

Beatriz Tupinambá Freitas

Determinação das estruturas terciária e quaternária de uma lectina de sementes de *Canavalia gladiata* (Jacq.)DC na sua forma nativa e complexada com alfa-metil-manosídeo

Tese apresentada à coordenação do Curso de Pós-Graduação em Bioquímica da Universidade Federal do Ceará como requisito parcial para obtenção do grau de Doutor em Bioquímica.

Orientador: Prof. Dr. Benildo Sousa Cavada

Fortaleza – CE

2005

Determinação das estruturas terciária e quaternária de uma lectina  
de sementes de *Canavalia gladiata* (Jacq.)DC na sua forma nativa e  
complexada com alfa-metil-manosídeo

Beatriz Tupinambá Freitas

Fortaleza, 24/02/2005

*BANCA EXAMINADORA*

---

Prof. Dr. Benildo Sousa Cavada

UFC - Orientador

---

Prof. Dr. Marcos Hikari Toyama

UNESP- Sao Vicente/SP - Conselheira

---

~~Prof. Dr. Jorge Luiz Martins~~

IQG, UFPel, Pelotas/RS

- Conselheiro

---

~~Prof. Dr. André Herzog Cardoso~~

URCA - Conselheiro

---

Prof. Dr. Gandhi Radis Baptista

UFC - Conselheiro

## AGRADECIMENTOS

Agradeço ao Prof. Dr. Benildo Sousa Cavada e ao Magnífico Reitor Dr. André Herzog pelo apoio e compreensão obtidos durante a elaboração desta tese.

Ao Prof. Benildo agradeço, além disso, pela amizade que já está se tornando familiaridade e vem desde o século passado e pela extrema paciência com minha teimosia.

A todos os companheiros de laboratório que conviveram comigo durante esta dura jornada, Tatiane e Carlos, Dárcio, Emmanuel, Kyria, Rolando, Edson, Kátia, Creuza, Iza, Fernanda, Nagano, Taianá, Gustavo, Victor, Lucas, Karol, Cecília, Ramon, Eudismar e tantos outros.

Agradeço aos companheiros de trabalho com quem passei os momentos mais desafiadores e intrigantes dos últimos quatro anos: Bruno Anderson, Emmanuel e Plínio. Ao Plínio os meus agradecimentos são mais que especiais pois, além de me guiar nas descobertas me brindou com uma amizade franca e aberta.

Aos professores e funcionários ligados ao curso de Pós-Graduação em Bioquímica pela colaboração ao longo do curso.

Ao Dr. Francisco Javier Medrano Martin, Dr. Walter Filgueira de Azevedo Jr. E a Fenanda Canduri pela ajuda durante as madrugadas nos experimentos do LNLS.

A meus companheiros da Soka Gakkai pelos incentivos e orientações sempre oportunas e bem colocadas.

A minha mãe que sempre me apoiou e incentivou com seu jeito meigo e rigoroso nas medidas certas. Ao meu pai que me ensinou lições importantíssimas sem as quais, com certeza, não teria conseguido chegar até aqui. A meus irmãos e irmãs que tiveram um papel único nesta etapa da minha vida, cada um deles. A minha avó Nini que sempre me disse desde

criança da importância dos estudos.

A meus filhos, Ana Clara, Simone, José Luis e Ana Luisa pela força e obstinação que me induziram a ter de levar a cabo esta empreitada.

A meu companheiro e amigo Ulisses que, em todos os momentos, esteve a meu lado me incentivando e ajudando na jornada.

**Este Trabalho foi Realizado com o Auxílio das Seguintes Instituições:**

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), através da bolsa de Doutorado por auxílios de pesquisa concedidos ao BioMol-Lab (Laboratório de Moléculas Biologicamente Ativas do Departamento de Bioquímica e Biologia Molecular do Centro de Ciências da Universidade Federal do Ceará.

Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP) pelos convênios estabelecidos ao Curso de Pós-Graduação em Bioquímica do Centro de Ciências da Universidade Federal do Ceará.

Laboratório de Moléculas Biologicamente Ativas do Departamento de Bioquímica e Biologia Molecular do Centro de Ciências da Universidade Federal do Ceará, onde se encontra o laboratório de cristalização.

Laboratório de Sistemas Biomoleculares do Departamento de Física da Universidade Estadual Paulista (IBILCE) em São José do Rio Preto.

Ministério de Ciência e Tecnologia através de treinamentos oferecidos pela Rede Nacional de Proteômica, através do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), além do financiamento aos Projetos de Cristalografia e Bioinformática.

Laboratório Nacional de Luz Síncrotron (LNLS), onde foram realizados os experimentos de difração de Raios X.

Universidade Regional do Cariri – URCA pelo apoio concedido nas viagens.

*Mais valioso que o tesouro do cofre  
é o tesouro do corpo, e o tesouro do coração  
é o mais valioso de todos.*

*Nitiren Daishonin*

## RESUMO

Lectinas são proteínas que têm a capacidade de reconhecer e se ligar não covalentemente a carboidratos e/ou glicoproteínas de forma específica. Lectinas de Leguminosas têm sido estudadas amplamente não só com respeito a suas propriedades físico/químicas e biológicas, também com respeito à estrutura tridimensional de suas moléculas. Lectina de sementes de *Canavalia gladiata* (CGL) foi purificada e cristalizada utilizando o método da matriz esparsa para a pesquisa das condições de cristalização, com o objetivo de se determinar sua estrutura tridimensional. Cristais protéicos cresceram em 4 dias a 20 ° C usando o método da difusão de vapor em gota suspensa. Os cristais foram utilizados também para obtenção da estrutura da CGL complexada com o açúcar alfa-metil manosídeo utilizando-se a técnica de soaking. Usando Luz Síncrotron, um conjunto de dados de difração completo foi coletado a uma resolução de 1,99 Å para a lectina nativa e resolução de 2,22 Å para a lectina complexada. A estrutura da CGL, determinada por substituição molecular utilizando a lectina de *Canavalia ensiformis* como modelo, apresentou coeficiente de correlação de 93,0 % com  $R_{\text{fator}}$  chegando a 19,83 % e  $R_{\text{free}}$  a 23,87 %. Em seu modelo final, 85,9 % dos resíduos estão em regiões amplamente favoráveis no gráfico de Ramachandran e 14,1 % em regiões adicionalmente permitidas. A estrutura da CGL complexada com alfa-metil-manosídeo foi determinada por substituição molecular utilizando a estrutura obtida anteriormente de CGL como modelo. A estrutura final apresentou  $R_{\text{fator}}$  de 18,59 % e  $R_{\text{free}}$  de 24,57 %. O coeficiente de correlação chegou a 93,9 %.

## ABSTRACT

Lectins are proteins that have the ability to binding not covalently to carbohydrates and/or glycoproteins in a specific way. Lectins from Leguminous have been studied not only concerning to their physic/chemistry and biologics proprieties but in their three-dimensional molecules structures too. *Canavalia gladiata* seeds lectin (CGL) has been purified and crystallized using Sparse Matrix Method to research crystallization conditions in order to establish its three-dimensional structure. Protein crystals grown into 4 days at 20 ° C using vapour diffusion in sitting drops. Crystals were used to obtain the CGL complexed with alpha-methyl-mannoside by soaking. Using Synchrotron Light as source of X-rays, were obtained a diffraction data set at 1.99 Å resolution to CGL and 2.22 Å resolution to CGL complexed with alpha-methyl-mannoside. CGL structure was established with *Canavalia ensiformis* lectin as research model and showed a correlation coefficient of 93.0 % with  $R_{\text{factor}}$  19.83 % and  $R_{\text{free}}$  23.87 %. At its final model 85.9 % residues are in widely allowed regions in Ramachandran and 14.1 % in additional allowed regions. CGL with alpha-methyl-mannoside structure were molecular replacement determined with CGL structure as research model. The final structure showed  $R_{\text{factor}}$  18.59 % and  $R_{\text{free}}$  24.57 % and correlation coefficient 93.9 %.

## Sumário

Lista de Figuras .....	12
1. Introdução.....	15
1.1 Planta.....	15
1.2 Lectinas .....	17
2. Cristalização.....	24
2.1. Estrutura cristalina .....	25
2.2 Grupos espaciais.....	27
2.2.1 Grupo espacial cúbico.....	27
2.2.2 Grupo Espacial Tetragonal .....	28
2.2.3 Grupo Espacial Ortorrômbico.....	28
2.2.4 Grupo Espacial Monoclínico .....	29
2.2.5 Grupo Espacial Triclínico.....	29
2.2.6 Grupo Espacial Romboédrico ou Trigonal .....	30
2.2.7 Grupo espacial Hexagonal.....	30
2. 3 Direções e planos cristalinos .....	31
2. 4 Simetria .....	32
2. 4. 1 Centro de simetria .....	32
2. 4. 2 Eixo de simetria .....	32
2. 4. 3 Plano de simetria .....	33
2. 4. 4 Centro de inversão.....	33
2. 4. 5 Translação .....	34
2. 5 Crescimento de cristais.....	34
2.6 Pureza da amostra.....	35
2. 7 Formação dos cristais – nucleação .....	35
2. 7. 1 Nucleação não-assistida.....	36

	10
2. 7. 2 Nucleação assistida .....	37
2. 8 Diagrama de fases .....	37
2. 9 Métodos de cristalização de proteínas .....	40
2. 9. 1 Difusão de vapor .....	40
2. 9. 1. 1 Gota suspensa .....	41
2. 9. 1. 2 Gota sentada .....	41
2. 9. 1. 3 Gota sanduíche .....	42
2. 10 Screening.....	42
2. 11 Observação dos cristais.....	44
2. 12 Otimização .....	49
2. 12. 1 Aditivos .....	49
2. 13 Soaking e co-cristalização .....	50
2. 14 Difração .....	50
2. 14. 1 Danos causados pela radiação .....	51
2. 14. 2 Seleção de crio-protetor .....	52
2. 15 Fontes de Raios X.....	53
2. 15. 1 Difratômetros .....	54
2. 15. 2 Detectores .....	55
2. 15. 3 Fontes de luz síncrotron.....	56
2. 15. 3. 1 Princípios Teóricos .....	56
2. 15. 3. 2 Câmaras de vácuo .....	58
2. 15. 3. 3 Ímãs defletores ou dipolares.....	59
2. 15. 3. 4 Ímãs quadrupolares e sextupolares .....	59
2. 15. 3. 5 Cavidade de radiofrequência.....	59
2. 15. 3. 6 Linhas de luz.....	59
2. 15. 3. 7 Estações experimentais.....	60
2. 15. 3. 8 Blindagens.....	61

	11
2. 15. 3. 9 Inserções e trechos retos .....	61
2. 15. 3. 10 Fonte de luz síncrotron do LNLS.....	61
2. 15. 3. 11 Características da emissão das fontes de luz síncrotron .....	63
2. 15. 3. 11 .1 Laboratório Nacional de Luz Síncrotron.....	63
2. 15. 3. 11 . 2 Linha CPr – Cristalografia de Proteínas .....	64
2. 16 Coleta dos dados .....	64
2. 16. 1 Caracterização do cristal .....	64
2. 17 Tratamento dos dados.....	65
2. 18 Determinação das estruturas protéicas .....	68
2. 18. 1 Construção do mapa de densidade eletrônica.....	74
2. 18. 2 Refinamento do modelo.....	78
2. 19 Análise da estrutura.....	84
2. 19. 1 Análise dos erros numa estrutura determinada experimentalmente.....	84
2. 19. 2 Qualidade do modelo .....	85
2. 19. 3 As informações contidas nos arquivos PDB .....	86
3. Materiais e métodos.....	89
3.1 Cristalização .....	89
3. 2 Coleta e processamento de dados de difração de Raios X.....	89
3. 3 Substituição Molecular.....	89
3. 4 Refinamento.....	90
3. 5 Avaliação da estrutura.....	92
4. Resultados .....	93
4.1 Lectina de sementes de <i>C.glabriata</i> nativa.....	93
4. 2 Lectina de sementes de <i>C.glabriata</i> complexada com alfa-metil-manosídeo.....	96
5. Discussão.....	99
6. Conclusões .....	111
Referências.....	112

## LISTA DE FIGURAS

Figura 1 - Hábito de <i>Canavalia gladiata</i> (Facq.)DC. ....	15
Figura 2 - Sementes de <i>Canavalia gladiata</i> (Facq.)DC. ....	16
Figura 3 – Configuração dos grupos hidroxila em torno do C-3 e C-4 do anel piranosídeo de açúcares utilizada para a classificação das lectinas por MAKELA & MAKELA (1954). ....	19
Figura 4 – Exemplo esquemático de um cristal bidimensional. ....	26
Figura 5 - Parâmetros da rede cristalina. ....	27
Figura 6 - Representação esquemática do grupo espacial cúbico.. ....	27
Figura 7 – Representação esquemática do grupo tetragonal ....	28
Figura 8 – Representação esquemática do grupo espacial ortorrômbica.....	28
Figura 9 – Representação esquemática do grupo espacial monoclinico ....	29
Figura 10 – Representação esquemática do grupo espacial triclinico.....	29
Figura 11 – Representação do grupo espacial trigonal. ....	30
Figura 12 - Representação esquemática do grupo espacial hexagonal. ....	30
Figura 13 - Exemplo de direção cristalina.....	31
Figura 14 - Exemplo de eixo de simetria.....	33
Figura 15 - Exemplo de plano de simetria ....	33
Figura 16 – Exemplo de um centro de inversão aplicado à mão esquerda. ....	34
Figura 17 - Exemplo de translação aplicado à mão esquerda. ....	34
Figura 18 - Representação esquemática da incidência e reflexão de um feixe sobre um cristal com alta mosaicidade. ....	36
Figura 19 - Diagrama de fases para uma proteína típica. ....	38
Figura 20 - Diagrama de energia para cristalização.....	39
Figura 21 - Diagrama de fases mostrando zonas para nucleação e crescimento de cristais e precipitação da proteína.....	39
Figura 22 - Diagrama de fases para cristalização de proteínas por difusão de vapor. ....	40

Figura 23 - Representação esquemática de cristalização por difusão de vapor, técnica gota suspensa.....	41
Figura 24 - Representação esquemática de cristalização por difusão de vapor, técnica gota sentada. ....	41
Figura 25 - Representação esquemática da técnica de cristalização por difusão de vapor gota sanduíche. ....	42
Figura 26 - Cristais de urease de sementes de <i>Canavalia ensiformis</i> . ....	44
Figura 27 – Visão microscópica de uma gota clara .....	45
Figura 28 – Da Esquerda para a direita: precipitado e formação de gel.....	45
Figura 29 – Da esquerda para a direita: Formação de película e esferilutes.....	46
Figura 30 - Formação de cristal ao lado da separação de fases.....	46
Figura 31 - Formação de óleo. ....	47
Figura 32 - Cristais em diversas formações, da esquerda para a direita, precipitado microcristalino, agulhas e sólidos tridimensionais (na forma de cristais geminados). ....	47
Figura 33 - Cristais de proteínas corados com IZIT.....	48
Figura 34 - Padrão de difração de um sal. ....	48
Figura 35 - Hábitos cristalinos. ....	51
Figura 36 - Formação de anéis no padrão de difração de raios X ocasionado pela presença de cristais de gelo na amostra. ....	52
Figura 37- Gerador compacto de raios X de tubo selado.....	54
Figura 38 - Difratorômetro para uso em linhas Síncrotron KM6-4. ....	55
Figura 39 - Detectores CCD: Onyx e MarCCD.....	55
Figura 40 - Emissão por um elétron de baixa energia e elétrons com energia $E \gg m_0c^2$ .....	57
Figura 41 - Maquete de uma fonte de luz síncrotron (LNLS, Campinas).....	58
Figura 42 - Laboratório Nacional de Luz Síncrotron - SP.....	63
Figura 43 - Cabeçalho típico de um arquivo MTZ do programa Mosflm. ....	66
Figura 44 - Os dados de difração após a redução. ....	67

Figura 45 - Ilustração da influência da resolução dos dados no detalhamento observável na estrutura.....	77
Figura 46 - Alinhamento das seqüências primárias das lectinas de <i>Canavalia ensiformis</i> (ConA), <i>C.brasiliensis</i> (ConBr), <i>Dioclea grandiflora</i> (DGL), <i>D.guianensis</i> (Dgui), <i>Cratylia mollis</i> (Cra) e <i>Canavalia gladiata</i> (CGL).....	99
Figura 47 - Substituição do resíduo Ser21 por Asn no monômero da CGL. ....	100
Figura 48 - Substituição do resíduo Asp56 por Gly no monômero da CGL.....	100
Figura 49 - Substituição do resíduo Ala70 por Gly no monômero da CGL.....	101
Figura 50 - Substituição do resíduo Glu192 por Asp no monômero da CGL.....	101
Figura 51 - Estrutura tridimensional completa do tetrâmero de CGL. ....	102
Figura 52 - Dímeros da DGL, à esquerda, dímeros canônicas; à direita, não-canônicos. ....	102
Figura 53 - Folhas-beta de 6 segmentos e de 7 segmentos do monômero de CGL. ....	103
Figura 54 - Ligações de hidrogênio entre os resíduos Asn131 e Gln132 de um monômero e Thr123 e Asn124 do outro monômero de CGL. ....	104
Figura 55 - Corte longitudinal de dímeros não-canônicos de CGL.....	105
Figura 56 - "Loops" da CGL. à esquerda o "loop" 149-151 e à direita "loop" 160-163.....	106
Figura 57 - Sítio de ligação de metais. ....	106
Figura 58 - Gráfico de Ramachandran dos resíduos da lectina de sementes de <i>C.gladiata</i> .....	108
Figura 59 - Sobreposição das cadeias A dos tetrâmeros de CGL e da ConA (5CNA) .....	108
Figura 60 - Sobreposição dos tetrâmeros de CGL (amarelo) e da ConA (5CNA, azul) .....	109
Figura 61 - Posicionamento do açúcar alfa-metil-manosídeo no sítio de ligação de carboidratos. ....	110

## 1. INTRODUÇÃO

### 1.1 Planta

*Canavalia gladiata* (Jacq.) DC pertence à família Fabaceae, subfamília Faboideae, tribo Phaseoleae subtribo Diocleinae, também é classificada como Papilionaceae. Conhecida por muitos nomes em vários países, *fava contra* (Brasil); *dolic en sabre, pois sabre* (França); *haba de burro, poroto sable* (Espanha); *sword bean* (Estados Unidos); *Dao Dou* (Índia), entre outros. Sua ocorrência é ampla, é encontrada na Ásia, África, Caribe e continente Americano.



Figura 1 - Hábito de *Canavalia gladiata* (Jacq.)DC.

A planta é uma trepadeira lenhosa (Figura 1). Suas sementes são grandes, lenhosas e duras, geralmente de cor marrom e tendem a ser reniformes ou cordiformes. Têm entre 2 e 3,5 cm de comprimento, 1-2 cm de largura e 0,5-1,2 cm de espessura (Figura 2). A superfície exterior de seus grãos é de cor vermelha pálida ou vermelha púrpura, levemente enrugada e lustrosa. Um hilo em forma de olho de cerca de 2 cm de comprimento e 3 finas estrias, ocorre em um dos lados. Tem 2 cotilédones e apresenta um óleo branco-amarelado. Não tem cheiro e o gosto é parecido ao de feijão preto. As vagens podem ter 2,5 cm de comprimento (tendo tanto órgãos femininos como masculinos) e são polinizados por insetos.



Figura 2 - Sementes de *Canavalia gladiata* (Facq.)DC.

*Canavalia ensiformis* e *Canavalia gladiata* são muito parecidas. Em ambas espécies as vagens e as sementes são grandes e mais ou menos ovais. As duas espécies diferem principalmente na extensão do hilo da semente (cicatriz). O hilo de *C.gladiata* é mais que metade do comprimento das sementes, enquanto que da *C.ensiformis* é somente cerca de um terço do comprimento da semente.

