sintéticos ligados a este projeto (Jean Lee, comunicação pessoal).

Devido à dificuldade em solubilizar os compostos em PEG não foi possível utilizar este composto na forma cristalina I, a qual só pode ser transferida para concentrações superiores a

40% (p/v) de PEG 6000. Os outros dois compostos da Figura 4.7 foram testados mas não apresentaram solubilidade nestas soluções contendo 44% de PEG 6000 e nem mesmo solubilidade significativa a 25% (p/v) de PEG 6000, o que não motivou novos experimentos de difusão com estes compostos.

Figura 4.10 - Mapa de densidade eletrônica do cristal da forma cristalina II no qual se difundiu o composto $C_{13}N_2H_{14}PO_6$. Mapa do tipo $2F_o-F_c$ e contorno de 1.5 rms do mapa.

4.4.4 Experimentos de Difusão do Composto $C_4NH_{10}PO_5$ na Forma Cristalina II

O grupo do Dr. Cornelis J. Witmans, do Centro Universitário para a Farmácia, da Universidade Estadual de Groningen, Holanda, enviou uma série de compostos fosfonatos lá sintetizados com potenciais atividades inibitórias. Alguns destes compostos foram ensaiados e há informações das suas constantes de inibição ou IC_{50} (Figura 4.8). Como já descrito anteriormente a TIM de *T. brucei* apresenta baixa afinidade aos fosfonatos quando comparados aos fosfatos. Por ainda não serem claras as causas destas diferenças, decidiu-se escolher dentre os compostos um que apresentasse além de uma constante de inibição razoável (ao redor de

10 mM), estabilidade e interesse quanto a novas informações da interação do composto com a proteína. O composto (5) tem um IC₅₀ de 13 mM e apresenta uma ramificação metil no carbono 1 (Figura 4.8), a qual pode introduzir novas informações de interações com a proteína. Os fosfonatos são frequentemente compostos quimicamente estáveis, o que não acontece com o composto (7). Os compostos (1) e (2) apresentam baixa solubilidade e os outros não introduzem maiores informações que as já obtidas nos complexos apresentados nas seções anteriores (Figura 4.8).

Tabela 4.7 - Dados de difração coletados da forma cristalina II na qual se difundiu o composto C₄NH₁₀PO₃.

Parâmetros de rede (em Å)	a=94,52 β=100,38°	b=48,18	c=130,44
Grupo espacial	C2 (2 monômeros por unidade assimétrica)		
Reflexões coletadas	52828 (até 2,3 Å)		
Reflexões únicas	22809 (até 2,3 Å)		
Completeza	93% (8,0-2,3 Å)		
última camada	84% (2,5-2,3 Å)		
R _{sym}	4,6%		
R _{sym}	7,8% (2,5-2,3 Å)		

O procedimento de difusão foi o mesmo descrito nas seções anteriores. O composto (5) apresenta alta solubilidade o que permitiu realizar o experimento de difusão em soluções contendo 44% de PEG 6000 (p/v). Devido à estabilidade do composto e sua solubilidade, experimentos de difusão mais longos e com maiores concentrações foram feitos. Os cristais foram transferidos para soluções contendo o potencial inibidor e mantidos por 48 horas, com 4 trocas de soluções, com concentrações crescentes do composto (5), de 50 mM até 200 mM, sendo que os cristais sobreviveram bem a este processo. Os cristais foram montados em capilares e os dados foram coletados no detector de área XENTRONICS Siemens/Nicolet X1000, usando radiação do gerador de anodo rotatório RU200 da Universidade de Washington, Seattle, USA. A estatística dos dados coletados estão resumidas na Tabela 4.7.

O refinamento foi iniciado de forma similar à descrita nas seções anteriores, utilizando o modelo de Noble (1992) da TIM de *T. brucei*, com um ciclo de refinamento por corpo rígido, minimização de energia e refinamento do fator de temperatura, com as duas estruturas em volta

flexível omitidas. Este refinamento não se comportou adequadamente na medida em que a análise dos mapas de diferença calculados da mesma forma descrita nas seções anteriores não mostraram densidade eletrônica para as duas estruturas em volta flexível, em ambas as subunidades.

Tabela 4.8 - Estatísticas do refinamento da estrutura da forma cristalina II na qual se difundiu o composto $C_4NH_{10}PO_4$.

Fator R	27,4%
Resolução	8.0-2.4 Å
Reflexões	18626
Átomos de proteína	3634
desvio da ligação covalente	0.014 Å
desvio dos ângulos dos átomos ligados	1.8°
desvio dos ângulos diédricos	24,0°
desvio dos ângulos impróprios	1,6°

Não ficou clara a razão da ausência de densidade eletrônica das estruturas em volta da subunidade 1 na medida em que esta encontra-se fixa por interações com um outro monômero relacionado por simetria cristalográfica. Provavelmente a interação deste composto tenha motivado mudanças conformacionais nas estruturas em volta flexível de ambas subunidades, as quais, apesar de desordenadas, não foram suficientes para a desintegração da ordem cristalina. No entanto a qualidade do modelo sofreu piora como pode ser observado pelas estatísticas do refinamento na Tabela 4.8. Através de uma análise óptica mais detalhada dos cristais, foram visíveis pequenas trincas, mas os dados cristalográficos apresentaram boa qualidade como pode ser observado na Tabela 4.7.

Refinamento por dinâmica molecular de resfriamento lento foi tentado, não apresentando resultados satisfatórios. O modelo atual apresenta um fator R muito elevado (27,4 % no intervalo de 8,0 a 2,4 Å). A análise não mostrou nenhum pico de densidade eletrônica no sítio de ligação do composto fosfonato. Neste estágio do refinamento não se pode afirmar com certeza que o composto fosfonato não se ligou à proteína, pois mapas de Fourier diferença em estruturas distantes da correta podem não ser conclusivos. Novas tentativas de refinamento foram feitas sem

sucesso satisfatórios com convergência para modelos mais aceitáveis. As estatísticas do refinamento podem ser vistas na Tabela 4.8.

4.4.5 Experimentos de Difusão do Composto $C_4NH_{10}PO_5$ na Forma Cristalina I

A forma cristalina I tem demonstrado os resultados mais satisfatórios para os experimentos de difusão, apesar de apresentar a maior dificuldade em manipulação, pois a transferência da condição de cristalização em alta concentração de sulfato de amônia (2,4 M) para soluções em PEG e livres do íon sulfato é bastante difícil. O processo de transferência descrito por Schreuder *et al.* (1988) foi utilizado. Vários cristais não resistiram mas a: sobreviveram apesar de apresentar trincas. Durante a transferência ficou claro a dependência com o pH, ou seja, os cristais deviam ser transferidos para soluções com exatamente o mesmo pH no qual foram crescidos.

Os cristais foram transferidos da condição de cristalização para uma gota contendo o mesmo tampão, substituindo o sulfato de amônia por 44% PEG 6000 (p/v) mais 100 mM de sulfato de amônia. Esta transferência inicial é feita para que o cristal possa passar para o novo tampão sem remover o sulfato do sítio ativo, o que fatalmente destruiria o cristal. A concentração de PEG pode ser ajustada para cada caso mas deve ser em torno de 44% p/v pois esta é a concentração de PEG que apresenta a mesma pressão de vapor que a de 2,4 M de sulfato de amônia, equilibrando desta forma a quantidade de água existente no cristal. Diferenças nestas pressões podem resultar em possíveis trincas ou mesmo a destruição do cristal. Vários cristais foram transferidos e o processo de otimização da condição destruiu muitos destes. Em alguns casos, depois de otimizada a condição com cristais pequenos, o processo de transferência não apresentava os mesmos resultados para cristais maiores, crescidos na mesma condição dos cristais menores.

Para os cristais que resistiram à transferência de solução descrita no item 1, experimentos visando substituir o sulfato de amônia pelo inibidor foram feitos em passos, diminuindo a concentração do sulfato de amônia e aumentando a concentração do inibidor. Em muitos casos o intervalo entre cada passo foi importante para a estabilidade do cristal. Cristais que resistiram a troca de inibidor foram usados para coleta de dados. No entanto, o limite máximo de resolução em alguns casos diminuiu bastante, sendo no máximo ao redor de 3,0 Å, resultando em imagens não processáveis devido à mudança no padrão de espalhamento, provavelmente afetando a

integridade dos cristais. Experimentos de difusão mais rápidos produziram boa qualidade de difração, bons dados mas após sua análise mostraram a presença do íon sulfato no sítio de ligação do fosfato, não indicando nenhuma densidade eletrônica que pudesse justificar a presença do inibidor, mesmo que parcial. Isto pode ser devido ao intervalo reduzido do experimento de difusão.

A troca de inibidores durante o experimento de difusão deve ser mais lenta, provavelmente devido ao íon sulfato apresentar uma constante de afinidade, no cristal, aparentemente aumentada. Isto talvez se deva às barreiras das mudanças conformacionais dentro do cristal para que possa permitir a saída do íon sulfato e a entrada do inibidor, que provavelmente devem ter suas cinéticas diminuídas. O decréscimo da concentração do íon sulfato para níveis abaixo da constante de inibição deve ser observado. No entanto deve-se levar em conta que a constante de inibição na forma cristalina pode ser menor que a observada em solução, a qual pode estar relacionada com as observações feitas acima. Também o tempo de difusão deve ser ajustado experimentalmente para se certificar que a troca tenha ocorrido.

4.5 Conclusão

Este projeto teve início em 1984 com a cristalização da triosefosfato isomerase de *T. brucei* (Wierenga *et. al.*, 1984). Anteriormente o grupo do Dr. Fred Oppenheimer fez estudos para identificação dos processos metabólicos mais importantes do tripanossomo. A estrutura foi determinada e refinada a alta resolução com dados coletados em sincrotron (Wierenga *et. al.*, 1991b). Neste estágio o projeto de desenho de drogas encontrava-se apenas no início. Dr. Verlinde então iniciou a modelagem em computador de um composto que apresentasse seletividade e inibição combinadas, a qual resultou na molécula DG-11 (Verlinde *et. al.*, 1992a). Durante o mesmo período projeto de determinação da estrutura de complexos de inibidores análogos ao substrato, derivados de fosfatos e fosfonatos, foram feitos como descrito neste Capítulo, nas seções anteriores. Mesmo com um grande número de complexos não só com a TIM de *T. brucei*, mas com várias outras (levedura, Lolis *et. al.*, 1990; humana, Mande *et. al.*, 1994; *Bacillus stearothermophilus*, Delboni *et. al.*, submetido) o entendimento de detalhes como a diferença de afinidade para fosfato e fosfonato ainda não foi obtido. Vários autores tem estudado não somente TIM mas a estrutura em barril que aparece em várias outras proteínas (Tabela 1.1). Portanto apesar desta proteína e o motivo estrutural ser altamente estudado, a obtenção de

resultados positivos, como a obtenção de inibidores que apresentem alguma seletividade ainda não foi conseguida. Isto mostra como em particular esta proteína não é a ideal para um trabalho de desenvolvimento de droga seletiva. A homologia estrutural é alta o que demonstrada no único tripeptídeo que apresenta diferença a aproximadamente 19 Å do sítio de ligação do fosfato. A grande maioria dos complexos obtidos foram com a volta flexível na conformação fechada e a modelagem tem sido concentrada nesta conformação da TIM, sendo que apenas a subunidade 2 tem sido usada já que na subunidade 1 a volta flexível encontra-se na conformação aberta devido a contatos cristalinos. Muito recentemente a estrutura de um complexo que apresenta a volta flexível na conformação aberta (Verlinde *et. al.*, 1992b) possibilitou novo enfoque na modelagem de inibidores (Capítulo 5 desta tese). Portanto apesar de ter-se um grande número de complexos o entendimento dos modos de ligação e das características do sítio ativo ainda não estão completamente entendidas. Novos complexos com diferentes compostos, sinte partir de informações de análise dos complexos já obtidos são extremamente necessários para progredir em direção ao entendimento das interações no sítio ativo para poder aumentar portanto a constante de afinidade, antes mesmo de obter seletividade.

Os resultados apresentados aqui fazem parte deste amplo projeto que encontrou suas dificuldades após a determinação da estrutura tridimensional da TIM de *T. brucei*, a qual foi conseguida com bastante facilidade, com apenas 4 mg de proteína extraídas diretamente de *T. brucei* (Wierenga *et al.*, 1984). As dificuldades surgidas com a proteína expressa em grandes quantidades, apesar de produzir quantidades expressivas em 1 litro de cultura (de 70 a 100 mg) estiveram concentradas na obtenção de cristais que pudessem difratar a alta resolução estudo de complexos com inibidores. O problema da idade da proteína foi deste projeto até o momento. Este problema foi resolvido com os experimentos de expressão sendo realizados imediatamente antes da cristalização. A conclusão sobre o problema que nós chamamos de envelhecimento custou longos meses de trabalhos e três anos ao grupo do Dr. Wim Hol para voltar a conseguir cristais com qualidades de difração além de 2,0 Å de resolução.

Como os resultados apresentados anteriormente, muitos complexos foram conseguidos sendo que alguns descritos não obtiveram sucesso. Dentre os complexos obtidos muitos não introduziram informações que pudessem viabilizar o estágio seguinte do projeto de desenvolvimento racional de drogas. Todos os inibidores estudados se ligam ao sítio do fosfato mas não se estendem em direção ao sítio de seletividade. Nenhum dos complexos pode introduzir novas informações para então se desenvolver novos compostos, através de análises das estruturas ou através do uso de programas como o LUDI (Böhm, 1992) que podem ser utilizados no modo LINK, no qual diferentes compostos ajustados em sítios próximos são ligados por fragmentos

para gerar compostos maiores que possam ser posteriormente sintetizados. Também todos os complexos obtidos até o momento são muito similares ao substrato, sendo derivados de grupos fosfatos ou fosfonatos, que não são ideais para suportar modificações químicas para estendê-los ao sítio seletivo. Novas buscas de compostos alternativos devem ser realizadas para impulsionar o projeto em direções mais otimistas.

A grande contribuição introduzida neste projeto foi a retomada com a obtenção de cristais de TIM de *T. brucei* expressa em grandes quantidades em *E. coli*, com boa qualidade de difração até abaixo de 2,0 Å de resolução, podendo assim permitir a continuação do trabalho de complexos e testes com possíveis inibidores modelados. Apesar de ter sido tentado a obtenção de complexos com possíveis inibidores (apesar de apresentarem pobre constante de inibição) não foi possível observar nenhum resultado positivo. Também contribuição importante foi a reprodutibilidade das transferências de cristais de soluções completamente diferentes (sulfato de amônia para PEG) a qual permite a difusão de inibidores sem a presença do inibidor competitivo natural, o íon sulfato. A grande vantagem introduzida foi a facilidade da obtenção de cristais. Em uma semana pode-se expressar a proteína, purificar e cristalizar. Em no máximo um mês, devido ao alto rendimento da expressão, centena ou até milhares de cristais podem ser obtidos. Isto é necessário para um projeto que visa a constante uso de cristais para os experimentos de difusão de novos inibidores.

O tempo de treinamento no grupo do Dr. Wim Hol, portanto, não foi suficiente para obtenção de um complexo. Apesar dos resultados científicos publicáveis não terem sido obtidos, os resultados observados durante todo o processo foram, de forma apreciável, parte da trajetória da formação científica do autor deste trabalho.

O projeto, que agora encontra-se em fase de produzir resultados publicáveis, está, no momento, sendo conduzido por um aluno de doutorado, Brad Bernstein, no grupo do Dr. Wim Hol, o qual muito provavelmente deve obter resultados a curto prazo, com novos complexos que possam introduzir novas informações dentro do projeto de desenho racional de drogas para a doença do sono.

Capítulo 5

Desenho Racional de Drogas contra a Doença do Sono: Procura por Novos Inibidores da TIM de *T. brucei*, na Conformação Aberta da Estrutura em Volta Flexível

5.1 Introdução

A bioquímica dos processos biológicos tem sido o alvo principal da cristalografia de proteínas. Com a elucidação da estrutura tridimensional de proteínas, os processos bioquímicos podem ser interpretados muitas vezes com o auxílio de outros resultados vindos de outras técnicas ou mesmo fornecidos pela cristalografia através de complexos da proteína com análogos dos substratos.

Mecanismos de muitas doenças puderam ser entendidos através da determinação da estrutura de proteínas a elas relacionadas. Estas informações estruturais acumuladas nas últimas décadas marca o início para uma meta inteiramente nova: o desenho racional de drogas baseado na estrutura tridimensional de proteínas. Este campo é atualmente bastante popular, tendo sido assunto de um livro por um dos maiores cristalógrafos de proteínas, Dr. M. Perutz (Perutz, 1992) como também várias revisões (White e Blundell, 1994, Verlinde e Hol, 1994, Greer *et al.*, 1994).

No caso do desenho racional de drogas baseado em estruturas macromoleculares, o passo mais difícil encontra-se em achar uma molécula pequena que seja inibidora da proteína usada como alvo. Este inibidor é chamado de guia ("lead") por apresentar algumas propriedades desejadas de uma eventual droga mas ainda bastante distante. Uma otimização do composto guia é subsequentemente requerida para prover a molécula inicial com suficiente grau de afinidade, seletividade, baixa toxicidade, disponibilidade biológica, estabilidade, solubilidade e mais alguns outros fatores importantes para uma droga.

Na busca do composto guia duas técnicas estão sendo aprimoradas rapidamente. A primeira delas usa imensos coquetéis de bibliotecas moleculares em solução. Através de ensaios enzimáticos destes coquetéis consegue-se isolar a pequena molécula mais ativa que pode então ser melhorada quanto às suas propriedades. Este processo é totalmente independente do conhecimento da estrutura tridimensional da proteína.

A segunda técnica é totalmente baseada na estrutura tridimensional da proteína de interesse. Inicia-se a partir da estrutura tridimensional da proteína e por observação direta ou através do uso de sofisticados programas computacionais tenta-se projetar 'ab initio' inibidores da proteína de interesse.

Entre os programas disponíveis encontra-se o LUDI (Böhm, 1992) que descreve o sítio ativo da proteína de interesse em termos de doadores e aceptadores de ligações de hidrogênio e superfícies hidrofóbicas. Estas descrições são então comparadas com as descrições das moléculas pequenas contidas em bancos de dados no intuito de encontrar moléculas que apresentem a maior

similaridade da descrição do sítio de interesse da proteína. A pontuação para cada molécula do banco de dados basicamente usa o número de interações favoráveis, ou seja, o número de descrições que são favoráveis entre a molécula pequena do banco de dados e a da proteína. Descrições aqui representam toda a estrutura do sítio ativo, seus grupos funcionais capazes de participar em ligações de hidrogênio como doadores e aceptadores e interações hidrofóbicas, assim como grupos funcionais dos compostos pertencentes aos bancos de dados de pequenas moléculas.

Este processo de pontuação ainda apresenta algumas desvantagens pois para o caso deste programa, não existe penalizações para descrições não satisfeitas pela proteína, como um aceptador ou doador de ligação de hidrogênio sem um par na molécula encontrada pelo programa.

Outro programa, que já apresentou resultados positivos para algumas proteínas (Shoichet *et al.*, 1993 , para uma revisão Kuntz, 1992), é o DOCK (Shoichet *et al.*, 1992). O primeiro passo do programa DOCK é essencialmente um encaixe entre a molécula pequena e a proteína baseado na complementariedade estérica. O programa primeiramente calcula a superfície da molécula de interesse e depois através de uma busca em bancos de dados tenta encontrar a molécula que melhor se encaixa no sítio da proteína. A pontuação para este programa é feita baseado na complementariedade estérica assim como na interação eletrostática da molécula com a proteína.

O próximo passo é a otimização do composto guia obtido. Aqui o conhecimento da estrutura da proteína de interesse em complexo com o composto guia é de grande importância. Em vários laboratórios no mundo esta estratégia tem sido seguida mostrando ganhos significativos em afinidade do ligante para a proteína de interesse, apesar de muitos resultados não serem reportados por indústrias farmacêuticas por razões estratégicas. A chave para o sucesso deste trabalho é a integração de várias disciplinas no chamado "ciclo do desenho de drogas" que foi descrito pela primeira vez por Hol (Hol, 1986) e que pode ser visto na Figura 5.1.

5.1. Biologia molecular, farmacologia, química médica, biologia estrutural e modelagem molecular têm que trabalhar em perfeita harmonia para chegar às moléculas pequenas que podem contribuir para o controle de doenças infecciosas, do câncer, ou de problemas por mutações que levem ao mal funcionamento de um processo biológico.

Também é óbvio que moléculas que se ligam muito bem às proteínas alvo não necessariamente podem ser desenvolvidas em drogas efetivas e seguras. As drogas não somente têm que atingir a molécula de interesse no organismo, como também devem atravessar muitas vezes várias membranas, não podem ser rapidamente digeridas ou processadas pelo organismo humano, não podem apresentar muitos efeitos colaterais, devem ser sintetizadas facilmente com grande pureza e em larga quantidade, não podem ser transformadas em compostos inócuos pelo organismo.

Figura 5.1 - Desenho esquemático do "ciclo de desenho racional de drogas"

Existem vários tipos de doenças as quais devem ser abordadas com procedimentos particulares no desenho racional de drogas. Existem as doenças causadas por organismos invasores do organismo humano como no caso bastante conhecido das doenças tropicais (malária, doença do sono, doença de Chagas, leishmaniose entre outras) que atingem milhões de pessoas e animais em todos os países tropicais, em alguns casos atingindo os países que se encontram fora desta região geográfica devido à transmissão por outros meios como no caso de transfusão de sangue contaminado (Eleventh Programme report of the UNDP/World Bank/WHO).

Para que se possa então desenvolver uma droga, primeiramente deve-se encontrar o processo bioquímico que se deseja afetar no organismo invasor. Em alguns casos pode-se encontrar proteínas que são essenciais para o organismo invasor e que o organismo invadido não apresente uma proteína homóloga. Neste caso o desenho racional da droga torna-se em princípio mais simples. No caso do processo bioquímico a ser atacado ser também encontrado no organismo invadido o problema se complica pois neste caso é necessário que o ataque do organismo causador da doença seja feito através de uma droga que apresente seletividade a ponto de afetar apenas a bioquímica deste organismo, deixando a bioquímica do hospedeiro inalterada ou apenas levemente afetada. Desta forma, caso a droga não seja suficiente para eliminar o organismo causador da doença, pode dar condições para o organismo afetado, através do sistema imunológico, de reagir a ponto de eliminar o organismo causador da doença. Este é o caminho que se tenta seguir no caso da doença do sono.

A doença do sono é causada por um pequeno número de subespécies do protozoário hemoflagelado *Trypanosoma brucei*, o qual é transmitido de hospedeiro para hospedeiro pela picada da mosca tsé-tsé. De acordo com a Organização Mundial da Saúde (OMS) 50 milhões de pessoas encontram-se em risco de serem contaminadas na África, sendo que ao redor de 15.000-20.000 casos são reportados por ano, mas o número real deve chegar ao redor de 200.000-300.000 por ano. A doença do sono é responsável por dezenas de milhares de mortes anualmente (Eleventh Programme report of the UNDP/World Bank/WHO).

No estágio inicial da doença o tripanossomo vive livremente no sangue do hospedeiro e no último estágio ultrapassa a barreira sanguínea do cérebro, resultando em coma e, se não tratado, na morte.

As drogas utilizadas atualmente estão muito longe de serem perfeitas. Muitas destas apresentam alta toxicidade e tem numerosos efeitos colaterais. Pantamidina e suramina são somente úteis nos estágios iniciais da infecção (Eleventh Programme report of the UNDP/World Bank/WHO). Melarsoprol é usada no estágio mais avançado mas exibe efeitos colaterais que podem ser fatais. A mais recente droga introduzida, eflornitina, apresenta também problemas (i) sendo menos efetiva contra a forma mais virulenta do parasita, que é a rhodesiense, (ii) resistência já foi reportada e (iii) o tratamento requer a administração intra venosa de quantidades altas e o paciente deve estar sob cuidados hospitalares. É clara a necessidade de drogas mais eficientes e seguras.

O tripanossomo exibe várias características não usuais que podem ser exploradas do ponto de vista do desenho de drogas. *T. brucei* possui uma organela chamada glicosomo (Borst e Borst, 1977) na qual ficam confinadas as enzimas da glicólise. Nesta organela a glicólise e a forma do tripanossoma que se encontra no sangue possui uma atividade 50 vezes maior que as taxas apresentadas nos eritrócitos humanos (Wang, 1984, Figura 1.1). Foi demonstrado que bloqueando a glicólise do parasita resulta em morte tanto *in vitro* (Clarkson e Brohn, 1976) como *in vivo* (Oppenheimer e Borst, 1977).

