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

**“ESTUDOS ESTRUTURAIS DE DM43: UM INIBIDOR DE METALOPROTEASE
DE VENENO DE SERPENTE EXTRAÍDO DO SORO DO GAMBÁ *Didelphis
marsupialis*”**

DEBORA LIKA MAKINO

Dissertação apresentada ao Instituto de Física de São Carlos, Universidade de São Paulo, para
obtenção do título de Mestre em Ciências “Física Aplicada”

Orientador: Prof. Dr. Richard Charles Garratt

São Carlos – São Paulo
2000

95
USP/IFSC/SBI

8-2-001332

IFSC-USP SERVIÇO DE BIBLIOTECA
INFORMAÇÃO

Makino, Débora Lika

Estudos estruturais de DM43: um inibidor de metaloprotease de veneno de serpente extraído do soro do gambá *Didelphis marsupialis*. - São Carlos, 2000.

151p.

Dissertação (Mestrado) - Instituto de Física de São Carlos, 2000.

Orientador: Prof. Dr. Richard Charles Garratt

1. DM43. 2. Cristalização 3. Modelagem Molecular por Homologia.
I. Título.

UNIVERSIDADE
DE SÃO PAULO

Instituto de Física de São Carlos

Av. Dr. Carlos Botelho, 1465
CEP 13560-250 - São Carlos - SP
Brasil

Fone (016) 273-9333

Fax (016) 272-2218

**MEMBROS DA COMISSÃO JULGADORA DA DISSERTAÇÃO DE MESTRADO DE
DEBORA LIKA MAKINO APRESENTADA AO INSTITUTO DE FÍSICA DE SÃO CARLOS,
DA UNIVERSIDADE DE SÃO PAULO, EM 26 DE JUNHO DE 2000**

COMISSÃO JULGADORA:

Prof. Dr. Richard Charles Garratt/IFSC-FFI-USP

Prof. Dr. Gilberto Barbosa Dumont/UFRJ

Prof. Dr. Otávio Henrique Thiemann/IFSC-FFI-USP

E mail: wladerez@if.sc.usp.br

Aos meus pais e irmãos

Agradecimentos

Ao prof. Dr. Richard Charles Garratt

Ao prof. Dr. Glaucius Oliva

Ao prof. Dr. Gilberto Barbosa Domont do Departamento de Bioquímica da UFRJ-RJ

Ao prof. Dr. Jonas Perales do Departamento de Fisiologia e Farmacodinâmica da FIOCRUZ-RJ

À Dra. Ana Gisele da Costa Neves-Ferreira do Departamento de Fisiologia e Farmacodinâmica da FIOCRUZ-RJ

À Dra. Susan Jones da University of College London - UK.

À Regiane Burger

À Cláudia Munte

Ao Carlos Henrique Tomich de Paula da Silva

Aos funcionários do Grupo de Cristalografia e do Instituto de Física de São Carlos

Aos amigos do Grupo de Cristalografia

Aos amigos do Instituto de Física de São Carlos

Ao meu namorado Ivan

Ao Instituto de Física de São Carlos - USP

Ao CNPq pelo apoio financeiro deste trabalho

À todos que ajudaram a tornar possível este trabalho.

Índice

Lista de Figuras	i
Lista de Tabelas	v
Lista de Abreviaturas	vii
Resumo	ix
Abstract	x

I.DM43, A Imunidade Natural Contra o Veneno da Serpente <i>Bothrops jararaca</i>	1
I.1.A Soroterapia	2
I.2.Venenos de Serpentes	3
I.3.Os Inibidores de Metaloproteases de Veneno da serpente <i>Bothrops jararaca</i>	4
I.3.1.A Atividade Inibitória	6
I.3.2 Características Físico-Químicas	8
I.3.3.Estrutura Primária de DM43	8
II.Objetivos	9

PARTE A

III.Considerações Teóricas	11
III.1.Interações Moleculares a Nível Macroscópico	11
III.1.1.Interações Hidrofóbicas	12
III.1.2.Força Iônica	14
III.1.3.Outros Fatores que Influenciam as Interações Moleculares	16
III.2.Cromatografia	17
III.2.1.Cromatografia por Interação Hidrofóbica	18
III.2.2.Cromatografia por Troca Iônica	20
III.3.Cristalografia	23
III.3.1.Cristalogênese de Proteínas	23
III.3.2.Difração de Raios-X	27
III.4.Desglicosilação Enzimática	30
IV.Materiais e Métodos	33
IV.1.DM43	33
IV.2.Coefficiente de Extinção Molar de DM43	33
IV.3.Cristalização	34
IV.4.Desglicosilação	37
IV.5.Cristalização da DM43 Desglicosilada	38
IV.6.Cromatografia	38
IV.7.Criocristalografia	40

IV.8.Coleta de Dados	41
IV.9.Análise dos Padrões de Difração	41
V.Resultados e Discussões	43
V.1.Cristalização de DM43 e Coletas de Dados	44
V.2.Desglicosilação	51
V.3.Purificação	55
V.4.Cristalização de DM43 Após Processo de Desglicosilação	57
V.5.Perspectivas	58
 PARTE B	
VI.Modelagem Molecular Estrutural	60
VI.1. MODELLER	65
VI.1.1.Derivação das Restrições Espaciais	66
VI.1.1.1.Cálculo da Distribuição de Probabilidade de um Esperso Conjunto de Dados	68
VI.1.1.2.Derivação das Restrições Estereoquímicas	70
VI.1.1.3.Restrições de Distâncias entre C α - C α e N - O	71
VI.1.1.4.Restrição da Cadeia Principal de um Resíduo	73
VI.1.1.5.Restrição da Cadeia Lateral de um Resíduo	74
VI.1.2.Satisfação das Restrições Espaciais	75
VI.1.2.1 Derivação da pdf Característica	75
VI.1.2.2.Derivação da pdf Molecular a partir da pdf Característica	77
VI.1.2.3.Otimização da pdf Molecular	78
VI.2. Avaliação dos Modelos	79
VI.3. PATCHES	81
VII. Materiais e Métodos	84
VII.1.BLAST	84
VII.2.THREADER	85
VII.3.MULTALIGN	86
VII.4.PROMOTIF	87
VII.5.MODELLER	87
VII.6.Visualização	92
VII.7.Avaliação dos Modelos	92
VII.8.GRASP	93
VII.9.PATCHES	93
VIII.Resultados e Discussões	95
VIII.1.Busca e Análise de Sequências Homólogas a DM43	95
VIII.2.Modelagem	104
VIII.3.Avaliação dos Modelos	109
VIII.4.Interpretação dos Modelos	118
VIII.5.As Prováveis Regiões de Interação de DM43	122
IX. Conclusões e Perspectivas	131
 Anexo-I	
Soluções do Fatorial I	134

Anexo-II

Protocolo de Detecção de Glicoproteínas em Membrana de Nitrocelulose_ 135

- B.1 Transferência_____ 135
 - B.1.1 Equipamento _____ 135
 - B.1.2 Tampão de Transferência _____ 136
 - B.1.3 Preparação da Cuba _____ 136
 - B.1.4 Condições para Transferência _____ 139
- B.2 Revelação da Membrana _____ 139
 - B.2.1 Reagentes _____ 140
 - B.2.2 Procedimentos _____ 142

Referências Bibliográficas _____ 144

Lista de Figuras

Figura	I- 1: Foto da serpente <i>Bothrops jararaca</i> _____	1
Figura	I- 2: Alguns exemplos dos efeitos da ação do veneno de serpentes do gênero <i>Bothrops</i> _____	2
Figura	I- 3: Gambá _____	5
Figura	III- 1: Camadas organizadas de moléculas de água em torno da região hidrofóbica da proteína. _____	13
Figura	III- 2: Curva do logaritmo da solubilidade da carbonomonoxihemoglobina com relação a força iônica de vários sais na região de salting-out _____	15
Figura	III- 3: Curva do logaritmo da solubilidade da carbonomonoxihemoglobina com relação a raiz quadrada da força iônica de vários sais na região de <i>salting-in</i> e <i>salting-out</i> . _____	15
Figura	III- 5: Grupos apolares utilizados na cromatografia por interação hidrofóbica. _____	18
Figura	III- 6: Etapas de uma interação entre regiões hidrofóbicas de proteína-resina (cromatografia hidrofóbica) ou eventualmente de proteína-proteína durante a cristalização. _____	19
Figura	III- 7: Variação na carga líquida da proteína com o pH. _____	20
Figura	III- 8: Representação dos dois tipos de interação iônica. _____	21
Figura	III- 9: Etapas da eluição de uma proteína de carga líquida negativa ligada em um trocador aniônico. _____	21
Figura	III-10: Incorporação de uma impureza no cristal alterando a disposição ordenada das proteínas ao seu redor. _____	24
Figura	III-11: Zona de depleção, ou gradiente de concentração, revelado pela área escura ao redor do cristal do vírus STMV (<i>satellite tobacco mosaic virus</i>) em crescimento. _____	25
Figura	III-12: Representação exagerada de um cristal real, mostrando detalhes da mosaicidade composta por blocos ou microcristais ideais. _____	29
Figura	III-13: Etapas do processo de detecção de carboidratos ligados a proteínas) _	31
Figura	V- 1: Cristais de DM43 crescidos à 18°C em 48% S.A. – pH5.0. _____	44
Figura	V- 2: Cristais antigos de DM43 em 54% S.A. pH5.0 e 54% S.A. pH7.0 _____	45
Figura	V- 3: Cristais de DM43 crescidos à 4°C em gota de 4µl contendo 56% S.A. - pH5.0 _____	45
Figura	V- 4: Diferentes hábitos cristalinos apresentados por DM43. _____	46
Figura	V- 5: Cristais crescidos após 8 meses em gota de 6µl - 50% S.A. - pH5.0, 18°C. Detalhe de três reflexões de uma imagem de difração por radiação síncrotron (LNLS) _____	47

Figura V- 6:	Gel de poliacrilamida e membrana de nitrocelulose indicando a desglicosilação completa de DM43 por processos enzimáticos	52
Figura V- 7:	Géis de poliacrilamida indicando a total desglicosilação de DM43 após tratamento químico com TFMS e ácido periódico Schiff para coloração de glicoproteínas.	53
Figura V- 8:	Gel de SDS-PAGE 15% revelando a desglicosilação parcial de DM43.	54
Figura V- 9:	Géis de SDS-PAGE 15% corado com prata, apresentando um padrão de eluição das DM43 nativa e parcialmente desglicosilada.	56
Figura V-10:	Detalhe do perfil cromatográfico de DM43 em coluna de interação aniônica fraca POROS 20HQ.	56
Figura VI- 1:	Número de sequências e estruturas de proteínas obtidas entre os anos 1986 e 1994.	60
Figura VI- 2:	Estatística anual do número total de sequências registrado na base de dados de SWISS-PROT	61
Figura VI- 3:	Estatística recente do número de estruturas depositadas no <i>Brookhaven Protein Data Bank</i> (PDB)	61
Figura VI- 4:	Modelagem molecular por satisfação das restrições espaciais	65
Figura VI- 5:	Diferentes distribuições W' de pares equivalentes de distâncias entre a proteína alvo e estruturas do banco de dados local, fitados pelos modelos Gaussianos .	73
Figura VI- 6:	Derivação da pdf característica a partir das pdfs base	76
Figura VI- 7:	Otimização da função objetiva	79
Figura VI- 8:	Fluxograma do algoritmo PATCHES, utliizado para predizer a região mais provável para uma interação proteína-proteína.	83
Figura VII- 1:	Alinhamentos iniciais de 1nkr contra D_0 e D_1 - D_2 de DM43, utilizados em MODELLER, na produção dos primeiros modelos das estruturas dos domínios D_0 e D_1 - D_2 .	88
Figura VII- 2:	Alinhamentos de 1nkr contra D_0 e D_1 - D_2 de DM43, utilizados em MODELLER, para produção dos modelos das estruturas dos domínios D_0 e D_1 - D_2 .	89
Figura VIII- 1:	Representação esquemática do processo de interação dos KIRs com HLAs.	97
Figura VIII- 2:	Distribuição dos domínios KIRs sugerindo a dos três domínios de DM43 e Oprina.	98
Figura VIII- 3:	Alinhamento para auxiliar na derivação do melhor alinhamento de DM43 contra 1nkr (KIR2DL1).	102
Figura VIII- 4:	Topologia bidimensional de dois enovelamentos tipo imunoglobulina.	105
Figura VIII- 5:	Diagramas de Ramachandran específicas para prolinas, obtidas através do programa PROCHECK..	105

Figura VIII- 6: Orientação intradomínios de imunoglobulinas, receptor do hormônio de crescimento prolactina, os KIRs 1nkr, 2dl2 e 1b6u. ____	106
Figura VIII- 7: Sobreposição dos dez modelos de DM43_D ₀ no traçado da cadeia principal e a representação estereográfica em sausicha. _____	107
Figura VIII- 8: Sobreposição dos dez modelos de DM43_D ₁ -D ₂ representados pelo traçado da cadeia principal e, na mesma orientação, os mesmos dez modelos no estilo sausicha. _____	108
Figura VIII- 9: Diagrama de Ramachandran para o modelo final de DM43_D ₀ . ____	110
Figura VIII-10: Diagrama de Ramachandran para o modelo final de DM43_D ₁ -D ₂ . ____	112
Figura VIII-11: Fator-G por resíduo segundo ângulos e distâncias do domínio DM43_D ₀ . _____	113
Figura VIII-12: Fator-G por resíduo segundo ângulos e distâncias dos dois domínios de DM43_D ₁ -D ₂ . _____	114
Figura VIII-13: Gráfico apresentando a variação da qualidade do contato atômico de cada resíduo segundo OLD-QUALITY de DM43_D ₀ _____	115
Figura VIII-14: Gráfico apresentando a variação da qualidade do contato atômico de cada resíduo segundo OLD-QUALITY de DM43_D ₁ -D ₂ _____	115
Figura VIII-15: Distribuição do <i>score</i> de qualidade de ambiente químico por resíduo segundo VERIFY_3D. _____	116
Figura VIII-16: Distribuição do <i>score</i> de qualidade de ambiente químico por resíduo segundo VERIFY_3D. _____	116
Figura VIII-17: Apresentação dos modelos finais de DM43. _____	117
Figura VIII-18: Visão estereográfica da sobreposição do modelo final de DM43_D ₀ e da estrutura 1nkr_D ₂ _____	118
Figura VIII-19: Visão estereográfica da sobreposição do modelo final de DM43_D ₁ -D ₂ e da estrutura 1nkr. _____	118
Figura VIII-20: Detalhe da região WSXWS <i>box</i> de DM43_D ₁ -D ₂ e sua localização relativa entre os domínios. _____	119
Figura VIII-21: Representação da junção entre os domínios DM43_D ₁ e DM43_D ₂ explicitando a formação de três ligações de hidrogênio com a fita G do primeiro domínio favorecido pela presença da prolina 273. ____	119
Figura VIII-22: Orientação esquemática dos três domínios de DM43, dada a presença do triptofano do motivo WSXWS <i>box</i> e os quatro sítios e glicosilação evitando a criação de impedimentos estéricos entre os domínios D ₀ e D ₂ de DM43. _____	120
Figura VIII-23: Localização de três sítios de glicosilação em D ₁ , com relação à região do motivo WSXWS, e um quarto sítio em D ₀ . _____	121
Figura VIII-24: Detalhe da região WSXWS <i>box</i> de DM43_D ₀ e sua localização relativa no domínio _____	121
Figura VIII-25: Representação de um complexo hormônio de crescimento e receptor de prolactina destacando a região do motivo WSXWS. ____	122

Figura VIII-26: Representação no modelo DM43_D ₁ -D ₂ das prováveis regiões de interação com a metaloprotease. _____	123
Figura VIII-27: Representação esquemática de dimerização de KIR2DL na presença do ligante. _____	124
Figura VIII-28: Representação em <i>ribbons</i> da provável superfície de dimerização. ____	124
Figura VIII-29: Superfície de potencial de DM43_D ₁ -D ₂ _____	125
Figura VIII-30: Superfície de potencial de DM43_D ₀ . _____	125
Figura VIII-31: Vista frontal de DM43_D ₁ -D ₂ com os <i>patches</i> destacados em verde, para o centrado em Ala 169 e em vermelho para o centrado em Leu 252. _____	127
Figura VIII-32: Representação das prováveis regiões de interação encontradas em DM43_D ₁ -D ₂ , pelo programa PATCHES. _____	127
Figura VIII-33: Visão completa de DM43 com o domínio D ₀ em posição de encaixe geométrico pelas regiões encontradas por PATCHES e a junção destes domínios pelo C-terminal de D ₀ e N-terminal de D ₁ . _____	128
Figura VIII-34: Hipótese sobre o processo de complexação de DM43 com a metaloprotease. _____	129
Figura VIII-35: Resumo das principais etapas de modelagem de DM43. _____	130

Lista de Tabelas

Tabela III- 1: Lista das constantes dielétricas de alguns solventes orgânicos utilizados em cristalização de proteínas _____	16
Tabela III- 2: Curvas de titulação de trocadores aniônicos e catiônicos fortes e fracos. _____	22
Tabela III- 3: Indicação de cada um dos fatores que influenciam na atenuação da intensidade da radiação difratada. _____	28
Tabela IV- 1: Experimentos de Cristalização da DM43 _____	35
Tabela IV- 2: Experimentos de Cristalização da DM43 Nativa e Desglicosilada. _____	36
Tabela IV- 3: Reagentes e proteínas utilizadas no experimento de desglicosilação enzimática, segundo manual da Bio-Rad Laboratories _____	37
Tabela IV- 4: Parâmetros utilizados na purificação de DM43 parcialmente desglicosilada com a coluna aniônica forte da PerSeptive BioSystems. _____	39
Tabela IV- 5: Parâmetros utilizados na purificação de DM43 parcialmente desglicosilada com a coluna aniônica fraca da PerSeptive BioSystems. _____	39
Tabela IV- 6: Parâmetros utilizados na purificação de DM43 parcialmente desglicosilada com a coluna aniônica fraca da Pharmacia LKB Biotechnology. _____	39
Tabela IV- 7: Parâmetros utilizados na purificação de DM43 parcialmente desglicosilada com a resina de interação hidrofóbica da Pharmacia LKB Biotechnology. _____	39
Tabela V- 1: Descrição das coletas de difração na estação PCr (LNLS), de alguns cristais de DM43 crescidos em Sulfato de Amônio. _____	48
Tabela VI- 1: Descrição de métodos de modelagem por homologia existentes para a obtenção teórica de estruturas de proteína. _____	64
Tabela VI- 2: Características espaciais para calcular os elementos da matriz de frequências relativas, W' . _____	67
Tabela VII-1: Tabela indicando as restrições especiais acrescentadas no MODELLER.. _____	91
Tabela VIII-1: Tabela de identidade interdomínios de DM43 contra α_1 B-glicoproteína humana. _____	96
Tabela VIII-2: Porcentagem de identidade entre DM43 e os membros da família KIR2DL e KIR3DL. _____	98
Tabela VIII-3: Porcentagem de identidade média interdomínios entre os DM43 e os dos membros da família KIR2 e KIR3. _____	98
Tabela VIII-4: Avaliação das três estruturas homólogas a DM43. _____	100

Tabela VIII-5: Valores da função objetiva dos dez modelos de saída do MODELLER escolhidos para modelar manualmente para se chegar na estrutura final. _____	109
Tabela VIII-6: Comparação das funções objetivas da estrutura de referência 1nkr e do modelo final DM43.. _____	109
Tabela VIII-7: Especificação dos prováveis resíduos envolvidos na interação com a metaloprotease. _____	123
Tabela VIII-8: Resíduos pertencentes às prováveis regiões de interação identificados pelo programa PATCHES. _____	126

Lista de Abreviaturas

3D	Tridimensional
Bis-Tris	Bis (2-hidroxietil) amino-tris (hidroximetil) metano
DM43	Inibidor de metaloprotease do veneno da serpente <i>Bothrops jararaca</i> , isolado do soro do gambá sul-americano <i>Didelphis marsupialis</i> . Peso molecular de 43KDa.
DMF	Dimetilformamida
DO ₂₈₀	Densidade Óptica a 280nm
EDTA	Ácido Etileno Diamino Tetra Acético
HEPES	Ácido N-(2-hidroxietil) piperazina-N'-2-etano-sulfônico
KIRs	Receptores Inibitórios de Células Assassinas
LNLS	Laboratório Nacional de Luz Síncrotron – Campinas/SP-Brasil
MES	Ácido 2-Morfolinoetanosulfônico Monohidrato
MPD	2-Metil-2,4-Pentadienol
NMR	Ressonância Magnética Nuclear
PBS	<i>Phosphate buffered saline</i>
PEGx	Polietileno glicol de massa molecular x
pdf	Função Densidade de Probabilidade
pl	Ponto Isoelétrico
PNGase F	Peptídeo-N4-(N-Acetil-β -D-Glucosaminil) Asparagina Amidase F

R.M.S.D	Desvio Médio Quadrático
S.A.	Sulfato de Amônio
SDS-PAGE	Eletroforese em gel de poliacrilamida contendo dodecil sulfato de sódio
TBS	<i>Tris Buffered Saline</i>
TFMS	Ácido Anihidro trifluorometanosulfonico
TRIS	Tris (hidroximetil) aminometano
VTFM	Método de Função Alvo Variável

Resumo

A resistência natural do gambá *Didelphis marsupialis* ao veneno de serpentes se deve a fatores antibotrópicos presentes em seu soro, do qual DM43 é um dos responsáveis pela inibição da atividade hemorrágica. Com o objetivo de entender melhor o mecanismo de ação desta proteína contra a metaloprotease contida no veneno pretendeu-se, neste trabalho, otimizar as condições de cristalização de DM43 na tentativa de determinar sua estrutura por difração de raios-X. Ensaio de cristalização revelaram que o sulfato de amônio é o precipitante mais eficiente na produção de cristais de DM43. A visualização e indexação dos padrões de difração destes cristais permitiram apenas a determinação dos parâmetros de sua cela unitária devido a baixa qualidade dos mesmos. Deste modo, os seguintes parâmetros de cela unitária foram encontrados: $a = b = 117,34 (0,03) \text{ \AA}$, $c = 193,39 (0,02) \text{ \AA}$, $\alpha = \beta = 90^\circ$ e $\gamma = 120^\circ$, correspondentes ao sistema cristalino trigonal ou hexagonal. A recente determinação da seqüência de aminoácidos de DM43, possibilitou modelar sua estrutura tridimensional por homologia utilizando o método de satisfação das restrições espaciais. Os três domínios tipo imunoglobulina de DM43 foram modelados em duas partes a partir da estrutura de referência KIR2DL1(1nkr) homóloga a dois domínios C-terminais de DM43, com apenas 27,72% de identidade. O domínio N-terminal denominado D₀, foi modelado separadamente contra o domínio C-terminal de KIR2DL1 com um grau de identidade igual a 28,89%. A partir do programa GRASP e PATCHES, foram detectadas as prováveis regiões de ligação entre as duas partes da molécula e uma possível região responsável pela dimerização no domínio D₂ de DM43. Interessantemente, resíduos polares em KIR2DLs foram substituídos por hidrofóbicos em DM43 concentrando-se em uma face do domínio D₂ ausente de sítios de glicosilação, o que coincide com as observações de dimerização em receptores do hormônio de crescimento. Um motivo estrutural característico de receptores hematopoéticos identificados como *WSXWS box*, foi encontrado em todos os domínios de DM43, especialmente no domínio D₂. Supõe-se que a sua existência esteja relacionada com a orientação relativa dos domínios por manter o primeiro triptofano do motivo posicionado exatamente na interface entre os domínios. Surge ainda uma fita extra (F'), no domínio D₂, não pertencente ao enovelamento imunoglobulina devido a presença da Pro273, muito bem conservada a três resíduos do motivo *WSXWS*. Esta fita parece desempenhar um papel importante na orientação do motivo pois além de formar ligações de hidrogênio com a fita G do domínio anterior, posiciona corretamente o primeiro triptofano do motivo na interface. Somando-se a isto, a presença de resíduos hidrofóbicos na interface dos domínios D₁ e D₂ pode estar contribuindo para a orientação angular relativa menor que 90° entre os dois domínios. Uma interpretação análoga a de hormônios de crescimento sugere que, os *loops* entre as fitas AB, CC' e EF do domínio 1, BC e FG do domínio 2 e o *linker* entre estes dois domínios, estejam envolvidos na interação da metaloprotease com DM43.

Abstract

The natural resistance of the opossum, *Didelphis marsupialis*, towards snake venom is due to antithrombotic factors present in the blood serum, of which DM43 is one. It is responsible for the inhibition of the haemorrhagic activity of the venom. With a view to better understanding the mechanism of action of this protein against venom metalloproteases, one of the aims of the present work was to optimise the crystallisation conditions of DM43 in order to determine its three-dimensional structure by X-ray diffraction. Crystallisation trials revealed that ammonium sulphate was the best precipitant. Visualisation and indexing of the diffraction patterns from such crystals only allowed for the determination of the unit cell parameters due to their inherently low quality. The values obtained were $a = b = 117.34 (0.03) \text{ \AA}$, $c = 193.39 (0.02) \text{ \AA}$, $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$, corresponding to the trigonal or hexagonal crystal systems. The recent determination of the amino acid sequence of DM43 permitted the homology modelling of its three-dimensional structure by satisfaction of spatial restraints. The three immunoglobulin-like domains of DM43 were modelled in two separate parts from the reference structure KIR2DL1(1nkr), homologous to the two C-terminal domains of DM43, with which it shares 27.72% sequence identity. The N-terminal domain, denominated D_0 , was separately modelled based on the C-terminal domain of KIR2DL1, with which it shares 28.89% identity. The programs PATCHES and GRASP were used to detect the probable interface region between the two separate parts of the molecule as well as a region probably responsible for dimerisation. Polar residues in KIR2DL1 which had been substituted by hydrophobic residues in DM43 are observed concentrated on one face of the molecule devoid of glycosylation sites (D_2) coinciding with the dimerisation interface observed in the growth hormone receptor. A structural motif characteristic of the haematopoietic receptor family, identified as the WSXWS box, was observed to some extent in all three domains of DM43 and most evident in domain D_2 . It is speculated that the existence of this motif is related to the relative orientation of the domains, by maintaining the first tryptophan of the motif positioned at the domain interface. An additional β -strand (F') in D_2 , which is not characteristic of the immunoglobulin fold, is observed due to the presence of the conserved Pro273, three-residues prior to the WSXWS box. This strand appears to play an important role in correctly positioning the motif as it forms hydrogen bonds with strand G of the previous domain thus orientating the first tryptophan of the motif at the interface. The presence of hydrophobic residues at the interdomain interface, may also contribute to stabilising the acute angular relationship between D_1 and D_2 . An analogous interpretation to that given for the growth hormone receptor, suggests that the loops between β -strands AB, CC' and EF of domain D_1 and between strands BC and FG of domain D_2 plus the linker region, are involved in the binding of metalloproteinase by DM43.

I.

**DM43,
A IMUNIDADE NATURAL CONTRA
O VENENO DA SERPENTE
*Bothrops jararaca*¹**

Segundo estimativas a nível mundial, existem por volta de cinco milhões de casos de acidentes ofídicos humanos por ano. No Brasil, são vinte mil casos de envenenamento anual e grande parte deles devido ao ataque de serpentes do gênero *Bothrops*.

Figura I-1: Foto da serpente *Bothrops jararaca* (CEVAP).

¹ Neves-Ferreira, A. G. C. , 1999

Muito comum na América do Sul, a *Bothrops jararaca* (família Viperidae, subfamília Crotalinae) (figura I-1) é vista em campos e em lavouras, sendo muito facilmente encontrada desde o estado da Bahia até o estado de Santa Catarina (Leitão, 1980). O veneno deste gênero de serpente provoca efeitos locais bastante graves como hemorragia, mionecrose, hiperalgesia, edema, necrose, equimose e formação de bolhas (figura I-2). Podem ocorrer também sangramentos espontâneos nas gengivas, nariz, pele, cérebro, trato gastrointestinal e tecidos retro-peritoneais. O quadro hemorrágico pode resultar em hipovolemia, da qual pode ser conferido o efeito de choque persistente, hipotensão e insuficiência renal aguda. Estes dois últimos e a hemorragia intracraniana são as causas mais frequentes de morte (CEVAP).

Figura I-2: Alguns exemplos dos efeitos da ação do veneno de serpentes do gênero *Bothrops* (a) Ação proteolítica. (b) Necrose. (c) Ação vasculotóxica. (d) Bolhas. (CEVAP)

I.1. A Soroterapia

A soroterapia contra venenos de serpentes foi introduzida no Brasil por Vital Brazil Mineiro da Campanha (Instituto Butantã - São Paulo), em 1897. É até hoje, o único tratamento aceito pela comunidade médica apesar das reações adversas após a administração do soro, como os efeitos pirogênicos, anafiláticos e anafilactóides (CEVAP).

A obtenção do soro é feita segundo a técnica desenvolvida por Calmette, Phisalix e Bertrand há mais de um século. Uma dose de veneno não tóxica é injetada de dez a cinquenta vezes no cavalo, em períodos regulares, durante três a quinze meses; o sangue do cavalo é retirado e tratado de forma a eliminar as proteínas que não estão relacionadas com a imunidade do animal. Devido a característica heteróloga deste soro, que também pode conter nele algumas endotoxinas bacterianas, podem provocar as já citadas reações adversas e, em raras ocasiões, a morte do paciente. Existem testes alérgicos à sensibilidade do

soro, mas desaconselhado pelo Manual do Ministério da Saúde para o tratamento de acidentes por animais peçonhentos. É também aconselhada uma prévia preparação das drogas dextroclorfeniramina, hidrocortisona e cimetidine, para um eventual tratamento das reações adversas imediatas após a administração do soro (CEVAP).

I.2. Venenos de Serpentes

Venenos de serpentes são como uma "sopa" de proteínas contendo sais orgânicos e inorgânicos, aminas biogênicas, aminoácidos livres e peptídeos¹. As hemorraginas (responsáveis pela indução de hemorragia tanto local quanto sistêmico, e indiretamente de mionecrose pela interrupção do irrigamento sangüíneo no músculo) e miotoxinas (peptídeos ou enzimas que atuam sobre fibras musculares esqueléticas) são componentes protéicos típicos de venenos de serpentes da família Viperidae.

As hemorraginas são metaloproteases que induzem hemorragia a partir da degradação das proteínas da membrana basal da parede dos vasos sanguíneos. Estudos mostraram que existe uma correlação direta entre a intensidade do grau de hemorragia com a atividade enzimática das metaloproteases sobre as proteínas da membrana basal². Algumas delas possuem um domínio tipo desintegrina que direciona a toxina para membranas de plaquetas e células endoteliais onde o domínio metaloprotease manifesta sua ação proteolítica sobre integrinas, fibrinogênios e proteínas de matriz. Artigo recente sobre o domínio tipo desintegrina da jararagina, mostrou que são inibidores eficientes da agregação plaquetária agravando, conseqüentemente, o quadro hemorrágico do paciente³.

¹ BIEBER, A. L. Metal and nonprotein constituents in snake venoms. *Handbook of Experimental Pharmacology-Snake Venoms*, Lee, C. Y., Ed. Berlin, Springer-Verlag, v. 52, p. 295-306, 1979.
MARKLAND, F. S. Snake venoms and the hemostatic system. *Toxicon*, v. 36, p. 1749-1800, 1998.

² BJARNASON, J. B.; HAMILTON, D.; FOX, J. W. Studies on the mechanism of hemorrhage production by five proteolytic hemorrhagic toxins from *Crotalus atrox* venom. *Biol. Chem. Hoppe-Seyler*, v. 369, p. 121-129, 1988.

³ MOURA-DA-SILVA, A. M. *et al.* Jararhagin ECD-containing disintegrin domain: expression in *Escherichia coli* and inhibition of the platelet-collagen interaction. *Arch. Biochem. Biophys.*, v. 369, p. 295-301, 1999.

O primeiro estudo sobre a resistência natural a venenos (Domont, Perales, Moussatché, 1991) data de 1781, quando verificaram que a víbora era resistente a seu próprio veneno. Desde então, vários estudos têm sido realizados com o interesse de isolar e caracterizar proteínas com propriedades antiofídicas¹. As proteínas antiofídicas são divididas em dois grupos: fatores antineurotóxicos e antimiotóxicos que são os inibidores de Fosfolipase A₂ (PLA₂) envolvidas nos efeitos de neurotoxicidade, miotoxidade, citotoxicidade e distúrbios hemostáticos; e os fatores anti-hemorragicos que correspondem aos inibidores de metaloprotease.

Foi na década de setenta que se iniciaram os primeiros estudos sobre a imunidade natural do gambá sul-americano *Didelphis marsupialis* sobre venenos de serpente, em especial de *Bothrops jararaca*². A quantidade de veneno por ele tolerado equivale à administração de 24g de veneno bruto em um homem de 60kg (Perales, *et al*, 1994).

I.3. Os Inibidores de Metaloproteases de Veneno da serpente *Bothrops jararaca*

O Laboratório do Departamento de Fisiologia e Farmacodinâmica do Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, isolou do soro do gambá Sul Americano *Didelphis marsupialis* (figura I-3) uma fração composta por duas glicoproteínas³ de baixa massa molecular com atividade antibotrópica, especialmente contra o veneno da serpente *Bothrops jararaca*. Por possuírem baixos valores de massa molecular, elas se enquadram na classe de inibidores de metaloprotease ou fatores anti-

¹ THWIN, M. M.; GOPALAKRISHNAKONE, P. Snake envenomation and protective natural endogenous proteins: a mini review on the recent developments (1991-1997). *Toxicon*, v. 36, p. 1471-1482, 1998.

FOX, J. W.; BJARNASON, J. B. Metalloproteinase inhibitors. *Enzymes from snake venoms*. Bailey, G. S., Ed. Colorado, Alaken, Inc. p. 599-632, 1998.

PÉREZ, J. C.; SÁNCHEZ, E. E. Natural protease inhibitors to hemorrhagins in snake venoms and their protential use in medicine. *Toxicon*, v. 37, p. 703 -728, 1999.

² MOUSSATCHÉ, H. *et al*. Mechanisms of resistance of the opossum to some snake venoms. *Toxicon*, v.17, p.130,1979.

³ Uma terceira glicoproteína com atividade antimiotóxica do veneno da serpente *Bothrops asper* foi recentemente detectada no soro de *D.marsupialis*. É uma proteína ácida de 15% de glicosilação e massa molecular de 63659 u.m.a. , denominada como DM64 (neves-Ferreira, 1999).

hemorrágicos mais abundantes, de 52 e 90 KDa (Domont, Perales, Moussatché, 1991).

Figura I-3: Gambá¹.