Nenhum dos dois feijões tem importância comercial nos EUA (Institute of Food and Agricultural Sciences, Universidade da Florida, EUA). Ambos são conhecidos como comestíveis em estágio tenro e imaturo de suas vagens cozidas ou somente suas sementes. Existem registros que indicam média toxicidade de seus feijões quando ingeridos em grande quantidade. Por causa disto, a água utilizada em seu cozimento deve ser descartada para remover quaisquer substâncias tóxicas retiradas dos feijões. Companhias farmacêuticas têm demonstrado algum interesse nestes feijões como uma possível fonte da enzima urease (LUCA & REIS, 2001).

A composição química (Tibetan Medicine Encyclopaedia) de suas sementes inclui glicosídeos como canavaliósídeos e oito novos glicosídeos flavonóis acilados, gladiatosídeos A1, A2, A3, B1, B2, B3, C1 e C2 que foram isolados das sementes juntamente com robinina, kaempferol 3-O-beta-D-galactopiranosil-7-O-alfa-L-ramnopiranosídeo e kaikasaponina (MURAKAMI *et al*, 2000), guanidino oxiamina (HAMANA & MATSUZAKI, 1985) e uma nova poliamina, canavalmina (FUJIHARA; NAKASHIMA & KUROGOCHI, 1982). Além

destes compostos também há a ocorrência, em suas sementes imaturas, das giberelinas GA22 (MUROFUSHI *et al.*, 1969) e GA59 (YOKOTA & TAKAHASHI, 1981). Giberelinas (GAs) são um grupo de ácidos diterpenóides que atuam como reguladores de crescimento das plantas, influenciando uma série de processos de desenvolvimento de plantas superiores, incluindo alongação, germinação, dormência, floração, expressão sexual, indução de enzimas e envelhecimento de folhas e frutos (TAMURA, 1990).

O valor nutricional da farinha de suas sementes maduras foi avaliado por EKANAYAKE e colaboradores (2000) utilizando ratos Sprague-Dawley machos e uma dieta referência. A farinha *in natura* mostrou-se menos eficiente que a dieta padrão, no entanto, o processamento prévio dos cotilédones (autoclavados a seco ou torrados) provocou um incremento significativo no aumento de peso dos ratos indicando maior digestibilidade da farinha processada. Isto pode ser um indicativo da presença de proteínas tóxicas em suas sementes.

As sementes de leguminosas são consideradas como fonte barata de boa proteína e calorias que são facilmente disponíveis. No entanto, mesmo com este conhecimento, a utilização da maioria das leguminosas é aparentemente baixa. *Canavalia gladiata* tem favoráveis condições agronômicas favoráveis para cultivo nos trópicos e um grande rendimento se comparada à soja (EKANAYAKE, JANSZ & NAIR, 2000).

Entre seus usos populares, as sementes de *C. gladiata* são utilizadas como remédio contra mordeduras de serpentes. As sementes são esmagadas e marinadas com rum, o líquido é usado topicamente e também bebido e a massa das sementes espremidas, aplicada à mordedura.

## 1.2 LECTINAS

Lectinas são um grupo de (glico) proteínas de peculiar heterogeneidade (VAN DAMME *et al.*, 1995) com respeito a sua estrutura molecular, que têm em comum a habilidade de se ligar reversível e especificamente a carboidratos (mono ou oligossacarídeos)

e/ou glicoproteínas. A característica de reconhecimento específico a carboidratos, primeiramente descrita para a concanavalina A (ConA) em 1952 (WATKINS & MORGAN), levou ao esclarecimento da natureza química do sistema ABH dos grupos sanguíneos e esta mesma característica foi a responsável pela denominação “lectina” (do latim *legere* que significa *escolher*) em 1954 (BOYD & SHAPLEIGH).

Estruturas tridimensionais resolvidas através de difração de raios X de cristais protéicos de membros de diferentes famílias de lectinas vegetais e animais têm revelado uma ampla variedade de enovelamento destas lectinas e de arquiteturas de sítios de ligação a carboidratos. A despeito desta grande variabilidade, vários casos interessantes de evolução, tanto convergente como divergente dentro de famílias de lectinas animais e vegetais, podem ser notadas. Estas similaridades existem no dobramento das proteínas, na arquitetura dos sítios de ligação a carboidrato, como também nas estruturas quaternárias e podem ter origem de necessidades funcionais similares (LORIS, 2002).

As primeiras tentativas de classificação das lectinas utilizaram sua especificidade por monossacarídeos. Em 1954, MAKELA & MAKELA sugeriram que as lectinas poderiam ser classificadas em quatro grupos de acordo com a configuração relativa dos grupos hidroxila em torno dos carbonos 3 e 4 do anel piranosídico do açúcar (Figura 3): lectinas que se ligam a L-fucose seriam membros do grupo I como a lectina de *Lotus tetragonolobus*; lectinas com especificidade por N-acetil-D-galactosamina e/ou D-galactose (grupo II), como a lectina de soja (SBA); lectinas específicas por D-manose e/ou D-glicose (grupo III), como a ConA e a lectina de *Canavalia gladiata*, objeto do presente estudo. Lectinas pertencentes ao grupo IV, que reconheceriam os açúcares idose, gulose, glicose e L-xilose ainda não foram descobertas (VAN DAMME *et al.*, 1998<sup>a</sup>).

Os grupos criados por Makela & Makela foram modificados posteriormente (GOLDSTEIN & PORETZ, 1986; VAN DAMME *et al.*, 1998<sup>a</sup>) para lectinas que se ligam a manose, manose/glicose, manose/maltose, galactose/N-acetil galactosamina, N-acetil glicosamina/(GlcNAc)<sub>n</sub>, fucose e ácido siálico.

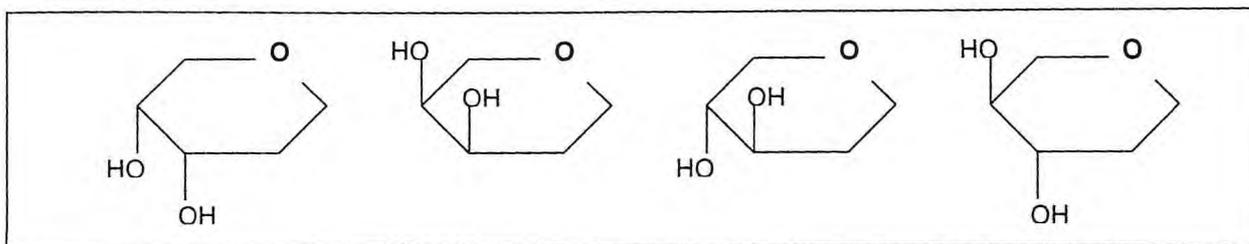


Figura 3 – Configuração dos grupos hidroxila em torno do C-3 e C-4 do anel piranosídico de açúcares utilizadas para a classificação das lectinas por MAKELA & MAKELA (1954).

Mais tarde começaram tentativas de classificação das lectinas não com base em sua especificidade, mas sim em sua estrutura. PEUMANS e colaboradores (1995) classificaram as lectinas em **merolectinas**, que consistem de um único domínio carboidrato ligante; **hololectinas**, contendo dois ou mais domínios de ligação a carboidrato iguais ou muito homólogos que se ligam ao mesmo açúcar ou a açúcares estruturalmente similares; **quimerolectinas**, consistindo de um ou mais domínios de ligação a carboidratos ligados a um domínio não relacionado que pode ser enzimático ou ter uma outra atividade biológica e que deve atuar independente do sítio de ligação ao carboidrato; em 1998 VAN DAMME e colaboradores introduziram uma 4ª categoria de lectinas, as **superlectinas**, que têm dois ou mais domínios de ligação a carboidratos, mas de especificidades diferentes entre si.

A abundância das lectinas no reino vegetal sugere que estas proteínas têm diversos papéis nestes organismos (GOLDSTEIN & PORETZ, 1986). Entre estes papéis podemos citar: fator de reconhecimento entre bactérias simbióticas fixadoras de nitrogênio e plantas hospedeiras (KARPUNINA, MEL'NIKOVA & KONNOVA, 2003); prevenção contra o ataque de fitopatógenos como fungos, insetos e animais (IP & LAU, 2004); como proteínas de armazenamento e como auxiliar na reprodução sexual em *Chlamydomonas* (SENGUPTA *et al.*, 1997), entre outros. Além disso, o fato da expressão de lectinas vegetais citoplasmáticas e/ou nucleares ser estimulada por estímulos bióticos ou abióticos indica que as lectinas estão envolvidas em interações endógenas proteínas /carboidratos (VAN DAMME *et al.*, 2004).

A análise das seqüências de lectinas de plantas disponíveis levou a sua classificação em

sete famílias de proteínas evolutivamente relacionadas: lectinas de leguminosas, lectinas de monocotiledôneas que se ligam a manose, lectinas contendo domínios heveínicos que se ligam à quitina, as RIP tipo 2, lectinas relacionadas a jacalina, lectinas da família das amarantinas e as lectinas do floema de Curcubitaceae (VAN DAMME *et al.*, 1986).

As lectinas de leguminosas constituem a família de lectinas de plantas mais estudada até hoje, seu estudo começou em 1890 com Power e Cambier (KOCOUREK, 1986) descrevendo o isolamento de uma lectina da casca de *Robinia pseudoacacia*. Em 1983, o primeiro gene de lectina isolado foi o de sementes de soja (VODKIN, RHODES e GOLDBERG, 1983). Novamente, a primeira lectina de planta a ser cristalizada (SUMMER e HOWELL, 1936) e ter suas estruturas primária e tridimensional resolvidas (EDELMAN, 1972; HARDMAN e AINSWORTH, 1972) foi de uma leguminosa, a lectina de sementes de *Canavalia ensiformis* (ConA).

Todas as lectinas de leguminosas são constituídas de protômeros de aproximadamente 30 kDa constituídos por uma única cadeia polipeptídica de aproximadamente 230 aminoácidos, conhecidas como lectinas “a uma cadeia”. Algumas vezes esta cadeia pode estar parcialmente clivada em dois peptídeos menores de tamanhos iguais ou não. Estas lectinas são chamadas lectinas de leguminosas “a duas cadeias”.

Estes protômeros mantêm entre si características semelhantes que permitem descrevê-los como formados por uma folha beta pregueada posterior quase plana, formada por cinco cadeias, uma outra folha beta dianteira que é curva e formada por sete cadeias e uma folha curta de cinco cadeias que se situa no topo da molécula. Vários “loops” fazem a conexão entre estas folhas beta (BANERJEE *et al.*, 1996).

A homologia entre as lectinas de leguminosas é grande, variando entre 28 a 99% de identidade de seqüência primária. Mesmo assim, a especificidade por açúcares entre estas lectinas é bem variável, compreendendo todas as classes de açúcares já descritas. As alterações na seqüência primária encontram-se principalmente nos “loops”.

Os aminoácidos dos sítios de ligação a metais e a carboidratos são altamente

conservados entre todos os membros desta família de lectinas. No entanto, pequenas alterações na seqüência de aminoácidos provocam grandes modificações na associação quaternária, embora não na estrutura terciária (SANZ-APARICIO *et al.* 1997).

O sítio de ligação a carboidratos de lectinas de leguminosas tem seus resíduos altamente conservados. As interações essenciais envolvem ligações de hidrogênio com três resíduos: um aspartato (Asp208, na ConA), que é precedido por uma ligação peptídica *cis* e mantido em sua posição através de ligação com um íon cálcio via água, uma asparagina (Asn14), que interage diretamente com o cálcio e o grupamento NH do esqueleto de Arg228 (na ConA e Gly em várias outras leguminosas), que também interage com o íon cálcio através de ponte com água e seu átomo de oxigênio do esqueleto. Interações de van der Waals entre resíduos aromáticos (Tyr12 na ConA, mas também Phe e Leu ou Cys em outras leguminosas) e o anel do açúcar também são essenciais para o reconhecimento. Outros tipos de interações envolvem um "loop" (Thr97-Glu102) cuja conformação e comprimento são variáveis dentro da família das leguminosas e não têm contato com o íon cálcio (SANZ-APARICIO *et al.* 1997).

Fazem parte da estrutura de lectinas de leguminosas dois sítios de ligação a metais que são essenciais para sua atividade de ligação a carboidrato (LORIS *et al.*, 1998). Nestes sítios metais divalentes,  $Mn^{2+}$  e  $Ca^{2+}$ , se ligam obrigando assim a cadeia polipeptídica a adotar uma conformação que ocasiona a construção do sítio de ligação a carboidrato. Esta conformação inclui a formação de uma ligação peptídica *cis* entre Ala207-Asp208. A distância entre os metais é de aproximadamente 4,5 Å e há a formação de uma ponte entre eles através do resíduo Asp10 (LORIS *et al.*, 1998).

Um sítio hidrofóbico também foi identificado em várias lectinas de leguminosas através da detecção da alteração do espectro de fluorescência de soluções de ácido 8-anilino naftaleno sulfônico (ANS) e ácido 2,6-toluidinil naftaleno sulfônico (TNS) quando adicionadas lectinas (ROBERTS & GOLDSTEIN, 1983). Titulações com TNS revelaram sítios de alta afinidade por este ligante em várias lectinas. A ligação de ANS a sítio hidrofóbico parece ser independente dos sítios de ligação a carboidrato já que a adição de açúcares não inibiu esta

ligação.

Algumas lectinas de plantas interagem com outras espécies de moléculas como, por exemplo, a fito hormônios, incluindo ácido indol acético (UMEKAWA, 1990) e com citocininas endógenas (PAPET *et al.*, 1990). Além desses, regiões de ligação a adenina também foram identificadas sobre a superfície de ConA e lectinas de *P.lunatus* (ROBERTS e GOLDSTEIN, 1983), entre outras. A exata função destes sítios de ligação é desconhecida, mas especula-se que as lectinas de leguminosas que se ligam à adenina/citocininas podem estar envolvidas no armazenamento de hormônios vegetais ou na regulação do crescimento. Estruturas de lectina de sementes de *Dolichos biflorus* (DBL), complexadas com adenina (1BJQ) (HAMELRYCK *et al.*, 1999) e da lectina dimérica de suas folhas [DB58 (BUTS, 2001), 1LUL (HAMELRYCK *et al.*, 1999)], dão os primeiros dados estruturais sobre estes sítios de ligação. Ambos os oligômeros possuem uma arquitetura não usual com uma alfa-hélice entre os dois monômeros. Em ambos os oligômeros esta alfa-hélice está diretamente envolvida na formação do sítio hidrofóbico (HAMELRYCK *et al.*, 1999).

Todas as lectinas de leguminosas são dímeros ou tetrâmeros constituídos de dímeros. A única exceção até então conhecida à regra é a arcelina 5a, uma proteína de defesa relacionada a lectina, de *Phaseolus vulgaris* da qual não se tem conhecimento de potencial ligação a carboidrato (HAMELRYCK *et al.*, 1996). A dimerização pode ser descrita em termos da disposição mútua das paredes constituídas por seis folhas beta pregueadas, dos dois monômeros. A presença de resíduos de características cruciais, como polaridade, hidrofobicidade e tamanho, em regiões de interface, determina o tipo de oligomerização que ocorrerá. Na maioria das vezes a dimerização envolve um arranjo lado a lado, resultando numa folha beta contígua de 12 membros com um eixo de simetria de ordem 2 perpendicular a folha beta (PRABU, SUGUNA, VIJAYAN, 1999). JONES e THORNTON (1995) descreveram este tipo de associação como sendo de tipo II. Oligomerização deste tipo foi primeiramente encontrada na ConA (SANZ-APARICIO *et al.*, 1997). Todas as lectinas que apresentam este tipo de oligomerização possuem resíduos com cadeias laterais pouco ou nada polares (Ser, Thr

ou Ala) na posição 66. Estes resíduos mantêm os dois monômeros unidos através de ligações fracas entre si como interações de Van der Waals ou ligações de hidrogênio. WBAI, GS4, PNA, que não formam este tipo de associação, têm resíduos carregados (Lys, Glu em WBAI e GS4, respectivamente) ou resíduos grandes e hidrofóbicos como Met, no caso de PNA (MANOJ e SUGUNA, 2001).

O dímero descrito acima se trata de uma estrutura conhecida como “dímero canônico das lectinas de leguminosas”. Duas lectinas diméricas de leguminosas de seqüências conhecidas não formam dímero canônico: *Erythrina coralloidendron* e a lectina IV de *Griffonia simplicifolia*. No caso da *E. coralloidendron*, a presença de um carboidrato covalentemente ligado a Asn17 causa o impedimento estérico que impede a formação do dímero canônico (SHAANAN, LIS e SHARON, 1991) No caso da GS-IV, a glicosilação também é o fator determinante. A lectina é formada por heterodímeros que diferem na ligação de carboidratos que se dá, em um dos dímeros, somente na posição Asn18, e no outro, também na posição Asn5 (DELBAERE *et al.*, 1993). A formação de um dímero canônico entre estas subunidades é impedida por um resíduo (Glu58) carregado posicionado na interface das subunidades.

Outro tipo de associação que ocorre entre monômeros é o que envolve associação das folhas beta pregueadas de seis cadeias dispostas frente a frente. Este tipo de associação pode ser classificada pelos tipos X1, X2, X3 e X4. Em todos os casos, a maioria dos contatos entre as subunidades se dá entre as cadeias 4<sup>a</sup>, 5<sup>a</sup> e 6<sup>a</sup> da folha beta posterior. Na ConA, além de dímeros tipo II, já vistos anteriormente, também são formados dímeros tipo X2. A presença de uma Arg na posição 210 evita a formação de interface X1, X3 e X4.

Esta associação se dá na formação dos tetrâmeros. No caso da ConA, é formada pela associação de dois dímeros canônicos com suas folhas beta pregueadas posteriores voltadas entre si. Cria-se, assim, uma grande concavidade que é preenchida por água. Vários resíduos carregados e cadeias laterais ionizáveis ficam escondidos pela disposição dos dímeros em tetrâmero e algumas podem formar pontes salinas.

A disposição do tetrâmero de ConA é levemente diferente da de ConBr. A única diferença na estrutura primária que ocorre na região de interface entre os dois dímeros é na posição 58 que é glicina na ConBr e Asp na ConA. Esta simples troca parece ser suficiente para fazer com que o tetrâmero seja mais aberto na ConBr. Uma ligação de hidrogênio entre Asp58 da subunidade A e Ser62 da subunidade C na ConA não se estabelece na ConBr e, assim, os resíduos correspondentes na ConBr têm seus carbonos alfa 1,5 Å mais próximos que na ConA (SANZ-APARICIO *et al.*, 1997). Afora esta substituição, só existe uma outra que se localiza num dos "loops" (67-71) que interage com a região 117-123 entre as subunidades AC e BD, Gly70 na ConBr que é alanina na ConA.

Tal diferença entre os oligômeros na ConA e ConBr leva a se propor que seria a responsável por diferenças nas atividades biológicas como indução de edema de pata (Bento *et al.*, 1993), liberação de macrófagos em ratos (RODRIGUEZ *et al.*, 1992) e estimulação de linfócitos humanos *in vitro* (BARRAL-NETTO *et al.*, 1992) e *in vivo* (BARRAL-NETTO *et al.*, 1996). De forma geral, lectinas de Diocleinae têm entre si alta homologia, mas apresentam diferenças em suas atividades biológicas.

As lectinas podem ser usadas como unidades reconhecedoras de glicoconjugados celulares apresentando assim, grande potencial para química médica já que estão envolvidas em uma vasta série de processos bioquímicos incluindo tráfico de glicoconjugados intra e extracelulares, iniciação do efeito de transdução de sinais como, por exemplo, regulação de crescimento e adesão celular nos animais (RUDIGER *et al.*, 2000). A relevância funcional das lectinas no desenho de reagentes de alta afinidade para reconhecimento de domínios carboidratos oferece a perspectiva de uma atrativa fonte de novas drogas.

## 2. CRISTALIZAÇÃO

Um cristal pode ser definido como um material homogêneo que apresenta um alto grau de simetria interna, na ordem de dimensões atômicas (Protein Crystallography Course). O cristal pode consistir de virtualmente qualquer composto químico único puro, embora possa

haver pequenas contaminações e mesmo co-cristalização de compostos. Os compostos podem ser inorgânicos ou orgânicos.