Outra espécie de tripanossomo, como *T. cruzi* e a espécie próxima, a *Leishmania mexicana*, são responsáveis por outras doenças tropicais. *T. cruzi* é o agente causador da doença de Chagas, 10% dos casos sendo fatais, geralmente como resultado de falhas cardíacas e inflamação severa do trato intestinal (Mengel e Rossi, 1992). Esta doença afeta 10 a 20 milhões de pessoas e encontra-se principalmente distribuída na América Central e do Sul, mas casos foram reportados em Maryland nos Estados Unidos (Kirchhoff, 1993, Mengel e Rossi, 1992, 1983). Não existe atualmente nenhuma droga disponível para o tratamento definitivo da doença (Eleventh Programme report of the UNDP/World Bank/WHO).

As várias espécies de leishmania que pertencem também a classe tripanosomatidae causa várias formas de leishmaniases. A Organização Mundial da Saúde estima que 350 milhões de pessoas em todo o mundo estão em risco de serem contaminadas e 12 milhões de pessoas encontram-se infectadas (Eleventh Programme report of the UNDP/World Bank/WHO). A forma mais comum é a *L. mexicana*. Este parasita encontra-se distribuído na grande maioria dos países sul-americanos, parte da América Central e no sul dos Estados Unidos.

Ainda que *T. cruzi* e *L. mexicana* são menos dependentes da glicólise para a geração de energia do que *T. brucei* (Brener, 1973), é esperado que a perturbação de um processo bioquímico tão vital como este, o organismo, que vive num meio hostil, não consiga sobreviver às constantes reações do sistema imunológico. Portanto, as enzimas glicolíticas neste organismos podem ser um alvo promissor para o desenho de drogas.

Esforços para o desenvolvimento de grupos de pesquisas em desenho racional de drogas em países em desenvolvimento como o Brasil e os países da America Latina é de extrema importância pois estas doenças afetam principalmente estes países que são na grande maioria pobres, os quais são de certa forma ignorados pelas grandes indústrias farmacêuticas.

5.2 Descrição do Método de Desenho Racional de Drogas

Como descrito no Capítulo 4 vários complexos da TIM de *T. brucei* foram obtidos (2-fosfoglicerato, 3-fosfoglicerato, glicerol-3-fosfato, 3 fosfonopropionato, mono hidrogênio fosfato, 2-(N-formil-n-hidroxiamino)etil-fosfonato) no qual a estrutura em volta flexível encontra-se na conformação fechada. Esta conformação foi utilizada e ainda encontra-se em análise pelo Dr. C. Verlinde na modelagem de compostos guia através da análise do sítio ativo e suas redondezas na procura de um composto que se estenda desde o sítio ativo até a região que apresenta diferenças possíveis de serem usadas como sítio seletivo entre a TIM humana e de *T. brucei*. Esta região encontra-se ao redor de 15 Å do sítio de ligação do fosfato, formado pelos resíduos Ala 100, Tyr 101 e Tyr 102 que são His-Val-Phe na enzima humana (Mande *et al.*, 1994). Como resultado de tal modelagem o composto DG-11 (Figura 4.4) foi proposto, o qual foi descrito no Capítulo 4 e parte deste composto já foi sintetizada e medidas da constante de inibição assim como experimentos de difusão foram realizados (Capítulo 4 desta tese). Também uma busca através do bando de dados de moléculas pequenas foi feita utilizando o programa LUDI (Böhm, 1992) o qual não forneceu nenhum resultado satisfatório (comunicação pessoal

com Dr. C. Verlinde). Atualmente o grupo do Dr. Hol está em estreita colaboração com o grupo do Dr. M. Kahn, também na Universidade de Washington, em Seattle, na busca de novos compostos que venham a ser inibidores da conformação fechada da estrutura em volta flexível.

Com a determinação da estrutura tridimensional do complexo da TIM de *T. brucei* com o composto N-hidroxi-fosfono-butanamida (Verlinde *et al.*, 1992b) abriu-se um novo ponto de análise para a modelagem de compostos inibidores desta enzima. Este composto é um (e em alguns casos dois) átomo de carbono mais longo que os compostos discutidos no capítulo 4 (Figura 4.3) e se liga à conformação aberta da estrutura em volta flexível (Verlinde *et al.*, 1992b). Desta forma é possível analisar a estrutura em volta flexível na conformação aberta para a modelagem de compostos que venham a se estender até o sítio seletivo da TIM de *T. brucei* (Mande *et al.*, 1994) usando como base a estrutura deste complexo determinada por Verlinde *et al.* (1992b).

Portanto, em vistas dos resultados já reportados do sucesso do programa DOCK (Shoichet *et al.*, 1993), decidiu-se pela utilização deste programa na busca de compostos que possam servir como guia para futuros testes de inibição e possíveis modificações que venham a direcionar o composto para apresentar uma seletividade de afinidade à TIM de *T. brucei*.

5.3 Descrição do Programa DOCK

Aqui será descrito a metodologia utilizada para o encaixe de moléculas através do programa DOCK (Shoichet *et al.*, 1992). Este programa é dividido em partes. Primeiramente determina-se descrições geométricas das reentrâncias e protuberâncias na superfície tanto da molécula receptora quanto da molécula que será ligada. Estas descrições geométricas são utilizadas para guiar a busca de orientações no espaço, sendo que o objetivo é encaixar uma protuberância de uma molécula em uma reentrância de outra. O programa olha para os pontos de esferas da primeira molécula que têm a mesma distância interna, dentro de uma certa tolerância, entre seus centros e os centros dos conjuntos de esferas da segunda molécula. Com base nos conjuntos que obedecem esta distância de coincidência ("match") . Estas coincidências ("matches") são então usadas para definir matrizes de rotação e translação que então levam a segunda molécula na região complementar da primeira. Uma vez encontrada uma orientação esta é submetida a um exame preliminar rápido das complementariedades baseado-se na simples análise dos contatos entre a molécula receptora e a molécula que é ligada. Também uma análise

das interações eletrostáticas é feita entre as duas moléculas. Estas duas análises definem uma pontuação para o encaixe de cada molécula no receptor. Orientações que colocam partes da molécula sobre regiões definidas como sendo do receptor são descartadas. As configurações que passam pelo filtro baseado no volume e que possuem bons contatos e boa interação eletrostática são mantidas para futuras análises.

5.3.1 Descrição Molecular: Esferas

O receptor é descrito geometricamente por esferas complementares às reentrâncias e protuberâncias da superfície molecular, gerando portando uma imagem negativa da superfície do receptor. Os centros destas esferas podem ser pensados como pseudo-átomos. Eles são usados como uma malha irregular para mapear o ligante no sítio de ligação. As esferas que definem a superfície do receptor são de diferentes tamanhos e algumas se superpõem. Estas moléculas que se superpõem são agrupadas em um conjunto ("cluster"). A superfície de um receptor normalmente tem dezenas de conjuntos ("clusters"), cada um dos quais definem um potencial sítio de interação. As moléculas ligantes são descritas da mesma forma sendo que os centros encontram-se localizados dentro da superfície. Estas esferas são geradas pelo programa SPHGEN (Kuntz *et al.*, 1982) que usa a superfície da proteína gerada pelo programa MS (Connolly, 1983).

Os centros das esferas podem ser especificados pelo programa para serem localizados dentro da superfície, quando o cálculo é feito para o ligante e fora da superfície, quando o cálculo é feito para o receptor. Também o usuário pode especificar um subgrupo de pontos da superfície ou todos os pontos, pode colocar limites para evitar gerar esferas muito grandes, esferas muito próximas e um limite mínimo para o raio das esferas.

5.3.2 O Problema da Complementariedade: Gráfico Bifurcado

O problema do encaixe pode ser colocado como um problema gráfico teórico. Se um ligante tem N_l descrições (descrição aqui é malha de complementariedade do sítio alvo, definido pela superfície obtida segundo descrito na seção anterior) e o receptor N_r , então o número de nós no gráfico de encaixe D é $N_l \times N_r$. Uma aresta existe entre dois vértices compostos por descrições

$(N_l)_i, (N_r)_i$ e $(N_l)_j, (N_r)_j$ (onde i e j são descrições do ligante e receptor) do ligante e do receptor, quando a distância $N_l(i,j)$ é a mesma que a $N_r(i,j)$, dentro de alguma tolerância. Uma complementação mínima entre o ligante e o receptor ocorre quando existe um subconjunto do gráfico D que é completamente conectado por arestas e que tem ao menos quatro vértices (e portanto seis arestas) (Figura 5.2).

Figura 5.2 - Distância interna de complementação. Distâncias internas do ligante e do receptor são comparadas. Se a distância interna não coincide em um dado vértice, a árvore de procura é interrompida neste vértice.

Quatro vértices devem ser especificados no mínimo para determinar uma matriz de rotação e translação que preserve informações quirais do composto a ser orientado no sítio alvo. É fisicamente impossível complementar, simultaneamente, a maioria das características do ligante e do receptor, ou seja, D é esparsamente conectado, o qual guia o método de busca para cortes no espaço de orientações. Isto quer dizer que um gráfico que falha no exame da distância em M vértices, ou seja, que não se encontra completamente conectado devido a adição de M vértices, falhará em todos os números maiores que M , portanto a busca pode ser interrompida em M . Este procedimento reduz drasticamente o número de vértices necessários a ser considerados. O custo de busca também é reduzido através da ponderação de interações de longo alcance mais que as de curto alcance, ou seja, informações locais.

Para controlar a busca do espaço de orientações, organiza-se as esferas do receptor baseando-se na distância interna entre pares de centro de esferas. Inicia-se com cada centro de

esfera, todos os outros centros de esfera dentro do conjunto são ordenados em camadas baseados nas suas distâncias ao centro da esfera inicial. Todas as distâncias dentro de um certo intervalo são agrupadas numa mesma camada. Estas camadas são passíveis de ajuste pelo usuário. Quanto maior o intervalo, maior o número de pontos dentro deste e menor o número de intervalos. Estes intervalos são contruídos para cada esfera receptora.

As camadas para o ligante são feitas da mesma forma. As camadas do ligante e do receptor para cada par de pontos iniciais são comparadas baseando-se nos intervalos de distâncias dos pontos dentro delas. Somente aqueles pontos nas camadas com intervalos similares serão usadas para gerar o gráfico. As primeiras N-1 camadas de receptores e ligantes (aquelas representando as distâncias mais longas) que coincidem dentro de uma certa tolerância são usadas para gerar o gráfico. Características de uma determinada camada são testadas em apenas um estágio na geração do gráfico; então, a característica da segunda camada sempre fornecerá o terceiro nó em um gráfico, o par original de pontos definindo o primeiro par de nós. Todos os centros são finalmente testados como pontos iniciais e todos os centros dentro de uma camada de mais longa distância são testadas na geração do gráfico de distâncias que coincidem.

Por causa do método da mais longa distância descrito acima, nem todas as camadas são testadas e portanto o método não é exaustivo. Com o cuidado deste método, no entanto, o processo é independente do caminho. O número de pontos em cada camada determina quantas comparações serão testadas. Em geral, quanto maior a camada maior o número de orientações geradas. A extensão da busca encontra-se sob controle do usuário.

A construção do gráfico é feita por algoritmos que podem ser escolhidos entre três que são: "Fan", "Cat's cradle" e centro de massa os quais são descritos no artigo que apresenta o programa DOCK (Shoichet *et al.*, 1992), que não é o objetivo deste trabalho.

5.3.3 Deteminação da Pontuação

A versão mais recente do programa DOCK (versão 3.5) apresenta dois tipos de pontuações: pontuação de contatos e energia de interações eletrostáticas ou energia de mecânica molecular. Inicialmente o programa apresentava apenas a pontuação por contato, nas versões até 2.0. A introdução da pontuação por interações eletrostáticas (que foi implementada a partir da versão 3.0) permite uma análise mais ampla dos resultados. Pelo fato dos dois sistemas de pontuação não apresentarem as mesmas unidades não é possível obter um resultado através de

uma pontuação combinada entre as duas. O programa apresenta duas listas com as melhores pontuações de contato e de interação eletrostática. Nem todos os compostos apresentados numa lista estão presentes na outra. Mas o que pode ser observado é que muitos que apresentam boa pontuação por contato também apresentam boa pontuação por interação eletrostática mostrando talvez que os dois tipos de pontuação são complementares. A análise dos resultados é deixada ao usuário que então toma as decisões. Isto faz do programa um processo não totalmente automático, permitindo para o usuário ainda exercitar seu bom senso na decisão dos passos seguintes.

Enquanto cada uma das opções para o cálculo da malha faz uso de uma rede cúbica, há diferenças nos detalhes de implementação. A malha de contatos é automaticamente construída para envolver os átomos incluídos, os quais podem ser parte ou todo o receptor. A malha eletrostática envolve um volume cúbico, o qual, devido à natureza dos cálculos, deveria incluir a molécula inteira do receptor.

O volume envolvido pela malha do campo de forças pode ter diferentes extensões em x , y e z . Todo os átomos do receptor são incluídos nos cálculos estando ou não dentro da malha definida pelo volume em análise. A malha do campo de força pode ser posicionada ou pela especificação direta de suas coordenadas ou pela centragem dentro de um aglomerado de esferas; desta forma pode-se definir uma caixa que eficientemente envolve o espaço que provavelmente será ocupado pela molécula a ser posicionada.

5.3.3.1 Pontuação por Contato

A pontuação por contato baseia-se no contato entre os átomos do ligante e da estrutura receptora. O programa DISTMAP (Shoichet *et al.*, 1991, 1992) produz uma malha para a análise da pontuação do posicionamento por contatos. O usuário especifica a resolução, dois limites de contato próximo (contatos próximos - "close contacts" - para receptores polares e apolares) e uma distância de corte para definir o intervalo de contatos relativo aos pares. Para cada átomo do receptor dentro do intervalo de contato, a soma em um ponto da malha é aumentada de uma unidade, a menos que um limite de contato próximo seja violado, neste caso sendo então adicionado um número negativo. Hidrogênios não são incluídos nos cálculos.

Os parâmetros de entrada para este programa são distância mínima para átomos próximos, polares e não-polares. Pontos da malha mais próximos do que a distância limite de qualquer

nitrogênio ou oxigênio recebem pontuação negativa para o caso polar. Pontos da malha mais próximos do que a distância limite de qualquer átomo que não seja nitrogênio, oxigênio, hidrogênio ou deutério recebem pontuação negativa para o caso não-polar. Distância limite para contatos "bons" e número de pontos da malha por Ångström são outros parâmetros de entrada para o programa.

5.3.3.2 Pontuação por Interação Eletrostática

A análise da pontuação das interações eletrostáticas se faz através da energia de interação eletrostática baseada no potencial calculado pelo programa Delphi (Klapper *et al.*, 1986, Gilson *et al.*, 1987). O program Delphi usa um algoritmo de diferenças finitas para resolver a equação de Poisson-Boltzmann. Os autores do programa acreditam que o potencial eletrostático resultante é mais realístico do que aquele de campos de força padrão (Fersht e Sternberg, 1989). Dielétricos externos e internos de diferentes magnitudes, força iônica diferente de zero e efeitos de exclusão iônica podem ser modelados (Klapper et al., 1986, Gilson *et al.*, 1987). O potencial é calculado na ausência do ligante.

O programa CHEMIGRID (Meng *et al.*, 1992) produz os valores para o cálculo da pontuação do campo de força. Estas pontuações, ou energia de interação de mecânica molecular, são calculados como uma soma de componentes de van der Waals:

$$E = \sum_{i=1}^{lig} \sum_{j=1}^{rec} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + 332,0 \frac{q_i q_j}{D \cdot r_{ij}} \right] \quad (5.1)$$

onde cada termo é uma soma dupla sobre os átomos do ligante *i* e os átomos do receptor *j*, *A_{ij}* e *B_{ij}* são os parâmetros de repulsão e atração de van der Waals, *r_{ij}* é a distância entre os átomos *i* e *j*, *q_i* e *q_j* são as cargas pontuais dos átomos *i* e *j*, *D* é a função dielétrica, e 332,0 é um fator que converte a energia eletrostática em quilocalorias por mol. A equação (5.1) contém o termo "intermolecular" presente na função de mecânica molecular do programa AMBER (Weiner *et al.*, 1984), exceto por um termo explícito de ligação de hidrogênio. Os autores do programa assumem que a energia das ligações de hidrogênio pode ser abrangida pelo termo eletrostático.

Pontuação baseada na malha pode ser completada eficientemente quando os termos do ligante e do receptor são separados nos cálculos das funções. Isto é geralmente verdade para a

parte eletrostática de uma função potencial. Para os termos de van der Waals, é necessário o uso de uma aproximação geométrica média (Pattabiraman et al., 1985, Hagler et al., 1977):

$$A_{ij} = \sqrt{A_{ii}} \cdot \sqrt{A_{jj}} \quad e \quad B_{ij} = \sqrt{B_{ii}} \cdot \sqrt{B_{jj}} \quad (5.2)$$

onde os parâmetros para um átomo são calculados a partir do raio de van der Waals, R, e a profundidade do poço, ϵ , de acordo com:

$$A = \epsilon (2.R)^{12} \quad e \quad B = 2.\epsilon (2.R)^6 \quad (5.3)$$

Por esta aproximação a equação (5.1) torna-se:

$$E = \sum_{i=1}^{lig} \left[\sqrt{A_{ii}} \sum_{j=1}^{rec} \frac{\sqrt{A_{jj}}}{r_{ij}^{12}} - \sqrt{B_{ii}} \sum_{j=1}^{rec} \frac{\sqrt{B_{jj}}}{r_{ij}^6} + 332,0 . q_i \sum_{j=1}^{rec} \frac{q_j}{D . r_{ij}} \right] \quad (5.4)$$

Três valores são estocados para cada ponto k da malha, cada um sendo uma soma sobre os átomos do receptor que se encontram dentro de um limite estipulado pelo usuário:

$$aval = \sum_{j=1}^{rec} \frac{\sqrt{A_{jj}}}{r_{jk}^{12}}, \quad bval = \sum_{j=1}^{rec} \frac{\sqrt{B_{jj}}}{r_{jk}^6} \quad e \quad esval = 332,0 \sum_{j=1}^{rec} \frac{q_j}{D . r_{jk}} \quad (5.5)$$

Estes valores, com ou sem interpolação, podem ser multiplicados por fatores apropriados do ligante para dar a energia de interação.

Os parâmetros de entradas para o programa CHEMIGRID (Meng *et al.*, 1992) incluem resolução da malha, localização e dimensões, a forma da função dielétrica (constante ou dependente da distância), fator de escala para função dielétrica, distância limite para interação para átomos não ligados. Dois arquivos de parâmetros são lidos durante a execução do programa: uma tabela contendo as cargas para cada tipo de interação de van der Waals para cada um dos 20 aminoácidos padrões e uma tabela contendo A^{12} e E interação de van der Waals.

5.4 Parâmetros Usados na Busca de Compostos a Serem Ajustados na TIM de *T. brucei*

Nesta seção serão apresentados todos os parâmetros usados na busca efetuada para o modelo da TIM de *T. brucei* com a estrutura em volta flexível na conformação aberta (Para maiores detalhes desta conformação veja Capítulos 4 e 6 desta tese).

5.4.1 Modelo da TIM de *T. brucei* Utilizado

O modelo do complexo entre TIM de *T. brucei* e o inibidor N-hidroxi-4-fosfonobutanamida (4PBH, já discutido no Capítulo 4) foi o usado para a busca através do programa DOCK (versão 3.0 - versão disponível durante a realização deste trabalho). Como o tratamento das moléculas de água torna o procedimento bastante complexo, todas as moléculas de água assim como a molécula do inibidor foram removidas. Somente a subunidade 2 foi utilizada, pois apesar do inibidor 4PBH também se ligar a subunidade 1, esta apresenta a estrutura em volta flexível fixa por interações com outras subunidades relacionadas por operação de simetria e qualquer ajuste devido à ligação do inibidor não é observada, além do inibidor apresentar baixo fator de ocupação para o sítio da subunidade 1. A subunidade 2, livre para se adaptar aos diferentes tipos de ligantes provavelmente resultou num melhor padrão para a TIM de *T. brucei* complexada com um inibidor na conformação aberta da estrutura em volta flexível. O modelo foi refinado por Verlinde *et al.* (1992b), atingindo um fator R de 11,5% para o intervalo de resolução de 6,0 a 2,84 Å, com boa geometria.

O inibidor 4PBH apresenta uma ampla rede de interações com a proteína: 8 ligações de hidrogênio diretamente com a proteína e 5 com moléculas de água. No estudo realizado com o programa DOCK deve-se optar por um sítio da proteína. Sabe-se que as moléculas de água fazem parte da estrutura do complexo, mas também sabe-se que a distribuição e orientação destas moléculas de água muitas vezes dependem do inibidor em questão. Desta forma para este estudo todas as moléculas de água foram removidas. Apenas interações entre a proteína e o inibidor foram analisadas e o mesmo sítio de ligação do inibidor 4PBH foi utilizado.

5.4.2 Cálculo da Superfície Através dos Programas MS e SPHGEN

Tanto a proteína como o inibidor são usados para o cálculo da superfície. Para o receptor, no caso a proteína, calcula-se uma imagem negativa, enquanto para o ligante uma imagem positiva é calculada.

As esferas para o receptor são calculadas usando a superfície molecular descrita por Richards (1977) calculada através do programa MS (Connolly, 1983). As esferas foram então calculadas através do programa SPHGEN, usando todos os pontos da superfície molecular, com

raios máximos e mínimos de 4,0 Å e 1,4 Å respectivamente. O programa lê a superfície molecular gerada pelo programa MS, sendo especificado se as esferas terão seus centros dentro da superfície (ligante) ou fora (receptor). O programa gerou 28 conjuntos ("clusters") para o sítio da TIM, sendo que o maior deles continha 56 esferas. Somente este conjunto foi utilizado na procura pelo banco de dados já que a inclusão de mais de um conjunto torna a procura muito lenta e considera-se que o primeiro conjunto é o que melhor representa o sítio de ligação.

5.4.3 Cálculo da Malha para a Pontuação

Duas malhas de pontuações foram calculadas como descrito nas seções anteriores: de contato e de interação eletrostática.

5.4.3.1 Cálculo da Malha para a Pontuação por Contato

O programa DISTMAP (Shoichet *et al.*, 1991, 1992) foi utilizado para o cálculo da malha para a pontuação por contato. DISTMAP produz um arquivo de pontuações pré-calculadas. Cada ponto é calculado em termos da distância de cada átomo do receptor incluído na malha. Limites de 2,4 Å e 2,8 Å foram usados para contatos próximos em interações polares e apolares respectivamente. O valor de 4,5 Å foi usado como limite para contatos considerados como "bom". O número de 3 pontos da malha por Å foi usado, tendo uma distância de 0,33 Å entre cada ponto da malha, valor este razoável para o cálculo da pontuação e não comprometendo o tempo computacional. Uma caixa com limites que envolvem o ligante cristalográfico 4PBH foi definida com dimensões de aproximadamente 34 Å. Esta caixa envolve um volume suficiente para incluir o sítio de seletividade de interesse para o desenho de algum inibidor que apresente seletividade, sendo incluídos neste volume 789 átomos não-hidrogênio.

5.4.3.2 Cálculo da Malha para Pontuação por Interação Eletrostática

O programa CHEMIGRID (Meng *et al.*, 1992) foi utilizado para o cálculo da malha para pontuação por interação eletrostática. O arquivo de coordenadas da proteína foi lido como dado de entrada. Previamente as moléculas do ligante 4PBH assim como as moléculas de água foram removidas do arquivo de coordenadas. Também o arquivo contendo os parâmetros do receptor, que podem ser "protein, AMBER united-atom charges, AMBER hydrogen names", "protein, AMBER united-atom charges, PBD hydrogen names" e " nucleic acid, AMBER united-atom charges", foi lido, sendo que para este caso o arquivo "protein, AMBER united-atom charges, PBD hydrogen names" foi usado. Os parâmetros contidos nos arquivos acima são da referência de 1984 do AMBER (Weiner *et al.*, 1984). O arquivo contendo os parâmetros de van der Waals trás os parâmetros $A^{1/2}$ e $B^{1/2}$. O centro de massa do conjunto ("cluster") foi escolhido para ser o centro da caixa usada no cálculo da malha. Somente um conjunto foi usado no cálculo, como mencionado na seção 5.4.2. A malha foi definida como tendo separação de 0,3 Å. Uma função dependente da distância para o dielétrico foi usada, sendo $D=4,0$ (equação (5.4)). Distância de corte de 10 Å foi usada, com distâncias de 2,3 e 2,8 Å para limites próximos para interações polares e apolares respectivamente.

5.4.4 Banco de Dados de Moléculas Pequenas

"Fine Chemical Directory" foi modificado para ter cargas para ser possível a utilização da pontuação por interação eletrostática através do pacote BIOGRAF (BIOGRAF 2.10, Biodesign Inc., 199 S. Los Robles Ave., Pasadena, CA 91191, 1989), o qual usa o programa DELPHI (Kappler *et al.*, 1986, Gilson *et al.*, 1987). Cerca de 40 000 compostos compõe o banco de dados.