Segundo experimentos de espectrometria de massa por MALDI-TOF, suas massas moleculares correspondem a 40.318 e 43.010 u.m.a., sendo que isoformas foram observadas nesta última, apresentando também uma massa de 42.373 u.m.a.. Embora outras técnicas revelassem massas moleculares diferentes (44,8 e 47,3 KDa por SDS-PAGE² e 97,0 e 72,5 KDa por filtração em gel³, respectivamente), foram dados a elas os nomes de DM40 e DM43 como referência ao nome científico do gambá *Didelphis marsupialis* e suas respectivas massas moleculares.

A obtenção do soro do gambá *Didelphis marsupialis* foi feita segundo uma adaptação da metodologia introduzida em 1977 por HILLAM *et al.* (1974)⁴, que substitui o método tradicional de punção cardíaca pela coleta do fluido seromatoso a partir de uma bola de golfe oca e perfurada, implantada subcutaneamente no animal.

¹ LEE, S. *Opossums: Naturebooks Series*. School & Library Binding, 1998.

² A presença de carboidratos pode ter diminuído a interação da proteína com o SDS. Isto resultaria em uma migração mais lenta se comparada a das proteínas sem glicosilação, induzindo a concluir um peso molecular acima do valor real.

³ Sabe-se que o mecanismo de separação por permeação em gel depende não apenas da massa mas também da forma da molécula que, quanto mais alongada e fina, é mais rapidamente eluída que as moléculas globulares de mesmo peso molecular. Possivelmente, tanto DM40 quanto DM43 adotem a conformação dimérica elipsoidal em condições nativas, resultando em massas de valores tão diferentes nos valores das massas.

⁴ HILLAM, R. P.; TENGEDY, R. P.; BROWN, G. D. Local antibody production against the murine toxin *Yersinia pestis* in a golf ball-induced granuloma. *Infect. Immun.*, v. 10, p. 458 – 463, 1974.

Do soro, foi isolada a fração antibotrópica¹ da qual foram separadas DM40 e DM43 através de métodos cromatográficos convencionais. O maior rendimento de DM43 em termos de quantidade recuperada do soro (0,93% contra 0,06% de DM40), possibilitou uma análise mais completa sobre suas propriedades biológicas e estruturais primária que DM40.

I.3.1. A Atividade Inibitória

A fração antibotrópica isolada do soro do gambá apresentou resistência a atividades hemorrágicas, necrosantes, edematogênicas, hiperalgênicas e letais induzidas por venenos de serpentes da família Bothrops. Teoricamente e do ponto de vista terapêutico, seria talvez possível que a caracterização de um único inibidor, no caso a fração antibotrópica de *D. marsupialis* fosse capaz de inibir a ação hemorrágica de diversas espécies de serpentes.

Testes indicaram que a fração antibotrópica do gambá possui uma capacidade de inibição da atividade hemorrágica seis vezes maior que os soros antiofídicos comerciais. Sem levar em conta que é mais interessante administrar uma dose de antídoto mais direcionada à atividade hemorrágica, por exemplo, do que proceder com a vítima uma dose de soro com proteínas de diversas funções, sendo que algumas delas podem causar reações imunológicas ao paciente.

Os efeitos hemorrágicos do veneno da *Bothrops jararaca* são inibidos pela fração antibotrópica do gambá *Didelphis marsupialis* através da sua intervenção nas atividades proteolíticas do veneno sobre caseína, fibrinogênio e proteínas da matriz extracelular (como a laminina, fibronectina, colágeno IV e gelatinas I e V). A inibição desta atividade hemorrágica foi também observada com EDTA (Ácido Etileno-Diamino-Tetra Acético), dando fortes indícios de que estão envolvidas na ação hemorrágica enzimas dependentes de íons, ou melhor, de metaloproteases de venenos de serpentes.

¹ Foi também isolada do leite do gambá uma fração equivalente a esta, apresentado as mesmas características de peso molecular, pl, sequência N-terminal, atividade antiletal e antigênica (Jurgilas, *et al*, 1999). Isto corresponde ao primeiro relato na literatura sobre presença de proteínas com propriedades antiveneno de serpentes em um meio que não o soro ou plasma de animais imunes. Fato este que se assemelha aos anticorpos.

Nos estudos de toxicidade e inibição em ratos, o fator antibotrópico (contendo DM40 e DM43) foi capaz de inibir totalmente a letalidade induzida por $2DL_{50}$ ¹ do veneno da *Bothrops jararaca*. Numa relação 1:1 (massa:massa), DM43 foi capaz de inibir apenas parcialmente a letalidade mesmo quando dobrada a sua quantidade. Interessantemente, a inibição da letalidade total foi obtida na presença do segundo inibidor, DM40, o que principalmente levou a propor a existência de um complexo formado por DM40 e DM43 (Perales, *et al*, 1994). Contrariamente, análises de interação entre estas duas proteínas revelaram que a ação é mais independente e sinérgico do que de complexo.

Quanto à resposta edematogênica e hiperalgésica, atribui-se parcialmente como sendo resultante da ação hemorrágica do veneno da serpente. A inibição de 53% e ~50% dos efeitos edematogênicos e hiperalgésicos do veneno, respectivamente, somado ao fato de que este último foi também inibido por EDTA, evidenciam, mais uma vez, que DM43 atua como um inibidor de metaloprotease do veneno de serpente.

Para estudos de interação entre a DM43 e a metaloprotease foi escolhida a jararagina, proteína responsável pela ação hemorrágica do veneno da *Bothrops jararaca*. É considerada, até hoje, como uma das metaloproteases de veneno de serpente mais bem caracterizadas. Proteína ácida de 52KDa é, segundo análise da sequência de cDNA, constituída de três domínios: catalítico (metaloprotease), tipo-desintegrina e rico em cisteínas.

Testes verificaram a formação de complexos solúveis entre a jararagina e DM43, em uma relação de 1:1 (mol:mol). A retirada do domínio catalítico não resultou na formação de complexos.

Esta interação pode ser reforçada pelo fato de EDTA ter inibido da mesma forma que DM43 os efeitos hemorrágicos do veneno de serpente, os mesmos causados pela jararagina, metaloprotease responsável por este efeito.

¹ DL_{50} : Dose Letal quando causa morte de 50% dos animais testados em um experimento de toxicidade.

I.3.2. Características Físico-Químicas

A DM43 possui características típicas de inibidores de metaloproteases, sendo uma glicoproteína ácida e bastante solúvel a amplos intervalos de pH e temperatura.

A partir da análise completa da composição dos aminoácidos e uma comparação de seu peso molecular por SDS-PAGE antes e após os ensaios de desglicosilação química, pôde-se estimar a quantidade de açúcares ligados a DM43. Obteve-se, então, que os carboidratos contribuem com cerca de 29,0% (espectrometria de massa - MALDI-TOF) e 25,9 / 26,7% (SDS-PAGE) na massa total de DM43, equivalente a aproximadamente 12KDa de massa molecular, líquida, de açúcares.

Com relação ao seu ponto isoelétrico, não foi possível determinar com exatidão devido ao seu alto grau de acidez. Porém, pôde-se certificar que DM43 possui pI abaixo de 3,5 correspondendo a um valor menor que outros inibidores de metaloprotease ou anti-hemorrágicos já descritos na literatura, como: oprina de *Didelphis virginiana* (pI 3,5); HSF de *T. flavoviridis* (pI 4,0); BaSAH de *Bothrops asper* (pI 5,2); antihemorráginas de *Vipera palaestinae* (pI 5,2), *Sigmondon hispidus* (pI 5,4) e *Neotoma micropus* (pI 4,1).

Até então, pouco se soube sobre as características físico-químicas e químicas de inibidores desta espécie em animais resistentes a metaloproteases, proteínas responsáveis pelos efeitos hemorrágicos de venenos de serpentes, e muito menos quanto aos seus aspectos estrutural-funcionais.

I.3.3. Estrutura Primária de DM43

Recentemente, foi determinada a estrutura primária completa da DM43 junto com uma ampla caracterização físico-química e funcional da mesma (Neves-Ferreira, 1999; Neves-Ferreira *et al.*, 2000). Em solução, DM43 forma homodímeros levando teoricamente a uma molécula de aproximadamente 86KDa de massa molecular. A sua seqüência de aminoácidos é composta por 291 resíduos, onde é possível identificar três domínios tipo imunoglobulina. Cada domínio é caracterizado pela presença de dois resíduos de cisteína (seis no total) que devem formar uma ponte dissulfeto entre as fitas B e F de cada domínio tipo imunoglobulina. Com o seqüenciamento, foram detectados quatro possíveis sítios de glicosilação do tipo N-linked, sendo um no primeiro domínio (D₀) e três no segundo domínio (D₁).

II.

Objetivos

Em trabalho conjunto com o Departamento de Fisiologia e Farmacodinâmica da FIOCRUZ-RJ e o Departamento de Bioquímica da UFRJ, o Grupo de Cristalografia de Proteínas e Biologia Estrutural do Instituto de Física de São Carlos - USP iniciou seus estudos acerca da determinação e predição da estrutura tridimensional de DM43, assunto deste trabalho, através de métodos experimentais de difração de raios-X e métodos teóricos de modelagem molecular, respectivamente.

A análise desta estrutura poderá, futuramente, ajudar a entender melhor não somente o mecanismo de ação deste inibidor em específico, mas também de outros inibidores de metaloproteases de venenos botrópicos. Pode-se ainda especular sobre a possibilidade da obtenção de uma terapêutica mais eficiente e menos prejudicial ao homem, que os soros antiofídicos comercialmente disponíveis contra os efeitos letais dos venenos botrópicos de serpentes.

Desta forma, os objetivos deste trabalho compreendem:

- a otimização das condições de cristalização obtidas em trabalhos anteriores por Oliveira *et. al.* (1993), aplicando métodos recentes de busca por condições de cristalização, coletas de dados a temperaturas criogênicas e radiação síncrotron;

- a modelagem molecular por homologia estrutural, segundo satisfação das restrições espaciais, para determinação da estrutura teórica de DM43.
- análise da estrutura de DM43 modelada, com o propósito de compreender e levantar hipóteses sobre o mecanismo de homodimerização de DM43 e complexação com a metaloprotease.

A apresentação deste trabalho está organizada em duas partes. A parte A, detalha a fase experimental do trabalho onde são apresentadas as tentativas de otimização das condições de cristalização e coleta de dados. A parte B descreve a fase teórica do trabalho delineando as etapas da predição da estrutura tridimensional de DM43, possível graças ao recente trabalho de Neves-Ferreira (1999) na determinação da seqüência completa de aminoácidos de DM43.

Parte A

III.

CONSIDERAÇÕES TEÓRICAS

Os conceitos envolvidos nos métodos utilizados na parte A deste trabalho estão brevemente descritos neste capítulo. Inicia-se com a descrição das interações existentes a nível molecular, voltados mais especificamente à cromatografia de interação hidrofóbica e de troca iônica, métodos utilizados neste trabalho. Uma introdução rápida sobre os conceitos básicos de cristalografia e desglicosilação enzimática encerram este capítulo.

III.1. Interações Moleculares a Nível Macroscópico

Os vários tipos de interações existentes entre as superfícies de moléculas podem ser usados como uma ferramenta poderosa para sua manipulação e, portanto, sua caracterização. Isto implica na determinação tanto das propriedades físico-químicas da molécula (usando técnicas tais como eletroforese, focalização isoelétrica, espectroscopia, cromatografia, etc.) e a obtenção dos vários níveis de estrutura molecular (através da difração de raios-X, ressonância magnética nuclear, ressonância paramagnética eletrônica, etc.), quanto da predição do mecanismo de

ação em seu ambiente natural e até o esclarecimento de seu papel em um organismo biológico.

Especificamente às estruturas de moléculas de proteínas, a sua caracterização físico-química (pl, massa molecular, atividade, solubilidade, estabilidade, etc.) consiste na interpretação da reação da proteína sob a ação de agentes externos, tais como campo elétrico, campo magnético, campo eletromagnético, temperatura, constante dielétrica do meio, pressão e pH. A partir destes resultados, é possível selecionar de forma mais crítica a condição de cristalização inicial. Por fim, uma subsequente estrutura tridimensional é derivada a partir da análise dos efeitos de interação do cristal com a radiação-X.

Introduz-se, a seguir, algumas considerações teóricas acerca dos tipos de interações existentes entre proteína-solvente, proteína-proteína e proteína-grupo inorgânico descrevendo ao final, métodos cromatográficos, cristalização e desglicosilação enzimática utilizados neste trabalho com DM43. Métodos estes que se baseiam em tais propriedades de interação, para isolar e agregar de maneira uniforme e ordenada as proteínas em meios estrategicamente selecionados.

III.1.1. Interações Hidrofóbicas

A hidrofobicidade é uma propriedade atribuída a moléculas ou regiões de macromoléculas que são solúveis em solventes apolares e praticamente insolúveis em solventes polares, como a água (Tanford, 1973).

As interações hidrofóbicas têm um papel significativo na estabilidade estrutural tridimensional das proteínas bem como em alguns fenômenos biológicos relacionados à formação de proteínas oligoméricas e complexos entre antígeno-anticorpo, receptores- hormônios e substrato-enzima.

A capacidade das proteínas de interagir através de regiões hidrofóbicas permite que a bioquímica e a físico-química utilizem-nas como ferramenta para isolar uma determinada proteína de uma solução heterogênea, ou então, obter cristais para difração de raios-X.

De fato, a hidrofobicidade não é um fenômeno de atração entre regiões apolares e sim de um processo de agregação espontânea devido a uma intensa força atrativa existente entre as moléculas de água.

Esta grande força de atração leva a formação de uma reentrância no solvente ao redor das superfícies apolares, que equivale ao comportamento visto na superfície de uma porção aquosa na parede de um recipiente, por exemplo. Pode-se dizer que a força que rege a estrutura isotrópica de moléculas de água que se forma ao redor das regiões hidrofóbicas da proteína (camada de solvatação) (figura III-1), nada mais é do que o fenômeno físico denominado tensão superficial.

Figura III-1: Camadas organizadas de moléculas de água em torno da região hidrofóbica da proteína. Longe dela, as moléculas dispõem-se de forma livre, desordenada.

Devido a alta força de coesão entre as moléculas de água, esta camada de solvatação não é facilmente desfeita senão na presença de um soluto que tenha uma força de interação forte o suficiente para rompê-la.

Acrescentando-se sais ou solventes orgânicos na solução, a estrutura interna da água se modificará e, conseqüentemente, o fenômeno de interação hidrofóbica começará a surgir entre as proteínas. Termodinamicamente, a desordem da água provoca um aumento na entropia ($\Delta S > 0$), e uma vez que a energia de interação (ΔH) é pequena (DeHoff, 1993), a energia livre pode se tornar negativa, implicando em uma reação de interação espontânea. A definição de energia livre, $\Delta G = \Delta H - T\Delta S$, também revela uma dependência da hidrofobicidade pela temperatura, significando que altas temperaturas resultam em valores negativos de energias livres, e conseqüentemente no favorecimento do fenômeno hidrofóbico.

Isto corresponde a um dos fenômenos de força iônica, o *salting-out*, fortemente dependente das interações hidrofóbicas, constituindo base para as várias técnicas de purificação (cromatografia por interação hidrofóbica e fase reversa) e cristalização de proteínas (difusão de vapor, diálise e batch).

III.1.2. Força Iônica

O *salting-out* pode ser alternativamente explicado a partir de um fenômeno completamente diferente do hidrofóbico: as interações hidrofílicas ou polares.

Os conceitos de força iônica têm base na teoria de Debye-Huckel (Blundell e Johnson, 1976), onde faz-se uma extensão da interpretação da solubilidade de íons pequenos para proteína, tratando-a como um grande íon polivalente.

Desta forma, as proteínas competem por moléculas de água do mesmo modo que os íons de sal. Porém, quando a presença de íons salinos é significativa, a camada de solvatação da proteína começa a se desfazer perdendo as moléculas de água para os sais. Isto implica em uma alteração local da constante dielétrica do meio (item III.1.3) ao redor da proteína e, conseqüentemente da força de atração eletrostática com outras proteínas próximas a ela, que também sofrem com a remoção de moléculas de água pelo sal. Conseqüentemente, as moléculas de proteína começam a se agregar formando sólidos amorfos ou não. Diz-se, com isto, que houve uma diminuição da solubilidade da proteína.

O fenômeno descrito acima pode ser resumido de maneira aproximada, através da seguinte equação:

$$S \approx S_o e^{-K_S \mu} \quad \text{Eq. (1)}$$

sendo que: S , é a solubilidade da proteína na presença de sais; S_o , a solubilidade da proteína na ausência de sais; $\mu = \frac{1}{2} \sum_j c_j z_j^2$, a força iônica do meio equivalente a metade do somatório do produto entre a concentração (c_j) do j -ésimo íon contido na solução e a carga (z_j) deste íon ao quadrado; e K_S , uma constante.

Observa-se na equação (1) um comportamento exponencial decrescente da solubilidade S com o aumento da força iônica do meio, completamente determinada por K_S e o tipo de íon contido na solução, representado por z_j . Esta interpretação é

mais bem visualizada na figura III-2, onde estão graficadas as curvas do logaritmo da solubilidade de uma proteína com relação à força iônica de vários tipos de íons.

Figura III-2: Curva do logaritmo da solubilidade da carbonomonoxihemoglobina com relação à força iônica de vários sais na região de salting-out (Blundell e Johnson, 1976).

O mesmo comportamento não é visto a baixas concentrações de sal onde, ao contrário, observa-se um aumento de solubilidade da proteína, com a força iônica (figura III-3). Este fenômeno é denominado *salting-in* e também interpretável pela teoria de Debye-Huckel.

Figura III-3: Curva do logaritmo da solubilidade da carbonomonoxihemoglobina com relação à raiz quadrada da força iônica de vários sais na região de *salting-in* (região crescente das curvas) e *salting-out* (região decrescente das curvas) (Blundell e Johnson, 1976).

A região de *salting-in* abrange aproximadamente o intervalo de 0,0-0,5M de sal. Ela é adotada tanto para fins cromatográficos como uma etapa do protocolo de purificação (uma diálise, por exemplo), quanto para cristalização de proteínas de membrana cujas extensas regiões hidrofóbicas induzem uma fácil precipitação. Sendo o *salting-in* uma região onde a solubilidade aumenta com o aumento da concentração de sal, ela pode ser utilizada como alternativa para se chegar a uma agregação mais lenta e, portanto, mais ordenada, deste tipo de proteína.

III.1.3. Outros Fatores que Influenciam as Interações Moleculares

A constante dielétrica do meio é um outro fator que contribui para a solubilidade de uma molécula. Segundo as leis físicas da eletrostática, a constante dielétrica pertence a uma função inversamente proporcional a força eletrostática de duas partículas carregadas. Portanto, quanto maior a constante dielétrica, mais fraca é a força de interação, seja ela atrativa ou repulsiva.

A água é uma das poucas substâncias que possuem constante dielétrica alta ($\epsilon_r = 78,00$ a 25°C) (Atkins, 1998), devido a isto, sólidos unidos por interações iônicas podem ser dissolvidos completamente em uma solução aquosa, uma vez que a água, devido a sua constante dielétrica alta, é capaz de enfraquecer esta interação e separar entre si os íons que constituem o sólido.

Existe, então, a possibilidade de aumentar ou diminuir as interações entre as moléculas (inclusive de proteínas) variando a constante dielétrica do meio. Os solventes orgânicos (ver tabela III-1) possuem a capacidade de diminuir grandemente a constante dielétrica de um meio aquoso aumentando a força Coulômbica ou iônica de interação entre as moléculas do meio, levando-as à conseqüente diminuição de sua solubilidade.

Tabela III-1: Lista das constantes dielétricas de alguns solventes orgânicos utilizados em cristalização de proteínas (McPherson, 1989).

Nome	Constantes Dielétricas
Acetonitrila	37,50
1,3-Propanediol	35,00
Metanol	32,80
Etanol	24,30
Acetona	20,70
Isopropanol	18,30
Butanol	17,10

A solubilidade ou a capacidade de interação com o solvente depende também do pH do meio. Há uma região no espaço dos pHs onde a carga elétrica líquida da superfície da proteína começa a ser neutralizada pelos íons H_3O^+ ou OH^- contidos na solução. Ao atingir o estado de completa neutralização, ou seja, carga líquida da proteína igual a zero, o pH do meio terá atingido o ponto isoelétrico da proteína. Frequentemente nesta condição, ela começa a se agregar com outras proteínas precipitando de modo ordenado ou não, dependendo da velocidade do processo.

III.2. Cromatografia

O processo de cristalização é tido como uma forma de separar proteínas, e a capacidade de cristalizar como um indicativo do seu grau de pureza. A tendência de formar cristais de melhor qualidade a cada recristalização, sugere fortemente que a pureza da solução de proteína é um dos principais fatores para constituir a perfeita formação e crescimento do cristal.

A falta de pureza pode resultar em: fracasso total nas tentativas de cristalização; formação de cristais muito pequenos impróprios para fornecer um padrão de difração adequada para análises cristalográficas; e até na formação de cristais grandes e morfologicamente bonitos, mas sem poder de difração. Existem várias técnicas cromatográficas (como adsorção, troca-iônica, interação hidrofóbica, cromatofocalização, gel filtração, cromatografia de afinidade e fase reversa) que separam as inúmeras espécies de proteína e impurezas de acordo com as diferenças nas suas características físico-químicas e morfológicas, que podem ser bem aproveitadas para aumentar as chances de uma cristalização bem sucedida.

A cromatografia envolve três elementos: a fase estacionária, que compreende uma matriz rígida e quimicamente estável (como celulose, agarose, dextran, poliacrilamida, sílica ou poliestireno), ligados a grupos funcionais que definem uma dentre as várias técnicas cromatográficas; a fase móvel ou solução com pH bem definido, podendo conter sal ou aditivos dependendo da propriedade cromatográfica escolhida; e a solução que contém a amostra a ser purificada.

As duas técnicas utilizadas neste trabalho (cromatografia hidrofóbica e de troca-iônica) estão descritas a seguir com mais detalhes.

III.2.1. Cromatografia por Interação Hidrofóbica¹

A superfície hidrofóbica de uma proteína quanto a sua abrangência e número varia de proteína para proteína. Estas diferenças possibilitam distinguir as várias espécies de proteínas dentro de uma solução heterogênea através da combinação de interações entre um meio sólido apolar e um meio líquido polar, privilegiando proteínas com características mais, ou menos hidrofóbicas.

A cromatografia por interação hidrofóbica utiliza uma fase líquida de alta força iônica e uma fase estacionária ligeiramente hidrofóbica. A ligação da proteína com os grupos apolares da matriz (figura III-5) é promovida pela alta força iônica do meio que reforça a propriedade hidrofóbica da proteína (figura III-6), removendo a camada de solvatação tanto da resina quanto das regiões apolares da superfície da proteína (*salting-out*). A eluição é feita com a diminuição da concentração de sal (gradiente de força iônica decrescente) de forma que as interações hidrofóbicas comecem a se enfraquecer e, conseqüentemente, desprender as proteínas da resina.

Figura III-5: Grupos apolares utilizados na cromatografia por interação hidrofóbica.

Este fenômeno já foi descrito na seção anterior, e segundo a definição termodinâmica da energia livre de Gibbs, a interação hidrofóbica pode ser

¹ Harris e Angal, 1993; Scopes, 1994; Harris e Angal, 1995

aumentada se a reação ocorrer a baixas temperaturas, lembrando que:
 $\Delta G = \Delta H - T\Delta S$.

Se a fase móvel possuir um pH próximo ao pI da proteína, a interação hidrofóbica com a resina será ainda maior, pois assim, a carga total da proteína será aproximadamente nula.

Figura III-6: Etapas de uma interação entre regiões hidrofóbicas de proteína-resina (cromatografia hidrofóbica) ou eventualmente de proteína-proteína durante a cristalização.

A Cromatografia por Fase-Reversa utiliza o mesmo princípio, mas contrariamente a Cromatografia por Interação Hidrofóbica, possui a fase estacionária extremamente hidrofóbica (tipicamente sílica ligado a grupos carbonetos C4, C8 ou C18) e a fase móvel ligeiramente polar (geralmente água ou solução aquosa contendo solventes orgânicos). (Harris & Angal, 1993; Perseptive Biosystems - Manual, 1996).

III.2.2. Cromatografia por Troca Iônica

O número relativo dos grupos polares positivos e negativos expostos na superfície da proteína e o pH do meio, definirão a carga líquida da molécula (fig. III-7). A interação iônica desta com os trocadores catiônicos (CM ou SP) (figura III-8(a)) ou aniônicos (DEAE ou QAE) (figura III-8(b)) ligados na matriz separarão as várias espécies de proteínas.

A eluição pode ser feita segundo um gradiente crescente de concentração de sal em que a força iônica da fase móvel rompe a ligação das proteínas com os grupos polares da matriz, através de seus contra-íons (figura III-9).

Figura III-7: Variação na carga líquida da proteína com o pH. A escolha do pH da fase móvel definirá, de certa forma, a escolha da cromatografia em catiônica ou aniônica (Texas A&M University-Kingsville - <http://ntri.tamuk.edu>).

Figura III-8: Representação dos dois tipos de interação iônica. (a) Interação entre grupos de trocadores catiônicos e proteína; (b) interação entre grupos de trocadores aniônicos e proteína. Figuras obtidas do manual de cromatografia da PerSeptive Biosystems.

Figura III-9: Etapas da eluição de uma proteína (em vermelho) de carga líquida negativa ligada em um trocador aniônico. A eluição se torna cada vez mais efetiva com o aumento da concentração de íons na solução (Texas A&M University-Kingsville - TAMUK).

Os trocadores catiônicos e aniônicos são classificados em “fortes” ou “fracos”, segundo a capacidade de manter constante sua carga efetiva, a longos ou curtos intervalos de pH da fase móvel (tabela III-2).

Tabela III-2: Curvas de titulação de trocadores aniônicos e catiônicos fortes e fracos. Os trocadores fortes (QAE e SP) conseguem manter suas cargas a grandes intervalos de pH, o que não ocorre com os trocadores fracos (DEAE e CM) (Pharmacia Biotech).

Grupos funcionais	Fórmulas	Curvas de Titulação
DEAE (Dietilaminoetil) Trocador Aniônico Fraco	$-C_2H_4NH(C_2H_5)_2$	
QAE (Dietil-2-hidroxi- propilamino- etil) Trocador Aniônico Forte	$-C_2H_4N^+(C_2H_5)_2$ $CH_2CH(OH)CH_3$	
CM (Carboximetil) Trocador Catiônico Fraco	$-CH_2COO^-$	
SP (Sulfopropil) Trocador Catiônico Forte	$-C_3H_6SO_3^-$	

III.3. Cristalografia

Existem, atualmente, várias técnicas que podem ser utilizadas para a determinação da estrutura tridimensional de uma molécula biológica. Pode-se listar a Cristalografia de Raios-X, Ressonância Magnética Nuclear e, muito recentemente, a Microscopia Eletrônica e difração de elétrons. Por enquanto, a mais antiga das técnicas, a Cristalografia de Raios-X, continua sendo a mais utilizada para a determinação de estruturas de proteínas globulares, DNA, RNA e até estruturas de grande porte, como vírus.

Entretanto, a obtenção de uma boa estrutura cristalográfica depende não somente da qualidade de todo o aparato experimental (fonte de raios-X, detector, sistema criogênico, entre outros), como também do cristal a ser utilizado.

Satisfeitos estes dois quesitos, parte-se para a escolha do método de determinação da estrutura podendo ser, em casos de estruturas inéditas, a substituição isomórfica múltipla ou espalhamento anômalo em diversos comprimentos de onda, e nos casos de similaridade seqüencial com uma estrutura conhecida superior a ~40%, a substituição molecular (Rossmann, 1990).

III.3.1. Cristalogênese de Proteínas

A determinação de estruturas por difração de raios-X depende intrinsecamente da obtenção de bons cristais. Para isto, é necessário que a proteína precipite de modo ordenado. Este precipitado de proteínas somente será possível quando elas estiverem no nível de “insolubilidade fraca”, onde a agregação é lenta o suficiente para não resultar em aglomerados amorfos. A busca por este mínimo ideal de solubilidade da proteína é muitas vezes árdua, uma vez que envolvem vários parâmetros como, concentração de íons na solução, pH, temperatura, quantidade de proteína na solução, pureza, homogeneidade estrutural da proteína, entre outros.

Esta busca é grandemente facilitada com a utilização de soluções de cristalização comercialmente disponíveis como Fatorial I e II da Hampton Research, entre outras, soluções estas desenvolvidas a partir de um levantamento das melhores condições de cristalização empiricamente encontradas em pesquisas anteriores (Jancarik, Kim, 1991). Por englobar uma vasta variedade de pHs,

precipitantes e aditivos, os métodos de *Crystal Screen* ou Fatorial proporcionam rapidez na busca preliminar da condição de cristalização e economia na quantidade de proteína a ser utilizada.

É muito comum em um experimento de cristalização obter precipitados amorfos, fibras, placas ou ainda cristais de tamanho insuficiente para gerar padrões de intensidade adequada, ou frágeis demais para serem montados em um difratômetro (Nave, 1998). Geralmente, estes problemas estão relacionados com a maneira em que as moléculas de proteínas empacotam sobre a superfície de um núcleo em crescimento. Este processo pode ser interrompido ou mal sucedido na presença de impurezas (figura III-10), concentrações inadequadas de sais ou precipitantes, pH do meio induzindo uma distribuição de cargas não favoráveis para uma interação mais estável, além da possibilidade de ocorrer uma microheterogeneidade (ou “heterogeneidade em macromoléculas puras” (Ducruix & Giegé, 1992)) na distribuição de cargas ou estrutura entre as moléculas de proteína.

Figura III-10: Incorporação de uma impureza no cristal alterando a disposição ordenada das proteínas ao seu redor. Este defeito causa uma instabilidade local que logo mais se propagará por toda a rede. (McPherson, 1999)

Métodos eficientes de monitoração do crescimento destes cristais têm surgido a fim de verificar a existência de problemas durante o crescimento. Desta forma, é possível identificar os parâmetros certos para eliminá-los. Listam-se, por exemplo:

- Microscopia de Força Atômica (AFM) (McPherson, A. *et al.*, 1995; Malkin, A. J. *et al.*, 1995a; Malkin, A. J. *et al.*, 1995b; Malkin, A. J. *et al.*, 1996; Kuznetsov, Y. G. *et al.*, 1999): faz uma varredura na superfície do cristal *in situ*, com resolução espacial da ordem de décimos de nanômetros, permitindo por exemplo visualizar partículas de proteínas na rede cristalina sem perturbar o crescimento do cristal;
- Espalhamento Quasi-Elastico (QELS) (McPherson, A. *et al.*, 1995): possibilita o estudo do processo de nucleação de macromoléculas na solução de cristalização, permitindo medir a taxa de crescimento do núcleo até atingir um tamanho suficiente para visualização direta por um microscópio de luz visível;
- Interferômetro de Michaelson (McPherson, A. *et al.*, 1995; Kuznetsov, Y.G. *et al.*, 1995; Kuznetsov, Y.G. *et al.*, 1996): mede a cinética de incorporação de uma molécula na superfície do cristal;
- Interferômetro de Mach-Zehnder (McPherson, A. *et al.*, 1995): corresponde a um tipo diferente de interferômetro utilizado para investigar de forma quantitativa o gradiente de concentração (McPherson, A. *et al.*, 1999) de proteína ao redor dos cristais em crescimento (*depletion zones*) (figura III-11).

Figura III-11: Zona de depleção, ou gradiente de concentração, revelada pela área escura ao redor do cristal do vírus STMV (*satellite tobacco mosaic virus*) em crescimento. Imagem obtida com a utilização do Interferômetro de Mach-Zehnder. (McPherson, 1999).

Desta forma, surgem a cada ano vários artigos propondo diversas otimizações das já existentes técnicas de cristalização. Dentre elas, podem ser citadas as seguintes propostas:

- cristalização sob a ação de um campo magnético de até 10 Teslas (Careri, 1977; Yamagishi, 1990; Ataka, 1997; Sazaki, 1997; Sakurazawa, 1999; Yanagiya, 1999);
- cristalização sob radiação laser, que utiliza o perfil gaussiano de sua intensidade para que a pressão de radiação seja capaz de forçar a formação de cristais de proteína no foco do feixe¹ (Ashkin, 1970; Ashkin, Dziedzic, 1975; Ashkin, 1980; Ashkin *et al.*, 1986; Ashkin, Dziedzic, Yamane, 1987; Ashkin, Dziedzic, 1987; Wright *et al.*, 1990; Block, 1992; Svoboda, Block, 1994; Bancel *et al.*, 1998; ; Bancel *et al.*, 1999)
- utilização de óleos (Chayen, 1997; Baldock, 1996; D'Arcy, 1996) e géis (Moreno, 1999) para diminuir a velocidade de evaporação do solvente da solução de cristalização e, portanto, levar à condição de supersaturação da proteína de forma mais lenta. Ou ainda, utilizá-los de maneira que permita o crescimento livre do cristal, sem entrar em contato com as paredes do recipiente (Lorber e Giegé, 1996);
- microgravidade (McPherson, DeLucas, 1999; McPherson *et al.*, 1999), onde devido à baixa força gravitacional, o gradiente de concentração de proteínas ao redor do núcleo ou semente (*depletion zone*) seja uniforme em todas as direções e a acomodação das proteínas na superfície do núcleo, mais lenta;
- eliminação das microeterogeneidades (McPherson, 1989) que têm origem na degradação proteolítica, glicoproteínas, oxidação das cadeias laterais, flexibilidade na conformação, entre outros.

Seguindo as dificuldades encontradas durante a cristalogênese, há ainda os problemas de danos no cristal causados pela radiação durante a coleta dos padrões de difração (Nave, 1995). Para tanto, a criocristalografia adota métodos de congelamento do cristal suspendendo-o por uma fina película de solvente vitrificado, o que melhora não apenas a intensidade relativa entre os picos e a radiação de fundo, mas principalemtno o tempo de vida do cristal (Garman, Schneider, 1997; Garman, Mitchell, 1996; Xiao, Gamblin, 1996).

¹ A pressão de radiação também tem sido utilizada como pinça óptica, que poupa os frágeis cristais de proteína da agressão mecânica durante sua manipulação.

III.3.2. Difração de Raios-X

O fenômeno de difração ocorre sempre que uma radiação atravessa um obstáculo ou fenda de dimensões da ordem do seu comprimento de onda. O resultado é um espalhamento de um padrão de intensidades construtivas e destrutivas, conseqüência de um somatório de amplitudes e fases das frentes de onda. A grande utilidade deste fenômeno é que as características da fenda e do padrão gerado estão relacionadas por uma Transformada de Fourier.

Partindo-se deste princípio, a cristalografia de macromoléculas consegue determinar, em princípio, as suas características estruturais (em termos de posições atômicas, etc.) a partir de um padrão de difração. Para isto, a radiação utilizada deverá ter um comprimento de onda da ordem de pelo menos uma ligação interatômica covalente, que é por volta de $1,2\text{\AA}$, assim compreendendo a região espectral dos raios-X.