Os cristais protéicos não existem sem “líquido-mãe”, sua solução de cristalização. Se acontecer de um cristal protéico secar, ele será destruído. Isso ocorre porque as forças que mantêm as moléculas protéicas ligadas umas às outras, são todas fracas e de curto alcance.

A maioria dos cristais protéicos é formada por cerca de 50 % de solvente embora esta porcentagem possa variar entre 25 a 90% (McPHERSON, 1990). Isto cria largos canais de solvente que podem então facilitar a difusão de pequenas moléculas no interior do cristal dando origem a uma técnica chamada “soaking” onde, após o cristal formado, um ligante de pequenas dimensões pode ser adicionado na forma de solução ao líquido-mãe do cristal. Pode ocorrer então a difusão através do interior do cristal e suas interações com a proteína podem ser determinadas cristalograficamente.

Entre as forças que mantêm as moléculas protéicas em uma conformação específica e, também, posicionadas entre si num cristal, pode-se citar interações fracas como de Van der Waals, as forças de London ou forças de dispersão, que ocorrem entre dois dipolos induzidos, interações hidrofóbicas, ligações de hidrogênio entre outras. Embora todas estas interações sejam de curto alcance e fracas, em conjunto conseguem manter as moléculas o suficiente para que haja a formação de cristal.

Logicamente o cristal protéico, então, não terá dureza comparável com a de um cristal inorgânico, de fato para desespero do cristalógrafo, uma simples trepidação ou variação de poucos graus na temperatura já é suficiente para destruir completamente um cristal protéico.

## **2.1. Estrutura cristalina**

Um cristal ideal pode ser descrito como sendo um conjunto de átomos dispostos numa rede definida por três vetores linearmente independentes  $a$ ,  $b$ ,  $c$  chamados vetores fundamentais de translação de tal forma que, dois conjuntos semelhantes de átomos podem ser relacionados através de dois pontos com vetores  $r$  e  $r'$ :

$$r' = r + ha + kb + lc$$

com  $h$ ,  $k$  e  $l$  inteiros. Com  $r$  fixo, o conjunto de pontos que se obtém variando  $h$ ,  $k$  e  $l$  na equação recebe o nome de rede cristalina ou rede de Bravais.

A rede cristalina é uma abstração onde cada ponto pode representar um conjunto de átomos ou mesmo moléculas inteiras. A rede cristalina deixa então clara a disposição destes pontos no espaço.

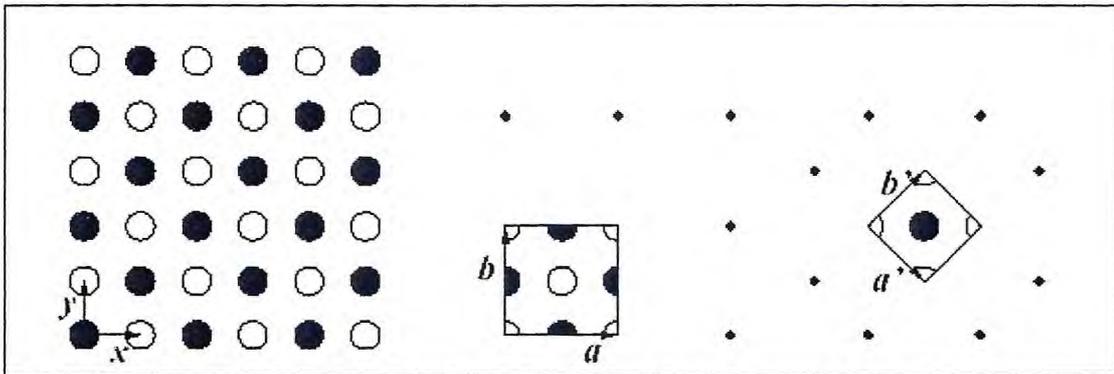


Figura 4 - Exemplo esquemático de um cristal bidimensional.

Na Figura 4 estão representados alguns vetores. Os vetores  $x$  e  $y$  não são fundamentais por que não ligam pontos equivalentes no cristal. Os vetores  $a$ ,  $b$ ,  $a'$  e  $b'$  são vetores da rede cristalina, pois ligam pontos equivalentes. Os vetores fundamentais são ditos primitivos quando conseguem representar, através de combinações lineares de seus coeficientes com números inteiros, quaisquer vetores da rede cristalina. Assim sendo,  $a$  e  $b$  não são vetores fundamentais ao passo que  $a'$  e  $b'$  o são. Pode-se obter o vetor  $b'$  a partir de  $a$  e  $b$ , mas só através de coeficientes fracionários:

$$b' = \frac{1}{2} a + \frac{1}{2} b$$

Ao paralelogramo obtido pelos vetores fundamentais primitivos dá-se o nome de **célula unitária primitiva**. A célula unitária primitiva copiada através de translações geradas por todos os vetores da rede, deve preencher todo o volume da rede cristalina, sem sobreposições ou espaços vazios. Deste modo, a célula unitária deve conter um, e apenas um, ponto da rede. A disposição da rede cristalina no espaço gera diferentes grupamentos espaciais.

## 2.2 Grupos espaciais

Definimos os módulos dos vetores fundamentais como  $a$ ,  $b$  e  $c$  e os ângulos entre eles como  $\alpha$ ,  $\beta$  e  $\gamma$ . A estes valores se dá o nome de parâmetros da rede cristalina (Figura 5). A importância destes vetores é que descrevem grupos de planos paralelos do retículo cristalino e, através de suas características, definem os grupamentos espaciais.

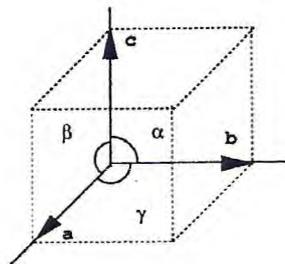


Figura 5 - Parâmetros da rede cristalina.

### 2.2.1 Grupo espacial cúbico

No grupo espacial cúbico os parâmetros da rede são todos iguais entre si, isto é:

$$a = b = c$$

$$\alpha = \beta = \gamma = 90^\circ$$

Há três subespécies de grupo cúbico: cúbico simples, cujos pontos estão dispostos como os vértices de cubos iguais; cúbico de corpo centrado, que ainda contém um ponto no centro do corpo de um dos cubos e o cúbico de faces centradas, formada pelos pontos que compõem o cúbico simples e ainda contém um no centro de cada uma das faces deste cubo (Figura 6).

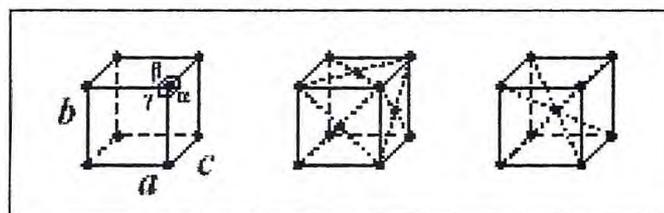


Figura 6 - Representação esquemática do grupo espacial cúbico. Da esquerda para a direita, grupo cúbico simples, de faces centradas e de corpo centrado.

### 2.2.2 Grupo Espacial Tetragonal

Se um dos eixos da célula unitária for alongado ou diminuído obtemos um grupo do tipo tetragonal.

Os pontos podem se dispor nos vértices dos prismas retos de base quadrada ou nos centros dos corpos destes prismas (grupo tetragonal de corpo centrado) (Figura 7). Assim sendo, seus parâmetros de rede são:

$$a = b \neq c$$

$$\alpha = \beta = \gamma = 90^\circ$$

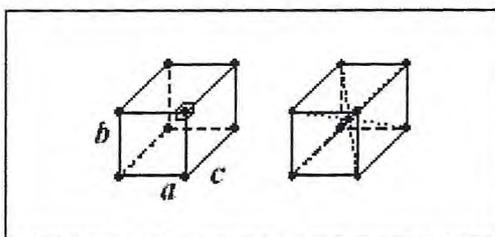


Figura 7 – Representação esquemática do grupo tetragonal. Da esquerda para a direita, simples e de corpo centrado.

### 2.2.3 Grupo Espacial Ortorrômbo

É obtido deformando-se não um, mas dois eixos da célula unitária. Os ângulos continuam perpendiculares entre si.

$$a \neq b \neq c$$

$$\alpha = \beta = \gamma = 90^\circ$$

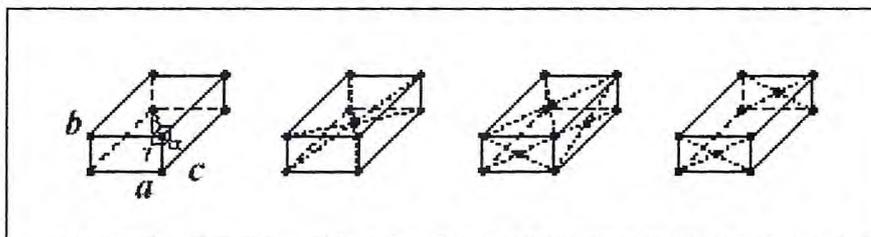


Figura 8 – Representação esquemática do grupo espacial ortorrômbo. Da esquerda para a direita: simples, de corpo centrado, de faces centradas e de bases centradas.

Este tipo de grupo espacial apresenta as três variantes já descritas, simples, de corpo

centrado e de faces centradas e ainda uma outra, de bases centradas, que é formada por pontos nos vértices de paralelepípedos iguais dispostos continuamente e dois pontos, nos centros de duas faces opostas (Figura 8).

#### 2.2.4 Grupo Espacial Monoclínico

Um grupo ortorrômbico deformado no valor de  $\gamma$  é conhecido como grupo espacial monoclínico que apresenta apenas as variantes simples e de bases centradas (Figura 9).

$$a \neq b \neq c$$

$$\alpha = \beta = 90^\circ \neq \gamma$$

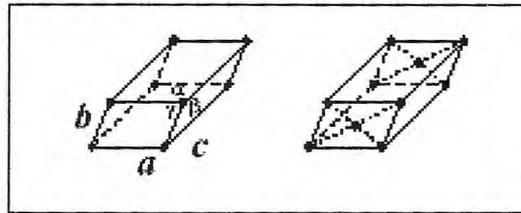


Figura 9 – Representação esquemática do grupo espacial monoclínico. Da esquerda para a direita, simples e de bases centradas.

#### 2.2.5 Grupo Espacial Triclínico

O grupo espacial triclínico é o que tem menos impedimentos a seus parâmetros de rede (Figura 10).

$$a \neq b \neq c$$

$$\alpha \neq \beta \neq \gamma$$

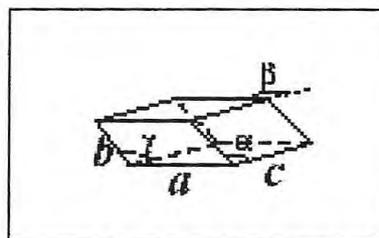


Figura 10 – Representação esquemática do grupo espacial triclínico.

### 2.2.6 Grupo Espacial Romboédrico ou Trigonal

Obtido pela deformação do grupo espacial cúbico na direção de uma das diagonais principais (Figura 11):

$$a = b = c$$

$$\alpha = \beta = \gamma < 120^\circ \text{ e } \neq 90^\circ$$

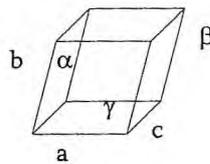


Figura 11 – Representação do grupo espacial trigonal.

### 2.2.7 Grupo espacial Hexagonal

São casos particulares do grupo espacial monoclinico em que  $\gamma = 120^\circ$  (Figura 12).

$$a = b \neq c$$

$$\alpha = \beta = 90^\circ, \gamma = 120^\circ$$

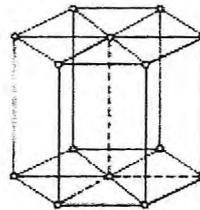


Figura 12 - Representação esquemática do grupo espacial hexagonal.

Quando a cela unitária possui unicamente pontos nos vértices, é chamada **primitiva**, sendo denotada pela letra P (exceção feita ao grupo espacial romboédrico ou trigonal, em que é empregada a letra R). Aos retículos P, juntam-se os sete sistemas **não-primitivos**, que possuem a mesma simetria, mas que comportam diversos pontos na cela unitária: são os retículos de corpo centrado, denominados I, os retículos de faces centradas, F, e os retículos de

bases centradas A, B ou C, estes últimos, segundo a natureza das faces que têm um ponto em seu centro. Todo este conjunto constitui o que é conhecido como “14 retículos de Bravais”, apresentados na Tabela 1.

Tabela 1 - Tabela Internacional para Cristalografia de raios X.

Grupo Espacial	Características das celas unitárias	Número de retículos Celas de Bravais	Tipo de retículos celas de Bravais
Cúbico	$a = b = c$ $\alpha = \beta = \gamma = 90^\circ$	3	P, I, F
Tetragonal	$a = b \neq c$ $\alpha = \beta = \gamma = 90^\circ$	2	P, I
Ortorrômbico	$a \neq b \neq c$ $\alpha = \beta = \gamma = 90^\circ$	4	P, C, I, F
Monoclínico	$a \neq b \neq c$ $\alpha = \gamma = 90^\circ \neq \beta$	2	P, C
Triclínico	$a \neq b \neq c$ $\alpha \neq \beta \neq \gamma$	1	P
Romboédrico ou Trigonal	$a = b = c$ $\alpha = \beta = \gamma < 120^\circ$ e $\neq 90^\circ$	1	R
Hexagonal	$a = b \neq c$ $\alpha = \beta = 90^\circ$ $\gamma = 120^\circ$	1	P

### 2. 3 Direções e planos cristalinos

Na cristalografia há uma notação, baseada na utilização de fases formadas com vetores fundamentais, que permite especificar posições, direções e planos num cristal.

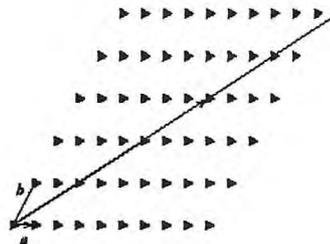


Figura 13 - Exemplo de direção cristalina.

Chamam-se direções cristalinas direções definidas por dois pontos da rede cristalina (Figura 13).

Considerando-se o vetor  $R$  como sendo o vetor desta direção cristalina,

$$R = ha + kb + lc$$

Na notação cristalográfica apresentam-se entre colchetes e sem qualquer caractere separador, os três valores,  $h$ ,  $k$  e  $l$ , que representam este vetor  $[hkl]$ . Se algum valor for negativo o sinal deve ser colocado sobre o índice respectivo  $[\bar{1}21]$ . Esta representação é conhecida como *índices de Miller*.

## 2. 4 Simetria

Materiais cristalinos têm simetria devido à ordenação de seus átomos ou conjunto de átomos, chamado *motivo*, na célula unitária e a maneira pela qual este motivo se repete a longo dos três eixos para formar o cristal inteiro. Elementos de simetria são centro, eixos e planos de simetria. Operadores de simetria que permitem definir toda a rede cristalina são os movimentos feitos em torno dos elementos de simetria.

### 2. 4. 1 Centro de simetria

O centro de simetria é um ponto ideal situado no interior do cristal que divide os eixos cristalográficos que passam por ele em duas partes iguais.

### 2. 4. 2 Eixo de simetria

É a reta que passa pelo centro de simetria e que, ao girar o cristal em  $360^\circ$ , se pode observar a mesma disposição de motivos em torno dela repetidos 2, 3, 4 ou 6 vezes (Figura 14). A ordem dos eixos é determinada por este número de repetições. Um eixo de ordem 2 existe quando há repetição do motivo 2 vezes durante o giro total do eixo; de ordem 3, quando se repete 3 vezes; de ordem 4, quando se repete 4 vezes e de ordem 6, quando se repete 6 vezes.

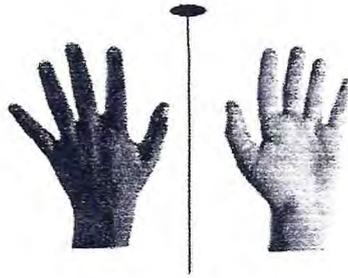


Figura 14 - Exemplo de eixo de simetria. Aplicando-se um eixo de simetria de ordem 2 à mão esquerda, a imagem produzida seria a própria mão esquerda, mas voltada com a palma para cima, resultado de um giro de  $180^\circ$  em torno do eixo de simetria.

### 2. 4. 3 Plano de simetria

Plano de simetria é o plano ideal que divide o cristal em duas partes iguais e simétricas (Figura 15).

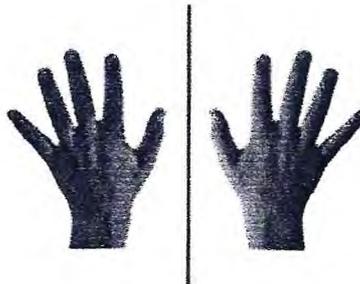


Figura 15 - Exemplo de operador no plano de simetria especular. As mãos direita e esquerda se relacionam através de um plano de simetria que cria uma imagem especular. Este operador de simetria não se aplica às proteínas devido à sua quiralidade.

### 2. 4. 4 Centro de inversão

Um centro de inversão relaciona ponto a ponto um objeto ou motivo com sua imagem equidistante de um ponto e invertida. Devido a quiralidade das moléculas protéicas o centro de inversão não se aplica a cristais protéicos (Figura 16).



Figura 16 – Exemplo de um centro de inversão aplicado à mão esquerda.

#### 2. 4. 5 Translação

A translação no espaço é outra operação de simetria que permite racionalizar redes periódicas como as de átomos em um cristal através de sua simetria (Figura 17).

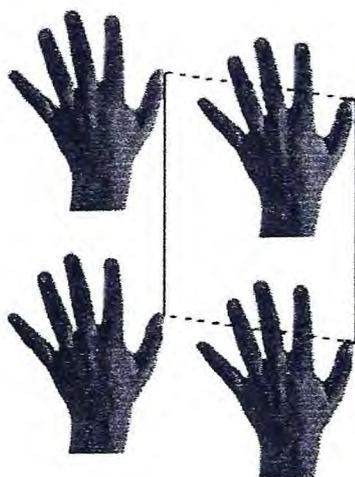


Figura 17 - Exemplo de translação aplicado à mão esquerda.

#### 2. 5 *Crescimento de cristais*

Em geral os cristais crescem somente em soluções supersaturadas. A concentração da

proteína deve ser alta inicialmente e sua solubilidade limite deve ser diminuída até abaixo desta concentração.

A solubilidade de uma proteína pode ser diminuída fazendo com que suas interações com o solvente se tornem menos favoráveis pela adição de agentes que competem com esta proteína pela água. As moléculas protéicas devem se ligar à água para se manterem em solução. Quando a interação destas moléculas protéicas com a água é diminuída elas irão se associar entre si e não com a água.

A cristalização pode ocorrer, após o fenômeno da nucleação, pela diminuição da solubilidade da proteína até um ponto suficientemente perto da supersaturação a fim de produzir a primeira aparência de uma rede cristalina (CREIGHTON, 1993)

## **2.6 Pureza da amostra**

A técnica de cristalização é utilizada para aumentar a pureza de substâncias já quase puras. A formação de cristais “empurraria” os contaminantes (pelo menos a maior quantidade destes), para fora da estrutura cristalina deixando-os então, em solução e forçando o soluto “limpo” a cristalizar.

Exatamente por isso é necessária uma amostra pura para que se possa obter um cristal. E, quando se fala de cristais protéicos, esta condição é uma das principais a serem atendidas para que se consiga obter um bom cristal.

A formação de cristais é ainda uma operação muito empírica, deste modo, às vezes, uma pequena impureza numa solução cristalina pode induzir a formação de cristais enquanto que a falta da pureza pode tornar a operação inviável.

## **2.7 Formação dos cristais – nucleação**

Os cristais começam a crescer através de um processo chamado de “nucleação”. A nucleação pode começar tanto com as próprias moléculas (chamada nucleação não-assistida), como pelo auxílio de alguma matéria sólida em solução (nucleação assistida).