5.4.5 Os Parâmetros Utilizados para o Programa DOCK

Após todos os passos acima completados, o programa DOCK3.0 foi utilizado. Como teste somente o ligante 4PBH foi utilizado para verificar a eficiência do programa em testar um

composto que apresenta a estrutura determina em complexo com a proteína. DOCK3.0 apresenta duas opções para a execução: "SINGLE" e "SEARCH". No modo "SINGLE" somente um ligante é usado contra a proteína. Neste caso vários conjuntos ("clusters") podem ser testados, pois o tempo computacional é bastante reduzido. Mas como o modo usado foi o "SEARCH" com o banco de dados descrito acima, usou-se apenas um conjunto. O ligante 4PBH foi encaixado na mesma posição do resultado cristalográfico, mas girado de 180°. Vários parâmetros foram alterados e o resultado apresentou sempre a mesma conformação. Não é simples explicar qual o motivo de tal resultado mas existe uma clara polaridade nos extremos da molécula 4PBH (Figuras 4.3 e 5.3). O programa provavelmente encontra valores de pontuação maiores para esta conformação girada de 180°. Talvez a alteração em parâmetros da definição da malha de pontuação poderia ser feita para observar o comportamento do resultado. Mesmo com o resultado obtido com 4PBH através do programa DOCK3.0 discordando do resultado cristalográfico, o programa foi usado na procura através do banco de dados. Os mesmos parâmetros descritos acima foram usados para o modo "SEARCH" para o banco de dados descrito.

A pontuação por contato e por campo de força foi escolhida como sugestão dos autores por ter apresentado os melhores resultados (Meng *et al.*, 1992), executadas no modo "SEARCH". Tolerância de 1,5 Å foi usada para as distâncias entre os centros do ligante e receptor. Cinco pares foram testados para gerar uma orientação. Nenhum ligante foi removido baseado na distância interna mais longa. DOCK deve encontrar no mínimo 4 pares de centros ligantes/receptores para gerar uma orientação, com camadas para centros de ligantes de 0,3 Å, tendo 0,1 Å de superposição entre as camadas dos centros dos ligantes. Camadas de 0,8 Å foram usadas para os centros do receptor (centros das esferas) com 0,2 Å de superposição entre as camadas dos centros do receptor. Somente 1 contato impróprio foi permitido. Somente ligantes com no mínimo 5 átomos (inclusive) e máximo de 100 átomos (inclusive) foram incluídos na procura. As 300 melhores soluções foram guardadas em um arquivo, sendo que apenas 100 foram mantidas após a normalização da pontuação. Nenhum refinamento de cada solução foi feito através da opção "zoom" devido ao grande número de ligantes a ser testado. Interpolação trilinear foi adotada.

5.5 Resultados

De todos os compostos no banco de dados, 39 100 foram utilizados através do programa DOCK3.0 para encontrar um composto que apresentasse interações com o sítio ativo da TIM de *T. brucei* na conformação aberta da estrutura em volta flexível. Com as definições acima, foram listadas as 300 melhores soluções segundo a pontuação por contato e 300 outras por interação eletrostática. Como já discutido, o passo que ainda é deficitário é o sistema de pontuação, que envolve definições de parâmetros de análise que possam representar o mais realisticamente as interações entre os ligantes e a proteína. Esta dificuldade é compartilhada por todos os programas disponíveis para o desenho racional de drogas.

O programa DOCK apresenta opções para até dois tipos de pontuação mas não apresenta uma pontuação combinada dos dois tipos. O programa apresenta apenas, desta forma, duas listas com os melhores resultados em cada pontuação. Os resultados presentes em uma das listas como os melhores não necessariamente encontram-se presentes na outra lista. Desta forma decidiu-se fazer uma análise primeiramente independente de cada lista de pontuação. Como uma lista de pontuação combinada poderia introduzir ligantes que satisfizessem os dois tipos de pontuação, decidiu-se gerar uma lista, na qual os melhores resultados desta combinação fossem listados.

5.5.1 Pontuação por Contato

O sítio ativo da TIM apresenta um sítio de ligação que pode ser descrito da seguinte forma: o sítio catalítico o qual é constituído pelos resíduos Asn11, Lys13, His95 e Glu167; um canal constituído de um lado pelos resíduos Gly211, Gly212, Ser213 e Val214 e pelo outro lado por Leu232, Val233, Gly234, Gly235 e Ala236; e finalmente pela estrutura em volta flexível, principalmente pela Gly 172 a qual é a única interação da estrutura em volta flexível com o ligante na conformação fechada. A interação do composto 4PBH com os resíduos do sítio de ligação pode ser vista na Figura 5.3. A parte polar formada pelo grupo carboxílico, de um lado interage com a Lys13 e His95, enquanto que o oxigênio do grupo nitroxi interage com o nitrogênio da Asn11. Não há carga positiva líquida neste sítio suficiente para satisfazer o grupo fosfato. Por outro lado os oxigênios do grupo fosfato não interagem com nenhum grupo carregado positivamente. Os únicos grupos com os quais os oxigênios interagem são nitrogênios de cadeia principal e moléculas de água, os quais circundam o grupo fosfato, desta forma estabilizando a

carga.

Segundo a lista dos compostos que apresentam os melhores resultados por contato, o valor máximo foi de 241 (unidade adimensional - veja seção 5.3.3.1). Apesar de serem apenas pontuados por contato os cinco melhores resultados apresentam um número razoável de ligações de hidrogênio. O Composto Z978, com a quinta melhor pontuação por contato apresenta 9 ligações de hidrogênio (cada interação é contada aqui, ou seja, um átomo que interage com dois outros átomos é contado como tendo duas ligações de hidrogênio). Primeiro e segundo compostos da lista de pontuação por contato (d792 e Z560) apresentam 7 ligações de hidrogênio, sendo que o composto Z560 apresenta um dos contatos com distância inferior ao observado para ligação de hidrogênio (distância de 1,92 Å entre um nitrogênio da cadeia principal e um oxigênio do grupo carboxílico do ligante). O composto e443 (terceiro melhor resultado) apresenta apenas 5 interações e o composto n189 (quarto melhor resultado) não apresenta nenhuma interação.

Figura 5.3 - Interação do composto 4PBH com resíduos do sítio ativo da TIM de *T. brucei*, como determinado por técnicas cristalográficas (Verlinde *et al.*, 1992b). Interações com moléculas de água foram removidas.

Todos os compostos apresentam boa interação hidrofóbica com os resíduos constituintes de ambos os lados do canal do sítio ativo, assim como a parte hidrofóbica dos resíduos catalíticos e da estrutura em volta flexível. O programa não oferece para esta lista a energia de interação eletrostática o que impede uma análise comparativa das energias de interação.

Figura 5.4 - Interações dos compostos (a) e443 e (b) Z978 com os resíduos do sítio ativo da TIM de *T. brucei*, obtidos da procura através do programa DOCK. Maiores detalhes no texto.

Apesar de apresentar um número menor de interações, o composto e443 é o que apresenta interações com o maior número de resíduos os quais participam de interações com o composto 4PBH (Lys13, His95, Gly234, Gly212, Val214), interagindo inclusive com o Glu167, mas não interagindo com Asn11. Os compostos d792 e Z978 apresentam interações com a Lys13 e mantém interações com os resíduos Gly232, Gly234, Gly 211, Gly212.

Este pode ser um bom composto guia pois oferece ainda opções para introduzir diferentes grupos nas outras posições do anel, possibilitando um aumento da afinidade ao sítio de ligação da proteína. A parte alifática do composto também oferece possibilidades para introduzir grupos

polares para interagir com mais resíduos, através de interações eletrostáticas. Desta forma este é um excelente candidato como um composto guia para ser utilizado em futuras modelagens. Este composto encontra-se aproximadamente 9,0 Å do sítio de seletividade, o que permitirá, após aumentar a afinidade ao sítio de ligação do composto, definir ligantes que possam se estender desta posição e então interagir com o sítio de seletividade, conferindo ao composto a seletividade necessária.

Os compostos e443 e Z978 são os que apresentam interações distribuídas por toda a molécula, em ambos os extremos e na região central, o que difere grandemente do composto Z792. Desta forma ambos são candidatos a compostos guia e devem ser testados quanto à afinidade ao sítio ativo da TIM de *T. brucei*.

5.5.2 Pontuação por Interação Eletrostática

O programa DOCK3.0 apresenta além da lista dos melhores resultados da pontuação por contato, uma lista com os melhores resultados da pontuação por interação eletrostática. A vantagem desta lista, diferentemente da lista de contatos, é a presença das pontuações tanto de interação eletrostática quanto as de contato. Isto permite uma análise comparativa. O melhor resultado da interação eletrostática apresenta uma energia de -76,1 Kcal/mol, com uma pontuação de 48 unidades para contato. Como pode ver visto a pontuação por contato é bastante insignificante quando comparada ao valor máximo obtido de 241 (apenas 20% do valor máximo obtido na lista de pontuação por contato - seção anterior). A energia de interação eletrostática mais baixa foi de -27,1 Kcal/mol.

O composto com a melhor pontuação em interação eletrostática, composto a895, apresenta duas interações com grupos que poderiam ser interpretadas como ligação de hidrogênio, sendo que uma delas apresenta uma distância entre um nitrogênio do composto e oxigênio da cadeia principal da proteína de apenas 1,0 Å. Existem 5 átomos de nitrogênio e 3 de oxigênio neste composto os quais fazem com que a interação eletrostática atinja os valores observados, mas pela análise gráfica não apresenta interações satisfatórias para ser considerado como um bom candidato para um composto guia.

Os compostos que apresentam a segunda e terceira melhor pontuação são constituídos de um anel benzênico com seis e cinco grupos carboxílicos ao redor, apresentando 12 e 10 átomos de oxigênio, respectivamente. Neste caso não existe interação hidrofóbica com a proteína pois

os átomos de oxigênio cobrem toda a superfície hidrofóbica da molécula restando apenas os dois lados do anel que pouco interagem com a proteína. Além disto os compostos são relativamente pequenos para se estenderem em direção ao sítio de seletividade da proteína.

Os outros resultados (quarto e quinto melhores compostos) são derivados de fosfatos. Um deles, o composto J392, apresenta dois grupos fosfato, um em cada extremidade, os quais não interagem adequadamente com a proteína, além de se estender em direção oposta ao sítio de seletividade.

O composto C395, derivado de fosfato, apresenta 3 grupos fosfato, os quais interagem com 4 resíduos da proteína mas encontram-se em torno de 7 Å da região de seletividade. Compostos derivados de grupos fosfato não são novidades como pode ser visto no Capítulo 4 desta tese, pois o substrato desta proteína é um composto com um grupo fosfato. As dificuldades com grupos fosfato podem ser enormes pois além de serem instáveis, são comuns as proteínas que ligam fosfato podendo ser no futuro um problema devido à possibilidade destes compostos ligarem-se à proteína de interesse, mas também ligarem-se à outras proteínas que possuem um sítio de ligação para grupos fosfato.

Uma observação comparativa mostra que estes compostos com grupos fosfato apresentam interações mais pobres que os complexos já analisados cristalograficamente (veja Capítulo 4 para maiores detalhes e referências dos complexos entre compostos contendo grupo fosfato e TIM).

5.5.3 Pontuação Combinada

Como dois compostos apresentaram-se promissores como compostos guia para futuras análises (seção 5.5.1), e a pontuação por interação eletrostática não apresentou compostos com características especiais, uma pontuação combinada foi utilizada. A combinação foi puramente a soma dos valores absolutos de cada tipo de pontuação. Existem várias maneiras de se combinar os resultados. Em princípio esta metodologia carece de sentido físico pois combina-se duas pontuações com unidades diferentes. O objetivo aqui não foi necessariamente obter um resultado justificável fisicamente mas obter os compostos que apresentassem valores médios altos de ambas as pontuações. Os melhores resultados encontram-se listados na Tabela 5.1

Tabela 5.1 - Lista dos compostos que apresentaram melhor resultado na pontuação combinada

Composto	Pontuação por	
	Contato	Interação Eletrostática
I355	197	-31.731
J475	165	-28.762
I835	159	-28.341
k938	153	-27.956
c963	151	-28.064

Figura 5.5 - Interação do composto I355 com resíduos do sítio ativo da TIM de *T. brucei*, obtido através da pontuação combinada. Maiores detalhes no texto

Este resultado foi analisado mostrando compostos muito mais razoáveis que os resultados apresentados nas seções 5.5.1 e 5.5.2. Estes compostos são mais estendidos ocupando mais amplamente o sítio ativo, interagindo com resíduos envolvidos no mecanismo de catálise assim como com resíduos do sítio de ligação do fosfato. Existe uma ampla interação hidrofóbica com a estrutura em volta flexível que em solução muito provavelmente adotaria uma conformação mais "fechada", aumentando a interação hidrofóbica, assim como complementando interações eletrostáticas. Resíduos expostos ao solvente são claramente polares permitindo interações com moléculas de água. O composto que apresentou melhor resultado pode ser visto na Figura 5.5.

Existe um oxigênio interagindo com a His95 e Asn11. O anel pode ainda ser estendido

parte intermediária um oxigênio e um nitrogênio interagem com grupos do canal, que ainda permitem modificações para ampliar a malha de interação. A outra extremidade apresenta interações com resíduos da extremidade do canal e ainda podem participar de interações com moléculas de água. Estes resíduos apresentam ótimas características e fornecem opções para serem modificados para aumentar a interação com a proteína.

5.6 Conclusão

Este projeto teve início com a obtenção por Verlinde *et. al.* (1992b) da estrutura de TIM de *T. brucei* em complexo com o composto 4PBH, o qual se liga a conformação aberta da estrutura em volta flexível. Anteriormente, devido a todos os resultados de complexos serem obtidos com a conformação fechada da estrutural em volta flexível, todas as modelagem foram feitas tendo esta estrutura como básica. Isto limitava o desenvolvimento de novos inibidores para apenas uma conformação além de introduzir impedimentos estéricos para se alcançar o sítio de seletividade. Um inibidor não necessariamente precisa ser competitivo. Se se obtiver um inibidor que impeça a estrutura em volta flexível de se mover a catálise enzimática desenvolvida pela TIM sofrerá decréscimo substancial. Portanto a procura por outras maneiras de conseguir que esta proteína tenha sua atividade alterada e ao mesmo tempo, conferindo ao composto que interfira com a atividade certa seletividade, possibilitará a obtenção de um inibidor potente e seletivo. A grande vantagem que se encontra neste enfoque é que o sítio a ser procurado pode ser diferente do sítio ativo, mais próximo do sítio de seletividade, portanto permitindo que estes compostos não sejam impraticáveis como o DG-11 (Verlinde *et. al.*, 1992a). Um grande esforço tem sido feito para obtenção de compostos que apresentem seletividade baseado na conformação fechada da estrutura em volta flexível, sendo que nenhum apresentou a menor contribuição para esta seletividade. Trabalhos também foram feitos com o objetivo de alterar a estabilidade do dímero de TIM, com a modelagem de oligopeptídeos mas sem apresentar resultados promissores (Dr. Rik Wierenga, comunicação pessoal). Apesar de TIM de *T. brucei* ter sido a primeira estrutura a ser determinada no projeto de desenho racional de drogas contra a doença do sono, com quantidade reduzida de proteína nativa (4 mg) e ser a que hoje apresenta a maior quantidade de proteína expressa em *E. coli* e de fácil cristalização (obedecido o tempo de envelhecimento de um mês) as dificuldades em obter seletividade são muito superiores às outras proteínas do projeto (um exemplo promissor é o da GAPDH de leishmania, Verlinde *et. al.*, 1994).

Cada proteína tem características particulares que conferem portanto maior ou menor dificuldade na obtenção de sistemas ideais para a modelagem de inibidores que venham a ser eficientes para o desenvolvimento de drogas seletivas. Portanto parece que TIM, na conformação fechada da estrutura em volta, não é o sistema ideal para a modelagem de drogas seletivas. A escolha da conformação aberta abre novas perspectivas, que se iniciaram com o trabalho apresentado neste Capítulo desta Tese. Os resultados aqui obtidos foram iniciais e serão portanto utilizados para tentar completar o ciclo de modelagem (Figura 5.1) e obtenção de inibidores, com possibilidades de serem posteriormente desenvolvidos como drogas.

Vários programas permitem a análise computacional destes compostos e particularmente aqui o programa DOCK foi escolhido por ter apresentado vários resultados positivos e por apresenta um tratamento das interações entre proteína e ligante de uma forma bastante generalizada, portanto representando de uma forma mais completa a interação entre ligante e proteína.

O programa DOCK3.0 não foi extensivamente explorado devido ao grande número de variáveis envolvidas em todo o processo de geração das superfícies, cálculo de malhas e o próprio programa DOCK. Nas seções anteriores foi apresentado um trabalho inicial na utilização do programa DOCK3.0 o qual parece ser bastante eficiente na procura de novos compostos que venham a ser utilizados como compostos guia para futuros inibidores mais potentes a nível de afinidade e seletividade.

Três compostos apresentaram características inicialmente promissoras para serem testados como possíveis compostos guias e posteriormente modificados para compostos que venham a apresentar valores de afinidade abaixo de μM e finalmente estendidos aos sítios de seletividade. Os compostos derivados de grupos fosfatos apareceram na lista dos melhores resultados das interações eletrostática. No entanto, no desenho de potentes inibidores que venham a ser desenvolvidos como drogas, deve-se preferencialmente buscar derivados de grupos funcionais alternativos para evitar possíveis problemas com estabilidade e efeitos colaterais que podem ocorrer com estes compostos.

Realmente é difícil entender porque o composto 4PBH foi modelado na conformação invertida pelo programa DOCK. O número de variáveis no programa DOCK permite uma vasta gama de combinações. Encontrar a perfeita combinação que descreva perfeitamente o sítio ativo é tarefa quase impossível. Portanto fica claro que a pontuação deve exercer uma função crucial nos programas de modelagem de compostos em sítios específicos. No entanto a pontuação ainda é um problema a ser melhorado neste procedimento. Novos sistemas de pontuações seriam talvez uma opção para os autores destes programas para introduzir descrições mais completas do sítio

das proteínas.

Com o que se encontra disponível e exercendo a importante função de análise crítica pode-se obter resultados iniciais que poderão ser desenvolvidos em compostos que apresentarão além de maior afinidade, também seletividade. Os resultados além de fornecerem informações valiosas para o desenvolvimento destes compostos, irão fornecer mais informações que poderão ser usadas no aprimoramento dos sistemas de pontuações que tornarão os programas mais efetivos.

Os três compostos sugeridos aqui podem ser adquiridos facilmente e testados na proteína TIM de *T. brucei* através de ensaio enzimático, e se mostrarem inibição significativa, serão co-cristalizados ou difundidos em cristais, trabalho este que será seguido futuramente pelo estudante de doutorado Brad Bernstein, do grupo do Dr. Wim Hol. Com os resultados cristalográficos, análises gráficas poderão sugerir novas modificações para a busca da seletividade.

Este ciclo será o utilizado na proteína gliceraldeído-3-fosfato desidrogenase de *T. cruzi*, projeto este em andamento no Grupo de Cristalografia de Proteínas da USP de São Carlos. Espera-se que à partir dos resultados já obtidos pelo grupo do Dr. Hol com a proteína homóloga de *T. brucei* poder-se-á avançar mais rapidamente através do uso de compostos guias já obtidos e testados quanto a afinidade e seletividade (Verlinde *et al.*, 1994).

Capítulo 6
Determinação da Estrutura Cristalográfica da
Triosefosfato Isomerase de *Bacillus*
***stearothermophilus* em Complexo com**
2-fosfoglicolato, um Inibidor Análogo ao Substrato,
à 2,8 Å de Resolução

6.1 Introdução

Triosefosfato isomerase (TIM) é um dímero de subunidades idênticas com peso molecular de aproximadamente 25 KDa que catalisa a interconversão do D-gliceraldeído-3-fosfato (GAP) e di-hidroxiacetona fosfato (DHAP) no ciclo da glicólise. Somente o dímero é ativo (Waley, 1973, Zabori *et al.*, 1980). Propriedades catalíticas da TIM têm sido estudadas em detalhes por numerosos investigadores, através de estudos cinéticos (Belasco e Knowles, 1980) análises estruturais dos complexos dos análogos dos estados de transição (Lolis and Petsko, 1990, Davenport *et al.*, 1991), experimentos de mutação sítio dirigida (Blacklow *et al.*, 1991) e estudos quânticos e de mecânica molecular (Bash *et al.*, 1991). Os resíduos fundamentais na catálise são Lys 10, His 94 e Glu 166, usando a numeração da sequência de aminoácidos da TIM de *Bacillus stearothermophilus*.

O mecanismo de catálise, bem conhecido, envolve uma mudança conformacional no sítio catalítico pelo movimento de uma estrutura em volta ("loop") chamada de flexível (já discutida nos Capítulos 4 e 5) a qual fecha o sítio ativo quando o substrato se liga. Várias estruturas de complexos de inibidores análogos ao substrato foram resolvidas cristalograficamente mostrando três diferentes conformações para a estrutura em volta flexível, "aberta" ("open") quando não existe nenhum ligante no sítio ativo (Noble, 1992), "parcialmente fechada" quando o íon sulfato liga-se ao sítio ativo (Wierenga *et al.*, 1991a), e "fechada" ("closed") quando inibidores análogos ao substrato encontram-se ligados ao sítio ativo (Lolis *et al.*, 1990, Noble *et al.*, 1991a, Noble *et al.*, 1991b). Há somente um caso reportado de um inibidor com um carbono a mais que o substrato (o substrato possui três átomos de carbono), N-hidroxi-4-fosfono-butanamida, o qual se liga à enzima com a estrutura em volta flexível na conformação aberta (Verlinde *et al.*, 1992b).

Até o momento 20 sequências de TIM são conhecidas, das quais 18 encontram-se depositadas no banco de sequências e foram utilizadas para o alinhamento múltiplo que pode ser observado na Tabela 6.1. Cinco sequências de triosefosfato isomerase têm estrutura tridimensional reportada: TIM de galinha (Banner *et al.*, 1975), TIM de levedura (Lolis *et al.*, 1990), TIM de *T. brucei* (Wierenga *et al.*, 1991b), TIM de *E. coli* (Noble *et al.*, 1993) e TIM humana (Mande *et al.*, 1994).

A estrutura descrita neste trabalho é a sexta a ser resolvida cristalograficamente. Uma comparação detalhada das estruturas de TIM de galinha, levedura e de *T. brucei* já foi feita (Wierenga *et al.*, 1992) mas nenhuma destas enzimas são termicamente estáveis. A estrutura descrita neste trabalho é a primeira TIM que apresenta estabilidade térmica aumentada.