Em objetos tridimensionais, o fenômeno é descrito segundo:

- a Lei de Bragg, que descreve o fenômeno de espalhamento na coordenadas do cristal (ou no espaço direto), segundo:

$$2 \cdot d \cdot \sin \theta = n \cdot \lambda \quad \text{Eq. (1)}$$

onde $d \rightarrow$ distância entre dois planos consecutivos do espaço direto (cristal);

$\theta \rightarrow$ ângulo de incidência do feixe com relação aos planos em questão;

$n \rightarrow$ número inteiro (1, 2,...) indicando a ordem da reflexão e ao mesmo tempo impondo a condição de interferência construtiva entre os feixes espalhados pelos planos pertencentes à mesma família;

$\lambda \rightarrow$ comprimento de onda da radiação.

- a construção de Ewald, uma variação da Lei de Bragg descrita nas coordenadas da Transformada de Fourier do espaço direto (ou espaço recíproco).

A construção ou esfera de Ewald comprova a Lei de Bragg utilizando a geometria do espaço recíproco. Para cada ponto da rede recíproca do cristal

que cruzar o perímetro de uma esfera de raio $1/\lambda$ satisfará a condição de difração dada pela equação (1).

- a condição de Laue, também corresponde a uma derivação da Lei de Bragg. Determina a direção do vetor de espalhamento S , a partir dos índices de Miller (h, k, l) que caracterizam a família de planos em questão, e dos vetores do espaço direto (a, b, c):

$$a \cdot S = h \cdot \lambda \qquad b \cdot S = k \cdot \lambda \qquad c \cdot S = l \cdot \lambda$$

Eq. (2)

Uma vez satisfeitas as condições para ocorrer difração, as únicas informações disponíveis de um experimento de difração por raios-X de um cristal, serão as intensidades e as posições relativas dos seus picos, nas coordenadas h, k e l .

A intensidade da radiação espalhada pelo cristal é descrita como (Giacovazzo *et al.*, 1995):

$$I_{hkl} = k_1 k_2 I_o L P T E |\vec{F}_{hkl}|^2$$

Eq. (3)

$$I_{hkl} = \left(\frac{e^4}{m^2 c^4} \right) \left(\frac{\lambda^3 \Omega}{V^2} \right) I_o \left(\frac{1}{\sin(2\theta)} \right) \left(\frac{1 + \cos^2 2\theta}{2} \right) \left(\frac{1}{V} \int_V e^{-i\mu(p+q)} dv \right) E |\vec{F}_{hkl}|^2$$

Eq. (4)

onde k_1, k_2, L, P, T e E correspondem, respectivamente, os termos delimitados em parênteses na equação (4).

Os termos L, P, T e E representam os fatores atenuadores da intensidade do feixe direto (I_o), devido as diferentes interações com o cristal (tabela III-3).

Tabela III-3: Indicação de cada um dos fatores que influenciam na atenuação da intensidade da radiação difratada.

Termos	Funções
k_1	Proveniente da indução eletromagnética em um elétron. e, m e c são constantes universais.
k_2	Ω , volume do cristal; V , volume da cela unitária; λ , comprimento de onda da radiação.
L	Correção de Lorentz. Tempo de permanência de cada ponto do espaço recípro na condição de difração.
P	Correção da polarização do feixe.
T	Correção na absorção da radiação pelo cristal. p e q , caminhos ópticos do feixe incidente e difratado, respectivamente. μ , coeficiente de absorção linear.
E	Coeficiente de extinção. Depende da mosaicidade do cristal, da extinção primária e secundária, relacionadas com a o grau de intensidade do feixe incidente recebido por cada um dos planos e a interferência destrutiva devido a múltiplas reflexões entre os planos da mesma família.

$|\vec{F}_{hkl}|^2$ corresponde ao termo que modula a intensidade do feixe incidente para ser difratado. Dado em termos do fator de forma atômico f_j pode-se dizer que o fator de estrutura $\vec{F}_{hkl}$ é um vetor resultante do somatório de todos os vetores de amplitude f_j , cuja direção e sentido são expressos como $\exp[2\pi i(hx_j+ky_j+lz_j)]$. Desta forma:

$$\vec{F}_{hkl} = \sum_{j=1}^N f_j e^{2\pi i(hx_j+ky_j+lz_j)} e^{-8\pi^2 \langle r_j^2 \rangle \sin^2 \theta / \lambda^2} \quad \text{Eq. (5)}$$

- N → número total de átomos contidos na cela unitária;
- $8\pi^2 \langle r_j^2 \rangle$ → fator de Debye-Waller, relacionado com a vibração isotrópica média do j -ésimo átomo;
- h, k, l → amplitudes do vetor do espaço recíproco;
- x_j, y_j, z_j → amplitudes do vetor do espaço direto correspondentes ao j -ésimo átomo.

Nota-se pelas equações (3) e (5) que a intensidade I_{hkl} do raio difratado com sua respectiva localização (h, k, l) depende também da distribuição atômica (x, y, z) na cela unitária.

A Transformada de Fourier da equação (5) permite conhecer a distribuição de densidade eletrônica da proteína que compõe o cristal:

$$\rho(x, y, z) = \frac{1}{V} \sum_{h,k,l=-\infty}^{+\infty} F_{hkl} e^{-2\pi i(hx+ky+lz)} \quad \text{Eq. (6)}$$

Curiosamente, é desejado um certo grau de mosaicidade (figura III-12) no cristal ponderando entre o não muito alto a ponto de assemelhar-se a um policristal (Nave, 1998) e o não muito baixo para que não ocorra como em um cristal perfeito, uma interferência destrutiva entre os raios difratados pela mesma família de planos (ver coeficiente de extinção, E na tabela III-3).

Figura III-12: Representação exagerada de um cristal real, mostrando detalhes da mosaicidade composta por blocos ou microcristais ideais.

III.4. Desglicosilação Enzimática

Os carboidratos ligados a proteínas fazem parte de uma lista de fontes de microeterogeneidades (como, por exemplo, desamidação de asparaginas ou glutaminas, proteólises, instabilidade conformacional e desnaturação parcial) que podem existir mesmo em soluções puras de proteína (McPherson, 1989; Ducruix & Giegé, 1992).

As glicoproteínas freqüentemente trazem problemas na cristalização não somente devido ao alto grau de flexibilidade e variabilidade no comprimento de suas cadeias oligossacarídicas, mas também devido à presença de ácidos siálicos (geralmente presente em glicoproteínas de origem animal) em seus terminais que podem modificar a distribuição líquida de cargas na superfície da proteína.

Há duas maneiras de eliminar completa ou parcialmente a microeterogeneidade em glicoproteínas: pela retirada do ácido siálico, ou de toda ou parte da cadeia de carboidratos.

A neuraminidase é comumente utilizada para a retirada do ácido siálico, eliminando suas cargas negativas a pH neutro. Dependendo do tipo de ligação da proteína com os carboidratos, a digestão completa pode ser obtida através das enzimas PNGaseF, para o caso em que estão ligados a Asparaginas, e O-Glicosidase DS quando ligadas a Serinas/Treoninas.

A PNGaseF¹ é uma endoglicosidase que remove toda a cadeia de carboidratos ligados em asparaginas transformando-as, neste processo, em ácidos aspárticos. Com esta desamidação, a proteína ganha uma carga negativa para cada oligossacarídeo removido.

A figura III-13 apresenta um esboço dos principais passos de detecção de glicoproteínas.

¹ Também denominada de Glicopeptídeo N-Glicosidase, ou N-Glicanase (Peptídeo-N4-(N-Acetil- β -D-Glucosaminil) Asparagina Amidase F), cuja inicial F provém do nome do organismo de origem *Flavobacterium meningosepticum*. Possui massa molecular de 34779 Da e 314 aminoácidos (Tarentino, 1990).

Figura III-13: Etapas do processo de detecção de carboidratos ligados a proteínas (Manual Bio-Rad, 1998/1999)

Após as etapas de digestão dos oligossacarídeos pela PNGase F (passo 2), a amostra é aplicada em um gel de SDS-PAGE (passo 3) para em seguida ser transferida para uma membrana de nitrocelulose (PVDF-polivinilideno difluoreto), sob ação de um campo elétrico transversal. A amostra é então fixada na

membrana, e as proteínas cujos carboidratos não foram completamente digeridos serão oxidados e marcados com biotina. O complexo estreptavidina e fosfatase alcalina identifica as posições destas biotinas e reage na presença de NBT (nitroblue tetrazólio) e BCIP (fosfato 5-bromo-4cloro-3-indolil) colorindo, deste modo, as bandas relativas às proteínas que não foram totalmente deglicosiladas.

Diferentemente dos métodos químicos de desglicosilação, a remoção dos carboidratos por processos enzimáticos aumenta as chances da não desnaturação da proteína, permitindo manter seu enovelamento original. Desta forma, a desglicosilação enzimática se torna uma ótima ferramenta para entender o papel da glicosilação no enovelamento e atividade de uma proteína, além resolver o problema de microeterogeneidades na cristalização.

IV.

MATERIAIS E MÉTODOS

IV.1.DM43

Dois lotes de 5,0mg de DM43 pura (>95%) e liofilizada utilizados neste trabalho foram fornecidos pelos profs. Drs. Gilberto Barbosa Domont e Jonas Perales do Departamento de Química da UFRJ e Departamento de Fisiologia e Farmacodinâmica da Fundação Oswaldo Cruz (FIOCRUZ-RJ), respectivamente. A purificação foi realizada segundo a metodologia descrita por Neves-Ferreira *et al.*, 2000.

IV.2.Coefficiente de Extinção Molar de DM43

O coeficiente de extinção molar e a absorbância a vários comprimentos de onda foram obtidos a partir da sequência de aminoácidos da DM43 (Neves-Ferreira *et al.*, a submeter) usando o programa "ProtParam tool" da ExPASy (Expert Protein Analysis System - <http://expasy.cbr.nrc.ca/tools/protparam.html>), onde os cálculos são feitos baseando-se na composição de aminoácidos da proteína (Gill, von Hippel, 1989).

O valor de $56120 \text{ M}^{-1}\text{cm}^{-1}$ foi encontrado para o coeficiente de extinção molar à 280nm com uma absorbância de 1,733, equivalente a uma solução 1mg/ml. Estes foram os valores utilizados durante toda a parte experimental para a determinar a concentração, em solução, de DM43.

IV.3.Cristalização

Reagentes dos laboratórios Aldrich, Hampton, Mallinckrodt, Merck e Sigma foram utilizados no preparo de todas as soluções utilizadas neste trabalho, inclusive em cristalização e em criocristalização. Suas soluções foram filtradas com Millex-GS $0,22\mu\text{m}$ estéril da Millipore. Do laboratório da Hampton Research foi utilizado o Crystal Screen I, pacote de 50 diferentes soluções de cristalização prontas para uso.

Na cristalização adotou-se o método de difusão de vapor por *hanging drop* em placas Linbro contendo 24 poços da Flow Laboratories Inc.. As gotas foram montadas em lamínulas (Dow Corning) siliconizadas para que, por tensão superficial, fiquem suspensas com sua face voltada para a solução do poço. Para o isolamento do ambiente de cristalização foi colocada uma camada de graxa de vácuo (Dow Corning) entre a borda plana do poço e a lamínula.

As condições de cristalização com sulfato de amônio foram determinadas através do teste de precipitação que consiste em adicionar frações de $1\mu\text{l}$ de precipitante em $5\mu\text{l}$ de solução de proteína até chegar ao ponto de precipitação. Placas de cristalização foram montadas utilizando uma série de concentrações ligeiramente inferiores ao valor encontrado, sob diferentes pHs (pH4,0, 5,0, 6,0, 7,0 e 8,0).

As concentrações de sulfato de amônio estimadas para a cristalização de DM43 foram muito semelhantes às de Oliveira (1993). As condições finais de cristalização estão detalhadas nas tabelas IV-1(a) - (d).

Tipicamente, a dissolução de DM43 foi em 0,02M de HEPES pH 7,4. Nas condições de cristalização descritas na tabela IV-2, a DM43 foi também dissolvida em água Milli-Q_{plus} (Millipore) e em 0,02M de Acetato de Sódio pH 5,0.

Tabela IV-1: Experimentos de Cristalização da DM43. Os pHs referem-se aos seguintes tampões: pH 4,0, Ácido Cítrico; pH 4,5, Citrato de Sódio; pH 5,0, Acetato de Sódio; pH 5,6, Citrato de Sódio; pH 6,0, Ácido Cacodílico; pH 6,5, MES; pH 7,0, HEPES; pH 8,0, TRIS; todos na concentração de 0,1M. O volume de 1,0ml foi aplicado na solução do poço, com exceção do experimento 3 (Fatorial I) que foi de 0,5ml. A relação 1:1 (v:v) de solução do poço para solução de proteína foi adotado em todos os experimentos de forma que a concentração inicial de precipitante na gota seja metade da concentração do poço.

(a) Sulfato de Amônio¹

Experimento	Proteína	[Proteína] ² (mg/ml)	Proteína dissolvida em	Intervalo de pH	Passo pH	Precipitante	Intervalo (%v/v)	Passo (%v/v)	Aditivos	Gota (µl)	Temp. (°C)
1	DM43 Nativa	5,11	0,02M de HEPES - pH7,4	4,0 - 8,0	1,0	Sulfato de Amônio	46-56	2	-	6,0	18,0
2	DM43 Nativa	5,11	0,02M de HEPES - pH7,4	4,0 - 8,0	1,0	Sulfato de Amônio	50-56	2	-	2,0 e 3,0	4,0

(b) Fatorial I

(b) Fatorial I				Fatorial I (50 condições diferentes de cristalização - Anexo-I)		Gota (µl)	Temp. (°C)
Experimento	Proteína	[Proteína] (mg/ml)	Proteína dissolvida em				
3	DM43 Nativa	8,00	0,02M de HEPES - pH7,4				

(c) Screening do Fatorial I nº 9

Experimento	Proteína	[Proteína] (mg/ml)	Proteína dissolvida em	Intervalo de pH	Passo pH	Precipitante	Intervalo (%w/v)	Passo (%v/v)	Aditivos	Gota (µl)	Temp. (°C)
4	DM43 Nativa	8,00	0,02M de HEPES - pH7,4	4,5/5,6/6,5	-	PEG4000	30	-	0,2M Acetato de amônio	4,0	18,0
5	DM43 Nativa	8,00	0,02M de HEPES - pH7,4	4,5/5,6/6,5	-	PEG4000	30	-	-	4,0	18,0

(d) Screening do Fatorial I nº 15

Experimento	Proteína	[Proteína] (mg/ml)	Proteína dissolvida em	Intervalo de pH	Passo pH	Precipitante	Intervalo (%w/v)	Passo (%v/v)	Aditivos	Gota (µl)	Temp. (°C)
6	DM43 Nativa	8,00	0,02M de HEPES - pH7,4	5,6/ 6,5	-	Sulfato de Amônio	30	-	0,2M Sulfato de Amônio	4	18
7	DM43 Nativa	8,00	0,02M de HEPES - pH7,4	5,6/ 6,5	-	Sulfato de Amônio	30	-	-	4	18

¹ Nos experimentos de Oliveira (1993) foram testadas as concentrações de 52%, 54% e 56% de sulfato de amônio em pHs 5,0, 6,0, 7,0 e 8,0.

² A concentração de proteína foi calculada baseando-se em medidas de absorbância em espectrofotômetros, e comparada com o valor teórico calculado pelo programa ProtParam (ExpAsy) (seção IV.2).

Tabela IV-2: Experimentos de Cristalização da DM43 Nativa e Desglicosilada. Volume da solução do poço: 1,0ml.

Sulfato de Amônio

Experimento	Proteína	[Proteína] ¹ (mg/ml)	Proteína dissolvida em	Intervalo de pH	Passo pH	Precipitante	Intervalo (%w/v)	Passo(%v/v)	Aditivos	Gota (µl)	Temp. (°C)
8	DM43 Nativa	8,11	H ₂ O	5,0	-	Sulfato de Amônio	50 - 56	2	-	6	18
9	DM43 Nativa	6,79	0,02M de Acetato de Sódio - pH5,0	5,0	-	Sulfato de Amônio	50 - 56	2	-	6	18
10	DM43 Deglicos.	5,90	0,02M de HEPES- pH7,4	5,0	-	Sulfato de Amônio	50 - 56	2	-	6	18
11	DM43 Deglicos.	9,34	Solução de Desglicosilação ²	5,0	-	Sulfato de Amônio	50 - 56	2	-	4	18
12	DM43 Deglicos.	2,22	Solução de Desglicosilação	5,0	-	Sulfato de Amônio	50 - 62	4	-	4	18

¹ A concentração de proteína foi calculada baseando-se em medidas de absorbância em espectrofotômetros, e comparada com o valor teórico calculado pelo programa ProtParam (ExpAsy) a partir da sequência de aminoácidos de DM43.

² 0,25M de Fosfato de Sódio, pH 6,0; 0,5M Fosfato de Sódio Dibásico; PNGaseF (seção IV-4).

IV.4.Desglicosilação

A desglicosilação de DM43 e a transferência (*blotting*) das diversas bandas de DM43 desglicosilada (SDS-PAGE) para a membrana de nitrocelulose foram feitas segundo protocolo, reagentes, materiais e enzimas da Bio-Rad Laboratories.

Os reagentes utilizados na desglicosilação estão apresentados na tabela IV-3:

Tabela IV-3: Reagentes e proteínas utilizadas no experimento de desglicosilação enzimática, segundo manual da Bio-Rad Laboratories.

Componentes	Concentrações
PNGase F (Enzima responsável pela clivagem de todos os oligossacarídeos ligados a asparaginas)	2,5U/ml (em 20mM Tris-HCl pH 7,5; 50mM NaCl; 1mM EDTA)
Fetuína Bovina (Glicoproteína utilizada como controle da eficiência de PNGase F e do procedimento)	500µg
5x Tampão de Reação	250mM Fosfato de Sódio pH 6,0
Tampão para Ajuste de pH	500mM Fosfato de Sódio Dibásico

Na preparação da amostra, 100µg de DM43 ou Fetuína Bovina foram dissolvidos em 12µl de água (Milli-Q_{plus}, Millipore), resultando em uma concentração de 8,3 mg/ml.

A seguir, 4µl do 5x Tampão de Reação foi adicionada à amostra seguida de 14µl de água e 10µl de Tampão de ajuste de pH. Por último, 2µl de PNGase F foram misturados na solução e levados para incubar por 24 horas a uma temperatura constante de 37°C.

Após a desglicosilação seguiram-se às etapas de detecção da porção desglicosilada. As amostras incubadas com PNGase F e as nativas foram aplicadas em gel de SDS-PAGE para então transferi-las em membrana de nitrocelulose. Nela foram detectadas, após tingimento, as bandas correspondentes às amostras que tiveram seus carboidratos totalmente clivados pela enzima.

Os equipamentos e o protocolo de preparação dos reagentes e estão detalhados no Anexo-II, bem como o procedimento completo de transferência e revelação da membrana de nitrocelulose.

IV.5.Cristalização da DM43 Desglicosilada

Foram necessários dialisar e concentrar a DM43 desglicosilada para utilização nos experimentos de cristalização. Para isto, utilizou-se uma membrana Microcon 10 Green, cut-off 10KDa da Amico, Inc., acoplada a uma centrífuga (MSE-MicroCentaur (Sanyo)) sob rotação de 10000 rpm à temperatura ambiente.

Os cálculos de concentração de DM43 foram feitos através dos dados de absorção a 280nm medidos em espectrofotômetro (U-2000 Spectrophotometer, Hitachi) e comparados com os obtidos a partir do valor teórico da absorbância de DM43 a 1mg/ml (seção IV.2).

190µl de 2,22mg/ml de DM43 desglicosilada foram concentradas na própria solução de desglicosilação (seção IV-4) para 9,34mg/ml. A diálise foi feita contra 0,02M de HEPES pH 7,4 com a mesma membrana utilizada na concentração.

As condições de cristalização da DM43 desglicosilada estão detalhadas na tabela IV-2, experimentos 10 -12.

IV.6.Cromatografia

O processo de desglicosilação enzimática por PNGase F não foi 100% nas tentativas seguintes de reproduzir o resultado da primeira desglicosilação. A intenção de realizar experimentos de cristalização mesmo com DM43 parcialmente desglicosilada, motivou sua purificação a partir de uma solução contendo moléculas ainda na forma nativa.

As tentativas de purificação da DM43 parcialmente desglicosilada foram feitas à temperatura ambiente no sistema automatizado de cromatografia líquida, composto por ÄKTA explorer 10 e programa de monitoração UNICORN® 2,30 (Pharmacia LKB Biotechnology).

Lista-se a seguir as resinas utilizadas e os seus respectivos parâmetros de purificação:

▪ Troca Aniônica

- Forte (POROS® 20PI - PerSeptive BioSystems)

Tabela IV-4: Parâmetros utilizados na purificação de DM43 parcialmente desglicosilada com a coluna aniônica forte da PerSeptive BioSystems.

Experimento	[DM43] (mg/ml)	Loop (μl)	Fluxo (ml/min)	Gradiente (ml)	Tampão	[NaCl] (M)	pH
1	2,22	25	2,0	25	0,05M Bis-Tris/HCl	1,0	6,4

- Fraca (POROS® 20HQ - PerSeptive BioSystems)

Tabela IV-5: Parâmetros utilizados na purificação de DM43 parcialmente desglicosilada com a coluna aniônica fraca da PerSeptive BioSystems.

Experimento	[DM43] (mg/ml)	Loop (μl)	Fluxo (ml/min)	Gradiente (ml)	Tampão	[NaCl] (M)	pH
2	0,37	100	5,0	25	0,05M Tris/HCl	1,0	8,5
3	0,37	100	5,0	50	0,05M Tris/HCl	0,5	8,5
4	0,37	100	5,0	50	0,05M Piperazina/HCl	0,5	5,0

- Fraca (MONOQ® HR 5/5 – Pharmacia LKB Biotechnology)

Tabela IV-6: Parâmetros utilizados na purificação de DM43 parcialmente desglicosilada com a coluna aniônica fraca da Pharmacia LKB Biotechnology.

Experimento	[DM43] (mg/ml)	Loop (μl)	Fluxo (ml/min)	Gradiente (ml)	Tampão	[NaCl] (M)	pH
5	0,37	100	1,0	30	0,02M Bis-Tris/HCl	1,0	6,4

▪ Interação Hidrofóbica

- (Phenyl Sepharose® CL-4B - Pharmacia LKB Biotechnology)

As colunas foram equilibradas com tampão contendo sal e a eluição com gradiente decrescente de sal.

Tabela IV-7: Parâmetros utilizados na purificação de DM43 parcialmente desglicosilada com a resina de interação hidrofóbica da Pharmacia LKB Biotechnology.

Experimento	[DM43] (mg/ml)	Loop (μl)	Fluxo (ml/min)	Gradiente (ml)	Tampão	[Sulfato de Amônio] (M)	pH
6	1,78	25	0,5	25	0,05M Fosfato de Sódio	1,7	7,0
7	0,48	50	0,5	30	0,05M Acetato de Sódio/Ácido Acético	2,2	4,5

O monitoramento do processo de purificação foi essencialmente através da absorção da proteína a 280nm. As frações correspondentes aos picos de absorção foram dialisadas contra água Milli-Q_{plus} (Millipore) em membranas de MWCO 6-8000 (Spectrum Medical Industries, Inc.), liofilizadas (SpeedVac-SC200, Savant) e analisadas em géis de poliacrilamida.

Os experimentos de eletroforese foram realizados segundo o método de Laemmli (1970) que faz uso de SDS (Sodium Dodecyl Sulphate) para desnaturar proteínas e separá-las segundo suas massas moleculares. As amostras de proteína foram aplicadas em géis constituídos por 15% de poliacrilamida na presença de β -mercaptoetanol, e migrados unidimensionalmente sob tensão constante de 120V (Pharmacia LKB Biotechnology; Gibco-BRL Life Technologies).

A identificação da proteína segundo sua massa molecular no gel de poliacrilamida é feita por comparação à migração dos padrões de massa molecular conhecidos, cujo logaritmo é diretamente proporcional aos valores de suas massas. Os seguintes padrões foram utilizados: Albumina de Soro Bovino (66,0 KDa) (Riedel-deHaën); Ovo Albumina (45,0 KDa) (Sigma); Anidrase Carbônica Bovina (29,0 KDa) (Sigma); Inibidor de Tripsina de Soja (20,1 KDa) (Sigma); e Citocromo C (12,4 KDa) (Sigma).

Os géis foram corados com *Coomassie blue* e prata (Morrissey, 1981; Merril, 1981).

IV.7. Criocristalografia

Os cristais de proteína costumam ser muito frágeis aos efeitos da radiação. Grande parte se deve ao alto conteúdo de solvente contido no cristal, que facilita a difusão e proliferação dos radicais livres, também produzidos pela radiação.

O resfriamento, entretanto, pode induzir a formação de cristais de moléculas de água que interferem no padrão de difração do cristal de proteína. Além disto, é possível que a rede cristalina sofra deformações devido à expansão das moléculas de água durante o congelamento. De forma a superar estes problemas, existem soluções crioprotetoras que inibem a formação de cristais de gelo.

A técnica de congelamento de cristais para difração de raios-X, ou a criocristalografia, possibilitou coletas de cristais de proteína suspensas em *loops*, onde uma fina película de solução criogênica sustenta o cristal apenas por tensão superficial. Desta forma, não há necessidade de inserí-lo em capilar onde o feixe de raios-X pode ter sua intensidade atenuada pelas paredes do capilar antes de atingir o cristal, e após a difração.

Soluções contendo diferentes concentrações de etilenoglicol e glicerol foram acrescentadas na solução de cristalização procurando manter as condições originais.

Para os testes, foi utilizado um sistema composto por uma unidade criogênica de nitrogênio líquido e outra de ar seco, ambas da Oxford Cryosystems. *Loops* de diâmetros entre 0,1 a 0,5 mm foram testados para cada solução de etilenoglicol e glicerol submetendo-as a um choque térmico sob um fluxo de nitrogênio gasoso de 100K. A verificação definitiva da propriedade criogênica da solução foi através da coleta de uma imagem *still* de 20 min. no sistema R-AXIS IIC - Rigaku.

IV.8. Coleta de Dados

Nas coletas de dados de difração foram utilizadas fontes de raios-X do tipo anodo rotatório (RU-200B, Rigaku) e Luz Síncrotron (LNLS-Campinas/SP) (Polikarpov *et al.*, 1998a; Polikarpov *et al.*, 1998b), e detectores de área R-AXIS IIC (Rigaku) e Mar345 Imaging Plate Detector System (MAR Research Detectors), respectivamente. Os comprimentos de onda utilizados foram 1,5418Å, relativo a média ponderada dos espectros de emissão característicos ($K_{\alpha 1}$ e $K_{\alpha 2}$) do Cobre (Drenth, 1995) e 1,387Å na Estação PCr (Cristalografia de Proteínas) do LNLS. Em ambos os casos foram utilizados o método de oscilação para coletas.

IV.9. Análise dos Padrões de Difração

Os padrões de difração foram analisados através do pacote HKL (Otwinowski, Minor, 1997) instalados em estações Silicon Graphics, que inclui: programas para

visualização (XdisplayF); indexação (Denzo); e análise estatística completa do conjunto dados de difração (Scalepack).

V.

RESULTADOS E DISCUSSÕES

Nesta primeira parte do trabalho, pretendeu-se otimizar as condições de cristalização anteriormente obtidas por Oliveira *et al.* (1993) através da utilização de métodos de busca por novas condições de cristalização, (Fatorial I da Hampton Research (Anexo-I)) coletas a temperaturas criogênicas e uso de radiação síncrotron, procedimentos ainda não disponíveis na época.

Partindo-se do conhecimento da característica glicoproteica de DM43 (Neves-Ferreira, 1999), a participação dos 29% dos carboidratos na massa total de DM43 não poderia ser desconsiderada nas análises dos experimentos de cristalização. Pouco se sabe, atualmente, sobre sua estrutura e composição, mas acredita-se na existência de quatro ligações covalentes na proteína, do tipo N-glicosídicas, além da possibilidade de existir grupos glicídicos de ácidos siálicos nos seus terminais.

Fazendo cálculos aproximados, há cerca de 12 KDa de carboidratos distribuídos entre os quatro sítios de ligação. Isto implica em uma extensa massa de carboidratos em contato com o solvente, proporcionando uma grande flexibilidade estrutural na superfície de DM43. Estas microeterogeneidades químicas e estruturais existentes em glicoproteínas (ver seção III.4) devem, às vezes, serem totalmente eliminadas de maneira a garantir uma estrutura menos

maleável e uma distribuição mais uniforme de cargas. Cria-se, com isto, a possibilidade de um aumento no nível de organização molecular do cristal, proporcionando uma coleta de dados de difração até a mais alta resolução.

Isto proporcionou ao presente trabalho, uma nova luz para um caminho que possivelmente levará a cristais com qualidade e estabilidade suficientes para a determinação inédita da estrutura tridimensional de DM43, através de novas coletas de dados por difração de raios-X e a subsequente análise cristalográfica utilizando, por exemplo, o método de substituição isomórfica múltipla.

V.1.Cristalização de DM43 e Coletas de Dados

O processo de nucleação e crescimento de DM43 é de certa forma rápida e possível a qualquer pH, dentro de um intervalo de 4,0 e 8,0. Observa-se uma tendência de DM43 em cristalizar na presença de sulfato de amônio ou acetato de amônio.

Os maiores e mais resistentes cristais foram obtidos em um tempo maior, de aproximadamente oito meses à 18°C e em soluções contendo apenas sulfato de amônio como precipitante, e tampão.

Figura V-1: Cristais de DM43 crescidos (após sete meses) à 18°C e em solução contendo 48% S.A. – pH 5,0. Dimensões médias (0,35×0,063×0,063)mm³.

A grande maioria dos cristais apresentou uma morfologia de prisma hexagonal contendo um eixo de ordem seis ao longo do comprimento do cristal, e afinados nas pontas. A mesma morfologia foi também observada no trabalho anterior (figura V-2).

(a)

(b)

Figura V-2: Cristais antigos de DM43 em 54% S.A. pH 5,0 (a) e 54% S.A. pH 7,0 (b). Gotas de 10 μ l, montadas em 31/01/1993. Dimensões aproximadas do maior cristal em (a) $(0,61 \times 0,06 \times 0,06) \text{ mm}^3$, e $(1,50 \times 0,34 \times 0,34) \text{ mm}^3$ em (b) (Oliveira *et al.*, 1993).

Diferentes hábitos cristalinos foram obtidos à 4°C (figura V-3) e à 18°C em soluções contendo PEG 4000 e 8000 como precipitantes e acetato ou sulfato de amônio como sais (figura V-4). Todos estes cristais são fisicamente frágeis a variações de temperatura e até mesmo a leves choques mecânicos. Por este motivo, não foi possível coletar dados de difração dos cristais obtidos em sulfato de amônio a 4°C e em PEG 4000, ilustrados nas figuras V-3 e V-4(b), respectivamente. Os cristais crescidos em PEG 8000 (figura V-4(d)) também não apresentaram sinais de reflexões.

Figura V-3: Cristais de DM43 crescidos à 4°C em gota de 4 μ l contendo 56% S.A. – pH 5,0. Dimensões aproximadas: $(0,11 \times 0,02 \times 0,02) \text{ mm}^3$

(a) Fatorial I nº 9, 18°C.

(b) Gota de 4µl; 0,1M Citrato de Sódio pH 5,6; 0,2 M Acetato de Amônio; 30%w/v PEG4000; 18°C. Dimensões médias dos cristais: (0,11×0,03×0,03)mm³.

(c) Fatorial I nº 15, 18°C.

(d) Gota de 4µl; 0,1 M MES pH 6,5; 0,2M Sulfato de Amônio; 30% w/v PEG 8000. Dimensões médias dos cristais (0,37×0,06×0,02)mm³.

(e) Esferulitas

Figura V-4: Diferentes hábitos cristalinos apresentados por DM43. Cristais (b) e (d) crescidos na presença de PEG 4000 e 8000, respectivamente, de morfologias diferentes das obtidas com sulfato de amônio (fig.1). Resultados estes de um *screening* a partir dos indícios de uma possível formação de cristais ao redor dos Fatoriais I nº 9 (a) e 15 (c), respectivamente a (b) e (d). A presença de esferulitas em soluções contendo PEG é ilustrada na figura (e).

As fotos que se seguiram apresentaram as gotas referentes as condições 9 (figura V-4(a)) e 15 (figura V-4(c)) do Fatorial I. A formação de precipitados quase cristalinos deu direção à montagem de um *screening* destas condições, detalhados na tabela IV-1(c) e (d) do capítulo IV. Todas as gotas do *screening* resultaram em precipitados do tipo esferulita e fibra (figura V-4-(e)), exceto em duas gotas que curiosamente têm as mesmas condições de cristalização do fatorial soluções números 9 e 15 (figuras V-4-(b) e (d), respectivamente). Provavelmente, diferenças na origem dos reagentes e o modo de preparo dos tampões podem ter influenciado na nucleação e crescimento destes cristais.

Uma coleta preliminar de um cristal da gota de 50% S.A.-pH5.0 - 18°C (figura V-5(a)) no R-AXIS IIC produziu um espalhamento a uma resolução bem maior ($\sim 3,1\text{\AA}$) do obtido até então, criando a possibilidade de ainda melhorar estes dados com uma coleta completa no Laboratório Nacional de Luz Síncrotron (LNLS-Campinas/SP) (Polikarpov *et al.*, 1998a; Polikarpov *et al.*, 1998b)

Figura V-5: (a) Cristais crescidos após 8 meses em gota de $6\mu\text{l}$ - 50% S.A. - pH5,0, 18°C. Dimensões aproximadas ($0,42 \times 0,09 \times 0,09$) mm³. (b) Detalhe de três reflexões de uma imagem de difração por radiação síncrotron (LNLS). Padrão obtido de um cristal da foto ao lado congelado a 85K em solução criogênica contendo 21% de Glicerol 87%.

De fato, uma coleta de dados bem sucedida foi possível com a utilização de raios-X gerados em fonte de luz síncrotron que, devido à alta intensidade de seu feixe, permitiu um espalhamento mais intenso e conseqüentemente, resoluções mais altas. Outra grande vantagem desta alta intensidade do feixe é a diminuição considerável, por um fator de ~ 4 , do tempo de coleta de uma imagem se comparada a fontes convencionais. Isto aumenta muito as chances de obter um conjunto completo de dados de difração antes mesmo do cristal sofrer os danos decorrentes da radiação.

As medidas realizadas no LNLS estão listadas na tabela V-1:

Tabela V-1: Descrição das coletas de difração na estação PCr (LNLS), de alguns cristais de DM43 crescidos em Sulfato de Amônio.