Uma vez a solução esteja saturada, material sólido começa a se formar. Se as moléculas

se agrupam em arranjo aleatório, elas não ocupam o espaço de maneira otimizada. No entanto, se as moléculas se agrupam em seqüência ordenada, elas se empacotam em um espaço muito menor. Assim sendo, o empacotamento adequado utiliza menos espaço e é também de menor energia, o que sempre leva a uma condição mais estável. Quando isto acontece, este padrão de seqüência ordenada se repete regularmente nas três dimensões.

As faces de um cristal resultam do fato de que certas direções desta seqüência são mais acessíveis à ligação de novas moléculas. Porém, para que haja a correta aproximação de nova molécula é necessário um pequeno tempo para que esta possa se alinhar propriamente na superfície de um cristal crescente e este continue a crescer uniformemente. Daí advém a necessidade da cristalização ser lenta. Se a solução se torna super saturada de maneira que o sólido se forma rapidamente, as moléculas não terão tempo para se alinharem corretamente e os cristais, se obtidos, serão menores e grosseiramente formados.

A imperfeição da deposição das faces cristalinas é chamada de mosaicidade. Os parâmetros aceitáveis em um cristal para coleta de dados de difração de raios X é de até  $1^\circ$  (Figura 18).

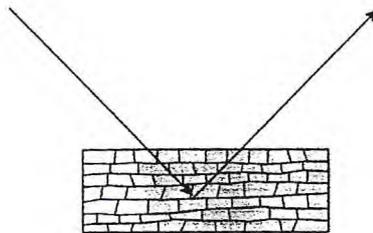


Figura 18 - Representação esquemática da incidência e reflexão de um feixe sobre um cristal com alta mosaicidade.

### 2. 7. 1 Nucleação não-assistida

Quando as moléculas do “soluto” (o material do qual os cristais são constituídos) estão

em solução, na maior parte do tempo só têm moléculas do solvente em torno de si. No entanto, ocasionalmente, podem se encontrar com outras moléculas de soluto. Se o composto é um sólido quando está puro, haverá algumas forças atrativas entre estas moléculas de soluto. A maior parte das vezes em que se encontrarem elas permanecerão juntas por algum tempo até que eventualmente outras forças as separem. Algumas vezes, duas moléculas se manterão juntas tempo suficiente para encontrarem uma terceira molécula de soluto e então uma quarta, e uma quinta etc.

Uma vez que estas cheguem a um certo número de moléculas juntas, o tamanho crítico, as forças atrativas combinadas entre as moléculas de soluto se tornam mais fortes que as outras forças na solução que tentam evitar a formação destes “agregados”. Esta é a ocasião onde este “protocristal” (um tipo de pré-cristal) se torna um sítio de nucleação.

Assim que este protocristal flutua na solução, encontra outras moléculas de soluto. Estas moléculas sentirão as forças atrativas do protocristal e se juntarão a ele. É assim que um cristal começa a crescer. Ele continua crescendo até que, eventualmente não possa mais permanecer “dissolvido” na solução e saia dela. Agora outras moléculas de soluto começam a crescer sobre a superfície do cristal e este continua a crescer até que um equilíbrio entre as moléculas de soluto no cristal e as que ainda estão em solução seja estabelecido.

### **2. 7. 2 Nucleação assistida**

É muito parecida com a nucleação não-assistida exceto que uma superfície sólida atua como um local para as moléculas de soluto se encontrarem. Uma molécula de soluto encontra uma superfície, se adsorve a esta superfície e permanece nela por um certo tempo antes que forças aleatórias da solução a retirem. Moléculas de soluto tenderão a se adsorver e agregar sobre a superfície. É aí que o protocristal se forma e o mesmo processo descrito anteriormente acontece.

### **2. 8 Diagrama de fases**

A solubilidade de uma proteína pode ser expressa em um gráfico chamado Diagrama

de Fases onde a concentração da proteína é grafada no eixo das ordenadas e nas abscissas é indicado um parâmetro que sofre variação, normalmente a concentração de um precipitante (Figura 19).

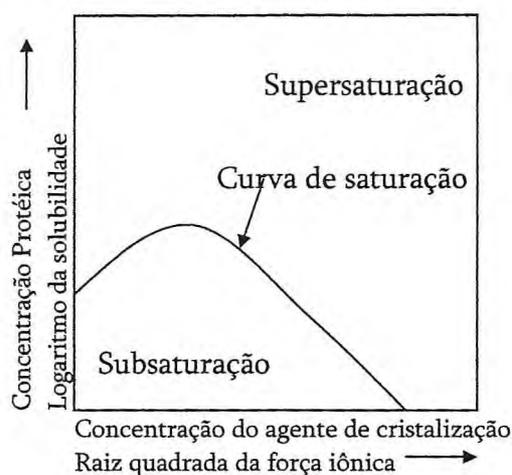


Figura 19 - Diagrama de fases para uma proteína típica.

A saturação ocorre quando a taxa de perda e ganho de ambas as fases, sólido e solução são iguais e o sistema está em equilíbrio. O fenômeno de salting-out é visto no lado direito do diagrama onde há uma redução na solubilidade da proteína conforme a concentração de sal aumenta. Salting-in (GREEN, 1931) é visto do lado esquerdo do diagrama onde há um aumento na solubilidade da proteína conforme a concentração de sal aumenta.

Um diagrama de fases também pode ser representado através do gráfico energia x tempo (Figura 20). Para que haja cristalização uma proteína deve ultrapassar uma barreira energética. Como o arranjo cristalino tridimensional é o arranjo de menor energia, o mais estável, a proteína se arranja em aglomerados até que se formam núcleos críticos que dão origem aos cristais diminuindo a energia do sistema.

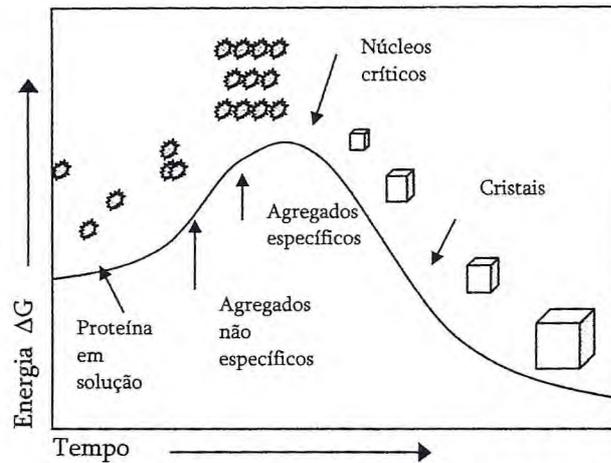


Figura 20 - Diagrama de energia para cristalização. Os núcleos críticos correspondem a maior energia intermediária. Quanto maior a maior barreira energética, menor taxa de nucleação.

Quanto mais supersaturada a solução protéica maior a probabilidade que um núcleo crítico se forme e menores são os núcleos necessários para induzir a formação do cristal.

Isto pode ser representado num diagrama de fases pela divisão da zona supersaturada em regiões de aumento da probabilidade de nucleação e precipitação (Figura 21).



Figura 21 - Diagrama de fases mostrando zonas para nucleação e crescimento de cristais e precipitação da proteína.

## 2. 9 Métodos de cristalização de proteínas

### 2. 9. 1 Difusão de vapor

No método de cristalização por difusão de vapor, pequenos volumes do precipitante e da proteína são misturados e, numa gota, equilibrados contra uma quantidade maior de solução contendo o precipitante ou outro agente desidratante. Com o passar do tempo, na tentativa de estabelecimento de um equilíbrio, a solução onde está a proteína começa a perder solvente aumentando assim tanto a concentração da proteína quanto a do precipitante na gota. A concentração da proteína entra então numa área metaestável e pode atingir a área de nucleação. Se o processo for lento, os cristais começam a se formar e a concentração de proteína cai. Se o processo for muito rápido, a área de nucleação é trespassada e ocorre a precipitação (Figura 22).

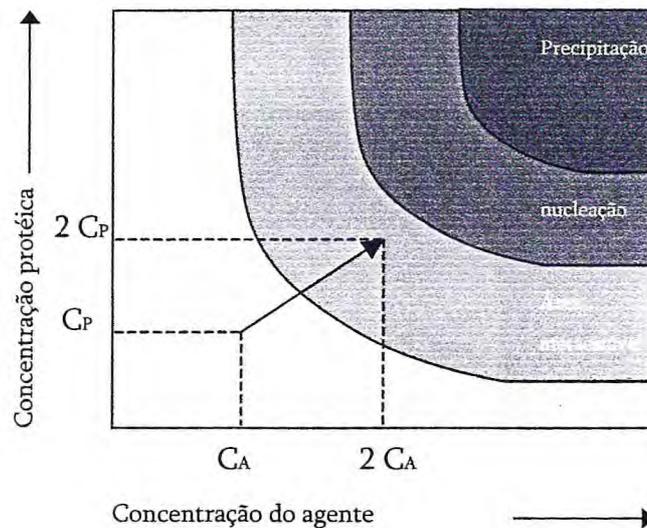


Figura 22 - Diagrama de fases para cristalização de proteínas por difusão de vapor.

Três técnicas são utilizadas para cristalização por difusão de vapor: gota suspensa, gota sentada e gota sanduíche.

### 2. 9. 1. 1 Gota suspensa

Na técnica da gota suspensa, a mistura de agentes precipitantes e/ou agentes desidratantes é colocada num reservatório. Uma gota contendo a solução protéica e uma quantidade de solução do poço é montada em uma lamínula siliconizada e esta é colocada invertida sobre o poço utilizando-se graxa de silicone para vedar suas beiradas (Figura 23).

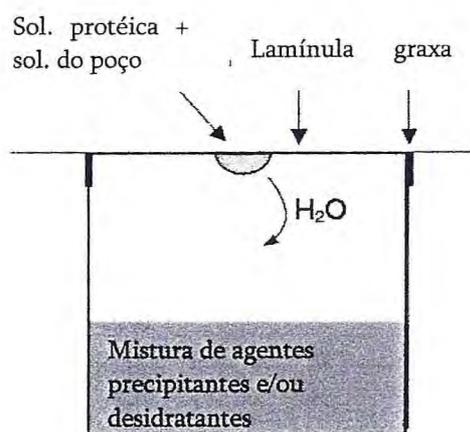
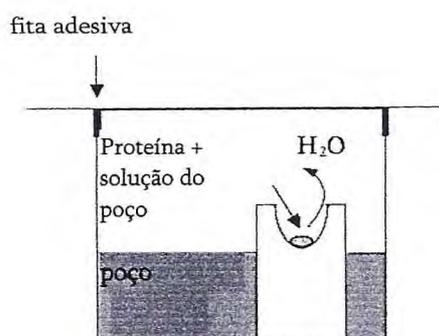


Figura 23 - Representação esquemática de cristalização por difusão de vapor, técnica gota suspensa.

### 2. 9. 1. 2 Gota sentada

Na técnica da gota sentada, a gota contendo a solução protéica e solução do poço é colocada num pequeno reservatório, este fica dentro do poço e, desta maneira, circundado pela solução. Os poços são cobertos por fita adesiva (Figura 24).

Figura 24 - Representação esquemática de cristalização por difusão de vapor, técnica gota sentada.



### 2. 9. 1. 3 Gota sanduíche

Na técnica de gota sanduíche a gota com a solução protéica e a solução do poço fica entre duas lamínulas tendo uma abertura para que a solução entre em equilíbrio com a solução do reservatório (Figura 25).

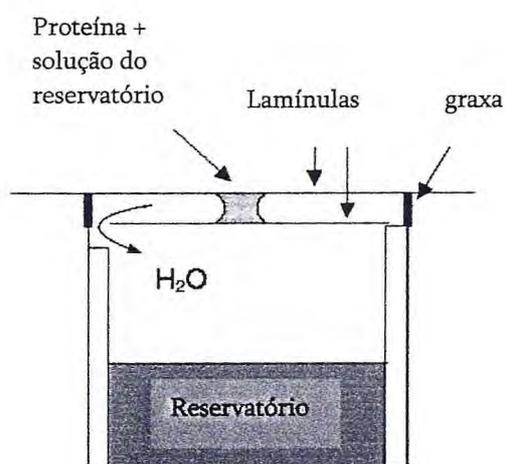


Figura 25 - Representação esquemática da técnica de cristalização por difusão de vapor gota sanduíche.

## 2. 10 Screening

A cristalização de macromoléculas ainda hoje tem bases principalmente empíricas (MCPHERSON, 1998; BERGFORS, 1999) e é um problema complexo que envolve a investigação de um grande número de condições físicas e químicas. Tampões, sais e aditivos usados para preparar a proteína estarão presentes em cada condição de cristalização. É imperativo que estas condições sejam definidas para o crescimento do cristal. Os métodos de

cristalização normalmente utilizados para os experimentos iniciais de identificação de condições de cristalização para proteínas são os de difusão de vapor de gota pendente e gota sentada (McREE, 1993).

É impossível sistematizar tentativas de uso de cada precipitante utilizado para cristalização de proteínas. Durante os últimos 20 anos numerosos esforços têm sido feitos em busca de condições de cristalização mais eficientes. O método de busca mais utilizado é o de matriz esparsa que foi desenvolvido através de dados disponíveis de condições favoráveis para cristalização (CUDNEY *et al.*, 1994; DeLUCAS *et al.*, 2003; JANCARIK & KIM, 1991).

No método de matriz esparsa três condições principais que afetam a cristalização foram escolhidas para serem utilizadas como variáveis: pH e tampões, aditivos e agentes precipitantes (JANCARIK & KIM, 1991). Estas condições são variadas gerando matrizes. A escolha de cada condição é feita levando-se em conta os casos positivos de cristalização expressos na literatura.

Existem alguns programas que preparam condições de cristalização de acordo com as características da proteína a ser cristalizada (SEGELKE, 2001; SEGELKE, 1995). Além disto, existe um banco de dados que permite a consulta das condições de cristalização de proteínas cristalizadas utilizando microgravidade em experimentos de cristalização no espaço e (Biological Macromolecule Crystallization Database – <http://wwwbmcdb.nist.gov:8080/bmcd/bmcd.html>) (GILLILAND *et al.*, 1994) está disponível na internet.

Outro método de busca de condições adequadas para cristalização de proteínas é o Método de Solubilização Ótima [Optimum Solubility Screen (JANCARIK *et al.*, 2004)]. Este método consiste de duas etapas. A primeira etapa é a seleção de um tampão melhor que o utilizado durante a purificação da proteína. Para isso são utilizados 24 tampões diferentes cobrindo uma faixa de pH que vai de 3 a 10. A solubilidade da proteína é primeiramente determinada pelo exame visual de pequenas quantidades de proteína misturadas a cada um dos tampões. Para tal avaliação pode-se utilizar um microscópio. Se os resultados deste

primeiro passo não forem satisfatórios, isto é, nenhum tampão lograr em solubilizar bem a amostra, um segundo passo explora uma variedade de aditivos químicos para aumentar a mono dispersividade da amostra protéica.

Alguns trabalhos foram feitos comparando vários métodos de busca de condições de cristalização de proteínas utilizando matriz esparsa. Comparação entre kits comerciais [Crystal Screen e Crystal Screen 2 (Hampton Research), Wizard Screens I e II (Emerald BioStructures) e Personal Structure Screens 1 e 2 (Molecular Dimensions)] foram feitos e quase todas as proteínas testadas (18 entre 19) foram cristalizadas (WOOH *et al.*, 2003). O estudo demonstrou que polietileno glicóis de alta massa molecular e misturas de diversos precipitantes foram os mais efetivos agentes cristalizantes.

## 2. 11 Observação dos cristais

As gotas com a solução protéica para cristalização devem ser inspecionadas não muito freqüentemente e, segundo alguns autores (McREE, 1993), a primeira leitura só deve ser feita uma semana após a preparação dos experimentos. Isto se faz necessário para evitar distúrbios que possam acarretar núcleos extras e malformação dos cristais.

Cristais podem ser identificados quando vistos por microscópio por serem brilhantes, pela característica de apresentarem birrefringência, ou seja, desviarem a luz incidente formando assim “imagens fantasmas”, e, na presença de luz polarizada, mudarem de cor quando girado o polarímetro ou a luz incidir em diferentes planos do cristal (Figura 26).

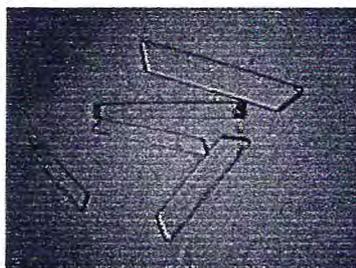


Figura 26 - Cristais de urease de sementes de *Canavalia ensiformis*.

Existem várias formações que podem ser vistas na ocasião, sozinhas ou em combinação.

Entre elas podem ser vistas: gotas claras, precipitados, géis, formação de películas, esferulitas, separação de fases, óleos e cristais.

As gotas claras indicam que a solução protéica não alcançou, ainda, a região de supersaturação (Figura 27). Neste caso pode ser tentada a adição de mais agente precipitante na solução do poço ou o aumento da concentração da proteína.

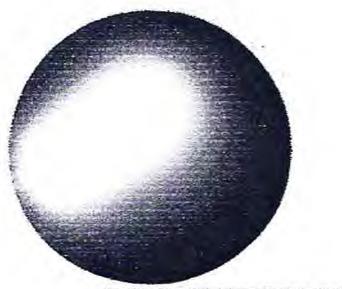


Figura 27 – Visão microscópica de uma gota clara. Foto de Johan Zeelen

A formação de precipitados indica que a zona de supersaturação foi alcançada com muita rapidez, não havendo tempo hábil para a formação de cristais ou, ainda, que a proteína está desnaturada (Figura 28, à esquerda).

Géis são formados por se formarem regiões muito densas na solução protéica. Podem se formar antes ou depois de se formarem cristais (Figura 28, à direita).

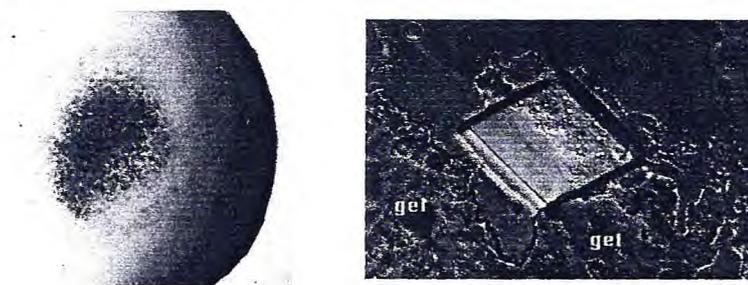


Figura 28 – Da Esquerda para a direita: precipitado e formação de gel. Fotos de Terese Bergfors.

O processo de formação de películas é um fenômeno ainda não compreendido. Cristais podem aparecer antes da formação das películas, mas não depois. (Figura 29, à esquerda)



Figura 29 – Da esquerda para a direita: Formação de película e esferulites. Foto de Terese Bergfors e Johan Zeelen, respectivamente.

Esferulitas são grupamentos birrefringentes transparentes e difíceis de se diferenciar dos géis (Figura 29, à direita).

A separação de fases é percebida pela formação de centenas de pequenas gotículas. Ocorre quando um dos agentes precipitantes se torna imiscível com outro agente (Figura 30). Os cristais podem crescer a partir da separação de fases.



Figura 30 - Formação de cristal ao lado da separação de fases. Foto de Terese Berfors.

A ocorrência da separação de fases na gota é como a precipitação no seguinte sentido: ambos são vistos na parte supersaturada do diagrama de fases. Não necessariamente deve ocorrer precipitação para que se possam obter os cristais, mas algumas vezes é necessário e mesmo inevitável que se tenha alguma precipitação para que os cristais ocorram. Lógico que é possível obter cristais sem que ocorra primeiro a precipitação e isto é preferível. O mesmo é verdade para a separação de fases. Embora geralmente não seja aconselhável encorajar a

formação de precipitados ou qualquer outra fase para se obter a nucleação, às vezes esta é a única maneira de se obter cristais (McPHERSON, 1999). Há casos relatados na literatura sobre deliberada indução à separação de fases para levar os cristais à nucleação (RAY AND BRACKER, 1986).

