

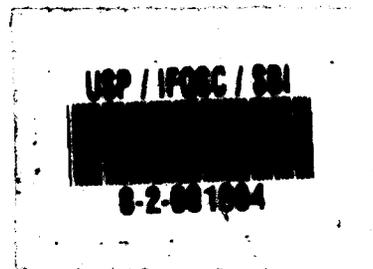
UNIVERSIDADE DE SÃO PAULO  
INSTITUTO DE FÍSICA DE SÃO CARLOS  
DEP. DE FÍSICA E INFORMÁTICA

**Determinação da estrutura cristalográfica da  
enzima Glucosamina-6-fosfato desaminase de  
*E. coli* K12 e seus complexos com ativador  
alostérico e inibidor**

**Por: Marcos Roberto de M. Fontes**

**Orientador: Prof. Dr. Glaucius Oliva**

*OL*



Tese apresentada ao Instituto de  
Física de São Carlos, para  
obtenção do título de **Doutor em  
Física Aplicada**

Agosto de 1995



MEMBROS DA COMISSÃO JULGADORA DA TESE DE DOUTORADO DE MARCOS ROBERTO DE MATTOS FONTES APRESENTADA AO INSTITUTO DE FÍSICA DE SÃO CARLOS, UNIVERSIDADE DE SÃO PAULO, EM 07/08/1995.

**COMISSÃO JULGADORA:**

Prof. Dr. Glaucius Oliva

  
-----

Prof. Dr. Oswaldo Baffa-Filho

Profa. Dra. Lia Queiroz do Amaral

Profa. Dra. Rosemary Sanches

Prof. Dr. Fernando de Castro Reinach

Este trabalho teve apoio financeiro:

CNPq  
PICD-CAPES  
FAPESP  
FINEP  
IFSC-USP  
IBB-UNESP

Aos meus pais, Marisa e Antonio  
Carlos, ao meu irmão Sérgio e à minha  
esposa Sonia.

## AGRADECIMENTOS

- Ao Prof. Dr. Glaucius Oliva, pela orientação, apoio e amizade.
- Ao Prof. Dr. Eduardo Horjales, pela co-orientação e amizade.
- Ao Prof. Dr. Richard C. Garratt, pelos ensinamentos, constante disposição em colaborar e amizade.
- Aos Profs. Drs. Eduardo E. Castellano, Yvonne P. Mascarenhas e Julio Zukerman-Schpector, pelos ensinamentos e amizade.
- Ao Profs. Drs. Mário L. Calcagno e Myriam M. Altamirano pelo fornecimento das amostras.
- Aos colegas de laboratório, Paulo, Valma, Ezequiel, Tereza, Walter, Rodrigo, pelas úteis discussões, ensinamentos e amizade.
- Aos colegas de laboratório, Bianca, Geraldo, Jorge, Rejane, João, Gisele, Dulce, Fernando, Cristina, Juan e Beatriz, pela amizade.
- Aos técnicos Wanda, Valdir, Augusto, Geraldo pela constante disposição em colaborar e amizade.
- À Maria Helena e a Vladerez
- Aos colegas do Departamento de Física e Biofísica do IB-UNESP, Botucatu, em especial, ao Prof. Dr. José Roberto C. Saglietti pelo apoio nas etapas finais de realização desta tese e pela amizade.
- A todas as pessoas que direta ou indiretamente me ajudaram na realização deste trabalho.

# SUMÁRIO

ABREVIATURAS.....	viii
LISTA DE TABELAS.....	ix
LISTA DE FIGURAS.....	x
RESUMO.....	xiii
ABSTRACT.....	xiv
INTRODUÇÃO.....	1
1 - A ENZIMA GLUCOSAMINA 6-FOSFATO DESAMINASE DE <i>Escherichia coli</i> .....	3
1.1 - Características principais.....	3
1.2 - Revisão do conhecimento a níveis estruturais e funcionais sobre a GlcN6P desaminase de <i>E. coli</i> .....	4
1.3 - Cristalização da enzima.....	7
1.4 - Coleta e análise preliminar dos dados de difração de raios-X da proteína nativa.....	8
2 - PREPARAÇÃO E COLETA DE DADOS DE DIFRAÇÃO DE RAIOS-X DERIVADOS ISOMORFOS.....	9
2.1 - Preparação dos derivados: difusão em solução ( <i>soaking</i> ).....	11
2.2 - Coleta de dados de difração de raios-X.....	14
2.2.1 - O sistema de coleta de dados (placa de imagem eletrônica).....	14
2.2.2 - Danos aos cristais devido à radiação ionizante, sistema de minimização de danos por resfriamento do cristal.....	16
2.2.3 - Escolha e montagem dos cristais.....	17
2.2.4 - Estratégias de coleta de dados de difração de raios-X.....	19
2.2.5 - Processamento inicial dos dados para a obtenção dos $F_{obs}$ .....	22
2.2.6 - Coleta de dados para obtenção de derivados isomorfos.....	32
3 - CÁLCULO INICIAL DAS FASES.....	34
3.1 - O método da substituição isomórfica múltipla (MIR) para o cálculo de fases.....	34
3.1.1 - Determinação das posições dos átomos pesados.....	34
3.1.2 - Cálculo das fases.....	35
3.2 - Metodologia empregada para construção do mapa de Patterson diferença.....	42
3.3 - Resolução do mapa de Patterson diferença do derivado 2-cloromercúrio 4-nitrofenol.....	44
3.4 - Determinação das posições dos átomos pesados do ácido mersálico e geração inicial das fases.....	51
3.5 - Determinação das posições dos átomos pesados do derivado $K_2PtCl_4$ e geração das fases MIR.....	53
3.6 - Localização dos sítios dos átomos pesados na proteína nativa.....	56

4 - CONSTRUÇÃO E REFINAMENTO DO MODELO DA PROTEÍNA NATIVA.....	59
4.1 - O método de nivelamento de solvente ( <i>solvent flattening</i> ).....	59
4.2 - Início da construção do modelo da GlcN6P desaminase.....	64
4.3 - Refinamento do modelo.....	65
4.3.1 - Geração da estrutura molecular: Topologia e Parâmetros.....	68
4.3.2 - Minimização de energia.....	69
4.3.2.1 - Minimização de corpo rígido.....	70
4.3.2.2 - Minimização de posição ( <i>positional refinement</i> ).....	70
4.3.3 - Dinâmica molecular com temperatura controlada ( <i>simulated annealing</i> ).....	70
4.3.4 - Refinamento dos fatores de temperatura (fatores B's).....	74
4.3.5 - Índice de discordância livre, <i>R<sub>free</sub></i> .....	75
4.3.6 - Inclusão de moléculas de água no modelo.....	76
4.3.7 - Refinamento da Glucosamina-6-fosfato desaminase.....	76
5. CO-CRISTALIZAÇÃO, COLETA DE DADOS, CONSTRUÇÃO E REFINAMENTO DOS COMPLEXOS: PROTEÍNA+ATIVADOR ALOSTÉRICO E PROTEÍNA+INIBIDOR.....	80
5.1 - Co-cristalização dos complexos.....	80
5.1.1 - Introdução.....	80
5.1.2 - Processo para cristalização de proteínas: <i>salting-out</i> , difusão de vapor e <i>hanging drop</i> .....	81
5.1.3 - Procedimento experimental.....	82
5.1.3.1 - Preparação da proteína nativa.....	82
5.1.3.2 - Determinação das condições para cristalização.....	82
5.1.3.3 - Montagem das placas de cristalização.....	83
5.1.3.4 - Cristais obtidos.....	84
5.2 - Coleta de dados.....	86
5.3 - Construção do modelo dos complexos.....	87
5.4 - Refinamento dos complexos e localização de moléculas de água.....	90
6. RESULTADOS E DISCUSSÃO.....	96
6.1 - Análise da qualidade da estrutura nativa.....	96
6.2 - Descrição da estrutura do monômero e dímero.....	98
6.3 - Descrição da estrutura do hexâmero.....	101
6.4 - Descrição da estrutura dos complexos co-cristalizados. Os sítios catalíticos e alostéricos.....	104
6.4.1 - Complexo enzima ativador-Pi: o sítio alostérico.....	104
6.4.2 - Complexo enzima inibidor-Pi: o sítio catalítico.....	106
6.4.3 - Complexo enzima 2-inibidor.....	108

7. CONCLUSÕES.....	109
7.1 - Evidências que a estrutura nativa e dos complexos estejam no estado conformacional R: os sítios alostérico e catalítico.....	109
7.2 - Proposição de um mecanismo catalítico.....	110
7.3 - Modelo de Monod-Wyman-Changeux (MWC) e sistema tipo K.....	113
7.4 - Comparações com outras estruturas.....	114
 CONSIDERAÇÕES FINAIS.....	 116
 REFERÊNCIAS BIBLIOGRÁFICAS.....	 117
 APÊNDICE A	
Arquivos de topologia e parâmetros adaptados para os refinamentos dos complexos usando o programa XPLOR: enzima ativador-Pi, enzima inibidor-Pi, enzima 2-inibidor.....	A1
 APÊNDICE B	
Gráficos de análise de qualidade estereoquímica feitos pelo programa <i>PROCHECK</i> para a estrutura nativa.....	B1
 APÊNDICE C	
Gráficos de análise da qualidade estereoquímica feitos pelo programa <i>PROCHECK</i> para os complexos: enzima ativador-Pi, enzima inibidor-Pi, enzima 2-inibidor.....	C1

## ABREVIATURAS

GlcN6P desaminase - Glucosamina-6-fosfato desaminase

GlcN - D-glucosamina

GlcNAc - N-acetil-D-glucosamina

GlcN6P - D-glucosamina-6-fosfato

GlcNAc6P - N-acetil-D-glucosamina-6-fosfato

GlcN-ol-6P - 2 desoxi-2-amino D-glucitol 6-fosfato

Fru6P - Frutose-6-fosfato

CMNF - 2-Cloromercúrio-4-nitrofenol

## LISTA DE TABELAS

Tabela 2.1 - Experimentos realizados para obtenção de derivados isomorfos.....	12
Tabela 2.2 - Resumo das coletas de dados para obtenção de derivados isomorfos.....	32
Tabela 3.1 - Vetores de Harker do grupo espacial R32.....	44
Tabela 3.2 - Sítios de ligação do derivado 2-Cloromercúrio 4-nitrofenol.....	50
Tabela 3.3 - Sítios de Ligação do derivado Ácido Mersálico.....	51
Tabela 3.4 - Sítios de ligação do derivado $K_2PtCl_4$ .....	54
Tabela 3.5 - Figura de mérito em função da resolução para reflexões centradas e não centradas.....	55
Tabela 3.6 - <i>Phasing Power</i> em função da resolução para reflexões centradas e não centradas para os derivados de ácido mersálico e $K_2PtCl_4$ .....	56
Tabela 3.7 - Ligações do derivado $K_2PtCl_4$ com a GlcN6P desaminase.....	57
Tabela 4.1 - Índices de discordância em algumas etapas do refinamento.....	77
Tabela 5.1 - Resumo da coleta de dados dos complexos co-cristalizados.....	87
Tabela 5.2 - Dados estatísticos para os complexos .....	91
Tabela 5.3 - Índices de discordância (R) em algumas etapas do refinamento dos complexos.....	91
Tabela 6.1 - Principais contatos entre o ativador alostérico e a enzima.....	105
Tabela 6.2 - Principais contatos entre o inibidor e a enzima.....	106

## LISTA DE FIGURAS

Figura 2.1 - Fluxograma de execução de uma busca por derivados isomorfo.....	10
Figura 2.2 - Transferência de cristal da solução que foi crescido para uma solução com átomos pesados ( <i>soaking</i> ).....	11
Figura 2.3 - Fotografia do sistema de coleta de dados R-AXIS IIC, da Rigaku Corp.....	13
Figura 2.4 - Fotografia do sistema de coleta de dados R-AXIS IIC, da Rigaku Corp., mostrando em detalhes o local onde é colocado o cristal para experimentos de difração de raios-X.....	14
Figura 2.5 - Esquema mostrando o funcionamento da placa de imagens do sistema R-AXIS IIC.....	16
Figura 2.6 - Esquema mostrando a montagem de cristal proteico dentro de um capilar de vidro, para posterior experimentos de difração de raios-X.....	19
Figura 2.7 - Volume necessário para uma coleta de dados de difração de raios-X completa para cristais do grupo espacial monoclinico.....	21
Figura 2.8 - Difratograma de um experimento de difração de raios-X, mostrando uma imagem de oscilação, vista com a utilização do programa PROCESS (foto da tela da estação gráfica).....	23
Figura 2.9 - Esfera de Ewald (estilizada com uma suposta espessura), mostrando uma reflexão entrando (A) e saindo (B) da condição de difração.....	26
Figura 2.10 - Metodologia para integração de reflexões.....	26
Figura 2.11 - Perfil médio das reflexões de um filme.....	27
Figura 3.1 - Diagrama mostrando uma representação dos vetores fatores de estrutura, $F_P$ (proteína nativa), $F_{PH}$ (proteína+átomo pesado) e $F_H$ (átomo pesado) e suas respectivas fases $\alpha_P$ , $\alpha_{PH}$ , $\alpha_H$ .....	36
Figura 3.2 - Construção de Harker para o cálculo de fases. No método da substituição isomórfica há duas possibilidades para o ângulo de fase da proteína $\alpha_P$ , correspondendo a dois vetores $F_P(1)$ e $F_P(2)$ .....	37
Figura 3.3 - A construção de Harker para determinação da fase pelo método da substituição isomórfica múltipla.....	38
Figura 3.4 - Situação real, onde o triângulo vetorial $F_{PH} = F_P + F_H$ não fecha exatamente. Os valores observados e calculados de $ F_{PH} $ diferem do erro $\epsilon$ .....	39
Figura 3.5 - A probabilidade $P_j(\alpha)$ para uma reflexão ter $\alpha$ como ângulo de fase correto obtida do derivado $j$ é mostrado como função de todos os ângulos entre 0 e $2\pi$ .....	39
Figura 3.6 - Exemplos da probabilidade total $P(\alpha)$ para o ângulo de fase $\alpha$ de uma reflexão obtida de mais de um derivado.....	40
Figura 3.7 - Seção de Harker para $Z=0$ ( $u, v, 0$ ), para o derivado CMNF.....	45
Figura 3.8 - Seção de Harker do tipo $-x, x, 2z$ (plano $uw$ ) para o derivado 2-cloromercúrio 4-nitrofenol.....	48
Figura 3.9 - A função de Patterson para um conjunto de $N$ átomos na cela unitária (neste caso A,B,C e D) que contém $N$ imagens do conjunto.....	49

Figura 3.10 - Seção de Harker $Z=0$ para o derivado de ácido mersálico.....	51
Figura 3.11 - Seção de Harker $-x, x, 2z$ para o derivado de ácido mersálico.....	52
Figura 3.12 - Seção de Harker $Z=0$ para o derivado $K_2PtCl_4$ .....	53
Figura 3.13 - Seção de Harker $-x, x, 2z$ para o derivado $K_2PtCl_4$ .....	54
Figura 4.1 - Fluxograma do método de nivelamento de solvente ( <i>solvent flattening</i> ).....	62
Figura 4.2a, 4.2b - Seções ( $Z=20$ e $33$ ) do mapa densidade eletrônica (com fases MIR até $3.0 \text{ \AA}$ ), mostrando o envelope molecular utilizado para o método de nivelamento de solvente.....	63
Figura 4.3 - Comparação do mapa densidade eletrônica com a cadeia principal traçada de uma região da GlcN6P desaminase, antes (à direita) e após (à esquerda) a utilização do método de nivelamento de solvente.....	64
Figura 4.4 - Fluxograma para refinamento do modelo.....	67
Figura 4.5 - Distribuição de velocidades de Maxwell-Boltzmann para água em várias temperaturas.....	73
Figura 4.6 - Mapa densidade eletrônica com o modelo refinado, mostrando parte de uma folha $\beta$ pregueada.....	77
Figura 4.7 - Mapa densidade eletrônica com o modelo refinado, mostrando parte de uma hélice- $\alpha$ .....	78
Figura 4.8 - Gráfico mostrando os fatores de temperatura ( $\text{\AA}^2$ ) versus carbono $\alpha$ de cada aminoácido.....	79
Figura 4.9 - Gráfico mostrando o valor médio dos fatores de temperatura ( $\text{\AA}^2$ ) versus átomos da cadeia lateral.....	79
Figura 5.1 - Método para cristalização: <i>hanging drop</i> (gota suspensa).....	81
Figura 5.2 - Arranjo utilizado para testes de cristalização (placas de cultura celular).....	83
Figura 5.3-Cristais de enzima+ativador alostérico crescidos em $K_2H/NaH_2 PO_4$ $0.72 \text{ M}$ , pH $8.2$ .....	85
Figura 5.4 - Cristais de enzima+inibidor crescidos em $K_2H/NaH_2 PO_4$ $0.64 \text{ M}$ , pH $8.2$ .....	85
Figura 5.5 - Cristais de enzima+inibidor crescidos em acetato de sódio, $0.8 \text{ M}$ , pH $8.1$ .....	86
Figura 5.6 - Mapa diferença ( $3\sigma$ ) mostrando o ativador (complexo de proteína e ativador em precipitante Na/K fosfato), localizado entre dois monômeros.....	88
Figura 5.7 - Mapa diferença ( $3\sigma$ ) mostrando o inibidor (complexo de proteína e inibidor em precipitante Na/K fosfato), localizado no interior de um monômero.....	88
Figura 5.8 - Mapa diferença ( $2.5\sigma$ ) mostrando o inibidor (complexo de proteína e inibidor em precipitante acetato de sódio), localizado entre dois monômeros.....	89
Figura 5.9 - Gráfico mostrando os fatores de temperatura ( $\text{\AA}^2$ ) versus carbono $\alpha$ de cada aminoácido do complexo enzima ativador-Pi.....	92
Figura 5.10 - Gráfico mostrando os fatores de temperatura ( $\text{\AA}^2$ ) versus carbono $\alpha$ de cada aminoácido do complexo enzima inibidor-Pi.....	92

Figura 5.11 - Gráfico mostrando os fatores de temperatura ( $\text{\AA}^2$ ) versus carbono $\alpha$ de cada aminoácido do complexo enzima 2-inibidor.....	93
Figura 5.12 - Modelo do ativador no sítio alostérico (complexo de proteína e ativador em precipitante Na/K fosfato), após o refinamento utilizando o programa X-PLOR.....	94
Figura 5.13 - Modelo do inibidor no sítio catalítico (complexo de proteína e inibidor em precipitante fosfato Na/K ), após o refinamento utilizando o programa X-PLOR.....	94
Figura 5.14 - Modelo do inibidor no sítio alostérico (complexo de proteína e inibidor em precipitante acetato de sódio), após o refinamento utilizando o programa X-PLOR.....	95
Figura 5.15 - Modelo do monômero da GlcN6P desaminase com as moléculas do ativador e do inibidor nos sítios alostérico e catalítico.....	95
Figura 6.1 - Gráfico de Ramachandran do modelo da Glucosamina 6-fosfato desaminase.....	97
Figura 6.2 - Resultado da análise de qualidade dos resíduos da GlcN6P desaminase feita pelo programa <i>Quality</i> .....	98
Figura 6.3 - Enovelamento do monômero da GlcN6P desaminase, mostrando as estruturas secundárias.....	99
Figura 6.4 - Esquema mostrando a topologia da GlcN6P desaminase.....	99
Figura 6.5 - Dímero da GlcN6P desaminase presente na unidade assimétrica.....	100
Figura 6.6 - Destaque da interface de ligação entre os dímeros da GlcN6P desaminase....	100
Figura 6.7 - Hexâmero da GlcN6P desaminase. À direita é mostrada a vista de cima e à esquerda a vista lateral.....	102
Figura 6.8 - Densidade eletrônica mostrando o resíduo Cys 219 de cada um dos monômeros do trímero formando uma possível ponte dissulfeto e SH livre em uma distribuição randômica ao longo das moléculas do cristal.....	102
Figura 6.9 - Representação do trímero da GlcN6P desaminase, mostrando a ponte dissulfeto entre estes e os íons fosfato (dois por monômero).....	103
Figura 6.10 - Destaque para a interseção entre os monômeros formadores do trímero da GlcN6P desaminase, mostrando a ponte dissulfeto entre as três Cys 219 e os íons fosfato na interface entre monômeros.....	103
Figura 6.11 - Representação do ativador alostérico e seu ambiente de ligação, destacando os aminoácidos mais importantes na sua ligação.....	105
Figura 6.12 - Representação do inibidor e seu ambiente de ligação.....	107
Figura 7.1 - Mecanismo catalítico proposto para a Glucosamina 6-fosfato desaminase.....	112
Figura 7.2 - Modelo para o substrato GlcN6P na forma de anel piranose no sítio catalítico da enzima, construído a partir da estrutura do inibidor GlcN-ol-6P.....	113
Figura 7.3 - Monômero da GlcN6P desaminase. A parte mostrada em <u>amarelo</u> tem enovelamento similar ao NAD <i>binding domain</i> , em <u>verde</u> é mostrado o N-terminal (1-35), em <u>vermelho</u> o C-terminal (243-266) e em <u>azul</u> o segmento 143-188.....	115

## RESUMO

A enzima Glucosamina-6-fosfato desaminase (GlcN6P desaminase) é envolvida na conversão reversível da D-glucosamina-6-fosfato (GlcN6P) em Fru6P e amônia, como parte do caminho metabólico de aminoaçúcares como fonte de energia celular. A enzima hexamérica (peso mol. 178200) exibe uma cooperatividade homotrópica intensa em direção à GlcN6P a qual é modulada alostericamente pelo ativador N-acetil-D-glucosamina 6-fosfato (GlcNAc6P). A GlcN6P desaminase foi cristalizada no grupo espacial R32, com parâmetros de rede  $a = b = 125.9 \text{ \AA}$  e  $c = 223.2 \text{ \AA}$  e um conjunto de dados à  $2.1 \text{ \AA}$  de resolução foi coletado usando radiação de luz síncrotron (Horjales *et al.*, 1992). A procura no banco de dados de seqüências OWL não mostrou homologia significativa com qualquer outra família de proteína, desta maneira a determinação da estrutura foi feita pela técnica de substituição isomórfica múltipla (MIR) a partir de dois derivados, um composto de platina, o  $K_2PtCl_4$  e um complexo de mercúrio, o ácido mersálico. O mapa MIR a  $3 \text{ \AA}$  de resolução mostrou contornos claros e utilizando técnicas de nivelamento de solvente (*solvent flattening*) estendeu-se as fases até  $2.5 \text{ \AA}$ . A enzima cristaliza-se com dois monômeros na unidade assimétrica. A densidade eletrônica final foi interpretada com o auxílio do programa gráfico 'O', sendo possível determinar sem ambigüidade 230 dos 266 resíduos de cada monômero; a partir daí foram usados subseqüentes mapas de Fourier diferença para a localização de todos os outros resíduos. O refinamento do modelo foi feito utilizando o programa X-PLOR (Brünger, 1993), usando a rotina *simulated annealing*, obtendo o fator R final de 17.4 % com 348 moléculas de água e quatro íons inorgânicos de fosfato.

O enovelamento do monômero tem uma estrutura do tipo  $\alpha/\beta$  com uma folha- $\beta$  pregueada paralela central com sete fitas com topologia 4x, 1x, 1x, -3x, -1x, -1x, envolvida por ambos os lados por oito hélices- $\alpha$  e uma hélice  $3_{10}$  com duas voltas. A sexta fita da folha- $\beta$  central tem um prolongamento no C-terminal que faz parte de uma segunda folha- $\beta$  antiparalela de três fitas com topologia 2, -1. O hexâmero tem uma simetria local 32, com dois trímeros empacotados frente-a-frente com uma rotação relativa de  $15^\circ$  em torno do eixo de ordem 3 e ligados por pontes salinas e algumas interações hidrofóbicas em torno do eixo não cristalográfico de ordem 2. As moléculas de cada trímero formam um contato não usual de três resíduos Cis 219 próximo ao eixo de ordem três.

Os complexos com ativador alostérico (GlcNAc6P) e inibidor competitivo (2-desoxi 2-amino glucitol 6-fosfato) foram co-cristalizados isomorficamente com a estrutura nativa. Os mapas Fourier diferença mostram claramente densidades para os ligantes, definindo sem ambigüidade os sítios ativo e alostérico. O refinamento dos complexos produziram a mesma conformação da proteína nativa, na margem de erro experimental. Os sítios alostéricos (seis) estão localizados na interface adjacente dos monômeros de cada trímero e os sítios ativos (ou catalíticos) no lado externo de cada monômero, no C-terminal da folha- $\beta$  central. O monômero tem uma topologia com enovelamento similar a um domínio de ligação de NAD, excluindo os segmentos de aminoácidos 1-35, 145-188 e 243-266. As estruturas dos complexos e da nativa estão em um estado alostérico R em concordância com o modelo MWC para um sistema do tipo K (Monod *et al.*, 1965). Um mecanismo alostérico similar ao da GlcN6P desaminase é encontrado na enzima fosfofrutoquinase (Evans, 1981). Um mecanismo catalítico é proposto para a reação de isomerização-desaminação da enzima GlcN6P desaminase a partir do mecanismo geral para aldose-cetona isomerases.

## ABSTRACT

The enzyme Glucosamine-6-phosphate deaminase (GlcN6P deaminase) is involved in the reversible conversion of D-glucosamine-6-phosphate (GlcN6P) into Fru6P and ammonia. The hexameric enzyme (mol.wt.=178200) exhibits an intense homotropic co-operativity towards GlcN6P which is allosterically modulated by the activator N-acetyl-D-glucosamine 6-phosphate (GlcNAc6P).

The GlcN6P deaminase was crystallized in space group R32, with cell parameters  $a = b = 125.9 \text{ \AA}$  and  $c = 223.2 \text{ \AA}$  and a native dataset was collected to  $2.1 \text{ \AA}$  resolution at a synchrotron source (Horjales *et al.*, 1992). A search of the OWL sequences database has shown no significant homology with any other known protein family. Therefore, the structure determination will have to be achieved through the Multiple Isomorphous Replacement technique from two isomorphous derivatives, a platinum compound  $\text{K}_2\text{PtCl}_4$  and a mercury complex, mersalyl acid. The MIR map at  $3 \text{ \AA}$  resolution showed clear molecular boundaries and solvent flattening techniques (Wang, 1985) were used to extend the phase set to  $2.5 \text{ \AA}$ . The final electron density map was interpreted with the aid of the graphic program 'O'. The enzyme crystallizes with a dimer in the asymmetric unit and 230 out of the total 266 residues of each crystallographically independent monomer could be unambiguously identified in the map. The remaining residues were located after subsequent difference Fourier maps. The refinement was made with program X-PLOR (Brunger, 1993), using the simulated annealing routine, obtained  $R=17.4 \%$  with 348 water molecules and four inorganic phosphate ions.

The monomer fold shows an  $\alpha/\beta$  structure with a central 7-stranded  $\beta$ -sheet with topology 4x, 1x, 1x, -3x, -1x, -1x, surrounded on both sides by eight  $\alpha$ -helices and 2-turn  $3_{10}$ -helix. The sixth strand of the central  $\beta$ -sheet is common to a second 3-stranded anti-parallel  $\beta$ -sheet with topology 2, -1. The hexamer has local 32 symmetry, with two trimers packed in a face-to-face arrangement with a relative rotation of  $15^\circ$  around the 3-fold axis, and linked together by salt-bridge and some hydrophobic contacts. The molecules of each trimer have extensive contacts and show an unusual feature of the three Cys219 residues closely clustered around the 3-fold axis.

The complexes with allosteric activator (GlcNAc6P) and inhibitor (2-deoxy-2-amino glucitol 6-phosphate) were co-crystallized isomorphously with the native structure. The difference Fourier maps show clear density for the ligands, unambiguously defining the active and allosteric sites. The complexes refinement produced the same conformation of the native, within experimental error. The allosteric sites are located at the interfaces of adjacent monomers from each trimer and the active sites (or catalytic) lie at the external side of each monomer, at the C-terminal end of the central parallel  $\beta$ -sheet. The monomer has a similar folding topology as a typical NAD binding domain, excluding the segments of aminoacids 1-35, 145-188 and 243-266. The native and complexes structures are at the allosteric state R concerted with MWC model for a K-system (Monod *et al.*, 1965). A similar allosteric mechanism is found in the enzyme phosphofructokinase (Evans, 1981). A catalytic mechanism is proposed for the isomerisation-deamination reaction of the enzyme from general mechanism for aldo-keto isomerases.

## INTRODUÇÃO

No desenvolvimento da ciência, encontramos hoje em uma nova realidade, a chamada interdisciplinaridade, ou seja, a união dos diversos ramos da ciência, incluindo aí as ciências exatas e biológicas.

O desenvolvimento do conhecimento em ciências biológicas, progrediu grandemente no momento em que iniciou-se o estudo do universo biológico a nível microscópico, começando este estudo com microscópios óticos. Concomitantemente, as ciências exatas iniciaram o estudo da matéria microscópica (não biológica) com o desenvolvimento de instrumentação para esta, utilizando espectroscopia e difração de raios-X (cuja a base teórica foi desenvolvida no início do século). Com o intuito de conhecer cada vez a nível mais aprofundado os sistemas biológicos, introduziu-se novos equipamentos aos experimentos, tal como microscópios eletrônicos e começou-se a utilizar a espectroscopia.

Porém, somente foi possível a 'visualização' de sistemas biológicos à nível atômico com o uso da difração de raios-X, que foi iniciado por Kendrew e Perutz na década de 50, culminando com a resolução da estrutura à baixa resolução da mioglobina de baleia (Kendrew, 1958) e da hemoglobina humana por Perutz (Perutz, 1960). Com o passar dos anos, a obtenção de modelos tridimensionais de estruturas de macromoléculas desenvolveu-se em uma velocidade muito rápida (há mais de 2500 estruturas macromoleculares determinadas atualmente), motivada por facilidades técnicas que se dispõem hoje como, estações gráficas de alta definição e fontes de raios-X de alta intensidade (radiação sincrotron), juntamente com novas técnicas bioquímicas e biológicas que foram desenvolvidas com o passar destes anos, tal como o sequenciamento de proteínas e tecnologia de DNA recombinante, sendo possível, hoje em dia, resolver estruturas com 1 Å de resolução, estruturas de ácidos nucleicos e vírus. Toda esta bagagem estrutural tornou-se determinante no desenvolvimento de diversas áreas da biologia e medicina, tais como a enzimologia, imunologia e biologia molecular etc.

O desenvolvimento destas técnicas resultou na interação de profissionais de diversos ramos da ciência, no qual é necessário que todos falem uma linguagem comum, de maneira a chegar a conclusões conjuntas. Desta maneira, a necessidade da interdisciplinaridade está ocorrendo em toda a comunidade científica onde os cientistas, sejam físicos, bioquímicos, químicos, biólogos e profissionais da área de computação trabalham resolvendo problemas em comum.

A cristalografia de proteínas é uma ferramenta poderosa na compreensão de mecanismos biológicos como aplicação acadêmica e é também poderosa nas aplicações tecnológicas, tais como no desenvolvimento de drogas e vacinas.

Assim, o trabalho a ser descrito nesta tese se insere neste contexto, no qual uma enzima que tem sido estudada a vários anos por bioquímicos, os quais

obtiveram várias conclusões, é estudada agora à nível estrutural onde se tentará provar várias suposições e obter novos resultados sobre esta proteína.

A enzima GlcN6P desaminase, cujas amostras de proteína purificada foram fornecidas pelos Drs. Mário L. Calcagno e Myriam Marlenne Altamirano, foi a primeira proteína cristalizada a ter sua a estrutura resolvida à nível atômico no Brasil (Dep. Física e Informática - IFSC-USP).

Desta maneira, devido ao pioneirismo deste trabalho na cristalografia de proteínas e não na cristalografia de pequenas moléculas (moléculas com menos de 200 átomos), cujo o trabalho é desenvolvido no FFI-IFSC/USP desde a década de 70 tendo várias dissertações e teses escritas, esta tese é centrada nos pontos onde é marcante a diferença entre as técnicas utilizadas para pequenas e macro moléculas.

A divisão da tese é feita com a ordem cronológica da resolução, isto é, o primeiro capítulo é uma revisão dos conhecimentos sobre a enzima até o momento do início do trabalho, o segundo mostra os experimentos para obtenção de derivados isomorfos, o terceiro descreve a obtenção das fases, o quarto refinamento da estrutura, quinto obtenção, coleta e refinamento dos complexos proteína com ativador e inibidor, no sexto são mostrados os resultados e discussão dos dados obtidos e no sétimo, as conclusões mais relevantes.

Tentou-se desta maneira, abordar sucintamente os assuntos teóricos e práticos da resolução de uma estrutura, exemplificando com uma situação real da estrutura resolvida por nós. Na parte teórica da tese, foi dada ênfase aos assuntos onde a teoria é muito escassa e para assuntos já amplamente discutidos, houve uma breve introdução ou somente uma indicação de uma referência de boa qualidade e fácil compreensão.

# CAPÍTULO 1

## A ENZIMA GLUCOSAMINA 6-FOSFATO DESAMINASE DE *Escherichia coli*

A glucosamina 6-fosfato desaminase é uma enzima com características bastante interessantes, tal como o seu mecanismo alostérico, que conduz a mudanças conformacionais com a ligação do ativador alostérico ou com o substrato da reação, possibilitando um estudo bastante amplo e interdisciplinar. Aliado a estes fatos, a proteína que foi clonada, permite um estudo mais fácil a nível experimental, pois a estrutura primária pode ser mais facilmente estabelecida e pode-se produzir proteína pura em grandes quantidades. Além disso, a proteína pode ser purificada e cristalizada com facilidade.

A seguir, serão mostradas as características principais desta enzima, bem como serão descritos resumidamente os experimentos de cristalização e coleta de dados de difração de raios-X feitos antes do início desta tese.

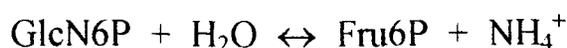
### 1.1 - CARACTERÍSTICAS PRINCIPAIS

Os açúcares aminados, N-acetil-D-glucosamina (GlcNAc) e D-glucosamina (GlcN) são essenciais para o crescimento bacteriano, pois são componentes essenciais na constituição da parede celular e membrana externa (Mirelman, 1976 e Raetz, 1987). Além disso, tanto GlcNAc como GlcN são excelentes fontes de carbono e nitrogênio para a bactéria (Bachmann, 1990).

Quando as células de *Escherichia coli* são cultivadas na presença de GlcNAc (ou GlcN), a síntese das proteínas necessárias para a conversão desta para Frutose 6-fosfato (Fru6P), é induzida; uma destas proteínas, codificada pelo gen *nagB* da *E. coli*, é a Glucosamina 6-fosfato desaminase.

De maneira resumida, as reações do metabolismo que ocorrem culminando com a síntese da GlcN6P desaminase são as seguintes:

1. Transporte de GlcNAc (ou GlcN) para dentro da célula, fazendo com que estes açúcares sejam fosforilados na posição 6.
2. A GlcNAc6P se desacetila pela ação da enzima N-acetil glucosamina-6-fosfato desacetilase que a converte em GlcN6P
3. A GlcN6P é convertida em Fru6P e amônia pela enzima GlcN6P desaminase, por meio da reação reversível (esta reação é uma isomerização aldose-cetona acoplada com uma desaminação (Hodge, 1955) ):



Na ausência de GlcNAc ou GlcN na cultura celular, a bactéria *Escherichia coli* sintetiza GlcN6P desaminase a partir de Fru6P e L-glutamina (Sarvas, 1971; Wu *et al.*, 1971).

A enzima apresenta uma cooperatividade homotrópica<sup>1</sup> em relação aos seus substratos, GlcN6P e Fru6P, na direção direta e reversa da reação, porém bem mais intensa com relação ao substrato GlcN6P (cinética descrita pelo modelo de Monod *et al.* (veja nota de rodapé da página 6)). O controle da regulação da enzima se dá tanto a nível transcricional (presentes no regulon *nag* do mapa cromossômico da *Escherichia coli*) como a nível enzimático feito pelo metabólito GlcNAc6P (que controla tanto a expressão como a atividade da enzima). GlcNAc6P é um ativador de ligação exclusivo da enzima, formando um mecanismo alostérico (mecanismo pelo qual a enzima muda de conformação tridimensional e o ativador (ou inibidor) alostérico, geralmente diferente do substrato, se liga em um sítio diferente do catalítico) (Midelfort e Rose, 1977; Calcagno *et al.*, 1984).

A GlcN6P desaminase é uma proteína oligomérica composta de seis cadeias polipeptídicas idênticas (Calcagno *et al.*, 1984) cuja estrutura primária é conhecida da seqüência de DNA do gene *nagB* da *Escherichia coli* (Rogers *et al.*, 1988), tendo 266 aminoácidos.

A enzima GlcN6P desaminase de *E. coli* foi purificada, clonada e extensivamente estudada (veja item 1.2). Foi também, purificada e clonada do fungo *Candida albicans* (Das e Datta, 1992), tendo sido ainda, identificada em diversos animais, fungos e bactérias (Noltmann, 1972).

## 1.2 - REVISÃO DO CONHECIMENTO A NÍVEIS ESTRUTURAIS E FUNCIONAIS SOBRE A GlcN6P DESAMINASE DE *E. coli*

A enzima Glucosamina 6-fosfato desaminase de *Escherichia coli* K12 vem sendo extensivamente estudada (centralizado nos estudos cinéticos e estruturais da mesma), desde o início da década de 80, pelos Drs. Mário L. Calcagno, Myriam Marlenne Altamirano e colaboradores da Faculdade de Medicina na Universidade Nacional Autônoma do México (cidade do México) e pela Dra. Jacqueline A. Prumbridge do IBPC (CNRS), Paris, França. Myriam Altamirano trabalhou com esta durante o seu mestrado e doutorado (com término em julho de 1993).

A purificação foi feita construindo-se uma coluna de afinidade, utilizando-se um análogo do ativador alostérico (N- $\epsilon$ -aminocaproil-GlcN6P) imobilizado em

---

<sup>1</sup> A cooperatividade ocorre quando a ligação de um ligante em um sítio da proteína altera a afinidade dos outros sítios. No caso dos sítios serem idênticos e com as mesmas propriedades intrínsecas de ligação, a cooperatividade é chamada de homotrópica.

agarose. A enzima se fixa a pH 7.5, temperatura ambiente e tampão de Tris-HCl 0.1 M. A enzima elui com uma solução de GlcNAc6P 0.01 M (Calcagno, *et al.*, 1984).

A clonagem do regulon *nag* do mapa cromossômico da *E. coli* por Plumbridge, *et al.*, 1989, permitiu obter a proteína em grandes quantidades (essencial para ensaios de cristalização) com menor tempo e custo, facilitando muito o estudo cinético e estrutural da enzima, e possibilitando determinar a estrutura primária (Rogers *et al.*, 1988) da mesma praticamente sem ambigüidade.

A proteína tem estrutura quaternária formada por seis subunidades idênticas de peso molecular 29.7 kDa cada uma, possui ponto isoelétrico de 6.1, volume parcial específico de 0.737 cm<sup>3</sup>/g e K<sub>m</sub> de 2 mM para GlcN6P desaminase (Calcagno *et al.*, 1984)<sup>2</sup>.

Um resumo com os resultados mais relevantes (do ponto de vista do trabalho a ser desenvolvido nesta tese) sobre a enzima é dado abaixo:

1. Estrutura primária é conhecida da seqüência de DNA do gen *nagB* da *Escherichia coli* (Rogers *et al.*, 1988).
2. A Enzima é um hexâmero composto de seis cadeias polipeptídicas idênticas (Calcagno *et al.*, 1984).
3. Estrutura quaternária é organizada como um trímero de dímeros (Hernández-Arana, 1993; Altamirano *et al.*, 1993).
4. A GlcN6P desaminase é ativada alostericamente por GlcNAc6P. (Comb e Roseman, 1958; Midelfort e Rose, 1977, Calcagno *et al.*, 1984).
5. Experimentalmente demonstrou-se a presença de pelo menos duas conformações da enzima: não ativa (T) e ativa (R) (Altamirano *et al.*, 1989)<sup>3</sup>.
6. Há um sítio alostérico por monômero ou seis por hexâmero, porém o mecanismo de alostericidade somente ocorre com a enzima hexamérica (Altamirano, 1993-1).
7. Inibição por Glc-ol-6P, comportando-se como um ligante homotrópico, causando a mudança conformacional para o estado mais ativo (R) a baixas concentrações deste (Midelfort e Rose, 1977; Altamirano, 1993-1).

<sup>2</sup> O mecanismo de cinética de reação proposto por L. Michaelis e M. L. Menten é um modelo simples para descrever as propriedades de muitas enzimas. Para tanto, pode-se considerar o seguinte esquema:



onde E é a enzima, S o substrato, ES é o complexo enzima-substrato, P o produto da reação e K<sub>S</sub> e k<sub>cat</sub> são as constantes de dissociação da reação. A constante de Michaelis-Menten (K<sub>m</sub>) é definida como K<sub>m</sub> = 1/2 V<sub>máx</sub>, onde V<sub>máx</sub> é a concentração máxima do substrato (ou de saturação) que altera a razão de formação de produtos de uma reação catalisada. Segundo a teoria de Michaelis-Menten, para vários mecanismos de catálise K<sub>M</sub> = K<sub>S</sub>, ou seja, K<sub>M</sub> é a constante de dissociação do complexo ES (veja Fersht, 1985, p.99-105).

## 8. Grupos SH:

- Das quatro cisteínas presentes na cadeia (Cys 118, 219, 228, 239), três formam grupos sulfidrilo, das quais dois SH aparecem expostos e reativos (Cys 118 e 239) e um (Cys 228) está aparentemente oculto ao solvente e somente reage na presença de agentes desnaturantes. A cisteína que não forma grupo sulfidrilo (Cys 219) corresponde a uma ponte dissulfeto interpretada como sendo intercatenária (Altamirano *et al.*, 1989).
- A transição alostérica do conformero não ativo (T) para o ativo (R) protege completamente as Cys 118 e 239 contra ataques de agentes químicos, isto é, tornam-se ocultas ao solvente (Altamirano *et al.*, 1989, 1992).
- Foram construídas três formas mutadas de Cys 118 e 239 (cada uma das cisteínas e ambas) para serina, por mutagênese sítio dirigida; onde concluiu-se que as mutações simples ou dupla afetam a eficiência catalítica de maneira similar (mesmo resultado tinha sido obtido por bloqueio químico aos SH), porém somente o mutante do resíduo Cys 239 por Ser modificou substancialmente o mecanismo de alostericidade da enzima. A enzima mutada de Cys 239 por Ser tem uma possível conformação intermediária entre T e R (Altamirano *et al.*, 1992).

## 9. Tyr 121:

- Evidências da presença de um resíduo tirosina no sítio alostérico de GlcN6P desaminase (usando Dicroísmo Circular) (Altamirano, 1993-2).
- A substituição, por mutagênese sítio-dirigida, de Tyr 121 (aminoácido Tyr com maior probabilidade de estar na superfície) por Thr e Trp, produz diferenças nas propriedades catalíticas e alostéricas em relação à proteína nativa, demonstrando a participação fundamental deste aminoácido na estrutura e função da enzima (Altamirano, 1993-4).

---

<sup>3</sup> A atividade alostérica da GlcN6P desaminase é precisamente descrita pelo modelo de Monod-Wyman-Changeux (MWC) em um sistema do tipo K (Altamirano *et al.* 1989). O modelo de Monod (Monod *et al.*, 1965), descreve o mecanismo alostérico de enzimas com as seguintes características: as enzimas são oligômeros; têm dois estados conformacionais estáveis, T (=tenso) quando a proteína está desligada e R (=relaxado); o estado T tem mais baixa afinidade por ligantes; todos os sítios em cada estado são equivalentes e têm constantes de ligação idênticas. As enzimas alostéricas descritas pelo mecanismo de Monod *et al.*, podem pertencer a dois tipos de sistemas, do tipo K ou do tipo V. Em sistemas do tipo K (=ligação), as moléculas do substrato e do ativador têm diferentes afinidades pelos estados T e R e a ligação do ativador altera a afinidade da enzima pelo substrato devido às mudanças conformacionais provocadas por este; além disso, os estados R e T têm o mesmo valor intrínseco de  $k_{cat}$  sendo sua atividade modulada por mudanças de afinidade pelo substrato. Enquanto para sistema V (=razão), o substrato tem a mesma afinidade pelos dois estados (T e R), porém um dos estados tem um valor de  $k_{cat}$  muito mais alto. O ativador liga preferencialmente a um dos estados e então modula a atividade por mudanças na posição de equilíbrio entre os dois estados. Como o substrato se liga bem em ambos os estados, não há cooperatividade pela ligação do substrato (veja Fersht, 1985, p.269).

10. Estrutura secundária: por dicroísmo circular (CD) e predição de estrutura secundária empírica baseada em estrutura primária. Conteúdo de hélices- $\alpha$  (37% predição e 34% CD) e de folhas  $\beta$  (22% predição e 20% CD): estrutura tipo  $\alpha/\beta$  (Altamirano *et al.*, 1991).

### 1.3 - CRISTALIZAÇÃO DA ENZIMA (Horjales, E. *et al.*, 1992)

Os cristais da Glucosamina 6-fosfato desaminase nativa foram obtidos no laboratório de cristalografia de proteínas do IFSC pelo pesquisador visitante Dr. Eduardo Horjales, no final de 1991.

A proteína purificada foi armazenada em uma solução estoque de sulfato de amônia 2.5 M e 5 mM de EDTA dissódico. Para cristalização, a proteína estocada foi dialisada contra um tampão de HEPES/NaOH 10mM, pH 7.4. Em seguida, foi feita uma centrifugação retirando-se o sobrenadante e logo após, uma outra centrifugação com membrana semi-permeável (centricon nº 100). A última etapa para preparação é a medida aproximada da concentração da proteína usando um espectrofotômetro (veja metodologia no capítulo 5). A concentração utilizada na cristalização foi da ordem de 10 mg/ml.

A cristalização é feita por difusão de vapor, método da gota suspensa (*hanging drop*) a temperatura de 20°C (para mais detalhes veja cap.5).

Foram obtidos cristais sob diversas condições:

1. Precipitante e tampão de fosfato sódio/potássio ( $\text{NaH}_2\text{PO}_4$  e  $\text{K}_2\text{HPO}_4$ ) 1.4 M (pH 6, 7 e 8).
2. Precipitante de sulfato de amônia ( $(\text{NH}_4)_2\text{SO}_4$ ) 42 % (p/v), tampão Tris/HCl 200 mM (pH 8 e 9).
3. Precipitante de sulfato de amônia ( $(\text{NH}_4)_2\text{SO}_4$ ) 42% (p/v), tampão HEPES/NaOH 200 mM (pH 7).
4. Precipitante de acetato de sódio 42 %, tampão Tris/NaOH 200mM (pH 8).

Os cristais cresceram a partir de cinco dias e completaram o crescimento em aproximadamente três semanas, com dimensões de até 1.0 mm x 0.6 mm x 0.3 mm.

Posteriormente, o material cristalino foi submetido a uma eletroforese em gel de SDS/poliacrilamina na presença de  $\beta$ -mercaptoetanol, confirmando que os monômeros da enzima estavam intactos e migravam identicamente com a enzima antes da cristalização.

### 1.4 - COLETA E ANÁLISE PRELIMINAR DOS DADOS DE DIFRAÇÃO DE RAIOS-X DA PROTEÍNA NATIVA (Horjales *et al.*, 1992).

Para a coleta de dados foram utilizados os cristais crescidos em fosfato de Na/K com pH 8.

Os cristais foram inicialmente caracterizados por fotografias de precessão em uma fonte de raios-X Rigaku RU200B com ânodo rotatório ( $\lambda=1.54 \text{ \AA}$ ) no qual foi possível determinar o grupo espacial, R32, e determinar aproximadamente os parâmetros de rede dos cristais (experimentos feitos no laboratório de cristalografia de proteínas - IFSC).

Um conjunto de dados completo a  $2.1 \text{ \AA}$  foi coletado de um cristal utilizando radiação síncrotron no EMBL Protein Crystallography Station X31 em HASYLAB em DESY, Hamburgo - Alemanha, com um detetor com placa de imagens. Durante a coleta de dados (que demorou mais de 24 h) utilizou-se comprimento de onda de  $1 \text{ \AA}$ , a fim de minimizar os efeitos da radiação ionizante, pois utilizando  $\lambda=1.54 \text{ \AA}$  (fonte convencional) os cristais mostraram ser bem mais sensíveis à radiação.

Foram coletadas 200692 reflexões, sendo 38368 independentes com um  $R_{\text{sym}}=10 \%$  ( $R_{\text{sym}} = \sum_h [\sum_i |I_{hi} - \langle I_h \rangle| / \sum_i I_{hi}]$ ) a  $2.1 \text{ \AA}$ , com uma completude de 99 % dos dados (para mais detalhes: Horjales *et al.*, 1992).

Os parâmetros de rede dos cristais são (sistema hexagonal):  $a=b=125.9$ ,  $c=223.2 \text{ \AA}$ .

Para determinar a simetria não cristalográfica entre os dois monômeros independentes da unidade assimétrica foi utilizada uma função de auto rotação na qual foi observada a presença de um eixo não cristalográfico perpendicular ao eixo cristalográfico de ordem três (eixo c) e deslocado  $15^\circ$  do eixo cristalográfico de ordem dois (eixo a). Indicando que o hexâmero tem simetria pontual 32, centralizado no eixo de ordem três e deslocado da origem.

## CAPÍTULO 2

### PREPARAÇÃO E COLETA DE DADOS DE DIFRAÇÃO DE RAIOS-X DE DERIVADOS ISOMORFOS

A procura no banco de dados de seqüências OWL com 34100 entradas, (Bleasby & Wootton, 1990) não mostrou nenhuma homologia significativa da Glucosamina-6-fosfato desaminase com qualquer outra família de proteínas conhecidas.

Desta maneira, para determinar as fases do conjunto de dados coletado (capítulo 1), podem ser utilizados normalmente os seguintes métodos (Drenth, p.129, 153 e 205):

1. Método da substituição isomórfica (MIR), no qual é necessário a obtenção de cristais com átomos pesados (com P.M., normalmente, acima de 150) ligados à proteína que sejam isomórficos com os nativos, isto é que a ligação destes não alterem o empacotamento cristalino ou a estrutura da proteína.
2. Método do espalhamento anômalo, utilizado em conjunto com o método da substituição isomórfica.
3. Método do espalhamento anômalo com múltiplos comprimentos de onda (MAD), requer o uso de radiação síncrotron sintonizável (no qual possa se utilizar diversos comprimentos de onda) e a presença de átomos espalhadores com P.M. maior que 34 na molécula proteica ou ligado a esta (Hendrickson, 1990).

Nesta tese escolheu-se usar o método da substituição isomórfica múltipla (MIR) (uma breve explicação sobre este método é dada no capítulo 3). Para o uso deste método é necessário a ligação de átomos pesados à proteína, o qual usualmente é feito por co-cristalização (proteína com átomo pesado) ou soaking (adição de átomos pesados por difusão através dos canais do cristal proteico nativo). A obtenção destes derivados é por tentativa e erro onde, para se determinar se os átomos estão ligados é necessário realizar coletas de dados de difração de raios-X destes e comparar as intensidades de difração dos cristais da proteína nativa com os cristais com possíveis átomos pesados.

A figura 2.1 representa um fluxograma de execução de uma busca por derivados isomorfos. Neste processo, tenta-se ligar átomos pesados à proteína (por meio de difusão através do cristal nativo); avalia-se se houve ligação fazendo uma coleta de dados preliminar verificando-se se há diferenças isomorfos (relativo as intensidades do cristais nativos) (veja cap.3).

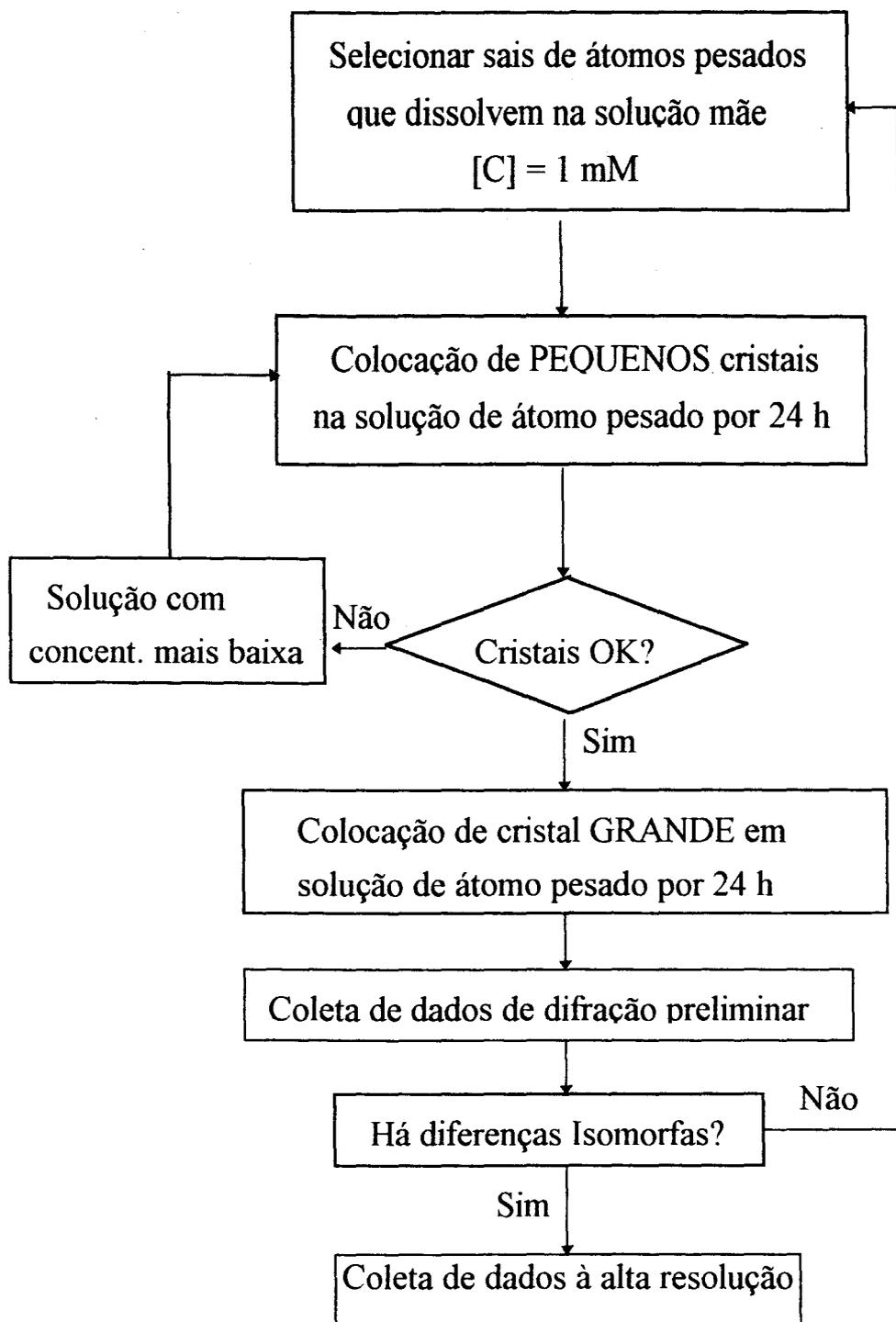


Figura 2.1 - Fluxograma de execução de uma busca por derivados isomorfos

Se há evidências de diferenças isomorfias, pode se fazer uma coleta à alta resolução, caso contrário, pode-se mudar tempo e concentração do sal de átomo

pesado ou tentar-se outro sal e assim sucessivamente. Este processo será discutido em detalhes neste e no próximo capítulo.

## 2.1 - PREPARAÇÃO DOS DERIVADOS: DIFUSÃO EM SOLUÇÃO (SOAKING)

O método de difusão em solução (*soaking*) utilizado nesta tese para obtenção de derivados isomorfos, se baseia no fato que cristais proteicos possuem grande quantidade de solvente (usualmente 50-60% do seu volume), devido ao seu empacotamento cristalino, tendo grandes canais em seu interior preenchidos por solvente. Por estes canais pode se difundir átomos pesados, possibilitando a ligação em sítios específicos da proteína (sem fazer ligações aleatórias com a proteína, que acarreta a destruição da natureza cristalina).

Este experimento pode ser realizado colocando-se um cristal de proteína nativa em uma solução idêntica a utilizada para sua cristalização (solução mãe), porém com um sal de átomo pesado a baixa concentração. A figura 2.2 sumariza o método.

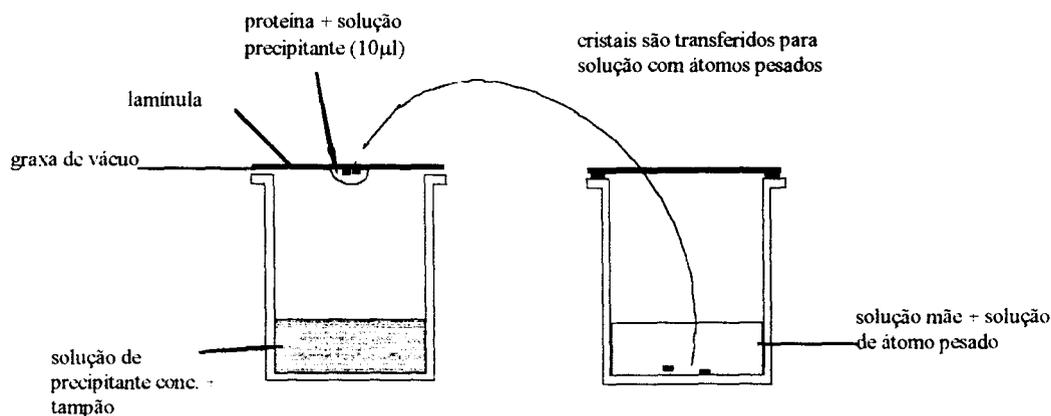


Figura 2.2 - Transferência de cristal da solução em que foi crescido para uma solução com átomos pesados (*soaking*).

Tabela 2.1- Experimentos realizados para obtenção de derivados isomorfos.

Sais de Átomos Pesados	Dissolução	Condições de soaking(mM,h)	Resultado após coleta
ácido benzóico 4-cloromercúrio	não dissolveu totalmente	[C]=1; cristais dissolveram	-
3-Cloromercúrio 2metóxido propil uréia	dissolve com dificuldade.	[C]=0.1, cristais trincaram	-
2-cloromercúrio-4-nitrofenol	dissolve após colocação de $\text{HPO}_4^{2-}$ NaH/K <sub>2</sub>	[C]=0.1; 24,48 h	2 coletas de dados, há diferença isomorfa
ácido tiosalicílico etilmercúrio	não dissolve totalmente	[C]=1; 24 h	coleta de dados, não há diferença isomorfa
ácido sódio benzóico p-hidroximercúrio	dissolve, com agitação	[C]=0.1,0.2, formação de amorfo	-
acetato de mercúrio (II)	há formação de mercúrio metálico	-	-
tiocianato de mercúrio(II)	não dissolve	-	-
tetratiocianetocobaltato(II) mercúrio	dissolve colocando 1º água	[C]=0.1, 24 h cristal explodiu	-
cloreto de mercúrio(I) ácido mersálico	não dissolve	-	-
acetato de fenilmercúrio	dissolve após colocação de $\text{HPO}_4^{2-}$ NaH/K <sub>2</sub>	[C]=0.1; 48,52 h	5 coletas, há diferença isomorfa
acetato de chumbo(II) Trihidratado	dissolve com agitação	[C]=0.2; 24h	-
óxido de chumbo	não dissolve totalmente	[C]=0.1; 24 h cristais trincam	-
tetracloro-aurato(III) de potássio hexacloro	não dissolve totalmente	[C]=0.1,1; 24 h	coleta de dados, não há diferença isomorfa
platinato (IV) de potássio tetracloro	não dissolve totalmente	[C]=1; 24 h	coleta de dados, não há diferença isomorfa
platinato (II) de potássio tetratiociano	dissolve colocando 1º água	[C]=0.5,1; 24 h	coleta de dados, não há diferença isomorfa
platinato (II) potássio dicloreto trans-platina-(II) diamina	precipita ao colocar $\text{HPO}_4^{2-}$ NaH/K <sub>2</sub>	-	-
hexabromo de platina (IV) de potássio (K <sub>2</sub> PtBr <sub>6</sub> )	dissolve colocando 1º água	[C]=1; 24,48	2 coletas. há diferença isomorfa
acetato de európio (III) hidratado	não dissolve	-	-
sulfato de európio (III) hidratado	não dissolve	-	-
acetato de gadolínio(III) hidratado	não dissolve	-	-
sulfato de gadolínio(III) hidratado	dissolve facilmente	[C]=0.1,0.3,1; 24 h	5 coletas. cristais difratam muito pouco
acetato de samário(III) hidratado	dissolve em (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	[C]=1; 24,48 h	-
acetato de uranila	dissolve em (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	[C]=1,5; 22,30 h	2 coletas. não há diferença isomorfa.
oxalato de uranila	dissolve em (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	[C]=5; 24 h	-
sulfato de uranila	dissolve em (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	[C]=5; 24 h	1 coleta. não há diferença isomorfa.
	não dissolve	-	-
	não dissolve	-	-
	dissolve colocando 1º água	[C]=1, 42 h	1 coleta. não há diferença isomorfa.

Os principais problemas enfrentados neste processo são os seguintes:

- Insolubilidade (parcial ou total) do sal na solução mãe.
- Ataque do sais de átomos pesados ao cristal, causando quebra das ligações da rede cristalina (mesmo a baixa concentração e pouco tempo), acarretando trincas, formação de precipitados ou destruição do cristal.

Na tabela 2.1 são sumarizados os testes que foram feitos com os sais de átomos pesados, mostrando quais sais são solúveis na solução mãe, condições de solubilidade, quais suportam o processo de difusão, suas condições (concentração e tempo) e o resultado da coleta de dados (nas que foram possível fazê-lo). Para os experimentos com sais de lantanídeos (Eu, Gd, Sm, U) foram usados cristais crescidos em sulfato de amônia, pois estes sais não eram solúveis em fosfato. Supunha-se que os cristais crescidos em sulfato eram isomorfos aos crescidos em fosfato, pois com a coleta de dados de ambos (sulfato e fosfato nativos) obteve-se os mesmos parâmetros de rede (esta suspeita confirmou-se após refinamento das duas estruturas em fosfato e sulfato). Na coluna dissolução (tabela 2.1), descreve-se a maneira com que foi dissolvido o sal de átomo pesado, isto é, há casos em que o sal dissolve somente com água e não com a solução de Na/K fosfato; em outros casos, a dissolução ocorre com intensa agitação e/ou aquecimento da solução.

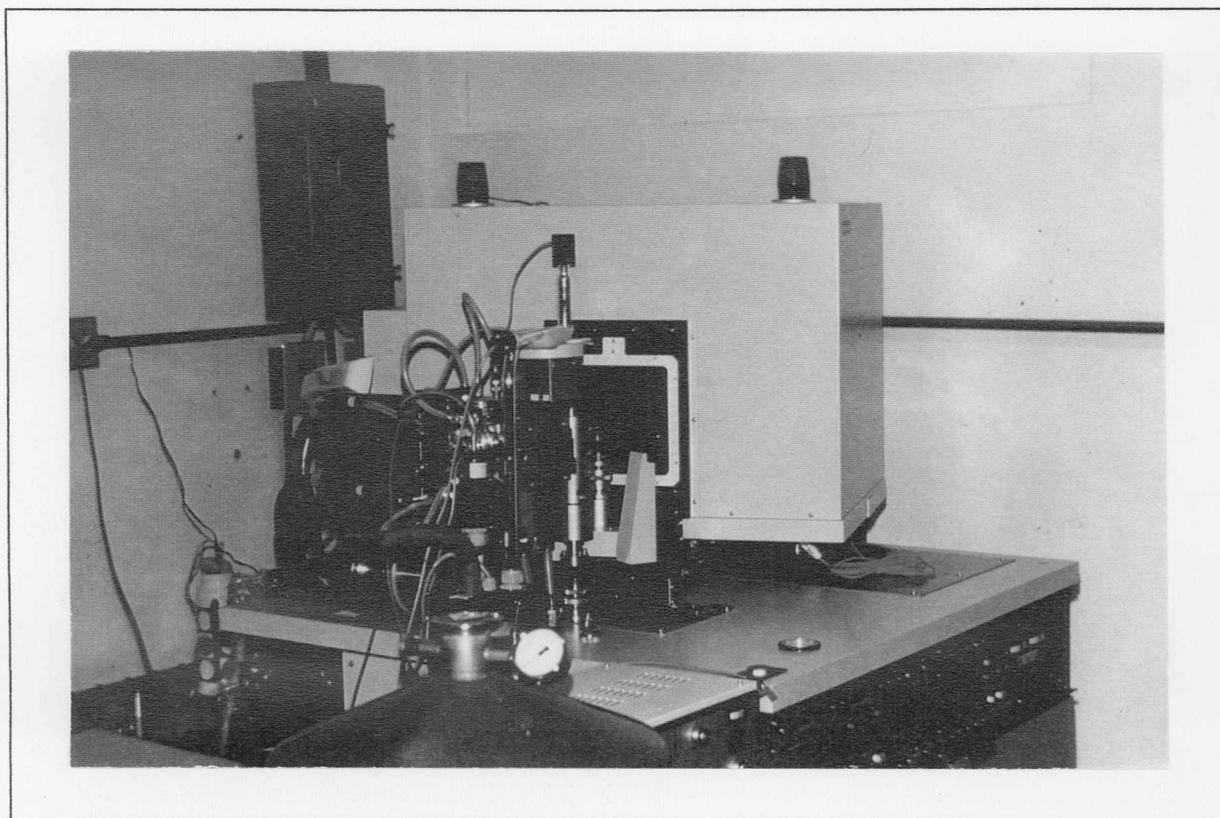


Figura 2.3 - Fotografia do sistema de coleta de dados R-AXIS IIC, da Rigaku Corp.



Figura 2.4 - Fotografia do sistema de coleta de dados R-AXIS IIC, da Rigaku Corp., mostrando em detalhes o local onde é colocado o cristal para experimentos de difração de raios-X.

## 2.2 - COLETA DE DADOS DE DIFRAÇÃO DE RAIOS-X

### 2.2.1 - O Sistema de Coleta de Dados (Placa de Imagem Eletrônica)

O aparelho utilizado para a coleta de dados foi o sistema R-AXIS IIC da Rigaku Corporation, constituído de: fonte de raios-X com ânodo rotatório, monocromatizado por um cristal de grafite (comprimento de onda 1.5418 Å), câmara de oscilação com um grau de liberdade (ângulo  $\phi$ ) e placa de imagem reutilizável com leitura à laser, ligada através de uma interface a uma estação gráfica na qual pode-se visualizar as imagens e processar os dados para obtenção das intensidades integradas (item 2.2.4).

A figura 2.3 é uma foto do sistema e a figura 2.4 é uma foto mostrando, em detalhes, o local onde deve ser colocado o cristal (chamado de cabeça goniométrica), para submetê-lo a difração de raios-X.

A placa de imagens do R-AXIS IIC é um detetor de raios-X bidimensional, com área de  $200 \times 200 \text{ mm}^2$  e possui 1900 por 1900 *pixels* de  $105 \times 105 \text{ }\mu\text{m}^2$  de área cada um destes. É feita de material semicondutor, haleto de bário dopado com európio ( $\text{BaFBr:Eu}^{2+}$ ), que possui características de armazenagem de 'imagens latentes' e 'leitura' posterior (veja explicação abaixo).

O processo de exposição/armazenagem/leitura dos dados pelo equipamento, de maneira simplificada, é o seguinte:

O cristal deve ser montado na cabeça goniométrica (item 2.2.3) e esta, por sua vez, no eixo de oscilação do aparelho (ângulo  $\phi$ ). Durante o experimento de difração o goniômetro oscilará (por ângulo determinado pelo usuário, ver item 2.2.4) enquanto submetido ao feixe de raios-X; os fótons difratados que chegam à placa, excitarão parte dos elétrons do semicondutor ocasionando a mudança de nível energético destes, e desta maneira, as informações do número de fótons que chegaram a placa e sua localização ficam 'armazenadas' (como imagens 'latentes') nos *pixels* desta.