Cristais		Coletas			
		Temperatura	Solução crio contendo	Fonte	Comentários
50% S.A. pH 5,0 18°C	Dimensões aproximadas dos cristais (0,42×0,09×0,09) mm ³	277K (4°C)	-	Síncrotron	O cristal suportou a radiação para a coleta de apenas uma imagem.
		85K (-188°C)	16% Etilenoglicol	Síncrotron	Algumas reflexões à baixa resolução
		85K (-188°C)	21% Glicerol 87%	Síncrotron	Coletado um conjunto de 75 imagens.
50% S.A. pH 5,0 18°C	Dimensões aproximadas dos cristais (0,47×0,09×0,09) mm ³	85K (-188°C)	13% Glicerol 87%	Síncrotron	Coletado um conjunto de 69 imagens.
48% S.A. pH 5,0 18°C	Dimensões aproximadas dos cristais (0,35×0,063×0,063) mm ³	85K (-188°C)	21% Glicerol 87%	Síncrotron	Reflexões fracas à baixa resolução.
48% S.A. pH 5,0 18°C		85K (-188°C)	20% Etilenoglicol	Síncrotron	Não difratou
48% S.A. pH 4,0 18°C	Dimensões aproximadas dos cristais (0,47×0,08×0,08) mm ³	85K (-188°C)	20% Etilenoglicol	Síncrotron	Poucas reflexões à baixa resolução.

A visualização das imagens de difração do cristal de 50%S.A.-pH5,0 (21% de glicerol 87%) pelo programa XdisplayF indicou um espalhamento preferencial na

direção x paralelo ao eixo de oscilação obtendo, em algumas imagens, a resolução máxima de 3,1Å, e 7,5Å na direção y. Um padrão anisotrópico semelhante foi observado em imagens espalhadas por um cristal de 50%S.A.-pH5,0 em uma solução crio contendo 13% de etilenoglicol. Interessantemente, o cristal estava com a orientação de seu “eixo maior” ao longo do eixo de oscilação (eixo x).

A observação mais cuidadosa das imagens revelou o aparecimento de reflexões duplas de intensidades muito próximas (figura V-5(b)) em intervalos periódicos de ϕ (ângulo de oscilação do cristal durante a coleta) mantendo sempre um comportamento linear crescente da duplicidade, paralelo à direção do eixo de rotação. Apesar disto, foi feita uma tentativa de processar todas as 75 imagens coletadas, considerando uma reflexão dupla como sendo uma única.

A indexação destas imagens no programa Denzo ocorreu dentro dos padrões ótimos de processamento mantendo o parâmetro estatístico de indexação χ^2 , sempre em torno do valor ideal 1,00. Entretanto, a avaliação estatística do programa Scalepack após o pós-refinamento, *merging* e escalonamento de todo o conjunto de dados, foi inconclusiva a uma resolução acima de 4,0Å. Um fator R de aproximadamente 8,5% foi obtido com a inclusão de reflexões pertencentes a região de até 7,5Å de resolução.

Porém, estas estatísticas não são muito confiáveis, pois resultam de cálculos a partir de reflexões fracas e com perfis duplos, não possibilitando uma conseqüente determinação do grupo espacial do cristal. Estes dados não permitiram nem ao menos diferenciar entre o sistema cristalino trigonal de um hexagonal. Apesar do erro associado a posição exata da reflexão devido ao seu caráter duplo, foi possível identificar uma cela para este cristal (figura V-5(a)), a partir da indexação da melhor imagem dos 75 coletados. Desta forma, foram calculados os seguintes parâmetros de cela e com os respectivos erros: $a = b = 117,34 (0,03) \text{ Å}$, $c = 193,39 (0,02) \text{ Å}$; $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$.

Dados anteriormente coletados à temperatura ambiente por Oliveira, P.S.L. (não publicado) no Síncrotron, resultaram em: $a = b = 120,73 (0,05) \text{ Å}$, $c = 189,73 (0,07) \text{ Å}$, $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$, também determinados a partir da escolha da melhor imagem após o processamento no programa Denzo. Observa-se neste, um aumento de 3,9% no volume da cela unitária. Isto é compreensível, uma vez que o ambiente vítreo das moléculas no interior do cristal congelado, não permite uma

movilidade atômica tão grande como em temperatura ambiente, o que leva normalmente a uma redução na cela unitária.

Curiosamente, os seus padrões de difração também revelaram um perfil duplo de intensidade na reflexão, de modo que não foi possível determinar o seu grupo espacial.

Fazendo uma análise dos fatos descritos até agora, pode-se concluir que:

- a obtenção de reflexões de baixa qualidade por cristais de DM43, não está relacionada com solução crio. Cristais crescidos na mesma condição foram coletados utilizando dois diferentes crioprotetores (glicerol e etilenoglicol) (tabela V-1), ambos apresentando efeitos semelhantes. Somando-se a isto o fato de que os padrões coletados à temperatura ambiente e, portanto, sem crioprotetores resultaram igualmente em um espalhamento à média e à baixa resolução, inclusive com as reflexões de intensidades duplicadas;
- as reflexões duplas (figura V-5(b)) não podem ser decorrentes de uma mosaicidade alta, pois do contrário, os máximos de intensidades teriam uma distribuição mais larga e espalhada;
- como a formação das reflexões duplas não foi gradual no decorrer da coleta de dados, é improvável que a radiação tenha sido responsável pela deformidade do cristal de DM43. Sob condições criogênicas, os cristais suportavam bem a ação dos raios-X, mantendo um padrão de reflexões constante mesmo após horas de coleta;
- descarta-se a possibilidade de *twinning* pois este consiste de dois ou mais cristais individuais de mesma espécie relacionados por uma operação de simetria. Esta deformidade não produziria um espalhamento de reflexões duplas, mas de duas redes recíprocas.
- pode ser provável que as reflexões duplas decorrem de cristais que são facilmente trincados ao longo de uma determinada direção.

Como ainda não há possibilidade de se fazer um estudo a nível microscópico do crescimento do cristal (seção III.3.1) as conclusões acima são apenas hipóteses firmadas nos resultados da difração dos cristais de DM43.

Entretanto, o alto conteúdo de carboidratos é uma razão forte de que devido a sua estrutura altamente flexível, compromete a qualidade dos cristais.

Foi proposto, então, a remoção enzimática das cadeias de oligossacarídeos como um primeiro passo na tentativa de superar o problema de cristalização da DM43.

V.2.Desglicosilação

A desglicosilação enzimática foi solução de problemas de cristalizações mal sucedidas de outras glicoproteínas, como por exemplo, a apo-lactoferrina humana (Norris *et al.*, 1989) que após sua desglicosilação, teve a resolução de seus dados de difração melhorada de 4,0Å para 2,8Å.

Ao contrário da desglicosilação química, o processo de desglicosilação enzimática tem maior probabilidade de preservar, até certo ponto, o enovelamento original da glicoproteína. Alternativamente, pode-se retirar somente os ácidos siálicos localizados nas extremidades das cadeias oligossacarídicas através de exoglicosidases (as neuraminidases), uniformizando a distribuição total de cargas na superfície da proteína. Esta seria uma boa alternativa para otimizar as condições de cristalização de proteínas contendo um baixo grau de glicosilação. No caso como o de DM43, o conteúdo glicídico chega a corresponder 29% de sua massa molecular total, e a simples retirada de ácidos siálicos pode não proporcionar um empacotamento adequado para a formação de cristais. Uma vez que eles tendem a formar extensas ramificações, podem acabar gerando uma flexibilidade extra, desfavorável à estabilização de uma estrutura sólida e ordenada.

Baseando-se nestas idéias, optou-se pela desglicosilação total de DM43 com a utilização de PNGase F, uma enzima do tipo endoglicosidase responsável pela remoção total da cadeia de carboidratos ligados em asparaginas (*N-linked*) (seção III.4).

A desglicosilação de DM43 resultou na produção de duas bandas de intensidades significativas. A figura V-6 ilustra o gel de SDS-PAGE e a membrana de nitrocelulose com as indicações das bandas de DM43 denominadas a partir de agora como parcial (seta tracejada) e totalmente desglicosilada (seta cheia). As

De acordo com o resultado obtido na membrana de nitrocelulose e em ordem decrescente de massa molecular, as três bandas mais intensas das raias 3 e 4 correspondem a: fração de DM43 não desglicosilada (1a. banda); fração de DM43 parcialmente desglicosilada (2a. banda); e fração de DM43 aparentemente, totalmente desglicosilada (3a. banda). Supõe-se que as bandas menos intensas correspondem a estágios intermediários de desglicosilação. Esta classificação foi feita segundo a perda da massa molecular e detecção de suas bandas correspondentes na membrana de nitrocelulose.

Figura V-6: Gel de poliacrilamida (15% SDS-PAGE) (a) e membrana de nitrocelulose (b) indicando a desglicosilação completa de DM43 por processos enzimáticos.

Cálculos de distâncias de migração segundo massa molecular apresentaram um valor de 47,7 KDa para a banda nativa, 43,5 KDa para a banda parcialmente desglicosilada e 37,1 KDa para a banda de DM43 suposta totalmente desglicosilada. No gel de desglicosilação química (figura V-7), a banda de DM43

Figura V-7: Géis de poliacrilamida (12% SDS-PAGE) indicando a total desglicosilação de DM43 após tratamento químico com TFMS (a) e ácido periódico Schiff para coloração de glicoproteínas (b). Colunas I e A indicando a banda de DM43 antes da desglicosilação; II e B, DM43 após desglicosilação; III, marcadores de peso molecular. (Neves-Ferreira *et al.*, 2000)

¹ Neves-Ferreira *et al.* (2000) em seu artigo cita a presença de duas bandas muito próximas, de DM43, após desglicosilação enzimática. Aleatoriamente, foi escolhida a banda que corresponde a perda de 26,7% de glicosilação, ou seja, 34,7KDa.

Figura V-8: Gel de SDS-PAGE 15% revelando a desglicosilação parcial de DM43. Faixas 1 e 6, padrões de massa molecular; 2, DM43 nativa; 3, DM43 após desglicosilação; 4, DM43 após segunda desglicosilação de 24 hs de incubação; 5, DM43 após segunda desglicosilação de 4 dias de incubação.

As subseqüentes tentativas de reproduzir a desglicosilação enzimática de DM43 em larga escala resultaram na retirada parcial dos carboidratos. Isto foi evidenciado pelo aparecimento da banda parcialmente desglicosilada com uma intensidade muito maior que a da DM43 “totalmente desglicosilada” (fig. V-8). Note que apenas parte da DM43 nativa foi desglicosilada, podendo ser uma indicativa de que a PNGase F não tem atuado com a mesma eficiência da primeira tentativa, talvez devido ao seu envelhecimento.

Esta hipótese foi de fato reforçada com a obtenção deste mesmo resultado para quais fossem as modificações empregadas nos parâmetros de desglicosilação: aumento do período de incubação de 24hs, recomendado pelo fabricante, para até 14 dias; indução uma agitação constante de 150, 250, 300 e 340 rpm durante a desglicosilação; e ainda, a realização de uma segunda desglicosilação (fig. V-8).

Não sendo possível concluir a total desglicosilação de DM43, levantou-se a possibilidade de trabalhar com a sua porção parcialmente desglicosilada (8,8%). A ligeira mudança estrutural entre esta e a espécie nativa de DM43 poderia ser suficiente para melhorar a ordem do empacotamento do cristal ou induzir a formação de uma nova forma cristalina. Para tanto, tem-se a purificação da DM43 desglicosilada como passo seguinte a desglicosilação.

V.3. Purificação

Supondo a existência de ácidos siálicos nas extremidades dos grupos oligoholosídicos e a conversão das asparaginas em ácidos aspárticos com a digestão dos carboidratos, é possível que DM43 após a desglicosilação esteja com sua superfície alterada não apenas geometricamente mas também, eletrostaticamente. Uma vez que as distâncias de migração das duas espécies de DM43 em SDS-PAGE são muito semelhantes (figura V-8), uma separação segundo sua massa molecular não seria possível. Felizmente, a desglicosilação é de certa forma responsável pelas mudanças na distribuição de cargas da superfície de DM43, e também da exposição de novas regiões apolares ampliando, com isto, as possibilidades da exploração cromatográfica tanto em termos de interações iônicas, quanto hidrofóbicas.

Várias foram as tentativas de separar a DM43 parcialmente desglicosilada da DM43 nativa (seção IV.6). A mais explorada foi a cromatografia por interação iônica e a tentativa inicial foi utilizar uma coluna de interação aniônica forte com um pH distante do pI da proteína (cujo valor está abaixo de 3,5). Teoricamente, a superfície de DM43 se tornaria um pouco mais negativamente carregada, promovendo uma interação iônica mais intensa com a matriz positivamente carregada.

Como não houve separação, foi necessária, nas tentativas seguintes, a alteração de vários parâmetros como resinas, pHs, fluxos, comprimento do gradiente e concentrações de sal para eluição. A purificação com uma resina de troca aniônica fraca e um gradiente de 50ml em um intervalo de 0-0,5M de NaCl (experimento 3, seção IV.6), resultou na eluição de um único pico (figura V-10) cujas frações apresentaram o perfil ilustrado na figura V-9.

Verifica-se uma separação bem sutil nas laterais do pico, eluindo primeiro a DM43 nativa (primeira banda), seguido de várias frações da mistura das duas espécies de DM43 e finalizando com a eluição de pouquíssima quantidade da DM43 parcialmente desglicosilada (segunda banda). Evidentemente a resolução foi insuficiente para serem observados dois picos na cromatografia e tampouco para ser útil na cristalização.

Figura V-9: Géis de SDS-PAGE 15% corado com prata, apresentando um padrão de eluição das DM43 nativa (seta tracejada) e parcialmente desglicosilada (seta cheia). As raias denominadas P_x apresentam as bandas dos padrões de massa molecular, as demais são frações colhidas durante a eluição da cromatografia.

Figura V-10: Detalhe do perfil cromatográfico de DM43 em coluna de interação aniônica fraca POROS 20HQ. Pico de DM43 eluído a 47% do gradiente de 0 a 0,5M de NaCl em pH 8,5.

Difícilmente seria atribuída uma interpretação precisa deste resultado, sem o conhecimento prévio da composição dos carboidratos e dos tipos de resíduos que estão expostos na superfície de DM43 antes e após a desglicosilação. Porém, se a banda correspondente a DM43 parcialmente desglicosilada desprendesse em condições de uma força iônica maior que a necessária para a eluição da banda de DM43 nativa, é de se supor que a superfície da proteína se tornou mais

Até o presente momento todas as gotas montadas segundo experimento 10-12 da tabela IV-3, precipitaram sem nenhuma formação posterior de um único cristal. Porém, é sabido que os “melhores” ou maiores cristais de DM43 surgiram meses após a preparação. É possível que ainda seja possível observar a formação de cristais de DM43 a partir do precipitado da mistura de DM43 não desglicosilada e parcialmente desglicosilada.

V.5.Perspectivas

Possivelmente, o resultado negativo da obtenção de cristais de qualidade satisfatória para a determinação de sua estrutura, se deve à característica glicoproteica de DM43. Além da desglicosilação enzimática, há outras alternativas de superar este problema, como a remoção dos ácidos siálicos ou expressão de DM43 recombinante em procariotos, agora possível graças ao recente trabalho (Neves-Ferreira, 1999) de determinação da sua estrutura primária completa.

Existe, porém, a possibilidade dos carboidratos estarem de alguma forma influenciando na estrutura da proteína em si, ou através da participação na estabilidade do seu enovelamento, ou ainda contribuindo na conformação de alguma estrutura secundária. A resistência de DM43 à ação proteolítica do veneno da jararaca pode ser um indício de que os 29% em massa de carboidratos (MALDI-TOF) estejam criando uma capa protetora envolvendo grande parte da superfície, proporcionando a estabilidade conformacional e química da estrutura de DM43, mesmo na presença de proteases.

A cristalização do complexo jararagina-DM43 seria uma proposta interessante, uma vez que a superfície do complexo poderia favorecer a formação de cristais de melhor ordem interna diferente do de DM43, e talvez nem seja necessário insistir na sua desglicosilação. Isto proporcionaria a determinação de duas estruturas cristalográficas inéditas, permitindo examinar a conformação nativa e ativa de DM43. Este resultado esclareceria o mecanismo de ação de DM43 e também a verificação de sua conformação monomérica ou dimérica na complexação com a jararagina.

Na tentativa de prever a estrutura tridimensional de DM43, propõe-se na parte B deste trabalho, a modelagem molecular por homologia de DM43 a partir de sua sequência de aminoácidos.

Parte B

VI.

MODELAGEM MOLECULAR
ESTRUTURAL

A obtenção de estruturas tridimensionais de proteínas é feita segundo métodos experimentais ou teóricos. A Cristalografia de Raios-X e a Ressonância Magnética Nuclear correspondem aos dois principais métodos experimentais. Entretanto, ainda possuem algumas limitações técnicas que dificultam a determinação de estruturas na mesma velocidade em que são obtidas as seqüências de aminoácidos. Este problema é, de certa forma, superado através de métodos teóricos que predizem a estrutura tridimensional de uma proteína a partir de sua seqüência de aminoácidos.

Figura VI-1: Número de seqüências e estruturas 3D de proteínas obtidas entre os anos 1986 e 1994 (Sternberg, 1996). (a) Número de seqüências registradas no SWISS-PROT (Bairoch e Apweiler, 2000) e dentre elas, as seqüências cuja homologia é menor que 30%. (b) Número de estruturas 3D experimentalmente determinadas e registradas no *Brookhaven Protein Data Bank* (PDB) (Berman *et al*, 2000), número de seqüências cuja homologia é menor que 35%, e número de estruturas 3D não homólogas.

Estatísticas publicadas em 1994 por Sternberg (1996) (figuras VI-1), revelaram que o número de deposições de seqüências é 10 vezes maior que o de estruturas¹ resolvidas experimentalmente. Após cinco anos, esta proporção passou para 85.785 seqüências² contra 10.743 estruturas de proteínas, peptídeos e vírus³, diferença ainda significativa se comparada com os dados anteriores (figuras VI-2 e VI-3).

Figura VI-2: Estatística anual do número total de seqüências registrado na base de dados de SWISS-PROT (Bairoch e Apweiler, 2000). Informações obtidas em <http://expasy.cbr.nrc.ca/cgi-bin/lists?relnotes.txt>.

Figura VI-3: Estatística recente do número de estruturas depositadas no Brookhaven Protein Data Bank (PDB) (<http://www.rcsb.org/pdb/holdings.html>) (Berman *et al*, 2000). As colunas em vermelho indicam o número de registros por ano, e em verde o número total de estruturas até então depositadas. Incluem-se estruturas de proteínas, peptídeos, vírus, complexos proteína-ácidos nucleicos, ácidos nucleicos e carboidratos.

¹ Quando não especificado, *estrutura* quer significar *estrutura tridimensional*.
² Dado de 20/05/2000 obtido em SWISS-PROT <http://expasy.cbr.nrc.ca/sprot/>.
³ Dado de 23/05/2000 obtido em PDB <http://www.rcsb.org/pdb/holdings.html>.

Os métodos teóricos de obtenção de estruturas 3D de proteínas poderiam amenizar esta discrepância uma vez que para a sua aplicação é necessário, *a priori*, apenas a informação da sequência de aminoácidos. Ressalta-se, porém, que a modelagem molecular é utilizada nem sempre com a intenção de substituir os métodos experimentais existentes, mas de antecipadamente à determinação de sua estrutura, fornecer informações úteis que podem ser utilizadas para explicar as observações experimentais anteriores a respeito do funcionamento da molécula. Com isto, é possível planejar novos experimentos de modo mais direcionado, e ainda tomá-lo como ponto de partida para a determinação experimental de sua estrutura por Difração de Raios-X ou NMR.

As técnicas de predição de estrutura tridimensional têm sido facilitadas não somente pelo surgimento de computadores com capacidade de processamento cada vez maior, mas também devido ao aperfeiçoamento dos algoritmos de modelagem que estão cada vez mais automatizados, proporcionando a obtenção de estruturas ainda mais confiáveis em um curto espaço de tempo.

Dentre os métodos teóricos de obtenção de estrutura 3D listam-se os métodos físicos e os métodos empíricos. Os métodos físicos baseiam-se nas interações entre átomos utilizando-se as teorias de dinâmica molecular e minimização de energia, enquanto que os métodos empíricos dependem de informações de estruturas já experimentalmente determinadas. Neste último incluem as modelagens combinatoriais e comparativas (modelagem por homologia).

A modelagem comparativa apóia-se no fato de que uma pequena variação na sequência de aminoácidos resulta geralmente em uma pequena mudança na estrutura tridimensional. Portanto, é possível predizer a estrutura de uma proteína a partir da similaridade seqüencial desta, com as de outras de proteínas cujas estruturas já foram experimentalmente determinadas. O método teórico por comparação tem conseguido predizer estruturas com precisão maior que 1Å de RMSD nos casos onde o grau de identidade é maior que 40% (Šali *et al.*, 1995), enquanto que os métodos físicos e combinatórios carregam erros de pelo menos 3.5Å (Cohen e Kuntz, 1989; Wilson e Doniach, 1989; *apud* Šali e Blundell, 1993¹).

¹ COHEN, F. E.; KUNTZ, I. D. Tertiary structure prediction in "*Prediction of protein structure and the principles of protein conformation*", Fasman, G. D., ed., Plenum Press, p. 647-705, 1989.

WILSON, D.; DONIACH, S. A computer model to dynamically simulate protein folding: Studies with Crambin. *Proteins*, v. 6, p. 193-209, 1989.

Entretanto, os métodos experimentais superam em resolução podendo obter estruturas com RMSD de até 0.3Å por Cristalografia de Raios-X, e 0.5Å por NMR, dependendo da resolução dos dados e do número de NOE's, respectivamente (Clore e Gronenborn, (1991) *apud* Šali e Blundell (1993)¹).

Na tabela VI-1 estão listados os vários métodos de predição de estruturas 3D de proteínas por homologia.

A qualidade das estruturas preditas pelos métodos citados na tabela VI-1, depende fortemente do grau de similaridade e identidade entre as seqüências da proteína alvo e das proteínas de estruturas 3D conhecidas. Algoritmos de programas para modelagem por comparação estão sendo aperfeiçoados nesta direção, de maneira que resulte em modelos precisos de estruturas de proteínas de baixo grau de similaridade com estruturas experimentalmente determinadas, num processo completamente automatizado.

Andrej Šali foi o principal desenvolvedor de MODELLER, um programa automatizado que calcula modelos de proteínas impondo restrições espaciais obtidas de estruturas homólogas. Ele tem sido avaliado com sucesso em todos os encontros do *Critical Assessment of Techniques for Protein Structure Prediction* (CASP), que ocorre desde 1994 com o objetivo de avaliar métodos de predição de estruturas 3D de proteínas.

A praticidade e rapidez do programa foram conferidas em CASP2-1996 (Sánchez e Šali, 1997) com a produção de vários modelos a partir de alinhamentos e realinhamentos iterativamente modificáveis pelo usuário, melhorando a qualidade dos resultados obtidos em CASP1 -1994 (Šali *et al.*, 1995). Resultados de CASP3-1998 (Burke *et al.*, 1999) verificaram a utilidade de MODELLER mesmo em proteínas que possuem um grau de identidade muito baixo com relação às estruturas 3D conhecidas.

¹ CLORE, G. M.; GRONENBORN, A. M. Two-, three-, and four-dimensional NMR methods for obtaining larger and more precise three-dimensional structures of protein in solution. *Annu. Rev. Biophys. Chem.*, v. 20, p. 29-63, 1991.

Tabela VI-1: Descrição de métodos de modelagem por homologia existentes para a obtenção teórica de estruturas 3D de proteína.

Métodos de modelagem molecular por homologia, segundo Šali e Blundell, 1993, e Sternberg, 1996		
Método	Comentários	Referências ¹
Comparação Manual	Substituição das cadeias laterais do arquivo de coordenadas da estrutura correspondente, pelas da sequência da proteína que se deseja modelar.	Browne <i>et al.</i> , 1969 Warne <i>et al.</i> , 1974
Corpo Rígido	Baseado na construção de um modelo a partir da combinação de regiões de estruturas sequencialmente relacionadas, como <i>loops</i> , cadeia lateral, e <i>core</i> .	Jones e Thirup, 1986 Unger <i>et al.</i> , 1989
Sobreposição de Segmentos	Construção de modelos por combinação de segmentos de estruturas utilizando o critério de energia	Levitt, 1992
Satisfação das Restrições Espaciais	Aplicação das restrições espaciais na sequência a que se deseja prever sua estrutura. Estas restrições são obtidas a partir de um alinhamento entre a sequência alvo e as sequências de proteínas similares a ela.	Šali <i>et al.</i> , 1990
Geometria de Distâncias	Diferentemente das restrições espaciais, este método baseia-se nas restrições espaciais que são caracterizadas por possuírem delimitações rígidas das distâncias entre os átomos.	Havel e Snow, 1991
Posição dos átomos C _α	Gera as coordenadas cartesianas da cadeia principal da proteína alvo utilizando apenas as coordenadas dos átomos C _α da estrutura sequencialmente análoga a ela.	Correa, 1990
Redes Neurais	Utilização de redes neurais para calcular modelos a partir de um mapa de distâncias C _α de uma estrutura correlata.	Bohr <i>et al.</i> , 1990
Otimização da Função Potencial	Busca de um mínimo global de um tipo de função de energia potencial da molécula obtida através de parâmetros retirados da estrutura de uma proteína similar conhecida.	Snow, 1993

¹ BOHR, H. *et al.* A novel-approach to prediction of the 3-dimensional structures of protein backbones by neural networks. *FEBS Lett.*, v. 261, n. 1, p. 43-46, 1990.

BROWNE, W. J. *et al.* A possible three-dimensional structure of bovine α -lactalbumin based on that of hen's egg-white lysozyme. *J. Mol. Biol.*, v. 42, p. 65-86, 1969.

CORREA, P. E. The building of protein structures from alpha-carbon coordinates. *Proteins*, v. 7, n. 4, p. 366-377, 1990.

HAVEL, T. F.; SNOW, M. E. A new method for building protein conformations from sequence alignments with homologs of known structure. *J. Mol. Biol.*, v. 217, p. 1-7, 1991.

LEVITT, M. Accurate modeling of protein conformation by automatic segment matching. *J. Mol. Biol.*, v. 226, p. 507-533, 1992.

ŠALI, A. *et al.* From comparisons of protein sequences structures to protein modelling and design. *TIBS*, v. 15, p. 235-240, 1990.

UNGER, R. *et al.* A 3D building-blocks approach to analyzing and predicting structure of proteins. *Proteins*, v. 5, n. 4, p. 355-373, 1989.

WARME, C. M. *et al.* Computation of structures of homologous proteins: α -lactalbumin from lysozyme. *Biochemistry*, v. 13, p. 768-782, 1974.

VI.1. MODELLER

O método de satisfação das restrições espaciais utilizado por MODELLER é dividido em três etapas: alinhamento da sequência a ser modelada com as das estruturas homólogas; extração das restrições espaciais a partir do alinhamento e das características das estruturas; e satisfação destas restrições para obter o modelo tridimensional da proteína em questão (figura VI-4).

Figura VI-4: Modelagem molecular por satisfação das restrições espaciais (Šali, 1997). 1. A sequência alvo (SEQ) é alinhada com as sequências de proteínas com estruturas 3D conhecidas. 2. Características espaciais como distâncias C α -C α , ligações de hidrogênio, ângulos diedros são transferidos das estruturas para a proteína alvo. 3. Obtenção da estrutura 3D através da satisfação das características espaciais obtidas em 2.

O programa utilizado para a primeira etapa da modelagem pode ser outro¹ que não o MODELLER. Os dois últimos estágios correspondem ao procedimento principal do programa. As restrições espaciais são obtidas por meio de uma análise estatística das características das estruturas homólogas e representadas por uma função densidade de probabilidade condicional (ou pdf). Uma segunda fonte de restrições espaciais são as restrições estereoquímicas que não dependem das estruturas homólogas, mas simplesmente descrevem a geometria esperada para

¹ O programa de alinhamento utilizado neste trabalho foi o MULTALIGN do pacote AMPS (Barton e Sternberg, 1987)

uma cadeia polipeptídica. O modelo tridimensional da proteína é obtido pela otimização da pdf molecular, que inclui contribuições de todas as pdf's individuais, através do *Variable Target Function Method* (VTFM) utilizando o método de gradiente conjugado, e dinâmica molecular com *simulated annealing*.

VI.1.1. Derivação das Restrições Espaciais

Para se chegar a uma estrutura, a proteína alvo deverá satisfazer o maior número de restrições espaciais homólogas possíveis, como por exemplo, distâncias interatômicas, ângulos diedros, distâncias de ligações não covalentes, entre outras.

Estas restrições são definidas em termos de densidade de probabilidade que matematicamente e fisicamente, fornecem muito mais informações do que o simples valor médio das medidas. Desta forma, a restrição de uma característica x é definida como:

$$p(x/z_1, z_2, \dots, z_{m-1}, z_m) \quad \text{Eq. (1)}$$

onde p é a função densidade de probabilidade (pdf); z , as variáveis independentes previamente conhecidas; e x , a característica espacial da sequência da proteína alvo a ser restringida, dependente de z .

A função acima é chamada de função densidade de probabilidade **condicional**, pois sua distribuição em termos de x depende dos parâmetros z . Por exemplo, se x é o ângulo diedro χ_1 é provável que dependa do tipo de resíduo e dos ângulos diedros da cadeia principal Φ e Ψ e, portanto: $p(\chi_1/\text{tipo de resíduo}, \Phi, \Psi)$.

Para possuir um caráter probabilístico, a equação 1 deve satisfazer os seguintes critérios: $p(x/z_1, z_2, \dots, z_{m-1}, z_m) \geq 0$; a integral em todo o espaço x ser igual a 1; e a probabilidade de ocorrer um determinado evento entre o intervalo x_1 e x_2 , ser obtido por: $\int_{x_1}^{x_2} p(x/z_1, z_2, \dots, z_{m-1}, z_m) dx$.

A pdf, entretanto, corresponde a uma função teórica perfeita que não é possível obter na prática. Porém, uma aproximação pode ser obtida a partir de um banco de dados suficientemente grande. Com isto:

$$p(x/z_1, z_2, \dots, z_{m-1}, z_m) \approx W'_{x,z_1,z_2,\dots,z_m} \quad \text{Eq. (2)}$$

ou seja, faz-se uma aproximação da função p para uma matriz multidimensional W' que, como a pdf, possui características de uma função probabilidade satisfazendo os critérios citados acima.

A $W'_{x,z_1,z_2,\dots,z_m}$ é uma “função” construída por informações geradas empiricamente, contendo as frequências relativas das ocorrências de x , dado $z_1, z_2, \dots, z_{m-1}, z_m$. Têm-se então:

$$W'_{x,z_1,z_2,\dots,z_m} = \frac{W_{x,z_1,z_2,\dots,z_m}}{\sum_x W_{x,z_1,z_2,\dots,z_m}}$$

Eq. (3)

onde $W_{x,z_1,z_2,\dots,z_m}$ é uma matriz $(m+1) \times (m+1)$ cujos elementos representam as frequências absolutas, ou a contagem direta do número de ocorrências de cada combinação de $(x, z_1, z_2, \dots, z_{m-1}, z_m)$ obtidas a partir de um banco de estruturas alinhadas.

Uma lista das variáveis consideradas pelo programa é dada na tabela VI-2.

Tabela VI-2: Características espaciais que podem ser utilizadas pelo programa para calcular os elementos da matriz de frequências relativas, W' .

Índice	Variável	Característica
1	r	tipo de resíduo
2	Φ	ângulo diedro da cadeia principal Φ
3	Ψ	ângulo diedro da cadeia principal Ψ
4	ω	ângulo diedro da cadeia principal ω
5	ω_c	classe do ângulo diedro da cadeia principal ω . ("0°": cis, "180°": trans)
6	t	classe da estrutura secundária ("+" Φ , "H": hélice- α , "G": hélice-3 ₁₀ , "I": hélice- π , "E": fita- β , "B": saliência- β , " " , cadeia estendida, "T": volta, "S": dobra) a que pertence um resíduo
7	M	classe de conformação da cadeia principal de um resíduo ("A": hélice- α de mão direita, "B": fita- β , "P": conformação poliprolina, "G": região acessível, <i>a priori</i> , pelas glicinas com Φ positivo, "L": hélice- α de mão esquerda, "E": conformação estendida)
8	α	conteúdo de estrutura secundária (α , β , $\alpha+\beta$, α/β)
9	χ_i	ângulos diedros da cadeia lateral, $i = 1, 2, 3, 4, 5$.
10	c_i	classe a que pertencem os ângulos diedros da cadeia lateral ("+" : gauche +, "t": trans, " - " gauche -), $i = 1, 2, 3, 4, 5$.
11	$\underline{a}$	acessibilidade do resíduo ao solvente
12	$\overline{a}$	acessibilidade média de dois resíduos em uma proteína
13	$\underline{s}$	diferença local em termos de resíduos entre duas proteínas
14	$\overline{s}$	diferença média local em termos de resíduos entre duas proteínas
15	i	identidade seqüencial relativa entre duas proteínas
16	d	distância C $_{\alpha}$ - C $_{\alpha}$
17	Δd	diferença entre duas distâncias C $_{\alpha}$ - C $_{\alpha}$ de duas proteínas
18	h	distância N-O da cadeia principal
19	Δh	diferença entre duas distâncias N-O da cadeia principal de duas proteínas
20	b	fator de temperatura isotrópica (B _{isso}) média do resíduo
21	R	resolução da estrutura cristalográfica
22	$\underline{g}$	distância, em um alinhamento, em que o resíduo se encontra do <i>gap</i>
23	$\overline{g}$	distância média, em um alinhamento, em que o resíduo se encontra do <i>gap</i>

VI.1.1.1. Cálculo da Distribuição de Probabilidade de um Esperso Conjunto de Dados

Usualmente, o número de estruturas de boa qualidade e de razoável similaridade com a seqüência de proteína a ser modelada, é muito pequeno e passível de carregar erros significativos nos cálculos de probabilidade W' , a partir das frequências W .

Lida-se este problema inserindo $W'(x)$ como uma pequena perturbação da distribuição uniforme de probabilidade (A). Supõe-se, a seguir, que a “função” $W'(x)$ da equação (3) seja uma distribuição de probabilidades não contínua, mas construída por passos, sendo que cada passo x_i corresponda a uma medida de valor igual a $W(x_i)$. A distribuição resultante é novamente perturbada até que se esgote o número total de medidas, chegando finalmente na distribuição de probabilidade p . Portanto:

$$p(x_i) = w_1 A(x_i) + w_2 W'(x_i) \quad \text{Eq. (4)}$$

cujos pesos w_1 e w_2 são iguais a

$$w_1 = [1 + N/(\sigma n_x)]^{-1} \quad w_2 = 1 - w_1 \quad \text{Eq. (5)}$$

onde N é o número total de pontos ou dados em $W'(x)$; n_x o número de divisões da curva de $A(x)$; e σ , parâmetro que determina a contribuição relativa de A e W' na distribuição final p .