A formação de óleos se dá quando a proteína se separa da fase aquosa formando gotículas menores que as da separação de fases e menos agrupadas (Figura 31).

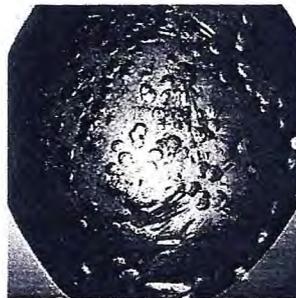


Figura 31 - Formação de óleo. Foto de Terese Berfors.

Os cristais podem se formar em 24h ou levar meses ou anos para aparecerem ou simplesmente nunca aparecerem. Podem se apresentar de várias formas, p.ex., como precipitado microcristalino, agulhas, placas ou sólidos tridimensionais (Figura 32).



Figura 32 - Cristais em diversas formações, da esquerda para a direita, precipitado microcristalino, agulhas e sólidos tridimensionais (na forma de cristais geminados).

Quando um cristal é produzido deve-se proceder à repetição das condições para obtenção de mais cristais. Se o cristal é reprodutível, deve-se sacrificar um para verificar se o cristal é de sal ou de proteína. Se não for reprodutível pode-se usá-lo para semear novos

cristais ou utilizá-lo diretamente para coleta de dados.

Para a verificação da natureza do cristal, se salino ou protéico, vários métodos podem ser utilizados. O mais fácil e mais definitivo deve ser tocar o cristal com um fino bastão de vidro. Se o cristal for protéico deve ser destruído facilmente. Outra maneira seria corá-lo com IZIT (Hampton Research Corp.) ou azul de metileno. Se o cristal for de proteína estes corantes, que são moléculas pequenas, se difundirão pelos canais de solvente sendo adsorvidos pelas moléculas protéicas (Figura 33). Como cristais de sais não têm canais de solvente grandes o suficiente para que as moléculas dos corantes se difundam, não são corados.

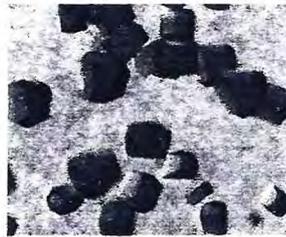


Figura 33 - Cristais de proteínas corados com IZIT.

A verdadeira natureza do cristal será conhecida na ocasião da coleta de dados. O padrão de difração de um cristal de sal é característico, poucos pontos e bem distantes e intensos, diferindo do padrão de proteína que mostra vários pontos próximos uns dos outros (Figura 34).

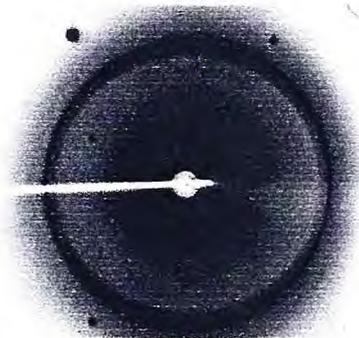


Figura 34 - Padrão de difração de um sal mostrando também anéis devido à formação de gelo no cristal.

## 2. 12 Otimização

No caso de serem obtidos micro cristais ou mesmo cristais não bem formados, pode ser tentada a otimização das condições de cristalização. Estas otimizações podem ser feitas variando levemente as condições que produziram os primeiros cristais tais como:

- Concentração do precipitante
- pH – em alguns casos a variação de 0,1 unidade de pH é suficiente para impedir a cristalização.
- Concentração final da proteína na gota – pode ser variada simplesmente variando-se as quantidades relativas de solução protéica e solução do poço na gota.
- Temperatura – normalmente trabalha-se a 4, 8 e 20 ° C.
- Método – gotas maiores podem dar origem a cristais maiores já que há mais proteína na gota e a velocidade de equilíbrio é mais baixa.

Uma outra estratégia que pode ser tentada é a utilização de óleo sobre a gota montada no intuito de diminuir a velocidade de equilíbrio (CHAYEN, 1997).

### 2. 12. 1 Aditivos

A maneira como pequenas moléculas afetam a cristalização de macromoléculas é, ainda, totalmente empírica, não há como se saber ao certo seu mecanismo de atuação (CUDNEY *et al.*, 1994). O que parece acontecer é que essas pequenas moléculas perturbam as interações proteína-proteína e proteína-solvente além das interações entre as moléculas de água.

Algumas substâncias podem ser adicionadas com o intuito de melhorar os cristais obtidos. Entre estas substâncias estão:

- Glicerol – pode diminuir a quantidade de núcleos formados dando origem a cristais maiores e com a vantagem adicional de ser um crioprotetor. Normalmente usa-se de 1 a 25 % de glicerol (Sousa, 1995).
- etanol ou dioxano – envenenam os cristais e cessam a nucleação excessiva

(DUCRUIX e GIEGE, 1992).

- Íons metálicos divalentes (TRAKHANOV e QUIOCHO, 1995).

## **2. 13 Soaking e co-cristalização**

Antes da coleta de dados de raios X, os cristais protéicos são muitas vezes embebidos em soluções para estabilizá-los ou modificá-los. A técnica de soaking é utilizada para preparar derivativos isomórficos, ligar pequenas moléculas a um sítio ativo ou para proteger o cristal de congelamento. Este processo pode induzir mudanças de parâmetros conformacionais e/ou de sua rede cristalina que podem levar a quebra do cristal, degradação da qualidade de difração ou, menos freqüentemente, melhora da difração.

Quando o ligante ou inibidor for uma molécula muito grande ou mesmo a proteína sofrer grandes mudanças conformacionais com sua ligação, é preferível tentar a co-cristalização, isto é, cristalizar a proteína já complexada com o ligante ou inibidor.

## **2. 14 Difração**

A qualidade da difração de um cristal é o passo decisivo para o estabelecimento da estrutura de uma proteína. Se o cristal não difratar bem os dados não serão bons e, conseqüentemente, a estrutura será prejudicada.

Não há maneira de se saber de antemão se um cristal protéico vai difratar bem (Figura 35). A qualidade de seu padrão de difração dependerá de vários pontos:

- tamanho do cristal – o espalhamento será proporcional ao número de células unitárias no cristal e este número é proporcional ao volume do cristal de tal forma que, dobrando todas as dimensões de um cristal cúbico, a difração será 8 vezes mais intensa.

- ordenação – a qualidade do padrão de espalhamento, ou difração, dependerá de quão idênticas cada célula unitária seja. Quanto mais parecidas mais intenso será o espalhamento.

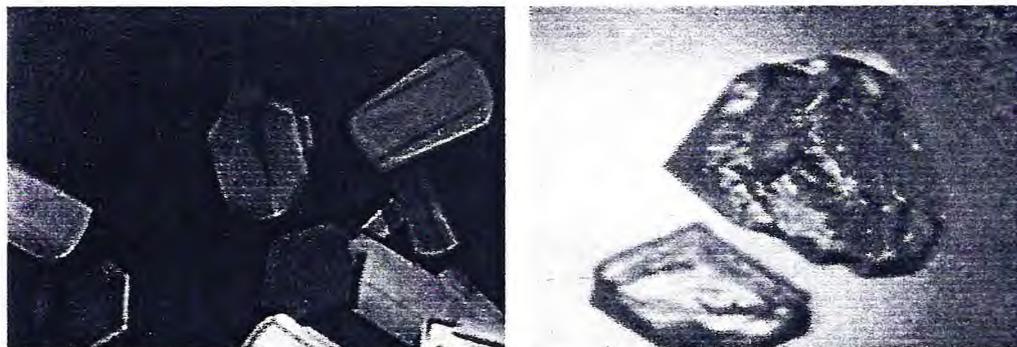


Figura 35 – Hábitos cristalinos. À esquerda cristais que não difrataram e à direita cristal que difratou a 1,6 Å.

Com isto fica claro que o tamanho do cristal não é um fator decisivo e sim, o grau de ordenação interna de suas moléculas na célula unitária.

### 2. 14. 1 Danos causados pela radiação

A incidência de radiação sobre um cristal protéico pode causar sua destruição (GONZALES AND NAVE, 1994) ou desordenamento devido ao aumento de temperatura na superfície do cristal. Determinada quantidade de raios X deve passar através de um cristal, a despeito de seu tamanho, para produzir pontos de difração de intensidade suficiente (HENDERSON, 1990). Quanto menor o tamanho de um cristal, maior a exposição de raios X por célula unitária. Se o cristal for muito pequeno a radiação pode destruí-lo mesmo antes de se obter uma exposição adequada para a coleta de dados.

Uma maneira de se evitar a destruição do cristal pela radiação é utilizando-se temperaturas baixas para a coleta de dados (HOPE, 1988). No entanto, se um cristal protéico for congelado, sua molécula pode se desnaturar graças à formação de cristais de gelo no laço utilizado para montar o cristal para coleta ou mesmo no solvente que se encontra no interior da molécula protéica.

Água congelada pode ter diferentes conformações ou estruturas cristalinas dependendo da pressão e temperatura. A maioria dos experimentos de crio-cristalografia de proteínas acontece à pressão atmosférica e entre 90-110 K, nestas condições a forma da água congelada é de gelo hexagonal  $I_h$ .

Estes cristais provocarão o aparecimento de anéis nos padrões de difração (Figura 36) o que, mesmo sem a destruição da estrutura protéica, dificultará a coleta de alguns pontos de difração do padrão. Além disso, os parâmetros de célula do cristal protéico podem ser alterados pelo gelo.



Figura 36 - Formação de anéis no padrão de difração de raios X ocasionado pela presença de cristais de gelo na amostra (Disponível em <http://www.riboworld.com/jmh/jmhs-hexaice.html#anhiceeff>).

O uso de crio-protetores ou de resfriamento instantâneo pode evitar a formação de gelo no solvente. A água no solvente se encontrará num estado chamado “gelo vítreo”, um material amorfo como vidro e que não difrata (BOUTRON, 1987).

## 2. 14. 2 Seleção de crio-protetor

A escolha de um crio-protetor, como tudo mais em cristalografia de proteínas, envolve tentativa e erro. No caso de um polímero ser utilizado no líquido mãe da cristalização, pode-se aumentar sua concentração para 15-30 % e assim utilizá-lo como crio-protetor. Se solventes orgânicos voláteis, tais como pentanodiol, butanodiol (MEHL, 1989), propanol ou butanol, são utilizados, pode-se experimentar aumentar sua concentração.

Se nenhuma destas substâncias é utilizada, o primeiro crio-protetor a ser testado deve

ser o glicerol já que é o mais citado na literatura (GARMAN & MITCHELL, 1996). Polímeros, como polietileno glicóis; ácidos orgânicos, como malonatos; sais de halogênios em alta concentração e solventes orgânicos voláteis e não voláteis podem ser utilizadas como crioprotetores. No comércio são vendidos kits com soluções de crioprotetores já preparadas para serem utilizados nos experimentos de cristalização de proteínas (CryoPro - Hampton Research Corp., por exemplo).

Uma das maneiras de se testar um crioprotetor é misturando-o com o líquido mãe, emergindo um laço nesta mistura e colocando-o no spray criogênico para verificar se o crioprotetor foi eficiente e evitou o aparecimento de cristais de gelo (PARKIN & HOPE, 1998). Esta verificação se dá simplesmente montando-se o laço no goniômetro e permitindo-se que o feixe de raios X incida sobre ele. O aparecimento dos anéis característicos de gelo no detector indica que a concentração do crioprotetor não foi suficiente.

Outro crioprotetor é óleo (KWONG & LIU, 1999), Paratone ou Paraffin (Hampton Research Corp.) aonde o cristal será mergulhado para remoção de água de sua superfície prevenindo assim a formação de gelo na camada aquosa sobre a superfície do cristal e possíveis fraturas. O procedimento é mergulhar o cristal numa pequena gota de óleo. O óleo recobre a superfície protegendo o cristal de perda de água interna e desidratação excessiva (PETCOCK *et al.*, 2001). Deve ser evitado se deixar o cristal em contato com o óleo por tempo excessivo, pois o óleo pode absorver a água e desidratar o cristal. Normalmente não é necessário se remover o excesso de óleo do cristal para que se proceda à coleta de dados (SCHNEIDER, 1997).

## **2. 15 Fontes de Raios X**

Os raios X podem ser gerados através de dois métodos: acelerando elétrons com alta diferença de potencial contra um anteparo metálico (radiação de frenagem) ou usando radiação síncrotron emitida por elétrons ou pósitrons em anéis de armazenamento de alta voltagem.

No primeiro método podem ser utilizados dois tipos de geradores: fonte de tubo selado

e ânodo rotatório. Ambos produzem raios X acelerando elétrons a uma voltagem de 40-50 KV contra um anteparo metálico. A limitação destas fontes é a produção de calor no anteparo. Um tubo selado típico não pode operar a mais que 20 mA de corrente ou 0,8 KV a 40 KV. Um anodo rotatório pode alcançar 8 KW.

### 2. 15. 1 Difratoômeros

Os difratômetros (Figuras 37 e 38) utilizam uma geometria equatorial no qual o raio difratado se mede sempre em um plano horizontal definido pelo raio incidente e a rotação ao redor de um eixo que atravessa o cristal. O detector só se move neste plano e forma um ângulo de  $2\theta$  (entre 0-160°, dentro do limite do instrumento) com o raio incidente. O goniômetro mais utilizado é o de geometria euleriana que dá lugar ao difratômetro de quatro círculos com três rotações para o cristal ( $\omega$ ,  $\chi$ ,  $\phi$ ) e uma para o detector ( $2\theta$ ). Assim, se  $\chi$  é perpendicular ou não a  $\omega$ , o goniômetro se denomina de geometria Euler ou Kappa, respectivamente.

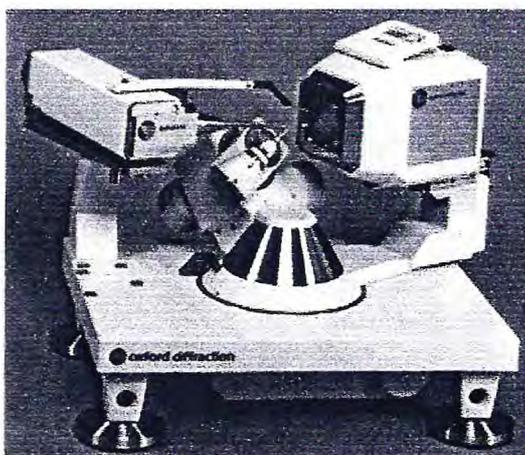


Figura 37- Gerador compacto de raios X de tubo selado com anteparo de cobre Cu Enhance, montado numa base Xcalibur 3 com CCD Sapphire acoplado. Ao centro vê-se o aparelho goniométrico <http://www.oxford-diffraction.com/>

O goniômetro gira e orienta o cristal de forma que cada família de planos cumpra a condição de difração de Bragg. A cabeça goniométrica que produz a rotação  $\phi$  (tipo spin) no cristal, descansa sobre o bloco  $\kappa$  suportado pelo bloco  $\omega$  que por sua vez é colinear com  $2\theta$ . O

centro do sistema (onde se coloca o cristal) e o detector estão em um plano perpendicular ao eixo  $\omega/2\theta$ .

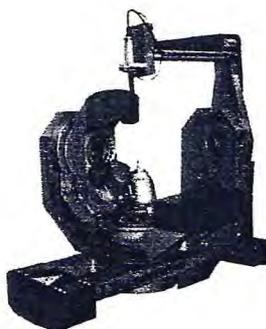


Figura 38 - Difratorômetro para uso em linhas Síncrotron KM6-4.

## 2. 15. 2 Detectores

Atualmente o dispositivo mais moderno e prático para a detecção e registro dos pontos de reflexão são as placas de imagens (Figura 39). Quando estas placas de imagens são expostas aos raios, os fótons provocam uma mudança química no revestimento da placa que libera uma fluorescência que, por sua vez, é detectada por um foto-amplificador quando este é iluminado com luz de comprimento de onda apropriado. Estes dados então são coletados em formato de imagens.

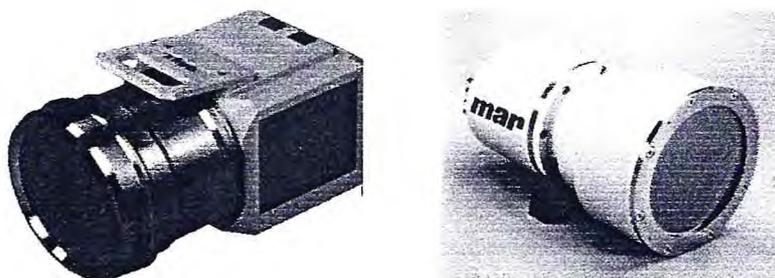


Figura 39 - Detectores CCD: Onyx e MarCCD da esquerda para a direita, respectivamente.

## 2. 15. 3 Fontes de luz síncrotron

Luz síncrotron é a intensa radiação eletromagnética produzida por elétrons de alta energia num acelerador de partículas. Abrange uma ampla faixa do espectro eletromagnético: Raios-X, Luz Ultravioleta e Infravermelha, além da Luz Visível.

A radiação síncrotron é altamente vantajosa para uso na cristalografia de proteínas por várias razões; uma resolução alta como resultado da alta intensidade do feixe, a facilidade de mudança de comprimento de ondas de um feixe monocromático, linha de fácil uso e controle.

### 2. 15. 3. 1 Princípios Teóricos

Sempre que uma carga elétrica é acelerada, da mesma forma que elétrons oscilando numa antena, ela emite radiação eletromagnética. Quando por exemplo um elétron sofre, pela ação de um campo magnético constante, uma aceleração perpendicular à sua velocidade, sua direção de movimento varia adotando uma trajetória circular. Se a energia cinética do elétron é bastante menor que sua energia de repouso,  $m_0c^2$  ( $m_0$ : massa em repouso do elétron,  $c$ : velocidade da luz no vácuo), a emissão de radiação é feita de acordo com uma distribuição angular esquematizada na parte superior da Fig. 40 (WINICK, 1994). Quando a energia cinética, ou a total, do elétron,  $E$ , é bastante superior à energia em repouso ( $E \gg m_0c^2$ ), a emissão de radiação se torna muito intensa e se concentra dentro de um cone estreito ao redor da direção do vetor velocidade (Figura 40), com uma abertura angular

$$\alpha = \frac{1}{\gamma} \quad \text{onde}$$

$$\gamma = \frac{E}{m_0c^2} \quad \text{e} \quad m_0c^2 = 0,511 \text{ MeV}$$

O ângulo alfa é da ordem de apenas 0,5 mrad para a parte central do espectro de emissão de um elétron com energia de 1 GeV.

A radiação ou luz síncrotron é produzida por aceleradores anulares nos quais circula uma corrente pulsada de elétrons de alta energia ( $E \gg m_0c^2$ ) com velocidade praticamente igual à velocidade da luz. A energia típica dos elétrons nas fontes de luz síncrotron

atualmente em funcionamento (aproximadamente 50 no mundo) vai de 400 MeV até 7 GeV. A radiação produzida em qualquer uma dessas fontes é bastante colimada já que, de acordo à equação anterior, a divergência vertical (ângulo alfa) é menor de 1m.rad. A 10 m da fonte a altura do feixe é de aproximadamente 10 mm para elétrons de 500 MeV e de menos de 1 mm para elétrons de 7 GeV.

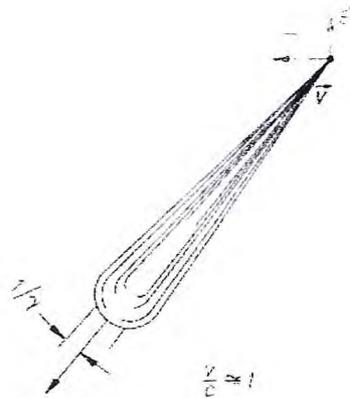


Figura 40 - Na parte superior se esquematiza a emissão por um elétron de baixa energia e na parte inferior a emissão por elétrons com energia  $E \gg mc^2$ .