A 'leitura' da placa é feita pela varredura (por meio de um arranjo bidimensional ortogonal, figura 2.5) de um feixe de excitação, um laser de He-Ne (633 nm), o qual leva o semicondutor da placa ao estado intermediário de energia, produzindo luminescência em 390 nm (azul), que é proporcional ao número de fótons de raios-X incidentes na placa, passando em seguida por uma fotomultiplicadora que mede o número de fótons fluorescentes emitidos. Este dados (imagem digitalizada) são transmitidos por uma interface a uma estação gráfica onde as imagens são analisadas.

Após este processo a placa é submetida a uma luz branca que permite o retorno do sistema ao estado fundamental, isto é, a placa é 'apagada' ficando pronta para a próxima coleta. No equipamento há duas placas de imagens, desta maneira, enquanto o processo de leitura de dados é feito com uma, a outra já está sendo submetida a uma nova coleta (Sato, 1992).

Um esquema mostrando o sistema de placa de imagens capaz de armazenar/ler os dados de difração, é dado na figura 2.5.

A placa de imagens eletrônica cumpre a mesma função de um 'filme de raios-X' usado antigamente, porém tem muitas vantagens em relação a este, tais como:

- Rapidez na coleta de dados
- Reutilização
- É pelo menos dez vezes mais sensível do que o filme convencional, sua faixa dinâmica é muito maior ( $1:10^4-10^5$ ), ou seja, permite a aquisição de reflexões mais fracas e mais fortes.

- Tem bastante sensibilidade inclusive para pequenos comprimentos de onda (ex.  $0.7 \text{ \AA}$ ), usado para radiação sincrotron.
- Apresenta baixa radiação de fundo (*background*) intrínseca (os filmes convencionais apresentam um ruído de fundo denominado 'fog químico' decorrente do processo de revelação).

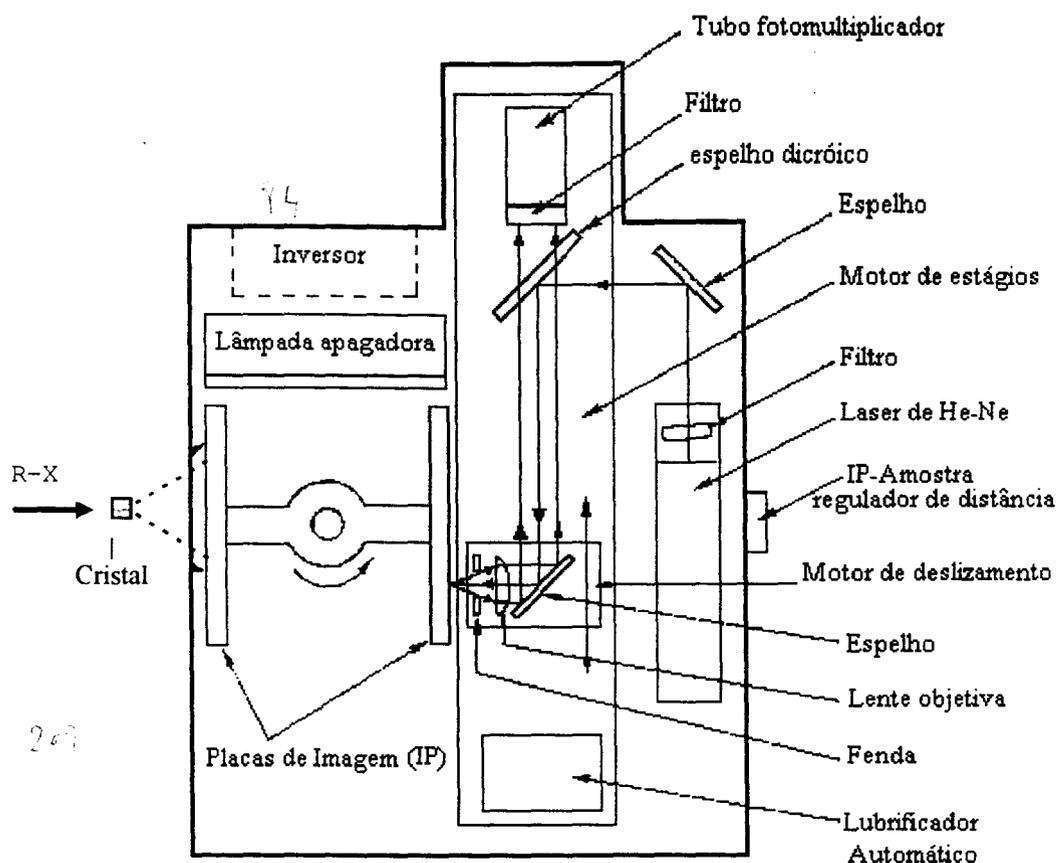


Figura 2.5 - Esquema mostrando o funcionamento da placa de imagens do sistema R-AXIS IIC (Sato, 1992).

### 2.2.2 - Danos aos cristais devido à radiação ionizante, sistema de minimização de danos por resfriamento do cristal

Um dos maiores problemas da difração de raios-X de materiais biológicos é que estes são suscetíveis a danos devido à radiação ionizante. Cristais proteicos são particularmente sensíveis pois os contatos de rede cristalina são relativamente fracos e facilmente quebrados por produtos radiolíticos resultantes de fótons absorvidos. Há dois componentes básicos para os danos de radiação: Danos primários: causados por elétrons energéticos resultantes principalmente dos efeitos fotoelétricos e Compton.

Danos secundários: causados pelas reações dos produtos radiolíticos que se propagam pelo cristal e causam reações posteriores e danos nos contatos de rede das proteínas.

Os processos secundários variam muito entre diferentes proteínas e formas cristalinas diferentes da mesma proteína. Porém, trabalhos recentes (Gonzalez, 1994) vêm mostrando que os efeitos secundários são muito reduzidos quando as coletas de dados de difração de raios-X são feitas à baixa temperatura e é obtido um grande aumento no tempo de vida do cristal.

Além disso os processos secundários se desenvolvem ao longo do tempo, iniciando-se no momento da primeira exposição do cristal aos raios-X, desta maneira, uma coleta rápida é essencial para prevenir estes danos. Assim, o sistema com placa de imagem reutilizável, descrito no item anterior, leva grande vantagem com relação, por exemplo, a um difratômetro de quatro círculos. Similarmente, a utilização de uma fonte de raios-X sincrotron, que possui um feixe de alta intensidade, possibilita uma coleta muito mais rápida que em fontes convencionais.

No sistema utilizado para coleta de dados de difração de raios-X nesta tese, há um sistema de resfriamento de cristais simples e eficiente. O resfriamento é feito submetendo o capilar, onde está o cristal, a um fluxo de nitrogênio gasoso (verticalmente ao capilar) a uma temperatura controlada por uma termopar (ver figura 2.4).

Nos experimentos realizados nesta tese, o processo de resfriamento de cristais mostrou ser bastante eficiente, prolongando o tempo útil de coleta (período no qual o cristal ainda difrata a uma resolução razoável).

### 2.2.3 - Escolha e montagem dos cristais

A escolha do cristal conveniente para coleta deve levar em consideração alguns aspectos:

1. Deve ser uma substância cristalina. A observação do suposto cristal sob microscópio polarizador de luz pode dar uma indicação destas características. Se o material sob observação não mostrar propriedades de birrefringência, o mesmo pode pertencer ao sistema cristalino cúbico, ou pode estar orientado paralelo ao eixo ótico ou simplesmente pode não ser uma substância cristalina.
2. Deve ser um monocristal (um cristal único, e não um aglomerado de cristais). Com o uso do microscópio polarizador de luz, pode-se ter uma indicação se o cristal é único ou um policristal, pois se o cristal em questão sofrer extinção (quando rodado  $360^\circ$ ) por completo, em vez de somente partes deste, pode ser um monocristal.
3. Tamanho conveniente (experimentalmente este tamanho deve ser entre 0.3 e 1.0 mm). Deve ser dada preferência a cristais maiores pois devido ao maior

número de moléculas que podem difratar no cristal, a coleta é feita com maior rapidez, diminuindo desta maneira, os danos devidos à radiação ionizante. (A princípio, há uma limitação ao tamanho dos cristais, pois em cristais muito grandes a absorção da radiação pelo cristal toma valores críticos, comprometendo a coleta).

4. Morfologia e transparência do cristal. Experimentalmente cristais com contornos bem definidos, apresentando transparência, boa aparência e mostrando extinções bem definidas sob microscópio polarizador, costumam ser mais convenientes para experimentos de difração de raios-X.

Para montar cristais proteicos e submetê-los a experimentos de difração de raios-X (tanto no aparelho descrito no item 2.2.1, como em outros equipamentos como a câmara de precessão) deve-se considerar que os cristais de macromoléculas (justamente por serem moléculas de grande porte) possuem grandes canais com solvente, de maneira que suas condições de equilíbrio podem ser facilmente modificadas, pela mudança na concentração ou evaporação da solução dos canais.

Desta maneira, ao manipular os cristais e alterar as condições em que estes estão no momento da retirada do poço de cristalização, os mesmos podem sofrer danos irreversíveis que serão refletidos na qualidade dos dados a serem medidos. Assim, estes são montados de maneira que fiquem sob pressão de vapor com a solução mãe, evitando evaporação desta, e uma eventual 'secagem' e perda do cristal.

Para tanto, os passos utilizados nesta tese para a montagem de cristais foram as seguintes:

1. Retirada da lamínula da placa de cristalização (ver item 5.1.3.3)
2. Colocação de mais solução mãe na lamínula (a fim de evitar que o cristal fique seco devido a evaporação da solução).
3. Usando capilares de vidro com de 1.0 ou 1.5 mm de diâmetro com espessura muito fina (a fim de minimizar a absorção da radiação pelo vidro) acoplados a uma seringa, suga-se o cristal escolhido para dentro do capilar.
4. Coloca-se (sugando novamente com a seringa) uma coluna de solução mãe, fechando-se esta ponta do capilar com cera de abelhas ou cola.
5. O cristal deve ser parcialmente seco. Isto é feito, por exemplo, com pequenas fibras de papel filtro que devem ser inseridas dentro do capilar, evitando tocar o cristal. Este procedimento é feito, pois quando o cristal é sugado para dentro do capilar, traz consigo uma coluna de solução, a qual quando submetida a radiação produz forte espalhamento, aumentando sensivelmente a radiação de fundo (*background*) da coleta de dados. Deve-se ter em mente, que o cristal não deve ser seco demasiadamente, pois é necessário um pouco de solução para mantê-lo preso à parede do capilar por tensão superficial.
6. O capilar deve ser cortado do lado que ainda não foi fechado, com tamanho conveniente para a colocação no aparelho. Para tanto, após a medição de uma

distância entre 1.0 e 1.5 cm do cristal até a ponta aberta do capilar, pinga-se uma gota de cera quente na posição a ser cortada e quebra-se o capilar com uma pinça, evitando-se a trinca do capilar.

7. Coloca-se uma outra coluna de solução mãe (do lado que está aberto do capilar) fecha-se este com cera e acopla-se (usando cera de abelhas) o capilar na cabeça goniométrica (suporte que é colocado no aparelho para coleta de dados).

A figura 2.6 mostra um cristal montado dentro de um capilar.

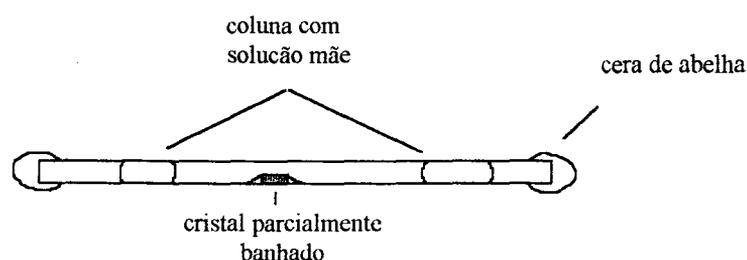


Figura 2.6 - Esquema mostrando a montagem de cristal proteico dentro de um capilar de vidro, para posterior experimentos de difração de raios-X.

A cabeça goniométrica com o capilar e cristal deve ser acoplada ao aparelho e o cristal deve ser alinhado ao feixe de raios-X (a fim de que durante todo o movimento de oscilação do cristal, este fique exposto de maneira igual à radiação). O alinhamento é feito por alguns ajustes presentes na cabeça goniométrica observando-se o cristal em várias posições com auxílio de um microscópio (aumento de 40 vezes) presente no equipamento.

#### 2.2.4 - Estratégias de coleta de dados de difração de raios-X

A fim de obter os melhores dados possíveis do cristal que se dispõe, antes que este sofra danos sérios devido a exposição à radiação ionizante, deve-se fazer algumas considerações, dadas abaixo.

No caso de uma coleta de dados de um cristal proteico no qual nunca foi feito este experimento, não há muitas estratégias a serem definidas, pois não se

conhece seu grupo espacial, sua sensibilidade a radiação ionizante, a qual resolução que este difrata. Resta-se neste caso, devido a experiência com outras proteínas, tentar prever estes fatores experimentais (por exemplo, sabe-se que normalmente quanto maior o peso molecular da macromolécula mais esta sofre danos devido à radiação ionizante, ou seja, o cristal de um vírus sofre muito mais com a radiação que uma pequena enzima).

No caso de cristais de proteínas já manipulados (onde se conhece grupo espacial, parâmetros de rede, sensibilidade do cristal à radiação ionizante), pode-se usar este conhecimento prévio para fazer coletas onde se obtenha o melhor conjunto de dados possível, que implica em obter dados mais completos em menor tempo, com uma redundância adequada para o escalonamento entre os filmes (item 2.2.4.4).

Assim os fatores que se devem levar em consideração para a maximização dos dados, são os seguintes:

1. Grupo Espacial do cristal: (Stout & Jensen, 1968, pp.154-156)

Devido às simetrias dos grupos espaciais e pela lei de Friedel ( $I_{hkl}=I_{-h-k-l}$ , Blundell e Johnson, 1976, p.133), somente uma parte dos dados possíveis de serem medidos serão necessários medir, estes são chamados de dados únicos. Por exemplo, no caso de cristais pertencentes ao grupo espacial triclinico, só é necessário medir metade dos dados (usando lei de Friedel), se o grupo espacial tem uma simetria maior, elementos de simetria adicionais devem ser levados em conta. No caso do grupo espacial monoclinico, há sempre uma simetria 2/m e as reflexões  $h$ ,  $k$ ,  $l$ , podem ser divididas da seguinte maneira (eixo de ordem 2 paralelo a  $b^*$ ):

$$I_{hkl} = I_{-h-k-l} = I_{hk-l} = I_{h-kl} \quad ; \quad I_{-hkl} = I_{h-k-l} = I_{hk-l} = I_{h-kl} \quad \text{e} \quad I_{hkl} \neq I_{-hkl}$$

Desta maneira, se for pensado em termos do espaço recíproco, será necessário um quarto da esfera de reflexões para obter um conjunto completo, como é mostrado na figura 2.7.

Assim, este cálculo pode ser estendido para grupos espaciais com maior simetria e assim, pode-se saber previamente qual a região necessária para uma coleta completa de determinado cristal.

2. Parâmetros de rede de cristal: cristais com parâmetros de rede maiores necessitam de ângulos de oscilação menores (veja explicação mais abaixo).

3. Tamanho do cristal: quanto maior o cristal montado, mais difração ocorrerá em um tempo mais curto, tem-se assim mais condições de obter uma coleta completa antes que o cristal tenha decaimento de difração devido à danos de radiação.

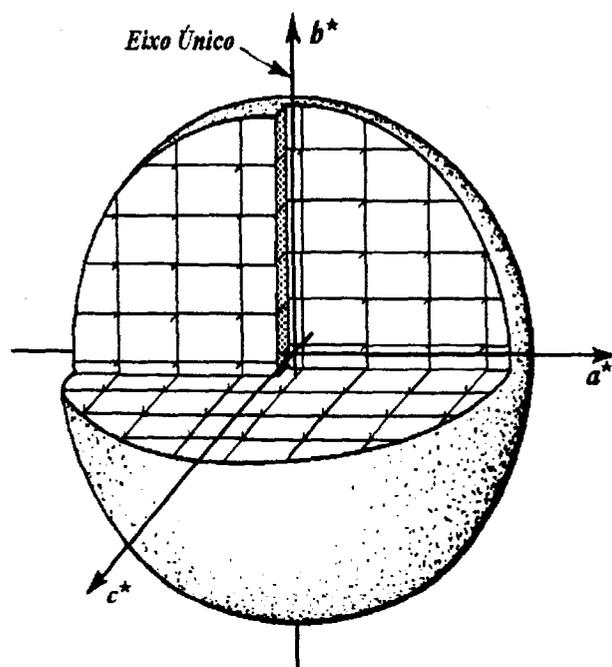


Figura 2.7 - Volume necessário para uma coleta de dados de difração de raios-X completa para cristais do grupo espacial monoclinico (Stout e Jensen, 1989, p.161) .

4. Estado do cristal: durante a montagem, muitas vezes o cristal pode sofrer trincas, ter pontas quebradas, etc; principalmente quando este foi tocado por algum objeto, ou ficou sem solução mãe por algum momento. Este cristal difratará menos e sofrerá danos por radiação mais rapidamente que um cristal em boas condições.

O equipamento utilizado permite alguns ajustes que levam em consideração os itens anteriores, estes são dados abaixo:

1. Ajuste da distância Cristal-Detetor.

Com este ajuste o usuário pode definir em qual resolução será a coleta de dados, ou seja, quanto mais perto o cristal do detetor pode-se gravar reflexões com maior ângulo de Bragg (maior resolução), e vice-versa; porém como desvantagem, quanto mais perto (cristal-detetor) mais próximos os pontos de difração (maior superposição), o que torna-se mais crítico para cristais com celas unitárias grandes (o aparelho permite ajuste de distância entre 55 e 200 mm).

2. Ângulo de oscilação:

Uma coleta com ângulos maiores (p.ex.  $2^\circ$ ) acarreta em uma maior possibilidade de superposição entre os pontos de difração, porém haverá menor número de reflexões parciais que provocam erros no processo de integração dos filmes (veja item 2.2.5.2) e maior número de dados num tempo menor (pois não há necessidade de troca de filme); e vice-versa para ângulos pequenos (p.ex.  $1^\circ$ ).

Assim, como se pode notar, estes ajustes devem ter um compromisso, entre resolução, sobreposição dos pontos de difração, rapidez de coleta e devem ser escolhidos pelo usuário usando, também, sua experiência.

Um recurso para obter um conjunto de dados mais completo possível em menor espaço de tempo, é usar o programa de processamento de dados do equipamento (opção *simulation*, dentro da rotina *oscill* do programa PROCESS, item 2.2.5). Após a obtenção dos filmes *still* e de se calcular a matriz de orientação do cristal (item 2.2.5.2), pode-se fazer uma previsão de coleta de dados conforme região escolhida para a coleta (p.ex. coleta de dados entre os ângulos de oscilação  $\phi=0-30^\circ$ ), ou seja, com a matriz orientação pode-se calcular qual a região para uma coleta de dados mais completa possível, variando-se o ângulo  $\phi$  de oscilação do goniômetro.

### 2.2.5 - Processamento Inicial dos dados para a obtenção dos $F_{obs}$

Durante uma coleta de dados obtém-se filmes (ou imagens) contendo pontos de difração (*spots*) que têm uma certa intensidade (proporcionais aos fatores de estrutura,  $F_{obs}$ ), os quais podem ser indexados pelos *índices de Miller*  $h, k, l$ . Para a indexação, deve-se conhecer os parâmetros de rede do cristal e este deve estar com uma orientação conhecida em relação a um sistema de referência (no caso da câmara de oscilação, o eixo  $z$  está no eixo de oscilação do cristal (ângulo  $\phi$ ) e o plano  $xy$  fica no plano de difração).

Em equipamentos como a câmara de precessão é necessário fazer esta orientação manualmente, fazendo vários filmes *still* (filmes sem precessão), o que tem a desvantagem que durante esta orientação o cristal vai se degradando devido aos danos da radiação ionizante de raios-X.

O sistema câmara de oscilação-placa de imagem eletrônica que foi utilizado realiza esta orientação automaticamente com ajuda do programa PROCESS (programa de processamento de dados do sistema R-AXIS IIC) que acompanha o equipamento. Este programa permite a indexação dos filmes sem conhecimento prévio dos parâmetros de rede, inclusive a partir de filmes de oscilação (porém a partir destes, a indexação torna-se bem mais difícil, de pior qualidade, e em vários casos não se consegue fazê-la).

O processo padrão, consiste em obter filmes *still* (normalmente dois, perpendiculares entre si), calcular a orientação em relação ao sistema de coordenadas e indexar os pontos de difração (item 2.2.5.1). É possível determinar usando este mesmo programa, os parâmetros de rede do cristal, se não foi determinado até então.

Em seguida, pode-se fazer a coleta de dados, propriamente dita, obtendo-se filmes de oscilação os quais devem ser escalados entre si após a coleta (item 2.2.5.4), colocados num único arquivo fazendo as devidas estatísticas para estes (item 2.2.5.6) e finalmente, obtendo os fatores de estrutura para cada reflexão (item 2.2.5.7).

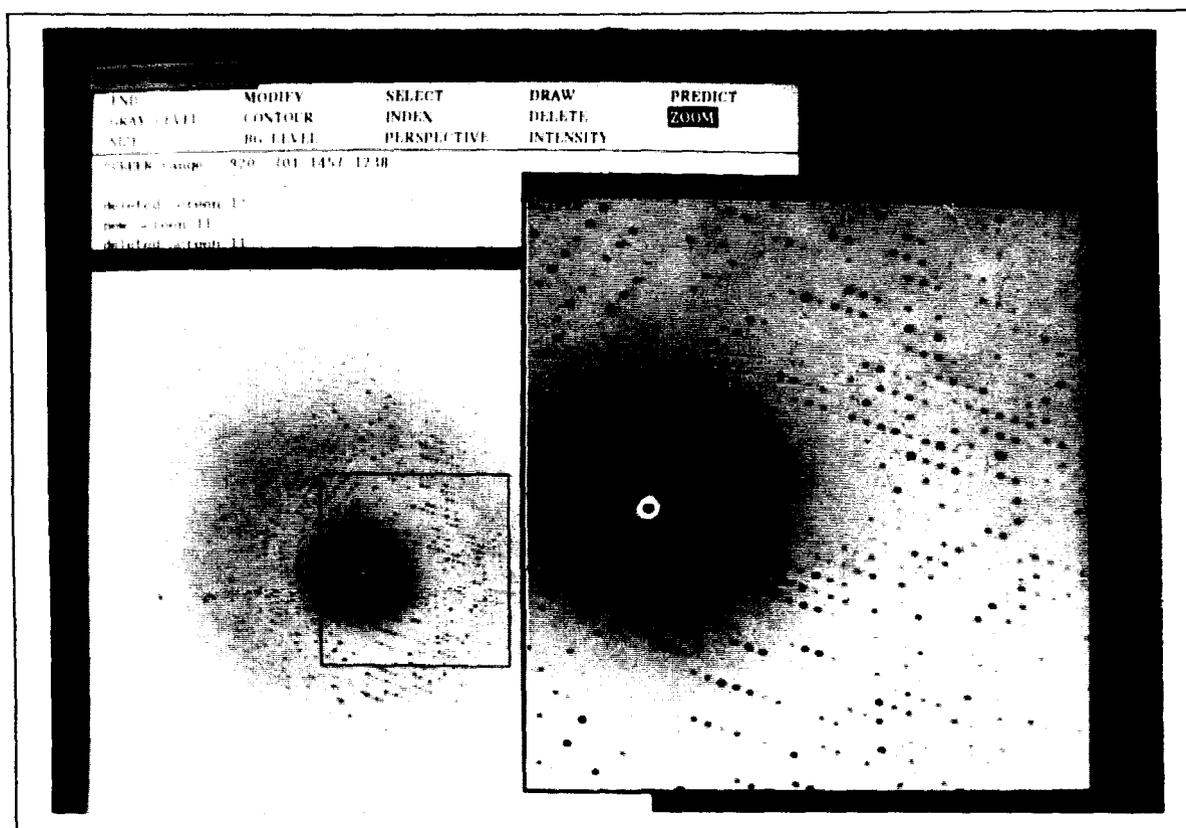


Figura 2.8 - Difratoograma de um experimento de difração de raios-X, mostrando uma imagem de oscilação, vista com a utilização do programa PROCESS (foto da tela da estação gráfica).

Na figura 2.8 é mostrada uma imagem de oscilação, vista com a utilização do programa PROCESS (foto da tela da estação gráfica).

As etapas no processamento dos dados (que são seguidas pelo programa) são dadas abaixo resumidamente:

*1 - Indexação dos Filmes e determinação dos parâmetros de rede (se for o caso) (Higashi, 1990):*

O método desenvolvido por Higashi permite a indexação de filmes oscilação ou *still* sem o conhecimento prévio dos parâmetros de rede do cristal; no caso destes já serem conhecidos o método fica bastante simplificado e será também descrito.

A auto-indexação é equivalente a achar, sem intervenção manual, três vetores de base de um conjunto de vetores espalhamento ( $S$ ) observados. Este método é bastante similar ao usado pelo difratômetro de 4 círculos (Busing e Levi, 1967 e manual do difratômetro CAD-4 da Enraf-Nonius), porém para um difratômetro a tarefa é simplificada, desde que neste equipamento está disponível praticamente todo o espaço recíproco durante a coleta de dados (há muitos graus de liberdade), porém no caso de um filme de oscilação tem-se um espaço muito restrito da oscilação (um filme de oscilação tem normalmente entre  $1^{\circ}$  a  $2^{\circ}$  de oscilação).

A primeira etapa é a localização dos pontos em relação ao centro do filme, atribuindo-as as coordenadas (Y,Z); estas juntamente com o valor de D (distância cristal-filme) podem ser expressas por meio de um vetor  $\mathbf{X} = (D, Y, Z)$ . Em seguida, o vetor  $\mathbf{X}$  é transformado para o espaço recíproco (Kim, 1989) o qual será chamado  $\mathbf{X}'$  e pode ser expresso em termos de índices de Miller, pela equação matricial (tomando o cristal no ângulo de oscilação  $\varphi=0$ ):

$$\mathbf{X}' = \mathbf{A} \mathbf{h} \quad (2.1)$$

onde  $\mathbf{A}$  é a matriz de orientação cristalina que representa os três vetores de base a serem obtidos por este algoritmo.

Calcula-se, a seguir, vetores diferença de  $\mathbf{X}'$  para uma resolução selecionada, e estes são ordenados segundo seus comprimentos (para os vetores equivalentes, em termos de comprimento, é feita uma média entre estes). No próximo passo os vetores são selecionados como vetores de base, usando os critérios:

$$\begin{aligned} & - \text{ângulo } (\mathbf{v}_1, \mathbf{v}_2) > C_1 \\ & - \text{ângulo } (\mathbf{v}_1 \times \mathbf{v}_2, \mathbf{v}_3) > C_2 \end{aligned} \quad (2.2)$$

Os valores para  $C_1$  e  $C_2$  são escolhidos pelo usuário, valores típicos são:  $C_1 = 30^\circ$  e  $C_2 = \theta_{\text{máx}}$  (da coleta de dados). A partir destes vetores tem-se a matriz  $\mathbf{A}$ , que na realidade é uma escolha arbitrária (pois, há outros valores possíveis de base que também são soluções) e na maioria das vezes é necessário uma redução para a cela apropriada (Gruber, 1973 e 1989).

A matriz  $\mathbf{A}$  obtida pode ser refinada por mínimos quadrados, baseado no fato que os índices de Miller devem ser números inteiros.

Tem-se, desta maneira, a partir da matriz  $\mathbf{A}$ , obtido os parâmetros de rede a partir dos quais podem ser calculados os índices de Miller de cada reflexão pela equação (2.1).

No caso dos parâmetros de rede já conhecidos, deve-se orientar o cristal pelos três ângulos a partir dos eixos ortogonais  $\phi_x, \phi_y, \phi_z$  (Arndt, U.W. e Wonacott, A.J., 1977 p.78). Da mesma maneira que no caso anterior, deve-se calcular a matriz  $\mathbf{A}$  (definida acima), que pode ser obtida pelas contribuições dos parâmetros de rede e das operações de rotação que são dadas pela equação:

$$\mathbf{A} = \mathbf{R} \mathbf{A}_0 \quad (2.3)$$

onde  $\mathbf{R}$  é uma matriz de rotação que pode ser dada em termos das rotações em função de cada ângulo, pela equação:

$$\mathbf{R} = \mathbf{R}_x(\phi_x) \mathbf{R}_y(\phi_y) \mathbf{R}_z(\phi_z) \quad (2.4)$$

$$\text{e } \mathbf{A}_0 \text{ é dado pela matriz: } \mathbf{A}_0 = \lambda \begin{pmatrix} a^* & b^* \cos \gamma^* & c^* \cos \beta^* \\ 0 & b^* \sin \gamma^* & -c^* \sin \beta^* \cos \alpha \\ 0 & 0 & 1/c \end{pmatrix} \quad (2.5)$$

A matriz  $\mathbf{A}_0$  pode ser calculada pelos parâmetros de rede já conhecidos, e para calcular a matriz  $\mathbf{A}$  o melhor procedimento é conhecido como *vector-matching procedure* (procedimento de comparação de vetores). O método consiste em indexar alguns vetores (normalmente de 3 a 5) da lista de vetores diferença por comparação com vetores que são gerados dos parâmetros de rede conhecidos (por meio da matriz  $\mathbf{A}_0$ ). Destes vetores e dos seus índices obtidos, a matriz  $\mathbf{A}$  pode ser calculada a partir de uma equação equivalente a (2.1) (onde esta é pesada, baseado no fato que os índices de Miller tem que ser números inteiros (Higashi, 1990)). Desta maneira, obtém-se estas duas matrizes e usando as equações (2.3) e (2.4) pode-se calcular a orientação do cristal (ângulos  $\phi_x, \phi_y, \phi_z$ ) (Kabsch, 1988).

O programa, após o processo de indexação e determinação dos parâmetros de rede do cristal, faz um refinamento destes parâmetros incluindo parâmetros de câmara (tais como, distância cristal-filme, coordenadas do centro do filme ( $X_{cen}, Y_{cen}$ ), ângulo ômega (deslocamento angular ( $^\circ$ ) do eixo de rotação do cristal em relação ao eixo vertical de leitura na placa de imagem), etc).

## 2 - Integração dos filmes de oscilação

Após a indexação dos pontos de difração, a próxima etapa será calcular a intensidade de cada um destes, ou seja, é necessário saber a quantos fótons equivale cada um dos pontos de difração. A maior dificuldade no processo de integração surge do fato de cada ponto de difração não ser, na realidade, um único ponto e sim uma esfera.

Este fato é devido a alguns fatores experimentais, tais como (Arndt, U.W. e Wonacott, A.J., 1977, pp.6-7):

1. Mosaicidade do cristal. (veja Stout e Jensen, 1968, p.67).
2. O efeito do uso de uma faixa de comprimento de onda,  $\Delta\lambda$ , e não um valor único como seria o ideal neste experimento.
3. O fato do feixe de raios-X não ser colimado perfeitamente ou divergência dos raios-X.
4. Tamanho finito da fonte e do cristal.

Se for pensado em termos da Esfera de Ewald, os fatores citados acima resultam em uma certa espessura na esfera, fazendo com que os pontos da rede recíproca tenham um volume associado. Deste modo, a reflexão em dado

momento começa a entrar em condição de difração, fica nesta condição (por um certo ângulo de rotação) e sai (figura 2.9).

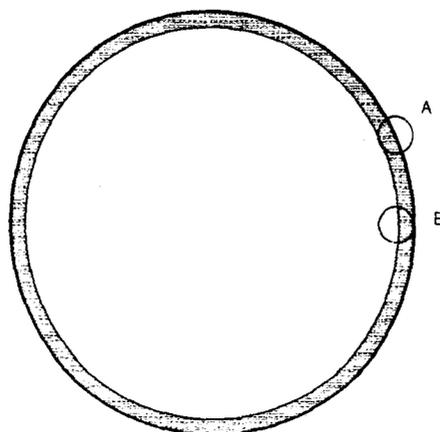


Figura 2.9 - Esfera de Ewald (estilizada com uma suposta espessura), mostrando uma reflexão entrando (A) e saindo (B) da condição de difração.

No caso de placas de imagem que são divididas em pequenas áreas (*pixels*), da ordem de  $\mu\text{m}$  (veja seção 2.2.1), o procedimento a ser feito é somar estes a fim de se obter o número total de fótons para cada reflexão. Deve-se porém, levar em consideração a radiação de fundo (*background*) presente em todo o detetor, além de ser necessário diferenciar uma reflexão da outra. (Rossmann, 1985, p.237, cap.20, Mét. Enz.).

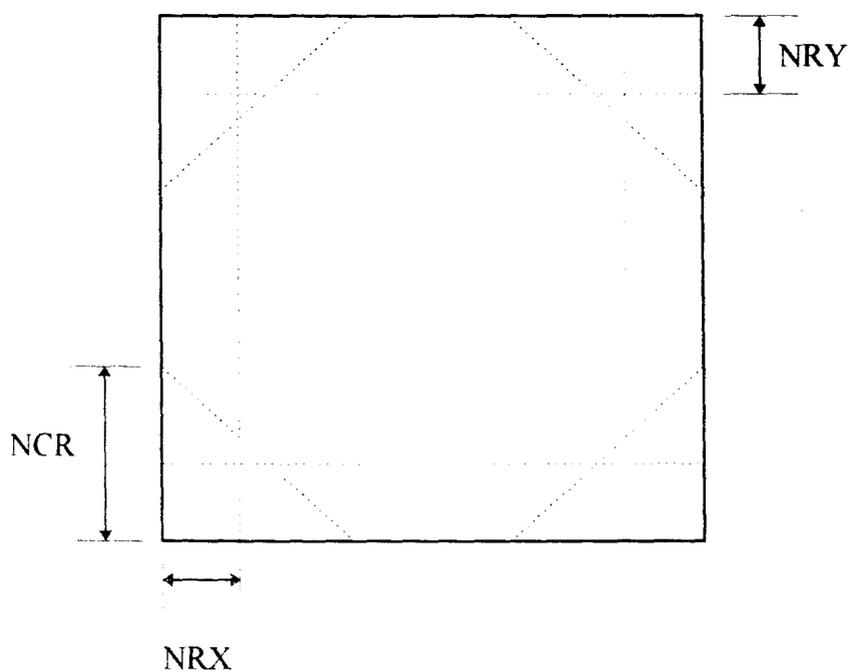


Figura 2.10 - Metodologia para integração de reflexões.

Estes problemas são tratados no programa de processamento de dados calculando-se um perfil médio para as reflexões (figura 2.11) usando-se caixas centradas na posição calculada pela matriz de orientação para cada reflexão (item 2.2.5.1). Desta caixa, é retirada uma certa área na altura (NRY), largura (NRX) e diagonais (NCR) (representadas por linhas pontilhadas na figura 2.10), formando assim duas áreas na caixa, uma interna às linhas pontilhadas e uma externa. A externa é considerada como devido somente a radiação de fundo e a interna como reflexão e radiação de fundo, assim a radiação devida unicamente à reflexão é obtida pela equação:

$$I_{hkl} = \text{Área Interna} - \text{Área externa}$$

A figura 2.11 representa uma caixa de perfil médio das reflexões do filme, onde se atribui números conforme varia a intensidade de cada *pixel*.

O programa usado permite ainda, o ajuste do tamanho da caixa (a caixa da figura 2.11 é de 7x9). Este ajuste é útil pois, para pontos de difração maiores, obtidos por cristais com maior volume ou que ficam mais tempo expostos à radiação, necessitam de caixas maiores e vice-versa.

0	0	0	0	0	0	0	0	0
0	0	1	4	8	4	1	0	2
0	1	7	76	147	82	12	1	0
0	2	25	189	255	175	32	3	1
1	2	16	80	123	65	14	3	1
0	1	4	10	13	8	3	1	0
0	0	1	1	2	1	1	0	0

Figura 2.11 - Perfil médio típico das reflexões de um filme.

Na realidade, o cálculo do perfil médio para todo o filme é feito somente na primeira etapa para a escolha da caixa pelo usuário; após esta escolha, o programa divide o filme em cinco regiões, uma central e quatro periféricas, é feito um refinamento de cada uma destas regiões independentemente adequando as coordenadas do centro do filme ( $X_{cen}$ ,  $Y_{cen}$ ),  $D$  (distância filme-detetor), ângulo  $\omega$  (veja item anterior) e parâmetros de distorção da câmara de acordo com a localização dos pontos de difração em termos dos índices de Miller. Com

este refinamento por regiões, melhora-se a qualidade dos resultados, pois pode-se levar em consideração as características de cada região.

O programa possui um critério de inclusão ou não de reflexões de acordo com a sua intensidade por área. Este valor deve ser colocado pelo usuário como parâmetro, havendo a possibilidade de colocar valores diferentes para as regiões central e periféricas do filme.

Surge porém, do fato dos filmes de oscilação serem feitos por ângulos limitados de coleta (geralmente são oscilações de 1° a 2°) e da Esfera de Ewald ter uma espessura (figura 2.9), as chamadas **reflexões parciais**. São reflexões que começam a entrar em condição de difração, mas saem desta condição antes que tenham atravessado totalmente a esfera de reflexão (Ewald) (devido ao ângulo de oscilação limitado) e são desta maneira, gravadas apenas parcialmente no filme. Porém, no filme seguinte a continuação desta reflexão será gravada, sendo então, possível utilizar estas reflexões (para tanto, o sistema mecânico do aparelho deve definir com grande precisão o ângulo que um filme termina e o outro começa; na prática este valor é da ordem de  $10^{-3}$  graus).

Para o uso das reflexões parciais pelo programa, é necessário primeiro que este saiba localizar quais são as reflexões parciais (veja item 7.5 e figuras 7.10 e 7.11 do Arndt, U.W. e Wonacott, A.J., 1977, pp.87-89), e que estas estejam na mesma escala (pois, de um filme para outro, há normalmente uma diminuição da intensidade e qualidade devido a degradação do cristal), cálculo que é feito no item 4 abaixo.

O programa faz neste passo ainda, a correção de Lorentz e Polarização (Wonacott, p.58 e 85).

### 3 - Merge

Este procedimento apenas junta as reflexões integradas no item anterior em um mesmo arquivo (arquivo *merge*) e coloca estas em ordem crescente em termos dos índices de Miller (*sorting*)

### 4 - Escalonamento (*Scaling*)

Após a integração de cada reflexão e estas estarem em um arquivo (*merge*), o próximo passo é colocar as reflexões de diferentes filmes em uma mesma escala a fim de que estas possam ser manipuladas. Deve-se lembrar que durante a coleta de dados o cristal vai se degradando; este fato reflete-se na intensidade das reflexões, de maneira que uma reflexão medida no primeiro filme, não terá a mesma intensidade se medida novamente no décimo quinto filme, por exemplo. Assim, deve-se aplicar um fator de correção para estas. É útil lembrar que há o problema das reflexões parciais, as quais podem se tornar

completas encontrando e colocando na mesma escala a outra parte desta, que estará em outro filme.

O procedimento dado abaixo (Hamilton, 1965; Fox, 1966; Blundell e Johnson, 1976 pp. 331 e 332) coloca na mesma escala, um conjunto de dados de diferentes filmes levando em conta reflexões comuns entre estes.

Seja  $I(\mathbf{h})_i$  uma intensidade observada de uma reflexão  $\mathbf{h}$  em um filme  $i$ . O que será feito neste método é obter fatores de escala  $K_i$ , para cada filme, para ser aplicado a cada  $I(\mathbf{h})_i$ , de tal maneira que se possa determinar o melhor valor de intensidade  $I(\mathbf{h})$  de uma reflexão  $\mathbf{h}$ , por mínimos quadrados.

Nas equações dos mínimos quadrados é mais conveniente ter os parâmetros não conhecidos ( $K_i$  e  $I(\mathbf{h})$  neste caso) separados dos parâmetros conhecidos ( $I(\mathbf{h})_i$ ) tal que a função de minimização não se altere se todo  $K_i$  é multiplicado por um fator de escala. Assim as equações observacionais podem ser escritas como:

$$\varphi(\mathbf{h})_i = w(\mathbf{h})_i^{1/2} (I(\mathbf{h})_i - I(\mathbf{h})/K_i) \quad (2.6)$$

onde  $w(\mathbf{h})_i^{1/2}$  é o peso para cada  $i$ -ésima observação e é proporcional ao recíproco do desvio padrão de  $I(\mathbf{h})$ . Com o objetivo de minimizar a equação acima, a fim de obter o melhor valor de  $I(\mathbf{h})$ , será feito:

$$\psi = \sum_{\mathbf{h}} \sum_{i=1}^N [\varphi(\mathbf{h})_i]^2 \quad (2.7)$$

onde  $N$  é o número de observações da reflexão  $\mathbf{h}$ . O melhor valor do ajuste por mínimos quadrados é achado no ponto de máxima da função  $\psi$ , isto é, quando  $\partial\psi/\partial I(\mathbf{h}) = 0$ . Assim substituindo (2.6) em (2.7), derivando em função de  $I(\mathbf{h})$ , fazendo igual a zero e rearranjando, obtém-se:

$$I(\mathbf{h}) = \frac{\sum_{i=1}^N w(h)_i \frac{I(h)_i}{K_i}}{\sum_{i=1}^N w(h)_i \left(\frac{1}{K_i}\right)^2} \quad (2.8)$$

Esta equação calcula qual deve ser a 'intensidade de uma reflexão  $I(\mathbf{h})$ ' em função das intensidades medidas nos diversos filmes ( $i$  filmes), usando fatores de escala ( $K_i$ ) que também serão calculados no processo de minimização. (Deve-se notar porém que a solução desta equação não linear (2.8) (observe que o fator  $K_i$  no denominador é quadrático) é mais facilmente resolvida em termos computacionais por uma expansão em série de Taylor).

O método descrito acima, pode ser usado também no caso de escalar dados de cristais diferentes em vez de filmes diferentes como foi mostrado.

Na rotina *Scaling* há também a possibilidade de refinar, em vez de um único fator de escala linear ( $K$ ), um outro dependente da resolução. Ou seja, uma função dada por  $A \exp(-b \operatorname{sen}\theta/\lambda)$ , onde  $A$  e  $b$  são fatores de escala a serem calculados por método semelhante ao descrito acima (mínimos quadrados).

### 5 - Pós-refinamento

A rotina de pós-refinamento permite o refinamento (utilizando método dos mínimos quadrados) dos parâmetros calculados no início da manipulação dos dados (item 1), tais como, parâmetros de orientação do cristal (ângulos  $\varphi_x$ ,  $\varphi_y$ ,  $\varphi_z$ ), parâmetros de rede e de câmara, além dos parâmetros mosaicidade que não foram refinados ainda, incluindo agora as **reflexões parciais**.

Nesta rotina é calculado o grau de parcialidade,  $p_{\text{calc}}$  a partir do conjunto de parâmetros cristalinos, orientação do cristal e mosaicidade deste e mede o grau de parcialidade observada,  $p_{\text{obs}}$ . Estes dois valores,  $p_{\text{calc}}$  e  $p_{\text{obs}}$ , são minimizados por uma função:

$$E = \sum w(p_{\text{calc}} - p_{\text{obs}})^2 \quad (2.9)$$

onde o somatório é sobre todas as reflexões parciais para as quais uma ou mais reflexões completas (*full reflections*) são também observadas e  $w$  é uma função peso. As expressões utilizadas no cálculo de  $p_{\text{calc}}$  e  $p_{\text{obs}}$  são dadas abaixo (Rossmann *et al*, 1979):

$$p_{\text{obs}} = \frac{F_{hi}^2}{G_i F_h^2} \quad p_{\text{calc}} = 3q^2 - 2q^3 \quad (2.10)$$

onde  $F_{hi}^2$  é uma reflexão parcial  $hkl$ ,  $F_h$  é a melhor estimativa de uma reflexão completa  $hkl$  obtida do mesmo filme, de outros filmes ou de outros cristais,  $G_i$  ( $=1/K_i$ ) é o inverso do fator de escala para um filme  $i$  e  $q$  é o grau de penetração (fração do volume de uma reflexão completa). O valor de  $q$  pode ser calculado utilizando os parâmetros de rede cristalino, orientação do cristal e mosaicidade deste (Rossmann *et al*, 1979); e pela equação 2.10 observa-se que a única dependência de  $p_{\text{calc}}$  é  $q$ , assim quando é feita a minimização determinada pela equação 2.9, faz-se na realidade um refinamento dos parâmetros de cela, orientação e mosaicidade com respeito às reflexões parciais.

O programa permite o refinamento destes parâmetros separadamente ou em conjunto, e tenta, chegar a um mínimo; isto nem sempre é possível, principalmente com dados de pior qualidade, onde a inclusão de reflexões parciais podem gerar erros maiores.

## 6 - Promediação (Averaging)

Faz a promediação ('média') entre reflexões equivalentes (para reflexões completas, parciais e ambas), calcula  $R_{\text{sym}}$  ( $R_{\text{sym}} = \frac{\sum_{hkl} \sum_i |F_i(hkl)| - \langle |F(hkl)| \rangle}{\sum_{hkl} \sum_i |F_i(hkl)|}$  onde

$F_i(\mathbf{h})^2$  é a intensidade do  $i$ -ésimo filme,  $\langle F_i(\mathbf{h}) \rangle^2$  é a intensidade média do  $i$ -ésimo filme), calcula a intensidade média das reflexões por filme e por faixa de resolução e outras estatísticas, além de fazer (se optado pelo usuário) cálculos do efeito da dispersão anômala.

O cálculo de  $R_{\text{sym}}$  (ou  $R_{\text{merge}}$ ), se for feito sobre um número grande de reflexões, é um indicativo bastante útil da qualidade dos dados coletados; pois se este número tiver um valor grande pode indicar que os dados possuem muitos erros devido a fatores como absorção dos raios-X pelo cristal, espalhamento anômalo e determinação incorreta de parâmetros cristalinos (um valor típico para  $R_{\text{sym}}$  é de 5 %).

Esta rotina do programa PROCESS possui alguns critérios de rejeição de reflexões os quais devem ser atribuídos pelo usuário. Um dos critérios leva em consideração a relação: intensidade integrada da reflexão/desvio padrão. Outro critério leva em consideração a diferença entre duas reflexões equivalentes, que é comparado com a intensidade média das reflexões e a intensidade média das reflexões equivalentes.

## 7 - Saída dos dados (Output)

Gera o arquivo de saída com fatores de estrutura para cada reflexão  $h, k, l$  e calcula completeza dos dados com a resolução.

### 2.2.6 - Coleta de dados para obtenção de derivados isomorfos

A coleta de dados para obtenção de derivados isomorfos deve levar em consideração que os cristais, quando colocados para *soaking* na presença de sais de átomos pesados (item 2.1), sofrem danos devido a presença destes na solução (as ligações de rede do cristal são afetadas com a presença de átomos pesados ligados às moléculas do cristal), assim normalmente, os dados coletados são de pior qualidade do que para uma coleta de cristais nativos.

Na tabela 2.2 são mostrados detalhes das coletas de dados realizadas (da melhor coleta para cada derivado) para os derivados onde foi possível fazê-lo (derivados que possuem a última coluna na tabela 2.1). Os cristais pertencentes ao grupo espacial R32, estão representados no sistema de coordenadas hexagonal (International Table for X-Ray Crystallography vol. I, 1987).

Tabela 2.2 - Resumo das coletas de dados para obtenção de derivados isomorfos

	K <sub>2</sub> PtCl <sub>4</sub>	K <sub>2</sub> PtBr <sub>6</sub>	Mers	CMNF	EHgS	Au	PbAc	PbO	Usulf	EuSf	Smac	GdSf
a, b (Å)	125.94	125.89	125.95	125.96	125.76	126.31	125.74	125.71	125.77	126.07	125.99	125.98
c (Å)	220.76	221.86	222.72	224.50	223.38	222.11	223.16	223.25	222.95	220.79	222.90	222.79
dist. filme-cristal	130.31	130.35	130.45	130.12	130.49	130.70	150.17	150.21	150.39	130.32	130.18	130.21
R <sub>sym</sub> (%)	8.2	6.5	9.0	8.3	8.1	7.4	5.3	4.7	4.8	4.8	7.6	7.1
Refl. independ.	22611	6559	10327	7155	13150	13064	11384	3121	10072	20914	14300	18050
Refl. medidas	52005	9391	23135	11850	25100	23121	16957	3788	17376	36558	23979	29342
n <sup>o</sup> fotos	15	10	15	15	20	3	17	3	10	15	15	14
n <sup>o</sup> coletas	2	5	5	3	1	2	1	1	2	2	2	2
Completeza (%)	95.5	50.1	78.4	51	55.4	57.7	86.4	23.7	76.6	88.1	62.9	82.5
Resolução (Å)	2.5	3.0	3.0	3.0	2.5	2.5	3.0	3.0	3.0	2.5	2.5	2.5
R <sub>deriv</sub> (%)	26.8	16	16.3	11.3	10.4	10.1	7.7	7.1	7.4	9.7	10.5	14.5

Abreviaturas: **dist. filme-cristal**: distância cristal-filme (após pós-refinamento), **R<sub>sym</sub>** índice de consistência em relação as reflexões do mesmo conjunto de dados, **Refl. independ**: Reflexões independentes, **Refl. medidas**: reflexões medidas, **n<sup>o</sup> coletas**: coletas realizadas com diferentes cristais para o mesmo derivado de átomo pesado, **Completeza**: porcentagem dos dados medidos dos possíveis teoricamente, **R<sub>deriv</sub>** índice de consistência do derivado em relação ao nativo. Os resultados apresentados se referem ao melhor conjunto de dados medidos. **K<sub>2</sub>PtCl<sub>4</sub>**: tetracloro platinato (IV) de potássio, **K<sub>2</sub>PtBr<sub>6</sub>**: hexabromo de platina (IV) de potássio, **Mers**: ácido mersálico, **CMNF**: 2-cloro mercúrio nitrofenol, **EHgS**: ácido tiosalicílico etilmercúrio, **Au**: tetracloro-aurato (III) de potássio, **PbAc**: acetato de chumbo, **PbO**: óxido de chumbo, **Usulf**: sulfato de uranila, **EuSf**: sulfato de európio (III), **Smac**: acetato de samário (III), **GdSf**: sulfato de gadolínio (III).

A comparação do  $R_{sym}$  (veja item 2.2.5.6) com o  $R_{deriv}$  ( $R_{deriv} = \frac{\sum ||F_{PH}(\mathbf{h})| - |F_P(\mathbf{h})||}{\sum |F_P(\mathbf{h})|}$ ) do derivado, pode indicar a presença de um derivado promissor se a diferença entre estes for grande; porém a decisão final é obtida após o mapa de Patterson diferença (item 3.1), onde devem haver picos significativos. Da tabela 2.2 pode-se notar que os derivados das quatro primeiras e da última colunas têm  $R_{deriv}$  com valores bem superiores aos valores dos outros derivados.

Desta maneira, para a resolução da estrutura da GlcN6P desaminase, utilizou-se o derivado K<sub>2</sub>PtCl<sub>4</sub> com dados à 2.5 Å e também derivados de Mercúrio, o ácido mersálico (C<sub>13</sub>H<sub>17</sub>HgNO<sub>6</sub>) e 2-Cloromercúrio-4-nitrofenol (C<sub>6</sub>H<sub>4</sub>ClHgNO<sub>3</sub>) com dados à 3.0 Å de resolução (que tem sítios de ligação de átomos pesados equivalentes, sendo considerados um só derivado, veja capítulo 3).

Com a finalidade de conseguir-se um terceiro derivado e desta maneira, obter fases de melhor qualidade (veja capítulo 3), continuou-se a busca por

outros derivados, mesmo após a obtenção destes. Não sendo possível porém, obter um terceiro derivado, resolveu-se a estrutura com estes dois e por conseguinte, as fases foram melhoradas com métodos de modificação de densidade e refinamento (capítulo 4).

## CAPÍTULO 3

### CÁLCULO INICIAL DAS FASES

#### 3.1 - O MÉTODO DA SUBSTITUIÇÃO ISOMÓRFICA MÚLTIPLA (MIR) PARA O CÁLCULO DE FASES

##### 3.1.1 - Determinação das posições dos átomos pesados

Após a obtenção dos derivados isomorfos, a próxima etapa no método da substituição isomórfica é a determinação das posições dos átomos pesados, que nesta tese foi feita pelo método de Patterson (Patterson, 1934). Este método será discutido brevemente a seguir (Blundell e Johnson, 1976 e Patterson, 1934).

A função de Patterson é definida como:

$$P(\mathbf{uvw}) = \int_V \rho(\mathbf{xyz})\rho(\mathbf{x} + \mathbf{u}, \mathbf{y} + \mathbf{v}, \mathbf{z} + \mathbf{w})d\mathbf{x} \quad (3.1)$$

Esta integral é conhecida da teoria de série de Fourier como auto-correlação de  $\rho(\mathbf{x})$ . Assim, o valor desta função em um particular ponto  $\mathbf{u}(\mathbf{uvw})$  é calculado pelo produto dos vetores de densidade eletrônica nas posições  $\mathbf{x}$  e  $\mathbf{x}+\mathbf{u}$  (ou seja, separados por um vetor  $\mathbf{u}$ ) somado sobre toda a cela unitária.

A distribuição de densidade pode ser expressa na seguinte forma:

$$\rho(\bar{\mathbf{x}}) = \frac{1}{V} \sum_{\mathbf{h}=-\infty}^{+\infty} \bar{\mathbf{F}}(\bar{\mathbf{h}}) e^{-2\pi i \bar{\mathbf{h}} \cdot \bar{\mathbf{x}}} \quad (3.2)$$

Substituindo (3.2), para as posições  $\mathbf{x}$  e  $\mathbf{x}+\mathbf{u}$ , em (3.1), obtém-se:

$$P(\bar{\mathbf{u}}) = \frac{1}{V} \sum_{\bar{\mathbf{h}}} |\bar{\mathbf{F}}(\bar{\mathbf{h}})|^2 e^{-2\pi i \bar{\mathbf{h}} \cdot \bar{\mathbf{u}}} \quad (3.3)$$

É útil notar que a função de Patterson é a transformada de Fourier de  $|\mathbf{F}(\mathbf{h})|^2$ , em vez de  $\mathbf{F}(\mathbf{h})$ , que pode sempre ser calculado das intensidade de difração.

Pode-se ver que as principais contribuições para  $P(\mathbf{u})$  serão quando  $\rho(\mathbf{x})$  e  $\rho(\mathbf{x}+\mathbf{u})$  têm grandes valores, então um pico na curva  $P(\mathbf{u})$  para um dado valor de  $\mathbf{u}=\mathbf{u}_1$ , significa que pode-se ter dois átomos separados pela distância  $\mathbf{u}_1$ . Deste modo, a função de Patterson representa um mapa de vetores interatômicos. É útil notar ainda, que sempre haverá um pico na origem, isto é, para  $\mathbf{u}=0$ .

Em geral, se um átomo  $i$  contém  $Z_i$  elétrons e um átomo  $j$  contém  $Z_j$  elétrons, então o pico no mapa de Patterson que representa o vetor entre estes átomos terá um peso proporcional a  $Z_i Z_j$ . Assim, os átomos mais pesados resultarão em picos mais relevantes no mapa.

A função de Patterson contém um centro de simetria na origem, embora a estrutura em si possa não ser centro-simétrica. A natureza centro-simétrica da síntese de Patterson implica que o grupo espacial é diferente daquele da estrutura real; este pode ser derivado da estrutura real pela adição de um centro de simetria e perda de elementos de simetria translacionais (ver International Table for X-ray Crystallography, Vol. I, 1987) formando o chamado grupo de simetria de Laue.

A função a ser utilizada para o cálculo do mapa de Patterson para determinação de átomos pesados no método MIR, dada pela equação 3.4, deve usar coeficientes  $(|F_{PH}| - |F_P|)^2$ . Esta função é conhecida como Patterson diferença.

O mapa de Patterson possui picos de densidade eletrônica relativos a vetores interatômicos, desta maneira o objetivo é, a partir destes picos, calcular as coordenadas atômicas.

Para tanto, o primeiro passo é determinar as *linhas ou seções de Harker*. Nestas posições especiais do mapa de Patterson há uma grande densidade de picos relacionados por simetria correspondentes ao Grupo Espacial da estrutura a ser resolvida. As seções ou linhas de Harker de um dado grupo espacial, podem ser obtidas fazendo subtrações entre operações de simetria deste grupo espacial.

Nestas linhas e seções de Harker de um mapa de Patterson, os picos com coordenadas  $u, v, w$  corresponderão a uma diferença entre coordenadas de átomos relacionados por simetria. Estes picos correspondem a auto-vetores (também chamados de picos de Harker), pois são vetores diferença entre mesmos átomos (mesmo sítio) relacionados por simetria. No mapa de Patterson há outros picos que podem estar em qualquer posição neste mapa, estes correspondem a vetores cruzados, que são vetores entre átomos diferentes (estão em sítios diferentes e não estão relacionados por operações de simetria).

### 3.1.2 - Cálculo das fases

Conhecidas as posições dos átomos pesados pode-se calcular o fator de estrutura dos átomos pesados  $F_H$  pela equação  $F_H(hkl) = \sum f_j \exp 2\pi i (hx_j + ky_j + lz_j)$  (onde  $f_j$  é o fator de espalhamento atômico do átomo  $j$  e  $x_j, y_j, z_j$  são as coordenadas do átomo  $j$ ) (Blundell e Johnson, 1976, p.127).

Seja  $F_{PH}$  e  $F_P$  os fatores de estrutura da proteína + átomo pesado e da proteína, respectivamente (os módulos dos fatores de estrutura proteína+átomo pesado e da proteína nativa,  $|F_{PH}|$  e  $|F_P|$  são mensuráveis), pode-se escrever a seguinte equação:

$$F_{PH} = F_P + F_H \quad (3.4)$$

Aplicando a lei dos cosenos na figura 3.1, pode-se escrever que a fase da proteína é (Blundell e Johnson, 1976, p.156):

$$\alpha_P = \alpha_H + \cos^{-1} \left( \frac{F_{PH}^2 - F_P^2 - F_H^2}{2F_P F_H} \right) = \alpha_H \pm \alpha' \quad (3.5)$$

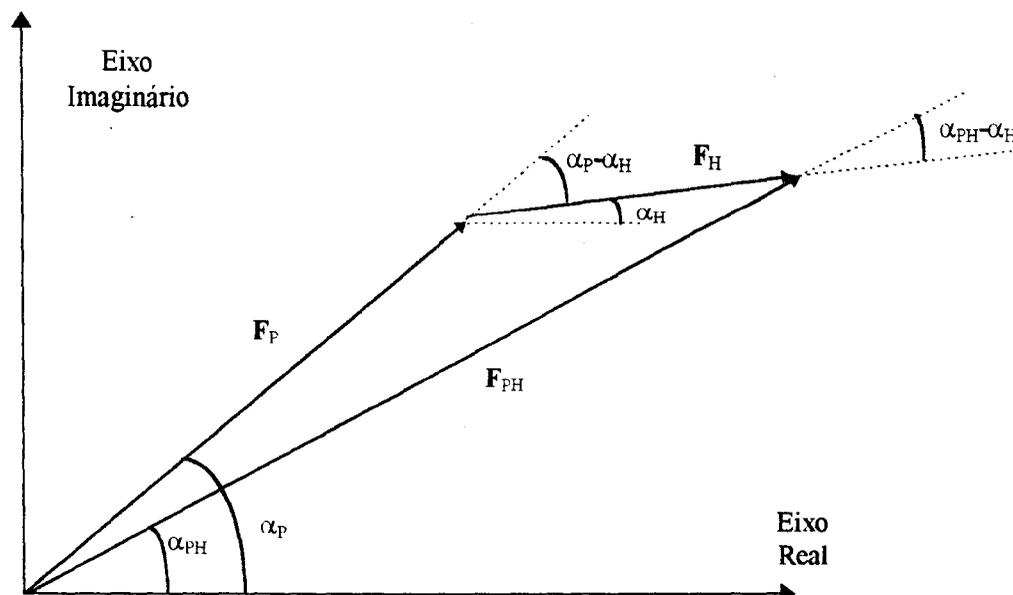


Figura 3.1 - Diagrama mostrando uma representação dos vetores fatores de estrutura,  $F_P$  (proteína nativa),  $F_{PH}$  (proteína+átomo pesado) e  $F_H$  (átomo pesado) e suas respectivas fases  $\alpha_P$ ,  $\alpha_{PH}$ ,  $\alpha_H$  (Blundell e Johnson, 1976, p.156).

A equação 3.5 mostra que há uma ambigüidade na solução da fase que poderá ser resolvida utilizando mais um derivado isomorfo. Esta ambigüidade pode ser ilustrada graficamente pela construção de Harker (Harker, 1956), mostrada na figura 3.2. Nesta construção, faz-se um círculo de raio  $|F_P|$  e do centro deste por sua vez, desenha-se um vetor  $-F_H$ . Do final deste, desenha-se outro círculo de raio  $|F_{PH}|$ . As interseções dos dois círculos correspondem aos dois ângulos de fase para a proteína ( $\alpha_P$ ) que são igualmente prováveis, pois o triângulo vetorial  $F_{PH} = F_P + F_H$  (equação 3.4) deve, teoricamente, fechar exatamente.

Com um segundo derivado isomorfo pode-se, em princípio, distinguir entre as duas possibilidades. Este fato pode ser representado graficamente a partir da figura 3.2, desenhando-se um vetor  $-F_H(2)$  (que ilustra o segundo derivado isomorfo), do ponto O; e a partir da extremidade do vetor  $F_H(2)$ , desenha-se um círculo de raio  $|F_{PH}(2)|$  (figura 3.3). Nota-se que os três círculos se interceptarão em um único ponto, o qual será a solução única para  $F_P$ , isto é, estará determinada a fase  $\alpha_P$ . Da figura 3.3, pode ser visto ainda, que quando  $F_{H1}$  e  $F_{H2}$  são colineares, o segundo derivado não ajudará em nada para a resolução da

ambigüidade do cálculo. O caso de dois derivados colineares ocorreu para os derivados de ácido mersálico e 2-cloro mercúrio nitrofenol no caso desta tese, nos quais haviam átomos pesados de mercúrio nos mesmos sítios de ligação.

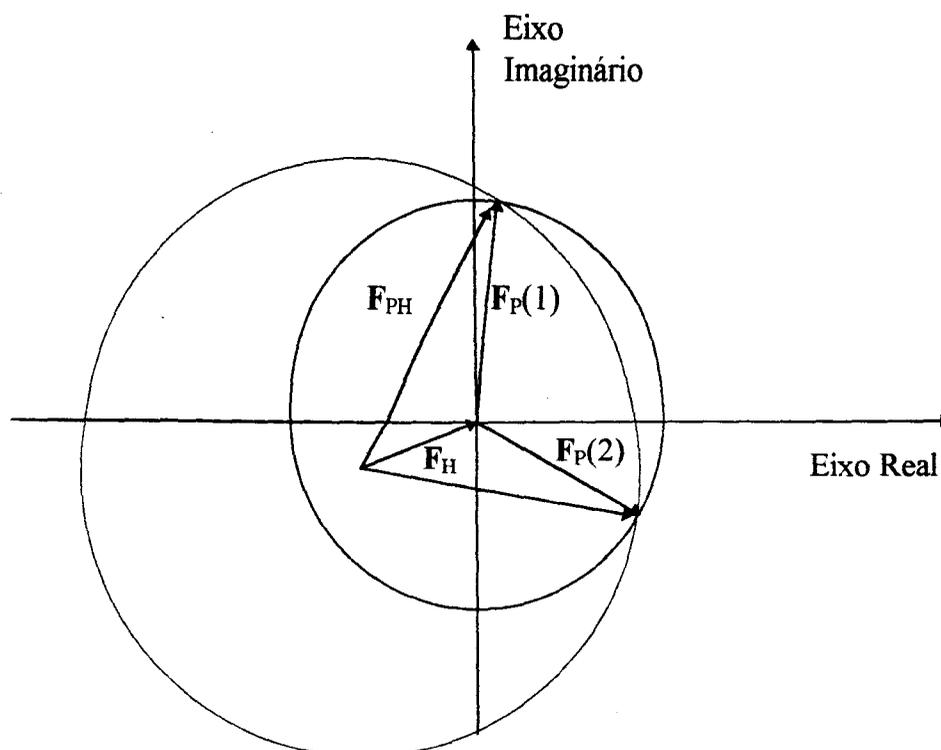


Figura 3.2 - Construção de Harker para o cálculo de fases. No método da substituição isomórfica cada derivado de átomo pesado contribui com duas possibilidades para o ângulo de fase da proteína  $\alpha_P$ , correspondendo a dois vetores  $F_P(1)$  e  $F_P(2)$  (Drenth, 1994, p.171)

Porém, normalmente, a interseção dos círculos da figura 3.3 não será em um único ponto, devido a erros na coleta de dados de difração de raios-X e a perda de isomorfismo dos cristais com átomos pesados. Assim, permanecerão incertezas com respeito ao real valor das fases da proteína ( $\alpha_P$ ). Uma maneira para minimizar estes erros é utilizar mais do que dois derivados isomorfos.

Desta maneira, o método que usa vários derivados isomorfos é chamado de método da substituição isomórfica múltipla.

Os erros no cálculo das fases da proteína ( $\alpha_P$ ), podem ser calculados pelo método dado a seguir (Blow e Crick, 1959). Blow e Crick discutem em seu trabalho que estes erros tem duas causas principais e distintas. Uma delas reside em  $F_H^{calc}$ , devido ao erro no cálculo das posições do átomos pesados pela relação aproximada para reflexões não centrosimétricas:  $|F_{PH}| = |F_P| + |F_H|$ , perda de isomorfismos dos cristais derivados, ocupação e parâmetros térmicos atribuídos aos átomos pesados. A outra causa de erros está nas imprecisões experimentais na determinação das amplitudes de  $|F_P|$ ,  $|F_{PH}|$ .

Blow e Crick mostraram também, que estes erros podem ser representados por duas funções probabilidade com erros devidos a  $F_H$  e  $F_{PH}$ . Estas duas funções

probabilidade podem ser combinadas em uma só e mostra-se então, que esta função probabilidade tem maior peso devido a variação do módulo de  $|F_{PH}|^{calc}$  (veja Blundell e Johnson, 1976, p.365).