Segundo equações (4) e (5), um grande número N de informações contidas em $W'(x)$ fará com que $p(x_i)$ seja quase que completamente dominado pelos dados experimentais expressos em $W'(x_i)$. Por outro lado, se N é pequeno e diferente de zero, a contribuição de $A(x_i)$ terá um peso tão significativo quanto a de $W'(x_i)$, não induzindo grandes alterações das informações experimentais para o cálculo de $p(x_i)$. Um caso limite ocorre quando o número médio de dados por divisões (N/n_x) é igual a σ . Desta forma, $w_1 = w_2 = 1/2$ e $p(x_i) = 1/2 \cdot [A(x_i) + W'(x_i)]$, ou seja, a contribuição de $A(x_i)$ na função $p(x_i)$ será exatamente igual a de $W'(x_i)$.

A equação (4) está um tanto simplificada, uma vez que o algoritmo lida com W' multidimensional, num espaço de $(m+1)=N$ dimensões, lembrando que m corresponde ao número total de variáveis independentes. Desta forma, são

necessários N ciclos para se chegar à função densidade de probabilidade final p^N . Sendo n um estágio qualquer entre 1 e N , a função densidade de probabilidade correspondente consiste na seguinte operação:

$$p^n(x_i / z_j^1, z_k^2, \dots, z_l^{n-1}) = w_1^{jk\dots l} A^n(x_i / z_j^1, z_k^2, \dots, z_l^{n-1}) + w_2^{jk\dots l} W^n(x_i / z_j^1, z_k^2, \dots, z_l^{n-1}) \quad \text{Eq. (6)}$$

onde p^n é a função densidade de probabilidade condicional para x_i dado $z_j^1, z_k^2, \dots, z_l^{n-1}$, sendo que os índices superiores diferenciam as várias características espaciais, e os índices inferiores os tipos de cada característica, como por exemplo, os 20 diferentes aminoácidos da característica resíduos; e W^n a matriz que representa as frações relativas da contagem dos dados calculados em W^n . Analogamente a equação (3), tem-se:

$$W^n(x_i / z_j^1, z_k^2, \dots, z_l^{n-1}) = \frac{\sum_{z^n, z^{n+1}, \dots, z^{N-1}} W^n(x_i, z_j^1, z_k^2, \dots, z_l^{N-1})}{\sum_{x_i, z^n, z^{n+1}, \dots, z^{N-1}} W^n(x_i, z_j^1, z_k^2, \dots, z_l^{N-1})} \quad \text{Eq. (7)}$$

A distribuição A^n é calculada a partir da média ponderada de todos os p^{n-1} :

$$A^n(x_i / z_j^1, z_k^2, \dots, z_l^{n-1}) = \sum_c \rho_c p_c^{n-1} \quad \sum_c \rho_c = 1 \quad \text{Eq. (8)}$$

onde c refere-se ao número total das diferentes combinações possíveis entre as $n-2$ variáveis independentes z , ou seja:

$$c = (N-1)! / \{ [(N-1)-(n-2)]! \cdot (n-2)! \} \quad \text{Eq. (9)}$$

Caso as variáveis independentes não fornecerem nenhuma contribuição adicional a variável dependente x , p_c^{n-1} não será incluído nos cálculos de A^n . Do contrário, quanto maior a dependência de x pelas variáveis independentes, maior a contribuição de p_c^{n-1} em A^n . Uma medida desta dependência é feita segundo o cálculo da entropia de p^{n-1} :

$$S_c(p^{n-1}) = - \sum_i p^{n-1}(x_i / z_j^1, z_k^2, \dots, z_l^{n-2}) \log p^{n-1}(x_i / z_j^1, z_k^2, \dots, z_l^{n-2}) \quad \text{Eq. (10)}$$

cujo máximo valor da entropia, $S_{\max}$, é igual a $\log(n_x)$, uma distribuição uniforme de p , o que significa que x é completamente independente das variáveis z . Portanto, o peso ρ_c pode ser definido como:

$$\rho_c = \frac{(S_{\max} - S_c)}{\sum_c (S_{\max} - S_c)} \quad \text{Eq. (11)}$$

Contrariamente, se a entropia resultar em um valor mínimo, este conjunto de variáveis independentes serão as que melhor definirão x . A seguinte relação quantifica a dependência de x pelas variáveis independentes z :

$$U(x/z_1, z_2, \dots, z_n) = \frac{S[p(x)] - S[p(x/z_1, z_2, \dots, z_n)]}{S[p(x)]} \quad \text{Eq. (12)}$$

Este coeficiente de incerteza de x fornece um valor entre 0 e 1, significando que: se x é completamente determinado por z , $S[p(x/z_1, z_2, \dots, z_n)]$ é mínimo de maneira que $U(x/z_1, z_2, \dots, z_n)$ tende a 1; se x é completamente independente das variáveis z , $S[p(x/z_1, z_2, \dots, z_n)]$ é igual a $S[p(x)]$ e portanto, $U[p(x/z_1, z_2, \dots, z_n)]$ igual a zero, segundo dita a equação 11.

VI.1.1.2. Derivação das Restrições Estereoquímicas

Diferentemente das restrições obtidas empiricamente a partir de uma estrutura homóloga, as restrições estereoquímicas são analiticamente retiradas da seqüência de aminácidos da proteína. Os valores médios e os desvios padrões de comprimentos e ângulos de ligações são determinados a partir de um conjunto de parâmetros derivados dos estudos espectroscópicos, cristalográficos e teóricos de pequenas moléculas.

As restrições estereoquímicas consideradas pelo MODELLER compreendem: as distâncias interatômicas; os ângulos diedros; planaridade de grupos peptídicos e anéis da cadeia lateral; quiralidade dos $C\alpha$ e dos C da cadeia lateral; distância de contato de van der Waals; comprimento de ligação, ângulos de ligação e ângulos diedros de pontes dissulfeto.

Estas restrições são descritas conforme o modelo de oscilador harmônico, cuja energia é dada por:

$$E(b) = \frac{1}{2} k_b (b - b_o)^2 \quad \text{Eq. (13)}$$

onde b é uma variável qualquer, podendo ser distância ou ângulo; b_o , o comprimento ou ângulo do sistema em repouso; e k_b , sua constante elástica.

Desta forma, a função densidade de probabilidade de b é, segundo a mecânica estatística, uma função Gaussiana do tipo:

$$p^b(b) = \frac{1}{\sigma_b \sqrt{2\pi}} e^{\left[-\frac{1}{2} \left(\frac{b-\bar{b}}{\sigma_b}\right)^2\right]} = N(\bar{b}, \sigma_b) \quad \text{Eq. (14)}$$

onde $\sigma_b = \sqrt{k_B T / k_b}$, sendo k_B , a constante de Boltzmann e T , a temperatura; e $\bar{b} = b_o$. Note que para definir esta distribuição, são necessários apenas dois parâmetros: o valor médio b , e o seu desvio σ_b . $N(\bar{b}, \sigma_b)$ corresponde a uma notação resumida de uma função Gaussiana indicando que sua distribuição está centrada em $\bar{b}$, com um desvio igual a σ_b .

Portanto, para permitir uma conformação *cis*, por exemplo, o ângulo diedro ω deve ser restringido por uma soma de duas funções Gaussianas: uma representando a distribuição da conformação *trans* com o valor médio 180° e o desvio de 6° ; e a outra distribuição representando a conformação *cis* com ângulo médio de 0° e desvio de 6° . Deste modo:

$$p^\omega = w_1 \cdot N(180^\circ, 6^\circ) + w_2 \cdot N(0^\circ, 6^\circ); \quad w_1 + w_2 = 1 \quad \text{Eq. (15)}$$

onde w_1 e w_2 são pesos atribuídos às duas conformações. Caso a conformação *cis* não tenha sido especificado, o peso w_2 será 0 e a função distribuição de probabilidade p^ω será determinada somente pela conformação *trans* $N(180^\circ, 6^\circ)$.

VI.1.1.3. Restrições de Distâncias entre C α - C α e N - O

Considera-se a modelagem de uma proteína A, baseando-se numa única estrutura 3D homóloga B. De um modo geral, é para se esperar que a distância entre dois C α quaisquer da estrutura modelada de A concorde com o valor encontrado na estrutura B.

Sendo respectivamente d e d' , as distâncias C α - C α da proteína alvo (A) e da estrutura homóloga conhecida (B), a restrição x neste caso será a diferença entre estas distâncias ($d-d'$). Deste modo, o MODELLER irá tabular as frequências relativas $W'(d-d' / \bar{g}, i, \bar{a}', d')$ obtendo a função densidade de probabilidade de ($d-d'$) condicionada a quatro variáveis independentes, onde: d' , é a distância C α - C α da estrutura correspondente; i , a identidade seqüencial relativa entre duas proteínas; $\bar{a}'$, acessibilidade média de dois resíduos da proteína de estrutura conhecida, ao solvente; $\bar{g}'$, distância média que um resíduo se localiza do *gap* (tabela VI-2). Em

geral, espera-se que $(d-d')$ tenha algum valor próximo de zero, pois as diferenças seqüenciais de A e B não permitiria que a estrutura se acomodasse rigidamente em zero.

Se a quantidade de valores possíveis de $(d-d')$ são grandes, haverá uma incerteza maior sobre o seu melhor valor. Um valor mais preciso é obtido com a imposição de condições para a característica a ser restringida, como por exemplo a identidade, a acessibilidade, ou distância ao *gap*. Deste modo, apenas alguns valores de $(d-d')$ se enquadrarão nestas condições, aumentando a precisão da informação.

Os histogramas da figura VI-5 demonstram que as distribuições de probabilidade ou de freqüência podem ser aproximadas por uma função Gaussiana, onde o valor médio é, desta forma, igual a zero e o desvio padrão definido a partir das variáveis independentes.

$$p^d(d/\bar{g}, i, \bar{a}', d') = \frac{1}{\sigma(\bar{g}, i, \bar{a}', d')\sqrt{2\pi}} e^{\left[-\frac{1}{2}\left(\frac{d-d'}{\sigma(\bar{g}, i, \bar{a}', d')}\right)^2\right]} \quad \text{Eq. (16)}$$

A função densidade de probabilidade calculada segundo este modelo é determinada baseando-se apenas no alinhamento entre a seqüência da proteína alvo e a da estrutura homóloga, enquanto que o desvio padrão é obtido através dos valores das variáveis independentes e pelo ajuste por mínimos quadrados da função modelo (equação 16), sobre os dados experimentais representados pelos histogramas de W na figura VI-5.

A figura VI-5 ilustra as distribuições de distâncias $C\alpha-C\alpha$ ($d-d'$) e $N-O$ ($h-h'$) em dois casos extremos. Observa-se que restringindo os dados para aqueles que provém de estruturas de maior identidade seqüencial (i), pouca ou nenhuma acessibilidade ao solvente (a') e mais bem conservadas (g), a distribuição se torna mais estreita e “aguda”, limitando a intervalo(s) de diferença de distância (σ) menor(es) e mais próximo(s) de zero.

Figura VI-5: Diferentes distribuições W' de pares equivalentes de distâncias entre a proteína alvo e estruturas do banco de dados local, fitados pelos modelos Gaussianos (linha contínua). Nota-se que dados provenientes de estruturas com porcentagem de identidade maior, átomos mais conservados e localizados no interior da molécula, converge grande parte das ocorrências para $(d-d')=0$, resultando em um desvio de $\sigma \sim 7$ vezes menor que em (a) ou (c). g , a' , d' e i correspondem às variáveis independentes de $(d-d')$ ou $(h-h')$; N , o número de distâncias que satisfazem as condições definidas pelas variáveis independentes; e σ , o desvio padrão do modelo Gaussiano.

A figura acima demonstra também que o modelo de distribuição Gaussiana pode descrever bem a dependência real da distância d pelas quatro variáveis independentes.

Analogamente, as distâncias entre os átomos de N e O da cadeia principal da proteína homóloga podem ser modeladas por (figura VI-5):

$$p^h(d | \bar{g}, i, \bar{a}', h') = \frac{1}{\sigma(\bar{g}, i, \bar{a}', h') \sqrt{2\pi}} e^{\left[-\frac{1}{2} \left(\frac{d - \bar{d}'}{\sigma(\bar{g}, i, \bar{a}', h')} \right)^2 \right]} \quad \text{Eq. (17)}$$

Esta restrição envolve ligações de hidrogênio da cadeia principal e, portanto, determinante em termos de estrutura secundária.

VI.1.1.4. Restrição da Cadeia Principal de um Resíduo

A conformação da cadeia principal de um resíduo é definida pelos ângulos torcionais Φ e Ψ . Apenas algumas combinações destas são estereoquimicamente

permitidas, e definem as seis classes de conformação M da cadeia principal: A, B, P, G, L e E (tabela VI-2). Desta forma, a função densidade de probabilidade dos ângulos de Φ e Ψ construído a partir da soma ponderada da função Gaussiana das seis possíveis conformações:

$$\begin{aligned} p^m(\Phi) &= \sum_{i=A, \dots, E} \omega_i N[\bar{\Phi}_i, \sigma_i(\Phi)] \\ p^m(\Psi) &= \sum_{i=A, \dots, E} \omega_i N[\bar{\Psi}_i, \sigma_i(\Psi)] \end{aligned} \tag{Eq. (18)}$$

A distribuição da função p^m no espaço Φ ou Ψ dependerá fortemente das probabilidades ω_i obtidas a partir das informações equivalentes no banco de dados local, onde foi demonstrado que o tipo de resíduo (r), a conformação do resíduo na estrutura conhecida (M') e a similaridade local (s) são as variáveis independentes que melhor definem as características da cadeia principal M .

VI.1.1.5. Restrição da Cadeia Lateral de um Resíduo

A cadeia lateral de um resíduo é definida pelos ângulos diedros χ e para cada ângulo diedro χ_i , a função densidade de probabilidade é analogamente descrita como:

$$p^s(\chi_i) = \sum_j \omega_{ij} N[\bar{\chi}_{ij}, \sigma_j(\chi_i)] \tag{Eq. (19)}$$

onde o somatório compreende todas as classes de χ_i (tabela VI-2). N é a função Gaussiana modelo, centrada em $\bar{\chi}_{ij}$ com desvio de $\sigma_j(\chi_i)$. Segundo os resultados do MODELLER, as distribuições que melhor definiram χ e, portanto, utilizadas para os cálculos de ω_{ij} são:

$$\chi_1 \rightarrow p(c_1 / r, r', c_1', s) \tag{Eq. (20)}$$

$$\chi_2 \rightarrow p(c_2 / r, r', c_1', c_2') \tag{Eq. (21)}$$

$$\chi_3 \rightarrow p(c_3 / r, r', c_3', t) \tag{Eq. (22)}$$

$$\chi_4 \rightarrow p(c_4 / r) \tag{Eq. (23)}$$

onde as variáveis independentes estão descritas na tabela VI-2.

VI.1.2. Satisfação das Restrições Espaciais

As funções densidade de probabilidade de todas as características geométricas e estereoquímicas da proteína alvo devem ser combinadas em uma pdf molecular. O modelo tridimensional da proteína alvo é derivada a partir da otimização desta pdf molecular, encontrando assim a combinação que melhor satisfaz o maior número de restrições.

Primeiramente, deriva-se a pdf característica descrita na base das funções densidade de probabilidade apresentadas nos itens VI.1.1.2-5. Cada uma destas k funções base é representada por $p_k^f(f)$, correspondendo a uma das “dimensões” base que definirão o $p^F(f)$. Esta última função característica de f , combina todas as informações possíveis de $p_k^f(f)$, que f pode assumir. Este procedimento está melhor explicado no item a seguir.

VI.1.2.1. Derivação da pdf Característica

Com a pdf característica, é possível construir uma função densidade de probabilidade de uma determinada característica a partir de informações simultâneas de duas ou mais estruturas.

Têm-se duas estruturas conhecidas e uma proteína alvo da qual se deseja determinar a distância d entre os átomos $C\alpha$. As seguintes funções base são consideradas: $p_1^d(d)=p(d/d')$ e $p_2^d(d)=p(d/d'')$, onde d' e d'' correspondem as distâncias equivalentes a d de cada uma das estruturas; restrição estereoquímica dos raios de van der Waals, dada por $p^v(d)$, estabelecendo que a distância interatômica deve ser maior que a soma das distâncias dos raios de van der Waals.

A combinação destas três informações deverá resultar em $p(d/d', d'', \bar{s}, \bar{s}'')$, onde a variável independente $\bar{s}$ (uma medida da similaridade local de resíduos de duas proteínas), é também considerada com o propósito de definir um peso $\omega(\bar{s})$ à contribuição de cada uma das individuais densidades de probabilidade $p(d/d')$ e $p(d/d'')$. Assim,

$$p(d/d', d'', \bar{s}, \bar{s}'') = \omega(\bar{s}) \cdot p(d/d') + \omega(\bar{s}'') \cdot p(d/d'') \quad \text{Eq. (24)}$$

Esta relação confere com comportamento dos histogramas derivados diretamente de $W(d/d', d'', \bar{s}, \bar{s}'')$, como pode ser verificado pelas curvas contínuas da figura VI-6.

Nota-se, também, que a soma ponderada das pdf's base representa bem $W(d/d', d'', \bar{s}', \bar{s}'')$, uma vez que o perfil das curvas depende claramente das variáveis $\bar{s}'$ e $\bar{s}''$.

Figura VI-6: Derivação da pdf característica a partir das pdfs base. Em todos os gráficos, $d'=18.0-18.5$ e $d''=22.5-23.0$. (a) $W(d/d', d'')$. (b) $W(d/d', d'', s', s'')$ onde $s'=0.2-0.4$ e $s''=0.2-0.4$. (c) $W(d/d', d'', s', s'')$ onde $s'=0.2-0.4$ e $s''=0.4-0.6$. (d) $W(d/d', d'', s', s'')$ onde $s'=0.4-0.6$ e $s''=0.2-0.4$. Os histogramas foram obtidos a partir dos dados da base local do MODELLER. As curvas tracejadas representam as pdf base $p_1^d(d)=p(d/d')$ e $p_2^d(d)=p(d/d'')$ dadas pela equação 16. As linhas contínuas são as pdf características $p^D(d/d', d'', s', s'')$ calculados segundo a equação 24.

Finalmente, obtém-se a pdf característica final acrescentando à pdf característica (equação 24) a restrição estereoquímica de van der Waals, que deve ser multiplicada, ao invés de somada, para satisfazer as informações de todas as estruturas. Portanto:

$$p^D(d) = p(d/d', d'', \bar{s}', \bar{s}'') p^v(d)$$

Eq. (25)

$$p^D(d) = [\omega(\bar{s}') \cdot p(d/d') + \omega(\bar{s}'') \cdot p(d/d'')] p^v(d)$$

A equação generalizada da pdf característica final de cada tipo de restrição é definida a seguir:

- Restrições de distância C α -C α

$$p^D(d) = p^v(d) \sum_i \omega_i p_i^d(d / z_1, z_2, \dots, z_{m-1}, z_m) \quad \text{Eq. (26)}$$

- Restrições de distância N-O da cadeia principal

$$p^H(h) = p^v(h) \sum_i \omega_i p_i^h(h / z_1, z_2, \dots, z_{m-1}, z_m) \quad \text{Eq. (27)}$$

- Restrições estereoquímicas

$$p^E(e) = p^e(e) \quad \text{Eq. (28)}$$

onde: i corresponde ao número de seqüências de estruturas homólogas a proteína alvo; e e , uma característica qualquer que pode ser comprimento de ligação, ângulo de ligação, ângulo diedro, contato de van der Waals, ou distância, ângulo e ângulo diedro de ponte dissulfeto.

- Restrições da conformação da cadeia principal

$$p^M(\theta) = \begin{cases} \sum_{i=1}^n \omega_i p_i^{m'}(\theta / z_1, z_2, \dots, z_m) & n > 0 \\ p^{m'}(\theta / R) & n = 0 \end{cases} \quad \text{Eq. (29)}$$

- Restrições da conformação da cadeia lateral

$$p^S(c) = \begin{cases} \sum_{i=1}^n \omega_i p_i^S(c / z_1, z_2, \dots, z_m) & n > 0 \\ p^S(c / R) & n = 0 \end{cases} \quad \text{Eq. (30)}$$

As duas últimas restrições podem ser aplicadas mesmo quando não houver resíduos equivalentes nas estruturas ($n=0$). Especificamente para este resíduo, utiliza-se restrições baseados em uma biblioteca de rotâmeros.

VI.1.2.2. Derivação da pdf Molecular a partir da pdf Característica

A pdf molecular corresponde a combinação de todas as pdfs características, da qual obtém-se o modelo tridimensional da proteína alvo encontrando a combinação mais provável dentre as várias características f . Deste modo, a pdf molecular é definida como o produtório de todas as pdfs características (equações 26-30):

$$P = \prod_i p^F(f_i) \quad \text{Eq. (31)}$$

Maximizando esta equação, encontra-se o modelo da proteína desconhecida que melhor satisfaz a imposição das restrições espaciais definidas pela correspondente combinação das pdfs características.

VI.1.2.3. Otimização da pdf Molecular

A pdf molecular é otimizada indiretamente através da função objetiva:

$$F = - \ln(P) \quad \text{Eq. (32)}$$

Esta é uma conversão técnica apenas para facilitar os cálculos de otimização que, de um produtório de termos passa a ser com a função objetiva, uma simples função de somas:

$$F = - \sum_i \ln[p^F(f_i)] \quad \text{Eq. (33)}$$

A otimização é feita segundo a aproximação da Função Alvo Variável (*Variable Target Function*) no espaço cartesiano, utilizando o Gradiente Conjugado como o algoritmo otimizador (Braun e Gõ, 1985). Desta forma, o melhor valor da pdf molecular é encontrado a partir de sucessivas otimizações da função alvo P . Inicialmente, apenas um número limitado de restrições locais é considerado na otimização de P . Nesta etapa, P é substituída por $P(\Delta r=1)$ onde $\Delta r=1,2,3,\dots,N$ (sendo N , o número total de resíduos da proteína alvo). A cada etapa de otimização, é englobada um intervalo cada vez maior de restrições culminando, finalmente, na otimização da verdadeira pdf molecular P com a incorporação de todas as restrições, de todos os resíduos ($\Delta r=N$).

Após esta otimização inicial da função objetiva pelo método da Função Alvo Variável, o processo de otimização é completado com a execução do método de Dinâmica Molecular com *Simulated Annealing*. Vários modelos ligeiramente diferentes resultarão desta otimização, dos quais podem ser utilizados para estimar os erros das regiões que apresentaram uma maior variabilidade no enovelamento. A figura a seguir ilustra uma das maneiras de otimização utilizadas por MODELLER, a partir de uma estrutura extendida.

No presente trabalho, a otimização não foi feita a partir de uma molécula extendida, mas de uma estrutura construída a partir das coordenadas da estrutura homóloga e completadas com informações estereoquímicas padrão (CHARMM). Antes mesmo de aplicar as restrições espaciais e iniciar os vários ciclos de

otimização (gradiente conjugado, *simulated annealing* com dinâmica molecular) são feitos deslocamentos aleatórios nas posições dos átomos em um intervalo de $\pm 4\text{\AA}$, por exemplo.

Figura VI-7: Otimização da função objetiva (equação 33). O número de iterações é também indicado a baixo de cada modelo. As primeiras ~2000 iterações correspondem ao método de Função Alvo Variável por Gradiente Conjugado. Este método satisfaz primeiramente as restrições locais, introduzindo nas iterações seguintes, restrições cada vez mais distantes até que toda a função objetiva é otimizada. As iterações finais correspondem ao refinamento do modelo pelo método de Dinâmica Molecular com *Simulated Annealing*.

VI.2. Avaliação dos Modelos

Dispõe-se, atualmente, de um grande número de programas de avaliação de estruturas de proteínas. A escolha da melhor estrutura, nas etapas finais da modelagem deste trabalho, foi feita a partir da análise das qualidades das propriedades a serem descritas abaixo, apesar de que a própria função objetiva ($-\ln(P)$) tenha sido utilizada como uma primeira medida de qualidade.

- **Estereoquímica** - Baseando-se em um banco de dados de estruturas experimentais bem refinadas, os seguintes parâmetros estereoquímicos do modelo são analisados: os ângulos torcionais Φ e Ψ (Diagrama de Ramachandran); os ângulos diedros ω ; a planaridade das ligações peptídicas; os comprimentos e ângulos de ligações dos átomos da cadeia principal; a geometria das ligações de hidrogênio; os ângulos diedros das cadeias laterais;

entre outros. O programa PROCHECK (Laskowski *et al.*, 1993) fornece uma medida geral média da qualidade do modelo, o fator-G, obtido a partir da avaliação destes parâmetros estereoquímicos. Dentro de uma dada faixa de resolução, cada um destes parâmetros é comparado com o valor médio e o desvio das estruturas experimentais. Assim, um fator-G pode ser calculado para um modelo gerado por modelagem comparativa ou modelagem por homologia, e determinar a “resolução” equivalente em termos de estereoquímica.

- **Ambiente Químico** (Bowie *et al.*, 1991) - A partir de informações de estruturas já conhecidas, constrói-se uma matriz 20×18 correspondendo aos 20 tipos de aminoácidos contra 18 diferentes classes de ambientes químicos definidos segundo a estrutura secundária, exposição ao solvente e a átomos polares. Cada um dos elementos s_{ij} da matriz corresponde a um *score* que indica a compatibilidade do resíduo i com o ambiente j . Desta forma, a medida da compatibilidade da seqüência do modelo com a sua estrutura é feita comparando o ambiente químico construído para cada resíduo baseando-se na matriz. Uma avaliação do modelo como um todo pode ser obtida a partir da soma desta. A seguinte equação empírica descreve a soma esperada S_{calc} para um determinado comprimento de seqüência L .

$$S_{calc} = \exp(-0.83 + 1.008 \times \ln(L)) \quad \text{Eq. (34)}$$

Um modelo aceitável obtém um valor maior que $0,45S_{calc}$. Valores menores que este limite indicam estruturas de má qualidade. O programa VERIFY_3D (Luthy *et al.*, 1992) utiliza esta metodologia para avaliar a compatibilidade da estrutura do modelo com a sua seqüência.

- **Contatos Atômicos** - Avalia o empacotamento da molécula segundo as interações átomo-átomo, pelo programa QUALITY do pacote WHATIF (Vriend e Sander, 1993). Para isto, é necessário conhecer as interações típicas de cada tipo de aminoácido. Os 20 aminoácidos são divididos em 80 fragmentos rígidos, individualmente centrados em um cubo de $16 \times 16 \times 16 \text{ \AA}^3$ delimitando, desta forma, o ambiente químico de cada fragmento. Os 80 fragmentos têm os seus átomos agrupados em 57 classes segundo sua natureza química. Distribuições de probabilidade são calculadas a partir do número de ocorrências de cada tipo de átomo em torno de cada fragmento, utilizando um banco de dados de estruturas bem resolvidas. O resultado é um conjunto de $80 \times 57 = 4560$

distribuições “padrão” que descrevem contatos atômicos típicos em proteínas, podendo ser utilizadas na avaliação de um modelo. A concordância entre os contatos observados para cada fragmento rígido de cada resíduo da estrutura do modelo com as distribuições probabilísticas, fornecerá o correspondente índice de qualidade dos contatos atômicos do modelo.

VI.3. PATCHES

Uma vez obtida a estrutura 3D (em muitos casos por modelagem) é interessante conhecer as possíveis interações desta com um outro monômero ou ligante. Existem regiões na superfície da proteína com características típicas de interface de dimerização ou complexação. Estas características podem ser analisadas para uma determinada região e comparadas com a superfície restante da proteína a fim de localizar a que tiver a maior propensão para um contato proteína-proteína. Esta abordagem é interessante quando não se conhece a forma de interação a partir da estrutura homóloga e, portanto, pretende-se prevê-la a partir da estrutura monomérica modelada.

O programa PATCHES desenvolvido por Jones e Thornton (1997a e 1997b) prediz a localização de uma região (*patch*) de dimerização ou complexação pela comparação das características desta com as das outras regiões da molécula. Cada resíduo pertencente à região em questão é analisado segundo sua hidrofobicidade, potencial de solvatação, índice de protrusão, planaridade, propensidade do resíduo à interface, e área superficial acessível ao solvente (ASA). Um *score* é atribuído a cada uma destas características e combinadas de forma a fornecer um valor da probabilidade da região em formar uma interação proteína-proteína.

Para cada resíduo da superfície da molécula, é atribuído um *patch* delimitado por um número n de resíduos. No entanto, apenas os $C\alpha$ destes resíduos são considerados pelo programa na identificação do *patch*. Para cada um destes resíduos, é obtido um valor relativo a cada uma das seis características e calculada uma média aritmética dos valores dos n resíduos pertencentes ao *patch* para uma determinada característica. Cada um dos seis valores de um *patch* é comparado com os dos *patches* restantes, e normalizado de forma que cada

parâmetro seja diferenciado por um *score* (S) de 1 a 100. Estes *scores* aumentam proporcionalmente ao valor de cada um dos seis parâmetros considerados. Os menores *scores* correspondem aos *patches* que possuem a maior hidrofobicidade, o menor potencial de solvatação, o maior índice de protrusão, o maior grau de planaridade, o maior índice de propensidade à interface e o maior valor de ASA. Estas correspondem geralmente às características típicas de um homodímero.

Um “*score* combinado” (P_j) é calculado pela média aritmética dos *scores* (S) das seis características. Entretanto, dependendo do tipo de complexação (homodímero, heterodímero, antígeno-anticorpo, etc.), algumas características que não contribuem significativamente na diferenciação do *patch* como superfície de dimerização, não são incluídos nos cálculos.

Como, por exemplo, a definição do “*score* combinado” para homodímeros inclui todos os termos, dado por:

$$P_j = (S_{sp} + S_{rp} + S_{hy} + S_{pi} + S_{asa} + S_{pl}) / N_{par} \quad \text{Eq. 35}$$

Os termos acima correspondem aos *scores* do *patch* P_j na distribuição: do potencial de solvatação (S_{sp}); dos resíduos propensos à interface (S_{rp}); da hidrofobicidade (S_{hy}); do índice de protrusão (S_{pi}); da acessibilidade ao solvente (S_{asa}); e planaridade (S_{pl}). N_{par} corresponde ao número de parâmetros considerados nos cálculos do “*score* combinado” P_j .

O diagrama a seguir resume as etapas de predição da superfície de dimerização de uma proteína.

Figura VI- 8: Fluxograma do algoritmo PATCHES, utilizado para prever a região mais provável para uma interação proteína-proteína.

VII.

MATERIAIS E MÉTODOS

Lista-se a seguir os programas utilizados nesta parte B do trabalho, executados em ambientes Windows 98 (PC), Linux (PC), IRIX 5.3 e 6.x (estações gráficas da Silicon Graphics).

VII.1. BLAST

O programa BLASTP v. 2.0.13 (Altschul, *et al.*, 1997) derivado do algoritmo de BLAST (*Basic Local Alignment Search Tool*) (Altschul, *et al.*, 1990), foi utilizado para encontrar seqüências de proteínas homólogas a DM43. A matriz *default* do programa, BLOSSUM62, foi empregada para calcular os alinhamentos no modo *pairwise* de DM43 contra dois bancos.

O primeiro banco utilizado foi o *nr* (*non redundant*) escolhido de modo que o programa retorne com resultados de alinhamentos de seqüências não redundantes, derivadas de diversos bancos de dados, incluindo: GenBank CDS translations (Benson, *et al.*, 1999), PDB (*Protein Data Bank*) (Berman, *et al.*, 2000), SwissProt (Bairoch, *et al.*, 2000), PIR (*Protein Information Resource*) (Barker *et al.*, 2000), PRF (*Protein Research Information* - <http://www.prf.or.jp/en/>).

A segunda busca foi feita especificamente contra o banco de dados PDB (*Protein Data Bank*) que contém seqüências de proteínas cujas estruturas já foram determinadas. Nesta etapa, pretendia-se encontrar estruturas homólogas a DM43 com o intuito de utilizá-las como referência para a predição da estrutura de DM43 por modelagem por homologia, segundo o método de satisfação das restrições espaciais.

VII.2. THREADER

THREADER v. 2.0 (Jones, D. T., *et al.*, 1992) é um programa de reconhecimento de enovelamentos pela comparação entre a seqüência de aminoácidos de entrada e um banco de estruturas representativas. Como resultado, o programa apresenta uma lista de estruturas com enovelamentos que mais se adequaram à seqüência.

O baixo grau de identidade entre os domínios de DM43 e proteínas pertencentes a estruturas do tipo imunoglobulina, segundo BLAST, seria por si só ainda insuficiente para garantir este enovelamento para DM43. Os resultados de THREADER, confirmando-o, foram importantes para auxiliar no alinhamento de DM43 com estruturas homólogas, fase esta, imprescindível para a próxima etapa: a modelagem.

Os melhores enovelamentos encontrados por THREADER, antes e após executar 50 randomizações, foram analisados no banco de dados do CATH (*Class Architecture Topology Homology*) (Orengo, 1997), onde as estruturas são qualificadas quanto a sua classe, arquitetura, topologia e homologia.

As coordenadas atômicas das melhores estruturas para se basear na construção de um modelo, segundo BLAST e THREADER, foram extraídas do PDB (*Protein Data Bank*) (Berman, *et al.*, 2000). Correspondem às três proteínas da família dos receptores inibitórios das células assassinas (KIRs-*Killer Cell Inhibitory Receptors*), constituídos de vários domínios, incluindo dois domínios tipo imunoglobulina: KIR2DL1 – 1nkr (Fan, *et al.*, 1997), KIR2DL2 – 2dl2, pelo (Snyder, *et al.*, 1999) e KIR2DL3 – 1b6u (Maenaka, *et al.*, 1999), onde as letras em minúsculo correspondem ao código pdb destas proteínas.

VII.3. MULTALIGN

O programa MULTALIGN do pacote AMPS - *Alignment of Multiple Protein Sequences* (v. 2.0) (Barton e Sternberg, 1987) foi utilizado para obter o alinhamento inicial de DM43 contra os KIRs2 (KIR2DL1, KIR2DL2 e KIR2DL3), e mais três proteínas sem estruturas conhecidas: a oprina e KIRs3 (KIR3DL1 e KIR3DL2) (Colonna e Samaridis, 1995). Os KIRs também são conhecidos como NKAT1, NKAT2, NKAT3 e NKAT4¹, respectivamente a KIR2DL1, KIR2DL3, KIR3DL1 e KIR3DL2. Os KIR3DL diferem dos KIR2DL, principalmente por possuírem um domínio extra em seu N-terminal, adotando também a conformação tipo imunoglobulina.

Baseado no algoritmo de Needleman e Wunsch (1970), o programa foi utilizado com a matriz de Dayhoff (1978) (PAM250) somando-se seus elementos a uma constante de valor igual a 8, e penalidade 8 para ocorrência de *gaps*.