	1			50		100								
Copja	..MSRKEFEVGG	NWKMN.GTGE	EVEKIVTILN	EA.....EV	IEELDVEEVV	VSEFFVFLFP	VKNLLR.A..	DFHVAAQNCW	VKKGGAGTGE	VSAEMLVNLG	IPWVILGHSE	RRALLNESNE	FVGDKTAYAL	SQGLKVIACV
Maize	..GRKEFEVGG	NWKMN.GTTD	QVEKIVTILN	EQ.....QV	IES.DVEEVV	VSEFFVFLFP	VKSQLR.Q..	EFHVAAQNCW	VKKGGAGTGE	VSAEMLVNLG	VPWVILGHSE	RRALLGESNE	FVGDKVAYAL	SQGLKVIACV
Human	AISRKKEFEVGG	NWKMN.GRQK	SLGELIGTILN	AA.....KV	IAD.T..EVV	CAFFTAYIDF	ARQKLD.P..	KIAVAAQNCY	KVTNGAGTGE	ISPGMIKDCG	ATWVVLGHSE	RRHVFGESE	LIGQKVAHAL	AEGLGVIACI
Macmu	AISRKKEFEVGG	NWKMN.GRQK	NIGELIGTILN	AA.....KV	IAD.T..EVV	CAFFTAYIDF	ARQKLD.P..	KIAVAAQNCY	KVTNGAGTGE	ISPGMIKDCG	ATWVVLGHSE	RRHVFGESE	LIGQKVAHAL	AEGLGVIACI
Rabit	AISRKKEFEVGG	NWKMN.GRKK	NIGELIGTILN	AA.....KV	IAD.T..EVV	CAFFTAYIDF	ARQKLD.P..	KIAVAAQNCY	KVTNGAGTGE	ISPGMIKDCG	ATWVVLGHSE	RRHVFGESE	LIGQKVAHAL	SEGLGVIACI
Mouse	AETREKEFEVGG	NWKMN.GRKK	CLGELIGTILN	AA.....NV	EAG.T..EVV	CAFFTAYIDF	ARQKLD.P..	KIAVAAQNCY	KVTNGAGTGE	ISPGMIKDLG	ATWVVLGHSE	RRHVFGESE	LIGQKVSHAL	AEGLGVIACI
Chick	..APRKEFEVGG	NWKMN.GDKK	SLGELIGTILN	GA.....KL	SAD.T..EVV	CGAPSIYLD	ARQKLD.A..	KIGVAAQNCY	KVPKAGTGE	ISPMIKDIG	AAWVILGHSE	RRHVFGESE	LIGQKVAHAL	AEGLGVIACI
Latch	..APRKEFEVGG	NWKMN.GDKK	SLGELIGTILN	AA.....KV	IEF.G..EIV	CAFFEAYLDF	ARLKVD.P..	KFGVAAQNCY	KVSKGAGTGE	ISPMIKDCG	VTWVILGHSE	RRHVFGESE	LIGQKVSHAL	SEGLGVVACI
Culta	..GRKEFEVGG	NWKMN.GDKA	SIADLKIVIT	TG.....IT	NAD.T..EVV	VGFEAYITL	ARSQLP.D..	SVCVAAQNCY	KVPKAGTGE	ISFAMKDLN	IGWVILGHSE	RRHVFGESE	LIGQKVSHAL	SEGLGVVACI
Drome	..MSRKEFEVGG	NWKMN.GRQK	SIADLKIVIT	SA.....AL	IEF.G..EIV	ICGPAIYIMY	ARNLLP.C..	ELGLAGQNAV	KVAKGAGTGE	ISFAMKDLN	ADWVILGHSE	RRHVFGESE	LIGQKVSHAL	SEGLGVVACI
Emeni	..MPRKEFEVGG	NWKMN.GNAE	STTSIIKILN	SA.....NI	IES..EIVV	VSEFPAIYI	LQ	AREVANK..	EIGVAAQNVF	DKPNGAGTGE	ISVQQLREAN	IDWTILGHSE	RRVILKETDE	FIARKTKAAI
Schpo	..ARKEFEVGG	NWKMN.GSLE	SMITIEGLN	TT.....KL	IEF.G..EIV	ICGPAIYIMY	ARNLLP.C..	ELGLAGQNAV	KVAKGAGTGE	ISFAMKDLN	ADWVILGHSE	RRHVFGESE	LIGQKVSHAL	SEGLGVVACI
Yeast	..ARTFEVGG	NWKMN.GSRK	SIADLKIVIT	TG.....IT	NAD.T..EVV	VGFEAYITL	ARSQLP.D..	SVCVAAQNCY	KVPKAGTGE	ISFAMKDLN	IGWVILGHSE	RRHVFGESE	LIGQKVSHAL	SEGLGVVACI
Trybb	..MKPQPIAAG	NWKMN.GSQQ	SLGELIGTILN	ST.....SI	NHD..EIVV	VSEFPAIYI	LQ	AREVANK..	EIGVAAQNVF	DKPNGAGTGE	ISVQQLREAN	IDWTILGHSE	RRVILKETDE	FIARKTKAAI
Bacst	..MRKEPIAAG	NWKMN.KILA	EAVQFVEDVK	GH.....KI	IEF.G..EIV	ICGPAIYIMY	ARNLLP.C..	ELGLAGQNAV	KVAKGAGTGE	ISFAMKDLN	ADWVILGHSE	RRHVFGESE	LIGQKVSHAL	SEGLGVVACI
Ecoli	..MRHPLVAG	NWKMN.GSRH	MVHELVSILR	KE.....LA	GV..AGGAVA	IAPPEMYIDM	AKREAEGS..	HIMLGAQNV	LNLGAGTGE	TSAAMLKDIG	ACQYIIIGHSE	RRYHKSDE	LIADKVAHAL	AEGLKVIACI
Corgl	..MRKPLIAG	NWKMN.LDHQ	QAIGTVQKLA	FA.....LF	HEFEEKVIVA	VTVPETIRS	VQTLVEGDKL	EVTFGAQDVS	QHESGAYTGE	VSASMLAKLN	CSWVVVGHSE	RREYHNESDE	LVAACKAAAL	SNGLSPIVCV
Mersp	..MQAWVIG	NWKMN.PATSH	DVNALLDPLC	AAITTEQMS	EDNSARQIM	VAFSLIHLAA	VNGRLKDTSI	LCAAQDVSGH	SASVGAAYTGE	CSAQQIVDAG	ATWTILGHSE	RRQYHQESND	ALLQKMIHAL	SQDLGVVFCI
Qunso	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****

	150					200					250				
Copja	GETLEQREAG	STISVVAAGT	KATAE.....	.EV..SDWIN	IVVAYEPVWA	IGTGKVASFA	QAQEVHFEIR	KWIKENVGAD	VAGSVRIIYG	GSVNGANSKE	IAGQPDIDGF	LVGGASLKPE	F.VDIISAT	VKSS.	
Maize	GETLEQREAG	STISVVAAGT	KATAE.....	.KI..KDWIN	IVVAYEPVWA	IGTGKVATPA	QAQEVHASLR	DWLKTNASPE	VAESTRIIYG	GSVTAANCKE	IAAQPDVDGF	LVGGASLKPE	F.IDIINAAT	VKSA.	
Human	GEKLDEREAG	ITEKVVEEQT	KVIAD.....	.NV..KDWSK	VVAYEPVWA	IGTGKTATPQ	QAQEVHEKLR	GWLKSNSVDA	VAQSTRIIYG	GSVTGATCKE	LASQPDVDGF	LVGGASLKPE	F.VDIINAKQ		
Macmu	GEKLDEREAG	ITEKVVEEQT	KVIAD.....	.NV..KDWSK	VVAYEPVWA	IGTGKTATPQ	QAQEVHEKLR	GWLKSNSVDA	VAQSTRIIYG	GSVTGATCKE	LASQPDVDGF	LVGGASLKPE	F.VDIINAKQ		
Rabit	GEKLDEREAG	ITEKVVEEQT	KVIAD.....	.NV..KDWSK	VVAYEPVWA	IGTGKTATPQ	QAQEVHEKLR	GWLKSNSVDA	VAQSTRIIYG	GSVTGATCKE	LASQPDVDGF	LVGGASLKPE	F.VDIINAKQ		
Mouse	GEKLDEREAG	ITEKVVEEQT	KVIAD.....	.NV..KDWSK	VVAYEPVWA	IGTGKTATPQ	QAQEVHEKLR	GWLKSNSVDA	VAQSTRIIYG	GSVTGATCKE	LATPADVDGF	LVGGASLKPE	F.VDIINAKQ		
Chick	GEKLDEREAG	ITEKVVEEQT	KAIAD.....	.NV..KDWSK	VVAYEPVWA	IGTGKTATPQ	QAQEVHEKLR	GWLKSHVSDA	VAQSTRIIYG	GSVTGGNCKE	LASQHDVDGF	LVGGASLKPE	F.VDIINAKH		
Latch	GEKLDEREAG	ITEGVVFEVT	EVIAD.....	.NV..KDWSK	VVAYEPVWA	IGTGKTASPO	QSQELHGKLR	KWLKENVSET	VADSVRIIYG	GSVTGATCKE	LASEPDVDGF	LVGGASLKPE	F.VEYKDVQR		
Culta	GETLQEREAG	QTEAVCFRQT	KAJAD.....	.KV..KDWSN	VVAYEPVWA	IGTGKTASPE	QAQEVHAAIR	KWFTENVSDA	VSAAIRIQYG	GSVTAANCRE	IAAKPDIDGF	LVGGASLKPE	F.IQIVNARQ		
Drome	GETLEEREAG	KINEVVARQM	CATAQ.....	.KI..KDWIN	IVVAYEPVWA	IGTGKTATPD	QAQEVHASLR	QWLSDNISKE	VSASLRIQYG	GSVTAANCKE	LAKKPDIDGF	LVGGASLKPE	F.LDIINARQ		
Emeni	GETLEEREAG	KTIDVVRQL	NAAAK.....	.ELSKEQWAK	VVAYEPVWA	IGTGKVATTE	QAQEVHSAIR	KWLKDAISAE	AAENTRIIYG	GSVSEKNCKD	LAKADIDGF	LVGGASLKPA	F.VDIIVNARL		
Schpo	GETLADREAN	ETITVVVRQL	NAIAD.....	.KV..QNWSK	IVVAYEPVWA	IGTGKTGTPE	EAQEVHAEIR	KWATNKLGA	VAEGLRVIYG	GSVTGGNCKE	FLKFHDIDGF	LVGGASLKPE	FPTNIVNVHS	L.....	
Yeast	GETLEEKERG	KTLDVVERQL	NAVLE.....	.EV..KDWTN	IVVAYEPVWA	IGTGLAATPE	DAQDIIHASIR	KFLKSLGDKD	AASELRILYG	GSANGSNART	FKDKADVDFG	LVGGASLKPE	F.VDIINSNR		
Trybb	GETLQERESG	RTAVVVLTQI	AAIAK.....	.KLKKADWAK	IVVAYEPVWA	IGTGKVATPD	QAQEAHALIR	SWVSSKIGAD	VRGELRILYG	GSVNGKNVAT	LYQQRDVNGF	LVGGASLKPE	F.VDIIKATQ		
Bacst	GESLEEREAG	QTNVAVVASQV	EKALA.....	.GLTPQOVQK	AVIAYEPIWA	IGTGKSSTPE	DANSVCGHIR	SVVSRLFGPE	AAEAIRIQYG	GSVKPDNIJD	FLOQRQIDGP	LVGGASLEPA	SFLQVLEAGR	HE.....	
Ecoli	GETEAENEAG	KTEFVCARQI	DAVLK.....	.TQGAAAFEG	AVIAYEPVWA	IGTGKSATPA	QAQAVHKFIR	DHIAKV..DAN	IAEQVIIQYG	GSVNASNAAE	LFAQPDIDGA	LVGGASLKAD	AFAVIVKAAE	AAKQA	
Corgl	GETLEIHEAG	THVEYVVEQT	RKSLA.....	.GLDAAELAN	TVIAYEPVWA	IGTGKVASAA	DAQEVCKAIR	GLIVELAGDE	VAEGLRILYG	GSVKAETVAE	IVGQPDVDGG	LVGGASLDGE	ALPKL..AAN	AASVA	
Mersp	GESQEYQYTK	QTLTVIDNQL	VVIKEFITQQ	PELIDALFTR	LIIVAYEPVWA	IGTGKVPTVA	EVSATHEHIK	QTLAGEF..AD	SLSNMFTVLYG	GSVNADNANS	FAADPMIDGA	LVGGASLKAD	SFLTITATFS	QASI.	
Qunso	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	

Tabela 6.1 - Alinhamento de seqüências através do programa GCG (Genetic Computer Group, Inc., Copyright 1982-1992, Wisconsin, USA) de todas as seqüências de TIM conhecidas. Em ordem: Copja, *coptis japonica* (Okada *et al.*, 1989); Maize, *zea mays* (maize) (Marchionni and Gilbert, 1986); Human, *homo sapiens* (Maquat *et al.*, 1985); Macmu, *macaca mulatta* (*rhesus macaque*) (Old and Mohrenweiser, 1988); Rabit, *oryctolagus cuniculus* (rabbit) (Corran and Waley, 1975); Mouse, *mus musculus* (mouse) (Cheng *et al.*, 1990); Chick, *gallus gallus* (chicken) (Straus and Gilbert, 1985); Latch, *latimeria chalumnae* (latimeria) (Kolb *et al.*, 1974); Culta, *culex tarsalis* (mosquito) (Tittiger *et al.*, 1993); Drome, *drosophila melanogaster* (fruit fly) (Shaw-Lee *et al.*, 1991); Emeni, *emerella nidulans* (aspergillus nidulans) (Mcknight *et al.*, 1986); Schpo, *schizosaccharomyces pombe* (fission yeast) (Russell 1985); Yeast, *saccharomyces cerevisiae* (baker's yeast) (Alber and Kawasaki, 1982); Trybb, *trypanosoma brucei brucei* (Swinkels *et al.*, 1986); Bacst, *Bacillus stearothermophilus* (Rentier-Delrue *et al.*, 1993); Ecoli, *escherichia coli* (Pichersky *et al.*, 1984); Corgl, *corynebacterium glutamicum* (Eikmanns, 1992); Mersp, *moraxella sp* (Rentier-Delrue *et al.*, 1993). A última linha de cada seqüência de alinhamento mostra, através de asteriscos, os resíduos que são conservados em todas as seqüências de TIM.

TIM de *B. stearothermophilus* é um dímero formado por dois monômeros de 252 aminoácidos cada. TIM de *B. stearothermophilus* é ativa até 65° C, significativamente mais alta que a enzima de *E. coli* a qual é ativa até ao redor de 50°C (Rentier-Delrue *et al.*, 1993). Estas proteínas são então denominadas como termofílica e mesofílica respectivamente. A comparação da sequência e estrutura da TIM de *B. stearothermophilus* com suas homólogas pode mostrar as razões para a maior estabilidade térmica. No entanto não se espera obter diferenças singularizadas que justifiquem por elas próprias a estabilidade térmica aumentada observada na TIM de *B. stearothermophilus* já que apesar de as temperaturas serem de diferença significativa, pequenas modificações podem ser responsáveis por tal diferença como verificado por substituições de apenas um aminoácido na lisozima em várias posições moleculares diferentes (Zhang, 1991). Também hoje sabe-se que *B. stearothermophilus* é um organismo de estabilidade térmica moderada, existindo organismos com estabilidade térmica extrema, sendo estáveis a temperaturas maiores que 100°C (para uma revisão veja Adams, 1993). Como a homologia das sequências de TIM com estrutura resolvida relativas à sequência da TIM de *B. stearothermophilus* é de aproximadamente 40%, é muito mais razoável que várias pequenas diferenças acumularam uma estabilidade térmica que resultou nos valores observados acima entre a TIM termofílica e as TIM mesofílicas. Fatores como ligações de hidrogênio, pontes salinas e interações hidrofóbicas assim como substituições preferenciais por resíduos conhecidos como estabilizantes têm sido estudados por diversos autores (Freire, 1993, Fontana, 1991, Creighton, 1991, Zhang, 1991) como sendo importantes fatores de estabilidade térmica. As razões para estabilidade térmica variam de proteína para proteína o que significa que cada proteína deve ter atingido a estabilidade térmica por diferentes combinações de pequenas modificações as quais aumentaram a estabilidade das estruturas secundária, terciária e quaternária.

Neste trabalho descreve-se a estrutura cristalina refinada da TIM de *B. stearothermophilus* em complexo com 2-fosfoglicolato (2PG), um inibidor análogo ao substrato e uma comparação detalhada desta estrutura com todas as outras enzimas TIM com estrutura conhecida na procura de fatores que justifiquem a estabilidade térmica aumentada desta proteína.

6.2 Cristalização da TIM de *Bacillus stearothermophilus*

Triosefosfato isomerase foi expressa em grande quantidade e purificada por Rentier-Delrue *et al.* (1993). Cristais foram obtidos pelo Dr. Shekhar Mande através de experimentos

de difusão de vapor através da técnica da gota pendurada à temperatura ambiente. Solução de proteína de 4 mg/ml em 5 mM MES¹, pH 6,5, 1 mM EDTA, 1 mM azida sódica e 2 mM de 2PG foi cristalizada de 25% p/v PEG 4K, 8% 2-propanol, 100 mM acetato pH 5,23, 2 mM DTT, 1 mM EDTA, 1 mM azida sódica. A mistura da gota foi de 3 µl de solução de proteína mais 3 µl de solução do poço.

Cristais apareceram depois de um mês. Um cristal de dimensões 0.75mmx0.25mmx0.10mm foi montado em tubo capilar por Tjaard Pjining, na Universidade de Groningen, Holanda e enviado à Universidade de Washington, Seattle, WA, USA, onde a coleta de dados foi feita.

6.3 Coleta de Dados de Difração de Raios-X

Os dados foram coletados no detetor de área XENTRONICS Nicolet/Siemens X100 usando radiação do gerador com anodo rotatório Rigaku RU200 na Universidade de Washington, Seattle, WA, USA. Os dados foram processados através do programa XENGEN (Software Copyright 1988, Genex Corp, Garthersburg, MD USA). Detalhes da cristalização, coleta e processamento dos dados podem ser acompanhados na Tabela 6.2.

O cristal difratou até 2,6 Å com boa completeza de dados. Os dados compreendidos na camada de resolução de 2,8 Å a 2,6 Å apresentaram um R_{sym} muito elevado (44,0%). De forma, apesar dos dados apresentarem boa completeza até 2,6 Å, somente dados máxima de 2,8 Å foram utilizados no refinamento. Dados até 2,8 Å de resolução têm 82,0% de completeza, sendo que a camada de 3,0 até 2,8 Å tem 77,7% de completeza. O grupo espacial foi determinado através de fotografias de precessão e confirmado posteriormente durante a solução da função de translação no processo de substituição molecular. Este grupo é $P2_12_12_1$. O volume da cela unitária é de $6,09 \times 10^5 \text{ Å}^3$ o que dá um V_m (Mattheus, 1968) igual a $2,81 \text{ Å}^3 \text{ Da}^{-1}$ para duas subunidades por unidade assimétrica e $5,62 \text{ Å}^3 \text{ Da}^{-1}$ para uma subunidade por unidade assimétrica. Baseado nestes resultados, ficou evidente que a unidade assimétrica compõe-se de duas subunidades ou um dímero.

¹ Veja Lista de Abreviações

Tabela 6.2 - Cristalização, coleta e processamento dos dados de difração da TIM de *Bacillus stearothermophilus*

<u>Cristal</u>	método:	Difusão de vapor, gota pendurada, temperatura ambiente		
	solução de proteína:	4 mg/ml em 5 mM Mes pH 6.5, 1 mM EDTA, 1 mM de azida sódica e 2 mM de 2PG		
	solução do poço:	8% 2-propanol, 100 mM tampão acetato pH 5.23, 2 mM DTT, 25% PEG 4 K, 1 mM EDTA, 1 mM azida sódica		
	gota:	3 µl solução de proteína, 3 µl solução do poço		
	dimensão:	0,75mmx0,25mmx0,10mm		
<u>Coleta de dados:</u>		Detetor de área Nicolet/Siemens X100		
<u>Processamento dos dados:</u>		XENGEN		
<u>Parâmetros de rede (em Å):</u>		a = 78.61	b = 108.42	c = 71.21
		Grupo espacial P2 ₁ 2 ₁ 2		
<u>Número de reflexões:</u>				
	coletadas	64784		
	únicas	16996		
<u>Completeza:</u>		85.3% (dados até 2.6 Å)		
		89.6% (dados até 2.8 Å)		
	Camada 3,0-2.8 Å	77.7%		
	Camada 2.8-2.6 Å	66.7%		
<u>Qualidade dos Dados:</u>				
	R _{sym}	9.3% (em intensidades - dados até 2.8 Å)		
	R _{sym} (3,0-2.8 Å)	22.8%		
	R _{sym} (2.8-2.6 Å)	44.0%		

6.4 Determinação da Estrutura

A estrutura foi determinada por Substituição Molecular (Molecular Replacement, veja Capítulo 2) usando a estrutura da TIM de *T. brucei* refinada até 1,83 Å de resolução (Wierenga *et al*, 1991b), através do programa XPLOR (Brünger, 1990a) (Tabela 6.3).

Figura 6.1a - Seção da função de rotação cruzada. Pico com solução $\theta_1=217.57^\circ$, $\theta_2=75.00^\circ$, $\theta_3=60.90^\circ$, seção $\theta_2=75.00^\circ$. Níveis dos contornos com intervalo de 1σ . Valor mínimo 1σ

A identidade entre a sequência da TIM de *Bacillus stearothermophilus* e as outras enzimas TIM de estruturas conhecidas encontra-se entre 38,9% e 43,3%. Desta forma a escolha do modelo para a substituição molecular foi feita devido à qualidade do modelo mais do que da identidade das sequências primárias.

A função rotação cruzada foi feita através do programa XPLOR usando a estrutura dimérica da TIM de *T. brucei* (41,5% de identidade). Diferentes vetores de Patterson como

também diferentes intervalos de resolução foram testados. Duas soluções mostraram-se frequentes em todos os testes ($\theta_1=217,57^\circ$, $\theta_2=75,00^\circ$, $\theta_3=60,90^\circ$ e $\theta_1=160,79^\circ$, $\theta_2=52,50^\circ$, $\theta_3=75,53^\circ$) apresentando o melhor resultado com raio de integração igual a 35 Å e intervalo de resolução de 8,0 a 3,5 Å. Estas soluções tinham 7,3 σ e 7,1 σ respectivamente acima do valor médio (Tabela 6.3), onde σ é o desvio quadrático médio da densidade eletrônica. Os próximos picos mais alto tinham 4,5 σ e 4,1 σ acima do valor médio, respectivamente. A seção dos picos obtidos podem ser vistas na Figura 6.1.

Figura 6.1b - Seção da função de rotação cruzada. Pico com solução $\theta_1=160.79^\circ$, $\theta_2=52.50^\circ$, $\theta_3=75.53^\circ$, seção $\theta_2=52.50^\circ$. Níveis dos contornos com intervalo de 1 σ . Valor mínimo 1 σ .

Tabela 6.3 - Determinação da estrutura da TIM de *Bacillus stearothermophilus* por substituição molecular usando a estrutura da TIM de *T. brucei* como modelo inicial

Substituição Molecular através do programa XPLOR					
Função de Rotação Cruzada:		Modelo inicial: TIM de <i>T. brucei</i> P1, a=b=c=128 Å Raio de integração: 35 Å Limites de resolução: 8,0-3,5 Å			
Soluções:		θ_1	θ_2	θ_3	n° of σ
	1	160,79°	52,50°	75,53°	7,1
	2*	217,57°	75,00°	60,90°	7,3
PC-Refinamento**:					
	1	160,31°	52,89°	75,61°	
	2	215,80°	76,01°	59,18°	
Função translação (usando a solução número 1 do PC-Refinamento)					14,1
* Solução relacionada a solução 1 por uma rotação de 180°					
** Veja texto e Capítulo 2 para maiores detalhes					

Estas duas soluções estão relacionadas por uma rotação de 180° (cálculo feito através de todos os ângulos equivalentes pelas operações de simetria do espaço de rotação segundo Narasinga Rao *et al.*, 1980).

Figura 6.2a - Coeficiente de correlação das soluções da função de rotação cruzada obtido através do PC-refinamento do programa XPLOR. Coeficiente de correlação da lista de soluções da função de rotação cruzada. As duas soluções se sobressaem em relação às outras obtidas pelo programa.

Figura 6.2b - Melhora do coeficiente de correlação das soluções da função de rotação cruzada através do PC-refinamento do programa XPLOR. Refinamento do dímero. Pode ser observado o aparecimento de mais duas soluções com alto coeficiente de correlação.

Figura 6.2c - Melhora do coeficiente de correlação das soluções da função rotação cruzada através do PC-refinamento do programa XPLOR. Refinamento dos monômeros. Pode ser observado o aparecimento de mais uma solução e a convergência dos valores da correlação das novas soluções para valores próximos aos valores observados para as duas primeiras soluções. Ver discussão no texto.

PC-refinamento (veja Capítulo 2 para maiores detalhes) do pacote XPLOR (Brünger, 1990b) foi aplicado à todas as soluções de substituição molecular, listadas pelo programa

XPLOR.

Como pode ser visto pela Figura 6.2 as duas soluções iniciais, mostradas acima, já apresentavam um valor da correlação que se sobressaía aos valores das correlações das outras soluções encontradas pelo programa. As duas soluções tiveram uma melhora acentuada na correlação quando o dímero foi refinado, com as subunidades tratadas independentemente.

Figura 6.3a - Seção da função translação da solução 1 da função de rotação cruzada (Tabela 6.3) para o maior pico. 14.1σ acima da media. Este pico está isolado nesta seção e o próximo pico mais alto encontra-se noutra seção com apenas 4.4σ acima do valor médio.

Os diferentes picos que tiveram uma melhora no fator de correlação são nada mais que convergência para as soluções já obtidas inicialmente, sendo que chegaram a convergir para a solução estando a mais de 10° deslocadas da posição das soluções iniciais (Figura 6.2b e 6.2c).

Figura 6.3b - Seção da função translação da solução 2 da função de rotação cruzada (Tabela 6.3) para o maior pico. 14.1σ acima da média. Este pico está isolado nesta seção e o próximo pico mais alto encontra-se noutra seção com apenas 4.4σ acima do valor médio.

Isto mostra o poder do PC-refinamento para o resultado da função rotação cruzada. Desta forma não faz diferença qual das duas soluções seja usada na função translação pois são

equivalentes devido ao fato do modelo usado na procura ser um dímero. Em princípio as duas soluções poderiam ser usadas para confirmar o resultado mas como já foi verificado a relação de 180° entre as duas soluções, escolheu-se apenas uma delas para continuar com a solução do problema de substituição molecular. Para uma verificação mais criteriosa o segundo pico foi utilizado na solução da função translação mostrando que os dois apresentam a mesma solução e somente uma solução foi escolhida para o refinamento. Um resumo da solução por substituição molecular pode ser acompanhado na Tabela 6.3.