Os elétrons são injetados no acelerador anular possuindo já uma certa energia que na prática varia, dependendo da fonte de luz, de um valor da ordem de 10 % da energia final (anel com injeção a baixa energia) até um valor igual à energia final (injeção a energia completa). Em geral os elétrons adquirem essa energia inicial num acelerador linear (LINAC), ou num microtron, para depois passar a um acelerador circular intermediário (“booster”) onde os elétrons ganham energia adicional antes de ser injetados no anel. Em fontes como a do LNLS (Laboratório Nacional da Luz Síncrotron), a injeção de elétrons é feita à baixa energia mediante um acelerador linear injetor e uma linha de transporte que os encaminha diretamente ao anel de armazenamento onde geram, nas partes curvas da trajetória, a luz síncrotron.

Na Figura 41 está a maquete da fonte de luz do LNLS, na qual se ilustram as suas partes principais: o LINAC injetor de 100 MeV no subsolo, a linha de transporte desde o

subsolo até o térreo, o anel de armazenamento de 1,37 GeV com doze ímãs defletores (em azul), uma linha de luz (na parte inferior da maquete) com sua correspondente estação experimental para a utilização pelos usuários e duas saídas de luz vizinhas para a instalação de estações adicionais.

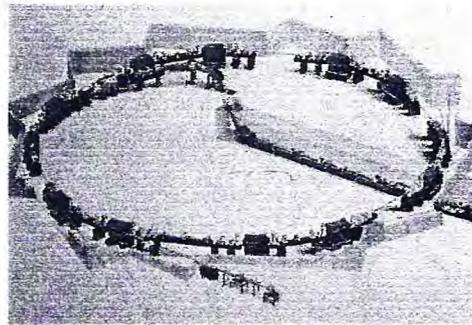


Figura 41 - Maquete de uma fonte de luz síncrotron (LNLS, Campinas). Observa-se o LINAC injetor de 110 MeV (no subsolo), a linha de transporte e o anel de armazenamento de elétrons de 1,37 GeV. As peças maiores do anel são os ímãs defletores (em número de 12). Em parte dos trechos retos do anel se vêem os ímãs quadrupolares e sextupolares. Um dos trechos retos (no ângulo superior esquerdo da figura) é utilizado para a injeção (que se faz desde o interior do anel para maximizar o espaço disponível para as estações experimentais). No trecho reto inferior da figura está instalada a cavidade de radiofrequência. Na parte inferior está também esquematizada uma das estações experimentais planejadas. O anel está rodeado por uma blindagem de concreto.

Os componentes principais das fontes de luz síncrotron são os seguintes:

### 2. 15. 3. 2 Câmaras de vácuo

Desde o acelerador injetor até o próprio anel de armazenamento os elétrons devem se deslocar em condições de alto vácuo (no anel a pressão é da ordem de  $10^{-9}$  -  $10^{-10}$  Torr). Isto reduz significativamente o espalhamento dos elétrons pelas moléculas de ar aumentando assim a vida média do feixe. As câmaras de vácuo do anel são simples tubos nos trechos retos e têm formas especiais na parte curva da trajetória, quando os elétrons sofrem a ação dos ímãs defletores (ou dipolares). É nesta região curva da trajetória que os elétrons, acelerados centripetamente, produzem a luz ou radiação síncrotron. Na parte da câmara interior aos ímãs defletores estão situadas as saídas de luz (tubos tangentes à trajetória no trecho circular da órbita dos elétrons) que conduzem os feixes de fótons pelas linhas até as estações de trabalho

ou estações experimentais, onde são utilizados pelos usuários nos diversos tipos de aplicações. No anel do LNLS há duas saídas de luz em cada um dos 12 ímãs dipolares.

### **2. 15. 3. 3 Ímãs defletores ou dipolares**

Como foi dito, estes ímãs defletem os elétrons quando eles circulam no seu interior. Nessa parte da trajetória, devido à ação do campo magnético constante perpendicular à órbita, os elétrons descrevem arcos de círculo sofrendo um desvio igual a  $360^\circ/N$ , onde N é o número de ímãs dipolares supostos iguais. No caso do anel do LNLS os ímãs defletores são 12 e, conseqüentemente, o ângulo de deflexão é de  $30^\circ$  por ímã.

### **2. 15. 3. 4 Ímãs quadrupolares e sextupolares**

Estão situados nos trechos retos do anel e servem para focalizar os elétrons e reduzir as aberrações cromáticas, respectivamente.

### **2. 15. 3. 5 Caverna de radiofrequência**

Estão situadas num ou em mais de um dos trechos retos. A suas funções são duas:

1. acelerar os elétrons injetados a baixa energia até a energia final e
2. fornecer aos elétrons, em cada passagem pelo interior da mesma, a energia perdida em cada volta devido à radiação produzida nas partes curvas de suas trajetórias. No anel do LNLS utiliza-se uma caverna de 476 MHz e de 60 kW de potência.

### **2. 15. 3. 6 Linhas de luz**

Denomina-se Linha de Luz a instrumentação que é acoplada ao anel de armazenamento de elétrons. Em cada linha de luz chegam os feixes de fótons (luz síncrotron), gerados pelos elétrons que circulam no anel de armazenamento.

É na linha de luz que os feixes são "preparados" para ter utilidade nas estações experimentais utilizadas pelos cientistas. Ou seja, em cada linha há um componente chamado monocromador que define a característica da luz que o pesquisador utilizará em seu

experimento científico. Cada linha de luz tem características próprias e nela é feito determinado tipo de trabalho.

As Linhas de Luz funcionam em condições de alto vácuo. Nas linhas, os feixes, inicialmente brancos e, como foi dito, com apenas uma pequena divergência vertical, são "preparados" para serem utilizados nas estações experimentais. O feixe incidente é

- i) monocromatizado mediante a difração por grades no caso de luz visível ou ultravioleta ou mediante a difração de Bragg por cristais perfeitos (Si, Ge, berilo) no caso de raios X e
- ii) focalizado mediante reflexão total por espelhos ou utilizando cristais com curvatura apropriada. As linhas de luz têm em geral sistemas de segurança com válvulas rápidas para evitar entradas acidentais de ar no anel, obturador primário de radiação síncrotron e barreira de raios gama para impedir a saída desta radiação gerada pela frenagem dos elétrons conseqüente ao espalhamento pelas moléculas do gás residual durante a injeção.

### 2. 15. 3. 7 Estações experimentais

A Estação Experimental é instalada numa linha de luz. Cada estação é formada por equipamentos e instrumentação que permitem ao cientista realizar pesquisas utilizando a luz síncrotron. Uma estação experimental inclui o sistema porta-amostra (no qual é colocada a amostra do material que se quer analisar), aparelho goniométrico (que permite posicionar a amostra para receber adequadamente a incidência da luz síncrotron) e detectores de feixes de fótons espalhados ou transmitidos (que registram os acontecimentos físicos que ocorrem nos átomos e moléculas do material em estudo). Em certos casos, em particular quando se trabalha com luz ultravioleta ou raios X moles (ou em geral em física e química de superfícies), a amostra deve ser situada numa câmara com alto ou ultra-alto vácuo. Nas linhas de luz UV e de raios X moles do LNLS, os feixes de fótons chegam até a amostra sem atravessar nenhuma janela, ou seja, a estação experimental está diretamente conectada com o anel e em condições de alto vácuo.

### 2. 15. 3. 8 Blindagens

Dois tipos de blindagens são em geral utilizados:

- i) blindagem principal, para proteger a área ocupada pelos usuários (no caso do LNLS é uma parede de blocos de concreto de 60 cm de espessura) e
- ii) blindagem secundária, somente necessária nas linhas de luz que trabalham com raios X para isolar o usuário da estação experimental. No caso das linhas de raios X do LNLS, são cabanas ou hutchs de aço com espessura de 1 mm.

### 2. 15. 3. 9 Inserções e trechos retos

As inserções são dispositivos instalados nas partes retilíneas (trechos retos) livres do anel cuja função é modificar as características do espectro de emissão. As inserções mais comuns são os onduladores e os wigglers. Os onduladores têm como função concentrar a radiação branca em faixas de energia estreitas. Os onduladores aumentam o fluxo monocromático para alguns valores da energia que podem ser variados. Os wigglers têm como função incrementar o fluxo de fótons na faixa de maior energia (raios X duros). Isso se consegue mediante o uso de ímãs dipolares de alto campo (usualmente supercondutores) inseridos nos trechos retos livres. Estes ímãs atuam da mesma forma que o ímã defletor básico, mas, devido ao alto valor do campo magnético (5 a 8 Teslas em lugar de aproximadamente 1,0 a 1,7 Tesla nos ímãs dipolares), eles produzem o enriquecimento da parte "dura" do espectro. Se o wiggler tiver ímãs normais e um elevado número de pólos, obtém-se um feixe com um espectro homogeneamente incrementado (o espectro de um dipolo é multiplicado por um fator igual ao número de pólos do wiggler).

### 2. 15. 3. 10 Fonte de luz síncrotron do LNLS

As diversas fontes de luz síncrotron existentes no mundo são bastante variadas quanto a seus espectros de emissão e outras características óticas. Os parâmetros principais da fonte do LNLS (Laboratório Nacional de Luz Síncrotron – Campinas/SP – Brasil), são (UVX parameter list):

Injeção: a baixa energia mediante acelerador linear de elétrons (LINAC).

Energia de injeção: 110 MeV

Diâmetro médio do anel de armazenamento: 28 m

Número de ímãs defletores: 12

Número de saídas de luz dos ímãs defletores: 24 (12 após deflexão dos elétrons de 12 graus e 4 após deflexão de 15 graus )

Número de trechos retos livres para inserções: 4 de 3m de comprimento cada um (nos dois trechos retos restantes estão instaladas a cavidade de radiofrequência e os componentes acessórios de injeção).

Campo magnético do ímã defletor: 1,67 Tesla

Energia final dos elétrons no anel: 1,37 GeV

Corrente eletrônica nominal: 100 mA

Emitância: horizontal 63,9 nm rad e vertical 6, 4 nm rad (no modo normal de operação)

Energia crítica: 2,16 KeV

Limite superior em energia do espectro "útil" de fótons: 12 KeV (este valor se refere particularmente às experiências de difração e espalhamento de raios X. Para as experiências espectroscópicas, como EXAFS, o espectro útil se estende até aproximadamente 25 KeV). O valor limite de 12 KeV equivale a um comprimento de onda mínimo dos fótons de aproximadamente 1 Å.

Tempo de vida dos elétrons circulando no anel: dependendo das condições de uso, aproximadamente 7 horas

Inserções:

- i) um wiggler multipolar (10 pólos) com ímãs normais de 2,0 Tesla e
- ii) um wave-shifter (1 polo) com ímã supercondutor de 6,0 a 7,5 Tesla. Estas inserções permitem acrescentar o fluxo de fótons particularmente na faixa de energia acima de 6 KeV (faixa dos raios X duros).

### 2. 15. 3. 11 Características da emissão das fontes de luz síncrotron

Os feixes de luz síncrotron produzidos pelos anéis de armazenamento de elétrons têm características singulares, o espectro de emissão é "branco" e se estende do infravermelho até os raios X. O limite superior da energia dos fótons emitidos depende da energia  $E$  dos elétrons circulando no anel e do valor do campo magnético  $B$  do ímã defletor ou do wiggler. São extremamente intensos se comparados com os das fontes clássicas de UV e de raios X e com um tamanho de fonte bastante reduzido (a dimensão vertical é de aproximadamente 0,3 mm no caso da fonte do LNLS).

Os feixes de elétrons estão praticamente concentrados no plano da órbita. A divergência vertical dos feixes emitidos pela fonte do LNLS, na parte central do espectro, é de aproximadamente 0,5 mrad, ou seja que, a 10 m de distância, a dimensão vertical do feixe é de 5 mm.

#### 2. 15. 3. 11.1 Laboratório Nacional de Luz Síncrotron

O Laboratório Nacional de Luz Síncrotron (LNLS), instalado em Campinas, é o único deste gênero existente no Hemisfério Sul. O LNLS, em 1987, começou a realizar um ambicioso projeto: colocar o Brasil num seleto grupo de países capazes de produzir luz síncrotron. Este objetivo foi alcançado e, desde julho de 1997, centenas de pesquisadores, do Brasil e do Exterior, utilizam a fonte brasileira de luz síncrotron para fazer pesquisas.

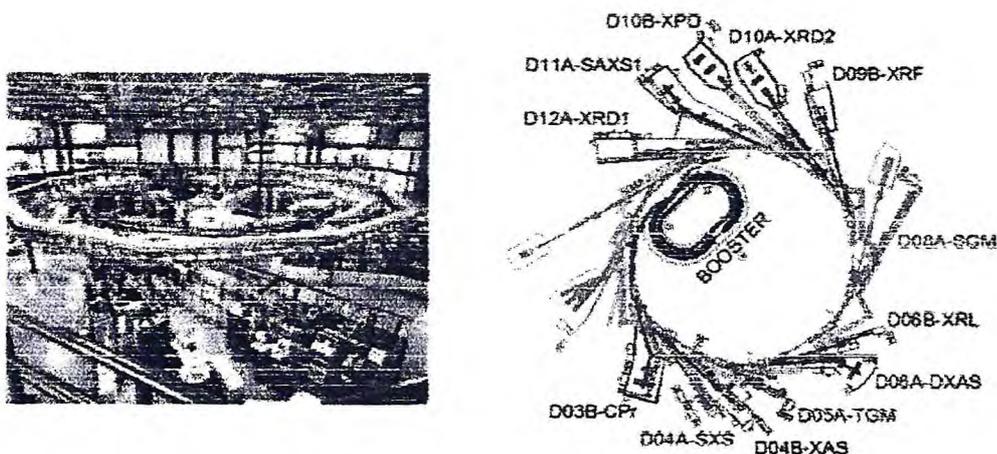


Figura 42 - Laboratório Nacional de Luz Síncrotron - SP. Da esquerda para a direita, vista superior e diagrama das linhas de luz e estações de trabalho.

### 2. 15. 3. 11. 2 Linha CPr – Cristalografia de Proteínas

Esta linha é a utilizada para estudos de alta resolução na coleta de dados de monocristais a um único comprimento de onda para análise de estruturas tridimensionais de macromoléculas biológicas. O comprimento de onda utilizado situa-se na faixa de 1-2 Å e a faixa de energia dos fótons é de 6-12 KeV. Utiliza como fonte um imã dipolo D03B (15 °) e como monocromador um cristal curvo de silício, com corte assimétrico. Os elementos de focalização desta linha constituem um espelho curvo cilíndrico para focalização vertical e cristal-monocromador curvo para focalização no plano horizontal. Seu detector é um CCD (MarResearch).

## 2. 16 Coleta dos dados

### 2. 16. 1 Caracterização do cristal

Cristais de tamanho suficiente (acima de 0,1 x 0,1 x 0,1 mm) devem ser caracterizados pela sua qualidade de difração, dimensões de célula unitária, grupo espacial (uma descrição da simetria cristalográfica) e o número de moléculas protéicas por célula unitária. Todos estes parâmetros podem ser determinados direta ou indiretamente das imagens de difração do cristal.

O número de moléculas protéicas por célula unitária pode ser estimado pelo método proposto por MATTHEWS (1968), pelo cálculo da relação entre o volume da célula unitária e a massa molecular da proteína. Matthews verificou que esta relação (chamada  $V_M$ ) está entre 1,7 e 3,5 Å<sup>3</sup>/Da para a maioria das proteínas.

O conhecimento do grupo espacial e do número de moléculas na célula unitária permite a determinação do número de moléculas na unidade assimétrica. Esta é a estrutura mínima que o cristalógrafo deve elucidar para poder determinar a estrutura tridimensional da proteína e estas são as coordenadas que são normalmente depositadas no Protein Data Bank (PDB, BERMAN *et al*, 2000). O número de moléculas na unidade assimétrica não representa, necessariamente, o número de moléculas que constituem a proteína na solução. Por exemplo,

uma proteína dimérica, composta de duas subunidades idênticas relacionadas por um eixo de simetria 2 (uma rotação de  $180^\circ$ ) pode cristalizar como somente uma subunidade na unidade assimétrica se a simetria cristalográfica inclui um eixo 2 sobre o qual o dímero pode estar situado. Em geral, não há uma relação simples entre o conteúdo da unidade assimétrica (e, assim sendo, as coordenadas depositadas do Protein Data Bank) e o estado oligomérico da proteína em solução.

## **2. 17 Tratamento dos dados**

A rede cristalina tridimensional dá origem a uma rede tridimensional de pontos de difração chamada de rede recíproca. Pelo padrão de difração ser um objeto tridimensional, fica uma vez mais clara a necessidade de girar o cristal para coletar todos os dados de difração, uma vez que a rotação do cristal gira a rede recíproca que então descreve sobre o detector os pontos de difração.

A rede recíproca deriva seu nome do fato de que tem parâmetros que são relacionados ao inverso daqueles da rede cristalina (CREIGHTON, 1993). Assim, após a coleta dos dados de difração, as reflexões são indexadas (seus índices de Miller determinados) e suas intensidades colocadas na mesma escala. Vários programas de computador fazem esta tarefa, entre eles HKL e Mosflm.

Mosflm é distribuído com o pacote de programas CCP4 (Collaborative Computational Project, Number 4) e escreve seus dados de reflexão diretamente em um arquivo MTZ, que é o formato binário de dados usado pelos programas do CCP4 (Figura 43). Os dados nestes arquivos são organizados para que possam ser eficientemente utilizados pelos programas.

```

* Number of Datasets = 1

* Dataset ID, project/crystal name, dataset name, cell dimensions, wavelength:

  1 lys_fine /
  02_05_02:13:53:50
  78.1485 78.1485 37.0149 90.0000 90.0000 90.0000
  1.54180

* Number of Columns = 16

* Number of Reflections = 13368

* Missing value set to NaN in input mtz file

* Number of Batches = 50

* HISTORY for current MTZ file :

  From MOSFLM run on 2/ 5/02

<snip> * Resolution Range :

  0.00089 0.17159 ( 33.520 - 2.414 A )

* There is no sort order recorded in the MTZ header

* Space group = P43212 (number 96)

```

Figura 43 - Cabeçalho típico de um arquivo MTZ do programa Mosflm.

O pacote de programas CCP4 é uma criação em conjunto da comunidade científica do Reino Unido. O projeto é coordenado pelo CCLRC Daresbury Laboratory e atualmente financiado pelo Biotechnology and Biological Sciences Research Council (BBSRC). Atualmente este pacote é distribuído gratuitamente para fins acadêmicos e comerciais.

Um exemplo do cabeçalho de informações contido num arquivo MTZ típico de Mosflm é apresentado na figura 43. Este arquivo foi processado de uma única série de dados (seguido de escalonamento). Além do nome do arquivo e data, as dimensões refinadas da célula unitária, o comprimento de onda utilizado no experimento, o número total de reflexões integradas, o número de imagens processadas, a faixa de resolução das reflexões no arquivo e o grupo espacial utilizado também são gravados. Há 16 colunas referentes às reflexões num arquivo MTZ:

H K L – índices da reflexão

M/ISYM – marcador para código de operador de simetria total/parcial

BATCH – grupo ao qual a imagem pertence

I SIGI – Intensidade e sigma  
 IPR SIGIPR – perfil da intensidade fixado e sigma  
 FRACTIONCALC – fração calculada da reflexão nesta imagem  
 XDET YDET – coordenadas do detector (pixels)  
 ROT WIDTH centróide phi e distância angular da reflexão  
 LP – fator de Lorentz e de polarização  
 MPART – marcador de parcialidade

Cada imagem em uma série provavelmente foi gravada com suas intensidades em diferentes escalas por várias razões, por exemplo, variação na intensidade da radiação incidente, absorção de raios difratados, mudanças no volume difratante etc.

As escalas devem ser aplicadas corretamente se o processo de indexação é de reflexões simétricas equivalentes de diferentes imagens. Esta parte do processo é feita pelo programa SCALA.

O resultado é um arquivo identificando cada reflexão por seus índices de Miller,  $hkl$ , juntamente com as intensidades das reflexões e seus erros (um exemplo deste arquivo é mostrado na Figura 44).