A oprina está seqüencialmente incompleta no primeiro domínio, mas corresponde a uma proteína equivalente à DM43 (86.08% de identidade) encontrada em *Didelphis virginiana*, um gambá norte-americano (Catanese e Kress, 1992). Diferentemente dos KIRs2, os KIRs3, oprina e DM43 são constituídos de três domínios, enquanto que os primeiros possuem apenas dois.

Um padrão esperado de duas cisteínas foi encontrado por domínio uma vez que o enovelamento de todas elas, segundo seus autores (Fan, *et al.* (1997), Snyder, *et al.* (1999), Maenaka, *et al.* (1999), Colonna e Samaridis (1995)) e os resultados em BLAST e THREADER, revelaram a conformação imunoglobulina.

A partir do alinhamento automático por MULTALIGN, foram feitos ajustes manuais destas seqüências, baseando-se nos resultados do programa PROMOTIF v.2.0 para definição de regiões de estruturas secundárias (Hutchinson e Thornton, 1996), e o conhecimento prévio das várias propriedades comuns de domínios tipo imunoglobulina da família dos receptores hematopoiéticos.

Os receptores hematopoiéticos caracterizam-se por possuírem um motivo WSXWS box próximo a interface dos seus dois domínios tipo imunoglobulina. O

primeiro triptofano do motivo acomoda-se estrategicamente no centro desta interface, criando um contato que obriga a formação de um ângulo relativo entre os domínios de aproximadamente 90°.

Os KIRs2DL apresentaram o enovelamento imunoglobulina em seus domínios além do mesmo motivo WSXWS anterior à fita G, também encontrado em DM43.

VII.4. PROMOTIF

A versão 2.0 de PROMOTIF (Hutchinson e Thornton, 1996) foi utilizada para identificar na estrutura de KIR2DL1, ou 1nkr, a localização dos motivos estruturais quanto a fitas ou folhas betas, voltas betas e gamas, saliências betas, entre outros. Essencialmente, foi utilizado apenas como uma referência para o alinhamento manual, servindo de guia para as possíveis posições de uma deleção ou inserção.

VII.5. MODELLER

Foi utilizada a versão 4.0 do MODELLER (Šali e Blundell, 1993) para a modelagem da estrutura de DM43 a partir da sua seqüência de aminoácidos (Neves-Ferreira, 2000) e de pelo menos uma estrutura cristalográfica. Resultados de análises estruturais de qualidade, refinamento, resolução e identidade com DM43, o KIR2DL1 (1nkr.pdb) apresentou os melhores parâmetros, descartando-se assim, KIR2DL2 (2dl2.pdb) e KIR2DL3 (1b6u.pdb).

A disponibilidade de estruturas cristalográficas constituídas de apenas dois domínios imunoglobulinas (KIRs2DL) (denominadas de D₁ e D₂) contra os três (D₀, D₁ e D₂) sugeridos pela seqüência de aminoácidos de DM43, obrigou sua modelagem contra KIR2DL1 (1nkr.pdb), em duas porções: DM43_D₀ contra 1nkr_D₂ e DM43_D₁-D₂ contra 1nkr_D₁-D₂. Estas divisões foram decididas segundo o critério de maior identidade interdomínios.

¹ NKAT# (Natural Killer Associated Transcript #)

D ₀	>P1;1nkr_D2	YEKPSLSAQP GPTVLAGENV TLSCSSRSS- --YDMYHLNR EGEAHERRLP AGPKVNGTFQ ADFPLGPAT-	
	>P2;DM43_D0	DPTPPLWIKT ESPSTPWTNV TLLCVATNTE EL--SFQVWK DGEL-LSTLP V-----VGLV GKFWLGPVTA	.60
		.5 .20 .30 .40 .50	
	>P1;1nkr_D2	-HGGTYRCFG SFHDSPEWS KSSDPLLVSF T-***	
D ₁ -D ₂	>P2;DM43_D0	DNRGIYRCRI LTSN--DWT PLSAPVEVTG KEPL*	
		.70 .80 .90 .98	
	>P1;1nkr_D1-D2	RKPSLLAHPG PLVKSEETVI LQCWSDVMFE HFLHREGMF NDTLRLIGEY HDGVSKANFS ISRMTQDLAQ TYRCYGSVTH	
	>P2;DM43_D1-D2	PAPSLHAEPG PWILPGLQTK LHCRGMLLGM IFDLYQEGE- QEPVKS--- QTPSAEATFI VNS-----TG NYSCLYRAPA	.160
D ₁ -D ₂		.99 .110 .120 .130 .140 .150	
	>P1;1nkr_D1-D2	S-PYQVSAPS DPLDIVIIGL YEKPSLSAQP GPTVLAGENV TLSCSSR--- SSYDM-YHLS REGEAHERRL PAGPKVNGTF	
	>P2;DM43_D1-D2	SAPSVNSTPS ETIHVVIPDF LPKANFYILN DRDFRPGDIV TFSCWARFSE REYDLEFKLF KDQETPVEV VPISDDPMKVF	.240
		.170 .180 .190 .200 .210 .220 .230	
D ₁ -D ₂	>P1;1nkr_D1-D2	QADFFPLGPAT HGGTYRCFGS FHDSPYEWSK SSDPLLVSF*	
	>P2;DM43_D1-D2	FDLTAVGP-K DGGKYSCRYR FRNGPPIWSE DSNVLELDLS*	.288
		.250 .260 .270 .280	

Figura VII-1: Alinhamentos iniciais de 1nkr contra D₀ e D₁-D₂ de DM43, utilizados em MODELLER, na produção dos primeiros modelos das estruturas dos domínios D₀ e D₁-D₂, respectivamente. Os grupos de resíduos marcados em verde correspondem a regiões cujo alinhamento foi modificado durante o processo de modelagem, de tal forma a liberar algumas restrições causadoras de maus contatos estereoquímicos no modelo produzido. O alinhamento final está reproduzido na figura VII-2.

Figura VII-2: Alinhamentos de 1nkr contra D₀ e D₁-D₂ de DM43, utilizados em MODELLER, para produção dos modelos das estruturas dos domínios D₀ e D₁-D₂, respectivamente. Os grupos de resíduos marcados em verde correspondem a regiões livres da imposição de qualquer restrição, ou seja, as características espaciais de 1nkr não serão impostas em DM43 nas regiões correspondentes, permitindo-as adotarem uma conformação que dependerá apenas dos resíduos vinculados e sequencialmente próximos.

A figura VII.1 apresenta o primeiro alinhamento utilizado em MODELLER, derivado da execução do programa MULTALIGN contra os KIRs2, KIRs3 e oprina. Houve ainda a necessidade de intervir no alinhamento através de modificações manuais, baseando-se nos conhecimentos prévios dos motivos estruturais dos KIR2DLs (seções VII.3 e VII.4).

Para a execução do programa foram necessários:

- um arquivo contendo alinhamentos da sequência da proteína alvo com pelo menos uma estrutura, no formato exigido pelo MODELLER;
- um arquivo contendo as coordenadas atômicas da estrutura, no formato pdb;
- um script de execução contendo, por exemplo, indicações dos arquivos de entrada, número de modelos desejados, tipos de otimização, tipos de restrições e parâmetros de refinamento.

Ao final da execução, o programa fornece um arquivo .log listando o desempenho e os valores da Função Objetiva ($-\ln(P)$, $P \rightarrow$ pdf molecular) otimizada, de cada modelo, além de arquivos contendo as coordenadas atômicas destes modelos.

Os parâmetros utilizados nas modelagens de D₀ e D₁-D₂ foram os seguintes:

- Execução completa do Método de Função Variável (VTFM);
- Máximo de 300 iterações para os ciclos do VTMF;
- Otimização completa por Dinâmica Molecular utilizando *simulated annealing*;
- Otimização de todos os átomos no refinamento;
- Inclusão das restrições estereoquímicas e diédricas;
- Aplicação de quatro ciclos de todo o processo de otimização que inclui VTFM e refinamento;
- Valores máximos para restrições de várias distâncias:
 - C α -C α : 14.0 Å

- N - O : 11.0 Å
- Cadeia Lateral - Cadeia Principal: 5.5 Å
- Cadeia Lateral - Cadeia Principal: 5.0 Å

Foram acrescentadas restrições extras no modelo, baseando-se no conhecimento prévio das características estruturais de proteínas da família dos receptores inibitórios da célula assassina (KIRs):

Tabela VII-1: Tabela indicando as restrições especiais acrescentadas no MODELLER.

Restrições	Razão	DM43_D ₀	DM43_D ₁ -D ₂
Prolinas	Redução da flexibilidade na restrição do ângulo diedro Φ das Prolinas, levando-as para uma conformação estereoquimicamente permitida. Isto foi feito para:	Todas as Prolinas	Todas as Prolinas
Ligações de Hidrogênio entre as fitas C-C'	Fixação do alinhamento entre as fitas C e C' com a imposição da formação de ligações de hidrogênio entre:	O:Phe34 e N:Leu46 O:Leu46 e N:Phe34 O:Leu32 e N:Val48	O:Leu34 e N:Val43 O:Val43 e N:Leu34 O:Phe32 e N:Ser45 O:Ser45 e N:Phe32
Serinas do motivo WSXWS box	Imposição da conhecida configuração entre estruturas que possuem o motivo WSXWS, com a criação de duas ligações de hidrogênio entre os seguintes átomos:	O γ :Ser179 e N:Tyr 169 O γ :Ser182 e N:Cys167	O γ :Thr80 e N:Ile72 O γ :Ser83 e N:Cys70

Uma vez que o programa utiliza o alinhamento como referência para derivar as restrições espaciais para impor na proteína alvo (no caso, a DM43) uma pequena modificação poderá melhorar ou não a concordância entre a restrição das características de uma determinada região da estrutura com as da proteína alvo.

Desta forma, diferentes alinhamentos foram testados baseando-se na avaliação anterior dos modelos pela utilização de três parâmetros de qualidade (seção VII.7). A praticidade do MODELLER permitiu a produção de cerca de 500 modelos na taxa de cinco a dez por alinhamento.

O alinhamento que resultou nos melhores modelos está apresentado na figura VII.2. Os resíduos destacados em verde correspondem exatamente às regiões de *loops* que, por pertencer a uma região geralmente não conservada, foram deslocadas de modo que o programa não aplique, neste trecho, qualquer

restrição obtida de 1nkr em DM43. Interessantemente, as regiões mais afastadas de *loops* (no *core*) não necessitaram ser desvinculadas, apesar do baixo grau de identidade, ~28%.

VII.6. Visualização

Os programas TOM v.3.2 (Jones, T. A., 1978), O v.5.10.3 e v.6.2.2 (Jones, T. A. *et al.*, 1991) e GRAFIC/WHATIF (Vriend, 1990) foram utilizados para visualização dos modelos e ajustes manuais de alguns resíduos durante a etapa final de modelagem. Figuras de DM43 e 1nkr foram preparadas em Swiss-PDBViewer v.3.6b2 (Guex e Peitsch, 1996) e figuras *sausage* de DM43 com MOLMOL v.2.3 (Koradi, *et al.*, 1996), ambos executáveis em computadores pessoais e estações Silicon Graphics.

VII.7. Avaliação dos Modelos

As estruturas foram analisadas segundo os seguintes parâmetros de qualidade:

- PROCHECK

A qualidade estereoquímica dos modelos obtidos neste trabalho foi analisada segundo resultados da avaliação do programa PROCHECK v. 2.0 (Laskowski, *et al.*, 1993). Um breve comentário sobre o programa foi apresentado no capítulo VI, seção 2.

O programa compara a estereoquímica dos modelos de DM43 com os parâmetros típicos de estruturas resolvidas a 1.8Å (resolução da estrutura cristalográfica de 1nkr), possibilitando assim análises mais críticas dos modelos.

- QUALITY

A avaliação da qualidade dos modelos segundo contatos atômicos foi feita em WHATIF (Vriend, 1990) utilizando o método antigo do programa QUALITY. É desejado um *score* acima de -1.5.

- **VERIFY_3D**

A qualidade dos modelos segundo seus ambientes químicos foi calculada pelo programa VERIFY_3D v. 2.3 (Luthy, *et al.*, 1992).

VII.8. GRASP

As superfícies de potencial eletrostático dos domínios D_0 e D_1 - D_2 de DM43 foram gerados no programa GRASP v. 1.04 (Nichells e Honig, 1991). A partir dos quais pode-se estimar as possíveis regiões de interação entre DM43-metaloprotease, (D_1 - D_2) - D_0 e DM43 - DM43 (interface dimérica) verificando a existência ou não de grupos carregados distribuídos na sua superfície. Estes resultados sugeriram a existência de prováveis regiões de dimerização e de ligação do domínio D_0 com os domínios D_1 e D_2 , levantando a possibilidade de utilizar o programa PATCHES com o intuito de obter a confirmação das hipóteses sobre as prováveis localizações dos sítios de interação domínio-domínio e proteína-proteína.

VII.9. PATCHES

As regiões de interface entre domínios foram teoricamente identificadas pelo programa PATCHES (Jones, S. e Thornton, J. M., 1996, 1997a, 1997b). Testes de eficiência têm comprovado acertos de no máximo 66% das predições.

O programa foi executado pela Dra. Susan Jones¹ para os domínios D_0 e D_1 - D_2 de DM43 utilizando os parâmetros *defaults* do programa:

- Tamanho do *patch*: 15 resíduos;
- *Cut off*, ou o intervalo (x) em porcentagem dos resultados que caem entre (100-x)% e 100% do melhor *score* obtido: 10%;

¹ Departamento de Bioquímica e Biologia Molecular - *University of College London* - Reino Unido.

- Características consideradas: potencial de solvatação, planaridade, propensão do resíduo à interface, hidrofobicidade, protrusão e acessibilidade ao solvente. As características que obtiveram altos valores foram atribuídas *scores* proporcionalmente altos, com exceção de planaridade que recebeu *scores* altos aos seus menores valores (mais planares).

VIII.

RESULTADOS E DISCUSSÕES

VIII.1. Busca e Análise de Sequências Homólogas a DM43

Pela primeira vez na literatura, um inibidor de metaloprotease de veneno de serpente (de mamífero) tem sua sequência de aminoácidos completamente determinada (Neves-Ferreira, 1999). Isto tornou possível a previsão do modelo estrutural tridimensional de DM43, através do método por satisfação das restrições espaciais que requer, além da sequência de aminoácidos da proteína alvo, a sequência e estrutura de uma proteína homóloga.

Desta forma, foi feita uma busca por sequências de proteínas homólogas a DM43 em bancos de dados. A opção *non-redundant* do programa BLASTP, contendo um número total de 508.464 submissões em cinco bancos de dados (GenBank, SwissProt, PDB, PIR e PRF), identificou os seguintes membros da superfamília das imunoglobulinas, apresentadas em ordem decrescente de similaridade com DM43: oprina, α 1-B-glicoproteína humana, MIRs (*Monocyte/Macrophage Immunoglobulin-like Receptors*) (Wagtmann, *et al.*, 1997),

PIRs (*Paired Ig-like Receptors*) (Kubagawa, Burrows e Cooper, 1997) e KIRs (*Killer Cell Inhibitory Receptors*).

A oprina é, como a DM43, uma glicoproteína inibidora de metaloprotease de veneno de serpente (Catanese e Kress, 1992) isolada do soro do gambá norte-americano *Didelphis virginiana* e sem estrutura conhecida. Apesar da seqüência do domínio D₀ da oprina estar incompleta, DM43 é segundo o programa MULTALIGN 86,08% idêntica a oprina.

Seguindo a oprina, a proteína que apresentou maior identidade seqüencial com DM43 foi a α_1 B-glicoproteína humana (Ishioka, *et al.*, 1986) com 36,71%. O alinhamento interdomínios de DM43 e α_1 B-glicoproteína pelo programa MULTALIGN, resultou em uma porcentagem de identidade de até 39,56% (tabela VIII-1). α_1 B-glicoproteína humana possui cinco domínios do tipo imunoglobulina, sem estrutura e função conhecidas.

Tabela VIII-1: Tabela de identidade interdomínios de DM43 contra α_1 B-glicoproteína humana. As células em destaque indicam os domínios de DM43 que apresentaram a maior porcentagem de identidade com os domínios de α_1 B-glicoproteína.

% de identidade		α_1 B-glicoproteína humana				
		D ₀	D ₁	D ₂	D ₃	D ₄
DM43	D ₀	32,98	24,72	29,21	16,85	23,96
	D ₁	22,22	39,56	23,86	32,93	29,67
	D ₂	28,57	20,65	34,04	20,22	24,74

MIRs e PIRs são membros da superfamília das imunoglobulinas relacionados a KIRs, mas sem estruturas conhecidas.

Os KIRs correspondem a proteínas inibidoras da ação das células assassinas contra as células deficientes nas proteínas HLA (*Human Leukocyte Antigens*) expostas em sua superfície. A figura VIII-1 ilustra a participação das células assassinas no sistema de defesa do organismo.

Figura VIII-1: Representação esquemática do processo de interação dos KIRs com HLAs. As células assassinas ou *Killer Cells* responde por um dos dois meios de defesa do organismo através de proteínas presentes na superfície da célula do hospedeiro, HLA. Por meio destas moléculas, as células assassinas e as T citotóxicas conseguem identificar e eliminar a célula infectada. Os KIRs atuam como reguladores da ação das células assassinas identificando ou não a presença de HLA em células saudáveis. Em células infectadas, os KIRs não são capazes de detectar HLA, permitindo que as células assassinas entrem em ação.

Diferentemente das células citotóxicas, as células assassinas atuam “naturalmente” com a secreção de substâncias citotóxicas (perforinas), sem necessitar de um contato prévio com o anticorpo, ou seja, não possui memória imunológica. A ação desenfreada das células assassinas é controlada pelos receptores localizados em sua superfície, através do reconhecimento das moléculas HLA na célula do hospedeiro (Lenvinson e Jawetz, 1996). Estes receptores são denominados de receptores inibitórios das células assassinas, KIRs (*Killer Cells Inhibitory Receptors*), classificados segundo suas massas moleculares em p58 e p70, ou KIR2DL e KIR3DL, respectivamente.

Uma busca mais direcionada à seqüências que possuem estruturas conhecidas, ou seja, ao banco de dados PDB, resultou em KIR2DLs constituídos de apenas dois domínios tipo imunoglobulina. Suas estruturas tridimensionais são denominadas de KIR2DL1 (1nkr), KIR2DL2 (2dl2) e KIR2DL3 (1b6u), e apresentam identidades seqüenciais com os dois domínios C-terminais de DM43_D₁-D₂ menores que 28%.

A distribuição extracelular dos dois domínios tipo imunoglobulina de KIR2DL (D₁ e D₂), assemelha-se a KIR3DL diferindo apenas de um domínio N-terminal extra e, portanto, denominado de D₀. Por serem proteínas de superfície, possuem ainda porções transmembrana, intracelular e haste, que conecta o domínio

transmembrana com o domínio extracelular D₂, além de um domínio sinalizador (figura VIII-2).

Figura VIII-2: Distribuição dos domínios KIRs sugerindo a dos três domínios de DM43 e Oprina. Por enquanto, não se sabe muito a cerca da estrutura da oprina. Um alinhamento da sequência incompleta de oprina contra DM43 apresentou alto grau de identidade. A suposição da existência de três domínios em oprina é incerta, uma vez que a sua massa molecular (52 KDa) e o conteúdo de pontes dissulfeto sugerem a presença de quatro domínios (Catanese e Kress, 1992). A nomenclatura dos domínios é apenas representativo, não implicando na sua similaridade.

Devido ao caráter solúvel de DM43 e a similaridade com os domínios tipo imunoglobulina de KIRs, supõe-se uma distribuição de domínios como os das regiões extracelulares de KIR3DL (figura VIII-2).

Tabela VIII-2: Porcentagem de identidade entre DM43 e os membros da família KIR2DL (2 domínios) e KIR3DL (3 domínios). Contra KIR2DLs foram alinhados DM43_D₁-D₂.

% de identidade	KIR2DL1 (1nkr.pdb) NKAT1	KIR2DL2 (2dl2.pdb) -	KIR2DL3 (1b6u.pdb) NKAT2	KIR3DL1 NKAT3	KIR3DL2 NKAT4
DM43	27,72	26,49	27,03	25,80	25,26

Tabela VIII-3: Porcentagem de identidade média interdomínios entre os DM43 e os dos membros da família KIR2 e KIR3. As células em destaque indicam os domínios que apresentaram o grau de identidade que melhor ajustará ao alinhamento de DM43 para modelagem.

% de identidade	D ₀ <KIR3DL> -	D ₁ <KIR3DL> <KIR2DL>	D ₂ <KIR3DL> <KIR2DL>
DM43_D ₀	22,64 -	18,62 25,46	25,18 28,05
DM43_D ₁	25,82 -	29,12 27,20	26,92 28,68
DM43_D ₂	24,21 -	21,87 19,26	25,00 25,94

As porcentagens de identidade entre as seqüências completas de DM43 e KIRs, e os domínios separados de DM43 contra os dos membros da família KIR, estão apresentadas nas tabelas VIII-2 e VIII-3, respectivamente.

O baixo grau de identidade entre DM43, α 1B-glicoproteína humana e KIRs, não poderiam garantir a DM43 o enovelamento do tipo imunoglobulina, ou a localização mais precisa da conformação das fitas betas quanto a sua topologia. Sabendo-se que bons modelos são obtidos a partir de alinhamentos confiáveis, buscou-se em THREADER a confirmação do enovelamento imunoglobulina de DM43.

A execução de THREADER a partir da seqüência de DM43, separada em três domínios, resultou basicamente em: folhas betas com arquitetura do tipo sanduíche; topologia tipo imunoglobulina; e homologia com 1hnf domínio 1, uma glicoproteína de adesão do linfócito T (estrutura contida no banco de dados do programa THREADER). Estas características confirmaram o enovelamento tipo imunoglobulina para todos os três domínios de DM43.

A fim de se chegar no modelo da estrutura tridimensional de DM43, o passo seguinte consistiu na escolha de uma ou mais estruturas disponíveis, que são: KIR2DL1 (1nkr.pdb) (Fan *et al.*, 1997), KIR2DL2 (2dl2.pdb) (Snyder, Brooks e Sun, 1999) e KIR2DL3 (1b6u.pdb) (Maenaka *et al.*, 1999). Este passo foi imprescindível para que o programa MODELLER pudesse, a partir delas, derivar as restrições espaciais que serão impostas na seqüência de DM43.

Deste modo, foram feitas avaliações quanto a resolução dos dados, a qualidade do refinamento, estereoquímica, contatos atômicos e ambiente químico de cada uma das estruturas disponíveis. Isto corresponde a uma etapa importante da modelagem por satisfação das restrições espaciais, pois a qualidade e homologia da estrutura de referência influem muito na confiabilidade do modelo. De forma geral, é desejado um número razoável de estruturas para contribuir em uma estatística mais precisa para calcular a Função Objetiva (F) (ver capítulo VI) e assim impor os valores de suas características espaciais na proteína por MODELLER. Entretanto, a avaliação apresentada a seguir na tabela VIII-4 revela que, das três estruturas disponíveis, 1nkr possui a estrutura mais confiável em termos de resolução, qualidade de refinamento e em todos os três índices de qualidade (Procheck, Quality e Verify_3D). Em função destes resultados, o 1nkr (ou

KIR2DL1) foi selecionado como a estrutura de referência para a modelagem de DM43, que em particular possui uma resolução muito melhor que as outras estruturas.

Tabela VIII-4: Avaliação das três estruturas homólogas a DM43. O fator-G do PROCHECK indica a qualidade esteroquímica geral da estrutura, onde índices superiores a -0,5 são desejados. Para estruturas, o índice de WHATIF é ideal acima de -1,0. Em VERIFY_3D, quanto mais próximo o primeiro número for do segundo, mais a estrutura em questão estará próxima do *score* médio de estruturas com o mesmo número de resíduos.

	Resolução	R	R _{free}	Fator G (Procheck)	Identidade sequencial com DM43	Qualidade segundo WHATIF (Old-Quality)	Qualidade segundo Verify_3D
1nkr (KIR2DL1)	1,7 Å	20,5	25,5	0,12	27,72%	-0,54	82,5 / 88,7
2dl2 (KIR2DL2)	3,0 Å	21,9	30,9	-0,20	26,49%	-1,53	86,6 / 89,6
1b6u (KIR2DL3)	3,0 Å	24,8	32,0	0,00	27,03%	-1,06	77,9 / 90,5

Como visto na tabela VIII-3, DM43_D₁ e DM43_D₂ apresentaram, respectivamente, uma porcentagem de identidade média maior com os domínios D₁ e D₂ de KIRs, enquanto que o domínio D₀ de DM43 (ou DM43_D₀), mostrou-se mais idêntico ao domínio D₂ de KIRs.

Por existirem apenas estruturas de KIRs contendo dois domínios, a comparação interdomínios de DM43 e KIRs permitiu definir qual domínio de 1nkr será a estrutura de referência para um determinado domínio de DM43. Desta forma, dividiu-se a modelagem em duas partes, sendo: a modelagem do domínio D₀ de DM43 a partir do domínio D₂ de 1nkr; e a modelagem dos domínios ligados D₁ e D₂ a partir da estrutura completa de 1nkr, compreendendo os domínios D₁ e D₂.

Um alinhamento geral englobando os KIRs, oprina e DM43, foi construído através do programa MULTALIGN a fim de aproveitar todas as informações estruturais primárias das seqüências homólogas a DM43. Isto possibilitou uma melhor análise da seqüência de DM43 quanto a substituições, deleções e inserções. A localização mais precisa da estrutura secundária no alinhamento foi obtida segundo a estrutura de 1nkr através do programa PROMOTIF. O próximo

passo consistiu na análise dos motivos estruturais dos KIRs com o intuito inicial de encontrar o melhor alinhamento de DM43 e KIRs (figura VIII-3).

Desta forma, as estruturas cristalográficas e seqüências dos KIR2DLs, revelaram possuir características semelhantes aos dos membros da família dos receptores hematopoiéticos. Essencialmente, pela presença de um motivo *WSXWS box* precedendo imediatamente a fita G do enovelamento imunoglobulina, e da conseqüente orientação relativa aos dois domínios. O segundo triptofano do motivo não costuma ser estritamente conservado podendo, às vezes, ser substituído por outros resíduos. Não se sabe ao certo a razão da existência deste motivo e a conservação das posições das duas serinas. Entretanto, levanta-se a hipótese de que a presença do triptofano na primeira posição do motivo seja responsável por uma redução no ângulo relativo aos dois domínios D_1 e D_2 quando comparado com imunoglobulinas, que costumam apresentar ângulos que variam entre 135° a 175° .

Apesar do baixo grau de identidade entre DM43 e KIRs, o alinhamento da figura VIII-3 apresenta poucas inserções ou deleções mantendo-se conservados os pares de cisteínas em todos os domínios tipo imunoglobulina de DM43, oprina, KIR3DLs e KIR2DLs. Observa-se também a presença do motivo *WSXWS box* em todos os domínios de DM43, além de um sítio de glicosilação no domínio DM43_ D_0 e mais três sítios no domínio DM43_ D_1 . Interessantemente, o motivo *WSXWS box* do domínio D_2 é o que melhor conserva o motivo com a presença de um triptofano na primeira posição. Contrariamente, no domínio D_1 , este triptofano é substituído por uma asparagina e a posição X por uma treonina, satisfazendo a condição para a existência de um sítio de glicosilação. As serinas mantêm-se quase totalmente conservadas com exceção do motivo do domínio (D_0) que, em DM43, é substituída por uma treonina.

A DM43 possui ainda os 4 resíduos padrões do motivo estrutural das superfamílias das imunoglobulinas: cisteínas das fitas B e F formando uma ponte dissulfeto; tirosina posicionada a dois resíduos anteriores a cisteína em F; e presença de um resíduo aromático no início da fita C posicionado sobre a ponte dissulfeto.

Figura VIII-3: Alinhamento para auxiliar na derivação do melhor alinhamento de DM43 contra 1nkr (KIR2DL1). Lembrando que KIR2DL2 = 2dl2 e KIR2DL3 = 1b6u. As pontes dissulfetos estão identificadas em amarelo. As caixas verdes destacam o motivo WSXWS, presente em todos os domínios de KIRs, oprina e DM43. CHO identifica a localização do sítio de glicosilação de DM43. As setas verticais pretas, delimitam os domínios.

VIII.2. Modelagem

Os resultados do MODELLER foram analisados segundo o valor mínimo da Função Objetiva, melhoria da qualidade estereoquímica, contato átomo-átomo e ambiente químico.

As restrições de 1nkr impostas em DM43 também não resultaram sempre em uma conformação local razoável, devido ao baixo grau de identidade com 1nkr. Nestas regiões, houve necessidade de intervenção manual no modelo final de DM43. Antes disto, tentou-se ao máximo melhorar a qualidade dos modelos através do próprio programa MODELLER, deslocando o alinhamento de maneira que um determinado trecho sofresse ou não as restrições derivadas de 1nkr. Houve ainda a necessidade de adicionar restrições sobre um ou mais resíduos específicos.

O alinhamento de DM43 contra 1nkr apresentado no capítulo anterior (figura VII-2), indica as posições que necessitaram ser liberadas das restrições de 1nkr. Uma vez que o alinhamento é desfeito, perde-se o vínculo das características espaciais de DM43 com 1nkr. Todos os trechos desvinculados corresponderam a regiões de *loops*, onde normalmente ocorrem as variações conformacionais da proteína.

Os modelos produzidos por MODELLER resultaram em uma topologia do tipo Ig-I (Clarke *et al.*, 1999) nos domínios DM43_D₁ e DM43_D₂ (figura VIII-4), e um misto de Ig-I com Ig-s (Bork, Holm e Sander, 1994) na conformação do domínio DM43_D₀. Particularmente, as fitas C' e D dos domínios DM43_D₁ e DM43_D₂ são, respectivamente, maior e menor que o esquematizado na figura VIII-4(a). O enovelamento do tipo Ig-I caracteriza-se por possuir uma fita C' estendida na segunda folha como uma fita D, e a fita A dividida em A e A', sendo que cada uma delas fica pareada com uma folha diferente (figura VIII-4(a)).

Figura VIII-4: Topologia bidimensional de dois enovelamentos tipo imunoglobulina. O enovelamento imunoglobulina do tipo I é apresentado pelos domínios DM43_D₁-D₂, enquanto que o domínio DM43_D₀ é um misto de imunoglobulina-I e s, sem a presença da fita D.

A ligação entre as fitas A e A' ocorre por um grampo de cabelo β cuja ponta, no final da fita A, é ocupada por um resíduo na conformação *cis*-prolina. Isto corresponde a uma característica específica das estruturas KIRs (Snyder, Brooks e Sun, 1999).

As restrições impostas por 1nkr em DM43 nestas regiões, forçaram uma ligação peptídica do tipo *cis* para Thr14, Pro107 e Asn199, nos domínios D₀, D₁ e D₂, respectivamente, todas posicionadas no final da fita A. Foi necessária uma intervenção manual para mudar a conformação *trans* de Asn199 e Thr14 para *cis* melhorando, com isto, sua estereoquímica em termos dos ângulos Φ e Ψ .

Nos primeiros modelos, todas as prolinas de DM43 estavam com seu ângulo diedro Φ afastado da conformação usual de $\sim -65^\circ$. Isto foi facilmente resolvido restringindo a variância de Φ para 10% do valor original (figura VIII-5).

Figura VIII-5: Diagramas de Ramachandran específicas para prolinas, obtidas através do programa PROCHECK. Ramachandran das prolinas do modelo de DM43_D₀ (a) e DM43_D₁-D₂ (b) após restrição da variância do ângulo Φ de todas as prolinas. O valor entre parênteses indica o número de prolinas existentes no modelo.

Houve ainda uma certa dificuldade do programa em criar vínculos entre as fitas C e C' dos três domínios predizendo uma região para DM43, estereoquimicamente desfavorável. Isto foi melhorado com a imposição de ligações de hidrogênio nestas fitas, forçando C' incorporar na folha A'GFC.

Finalmente, uma das principais restrições impostas nos modelos de DM43, refere-se a formação de ligações de hidrogênio entre as serinas do motivo *WSXWS box* com a cadeia principal da fita F. As serinas possuem um papel importante na estrutura de um modo geral, devido a sua estrita conservação em DM43, KIRs e em receptores da família hematopoiética.

Acredita-se que as ligações de hidrogênio pelas serinas forcem os resíduos freqüentemente hidrofóbicos nas posições dos triptofanos, a orientarem-se de maneira a formar um *core* hidrofóbico na interface intradomínios e, portanto, forçando uma orientação angular aguda, em torno de 90° ou menor. Isto é visto nos receptores de hormônios de crescimento (90°), eritropoietina (que também pertence à superfamília dos receptores hematopoiéticos, com orientação intradomínio de 90°) e nos KIRs, mas diferem das demais estruturas pertencentes à superfamília das imunoglobulinas ($>90^\circ$) (figura VIII-6).

Figura VIII-6: Orientação intradomínios de (a) imunoglobulinas, (b) receptor do hormônio de crescimento e prolactina, os KIRs 1nkr (c), 2dl2 (d) e 1b6u (e).

A partir do melhor alinhamento e da imposição de todas as restrições descritas até então, dez modelos foram produzidos por MODELLER. Os modelos

sobrepostos de DM43_D₀ e DM43_D₁-D₂, estão respectivamente ilustrados nas figuras VIII-7 e VIII-8.

Figura VIII-7: Sobreposição dos dez modelos de DM43_D₀ no traçado da cadeia principal (a) e a representação estereográfica em sausicha (b). As regiões em laranja correspondem a R.M.S.D. superiores a 1,0Å.

Figura VIII-8: Sobreposição dos dez modelos de DM43_D₁-D₂ representados pelo traçado da cadeia principal (a) e, na mesma orientação, os mesmos dez modelos no estilo sausicha. As regiões em laranja representam as variações relativas superiores a 1,0 Å (b).

As regiões de maior variabilidade ou imprecisão na construção dos modelos correspondem àquelas que sofreram inserções e foram desvinculadas das restrições de 1nkr (ver alinhamento na figura IV-2). Claramente, isto é devido ao número reduzido de restrições para tais regiões.

VIII.3. Avaliação dos Modelos

Na primeira avaliação, escolheram-se modelos que simultaneamente resultaram nos menores valores da função objetiva (equação 32 - cap.VI) e nos que menos tiveram sua qualidade estereoquímica comprometida, segundo PROCHECK. Os valores da função objetiva para os dez modelos de DM43_D₀ e DM43_D₁-D₂ estão apresentados na tabela VIII-5. Em destaque, estão os modelos escolhidos para derivar a partir de ajustes manuais, os modelos finais. Suas funções objetivas são comparadas com as das estruturas de referência, 1nkr (1nkr_D₂ e 1nkr_D₁-D₂), na tabela VIII-6. Diferentemente da tabela anterior, foram consideradas apenas as propriedades estereoquímicas de DM43 e de 1nkr nos cálculos de suas Funções Objetivas. Estes números indicam que a qualidade estereoquímica dos modelos é comparável com a estrutura cristalográfica .