A função de translação foi feita no grupo espacial $P2_12_12$ que foi determinado como descrito na seção 6.3, através do programa XPLOR (Brünger, 1990a) usando a solução 1 do resultado da função rotação cruzada (Tabela 6.3). Somente uma solução foi obtida com 14.1σ acima da média. O segundo pico esteve a 4.4σ acima da média. A seção do pico mais alto pode ser vista na Figura 6.3.

Neste estágio todo o modelo foi transformado em poli-alanina. Como pode ser visto na Tabela 6.1 a similaridade entre TIM de *Bacillus stearothermophilus* e *T. brucei* é de apenas 41.5%. Uma tentativa de refinamento usando o modelo original mostrou altos contatos entre átomos de diferentes subunidades. Isto pode ter sido devido ao diferente empacotamento pois o modelo usado da TIM de *T. brucei* pertence ao grupo espacial $P2_12_12$, diferente da TIM de *B. stearothermophilus*, e também devido à estrutura em volta flexível que ocupa diferentes conformações nas duas estruturas. Devido a dúvida inicial em relação à conformação da curva flexível, esta foi mantida como no modelo inicial. Maiores detalhes do refinamento desta estrutura em volta serão descrito nas próximas seções.

Este modelo de poli-alanina foi usado para um primeiro estágio de refinamento de corpo rígido usando o programa XPLOR (Brünger, 1990a). Primeiramente o dímero foi refinado e subsequentemente cada subunidade foi tratada independentemente. Este modelo produziu um fator de concordância R inicial igual a 49.2% (resolução 11,0-5,0 Å) o qual convergiu para 47,8%. Novo refinamento de corpo rígido foi feito em diferente intervalo de resolução (6,0-3,0 Å) com o fator de concordância R iniciando em 52,9% e convergindo para 51,8%. Este modelo foi submetido ao refinamento com minimização de energia o qual iniciou em um fator R igual a 52,7% (8,5-3,0 Å) e convergindo para 42.7% depois de 120 ciclos de refinamento. Um mapa de densidade eletrônica foi então calculado sendo que os fatores de estrutura e fases foram obtidos através do programa XPLOR (Brünger, 1990a), escalonados por KBRANI (programa para escalonamento, parte do pacote BIOMOL de programas para tratamento de dados do Departamento de Química, da Universidade de Groningen, Holanda) e ponderado através de SIGMAA (Read, 1986) (ver Capítulo 2 para detalhes).

Mapas diferença assim como $2F_o - F_c$ foram obtidos. Estes mapas foram analisados através do programa O (Jones *et al.*, 1991) numa estação gráfica Silicon Graphics. A qualidade do mapa mostrou várias cadeias laterais mas a densidade eletrônica apresentou-se muito pobre em várias regiões. Devido à qualidade do mapa e devido ao fato desta estrutura ter um eixo de simetria não-cristalográfico de ordem 2, por ser um homo-dímero, decidiu-se aplicar as técnicas de promediação da densidade eletrônica e nivelamento do solvente através do programa DEMON (**D**ensity **M**odification, Vellieux, 1993).

Tabela 6.4 - Refinamento do modelo da TIM de *B. stearothermophilus* por XPLOR

<u>Refinamento de corpo rígido do modelo de poli-alanina:</u>			
Resolução: 11.0 - 5.0 Å		Fator R	
	Inicial		Final
Dímero	49.2%		48.8%
Monômeros	48.8%		47.8%
Resolução: 6.0 - 3.0 Å			
Dímero	52.9%		52.2%
Monômeros	52.2%		51.8%
<u>Refinamento minimização de energia:</u>			
Resolução: 8.5 - 3.0 Å			
	52.7%		42.7%
<u>Refinamento de corpo rígido do dímero</u>			
(subunidade 2 está relacionada à subunidade 1 por $\kappa=180^\circ$)			
Resolução: 11.0-5.0 Å			
	45.1%		45.0%
Resolução: 6.0-3.0 Å			
	47.9%		47.4%

A simetria não-cristalográfica entre as subunidades é própria, isto é, rotação pura sem translação. Como o refinamento inicial permitiu que as subunidades fossem refinadas independentemente, então neste estágio a subunidade 1 foi usada para gerar a subunidade 2 através de uma rotação de 180° em torno do eixo de simetria de ordem 2 não-cristalográfico. Desta forma ficou estabelecido que entre as duas subunidades havia uma simetria não-cristalográfica de ordem 2 e própria (sem translação na direção do eixo de rotação). O programa SUPPOS do pacote BIOMOL (pacote de programas do Departamento de Química da Universidade Estadual de Groningen, Holanda) foi usado para calcular a matriz de transformação. Este novo modelo foi então submetido ao refinamento de corpo rígido apenas do

dímero para evitar que houvesse refinamento da simetria não-cristalográfica.

O fator R não mostrou significativa alteração [de $R=45,1\%$ ($11,0-5,0 \text{ \AA}$) para $R=45,0\%$ ($11,0-5,0 \text{ \AA}$) e de $R=47,9\%$ ($6,0-3,0 \text{ \AA}$) para $R=47,4\%$ ($6,0-3,0 \text{ \AA}$)]. Este foi então o modelo usado para a promediação da densidade eletrônica e nivelamento do solvente. Devido à baixa ordem da simetria não-cristalográfica (ordem 2) não se esperava expressiva melhora da densidade eletrônica mas devido a qualidade do mapa de densidade eletrônica qualquer melhora seria de grande importância.

Naturalmente, a remoção das cadeias laterais de todos os resíduos, mesmo os que eram conservados, contribuiu para a piora do mapa. Devido a este fato, o resultado desta promediação e nivelamento do solvente mostrariam o poder da técnica como também do programa usado para tanto. Todos os detalhes deste estágio do refinamento podem ser vistos na Tabela 6.4.

6.5 Promediação da Densidade Eletrônica e Nivelamento do Solvente

Com a colaboração do Dr. Fred Vellieux, a promediação da densidade eletrônica e o nivelamento do solvente foram aplicados. Como a simetria entre as duas subunidades é própria (rotação pura sem translação) este foi o procedimento seguido na aplicação do programa de modificação da densidade eletrônica. Seis ciclos de promediação e nivelamento do solvente foram aplicados na faixa de resolução $15,0-3,2 \text{ \AA}$. Extensão da resolução de $3,20 \text{ \AA}$ para $2,81 \text{ \AA}$ foi feita em passos de $0,03 \text{ \AA}$ com apenas um ciclo de promediação e nivelamento do solvente entre cada passo. Este passo foi determinado baseado na função G de Rossmann e Blow (Rossmann e Blow, 1962, vide Capítulo 2 para maiores detalhes). A densidade obtida depois da extensão foi submetida a seis ciclos de promediação e nivelamento do solvente na faixa de resolução $15,0-2,81 \text{ \AA}$. Um resumo da promediação da densidade eletrônica e nivelamento do solvente podem ser vistos na Tabela 6.5.

Neste estágio, mapas ponderados SIGMAA $m(2Fo-DFc)$ e $m(Fo-Fc)$ (Read, 1986) foram calculados e uma análise comparativa dos mapas antes e depois da modificação da densidade eletrônica por promediação e nivelamento do solvente foi feita através do programa O (Jones *et al*, 1991) o que mostrou que muitas das cadeias laterais apresentaram clara densidade eletrônica como pode ser visto na Figura 6.4.

Ficou clara também a evolução de algumas cadeias laterais que, devido à predominância das fases do modelo anterior, apresentavam densidade eletrônica da cadeia lateral do modelo da

TIM de *T. brucei* e depois da promediação e nivelamento do solvente apresentaram densidade da cadeia lateral do modelo da TIM de *B. stearothermophilus*.

Tabela 6.5 - Promediação e nivelamento do solvente do modelo da TIM de *B. stearothermophilus*

<u>Promediação e nivelamento do solvente (DEMON)</u>	
6 ciclos no intervalo de resolução 15.0-3.2 Å	
Extensão em 13 ciclos de 3.2 Å a 2.81 Å em passos de 0.03 Å	
6 ciclos no intervalo de resolução 15.0-2.81 Å	
<u>Modificação do modelo:</u> 60% das cadeias laterais foram visíveis e adicionadas	
<u>Refinamento de corpo rígido, minimização de energia e dinâmica molecular de resfriamento lento</u>	
Fator de temperatura 30 Å ² , restrições não cristalográficas	
R = 33.5% (8.0-2.8 Å)	
<u>Modificação do modelo:</u> 25% das cadeias laterais foram visíveis e adicionadas, totalizando 85% do modelo	
<u>Refinamento de corpo rígido, minimização de energia e dinâmica molecular de resfriamento lento</u>	
Fator de temperatura 30 Å ² , subunidades tratadas independentemente	
R = 30.5% (8.0-2.8 Å)	
<u>Refinamento do fator de temperatura B e minimização de energia</u>	
Subunidades tratadas independentes	
R = 26.2% (10-2.8 Å)	

A construção do modelo da TIM de *B. stearothermophilus* foi feita através da utilização do alinhamento das sequências da TIM de *T. brucei* e de *B. stearothermophilus* além da utilização dos mapas de densidade eletrônica antes e depois da modificação da densidade eletrônica já que algumas partes da superfície sofreram piora da densidade eletrônica provavelmente pela dificuldade em definir um envelope ao redor do modelo que englobe todos os resíduos da superfície.

Todos os mapas e modelos foram analisados através do programa O (Jones *et al.*, 1991) em uma estação gráfica Silicon Graphics. As mutações foram feitas usando a opção MUTATE em O (Jones *et al.*, 1991). As cadeias laterais adicionadas foram então ajustadas na densidade eletrônica através da opção TORSION e refinadas através da opção REFINE. Em alguns casos a opção LEGO LOOP foi usada para ajustar uma região da cadeia principal no mapa de densidade eletrônica, principalmente em estruturas em volta, onde a flexibilidade e diferença entre os modelos eram maiores.

Figura 6.4a - Densidade eletrônica do modelo (acima) antes da promediação e nivelamento do solvente e (abaixo) depois, mostrando comparativamente a melhora do mapa de densidade eletrônica. Contorno de 1.5σ .

Figura 6.4b - Densidade eletrônica do modelo final.

A estrutura em volta flexível apresentou uma péssima densidade eletrônica sendo impossível definir a cadeia principal. Através deste procedimento 60% das cadeias laterais puderam ser ajustadas no mapa de densidade eletrônica, 20% a mais que a homologia das estruturas. Somente a subunidade 1 foi submetida às mutações sendo que análises foram feitas também na subunidade 2.

Através da simetria não-cristalográfica, a subunidade 2, com as modificações, foi gerada a partir da subunidade 1. Os resíduos que não mostraram clara densidade eletrônica foram mantidos como Ala.

Neste estágio, refinamento de corpo rígido, minimização de energia e dinâmica molecular de resfriamento lento (slow cooling protocol, Brünger *et al.*, 1990) foram aplicados ao modelo mantendo restrições de simetria não-cristalográfica entre as subunidades através da opção NCS em XPLOR (Brünger, 1990a) e um fator de temperatura igual 30 Å^2 . O fator R convergiu para 33,5% ($8,0\text{-}2,8 \text{ Å}$). Um novo mapa foi calculado e mais 25% das cadeias laterais puderam ser adicionadas ao modelo. No entanto as estruturas em volta ainda apresentavam pobre densidade eletrônica.

Um novo estágio de refinamento por corpo rígido, minimização de energia e dinâmica

molecular de resfriamento lento (Brünger *et al.*, 1990, veja Capítulo 2) foram aplicados com a diferença de não impor restrições não-cristalográficas entre as subunidades mas ainda mantendo um fator de temperatura constante para todos os átomos de 30 Å². O próximo estágio foi o refinamento do fator de temperatura individual para cada átomo o que convergiu o fator R para 26,2% (10,0-2,8 Å). Um novo mapa mostrou uma melhora apreciável mas ainda as estruturas em volta na superfície mostraram pobre densidade eletrônica principalmente a estrutura em volta flexível em ambas as subunidades.

O resultado que seria esperado com a dinâmica molecular de resfriamento lento no refinamento de levar as partes flexíveis como as estruturas em volta para as posições que deveriam ocupar no modelo da TIM de *B. stearothermophilus* não ocorreu devido ao uso do peso dos termos da energia de raios-X no programa XPLOR dividido por 2,5, que deu um peso maior a geometria, impedindo que as estruturas em volta se movessem para posições que satisfizessem mais o mapa de densidade eletrônica (para maiores detalhes da função minimizada na dinâmica molecular veja Capítulo 2). Isto pode justificar a qualidade do mapa nas estruturas em volta e principalmente na flexível, a qual moveu-se muito pouco em direção à conformação fechada.

Mapas de omissão (veja Capítulo 2) foram então calculados para todas as estruturas em volta com pouca definição na densidade eletrônica e as cadeias principais foram ajustadas. O resultado mais expressivo foi o da estrutura em volta flexível que moveu-se da conformação "aberta" do modelo da substituição molecular aparecendo no mapa de omissão claramente na conformação "fechada" (Figura 6.5). Gly 172 neste ajuste moveu-se mais de 8 Å. Alguns ciclos a mais de refinamento de minimização de energia e fator de temperatura individual intercalados com reconstrução do modelo convergiu o fator R para 18,7% (12119 reflexões no intervalo de 8,0 a 2,8 Å).

Neste estágio ficou clara a identificação, no sítio ativo, densidade eletrônica para o inibidor 2-fosfoglicolato (2PG) em ambas as subunidades, sendo visível ao nível de 3σ para os átomos de carbono e oxigênio e ao nível de até 7σ para o átomo de fósforo e grupo fosfato. Depois da inclusão das moléculas dos inibidores nos dois sítios ativos, mais alguns ciclos de refinamento foram feitos, intercalados com mais alguns ajustes finais das cadeias laterais em geral, e principalmente ao redor das moléculas dos inibidores. O fator de concordância R convergiu para 17,6% com excelente geometria. A estatística do modelo refinado pode ser acompanhada na Tabela 6.6. Nenhuma molécula de água foi introduzida devido à resolução dos dados, apesar de densidades expressivas terem sido observadas ao redor do sítio ativo.

Figura 6.5 - Densidade eletrônica do modelo da TIM de *B. stearothermophilus*, após a omissão dos resíduos da estrutura em volta flexível (acima) e densidade eletrônica do modelo final (abaixo). Contorno de $1,5 \sigma$.

Tabela 6.6 - Estatística do refinamento da estrutura da TIM do *B. stearothermophilus*

Parâmetro	Valor
Fator R	17,6%
Limite de resolução	8,0-2,8 Å
Refleções	12119
Modelo	
Átomos da proteína	3769
Átomos do ligante	18
Desvio RMS do ideal para:	
Comprimento das ligações covalentes	0,010 Å
Ângulos de átomos ligados	1,647°
Ângulos diédricos	24,298°
Ângulos impróprios	1,75°

6.6 Qualidade da Estrutura

As estatísticas do refinamento da estrutura da TIM de *B. stearothermophilus* podem ser vistas na Tabela 6.6. Existem oito resíduos fora das regiões permitidas no diagrama Ramachandran (Figura 6.5). Resíduos Thr 64, Met 65, Ala 102, Ser212, Met 365 e Ser 366 estão muito próximos das regiões permitidas (numeração dos resíduos de 1 a 252 referem-se à "primeira subunidade" e 301 a 552 referem-se à "segunda subunidade"). Em contraste Lys 10 e Lys 310 encontram-se longe de qualquer combinação ϕ , ψ permitida. No entanto estas lisinas têm a mesma combinação ϕ , ψ em todas as estruturas de TIM já resolvidas até então e baixo fator de temperatura. Também é bem sabido que esta lisina é essencial para a atividade da TIM (Straus *et al.*, 1985). Lys 10 (e Lys 310) faz uma ligação de hidrogênio muito forte com o grupo carboxílico do inibidor 2PG (Figura 6.6).

Resíduos Ala 102, Ser 212, Ala 402, Ser 512 e Arg 550 estão localizados ou dentro ou próximos à região permitida para hélice de mão esquerda.

Figura 6.6 - Diagrama de Ramachandran das duas subunidades da TIM de *B. stearothermophilus* mostrando os resíduos que encontram combinações de ϕ , ψ fora das regiões permitidas.

Resíduos Ala 102 e Ala 402 estão a cerca de 20 Å do sítio ativo e estão expostos ao solvente, longe de qualquer contato cristalino. Não há razão clara para a combinação de hélice de mão esquerda para estes resíduos. Arg 550 também está na região de hélice de mão esquerda no diagrama de Ramachandran mas a densidade eletrônica deste resíduo próximo ao C-terminal da cadeia polipeptídica não é bem definida no mapa.

Os resíduos da região C-terminal, Glu 252 e Glu 552, não têm densidade eletrônica e não foram incluídos no modelo. Cadeias laterais dos resíduos Lys 175, Gln 225, His 251 da subunidade 1 e Gln 320 e Glu 538 da subunidade 2 não têm densidade eletrônica para as cadeias laterais e foram mantidos no modelo como Ala.

Figura 6.7 - Densidade eletrônica do inibidor 2PG no sítio ativo, contorno de $1,5 \sigma$ (acima) e interações observadas entre o inibidor 2PG e os resíduos do sítio ativo da subunidade 1 da TIM de *B. stearothermophilus* (abaixo).

Os resíduos equivalentes nas outras subunidades têm clara densidade eletrônica devido ao contato cristalino. Gln 20 é o único resíduo que está exposto ao solvente e que não faz nenhuma interação com outros átomos e ainda sim é visível na densidade eletrônica.

6.7 Características Estruturais da TIM de *B. stearothermophilus*

Nesta seção, características da estrutura da TIM de *B. stearothermophilus* e comparação com as estruturas de TIM conhecidas serão discutidas.

6.7.1 Comparação entre os Monômeros

A estrutura da subunidade 1 da TIM de *B. stearothermophilus* é mostrada na Figura 6.8 e é muito similar ao de outras estrutura de TIM já reportadas. Diferenças menores ocorrem nas regiões das estruturas em volta como nos seguintes casos: (i) estrutura em volta definida pelos resíduos 25 a 36, (ii) estrutura em volta 5, a qual compreende resíduos 127 a 137, que inclui também uma pequena hélice e (iii) resíduos 152 a 155, que são regiões não conservadas entre as sequências de TIM (Tabela 6.1). A estrutura em volta flexível (estrutura em volta 6) costuma apresentar diferentes conformações devido à diferenças nas condições de cristalização e presença ou ausência de inibidores. Estruturas que já foram resolvidas na presença de inibidores mostram a estrutura em volta flexível na conformação "fechada". Aquelas que foram ou cristalizada na ausência de inibidores ou onde a estrutura em volta flexível estava envolvida em contato cristalino no cristal mostram-na na conformação "aberta" (veja Capítulo 4 e 5 desta tese para maiores discussões da estrutura em volta flexível da TIM).

Os dois monômeros refinados independentemente apresentam uma diferença de 0,44 Å (desvio padrão médio) quando 1003 átomos equivalentes da cadeia principal são superpostos e 0,29 Å quando somente 972 átomos equivalentes da cadeia principal são comparados (31 átomos da cadeia principal que apresentam um deslocamento maior que 3 vezes o desvio padrão médio total foram removidos) (Figura 6.9).

Figura 6.8 - Estrutura do modelo da subunidade 1 da TIM de *B. stearothermophilus* feita através do programa MOLSCRIPT (Kraulis, 1991) mostrando em detalhe os resíduos do sítio ativo, o inibidor 2PG e a estrutura em volta flexível.

As regiões que diferem mais que 1,5 Å são ou em diferentes contatos cristalinos ou na região do C-terminal. O C α do resíduo Asp 54 está a 1,9 Å distante do resíduo equivalente da outra subunidade, Asp 354. O resíduo Asp 54 faz um extensivo contato cristalino com a subunidade 2 relacionada por uma operação de simetria através de várias ligações de hidrogênio bem definidas (ligações de hidrogênio com resíduos Arg 495, His 488 e Ser 491) e fatores de temperatura B relativamente baixos para os oxigênios O δ 1 e O δ 2 (27,9 Å² e 19,2 Å² respectivamente). Por outro lado O δ 2 do Asp 354 na outra subunidade apresenta apenas uma ligação de hidrogênio com N ζ da Lys326 da mesma subunidade e o outro oxigênio aponta para o solvente. Também os valores do fator de temperatura B para O δ 1 e O δ 2 são significativamente

altos ($65,9 \text{ \AA}^2$ e $66,9^2 \text{ \AA}$ respectivamente). Os resíduos da região C-terminal diferem por mais de $4,5 \text{ \AA}$, refletindo sua flexibilidade relativamente alta, resultando em densidade eletrônica bastante fraca.

Figura 6.9 - Gráfico do desvio entre as duas subunidades 1 e 2 quando superpostas. Diferenças maiores são visíveis na região dos resíduos Asp 54, Asp 354 e C-terminal.

O valor médio dos fatores de temperatura da cadeia principal para as duas subunidades são $18,9 \text{ \AA}^2$ e $15,9 \text{ \AA}^2$ respectivamente e os valores médios para os fatores de temperatura das cadeias laterais para as duas subunidades são $23,0$ e $19,9 \text{ \AA}^2$ respectivamente. Os fatores de temperatura das duas subunidades, cadeia principal e cadeia lateral podem ser vistos na Figura 6.10.

Na Figura 6.10 há uma significativa diferença no valor médio do fator de temperatura das regiões N-terminal e C-terminal da subunidade 1. O valor médio dos fatores de temperatura dos 120 resíduos da região N-terminal é $13,8 \text{ \AA}^2$ enquanto o valor médio dos fatores de temperatura da região C-terminal é $23,6 \text{ \AA}^2$. Esta diferença não é observada na subunidade 2. As regiões N-terminal das duas subunidades estão envolvidas em contatos inter-subunidades, portanto apresentando baixos fatores de temperatura. Pela análise gráfica a região C-terminal da subunidade 1 mostra baixo número de contatos cristalinos, a qual apresenta a mais alta área exposta ao solvente da proteína nesta forma cristalina. Portanto esta poderia ser a explicação para esta diferença no fator de temperatura da subunidade 1.

Figura 6.10a - Gráfico dos fatores de temperatura em função dos resíduos das cadeias principal e lateral da subunidade 1.

Figura 6.10b - Gráfico dos fatores de temperatura em função dos resíduos das cadeias principal e lateral da subunidade 2.

Os dois monômeros estão relacionados por um eixo de simetria não-cristalográfico de

ordem 2 quase perfeito, com um ângulo de rotação de $179,2^\circ$ sem translação. Há três inserções na sequência da TIM de *B. stearothermophilus* quando comparada a sequência da TIM de *T. brucei*, modelo este usado na substituição molecular. Todas as três inserções são em regiões de estruturas em volta. Duas inserções foram bastante fáceis de serem modeladas na densidade eletrônica após o mapa modificado por promediação e nivelamento do solvente mas a última inserção só foi possível de ser incluída mais tarde no refinamento, através da análise baseada no mapa de omissão.

6.7.2 Sítio Ativo e Ligação do Inibidor 2PG

Os sítios ativos das duas subunidades da TIM de *B. stearothermophilus* contêm o inibidor 2PG (Figura 6.7). Os resíduos Lys 10 e His 94 do sítio ativo estão bem posicionados para executar a reação enzimática. O terceiro resíduo essencial para a atividade da TIM, Glu 166, adota a conformação conhecida como "swung out" ($\chi_1 = -55^\circ$ na subunidade 1 e $\chi_1 = -58^\circ$ na subunidade 2) em ambas as subunidades, a qual é reportada como não sendo a conformação requerida para catálise (Wierenga *et al.*, 1991b). A estrutura em volta flexível adota a conformação fechada ("closed") em ambas as subunidades em acordo com resultados prévios (Lolis *et al.*, 1990, Wierenga *et al.*, 1991b).

A posição e interação da molécula de 2PG é muito similar em ambas as subunidades, e também com aquelas observadas nas estruturas de seus complexos previamente reportadas (Lolis *et al.*, 1990, Mande *et al.*, 1994) (Tabela 6.7 e Figura 6.7).

Os oxigênios do fosfato fazem ligações de hidrogênio com o grupo amino da cadeia principal dos resíduos Gly 172, Ser 212, Gly 233 e Gly 234 (Figura 6.7). Como mencionado anteriormente na seção 6.6, Ser 212 tem a combinação ϕ , ψ na região de hélice de mão esquerda no diagrama de Ramachandran. O grupo amino do resíduo Ser 212 (e Ser 512 na subunidade 2) faz também ligações de hidrogênio com o grupo carbonil do resíduo Ala 170. A Ser equivalente da TIM de levedura em complexo com 2PG (Lolis *et al.*, 1990), subunidade 2 da TIM de *T. brucei* em complexo com 2-fosfoglicerato (Noble *et al.*, 1991a) e glicerol-3-fosfato (Wierenga *et al.*, 1991b) tem a mesma combinação ϕ , ψ . Todas estas estruturas acima mostram a mesma ligação de hidrogênio com o inibidor mas somente a estrutura da TIM de *T. brucei* em complexo com 2-fosfoglicerato mostra a ligação de hidrogênio com a correspondente Ala 170 como observado na estrutura da TIM de *B. stearothermophilus*.