Uma avaliação da qualidade dos dados processados pode ser feita usando vários critérios objetivos. Um destes parâmetros é chamado  $R_{merge}$ , que mede a discordância entre reflexões equivalentes.

h	k	l	I	$\sigma(I)$
0	1	1	1136.84	30.00
0	1	2	676.69	13.96
0	1	3	578.05	17.66
0	1	4	531.10	22.67
0	1	5	347.97	21.41
.	.	.	.	.
.	.	.	.	.

Figura 44 – Os dados de difração após a redução. Cada reflexão é caracterizada por seus índices de Miller (hkl) seguidos por sua intensidade e desvio padrão.

Existem reflexões que são medidas mais que uma vez durante a coleta de dados ou são reflexões relacionadas pela simetria do padrão de difração que é determinado pela simetria cristalográfica. Em ambos casos as intensidades das reflexões devem ser idênticas e os desvios de um valor médio (valor "merged") podem ser utilizados para estimar a qualidade dos dados (DRENTH, 1994).

Um bom valor para  $R_{merge}$  depende da redundância dos dados, isto é, quantas reflexões equivalentes tiverem sido medidas e com que frequência. Uma boa regra é que os dados devem apresentar um  $R_{merge}$  total de menos de 10 %, mas que este valor pode ser significativamente maior para dados de maior resolução.

Um outro parâmetro usado na avaliação dos dados processados, denominado completeza, é a relação entre o número observado de reflexões (isto é, aquelas realmente medidas) e aquelas teoricamente passíveis de medição.

Claramente a redundância e completeza devem ser tão grandes quanto possível e o  $R_{merge}$  tão pequeno quanto possível.

## **2. 18 Determinação das estruturas protéicas**

A determinação da estrutura tridimensional de uma proteína por difração de raios X implica em encontrar os ângulos de fase para todas as reflexões geradas pelos raios X incidentes sobre o cristal. Na cristalografia de proteínas há basicamente três métodos para determinação das estruturas (isto é, resolver o problema das fases) que são utilizados de acordo com o tipo de problema a ser resolvido: substituição molecular (MR), substituição por múltiplas isoformas (MIR) e dispersão anômala de múltiplos comprimentos de onda (MAD) (BLUNDELL AND JOHNSON, 1976; GIACOVAZZO *et al.*, 1992; DRENTH, 1994).

Substituição molecular, desenvolvida por ROSSMANN E BLOW (1962), é o método mais rápido para determinação de estruturas protéicas, mas requer a disponibilidade de uma estrutura-modelo previamente conhecida, isto é, uma proteína similar. O método é baseado na observação de que proteínas similares compartilham idênticos dobramentos e, dependendo

do grau de identidade na seqüência, estruturas terciárias similares.

Neste método, a molécula de estrutura conhecida, denominada modelo de busca, é considerada como um modelo inicial a partir do qual se pode calcular a primeira estimativa dos ângulos de fase (relembrando que uma vez as coordenadas atômicas para uma estrutura sejam conhecidas, é possível calcular tanto as amplitudes como os ângulos de fase para um dado cristal daquela estrutura). Para fazer isto, no entanto, é um pré-requisito que o modelo de busca seja corretamente posicionado dentro da célula unitária da estrutura que se deseja determinar. Somente então, os átomos do modelo de busca estarão suficientemente próximos as reais posições de interesse e darão as fases com precisão suficiente.

O modelo de busca é posicionado na célula unitária da estrutura desconhecida através da aplicação de seis variáveis, três rotacionais e três translacionais. Esta transformação pode ser representada por (GIACOVAZZO *et al.*, 1992):

$$X' = [M]X + t$$

onde  $X$  é o grupo de posições vetoriais dos átomos da molécula original,  $[M]$  é a matriz que gira as coordenadas  $X$  para uma nova posição e  $t$  é uma translação vetorial. Resolver uma estrutura por substituição molecular reduz o problema a sobrepor duas estruturas protéicas, isto é, a achar  $[M]$  e  $t$ . No entanto, somente uma é conhecida. A outra é exatamente aquela que desejamos determinar. A solução é usar duas funções matemáticas, denominadas funções de rotação e translação, que não dependem do conhecimento da estrutura desconhecida.

A estrutura será substituída por um mapa de vetores interatômicos ao invés de posições atômicas. A vantagem de tal mapa (chamado mapa de Patterson) é que pode ser calculado diretamente das intensidades de difração sem qualquer conhecimento dos ângulos de fase. Mesmo sendo este um mapa vetorial, pode ser usado para orientar inicialmente o modelo de busca, através da função de rotação e então sobrepô-lo na estrutura desconhecida, usando a função translação. Tais funções podem ser facilmente computadas atualmente com o auxílio de programas de computador tais como AmoRe (NAVAZA, 1994) e MolRep (VAGIN & TEPLYAKOV, 2000) que também são programas que integram o pacote CCP4.

Utilizando o método de substituição molecular, um modelo de busca é posicionado dentro da célula da molécula cuja estrutura é desconhecida. A aplicação das matrizes dos vetores de rotação e translação ao modelo de busca leva a um modelo inicial passível de refinamento para a estrutura sob investigação. Após a substituição do modelo de busca pela estrutura real, fazendo as necessárias substituições de cadeias laterais, inclusão e retirada de resíduos de aminoácidos, as posições atômicas podem ser refinadas de acordo com os métodos descritos mais tarde.

No método de substituição isomórfica múltipla (MIR) as fases das reflexões são determinadas a partir do conhecimento da posição (conseqüentemente, das fases) de átomos pesados inseridos dentro do cristal protéico. Neste método, que tem sido extensivamente utilizado na determinação de proteínas para as quais não existem estruturas similares, é necessário coletar dados de difração do cristal da proteína nativa e também de, no mínimo, um derivado com átomos pesados. Uma vez que os elétrons é que são responsáveis pela difração de raios X, a introdução de um metal pesado dentro do cristal resulta em pequenas diferenças nas intensidades para as reflexões quando comparadas com aquelas do cristal nativo (sem o metal pesado). A partir destas pequenas diferenças a posição do átomo pesado pode ser obtida e este é o ponto de partida para a determinação dos ângulos de fase para as reflexões do cristal nativo (DRENTH, 1994). Infelizmente este procedimento leva a duas diferentes e igualmente prováveis soluções para o problema das fases de cada reflexão e, sendo assim, o problema ainda não foi totalmente resolvido. Para distinguir entre estas duas possíveis soluções, um segundo derivado com átomo pesado deve ser usado, o que também dará origem a dois ângulos de fase possíveis. Somente um destes deve ser igual a um daqueles dois previamente determinados do primeiro derivado e corresponderá à solução correta.

O sucesso do método depende de vários fatores incluindo o tamanho dos erros experimentais, o grau de similaridade, a ocupância dos átomos pesados e o número de sítios. Os erros experimentais, juntamente com o fato de que o sinal de átomos pesados é relativamente pequeno quando comparado com o resto da proteína, faz com que

normalmente mais que dois derivativos sejam necessários para se obter estimativas suficientemente boas das fases para a produção de um mapa de densidade eletrônica interpretável.

Isomorfismo se refere à extensão das estruturas da célula unitária dos cristais nativo e derivado que são idênticas. Idealmente a única diferença deve ser a presença dos átomos pesados ligados a um pequeno número de sítios altamente específicos sobre a superfície da proteína dos cristais derivativos. Se eles são vários sítios ou se são fracamente ocupados (isto é, nem todas as células unitárias têm átomos pesados ligados) então o processo de determinação das posições destes átomos pesados será mais difícil. A noção de isomorfismo pode ser esclarecida procurando-se mudanças nos parâmetros de célula unitária do cristal derivativo. No entanto, mesmo se estes não mudarem não há absoluta garantia que as posições atômicas não terão sido alteradas no cristal derivativo.

Na prática, o perfeito isomorfismo é difícil de ser alcançado devido a distorções da estrutura quando o átomo pesado é introduzido, normalmente por imersão de um cristal nativo pelo período de várias horas numa solução diluída de um composto apropriado do átomo pesado. Pequenas imperfeições no isomorfismo não impedem a determinação da estrutura porque os cristais podem ser considerados efetivamente isomorfos numa resolução que seja suficiente para resolver o problema das fases. No entanto, significantes alterações nos parâmetros de célula, por exemplo, eliminam o uso do método.

A escolha dos compostos metálicos a serem testados é de fundamental importância nos experimentos. Embora a busca por derivativos de metais pesados seja basicamente empírica e muitos reagentes devam ser testados antes de se conseguir resultados, deve-se empregar todo o conhecimento químico e bioquímico da proteína e suas condições de cristalização. A presença de alguns grupos químicos na proteína pode favorecer o uso de certos metais, por exemplo, compostos de mercúrio freqüentemente se ligam a átomos de enxofre de cisteínas livres. Sítios de ligações de metais de ocorrência natural na proteína de interesse ( $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ) podem algumas vezes ser substituídos por metais mais pesados tais como terras raras, obtendo-

se assim derivativos de maior ocupância.

Substituição isomórfica por átomos pesados pode ser feita tanto por imersão do cristal protéico numa solução do reagente como por co-cristalização. O método mais comumente utilizado é a imersão porque a introdução de novos reagentes para co-cristalização quase sempre interfere no processo de cristalização e pode mesmo impedir o crescimento do cristal. A imersão, no entanto, tem a desvantagem de preencher as cavidades de solvente do cristal com composto do metal pesado que pode aumentar o ruído do padrão de difração e, assim, reduzir a razão sinal/ruído. Isto pode ser resolvido por mais uma imersão na solução tampão para remover compostos do átomo pesado retidos nos canais de solvente ou somente não especificamente ou fracamente ligados à superfície da proteína. Uma útil base de dados pode ser acessada no endereço <http://bonsai.lif.icnet.uk/had/> a qual contém informação sobre um grande número de compostos de átomos pesados que têm sido utilizados com sucesso no processo.

O método de dispersão anômala de múltiplos comprimentos de onda (MAD) foi primeiramente utilizado com sucesso por Hendrickson e colaboradores na determinação de estruturas protéicas (HENDRICKSON *et al.*, 1988; KRISHNAMURTH *et al.*, 1988; HENDRICKSON, 1991). O método é provavelmente a mais elegante das técnicas de determinação de estruturas disponíveis para proteínas e explora a propriedade de espalhamento anômalo pelos átomos próximos e seus limites de absorção para resolver o problema das fases.

Elétrons internos do átomo são fortemente ligados ao núcleo, um fato que é especialmente real para os átomos pesados, que têm uma alta carga nuclear. Sob tais circunstâncias o feixe espalhado difere em fase, com respeito ao feixe incidente, não por exatamente  $180^\circ$ , como no caso dos elétrons de valência que podem ser considerados livres, mas a diferença de fases entre o feixe incidente e o feixe espalhado por elétrons internos é diferente de  $180^\circ$ , uma propriedade chamada de espalhamento anômalo. Uma das conseqüências deste fenômeno é que as intensidades das reflexões  $(hkl)$  e  $(ih, ik, il)$  já não são as mesmas como é normalmente o caso. Efeito anômalo somente toma importância se o

núcleo de elétrons contribui significativamente para o espalhamento e isto só ocorre a comprimentos de onda de raios X próximos dos limites de absorção dos elétrons internos. Uma característica do fenômeno é sua dependência ao comprimento de onda. Isto normalmente só tem uma contribuição efetiva no espalhamento se há um átomo relativamente pesado presente na estrutura e se os comprimentos de onda escolhidos são os de seus limites de absorção.

A aplicação da técnica MAD é muito similar a de MIR, só que, ao invés de usar diferentes cristais derivativos para medir várias séries de dados, todos os dados podem ser coletados de um mesmo cristal (ou tipo de cristal), mas usando vários comprimentos de raios X (HENDRICKSON, 1985; KARLE, 1980). Isto elimina os problemas de não-isomorfismo. Como consequência, fases e mapas de densidade eletrônica podem ser calculados a maior resolução deste o início do processo de determinação da estrutura, facilitando grandemente sua interpretação.

Devido às propriedades de espalhamento alteradas do átomo de espalhamento anômalo (átomo pesado) como uma função do comprimento de onda, as intensidades das reflexões mudarão marcadamente de uma série de dados (comprimento de onda) para outra.

A posição do espalhador anômalo pode ser determinada através destes dados combinados em uma maneira análoga à usada para MIR. Esta informação é subsequente usada para obtenção das fases para todas as reflexões.

Uma das limitações deste método é a necessidade obrigatória de uma fonte de raios X tal como um síncrotron para se ter suficiente intensidade disponível nos comprimentos de onda apropriados para o espalhador anômalo presente na amostra.

Uma outra limitação da técnica é a necessidade de um átomo de espalhamento anômalo na proteína. Tradicionalmente íons metálicos têm sido usados para este fim, como aqueles naturalmente presentes em metalo-proteínas ou podem ser introduzidos por imersão. Mais recentemente, selênio tem sido usado na forma de selenometionina (YANG *et al.*, 1990), que pode ser incorporado em proteínas recombinantes no lugar da metionina através do

crescimento da bactéria em meio apropriado.

## 2. 18. 1 Construção do mapa de densidade eletrônica

Após a determinação da fase por um dos métodos mencionados anteriormente, e um ciclo inicial de refinamento, um mapa de densidade eletrônica da célula unitária (ou pouco mais) pode ser calculado pela transformada de Fourier a partir dos fatores de estrutura ( $F(h)$ ). Este mapa é normalmente visualizado após contorno dos pontos de densidade eletrônica iguais nas regiões de maior densidade de elétrons. Estes correspondem às posições atômicas, mas normalmente não aparecem como ilhas de densidades eletrônicas isoladas devido a limitações na resolução, ao invés disso, aparecem como trechos contínuos de densidade eletrônica, correspondentes ao esqueleto protéico e são uma boa indicação de um mapa interpretável. A interpretação inicial do mapa de densidade eletrônica pode não ser trivial, a menos que boas fases tenham sido obtidas em alta resolução.

Vários fatores podem dificultar a análise dos mapas. Interrupções na densidade eletrônica (“mapas quebrados”) são comuns como uma perda de densidade eletrônica nos terminais N ou C causada pela normal flexibilidade da estrutura nestas regiões. Dificuldades na interpretação do mapa podem ser causadas pela fraca resolução, erros na determinação das fases ou dados perdidos. Se os dados estão incompletos em uma faixa angular porque o cristal “morreu” durante a coleta de dados, por exemplo, isto pode levar a artefatos no mapa, tais como picos artificiais, que confundem o processo de interpretação.

Se, por outro lado, os dados de baixa resolução estão perdidos porque estas reflexões, que são fortes na intensidade, saturaram o detector durante a coleta de dados, então pode haver dificuldades na distinção entre regiões da proteína e os canais de solventes.

Muito freqüentemente, no entanto, os limites das moléculas protéicas são facilmente discerníveis no mapa mesmo se a interpretação da densidade dentro da molécula está difícil. Em tais casos, técnicas de melhoramento (solvent flattening (TERWILLIGER, 1990)) podem ser usadas para intensificar um mapa não interpretável.

Outras técnicas de modificação de densidade são disponíveis e permitem que a aplicação de propriedades previsíveis do mapa possam ser utilizadas para melhorar sua qualidade. Tais técnicas incluem combinação de histograma ("histogram matching" (TERWILLIGER, 2000)) e o uso de simetria não cristalográfica (TONG, 2001).

Simetria não-cristalográfica é empregada a casos onde há mais que uma molécula idêntica na unidade assimétrica. Há uma boa razão para se acreditar que todas as cópias se apresentarão com estruturas idênticas (embora elas possam apresentar pequenas diferenças conformacionais) e, assim sendo, a densidade eletrônica de todas as moléculas pode ser a média entre elas se as transformações que as relacionam tiverem sido previamente determinadas. Em geral, este procedimento leva a uma melhoria na razão sinal/ruído do mapa, facilitando grandemente a interpretação.

O prévio conhecimento da estrutura primária da molécula é efetivamente essencial para uma análise de sucesso do mapa de densidade eletrônica. O primeiro passo para a interpretação é normalmente o traçado da cadeia principal. Isto é muito raramente feito seguindo-se o esqueleto peptídico seqüencialmente do N- ao C-terminal e em muitas ocasiões várias seções da molécula são mais ou menos simultaneamente analisadas. Uma das razões para isto é que quase sempre o resíduo terminal da cadeia não é visível no mapa devido à desordem dinâmica ou estática. Mais ainda, vários terminais falsos podem estar representados dentro da estrutura devido a descontinuidades da densidade como um resultado de desordem estática ou flexibilidade local.

Se a estrutura tiver sido resolvida por substituição molecular há menor ou nenhuma ambigüidade no traçado da cadeia. Se, por outro lado, é uma estrutura MIR ou MAD, então pontos de partida convenientes devem ser identificados na cadeia polipeptídica para iniciar o traçado. Estes pontos podem ser pontes di-sulfeto que são caracterizadas por grandes picos de densidade eletrônica ou íons metálicos ligantes que podem ser claramente visíveis. Alternativamente grandes, mas não raros aminoácidos, tais como triptofanos, podem ser freqüentemente reconhecidos por suas características de densidade de cadeia lateral.

A interpretação de mapas de densidade eletrônica faz uso intenso de telas gráficas onde o mapa pode ser manual e interativamente interpretado pelo cristalógrafo. Programas gráficos tais como FRODO (JONES, 1985), "O" (JONES AND KJELDGAARD, 1993) e XtalView (McREE, 1999) entre outros, são amplamente utilizados com este propósito. O processo envolve a construção da seqüência de aminoácidos na densidade enquanto preserva características estereoquímicas esperadas e bem conhecidas da cadeia principal e das cadeias laterais. O uso freqüente de base de dados de rotâmeros de cadeias laterais (PONDER AND RICHARDS, 1987) e estruturas de "loop" de cadeia principal (JONES AND THIRUP, 1986) é feito para auxiliar este processo.

Embora em grande extensão a interpretação de mapas de densidade eletrônica ainda seja um processo manual, algum grau de automação tem sido introduzido durante os últimos anos. De particular nota são os métodos que traçam automaticamente a direção do esqueleto pela busca de trechos contíguos de densidade (GREER, 1974). Tais métodos podem também ser usados para indicar posições potenciais de cadeias laterais. A primeira aproximação de uma estrutura, ou esqueleto, pode então ser utilizada juntamente com bases de dados para produzir um modelo completo. O grau de sucesso claramente depende da qualidade do mapa original.

O grau de detalhamento a ser observado na estrutura depende da qualidade dos dados de raios X e, particularmente, de sua resolução. Isto é claramente ilustrado na Figura 45.

O quadro A representa o mapa de densidade eletrônica de dados coletados a 2,0 Å de resolução. O modelo estrutural resultante tem um alto grau de definição. A resolução da imagem é restringida a 2,5; 3,0 e 4,0 Å nos quadros B, C e D, respectivamente. Fica visível que o grau de detalhamento obtido nas imagens diminui com a resolução.



Figura 45 - Ilustração da influência da resolução dos dados no detalhamento observável na estrutura. Quadros A, B, C e D mostram o mapa de densidade eletrônica de uma região interna da gliceraldeído 3-fosfato desidrogenase de *Trypanosoma cruzi* calculado a 2,0; 2,5; 3,0 e 4,0 Å respectivamente (SOUZA; SELISTRE-DE-ARAÚJO; GARRATT, 2000).

Um outro fator que afeta a informação presente no mapa é o grau de ordem de regiões particulares da molécula. Regiões flexíveis da proteína ou mesmo a existência de duas ou mais conformações alternativas podem resultar no truncamento da densidade eletrônica ou bifurcação. Se os dados estão disponíveis à resolução suficientemente alta, pode ser possível interpretar com sucesso regiões desordenadas em termos de duas ou mais conformações alternativas de diferentes ocupâncias.

Algumas vezes, confusão pode surgir na distinção entre uma cadeia lateral desordenada sobre a superfície da proteína e a presença de moléculas de água estrutural. De outra maneira, é certamente esperado encontrar-se um significativo número de moléculas de água ordenadas interagindo com a superfície da proteína. Estas são caracterizadas por pequenas e geralmente esféricas regiões de densidade isolada, freqüentemente dentro de uma distância que permita ligações de hidrogênio de grupos polares da superfície da proteína. Elas são adicionadas à estrutura geralmente nos últimos estágios de refinamento.