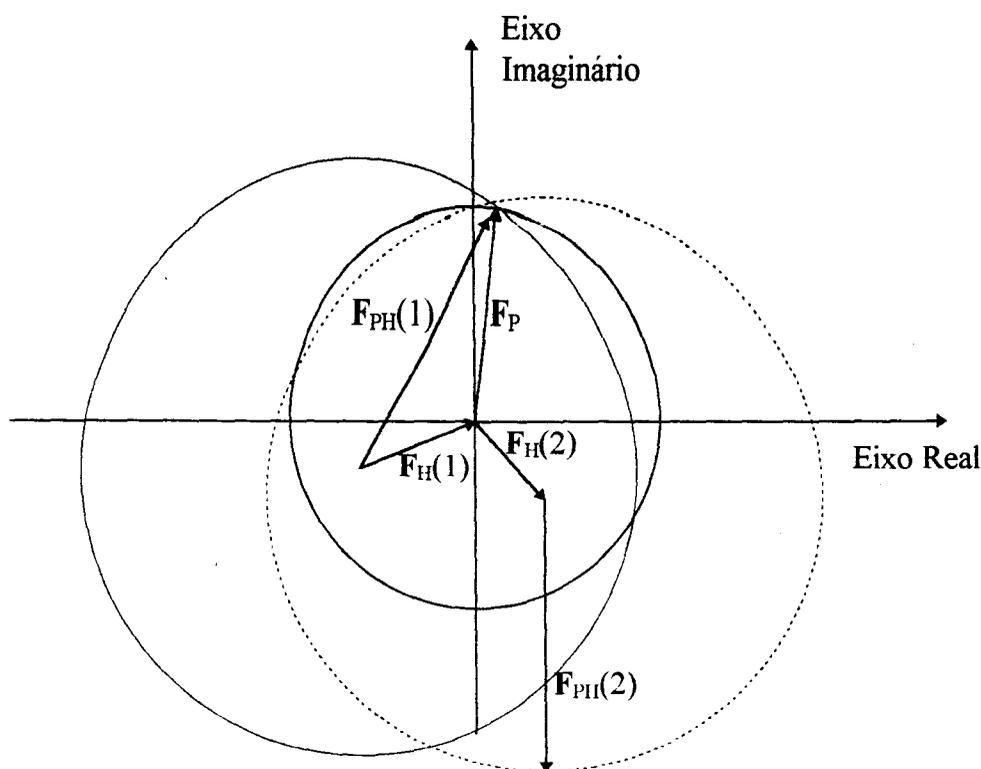


Figura 3.3 - A construção de Harker para determinação da fase pelo método da substituição isomórfica múltipla. Neste método são utilizados dois derivados  $F_H(1)$  e  $F_H(2)$  obtendo somente um ponto de interseção que determina o ângulo de fase da proteína  $\alpha_P$ . (Drenth, 1994, p.171; Blundell e Johnson, 1976, p.161)

O triângulo vetorial (figura 3.2)  $F_{PH} = F_P + F_H$ , somente se fechará na interseção dos dois círculos de raios  $|F_P|$  e  $|F_{PH}|$ ; estas interseções corresponderão aos valores de  $\alpha_P$ . Para os outros valores de  $\alpha_P$  o erro para fechar o triângulo será (veja figura 3.1):

$$\varepsilon(\alpha_P) = |F_{PH}|^{calc} - |F_{PH}|^{obs} \quad (3.6)$$

Para todos os valores de  $\alpha_P$ , o comprimento AC (ou  $|F_{PH}|^{calc}$ ) pode ser calculado pela regra do cosseno (fig. 3.4), pois  $|F_P|$  e  $F_H$  são conhecidos, desta maneira  $\varepsilon(\alpha_P)$  (chamado de *lack of closure*) é a diferença entre o valor calculado AC (ou  $|F_{PH}|^{calc}$ ) e o valor observado  $|F_{PH}|^{obs}$ . Se  $\varepsilon(\alpha_P)$  é pequeno, há maior probabilidade das fases da proteína estarem corretas, e vice-versa.

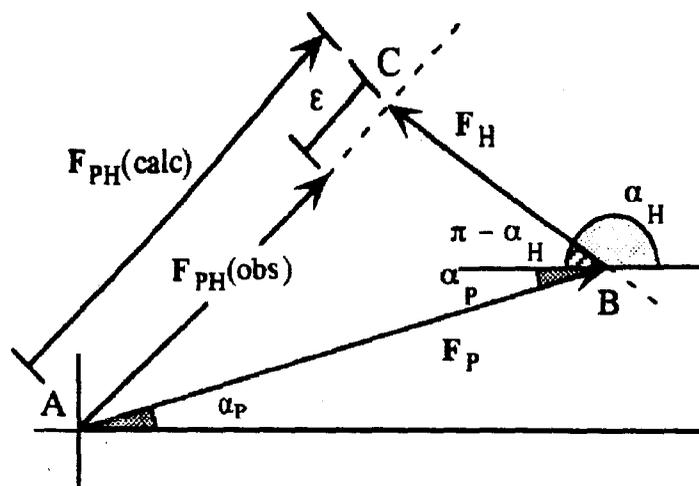


Figura 3.4 - Situação real, onde o triângulo vetorial  $F_{PH} = F_P + F_H$  não fecha exatamente. Os valores observados e calculados de  $|F_{PH}|$  diferem do erro  $\epsilon$  (Drenth, 1993, p.170).

Tomando a distribuição de probabilidade Gaussiana (Blow e Crick, 1959) para  $\epsilon(\alpha_P)$  para cada reflexão de cada derivado, tem-se:

$$P(\alpha_P) = P(\epsilon(\alpha_P)) = N \exp\left[-\frac{\epsilon^2(\alpha_P)}{2E^2}\right] \quad (3.7)$$

onde N é um fator de normalização, relacionado com o fato que  $\alpha_P$  deve ter valor entre 0 e  $2\pi$  (ou seja:  $\int_{\alpha_P=0}^{2\pi} P(\alpha_P) d\alpha_P = 1$ ),  $E^2$  é a média quadrática de  $\epsilon(\alpha_P)$ .

Se E é pequeno, a curva de probabilidade terá picos bem pronunciados e as fases da proteína serão bem determinadas, senão (E tem valor grande) as fases não serão bem determinadas. Para cada reflexão,  $P(\alpha_P)$  poderá ser calculada como função de  $\alpha_P$ . A curva obtida será simétrica (pois tem-se na equação 3.7 potência de 2) como dado na figura 3.5, em um ponto D.

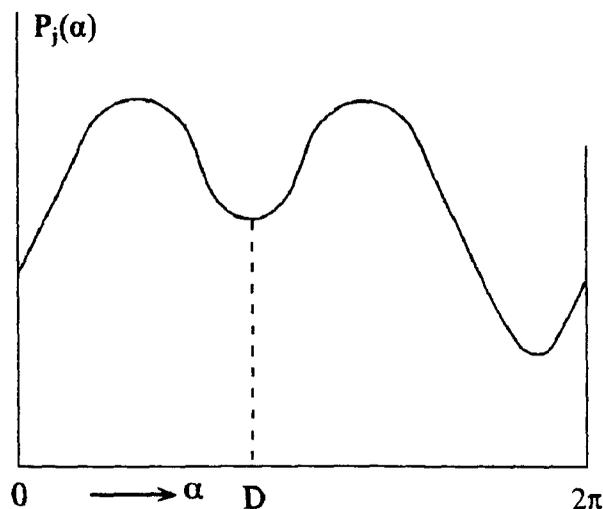


Figura 3.5 - A probabilidade  $P_j(\alpha)$  para uma reflexão ter  $\alpha$  como ângulo de fase correto obtida do derivado  $j$  é mostrada como função de todos os ângulos entre 0 e  $2\pi$  (Drenth, 1994, p.172).

Os dois picos correspondem aos dois valores possíveis de  $\alpha_p$  da figura 3.2. Esta figura pode ser calculada para cada uma das reflexões de cada um dos  $n$  derivados. A probabilidade total (todos os derivados) para cada uma das reflexões pode ser obtida multiplicando  $n$  vezes a equação 3.7:

$$P(\alpha_p) = \prod_{j=1}^n P_j(\alpha_p) = N' \exp \left[ - \sum_j \frac{\varepsilon_j^2(\alpha_p)}{2E_j^2} \right] \quad (3.8)$$

A probabilidade total  $P(\alpha_p)$  normalmente não será simétrica como em 3.5, e sim algo mostrado na figura 3.6.

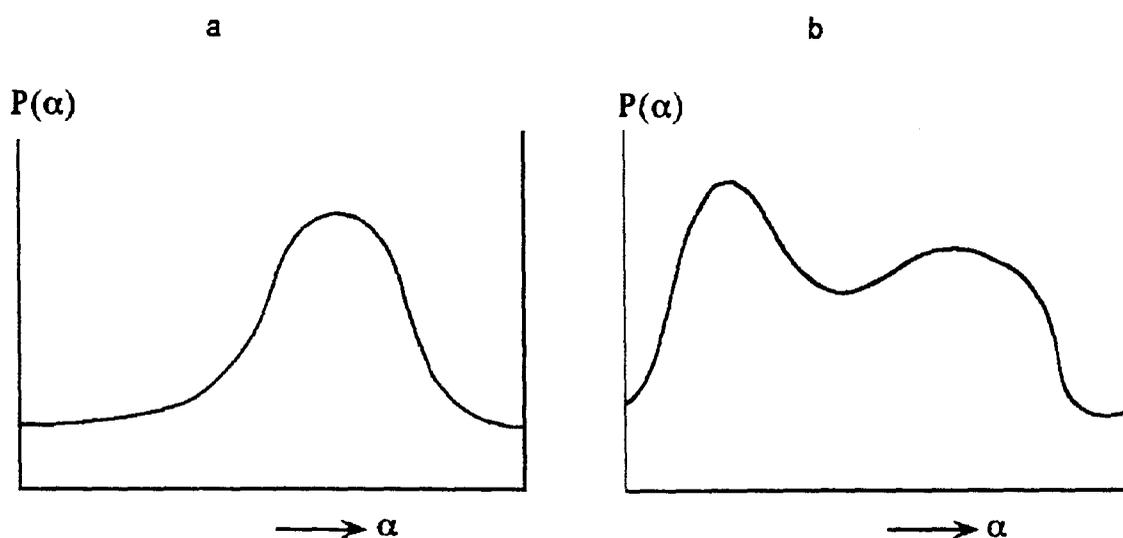


Figura 3.6 - Exemplos da probabilidade total  $P(\alpha)$  para o ângulo de fase  $\alpha$  de uma reflexão obtida de mais de um derivado (Drenth, 1994, p.173).

Assim, se for calculado um mapa densidade eletrônica usando a fase  $\alpha_{hkl}$ , a probabilidade para que este mapa esteja correto será dada pela equação 3.8, isto é,  $\prod_{hkl} P_{hkl}(\alpha_{hkl})$ . Porém, o objetivo é obter o melhor mapa de Fourier possível de ser calculado, que não é necessariamente o mais provável. O melhor mapa é definido como o mínimo erro da média quadrática na densidade devido a erros nos ângulos de fases. A razão para isto é que a função  $P_{hkl}(\alpha_p) = \prod_{j=1}^n P_j(\alpha_{hkl})$  não tem sempre um valor mínimo único, mas as vezes dois, ou é assimétrico ao redor do seu máximo. Calculando o mapa densidade eletrônica mais provável, isto é completamente desprezado, assim a melhor solução para o problema é a seguinte.

Seja o fator de estrutura  $F_{hkl} = |F_{hkl}| \exp[i\alpha_{hkl}]$  onde a amplitude é conhecida e a distribuição de probabilidade  $P_{hkl}(\alpha)$  também pode ser calculada. A melhor estimativa  $F_{hkl}(best)$  para o fator de estrutura real, que pode ser calculada

por meio do método dos mínimos quadrados com base em dados experimentais, é dada por (Drenth, 1994, p.167):

$$Q = \int_{\alpha} \{P_{hkl}(\alpha) |\bar{F}_{hkl}| \exp[i\alpha] - \bar{F}_{hkl}(best)\}^2 d\alpha \quad (3.9)$$

que deve ser mínima, assim:

$$\frac{dQ}{d\{\bar{F}_{hkl}(best)\}} = 0 \quad (3.10)$$

Resolvendo-se a equação 3.10, obtém-se:

$$\bar{F}_{hkl}(best) = \int_{\alpha} \{P_{hkl}(\alpha) |\bar{F}_{hkl}| \exp[i\alpha]\} d\alpha = |\bar{F}_{hkl}| \bar{m} \quad (3.11)$$

onde:

$$\bar{m} = \int_{\alpha} \{P_{hkl}(\alpha) \exp[i\alpha]\} d\alpha ; \quad 0 \leq |m| \leq 1$$

Ou seja, o melhor valor de  $F_{hkl}$  é obtido tomando a média ponderada sobre os possíveis  $F_{hkl}$ , e o ponto  $F_{hkl}(best)$  será o centro de gravidade da função distribuição de probabilidade.

Na prática, substitui-se a integração pela somatória, e a equação 3.11 torna-se:

$$\bar{F}_{best} = \frac{\sum P(\alpha) \bar{F}(\alpha)}{\sum P(\alpha)} \quad (3.12)$$

Fazendo  $m = m \exp[i\alpha_{best}]$  a equação 3.11 torna-se:

$$\bar{F}_{hkl}(best) = |\bar{F}_{hkl}| m \exp[i\alpha_{best}] \quad (3.13)$$

e

$$m = \frac{|\bar{F}_{hkl}(best)|}{|\bar{F}_{hkl}|} \quad (3.14)$$

$m$  é chamado de 'figura de mérito'.

Se os valores de  $F_{hkl}(best)$  são usados no cálculo da distribuição densidade eletrônica, é obtido um mapa que minimiza o erro quadrático médio devido a erros nas fases da proteína  $\alpha_P$ ; desta maneira a figura de mérito atua como uma função peso para as melhores reflexões.

A figura de mérito,  $m$ , pode ser interpretada como a probabilidade de  $\alpha_P$  estar correta (equação 3.7), podendo a sua magnitude variar entre 0 e 1. Um valor de 1 para a figura de mérito representa, teoricamente, uma fase correta e 0 significa que qualquer valor para a fase é igualmente provável (na figura 3.2 significa que qualquer valor de  $F_P$  é possível como solução).

O método descrito acima (Blow e Crick, 1959) tem como desvantagem em sua formulação que a cada vez que alguma nova informação é obtida (ex. um novo derivado, método de nivelamento de solvente (cap. 4), fases calculadas da estrutura parcial etc), a soma  $\sum_j [\epsilon_j(\alpha_p)^2/2E_j^2]$  tem que ser reavaliada e a distribuição probabilidade inteira (total) deve ser recalculada. Assim, deve-se armazenar todos os dados para cada um dos derivados para qualquer novo cálculo. Hendrickson e Lattman (1970) mostraram que a função probabilidade pode ser expressa por:

$$P(\alpha_p) = \exp[A \cos(\alpha_p) + B \sin(\alpha_p) + C \cos(2\alpha_p) + D \sin(2\alpha_p)] \quad (3.15)$$

onde os coeficientes  $A$ ,  $B$ ,  $C$ ,  $D$  constituem uma informação completa da fase para uma reflexão. Desta forma, a inclusão de uma nova informação obtida, somente requer adições para estes coeficientes, ou seja, a nova representação separa os parâmetros de informação da fase das funções de fase e é isto que facilita as combinações de diferentes ou novas informações de fase (veja Blundell e Johnson, 1976).

### 3.2 - METODOLOGIA EMPREGADA PARA A CONSTRUÇÃO DO MAPA DE PATTERSON DIFERENÇA

Todos os programas que serão usados neste capítulo estão incluídos no pacote de programas do CCP4 (Colaborative Computational Project, Number 4, 1994). Deve-se ressaltar porém, que durante o processamento dos dados desta tese, este programa foi trocado de versão (versão anterior: CCP4, 1985), desta maneira foram utilizadas no processamento de dados duas versões do referido pacote de programas. A nova versão apresenta várias vantagens em termos de praticidade em relação à antiga. Vários programas foram totalmente refeitos e outros mantidos com poucas modificações, no entanto praticamente para todos os programas da versão 'antiga' existe um equivalente na versão 'nova', assim pode-se fazer praticamente o mesmo com as duas versões.

Nesta tese, toda a vez que for citado programas do CCP4, será da versão nova (exceto quando não houver equivalência entre as duas versões), pois o atual e futuro usuário utilizará esta versão nos seus trabalhos.

O mapa de Patterson diferença a ser determinado utiliza  $(|F_{PH}| - |F_P|)$  como fatores de estrutura, ou seja, irá se determinar  $|F_H|$ . Na realidade, a igualdade  $|F_H| = |F_{PH}| - |F_P|$  somente é verdadeira quando o ângulo entre  $|F_{PH}|$  e  $|F_P|$ ,  $(\alpha_{PH} - \alpha_P)$  é pequeno (veja figura 3.1), porém para a maioria dos casos esta aproximação é satisfatória, principalmente quando  $|F_H|$  é pequeno quando comparado com  $|F_{PH}|$  e  $|F_P|$  (Drenth, 1994, p.149); além disso, estes erros são minimizados quando é feito o refinamento das posições dos átomos pesados (item 3.3).

Para a construção do mapa de Patterson diferença podem ser utilizadas as seguintes etapas:

1. Transformação para o formato MTZ (este é o formato utilizado para o processamento de dados no CCP4 'novo'): **F2MTZ**.
2. Ordenação das reflexões em ordem crescente de *hkl*: **SORTMTZ**.
3. Escalonamento das amplitudes do derivado em relação ao nativo: **SCALEIT**.
4. Transformação de Fourier: Fator de estrutura → Densidade eletrônica: **FFT** (opção Patterson).

Em seguida, será dado um pouco mais de detalhes sobre os dois últimos programas citados acima.

O programa **SCALEIT** calcula e aplica uma função de escala do derivado para o nativo. A metodologia empregada para este escalonamento é bastante parecida com a empregada no item 2.2.4.4, na qual se aplica uma função de escala aos dados do derivado, que pode ser uma constante ( $C$ ) ou dependente da resolução ( $C \exp(-B \sin \theta / \lambda)$ , onde  $C$  e  $B$  são constantes,  $\lambda$  o comprimento de onda e  $\theta$  o ângulo de Bragg). Após a aplicação destas funções há um processo de refinamento dos valores  $C$  e  $B$  (item 2.2.4.4). O programa possibilita a análise dos dados dos derivados com relação ao nativo comparando o  $R_{\text{deriv}} = \frac{\sum |F_{\text{PH}}| - |F_{\text{P}}|}{\sum |F_{\text{P}}|}$  inclusive por faixas de resolução, facilitando uma melhor decisão para o usuário se o derivado a ser utilizado é conveniente, ou ainda se é realmente um derivado.

O programa **FFT** (*Fast Fourier Transform*) faz, como o nome indica, uma transformação de Fourier rápida, pois este é um passo que consome muito tempo, principalmente durante as etapas de refinamento onde para cada ciclo deve haver uma transformação de Fourier. O programa pode ser usado para calcular mapas de Fourier, Fourier diferença, Patterson e Patterson diferença. Qualquer grupo espacial pode ser utilizado.

O método reduz o tempo da transformação de Fourier (cálculo de fatores de estrutura via densidade eletrônica e vice-versa) dividindo a cela unitária em pontos de uma grade (*grid points*) ao longo das três direções principais em múltiplos inteiros,  $nx$ ,  $ny$ ,  $nz$ , os quais devem ser múltiplos de números primos pequenos, isto é 2, 3, 5 e 7. Além disso, estes valores devem ser da ordem de três vezes o máximo valor de  $h$ ,  $k$ ,  $l$  para os fatores de estruturas de entrada (este valor pode ser calculado pela fórmula:  $nx = 3a/d_{\text{min}}$ , onde  $a$  é o comprimento da cela unitária e  $d_{\text{min}}$  é a resolução limite, ambos em Angstroms). Valores menores que duas vezes o máximo de  $h$ ,  $k$ ,  $l$  resultarão em mapas densidades eletrônicas incorretos (devido à amostragem das funções a serem transformadas ser insuficiente para descrever de maneira completa a função densidade eletrônica), valores muito superiores a três vezes consumirão muito tempo de máquina e os resultados não serão tão superiores.

Uma discussão completa sobre o procedimento teórico para cristalografia (pois a técnica já era usada em outros ramos da ciência desde 1965) é descrita no artigo de Eyck, 1973.

### 3.3 - RESOLUÇÃO DO MAPA DE PATTERSON DIFERENÇA DO DERIVADO 2-CLOROMERCÚRIO 4-NITROFENOL

Como descrito no capítulo anterior foram obtidos três derivados, os quais posteriormente mostraram ser somente dois, pois os derivados de Ácido Mersálico e 2-Cloromercúrio 4-nitrofenol (CMNF) tratam-se do mesmo, porém o de CMNF possui menos sítios de ligação de átomos pesados.

Devido ao derivado CMNF apresentar um mapa bem mais simples (mais 'limpo') que os outros dois, este será o mapa utilizado inicialmente para o cálculo de posições atômicas.

Antes de começar a analisar o caso da GlcN6P desaminase, cabem algumas ressalvas. Primeiramente, que a solução da função de Patterson não é única, há diferentes conjuntos de posições que podem explicar o mesmo Patterson; estas diferentes soluções caem em duas categorias: mudança de origem e escolha de enantiomorfo. A escolha da origem é arbitrária, e a escolha do enantiomorfo correto é importante porque somente a solução de mão direita é correta para proteínas. Não há como saber durante os cálculos para obtenção do mapa qual o enantiomorfo correto, ou seja, somente de posse do mapa densidade eletrônica da proteína é possível decidir qual o correto, porém se o enantiomorfo estiver errado, pode-se modificá-lo simplesmente invertendo os sinais de todas as posições de átomos pesados e recalculando as fases.

As seções e linhas de Harker podem ser obtidas fazendo a subtração entre todas as operações de simetria do Grupo Espacial R32, ao qual pertencem os cristais obtidos para a GlcN6P desaminase (tabela 3.1).

Tabela 3.1 - Vetores de Harker do grupo Espacial R32.

R 32	x, y, z	-y, x-y, z	-x + y, -x, z	y, x, -z	x-y, -y, -z	-x, -x+y, -z
x, y, z	0,0,0	x+y, -x+2y, 0	2x -y, x+y, 0	x-y, -x+y, 2z	y, 2y, 2z	2x, x, 2z
-x, x - y, z	-y-x, x-2y, 0	0,0,0	x-2y, 2x-y, 0	-2y, -y, 2z	-x, x, 2z	x-y, 2x-2y, 2z
-x + y, -x, z	-2x+y, -x-y, 0	-x+2y, 2x-y, 0	0,0,0	-x, 2x, 2z	-2x+2y, -x+y, 2z	y, -y, 2z
y, x, -z	y-x, x-y, -2z	2y, y, -2z	x, 2x, -2z	0,0,0	-x+2y, x+y	x+y, 2x-y, 0
x-y, -y, -z	-y, -2y, -2z	x, -x, -2z	2x -2y, x-y, -2z	x-2y, -x-y, 0	0,0,0	2x-y, x-2y, 0
-x, -x+y, -z	-2x, -x, -2z	-x+y, -2x+2y, -2z	-y, y, -2z	-x-y, -2x+y, 0	-2x+y, -x+2y, 0	0,0,0

Da tabela 3.1 pode-se notar que há dois tipos de seções de Harker, uma delas devido aos eixos de ordem 3, (seções do tipo  $u,v,0$ ) e a outra devida aos eixos de ordem 2 ( seções do tipo  $u,u,w$ ).

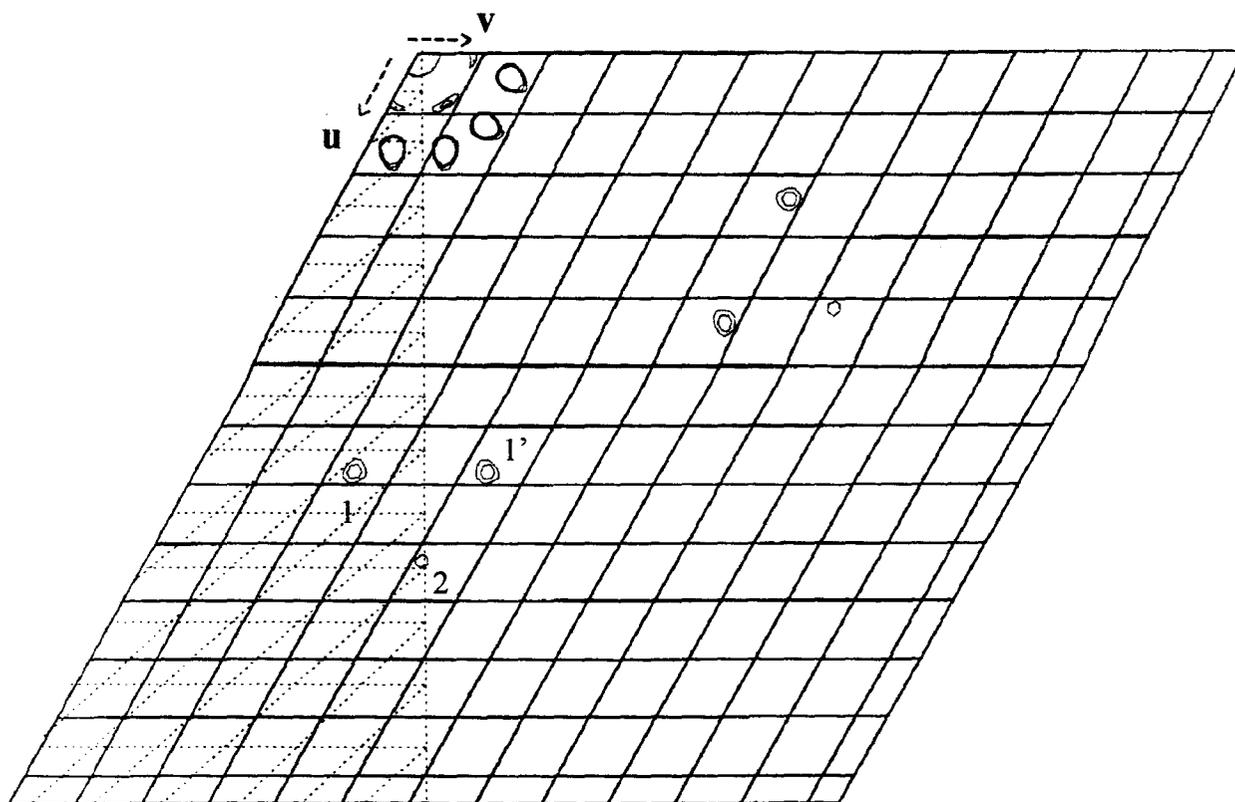


Figura 3.7 - Seção de Harker para  $Z=0$  ( $u, v, 0$ ), para o derivado 2-cloro mercúrio 4-nitrofenol (os números de 1 e 2 correspondem aos picos analisados e a área hachurada representa a unidade assimétrica do grupo espacial  $R\bar{3}m$ )

Deve-se notar que, devido à natureza centro-simétrica da síntese de Patterson, o grupo espacial do mapa de Patterson dos cristais da GlcN6P desaminase a ser analisado (simetria de Laue) é  $R\bar{3}m$  (grupo espacial nº 166 da International Table for X-ray Crystallography Vol I). Assim, a unidade assimétrica relativa a este grupo espacial é:  $0 \leq x \leq 2/3$ ;  $0 \leq y \leq 2/3$ ;  $0 \leq z \leq 1/6$ ;  $x \leq 2y$ , esta é desenhada de maneira hachurada na figura 3.7.

A figura 3.7 representa o plano  $xy$ , ou seja, representa qualquer uma das seções de Harker do tipo  $u,v,0$ . Deve-se levar em conta também, que a cada pico do mapa de Patterson diferença pode-se associar 36 operações de simetria do grupo de Patterson  $R\bar{3}m$ .

Por inspeção da figura 3.7, nota-se que há somente dois picos independentes, que são possíveis vetores interatômicos (indicados pelos números 1 e 2 na figura 3.7). As coordenadas dos átomos relativos a estes picos serão analisadas abaixo.

O cálculo das coordenadas dos átomos pesados requer alguns passos a seguir:

- Usando o pacote de programas CCP4, deve-se fazer o processamento de dados (item 3.1) obtendo-se as amplitudes de difração do derivado, das quais calcula-se um mapa de Patterson diferença (a seção  $Z=0$  é dada pela figura 3.7)
- Para calcular as coordenadas fracionárias de cada pico (figura 3.7) deve-se fazer uma simples regra de três, atribuindo-se 1/2 a ordenada e a abcissa, medidas em milímetros (86 a cada uma delas), medir  $u$  e  $v$  dos picos (em milímetros) no mapa e calcular as coordenadas fracionárias de  $u$  e  $v$ :

Pico 1) Coordenadas em milímetros: (58, 20.5)

$$\begin{array}{rcl} 1/2 & \text{---} & 86 \\ u & \text{---} & 58.0 \Rightarrow u = 0.337 \end{array}$$

$$\begin{array}{rcl} 1/2 & \text{---} & 86 \\ v & \text{---} & 20.5 \Rightarrow v = 0.119 \end{array}$$

Assim, as coordenadas fracionárias são: (0.337, 0.119)

Da mesma maneira, tem-se respectivamente as coordenadas em milímetros e as fracionárias para o pico 2:

$$\text{pico 2) } (71.5, 36) \Rightarrow (0.415, 0.209)$$

- Cálculo das coordenadas  $X$  e  $Y$ : Para tanto usa-se as seções de Harker do tipo  $u, v, 0$  da tabela 3.1, por exemplo:  $X+ Y, 2X - Y, 0$

$$\text{--- } u = X + Y$$

$$\text{--- } v = 2X - Y$$

Resolvendo o sistema, tem-se:

$$\text{pico 1) } X = 0.152 \quad Y = 0.186$$

$$\text{pico 2) } X = 0.208 \quad Y = 0.207$$

Determinada as coordenadas  $X$  e  $Y$  dos possíveis átomos pesados, resta agora calcular a coordenada  $Z$ , que pode ser obtida pelas seções de Harker do

tipo  $u,u,w$ , citada acima. Ou seja, devem ser calculadas as coordenadas  $x$ ,  $y$  e  $z$  pelas seções de Harker deste tipo e compará-las com as já calculadas pelas do tipo  $u,v,0$ .

As seções de Harker do tipo  $u,v,w$  podem ser divididas em três grupos: grupo 1:  $-x,x,2z$ ;  $y,-y,2z$ ;  $x-y,-x+y,2z$ ; grupo 2:  $2x,x,2z$ ;  $-2y,-y,2z$ ;  $-2x+2y,-x+y,2z$ ; grupo 3:  $-x,2x,2z$ ;  $y,2y,2z$ ;  $x-y,2x-2y,2z$

Na figura 3.8 é mostrada uma seção de Harker  $-x,x,2z$ , onde são mostrados os quatro possíveis picos desta seção de Harker representados pelas letras A,B,C.

Será analisado, a seguir, cada um dos possíveis picos:

O pico A tem valor muito próximo de  $Z=0$ , assim os valores para as coordenadas do átomo pesado são:

$$\text{pico A: } X \text{ ou } Y = 0.224, Z \cong 0$$

Para analisar os picos B e C, deve-se levar em conta que haverá três picos com o mesmo  $Z$  que pertencem ao grupo 1 de seções de Harker citadas acima:  $-x, x, 2z$ ;  $y,-y,2z$ ;  $x-y,-x+y,2z$ . Assim para um  dado valor de  $w$  ( $w=2z$ ) haverá três valores para   $u$ .

$$\text{pico B: } (w=0.067 \Rightarrow Z=0.033)$$

$$\begin{aligned} \text{projeções em } u: \quad u' &= 0.201 \\ u'' &= 0.840 - 1 = 0.160 \\ u''' &= 0.907 - 1 = 0.030 \end{aligned}$$

Estes valores correspondem a  $x$ ,  $y$ ,  $x-y$  respectivamente, quando comparados aos valores obtidos para as seções de Harker  $Z=0$ .

$$\text{pico C: } (w=0.044 \Rightarrow Z=0.022)$$

$$\begin{aligned} \text{projeções em } u: \quad u' &= 0.119 \\ u'' &= 0.366 \\ u''' &= 0.504 \end{aligned}$$

Somando-se  $2/3$  às projeções acima obtém-se os valores para  $u'$ ,  $u''$  e  $u'''$ : 0.207, 0.030, 0.174 que são os valores de  $y$ ,  $x-y$  e  $x$  respectivamente. Desta maneira, este pico é equivalente ao B pela operação de simetria  $(y,x,z)$  do grupo espacial  $R32$ .

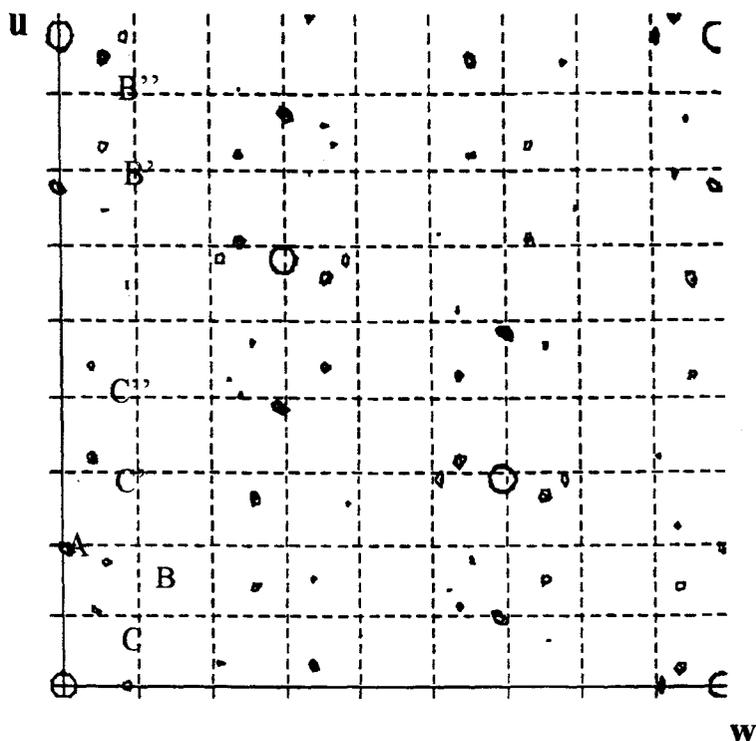


Figura 3.8 - Seção de Harker do tipo  $-x,x,2z$  (plano  $uw$ ) para o derivado 2-cloromercúrio 4-nitrofenol (com os picos A, B, C e as projeções dos picos B e C devido às outras seções de Harker do tipo 1:  $y,-y,2z$ ;  $x-y,-x+y,2z$ , B', B'', C' e C'' indicadas no mapa).

Assim comparando-se os dois tipos de Seções de Harker  $(u,v,0)$  e  $(u,u,w)$ , tem-se:

O pico A é o pico 2 no plano  $uv$ , e suas coordenadas são:  $(0.21, 0.21, 0.0)$

O pico B (ou C) é o pico 1 (ou 1') no plano  $uv$ , e suas coordenadas são:  $(0.19, 0.15, 0.03)$ .

Estes são os dois picos de Harker para este derivado.

Com o intuito de confirmar as posições dos sítios de átomos pesados obtidos e tentar localizar novos, pode-se utilizar o programa VECSUM (pertencente ao CCP4). O VECSUM usa para tanto o método da superposição vetorial (Stout & Jensen, 1968, pp. 339-352) que será discutido simplifadamente abaixo.

O método de Patterson (item 3.1) foi descrito em termos do conjunto de vetores entre todos os átomos constituintes da estrutura, tendo os picos relativos a estes vetores 'altura' proporcional ao produto dos números atômicos dos átomos. Uma maneira totalmente equivalente de 'olhar' para a função de Patterson considera esta como uma combinação de  $N$  imagens da estrutura cristalina original, onde cada uma delas é transladada por uma quantidade necessária para trazer cada uma dos átomos para a origem. A figura 3.9 exemplifica um mapa de Patterson obtido desta maneira para uma estrutura com quatro átomos.

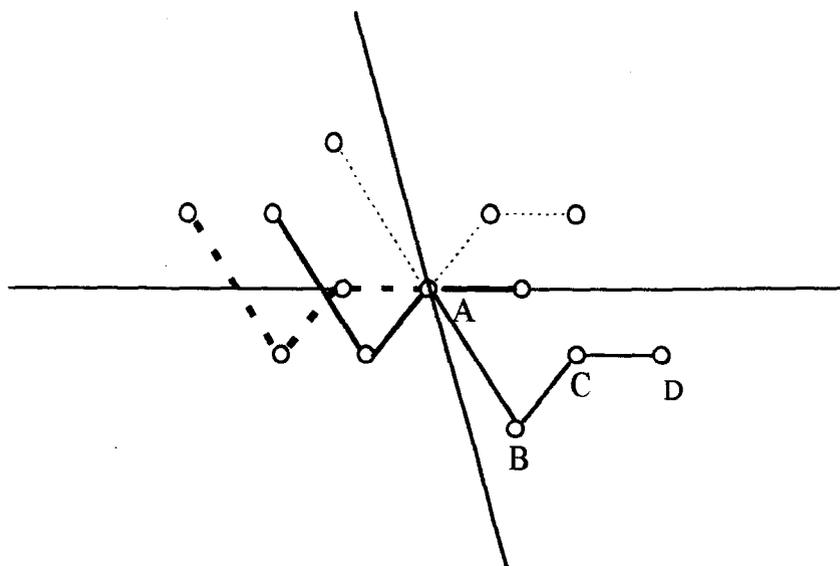


Figura 3.9 - A função de Patterson para um conjunto de  $N$  átomos na cela unitária (neste caso A,B,C e D) que contém  $N$  imagens do conjunto (Stout e Jensen, 1968, p.345).

O método da superposição, utilizado pelo VECSUM, faz a superposição de diversas cópias do mapa de Patterson original. Para compreender o princípio deste método, irá se supor um pico (AB) em um mapa de Patterson que representa o vetor de um átomo A até outro átomo B. O mesmo mapa de Patterson contém imagens da estrutura original com A colocado na origem e com B colocado nesta, as quais serão chamadas, respectivamente de  $I(A)$  e  $I(B)$  (para tanto considere a figura 3.9 com apenas dois átomos A e B e conseqüentemente, com duas imagens  $I(A)$  e  $I(B)$  representados na figura com as linhas — e .....). Considere agora o mapa de Patterson original, como mapa M, e seja um mapa N, que foi transladado de tal maneira que sua origem cai no pico (AB) do mapa M. Assim, o ponto pertencente a  $I(A)$  no mapa M coincidirá com o de  $I(B)$  no mapa N, ou seja, aqueles picos que pertencem a outros conjuntos não coincidirão, exceto por acidente. Então, em princípio, a estrutura ou uma aproximação desta, pode ser obtida selecionando-se somente os picos que se superpuserem nos dois mapas.

Há uma grande diferença deste método em relação ao feito manualmente por nós (utilizando seções de Harker) para localizar os sítios de ligação de átomos pesados, é que no caso da superposição vetorial são usados todos os picos interatômicos, ou seja, os autovetores (picos entre mesmos átomos relacionados por simetria) e os vetores cruzados (entre átomos diferentes). Em contrapartida, os picos utilizados por nós nas seções de Harker são autovetores (veja item 3.1).

Assim, na prática, com um ou mais picos inicialmente pode-se obter outros utilizando o VECSUM. Estes novos picos vão sendo colocados um a um numa lista de sítios do programa o qual vai refinando a ocupação destes. Os picos que

após o refinamento tiverem uma ocupação razoável são os sítios de ligação de átomos pesados.

O passo seguinte é utilizar o programa VECREF para refinar as posições destes sítios. Este programa faz basicamente um refinamento por mínimos quadrados no espaço real, ou seja, no espaço dos vetores interatômicos do mapa de Patterson diferença. A minimização é feita entre dois mapas de Patterson diferença utilizando como coeficientes os valores:  $|\mathbf{F}_H|^{\text{calc}}$  (obtido pela equação  $F_H(\text{hkl}) = \sum f_j \exp 2\pi i (\text{hx}_j + \text{ky}_j + \text{lz}_j)$ , onde  $x_j, y_j, z_j$  é o sítio do átomo pesado  $j$ ) e  $|\mathbf{F}_H|^{\text{obs}}$  ( $\approx |\mathbf{F}_{PH}|^{\text{obs}} - |\mathbf{F}_P|^{\text{obs}}$ ).

Esta maneira de minimização traz vantagens em relação a maneira convencional de refinar posições de átomos pesados, que é feita no espaço recíproco usando a relação  $E = \sum_{\text{hkl}} w(\text{hkl}) [ |F_H^{\text{obs}}|^2 - |F_H^{\text{calc}}|^2 ]^2$ . O refinamento no espaço dos vetores interatômicos (mapas de Patterson) é somente feito próximo dos picos dos vetores interatômicos, onde a densidade dos pontos é significativamente positiva. Ao passo que, no refinamento feito no espaço recíproco ocorre o contrário, pois a maior parte do espaço real está vazio e a probabilidade de superposição é pequena. Desde que os fatores de estrutura representam em maior parte este espaço vazio (cada fator de estrutura contém informação de todos os pontos do espaço real), e como os fatores de estrutura são o inverso do espaço real, eles são dominados por erros que originam-se de pequenas diferenças entre grandes quantidades. Como resultado, o refinamento de espaço recíproco de parâmetros de átomos pesados tem convergência bastante lenta e é insensível para distinguir entre sítios e erros no início do refinamento.

No caso do derivado CMNF foi colocado inicialmente o pico (0.19, 0.15, 0.03) no programa VECSUM com o qual obteve-se o outro pico, confirmando assim as suas coordenadas. Em seguida, as posições destes sítios foram refinadas (VECREF), obtendo-se os valores dados na tabela 3.2.

Tabela 3.2 - Sítios de ligação do derivado 2-Cloromercúrio 4-nitrofenol.

Sítios	X	Y	Z	Ocupação
1	0.19	0.15	0.05	0.12
2	0.45	0.11	0.01	0.17

Da tabela 3.2 nota-se que o sítio 2 é equivalente ao pico 2 (obtido usando as seções de Harker) somando-se  $-2/3, -1/3$  às coordenadas  $x$  e  $y$ . É útil observar, ainda, que para ambos os sítios, a ocupação é bastante baixa (menor que 20%).

### 3.4 - DETERMINAÇÃO DAS POSIÇÕES DOS ÁTOMOS PESADOS DO ÁCIDO MERSÁLICO E GERAÇÃO INICIAL DAS FASES

O cálculo das posições dos átomos pesados para o derivado de ácido mersálico deve seguir o mesmo procedimento utilizado para o derivado CMNF, no qual obtém-se as coordenadas  $x, y$  dos picos na seção de Harker  $Z=0$  ( $u, v, 0$ ) e deve-se compará-las com as coordenadas obtidas para a seção  $-x, x, 2z$  ( $u, u, w$ ). Nestas seções de Harker foram 5 picos (numerados de 1 à 5) nas figuras 3.10 e 3.11 dadas abaixo.

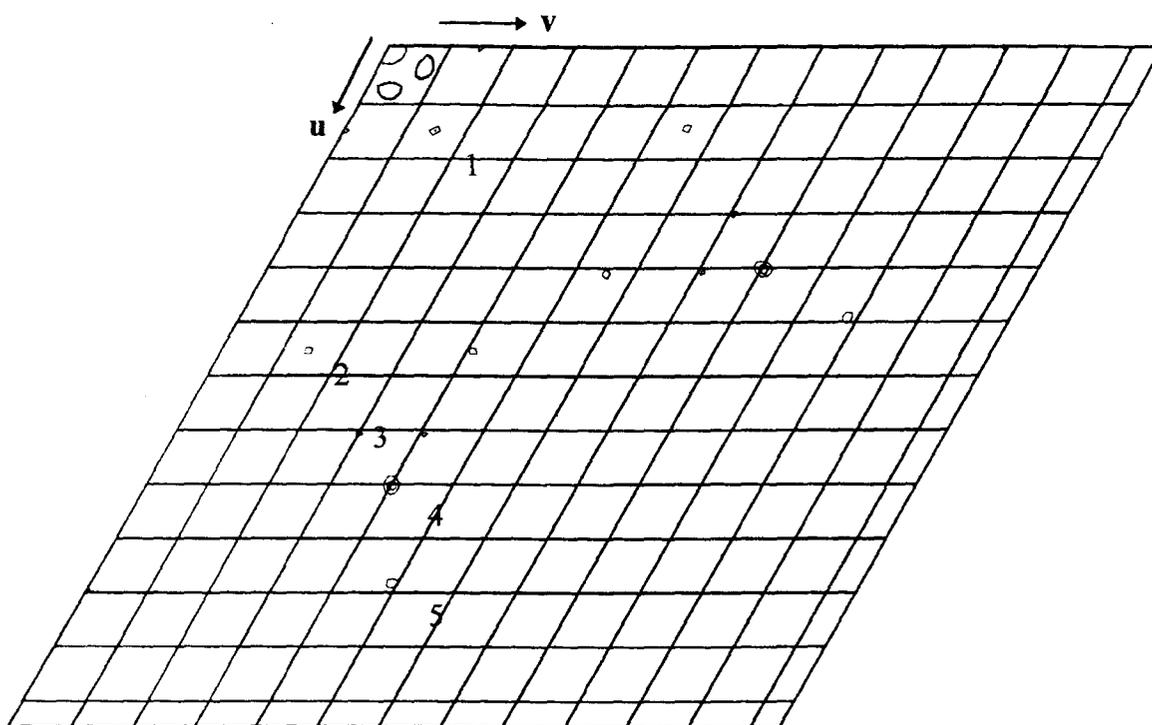


Figura 3.10 - Seção de Harker  $Z=0$  para o derivado de ácido mersálico.

Da figura 3.10 nota-se muita semelhança com relação à figura 3.7 do derivado CMNF, ou seja, há alguns picos a mais no Ácido Mersálico; este fato induz a pensar que no Ácido Mersálico há os mesmos sítios do CMNF e mais alguns.

Utilizando os programas VECSUM e VECREF da mesma maneira que para o CMNF, obtém-se os seguintes sítios de ligação para o derivado Ácido Mersálico (tabela 3.3).

Tabela 3.3 - Sítios de Ligação do derivado Ácido Mersálico.

Sítio	X	Y	Z	Ocupação
1	0.4953	0.1284	0.0207	0.4121
2	0.1787	0.1195	0.0534	0.4608
3	0.2124	0.1425	0.0641	0.2059
4	0.6197	0.3392	0.0499	0.2553
5	0.5181	0.1581	0.0103	0.2156

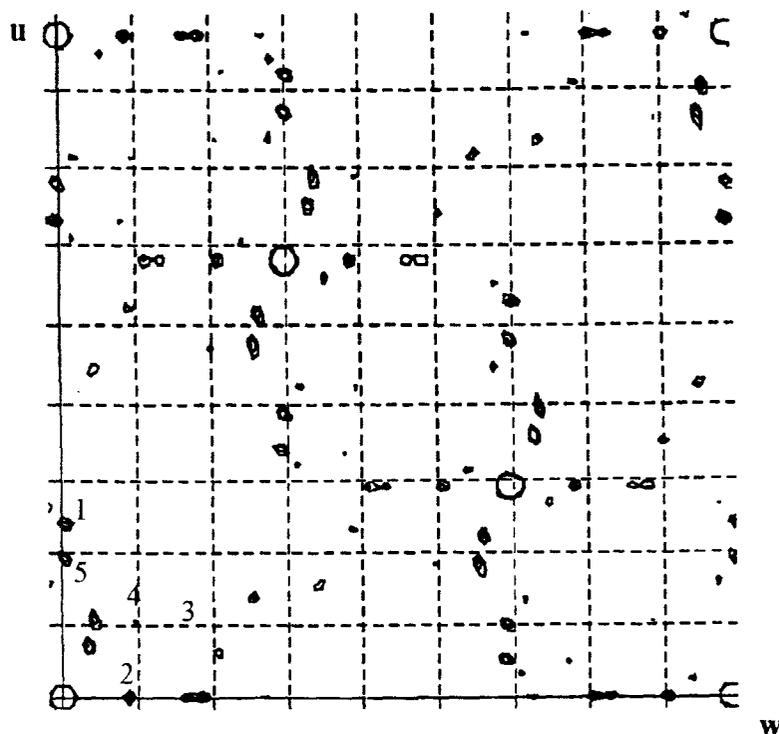


Figura 3.11 - Seção de Harker -x, x, 2z para o derivado de ácido mersálico

Da tabela 3.3 conclui-se ainda que os sítios 3 e 4 do ácido mersálico são equivalentes aos dois sítios do CMNF (tabela 3.2).

Assim, tem-se até aqui, na realidade, um único derivado isomorfo para ser utilizado para fases da proteína, o ácido mersálico que é um derivado com qualidade de dados bem superior ao CMNF (veja no capítulo 2, tabela 2.2, a completude dos derivados, ácido mersálico: 78.4%, CMNF: 51.1%).

Tendo-se obtido as posições dos átomos pesados (e por sua vez  $F_H$  e suas fases  $\alpha_H$ , lembrando que  $F_H = F_H \exp(i\alpha_H)$ ), pode-se obter, a partir da equação 3.5, a primeira aproximação de fases para o modelo, que são conhecidas como fases SIR (*Single Isomorphous Replacement*). Estas serão usadas para obter as posições dos átomos pesados dos outros derivados por Fourier diferença (neste caso, chamado de Fourier Cruzado ou *Cross Fourier*) usando os coeficientes:  $|F_{PH}| - |F_P|$ ,  $\alpha_{SIR}$ .

As fases SIR são obtidas como descrito na introdução do capítulo (equação 3.5), e as posições dos átomos pesados podem ser obtidas pela observação de um mapa densidade eletrônica utilizando a expressão (Blundell e Johnson, 1976, p.349; Drenth, 1994, p.151 e Stryer *et al.*, 1964)

$$\Delta\rho(\vec{r}) = \frac{1}{V} \sum_{\vec{h}} m[|F_{PH}(\vec{h})| - |F_P(\vec{h})|] \exp[-2\pi i \vec{h} \cdot \vec{r} + i\alpha_{SIR}] \quad (3.16)$$

onde  $m$  é a figura de mérito (equação 3.14).

Para obter um mapa Fourier diferença, tem-se os seguintes passos a seguir, após o procedimento 1 à 3 dado no item 3.2:

1. Usando o programa MLPHARE geram-se as fases iniciais e figuras de mérito associadas.
2. Com o programa FFT, pode-se fazer um mapa Fourier diferença com coeficientes ( $|F_{PH}| - |F_P|$ ), usando para  $|F_{PH}|$  os dados coletados do ácido mersálico.
3. Geração do mapa (pode ser feita com o programa MAPEXCHANGE) para a visualização na estação gráfica.

Observando o mapa (isto poder ser feito diretamente na Estação Gráfica), deve-se compará-lo com os picos previstos pelo mapa de Patterson para este derivado e observar-se os picos que ocorrem em ambos, a fim de obter a confirmação de cada um destes.

### 3.5 - DETERMINAÇÃO DAS POSIÇÕES DOS ÁTOMOS PESADOS DO DERIVADO $K_2PtCl_4$ E GERAÇÃO DAS FASES MIR.

Para a determinação das posições dos átomos pesados do derivado  $K_2PtCl_4$  foi utilizado o procedimento de Fourier cruzado (*Cross Fourier*), descrito no item anterior.

Nas figuras 3.12 e 3.13 tem-se respectivamente as seções  $Z=0$  e  $-x, x, 2z$  para o derivado de Pt, que deve ser comparado com o mapa Fourier diferença para a localização dos átomos pesados.

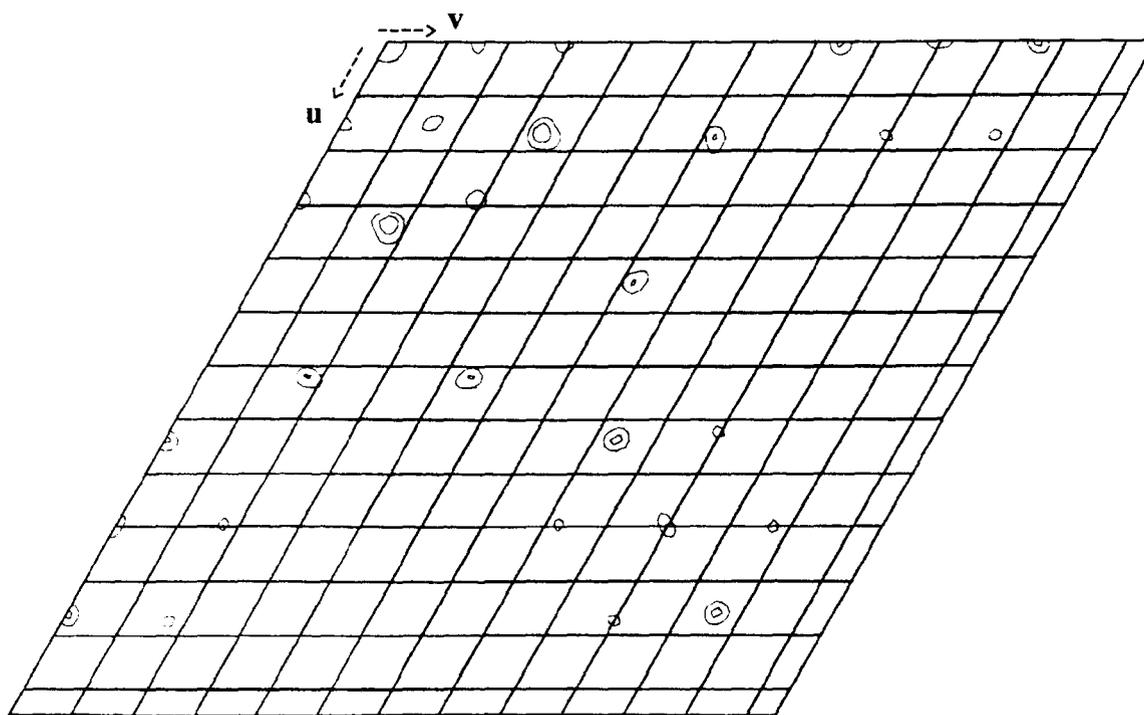


Figura 3.12 - Seção de Harker  $Z=0$  para o derivado  $K_2PtCl_4$

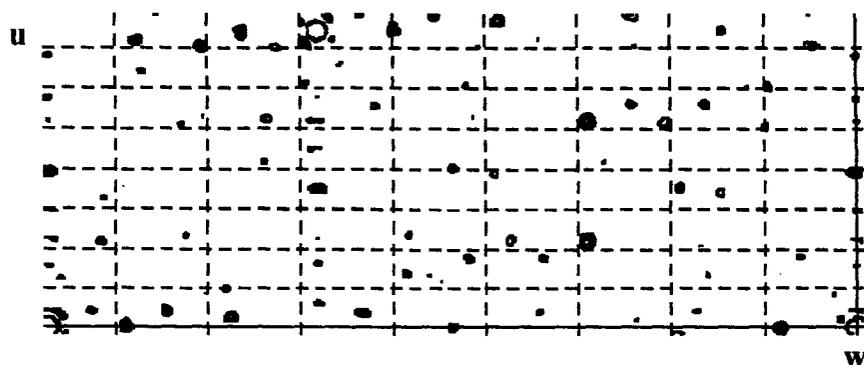


Figura 3.13 - Seção de Harker  $-x, x, 2z$  para o derivado  $K_2PtCl_4$

Na tabela 3.4 são dados os sítios de ligação para o derivado de Pt após o refinamento com VECREF e com o programa MLPHARE que utiliza refinamento no espaço recíproco, cuja metodologia é descrita a seguir.

Tabela 3.4 - Sítios de ligação do derivado  $K_2PtCl_4$

Sítios	X	Y	Z	Ocupação
1	0.7135	0.1101	0.0616	0.1587
2	0.6748	0.5744	0.2375	0.1176
3	0.5789	0.3944	0.3536	0.0740
4	0.6025	0.5123	0.1353	0.1358
5	0.5301	0.1679	0.3293	0.1145
6	0.5154	0.4654	0.2800	0.1554
7	0.3325	0.0347	0.3078	0.0855

O programa MLPHARE faz o refinamento dos sítios de átomos pesados e ocupação destes, porém este refinamento será diferente do método utilizado pelo programa VECREF, pois agora já se tem fases calculadas.

O método a ser descrito aqui (Drenth, 1994, pp.169-170) necessita de fases *preliminares* para proteína  $\alpha_P$  para poder desenhar o triângulo vetorial  $F_P + F_H = F_{PH}$  (figura 3.1) para cada reflexão. O módulo e direção aproximados de  $F_P$  e  $F_H$  já são conhecidos (a fase  $\alpha_P$  é uma fase preliminar e  $\alpha_H$  para o derivado de Pt pode ser calculado do mesmo modo que para o derivado CMNF) e  $F_{PH}$  seria

colocada desta maneira, a completar o triângulo da figura 3.4, que na prática porém, pode ser mais curto ou comprido a fim de que o triângulo da figura fique correto. Faz-se então, usando o método dos mínimos quadrados, a minimização entre os valores de  $|F_{PH}|^{obs}$  que foram medidos experimentalmente e os valores deste calculado pela expressão (usando a lei dos cossenos a partir da figura 3.4):

$$|F_{PH}^{calc}| = \{ |F_P|^2 + |F_H|^2 + 2|F_P| |F_H| \cos(\alpha_H - \alpha_P) \}^{1/2} \quad (3.17)$$

onde são usados para  $|F_P|$  valores obtidos experimentalmente e para  $|F_H|$  valores calculados a partir das posições dos átomos pesados do derivado (Pt no caso desta tese).

Desta maneira, a expressão a ser minimizada é:

$$E = \sum_{hkl} \{ k_j (|F_{PH}|^{obs})_j - (|F_{PH}|^{calc})_j \}^2 \quad (3.18)$$

onde j representa cada derivado e  $k_j$  é um fator de escala.

Note que os parâmetros que variam na equação acima são na realidade os relacionados com  $|F_H|$  (da equação 3.17), ou seja, coordenadas atômicas e ocupação dos átomos pesados, de maneira que o gráfico da figura 3.4 fique o mais próximo possível de fechar o triângulo.

Tem-se agora condições de calcular as fases MIR (cuja metodologia é descrita no item 3.1.2) utilizando o programa MLPHARE usando as fases do ácido Mersálico e do  $K_2PtCl_4$  até 3 Å de resolução.

A seguir, são dadas algumas tabelas mostrando a qualidade das fases obtidas. Na tabela 3.5 é mostrado a figura de mérito média por faixa de resolução para reflexões centradas e não centradas; e na tabela 3.6 mostra a *Phasing Power* média (definida por  $\sum_n |F_H|^{calc} / \sum_n |\epsilon|$ , onde  $\epsilon$  é definido na equação 3.6 e figura 3.4) em função da resolução para cada um dos derivados.

Tabela 3.5 - Figura de mérito em função da resolução para reflexões centradas e não centradas

Resolução (Å)	n° de reflexões não centradas	figura de mérito refl. não centradas	n° de reflexões centradas	figura de mérito refl. centradas
10.65	190	0.5504	48	0.7958
7.26	494	0.5784	80	0.8790
5.51	970	0.5656	113	0.7918
4.44	1807	0.5010	198	0.6872
3.72	2683	0.4521	232	0.6017
3.20	3698	0.4078	244	0.6120
2.81	4751	0.3641	237	0.5705
2.50	5815	0.3412	221	0.5719
TOTAIS	20408	0.4057	1373	0.6443

A figura de mérito ( $m$ ) é um número entre 0 e 1 que é atribuído a cada reflexão. O valor médio para fases MIR da GlcN6P desaminase é de  $\bar{m} = 0.521$  para reflexões centradas e não centradas.

Tabela 3.6 - *Phasing Power* (definição no texto) em função da resolução para reflexões centradas e não centradas para os derivados de ácido mersálico e  $K_2PtCl_4$ .

Resolução (Å)	<i>Phasing Power</i> refl. não centradas ácido mersálico	<i>Phasing Power</i> refl. centradas ácido mersálico	<i>Phasing Power</i> refl. não centradas $K_2PtCl_4$	<i>Phasing Power</i> refl. centradas $K_2PtCl_4$
10.54	1.6	1.1	1.2	1.0
7.26	1.9	1.2	1.4	1.3
5.51	1.5	1.3	1.8	1.3
4.44	1.1	0.8	1.4	1.3
3.72	1.0	0.6	1.3	0.8
3.20	0.9	0.5	1.3	0.9
2.81	0.8	0.4	1.4	1.1
2.50	-	-	1.5	0.9

O número *Phasing Power* representa a mais significativa estatística para se analisar a qualidade de um derivado. Um valor considerado 'bom' para *Phasing Power* é acima de 1 (McRae, 1994, p.141). A análise da tabela 3.6 mostra que os valores obtidos para *Phasing Power* para o derivado de Platina são bem melhores que para o derivado de mercúrio, fato devido à melhor qualidade dos dados do derivado de Pt.

Com os derivados isomorfos obtidos foi possível calcular as fases MIR até 3 Å de resolução e além disso, pode-se utilizar os dados de  $K_2PtCl_4$  para calcular as fases SIR no intervalo de 3.0 à 2.5 Å de resolução. Ambas serão utilizadas para construir o modelo.

Após a obtenção das fases MIR e posterior análise do mapa densidade eletrônica da proteína, principalmente das hélices- $\alpha$ , observou-se que havia sido escolhido o enantiomorfo incorreto. Como descrito no item 3.3, as fases corretas puderam ser obtidas simplesmente calculando-se novamente as coordenadas dos átomos pesados (invertendo o sinal destas) e recalculando as fases.

### 3.6 - LOCALIZAÇÃO DOS SÍTIOS DOS ÁTOMOS PESADOS NA PROTEÍNA NATIVA

Com o intuito de estudar as ligações dos átomos pesados com a proteína, foi feito um estudo posterior à construção do modelo da proteína.

A partir das coordenadas ortogonais dos átomos pesados colocadas juntas ao modelo da proteína nativa (dois monômeros da unidade assimétrica), observou-se quais são os aminoácidos que estão próximos dos átomos pesados fazendo possíveis ligações com estes.

Como visto no item 3.4, há cinco sítios de ligação para o derivado de ácido mersálico (tabela 3.3). Os sítios 2 e 3 estão muito próximos entre si (aproximadamente 3 Å) e ambos estão eqüidistantes da Cis 118. A hipótese mais razoável para este arranjo é que a Cis 118 possui duas conformações possíveis que estão distribuídas aleatoriamente ao longo do cristal, e assim, as moléculas de mersalil podem ligar-se a uma ou outra conformação. Quanto aos sítios 1 e 5, ocorre a mesma situação dos sítios 2 e 3, porém estes estão no outro monômero da unidade assimétrica (será chamado aqui de monômero 1). O sítio de ligação 4 do ácido mersálico está próximo da Cis 228 (à 3.55 Å da C $\beta$  da Cis 228) no mesmo monômero dos sítios 2 e 3 (monômero 1).

Tabela 3.7 - Ligações do derivado K<sub>2</sub>PtCl<sub>4</sub> com a GlcN6P desaminase.

Sítos	Resíduos de ligação	Distância aproximada (Å) (átomo pesado-enzima)	Monômero de ligação
1	Met 223	2.82	1
	Thr 251	2.91	
2	Met 223	3.21	2
	Thr 251	3.88	
3	Met 46	2.14	2
	Lis 50	2.66	
	Asn 92	3.33	
4	Met 245	4.19	2
	His 164	3.67	
	Asp 165	3.38	
5	Met 46	2.50	1
	Lis 50	3.35	
	Asn 92	3.63	
6	Met 245	3.58	1
	His 164	3.89	
	Asp 165	3.47	
7	Met 55	2.68	1
	Lys 14	2.98	
	Glu 60	3.49	

É importante notar que a ligação de ácido mersálico à Cis 228 se deu somente em um dos monômeros. Este fato pode ser justificado pela acessibilidade deste resíduo ao solvente e aos íons de átomos pesados as moléculas situadas no cristal. Observando o empacotamento das moléculas no cristal nota-se que o hexâmero é construído com os monômeros 2 frente a frente

com os monômeros 2 do outro hexâmero, enquanto os monômeros 1 ficam bastante acessíveis ao solvente.

Os sítios de ligação do derivado  $K_2PtCl_4$  à enzima são dados na tabela 3.7.

Nota-se da tabela 3.7 que os sítios 1 e 2, 3 e 5, 4 e 6 são equivalente nos dois monômeros, porém o sítio 7 só há no monômero 1 devido provavelmente ao empacotamento deste monômero. Por inspeção nota-se que o sítio 7 está na interface entre os monômeros 1 de dois hexâmeros.

Comparando com a literatura existente sobre ligação de átomos pesados (Blundell e Johnson, 1976, p. 199-235). Verifica-se que usualmente ocorre ligação de derivados de Hg com cisteínas livres na enzima, bem como é comum a ligação de derivados de platina com metioninas, fato que ocorreu em todos os sítios de Pt no caso deste tese (tabela 3.7).

## CAPÍTULO 4

# CONSTRUÇÃO E REFINAMENTO DO MODELO DA PROTEÍNA NATIVA

Após se ter obtido o primeiro conjunto de fases, o próximo passo será a interpretação do mapa densidade eletrônica em termos da cadeia polipeptídica. Se a maior parte desta pode ser traçada utilizando o mapa densidade eletrônica, pode-se iniciar o refinamento (item 4.3) se não, ou seja, a densidade eletrônica não tem qualidade suficiente para se traçar sem ambigüidade a cadeia, pode-se utilizar os chamados métodos de modificação de densidade (item 4.1) com intuito de melhorar a qualidade das fases. A tentativa de traçar a cadeia utilizando mapas densidade eletrônica com baixa qualidade, pode acarretar na introdução de erros no modelo os quais podem não ser fáceis de remover durante o refinamento, por isso é interessante melhorar ao máximo a qualidade destas antes de iniciar a interpretação da cadeia.

Deve-se ressaltar porém, que é importante avaliar a qualidade do mapa de densidade eletrônica antes de iniciar o refinamento ou utilizar os métodos de modificação de densidade eletrônica. A maneira de avaliar esta qualidade depende muito da experiência do usuário, porém a mais comum, é notar se há definição clara de fronteiras entre proteína e solvente.

### 4.1 - O MÉTODO DE NIVELAMENTO DE SOLVENTE (*SOLVENT FLATTENING*)

Este método, bem como o chamado *molecular averaging* (promediação da densidade eletrônica, veja Drenth, 1993 ou McRee, 1994), são métodos de modificação de densidade que têm por objetivo o melhoramento de fases, obtido pela imposição de restrições na densidade no espaço real e então usando as fases do mapa modificado para filtrar ou substituir as fases experimentais.

Será feita a seguir uma breve descrição do método de nivelamento de solvente (*solvent flattening*), desenvolvido por Wang, 1985, o qual foi utilizado nesta tese.

O método assume que a densidade eletrônica na região do solvente é aproximadamente plana (este fato pode ser comprovado em estruturas resolvidas a alta resolução e muito bem refinadas). A razão para tal fato é que as moléculas de solvente estão aleatoriamente distribuídas dentro do cristal protéico, ou seja, não têm estrutura definida, exceto as moléculas de solvente que estão localizadas internamente à proteína e as primeiras e/ou segundas camadas de hidratação da superfície desta. Assim, se a região do solvente puder ser identificada, os picos

de densidade eletrônica nesta região que serão devidos a ruídos (incorreções nas fases) poderão ser removidos.

Os passos do método são dados a seguir:

### 1. Construção do envelope de Wang:

Uma grade tridimensional (com pontos descritos simbolicamente por  $j$ ) é superposto à grade do mapa original (pontos  $i$ ) na unidade assimétrica. Para cada ponto da grade do mapa original ( $i$ ) sua densidade eletrônica ( $\rho_i$ ) é substituída por um novo valor de densidade eletrônica ( $\rho_j$ ) em uma esfera de raio  $R$  (tipicamente 10 Å), segundo o peso:

$$\rho_j = K \sum_i^R w_i \rho_i \quad \begin{array}{l} w_i = 1 - r_{ij} \quad (r_{ij} < R \text{ ou } \rho_i > 0) \\ w_i = 0 \quad (r_{ij} > R \text{ ou } \rho_i < 0) \end{array} \quad (4.1)$$

onde  $r_{ij}$  é a distância entre os pontos das grades (*grid points*),  $w_i$  é a função peso e  $K$  é uma constante.

A aplicação deste peso é baseada no fato que exceto em casos desfavoráveis a densidade eletrônica para a região da proteína é muito maior do que na região do solvente e devido a que os picos existente na região do solvente estão distribuídos aleatoriamente e se cancelarão fazendo a somatória em toda região da esfera  $R$ .

Desta maneira, o resultado do método é a apresentação de uma região mais contínua formando uma espécie de envelope em volta da proteína, revelando assim os contornos desta.

### 2. Aplicação do filtro no espaço real (nivelamento de solvente).

A densidade eletrônica na região do solvente é substituída por:

$$\text{densidade na região do solvente} = \langle \rho_{sol} \rangle + \rho_c$$

onde  $\langle \rho_{sol} \rangle$  é a densidade média esperada do solvente e  $\rho_c$  é uma constante. Na região da proteína a sua densidade é somada da constante  $\rho_c$  e qualquer densidade negativa nesta região é considerada um erro e transformada para zero.

3. Com este novo mapa de densidades eletrônicas calcula-se seus fatores de estrutura e fases por uma transformada de Fourier inversa.