Tabela 6.7 - Relação de átomos da molécula do inibidor e da TIM de *B. stearothermophilus* que interagem e as respectivas distâncias em Å.

	2PG	Resíduo	Distance (Å)
Subunidade 1			
	O11	His 94 NE2	2,98
	O11	Lys 10 NZ	2,76
	O12	His 94 NE2	2,89
	O12	Glu 166 OE1	2,77
	O13	Gly 233 N	3,28
	O14	Ser 212 N	2,91
	O14	Gly 172 N	2,66
	O15	Gly 234 N	3,00
	O16	Gly 233 N	2,97
Subunidade 2			
	O21	His 94 NE2	2,82
	O21	Lys 10 NZ	2,70
	O22	His 94 NE2	2,99
	O22	Glu 166 OE1	3,10
	O23	Gly 233 N	3,31
	O24	Ser 212 N	2,79
	O24	Gly 172 N	2,87
	O25	Gly 234 N	3,66
	O26	Gly 233 N	3,19

Aparentemente esta é a conformação requerida para a Ser 211 na ligação do inibidor. O grupo carboxílico da cadeia principal do resíduo Gly 211 (Gly 210 na sequência da levedura) está apontando para o grupo hidroxila da Tyr 209 (Tyr 208 na sequência da levedura) fazendo uma ligação de hidrogênio. Também o grupo hidroxila da Tyr 209 faz uma ligação de hidrogênio com o grupo amino da cadeia principal da Ser 177 (a qual é Ala 176 na TIM de levedura). Ambas as interações citadas são observadas na TIM de levedura em complexo com 2PG (como já mencionado), subunidade 2 da TIM de *T. brucei* em complexo com glicerol-3-fosfato, sulfato e 2-fosfoglicerato. Por outro lado estas interações não são observadas nas estruturas de TIM humana (Mande *et al.*, 1994; em ambas as subunidades, a despeito do fato de a subunidade 2 estar em complexo com 2PG), *E. coli* (Noble *et al.*, 1993) e galinha (Banner *et al.*, 1975) e subunidade 1 da TIM de *T. brucei* em complexo com 2-fosfoglicerato (Noble *et al.*, 1991a), sulfato e glicerol-3-fosfato (Wierenga *et al.*, 1991b). Importante salientar aqui que nas subunidades 1 destas estruturas citadas acima não se observa a ligação do inibidor devido ao contato cristalino que a estrutura em volta flexível da subunidade 1 faz com subunidade

relacionada por operação de simetria. No caso da estrutura da TIM humana apesar do cristal ter sido crescido na presença do inibidor, este foi montado em um líquido mãe sem o inibidor, o que provavelmente afetou o estado de ligação do mesmo (comunicação pessoal com Dr. Shekhar Mande). Em todas as estruturas nas quais Ser 211 não está na mesma conformação como na TIM de *B. stearothermophilus*, Tyr 209 não tem parceiro para a ligação de hidrogênio.

O átomo N ζ do resíduo Lys 10 do sítio ativo faz uma ligação de hidrogênio com um dos oxigênios do grupo carboxílico da molécula de 2PG. O átomo Ne $_2$ da His 94 faz duas ligações de hidrogênio com ambos os oxigênios do grupo carboxílico da molécula de 2PG e Oe $_2$ do Glu 166 faz uma ligação de hidrogênio com o oxigênio carboxílico da molécula 2PG.

Resíduo Gly 172 é o único aminoácido da estrutura em volta flexível que faz uma ligação de hidrogênio com a molécula 2PG. O C α da Gly 172 move cerca de 8 Å de sua posição da conformação "aberta" no modelo de substituição molecular da TIM de *T. brucei* para a conformação "fechada" no modelo aqui determinado e refinado da TIM de *B. stearothermophilus*. Outros contatos entre a estrutura em volta flexível e a proteína envolvem somente interações de van der Waals, mais extensivamente pelo resíduo Ile 171.

Também a estabilidade do grupo fosfato da molécula de 2PG é fornecida pelo dipolo de uma pequena hélice 3_{10} (Hol *et al.*, 1978), cujo N-terminal aponta diretamente para o fosfato.

O oxigênio carboxílico dos resíduos Gly 211 e Gly 511 apontam em direção contrária ao do fosfato da molécula de 2PG, em acordo com o reportado por Lolis *et al.* (Lolis *et al.*, 1990) para a estrutura da TIM de levedura em complexo com 2PG, que é a direção oposta da estrutura da TIM humana em complexo com 2PG (Mande *et al.*, 1994) e TIM de *T. brucei* em complexo com sulfato (Wierenga *et al.*, 1991b). Este oxigênio carboxílico da Gly 211 (e 511) aponta para o grupo hidroxila da Tyr 209 na estrutura da TIM de *B. stearothermophilus*, a qual está em distância de ligação de hidrogênio. A conformação que este resíduo deve adotar durante a isomerização do substrato ainda permanece desconhecida. Por outro lado a conformação do oxigênio carboxílico das Gly 211 e 511 podem estar na conformação correta quando a enzima TIM encontra-se em complexo com o substrato pois a estrutura da TIM humana apresenta fraca ocupação para o inibidor, que pode afetar a orientação deste grupo. Também a estrutura da TIM de *T. brucei* em complexo com sulfato não é similar ao complexo com o inibidor 2PG, não sendo então capaz de induzir a conformação observada para o oxigênio carboxílico da Gly 211 (e 511).

6.7.3 Comparação da TIM de *B. stearothermophilus* com as outras TIM de Estruturas Conhecidas

No geral a estrutura da TIM de *B. stearothermophilus* é bastante similar à todas as cinco estruturas de TIM conhecidas (humana, galinha, *T. brucei*, *E. coli* e levedura). Diferenças estão presentes em regiões de estruturas em volta. O alinhamento das sequências das TIM conhecidas pode ser visto na Tabela 6.1. Uma lista das estruturas de TIM reportadas pode ser vista na Tabela 6.8.

Tabela 6.8 - Estruturas de TIM de organismos mesofílicos, descritas na literatura.

Fonte	Resolução (Å)	Fator R	Solvente	Inibidor	Protocolo de Refinamento
Humana	8,0-2,8	16,7%	-	9	XPLOR
Galinha	2,5	17,0%	-	-	-
<i>T. brucei</i>	6,0-1,83	18,3%	278	18	PROLSQ/TNT
Levedura	10,0-2,5	18,0%	24	18	Hendrickson and Konnert/XPLOR
<i>E. coli</i>	25,0-2,6	11,9%	172	-	XPLOR/TNT

Tabela 6.9a - Razão Arg/(Arg+Lys) para as enzimas TIM de estruturas conhecidas. Como pode ser visto, TIM de *B. stearothermophilus* apresenta a maior razão.

Fonte	R/(R+K)
Humana (hu)	0,286
Galinha (ch)	0,259
<i>T brucei</i> (tb)	0,370
Levedura (ye)	0,276
<i>E. coli</i> (ec)	0,310
<i>B. stearothermophilus</i> (bs)	0,478

Tabela 6.9b - Composição em porcentagem de aminoácidos por tipo nas seis seqüências de TIM de estruturas tridimensionais conhecidas.

	hu	ch	tb	ye	ec	bs
Hidrofóbicos	133 (53.6%)	133 (53.8%)	135 (54.2%)	128 (51.8%)	143 (56.3%)	139 (55.2%)
Carregados						
+ ¹	28 (11.3%)	29 (11.7%)	27 (10.8%)	29 (11.7%)	24 (9.4%)	23 (9.1%)
+ ²	32 (12.9%)	37 (15.0%)	32 (12.9%)	32 (13.0%)	32 (12.5%)	29 (11.5%)
-	29 (11.7%)	30 (12.1%)	21 (8.4%)	32 (13.0%)	31 (12.2%)	32 (12.7%)
1	-1	-1	+6	-3	-7	-9
2	+3	+7	+11	0	+1	-3
Polares	54 (21.8%)	47 (19.0%)	61 (24.5%)	55 (22.3%)	48 (18.8%)	52 (20.6%)

1 His is considerada sem carga, 2 His considerada carregada

Códigos: hu - humana, ch - galinha, tb - *T. brucei*, ye - levedura, ec - *E. coli*, bs - *B. stearothermophilus*

Tabela 6.9c - Composição em porcentagem de aminoácidos nas seis seqüências de TIM de estruturas conhecidas. Resultados relevantes estão em negrito.

	Hu	Ch	Tb	Ye	Ec	Bs
Gly	10,1	10,9	7,6	8,9	8,6	8,7
Ala	11,3	11,3	13,3	10,1	17,7	11,5
Val	10,1	8,9	10,0	10,5	7,5	9,9
Leu	6,1	6,9	6,8	7,7	6,7	7,5
Ile	6,1	6,9	7,6	6,1	7,5	6,4
Ser	4,8	5,3	7,2	6,5	4,3	5,2
Thr	5,7	4,1	4,8	4,9	3,5	4,8
Phe	3,2	3,2	3,2	4,5	2,4	3,6
Tyr	1,6	1,6	2,0	2,4	2,0	1,6
Trp	2,0	2,0	2,0	1,2	0,8	0,8
Lys	8,1	8,9	6,8	8,5	6,3	4,8
Arg	3,2	2,8	4,0	3,2	3,1	4,4
His	1,6	3,2	2,0	1,2	3,1	2,4
Asp	4,8	5,3	3,2	6,1	3,9	4,8
Glu	6,9	6,9	5,2	6,9	8,2	7,9
Asn	3,2	2,8	4,0	4,9	3,5	2,0
Gln	4,4	3,6	5,2	2,8	4,3	5,6
Cys	2,0	1,6	1,2	0,8	1,2	1,6
Met	0,8	0,8	0,8	0,0	2,8	1,6
Pro	4,0	2,8	2,8	2,8	2,8	5,2

Fatores que afetam a estabilidade térmica das proteínas têm sido objeto de numerosos estudos (para um revisão veja Ikai, 1980, Merker *et al.*, 1981, Vihinen, 1987, Menendez-Arias

and Argos, 1989, Zhang *et al.*, 1991, Eriksson *et al.*, 1992, Matthews, 1993, Vriend e Eijssink, 1993).

Uma discussão sobre estruturas de proteínas de organismos termofílicos e mesofílicos, as quais permitem uma comparação detalhada na procura de fatores de estabilidade térmica, não foi feita com um grande número de estruturas como no presente estudo, no qual cinco estruturas mesofílicas são conhecidas, uma das quais sendo próxima ao grupo termofílico (levedura) e uma termofílica. No entanto a comparação fica parcialmente prejudicada pela existência de apenas uma estrutura (a estrutura apresentada neste trabalho) de organismos termofílicos, sendo que este organismo é considerado atualmente como sendo de estabilidade térmica apenas moderada.

TIM de *B. stearothermophilus* é estável até a temperatura de 65°C. Atualmente, organismos que vivem em temperaturas mais altas são conhecidos (organismos de estabilidade térmica extrema), sobrevivendo a mais de 100°C (para uma revisão veja Adams, 1993). Por outro lado esta é a primeira oportunidade para se iniciar uma comparação deste tipo, o que pode servir como um padrão para novas investigações na comparação de estruturas da mesma proteína de organismos de extrema estabilidade térmica.

Portanto existe aqui uma desvantagem neste tipo de comparação devido ao fato do *B. stearothermophilus* ser um organismo de estabilidade térmica moderada, o que faz a análise muito difícil já que diferenças como uma simples ponte salina poderia justificar as diferenças de temperaturas de estabilidade observadas para este caso. Também a homologia entre todas as estruturas de TIM já conhecidas e a de *B. stearothermophilus* encontra-se ao redor de 40% o que permite uma vasta modificação de resíduos responsáveis pela estabilidade térmica. É claro que também limitações são aplicáveis na comparação estrutural, pois todas as estruturas de TIM foram determinadas por diferentes grupos, por diferentes métodos e refinada por diferentes protocolos e à diferentes resoluções (Tabela 6.8).

Desta forma, características de cada estrutura, como ligações de hidrogênio, que dependem consideravelmente do programa usado no refinamento, devem ser levadas em consideração no momento de uma comparação mais detalhada.

A composição de aminoácidos para todas as sequências de TIM com estruturas conhecidas pode ser vista na Tabela 6.9. A razão Arg/(Arg+Lys) foi referida como sendo elevada para proteínas com estabilidade térmica (Merker *et al.*, 1981). De todas as sequências conhecidas, a TIM de *B. stearothermophilus* apresenta a mais alta razão (Tabela 6.1 e 6.9a). No entanto esta razão é alta devido mais ao baixo número de Lys do que ao alto número de Arg (Tabelas 6.9b e 6.9c).

A análise de 74 sequências de proteínas de *B. stearothermophilus* mostrou um valor médio

de 0,44 para a razão $\text{Arg}/(\text{Arg}+\text{Lys})$, sendo que a mesma análise para 62 proteínas de *E. coli* mostrou um valor médio de 0,47 (Figura 6.11). Uma análise mais crítica do trabalho de Merker *et al.* (Merker *et al.*, 1980) mostra que a razão $\text{Arg}/(\text{Arg}+\text{Lys})$ não se sobressai significativamente nas proteínas de organismos mesofílicos quando comparadas com organismos termofílicos. Também o universo de sequências usado não foi significativo para uma generalização desta regra empírica.

Figura 6.11 - Análise de: (a) 62 sequências de proteínas de *E. coli* disponíveis no banco de sequências do programa GCG mostrando a distribuição da razão $\text{Arg}/(\text{Arg}+\text{Lys})$, a qual apresenta o valor médio de 0,47; (b) 74 sequências de proteínas de *B. stearothermophilus* disponíveis no banco de sequências do programa GCG mostrando a distribuição da razão $\text{Arg}/(\text{Arg}+\text{Lys})$, a qual apresenta o valor médio de 0,44. A escolha das proteínas foi feita para se ter um máximo de proteínas homólogas dentro da análise.

Figura 6.12 - Distribuição da razão Arg/(Arg+Lys) nas 20 seqüências de TIM conhecidas (todas as seqüências presentes na Tabela 6.1 e duas outras não incluídas no banco de dados, *B. subtilis* e *B. megaterium*). Esta distribuição mostra valor médio de 0,32, sendo aparente a diferença relativa à todas as razões das outras seqüências de TIM e a de *B. stearothermophilus*, que apresenta a razão de 0,478 (Tabela 6.9a).

Na Figura 6.12 pode ser vista a distribuição da razão Arg/(Arg+Lys) para as 20 seqüências de TIM conhecidas, sendo incluídas todas as seqüências de TIM utilizadas no alinhamento na Tabela 6.1 mais as duas outras não incluídas no banco de dados já reportadas, *B. megaterium* (Schlapfer e Zuber, 1992) e *B. subtilis* (Leyva-Vasquez e Setlow, 1994). Fica claro que a razão Arg/(Arg+Lys) na TIM de *B. stearothermophilus* sobressai às outras razões e este resultado é devido principalmente ao baixo número de Lys do que a um alto número de Arg quando comparada às outras seqüências de TIM, como pode ser visto na Tabela 6.9b e 6.9c

Uma descrição detalhada das interações de todas as Arg em ambas as subunidades pode ser vista na Tabela 6.10. Quase todas as Arg na TIM de *B. stearothermophilus* estão localizadas na superfície e fazendo ou ligações de hidrogênio com Asp ou Glu ou ao menos próximas destes resíduos, equilibrando desta forma a carga eletrônica. De 22 Arg em ambas as subunidades, 10 não fazem pontes salinas com Asp e Glu, 5 delas sem qualquer ligação de hidrogênio com átomos Ne, NH1 ou NH2, e somente 1 sem qualquer ligação de hidrogênio, mas encontra-se numa região próxima a Asp ou Glu. Também, somente Arg 190 (Arg 490) está completamente inacessível ao solvente na estrutura. Esta arginina faz ponte salina com Asp 228 e ligações de hidrogênio com os grupos carboxílicos dos resíduos Ala 202, Ile 205, e com Oδ1 da Gln 226. Há também interações hidrofóbicas com mais 6 resíduos (Tabela 6.10). Também a Arg 1 encontra-se muito próxima à Arg 190 a qual pode justificar a presença do Glu 203 próximo à Arg 1. Glu 547 de uma subunidade relacionada por simetria encontra-se próximo a Arg 190 provavelmente contribuindo

para a estabilização de todas as cargas nesta região.

Os fatores de temperatura para quase todas as argininas estão no intervalo de 2,8 a 46,0 Å². Somente Arg 195, Arg 219, Arg 250, Arg 495 e Arg 550 estão no limite superior dos valores para os fatores de temperatura. Todas estas argininas citadas têm apenas uma ligação de hidrogênio ou mesmo nenhuma.

Vários autores têm recentemente reportado uma estatística de interações de Arg em proteínas (Nandi *et al.*, 1993, Borders *et al.*, 1994, Flocco e Mowbray, 1994). Argininas na TIM de *B. stearothermophilus* estão em bom acordo com os resultados apresentados por Nandi *et al.*, 1993). A interação mais frequente é com o grupo carboxílico carregado negativamente (O⁻) o qual ocorre 12 vezes num total de 19 interações na subunidade 1 e 20 na subunidade 2. Todas as outras interações do grupo guanidínio da arginina são feitas com grupos carboxílicos de cadeia principal. A interação mais extensiva é feita pela Arg 190 a qual faz 3 ligações de hidrogênio com grupos carboxílicos de cadeia principal dos resíduos Ala 202, Ile 205 e Gln 226. Mais três argininas apresentam ligações de hidrogênio de um único tipo: Arg 97/Thr 374, Arg 98/Cys 125 e Arg 206/Lys 2. Todas estas interações são também encontradas na subunidade 2 com a adição de mais uma: Arg 348/Ala 352 (Tabela 6.10). Esta análise parece sugerir que Arg 190 pode ter uma função estrutural apresentando interações como as reportadas por Borders *et al.* (Borders *et al.*, 1994) mas esta arginina é conservada em todas as sequências de TIM, uma única exceção observada na TIM de *Moraxela*, onde é uma Lys (Tabela 6.1), a qual pode exercer também um papel estrutural como em alguns casos reportados por Borders *et al.* (1994).

TIM de *B. stearothermophilus* mostra a mesma interação descrita entre argininas e anéis aromáticos de cadeias laterais (Flocco e Mowbray, 1994), como Arg 97 e Phe 101, na qual o plano do grupo guanidínio da Arg 97 e o anel da Phe 101 estão fazendo um ângulo de aproximadamente 30° e uma forte interação de van der Waals com distâncias ao redor de 3,5 Å. Por outro lado esta interação não é observada entre Arg 133 e os resíduos Tyr 165 e Trp 169. O grupo OH da Tyr 165 está em distância de ligação de hidrogênio do NH₂ da Arg 133, não obedecendo, no entanto, a geometria de ligação de hidrogênio. O anel da Tyr 165 está quase no mesmo plano do grupo guanidínio da Arg 133. Já no caso da Trp 169, os planos estão quase perpendiculares entre si, fazendo interações de van der Waals com distâncias ao redor de 3,5 Å.

Tabela 6.10a.- Interações das Argininas da subunidade 1 na estrutura da TIM de *B. stearothermophilus*

Resíduo	Átomo	Acess. (Å ²)	Fator B (Å ²)	Lig. de H	Contatos
Arg1	NE NH1 NH2	5,4 ocluido 25,9	27,3 36,8 32,7	OD2 Asp228	Pro3,Arg190,Glu203, Ala204,Arg206,Asp228
Arg48	NE NH1 NH2	1,7 22,7 14,9	31,2 38,4 39,4	OD2 Asp347 OE1 Asp347 OE1 Gln351	Leu15,Val19,Phe45, Leu46,Val50,Gln51, Ala52,Asp347,Gln351
Arg97	NE NH1 NH2	ocluido 23,0 4,9	10,7 3,1 4,0	O Thr374 O Thr374 OE2 Glu376	Phe101,Glu103,Thr374 Gly375,Glu376
Arg98	NE NH1 NH2	7,0 5,2 6,0	8,3 11,6 12,5	O Cys125	Gly93,His94,Ser95,Glu96, Met100,Glu103,Asp105, Val108,Cys125,Gln145
Arg118	NE NH1 NH2	4,3 19,9 17,7	26,9 30,5 39,6		Asp69,Pro79,Val80, Lys83,Ala114,Ala115, Phe116,Leu120
Arg133	NE NH1 NH2	3,9 16,9 6,7	34,1 35,0 39,5	OE2 Glu128 OD1 Asp181 OE2 Glu128 OD2 Asp181	Glu128,Ser129,Leu130, Glu131,Ala135,Gly136, Thr138,Tyr165,Trp169, Asp181
Arg190	NE NH1 NH2	ocluido ocluido ocluido	19,6 27,1 29,8	OE1 Gln226 O Ala202 O Ile205 O Gln226 OD2 Asp228	Arg1,Gly187,His188, Val192,Val193,Ser194, Ala202,Ile205,Ile207, Gln226,Asp228
Arg195	NE NH1 NH2	11,5 19,8 21,9	71,0 82,5 75,9		Ser191,Val192,Val193, Phe197,Glu547*
Arg206	NE NH1 NH2	11,9 ocluido 12,3	16,7 30,9 33,4	O Lys2	Arg1,Lys2,Ile4,Ile36, Val162,Ile163,Asp228
Arg219	NE NH1 NH2	6,7 17,6 17,0	84,7 88,4 91,0	OD1 Asp220	Asn217,Phe221,Leu222, Ala223
Arg250	NE NH1 NH2	7,7 19,0 19,3	83,2 97,9 98,1		Glu34,Val35,Val246, Glu247,Ala248

* Relacionado por operação de simetria

Tabela 6.10b- Interações das Argininas da subunidade 2 na estrutura da TIM de *B. stearothermophilus*

Resíduo	Átomo	Acess. (Å ²)	Fator B (Å ²)	Lig. de H	Contatos
Arg301	NE NH1 NH2	5.7 1.4 17.6	29.1 39.4 32.5	OD2 Asp528	Arg490,Ala504,Ile505, Gln525,Asp528
Arg348	NE NH1 NH2	2.4 21.3 14.1	32.6 46.0 41.8	OD2 Asp47 OD1 Asp47 O Ala352	Asp47,Gln51,Leu315, Val319,Phe345,Leu346, Val350,Gln351,Ala352
Arg397	NE NH1 NH2	ocluido 17.9 17.8	23.6 2.0 4.7	O Thr74 O Thr74 OE2 Glu76	Thr74,Glu76,Thr364, His394,Ser395,Gln399, Phe401,Ala402,Glu403
Arg398	NE NH1 NH2	10.1 6.6 6.5	2.8 7.3 18.2	O Gly393	Gly393,His394,Ser395, Glu396,Met400,Glu403, Asp405,Val408,Cys425
Arg418	NE NH1 NH2	1.4 16.5 16.7	20.1 18.9 11.7	OD1 Asp369 OD2 Asp369	Asp369,Pro379,Val380, Lys383,Ala414,Ala415, Phe416,Leu420
Arg433	NE NH1 NH2	3.8 19.2 7.1	21.1 29.3 21.0	OE2 Glu428 OD1 Asp481 OD2 Asp481	Glu428,Ser429,Leu430, Glu431,Ala435,Gly436, Thr438,Tyr465,Asp481
Arg490	NE NH1 NH2	ocluido ocluido ocluido	8.3 4.4 7.9	OE1 Gln526 O Ala502 O Ile505 O Gln526 OD2 Asp528	Arg301,Cys486,Gly487, His488,Val492,Val493, Ser494,Ala502,Ile505, Ile507,Gln526,Asp528
Arg495	NE NH1 NH2	6.4 23.5 12.8	74.2 86.7 85.5	OD1 Asp54*	Asp54*,Ser491,Val492, Val493,Phe497
Arg506	NE NH1 NH2	10.6 ocluido 18.2	19.1 24.4 29.4	O Lys302	Lys302,Ile304,Ile336, Val462,Ile463,Asp528
Arg519	NE NH1 NH2	10.1 16.1 22.0	19.1 24.4 29.4		Asn517,Phe521,Leu522, Ala523,Ala548
Arg550	NE NH1 NH2	6.2 15.3 11.1	80.8 89.7 87.6		Glu334,Glu547,Ala548

* Relacionada por operação de simetria

Há na TIM de *B. stearothermophilus* duas argininas que não são observadas nas outras estruturas de TIM: Arg 48 e Arg 219. Arg 48 está fazendo uma ponte salina intersubunidade com Asp 347 (Tabela 6.10). Arg 219 está também fazendo uma ponte salina mas com o Asp 220, ambosexpostos ao solvente. Uma análise detalhada das mesmas regiões das pontes salinas Arg 48/Asp 347 na TIM de *B. stearothermophilus* foi feita para todas as outras estruturas de TIM. As cinco estruturas foram superpostas sobre a TIM de *B. stearothermophilus* através do programa O (Jones *et al.*, 1991). As estruturas da TIM humana, galinha e levedura mostram uma ponte salina na mesma região, entre os resíduos Lys 18 na sequência da TIM humana (Lys 17 nas sequências de levedura e de galinha) e Asp 349 (Asp 346 na sequência de levedura, Asp 348 na sequência de galinha). A função da Arg 48 na TIM de *B. stearothermophilus* é substituída pela Arg 18 em *E. coli*, que faz uma ponte salina com Asp 350. Somente a TIM de *T. brucei* não mostra qualquer ponte salina intersubunidade nesta região. Mesmo que quase todas as argininas na TIM de *B. stearothermophilus* estão fazendo um número extensivo de interações, sendo estas reportadas como frequentes em argininas, não se pode concluir que estas interações sejam os fatores determinantes para o aumento da estabilidade térmica. Portanto é muito difícil concluir neste caso como as argininas ou mesmo a razão Arg/(Arg+Lys) esteja contribuindo para o aumento na estabilidade térmica da TIM de *B. stearothermophilus* quando comparada com as outras estruturas. O aumento na razão neste caso é um resultado do baixo número de lisinas ao invés de um alto número de argininas quando comparado com as outras estruturas.