Tabela VIII-5: Valores da função objetiva dos dez modelos de saída do MODELLER. As células em destaque correspondem aos modelos iniciais de DM43_D₀ e DM43_D₁-D₂ escolhidos para modelar manualmente para se chegar na estrutura final. A tabela está organizada em ordem crescente dos valores da função objetiva.

Valores da Função Objetiva dos vários modelos de DM43			
nº do modelo	D ₀	nº do modelo	D ₁ .D ₂
5	338,58	6	992,16
2	348,04	2	999,28
3	349,69	4	1016,6
8	356,34	1	1024,7
6	357,58	5	1033,5
9	359,35	7	1059,2
1	360,50	9	1113,3
7	364,25	3	1159,1
4	386,76	8	1184,9
10	413,13	10	1259,8

Tabela VIII-6: Comparação das funções objetivas da estrutura de referência 1nkr e do modelo final DM43. Neste caso, as funções objetivas derivam apenas das restrições estereoquímicas das estruturas.

Valores finais da Função Objetiva de DM43			
1nkr_D ₂	826,72	1nkr_D ₁ -D ₂	1684,7
DM43_D ₀	515,37	DM43_D ₁ -D ₂	1982,8

A avaliação da estereoquímica da estrutura quanto a conformação da cadeia principal de DM43_D₀ apresentou 77,6% dos resíduos na conformação mais favorável (região demarcada em vermelho), 15,8% na conformação razoavelmente favorável (laranja), 6,6% na região generosamente permitida (amarelo) e 0,0% nas regiões desfavoráveis. Em termos da conformação da cadeia principal, este resultado é bastante razoável, lembrando que DM43_D₀ e 1nkr_D₂ estão relacionados por um baixo grau de identidade seqüencial.

Os cinco resíduos localizados na região generosamente permitida foram identificados pelo programa: Trp21, Glu34, Leu46, Asn81 e Asp82. Todos estes resíduos, com exceção do Trp21, pertencem a regiões de *loops* onde há inserções ou deleções no alinhamento com 1nkr_D₂ (figura VII-1). O Trp21 encontra-se em uma região fechada, distante de *gaps*. Corresponde a uma substituição de uma glicina em 1nkr_D₂ no *loop* que conecta as fitas A' e B. Sua localização, no diagrama de Ramachandran, coincide com a região de hélice alfa de mão esquerda, freqüentemente ocupada de forma quase exclusiva por glicinas. Um pequeno ajuste manual poderia levar o Trp21 à região permitida, em torno de $\Phi=60^\circ$ e $\Psi=50^\circ$. Esta região poderia ainda ser remodelada liberando as restrições da glicina de 1nkr sobre o triptofano substituinte. Esta alternativa não foi experimentada, pois esta não era uma região crítica para os propósitos do modelo.

Figura VIII-10: Diagrama de Ramachandran para o modelo final de DM43_D1-D2.

A figura VIII-10 apresenta o diagrama de Ramachandran para o modelo de DM43_D1-D2. Para estes domínios, o modelo apresentou um número menor de resíduos mal posicionados, possivelmente devido a um menor número de deleções e inserções com relação ao domínio DM43_D0.

Segundo as estatísticas do PROCHECK: 83,5% estão nas regiões favoráveis do diagrama de Ramachandran; 15%, nas regiões parcialmente favoráveis; 1,3%, nas quase não favoráveis; e 0,0% nas regiões não permitidas. Os 1,3% correspondem à Met246 e Lys258 identificadas no diagrama. Novamente,

estes resultados são bem razoáveis dada a baixa identidade seqüencial de DM43_D₁-D₂ com 1nkr.

A Met246 substitui uma asparagina que é seguida de uma glicina em 1nkr. Ambos definem uma volta β do tipo I', ou seja, adotam uma conformação de hélice de mão esquerda, não permitida para metionina. Apesar de retirá-la da região de mão esquerda, a solução não foi inteiramente satisfatória.

A Lys256 localiza-se no *loop* que une as duas folhas betas, entre as fitas E e F. Devido a deleção desta região no alinhamento, houve dificuldade em encontrar uma conformação melhor.

Quanto às outras deleções encontradas no alinhamento (figura VIII-3), foram possíveis ajustar suas conformações no próprio programa MODELLER (figura VI-1), ou manualmente em TOM, O ou WHATIF, de forma a adotarem uma estereoquímica satisfatória.

Um dos principais parâmetros do PROCHECK, o fator-G pode ser utilizado como um *score* que gradua a qualidade geral do modelo segundo os ângulos diedros, distâncias de ligações covalentes e a estrutura, de um modo geral. São consideradas boas estruturas, segundo o programa, aquelas que possuem um fator G acima de -0,5. Desta forma, a qualidade estereoquímica de DM43_D₀ e DM43_D₁-D₂ estão bem qualificadas, possuindo um fator-G de -0,39 e 0,27, respectivamente.

As figuras a seguir apresentam *scores* individuais por resíduo, com relação aos vários ângulos diedros, ligações e um *score*, em especial, considerando todas as variáveis. Uma cor mais avermelhada indica uma conformação não favorável ao resíduo em particular.

Figura VIII-11: Fator-G por resíduo segundo ângulos e distâncias do domínio DM43_D₀. À direita, é apresentada a pontuação média de todos os resíduos para cada característica. A figura indica também a existência de um resíduo na conformação *cis*, pertencente ao início do *link* entre as fitas A e A', característico das estruturas KIR.

Figura VIII-12: Fator-G por resíduo segundo ângulos e distâncias dos dois domínios de DM43_D₁-D₂. é apresentada a pontuação média de todos os resíduos para cada característica. A figura indica também a existência dois resíduos na conformação *cis*, pertencentes ao início do *link* entre as fitas A e A', característico das estruturas KIR.

As figuras acima identificam com a letra c os resíduos com conformação *cis*. Esta conformação provém das restrições de 1nkr para as posições do primeiro resíduo que une as fitas A e A'. A conformação *cis* foi naturalmente adotada pela Pro107 em DM43_D₁-D₂, enquanto que a Asn199 e a Thr14 (DM43_D₁) tiveram que ser, respectivamente, transformadas em *cis* utilizando os métodos manual (TOM) e automático (MODELLER).

Vale enfatizar que esta conformação distingue a topologia dos KIRs com relação ao restante das estruturas dos receptores hematopoiéticos, adotando o enovelamento imunoglobulina do tipo I (figura VIII-4(a)).

Apesar da Asn e principalmente da Thr não serem comumente encontrados na conformação *cis*, como as Pro, uma tentativa de remodelar esta região na conformação *trans* poderia, sistematicamente, produzir valores não permitidos de Φ e Ψ .

Ajustes mais detalhados da estrutura foram possíveis após identificar os resíduos que possuíam os piores contatos atômicos (*scores* < -3.0) de acordo com OLD-QUALITY (programa WHATIF). Os gráficos a seguir ilustram a variação por resíduo, dos modelos finais, sobre a qualidade de seus contatos atômicos segundo OLD-QUALITY/WHATIF.

Figura VIII-13: Gráfico apresentando a variação da qualidade do contato atômico de cada resíduo segundo OLD-QUALITY de DM43_D₀. A média dos valores contidos neste gráfico é igual a -1,342, situando DM43_D₀ na região de um modelo normal a muito bom.

Figura VIII-14: Gráfico apresentando a variação da qualidade do contato atômico de cada resíduo segundo OLD-QUALITY de DM43_D_{1-D2}. A média dos valores contidos neste gráfico é igual a -0,925, situando DM43_D_{1-D2} na região de um modelo muito bom.

O valor médio destes scores fornece uma idéia da qualidade do modelo como um todo. Desta forma, o score médio de DM43_D₀ é igual a -1,342 e equivale, segundo o autor do programa, a um modelo bom ou muito bom, enquanto que o score -0,925 de DM43_D_{1-D2} é típico de estruturas muito bem modeladas. Desta forma, as estruturas modeladas de DM43 parecem ser de boa qualidade em termos de estrutura terciária.

A avaliação de acordo com o ambiente químico (programa VERIFY_3D) revelou regiões cuja qualidade decresce acentuadamente (figura VIII-15 e VIII-16) coincidindo com àquelas já detectadas por PROCHECK, como por exemplo: a região 40-50 de DM43_D₀ onde há uma deleção no *loop* entre as fitas C e E com relação a 1nkr; a deleção de um resíduo no *loop* que conecta as fitas C e C' (136-

138) do domínio D_1 ; e a região 240-250 que compreendem a volta β do tipo I' onde há uma substituição de dois resíduos com conformação tipo hélice alfa de mão esquerda por dois outros resíduos raros em adotar um ângulo diedro Φ positivo.

A avaliação geral dos modelos por VERIFY_3D fornece o valor de qualidade igual a 25,81 para DM43_D₀ e 59,77 para DM43_D₁-D₂, classificando-os dentro da faixa das estruturas de boa qualidade. O valor mínimo necessário para considerar um modelo razoável é de 19,13 e 38,88 para proteínas de comprimentos iguais a de DM43_D₀ e de DM43_D₁-D₂, respectivamente.

Os gráficos de qualidade por resíduo são apresentados a seguir (figura VIII-15 e VIII-16). Os scores médios 3D-1D correspondem a scores promediados dentro de uma janela de 21 resíduos. Valores negativos indicam uma região da estrutura com ambientes "incomuns" e, portanto, possivelmente incorreta.

Figura VIII-15: Distribuição do score de qualidade de ambiente químico por resíduo segundo VERIFY_3D. Valores acima de zero correspondem a ambientes favoráveis.

Figura VIII-16: Distribuição do score de qualidade de ambiente químico por resíduo segundo VERIFY_3D. Valores acima de zero correspondem a ambientes favoráveis.

Após as avaliações e ajustes das regiões mais críticas dos modelos DM43_D₀ e DM43_D₁-D₂, chegou-se nas seguintes estruturas (figura VIII-17):

Figura VIII-17: Apresentação dos modelos finais de DM43. As fitas estão identificadas segundo convenção de enovelamento tipo imunoglobulina.

onde a identificação das fitas β foi feita segundo a convenção do enovelamento de proteínas tipo imunoglobulina I e s (figura VIII-4). Estas estruturas finais modeladas de DM43 estão representadas no estilo *ribbons* que permite distinguir facilmente as várias estruturas secundárias, neste caso, os *loops* (conexões em cinza) e as fitas betas (setas azuis).

Uma comparação com a estrutura de referência, 1nkr, é feita através da sobreposição e cálculo do RMSD. A sobreposição das estruturas de 1nkr e do modelo final de DM43 estão apresentadas nas figuras estereográficas a seguir (VIII-18 e VIII-19). O RMSD de DM43_D₀ e DM43_D₁-D₂ segundo o comando "lsq_improve" do programa O são, respectivamente, iguais a 1,872Å e 1,297Å. Este valor se deve, provavelmente, a grande variação estrutural entre os modelos de DM43 nos *loops*, regiões difíceis de modelar devido a falta de restrições. (figura VIII-7 e VIII-8).

Figura VIII-18: Visão estereográfica da sobreposição do modelo final de DM43_D₀ (em azul) e da estrutura 1nkr_D₂ (em branco). O valor do RMSD nos C α , segundo o programa O, foi de 1,872Å.

Figura VIII-19: Visão estereográfica da sobreposição do modelo final de DM43_D₁-D₂ (em azul) e da estrutura 1nkr (em branco). O valor do RMSD nos C α foi de 1,297Å, segundo programa O.

VIII.4. Interpretação dos Modelos

Apesar da estrutura de referência possuir baixa identidade seqüencial com relação a DM43, foram obtidos bons modelos segundo as avaliações feitas na seção anterior. No modelo de DM43_D₁-D₂, o motivo WSXWS aparentou manter sua posição com o primeiro triptofano do motivo estrategicamente posicionado na interface entre os domínios (figura VIII-20). A imposição das ligações de hidrogênio pelas serinas do motivo foi obedecida sem necessitar levar qualquer resíduo vizinho a uma conformação estericamente desfavorável. Entretanto, a distribuição “escada” de resíduos carregados e hidrofóbicos (+ ϕ + ϕ + ϕ +) na região de WSXWS, não foi observada, como em receptores de hormônio de crescimento e prolactina, membro da família dos receptores hematopoiéticos (Somers, *et al.*, 1994).

Figura VIII-20: Detalhe da região WSXWS *box* de DM43_D₁-D₂ (em amarelo) e sua localização relativa entre os domínios. As distâncias das ligações de hidrogênio são, a partir do triptofano, iguais a 2,91Å e 2,83Å.

Curiosamente, uma fita extra não pertencente a classe de enovelamentos do tipo imunoglobulina surge imediatamente antes do motivo WSXWS. Inclusive o primeiro triptofano do motivo faz parte desta fita da qual é iniciada por uma prolina completamente conservada em todas as seqüências de KIRs. Esta prolina desempenha um papel importante nesta posição, pois estabelece uma dobra brusca no *loop* que conecta as fitas F e G do domínio D₂, possibilitando a formação de três ligações de hidrogênio com a cadeia principal da fita G do domínio D₁. Esta nova fita denominada de F', aparentemente reforça o papel das serinas e do triptofano do WSXWS *box*, em manter a orientação relativa $<90^\circ$ dos dois domínios DM43_D₁ e DM43_D₂ (figura VIII-21).

Figura VIII-21: Representação da junção entre os domínios DM43_D₁ e DM43_D₂ explicitando a formação de três ligações de hidrogênio com a fita G do primeiro domínio favorecido pela presença da prolina 273. A nova fita denominada de F' não pertence à classe das estruturas de imunoglobulinas. Sua conformação se estende para a fita G com a ajuda das serinas do motivo WSXWS.

No alinhamento, nota-se que a conservação do motivo WSXWS no domínio D₂ é maior que nos outros domínios de DM43, em termos das duas serinas e do primeiro triptofano. Uma conservação semelhante em DM43_D₁ forçaria DM43_D₀ a uma orientação relativa a DM43_D₁, similar ao ângulo adotado por DM43_D₂ e DM43_D₁. Este tipo de conformação levaria a região N-terminal do domínio DM43_D₀ chocar-se com a porção C-terminal de DM43_D₂, produzindo um impedimento estérico (figura VIII-22).

Figura VIII-22: Orientação esquemática dos três domínios de DM43, dada a presença do triptofano (em vermelho) do motivo WSXWS *box* e os quatro sítios de glicosilação evitando a criação de impedimentos estéricos entre os domínios D₀ e D₂ de DM43. A representação vazada do domínio D₀ simula a sua provável orientação caso não houvesse os três sítios de glicosilação inclusive no WSXWS *box* do domínio D₁.

Com relação a orientação relativa dos outros domínios, observa-se que o triptofano do motivo WSXWS em DM43_D₁ é substituído por um sítio de glicosilação, dado por 'NSTPS'. Este motivo, equivalentemente a DM43_D₂, participaria da interface dos domínios D₀ e D₁. Como a asparagina substitui a posição do triptofano, a cadeia de carboidratos ocuparia a região de interface forçando o domínio DM43_D₀ a uma orientação relativa à DM43_D₁, em um ângulo provavelmente maior. Reforçando esta hipótese, existem mais dois sítios de glicosilação que estariam voltados para esta mesma região caso DM43_D₀ adotasse a conformação ilustrada na figura VIII-22 (domínio representado apenas pelo contorno em azul), que são: o único sítio de DM43_D₀ e o segundo sítio de glicosilação de DM43_D₁ (figura VIII-22 e VIII-23).

Figura VIII-23: Localização de três sítios de glicosilação (em vermelho) em D₁, com relação à região do motivo WSXWS, e um quarto sítio em D₀.

Quanto ao motivo WSXWS *box* em DM43_D₀ (figura VIII-24), é de se supor que não haja uma conservação tão rigorosa quanto em DM43_D₂, justamente por não haver significado estrutural algum em termos de orientação domínio-domínio. As ligações de hidrogênio são formadas por uma treonina e uma serina, ao invés de um par de serinas. O padrão escada de resíduos hidrofóbicos e carregados (+ ϕ + ϕ + ϕ +) na região de WSXWS, também não foi observado em DM43_D₀.

Figura VIII-24: Detalhe da região WSXWS *box* de DM43_D₀ e sua localização relativa no domínio. As distâncias das ligações de hidrogênio são, a partir do triptofano, iguais a 2,82Å e 2,80Å

VIII.5. As Prováveis Regiões de Interação de DM43

O modelo de DM43 pode ser utilizado para identificar as regiões de interação com a metaloprotease que, aliás, corresponde a um dos principais objetivos desta modelagem. A identificação destas regiões pode dar indicativas para experimentos futuros como, por exemplo, a mutagênese de sítio dirigido.

Certas regiões prováveis de estarem relacionadas com a interação da metaloprotease, não puderam ser muito bem previstas, pois correspondem exatamente às regiões de maior variação local, em termos de quantidade e tipo de resíduo.

Tais prováveis regiões de interação foram deduzidas em DM43 a partir da análise estrutural do complexo de receptores de hormônio de crescimento e prolactina, onde Somers *et al.* (1994) identificaram seis regiões distribuídas entre os dois domínios do receptor (figura VIII-25).

Figura VIII-25: Representação de um complexo hormônio de crescimento (em vermelho) e receptor de prolactina (em azul) destacando a região do motivo WSXWS em verde. As regiões do receptor envolvidas na interação estão nos *loops* que conectam as fitas: AB, CD, EG (domínio 1), GA', B'C' e F'G' (domínio 2).

Extendendo esta análise para o caso de DM43, as prováveis regiões de interação compreenderiam os seguintes intervalos de seqüências (tabela VIII-7):

Tabela VIII-7: Especificação dos prováveis resíduos envolvidos na interação com a metaloprotease.

D ₁	A'B	Pro113 - Gly114 - Leu115 - Glu116
	CC'	Gly136 - Glu137 - Gln138
	EF	Asn156 - Ser157 - Thr158
LINKER (dos domínios D ₁ - D ₂)		Asp188 - Phe189 - Leu190 - Pro191 - Lys192 - Ala193
D ₂	BC	Arg 116 - Phe117 - Ser118 - Glu119 - Arg120 - Glu121 - Tyr122 - Asp123 - Leu124
	FG	Arg270 - Asn271 - Gly272 - Pro273 - Pro274 - Ile275 - Trp276 - Ser277 - Glu278 - Asp279 - Ser280 -Asn281 - Val282

As localizações destes resíduos estão indicadas na figura VIII-27:

Figura VIII-26: Representação no modelo DM43_D1-D2 das prováveis regiões de interação com a metaloprotease.

Quanto a interação com proteínas de mesma espécie, Fan *et al.* (1997) levanta a hipótese de dimerização de KIR2DL na proporção de 2 KIR2DL para 1 HLA-C (proteína envolvida no sistema imune, localizada nas superfícies das células) através da interação entre as faces “externas” dos domínios D₂ (figura VIII-27). Cita-se ainda a possibilidade de KIR3DL também formar dímeros mesmo na ausência de ligantes.

Figura VIII-27: Representação esquemática de dimerização de KIR2DL na presença do ligante HLA-C. O complexo se forma através do contato entre as interfaces dos domínios D₂.

É possível que DM43 possa, de modo análogo, formar dímero com ou sem a presença de metaloprotease. O alinhamento entre DM43 e KIRs, revela uma substituição sistemática dos resíduos hidrofílicos por hidrofóbicos em DM43, distribuídos ao longo da face oposta ao domínio D₁. É interessante notar a ausência de qualquer sítio de glicosilação em sua proximidade (figura VIII-28).

Figura VIII-28: Representação em *ribbons* da provável superfície de dimerização. (a) Os resíduos destacados em verde correspondem aos hidrofóbicos que substituíram os resíduos hidrofílicos de KIRs. Em vermelho, as asparaginas indicando os sítios de glicosilação, e em amarelo a orientação do triptofano correspondente ao motivo WSXWS. (b) Vista frontal da interface de dimerização.

No artigo de Jones e Thornton (1996), são citados resultados estatísticos de que as superfícies envolvidas na formação de um complexo homodimérico costumam ser menos planares que as de um complexo heterodimérico. Esta

evidência reforçaria a hipótese de dimerização de DM43 na interface em questão, visto o seu caráter hidrofóbico e a sua concavidade (figura VIII-28).

Numa tentativa de confirmar os potenciais sítios de dimerização e de interação entre D_1 e D_0 , foram feitos cálculos do potencial eletrostático da superfície de DM43_ D_0 e DM43_ D_1 - D_2 . Eles revelaram que além de sua propriedade ácida (Neves-Ferreira, 2000) existem duas regiões não carregadas em DM43_ D_1 - D_2 (figura VIII-29) e uma terceira região, também neutra, em DM43_ D_0 (figura VIII-30).

Figura VIII-29: Superfície de potencial de DM43_ D_1 - D_2 . (a) Vista frontal esquerda de (b), e (c) vista frontal direita de (b), onde (b) ilustra a visão total de uma das faces de DM43_ D_1 - D_2 . A cor vermelha indica o caráter ácido da superfície, portanto, potencial negativo; o azul, as regiões de potencial positivo; e o branco, as regiões neutras.

Figura VIII-30: Superfície de potencial de DM43_ D_0 . (a) Vista lateral da região N-terminal do domínio D_0 . (b) Vista frontal de DM43_ D_0 . (c) Superfície da região C-terminal onde começa o domínio DM43_ D_1 - D_2 .

Um estudo mais detalhado na busca das regiões de dimerização e interação com o domínio D_0 foi feito através do programa PATCHES. O programa identificou

dois prováveis *patches* (ou regiões de interface) em DM43_D₁-D₂, centrados nos resíduos Ala169 e Leu252, e apenas um único *patch* em DM43_D₀ centrado em Thr49. A tabela a seguir lista os resíduos identificados nas posições das prováveis regiões de interação.

Tabela VIII-8: Resíduos pertencentes às prováveis regiões de interação identificados pelo programa PATCHES.

DM43_D ₀	DM43_D ₁ -D ₂	
Centrado em Thr 49	Centrado em Ala 169	Centrado em Leu 252
Gln 39	Leu 125	Tyr 196
Trp 41	Leu 126	Ile 197
Glu 45	Met 128	Leu 198
Leu 46	Ile 129	Thr 210
Leu 47	Tyr 165	Trp 214
Ser 48	Arg 166	Val 239
Thr 49	Ala 167	Val 240
Leu 50	Pro 168	Pro 241
Pro 51	Ala 169	Ile 242
Val 52	Ser 170	Phe 250
Trp 60	Ala 171	Asp 251
Arg 75	Pro 172	Leu 252
Leu 77	Ser 173	Thr 253
Thr 78	Val 174	Ala 254
Trp 83	Ser 176	Val 255

As superfícies correspondentes a estas regiões foram visualizadas em GRASP, cujas imagens estão apresentadas nas figuras VIII-31 e VIII-32. Interessantemente, as localizações destes *patches* concordaram com as regiões não carregadas identificadas pelos cálculos de superfície de potencial de GRASP.

O *patch* localizado no domínio D₁ de DM43_D₁-D₂ corresponderia à interface com o domínio D₀ que, por sua vez, possui um *patch* próximo ao seu C-terminal (regiões identificadas em verde nas figuras VIII-31 e VIII-32). O terceiro *patch* situado na face oposta ao domínio D₂ de DM43_D₁-D₂ (identificada em vermelho nas figuras VIII-31 e VIII-32), confirma a provável existência de uma superfície de dimerização, como foi anteriormente proposto.

A partir da identificação destas regiões, é possível sugerir uma conformação de DM43_D₀ com relação à DM43_D₁-D₂, encaixando geometricamente os *patches* identificados em DM43_D₀ (figura VIII-33).

Figura VIII-31: Vista frontal de DM43_D1-D2 com os *patches* destacados em verde, para o centrado em Ala 169 e em vermelho para o centrado em Leu 252.

Figura VIII-32: Representação das prováveis regiões de interação encontradas em DM43_D1-D2, pelo programa PATCHES. (a) A região provável de interação com o domínio D0 está destacada em verde. Corresponde ao *patch* centrado em Ala 169. (b) A região provável de dimerização está destacada em vermelho. Corresponde a um *patch* centrado em Leu 252.

Figura VIII-33: Visão completa de DM43 com o domínio D_0 em posição de encaixe geométrico pelas regiões encontradas por PATCHES (em verde) e a junção destes domínios pelo C-terminal de D_0 e N-terminal de D_1 (em azul). A provável superfície de dimerização encontrada por PATCH no domínio D_2 está indicada em vermelho.

A partir de toda esta análise, fecha-se a possibilidade de DM43 adotar uma configuração elipsoidal no seu estado monomérico e formar um dímero com seu idêntico através da face hidrofóbica de DM43_ D_2 . A localização dos possíveis *loops* envolvidos na inibição da metaloprotease em DM43_ D_1 - D_2 , sugere o seu atracamento na cavidade formada pelo dímero de DM43 (figura VIII-34).

A presença de um sítio de glicosilação em D_1 muito próximo aos *loops* envolvidos na ligação com a metaloprotease, sugere que os carboidratos deste sítio estejam reforçando sua ligação em uma interação direta. Além disto, podem estar ocupando o espaço restante para uma melhor complementaridade física da metaloprotease na interface de complexação com DM43.

Resumindo, sugere-se o seguinte mecanismo de dimerização e complexação de DM43. Na ausência da metaloprotease, DM43 apresenta-se na forma homodimérica devido à interação entre as superfícies hidrofóbicas dos domínios D_2 de cada monômero. Ao encontrar uma metaloprotease, esta se liga aos *loops* de interação dos domínios D_1 e D_2 . Neste ponto, a superfície de

dimerização de D_2 não mais está interagindo com a de outro monômero, mas com uma provável interface hidrofóbica da metaloprotease. Esta hipótese concorda com os resultados experimentais apresentados por Neves-Ferreira (1999) sobre a formação de homodímeros, e de complexos de DM43 na proporção de 1:1 com a metaloprotease. A figura VIII-34 ilustra este possível mecanismo de atuação de DM43.

Figura VIII-34: Hipótese sobre o processo de complexação de DM43 com a metaloprotease.

Na página seguinte (figura VIII-35), apresenta-se um resumo das principais etapas de modelagem de DM43, que corresponde à parte B deste trabalho. Apesar de não ter sido anteriormente mencionado neste capítulo, foi incluído uma tabela extra dos valores de porcentagem de identidade entre os três domínios de DM43. Nela, destacou-se a combinação de maior porcentagem de identidade sequencial, que foi entre os domínios D_0 e D_2 . Justificando, portanto, ainda mais a modelagem de DM43_ D_0 a partir da estrutura tridimensional do domínio D_2 de 1nkr.

IX.**CONCLUSÕES E PERSPECTIVAS**

O trabalho experimental completo compreendendo testes de cristalização, desglicosilação enzimática e purificação, despendeu menos de 10mg de proteína liofilizada.

Os maiores cristais foram coletados a 85K utilizando radiação síncrotron (LNLS/Campinas-SP). Apesar da desordem estrutural dos cristais de DM43, foi possível determinar os seus parâmetros de cela unitária, que segundo resultados de indexação de uma imagem de difração, resultaram em: $a = b = 117,34 (0,03) \text{ \AA}$, $c = 193,39 (0,02) \text{ \AA}$, $\alpha = \beta = 90^\circ$ e $\gamma = 120^\circ$. Correspondendo, portanto, ao sistema cristalino trigonal ou hexagonal.

A tentativa de obter melhores cristais de DM43 através da desglicosilação enzimática resultou na retirada parcial dos 29% em massa de carboidratos ligados a DM43. A grande semelhança nas características físico-químicas das espécies de DM43 (parcialmente desglicosilada e nativa) e a disponibilidade de uma pequena quantidade de proteína não permitiram resolução suficiente para distinguir os correspondentes picos de eluição cromatográficos.

As gotas de cristalização montadas a partir de uma mistura de DM43 nativa e desglicosilada não apresentaram cristais.

A modelagem possibilitou uma visão prévia da estrutura cristalográfica de DM43. Devido à existência de estruturas homólogas possuindo apenas dois domínios tipo imunoglobulina contra os três de DM43, foi necessário modelá-la em duas partes. Deste modo, o domínio D₀ de DM43 foi construído a partir da estrutura do domínio D₂ de KIR2DL2 (1nkr), enquanto que os dois domínios C-terminal restantes de DM43 foram baseados na estrutura completa de 1nk. O método utilizado foi a modelagem molecular por homologia pela satisfação das restrições espaciais. Apesar do baixo grau de identidade entre DM43 e a estrutura de referência 1nkr (~28%), o método obteve modelos que, segundo três parâmetros de qualidade (PROCHECK, VERIFY_3D e QUALITY), correspondem a estruturas bem modeladas.

A modelagem de DM43_D₀ e DM43_D₁-D₂ indicou, por exemplo, a importância de alguns resíduos na conformação espacial de sua estrutura que provavelmente refletem nos mecanismos de atuação inibitória de DM43 contra a metaloprotease do veneno de serpente.

Algumas características estruturais de DM43 são comuns aos receptores inibitórios de células assassinas (KIRs), como:

- enovelamento tipo imunoglobulina;
- existência do motivo WSXWS *box* em todos os três domínios;
- fita β extra não pertencente ao enovelamento imunoglobulina, posicionada na junção entre, pelo menos, dois últimos domínios de DM43;
- prolongamento da fita A na outra folha, com a presença de um resíduo na conformação *cis*, dividindo-a em duas fitas conectadas por um *loop* reto.

A existência de quatro sítios de glicosilação em DM43 pode estar estruturalmente relacionada com a orientação relativa dos três domínios e possivelmente, com a interação envolvendo a metaloprotease.

Análises de estruturas de receptores de hormônios de crescimento e prolactina indicaram a participação de seis *loops*, cujas seqüências foram utilizadas como referência para identificar, em DM43, os possíveis *loops* envolvidos na interação com a metaloprotease. A localização de um sítio de glicosilação muito próxima a estes *loops* pode indicar uma participação dos carboidratos na interação como uma ferramenta extra de DM43 para aumentar sua eficiência na “captura” e fixação da metaloprotease.

Os estudos de modelagem de DM43 concordaram com os resultados experimentais de formação de homodímeros devido à presença de uma interface hidrofóbica no domínio D₂, e à complexação de 1:1 pela interação da metaloprotease com a interface hidrofóbica e *loops*. A presença do motivo WSXWS *box* em todos os três domínios de DM43 permitiu levantar hipóteses sobre suas orientações relativas, sendo que em D₂ o motivo apresenta estar mais bem conservado que em outros domínios. Esta conservação maior em D₂, especialmente pela presença do primeiro triptofano e das serinas, pode estar ditando a orientação relativa com o domínio D₁, menor que 90°. Curiosamente, no domínio D₁ o motivo WSXWS é substituído por NSTPS, o que satisfaz a condição para a existência de um sítio de glicosilação. A presença deste sítio somada com a de outros dois (um em D₀ e outro em D₁) podem estar relacionados com a orientação do domínio D₀, com relação ao domínio D₁, em um ângulo maior que 90° (figura VIII-34).

Portanto, conclui-se que as posições dos sítios de glicosilação de uma forma geral são compatíveis com o modelo.

Propõe-se, no futuro, proceder com novos ensaios de desglicosilação de DM43, utilizando enzimas mais frescas na esperança de obter DM43 totalmente desglicosilada e em maior quantidade. Alternativamente, pode-se realizar experimentos de expressão de DM43 em procariotos através de processos recombinantes que garantem a obtenção de proteínas ausentes de estruturas glicídicas. Ainda existe a possibilidade de obter cristais de complexos com a jararagina¹ onde a DM43 pode, neste caso, estar ou não desglicosilada.

Com a identificação dos possíveis *loops* de interação com a metaloprotease e os resíduos hidrofóbicos da interface de dimerização, pode-se iniciar trabalhos envolvendo mutagênese de sítio dirigido que possibilitarão detectar os resíduos-chave envolvidos na inibição dos efeitos hemorrágicos do veneno de serpentes.

¹ Principal toxina hemorrágica de *Bothrops jararaca* muito bem caracterizada físico-quimicamente, mas sem estrutura conhecida.