4. Combinação das fases (originais e novas após nivelamento de solvente), chamada de filtro no espaço recíproco, feita utilizando-se a distribuição de probabilidade equivalente ao procedimento de peso de Sim (Sim, 1959), dada abaixo.

$$P_{SF}(\alpha) = N \exp \frac{|F_{obs}| |F_{calc}|}{|F_{obs}^2 - F_{calc}^2|} \cos(\alpha_p - \alpha_{calc}) \quad (4.2)$$

onde  $N$  é uma constante de normalização,  $\alpha_p$  e  $\alpha_{calc}$  são as fases da proteína antes e depois do processo de nivelamento de solvente respectivamente. O procedimento de peso de Sim é utilizado para ‘pesar’ uma determinada parte do modelo que é conhecida em relação a outra parte desconhecida (ou que não se tem certeza da sua veracidade). No caso do nivelamento de solvente, a parte conhecida é a parte do solvente que deve ser plana. Esta função probabilidade (equação 4.2) é usada também, no chamado ‘mapa de omissão’ (OMIT map) onde se deseja confirmar parte da densidade eletrônica de um modelo, e desta maneira, calcula-se o mapa a partir somente dos fatores de estrutura e fases da parte conhecida.

Da equação 4.2 pode-se calcular a ‘melhor’ fase e figura de mérito que pode ser combinada com a probabilidade de fases originais ( $P_{orig}$ ) produzindo finalmente uma nova probabilidade de fases ( $P_{comb}$ ) que darão as fases finais do processo, pela equação:

$$P_{comb}(\alpha) = P_{SF}(\alpha)P_{orig}(\alpha) \quad (4.3)$$

Obtém-se desta maneira as novas fases; o processo deve ser então novamente repetido até que se obtenha uma convergência para os valores de fase. É útil combinar as funções probabilidades relativas às fases originais (MIR) e após o nivelamento de solvente, pois é evitado que erros sejam introduzidos no modelo, isto é, é feita uma filtragem de possíveis erros ocorridos durante o processo (p.ex. inclusão de alguma parte relativa à proteína na porção relativa ao solvente). Além disso, com os vários ciclos que são feitos no processo de nivelamento de solvente, obtém-se uma convergência nas fases onde o peso maior na equação 4.3, será devido a  $P_{SF}$ .

Deve-se levar em conta que normalmente neste processo é necessário considerar que o cristal contém menos solvente do que o calculado (nos primeiros ciclos, 10-15% menos e nos últimos ciclos até 5% menos) para que não haja o perigo do envelope gerado para proteína cortar partes destas, já que o método é apenas uma aproximação.

Em 1987, Leslie publicou um aperfeiçoamento desta técnica, no qual usa basicamente o mesmo processo, porém minimiza o tempo computacional gasto.

Na figura 4.1 há um fluxograma mostrando como é o método apresentado.

Deve-se ressaltar que se obtém melhores resultados com o método de nivelamento de solvente no caso de cristais que apresentam grande quantidade de solvente. Isto ocorre, porque nestes casos, haverá um número maior de pontos da densidade eletrônica que podem ser modificados contribuindo para os fatores de estrutura recalculados, melhorando desta maneira, o processo de nivelamento de solvente.

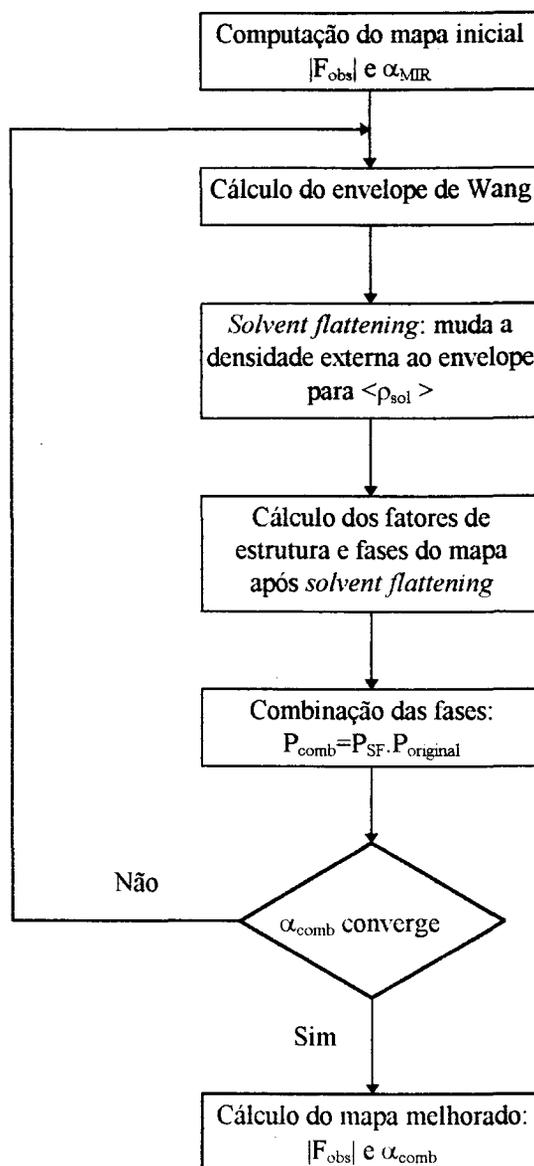


Figura 4.1 - Fluxograma do método de nivelamento de solvente (*solvent flattening*)

As figuras 4.2a e 4.2b mostram duas seções do mapa densidade eletrônica ( $Z=20$  e  $33$ ) da Glucosamina 6-fosfato desaminase com fases MIR até  $3.0 \text{ \AA}$  de resolução, indicando o envelope molecular utilizando para o método de nivelamento de solvente. Este envelope mostra fronteiras claras entre a região da proteína e a região de solvente na cela unitária.

Na figura 4.3 é mostrada uma comparação do mapa densidade eletrônica com a cadeia principal traçada de uma região da GlcN6P desaminase, antes (à esquerda) e após (à direita) da utilização do método de nivelamento de solvente.

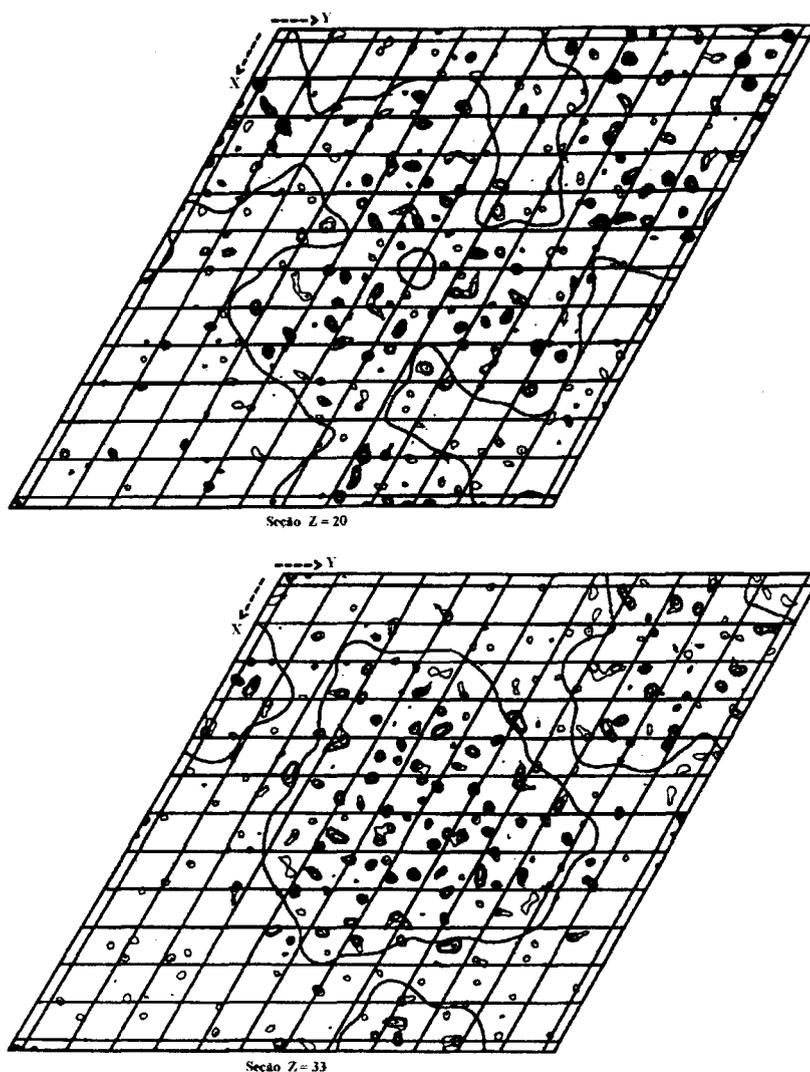


Figura 4.2a, 4.2b - Seções ( $Z=20$  e  $33$ ) do mapa densidade eletrônica (com fases MIR até  $3.0 \text{ \AA}$ ), mostrando o envelope molecular utilizado para o método de nivelamento de solvente.

O método de nivelamento de solvente foi utilizado no caso da GlcN6P desaminase para remoção da ambigüidade de fases SIR entre  $3.0$  e  $2.5 \text{ \AA}$ , faixa de resolução obtida somente pelo derivado de Pt. As fases nesta faixa de resolução foram adicionadas gradativamente à medida em que os ciclos de refinamento de fases do método de nivelamento de solvente eram feitos.

Do ponto de vista qualitativo houve uma melhora muito grande na figura de mérito média; de  $0.51$  antes do nivelamento de solvente para  $0.90$ . A figura de mérito tem como variável o valor de  $\epsilon$ , conhecido por *lack of closure* (equação 3.6) que por sua vez, é dependente de  $|F_{PH}|^{obs}$  e  $|F_{PH}|^{calc}$ , usada para medir a qualidade de  $F_H$  obtida após o refinamento desta. Assim, normalmente, pelo método de nivelamento de solvente tem-se valores muito próximos de  $F_P^{calc}$  e  $F_P^{obs}$ , praticamente fechando o triângulo vetorial  $F_{PH}=F_P+F_H$  (figura 3.4), obtendo-se valores muito bons para figura de mérito.

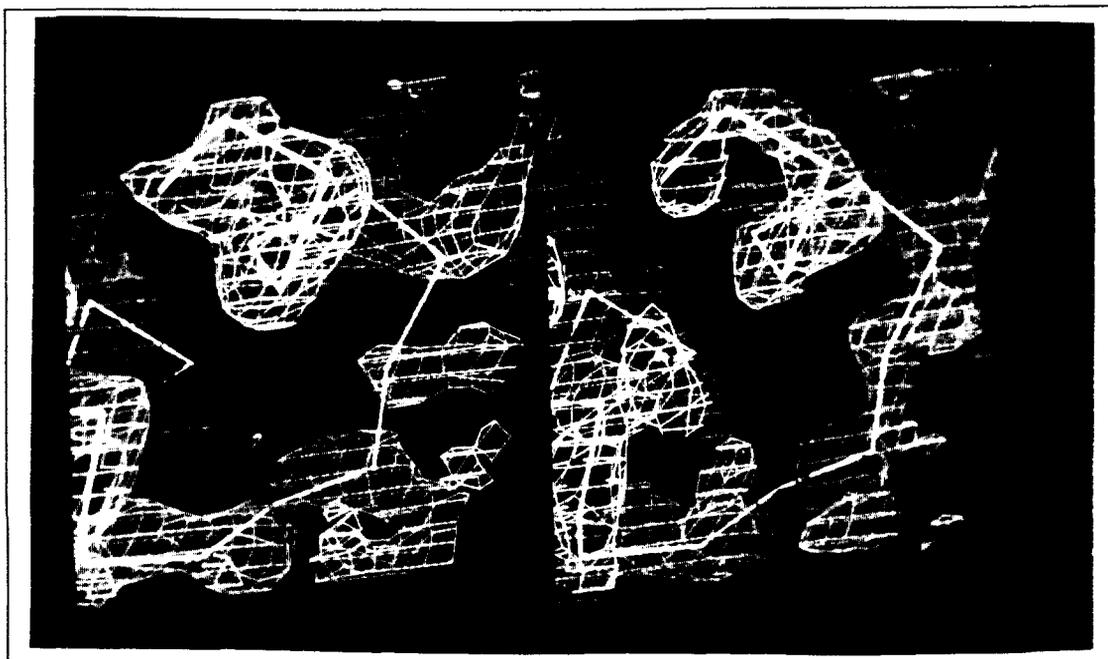


Figura 4.3 - Comparação do mapa densidade eletrônica com a cadeia principal traçada de uma região da GlcN6P desaminase, antes (à esquerda) e após (à direita) a utilização do método de nivelamento de solvente.

## 4.2 - INÍCIO DA CONSTRUÇÃO DO MODELO DA GlcN6P DESAMINASE

Do mapa densidade eletrônica com fases obtidas após os métodos de modificação de densidade eletrônica, pode-se normalmente interpretar pelo menos a maior parte da cadeia polipeptídica (cadeia principal), principalmente as estruturas secundárias. Para traçar a cadeia principal pode-se utilizar a rotina BONES do programa gráfico 'O' (Jones, *et al.*, 1991). A rotina BONES associa as maiores densidades eletrônicas a átomos de  $C_{\alpha}$  e tenta dar conectividade a estes por meio de uma esfera de raio equivalente a ligação peptídica convencional, a partir de um dado  $C_{\alpha}$  tentando encontrar outro  $C_{\alpha}$  nesta faixa. Se isto ocorre, os átomos são conectados formando assim um esqueleto, ou seja, uma cadeia principal mais provável.

O mapa densidade eletrônica inicial da Glucosamina 6-fosfato desaminase foi construído utilizando as fases MIR obtidas até 3.0 Å e as fases SIR, do derivado  $K_2PtCl_4$  (3.0-2.5 Å) após o refinamento destas pelo método de nivelamento de solvente. Deste mapa pode-se visualizar claramente os elementos de estruturas secundárias e foi possível traçar a cadeia polipeptídica utilizando a rotina BONES do programa 'O'. Deste mapa foi possível traçar 230 dos 266 aminoácidos para ambos os monômeros da unidade assimétrica. Os resíduos não encontrados eram do *loop* de resíduos 78-85, o segmento da cadeia de resíduos 161 a 182 e oito aminoácidos do C-terminal da molécula. Após cuidadosa inspeção do envelope molecular automático gerado pelo método de nivelamento

de solvente, conclui-se que este estava ‘cortando’ parte da molécula, justamente o *loop* 78-85 e os resíduos 161-182. Desta maneira, usando-se sucessivos Fourier diferença, foi possível traçar toda a cadeia principal exceto os oito resíduos do C-terminal que parecem estar bastante desordenados; estes resíduos somente puderam ser encontrados nos estágios finais do refinamento.

É útil salientar que, apesar de não termos feito promediação de densidade eletrônica utilizando os dois monômeros relacionados por simetria não cristalográfica, esta informação foi utilizada na construção do modelo. Durante a modelagem de um dos monômeros foram superpostas as densidades eletrônicas referentes aos dois monômeros. Desta maneira, quando haviam dúvidas na construção de um monômero utilizando uma das densidades, estas eram normalmente solucionadas, com a superposição da densidade equivalente do outro monômero.

Para a construção do modelo foi utilizado o programa gráfico ‘O’ (Jones *et al.*, 1991).

O próximo passo é iniciar o refinamento do modelo.

### 4.3 - REFINAMENTO DO MODELO (Drenth, 1993; Brünger, 1989, 1992 e 1993)

Para melhorar um modelo deve-se ‘refiná-lo’, ou seja, realizar um processo de ajuste do modelo afim de encontrar uma melhor concordância entre os fatores de estrutura observados e calculados (medido pelo fator de discordância R, definido na equação 4.4), levando em conta as características estereoquímicas do modelo.

$$R = \frac{\sum_{hkl} ||F_{obs} - k|F_{calc}||}{\sum_{hkl} |F_{obs}|} \quad (4.4)$$

onde  $k$  é um fator de escala (veja equação 4.6).

No início do refinamento o fator R costuma ter valores altos, de 40 a 50% (no caso de estruturas aleatórias o valor é de 59% (Wilson, 1959)), alcançando no final do processo valores típicos entre 15 e 20% para refinamentos bem sucedidos.

O ajuste do modelo consiste em mudanças nos parâmetros posicionais e de temperatura para todos os átomos da estrutura (exceto hidrogênios). Porém, para proteínas há muitos parâmetros a serem refinados (de 5000-50000 para proteínas médias) se comparados ao número de medidas experimentais tem-se um conjunto subdeterminado (a GlcN6P desaminase, coletada até 2.1 Å, têm aproximadamente 39169 reflexões independentes usadas no refinamento para 24736 parâmetros, ou seja, 1.6 observações por parâmetro).

Em contrapartida tem-se várias informações que podem ser adicionadas ao refinamento, tais como, informações estereoquímicas comuns às proteínas, incluindo distâncias e ângulos de ligações químicas, ângulos diédricos, contatos de Van der Waals, pontes de hidrogênio etc. Há duas maneiras de utilizar estas informações:

1. *Constraint* (construção, confinamento) é o caso onde se permite variar em uma faixa muito pequena somente alguns parâmetros. Por exemplo, tomando distâncias e ângulos de ligações químicas como rígidos e permitindo somente os ângulos diédricos variar. Neste caso reduz-se muito os parâmetros a serem refinados, porém é muito difícil mover partes pequenas da estrutura.
2. *Restraint* (restrição) é o caso onde os parâmetros estereoquímicos são permitidos variar dentro de uma faixa, controlados por um termo de energia. Assim, pode-se mover pequenas partes do modelo porém é difícil mover grandes partes da proteína, como uma molécula inteira ou domínio.

Os procedimentos computacionais de refinamento têm evoluído bastante nos últimos anos. No início, o refinamento se restringia ao ajuste dos fatores de estrutura observados e calculados pelo método dos mínimos quadrados, método ainda utilizado para pequenas moléculas, as quais possuem alta relação observações/parâmetros (Drenth, 1993, p.244). Hoje porém, inclui-se termos de energia fazendo uma minimização desta durante o refinamento e, mais recentemente, utiliza-se também a dinâmica molecular. Estes dois últimos procedimentos serão descritos abaixo.

Nesta tese utilizou-se o programa X-PLOR para refinamento de estruturas (veja Brünger, 1993). Este programa é bastante versátil, adequando-se a praticamente todos os problemas encontrados durante o refinamento de estruturas macromoleculares, usando dados de difração de raios-X ou ressonância magnética nuclear.

A figura 4.4 representa um esquema utilizado no refinamento.

Seguindo o fluxograma da figura 4.4, tem-se que no caso desta tese foi feito o seguinte processo:

1. De posse de um modelo inicial gera-se a informação molecular (rotina *topology*)
2. Faz-se um refinamento com minimização de energia de corpo rígido.
3. Rotina *check* para obter o peso  $W_A$  (veja item 4.3.2)
4. Refinamento por dinâmica molecular com temperatura controlada (*simulated annealing*).
5. Refinamento posicional com minimização de energia (*positional refinement*).
6. Refinamento dos fatores de temperatura (fatores B) em grupo.
7. Refinamento dos fatores de temperatura individuais.
8. Geração do novo mapa de densidade eletrônica. Faz-se modificações no modelo usando-se o programa gráfico 'O' tentando adequar da melhor maneira possível a cadeia principal e as cadeias laterais ao mapa. Nesta fase também.

adicionou-se os oito resíduos remanescentes em cada monômero e incluiu-se os íons de fosfato presente no modelo (quatro no dímero) através de um mapa Fourier diferença.

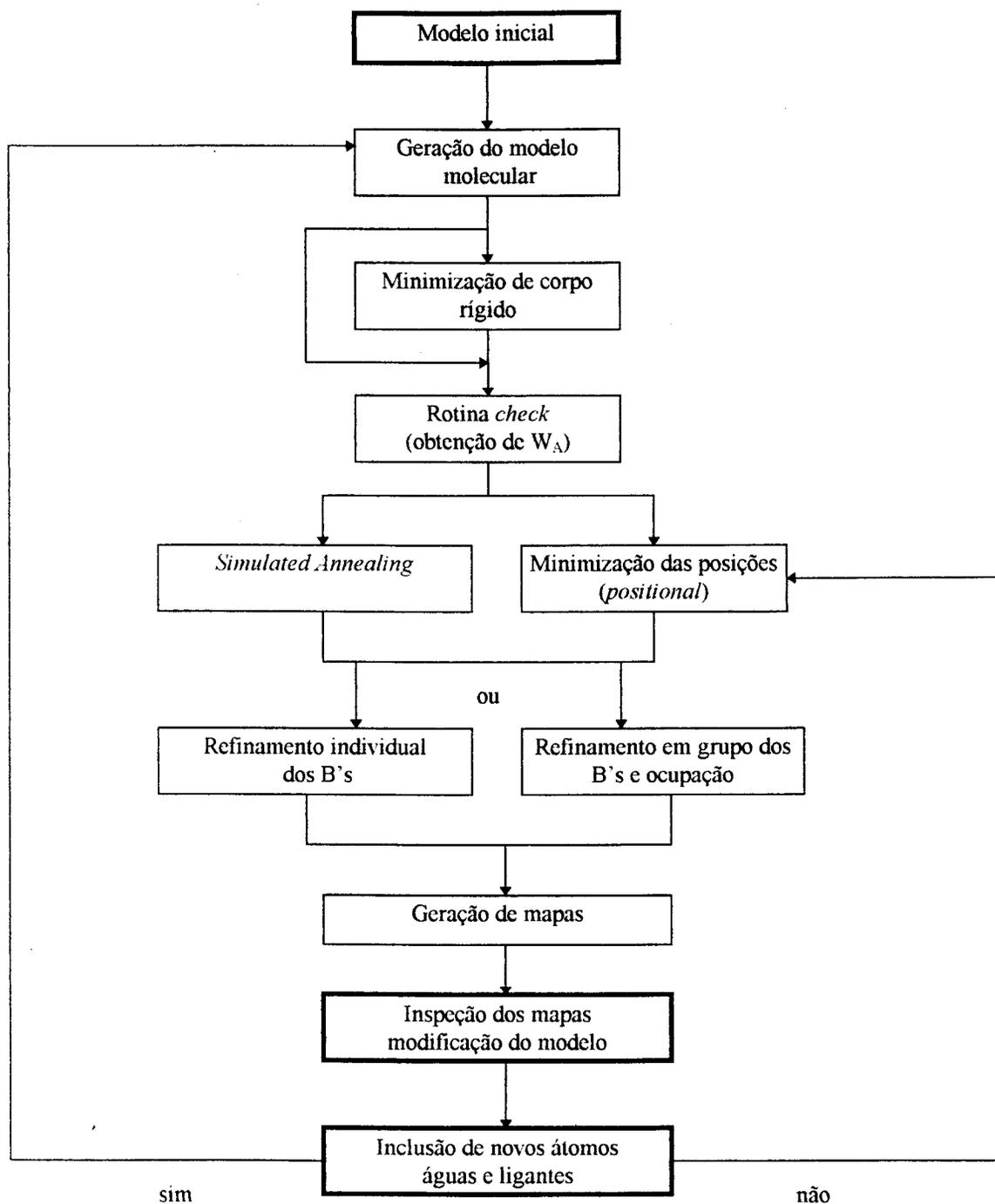


Figura 4.4 - Fluxograma para refinamento do modelo  
(quadros grifados são as etapas onde não se utiliza o programa X-PLOR)

9. Nova geração da informação molecular (rotina *topology*), incluindo os novos componentes do modelo.
10. Rotina *check* para obter o peso  $W_A$ .
11. Novo refinamento por dinâmica molecular com temperatura controlada (*simulated annealing*).
12. Refinamento com minimização das posições (*positional refinement*) (fazendo antes o *check*).
13. Refinamento dos fatores de temperatura individuais.
14. Geração do novo mapa, tentando adequar da melhor maneira possível o modelo ao mapa.
15. Repetição de 9-14 até obter os melhores resultados possíveis (o item 9 somente deve ser feito de for incluído novos átomos).
16. Inclusão de moléculas de água estruturadas ao modelo por meio de um mapa Fourier diferença, levando em conta a estereoquímica desta. Colocação destas no arquivo de coordenadas da proteína (arquivo PDB).
17. Novo processo de 9-14.
18. Inclusão de outras moléculas de água por análise do mapa Fourier diferença.
19. Novo processo de 9-14.

A seguir, serão dados mais detalhes sobre as características do processo de refinamento.

#### 4.3.1 - Geração da estrutura molecular: Topologia e Parâmetros (Brünger, 1992).

O bloco topologia do programa X-PLOR consiste de informações sobre os 'resíduos' de uma macromolécula, ou sobre os átomos que constituem substratos, cofatores e moléculas de água. Dentre estas informações têm-se: massa e carga do átomo, distância, ângulos e ângulos diédricos entre estes, quais são os átomos possíveis de doar ou receber elétrons e quais são os ângulos diédricos impróprios (restrições dadas as coordenadas com finalidade manter a planaridade dos anéis ou quiralidade) etc.

O bloco de parâmetros especifica as várias constantes para os termos de energia conformacional e não ligada, além de várias outras aplicações (veja Brünger, 1993) que serão utilizados durante o processo de refinamento.

O programa conta com arquivos que podem ser utilizados para a maioria dos tipos de proteína, para moléculas de água etc; porém no caso da presença de algum íon (no caso desta tese há quatro íons de fosfato no dímero na unidade assimétrica) ou cofator (no capítulo 5 será descrita a determinação da estrutura ligada com ativador e inibidor desta) os arquivos de topologia e parâmetros destes devem ser construídos pelo usuário tomando como base arquivos modelo

existentes no programa e conhecimentos prévios em termos estruturais dos íons, cofatores ou substratos a serem adicionados à estrutura (veja capítulo 5).

A rotina *topology* cria um arquivo com as coordenadas atômicas com formato PDB (*protein data bank*) que inclui os devidos átomos de hidrogênio que podem estar envolvidos em ligações de hidrogênio e cria um arquivo do tipo PSF (*protein structure file*) que inclui as informações estruturais e parâmetros a serem utilizados.

#### 4.3.2- Minimização de energia

Este é um refinamento por mínimos quadrados com restrições de estereoquímica. A função a ser minimizada é a energia potencial, na qual inclui-se termos de energia estereoquímicos e um termo cristalográfico (semelhante a função usada em um refinamento por mínimos quadrados convencional).

O programa X-PLOR faz uma minimização de energia que inclui um termo empírico (termos estereoquímicos de energia) e um termo efetivo que inclui o termo cristalográfico:

$$E_{\text{TOTAL}} = E_{\text{EMPÍRICO}} + E_{\text{EFETIVO}} \quad (4.5)$$

Os termos da energia potencial empírica incluem os potenciais devidos a (as equações que representam estes termos estão entre parênteses):

- ligações covalentes  $(E_{\text{LIGAÇÕES}} = \sum_{\text{ligações}} k_b(r-r_0)^2)$

- ângulos  $(E_{\text{ANGL}} = \sum_{\text{ângulos}} k_\theta(\theta-\theta_0)^2)$

- ângulos diédricos  $(E_{\text{DIED}} = \sum_{\text{diédricos}} k_\phi(\cos(n\phi+\delta))$

- ângulos impróprios  $(E_{\text{IMPR}} = \sum_{\text{quirais, planar}} k_\omega(\omega-\omega_0)^2)$

- contatos de van der Waals e eletrostáticos

$$(E_{\text{não ligados}} = \sum_{\text{pares de átomos}} (ar^{-12} + br^{-6} + cr^{-1}))$$

- contatos de van der Waals e eletrostáticos entre átomos relacionados por simetria.

(onde os  $k$ 's,  $\theta_0$ ,  $\phi_0$ ,  $\omega_0$ ,  $a$ ,  $b$  e  $c$  são constantes e  $\theta$ ,  $\phi$ ,  $\omega$ ,  $r$  são as variáveis dos potenciais (veja Brünger, 1989 ou 1993))

Os termos de energia potencial efetiva são compostos de um termo cristalográfico e termos de constrição de energia (*constraints*) (início do item 4.3 e Brünger, 1993). O termo cristalográfico usual é:

$$E_{\text{XREF}} = (W_A / N_A) \sum_h W_h [ |F_{\text{obs}}(\mathbf{h})| - k |F_{\text{calc}}(\mathbf{h})| ]^2 \quad (4.6)$$

onde  $N_A$  é um fator de normalização,  $W_h$  é uma função peso individual para cada reflexão (normalmente dependente de  $[\text{sen}\theta/\lambda]^n$ , no início do refinamento  $n=-1$  e no final  $n=1$ ),  $k$  é um fator de escala entre os fatores de estrutura observados e

calculados (comumente  $k = \frac{\sum_h |F_{obs}| |F_{calc}|}{\sum_h [|F_{calc}|]^2}$ ) e  $W_A$  é um peso adicional que relaciona

o termo cristalográfico com o termo empírico. Este peso é obtido através da rotina *check* do programa X-PLOR que faz uma breve dinâmica molecular (item 4.3.3) sem incluir o termo  $E_{XREF}$ . Normalmente, é utilizada uma porcentagem do peso  $W_A$  recomendado pela rotina *check*, isto é, dá-se um peso maior aos termos de energia empírico (termos estereoquímicos de energia). O peso para o termo  $E_{XREF}$  deve ser maior, quanto maior a relação observações/parâmetros refinados. Valores usualmente utilizados para  $W_A$  estão entre 30-80% do valor obtido no programa *check*.

#### 4.3.2.1 - Minimização de corpo rígido

Neste tipo de refinamento uma geometria rígida é determinada para partes da estrutura (um monômero por exemplo), e os parâmetros de orientação e posição destas partes são refinados, ao invés dos parâmetros atômicos individuais. Desta maneira, as mudanças nas posições para os grupos são calculadas pela combinação das mudanças para todos os átomos. Para cada grupo é possível refinar até seis parâmetros, três rotacionais e três translacionais.

A energia a ser minimizada é a  $E_{TOTAL}$  (equação 4.5) na qual podem ser incluídos ou excluídos termos.

Normalmente o refinamento de corpo rígido é o primeiro passo no procedimento de refinamento do modelo e é feito frequentemente com dados à baixa resolução, pois se são usados dados a muito alta resolução é comum que os grupos caiam em mínimos locais.

#### 4.3.2.2 - Minimização de posição (*positional refinement*)

O programa X-PLOR usa o método desenvolvido por Powell (1977) que utiliza a técnica de gradiente conjugado para fazer o refinamento das coordenadas dos átomos. A função de energia total  $E_{TOTAL}$  (equação 4.5) é utilizada.

#### 4.3.3 - Dinâmica molecular com temperatura controlada (*simulated annealing*)

O refinamento convencional por mínimos quadrados com restrições de estereoquímica e empacotamento das moléculas (na cela unitária) (item 4.3.2)

era, até o final da década de 80, a única maneira possível de refinar um modelo macromolecular. Porém, este procedimento muitas vezes não consegue alcançar o mínimo global, atingindo somente mínimos locais e desta maneira, necessitando de constantes ajustes manuais no modelo e assim pode-se não obter o resultado almejado, compatível com os dados experimentais disponíveis.

O método *simulated annealing* (têmpera simulada, em português), o qual será chamado nesta tese de dinâmica molecular com temperatura controlada, começou a ser utilizado para cristalografia de macromoléculas a partir de 1987 (Brünger, 1987) (já era utilizado desde o início da década de 80 em outras áreas). O nome 'têmpera' (*annealing*) vem do processo metalúrgico de aquecimento de metais até altas temperaturas e resfriamento até temperatura ambiente melhorando assim as suas características de condução e tornando-os mais resistentes a impactos.

Utilizando basicamente o mesmo mecanismo, o processo de *simulated annealing* faz uma simulação matemática que tem como alvo encontrar o mínimo global de uma função  $f(X) = f(x_1, x_2, \dots, x_n)$ , ( $x_1, x_2, \dots, x_n$  são parâmetros tais com posições atômicas, fatores de temperatura e ocupação) onde se tenta mover estes parâmetros de maneira que se possa escapar de mínimos locais, sendo possível aumentar  $f(X)$ , bem como diminuí-lo. Um aumento em  $f(X)$  é dado pela probabilidade  $\exp[-\Delta f(X)/kT]$  onde  $\Delta f(X)$  é a variação desta função antes e depois do aumento,  $T$  é a 'temperatura' do sistema e  $k$  é constante de Boltzmann. Na realidade  $T$  não deve ser entendido como uma temperatura física mas sim como um parâmetro de controle para que o sistema escape de mínimos locais. Assim, deve-se utilizar valores de  $T$  bastante elevados para superar as barreiras, seguido por uma lenta redução desta (resfriamento). Este processo pode ser repetido várias vezes até alcançar-se o mínimo global.

Este processo pode ser simulado com sucesso para cristalografia de macromoléculas utilizando-se o método da dinâmica molecular.

A dinâmica molecular resolve a equação de Newton de movimento  $F_i = m_i \delta^2 r_i / \delta t^2$ , onde  $F_i$  é a força e  $m_i$  a massa agindo sobre o átomo  $i$ . A força sobre cada átomo é proporcional a energia potencial agindo sobre este:  $F_i = -\delta E_{\text{pot}} / \delta r_i$ , ou seja:

$$m_i \delta^2 r_i / \delta t^2 = -\delta E_{\text{pot}} / \delta r_i \quad (4.7)$$

Em princípio então, com uma expressão adequada para a energia potencial para os átomos (no caso do X-PLOR é a energia total, dada pela equação 4.5) e conhecendo-se suas massas, será possível resolver a equação diferencial para futuras posições no tempo (a trajetória dos átomos). Porém, a mecânica clássica permite resolver exatamente sistemas com até duas partículas, ou seja, para sistemas com três ou mais partículas a solução da equação diferencial 4.7 não será exata, e pode ser representada pela seguinte série de Taylor, onde  $r(t)$  é a posição da partícula em um intervalo de tempo  $\Delta t$ :

$$\bar{r}(t + \Delta t) = \bar{r}(t) + \frac{\delta \bar{r}}{\delta t} \Delta t + \frac{\delta^2 \bar{r}}{\delta t^2} \frac{\Delta t^2}{2} + \dots \quad (4.8)$$

Desta maneira, a solução da equação de movimento depende do conhecimento da posição  $r(t)$ , da velocidade  $\delta r/\delta t$  e da aceleração  $\delta^2 r/\delta t^2$  e convenientes aproximações de termos de ordens mais altas. Pode-se obter coordenadas de uma minimização prévia ou da estrutura cristalina, velocidades são atribuídas aleatoriamente obedecendo uma distribuição de Maxwell-Boltzmann que correspondem a temperatura de interesse e a aceleração da equação 4.7; assim obtém-se  $r(t+\Delta t)$ . Então, as coordenadas originais são substituídas pelas novas coordenadas, as velocidades são calculadas pela subtração das coordenadas anteriores menos as novas e divididas pelo tempo do passo e a aceleração é corrigida computando-se os gradientes com as novas coordenadas. Um ciclo é completado e o processo é repetido milhares e talvez milhões de vezes.

Para realizar esta tarefa, o X-PLOR utiliza como base o algoritmo de Verlet (1967), que é dado abaixo.

Se  $v_{med}$  é a velocidade média durante um intervalo de tempo entre  $t$  e  $t+\Delta t$ , então a posição no final do intervalo é:

$$r(t+\Delta t) = r(t) + v_{med}\Delta t \quad (4.9)$$

Se for assumido que a velocidade varia linearmente no intervalo de tempo  $\Delta t$ , tem-se:

$$v_{med} = v(t+\Delta t/2) \quad (4.10)$$

Porém, a velocidade instantânea pode ser calculada da aceleração média durante o intervalo de tempo  $(t-\Delta t/2)$  e  $(t+\Delta t/2)$ :

$$v(t+\Delta t/2) = v(t-\Delta t/2) + a_{med} \Delta t \quad (4.11)$$

Se for assumido agora, que a aceleração varia linearmente entre  $(t-\Delta t/2)$  e  $(t+\Delta t/2)$ , então a aceleração média vale:

$$a_{med} = a(t) \quad (4.12)$$

Assim, substituindo 4.12 em 4.11 e 4.10 em 4.9 obtém-se as equações:

$$v(t+\Delta t/2) = v(t-\Delta t/2) + a(t)\Delta t \quad (4.13)$$

$$\mathbf{r}(t+\Delta t/2) = \mathbf{r}(t) + \mathbf{v}(t+\Delta t/2)\Delta t \quad (4.14)$$

Um ciclo para computar novas coordenadas, é dado pelos seguintes passos:

1. Calcula-se a aceleração  $(-1/m)dE_{\text{pot}}/dr$  no tempo  $t$ .
2. A partir de 4.13 obtém-se a velocidade no tempo  $(t+\Delta t/2)$  do seu valor em  $(t-\Delta t/2)$ .
3. Obtém-se as coordenadas em  $(t+\Delta t)$  a partir de 4.14

Um parâmetro chave no algoritmo de Verlet é o intervalo de tempo  $\Delta t$ . Se  $\Delta t$  tiver um valor grande o processo necessitará de menor tempo de uso de computação, porém há limitações quanto a grandes intervalos de tempo. A principal limitação é a mais alta frequência de movimento que deve ser simulado. Vibrações mais rápidas implicam em mudanças mais rápidas de velocidades e acelerações.

O período vibracional deve ser dividido por fatores de cinco a dez, para satisfazer as aproximações utilizadas no método de Verlet que assume velocidades e acelerações variando linearmente no tempo de integração. Supondo-se que a vibração mais rápida é uma ligação C-H, cujo o período é da ordem de  $10^{-14}$  segundos, assim o intervalo de tempo de integração deve ser da ordem de  $10^{-15}$  segundos.

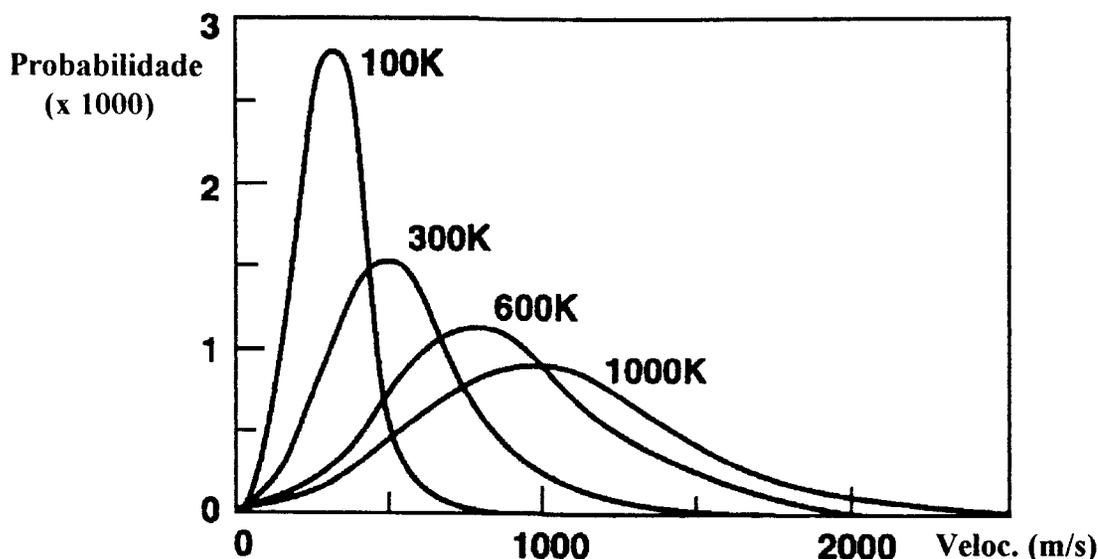


Figura 4.5 - Distribuição de velocidades de Maxwell-Boltzmann para água em várias temperaturas (Discover, 1993, p. 1/2-63)

O conceito de temperatura na simulação de dinâmica molecular é fundamental. Sabe-se que a temperatura é proporcional à energia cinética do

sistema, a qual pode ser expressa em termos das velocidades atômicas. A relação microscópica fundamental entre temperatura e velocidades moleculares é a equação de Maxwell-Boltzmann:

$$f(\mathbf{v})d\mathbf{v} = \left(\frac{m}{2\pi kT}\right)^{3/2} \exp(-m\mathbf{v}^2 / 2kT)4\pi v^2 d\mathbf{v} \quad (4.15)$$

Esta fórmula expressa a probabilidade  $f(\mathbf{v})$  que uma molécula de massa  $m$  tem de possuir uma velocidade  $\mathbf{v}$  quando sua temperatura é  $T$ . A figura 4.5 representa esta distribuição para uma molécula de água a várias temperaturas. Note que, quanto maior a temperatura maior a largura da curva de probabilidades, isto é, maior é a probabilidade que as moléculas tenham as mais diversas velocidades.

O controle de temperatura durante o cálculo da dinâmica molecular é obtido por reescalamento periódico e uniforme de velocidades  $V_i$ , ou seja:

$$V_i^{novo} = escala \times V_i^{antigo} \quad (4.16)$$

para todos os átomos  $i$ . O fator de escala é dado por:

$$escala = \frac{\left\langle \sum_i m V_i^{antigo}(t)^2 \right\rangle}{nkT} \quad (4.17)$$

O parâmetro  $n$  é o número de graus de liberdade do sistema,  $k$  é a constante de Boltzmann,  $T$  é a temperatura,  $V_i^{antigo}$  é a velocidade do átomo  $i$  no tempo  $t$  e  $\langle \rangle$  denota a trajetória média no intervalo de tempo entre o escalonamento de velocidades.

Nas versões novas do X-PLOR há outras maneiras de controle de temperatura (veja Brünger, 1993).

A grande vantagem do refinamento por *simulated annealing* é o raio de convergência que é significativamente maior do que no refinamento convencional (mínimos quadrados com restrições) e pode reduzir a necessidade de correções manuais.

#### 4.3.4 - Refinamento dos fatores de temperatura (fatores B's)

O refinamento dos fatores de temperatura isotrópicos (um parâmetro a ser refinado por átomo) ou anisotrópicos (seis parâmetros por átomo) podem ser feitos de duas maneiras pelo programa X-PLOR:

- Refinamento individual de todos os B's.

#### - Refinamento em grupo dos B's.

Este refinamento é feito usando a técnica de gradiente de minimização conjugado (item 4.3.2) utilizando como energia potencial somente o termo  $E_{XREF}$  (equação 4.6) porém substituindo o  $F_{calc}(\mathbf{h})$  por  $exp(-S^2B/4)F_{calc}(\mathbf{h})$  (onde  $S=2 \sin\theta/\lambda$  e B é o fator de temperatura) para o caso isotrópico (utilizado nesta tese).

O refinamento dos fatores de temperatura é feito após o de coordenadas. Na rotina de refinamento de B's é possível também refinar a ocupação dos átomos, se necessário.

#### 4.3.5 - Índice de discordância livre, $R_{free}$

Durante o processo de refinamento de um modelo um parâmetro importante utilizado para o acompanhamento da evolução do refinamento é o índice de discordância ou fator R (equação 4.4).

Porém, o fator R é estritamente relacionado com o termo de energia  $E_{XREF}$  a ser refinado (veja equação 4.6). Assim, este fator pode ser feito arbitrariamente pequeno aumentando-se o número de parâmetros do modelo e subsequentemente refinando  $E_{XREF}$ .

Como os dados cristalográficos são de alguma maneira redundantes, assim uma pequena porção dos dados podem ser omitidos sem afetar seriamente o resultado.

Sendo assim, o programa divide os dados cristalográficos em dois conjuntos, um conjunto de trabalho (A), o qual será usado no processo de refinamento, e um conjunto teste (T) (com usualmente 10% dos dados retirados aleatoriamente do conjunto total de dados), o qual não será incluído na função de minimização.

O refinamento é feito da maneira descrita anteriormente (minimização de energia, dinâmica molecular, etc), porém será possível acompanhar dois fatores R, ou seja, um para o conjunto de dados A (que está sendo realmente refinado) e um para o conjunto T. O fator R para o conjunto T é chamado de  $R_{free}$ .

Como consequência do refinamento, os fatores R e  $R_{free}$  terão valores diferentes ao curso deste, porém se as fases e outras informações têm boa precisão, o valor de  $R_{free}$  deve decrescer, mas com bem menor intensidade que o de R, ficando no final com valores superiores a este.

Porém, se o refinamento, apesar de instabilidades, estiver minimizando  $E_{total}$  e o modelo for fisicamente incorreto, o valor de  $R_{free}$  aumentará durante o processo. Assim, a monitorização de  $R_{free}$  é um indicador de confiabilidade do processo de refinamento da estrutura.

#### 4.3.6 - Inclusão de moléculas de água no modelo

A colocação de moléculas de água que estão estruturadas internamente e nas camadas de hidratação que envolvem a proteína, completa a resolução desta e permite muitas vezes tirar importantes conclusões com respeito às suas características funcionais.

O número de moléculas de água a serem incluídas no modelo depende de fatores como, a resolução que a estrutura está determinada e qualidade do modelo.

O procedimento a ser adotado para encontrar moléculas de água em uma dada estrutura utiliza o método de Fourier diferença (veja capítulo 3). Para tanto foi utilizado nesta tese o pacote de programas CCP4, da seguinte maneira:

1. Criar um mapa Fourier diferença com coeficientes ( $|F_{obs}| - |F_c|$ ) e fases calculadas do modelo usando o programa FFT (item 3.2).
2. Com o programa MAPSIG calcula-se os 500 maiores picos em função da relação sinal-ruído.
3. Com o programa NEWPICKS (E. Horjales) retira-se picos relacionados por simetria e escreve-se os picos restantes em um formato de leitura PDB.
4. Utilizando o programa 'O', usando a função WATER\_PEPKPIK, verifica-se cada pico quanto ao sentido estereoquímico, comparando as distâncias e geometria em relação a proteína e com os outros picos atribuídos como águas.
5. Os picos escolhidos como moléculas de água são colocados juntamente com a proteína e os íons fosfato, em um arquivo tipo PDB que pode ser lido pelo programa X-PLOR.
6. Assim, seguindo o fluxograma da figura 4.4 (programa X-PLOR), gera-se novo arquivo de topologia para o modelo (rotina *topology*, item 4.3.1), faz-se refinamento das posições (item 4.3.2.2) e refinamento dos B's (4.3.4).
7. De posse do arquivo de coordenadas atômicas final, pode-se analisar se estas moléculas de água estão com fatores de temperatura (B's) grandes demais (usualmente acima de 50) ou pequenos demais (abaixo de 6), os quais não fazem sentido físico e devem ser retirados deste.

É aconselhável, repetir o procedimento de 1 a 7 pelo menos mais uma vez para a inclusão de novas moléculas de água.

#### 4.3.7 - Refinamento da glucosamina 6-fosfato desaminase

Para o refinamento do modelo nativo foram utilizados dados a 2.1 Å de resolução, utilizando os passos descritos no início do item 4.3. No refinamento foram incluídos os dois monômeros (2 x 266 aminoácidos) da unidade assimétrica, que foram refinados independentemente, bem como quatro íons fosfato.

Foram feitas três etapas de colocação de moléculas de água. As águas que possuíam valores de fator de vibração térmica muito elevados foram retiradas, restando no final 348 moléculas de água com fator de vibração térmico médio de  $39.16 \text{ \AA}^2$ .

Na tabela 4.1 são dados os índices de discordância ( $R$  e  $R_{\text{free}}$ ) em algumas etapas do refinamento.

Tabela 4.1 - Índices de discordância em algumas etapas do refinamento

Etapa do refinamento (final de cada rotina)	R (%)	$R_{\text{free}}$ (%)
<i>Simulation annealing</i> (sem águas)	29.7	34.8
Minimização das posições (sem águas)	28.1	34.0
Refinamento dos B's (sem águas)	23.7	28.5
Minimização das posições (com 220 moléculas de águas)	19.8	23.2
Refinamento dos B's (com 220 águas)	18.7	22.4
Refinamento final	17.4	22.4

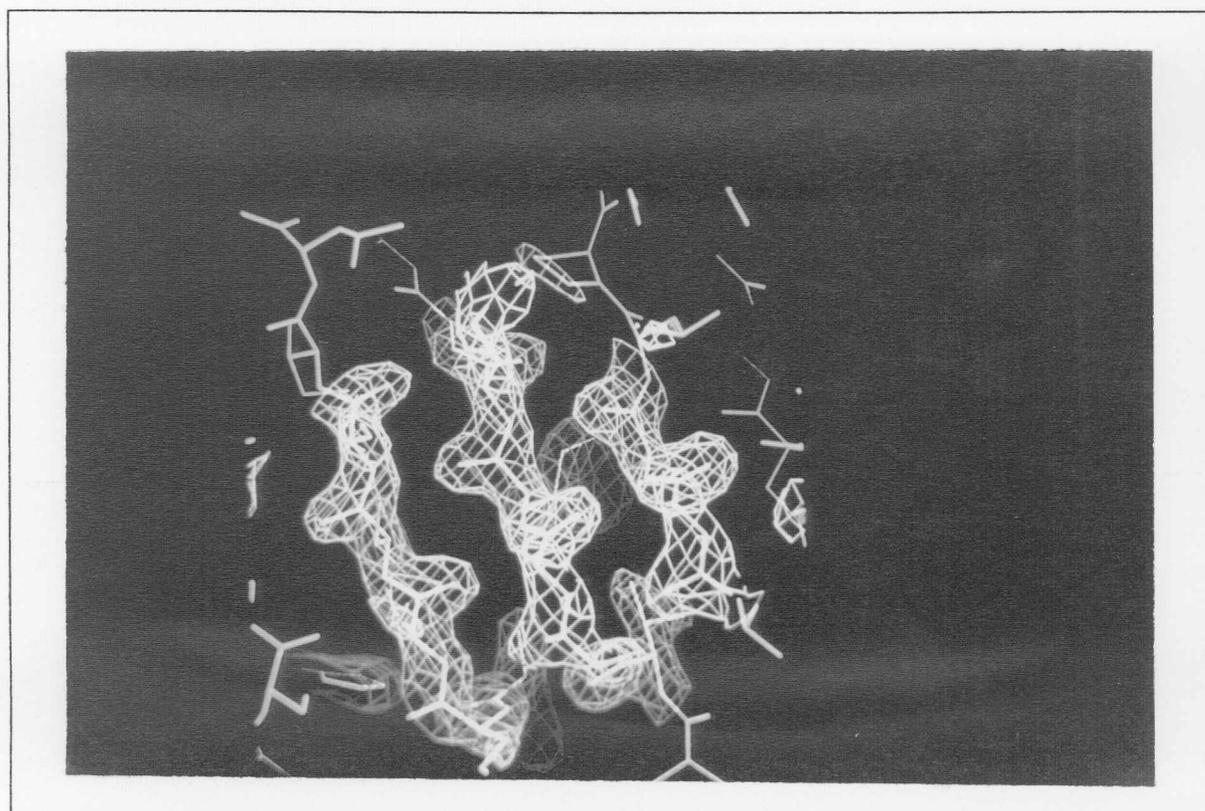


Figura 4.6 - Mapa densidade eletrônica com o modelo refinado, mostrando parte de uma folha  $\beta$  pregueada.

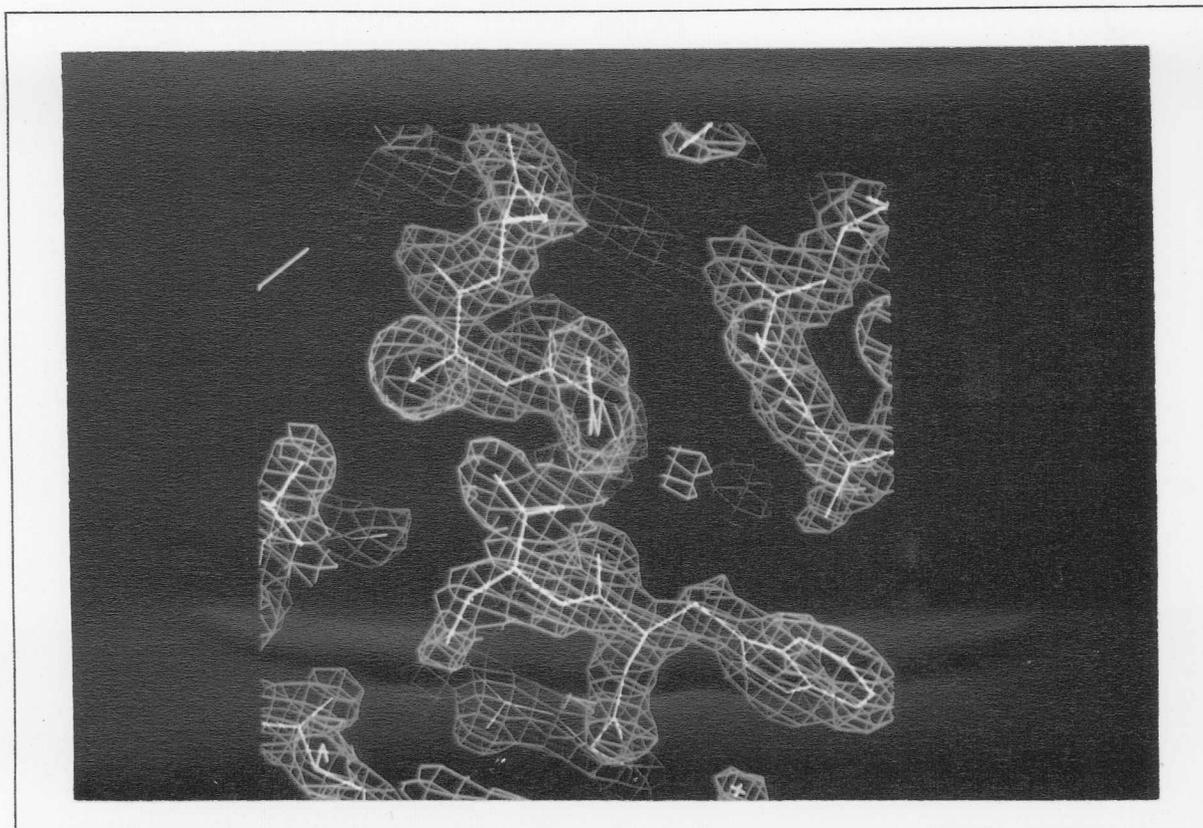


Figura 4.7 - Mapa densidade eletrônica com o modelo refinado, mostrando parte de uma hélice- $\alpha$

Foi utilizado nos refinamentos um peso de  $W_A = 200.000$  para a função  $E_{XREF}$ , aproximadamente de 30-40% o valor obtido com a rotina *check* do X-PLOR.

As figuras 4.6 e 4.7 mostram o mapa densidade eletrônica com o modelo refinado em duas regiões deste; a figura 4.6 mostra parte de uma folha  $\beta$  pregueada e na 4.7 parte de uma hélice- $\alpha$ .

Foi feito ainda, um estudo com valores de vibração térmica para os átomos do modelo, para observar o sentido físico deste (veja figuras 4.8 e 4.9). Foi observado que no C-terminal de cada um dos monômeros, os seus átomos apresentavam valores muito altos de vibração térmica. O C-terminal foi justamente a parte mais difícil de interpretar o mapa densidade eletrônica, desta maneira pode-se concluir que neste trecho da proteína havia grande desordem.

Na figura 4.8 pode-se ver um gráfico que representa o fator de vibração térmica com relação aos átomos  $C_\alpha$  dos aminoácidos dos dois monômeros; e na figura 4.9 o valor médio da vibração térmica para os átomos das cadeias laterais dos monômeros. O valor médio dos B's para os átomos  $C_\alpha$  é de  $19.64 \text{ \AA}^2$  e para os átomos das cadeias laterais é  $23.81 \text{ \AA}^2$ .

No capítulo 6 analisa-se a qualidade do modelo e descreve-se a topologia deste.

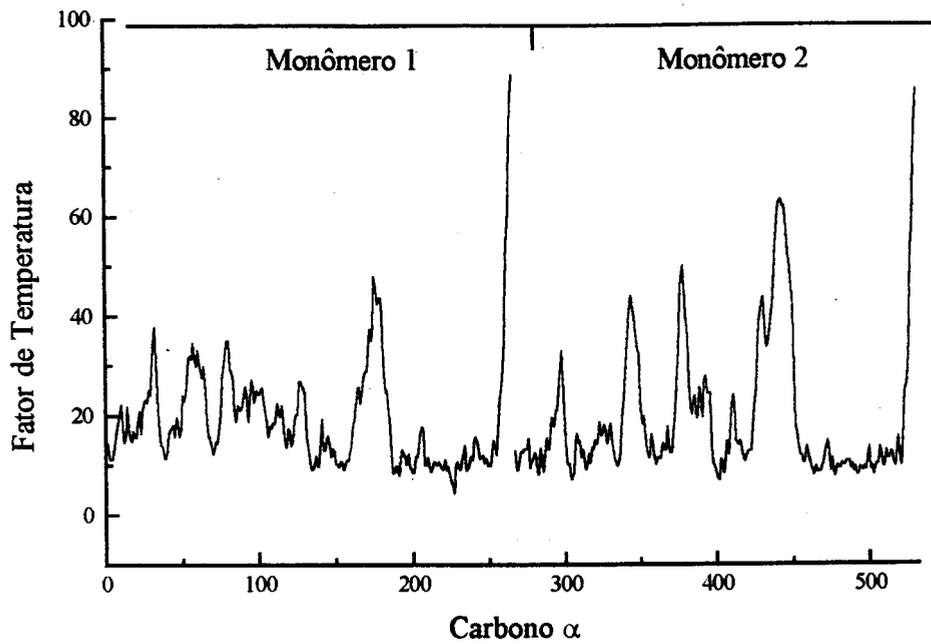


Figura 4.8 - Gráfico mostrando os fatores de temperatura ( $\text{\AA}^2$ ) *versus* carbono  $\alpha$  de cada aminoácido (1-266 monômero 1, 277-532 monômero 2).

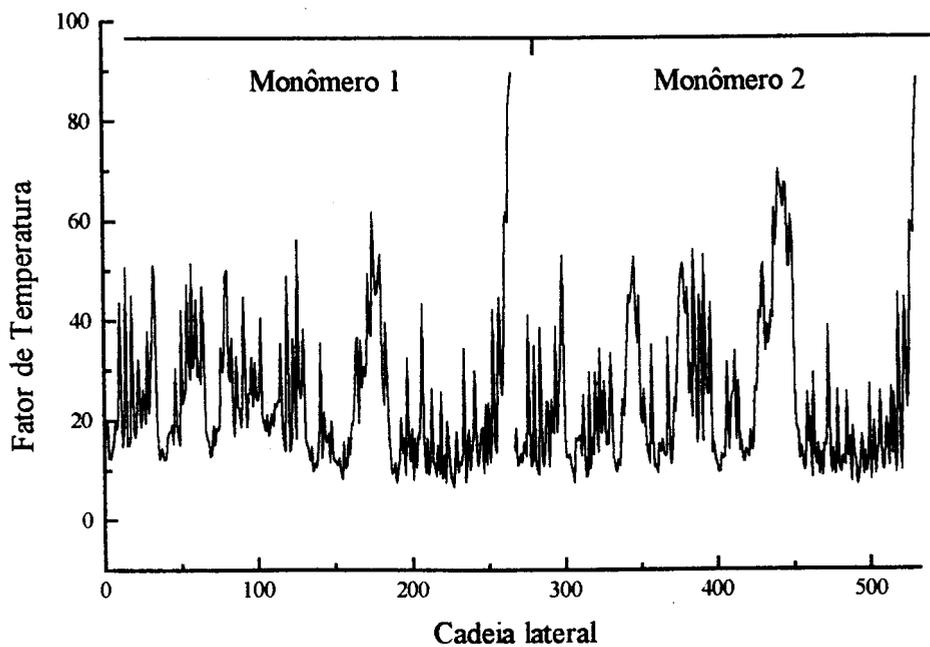


Figura 4.9 - Gráfico mostrando o valor médio dos fatores de temperatura ( $\text{\AA}^2$ ) *versus* átomos da cadeia lateral (1-266 monômero 1, 277-532 monômero 2).

## CAPÍTULO 5

# CO-CRISTALIZAÇÃO, COLETA DE DADOS, CONSTRUÇÃO E REFINAMENTO DOS COMPLEXOS: PROTEÍNA+ATIVADOR ALOSTÉRICO E PROTEÍNA+INIBIDOR.

O objetivo deste estudo é obter um maior conhecimento das características dos sítios ativo (ou catalítico) e alostérico da proteína Glucosamina-6-fosfato desaminase pela ligação do ativador alostérico (N-Acetil-D-Glucosamina-6-Fosfato) e de um inibidor competitivo (2-Desoxi-2-Amino-Glucitol-6-Fosfato) e desta maneira, tentar entender os seus mecanismos de catálise e de alostericidade.

### 5.1 - CO-CRISTALIZAÇÃO DOS COMPLEXOS

#### 5.1.1 - Introdução

O processo de co-cristalização consiste em crescer cristais da proteína nativa que já estavam ligadas (antes do momento da cristalização) com alguma molécula ou cofator por meio de algum tipo de ligação química (ligação iônica, pontes de hidrogênio, ponte salina etc).

No capítulo 2, foi descrito o processo de difusão em solução (*soaking*) que é a ligação de alguma molécula ou cofator com a proteína que já tinha sido cristalizada anteriormente.

Desta maneira, temos dois métodos para ligação de alguma molécula ou cofator com a proteína nativa. Porém, ambos apresentam características e finalidades diferentes. No método de difusão em solução (*soaking*), onde a proteína já está cristalizada, esta tem pouca mobilidade para mudanças conformacionais e assim é especialmente indicado nos casos em que se necessita da obtenção de cristais isomorfos (idênticos em termos de grupo espacial e celas unitárias) porém, apresenta a desvantagem para a ligação de cofatores que exigem mobilidade da proteína para que estes se liguem. O caso da co-cristalização é exatamente o contrário, isto é, a proteína possui total mobilidade para a ligação de cofatores; podendo ocorrer o crescimento de cristais em grupos espaciais diferentes em relação ao dos cristais nativos.

No caso desta tese queremos estudar justamente como é o comportamento, em termos estruturais, da proteína quando está ligada com o ativador, ou com o inibidor, ou seja, quando ocorrem estas ligações, como os sítios catalíticos e

alostéricos da proteína ficarão em termos estruturais; desta maneira, foi escolhido o método da co-cristalização.

### 5.1.2 -Processo para cristalização de proteínas: *salting-out*, difusão de vapor e *hanging drop*.

Para cristalização (ou co-cristalização) é utilizado o processo de *salting-out*. Este processo consiste na preparação de uma solução de um agente precipitante e de proteína, e posterior aumento gradativo da concentração do precipitante até alcançar o ponto de super-saturação da proteína, acarretando em sua precipitação ou na sua cristalização, se este processo for bastante lento. Deve-se lembrar que, em termos termodinâmicos, o estado de menor energia para a proteína é o estado cristalino, porém atingir este estado pode requerer um processo termodinamicamente complicado.

Um método bastante eficiente para o aumento gradativo da concentração do agente precipitante é por difusão de vapor, no qual, em recipiente fechado, é colocada uma solução de precipitante com alta concentração com uma solução formada pelo mesmo precipitante e proteína, podendo fazer trocas apenas por fase gasosa. Para tanto, pode-se utilizar o sistema chamado *hanging drop* (gota suspensa) figura 5.1.

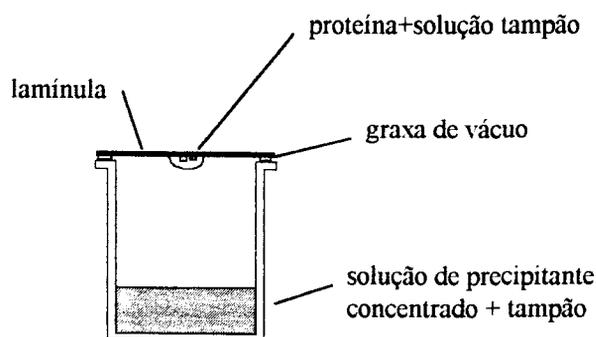


Figura 5.1 - Método para cristalização: *hanging drop* (gota suspensa)

Neste sistema, a solução de proteína e agente precipitante ficam na gota suspensa e o precipitante com alta concentração fica no recipiente. A gota fica pendurada devido à tensão superficial entre esta e a lamínula, pois nesta foi feita um processo de siliconização que aumenta a tensão superficial. Deve-se lembrar que o volume do precipitante no recipiente (normalmente 1 ml) é muito maior que o da lamínula (10  $\mu$ l), podendo desta maneira ser considerado um recipiente

infinito em relação a gota e assim a concentração final do precipitante da gota, após um tempo grande, será igual ao do precipitante do recipiente.

Deve-se ressaltar que nos experimentos de cristalização com proteínas um parâmetro muito importante é o pH da solução onde está a proteína, desta maneira, além do precipitante deve-se usar um tampão que mantenha o pH sob controle. Outro parâmetro importante no processo de cristalização é a temperatura que deve ser constante durante todos os experimentos. Além destes parâmetros há vários outros, tais como a força iônica que é caracterizada pelo tipo de precipitante utilizado e outros possíveis aditivos presentes no meio de cristalização.

### 5.1.3 - Procedimento experimental

#### 5.1.3.1 - Preparação da proteína nativa

Inicialmente, preparou-se a proteína fornecida pela Dra. Myriam Altamirano e Dr. Mário Calcagno da mesma maneira que foi preparada para obter os cristais nativos. Ou seja, a proteína que estava precipitada em sulfato de amônia 2.5 M deve ser, através de uma diálise, solubilizada em um tampão Tris/HCl 10mM pH 7.5. Após a diálise, faz-se uma centrifugação, retira-se o sobrenadante e finalmente, outra centrifugação com membrana semi-permeável (centricon nº 100). A última etapa para preparação é a medida aproximada da concentração da proteína utilizando um espectrofotômetro UV/Vis. Pois, como o coeficiente de extinção (para  $\lambda=278$  nm) já era conhecido por experimentos bioquímicos  $\epsilon_{278} = 20 \text{ M}^{-1}\text{cm}^{-1}$  (Altamirano, M.M., 1993), a concentração da proteína pode ser obtida pela relação  $[C]=A.\epsilon^{-1}l^{-1}$ , onde  $A$  é a absorvância da solução em 278 nm e  $l$  o caminho ótico do feixe de luz na cubeta. A concentração utilizada para obter os cristais nativos é da ordem de 10 mg/ml, desta maneira será utilizada a mesma ordem de grandeza para os experimentos de co-cristalização.

#### 5.1.3.2 - Determinação das condições para cristalização

O primeiro passo no processo de cristalização é saber quais devem ser as condições de precipitação para cada um dos agentes precipitantes que não são necessariamente as mesmas utilizadas para a proteína nativa.

Então, pode-se fazer um teste preliminar de precipitação para determinar qual deve ser a concentração aproximada do precipitante, da seguinte maneira:

1. Preparação dos complexos proteína+ativador e proteína+inibidor. Estes são preparados simplesmente adicionando inibidor ou ativador (com concentrações finais de 0.45 e 4.5 mM respectivamente) à proteína com pelo menos 10 minutos antes de fazer as etapas abaixo.

2. Sob o microscópio vai-se colocando gotas de aproximadamente 1  $\mu$ l do precipitante+tampão no complexo e observando se há precipitação da solução formada.
3. Ocorrida a precipitação, pode-se calcular a partir do volume colocado do tampão qual será a condição de precipitação para esta.
4. O ponto de precipitação, em geral, está muito próximo da condição de cristalização.

Outras condições de cristalização como pH e temperatura não serão alteradas em relação às da proteína nativa.

#### 5.1.3.3 - Montagem das Placas de Cristalização

Para os experimentos desta tese utilizou-se placas de cultura celular de acrílico que possuem trinta poços (a figura 5.2 mostra uma placa com quinze poços) as quais são seladas utilizando graxa de ultra-vácuo e lamínulas siliconizadas com a gota de proteína+tampão.