Desta forma a conhecida regra empírica a respeito da razão Arg/(Arg+Lys) realmente não se aplica para o caso da TIM de *B. stearothermophilus* em termos conclusivos. Possivelmente esta observação não seja válida nem mesmo de uma forma mais genérica, como já discutido e mostrado para o caso de uma média feita para proteínas de *B. stearothermophilus* e *E. coli* (Figura 6.11). Esta mesma conclusão foi também obtida pelo Dr. R. Jaenicke de uma forma mais geral (comunicação pessoal), ao analisar um grupo maior de sequências de organismos termofílicos.

Da Tabela 6.9 é claro que a TIM de *B. stearothermophilus* tem o mais alto número de prolina (5,2 %), sendo próxima à da humana (4,0 %) e sendo muito superior ao das outras TIM (2,8 %). Vários investigadores têm reportado que prolina são responsáveis por estabilidade térmica, decorrente do decréscimo dos graus de liberdade da cadeia principal (Hardy *et al.*, 1993, Nicholson *et al.*, 1992, Watanabe *et al.*, 1991). De 13 prolina na TIM de *B. stearothermophilus* uma é encontrada em N-terminal, 5 entre hélices α e fitas β , 4 em N-terminal de hélices α , 1 em fita β , 1 em C-terminal de fita β e 1 em uma longa hélice

dobrada (21 resíduos distantes do N-terminal da hélice e 5 resíduos distante do C-terminal da hélice) (Figura 6.13). Estas prolinas parecem estar exercendo um papel importante na manutenção de uma estrutura tridimensional mais rígida para cada monômero, contribuindo portanto para a estabilidade térmica da TIM de *B. stearotherophilus*.

Figura 6.13 - Distribuição das prolinas na estrutura da TIM de *B. stearotherophilus*. Nesta figura só estão representadas estruturas secundárias próximas às prolinas, que são predominantemente hélices.

TIM humana tem 10 prolinas das quais uma no N-terminal, 2 em N-terminal de hélices curtas e 7 entre hélices α e fitas β . Todas as outras estruturas de TIM têm apenas 7 prolinas. TIM

de *T. brucei* tem 2 prolina no N-terminal, 3 entre hélices α e fitas β , 1 no N-terminal de hélices e uma na posição 2 do N-terminal de uma hélice e na posição 5 do C-terminal da mesma hélice. TIM de *E. coli* tem 1 prolina no N-terminal, 4 entre hélices α e fita β , 1 em fita β e 1 em N-terminal de hélices. TIM de galinha tem 1 no N-terminal, 4 entre hélices α e fitas β e 2 em N-terminal de hélices. TIM de levedura mostra 6 prolina entre hélices α e fitas β e 1 em N-terminal de hélice. Portanto prolina aparecem como um fator de estabilidade, sendo que este fator encontra-se aumentado no caso da TIM de *B. stearothermophilus* quando comparada com todas as outras estruturas de TIM.

Um cálculo de todas as ligações de hidrogênio de todas as seis estruturas disponíveis mostra que a TIM de *B. stearothermophilus* não apresenta uma diferença significativa relativa às outras estruturas. As estruturas das enzimas TIM de *T. brucei* e *E. coli* mostram mais ligações de hidrogênio que a TIM de *B. stearothermophilus*. Estes resultados estão em bom acordo quando os programas WHATIF (Vriend, 1990) e XPLOR (Brünger, 1990a) são usados. A qualidade e energia das ligações de hidrogênio são de difícil comparação pois a resolução de cada estrutura é muito diferente (de 1,83 Å à 2,8 Å) ou o refinamento foi feito com diferentes programas para cada estrutura, dando então tratamento diferenciado às ligações de hidrogênio. Portanto uma análise detalhada da contribuição das ligações de hidrogênio à estabilidade térmica é difícil devido aos fatores descritos acima e não foi perseguida com mais detalhe.

Pontes salinas foram calculadas pelo programa WHATIF (Vriend, 1990). A TIM de *B. stearothermophilus* é a quinta em número de pontes salinas. As pontes salinas inter-subunidades são muito conservadas em todas as estruturas de TIM conhecidas (para mais detalhes nesta análise veja Wierenga *et al.*, 1992). Somente a enzima de *E. coli* apresenta uma ponte salina a mais quando histidinas são consideradas carregadas. Portanto o número de pontes salinas não pode ser considerado como um fator de estabilidade térmica positivo na TIM de *B. stearothermophilus*. Por outro lado as diferenças nas estruturas das pontes salinas devido à diferenças nos limites de resolução assim como no protocolo de refinamento tornam as conclusões difíceis.

A interação hidrofóbica tem sido mencionada por vários autores como sendo a mais importante interação para estabilidade de proteínas (para uma revisão bastante abrangente veja Dill, 1990, Sharp, 1991a,b) e tem sido reportada como sendo o fator principal em estabilidade térmica nas proteínas de organismos termofílicos de extrema estabilidade (Kirino *et al.*, 1994). A determinação da superfície hidrofóbica acessível de todas as enzimas TIM com estruturas conhecidas foi feita através da diferença entre a superfície acessível dos monômeros e do dímero. Como a estrutura em volta flexível ocupa diferentes conformações nas seis diferentes estruturas

de TIM e por encontrar-se próxima à interface de contato dos monômeros esta foi removida durante a determinação da superfície acessível. A determinação foi feita através do programa MCON (Connolly, 1983). A TIM de *B. stearothermophilus* mostra a mais baixa superfície hidrofóbica acessível por monômeros mas no entanto mostra a mais alta superfície hidrofóbica acessível dos monômeros que se torna inacessível ao solvente na formação dos dímeros: 124 a 262 Å² a mais de superfície hidrofóbica inacessível ao solvente do que as outras estruturas de TIM. Se for usado o valor estimado de 25 cal mol⁻¹ Å⁻² do ganho em energia por inacessibilidade de resíduos ao solvente (Cothia, 1974, Richards, 1977) há um ganho de 3,1 a 6,6 Kcal mol⁻¹ em estabilidade. Usando-se o valor estimado de 43 cal mol⁻¹ Å⁻² (Sharp *et al.*, 1991a, Sharp *et al.*, 1991b, De Young *et al.*, 1990) o ganho é até mesmo maior, sendo de 5,3 a 11,3 Kcal mol⁻¹ em estabilidade. Portanto mais um dos fatores que a TIM de *B. stearothermophilus* pode estar usando para o aumento de estabilidade térmica é a interação hidrofóbica entre os monômeros no processo de dimerização, o que portanto estabilizaria mais o dímero, em adição ao efeito das prolinas que dariam uma estabilidade maior aos monômeros.

Pela análise das sequências de todas as enzimas TIM disponíveis, Lys 13 em *B. stearothermophilus* é conservada em todas as estruturas como Gly, sendo somente Leu ou Ala em *Corynebacterium glutamicum* e *Moraxella sp.* Esta Lys tem uma densidade eletrônica muito bem definida e um baixo fator de temperatura (22,1 Å² para o Nζ) e faz uma forte ligação de hidrogênio com um grupo carboxílico da cadeia principal da Leu 237 da mesma subunidade (distância da ligação de hidrogênio de 2,64 Å) (Figura 6.14). No entanto a mesma Lys é encontrada nas sequências de TIM de *B. subtilis* e *B. megaterium*, duas proteínas mesofílicas reportadas recentemente (Leyva-Vasquez e Setlow, 1994, Schlapfer e Zuber, 1992). Portanto o que este resultado sugere é que a mutação Gly/Lys não é única na TIM de *B. stearothermophilus*.

O resíduo His 12 na TIM de *B. stearothermophilus* é conservado sendo Asn em todas as outras TIM de sequências conhecidas reportadas até agora incluindo as duas já mencionadas acima. Este resíduo tem uma densidade eletrônica bem definida e baixo fator de temperatura. Esta His faz interações hidrofóbicas na interface das subunidades. Este resíduo tem 23,1 Å² de superfície que se torna inacessível ao solvente durante a dimerização, contribuindo para uma estabilização de 0,6 Kcal mol⁻¹.

Uma conclusão clara deste resultado pode ser derivada do fator já discutido anteriormente da superfície hidrofóbica que se torna inacessível ao solvente durante o processo de dimerização. Devido à homologia de apenas 40% entre as enzimas TIM de estruturas conhecidas pode-se derivar a conclusão que as mutações nos resíduos da interface foram tais que contribuíram para

o aumento das interações hidrofóbicas e a mutação Asn/His é uma das que contribuiu mais, individualmente, para este resultado. Portanto uma análise mais detalhada dos resíduos que compõem a interface entre os monômeros poderia levar à mesma conclusão já observada, da estabilização do dímero por interações hidrofóbicas.

Figura 6.13 - Par de mutação Gly/Lys e Asn/His.

6.8 Conclusão

Como apresentado anteriormente há vários fatores que contribuem para a estabilidade térmica. Também em alguns casos certos autores tem reportado que somente pontes salinas foram as diferenças principais entre proteínas homólogas de organismos termofílicos e mesofílicos (Kelly *et al.*, 1993). No entanto nenhum autor discutiu até agora uma comparação de tão grande número de estruturas no intento de analisar a estabilidade térmica de proteínas, baseado na estrutura das proteínas. Desta forma, as conclusões observadas por certos autores restringem-se ao caso particular da estabilidade de uma proteína com sua homóloga e não um caso mais geral de desenvolvimento evolutivo do fator estabilidade térmica.

Portanto da análise feita aqui e baseado em todos os fatores de estabilidade térmica conclui-se que os fatores principais de estabilidade térmica para a TIM de *B. stearothermophilus* são o número aumentado de prolinas, que decresce a flexibilidade dos monômeros e a interação hidrofóbica entre as subunidades na dimerização estabilizando desta forma o dímero. Apesar da razão Arg/(Arg+Lys) ser significativamente elevada para a TIM de *B. stearothermophilus* não é claro, baseado numa análise detalhada das interações das argininas, de que forma estes resíduos estão contribuindo para a estabilidade. Esta razão é elevada na TIM de *B. stearothermophilus* devido a um decréscimo no número de Lys, mais do que um elevado número de argininas. Portanto esta razão não representa um resultado geral para estabilidade térmica de proteínas como defendido por Merker *et al.* (Merker *et al.*, 1981) e como foi concluído por análise feita neste trabalho e por outros autores (Dr. R. Jeanike, comunicação pessoal). O resultado de Merker *et al.* baseia-se em um número muito limitado de proteínas e ainda o valor da razão não se sobressai significativamente em seu trabalho para concluir que esta razão seja uma regra geral.

Uma análise evolutiva poderia ser aqui derivada baseada nas pressões seletivas que dirigiram o aparecimento de diferenças entre as proteínas mesofílicas e termofílicas. As bactérias como organismos menos evoluídos foram pressionadas a viverem em um ambiente de elevada temperatura no início da vida terrestre. Com a constante queda da temperatura terrestre para os patamares que observamos hoje, houve uma pressão da natureza para a passagem de organismos de extrema estabilidade térmica para organismos mesofílicos. Qual seria a justificativa desta pressão, ou melhor, como se justificaria esta necessidade de mudar de termofílicos para mesofílicos? Segundo certos autores as proteínas termofílicas apresentam baixa atividade enzimática à temperatura ambiente, apresentando picos de atividade enzimática a mais altas temperaturas (Dr. G. Petsko, comunicação pessoal). Devido à rigidez destas proteínas, requerida para a estabilidade térmica, estas se tornam menos ativas a mais baixas temperaturas. Desta forma

a pressão da natureza dirigiu as mutações de forma a diminuir a rigidez destas proteínas para manter o nível de atividade das enzimas.

Desta forma, como conclusão final, a TIM de *B. stearothermophilus* apresenta dois fatores principais de estabilidade térmica já mencionados acima somados a fatores relativamente sutis em relação às outras enzimas TIM com estrutura conhecida. Estes fatores foram obtidos provavelmente por pequenas mutações durante a evolução que neste caso foram decrescendo a estabilidade térmica e convergindo para a grande maioria dos organismos mesofílicos que hoje habitam o planeta. Vale lembrar aqui também que diferentes proteínas apresentam diferentes fatores de estabilidade térmica, inclusive no mesmo organismo, o que sugere que as diferentes proteínas usam diferentes processos para obter a estabilidade térmica. Um trabalho extensivo na análise de estabilidade de proteínas é feito pelo Dr. B. Matthews através da análise por mutação sítio dirigida em lisozima. Conclusões deste trabalho têm extrema importância para o entendimento dos fatores envolvidos na estabilidade protéica. Por outro lado muito dos fatores apresentados como estabilizantes pelo trabalho do Dr. Matthews não foram os usados pela TIM de *B. stearothermophilus*, mostrando desta forma que cada proteína encontra um estado de estabilidade através de mecanismos que dependem principalmente de sua particular estrutura tridimensional. Desta forma cada proteína deve ser tratada isoladamente no caso da análise de estabilidade térmica e comparada com suas homólogas para conclusões que possam justificar os fatores envolvidos na estabilidade. Para o caso particular da TIM, conclusões mais definitivas que possam permitir uma análise evolutiva dos fatores de estabilidade térmica para esta proteína devem ser obtidas quando da obtenção da estrutura da TIM de um organismo de alta estabilidade térmica.

Referências

- Adams, M. W. W. (1993). Enzymes and proteins from organisms that grow near and above 100°C. *Ann. Rev. Microbiol.*, **47**, 627-658.
- Alber, T., Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Rivers, P. S. e Wilson, I. A. (1981). On the three dimensional structure and catalytic mechanism of triosephosphate isomerase. *Phil. Trans. R. Soc. Ser. B.*, **293**, 159-171.
- Alber, T. e Kawasaki, G. (1982). Nucleotide sequence of the triose phosphate isomerase gene of *Saccharomyces cerevisiae*. *J. Mol. Appl. Genet.* **1**, 419-434.
- Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Pogson, C. I., Wilson, I. A., Corran, P. H., Furth, A. J., Milman, J. D., Offord, R. E., Priddle, J. D. e Waley, S. G. (1975). Structure of chicken muscle triosephosphate isomerase determined crystallographically at 2.5 Å resolution using amino acid sequence data. *Nature* **255**, 609-614.
- Bash, P. A., Field, M. J., Davenport, R. C., Petsko, G. A., Ringe, D. e Karplus, M. (1991). Computer simulation and analysis of the reaction pathway of triosephosphate isomerase. *Biochemistry* **30**, 5826-5832.
- Belasco, J. G. e Knowles, J. R. (1980). Direct observation of substrate distortion by triose phosphate isomerase using Fourier transform infrared spectroscopy. *Biochemistry*, **19(3)**, 472-477
- Belasco, J. G., Herihly, J. M. e Knowles, J. R. (1978). Direct observation of substrate distortion by triosephosphate isomerase. *Biochemistry*, **17**, 2971-2978.
- Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., Dinola, A. e Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, **81**, 3684-3690.
- Bernal, J. D. e Crowfoot, D. (1934). *Nature*, **133**, 794
- Blacklow, S.C., Liu, K.D. e Knowles, J.R. (1991). Stepwise improvements in catalytic effectiveness: independence and interdependence in combinations of point mutations of a sluggish triosephosphate isomerase. *Biochemistry* **30**, 8470-8476.
- Blundell, T. L. e Johnson, L. N. (1976). Protein Crystallography. Academica Press, New York.
- Böhm, H-J. (1992). The computer program LUDI: a new method for the novo design of enzyme inhibitors. *J. Comp-Aided Mol. Design*, **6**, 61-78.
- Borders Jr., C. L., Brodwater, J. A., Bekeny, P. A., Salmon, J. E., Lee, A. S., Eldrige, A. M. e Pett, V. B. (1994). A structural role for arginine in proteins: Multiple hydrogen bonds to backbone carbonyl oxygens. *Proteins Science* **3**, 541-548.

Borst, P. (1986). Discontinuous transcription and antigenic variation in trypanosomes. *Annu. Rev. Biochem.*, **55**, 55.

Brändén, C. I. Anatomy of α/β proteins. Current communications in molecular biology: computer graphics and molecular modelling. Fletterick and Zoller ed. 1986. Cold Spring Harbor Laboratories. Cold Spring Harbor, New York.

Brener, Z. (1973). Biology of *Trypanosoma cruzi*. *Ann. Rev. Microbiol.*, **27**, 347-382.

Bricogne, G. (1974). Methods and Programs for Direct-space exploitation of geometric redundancies. *Acta Cryst.*, **A32**, 832-847.

Brünger, A. T. (1990a). X-plor User Manual. Yale University Press, Yale.

Brünger, A. T. (1990b). Extension of molecular replacement: a new search strategy based on Patterson correlation refinement. *Acta Cryst.* **A46**, 46-57.

Brünger, A. T., Krukowski, A. e Erikson, J. (1990). Slow cooling protocols for crystallographic refinement by simulated annealing. *Acta Cryst.* **A46**, 585-593.

Brünger, A. T., Kuriyan, J. e Karplus, M. (1987). Crystallographic R factor refinement by molecular dynamics. *Science*, **236**, 458-593.

Buisson, G., Duee, E., Haser, R. e Payan, F. (1987). Three-dimensional structure of porcine pancreatic α -amylase at 2.9 Å resolution. Role of calcium in structure and activity. *EMBO J.*, **6**, 3909-3916.

Callens, M., Kuntz, D. A. e Oppendoes, F. R. (1991). Kinetic properties of fructose biphosphate aldolase from *Trypanosoma brucei* compared to aldolase from rabbit muscle and *Staphylococcus aureus*. *Mol. and Biochem. Parasit.*, **47**, 1-10

Cheng, J., Mielnicki, L. M., Pruitt, S. C. e Maquat, L. E. (1990). Nucleotide sequence of murine triosephosphate isomerase cDNA. *Nucleic. Acids. Res.* **18**, 4261

Chevalier, N., Callens, M e Michels, P. A. M. High-level expression of *Trypanosoma brucei* fructose-1,6-bisphosphate aldolase in *Escherichia coli* and purification of the enzyme. *Mol. Biol. Paras.*, submetido.

Chothia, C. (1974). Hydrophobic bonding and accessible surface area in proteins. *Nature* **248**, 338-339.

Clarkson, A. B. e Brohn, F. H. (1976). Trypanosomiasis: an approach by chemotherapy by the inhibition of carbohydrate catabolism. *Science*, **194**, 204-206.

Cohen, S. N., Chang, A., Boyer, H. e Helling, R. (1973). Construction of biologically functional bacterial plasmids in vitro. *Proc. Natl. Acad. Sci.*, **70**, 3240-3244.

Cohen, S. N., 1985. DNA cloning: historical perspectives. *Biogenetics of Neurohormonal Peptides*, 3-14, Academic Press.

Collins, K. D. (1974). An activated intermediate analogue. The use of phosphoglycolohydroxamate as a stable analogue of a transiently occurring dihydroxyacetone phosphate-derived enolate in enzymatic catalysis. *J. Biol. Chem.*, **249**, 136-142.

Connolly, M. L. (1983). Analytical molecular surface calculation. *J. Appl. Crystall.* **16**, 548-558.

Corran, P. H. e Waley, S. G. (1975). The amino acid sequence of rabbit muscle triose phosphate isomerase. *Biochem. J.* **145**, 335-344.

Creighton, T. E. (1991). Stability of folded conformation. *Curr. Opin. Struc. Biol.*, **1**, 5-16

Crowther, R. A. (1967). A linear analysis of non-crystallographic symmetry problem. *Acta Cryst.*, **22**, 758-764.

Crowther, R. A. (1969). The use of non-crystallographic symmetry for phase determination. *Acta Cryst.*, **25B**, 2571-2580.

Davenport, R. C., Bash, P. A., Seaton, B. A., Karplus, M., Petsko, G. A. e Ringe, D. (1991). Structure of the triosephosphate isomerase-phosphoglycolohydroxamate complex: an analogue of the intermediate on the reaction pathway. *Biochemistry* **30**, 5821-5826.

Delboni, L. F. e Hol. W. G. J. (1992). Comparison between *T. brucei* and human fructose-1,6-bisphosphate aldolases sequences based on 3-D structure of human muscle protein. *20th Microsymposium on glycosomes*, October 28th, 1992, Groningen.

Delboni, L. F., Mande, S. M., Rentier-Delrue, F., Mainfroid, V., Vellieux, F. M. D., Martial, J. A. e Hol, W. G. J. Crystal Structure of Recombinant Triosephosphate Isomerase from *Bacillus stearothermophilus* Complexed with 2-Phosphoglycolate at 2.8 Å Resolution, *submitted*.

DeYoung, L. R. e Dill, K. A. (1990). Partitioning of nonpolar solutes into bilayers and amorphous n-alkanes *J. Phys. Chem.* **94**, 801-809.

Dill, K. A. (1990). Dominant forces in Protein Folding. *Biochemistry* **29**, 7133-7155.

Eikmanns, B. J. (1992). Identification, sequence analysis, and expression of a *Corynebacterium glutamicum* gene cluster encoding the three glycolytic enzymes glyceraldehyde-3-phosphate dehydrogenase 3-phosphoglycerate kinase, and triosephosphate isomerase. *J. Bacteriol.* **174**, 6076-6086

Eriksson, A. E., Baase, W. A., Zhang, X.-J., Heinz, D. W., Blaber, M., Baldwin, E. P. e Matthews, B. W. (1992). Response of a Protein Structure to Cavity-Creating Mutation and Its Relation to the Hydrophobic Effect. *Science* **225**, 178-183.

Farber, G. K. e Petsko, G. A. (1990). The evolution of α/β barrel enzymes. *Trends Biochem. Sci.*, **15**, 228-234.

Fersht, A. R. e Sternberg, M. J.E., 1989. *Prot. Eng.*, **2**, 527-

Fontana, A. (1991). Analysis and modulation of protein stability. *Curr. Opin. Biotech.*, **2**, 551-560.

Freire, J. F. (1993). Structural thermodynamics: prediction of protein stability and protein binding affinities. *Arch. Biochem Biophys.*, **303(2)**, 181-184.

Flocco, M. M. e Mowbray, S. (1994). Planar interactions of Arginine and Aromatic Side-Chains in Proteins. *J. Mol. Biol.* **235**, 709-717.

Fothergill-Gilmore, L. A. e Michels-P-A (1993). Evolution of glycolysis. *Prog. Biophys. Mol. Biol.*, **59(2)**, 105-235.

Gamblin, S. J., Cooper, B. Millar, J. R., Davies, G. J., Littlechild, J. A. e Watson, H. C. (1990). The crystal structure of human muscle aldolase at 3,0 Å resolution. *FEBS Letters*, **262**, 282-286.

Gamblin, S. J., Davies, G. J., Grimes, J. M., Jackson, R. M., Littlechild, J. A. e Watson, H. C. (1991). Activity and specificity of human aldolases. *J. Mol. Biol.*, **219**, 573-576.

Gilliland, G. L. (1988). A biological macromolecule crystallization database: a basis for a crystallization strategy. *J. Cryst. Growth*, **90**, 51

Gilson, M. K., Sharp, K. A. e Honig, B. H., (1987). *J. Comp. Chem.*, **9**, 327-

Goldman, A., Ollis, D. L. e Steitz, T. A. (1987). Crystal structure of muconate lactonizing enzyme at 3 Å resolution. *J. Mol. Biol.*, **194**, 143-153.

Greer, J., Erickson, J. W., Baldwin, J. J. e Varney, M. D. (1994). Application os the three-dimensional strcutures of protein target molecules in structure-based drug design. *J. Med. Chem.*, **37**, 1233-1251.

Gutteridge, W. E. (1985). Existing chemotherapy and its limitations. *Br. Med. Bull.*, **41**, 162-168

Hagler, A. T., Huler, E. e Lifson, S., (1977). *J. Am. Chem. Soc.*, **96**, 5319-

Hardy, F., Vriend, G., Veltman, O. R., van der Vinne, B., Venema, G. e Eijnsink, V. G. (1993). Stabilization of *Bacillus stearothermophilus* neutral protease by introduction of prolines. *FEBS Letter* **317**, 89-92.