Anexo - I

Soluções do Fatorial I
(Hampton Research)

Crystal ScreenReagent Formulation

Tube Number	Salt	Tube Number	Buffer †	Tube Number	Precipitant
1.	0.02 M Calcium Chloride dihydrate	1.	0.1 M Sodium Acetate trihydrate pH 4.6	1.	30% v/v 2-Methyl-2,4-pentanediol
2.	None	2.	None	2.	0.4 M Potassium Sodium Tartrate tetrahydrate
3.	None	3.	None	3.	0.4 M mono-Ammonium dihydrogen Phosphate
4.	None	4.	0.1 M Tris Hydrochloride pH 8.5	4.	2.0 M Ammonium Sulfate
5.	0.2 M tri-Sodium Citrate dihydrate	5.	0.1 M HEPES - Na pH 7.5	5.	30% v/v 2-Methyl-2,4-pentanediol
6.	0.2 M Magnesium Chloride hexahydrate	6.	0.1 M Tris Hydrochloride pH 8.5	6.	30% w/v Polyethylene Glycol 4000
7.	None	7.	0.1 M Sodium Cacodylate pH 6.5	7.	1.4 M Sodium Acetate trihydrate
8.	0.2 M tri-Sodium Citrate dihydrate	8.	0.1 M Sodium Cacodylate pH 6.5	8.	30% v/v iso-Propanol
9.	0.2 M Ammonium Acetate	9.	0.1 M tri-Sodium Citrate dihydrate pH 5.6	9.	30% w/v Polyethylene Glycol 4000
10.	0.2 M Ammonium Acetate	10.	0.1 M Sodium Acetate trihydrate pH 4.6	10.	30% w/v Polyethylene Glycol 4000
11.	None	11.	0.1 M tri-Sodium Citrate dihydrate pH 5.6	11.	1.0 M mono-Ammonium dihydrogen Phosphate
12.	0.2 M Magnesium Chloride hexahydrate	12.	0.1 M HEPES - Na pH 7.5	12.	30% v/v iso-Propanol
13.	0.2 M tri-Sodium Citrate dihydrate	13.	0.1 M Tris Hydrochloride pH 8.5	13.	30% v/v Polyethylene Glycol 400
14.	0.2 M Calcium Chloride dihydrate	14.	0.1 M HEPES - Na pH 7.5	14.	28% v/v Polyethylene Glycol 400
15.	0.2 M Ammonium Sulfate	15.	0.1 M Sodium Cacodylate pH 6.5	15.	30% w/v Polyethylene Glycol 8000
16.	None	16.	0.1 M HEPES - Na pH 7.5	16.	1.5 M Lithium Sulfate monohydrate
17.	0.2 M Lithium Sulfate monohydrate	17.	0.1 M Tris Hydrochloride pH 8.5	17.	30% Polyethylene Glycol 4000
18.	0.2 M Magnesium Acetate tetrahydrate	18.	0.1 M Sodium Cacodylate pH 6.5	18.	20% Polyethylene Glycol 8000
19.	0.2 M Ammonium Acetate	19.	0.1 M Tris Hydrochloride pH 8.5	19.	30% v/v iso-Propanol
20.	0.2 M Ammonium Sulfate	20.	0.1 M Sodium Acetate trihydrate pH 4.6	20.	25% w/v Polyethylene Glycol 4000
21.	0.2 M Magnesium Acetate tetrahydrate	21.	0.1 M Sodium Cacodylate pH 6.5	21.	30% v/v 2-Methyl-2,4-pentanediol
22.	0.2 M Sodium Acetate trihydrate	22.	0.1 M Tris Hydrochloride pH 8.5	22.	30% w/v Polyethylene Glycol 4000
23.	0.2 M Magnesium chloride hexahydrate	23.	0.1 M HEPES - Na pH 7.5	23.	30% v/v Polyethylene Glycol 400
24.	0.2 M Calcium Chloride dihydrate	24.	0.1 M Sodium Acetate trihydrate pH 4.6	24.	20% v/v iso-Propanol
25.	None	25.	0.1 M Imidazole pH 6.5	25.	1.0 M Sodium Acetate trihydrate
26.	0.2 M Ammonium Acetate	26.	0.1 M tri-Sodium Citrate dihydrate pH 5.6	26.	30 % v/v 2-Methyl-2,4-pentanediol
27.	0.2 M tri-Sodium Citrate dihydrate	27.	0.1 M HEPES - Na pH 7.5	27.	20% v/v iso-Propanol
28.	0.2 M Sodium Acetate trihydrate	28.	0.1 M Sodium Cacodylate pH 6.5	28.	30% w/v Polyethylene Glycol 8000
29.	None	29.	0.1 M HEPES - Na pH 7.5	29.	0.8 M Potassium Sodium Tartrate tetrahydrate
30.	0.2 M Ammonium Sulfate	30.	None	30.	30% w/v Polyethylene Glycol 8000
31.	0.2 M Ammonium Sulfate	31.	None	31.	30% w/v Polyethylene Glycol 4000
32.	None	32.	None	32.	2.0 M Ammonium Sulfate
33.	None	33.	None	33.	4.0 M Sodium Formate
34.	None	34.	0.1 M Sodium Acetate trihydrate pH 4.6	34.	2.0 M Sodium Formate
35.	None	35.	0.1 M HEPES - Na pH 7.5	35.	0.8 M mono-Sodium dihydrogen phosphate 0.8 M mono-Potassium dihydrogen phosphate
36.	None	36.	0.1 M Tris Hydrochloride pH 8.5	36.	8% w/v Polyethylene Glycol 8000
37.	None	37.	0.1 M Sodium Acetate trihydrate pH 4.6	37.	8% w/v Polyethylene Glycol 4000
38.	None	38.	0.1 M HEPES - Na pH 7.5	38.	1.4 M tri-Sodium Citrate dihydrate
39.	None	39.	0.1 M HEPES - Na pH 7.5	39.	2% v/v Polyethylene Glycol 400, 2.0 M Ammonium Sulfate
40.	None	40.	0.1 M tri-Sodium Citrate dihydrate pH 5.6	40.	20% v/v iso-Propanol, 20% w/v Polyethylene Glycol 4000
41.	None	41.	0.1 M HEPES - Na pH 7.5	41.	10% v/v iso-Propanol, 20% w/v Polyethylene Glycol 4000
42.	0.05 M mono-Potassium dihydrogen Phosphate	42.	None	42.	20% w/v Polyethylene Glycol 8000
43.	None	43.	None	43.	30% w/v Polyethylene Glycol 1500
44.	None	44.	None	44.	0.2 M Magnesium Formate
45.	0.2 M Zinc Acetate dihydrate	45.	0.1 M Sodium Cacodylate pH 6.5	45.	18% w/v Polyethylene Glycol 8000
46.	0.2 M Calcium Acetate hydrate	46.	0.1 M Sodium Cacodylate pH 6.5	46.	18% w/v Polyethylene Glycol 8000
47.	None	47.	0.1 M Sodium Acetate trihydrate pH 4.6	47.	2.0 M Ammonium Sulfate
48.	None	48.	0.1 M Tris Hydrochloride pH 8.5	48.	2.0 M mono-Ammonium dihydrogen Phosphate
49.	1.0 M Lithium Sulfate monohydrate	49.	None	49.	2% w/v Polyethylene Glycol 8000
50.	0.5 M Lithium Sulfate monohydrate	50.	None	50.	15% w/v Polyethylene Glycol 8000

† Buffer pH is that of a 1.0 M stock prior to dilution with other reagent components. pH with HCl or NaOH.

Crystal Screen contains fifty unique reagents. To determine the formulation of each reagent, simply read across the page.

27632 El Lazo Road, Suite 100
Laguna Niguel, CA 92677-3913 U.S.A.
Tel: (949) 425-1321 • Fax: (949) 425-1611
Technical Support e-mail: tech@hmail.com
Internet: www.hamptonresearch.com

Anexo - II

Protocolo de Detecção de Glicoproteínas em Membrana de Nitrocelulose

(Bio-Rad Laboratories)

B.1 Transferência

B.1.1 Equipamento

Descrição do aparato experimental utilizado na transferência das amostras do gel de poliacrilamida em membrana de nitrocelulose.

Figura A- 1: Mini-Trans Blot® Eletrophoretic Transfer Cell (Intruccion Manual. Cod.: 170-3930/3935)

B.1.2 Tampão de Transferência

Foi preparado 500 ml de tampão de transferência, contendo 25mM Tris, 192mM glicina e 20% v/v metanol. O pH desta solução, sem ajuste, deverá ser de pH 8,3.

B.1.3 Preparação da Cuba

Fill the Bio-Ice cooling unit with water and store it in your laboratory freezer at -20 °C until ready to use. After use, return the cooling unit to the freezer for storage.

1. Prepare the transfer buffer. (See Section 3.3 for buffer formulation. Using buffer chilled to 4 °C will improve heat dissipation.)

2. Cut the membrane and the filter paper to the dimensions of the gel. Always wear gloves when handling membranes to prevent contamination. Equilibrate the gel and soak the membrane, filter paper, and fiber pads in transfer buffer (15 min–1 hour depending on gel thickness).
3. Prepare the gel sandwich.
 - Place the cassette, with the gray side down, on a clean surface.
 - Place one pre-wetted fiber pad on the gray side of the cassette.
 - Place a sheet of filter paper on the fiber pad.
 - Place the equilibrated gel on the filter paper.*
 - Place the pre-wetted membrane on the gel.*
 - Complete the sandwich by placing a piece of filter paper on the membrane.*

Add the last fiber pad.

* Removing any air bubbles which may have formed is very important for good results. Use a glass tube to gently roll air bubbles out.

4. Close the cassette firmly, being careful not to move the gel and filter paper sandwich. Lock the cassette closed with the white latch.
5. Place the cassette in module. Repeat for the other cassette.

6. Add the frozen Bio-Ice cooling unit. Place in tank and completely fill the tank with buffer.

7. Add a standard stir bar to help maintain even buffer temperature and ion distribution in the tank. Set the speed as fast as possible to keep ion distribution even.

8. Put on the lid, plug the cables into the power supply, and run the blot. Refer to Section 3 for run times and voltage settings with various buffers.

9. Upon completion of the run, disassemble the blotting sandwich and remove the membrane for development. Clean the cell, fiber pads, and cassettes with laboratory detergent and rinse well with deionized water.

B.1.4 Condições para Transferência

Aplica-se uma tensão constante de 100V para uma hora de transferência.

B.2 Revelação da Membrana

Os procedimentos descritos a seguir foram retirados do manual de revelação de glicoproteínas da Bio-Rad Laboratories (Cod.: 170-6490).

B.2.1 Reagentes

Description	Reagents/Quantities		Directions
1. TBS (50 mM Tris, 27 mM sodium chloride, pH 7.2)	6.05 g 1.6 g	Tris NaCl	Dissolve reagents in 900 ml ddH ₂ O. Adjust pH to 7.2 with 1 N HCl. Q.S. to 1,000 ml with ddH ₂ O.
2. 200 mM Sodium Acetate Buffer, pH 5.5	440 ml 60 ml	200 mM sodium acetate (12.0 g sodium acetate trihydrate in 440 ml ddH ₂ O) 200 mM acetic acid (690 µl glacial acetic acid in 60 ml ddH ₂ O)	Combine stock solutions and check pH is 5.5.
4. Color Develop- ment Buffer	1.21 g 1.01 g 0.58 g 90 ml	Tris MgCl ₂ •6H ₂ O NaCl ddH ₂ O	Dissolve reagents in 90 ml ddH ₂ O. Adjust pH to 9.5 with 0.1 N HCl and Q.S. to 100 ml total volume with ddH ₂ O.
5. DMF	3.0 ml	Dimethylfor- mamide	Ready to use.
6. PBS (9 mM sodium phosphate, 27 mM sodium chloride, pH 7.2)	575 mg 100 mg 800 mg	Na ₂ HPO ₄ NaH ₂ PO ₄ NaCl	Dissolve reagents in 500 ml ddH ₂ O.
7. 100 mM Sodium Acetate/5 mM EDTA	1.14 g 300 ml	EDTA (tetra sodium salt) 200 mM sodium acetate, pH 5.5 (solution 2)	Dissolve EDTA in 300 ml of solution 2. Q.S. to 600 ml total volume with ddH ₂ O.

Reagent Contents		Reagent Preparation	Storage/Shelf Life After Preparation
A	Periodate Protocol 1	Dissolve 21.4 mg of vial A in 10 ml sodium acetate/EDTA (solution 7) to make a 10 mM solution.	Prepare just prior to use. Light sensitive reagent.
B	Hydrazide	Dissolve vial B contents in 500 µl DMF (solution 5).	3 months at 4 °C.
C	Blocking Reagent	Dissolve 0.5 g of vial C in 100 ml TBS (solution 1); if necessary heat at 60 °C for 45 minutes to dissolve.	3 months at 4 °C.
D	Streptavidin-Alkaline Phosphatase	Ready to use.	3 months at 4 °C. Do not freeze.
E	NBT	Add 1.3 ml 70% (v/v) DMF (solution 5) to the contents of vial E and dissolve.	3 months at -20 °C. Light sensitive reagent.
F	BCIP	Add 1 ml DMF (solution 5) to the contents of vial F and dissolve.	3 months at -20 °C. Light sensitive reagent.

B.2.2 Procedimientos

1. Immobilize the protein sample onto the membrane by performing either a dot blot manually or with the Bio-Dot® or Bio-Dot SF microfiltration apparatus, or by performing SDS-PAGE with the Mini-PROTEAN® II cell followed by a Western blot transfer with the Trans-Blot®, Trans-Blot SD, or Mini Trans-Blot® cell.
2. Following transfer, wash the membrane with 10 ml PBS (solution 6) at room temperature for 10 minutes with gentle agitation. **Do not use TBS as Tris will interfere with the subsequent biotinylation reactions.**
3. Immerse the membrane in 10 ml of 10 mM sodium periodate in sodium acetate/EDTA buffer (reagent A). Incubate in the dark (cover dish with aluminum foil) at room temperature for 20 minutes with gentle agitation.
4. Wash the membrane 3 times with 10 ml of PBS (solution 6) at room temperature for 10 minutes with gentle agitation. Use fresh buffer each time.
5. Prepare the biotinylation solution immediately before use by adding 2 µl hydrazide solution (reagent B) to 10 ml sodium acetate/EDTA buffer (solution 7). Immerse the membrane in this solution and incubate for 60 minutes at room temperature with gentle agitation.
6. Wash the membrane 3 times with 10 ml of TBS (solution 1) at room temperature for 10 minutes with gentle agitation. Use fresh buffer each time. (Note: TBS will no longer interfere with the process and may be used in subsequent steps.)
7. Immerse the membrane in 10 ml of blocking solution (reagent C). Incubate for at least 30 minutes at

room temperature with gentle agitation or overnight at 4 °C.

8. Wash the membrane 3 times with 10 ml of TBS (solution 1) at room temperature for 10 minutes with gentle agitation. Use fresh buffer each time.
9. Prepare the conjugate solution by adding 5 µl Streptavidin-Alkaline Phosphatase conjugate (reagent D) to 10 ml TBS (solution 1). Immerse the membrane in this solution and incubate for 60 minutes at room temperature with gentle agitation.
10. Wash the membrane 3 times with 10 ml of TBS (solution 1) at room temperature for 10 minutes with gentle agitation. Use fresh buffer each time.
11. Prepare the color development solution immediately before use. Add 50 µl NBT (reagent E) and 37.5 µl BCIP (reagent F) to 10 ml of color development buffer (solution 4) and mix well. Immerse the membrane in the color development solution at room temperature **without agitation**. Develop to desired intensity (3-60 minutes).
12. To stop the development, rinse the membrane several times in ddH₂O and allow membrane to air dry.

Referências Bibliográficas

REFERÊNCIAS BIBLIOGRÁFICAS

- ABERGEL, C. *et al.* The effect of protein contaminants on the crystallization of turkey egg white lysozyme. *J. Crystal Growth*, v. 110, p. 11-19, 1991.
- AL-LAZIKANI, B.; LESK, A. M.; CHOTHIA, C. Standard conformations for the canonical structures of immunoglobulins. *J. Mol. Biol.*, v. 273, p. 927-948, 1997.
- ALTSCHUL, S. F., *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database searchprograms", *Nucl. Acids Res.*, v. 25, p. 3389-3402, 1997.
- ALTSCHUL, S. F.; GISH, W.; MILLER, W.; MYERS, E. W.; LIPMAN, D. J. Basic local alignment search tool. *J. Mol. Biol.*, v. 215, p. 403-410, 1990.
- ASHKIN, A. Acceleration and trapping of particles by radiation pressure. *Phys. Rev. Lett.*, v.24, n. 4, p. 156-159, 1970.
- ASHKIN, A. Applications of laser radiation pressure. *Science*, v. 210, n. 4474, p. 1081-1088, 1980.
- ASHKIN, A. *et al.* Observation of a single-beam gradient force optical trap for dielectric particles. *Opt. Lett.*, v. 11, n. 5, p. 288-290, 1986.
- ASHKIN, A.; DZIEDZIC, J. M. Optical levitation of liquid drops by radiation pressure. *Science*, v. 187, p. 1073-1075, 1975.
- ASHKIN, A.; DZIEDZIC, J. M. Optical trapping and manipulation of viruses and bacteria. *Science*, v. 235, p. 1517-1520, 1987.
- ASHKIN, A.; DZIEDZIC, J. M.; YAMANE, T. Optical trapping and manipulation of single cells using infrared laser beams. *Nature*, v. 330, p. 769-771, 1987.

- ATAKA, M.; KATOH, E.; WAKAYAMA, N. I. Magnetic orientation as a tool to study the initial stage of crystallization of lysozyme. *J. Cryst. Growth*, v. 173, p. 592-596, 1997.
- ATKINS, P. W. *Physical chemistry*. Oxford University Press, 6th ed., p. 655, 1998.
- BAIROCH, A.; APWEILER, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucl. Acids Res.*, v. 28, p. 45-48, 2000.
- BALDOCK, P.; MILLS, V.; STEWART, P. S. A comparison of microbatch and vapour diffusion for initial screening of crystallization conditions. *J. Crystal Growth*, v. 168, p. 170-174, 1996.
- BANCEL, P. A. *et al.* Laser seeding for biomolecular crystallization. *J. Crystal Growth*, v. 191, p. 537-544, 1998.
- BANCEL, P. A.; CAJIPE, V. B.; RODIER, F. Manipulating crystals with light. *J. Cryst. Growth*, v. 196, p. 685-690, 1999.
- BARKER, W.C. *et al.* The Protein Information Resource (PIR). *Nucleic Acids Research*, v. 28, n. 1, p. 41-44, 2000.
- BARTON, G. J.; STERNBERG, M. J. A strategy for the rapid multiple alignment of protein sequences. Confidence levels from tertiary structure comparisons. *J. Mol. Biol.*, v. 198, n. 2, p. 327-337, 1987.
- BENSON, D. A. *et al.* GenBank. *Nucl. Acids Res.*, v. 27, p. 12-17, 1999.
- BERMAN, H.M. *et al.* The Protein Data Bank. *Nucl. Acids Res.*, v. 28, p. 235-242, 2000.
- BHAMIDI, V. *et al.* The influence of a homologous protein impurity on lysozyme crystal growth. *J. Crystal Growth*, v. 204, p. 542-552, 1999.
- BIO-RAD LABORATORIES Catalogue. p. 156-157, 1998/1999.
- BLOCK, S. M. Making light work with optical tweezers. *Nature*, v. 360, p. 493-495, 1992.
- BLUNDELL, T. L.; JOHNSON, L. N. *Protein crystallography*, Academic Press, 565 p., 1976.
- BORK, P.; HOLM, L.; SANDER, C. The immunoglobulin fold. *J. Mol. Biol.*, v. 242, p. 309-320, 1994.
- BOWIE, J. U.; LUTHY, R.; EISENBERG, D. A method to identify protein sequences that fold into a known 3-Dimensional Structure. *Science*. v. 253, n. 5016, p. 164-170, 1991.
- BRAUN, W.; GÖ, N. Calculation of protein conformations by proton-proton distance constraints. *J. Mol. Biol.*, v. 186, p. 611-626, 1985.
- BURKE, D. F. *et al.* An iterative structure-assisted approach to sequence alignment and comparative modeling. *Proteins*, p. 55-60, 1999. Suplemento 3.

- CARERI, G. *et al.* Magnetic susceptibility of lysozyme. *Phys. Lett.*, v. 60A, n. 5, p. 490-491, 1977.
- CATANESE, J. J.; KRESS, L. F. Isolation from opossum serum of a metalloproteinase inhibitor homologous to human α 1B-glycoprotein. *Biochemistry*, v. 31, p. 410-418, 1992.
- CEVAP - Centro de Estudos de Venenos e Animais Peçonhentos. Emergências médicas-Serpentes. [http://www.botunet.com.br/cevap/portugues/Medicas/serpentes.htm#Acidente botrópico](http://www.botunet.com.br/cevap/portugues/Medicas/serpentes.htm#Acidente_botropico).
- CHAYEN, N.E. The role of oil in macromolecular crystallization. *Structure*, v. 5, p. 1269-1274, 1997.
- CLARKE, J. *et al.* Folding studies of immunoglobulin-like β -sandwich proteins suggest that they share a common folding pathway. *Structure*, v. 7, p. 1145-1153, 1999.
- COLONNA, M.; SAMARIDIS, J. Cloning of immunoglobulin-superfamily members associated with HLA-C and HLA-B recognition by human natural-killer-cells. *Science*, v. 268, n. 5209, p. 405-408, 1995.
- D'ARCY, A. *et al.* A novel approach to crystallising proteins under oil. *J. Crystal Growth*, v. 168, p. 175-180, 1996.
- DAYHOFF, M. O.; SCHWARTZ, R. M.; ORCUTT, B. C. Atlas of protein sequence and structure. National Biomedical Research Foundation. v. 5, supl. 3, p. 345-352, 1978.
- DEHOFF, R. T. *Thermodynamics in materials science*, McGraw-Hill, 532 p., 1993.
- DOMONT, G. B.; PERALES, J. ; MOUSSATCHÉ, H. Natural anti-snake venom proteins. *Toxicon*, v. 29, p. 1183-1194, 1991.
- DRENTH, J. *Principles of protein X-ray crystallography*. Springer-Verlag, 1995.
- DUCRUIX, A.; GIEGÉ, R. *Crystallization of Nucleic Acids and Proteins, a practical approach*. Irl Press, 331 p., 1992.
- ExPASy (Expert Protein Analysis System). <http://www.expasy.ch/cgi-bin/niceprot.pl?P21163>.
- FAN, Q. R. *et al.* Structure of the inhibitory receptor for human natural killer cells resembles haematopoietic receptors. *Nature*, v. 389, p. 96-100, 1997.
- GARMAN, E. F.; MITCHELL, E. P. Glycerol concentrations required for cryoprotection of 50 typical protein crystallization solutions. *J. Appl. Cryst.*, v. 29, p. 584-587, 1996.
- GARMAN, E. F.; SCHNEIDER, T. R.; *Macromolecular Cryocrystallography*. *J. Appl. Cryst.*, v. 30, p. 211-237, 1997.

- GIACOVAZZO, C. *et al.* *Fundamentals of crystallography*. Oxford University Press, 654 p., 1995.
- GILL, S. C.; VON HIPPEL, P. H. Calculation of protein extinction coefficients from amino-acid sequence data. *Anal. Biochem.*, v. 182, p. 319-326, 1989.
- GLYKOMART - Tools for glycobiology. <http://www.glykomart.com/glykomart/e-5006.html>.
- GUEx, N.; PEITSCH, M.C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* 18, 2714-2723. 1997.
- HARRIS, E. L. V.; ANGAL, S. *Protein purification methods, a practical approach*. Irl Press, 4th ed., 317 p., 1993.
- HARRIS, E.L.V.; ANGAL, S. *Protein purification applications, a practical approach*. Irl press, 179 p., 1995.
- HUTCHINSON, E. G.; THORNTON, J. M. PROMOTIF - A program to identify and analyse structural motifs in proteins. *Protein Science*, v. 5, p. 212-220, 1996.
- ISHIOKA, N.; TAKAHASHI, N.; PUTNAM, F.W. Amino acid sequence of human plasma α_1 B-glicoprotein: homology to the immunoglobulin supergene family. *Proc. Natl. Acad. Sci. USA*, v. 83, p. 2363-2367, 1986.
- JANCARIK, J.; KIM, S. H. Sparse matrix sampling: a screening method for crystallization of proteins. *J. Appl. Cryst.*, v. 24., p. 204-411, 1991.
- JONES, T. A. A graphics model building and refinement system for macromolecules. *J. Appl. Cryst.* v. 11, p. 268-272, 1978.
- JONES, T. A.; THIRUP, S. Using known substructures in protein model building and crystallography. *EMBO J.*, v. 5, p. 819-822, 1986.
- JONES, T. A.; ZON, J. Y.; CAVAN, S. W.; KJELGAARD, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst.*, v. 47A, p. 110-119, 1991.
- JONES, D. T. *et al.* A new approach to protein fold recognition. *Nature*, v. 358, p. 86-89, 1992.
- JONES, S.; THORNTON, J. M. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. USA*, v. 93, p. 13-20, 1996.
- JONES, S.; THORNTON, J. M. Analysis of protein-protein interaction sites using surface patches. *J. Mol. Biol.*, v. 272, p. 121-132, 1997a.
- JONES, S.; THORNTON, J. M. Prediction of protein-protein interaction sites using patch analysis. *J. Mol. Biol.*, v. 272, p. 133-143, 1997b.
- JURGILAS, P. B. *et al.* Detection of an antithrombotic fraction in opossum (*Didelphis marsupialis*) milk that neutralizes *Bothrops jararaca* venom. *Toxicon*, v. 37, p. 167-172, 1999.

- KORADI, R.; BILLETER, M.; WÜTHRICH, K. MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graphics*, v. 14, p. 51-55, 1996.
- KUBAGAWA, H.; BURROWS, P. D.; COOPER, M. D. A novel pair of immunoglobulin-like receptors expressed by B cells and myeloid cells. *Proc. Natl. Acad. Sci. USA*, v. 94, p. 5261-5266, 1997.
- KUZNETSOV, Y. G.; MALKIN, A. J.; GREENWOOD, A.; MCPHERSON, A. Interferometric studies of growth-kinetics and surface-morphology in macromolecular crystal-growth-canavalin, thaumatin, and turnip yellow mosaic-virus. *J. Struct. Biol.*, v. 114, n.3, p. 184-196, 1995.
- KUZNETSOV, Y. G.; MALKIN, A. J.; GREENWOOD, A.; MCPHERSON, A. Michelson interferometric studies of protein and virus crystal growth. *J. Crystal Growth*, v. 166, n.1- 4, p. 913-918, 1996.
- KUZNETSOV, Y. G.; MALKIN, A. J.; MCPHERSON, A. AFM studies of the nucleation and growth mechanisms of macromolecular crystals. *J. Crystal Growth*, v. 196, p. 489-502, 1999.
- LAEMMLI, U. K. Cleavage of structural Proteins during the assembly of the head of bacteriophage T4. *Nature*, v. 227, p. 680-685, 1970.
- LASKOWSKI, R. A.; MACARTHUR, M. W.; MOSS, D. S.; THORNTON, J. M.. Procheck: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, n. 26, p. 283-291, 1993.
- LEITÃO, C. M. *Zoogeografia do Brasil*. 3.ed. Rio Grande do Norte, Fundação Guimarães Duque, 1980. v. 111, p. 551, 1980.
- LEVINSON, W.; JAWETZ, E. *Medical microbiology & immunology*. Appleton & Lange, 4^a ed., 1996
- LORBER, B.; GIEGÉ, R. Containerless protein crystallization in floating drops: application to crystal growth monitoring under reduced nucleation conditions. *J. Crystal Growth*, v. 168, p. 204 -215, 1996.
- LUTHY, R.; BOWIE, J. U.; EISENBERG, D. Assessment of protein models with three-dimensional profiles. *Nature*, v. 356, n. 5, p. 83-85, 1992.
- MAENAKA, K.; JUJI, T.; STUART, D. I.; JONES, E. Y. Crystal structure of the human p58 killer cell inhibitory receptor (KIR2DL3) specific for HLA-Cw3-related MHC class I. *Structure*, v. 7, p. 391-398, 1999.
- MALKIN, A. J.; KUZNETSOV, Y. G.; LAND, T. A.; DEYOREO, J. J.; MCPHERSON, A. Mechanisms of growth for protein and virus crystals. *Nat. Struct. Biol.*, v. 2, n. 11, p. 956-959, 1995a.
- MALKIN, A. J.; KUZNETSOV, Y. G.; MCPHERSON, A. Incorporation of microcrystals by growing protein and virus crystals. *Proteins*, v. 24, n. 2, p. 247-252, 1996.

- MALKIN, A. J.; LAND, T. A.; KUZNETSOV, Y. G.; MCPHERSON, A.; DEYOREO, J. J. Investigation of virus crystal growth mechanisms by *In Situ* Atomic Force Microscopy. *Phys. Rev. Letters*, v. 75, n. 14, p. 2778-2781, 1995b.
- MCPHERSON, A. *Preparation and analysis of protein crystals*. Krieger Publishing, 371 p., 1989.
- MCPHERSON, A.; MALKIN, A. J.; KUZNETSOV, Y. G. The science of macromolecular crystallization. *Structure*, v. 15, n. 3, p. 759-568, 1995.
- MCPHERSON, A. *Crystallization of biological macromolecules*, CSHL Press, 586 p., 1999.
- MCPHERSON, A. *et al.* The effects of microgravity on protein crystallization: evidence for concentration gradients around growing crystals. *J. Crystal Growth*, v. 196, p. 572-586, 1999.
- MCPHERSON, A.; DELUCAS, L. Crystal-growing in space. *Science*, v. 283, n. 5407, p. 1459-1459, 1999.
- MERRIL, C. R. *et al.* Ultrasensitive stain for proteins in polyacrylamide gels shows regional variation in cerebrospinal fluid proteins. *Science*, v. 211, p. 1437-1438, 1981.
- MORENO, A. *et al.* Investigations on gravity influence upon protein crystallization by the gel acupuncture technique. *J. Crystal Growth*, v. 196, p. 587 – 594, 1999.
- MORRISSEY, J. H. Silver stain for proteins in polyacrylamide gels: a modified procedure with enhanced uniform sensitivity. *Analytical Biochemistry*, v. 117, p. 307-310, 1981.
- NAVE, C. A description of imperfections in protein crystals. *Acta Cryst.*, v. D54, p. 848-853, 1998.
- NAVE, C. Radiation damage in protein crystallography. *Radiat. Phys. Chem.*, v. 45, n. 3, p. 483-490, 1995.
- NEEDLEMAN, S. B.; WUNSCH, C. D.; A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.*, v. 48, n. 3, p. 443-453, 1970.
- NEVES-FERREIRA, A. G. C. *et al.* Isolation and characterization of DM40 and DM43, two snake venom metalloproteinase inhibitors from *Didelphis marsupialis* serum. *Biochim. Biophys. Acta*, v. 24987, p. 1-12, 2000.
- NEVES-FERREIRA, A. G. C. *et al.* New methodology for the obtainment of antithrombotic factors from the South American opossum (*Didelphis marsupialis*) and jararaca snake (*Bothrops jararaca*). *Toxicon*, v. 37, p. 1417-1429, 1999.
- NEVES-FERREIRA, A. G. C. *Imunidade natural: inibidores de metaloproteases de veneno de serpentes presentes no soro de Didelphis marsupialis (gambá)*. Rio de Janeiro, 1999. 130 p. Tese (Doutorado) – Departamento de Fisiologia e Farmacodinâmica, Fundação Oswaldo Cruz (FIOCRUZ).

- NICHELLS, A.; HONIG, B. A rapid finite difference algorithm utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *J. Comput.-Aided Mol. Design.*, v. 12, p. 345-440, 1991.
- NORRIS, G. E.; BAKER, H. M.; BAKER, E. N. Preliminary crystallographic studies on human apo-lactoferrin in its native and deglycosylated forms. *J. Mol. Biol.*, v. 209, p. 329-331, 1989.
- OLIVEIRA, P. S. L. *et al.* Crystallization of an anti-biothropic factor from *Didelphis marsupialis* (opossum) serum. *Resumos da XXII Reunião Anual da SBBq*, M-31, 1-4 de maio de 1993.
- ORENGO, C. A. *et al.* CATH - A hierarchic classification of protein domain structures. *Structure*, v. 5, n. 8, p. 1093-1108, 1997.
- OTWINOWSKI, Z; MINOR, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology*, v. 276A, p. 307-326, 1997.
- PERALES, J. *et al.* Isolation and partial characterization of an anti-biothropic complex from the serum of South American Didelphidae. *Toxicon*, v. 32, p. 1237-1249, 1994.
- PERSEPTIVE BIOSYSTEMS. The busy researcher's guide to biomolecule chromatography, 1996.
- PHARMACIA BIOTECH. *Ion exchange chromatography, principles and methods*, 1994.
- POLIKARPOV, I. *et al.* Set-up and experimental parameters of the protein crystallography beamline at the Brazilian National Synchrotron Laboratory. *J. Synchr. Rad.*, v.5, p. 72-76, 1998a.
- POLIKARPOV, I. *et al.* The protein crystallography beamline at LNLS, the Brazilian National Synchrotron Light Source. *Nucl. Instr. Meth.A*, v. 405, n.1, p. 159-164, 1998b.
- PROZYME – Specialty proteins & enzymes for diagnostics and research. <http://www.prozyme.com/glycopro/ge41tech.html>.
- ROSSMANN, M. G. The molecular replacement method. *Acta Cryst.*, v. A46, p. 73-82, 1990.
- SAKURAZAWA, S.; KUBOTA, T.; ATAKA, M. Orientation of protein crystals grown in a magnetic field. *J. Crystal Growth*, v. 196, p. 325-331, 1999.
- ŠALI, A. *et al.* Evaluation of comparative protein modeling by MODELLER. *Proteins*, v. 23, p. 318-326, 1995.
- ŠALI, A.; BLUNDELL, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, v. 234, p. 779-815, 1993.
- ŠALI, A.; SÁNCHEZ, R.; BADRETDINOV, A. MODELLER, a program for protein structure modeling, release 4. Manual, 1997.

- SANCHES, R.; ŠALI, A. Evaluation of comparative protein structure modeling by MODELLER-3. *Proteins*, p. 50-58, 1997. Suplemento 1.
- SAZAKI, G. *et al.* Effects of a magnetic field on the nucleation and growth of protein crystals. *J. Crystal Growth*, v. 173, p. 231-234, 1997.
- SCOPES, R. K. *Protein purification, principles and practice*. Springer-Verlag, 3rd ed., 380 p., 1994.

- SANCHES, R.; ŠALI, A. Evaluation of comparative protein structure modeling by MODELLER-3. *Proteins*, p. 50-58, 1997. Suplemento 1.
- SAZAKI, G. *et al.* Effects of a magnetic field on the nucleation and growth of protein crystals. *J. Crystal Growth*, v. 173, p. 231-234, 1997.
- SCOPES, R. K. *Protein purification, principles and practice*. Springer-Verlag, 3rd ed., 380 p., 1994.
- SNOW, M. E. A novel parameterization scheme for energy equations and its use to calculate the structure of protein molecules. *Proteins*, v. 15, n. 2, p. 183-190, 1993.
- SNYDER, G. A.; BROOKS, A. G.; SUN, P. Crystal structure of the HLA-Cw3 allotype-specific killer cell inhibitory receptor KIR2DL2. *Proc. Natl. Acad. Sci. USA*, v. 96, p. 3864-3869, 1999.
- SOMERS, W.; ULTSCH, M.; DE VOS, A. M.; KOSSIAKOFF, A. A. The X-ray structure of a growth hormone-prolactin receptor complex. *Nature*, v. 372, p. 478-481, 1994.
- STERNBERG, M. J. E. *Protein structure prediction, a practical approach*. Irl Press, 298 p., 1996.
- SVOBODA, K.; BLOCK, S. M. Biological applications of optical forces. *Annu. Rev. Biophys. Biomol. Struct.*, v. 23, p. 247-285, 1994.
- TANFORD, C. *The hydrophobic effect: formation of micelles and biological membranes*. John Wiley & Sons, 200 p., 1973.
- TARENTINO, A.L. *et al.* Molecular cloning and amino acid sequence of peptide-N4-(N-acetyl- beta-D-glucosaminy) asparagine amidase from flavobacterium meningosepticum. *J. Biol. Chem.*, v. 265, p. 6961-6966, 1990.
- VRIEND, G. What if: a molecular modeling and drug design program. *J. Mol. Graph.*, v. 8, n. 1, p. 52-56, 1990.
- VRIEND, G.; SANDER, C. Quality control of protein models: directional atomic contact analysis. *J. Appl. Cryst.*, v. 26, p. 47-60, 1993.
- WAGTMANN, N. *et al.* A new human gene complex encoding the killer cell inhibitory receptors and related monocyte/macrophage receptors. *Curr. Biol.*, v. 7, p. 615-618, 1997.
- WRIGHT, W. H. *et al.* Laser trapping in cell biology. *IEE J. Quantum Elect.*, v. 26, n. 12, p. 2148-2157, 1990.
- XIAO, B.; GAMBLIN, S. J. The effects of cryoprotectant on crystal stability. *J. Crystal Growth*, v. 168, p. 244-247, 1996.
- YAMAGISHI, A. Biological systems in high magnetic field. *J. Magn. Magn. Mater.*, v. 90-91, p. 43-46, 1990.
- YANAGIYA, S. *et al.* Effect of a magnetic field on the orientation of hen egg-white lysozyme crystals. *J. Crystal Growth*, v. 196, p. 319-324, 1999.