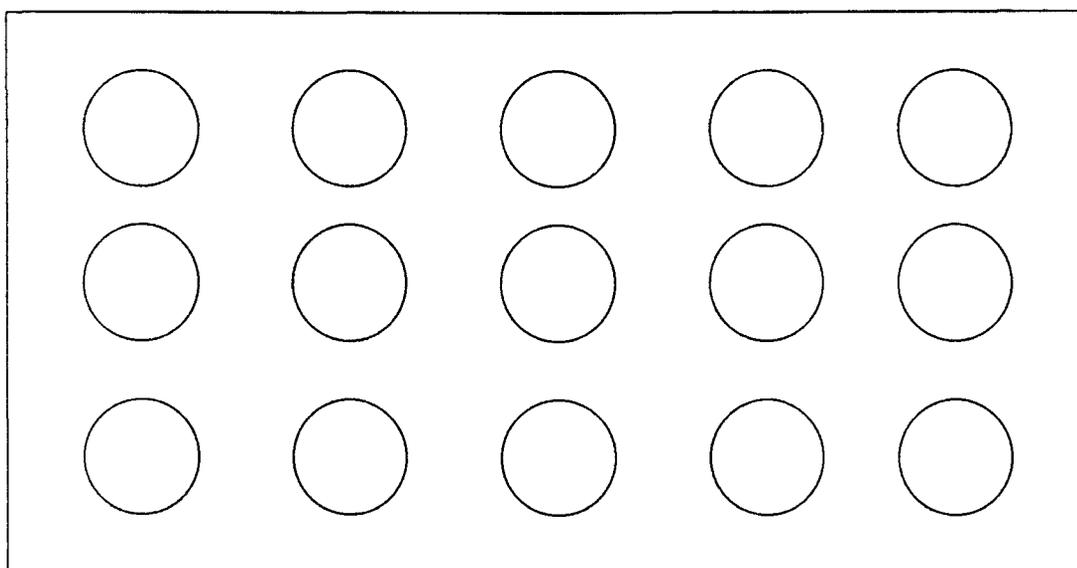


Figura 5.2 - Arranjo utilizado para testes de cristalização (placas de cultura celular).

Foram montadas placas de cristalização com os seguintes complexos:

1. Proteína+Ativador (concentração final do ativador na gota: 4.5 mM).
2. Proteína+Inibidor (concentração final do inibidor na gota: 0.45 mM).
3. Proteína+Ativador+Inibidor (concentração final na gota: ativador: 4.5 mM inibidor: 0.45 mM).

Fez-se inicialmente três placas (usando os três complexos em cada uma delas) utilizando para cada uma destas os seguintes agentes precipitantes e tampões com as seguintes faixas de concentração:

1. Fosfato de sódio/potássio ( $K_2HPO_4$  e  $NaH_2PO_4$ ) pH 8.2, como agente precipitante e tampão (este sal atua com as duas finalidades), faixas de concentração: 0.76 M-0.52 M
2. Acetato de sódio (AcNa) como sal precipitante em tampão de Tris/HCl pH 8.1 200mM, faixas de concentração: 0.8 M-1.36 M
3. Citrato de sódio como sal precipitante e tampão, faixas de concentração: 35%-50 % (porcentagem da solução saturada)

A montagem das placas consiste das seguintes etapas:

1. Montagem das gotas: 5  $\mu$ l um dos complexos + 5  $\mu$ l da solução tampão.
2. Montagem dos poços: 1 ml da composição de um dos agentes precipitantes+tampão e água.

#### 5.1.3.4 - Cristais obtidos

Todos os cristais obtidos para os complexos tinham uma morfologia aproximadamente cúbica, bastante diferente da morfologia dos cristais nativos, com os seguintes tamanhos máximos aproximados e respectiva concentração de precipitante:

- a) Fosfato:
  - ativador: 0.5x0.4x0.4 mm (0.72 M)
  - inibidor: 0.2x0.2x0.1 mm (0.64 M)
  - ativador+inibidor: 0.1x0.1x0.1 mm (0.56 M)
- b) Acetato de sódio:
  - ativador: microcristais menores que 0.1 mm (em toda faixa).
  - inibidor: 0.15x0.15x0.1mm (0.8 M)
  - ativador+inibidor: 0.1x0.1x0.1 mm (0.8 M)
- c) Citrato de sódio:
  - ativador: 0.3x0.2x0.2 mm (0.5 sat)
  - inibidor: 0.2x0.2x0.1 mm (0.4 sat)
  - ativador+inibidor: 0.1x0.1x0.1 mm (em toda faixa)

Nas figuras 5.3, 5.4, 5.5 são mostradas fotos dos cristais obtidos de ativador em fosfato, inibidor em fosfato e inibidor em acetato com as concentrações descritas acima.

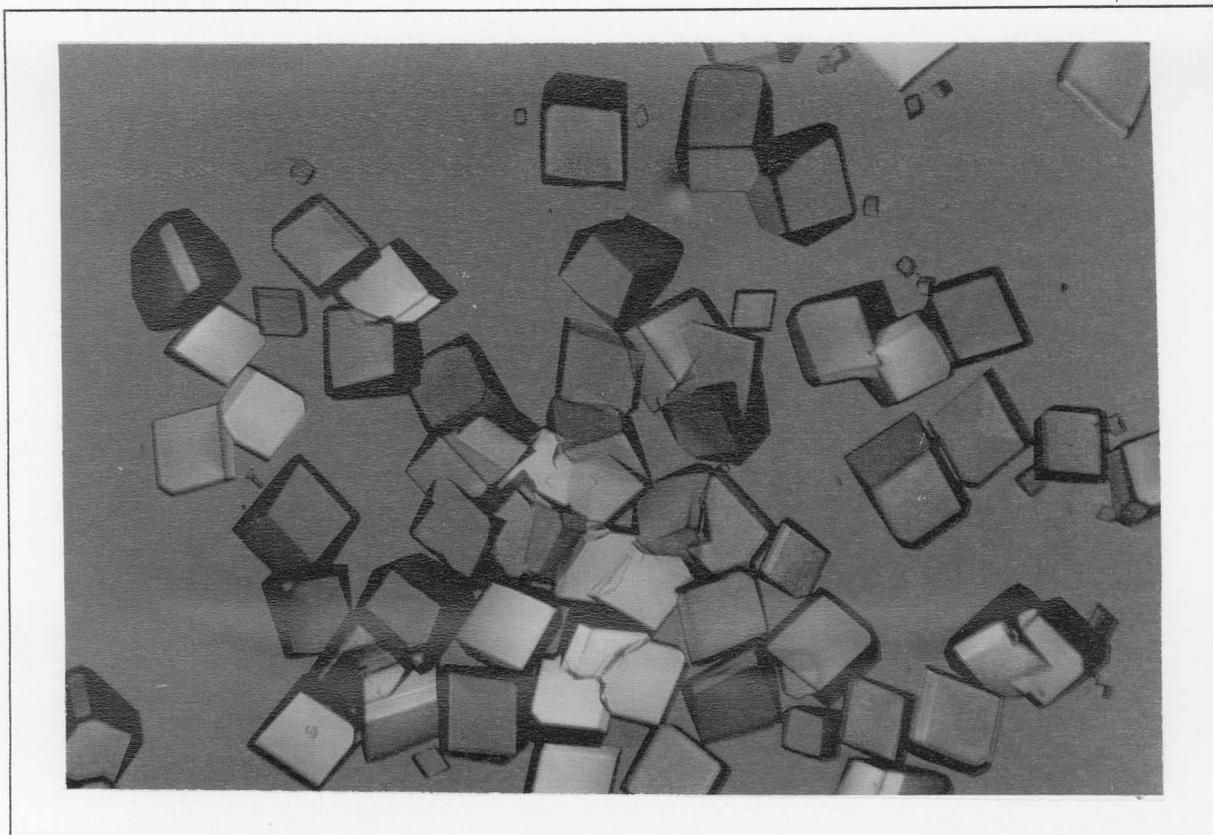


Figura 5.3-Cristais de enzima+ativador alostérico crescidos em  $K_2H/NaH_2PO_4$  0.72 M, pH 8.2.

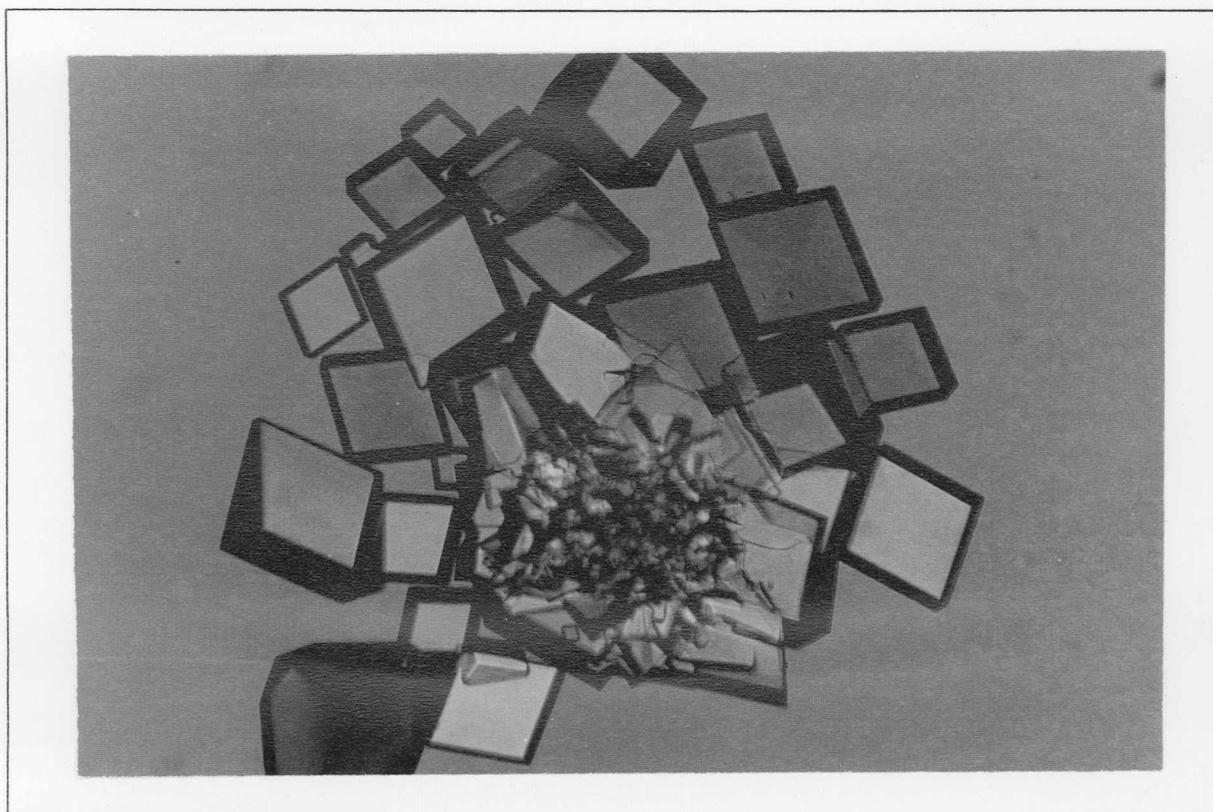


Figura 5.4 - Cristais de enzima+inibidor crescidos em  $K_2H/NaH_2PO_4$  0.64 M, pH 8.2.

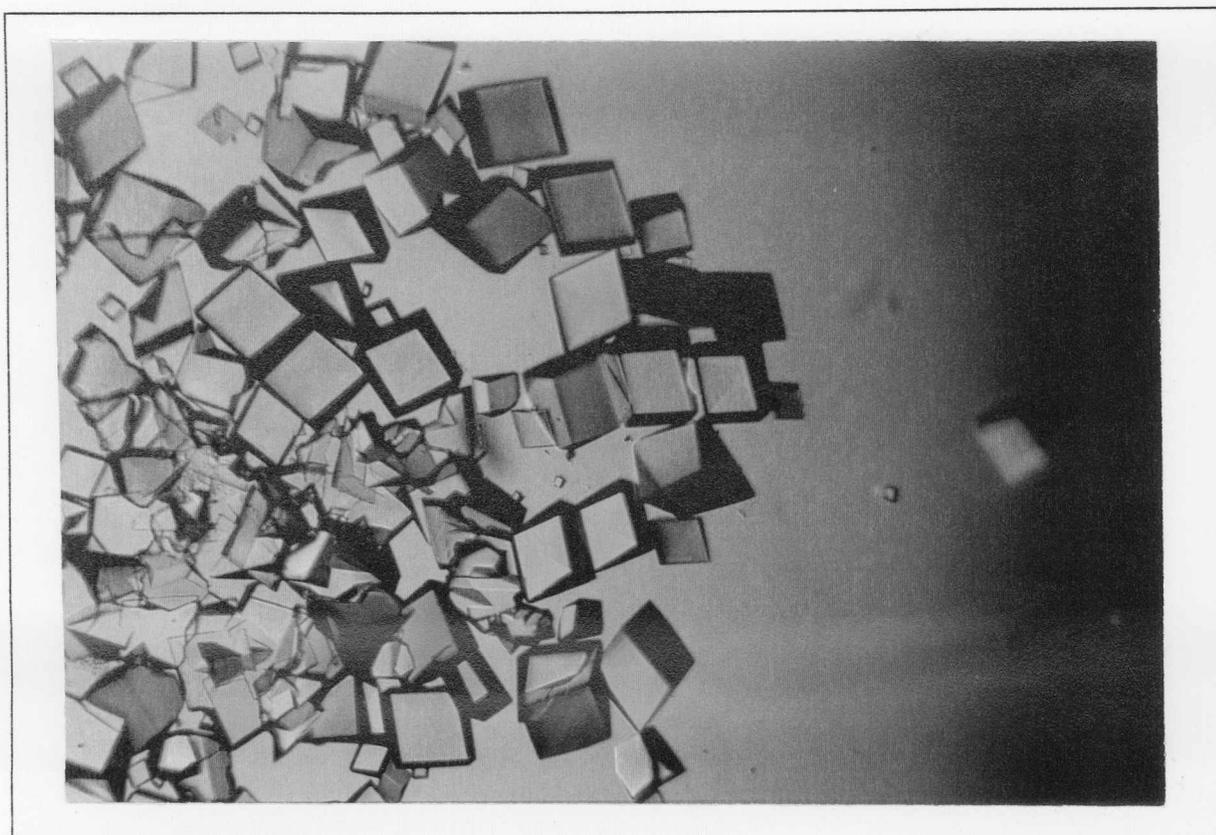


Figura 5.5 - Cristais de enzima+inibidor crescidos em acetato de sódio, 0.8 M, pH 8.1.

## 5.2 - COLETA DE DADOS

A coleta de dados foi feita da mesma maneira utilizada para os derivados isomorfos (capítulo 2), ou seja, usando uma fonte de raios-X convencional com ânodo rotatório e detetor de área com placa de imagem RAXIS-IIC da RIGAKU Co.

Todas as coletas de dados foram feitas utilizando o sistema de resfriamento do equipamento que mantém durante toda a coleta de dados um fluxo de nitrogênio sobre o cristal a 4° C.

Devido ao tamanho dos cristais obtidos e interesse experimental, foram feitas coletas de dados para cristais co-cristalizados com ativador e inibidor utilizando como precipitante fosfato Na/K e Acetato de Sódio.

Os dados das coletas de dados são sumarizados na tabela 5.1.

Da tabela 5.1, é importante notar que para todos os complexos tem-se o grupo espacial R32 e parâmetros de rede iguais entre si e com os dados do cristal nativo, dentro da margem de erro experimental.

Os cristais para ativador em acetato de sódio eram muito pequenos e foi possível indexar os parâmetros de rede apenas aproximadamente e foi impossível processar os seus dados.

Os dados dos complexos foram processados da mesma maneira descrita no capítulo 2.

Tabela 5.1 - Resumo da coleta de dados dos complexos co-cristalizados (os dados para ativador em acetato sódio são aproximados)

	Ativador (fosfato Na/K)	Ativador (acetato de sódio)	Inibidor (fosfato Na/K)	Inibidor (acetato de sódio)
Grupo espacial	R32	R32	R32	R32
Parâmetros de rede				
a=b (Å)	125.7	125.7	125.8	125.9
c	223.0	223.1	223.1	222.5
$\alpha=\beta$ (°)	90	90	90	90
$\gamma$	120	120	120	120
n° de moléculas na unidade assimétrica	2 moléculas	-	2 moléculas	4 moléculas
n° de reflexões	43420	-	37161	27906
(observadas/independentes)	25833	-	22131	16550
resolução/completeza	3.0 - 91.6	-	3.0 - 89.1	3.0 - 86.0
(Å,%)	2.5 - 88.1	-	2.5 - 81.1	2.5 - 78.5
	2.1 - 64.8	-	2.1 - 60.7	2.1 - 57.7
$R_{\text{merge}}$ (%)	5.14	-	5.94	7.12

### 5.3 - CONSTRUÇÃO DOS MODELOS DOS COMPLEXOS

Para a construção do modelo tem-se que fazer um mapa Fourier diferença utilizando como coeficientes: *fatores de estrutura observados dos complexos - fatores de estrutura calculados do nativo*. Para tanto utilizou-se o programa FFT do pacote de programas CCP4 (veja itens 3.1 e 3.3).

A partir do mapa Fourier diferença do complexo com ativador (em precipitante fosfato) foi possível localizar a densidade eletrônica relativa à molécula de ativador alostérico, localizado na interface dos monômeros, com o fosfato desta molécula exatamente no local onde havia um íon fosfato na molécula nativa. Identificaremos a partir de agora este complexo por enzima ativador-Pi.

Da mesma maneira, a partir do mapa Fourier diferença para o complexo com inibidor em Na/K fosfato, havia densidade eletrônica relativa a molécula de inibidor, com o fosfato desta no local onde havia o outro íon fosfato (não o sítio da interface) na molécula nativa. Este complexo será identificado como: enzima inibidor-Pi.

Quanto ao complexo com inibidor em acetato de sódio havia inibidores em ambos os sítios de íons fosfato na molécula nativa, porém com conformações diferentes entre si. Este complexo será identificado como: enzima 2-inibidor.

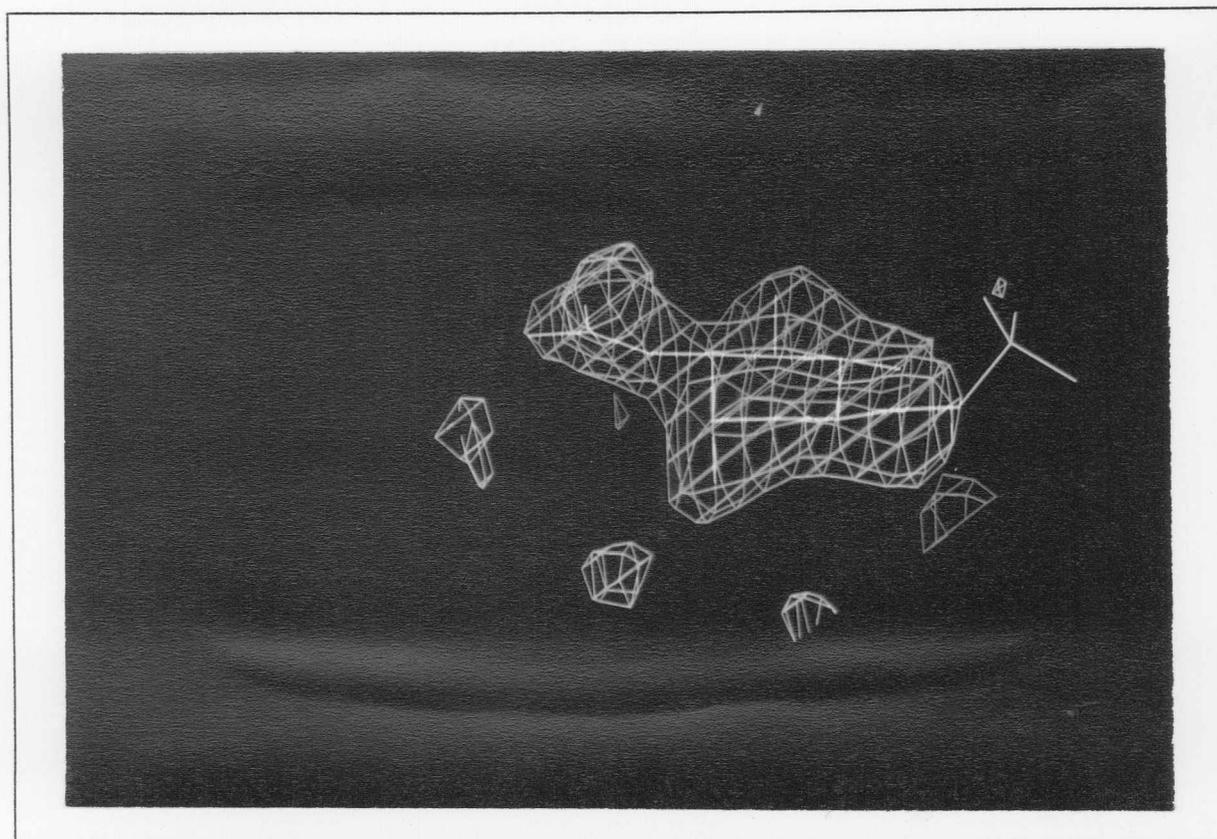


Figura 5.6 - Mapa diferença ( $3\sigma$ ) mostrando o ativador (complexo de proteína e ativador em precipitante fosfato Na/K), localizado entre dois monômeros.

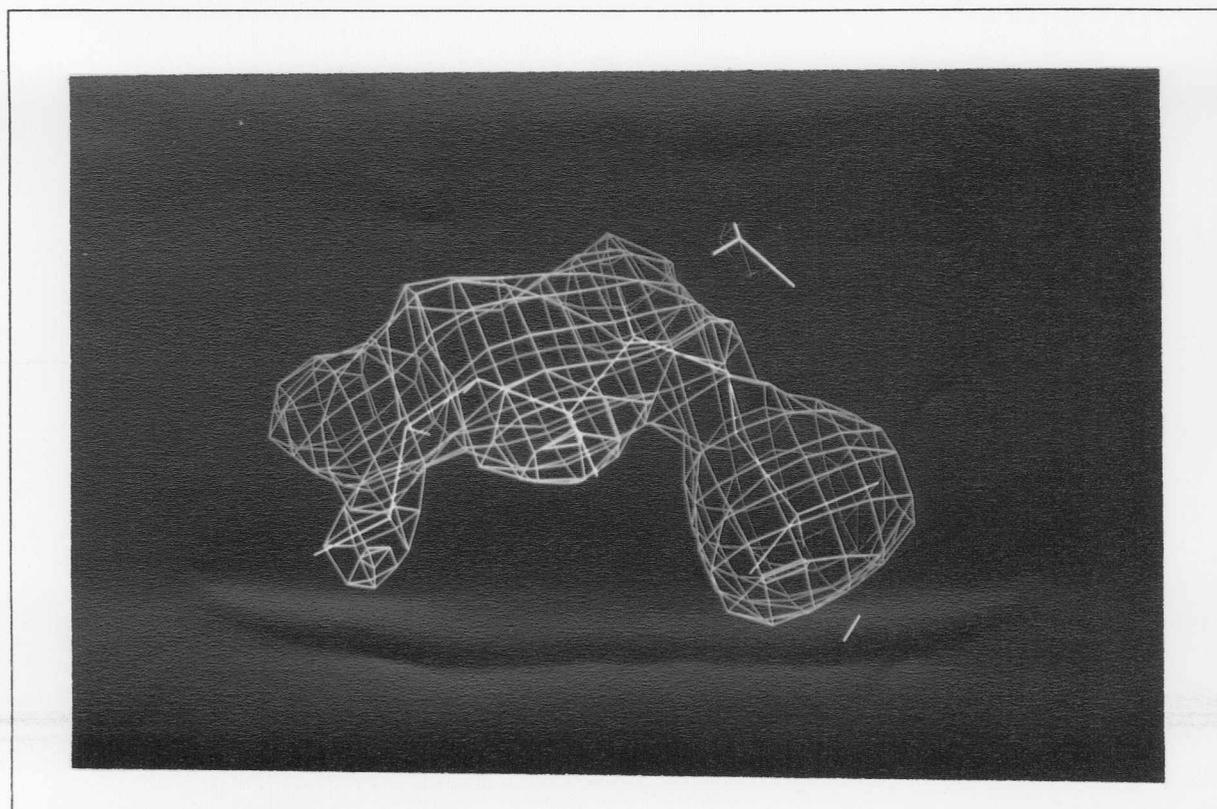


Figura 5.7 - Mapa diferença ( $3\sigma$ ) mostrando o inibidor (complexo de proteína e inibidor em precipitante fosfato Na/K), localizado no interior de um monômero.

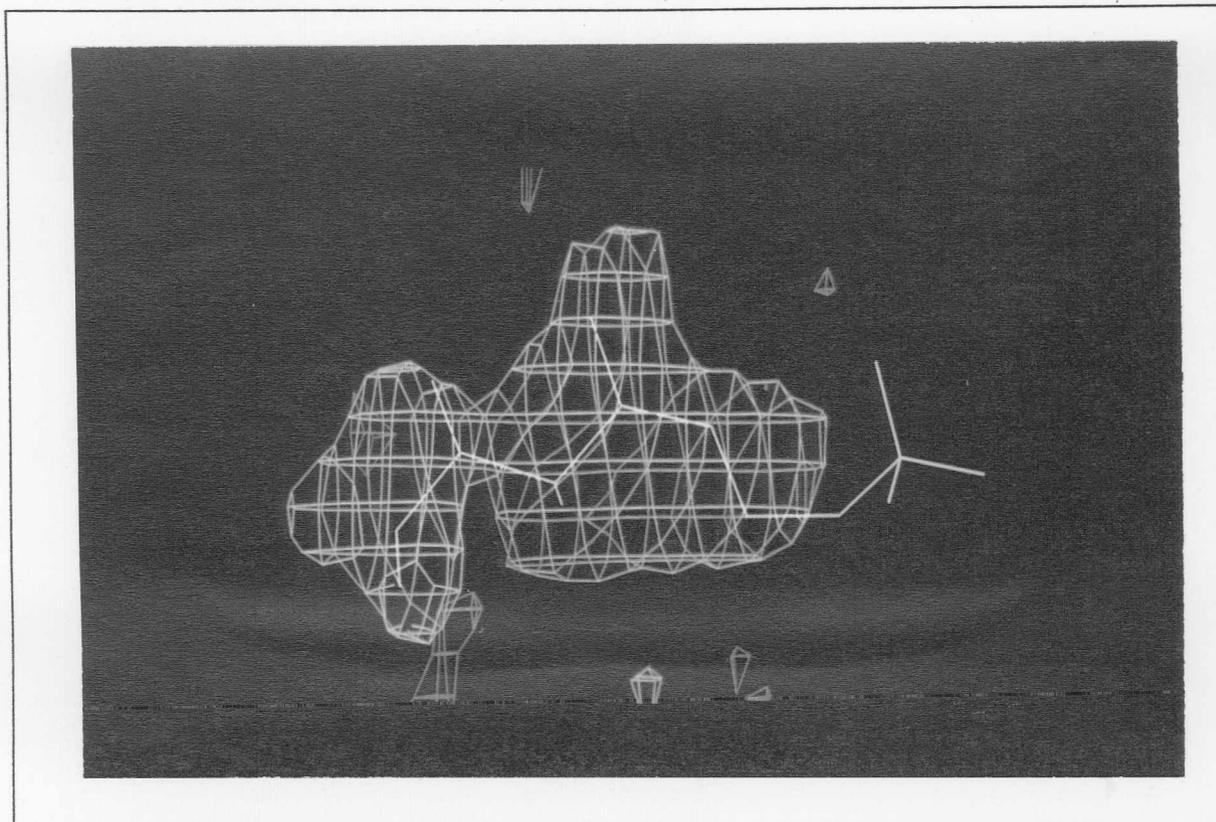


Figura 5.8 - Mapa diferença ( $2.5\sigma$ ) mostrando o inibidor (complexo de proteína e inibidor em precipitante acetato de sódio), localizado entre dois monômeros.

Foi possível visualizar com muita clareza as moléculas dos complexos com ativador-Pi e inibidor-Pi com mapas de  $3\sigma$ , e do complexo com 2-inibidor com mapas de  $2.5\sigma$  (onde  $\sigma$  é *rms*, valor médio quadrático).

As figuras 5.6, 5.7 e 5.8 mostram os mapas densidade eletrônica (mapas Fourier diferença) relativos ao ativador no sítio de ligação entre os monômeros, inibidor no outro sítio de ligação e inibidor no sítio entre os monômeros (complexo com precipitante acetato de sódio), com os modelos relativos a estes construídos tomando como base estes mapas (veja a seguir).

Com base nestes mapas Fourier diferença, a próxima etapa é a construção de modelos representando estes.

A construção dos modelos do ativador e inibidor (diferentes para os dois sítios) se deu da seguinte maneira:

1. Construção de um modelo inicial utilizando o programa INSIGHT II (*Insight II*, 1993)
2. Modelagem a partir do modelo inicial, levando em conta a densidade eletrônica utilizando o programa TOM (Cambillan, C.C. e Horjales, E., 1987)
3. Refinamento das posições do modelo (proteína e ativador ou inibidor), item 5.4.

Do mapa densidade eletrônica do ativador foi possível ver claramente que este está na conformação do enantiômero  $\alpha$  (o enantiômero  $\alpha$  significa que a

hidroxila do carbono 1 está em posição *cis* em relação ao grupo NAcetil do carbono 2 - figura 5.12).

#### 5.4 - REFINAMENTO DOS COMPLEXOS E LOCALIZAÇÃO DE MOLÉCULAS DE ÁGUA

Para o refinamento dos complexos, utilizou-se o programa X-PLOR (veja capítulo 4).

Foram usadas inicialmente as fases calculadas do modelo da proteína nativa, excluindo, conforme o caso (complexos enzima ativador-Pi, inibidor-Pi ou 2-inibidor) um ou dois íons fosfato do modelo da proteína nativa.

Como os parâmetros de geração da estrutura molecular destes complexos (item 4.3.1) não estão disponíveis nos arquivos do programa, estes devem ser adaptados utilizando arquivos já existentes no programa X-PLOR para complexos parecidos com o ativador e inibidor, fazendo-se várias modificações condizentes com a estrutura e carga destes (utilizou-se os seguintes arquivos de parâmetros existentes no X-PLOR: param3\_mod.cho e generateco.inp; e para arquivos de topologia: toph3.cho e generateco.inp). No apêndice A são listados os arquivos de topologia e parâmetros adaptados para os refinamentos dos complexos.

O procedimento para o refinamento dos complexos foi o seguinte:

1. Geração da informação molecular usando os arquivos criados acima (rotina *topology*).
2. Refinamento das posições com minimização de energia (*positional refinement*).
3. Refinamento dos fatores de temperatura em grupo.
4. Refinamento dos fatores de temperatura individuais.
5. Geração e análise do mapa densidade eletrônica.
6. Repetição de 1 a 5 (faz-se 1 somente se forem incluídos novos átomos).
7. Inclusão de moléculas de água, utilizando um mapa Fourier diferença (item 4.3.6).
8. Novo processo de 5 a 7.

Foram refinados além dos dois monômeros, dois íons fosfato, duas moléculas de ativador e 189 moléculas de água para o complexo enzima ativador-Pi; para o complexo enzima inibidor-Pi, dois íons fosfato, duas moléculas de inibidor e 216 moléculas de água; para o complexo enzima 2-inibidor, quatro moléculas de inibidor e 174 moléculas de água.

Analisando os quatro modelos obtidos após o refinamento (nativa, ativador-Pi, inibidor-Pi, 2-inibidor), observou-se somente pequenas mudanças próximas aos sítios alostéricos e catalíticos.

Devido a grande semelhança entre os modelos após o refinamento, estes foram superpostos na estação gráfica; em seguida, verificou-se cada molécula de

água dos modelos, onde haviam moléculas de água muito próximas referentes a três dos modelos, foi colocada a molécula correspondente no quarto modelo, com exceção das águas próximas aos sítios alostéricos e catalíticos. Após este processo, os modelos foram refinados novamente.

As moléculas de água que apresentavam fatores de temperatura (B's) superiores a  $60 \text{ \AA}^2$  foram retiradas e feito novo refinamento dos modelos. No final os valores de B's médio para os complexos ativador-Pi, inibidor-Pi e 2-inibidor foram 35.81, 34.59,  $34.12 \text{ \AA}^2$ .

A tabela 5.2 sumariza dados obtidos durante o refinamento dos complexos, refinados à  $2.1 \text{ \AA}$  de resolução.

Tabela 5.2 - Dados estatísticos para os complexos

	Ativador-Pi (Fosfato Na/K)	Inibidor-Pi (Fosfato Na/K)	2-Inibidor (Acetato de sódio)
R (%)	16.1	16.5	17.0
R <sub>free</sub> (%)	21.2	22.5	22.7
rms-distâncias(Å)	0.005	0.005	0.005
rms-ângulos(°)	1.21	1.21	1.20
rms-âng. diédricos(°)	23.81	23.75	23.51

Na tabela 5.3 é mostrada a evolução dos fatores R dos complexos durante o refinamento. Foi utilizado como peso nos refinamentos finais dos complexos o valor de  $W_A = 400.000$ , que é aproximadamente 50-60 % do valor obtido na rotina *check*.

Tabela 5.3 - Índices de discordância (R) em algumas etapas do refinamento dos complexos

Etapa do Refinamento (final de cada rotina)	R(%) Ativador-Pi	R(%) Inibidor-Pi	R(%) 2-Inibidor
Valor inicial	24.0	24.1	25.2
Refinamento das posições (sem água)	21.7	21.9	23.8
Refinamento dos B's (sem água)	21.3	21.5	22.4
Refinamento das posições (com água)	19.1	19.3	20.0
Refinamento dos B's (com água)	18.6	18.8	19.4
Refinamento das posições (com água, após superposição)	16.5	16.8	17.3
Refinamento dos B's (com água, após superposição)	16.1	16.5	17.0

Nas fases iniciais do refinamento foi usado o peso  $W_A=200.000$  e nas últimas etapas (após superposição dos modelos)  $W_A = 400.000$ .

Os valores de fatores B's médio para os átomos  $C_\alpha$  e para os átomos das cadeias laterais são: 22.56, 26.21; 20.44, 23.57; 20.56, 23.88  $\text{Å}^2$  para ativador-Pi, inibidor-Pi e 2-inibidor.

Nas figuras 5.9, 5.10 e 5.11 são mostrados os gráficos representativos dos fatores de vibração térmica dos átomos de  $C_{\alpha}$  dos aminoácidos dos dois monômeros.

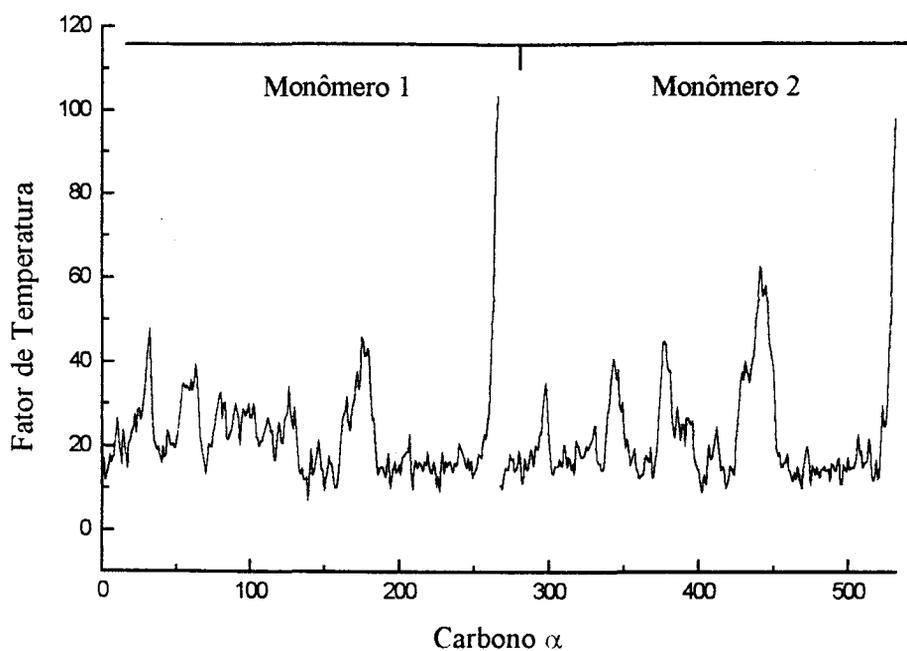


Figura 5.9 - Gráfico mostrando os fatores de temperatura ( $\text{\AA}^2$ ) versus carbono  $\alpha$  de cada aminoácido do complexo enzima ativador-Pi (monômero 1: 1-266, monômero 2: 267-532).

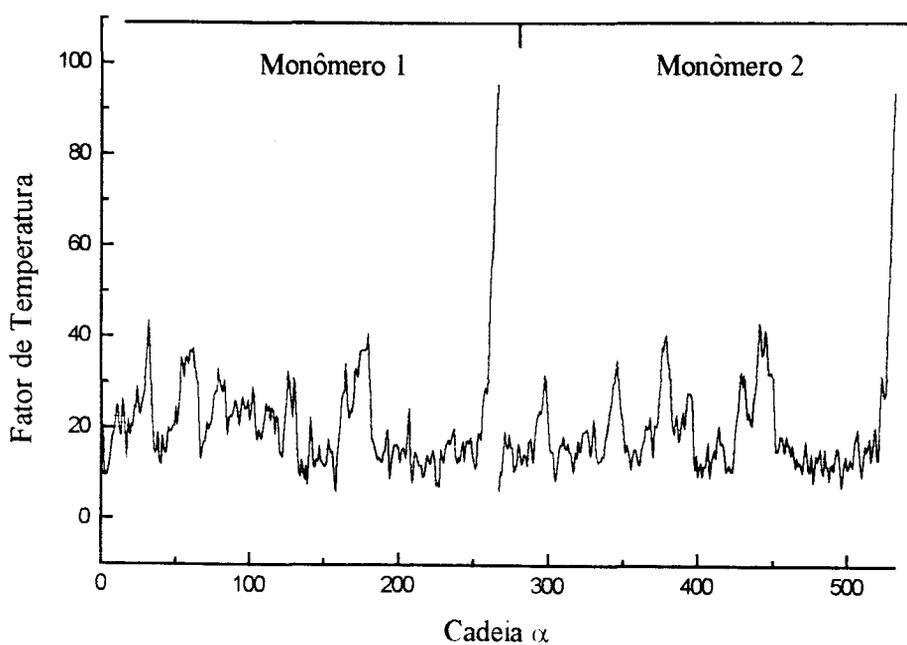


Figura 5.10 - Gráfico mostrando os fatores de temperatura ( $\text{\AA}^2$ ) versus carbono  $\alpha$  de cada aminoácido do complexo enzima inibidor-Pi (monômero 1: 1-266, monômero 2: 267-532).

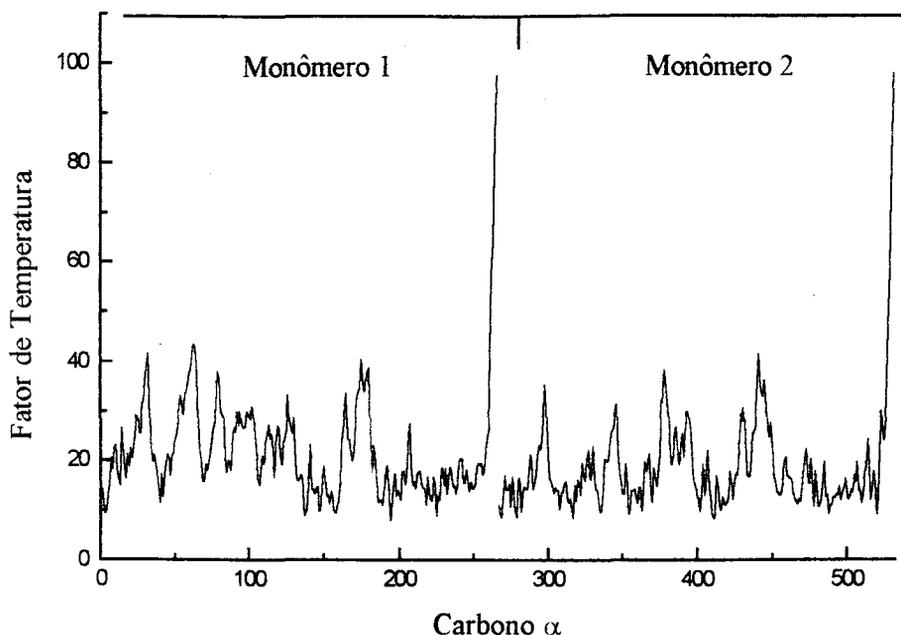


Figura 5.11 - Gráfico mostrando os fatores de temperatura ( $\text{\AA}^2$ ) versus carbono  $\alpha$  de cada aminoácido do complexo enzima 2-inibidor (monômero 1: 1-266, monômero 2: 267-532).

Estes gráficos são bastante similares entre si e com o gráfico 4.8 da enzima nativa, apresentando valores de B's muito altos para os átomos do C-terminal de ambos os monômeros.

As figuras 5.12, 5.13 e 5.14 mostram os modelos refinados com suas respectivas densidades eletrônicas para as molécula de ativador-Pi no sítio de ligação entre os monômeros (sítio alostérico, veja capítulo 6), inibidor-Pi no sítio no interior do monômero (sítio catalítico, veja capítulo 6) e 2-inibidor no sítio entre os monômeros.

A figura 5.15 representa um modelo do monômero da GlcN6P desaminase com as moléculas do ativador e do inibidor nos sítios alostérico e catalítico respectivamente.

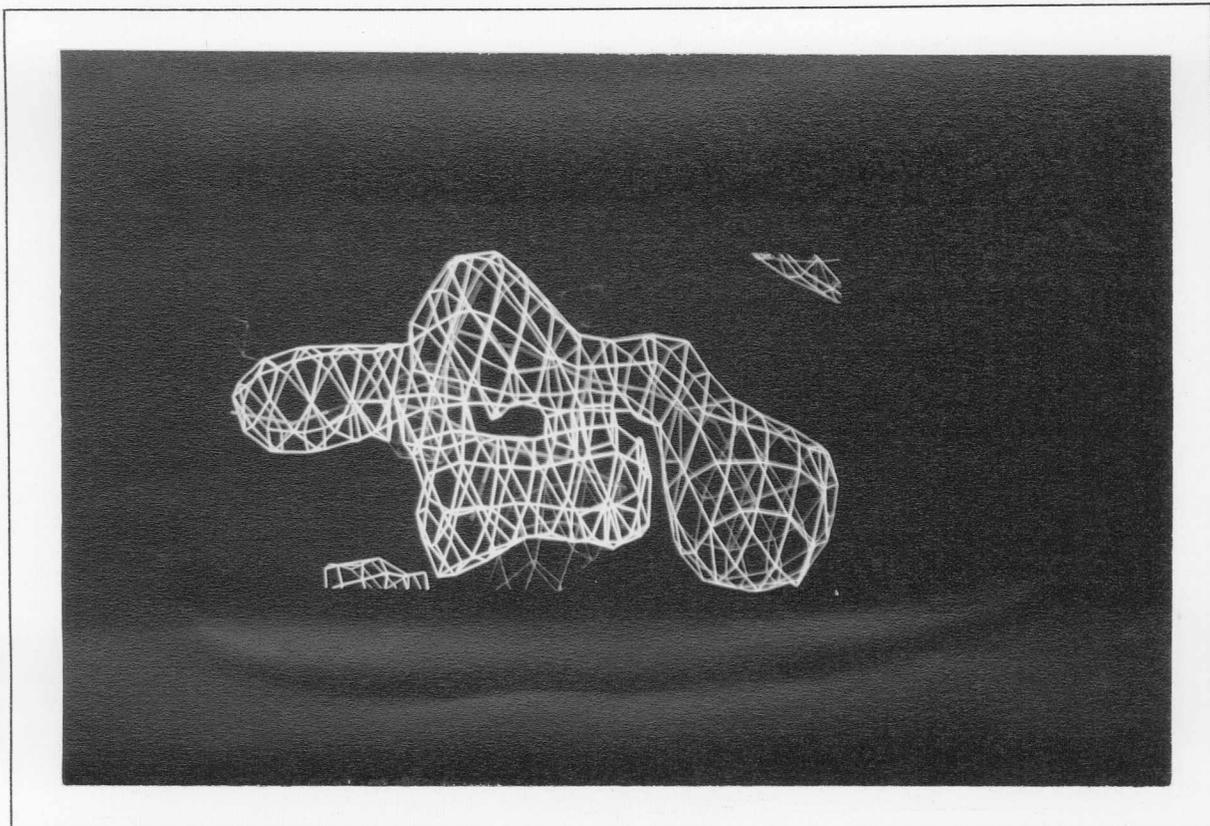


Figura 5.12 - Modelo do ativador no sítio alostérico (complexo de proteína e ativador em precipitante fosfato Na/K), após o refinamento utilizando o programa X-PLOR.

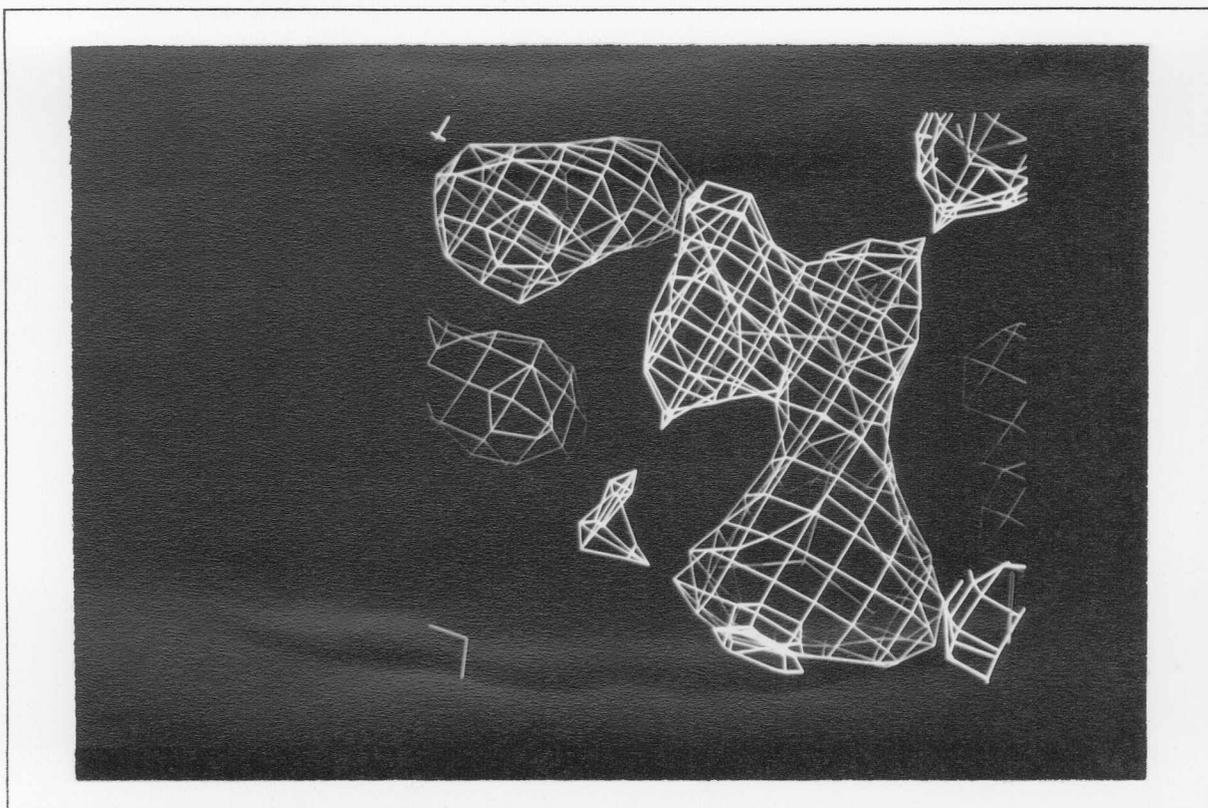


Figura 5.13 - Modelo do inibidor no sítio catalítico (complexo de proteína e inibidor em precipitante fosfato Na/K), após o refinamento utilizando o programa X-PLOR.

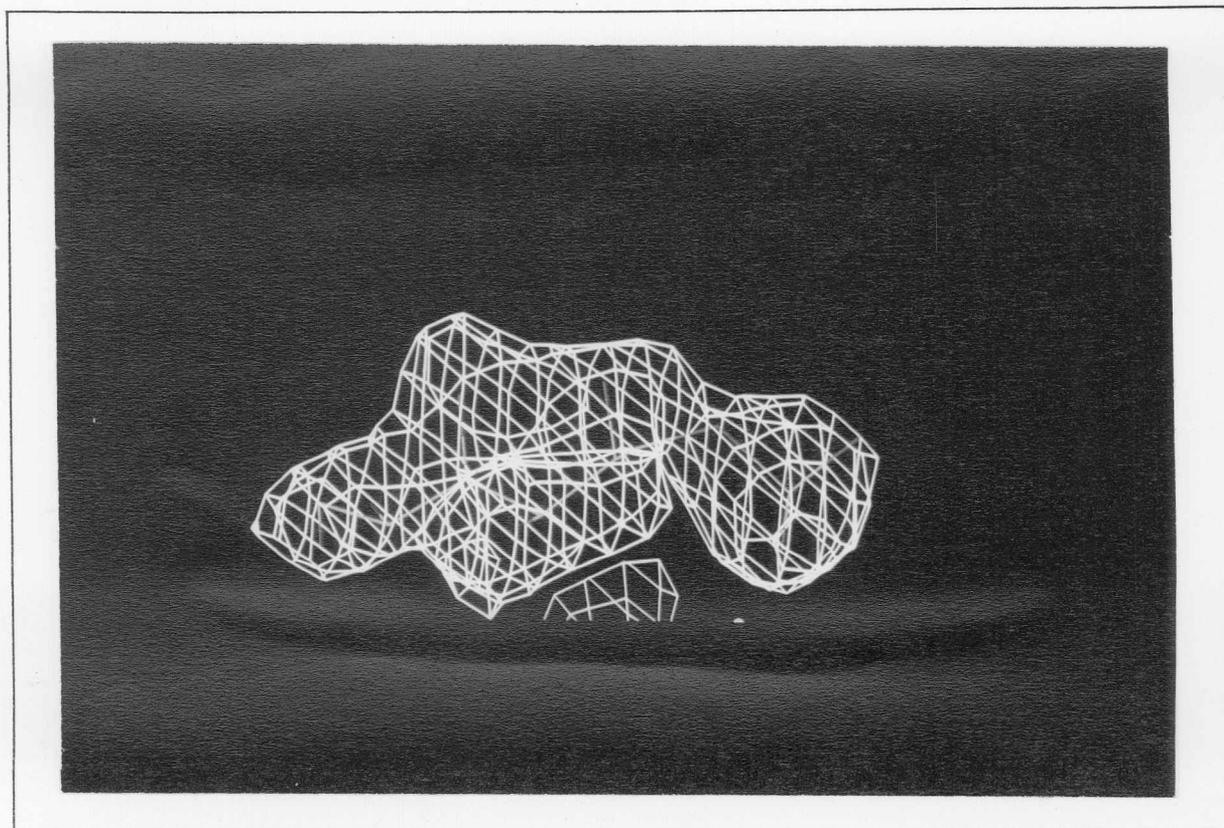


Figura 5.14 - Modelo do inibidor no sítio alostérico (complexo de proteína e inibidor em precipitante acetato de sódio), após o refinamento utilizando o programa X-PLOR.

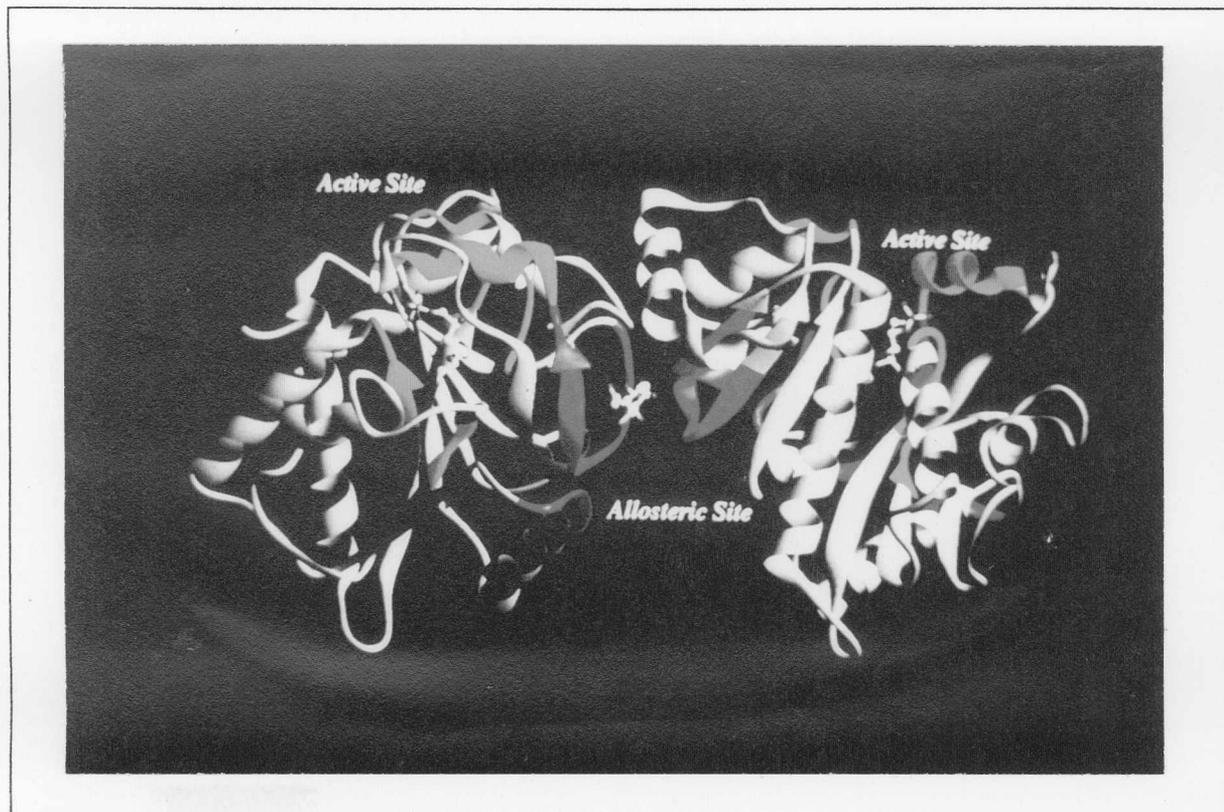


Figura 5.15 - Modelo da GlcN6P desaminase com as moléculas do ativador e do inibidor nos sítios alostérico e catalítico.

## CAPÍTULO 6

### RESULTADOS E DISCUSSÃO

A estrutura do complexo enzima com íons fosfato (ou enzima-2Pi) será chamada de estrutura nativa neste capítulo. No capítulo 7 será visto que a estrutura resolvida é possivelmente uma das conformações encontradas *in vivo*.

#### 6.1 -ANÁLISE DA QUALIDADE DA ESTRUTURA NATIVA

Para analisar se uma estrutura cristalográfica foi corretamente resolvida, pode-se notar, entre outros dados, os seguintes: qualidade dos mapas densidade eletrônica por inspeção, estereoquímica da estrutura, se a localização dos resíduos estão coerentes (hidrofóbicos no interior e hidrofílicos na parte externa), ambiente de cargas dos resíduos, bem como, analisar se a estrutura faz sentido com sua função e comparar os dados obtidos com os de outras técnicas. Serão feitos alguns destes testes a seguir.

Alguns dados estatísticos relativos ao refinamento são:  $R=17.4\%$ ,  $R_{\text{free}}=22.4\%$  (para 10% dos dados), para dados à 2.1 Å de resolução, incluindo o dímero (2 x 266 resíduos), quatro íons fosfato e 348 moléculas de água. Os desvios do *rms* (*root mean square*, raiz quadrática média) para valores geométricos ideais são: para distância das ligações de 0.015 Å, para ângulos 2.1° e para ângulos torcionais 1.0°.

O programa *Procheck* (Moris, A.L. *et al.*, 1992) analisa a estereoquímica da estrutura comparando com valores de outras estruturas resolvidas a alta resolução (no mínimo 2 Å) e de boa qualidade (com fator R menor que 20%) gerando vários gráficos. O gráfico mais interessante para se analisar é o diagrama de Ramachandran que analisa se os ângulos diédricos *phi* e *psi* estão em regiões usualmente favoráveis para proteínas. A figura 6.1 mostra o diagrama de Ramachandran para a estrutura nativa. Neste gráfico pode-se ver que não há nenhum aminoácido em regiões desfavoráveis e somente um resíduo em uma região 'generosamente' permitida (estas regiões estão indicadas no mapa por ~a, ~b, ~l, ~p), o resíduo Ala 145. Os triângulos representam resíduos de glicinas. Outros gráficos gerados pelo programa estão no apêndice B (estes gráficos levam em conta que a estrutura foi resolvida com resolução de 2.1 Å).

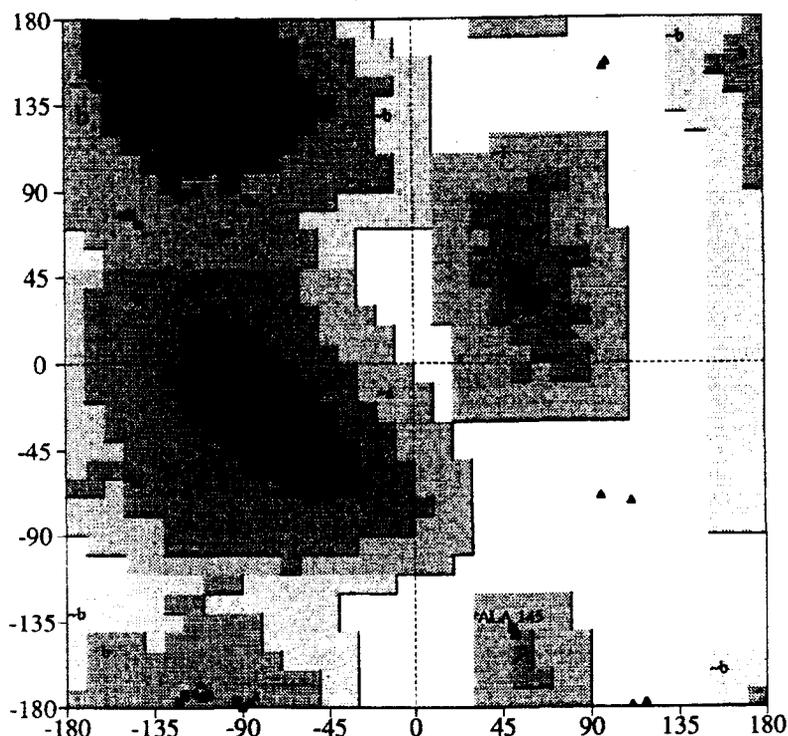


Figura 6.1 - Gráfico de Ramachandran do modelo da glucosamina 6-fosfato desaminase feito pelo programa *Procheck* (Moris, A.L. *et al*, 1992). O gráfico mostra que não há nenhum resíduo em regiões não permitidas e somente um resíduo em uma região 'generosamente' permitida, o resíduo Ala 145. Os triângulos representam resíduos de glicinas.

O programa *Quality* (Vriend, G. e Sander, C., 1993) do *What if* (Vriend, G., 1990) analisa a vizinhança dos resíduos da estrutura (principalmente em termos das cargas desta vizinhança e resíduos mais favoráveis) e compara com estruturas já resolvidas atribuindo notas a estes; as notas variam entre -10 a 10. Em um ambiente perfeitamente adequado, os resíduos terão notas próximas de 0, e são consideradas notas aceitáveis, valores abaixo de  $|6|$  ou  $|7|$ .

Para a estrutura de GlcN6P desaminase foi feito este teste para o dímero, que serviu inclusive para algumas correções na estrutura, resultando num modelo bastante razoável com apenas 15 resíduos com nota acima de  $|5|$  e 6 resíduos com nota acima de  $|6|$ , sendo dois destes C-terminais dos monômeros. Obteve-se um valor médio de -0.52, próximo ao valor médio de proteínas escolhidas do *Protein Data Bank* (PDB) para formar um banco de dados.

A figura 6.2 mostra um gráfico das notas atribuídas a cada aminoácido pelo programa *Quality*.

Vários dados experimentais de outras técnicas (veja itens 1.1 e 1.2) e cristalográficos estão em concordância, tais como:

1. Estrutura primária obtida por seqüenciamento e a densidade eletrônica obtida estão coerentes.
2. Estrutura quaternária hexamérica de subunidades idênticas.
3. Estrutura organizada como um trímero de dímeros.
4. Estrutura da enzima tipo  $\alpha/\beta$ .

5. Presença de sítios alostéricos e catalíticos, que podem ligar o ativador alostérico GlcNAc6P e o inibidor GlcN-ol-6P.
6. Um sítio alostérico por monômero ou seis por hexâmero.
7. Um sítio catalítico por monômero ou seis por hexâmero.
8. A Cys 219 forma ponte dissulfeto.

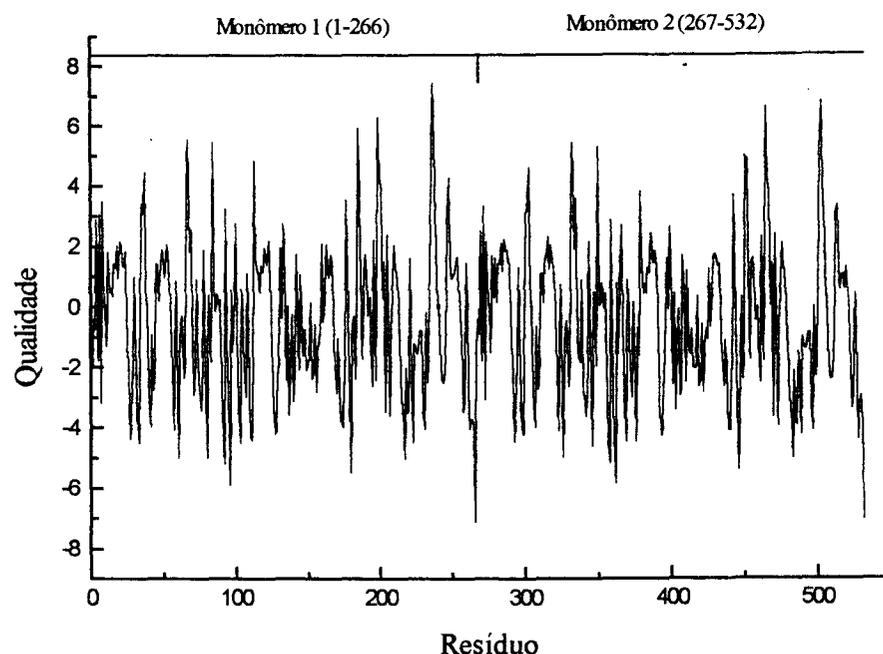


Figura 6.2 - Resultado da análise de qualidade dos resíduos da GlcN6P desaminase feita pelo programa *Quality* (Vriend, G. e Sander, C., 1993) do *What if* (Vriend, G., 1990).

## 6.2 - DESCRIÇÃO DA ESTRUTURA DO MONÔMERO E DÍMERO

O enovelamento do monômero (figura 6.3) é do tipo  $\alpha/\beta$ , isto é, apresenta estruturas secundárias  $\alpha$  e  $\beta$  (consistente com dados experimentais obtidos por Altamirano *et al.*, 1991). A estrutura tem uma folha  $\beta$  pregueada paralela formada por sete fitas com a seguinte topologia: 4x, 1x, 1x, -3x, -1x, -1x situada no centro da estrutura envolvida em ambos os lados por 8 hélices- $\alpha$  e uma hélice  $3_{10}$  de duas voltas.

A topologia do monômero é mostrada na figura 6.4, onde as hélices são numeradas de 1 à 8 e as fitas da folha  $\beta$  paralela de A à G, sendo que a fita C tem um prolongamento no seu C-terminal, o qual faz parte de uma segunda folha  $\beta$  antiparalela de três fitas, chamadas de A', B' e C', com topologia 2, -1.

A hélice- $\alpha$  5, situada entre as fitas B' e C', é muito flexível e tem poucos contatos com o resto da estrutura. A hélice- $\alpha$  4 tem contatos com as duas folhas- $\beta$ , é bastante livre e hidrofílica, tendo os aminoácidos Cys 118 e Tyr121 como os únicos contatos com o restante da estrutura.

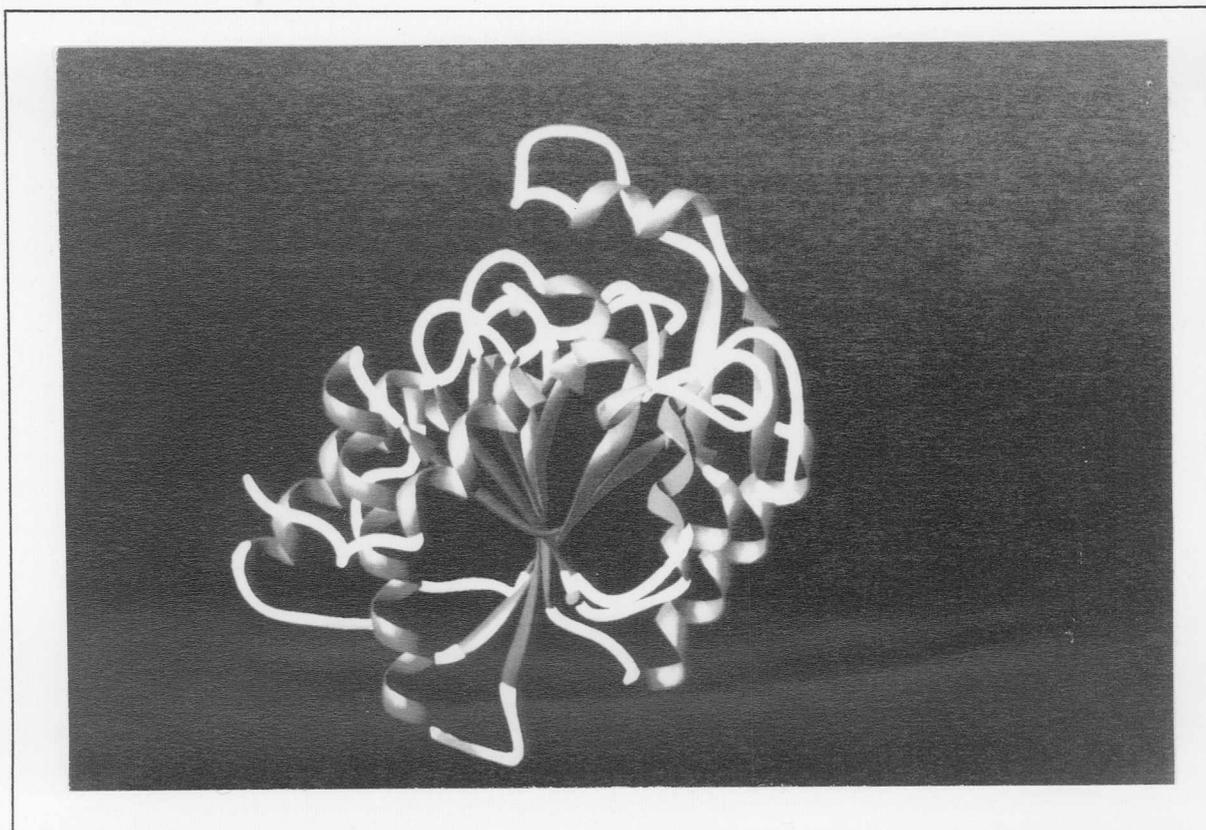


Figura 6.3 - Enovelamento do monômero da GlcN6P desaminase, mostrando as estruturas secundárias

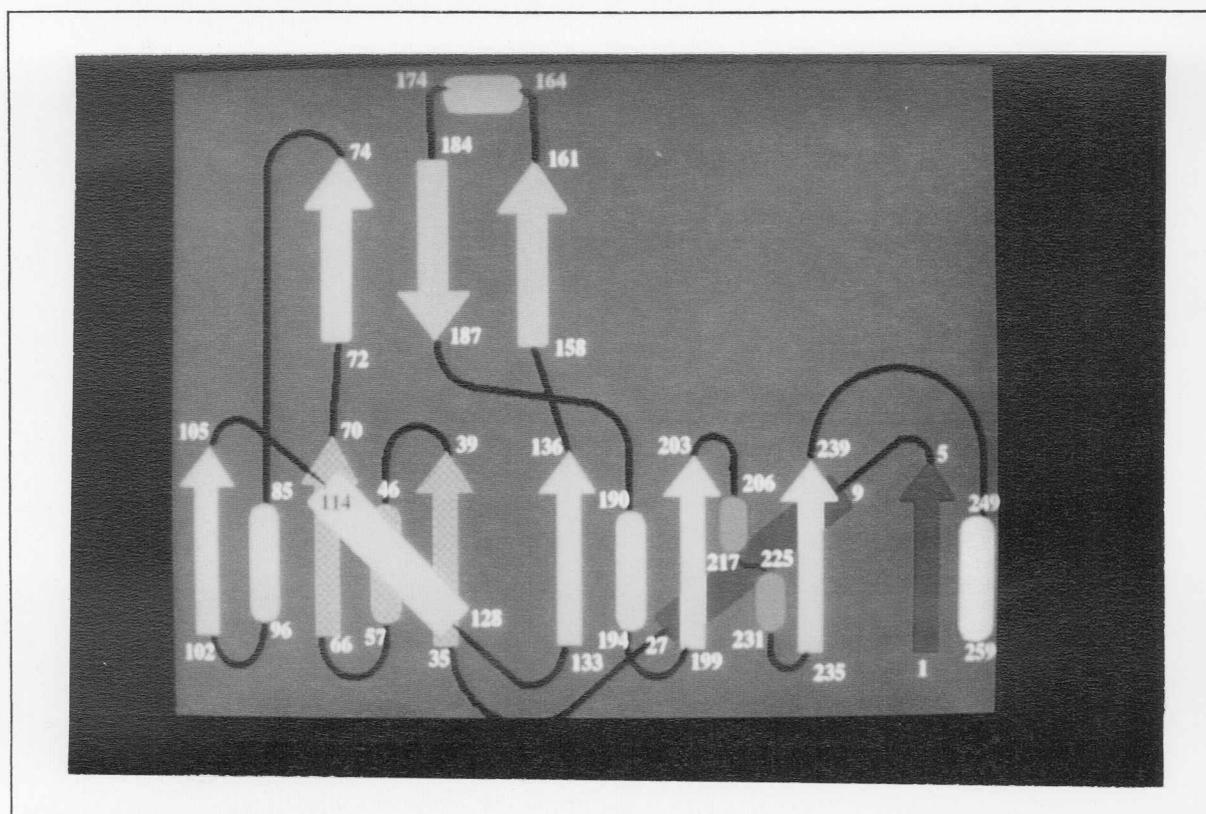


Figura 6.4 - Esquema mostrando a topologia da GlcN6P desaminase. As hélices- $\alpha$  são numeradas de 1 à 8, as fitas da folha  $\beta$  paralela de A à G e a fitas da folha  $\beta$  antiparalela de A' à C'.