Hauptman, H. (1982). On integrating the thecniques of direct methods and isomorphous replacement. I The theoretical basis. *Acta Cryst.*, **A38**, 289-294.

Henrick, K., Collyer, C. A. e Blow, D. M. (1989). Structure of Dxylose isomerase from *Arthobacter* strain B3728 containing the inhibitors xylitol and D-sorbitol at 2.5 and 2.3 Å resolution, respectively. *J. Mol. Biol.*, **208**, 129-157.

Hester, G., Brenner-Holzach, O., Rossi, F. A., Struck-Donatz, M., Winterhalter, K. H., Smith, J. D. G. e Piontek, K. (1991). The crystal structure of fructose-1,6-bisphosphate aldolase from *Drosophila melanogaster* at 2.5 Å resolution. *FEBS Letters*, **292**, 237-242.

Higgins, D. G., Bleasby, A. J. e Fuchs, R. (1992). CLUSTAL V: improved software for multiple sequence alignment. *Comput. Appl. Biosci.*, **8(2)**, 189-191.

Higgins, D. G. (1994). CLUSTAL V: multiple alignment of DNA and protein sequences. *Methods Mol. Biol.*, **25**, 307-318.

Hol, W. G. J., van Duijnen, P. T. e Berendsen, H. J. C. (1978). The α -helix dipole and the properties of proteins. *Nature (London)*, **273**, 443-446.

Horecker, B. L., Tjolas, O. e Lai, C. Y. (1972). em *The Enzymes* (ed. Byer, P. D.), **7**, 213-258, Academic Press, New York.

Hyde, C. C., Ahmed, S. A., Padlan, E. A., Miles, E. W. e Davies, D. R. (1988). Three-dimensional structure of the tryptophan synthase $\alpha 2\beta 2$ multienzyme complex from *Salmonella typhimurum*. *J. Biol. Chem.*, **263**, 17857-17871.

Ikai, A (1980). Thermostability and aliphatic index of globular proteins. *J. Biochem. (Tokyo)*, **88**, 1895-1898.

Jackson, D. A., Symons, R. H. e Berg, P. (1972). Biochemical method for inserting new genetic information into DNA of simian virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of *Escherichia coli*. *Proc. Natl. Acad. Sci.*, **69**, 2904-2909.

Jones, T. A., Zou, J.-Y., Cowan, S. W. e Kjeldgaard, M. (1991). Improved methods for building models in electron density maps and the location of errors in these models. *Acta Cryst.* **A47**, 110-119.

Joseph, D., Petsko, G. A. e Karplus, M. (1990). Anatomy of a conformational change: hinged "lid" motion of the triosephosphate isomerase loop. *Science*, **249**, 1425-1428.

Kelly, C. A., Nishiyama, M., Ohnishi, Y., Beppu, T. e Birktoft, J. (1993). Determinants of Protein thermostability Observed in the 1.9 Å Crystal structure of Malate dehydrogenase from the Thermophilic Bacterium *Thermus flavus*. *Biochemistry* **32**, 3913-3922.

Kirchhoff, L. V. (1993). American trypanosomiasis (Chagas' disease) - A tropical disease now in the United States. *New Engl. J. Med.*, **329**, 639-644.

Kirino, H., Aoki, M., Aoshima, M., Hayashi, Y., Ohba, M., Yamagishi, A., Wakagi, T. e Oshima, T. (1994). Hydrophobic interaction at the subunit interface contributes to the thermostability of 3-isopropylmalate dehydrogenase from an extreme thermophile, *Thermus thermophilus*. *Eur. J. Biochem.*, **220**(1), 275-81.

Kirkpatrick, S., Gelatt Jr., C. D. e Vecchi, M. P. (1983). Optimization by simulation annealing. *Science*, **220**, 671-680.

Klapper, I., Hagstom, R., Fine, R., Sharp, K. e Honig, B., (1986). *Proteins*. **1**, 47-59.

Klein, C. e Schulz G. E. (1991). Structure of cyclodextrin glycosyl transferase refined at 2,0 Å resolution. *J. mol. Biol.*, **217**, 737-750.

Knowles, J. R. e Alber, W. J. (1977). Perfection in enzyme catalysis: The energetics of triosephosphate isomerase. *Acc. Chem. Res.*, **10**, 105-111.

Kolb, E., Harris, J. I. e Bridgen-J. (1974). Triose phosphate isomerase from the coelacanth. An approach to the rapid determination of an amino acid sequence with small amounts of material. *Biochem. J.* **137**, 185-197.

Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystall.* **24**, 946-950.

Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R. e Ferrin, T. E. (1982). A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.*, **161**, 269-288.

Kuntz, I. D. (1992). Structure-based strategies for drug design and discovery. *Science*, **257**, 1078-1082.

Lambeir, A.-M., Oppendoorn, F. R. e Wieringa, R. K. (1987). Kinetic properties of triosephosphate isomerase from *Trypanosoma brucei brucei*: a comparison with the rabbit and yeast enzymes. *Eur. J. Biochem.*, **168**, 69-74.

Leiboda, L., Hatada, M. H., Tulinsky, A. e Maravadi, I. M. (1982). Comparison of the folding of 2-keto-3-deoxy-6-phosphogluconate aldolase, triose phosphate isomerase and pyruvate kinase. *J. Mol. Biol.*, **162**, 445-458.

Lim, L. W., Shamala, N., Mathews, F. S., Steenkamp, D. J., Hamlin, R. e Xuong, N. (1986). Three-dimensional structure of the iron-sulfur flavoprotein trimethyl amine dehydrogenase at 2.4 Å resolution. *J. Biol. Chem.*, **261**, 15140-15146.

Lindqvist, Y. (1989). Refined structure of spinach glycolate oxidase at 2.0 Å resolution. *J. Mol. Biol.*, **209**, 151-166.

Littlechild, J. A. e Watson, H. C. (1993). A data-based reaction mechanism for type I fructose biphosphate aldolase. *TIBS*, **18**, 36-39.

- Leyva-Vazquez, M. A. e Setlow, P. (1994). Cloning and nucleotide sequences of the genes encoding triose phosphate isomerase, phosphoglycerate mutase, and enolase from *Bacillus subtilis*. *J. Bact.* **176**(13), 3903-3910.
- Lolis, E., Alber, T., Davenport, R. C., Rose, D., Hartman, F. C. e Petsko, G. A. (1990). Structure of yeast triosephosphate isomerase at 1.9 Å resolution. *Biochemistry* **29**, 6609-6618.
- Lolis, E. e Petsko, G. A. (1990). Crystallographic analysis of the complex between triosephosphate isomerase and 2-phosphoglycolate at 2.5 Å resolution: implications for catalysis. *Biochemistry* **29**, 6619-6625.
- Main, P. e Rossmann, M. G. (1966). Relationships among structure factors due to identical molecules indifferent crystallographic environments. *Acta Cryst.*, **21**, 67-72.
- Main, P. (1967). Phase determination using non-crystallographic symmetry. *Acta Cryst.*, **23**, 50-54.
- Mande, S. C., Mainfroid, V., Kalk, K. H., Goraj, K., Martial, J. A. e Hol, W. G. J. (1994). Crystal structure of recombinant human triosephosphate isomerase at 2.8 Å resolution. Triosephosphate isomerase-related human genetic disorders and comparison with the trypanosomal enzyme. *Protein Science* **3**, 810-821.
- Maquat, L. E., Chilcote, R. e Ryan, P. M. (1985). Human triosephosphate isomerase cDNA and protein structure. Studies of triosephosphate isomerase deficiency in man. *J. Biol. Chem.* **260**, 3748-3753.
- Marchionni, M. e Gilbert, W. (1986). The triosephosphate isomerase gene from maize: introns antedate the plant-animal divergence. *Cell* **46**, 133-141.
- Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.*, **33**, 491-497.
- Matthews, B. W. (1993). Structural and genetic analysis of protein stability. *Ann. Rev. Biochem.* **62**, 139-160.
- Mavridis, I. M., Hadata, M. H., Tulinsky, A. e Lebioda, L. (1982). Structure of 2-keto-3-deoxy-6-phosphogluconate aldolase at 2,8 Å resolution. *J. Mol. Biol.*, **162**, 419-444.
- McKnight, G. L., O'Hara, P. J. e Parker, M. L. (1986). Nucleotide sequence of the triosephosphate isomerase gene from *Aspergillus nidulans*: implications for a differential loss of introns. *Cell* **46**, 143-147.
- Menendez-Arias, L. e Argos, P. (1989). Engineering Protein thermal stability. Sequence Statistics Point to residue Substitutions in α -helix. *J. Mol. Biol.* **206**, 397-406.
- Meng, E. C., (1992). Automated docking with grid-based energy evaluation. *J. Comp. Chem.*, **13**, 505-524.

- Mengel, J. O. e Rossi, M. A. (1992). Chronic chagasic myocarditis pathogenesis: dependence on autoimmune and microvascular factors. *Amm. Heart. J.*, **124**, 1052-1057.
- Merkler, D. J., Farrington, G. K. e Becktel, W. J. (1981). Protein thermostability. Correlation between macroscopic parameters and growth temperature for closely related thermophilic and mesophilic bacilli. *Int. J. Pept. Protein Res.*, **18**, 430-442.
- Misset, O., Bos, O. J. M. e Opperdoes, F. R. (1986). Glycolytic enzymes of *Trypanosoma brucei*. Simultaneous purification, intraglycosomal concentration and physical properties. *Eur. J. Biochem.*, **157**, 441-453.
- Molyneux, D. H. e Ashford, R. W. (1983). The biology of trypanosoma and leishmania, parasites of man and domestic animals. Eds., Laylor and Francis, London.
- Muirhead, H., Clayden, D. A., Barford, D., Lorimer, G., Fothergill-Gilmore, L. A., Schiltz, E. e Schmitt, W. (1986). The structure of cattle muscle pyruvate kinase. *EMBO J.*, **5**, 475-481.
- Nandi, C. L., Singh, J. e Thornton, J. M. (1993). Atomic environments of arginine residues in proteins. *Protein Engineering* **6**, 247-259.
- Narasinga Rao, S., Jih J.-H. e Hartsuck, J. A. (1980). Rotation-function studies of myoglobin. *Acta Cryst.*, **A36**, 878-884.
- Neidhart, D. J., Kenyon, G. L., Gerlt, J. A. e Petsko, G. A. (1990). Mandelate racemase and muconate lactonizing enzyme are mechanically distinct and structurally homologous. *Nature (London)*, **347**, 692-694.
- Nicholson, H., Tronrud, D. E., Becktel, W. J. e Matthews, B. W. (1992). Analysis of the effectiveness of proline substitutions and glycine replacements in increasing the stability of phage T4 lysozyme. *Biopolymers* **32**, 1431-1441.
- Noble, M. E. M., Zeelen, J. Ph., Wierenga, R. K., Mainfroid, V., Goraj, K., Gohimont, A. C. e Martial, J. A. (1993). Structure of triosephosphate isomerase from *Escherichia coli* determined at 2.6 Å resolution. *Acta Cryst.*, **D49**(4), 403-417.
- Noble, M. E. M. (1992) X-Ray Crystallographic Studies of Triosephosphate Isomerase. *PhD Thesis*, State University of Groningen, Chapter 5.
- Noble, M. E., Verlinde, C. L., Groendijk, H., Kalk, K. H., Wierenga, R. K. e Hol, W. G. J. (1991a). Crystallographic and Molecular Modeling Studies on Trypanosomal Triosephosphate Isomerase: A Critical Assessment of the Predicted and Observed Structures of the 2-phosphoglycerate. *J. Med. Chem.* **34**, 2709-2718.
- Noble, M. E. M., Wierenga, R. K., Lambeir, A. M., Opperdoes, F. R., Thunnissen, A. M., Kalk, K. H., Groendijk, H. e Hol, W. G. J. (1991b). The Adaptability of the Active Site of Trypanosomal Triosephosphate Isomerase as observed in Crystal Structures of Three Different Complexes. *Proteins: Struct. Funct. Genet.* **10**(1), 50-69.

- Okada, N., Koizumi, N., Tanaka, T., Ohkubo, H., Nakanishi, S. e Yamada, Y. (1989). Isolation, sequence, and bacterial expression of a cDNA for (S)-tetrahydroberberine oxidase from cultured berberine-producing *Coptis japonica* cells *Proc. Natl. Acad. Sci. U.S.A.* **86**, 534-538.
- Old, S. E. e Mohrenweiser, H. W. (1988). Nucleotide sequence of the triosephosphate isomerase gene from *Macaca mulatta*. *Nucleic. Acids. Res.* **16**, 9055.
- Opperdoes, F. R. e Borst, P. (1977). Localization of nine glycolytic enzymes in a microbody-like organelle in *Trypanosoma brucei*: the glycosome. *FEBS Lett.*, **80**, 360-364.
- Opperdoes, F. R. (1983) in *Mechanism of Drug Action*, eds. Singer, T. P., Mansor, T. E. e Undarza, R. N. (Academic Press, New York), 121-151.
- Pattabiraman, N, Levitt, M, Ferrin, T. E. e Langridge, R., (1985). *J. Comp. Chem.*, **6**, 432.
- Perutz, M. (1992). *Protein Structure: New Approaches to disease and therapy*. New York, W. H. Freeman & Co.
- Pichersky, E., Gottlieb, L. D. e Hess, J. F. (1984). Nucleotide sequence of the triosephosphate isomerase gene of *Escherichia coli*. *Mol. Gen. Genet.*, **195**, 314-319.
- Priestle, J. P., Gruetter, M. G., White, J.L., Vincent, M. G., Kania, M., Wilson, E., T. S., J., Kirschner, K. e Jansonius, J. N. (1991). Three dimensional structure of the bifunctional enzyme N-(5'-phosphoribosyl)-anthranilate isomerase-indole-3-glycerol-phosphate synthase from *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.*, **84**, 5690-5694.
- Read, R. J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst.* **A42**, 140-149.
- Rentier-Delrue, F., Mande, S. C., Moyens, S., Terpstra, P., Mainfroid, V., Goraj, K., Lion, M., Hol, W. G. J. e Martial, J. A. (1993). Cloning and Overexpression of the triosephosphate Isomerase Genes from Psychrophilic and thermophilic Bacteria: Structural comparison of the Predicted Protein Sequences. *J. Mol. Biol.* **229**, 85-93.
- Richards, F. M. (1977). Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* **6**, 151-176.
- Rondeau, J.-M., Tete-Favier, F., Podjarny, A. Reymann, J.-M., Barth, P., Biellmann, J.-F. e Moras, D. (1992). Novel NADPH-binding domain revealed by crystal structure of aldolase reductase. *Nature (London)*, **355**, 469-472.
- Rossmann, M. G. e Blow, D. M. (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Cryst.*, **15**, 24-31
- Rossmann, M. G., Blow, D. M., Harding, M. M. e Collier, E. (1964). The relative positions of independent molecules within the same asymmetric unit. *Acta Cryst.*, **17**, 338-342.

- Rossmann, M. G. (1990). The molecular replacement method. *Acta Cryst.*, **A46**, 73-82.
- Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J. K. C. e Jones, T. A. (1990). Three dimensional structure of cellobiohydrolase II from *Trichoderma reesi*. *Science*, **249**, 380-386.
- Russell, P. R. (1985). Transcription of the triosephosphate isomerase gene of *Schizosaccharomyces pombe* initiates from a start point different from that in *Saccharomyces cerevisiae*. *Gene* **40**, 125-130.
- Schlapfer, B. S. e Zuber, H. (1992) Cloning and sequencing of the gene encoding glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerate kinase and triosephosphate isomerase (gap operon) from mesophilic *Bacillus megaterium*: comparison with corresponding sequences from thermophilic *Bacillus stearothermophilus*. *Gene* **122(1)**, 53-62.
- Schneider, G., Lindqvist, Y. e Lundqvist, T. (1990). Crystallographic refinement and structure of ribulose-1,5-bisphosphate carboxylase from *Rhodospirillum rubrum* at 1,7 Å resolution. *J. Mol. Chem.*, **261**, 15140-15146.
- Schreuder, H. A., Groendijk, H., van der Laan, J. M. e wierenga, R. K. (1988). The transfer of protein crystals from their original mother liquor to a solution with a completely different precipitant. *J. Appl. Cryst.*, **21**, 426-429.
- Sharp, K. A., Nicholls, A., Fine, R. F. e Honing, B. (1991a). Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects. *Science* **252**, 106-109.
- Sharp, K. A., Nicholls, A., Friedman, R. e Honing, B. (1991b). Extracting hydrophobic free energies from experimental data: relationship to protein folding and theoretical models. *Biochemistry* **30**, 9686-9697.
- Shaw-Lee, R. L., Lissemore, J. L. e Sullivan, D. T. (1991). Structure and expression of the triosephosphate isomerase(Tpi) gene of *Drosophila melanogaster*. *Mol. Gen. Genet.* **230**, 225-2
- Schoichet, B. K. e Kuntz, I. D., (1991). Protein Docking and Complementarity. *J. Mol. Biol.*, **221**, 327-346.
- Shoichet, B. K., Stroud, R. M., Santi, D. V., Kuntz, I. D. e Perry, K. M. (1993). Structure-based discovery of inhibitors of thymidylate synthase. *Science*, **259**, 1445-1450.
- Shoichet, B. K., Bodian, D. L. e Kuntz, I. D. (1992). Molecular Docking Using Shape descriptors. *J. Comp. Chem.*, **13**, 380-397.
- Stec, B e Lebioda, L. (1990). Refined structure of yeast apo-enolase at 2,25 Å resolution. *J. Mol. Biol.*, **211**, 235-248.
- Straus, D. e Gilbert, W. (1985). Chicken triosephosphate isomerase complements an *Escherichia coli* deficiency. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2014-2018.

Straus, D., Raines, R., Kawashima, E. Knowles, J. R. e Gilbert, W. (1985). Active site of Triosephosphate isomerase: *in vitro* mutagenesis and characterization of an altered enzyme. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2272-2276.

Studier, F. W., Rosenberg, A. H., Dunn, J. J. e Dubendorff, J. W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods in Enzimology*, **185**, 60-89.

Sygush, J., Beaudry, D. e Allarie, M. (1987). Molecular Architecture of rabbit skeletal muscle aldolase at 2,7 Å resolution. *Proc. Nat. Acad. Sci., U.S.A.*, **84**, 7846-7850.

Swinkels, B. W., Gibson, W. C., Osinga, K. A., Kramer, R., Veeneman, G. H., van Boom, J. H. e Borst, P. (1986). Characterization of the gene for the microbody (glycosomal) triosephosphate isomerase of *Trypanosoma brucei*. *Embo J.* **5**, 1291-1298.

Takahashi, I., Takasaki, Y. e Hori-K. (1989). Site-directed mutagenesis of human aldolase isozymes: the role of Cys-72 and Cys-338 residues of aldolase A and of the carboxy-terminal Tyr residues of aldolases A and B. *J. Biochem. Tokyo*, **105(2)**, 281-286.

Tittiger, C., Whyard, S. e Walker, V. K. (1993). A novel intron site in the triosephosphate isomerase gene from the mosquito *Culex tarsalis*. *Nature* **361**, 470-472.

Vellieux, F. M. D. (1993). Density Modification (DEMON). Grenoble, FRA.

Verlet, L. (1967). Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.*, **159**, 98-105.

Verlinde, C. L. M. J., Rudenko, G. e Hol, W. G. J. (1992a). In search of new lead compounds for trypanosomiasis drug design; a protein structure-based linked fragment approach. *J. Comp. Aided Mol. Design*, **46**, 131-147.

Verlinde, C. L. M. J., Noble, M. E. M., Kalk, K. H., Groendijk, H., Wierenga, R. K., e Hol, W. G. J. (1991). Anion binding to the active site of trypanosomal triosephosphate isomerase: monohydrogen phosphate does not mimic sulphate. *Eur. J. Biochem.*, **198**, 53-57.

Verlinde, C. L. e Hol, W. G. J. (1994). Structure-based drug design: progress, results and challenges. *Structure*, **2(7)**, 577-587.

Verlinde, C. L., Witmans, C. J., Pijning, T., Kalk, K. H., Hol, W. G., Callens, M. e Opperdoes, F. R. (1992b). Structure of the complex between trypanosomal triosephosphate isomerase and N-hydroxy-4-phosphono-butanamide: binding at the active site despite an "open" flexible loop conformation. *Protein Sci.* **1(12)**, 1578-1584.

Verlinde, C. L. M. J., Merritt, E. A., van den Akker, F., Kim, H., Feil, I., Delboni, L. F., Mande, S. C., Sarfaty, S., Petra, P. H. e Hol, W. G. J., (1994). Protein crystallography and infectious diseases. *Protein Science*, **3**, 1670-1686.

- Vihinen, M. (1987). Relationship of protein flexibility to thermostability. *Protein Engineering* **1**, 477-480.
- Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.*, **8**, 52-56.
- Vriend, G e Eijssink, V. (1993). Prediction and analysis of structure, stability and unfolding of thermolysin-like proteases. *J. Comput. Aided Mol. Des.* **7**(4), 367-396.
- Waley, S. G. (1973). Refolding of triosephosphate isomerase. *Biochem. J.* **135**, 165-172.
- Wang, C. C. (1984). Parasite enzymes as potential target for antiparasitic chemotherapy. *J. Med. Chem.*, **27**, 1-9.
- Wang, B. C. (1985). Resolution of phase ambiguity in macromolecular crystallography. *Methods in Enzymology*, **115**, 90-112.
- Watanabe, K., Chishiro, K., Kitamura, K. and Suzuki, Y. (1991). Proline Residues Responsible for Thermostability Occur with High Frequency in the Loop Regions of an Extremely Thermostable Oligo-1,6-glucosidase from *Bacillus thermooglucosidasius* KP1006. *J. Biol. Chem.* **266**(36), 24287-24294.
- Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., Jr. e Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.*, **106**, 765-784.
- White, A., Allarie, M., Beaudry, D. e Sygush, J. (1991). Crystal structure of rabbit liver aldolase. American Crystallographic Association, Annual Meeting Abstracts, PG20.
- White, P.J. e Blundell, T. L. (1994). Protein structure-based drug design. *Ann. Rev. Biophys. Biomol. Struct.*, **23**, 349-375.
- Wierenga, R. K., Noble, M. E. M. e Davenport, R. C. (1992). Comparison of the Refined crystal Structure of Liganded and Unliganded Chicken, Yeast and Trypanosomal Triosephosphate Isomerase. *J. Mol. Biol.* **224**, 1115-1126.
- Wierenga, R. K., Noble, M. E. M., Postma, J. P. M., Groendijk, H., Kalk, K. H., Hol, W. G. J. e Opperdoes, F. R. (1991a). The crystal structure of the "open" and the "closed" conformation of the flexible loop of trypanosomal triosephosphate isomerase. *Proteins Struct. Funct.* **10**, 33-49.
- Wierenga, R. K., Noble, M. E. M., Vriend, G., Nauche, S. e Hol, W. G. J. (1991b). Refined 1.83Å structure of trypanosomal triosephosphate isomerase crystallized in the presence of 2.4 M-ammonium sulphate. A comparison with the structure of the trypanosomal triosephosphate isomerase-glycerol-3-phosphate complex. *J. Mol. Biol.* **220**, 995-1015.

Wierenga, R. K., Kalk, K. H. e Hol, W. G. J. (1987). Structure determination of the glycosomal triosephosphate isomerase from *Trypanosoma brucei brucei* at 2.4 Å resolution. *J. Mol. Biol.*, **198**, 109-121.

Wierenga, R. K., Kalk, K. H., Hol, W. G. J., Misset, O. e Opperdoes, F. R. (1984). Preliminary crystallographic studies of triosephosphate isomerase from the blood parasite *Trypanosoma brucei brucei*. *J. Mol. Biol.*, **178**, 487-490.

Wilson, D. K., Rudolph, F. B. e Quioco, F. A. (1991). Atomic structure of adenosine deaminase complexed with a transition-state analog: Understanding catalysis and immunodeficiency mutation. *Science*, **252**, 1278-1284.

Wolfenden, R. (1969). *Nature*, **223**, 704-705.

Xia, Z. e Mathews, F. S. (1990). Molecular structure of flavocytochrome b2 at 2.4 Å resolution. *J. Mol. Biol.*, **212**, 837-863.

Zabori, S., Rudolph, R. e Jaenicke, R. (1980). Folding and association of triosephosphate isomerase from rabbit muscle. *Z. Naturforsch.* **35C**, 999-1004.

Zhang, X.-J., Baase, W. A. e Matthews, B. W. (1991). Toward a Simplification of the Protein folding Problem: A Stabilizing Polyalanine α -Helix Engineered in T4 Lysozyme. *Biochemistry* **30**, 2012-2017.