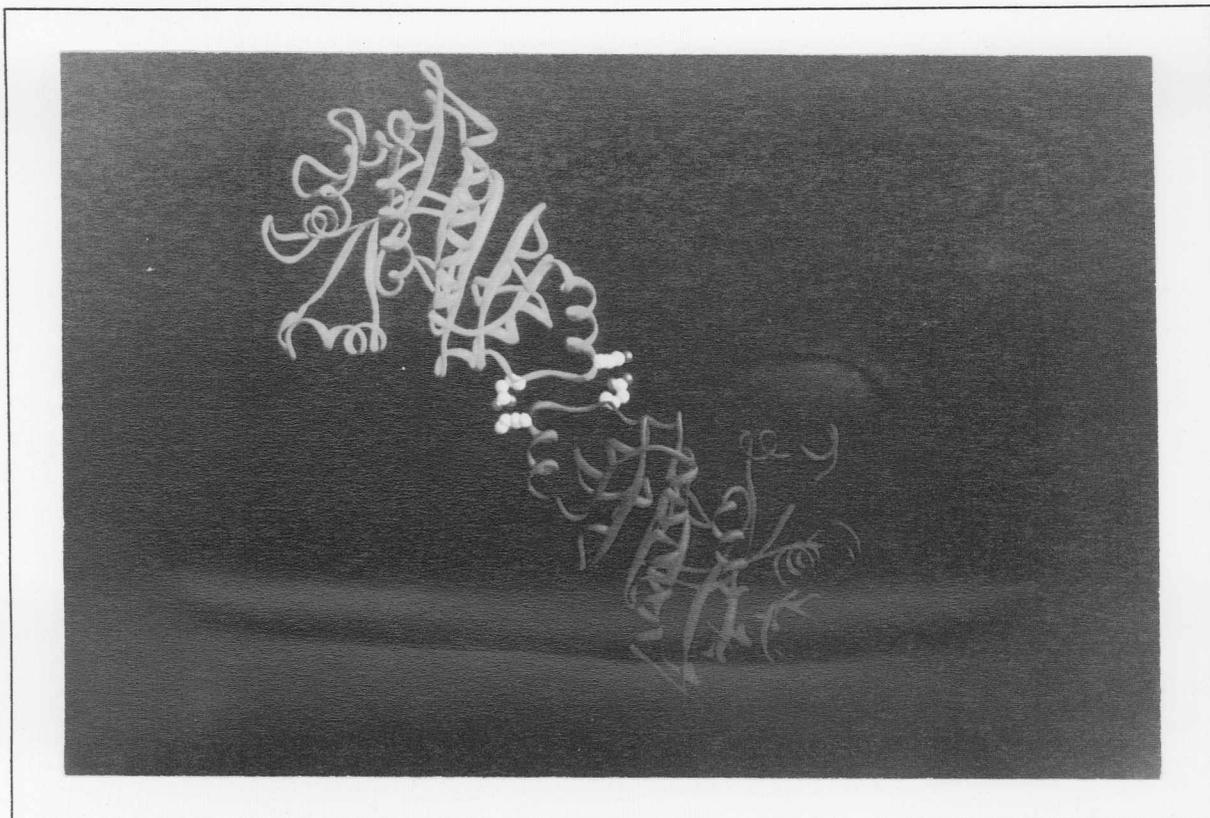


Figura 6.5 - Dímero da GlcN6P desaminase presente na unidade assimétrica.

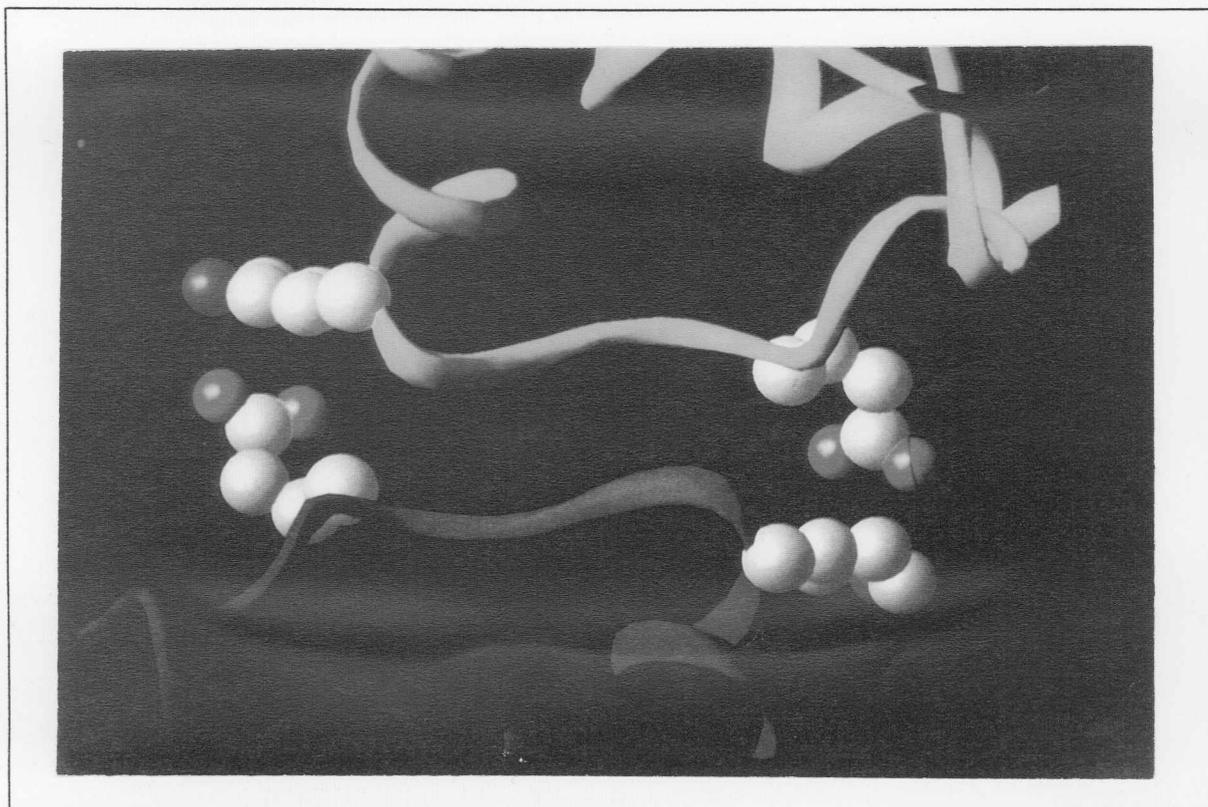


Figura 6.6 - Destaque da interface de ligação entre os dímeros da GlcN6P desaminase (as pontes salinas são feitas entre os resíduos Glu 246 e Lys 250 de ambos os monômeros).

Os dois íons fosfatos do monômero fazem as seguintes ligações: o fosfato situado na interface dos monômeros é ligado por pontes salinas aos resíduos Arg 158 e Lys 160 e ao grupo N-terminal do monômero vizinho; o outro íon fosfato é ligado por pontes salinas aos resíduos Lys 208 e Arg 172 e também faz ligações por pontes de hidrogênio com o *loop* formado pelos resíduos 40 à 44.

Na figura 6.5 é mostrada a estrutura do dímero presente na unidade assimétrica. Entre os dois monômeros está o eixo de ordem 2 não cristalográfico. Os contatos entre monômeros são feitos pelos resíduos 244 à 250 por seis ligações de hidrogênio intermonômeros. Porém, os principais contatos intermonoméricos são feitos pelas duas pontes salinas entre os resíduos Glu 246 e Lys 250 de cada um dos monômeros. A figura 6.6 mostra a interface entre dois monômeros com as pontes salinas referentes aos resíduos Glu 246 e Lys 250 em destaque.

### 6.3 - DESCRIÇÃO DA ESTRUTURA DO HEXÂMERO

O empacotamento cristalino apresenta as cadeias monoméricas agrupadas em hexâmeros (figura 6.7). As interações interhexâmeros são muito mais fracas do que as entre monômeros, formadores de hexâmeros.

A observação do agrupamento hexamérico da enzima é consistente com evidências experimentais obtidas por Calcagno *et al.*, 1984.

O hexâmero tem um eixo de ordem 3 cristalográfico (figura 6.7) e três eixos de ordem 2 não cristalográficos e perpendiculares entre si (figura 6.4), tendo a estrutura, desta maneira, uma simetria interna 32. Os dois trímeros são empacotados frente a frente com uma rotação relativa de 15° em torno do eixo de ordem 3.

O hexâmero tem uma cavidade central comunicada com o solvente externo. Desta maneira, os únicos contatos intermonoméricos estão entre o eixo de ordem 3 e entre cada um dos três eixos de ordem 2.

Próximo ao eixo de ordem 3, as três Cys 219 (de cada um dos monômeros) estão em posição apropriada para formar uma ponte dissulfeto com cada uma das correspondentes cisteínas dos outros monômeros. Observando-se a densidade eletrônica mostrada na figura 6.8, há uma grande evidência que haja uma ponte dissulfeto e um grupo SH livre, em uma distribuição aleatória ao longo das moléculas do cristal, preservando a simetria cristalográfica de ordem 3 (figuras 6.9 e 6.10).

A existência desta estrutura trigonal envolvendo as três Cys 219 é consistente com dados experimentais prévios, nos quais foram feitos estudos químicos dos grupos sulfidril em GlcN6P desaminase e mostraram que na titulação de tióis na presença de desnaturantes somente é detectada as cisteínas 118, 228 e 239 (Altamirano *et al.*, 1989 e 1992). Ou seja, o grupo sulfidril da Cys 219 comporta-se como se estivesse oxidado.

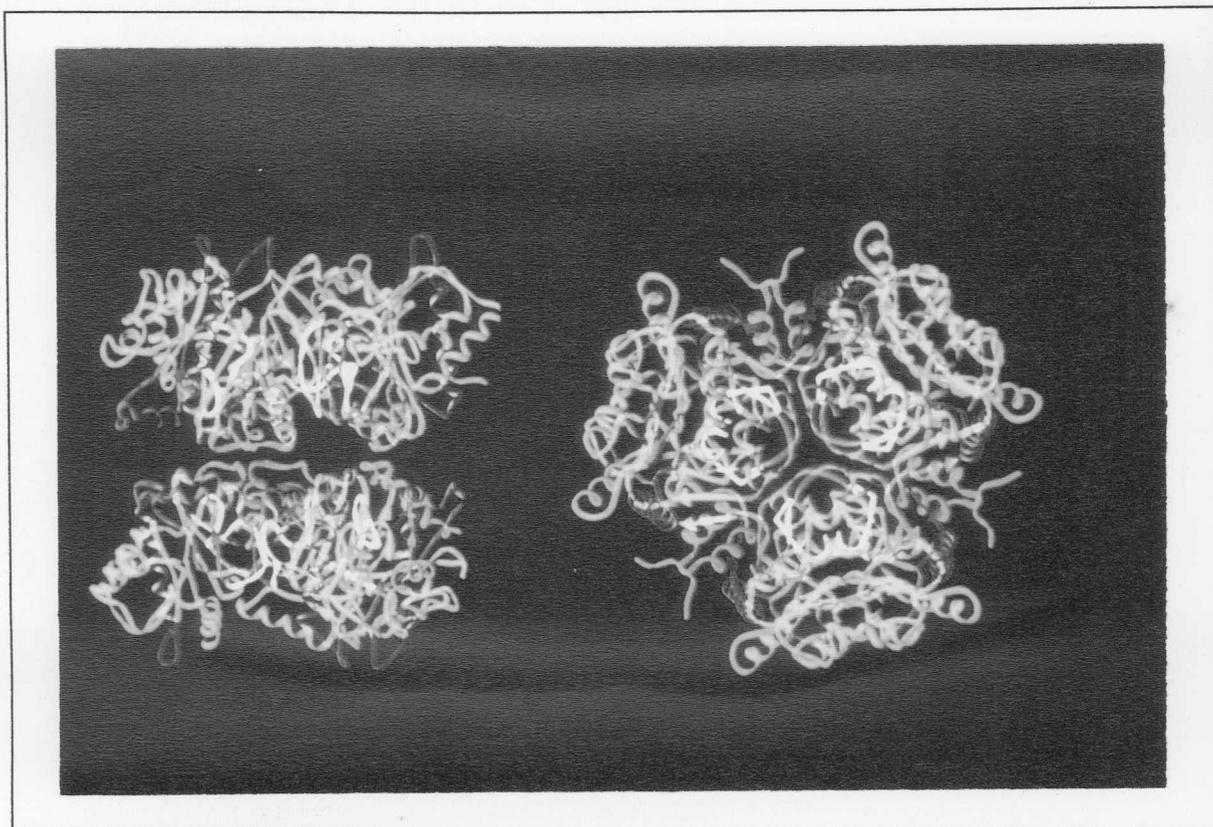


Figura 6.7 - Hexâmero da GlcN6P desaminase. À direita é mostrada a vista de cima e à esquerda a vista lateral.

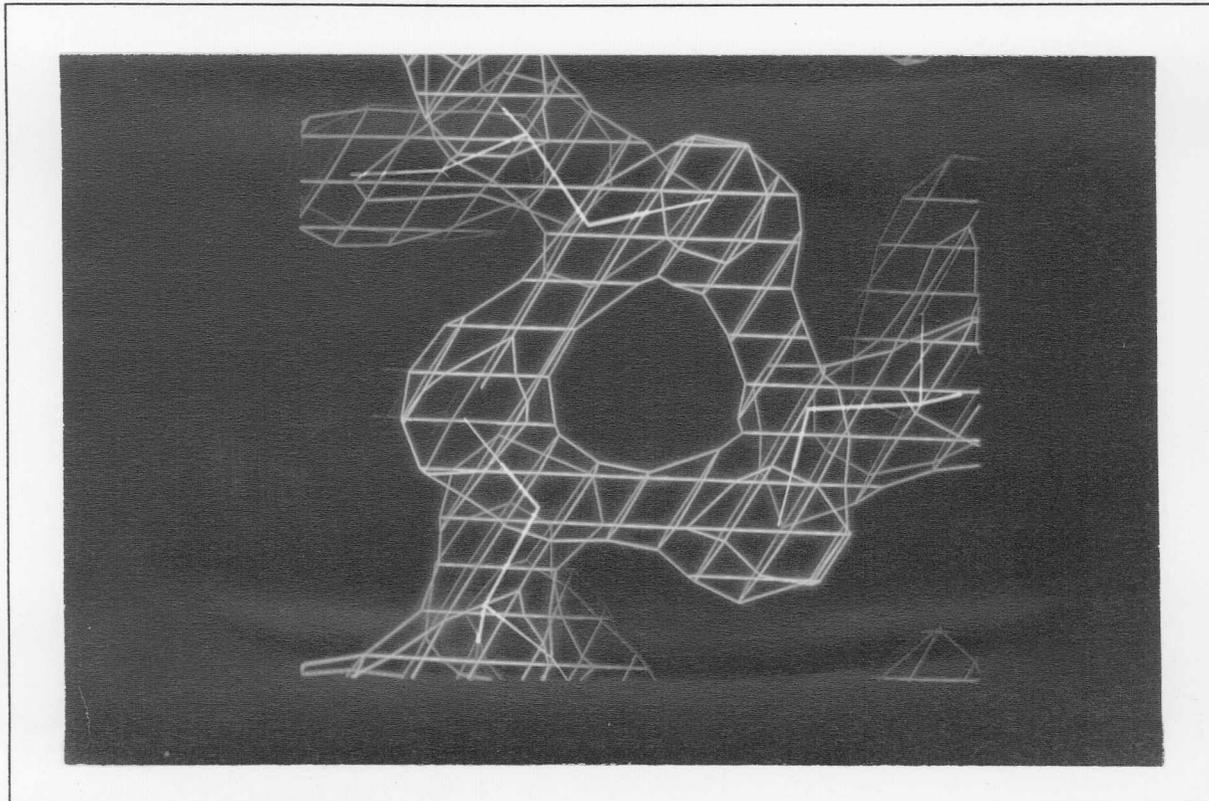


Figura 6.8 - Densidade eletrônica mostrando o resíduo Cys 219 de cada um dos monômeros do trîmero formando uma possível ponte dissulfeto e SH livre em uma distribuição aleatória ao longo das moléculas do cristal.

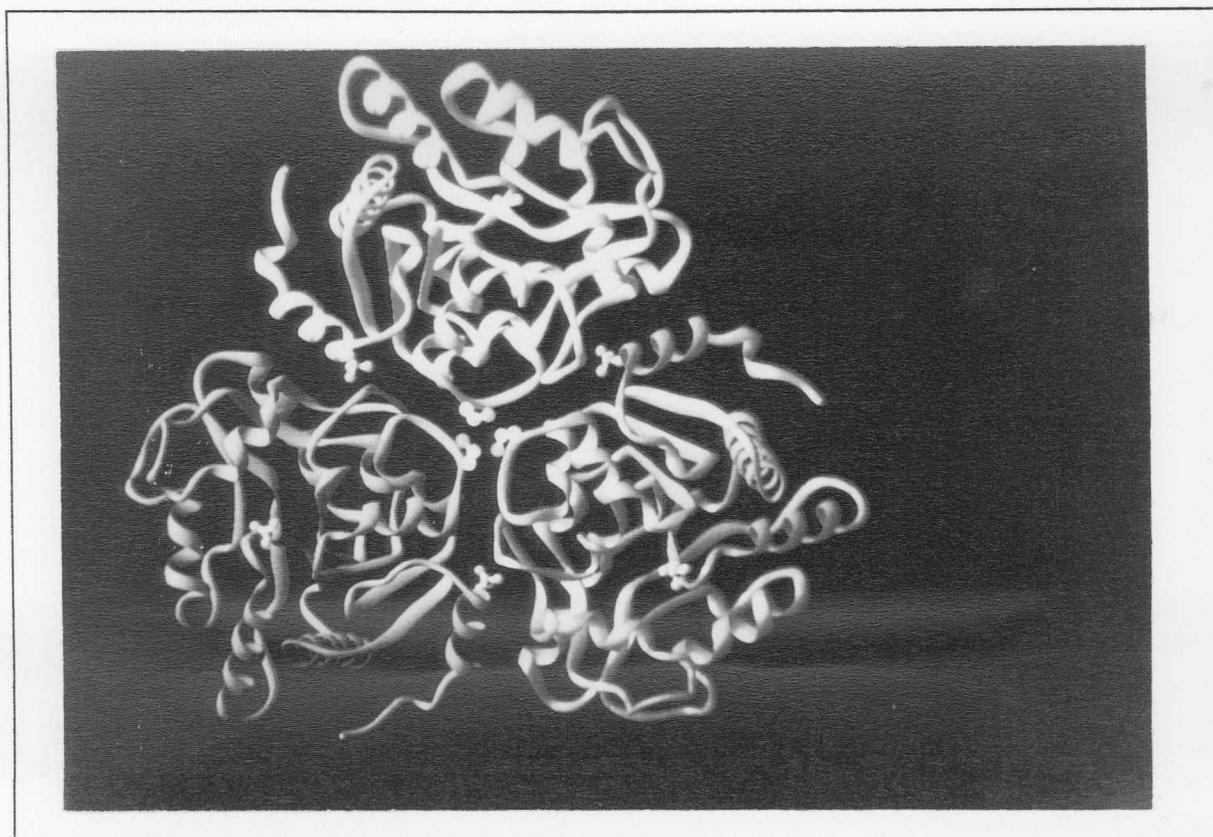


Figura 6.9 - Representação do trímero da GlcN6P desaminase, mostrando a ponte dissulfeto entre estes e os íons fosfato (dois por monômero).

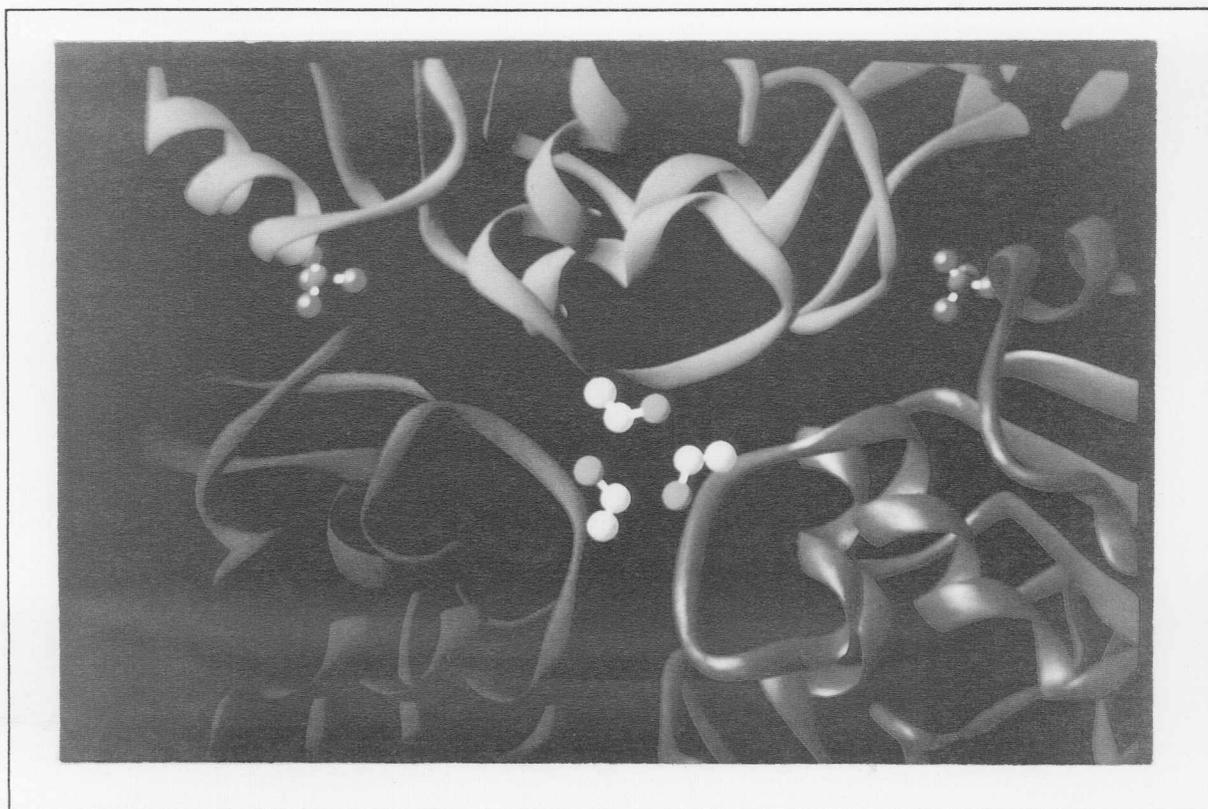


Figura 6.10 - Destaque para a interseção entre os monômeros formadores do trímero da GlcN6P desaminase, mostrando a ponte dissulfeto entre as três Cys 219 e os íons fosfato na interface entre monômeros.

Todavia, segundo Altamirano *et al.* (1992), o arranjo particular das Cys 219 não tem importância fundamental para a estabilidade do hexâmero ou para sua função (veja capítulo 1).

Outros contatos intermonômeros próximos ao eixo de ordem 3, incluem interações entre resíduos do *loop* 216-223 o qual contém a Cys 219 e os resíduos 230-232 da hélice  $3_{10}$ . Estas interações envolvem ligações de hidrogênio e contatos de van der Waals, os quais estabilizam ambas as metades do hexâmero (os trimeros).

## 6.4 - DESCRIÇÃO DA ESTRUTURA DOS COMPLEXOS CO-CRISTALIZADOS. OS SÍTIOS CATALÍTICOS E ALOSTÉRICOS

Os refinamentos das estruturas dos complexos co-cristalizados mostraram que, dentro da margem de erro experimental, as suas cadeias polipeptídicas têm a mesma estrutura da proteína nativa, discutida nos itens 6.2 e 6.3. Sendo assim, deve-se analisar apenas como se deram as ligações dos complexos de ativador alostérico e inibidor à enzima. No apêndice C são dados os gráficos de análise de qualidade estereoquímica feitos pelo programa PROCHECK (Moris, A.L. *et al.*, 1992) para os complexos. Estes gráficos são bastantes similares com os do modelo nativo.

### 6.4.1- Complexo enzima ativador-Pi : o sítio alostérico

O ativador alostérico se ligou na interface dos monômeros, com o fosfato situado na mesma posição do íon inorgânico fosfato presente na estrutura da proteína nativa, enquanto que o outro sítio de ligação de fosfato da proteína nativa continuava da mesma maneira, isto é, somente com o íon fosfato.

Os principais contatos do ativador alostérico com a enzima são feitos pelos átomos de oxigênio do grupo fosfato ( $-OPO_3H_2$ ), entre os quais com os resíduos 151 e 152 e o segmento de 158 à 160. Há também, contatos do ativador com o monômero vizinho, que incluem uma ponte salina entre o átomo O5 do ativador com Leu 258 e do N-terminal com o grupo fosfato. O ativador faz também, várias pontes de hidrogênio com moléculas de água. A tabela 6.1 mostra os principais contatos feitos pelo ativador.

A figura 6.11 mostra uma representação do ativador alostérico entre os monômeros da enzima com os aminoácidos mais importantes para sua ligação.

Por evidências bioquímicas, mostrou-se que GlcNAc6P (ativador alostérico) é um ligante exclusivo da enzima no estado R (veja capítulo 1), assim pode se inferir que o sítio de ligação de GlcNAc6P com esta seja o sítio alostérico (veja discussão no item 7.1).

Tabela 6.1 - Principais contatos entre o ativador alostérico e a enzima

	Átomo	Resíduo	Distância(Å)
OP1	N(1)	Cad. principal	2.6
	N(152)	Cad. principal	3.5
	OH2(120)	Água	3.8
	OH2(114)	Água	2.6
	OH2(5)	Água	3.3
OP2	N(152)	Cad. principal	3.5
	OG(151)	Ser	3.8
	NZ(160)	Lys	2.9
	NH2(158)	Arg	3.4
OP3	NE(158)	Arg	2.7
	OH2(5)	Água	2.7
O6	N(1)	Cad. principal	2.9
O5	CD2(258)	Leu	3.9
	OH2(11)	Água	2.9
O3	O(159)	Cad. principal	3.3
O1	CG1(159)	Cad. principal	3.6

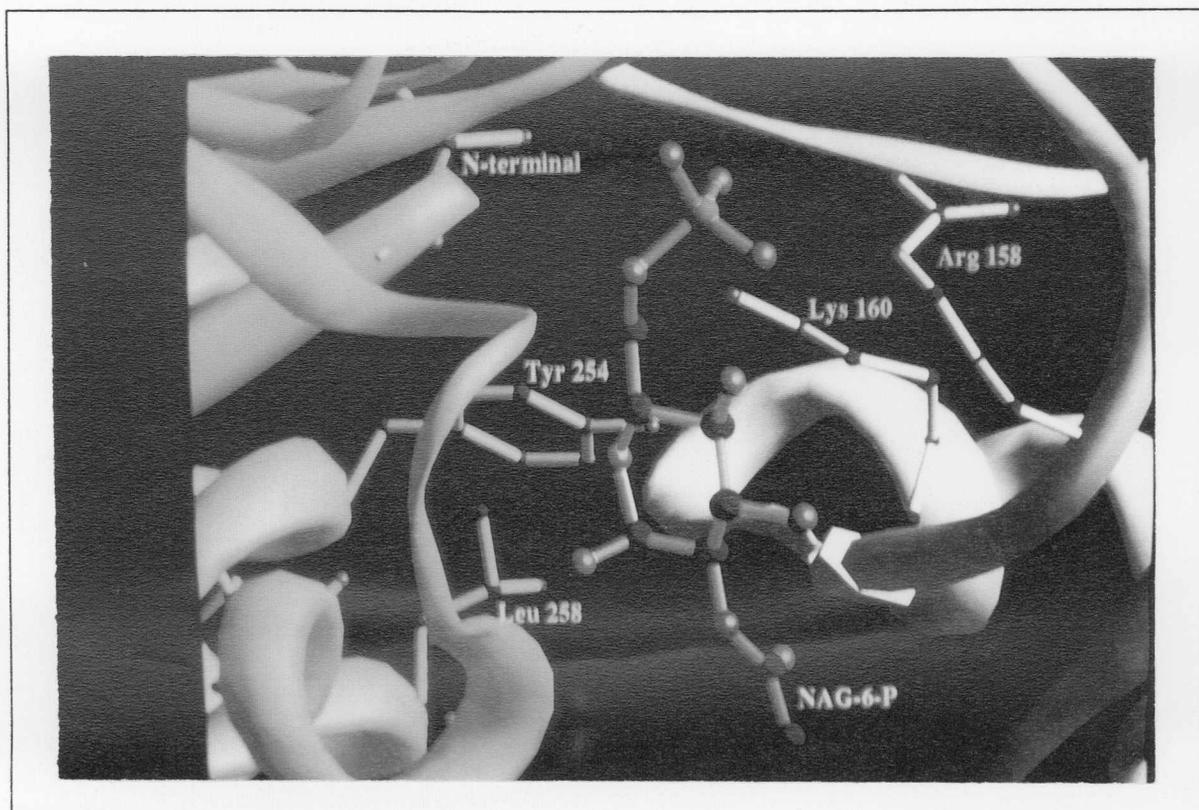


Figura 6.11 - Representação do ativador alostérico e seu ambiente de ligação, destacando os aminoácidos mais importantes na sua ligação.

#### 6.4.2- Complexo enzima inibidor-Pi (cristais co-cristalizados com precipitante fosfato): o sítio catalítico

O inibidor utilizado para a co-cristalização, o 2-desoxi-2-amino-glucitol-6-fosfato (GlcN-ol-6P), tem uma constante de inibição que é três ordens de magnitude mais baixa do que a constante de ligação para os substratos GlcN6P ou Fru6P, ou seja, a enzima tem uma afinidade muito maior pelo inibidor do que pelo substrato. Desta maneira, esta alta afinidade da enzima por GlcN-ol-6P, pode ser considerada como uma evidência para sua similaridade estrutural com estado intermediário de transição ocorrido na reação de catálise, podendo assim auxiliar na elucidação deste mecanismo.

A estrutura mostra o inibidor ligado no interior dos monômeros, com o seu íon fosfato situado na mesma posição do íon fosfato presente na estrutura nativa, enquanto no outro sítio de ligação de fosfato (o que está na interface), havia somente a presença do íon fosfato inorgânico, ou seja, houve ligação do inibidor em apenas um dos sítios (no interior do monômero). Este é possivelmente o sítio catalítico da enzima (veja discussão no item 7.1), permitindo-se assim, inferir univocamente que o sítio com a presença do inibidor é o sítio catalítico ou ativo.

Tabela 6.2 - Principais contatos entre o inibidor e a enzima

	Átomo	Resíduo	Distância(Å)
OP1	NH2(172)	Arg	3.0
	N(43)	Cad. principal	2.9
OP2	NZ(208)	Lys	2.9
	OG1(44)	Thr	3.2
OP3	OG1(44)	Thr	2.7
	N(44)	Cad. principal	3.2
C6	NZ(208)	Lys	3.3
	O(138)	Cad. principal	3.2
O5	NE2(143)	His	2.8
	CE1(143)	His	3.6
O4	O(41)	Cad. principal	3.7
	CA(137)	Cad. principal	3.5
O3	O(137)	Cad. principal	3.5
N2	O(41)	Cad. principal	3.6
	OH(85)	Tyr	3.8
C1	O(41)	Cad. principal	3.0
	N(41)	Cad. principal	3.3
	CB(40)	Cad. principal	3.4
O1	OG1(41)	Thr	3.3
	N(41)	Cad. principal	2.7
	OD1(72)	Asp 72	3.5

O inibidor situa-se entre dois *loops* dos C-terminal das fitas B e E da folha- $\beta$  paralela. Os principais contatos do inibidor são dados na tabela 6.2, os quais incluem vários contatos com os átomos de oxigênio do fosfato. O *loop* composto dos resíduos 41 à 44, situado entre as fitas B e a hélice- $\alpha$  2, forma ligações de hidrogênio com o fosfato do inibidor. Outras interações do inibidor com a enzima envolvem os resíduos Arg 172, Lys 208 e Thr 44.

Outro contato possivelmente importante é o do O5 do inibidor com o NE2 da His 143, pois este aminoácido faz contato com os Asp 141 e Glu 148 em um arranjo similar a um relê de prótons comum em muitas proteínas (Fink, A.L., 1987).

A figura 6.12 mostra uma representação do inibidor e seu ambiente de ligação com a enzima, bem como os aminoácidos mais importantes para a sua ligação.

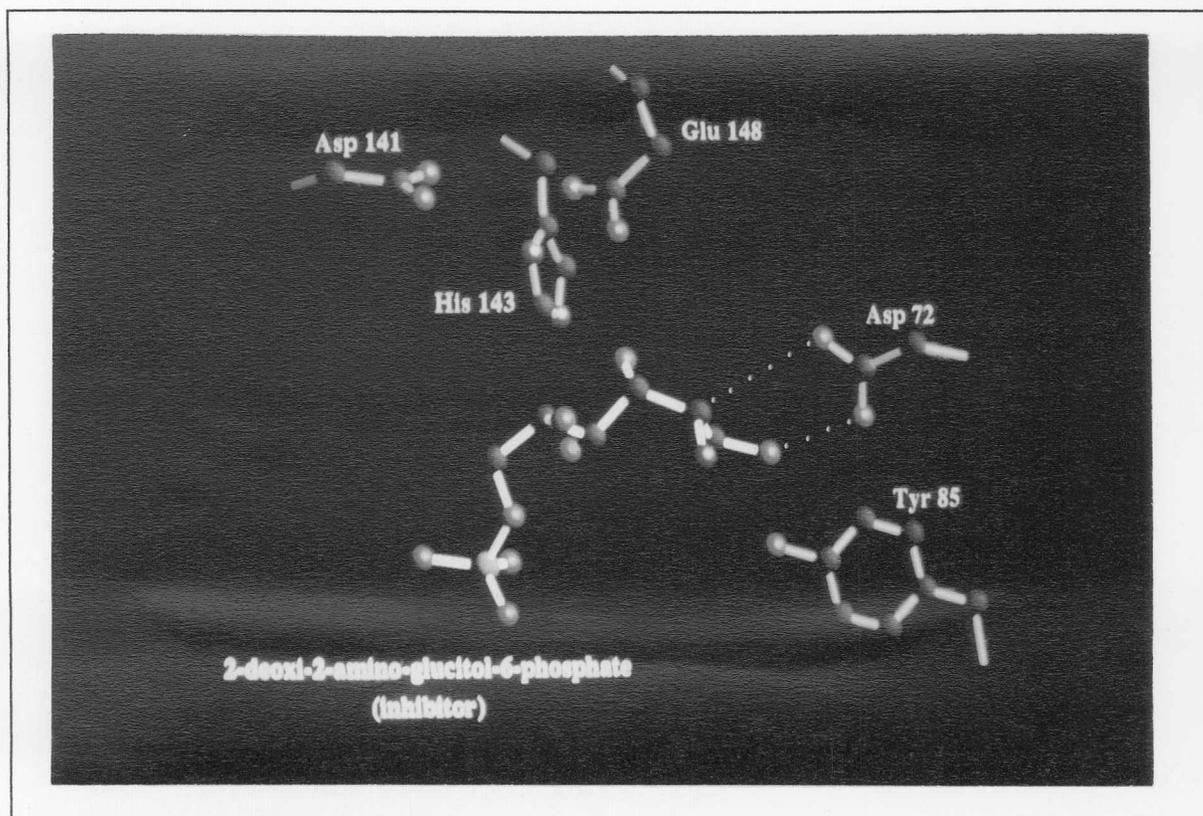


Figura 6.12 - Representação do inibidor e seu ambiente de ligação

### **6.4.3- Complexo enzima 2-inibidor (cristais co-cristalizados com precipitante acetato de sódio)**

A estrutura da enzima co-cristalizada com inibidor em precipitante acetato de sódio (ou seja, na ausência de íons fosfato na solução) apresenta ambos os sítios onde haviam íons fosfatos na estrutura nativa ocupados pelo inibidor (sítios alostéricos e catalíticos), porém com conformações diferentes nestes sítios.

No sítio catalítico a conformação é a mesma do complexo descrito no item anterior (item 6.4.2), porém no sítio alostérico (na interface dos monômeros) esta conformação torna-se de alguma maneira mais parecida com a do ativador alostérico, apesar da cadeia do inibidor ser aberta e a do ativador alostérico ser fechada. Assim, os contatos do inibidor no sítio alostérico e a enzima são essencialmente os mesmos descritos na tabela 6.1.

Como a estrutura da enzima mantém uma conformação praticamente idêntica em todas as estruturas descritas neste trabalho, as quais seus cristais foram crescidos em precipitante fosfato Na/K ou acetato de sódio, pode-se concluir que a configuração obtida em todas estas estruturas é independente das condições de cristalização.

## CAPÍTULO 7

### CONCLUSÕES

A resolução de uma nova estrutura protéica à nível atômico, que não tem homologia significativa com outra proteína e não tem características funcionais bem determinadas, é apenas o passo inicial para uma série de conclusões que podem ser obtidas. Para tanto, pode-se simplesmente fazer um cuidadosa análise da estrutura, ou fazer novos experimentos utilizando cristalografia, aliada ou não a outras técnicas experimentais, bem como pode-se usar modelagem molecular a partir das estruturas conhecidas.

No caso da GlcN6P desaminase, os mais eminentes interesses são os estudos dos mecanismos de alostericidade e catalítico, para os quais a resolução das estruturas nativa e dos complexos descritos neste trabalho são de grande importância. Algumas conclusões obtidas por nós a partir da análise das estruturas são dadas abaixo.

#### 7.1 - EVIDÊNCIAS QUE A ESTRUTURA NATIVA (ENZIMA 2-P<sub>i</sub>) E DOS COMPLEXOS ESTEJAM NO ESTADO CONFORMACIONAL R: OS SÍTIOS ALOSTÉRICO E CATALÍTICO

Como descrito no capítulo 1, os experimentos de cinética de reação evidenciam que a GlcNAc6P é um ativador de ligação exclusiva à GlcN6P desaminase, sendo responsável pelo mecanismo de alostericidade conformacional desta, de maneira que, quando se dá a ligação de GlcNAc6P à enzima, esta adquire uma conformação alostérica R (Midelfort e Rose, 1977; Calcagno *et al.*, 1984).

Assim, ao se obter um mapa Fourier diferença com uma densidade eletrônica na qual pode-se construir com relativa facilidade uma molécula de GlcNAc6P, é bem razoável supor que a estrutura apresentada nesta tese para os cristais crescidos a partir de proteína e ativador (enzima ativador-P<sub>i</sub>), esteja na conformação alostérica R e além disso, o local de ligação da GlcNAc6P a enzima seja o sítio alostérico.

Quanto ao inibidor GlcN-ol-6P, demonstrou-se por meio de experimentos de cinética de reação (item 1.2) que este comporta-se como um ligante homotrópico a enzima, causando mudança conformacional desta para o estado R, mesmo a baixas concentrações (note que a concentração do GlcN-ol-6P na solução para cristalização era muito baixa, 0.45 mM, dez vezes menor que a utilizada para o ativador). Similarmente ao modelo enzima ativador-P<sub>i</sub>, pode-se concluir que a estrutura apresentada para os cristais de enzima e GlcN-ol-6P está na conformação alostérica R.

Deve-se notar ainda que, apesar da baixa concentração de inibidor frente aos íons fosfato no complexo co-cristalizado com tampão fosfato Na/K, houve ligação em apenas um dos sítios (ficando o outro ocupado pelo íon fosfato) do monômero, demonstrando assim a grande afinidade deste por tal sítio. Por estas evidências, pode-se supor que este sítio (o sítio ocupado pelo inibidor nos cristais co-cristalizados com tampão fosfato Na/K) é o sítio catalítico da enzima GlcN6P desaminase.

Como os cristais co-cristalizados com ativador e inibidor (crescidos em fosfato Na/K ou acetato de sódio) tinham os mesmos parâmetros de rede da proteína nativa e como os modelos construídos a partir destes são idênticos dentro da margem de erro experimental (exceto na vizinhança dos sítios catalíticos e alostéricos, onde há pouquíssimas diferenças) pode-se inferir que o modelo para a proteína nativa (enzima 2-Pi) também está na conformação alostérica R ou em uma conformação bem próxima desta.

Foi concluída recentemente pelo Dr. Eduardo Horjales a estrutura da GlcN6P desaminase de cristais crescidos em agente precipitante acetato de sódio, os quais pertencem ao grupo espacial P6<sub>3</sub>22.

Esta estrutura não apresenta qualquer íon ou ligante nos sítios alostérico e catalítico. Desta maneira, pode-se inferir que a estrutura está provavelmente no estado conformacional T (veja capítulo 1, p. 6). Tal modelo, apresenta grandes mudanças conformacionais em relação aos obtidos nesta tese: enzima-2Pi, enzima ativador-Pi, enzima inibidor-Pi e enzima 2-inibidor, principalmente na região do sítio alostérico.

## 7.2 - PROPOSIÇÃO DE UM MECANISMO CATALÍTICO

Um possível mecanismo para a GlcN6P desaminase é proposto a partir do mecanismo geral para isomerização aldose-cetona (Fersht, 1985, p.227), dos dados cinéticos apresentados por Midelfort e Rose (1977) e das estruturas descritas nesta tese.

Três fatos contribuem para a proposição do modelo:

- O inibidor competitivo 2-desoxi-2-amino-glucitol-6-fosfato (GlcN-ol-6P), é análogo a forma aberta do substrato GlcN6P (a diferença está na presença de uma hidroxila do carbono C1 no inibidor, em vez de um aldeído em C1 no substrato).
- Como citado no item 6.4.2, GlcN-ol-6P tem uma constante de inibição que é três ordens de magnitude mais baixa do que a constante de ligação para os substratos GlcN6P (forma fechada) ou Fru6P, ou seja, a enzima tem uma afinidade muito maior pelo inibidor do que pelo substrato.

As duas observações acima indicam que o sítio ativo tem uma alta afinidade pela forma aberta da GlcN6P.

- Segundo Midelfort e Rose (1977), o substrato GlcN6P se liga a enzima GlcN6P desaminase na forma piranose cíclica, no enantiômero  $\alpha$ .

Assim, pode-se inferir que a seqüência de eventos que ocorrem no mecanismo de catálise incluem a abertura do anel piranose da  $\alpha$ -D-GlcN6P e posteriormente a reação de desaminação e isomerização que ocorrem após a abertura do anel piranose do substrato.

A partir da estrutura do GlcN-ol-6P foi feito um modelo da estrutura do substrato na sua conformação fechado no sítio ativo (figura 7.2). O modelo mostra o contato do átomo O5 do substrato do anel piranose com o N $\epsilon$ 2 da His 143 (resíduo carregado positivamente) (Figura 7.2). Este resíduo pode ser importante na abertura do anel do substrato que é o primeiro passo na reação catalítica (figura 7.1). O modelo mostra ainda, o contato do grupo amina da GlcN6P com o resíduo Asp 72; este arranjo pode estabilizar a carga negativa entre o Asp 72 e o grupo amino do substrato. Com base nas informações dadas acima, é proposto o seguinte mecanismo catalítico (veja figura 7.1, passos de I à IX):

1. O anel piranose é aberto a partir do átomo O5 do substrato (I), com a doação de um próton da His 143 para o átomo O5. Concomitantemente, é formado um aldeído a partir do C1 (II).
2. O substrato se liga com seu grupo amino protonado (II) pelo contato com Asp 72.
3. O GlcN6P é enolisado pela captura de um átomo de hidrogênio do átomo C2 como um próton, pela Asp 72 (III).
4. É formado um intermediário *cis*-enol-amônio pela transferência de um próton de Asp72 para o oxigênio do átomo C1 (IV).
5. A ligação  $\pi$  de C1-C2 do *cis*-enol-amônio é atacada por uma molécula de água vinda da *re*-face do intermediário, isto é, do mesmo lado onde está o Asp 72. O novo intermediário formado (V) tem uma estrutura que é similar à do inibidor competitivo GlcN-ol-6P e provavelmente é muito instável (a diferença está na presença de uma hidroxila ligada ao átomo C2).
6. O intermediário formado anteriormente perde o próton da hidroxila em C2 que é novamente capturado pelo Asp 72 (VI). Ao mesmo tempo, o grupo amônia é perdido como amônia (NH<sub>3</sub>)(VII). Evidências cinéticas indicam que o íon em vez da amônia desprotonada é o verdadeiro substrato da reação reversa da GlcN6P desaminase (Midelfort e Rose, 1977), nós propomos que a Asp 72 doa este próton (VIII).
7. A His 143 age como uma base geral catalítica para a formação de Fru6P (IX).

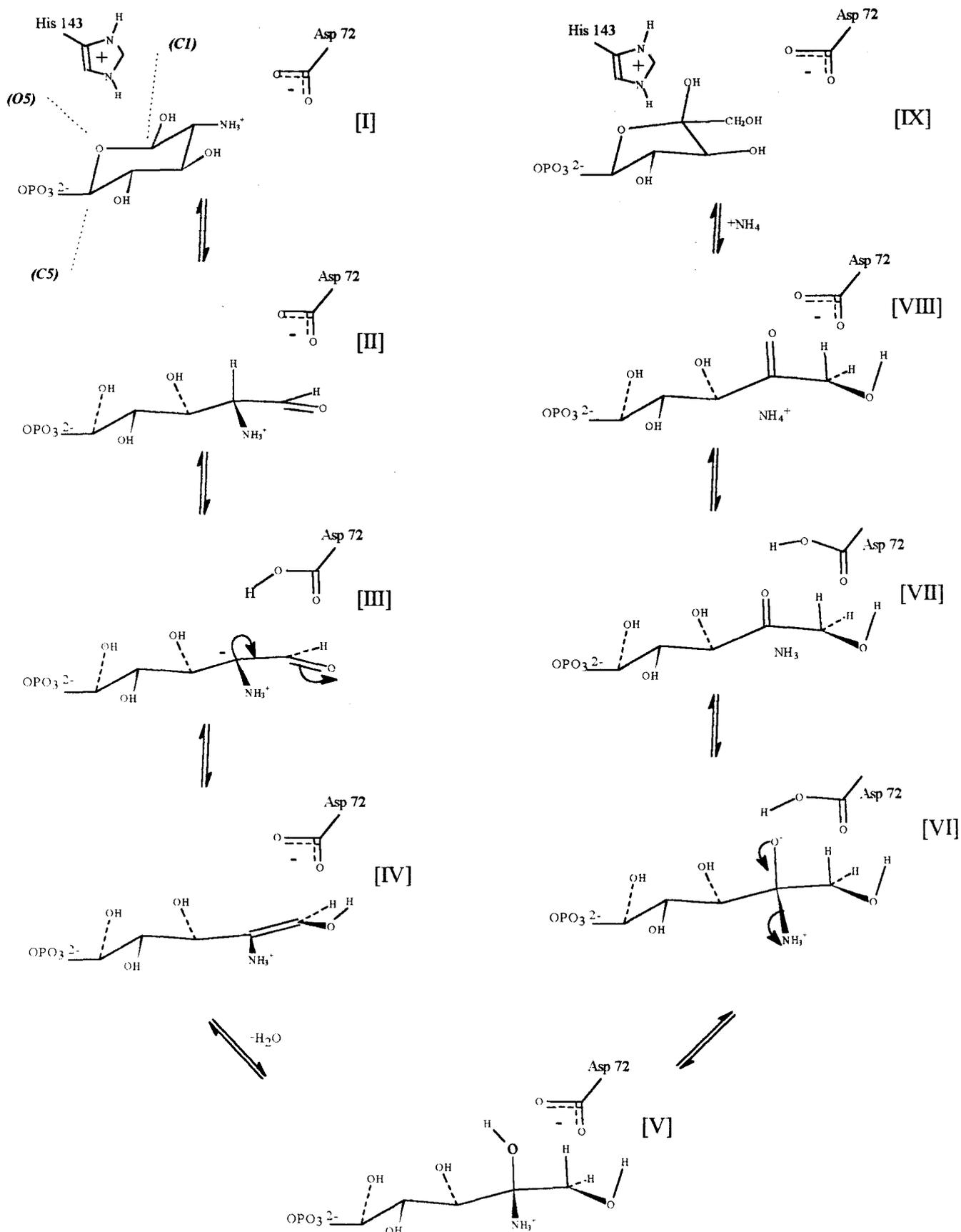


Figura 7.1 - Mecanismo catalítico proposto para a Glucosamina 6-fosfato desaminase

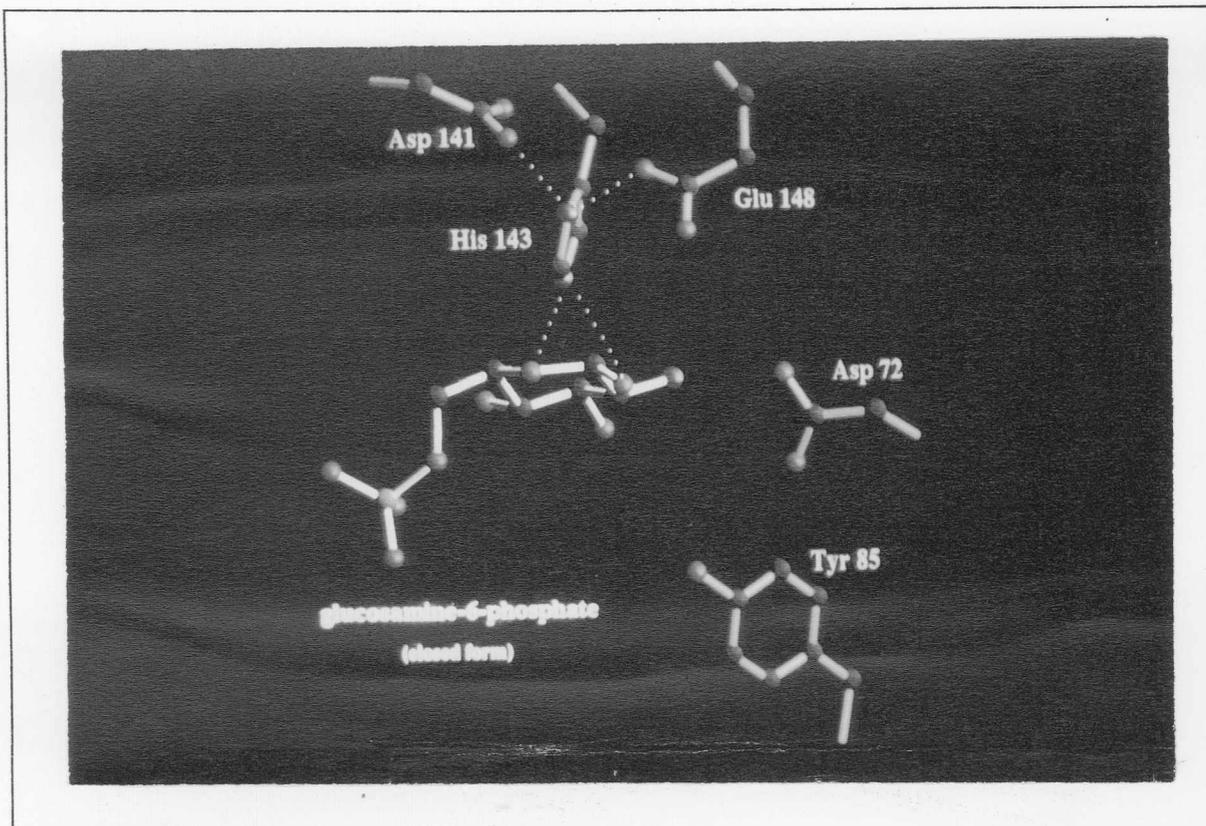


Figura 7.2 - Modelo para o substrato GlcN6P na forma de anel piranose no sítio catalítico da enzima, construído a partir da estrutura do inibidor GlcN-ol-6P

### 7.3 MODELO DE MONOD-WYMAN-CHANGEUX (MWC) E SISTEMA TIPO K

A estrutura resolvida está em concordância com o modelo MWC (Monod *et al.*, 1965), pois as suas características estruturais estão de acordo com este, tais como: é um oligômero, possui os estados conformacionais T e R, tem todos os sítios de ligação equivalentes (os seis sítios catalíticos do hexâmeros são equivalentes entre si e os seis alostéricos são equivalentes entre si).

As características estruturais da enzima implicam em um mecanismo de mudanças alostéricas de um sistema do tipo K (veja capítulo 1, p.6) para a GlcN6P desaminase, em concordância com dados obtidos por experimentos bioquímicos (cinética de reação catalítica, capítulo 1).

As características estruturais são:

1. O posicionamento do sítio alostérico na interface das subunidades significa que a ligação do ativador é equilibrada por mudanças na estrutura quaternária pois, uma mudança na estrutura quaternária significa que as subunidades se movem uma contra a outra (em sistemas do tipo K, a ligação de moléculas do ativador altera a afinidade da enzima pelo substrato devido às mudanças conformacionais provocadas por estas).

2. A localização dos resíduos catalíticos dentro das subunidades, significa que o valor intrínseco de  $k_{cat}$  não é afetado por mudanças na estrutura quaternária (em sistemas tipo K ambos os estados R e T têm o mesmo valor intrínseco de  $k_{cat}$ ).

#### 7.4 - COMPARAÇÕES COM OUTRAS ESTRUTURAS

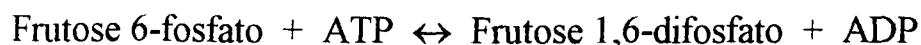
Quando uma nova estrutura proteica é resolvida, é muito importante a comparação desta com outras que tenham características semelhantes, sejam em termos estruturais ou em termos funcionais. Tais comparações, situam o novo modelo dentro do contexto da natureza, onde se sabe que durante a evolução das espécies, a maioria das características foram conservadas e outras foram 'aperfeiçoadas'. É útil também, para confirmar características já conhecidas ou propor novas.

A interessante ponte de dissulfeto e SH livre (das Cis 219) presente próxima ao eixo de ordem três da GlcN6P desaminase é, pela literatura pesquisada por nós, única até o momento, porém algo similar é encontrado no vírus da febre aftosa (Acharya, *et al.*, 1989), a qual tem próximo ao eixo de ordem cinco um arranjo semelhante de cisteínas, com cinco neste caso.

A topologia do monômero da GlcN6P desaminase tem um padrão de enovelamento semelhante ao da estrutura ligadora de dinucleotídeo, o NAD *binding domain* (domínio de ligação de NAD), com três inserções (na GlcN6P desaminase); no N-terminal (aminoácidos 1-35), C-terminal (243-266) e no segmento 143-188 (Creighton, 1993, p.250). Apesar deste fato, não há nenhuma similaridade seqüencial entre a GlcN6P desaminase e o NAD *binding domain*.

A figura 7.3 denota a parte do monômero da GlcN6P desaminase que tem enovelamento similar ao NAD *binding domain*, esta parte é mostrada em amarelo na figura 7.3. Em verde é mostrado o N-terminal (1-35), em vermelho o C-terminal (243-266) e em azul o segmento 143-188.

Em termos de mecanismos de alostericidade, a GlcN6P desaminase tem características semelhantes com a enzima fosfofrutoquinase (Fersht, 1985, p.283). Esta enzima também faz parte do caminho metabólico, onde se sabe que várias enzimas chaves são reguladas pelo controle de sua atividade. A fosfofrutoquinase catalisa a conversão:



e tem sido estudada em vários organismos, tais como os eucariontes: *Escherichia coli* e *Bacillus stearothermophilus*. As estruturas cristalográficas da fosfofrutoquinase de *B. stearothermophilus* e de *E. coli* foram resolvidas à 2.4 Å de resolução. Ambas são muito semelhantes e são formadas por um tetrâmero,

com peso molecular de aproximadamente 35000 *Da* cada um dos monômeros (Evans *et al.*, 1981 e Shirakihara e Evans, 1988). Estudos de cinética de reação mostraram que a enzima é qualitativamente consistente com o modelo de MWC, tendo dois estados R e T, e age como um sistema do tipo K. A enzima foi cristalizada com a presença de Fru6P, ADP e AMP-PNP (um análogo de ADP, estável sob condições de cristalização), para ambas as bactérias.

As estruturas da enzima possuem quatro sítios de ligação do substrato (ATP+Fru6P) no interior dos monômeros e quatro sítios alostéricos nas interfaces dos monômeros (ligação de ADP). Ambas as estruturas estão presumivelmente no estado R (pois ligam o ativador alostérico ADP).

As estruturas descritas acima possuem características bastante semelhantes à GlcN6P desaminase, principalmente com respeito ao seu mecanismo de alostericidade. Desta maneira, estes resultados reforçam as suposições feitas no trabalho apresentado nesta tese (itens 7.1 e 7.3), mostrando que este mecanismo talvez seja utilizado por várias enzimas, as quais devem ter sua atividade regulada alostericamente.



Figura 7.3 - Monômero da GlcN6P desaminase. A parte mostrada em amarelo tem enovelamento similar ao NAD *binding domain*, em verde é mostrado o N-terminal (1-35), em vermelho o C-terminal (243-266) e em azul o segmento 143-188.

## CONSIDERAÇÕES FINAIS

Nesta dissertação procurou-se abordar um problema novo no Brasil, a resolução a nível atômico de uma macromolécula proteica, de alguma maneira abrindo portas a trabalhos que estão sendo desenvolvidos no momento e que virão a ser feitos.

Pode-se se dizer que o trabalho teve êxito, que deve ser atribuído em grande parte aos pesquisadores que nos forneceram as amostras da enzima purificada e em grande quantidade (suficiente para as várias tentativas que foram feitas).

Com os experimentos realizados, foi possível obter derivados isomorfos que difratam a razoável resolução, permitindo obter um bom modelo onde foi possível visualizar todos os aminoácidos da proteína e desta maneira, analisar as características do monômero e do hexâmero da enzima. Obteve-se cristais dos complexos proteína+ativador e proteína+inibidor (que difratam até 2.1 Å) sendo possível determinar suas conformações tridimensionais e localizar os sítios alostéricos e catalíticos do hexâmero e onde são as ligações mais importantes da proteína com estes ligantes. É proposto ainda, um possível mecanismo catalítico de atuação da enzima. Algumas comparações com outras estruturas anteriormente resolvidas, reforçam a veracidade dos dados e conclusões obtidas.

Como planos para a próxima etapa, destacamos os mais importantes trabalhos a serem realizados:

- Obtenção e análise estrutural de cristais de co-cristalizados com ativador, crescidos em solução de acetato de sódio.
- Coleta de dados de cristais co-cristalizados com ativador e inibidor, crescidos em agente precipitante fostato de sódio/potássio e acetato de sódio, onde poderá se obter uma estrutura enzima ativador-inibidor.
- Análise da estrutura com cristais crescidos em diferentes pHs.
- Resolução das estruturas com cristais crescidos em outros agentes precipitantes: citrato de sódio e tartrato.
- Uso de outras técnicas bioquímicas e biofísicas para estudar as características estruturais-funcionais da enzima.

## REFERÊNCIAS BIBLIOGRÁFICAS

- Acharya, R.; Fry, E.; Stuart, D.; Fox, G. Rowlands, D. e Brown, F. (1989), *Nature*, **327**, 709.
- Altamirano, M.M. (1993) Tese de doutorado, Universidad Nacional Autónoma de México, UACPyP.
- Altamirano, M.M. e Calcagno, M. (1990), *Biochim. Biophys. Acta*, **1038**, 291.
- Altamirano, M.M.; Lara-Lemus, R.; Libreros-Minotta, C.A. e Calcagno, M. (1989), *Arch. Biochim. Biophys.*, **269**, 555.
- Altamirano, M.M.; Mulliert, G. e Calcagno, M. (1987), *Arch. Biochim. Biophys.*, **258**, 95
- Altamirano M.M.; Plumbridge, J.A.; Barba, H.A. e Calcagno, M. (1993), *Biochem. J.* In press.
- Altamirano M.M.; Plumbridge, J.A e Calcagno, M. (1992), *Biochemistry*, **31**, 1153.
- Altamirano M.M.; Plumbridge, J.A.; Hernández-Arana, A. e Calcagno, M. (1991), *Biochim. Biophys. Acta*, **1076**, 266.
- Arndt, U.W. e Wonacott, A.J., (1977) eds. *The Rotation Method in Crystallography*. North-Holland, Amsterdam.
- Bachmann, B.J. (1990), *Microbiological Reviews*, **54**, 455.
- Bleasby, A. e Wootton, J. (1990), *Protein Eng.*, **3**, 153.
- Blow, D.M. e Crick, F.H.C. (1959), *Acta Crystallogr.*, **12**, 794..
- Blundell, T.L. e Johnson, L.N. (1976), *Protein Crystallography*. Academic Press Inc., London, UK.
- Brünger, A.T.; Kuriyan, K. e Karplus, M., (1987), *Science.*, **235**, 458.
- Brünger, A.T.; Karplus, M. e Petsko, G.A., (1989), *Acta. Cryst.*, **A45**, 50.
- Brünger, A.T. (1992), *Nature*, **355**, 472.
- Brünger, A.T. (1993), *X-PLOR Version 3.1 manual*, Yale Univ., USA
- Busing, W.R. e Levy, H.A. (1957) *Acta Crystallogr.*, **10**, 180.
- Calcagno, M.; Campos, P., Mulliert, G.Y. e Suástegui, J. (1984), *Biochim. Biophys. Acta*, **787**, 165.
- Cambillau, C.C. e Horjales, E. (1987), *J. Mol. Graph.*, **5**, 174.
- Colaborative Computacional Project in Protein Crystallography (1985). SERC. Daresbury Laboratory, Warrington, UK.
- Colaborative Computacional Project, Number 4 (1994), *Acta Crystallogr.*, **D50**, 760. 'The CCP4 Suite: Programs for Protein Crystallography'.
- Comb, D.G. e Roseman, S. (1958), *J. Biol. Chem.*, **232**, 807.
- Creighton, T. E. (1993), *Proteins, Structures and Molecular Principles*, W. H. Freeman and Company, New York, USA.
- Das, M. e Datta, A. (1982), *Biochem. Int.*, **5**, 735.
- Discover, User Guide, Version 2.3.0. San Diego: Biosym Technologies, 1993.
- Drenth, J. (1994), *Principles of protein X-ray crystallography*, Springer-Verlag, New York. USA.
- Enraf-Nonius (1975), manual do difratômetro CAD-4.
- Evans, P.R.; Farrants, G.W. e Hudson, P.J. (1981), *Philos. Trans. R. Soc. London.*, **293**, 53.
- Eyck, L.F.T. (1973), *Acta Crystallogr.*, **A29**, 183.

- Fersht, A. (1985), *Enzyme structure and mechanism*, W. H. Freeman and Company, New York.
- Fink, A.L. (1987) Acyl group transfer: the serine proteases. In: *Enzyme Mechanisms* (Page, M.I. & Williams, A., eds.) pp. 159-177, The Royal Society of Chemistry, London.
- Fontes, M.R.M. (1992), *Dissertação de Mestrado*, Inst. Fis. Quim. São Carlos-USP
- Fox, G.C. e Holmes, K.C., (1966), *Acta Cryst.*, **20**, 886.
- Gonzalez, A. e Nave, C. (1994), *Acta Crystallogr.*, **D50**, 874.
- Gonzalez, A.; Denny, R. e Nave, C. (1994), *Acta Crystallogr.*, **D50**, 276.
- Gruber, B. (1973), *Acta Crystallogr.*, **A29**, 433.
- Gruber, B. (1989), *Acta Crystallogr.*, **A45**, 123.
- Hamilton, W.C.; Rollett, J.S. e Sparks, R.A. (1965), *Acta Crystallogr.*, **18**, 129.
- Harker, D., (1956), *Acta Crystallogr.*, **9**, 1.
- Hernández-Arana, A.; Rojo-Domínguez, A.; Altamirano, M.M. e Calcagno, M.L. (1993), *Biochemistry*, **32**, 2644.
- Higashi, T. (1990), *J. Appl. Cryst.*, **23**, 253.
- Hodge, J.E. (1955), *Advan. Carbohyde. Chem.*, **10**, 169.
- Horjales, E.; Altamirano, M.M.; Calcagno, M.L.; Dauter, Z.; Wilson, K.; Garratt, R.C. e Oliva, G. (1992), *J. Mol. Biol.*, **226**, 1283.
- Howell, P.L. e Smith, G.D. (1992), *J. Appl. Cryst.*, **25**, 81.
- Insight II User Guide, Version 2.3.0. San Diego: Biosym Technologies, 1993.
- *International Table for X-Ray Crystallography* (1987), Vol. I. The Kynoch Press, Birmingham, England.
- Jones, T.A.; Zou, J.Y.; Cowan, S.W. e Kjeldgaard, M. (1991), *Acta Crystallogr.*, **A47**, 110.
- Kabsch, W., (1988), *J. Appl. Cryst.*, **21**, 67.
- Kendrew, J.C.; Bodo, G.; Dintzis, H.M.; Parrish, R.G.; Wykoff, H. e Phillips, D.C. (1958), *Nature*, **181**, 662.
- Kim, S. (1989), *J. Appl. Cryst.*, **22**, 53.
- Leslie, A.G.W., (1987), *Acta Crystallogr.*, **A43**, 134.
- Lüthy, R.; Bowie, J.U. e Eisenberg, D. (1992), **356**, 83.
- McRee, D.E. (1994), *Practical Protein Crystallography*. Academic Press, Inc., San Diego, USA
- Midelfort, C. & Rose, I.A. (1977), *Biochemistry*, **16**, 1590-1596.
- Mirelman, D. (1976), in *Bacterial outer membranes*. Inouye, M ed., John Wiley & Sons, Inc. New York.
- Monod, J.; Wyman, J. e Changeux, J.P. (1965), *J. Mol. Biol.*, **12**, 88.
- Moris, A.L.; MacArthur, MW.; Hutchinson, E.G. e Thornton, J.M. (1992) *Proteins: Structure, Function and Genetics*. **12**, 345.
- Noltmann, E.A. (1972), In *The Enzymes* (Boyer, P. D., ed.), 3<sup>rd</sup> ed., vol. 6. 314-318. Academic Press, New York.
- Patterson, A.L. (1934), *Phys. Rev.*, **46**, 372.
- Plumbridge, J.A. (1989), *Mol. Microbiol*, **3**, 506
- Powell, M.J.D. (1977), *Mathematical Programming*, **12**, 241
- Raetz, C.R. (1987), In: *Escherichia coli and Salmonella typhimurium: cellular and molecular biology*. American Society for Microbiology, Washington, D.C.
- Rypniewski, W.R. e Evans, P.R. (1989), *J. Mol. Biol.*, **207**, 805.
- Rogers, M.J., Ohgi, T., Plumbridge, J.A. e Söll, D. (1988), *Gene*, **62**, 197.
- Rossmann, M.G. (1979), *J. Appl. Cryst.*, **12**, 225

- Rossmann, M.G.; Leslie, A.G.W.; Abdel-Meguid, S.S. e Tsukihara, T. (1979), *J. Appl. Cryst.*, **12**, 570.
- Rossmann, M.G. (1985), *In: Methods in Enzymology, Diffraction Methods for Biological Macromolecules v. 114*.
- Sarvas, M. (1971), *J. Bacteriol*, **105**, 467.
- Sato, M.; Yamamoto, M.; Imada, K.; Katsube, Y.; Tanaka, N.; Higashi, T. (1992), *Acta Crystallogr.*, **25**, p.348
- Shirakihara, Y. e Evans, P.R. (1988), *J. Mol. Biol.*, **204**, 973.
- Sim, G.A. (1959), *Acta Crystallogr.*, **12**, p.813
- Stephen, (1985) *In: Methods in Enzymology, Diffraction Methods for Biological Macromolecules v. 114*, 236.
- Stout, G.H. e Jensen L.H., (1968), *X-ray structure determination*, The Macmilan Company, New York.
- Stryer, L. (1992), *Bioquímica*, Ed. Guanabara Koogan S.A., Rio de Janeiro, BR.
- Verlet, L. (1967), *Phys. Rev.*, **159**, p.98.
- Vriend, G. (1990), *J. Mol. Graphics*, **8**, 52.
- Vriend, G. e Sander, C. (1993), *J. Appl. Cryst.*, **26**, 47.
- Wang, B.C. (1985), *Methods Enzymol.*, **115**, p.90.
- Wilson, A.J.C. (1950), *Acta Crystallogr.*, **3**, p.397.
- Winkler, F.K.; Schutt, C.E. e Harrison, S.C. (1979), *Acta Crystallogr.*, **A35**, 901.
- Wu, H.C. e Wu, T.C. (1971), *J. Bacteriol*, **105**, 455.

## APÊNDICE A

Arquivos de topologia e parâmetros adaptados para os refinamentos dos complexos usando o programa XPLOR: enzima ativador-Pi, enzima inibidor-Pi, enzima 2-inibidor

- Arquivo de topologia do ativador..... A1
- Arquivo de parâmetros do ativador..... A3
- Arquivo de topologia do inibidor..... A7
- Arquivo de parâmetros do inibidor..... A9

## ATTIVADOR - ARQUIVO DE TOPOLOGIA

```
set echo=false end
AUTOGENERATE ANGLES=TRUE END
{*****}
```

```
{default masses -- only those not in protein topology TOPH19X.PRO}
```

```
MASS PN 30.97400! phosphate phosphor
MASS OPN 15.99940! phosphate oxygen, P=O
MASS ON 15.99940! phosphate oxygen, P-O
MASS CC 12.01100! carbohydrate carbon
MASS CC6 12.01100! carbohydrate carbon, exocyclic C6
MASS CCE 12.01100! carbohydrate carbon, equatorial C1
MASS CCA 12.01100! carbohydrate carbon, axial C1
MASS OE 15.99940! ether oxygen
MASS OA 15.99940! ether oxygen in glycosidic bond
MASS HA 1.00800! aliphatic hydrogen
```

```
RESidue NAGP !2-N-Acetyl-beta-d-glucose-6P,pyranoseform
!(ALTERADO C1=CCE P/ CCA)
```

```
GROUP
```

```
ATOM C1 TYPE=CCA CHARGE= 0.350 END
ATOM H1 TYPE=HA CHARGE= 0.100 END
ATOM O1 TYPE=OH1 CHARGE=-0.650 END
ATOM HO1 TYPE=H CHARGE= 0.400 END
ATOM C2 TYPE=CC CHARGE= 0.000 END
ATOM H2 TYPE=HA CHARGE= 0.100 END
ATOM N2 TYPE=NH1 CHARGE=-0.350 END
ATOM HN2 TYPE=H CHARGE= 0.250 END
ATOM C7 TYPE=C CHARGE= 0.550 END
ATOM O7 TYPE=O CHARGE=-0.550 END
ATOM C8 TYPE=CC CHARGE=-0.300 END
ATOM H81 TYPE=HA CHARGE= 0.100 END
ATOM H82 TYPE=HA CHARGE= 0.100 END
ATOM H83 TYPE=HA CHARGE= 0.100 END
ATOM C3 TYPE=CC CHARGE= 0.150 END
ATOM H3 TYPE=HA CHARGE= 0.100 END
ATOM O3 TYPE=OH1 CHARGE=-0.650 END
ATOM HO3 TYPE=H CHARGE= 0.400 END
ATOM C4 TYPE=CC CHARGE= 0.150 END
ATOM H4 TYPE=HA CHARGE= 0.100 END
ATOM O4 TYPE=OH1 CHARGE=-0.650 END
ATOM HO4 TYPE=H CHARGE= 0.400 END
ATOM C5 TYPE=CC CHARGE= 0.100 END
ATOM H5 TYPE=HA CHARGE= 0.100 END
ATOM O5 TYPE=OE CHARGE=-0.400 END
ATOM C6 TYPE=CC6 CHARGE= 0.050 END
ATOM H61 TYPE=HA CHARGE= 0.100 END
ATOM H62 TYPE=HA CHARGE= 0.100 END
ATOM O6 TYPE=ON CHARGE=-0.250 END
ATOM P TYPE=PN CHARGE= 1.200 END
ATOM OP1 TYPE=OPN CHARGE=-0.300 END
ATOM OP2 TYPE=OPN CHARGE=-0.300 END
ATOM OP3 TYPE=OPN CHARGE=-0.300 END
```

```
BOND C1 O1
BOND C1 H1
BOND O1 HO1
BOND C1 O5
BOND C1 C2
BOND C2 H2
BOND C2 N2
BOND N2 HN2
BOND N2 C7
BOND C7 O7
BOND C7 C8
```

BOND C8 H81  
BOND C8 H82  
BOND C8 H83  
BOND C2 C3  
BOND C3 H3  
BOND C3 O3  
BOND O3 HO3  
BOND C3 C4  
BOND C4 H4  
BOND C4 O4  
BOND O4 HO4  
BOND C4 C5  
BOND C5 H5  
BOND C5 C6  
BOND C6 H61  
BOND C6 H62  
BOND C6 O6  
BOND C5 O5  
BOND O6 P  
BOND P OP1  
BOND P OP2  
BOND P OP3

DIHEDRAL C1 C2 C3 C4  
DIHEDRAL C2 C3 C4 C5  
DIHEDRAL C3 C4 C5 O5  
DIHEDRAL C4 C5 O5 C1  
DIHEDRAL C5 O5 C1 C2  
DIHEDRAL O5 C1 C2 C3  
DIHEDRAL H1 C1 O1 HO1  
DIHEDRAL H2 C2 N2 HN2  
DIHEDRAL C2 N2 C7 C8  
DIHEDRAL N2 C7 C8 H81  
DIHEDRAL H3 C3 O3 HO3  
DIHEDRAL H4 C4 O4 HO4  
DIHEDRAL H5 C5 C6 O6  
DIHEDRAL C5 C6 O6 P  
DIHEDRAL C6 O6 P OP1

IMPRoper C1 O5 H1 C2  
! IMPRoper C1 O1 O5 C2  
IMPRoper C2 C1 N2 C3  
IMPRoper N2 C7 C2 HN2 {planar N}  
IMPRoper C7 C8 N2 O7 {planar C}  
IMPRoper C3 O3 C2 C4  
IMPRoper C4 C3 O4 C5  
IMPRoper C5 C6 C4 O5  
DONOr HO1 O1  
DONOr HN2 N2  
DONOr HO3 O3  
DONOr HO4 O4  
ACCEptor O1 C1  
ACCEptor O7 C7  
ACCEptor O3 C3  
ACCEptor O4 C4  
ACCEptor O5 C5  
ACCEptor O6 C6  
ACCEptor OP1 P  
ACCEptor OP2 P  
ACCEptor OP3 P

END

□

## ATIVADOR - ARQUIVO DE PARAMETROS

REMARKS Parameter file for pyranose sugars

set echo=false end

```

bond CC OE 1780.2 1.438 ! closer to crystal structures
bond CCE OE 1780.2 1.428
bond CCA OE 1780.2 1.419
bond CC CC 1284.0 1.526 ! adjusted to Jeffries
bond CC CC6 1284.0 1.516
bond CCE CC 1284.0 1.526
bond CCA CC 1284.0 1.526
bond CC HA 337.3 1.099
bond CC6 HA 337.3 1.099
bond CCE HA 337.3 1.099
bond CCA HA 337.3 1.099
bond CC OH1 2005.8 1.420
bond CC6 OH1 2005.8 1.420
bond CCE OH1 2005.8 1.385
bond CCA OH1 2005.8 1.398
bond CCE OA 2005.8 1.385
bond CCA OA 2005.8 1.398
{ Note elongated non-anomeric C-O bond for bridges }
bond CC OA 2005.8 1.44
bond CC6 OA 2005.8 1.44
bond CC NH1 2631.111 1.45 !as extended atom carbon-N
bond CCE NH1 2631.111 1.45 !as extended atom carbon-N
bond CCA NH1 2631.111 1.45 !as extended atom carbon-N
bond CC C 1342.404 1.52 !" " " " -C
bond CCE C 1342.404 1.52
bond CCA C 1342.404 1.52
bond HN ON 450.0 0.96
bond OPN PN 528.0 1.48
bond ON PN 237.0 1.61
bond ON CC6 292.0 1.43
bond OPN CC6 292.0 1.43
{ Angle force constants increased 10x, except angle around bridging }
{ link oxygen increased 5 }
angle H OH1 CC 53.62 109.35
angle H OH1 CC6 53.62 109.35
angle H OH1 CCE 53.62 109.35
angle H OH1 CCA 53.62 109.35
angle HA CC HA 33.58 107.85
angle HA CC6 HA 33.58 107.85
angle HA CC CC 43.04 108.72
angle HA CC6 CC 43.04 108.72
angle HA CC CC6 43.04 108.72
angle HA CC CCE 43.04 108.72
angle HA CC CCA 43.04 108.72
angle HA CCE CC 43.04 108.72
angle HA CCA CC 43.04 108.72
angle HA CC OH1 45.89 109.89
angle HA CC6 OH1 45.89 109.89
angle HA CCE OH1 45.89 109.89
angle HA CCA OH1 45.89 109.89
angle HA CC OA 45.89 109.89

```

angle HA CC6 OA 45.89 109.89  
 angle HA CC6 OPN 45.89 109.89  
 angle HA CC6 ON 45.89 109.89  
 angle HA CCE OA 45.89 109.89  
 angle HA CCA OA 45.89 109.89  
 angle HA CC OE 45.16 107.24  
 angle HA CCE OE 45.16 107.24  
 angle HA CCA OE 45.16 107.24  
 angle CC CC CC 380.3 110.40  
 angle CC CC CC6 380.3 112.50  
 angle CCE CC CC 380.3 110.40  
 angle CCA CC CC 380.3 110.40  
 angle CC CC6 OH1 756.8 111.20  
 angle CC CCE OH1 756.8 109.70  
 angle CC CCA OH1 756.8 109.70  
 angle CCE CC OH1 756.8 109.70  
 angle CCA CC OH1 756.8 109.70  
 angle CC CC OH1 756.8 109.70  
 angle CC CC OE 809.5 110.00  
 angle CC6 CC OE 809.5 107.20  
 angle CC CCE OE 809.5 110.00  
 angle CC CCA OE 809.5 110.00  
 angle CC CC OA 809.5 109.70  
 angle CC CC6 OA 809.5 109.70 !guess for 1-6 links  
 angle CC CCE OA 809.5 109.70  
 angle CC CC6 OPN 809.5 109.70  
 angle CC CC6 ON 809.5 109.70  
 angle CC CCA OA 809.5 109.70  
 angle CCE CC OA 809.5 109.70  
 angle CCA CC OA 809.5 109.70  
 !angle OE CC OH1 925.6 111.55  
 !angle OE CC OE 925.6 110.70 { C5 - O5 - C1 }  
 angle CC OE CCE 906.9 112.0  
 angle CC OE CCA 906.9 114.0 { O5 - C1 - O1 }  
 angle OE CCE OH1 925.6 108.0  
 angle OE CCE OA 925.6 108.0  
 angle OE CCA OH1 925.6 112.1  
 angle OE CCA OA 925.6 112.1  
 angle CCE OA CC 453.45 116.  
 angle CCA OA CC 453.45 115.  
 angle CCE OA CC6 453.45 116.  
 !angle CCA OA CC6 453.45 115.  
 angle CCA OA CC6 453.45 111.5 !alpha 1,6 link  
 { The following are guesses, generally by analogy to extended atom }  
 { Carbons.Note equilibrium values not based on crystal structures }  
 angle CC CC NH1 650.0 110.0  
 angle CCE CC NH1 650.0 110.0  
 angle CCA CC NH1 650.0 110.0  
 angle CC CCE NH1 650.0 110.0  
 angle CC CCA NH1 650.0 110.0  
 angle HA CC NH1 45.0 109.5  
 angle HA CCE NH1 45.0 109.5  
 angle HA CCA NH1 45.0 109.5  
 angle CC NH1 H 35.0 120.0  
 angle CCE NH1 H 35.0 120.0

angle CCA NH1 H 35.0 120.0  
 angle CC NH1 C 775.0 120.0  
 angle CCE NH1 C 775.0 120.0  
 angle CCA NH1 C 775.0 120.0  
 angle NH1 CC OE 750.0 110.7 !A complete guess  
 angle NH1 CCE OE 750.0 110.7 !A complete guess  
 angle NH1 CCA OE 750.0 110.7 !A complete guess  
 angle NH1 C CC 200.0 117.5  
 angle C CCE OH1 700.0 109.5 !A complete guess for SIA O2  
 angle C CCA OH1 700.0 109.5 !A complete guess for SIA O2  
 angle C CCE OA 700.0 109.5 !A complete guess for SIA O2  
 angle C CCA OA 700.0 109.5 !A complete guess for SIA O2  
 angle O C CCE 850.0 121.5  
 angle O C CCA 850.0 121.5  
 angle O C CC 850.0 121.5  
 angle C CCE HA 70.0 109.5  
 angle C CCA HA 70.0 109.5  
 angle C CC HA 70.0 109.5  
 {This is (360-the usual OC-C-OC angle of 122.5)/2 }  
 angle CCE C OC 850.0 118.75  
 angle CCA C OC 850.0 118.75  
 {Use the same angles to HA in the C1 position of NANA }  
 angle C CCE OE 850.0 107.24  
 angle C CCA OE 850.0 107.24  
 angle C CCE CC 700.0 108.72  
 angle C CCA CC 700.0 108.72  
 angles HN ON PN 47.0 107.3  
 angles ON PN ON 48.1 102.0  
 angles OPN PN ON 100.0 108.0  
 angles OPN PN OPN 140.0 120.0  
 angles PN ON CC6 47.0 120.0  
 angles PN OPN CC6 47.0 120.0  
 dihedral X CC CC X 3.0630 3 0.0  
 dihedral X CC CC6 X 3.0630 3 0.0  
 dihedral X CCE CC X 3.0630 3 0.0  
 dihedral X CCA CC X 3.0630 3 0.0  
 dihedral X CC OH1 X 1.3293 3 0.0  
 dihedral X CC6 OH1 X 1.3293 3 0.0  
 dihedral X CCE OH1 X 1.3293 3 0.0  
 dihedral X CCA OH1 X 1.3293 3 0.0  
 dihedral X CC OA X 1.3293 3 0.0  
 dihedral X CC6 OA X 1.3293 3 0.0  
 dihedral X CCE OA X 1.3293 3 0.0  
 dihedral X CCA OA X 1.3293 3 0.0  
 dihedral X CC OE X 2.7834 3 0.0  
 dihedral X CCE OE X 2.7834 3 0.0  
 dihedral X CCA OE X 2.7834 3 0.0  
 {The following are guesses, generally by analogy to extended atom Carbons }  
 dihedral HA CC NH1 H 0.3 3 180.0 !for NAG and NANA  
 dihedral X CC NH1 X 0.9 3 0.0 !beta link Asn Nd1  
 dihedral X CCE NH1 X 0.9 3 0.0 !beta link Asn Nd1  
 dihedral X CCA NH1 X 0.9 3 0.0 !beta link Asn Nd1  
 dihedral X C CCE X 0.0 3 0.0  
 dihedral X C CCA X 0.0 3 0.0  
 dihedral X C CC X 0.0 3 0.0

```

dihedral X ON CC6 X 0.5 3 0.0      {*}Threefold.*}
dihedral X PN ON X 0.5 3 0.0      {*}Threefold.*}
dihedral X PN OPN X 0.5 3 0.0     {*}Threefold.*}

```

```
{These high force constants maintain chiral centers on the rings }
```

```

improper CC X X CC 300.0 0 35.26439
improper CCE X X CC 300.0 0 35.26439
improper CCA X X CC 300.0 0 35.26439
improper OH1 X X CC 300.0 0 35.26439
improper OH1 X X CCE 300.0 0 35.26439
improper OH1 X X CCA 300.0 0 35.26439
improper OA X X CC 300.0 0 35.26439
improper OA X X CCE 300.0 0 35.26439
improper OA X X CCA 300.0 0 35.26439
improper CC X X OE 300.0 0 35.26439
improper CCE X X OE 300.0 0 35.26439
improper CCA X X OE 300.0 0 35.26439

```

```
{ This is as other acids }
```

```

improper C X X CCE 300.0 0 0.0 !NANA acid planarity
improper C X X CCA 300.0 0 0.0 !NANA acid planarity

```

```
!!for use with:
```

```
!!NBXMOD=5 ATOM CDIEL SHIFT vswitch
```

```
!! CUTNB=8.0 CTOFNB=7.5 CTONNB=6.5 EPS=1.0 E14FAC=0.4 WMIN=1.5
```

```

!NONBonded HA 0.0045 2.6157 0.0045 2.6157
NONBonded CC 0.0903 3.2072 0.0903 3.2072
NONBonded CC6 0.0903 3.2072 0.0903 3.2072
NONBonded CCE 0.0903 3.2072 0.0903 3.2072
NONBonded CCA 0.0903 3.2072 0.0903 3.2072
NONBonded OE 0.1591 2.8509 0.1591 2.8509
NONBonded OA 0.1591 2.8509 0.1591 2.8509

```

```

{* Lennard-Jones parameters      *}
{*      -----1-4-----      *}
{* epsilon sigma   epsilon sigma *}
{* (kcal/mol) (A)   (kcal/mol) (A) *}

```

```

nonbonded HN 0.0498 1.4254 0.0498 1.4254
nonbonded ON 0.1591 2.8509 0.1591 2.8509
nonbonded OPN 0.1591 2.8509 0.1591 2.8509
nonbonded PN 0.5849 3.3854 0.5849 3.3854

```

```
set echo=true end
```

```
□
```

## INIBIDOR - ARQUIVO DE TOPOLOGIA

set echo=false end

AUTOGENERATE ANGLES=TRUE END

{\*=====\*

{default masses -- only those not in protein topology TOPH19X.PRO}

MASS PN 30.97400! phosphate phosphor  
MASS OPN 15.99940! phosphate oxygen, P=O  
MASS ON 15.99940! phosphate oxygen, P-O  
!MASS CC 12.01100! carbohydrate carbon  
!MASS CC6 12.01100! carbohydrate carbon, exocyclic C6  
!MASS CCE 12.01100! carbohydrate carbon, equatorial C1  
!MASS CCA 12.01100! carbohydrate carbon, axial C1  
MASS OE 15.99940! ether oxygen  
MASS OA 15.99940! ether oxygen in glycosidic bond  
!MASS HA 1.00800! aliphatic hydrogen

RESidue INH !2-N-Acetyl-beta-d-glucose-6P,pyranose form  
GROUP

ATOM C1 TYPE=CH2E CHARGE= 0.250 END!(.35)  
ATOM O1 TYPE=OH1 CHARGE=-0.650 END!(-.65)  
ATOM HO1 TYPE=H CHARGE= 0.400 END!(.4)  
ATOM C2 TYPE=CH1E CHARGE= 0.100 END!(0)  
ATOM N2 TYPE=NH1 CHARGE=-0.350 END!(-.35)  
ATOM HN2 TYPE=H CHARGE= 0.250 END!(.25)  
ATOM C3 TYPE=CH1E CHARGE= 0.250 END!(.15)  
ATOM O3 TYPE=OH1 CHARGE=-0.650 END!(-.65)  
ATOM HO3 TYPE=H CHARGE= 0.400 END!(.4)  
ATOM C4 TYPE=CH1E CHARGE= 0.250 END!(.15)  
ATOM O4 TYPE=OH1 CHARGE=-0.650 END!(-.65)  
ATOM HO4 TYPE=H CHARGE= 0.400 END!(.4)  
ATOM C5 TYPE=CH1E CHARGE= 0.250 END!(.1)  
ATOM O5 TYPE=OH1 CHARGE=-0.650 END!(-.4)  
ATOM HO5 TYPE=H CHARGE= 0.400 END!(.4)  
ATOM C6 TYPE=CH2E CHARGE= 0.250 END!(.05)  
ATOM O6 TYPE=ON CHARGE=-0.250 END  
ATOM P TYPE=PN CHARGE= 1.200 END  
ATOM OP1 TYPE=OPN CHARGE=-0.300 END  
ATOM OP2 TYPE=OPN CHARGE=-0.300 END  
ATOM OP3 TYPE=OPN CHARGE=-0.300 END  
BOND C1 O1  
BOND O1 HO1  
BOND C1 C2  
BOND C2 N2  
BOND N2 HN2  
BOND C2 C3  
BOND C3 O3  
BOND O3 HO3  
BOND C3 C4  
BOND C4 O4  
BOND O4 HO4  
BOND C4 C5

BOND C5 O5  
BOND O5 HO5  
BOND C5 C6  
BOND C6 O6  
BOND O6 P  
BOND P OP1  
BOND P OP2  
BOND P OP3  
DIHEDRAL HO1 O1 C1 C2  
DIHEDRAL O1 C1 C2 C3  
DIHEDRAL C1 C2 C3 C4  
DIHEDRAL C2 C3 C4 C5  
DIHEDRAL C3 C4 C5 C6  
DIHEDRAL C4 C5 C6 O6  
! DIHEDRAL C3 C4 C5 O5  
! DIHEDRAL C4 C5 O5 C1  
! DIHEDRAL C5 O5 C1 C2  
! DIHEDRAL O5 C1 C2 C3  
! DIHEDRAL H1 C1 O1 HO1  
DIHEDRAL C1 C2 N2 HN2  
! DIHEDRAL C2 N2 C7 C8  
! DIHEDRAL N2 C7 C8 H81  
DIHEDRAL C2 C3 O3 HO3  
DIHEDRAL C3 C4 O4 HO4  
DIHEDRAL C4 C5 O5 HO5  
DIHEDRAL C5 C6 O6 P  
DIHEDRAL C6 O6 P OP1  
! IMPRoper C1 O5 H1 C2  
! IMPRoper C1 O1 O5 C2  
IMPRoper C2 C1 N2 C3  
! IMPRoper N2 C7 C2 HN2 {planar N}  
! IMPRoper C7 C8 N2 O7 {planar C}  
IMPRoper C3 O3 C2 C4  
IMPRoper C4 C3 O4 C5  
IMPRoper C5 C6 C4 O5  
DONOr HO1 O1  
DONOr HN2 N2  
DONOr HO3 O3  
DONOr HO4 O4  
DONOr HO5 O5  
! ACCEptor O7 C7  
ACCEptor O1 C1  
ACCEptor O3 C3  
ACCEptor O4 C4  
ACCEptor O5 C5  
ACCEptor O6 C6  
ACCEptor OP1 P  
ACCEptor OP2 P  
ACCEptor OP3 P

END

□

## INIBIDOR - ARQUIVO DE PARAMETROS

```

set echo=false end
{bond force constants 6x previous.Some adjustments made to}
{equilibrium values }
bond CC OE 1780.2 1.438 ! closer to crystal structures
bond CCE OE 1780.2 1.428
bond CCA OE 1780.2 1.419
bond CC CC 1284.0 1.526 ! adjusted to Jeffries
bond CC CC6 1284.0 1.516
bond CCE CC 1284.0 1.526
bond CCA CC 1284.0 1.526
bond CC HA 337.3 1.099
bond CC6 HA 337.3 1.099
bond CCE HA 337.3 1.099
bond CCA HA 337.3 1.099
bond CC OH1 2005.8 1.420
bond CC6 OH1 2005.8 1.420
bond CCE OH1 2005.8 1.385
bond CCA OH1 2005.8 1.398
bond CCE OA 2005.8 1.385
bond CCA OA 2005.8 1.398
{Note elongated non-anomeric C-O bond for bridges }
bond CC OA 2005.8 1.44
bond CC6 OA 2005.8 1.44
bond CC NH1 2631.111 1.45 !as extended atom carbon-N
bond CCE NH1 2631.111 1.45 !as extended atom carbon-N
bond CCA NH1 2631.111 1.45 !as extended atom carbon-N
bond CC C 1342.404 1.52 !" " " " -C
bond CCE C 1342.404 1.52
bond CCA C 1342.404 1.52
bond HN ON 450.0 0.96
bond OPN PN 528.0 1.48
bond ON PN 237.0 1.61
bond ON CH2E 292.0 1.43 !! {COLOCADO PARA INIBIDOR}
bond ON CC6 292.0 1.43
bond OPN CC6 292.0 1.43
{Angle force constants increased 10x, except angle around bridging}
{link oxygen increased 5x }
angle H OH1 CC 53.62 109.35
angle H OH1 CC6 53.62 109.35
angle H OH1 CCE 53.62 109.35
angle H OH1 CCA 53.62 109.35
angle HA CC HA 33.58 107.85
angle HA CC6 HA 33.58 107.85
angle HA CC CC 43.04 108.72
angle HA CC6 CC 43.04 108.72
angle HA CC CC6 43.04 108.72
angle HA CC CCE 43.04 108.72
angle HA CC CCA 43.04 108.72
angle HA CCE CC 43.04 108.72
angle HA CCA CC 43.04 108.72
angle HA CC OH1 45.89 109.89
angle HA CC6 OH1 45.89 109.89
angle HA CCE OH1 45.89 109.89

```

angle HA CCA OH1 45.89 109.89  
 angle HA CC OA 45.89 109.89  
 angle HA CC6 OA 45.89 109.89  
 angle HA CC6 OPN 45.89 109.89  
 angle HA CC6 ON 45.89 109.89  
 angle HA CCE OA 45.89 109.89  
 angle HA CCA OA 45.89 109.89  
 angle HA CC OE 45.16 107.24  
 angle HA CCE OE 45.16 107.24  
 angle HA CCA OE 45.16 107.24  
 angle CC CC CC 380.3 110.40  
 angle CC CC CC6 380.3 112.50  
 angle CCE CC CC 380.3 110.40  
 angle CCA CC CC 380.3 110.40  
 angle CCA CC CC 380.3 110.40  
 angle CC CC6 OH1 756.8 111.20  
 angle CC CCE OH1 756.8 109.70  
 angle CH1E CH1E CH1E 672.5 110.40  
 !{COLOCADO PARA INIBIDOR, USANDO C2 C1 C1}  
 angle CCE CC OH1 756.8 109.70  
 angle CCA CC OH1 756.8 109.70  
 angle CC CC OH1 756.8 109.70  
 angle CH2E CH1E OH1 756.8 109.70!{COLOCADO PARA INIBIDOR}  
 angle CC CC OE 809.5 110.00  
 angle CC6 CC OE 809.5 107.20  
 angle CC CCE OE 809.5 110.00  
 angle CC CCA OE 809.5 110.00  
 angle CC CC OA 809.5 109.70  
 angle CC CC6 OA 809.5 109.70 !guess for 1-6 links  
 angle CC CCE OA 809.5 109.70  
 angle CC CC6 OPN 809.5 109.70  
 angle CC CC6 ON 809.5 109.70  
 angle CH1E CH2E ON 809.5 109.70!{COLOCADO PARA INIBIDOR}  
 angle CC CCA OA 809.5 109.70  
 angle CCE CC OA 809.5 109.70  
 angle CCA CC OA 809.5 109.70  
 !angle OE CC OH1 925.6 111.55  
 !angle OE CC OE 925.6 110.70{ C5 - O5 - C1 }  
 angle CC OE CCE 906.9 112.0  
 angle CC OE CCA 906.9 114.0{O5 - C1 - O1}  
 angle OE CCE OH1 925.6 108.0  
 angle OE CCE OA 925.6 108.0  
 angle OE CCA OH1 925.6 112.1  
 angle OE CCA OA 925.6 112.1  
 {bridge angle - note a fairly large spread in the literature axial generally}  
 {smaller than equatorial.Increased only 5x to reflect limited data}  
 angle CCE OA CC 453.45 116.  
 angle CCA OA CC 453.45 115.  
 angle CCE OA CC6 453.45 116.  
 !angle CCA OA CC6 453.45 115.  
 angle CCA OA CC6 453.45 111.5 !alpha 1,6 link  
 {The following are guesses, generally by analogy to extended atom}  
 {Carbons.Note equilibrium values not based on crystal structures.}  
 angle CC CC NH1 650.0 110.0  
 angle CCE CC NH1 650.0 110.0

angle CCA CC NH1 650.0 110.0  
 angle CC CCE NH1 650.0 110.0  
 angle CC CCA NH1 650.0 110.0  
 angle HA CC NH1 45.0 109.5  
 angle HA CCE NH1 45.0 109.5  
 angle HA CCA NH1 45.0 109.5  
 angle CC NH1 H 35.0 120.0  
 angle CCE NH1 H 35.0 120.0  
 angle CCA NH1 H 35.0 120.0  
 angle CC NH1 C 775.0 120.0  
 angle CCE NH1 C 775.0 120.0  
 angle CCA NH1 C 775.0 120.0  
 angle NH1 CC OE 750.0 110.7 !A complete guess  
 angle NH1 CCE OE 750.0 110.7 !A complete guess  
 angle NH1 CCA OE 750.0 110.7 !A complete guess  
 angle NH1 C CC 200.0 117.5  
 angle C CCE OH1 700.0 109.5 !complete guess for SIA O2  
 angle C CCA OH1 700.0 109.5 !complete guess for SIA O2  
 angle C CCE OA 700.0 109.5 !complete guess for SIA O2  
 angle C CCA OA 700.0 109.5 !complete guess for SIA O2  
 angle O C CCE 850.0 121.5  
 angle O C CCA 850.0 121.5  
 angle O C CC 850.0 121.5  
 angle C CCE HA 70.0 109.5  
 angle C CCA HA 70.0 109.5  
 angle C CC HA 70.0 109.5  
 {This is (360-the usual OC-C-OC angle of 122.5)/2}  
 angle CCE C OC 850.0 118.75  
 angle CCA C OC 850.0 118.75  
 {Use the same angles to HA in the C1 position of NANA}  
 angle C CCE OE 850.0 107.24  
 angle C CCA OE 850.0 107.24  
 angle C CCE CC 700.0 108.72  
 angle C CCA CC 700.0 108.72  
 angles HN ON PN 47.0 107.3  
 angles ON PN ON 48.1 102.0  
 angles OPN PN ON 100.0 108.0  
 angles OPN PN OPN 140.0 120.0  
 angles PN ON CC6 47.0 120.0  
 angles PN ON CH2E 47.0 120.0!{COLOCADO PARA INIBIDOR}  
 angles PN OPN CC6 47.0 120.0  
 dihedral X CH2E CH1E X 4.8 3 0.0!{COLOCADO PARA INIBIDOR}  
 dihedral X CC CC X 3.0630 3 0.0  
 dihedral X CC CC6 X 3.0630 3 0.0  
 dihedral X CCE CC X 3.0630 3 0.0  
 dihedral X CCA CC X 3.0630 3 0.0  
 dihedral X CC OH1 X 1.3293 3 0.0  
 dihedral X CC6 OH1 X 1.3293 3 0.0  
 dihedral X CCE OH1 X 1.3293 3 0.0  
 dihedral X CCA OH1 X 1.3293 3 0.0  
 dihedral X CC OA X 1.3293 3 0.0  
 dihedral X CC6 OA X 1.3293 3 0.0  
 dihedral X CCE OA X 1.3293 3 0.0  
 dihedral X CCA OA X 1.3293 3 0.0  
 dihedral X CC OE X 2.7834 3 0.0

```

dihedral X CCE OE X 2.7834 3 0.0
dihedral X CCA OE X 2.7834 3 0.0
{The following are guesses, generally by analogy to extended atom Carbons}
dihedral HA CC NH1 H 0.3 3 180.0 !for NAG and NANA
dihedral X CC NH1 X 0.9 3 0.0 !beta link Asn Nd1
dihedral X CCE NH1 X 0.9 3 0.0 !beta link Asn Nd1
dihedral X CCA NH1 X 0.9 3 0.0 !beta link Asn Nd1
dihedral X C CCE X 0.0 3 0.0
dihedral X C CCA X 0.0 3 0.0
dihedral X C CC X 0.0 3 0.0
dihedral X ON CC6 X 0.5 3 0.0          {*Threefold.*}
dihedral X ON CH2E X 0.5 3 0.0 !{COLOCADO P/INIBIDOR}
dihedral X PN ON X 0.5 3 0.0          {*Threefold.*}
dihedral X PN OPN X 0.5 3 0.0        {*Threefold.*}
{These high force constants maintain chiral centers on the rings}
improper CC X X CC 300.0 0 35.26439
improper CCE X X CC 300.0 0 35.26439
improper CCA X X CC 300.0 0 35.26439
improper OH1 X X CC 300.0 0 35.26439
improper OH1 X X CCE 300.0 0 35.26439
improper OH1 X X CCA 300.0 0 35.26439
improper OH1 X X CH1E 300.0 0 35.26439
!{COLOCADO PARA INIBIDOR}
improper OA X X CC 300.0 0 35.26439
improper OA X X CCE 300.0 0 35.26439
improper OA X X CCA 300.0 0 35.26439
improper CC X X OE 300.0 0 35.26439
improper CCE X X OE 300.0 0 35.26439
improper CCA X X OE 300.0 0 35.26439
{This is as other acids}
improper C X X CCE 300.0 0 0.0!NANA acid planarity
improper C X X CCA 300.0 0 0.0!NANA acid planarity
!!for use with:
!!NBXMOD=5 ATOM CDIEL SHIFT vs witch
!CUTNB=8.0 CTOFNB=7.5 CTONNB=6.5 EPS=1.0 E14FAC=0.4 WMIN=1.5
!NONBonded HA 0.0045 2.6157 0.0045 2.6157
NONBonded CC 0.0903 3.2072 0.0903 3.2072
NONBonded CC6 0.0903 3.2072 0.0903 3.2072
NONBonded CCE 0.0903 3.2072 0.0903 3.2072
NONBonded CCA 0.0903 3.2072 0.0903 3.2072
NONBonded OE 0.1591 2.8509 0.1591 2.8509
NONBonded OA 0.1591 2.8509 0.1591 2.8509

{* Lennard-Jones parameters *}
{* -----1-4----- *}
{* epsilon sigma epsilon sigma *}
{* (kcal/mol) (A) (kcal/mol) (A) *}
nonbonded HN 0.0498 1.4254 0.0498 1.4254
nonbonded ON 0.1591 2.8509 0.1591 2.8509
nonbonded OPN 0.1591 2.8509 0.1591 2.8509
nonbonded PN 0.5849 3.3854 0.5849 3.3854
set echo=true end

```

## APÊNDICE B

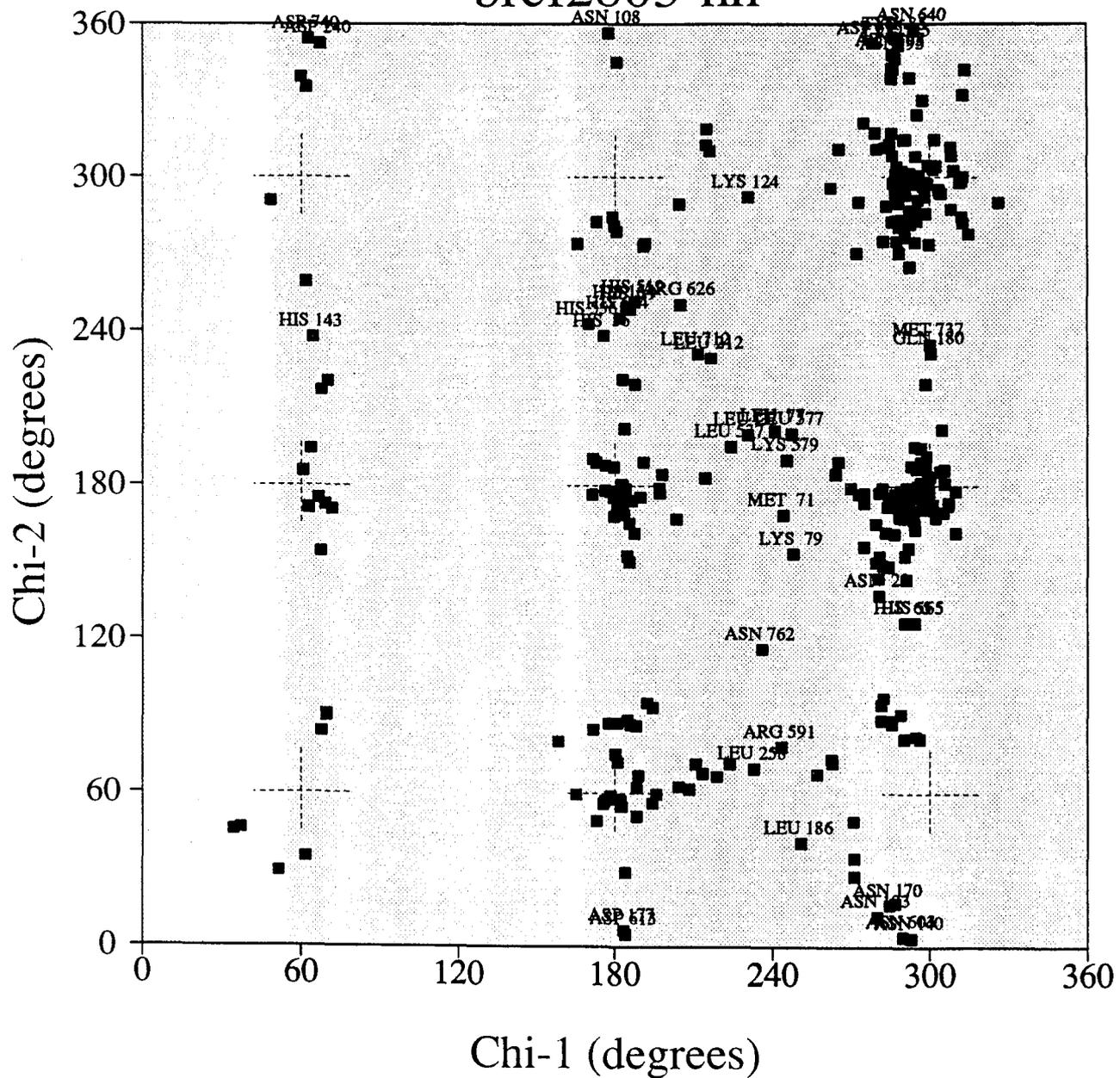
Gráficos de análise de qualidade estereoquímica feitos pelo programa  
*PROCHECK* (Moris, A.L. *et al.*, 1992) para a estrutura nativa

- Gráfico dos ângulos Qui1 *versus* Qui2 dos resíduos..... B1
- Gráficos que analisam os parâmetros da cadeia principal: qualidade do gráfico de Ramachandran, planaridade da ligação peptídica, medida dos maus contatos de átomos não ligados, distorção dos ângulos do carbono  $\alpha$ , energias das pontes de hidrogênio.....B2
- Gráficos que analisam os ângulos diédricos da cadeia lateral..... B3
- Gráficos que analisam propriedades dos resíduos: desvio do valor médio do ângulo Qui 1, desvio do valor médio do ângulo ômega 1, planaridade do carbono  $\alpha$ ..... B4

# Plot of Chi-1 vs Chi-2

bref2803-nh

B-1



Total number of residues = 532

Number of residues plotted = 322

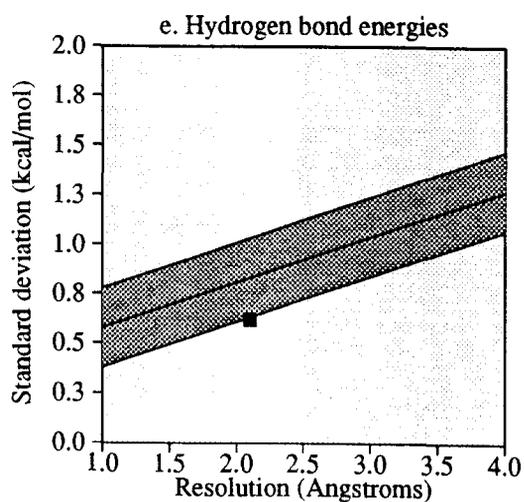
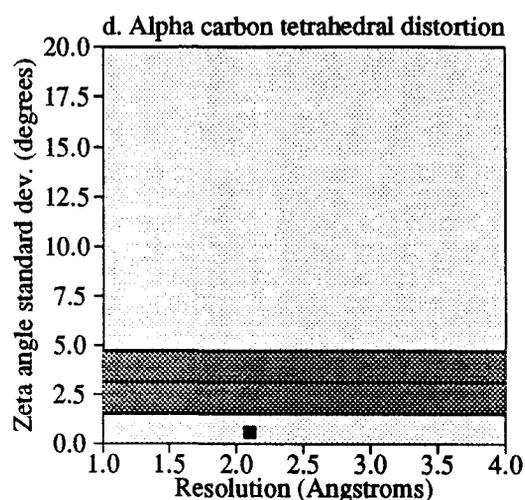
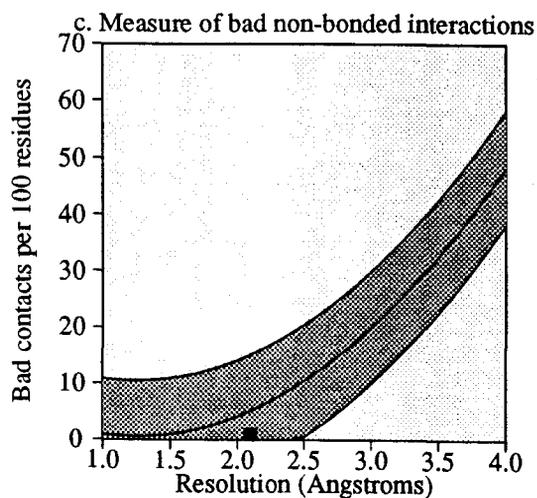
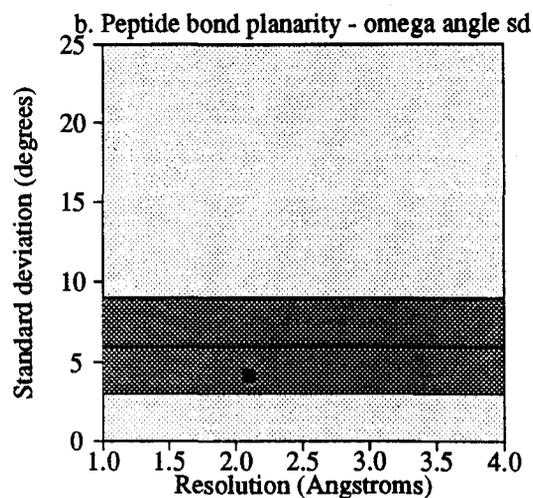
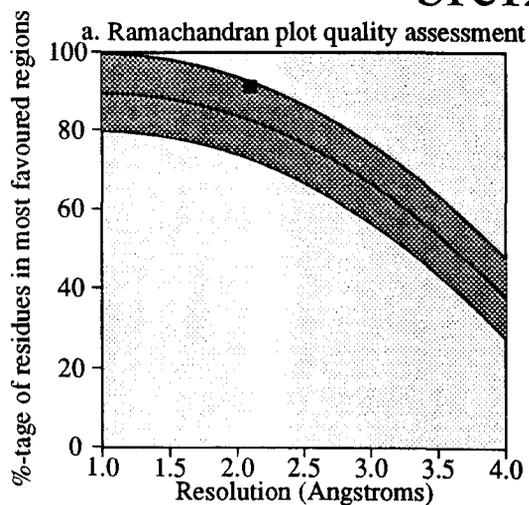
Number of labelled residues = 42

The dashed crosses represent the gauche minus, trans, and gauche plus regions for the Chi-1 and Chi-2 dihedral angles. The width of each cross is approximately one standard deviation. The points should therefore cluster around these crosses. Points more than 2.5 standard deviations from the ideal are labelled

# Main-chain parameters

## bref2803-nh

B-2



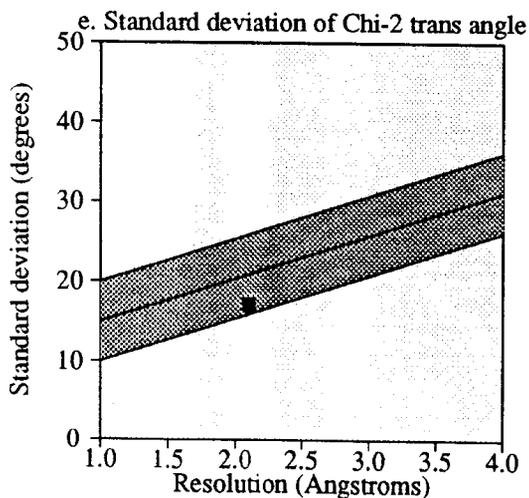
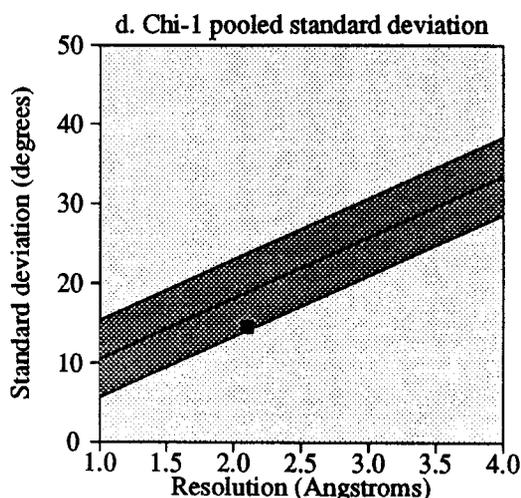
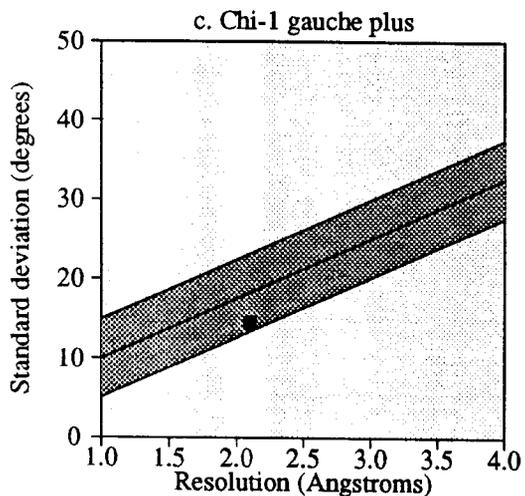
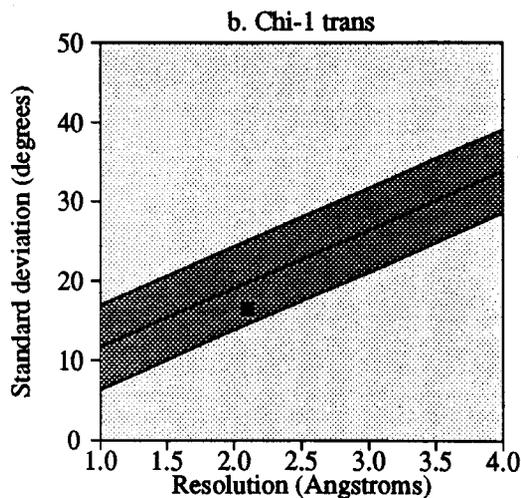
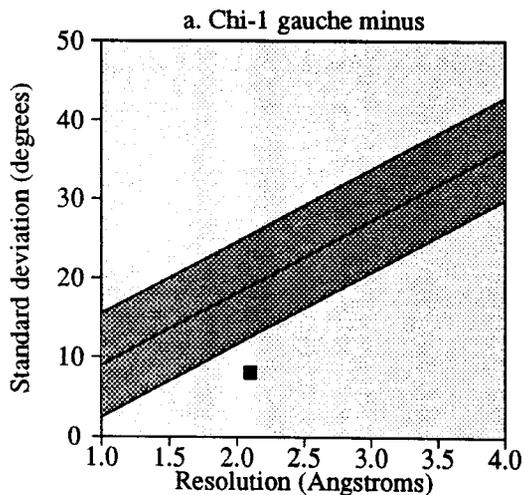
bref2803-nh

### Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values		No. of band widths from mean	
			Typical value	Band width		
a. %-tage residues in A, B, L	468	91.2	82.6	10.0	0.9	Better
b. Omega angle st dev	530	4.1	6.0	3.0	-0.6	Better
c. Bad contacts / 100 residues	5	0.9	5.2	10.0	-0.4	Better
d. Zeta angle st dev	498	0.6	3.1	1.6	-1.6	BETTER
e. H-bond energy st dev	339	0.6	0.8	0.2	-1.1	BETTER

# Side-chain parameters bref2803-nh

B-3



bref2803-nh

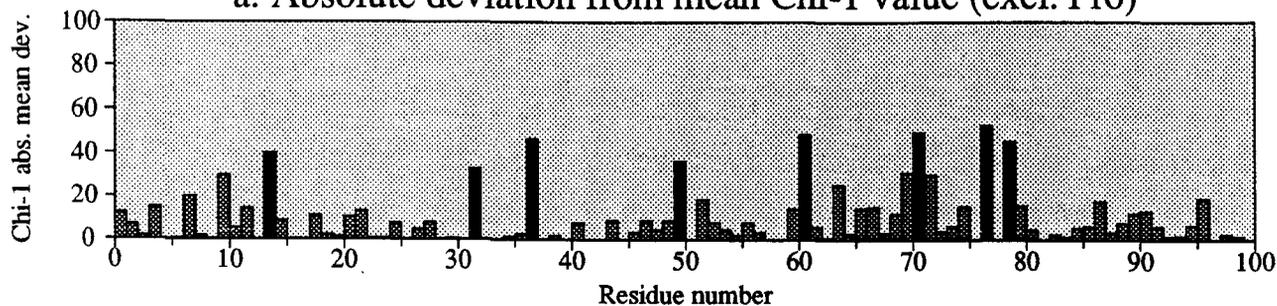
## Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values		No. of band widths from mean	
			Typical value	Band width		
a. Chi-1 gauche minus st dev	55	8.1	19.0	6.5	-1.7	BETTER
b. Chi-1 trans st dev	135	16.4	19.7	5.3	-0.6	Better
c. Chi-1 gauche plus st dev	238	14.5	18.3	4.9	-0.8	Better
d. Chi-1 pooled st dev	428	14.5	18.9	4.8	-0.9	Better
e. Chi-2 trans st dev	135	17.1	20.9	5.0	-0.8	Better

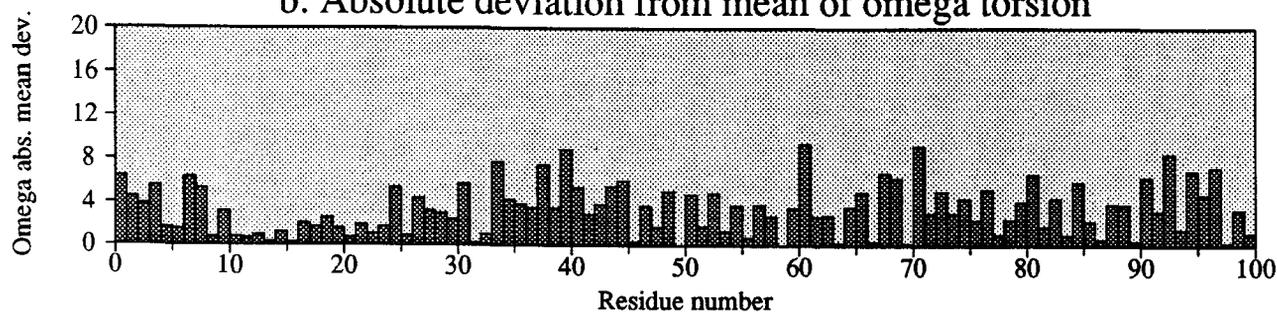
# Residue properties

## bref2803-nh

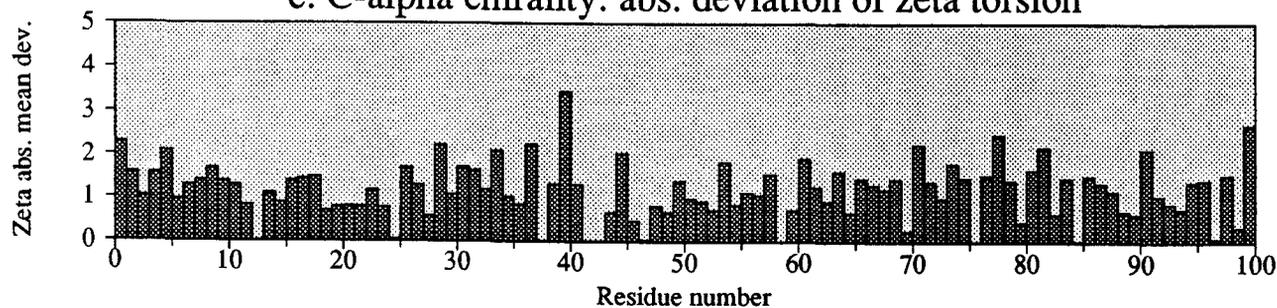
a. Absolute deviation from mean Chi-1 value (excl. Pro)



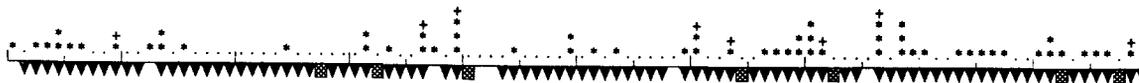
b. Absolute deviation from mean of omega torsion



c. C-alpha chirality: abs. deviation of zeta torsion



d. Max. deviation and Ramachandran plot region (see listing)



e. Sequence

MRLIPLTTAEQVGKWAARHIVNRINAFKPTADRPFVLGLPTGGTPMTTYKALVEMHKAGQVSFKHVVTFNMDEYVGLPKEHPESYYSFMHRNFFNHVDIP

f. Secondary structure



Key:-

~ Helix

□ Beta strand

— Random coil

Ramachandran plot regions:

▼ Most favoured

■ Allowed

▨ Generous

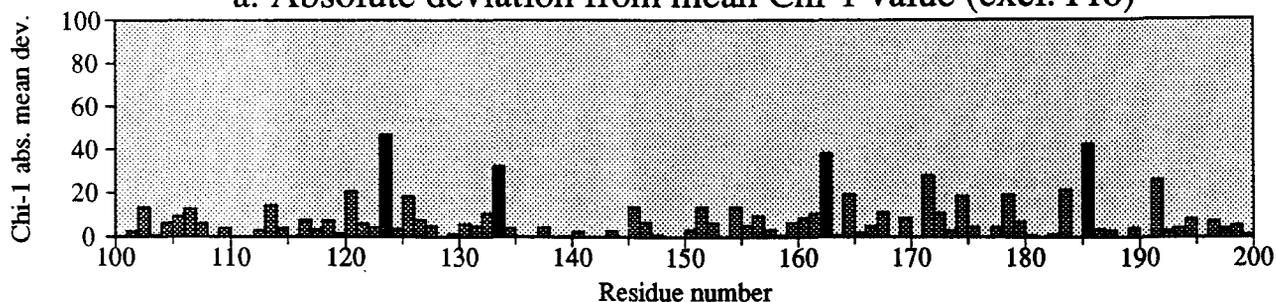
■ Disallowed

Highlighted residues in top graphs  
are those that deviate by more  
than 2.0 st. devs. from ideal

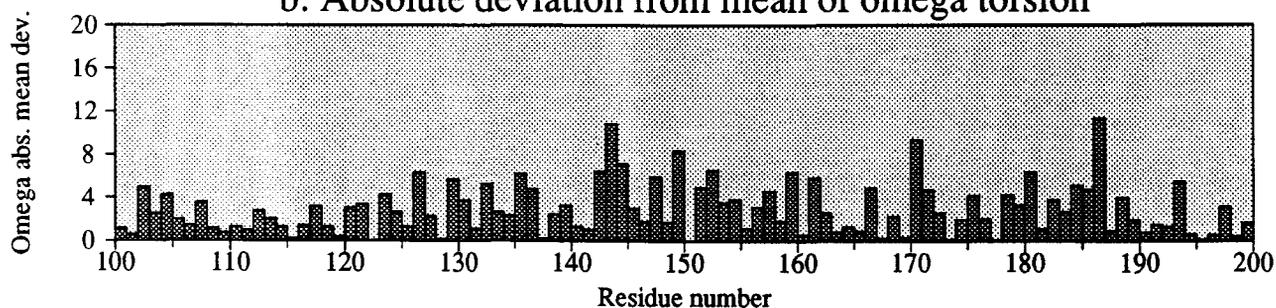
# Residue properties

## bref2803-nh

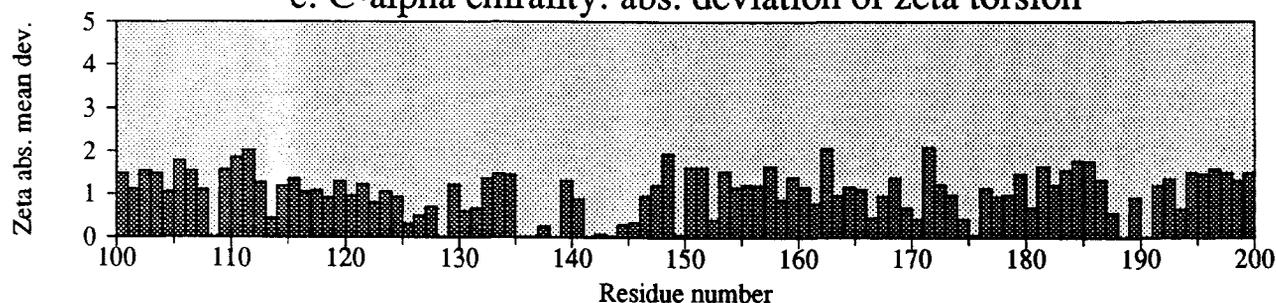
a. Absolute deviation from mean Chi-1 value (excl. Pro)



b. Absolute deviation from mean of omega torsion



c. C-alpha chirality: abs. deviation of zeta torsion



d. Max. deviation and Ramachandran plot region (see listing)



e. Sequence

AENINLLNGNAPDIDAECRQYEEKIRSYGKIHLFMGGVNDGHI AFNEPASSLASRTRIKTLTHDTRVANSRFFDNDVNQVPKYALTVGVGTLILDABEVM

f. Secondary structure



Key:-

Helix  
 Beta strand  
 Random coil

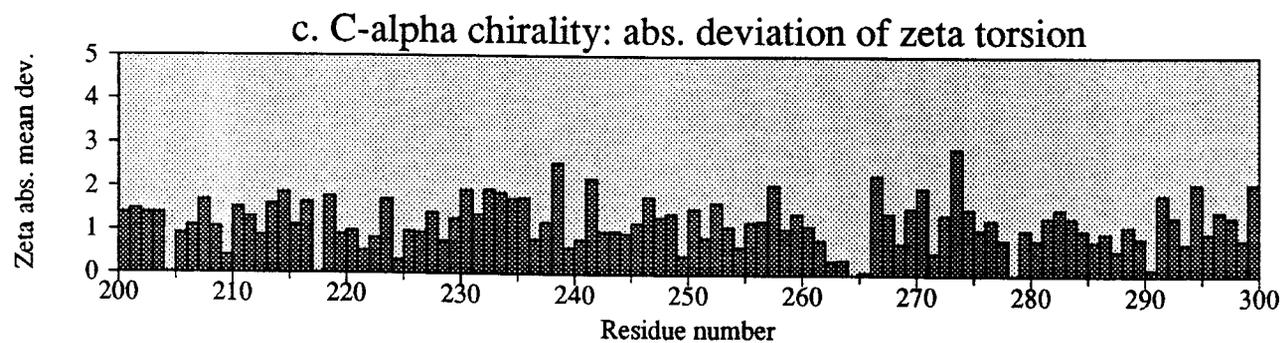
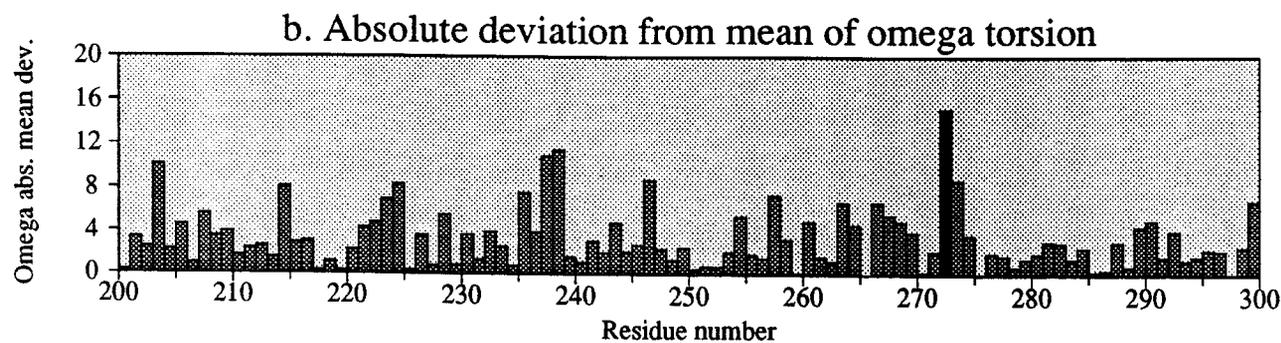
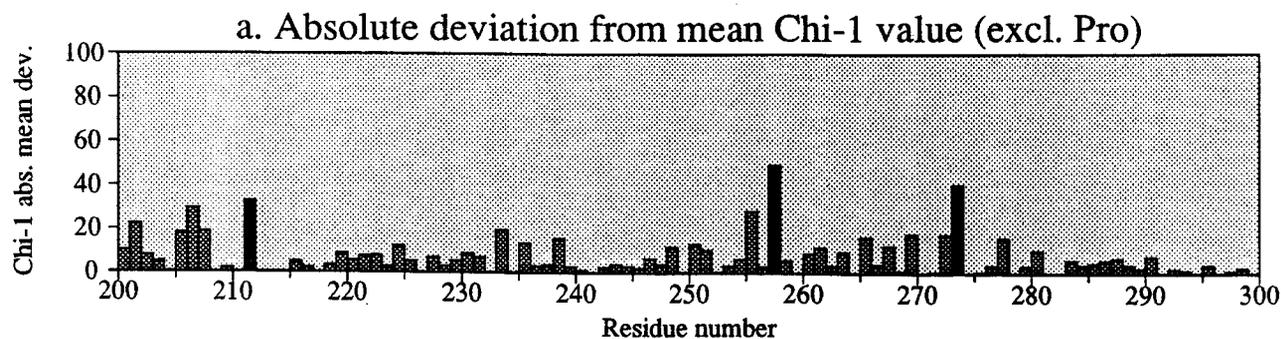
Ramachandran plot regions:

Most favoured  
 Allowed  
 Generous  
 Disallowed

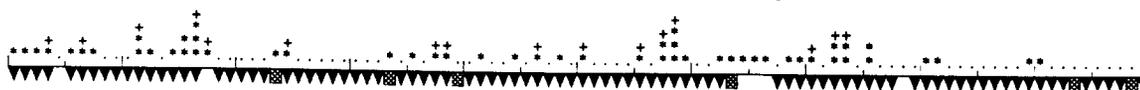
Highlighted residues in top graphs are those that deviate by more than 2.0 st. devs. from ideal

# Residue properties

## bref2803-nh



d. Max. deviation and Ramachandran plot region (see listing)



e. Sequence

ILVLGSQKALALQAAVEGCVNHMTIISCLQLHPKAIMVCEPSTMELKVKTLRYFNELEAENIKGLMRLIPLTTAEQVGKWAARHIVNRIINAFKPTADRP

f. Secondary structure



Key:-

Helix  
 Beta strand  
 Random coil

Ramachandran plot regions:

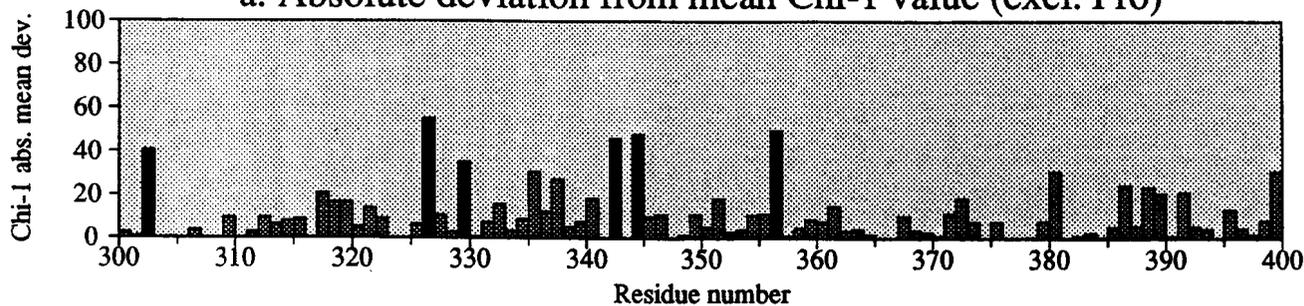
Most favoured  
 Allowed  
 Generous  
 Disallowed

Highlighted residues in top graphs are those that deviate by more than 2.0 st. devs. from ideal

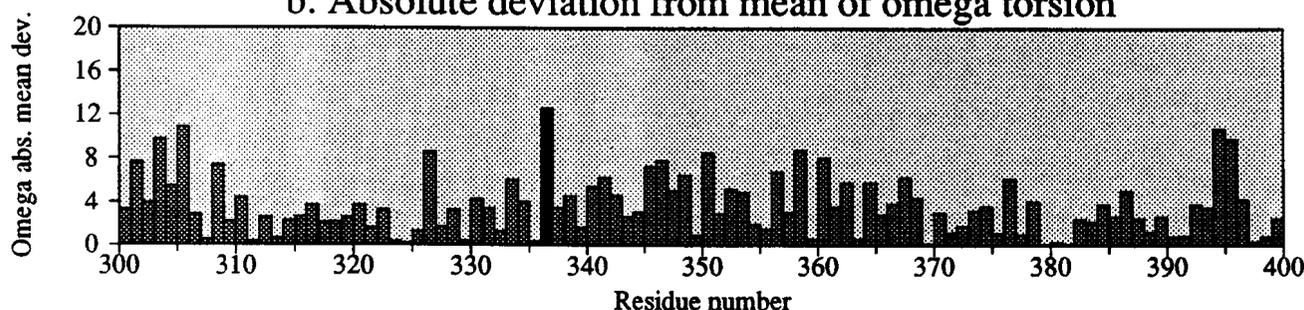
# Residue properties

## bref2803-nh

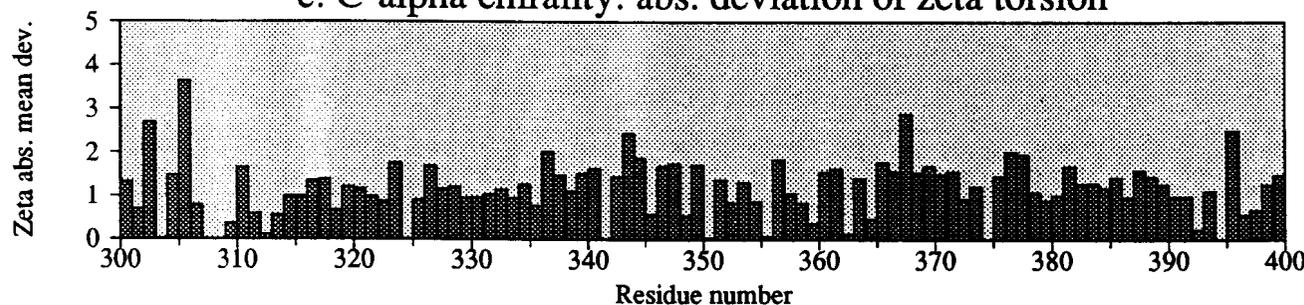
a. Absolute deviation from mean Chi-1 value (excl. Pro)



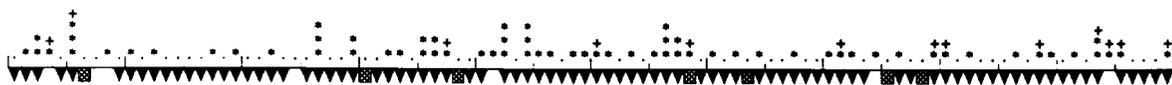
b. Absolute deviation from mean of omega torsion



c. C-alpha chirality: abs. deviation of zeta torsion



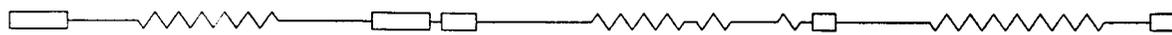
d. Max. deviation and Ramachandran plot region (see listing)



e. Sequence

FVLGLPTGGTPMTTYKALVEMHKAGQVSFKHVVTFTNMDYVGLPKEHPESYYSFMRNFFNHVDI PAENINLLNGNAPDIDAECRQYEEKIRSYGKIHLF

f. Secondary structure



Key:-

Helix  
 Beta strand  
 Random coil

Ramachandran plot regions:

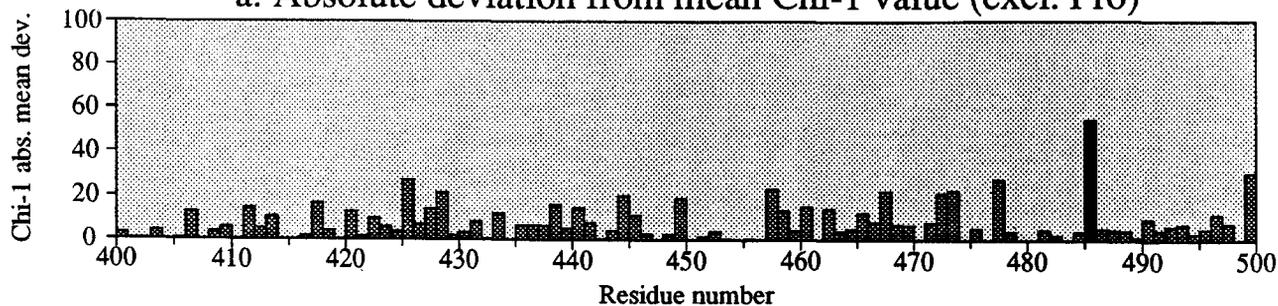
Most favoured  
 Allowed  
 Disallowed

Highlighted residues in top graphs are those that deviate by more than 2.0 st. devs. from ideal

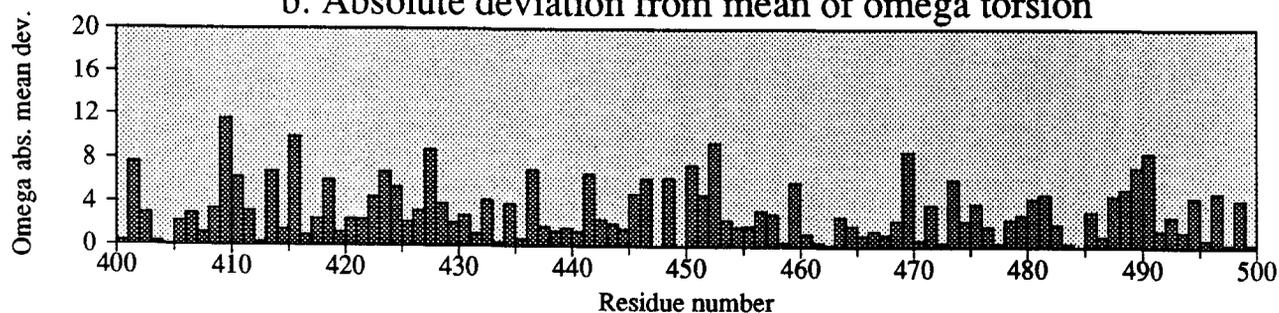
# Residue properties

## bref2803-nh

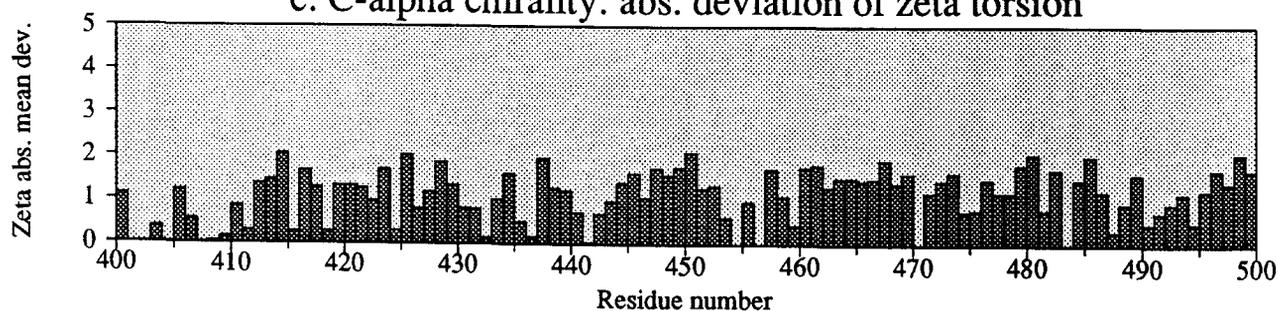
a. Absolute deviation from mean Chi-1 value (excl. Pro)



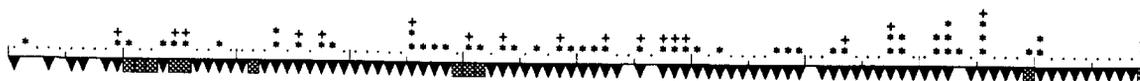
b. Absolute deviation from mean of omega torsion



c. C-alpha chirality: abs. deviation of zeta torsion



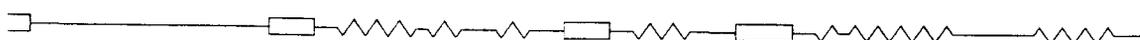
d. Max. deviation and Ramachandran plot region (see listing)



e. Sequence

MGGVGN<sup>+</sup>DGHI<sup>+</sup>AFNEPAS<sup>+</sup>SLAS<sup>+</sup>TRIK<sup>+</sup>TLTHD<sup>+</sup>TRVANS<sup>+</sup>RFFD<sup>+</sup>NDVN<sup>+</sup>QVPKY<sup>+</sup>ALTVGV<sup>+</sup>GTL<sup>+</sup>LDAA<sup>+</sup>EEVMI<sup>+</sup>LVLGS<sup>+</sup>QKALAL<sup>+</sup>QAA<sup>+</sup>VEGCV<sup>+</sup>NHMWT<sup>+</sup>ISCL<sup>+</sup>QLHPK

f. Secondary structure



Key:-

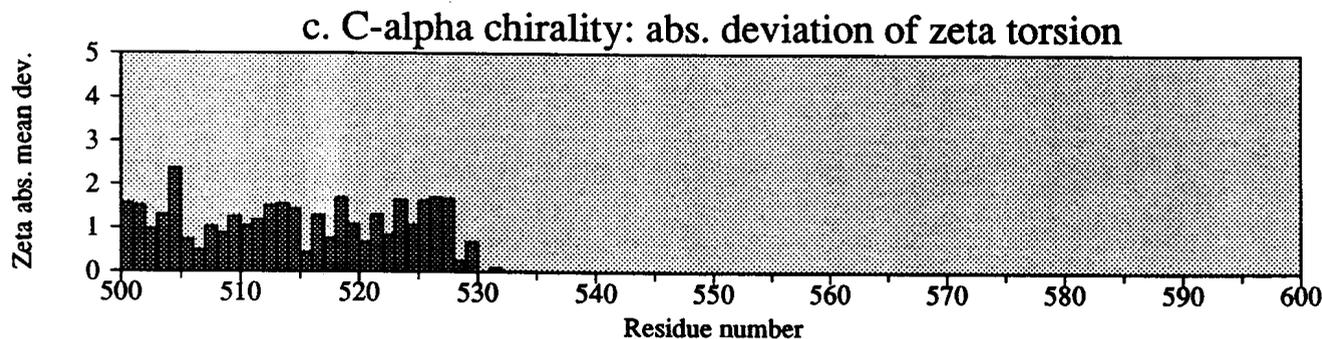
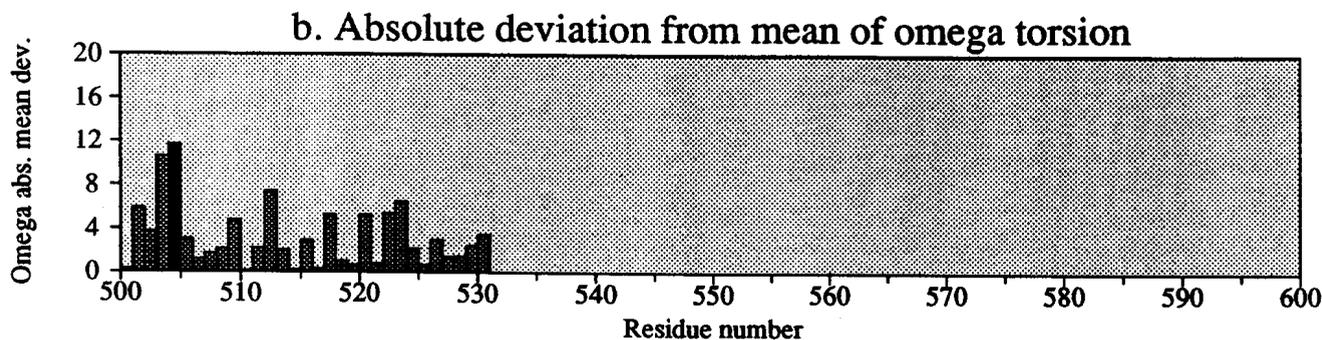
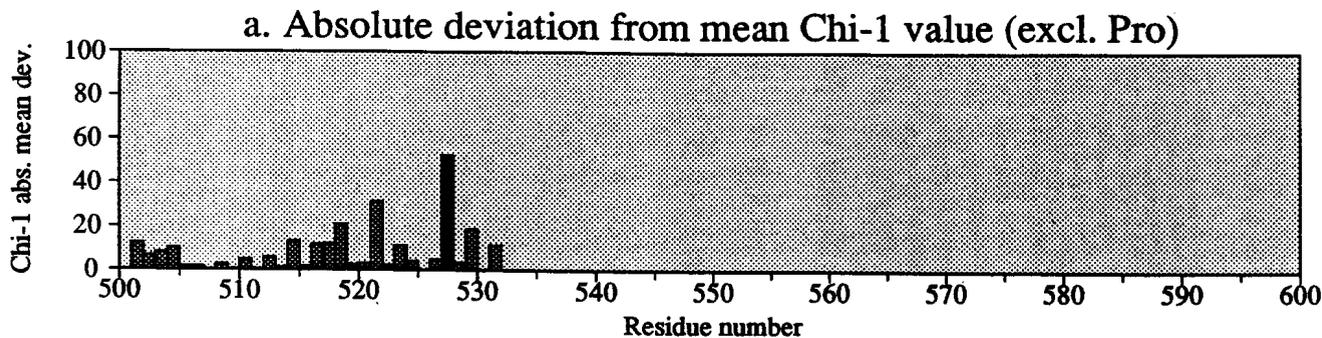
Helix  
 Beta strand  
 Random coil

Ramachandran plot regions:

Most favoured  
 Allowed  
 Generous  
 Disallowed

Highlighted residues in top graphs are those that deviate by more than 2.0 st. devs. from ideal

# Residue properties bref2803-nh



d. Max. deviation and Ramachandran plot region (see listing)



e. Sequence

A I M V C D E P S T M E L K V K T L R Y F N E L E A E N I K G L

f. Secondary structure



Key:-

Helix  
 Beta strand  
 Random coil

Ramachandran plot regions:

Most favoured  
 Allowed  
 Generous  
 Disallowed

Highlighted residues in top graphs are those that deviate by more than 2.0 st. devs. from ideal

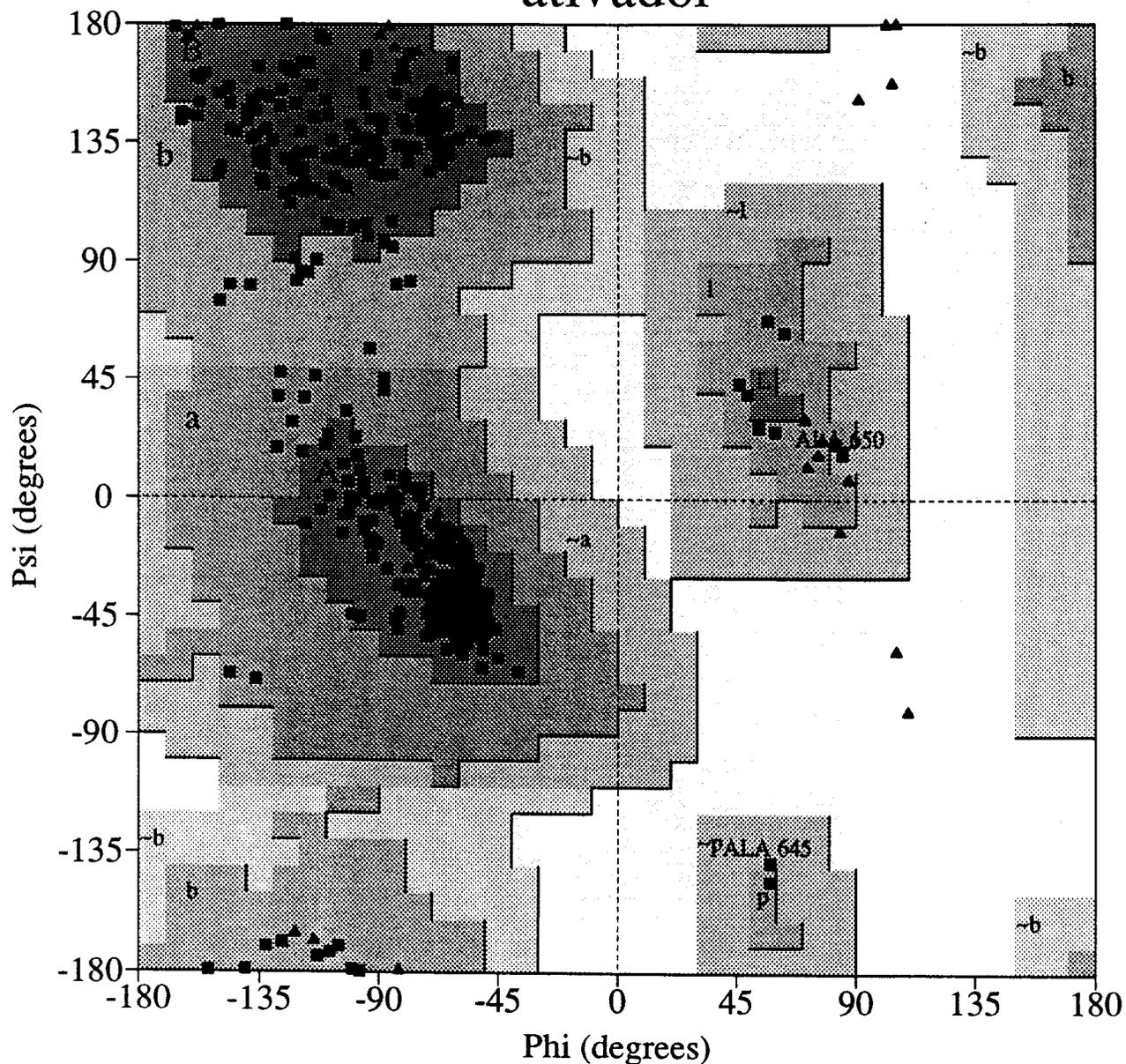
## APÊNDICE C

Gráficos de análise da qualidade estereoquímica feitos pelo programa *PROCHECK* (Moris, A.L. *et al.*, 1992) para os complexos: enzima ativador- $P_i$ , enzima inibidor- $P_i$  e enzima 2-inibidor.

- Gráfico de Ramachandran da enzima ativador- $P_i$  ..... C1
- Gráfico dos ângulos  $Qui1$  versus  $Qui2$  dos resíduos da enzima ativador- $P_i$ ..... C2
- Gráficos que analisam os parâmetros da cadeia principal: qualidade do gráfico de Ramachandran, planaridade da ligação peptídica, medida dos maus contatos de átomos não ligados, distorção dos ângulos do carbono  $\alpha$ , energias das pontes de hidrogênio da enzima ativador- $P_i$ ..... C3
- Gráficos que analisam os ângulos diédricos da cadeia lateral da enzima ativador- $P_i$ ..... C4
- Gráfico de Ramachandran da enzima inibidor- $P_i$  ..... C5
- Gráfico dos ângulos  $Qui1$  versus  $Qui2$  dos resíduos da enzima inibidor- $P_i$ ..... C6
- Gráficos que analisam os parâmetros da cadeia principal: qualidade do gráfico de Ramachandran, planaridade da ligação peptídica, medida dos maus contatos de átomos não ligados, distorção dos ângulos do carbono  $\alpha$ , energias das pontes de hidrogênio da enzima inibidor- $P_i$ ..... C7
- Gráficos que analisam os ângulos diédricos da cadeia lateral da enzima inibidor- $P_i$ ..... C8
- Gráfico de Ramachandran da enzima 2 inibidor..... C9
- Gráfico dos ângulos  $Qui1$  versus  $Qui2$  dos resíduos da enzima 2-inibidor..... C10
- Gráficos que analisam os parâmetros da cadeia principal: qualidade do gráfico de Ramachandran, planaridade da ligação peptídica, medida dos maus contatos de átomos não ligados, distorção dos ângulos do carbono  $\alpha$ , energias das pontes de hidrogênio da enzima 2-inibidor..... C11
- Gráficos que analisam os ângulos diédricos da cadeia lateral da enzima 2-inibidor..... C12

# Ramachandran Plot

## ativador

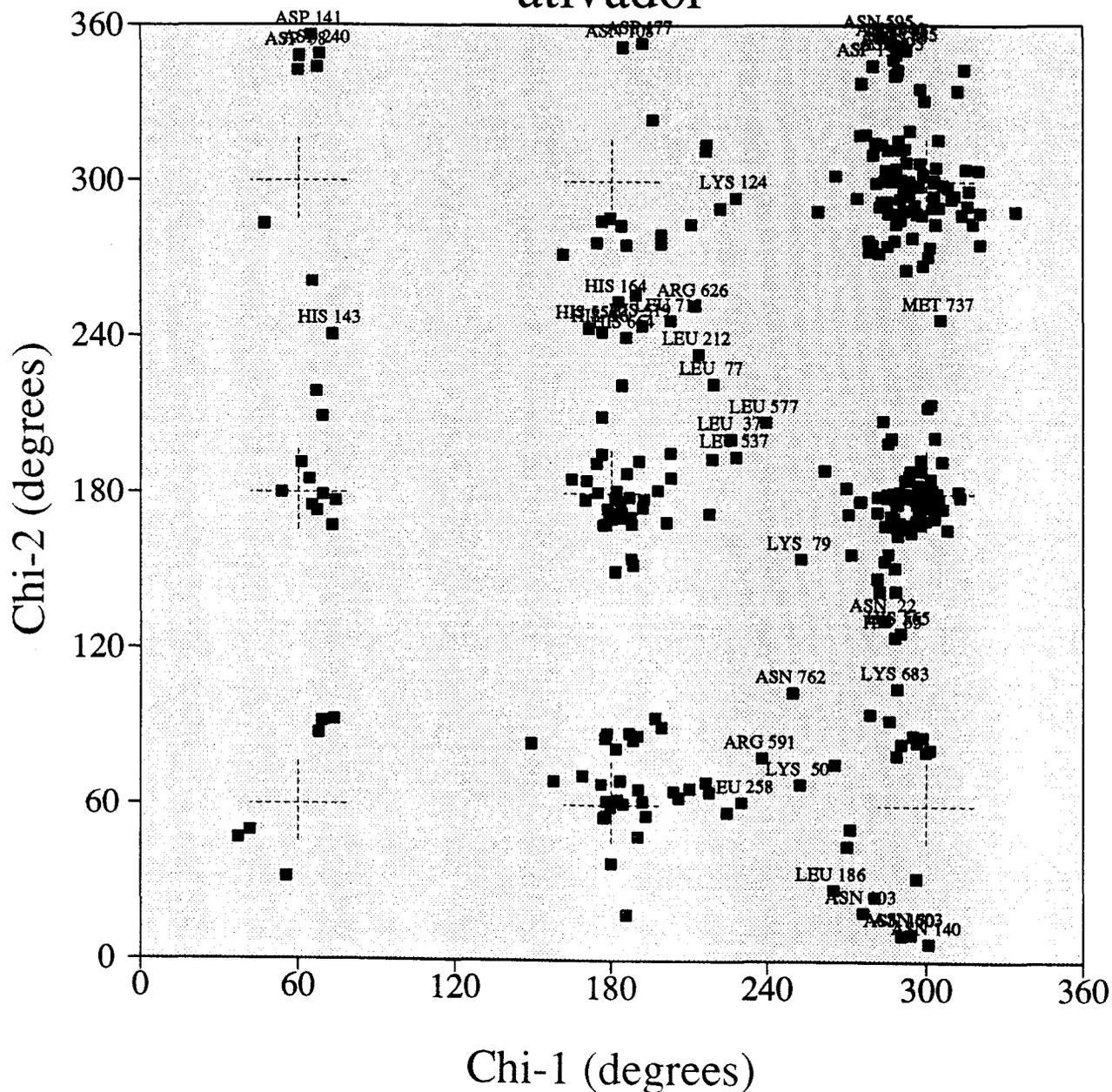


### Plot statistics

Residues in most favoured regions [A,B,L]	422	90.2%
Residues in additional allowed regions [a,b,l,p]	44	9.4%
Residues in generously allowed regions [-a,-b,-l,-p]	2	0.4%
Residues in disallowed regions	0	0.0%
-----		
Number of non-glycine and non-proline residues	468	100.0%
Number of end-residues	4	
Number of glycine residues (shown as triangles)	34	
Number of proline residues	26	
-----		
Total number of residues	532	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

# Plot of Chi-1 vs Chi-2 ativador



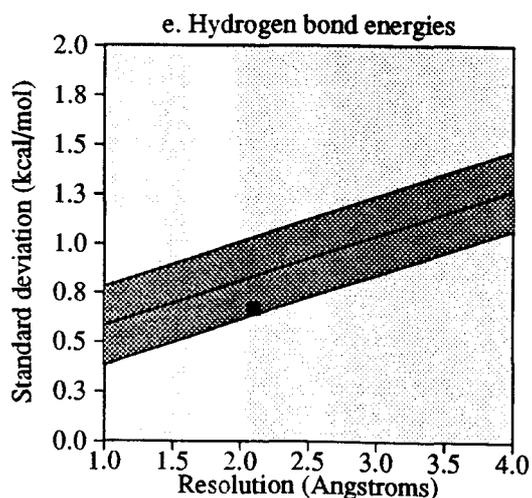
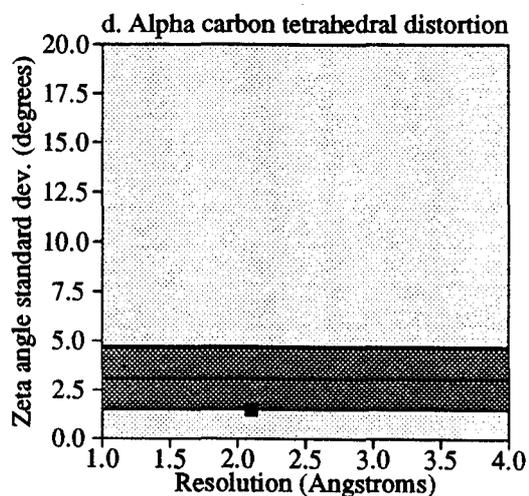
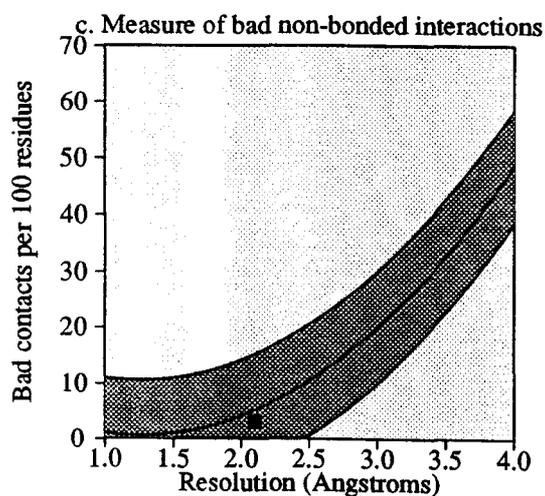
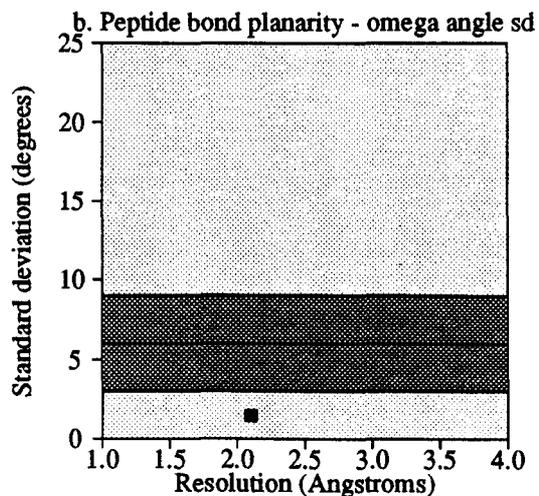
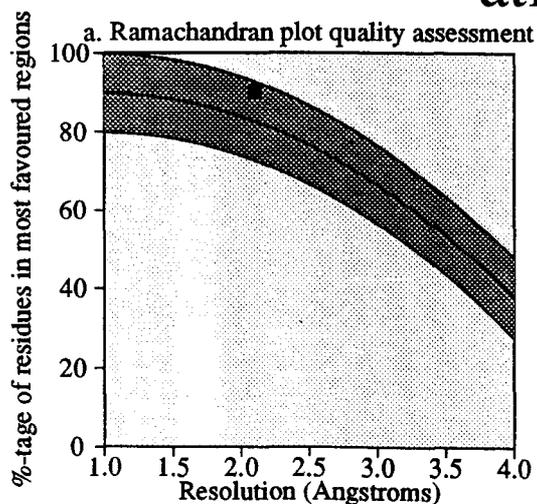
Total number of residues = 532

Number of residues plotted = 322

Number of labelled residues = 41

The dashed crosses represent the gauche minus, trans, and gauche plus regions for the Chi-1 and Chi-2 dihedral angles. The width of each cross is approximately one standard deviation. The points should therefore cluster around these crosses. Points more than 2.5 standard deviations from the ideal are labelled

# Main-chain parameters ativador

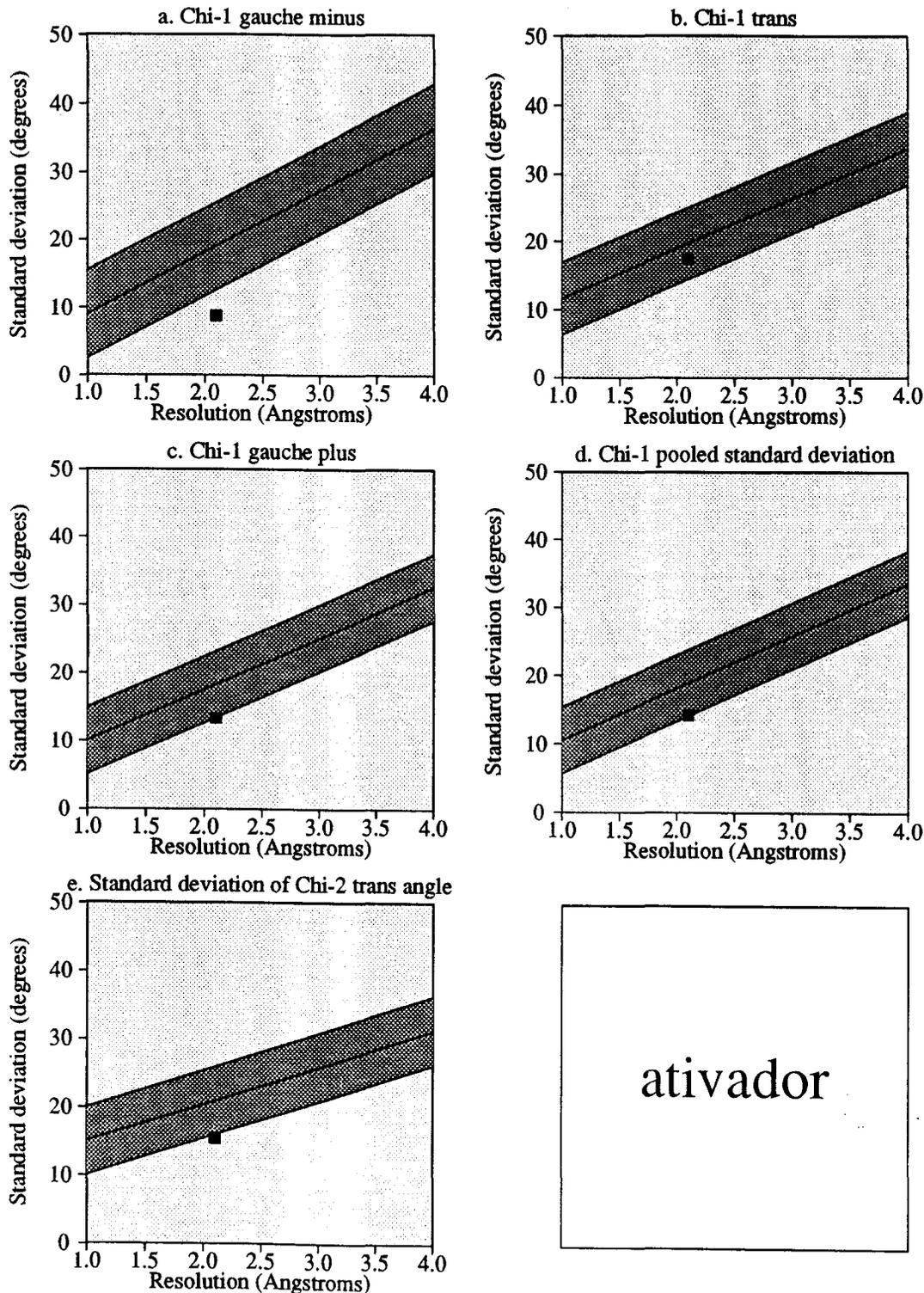


ativador

## Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values		No. of band widths from mean	
			Typical value	Band width		
a. %-tage residues in A, B, L	468	90.2	82.6	10.0	0.8	Better
b. Omega angle st dev	530	1.4	6.0	3.0	-1.5	BETTER
c. Bad contacts / 100 residues	16	3.0	5.2	10.0	-0.2	Better
d. Zeta angle st dev	498	1.5	3.1	1.6	-1.0	BETTER
e. H-bond energy st dev	342	0.7	0.8	0.2	-0.8	Better

# Side-chain parameters ativador



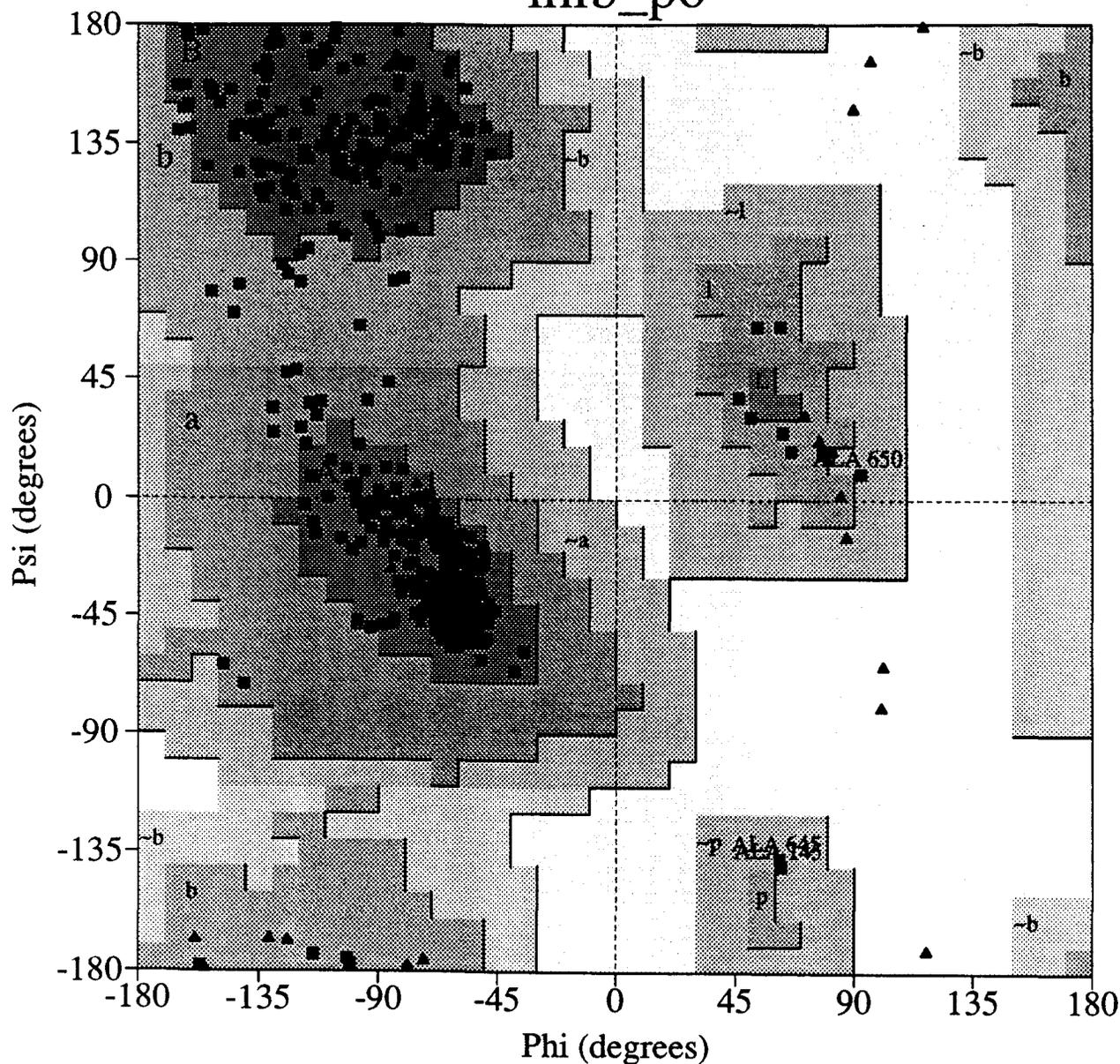
ativador

## Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values		No. of band widths from mean	
			Typical value	Band width		
a. Chi-1 gauche minus st dev	55	8.8	19.0	6.5	-1.6	BETTER
b. Chi-1 trans st dev	137	17.4	19.7	5.3	-0.4	Better
c. Chi-1 gauche plus st dev	236	13.3	18.3	4.9	-1.0	BETTER
d. Chi-1 pooled st dev	428	14.2	18.9	4.8	-1.0	Better
e. Chi-2 trans st dev	132	15.4	20.9	5.0	-1.1	BETTER

# Ramachandran Plot

## inib\_po



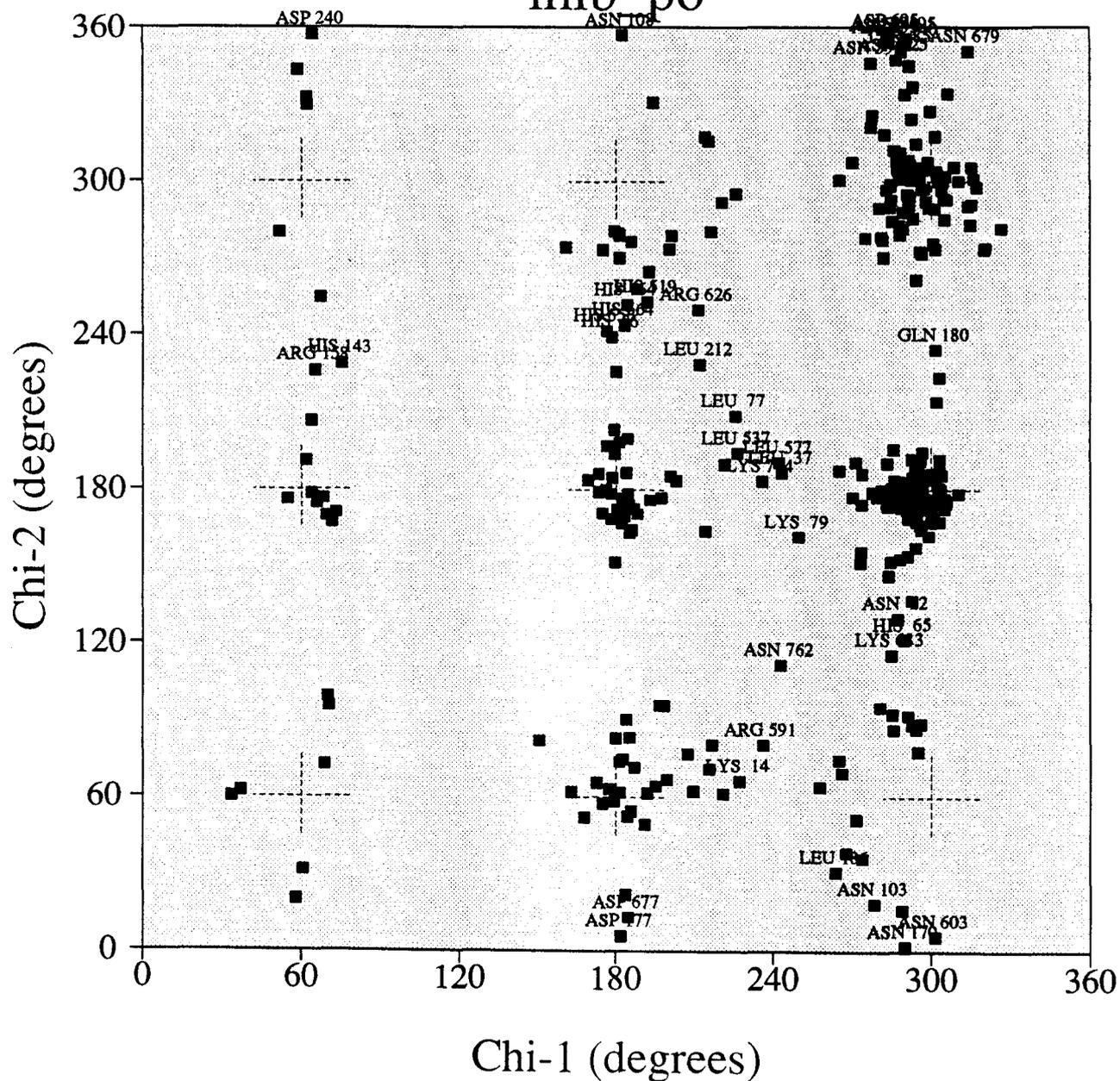
### Plot statistics

Residues in most favoured regions [A,B,L]	423	90.4%
Residues in additional allowed regions [a,b,l,p]	42	9.0%
Residues in generously allowed regions [~a,~b,~l,~p]	3	0.6%
Residues in disallowed regions	0	0.0%
-----		
Number of non-glycine and non-proline residues	468	100.0%
Number of end-residues	4	
Number of glycine residues (shown as triangles)	34	
Number of proline residues	26	
-----		
Total number of residues	532	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

# Plot of Chi-1 vs Chi-2

inib po



Total number of residues = 532

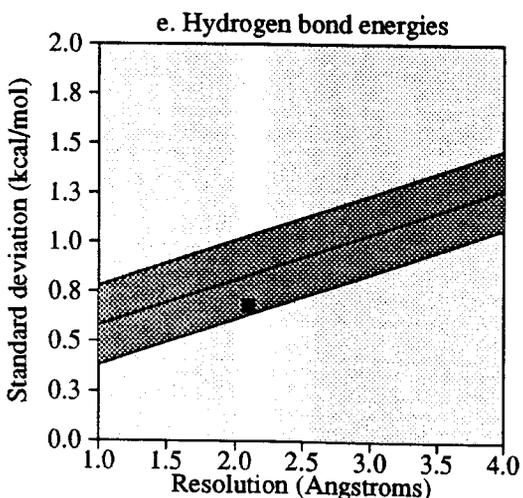
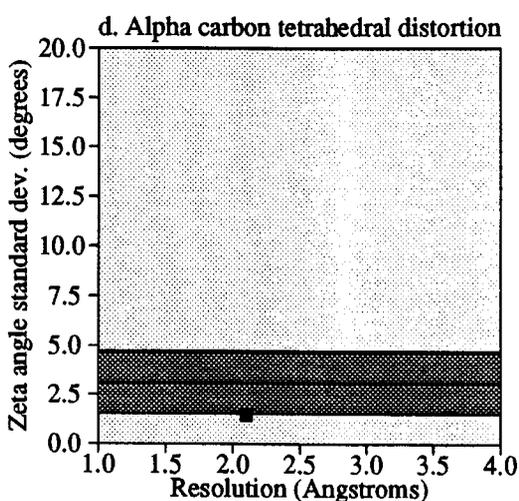
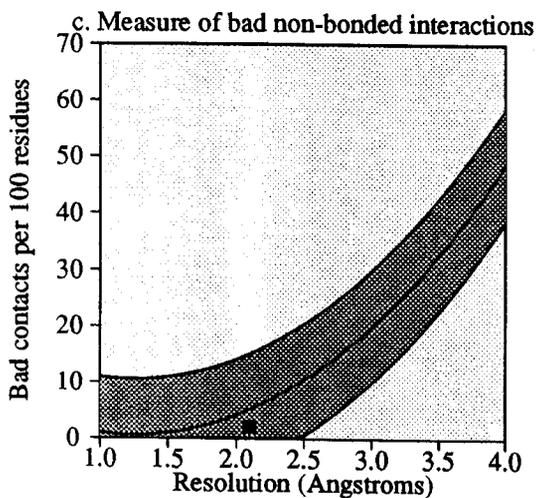
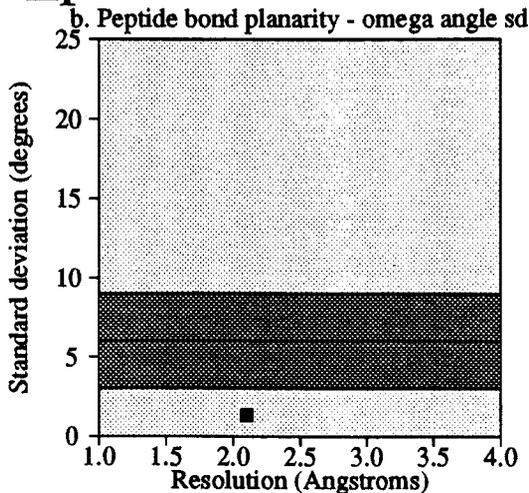
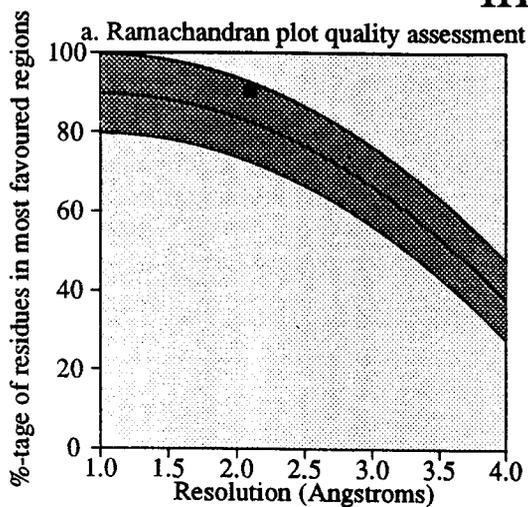
Number of residues plotted = 322

Number of labelled residues = 39

The dashed crosses represent the gauche minus, trans, and gauche plus regions for the Chi-1 and Chi-2 dihedral angles. The width of each cross is approximately one standard deviation. The points should therefore cluster around these crosses. Points more than 2.5 standard deviations from the ideal are labelled

# Main-chain parameters

## inib\_po



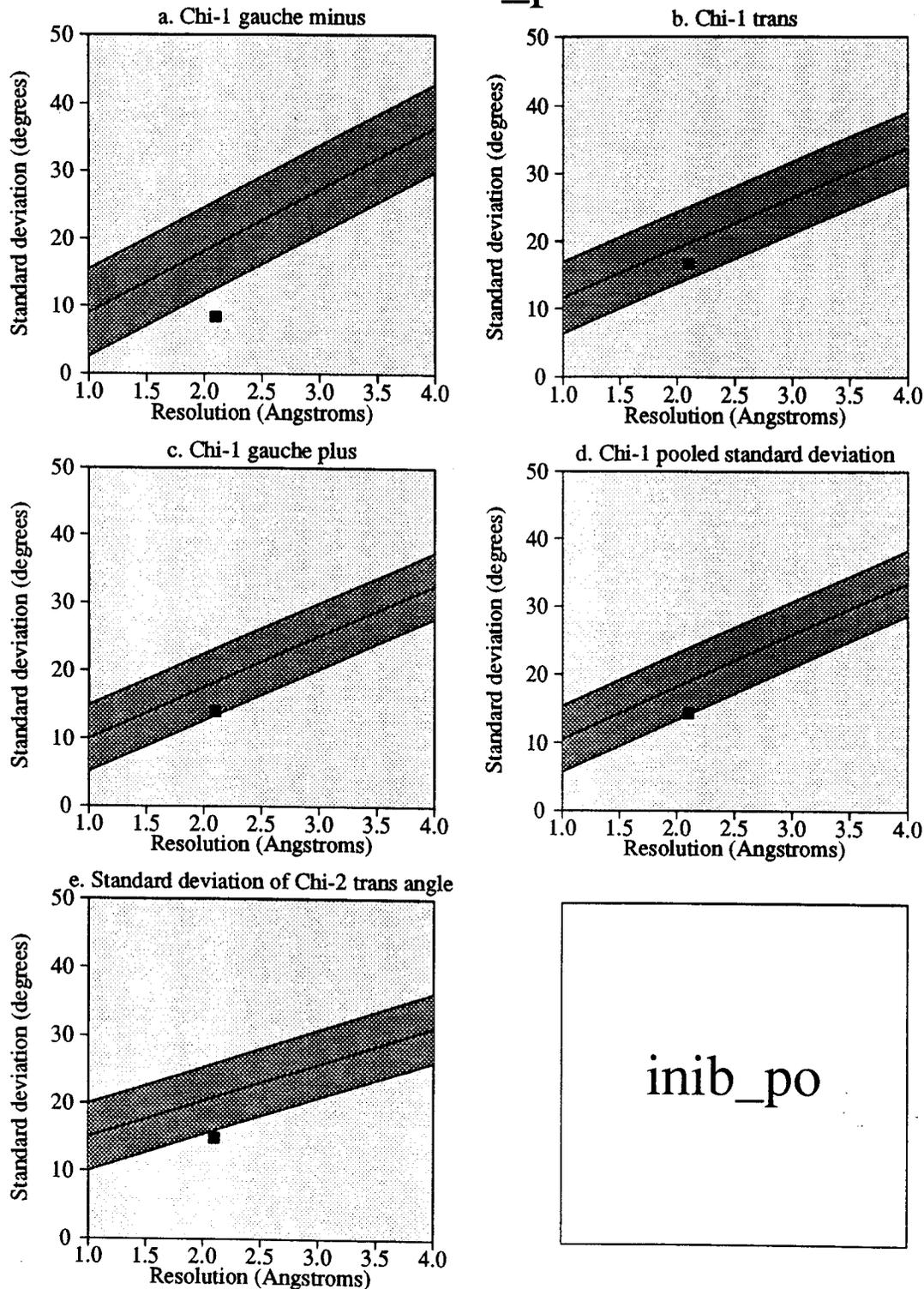
inib\_po

### Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values		No. of band widths from mean	
			Typical value	Band width		
a. %-tage residues in A, B, L	468	90.4	82.6	10.0	0.8	Better
b. Omega angle st dev	530	1.3	6.0	3.0	-1.6	BETTER
c. Bad contacts / 100 residues	11	2.1	5.2	10.0	-0.3	Better
d. Zeta angle st dev	498	1.4	3.1	1.6	-1.0	BETTER
e. H-bond energy st dev	342	0.7	0.8	0.2	-0.8	Better

# Side-chain parameters

## inib\_po

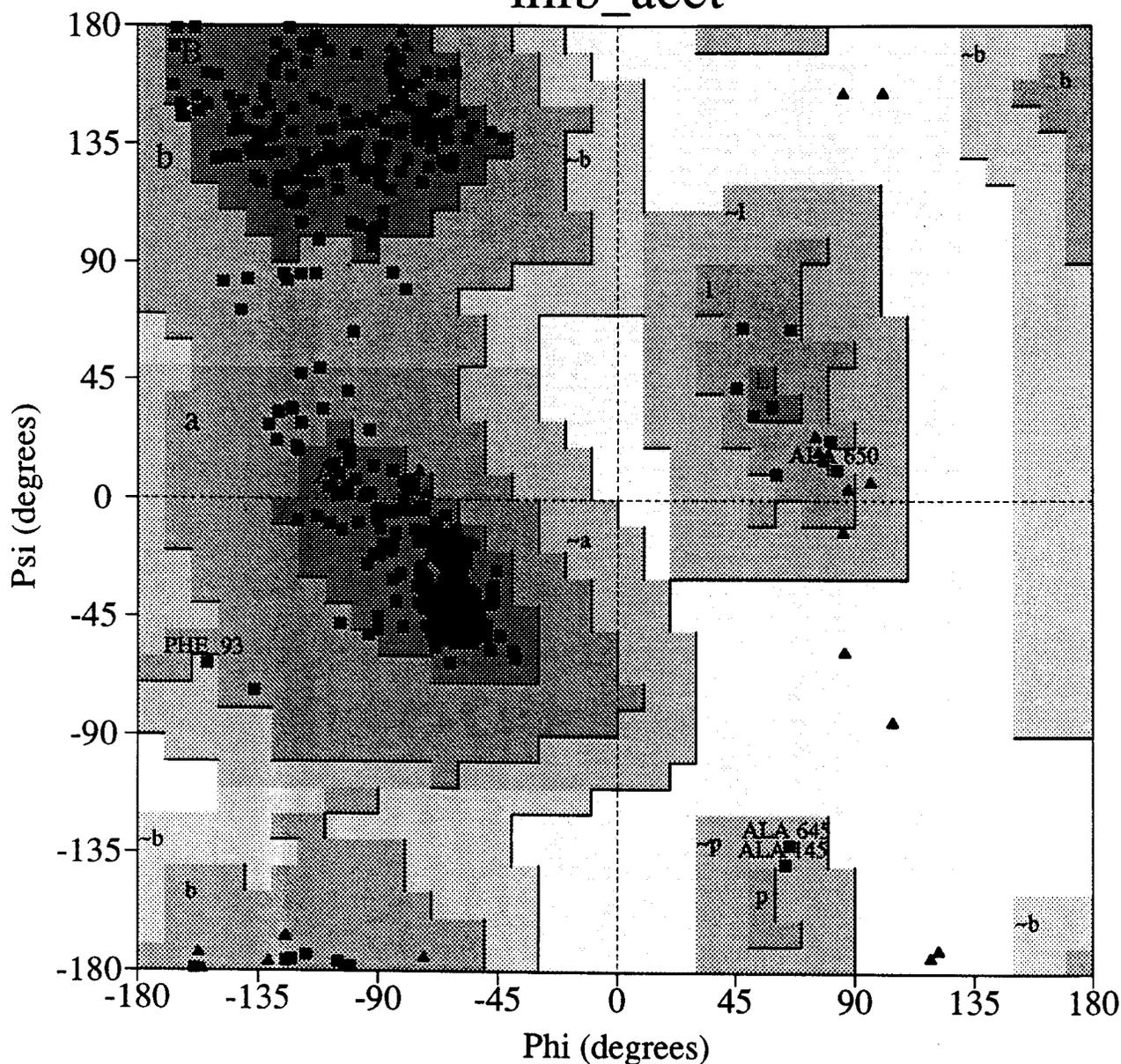


Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values		No. of band widths from mean	
			Typical value	Band width		
a. Chi-1 gauche minus st dev	55	8.4	19.0	6.5	-1.6	BETTER
b. Chi-1 trans st dev	135	16.7	19.7	5.3	-0.6	Better
c. Chi-1 gauche plus st dev	238	14.0	18.3	4.9	-0.9	Better
d. Chi-1 pooled st dev	428	14.3	18.9	4.8	-1.0	Better
e. Chi-2 trans st dev	133	14.8	20.9	5.0	-1.2	BETTER

# Ramachandran Plot

## inib\_acet

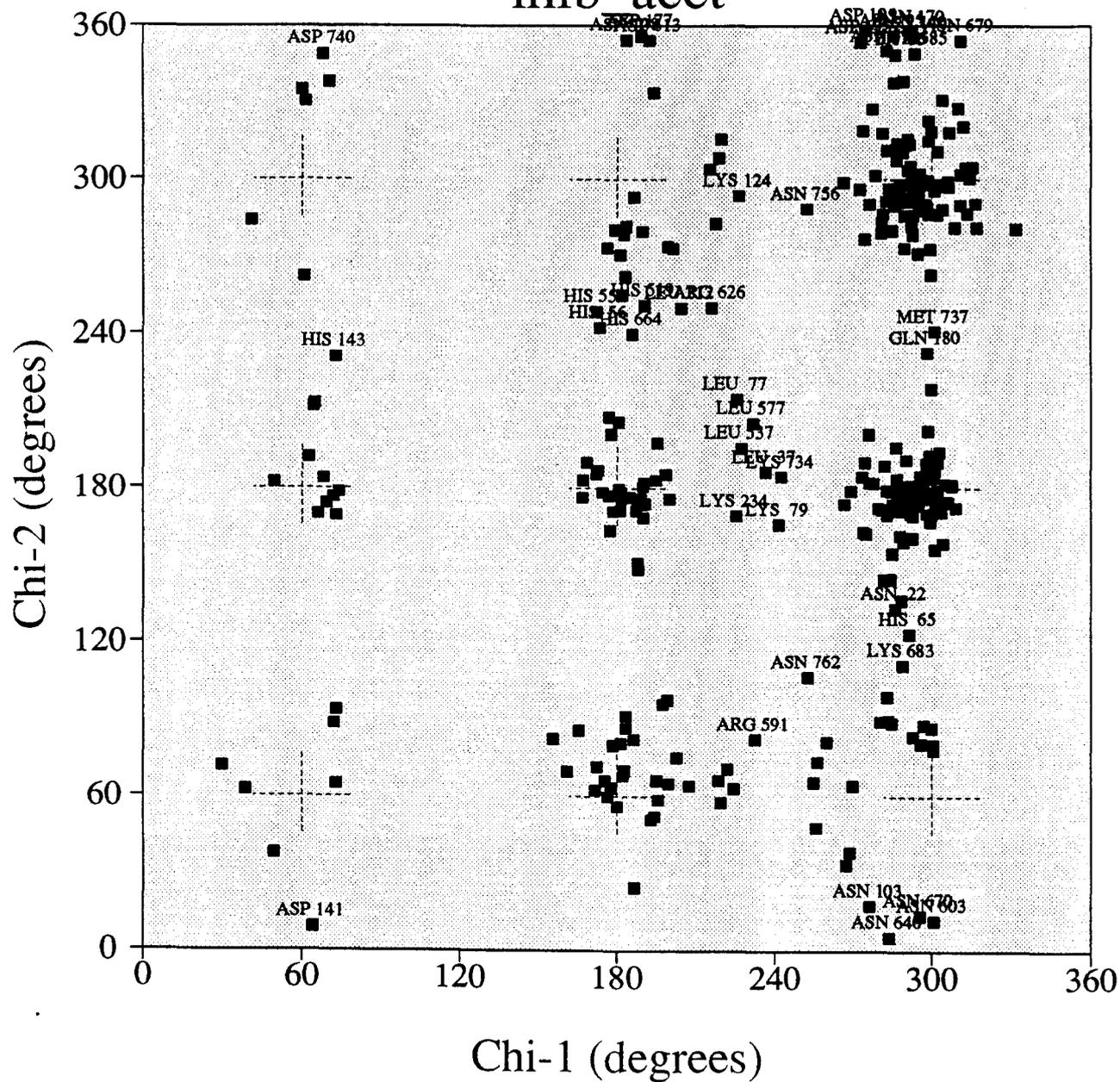


### Plot statistics

Residues in most favoured regions [A,B,L]	424	90.6%
Residues in additional allowed regions [a,b,l,p]	40	8.5%
Residues in generously allowed regions [~a,~b,~l,~p]	4	0.9%
Residues in disallowed regions	0	0.0%
-----		
Number of non-glycine and non-proline residues	468	100.0%
Number of end-residues	4	
Number of glycine residues (shown as triangles)	34	
Number of proline residues	26	
-----		
Total number of residues	532	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

# Plot of Chi-1 vs Chi-2 inib acet



Total number of residues = 532

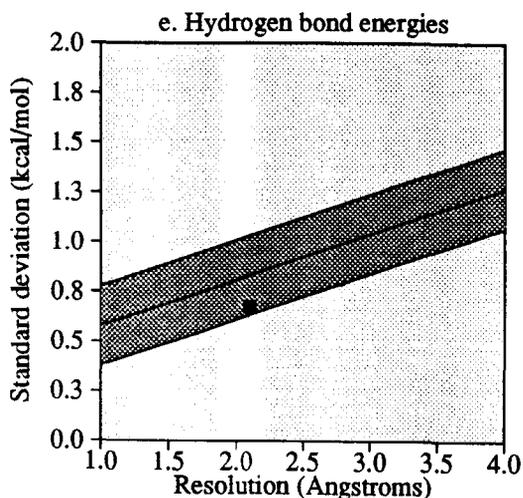
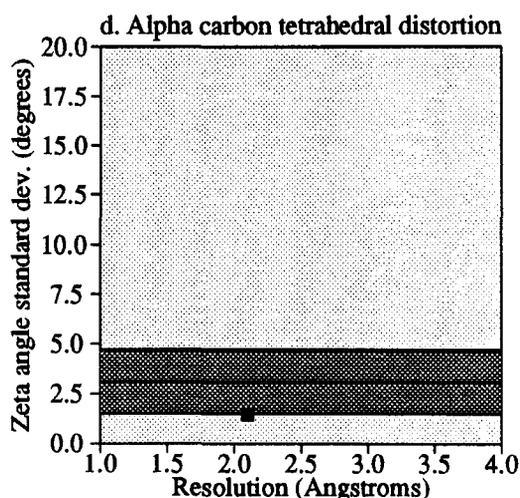
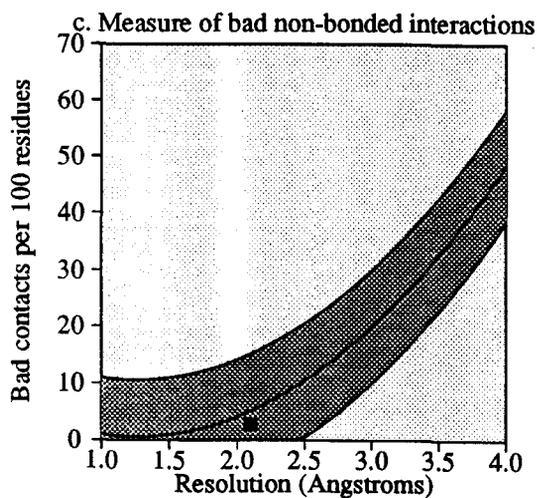
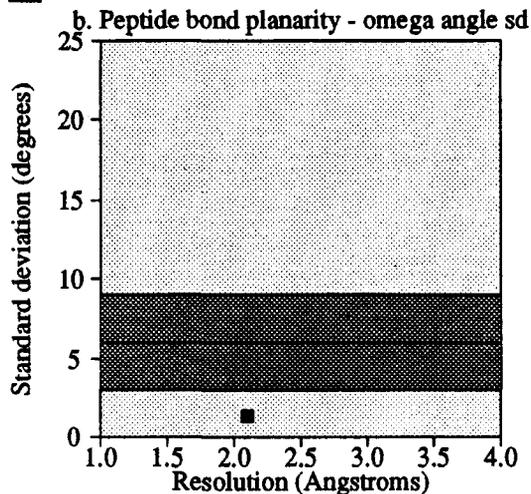
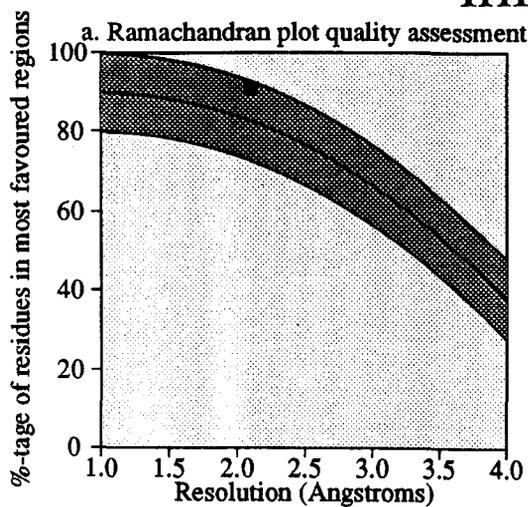
Number of residues plotted = 322

Number of labelled residues = 41

The dashed crosses represent the gauche minus, trans, and gauche plus regions for the Chi-1 and Chi-2 dihedral angles. The width of each cross is approximately one standard deviation. The points should therefore cluster around these crosses. Points more than 2.5 standard deviations from the ideal are labelled

# Main-chain parameters

## inib\_acet



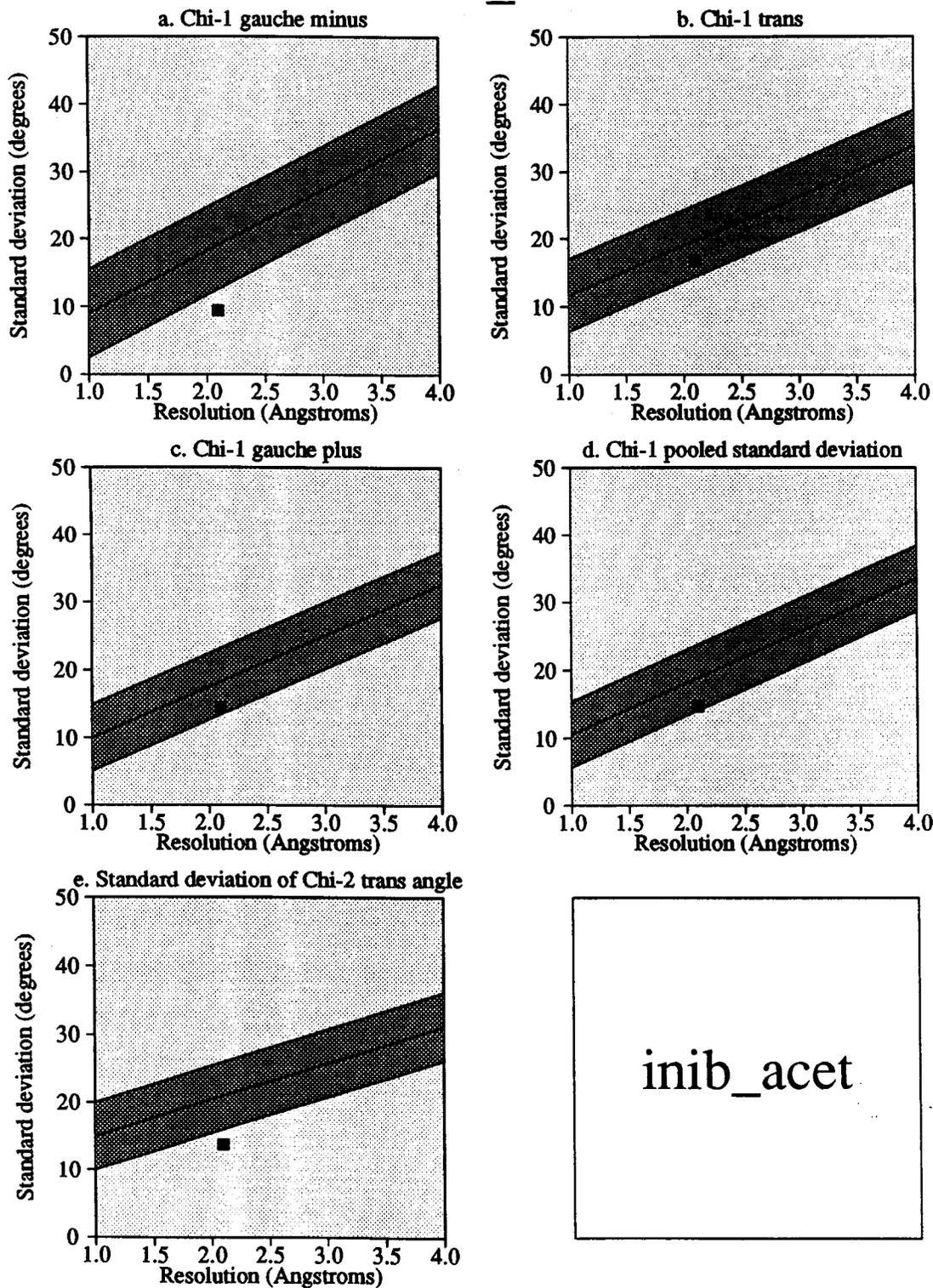
inib\_acet

### Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values		No. of band widths from mean	
			Typical value	Band width		
a. %-tage residues in A, B, L	468	90.6	82.6	10.0	0.8	Better
b. Omega angle st dev	530	1.3	6.0	3.0	-1.6	BETTER
c. Bad contacts / 100 residues	15	2.8	5.2	10.0	-0.2	Better
d. Zeta angle st dev	498	1.5	3.1	1.6	-1.0	BETTER
e. H-bond energy st dev	340	0.7	0.8	0.2	-0.8	Better

# Side-chain parameters

## inib\_acet



### Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values		No. of band widths from mean	
			Typical value	Band width		
a. Chi-1 gauche minus st dev	55	9.4	19.0	6.5	-1.5	BETTER
b. Chi-1 trans st dev	135	16.7	19.7	5.3	-0.6	Better
c. Chi-1 gauche plus st dev	238	14.4	18.3	4.9	-0.8	Better
d. Chi-1 pooled st dev	428	14.7	18.9	4.8	-0.9	Better
e. Chi-2 trans st dev	131	13.7	20.9	5.0	-1.4	BETTER