## 2. 18. 2 Refinamento do modelo

O modelo estrutural inicial conterá erros que podem ser minimizados através de refinamento. Este é um processo de ajuste das coordenadas atômicas do modelo para minimizar as diferenças entre as amplitudes de fatores de estrutura observados ( $F_{obs}$ ) e aquelas calculadas de um cristal hipotético contendo o modelo ( $F_{calc}$ ). Em outras palavras é uma tentativa de fazer coincidir o padrão de difração calculado do modelo com aquele medido experimentalmente.

Para cristais protéicos, a razão entre o número de observações experimentais (intensidades das reflexões) e o número de parâmetros a ser refinados (coordenadas atômicas (x, y e z), fatores de temperatura e ocupâncias) raramente é alta o suficiente para permitir o refinamento livre do modelo. No entanto, restrições determinadas empiricamente que correspondem a parâmetros estereoquímicos conhecidos (tais como comprimentos de

ligações, ângulos etc) e para interações atômicas dentro da molécula (como interações não ligantes) podem ser introduzidas de tal maneira a efetivamente aumentar esta razão.

Há duas maneiras de se conduzir o refinamento do modelo impondo tais restrições: diminuindo o número de variáveis a serem refinadas (um processo chamado refinamento de corpo rígido) ou artificialmente aumentando o número de observações experimentais, refinamento por restrições (restrained refinement) (GIACOVAZZO *et al.*, 1992).

No refinamento de corpo rígido, um grupo inteiro de átomos é tratado como um corpo rígido. Todos os átomos do corpo rígido se movem de maneira concertada durante o refinamento fazendo com que estes átomos sempre ocupem as mesmas posições relativas. Podemos tomar como exemplo o anel fenólico do aminoácido fenilalanina. Ignorando os átomos de hidrogênio, dezoito posições variáveis devem ser refinadas (um x, y e z para cada um dos átomos de carbono do anel). Se o anel é considerado como uma única entidade, somente seis variáveis serão refinadas (três rotacionais e três translacionais).

Esta estratégia é possível porque os comprimentos de ligações e ângulos dos aminoácidos foram bem determinados de pequenas moléculas previamente conhecidas (ENGH AND HUBER, 1991). No entanto, com o atual aumento de estruturas protéicas resolvidas a altas resoluções (DAUTER *et al.*, 1995), espera-se que os parâmetros obtidos das próprias proteínas em si sejam utilizados e continuamente atualizados no futuro.

O conceito de corpo rígido pode ser aplicado, ao menos no início do refinamento, não somente a pequenos grupos de átomos, mas a molécula inteira. Este é normalmente o caso para estruturas resolvidas por substituição molecular, onde as funções de rotação e translação deixam o modelo de busca posicionado somente aproximadamente correto dentro da célula unitária.

Uma alternativa para se reduzir o número de parâmetros (de aproximadamente dez vezes) é por dinâmica de ângulos de torção (RICE AND BRUNGER, 1994).

No refinamento por restrições, os comprimentos e ângulos de ligação podem variar em algum grau em torno de seus valores padrões. Esta informação pode ser introduzida no

refinamento do modelo incluindo-se termos estereoquímicos na função alvo de tal forma que a coincidência entre as amplitudes estruturais, calculada e observada, não seja mais o único fator a ser levado em conta. Isto pode ser introduzido dentro de procedimentos de mínimos quadrados.

Os dois métodos podem ser combinados, por exemplo, considerando-se a molécula composta por grupos rígidos e aplicando restrições para distâncias aceitáveis entre elas (FLOHIL, VRIEND, BERENDSEN, 2002; ERIKSSON & LILJAS, 1993; BINSTED, STRANGE, HASNAIN, 1992).

Uma outra maneira de diminuir o número de parâmetros a serem refinados é a imposição de simetria não cristalográfica (NCS) onde aplicável (ROSSMANN, 1995). A presença de NCS significa que há redundância dentro da unidade assimétrica, por exemplo, duas ou mais cópias da mesma subunidade protéica.

Na aplicação de NCS a protocolos de refinamento, presume-se que todas as moléculas pertencentes à unidade assimétrica são estruturalmente idênticas. Isto significa que o número de parâmetros que se necessita refinar diminui dramaticamente. No caso de duas moléculas por unidade assimétrica o refinamento é feito em uma só molécula e nos parâmetros de simetria que descrevem a relação entre as duas cópias independentes. Quase ao final do refinamento a NCS é liberada para se examinar possíveis pequenas diferenças conformacionais entre subunidades idênticas quimicamente. No entanto, a validade de tal estratégia depende da razão observações/parâmetros e, mais ainda, da resolução e completeza dos dados.

Dificuldades podem surgir no refinamento da estrutura protéica devido a grande número de parâmetros atômicos a serem refinados. Isto dá origem a uma função multidimensional (função alvo) que apresenta um grande número de mínimos locais além do mínimo global que é o objetivo.

Há duas considerações gerais a serem levadas em conta nos protocolos de refinamento. O primeiro é a função alvo a ser adotada, isto é, o que é que se tenta minimizar. Esta função normalmente conterá um termo experimental (o qual descreve quão bem o modelo reproduz

os dados) como também um outro termo que consiste de componentes que descrevem a estereoquímica. De qualquer maneira, ainda há uma escolha a ser feita com relação à exata forma do termo experimental. Tradicionalmente este tem sido um termo de mínimos quadrados (SUSSMAN & PODJARNY, 1983) que é a soma ao quadrado de todas as diferenças entre  $F_{obs}$  e  $F_{calc}$ .

Mais recentemente o conceito de maximum-likelihood tem sido introduzido na função alvo e se mostrado ser superior (PANNU AND READ, 1996). No último caso os efeitos de erro de modelo e incompleteza, além de erros nos dados, são levados em conta.

Uma vez que a função alvo tenha sido definida o segundo problema é como achar seu mínimo global. Tradicionalmente quadrados mínimos e métodos de gradiente conjugados têm sido empregados. Estes têm a limitação de necessariamente seguir o gradiente local e assim caem em mínimo local. Uma das desvantagens de tais métodos é seu limitado raio de convergência. Esta é uma medida da capacidade da técnica corrigir erros na estrutura não refinada, mas depende da resolução e por isto é quase sempre conveniente começar o refinamento utilizando somente os dados a uma limitada resolução máxima.

Se as distâncias das posições atômicas iniciais a seu valor ideal é muito grande, procedimentos tais como quadrados mínimos e minimização por gradiente conjugado podem nunca corrigir o erro. Este problema pode ser minimizado com o uso de estratégias de refinamento alternativas.

A introdução de dinâmica molecular por gradiente de temperatura (simulated annealing) no procedimento de refinamento para macromoléculas, proposto no fim dos anos 80 e ainda é provavelmente a técnica de escolha para o refinamento de estruturas protéicas (BRUNGER *et al.*, 1987; 1997a,b; FUJINAGA *et al.*, 1989). Dinâmica molecular por gradiente de temperatura se refere ao processo que simula a dinâmica da proteína a alta temperatura (geralmente entre 3000 e 4000 K) seguida pelo gradual resfriamento. Em altas temperaturas os átomos adquirem suficiente energia cinética para se mover sobre barreiras energéticas na função alvo, o que permite grandes mudanças conformacionais na estrutura. Isto dá a

possibilidade da estrutura se mover de um mínimo energético local a um mais profundo que está separado por uma barreira energética, um fenômeno que não é possível com minimização de gradiente conjugado ou mínimos quadrados.

A técnica efetivamente aumenta o raio de convergência do refinamento, permitindo a correção automática de erros relativamente grandes no modelo inicial sem a necessidade de intervenção manual, acelerando, assim, o processo como todo.

Não se deve esquecer que a função “energia” usada contém não somente termos de energia real, tais como os atribuídos a distâncias de ligações, ângulos, contatos não ligantes etc, mas também um termo cristalográfico que auxilia a minimizar as diferenças entre  $F_{obs}$  e  $F_{calc}$ . Este termo serve para manter a estrutura dinâmica dentro da densidade eletrônica durante a dinâmica molecular. Se, por exemplo, uma região de “loop” se move dramaticamente de uma conformação incorreta para uma correta durante o refinamento, então haverá uma melhor coincidência entre as amplitudes de estruturas calculada e experimental ( $F_{calc}$  e  $F_{obs}$ ). O termo cristalográfico da função energia tenderá a manter a estrutura nesta posição durante o restante do refinamento, efetivamente reduzindo a chance do “loop” se mover para fora da densidade eletrônica novamente.

Os pesos relativos que podem ser atribuídos para o termo de energia real e o cristalográfico (pseudo-energia) podem ser controlados pelo usuário e alterados quando o refinamento converge para a estrutura final. Fazendo-se um paralelo é como se a temperatura total (energia cinética) do sistema fosse mantida acoplando-se o sistema a um banho-maria. Isto seria usado para remover excesso de energia cinética do sistema quando a energia potencial cai, prevenindo “saltos” da estrutura para fora do poço energético no qual a estrutura caiu.

Dinâmica molecular por gradiente de temperatura é mais comumente implementada usando-se algoritmos nos quais o comportamento dinâmico de um sistema de partículas (a estrutura molecular) é simulada dando-se a cada átomo da molécula velocidade e direção específicas, inicialmente aleatoriamente determinadas para uma dada temperatura

(BRUNGER, 1988).

Em geral, o processo de refinamento não é totalmente automatizado, mas deve ser interrompido periodicamente para o cálculo de novos mapas de densidade eletrônica baseados nas melhores fases resultantes das novas posições atômicas. Tais mapas precisam ser inspecionados usando gráficos moleculares e alterações manuais na estrutura.

Freqüentemente diferentes tipos de mapas são calculados a cada ciclo de refinamento. Estes incluem mapas de diferença ( $F_{obs} - F_{calc}$ ) nos quais os picos positivos aparecem em posições onde átomos estão perdidos no modelo e picos negativos onde átomos não-existent tenham sido erroneamente incluídos. Um par de picos (um positivo, um negativo) significa que o átomo ou grupo de átomos em particular foi levemente mal posicionado.

Mapas nos quais pequenas regiões da estrutura tiverem sido deliberadamente omitidas do modelo (mapas de omissão) são também freqüentemente usados para determinar se os mapas estão sujeitos a erros no cálculo das fases. Se a densidade eletrônica aparece claramente em um mapa no qual a região correspondente da estrutura tiver sido omitida, é uma boa indicação que a estrutura para aquela região está correta.

Durante o refinamento a qualidade do modelo resultante pode ser avaliada pelo uso de índices de discordância ( $R_{fator}$ ) entre as amplitudes de estrutura observadas ( $F_{obs}$ ) e aquelas calculadas do modelo ( $F_{calc}$ ) (DRENTH, 1994).

$$R = \frac{\sum_{hkl} |F_{obs}| - |F_{calc}|}{\sum_{hkl} |F_{obs}|}$$

Um valor aceitável para o  $R_{fator}$  depende da resolução da estrutura. Para estruturas determinadas a altas resoluções ( $< 2,0 \text{ \AA}$ ) o valor do  $R_{fator}$  final deve estar abaixo de 0,2 ou 20 %.

Se a razão observações/parâmetros for baixa, o  $R_{fator}$  pode ser artificialmente diminuído às expensas da estereoquímica do modelo. Para evitar este tipo de erro BRUNGER (1992) introduziu uma nova metodologia no processo de refinamento - *cross validation* - onde aproximadamente 10 % dos dados (um grupo teste de reflexões) é isolado do remanescente (o

grupo de trabalho) e não é usado para o refinamento. Assim o refinamento é feito utilizando somente as reflexões pertencentes ao grupo de trabalho. A cada ciclo de refinamento os fatores de estrutura são calculados para ambos os grupos. Enquanto o grupo de trabalho é usado para o cálculo do  $R_{\text{str}}$  como definido acima, o grupo teste é usado para calcular um novo parâmetro,  $R_{\text{free}}$ , de maneira exatamente análoga. A diferença fundamental entre estes dois índices de discordância é que o procedimento de refinamento “sabe” sobre o grupo de trabalho (quer dizer, este é utilizado no próprio processo de refinamento), mas nunca “vê” o grupo teste. A grande vantagem do  $R_{\text{free}}$  é que é um indicador sensível para super-refinamento, isto é, um protocolo de refinamento que tenta extrair mais informação que os dados são capazes de prover.

O pacote de programas CCP4 inclui programas de refinamento bem potentes como Refmac 5, wArp/Arp e outros.

## **2. 19 Análise da estrutura**

### **2. 19. 1 Análise dos erros numa estrutura determinada experimentalmente**

Estruturas determinadas experimentalmente por cristalografia estão sujeitas aos mesmos tipos de erros que podem afetar qualquer medida experimental. Estes erros podem ser tanto aleatórios como sistemáticos afetando as intensidades medidas de feixes de raios X difratados. Mais ainda, nem toda a informação necessária para reprodução da estrutura molecular pode ser diretamente medida. A parte perdida (as fases dos feixes difratados) tem que ser obtida indiretamente e há normalmente grandes erros associados com sua estimativa inicial.

No estágio de interpretação do mapa de densidade eletrônica considerável intervenção humana é necessária. Todos estes fatores podem levar a erros na estrutura final que podem ser pequenos erros aleatórios de coordenadas, rotâmeros de cadeia lateral incorretos, regiões onde a estrutura não coincide com a seqüência, direcionamento contrário dos elementos de estrutura secundária e conectividade incorreta de tais elementos levando a uma topologia

errônea (BRANDEN AND JONES, 1990). Tais erros são normalmente uma função da resolução. Além dos tradicionais índices de qualidade para estruturas cristalinas ( $R_{int}$ , estereoquímica etc), vários outros itens úteis podem ser feitos nos arquivos PDB para se ter uma apreciação da qualidade da estrutura como um todo.

## 2. 19. 2 Qualidade do modelo

Os programas atualmente disponíveis para a avaliação da qualidade e normalidade de estruturas protéicas incluem PROCHECK (LASKOWSKI *et al.*, 1993), VERIFY\_3D (LUTHY BOWIE & EISENBERG, 1992), WHATIF (VRIEND, 1990; VRIEND & SANDER, 1993; HOLM AND SANDER, 1992), entre outros.

PROCHECK faz uma completa série de avaliações estereoquímicas incluindo ângulos e distâncias de ligações, planaridade de ligação peptídica, se os carbonos estão tetraédricos, energias de ligações de hidrogênio, rotâmeros de cadeias laterais e gráficos de Ramachandran. O último é algumas vezes, suficiente por si só para se distinguir entre uma estrutura bem determinada e uma que apresenta problemas estereoquímicos. Idealmente uma estrutura bem determinada terá mais que 90 % de seus resíduos nas regiões mais favoráveis do gráfico de Ramachandran. Dado que estas regiões são resíduo-dependentes, é importante que tais gráficos sejam também produzidos individualmente pelos 20 aminoácidos e comparados com as distribuições padrão.

WHATIF e VERIFY\_3D avaliam a qualidade das estruturas tridimensionais com referência a adequação do ambiente químico em torno de cada resíduo de aminoácido. No método empregado pelo WHATIF, a comparação é feita entre a distribuição atômica em torno de cada sub-fragmento rígido de cada resíduo e aquele encontrado numa distribuição padrão calculada de estruturas pré-determinadas. VERIFY\_3D adota uma maneira similar, mas mais simples na qual explicita as distribuições atômicas em torno de cada resíduo e oferece uma descrição do ambiente químico em termos de uma de dezoito classificações previamente definidas. Em ambos os casos um score pode ser atribuído à estrutura levando em conta o

quão comum a distribuição atômica ou ambiente químico é para um dado resíduo. Ambos métodos se referem à normalidade e não a qualidade em si. Em muitos casos resíduos de sítios ativos em enzimas, por exemplo, possuem ambientes químicos não usuais que são essenciais para sua atividade biológica. Eles podem ser assinalados por WATIF e VERIFY\_3D, mas não necessariamente estarem errados. Tais métodos devem ser tratados como guias para futura inspeção e não como avaliações absolutas.

### 2. 19. 3 As informações contidas nos arquivos PDB

O repositório padrão para os arquivos de coordenadas tridimensionais é o Protein Data Bank (PDB) (BERNSTEIN *et al.*, 1977). Um arquivo de coordenadas cristalográficas típico, identificado por seu código PDB (geralmente um número seguido por três caracteres alfanuméricos) consiste de um cabeçalho que inclui o seguinte tipo de informação:

- a fonte da proteína (ácido nucléico ou carboidrato) e se esta foi de alguma maneira modificada (por mutagênese sítio dirigida, complexada com um inibidor etc)
- A resolução da estrutura. Estruturas com resolução menor que 2 Å são normalmente consideradas “alta resolução”.
- O  $R_{factor}$ , que deve ser avaliado juntamente com a estereoquímica para o modelo.
- Parâmetros estereoquímicos que medem o desvio dos comprimentos de ligação, ângulos de ligação, ângulos de torção etc, de seu valor ideal (esperado ou alvo). Comprimento de ligações covalentes não devem desviar mais que 0,02 Å e ângulos de ligação e grupos planares por alguns graus. Ângulos de torção podem ser mais variáveis.
- A média dos erros de coordenadas é algumas vezes citada e indica a média do erro posicional esperado nas coordenadas. Este tem uma importância limitada, em algumas partes da estrutura (particularmente o núcleo) pode ter

muito maior precisão enquanto que, em "loops" flexíveis sobre a superfície, pode apresentar erros muito maiores. Provavelmente é mais útil olhar os fatores individuais de temperatura ao longo da estrutura.

- Comentários referentes a resíduos ou átomos perdidos indicam se não foi possível determinar a estrutura inteira. Dificuldades desta natureza normalmente aparecem por causa da desordem de certas partes da estrutura. Ocasionalmente pode ser devido a erros na seqüência de aminoácidos.
- A seqüência de aminoácidos. Esta não necessariamente deve ser idêntica aquela determinada pela proteína ou seqüenciamento do cDNA. Se a seqüência parece mais que uma vez, indica que há mais de uma molécula dentro da unidade assimétrica. Estas normalmente serão chamadas "cadeias" A, B etc e efetivamente representam independentes determinações da estrutura. Isto pode não ocorrer se existirem comentários sobre o uso de simetria não cristalográfica durante o refinamento.
- Comentários a respeito de padrões estruturais especiais, estrutura secundária, íons metálicos, inibidores etc.

Seguindo o cabeçalho estão os dados atômicos que consistem de 14 campos muitos dos quais auto-explicativos. Após o número de série do átomo vêm o nome do átomo (CA para carbono alfa etc), o código de 3 letras de aminoácido, identificador da cadeia e número de resíduo, as coordenadas cartesianas (x, y, z) seguidas pela ocupância atômica e o  $B_{\text{fator}}$  (fator de temperatura). A ocupância normalmente será de 1,000. Se não, pode significar que um átomo em particular pode ter uma de várias posições no cristal (devido a uma cadeia lateral desordenada, por exemplo) caso no qual mais que uma posição para o mesmo átomo pode estar representada. Se a ocupância é zero, significa que o átomo foi "modelado" provavelmente usando critérios estereoquímicos somente e não foi incluído no refinamento. Tais regiões devem ser tratadas com grande cuidado. O fator B representa o grau de desordem e corresponde a uma maior possibilidade de erro nas coordenadas. As partes bem

determinadas da estrutura tipicamente terão valores de  $B_{\text{fator}}$  abaixo de  $20 \text{ \AA}^2$  e onde os valores de fator B estão acima de  $50 \text{ \AA}^2$  a estrutura deve ser tratada com maior cuidado. Isto pode indicar outras conformações possíveis para a estrutura que não estão apresentadas.

Desde o começo de julho de 1999, o PDB é operado pelo Research Collaboratory for Structural Bioinformatics que pode ser acessado a <http://www.rcsb.org/pdb>. As buscas podem ser feitas pelo nome da proteína, código da estrutura, nome do autor, família protéica etc. e os resultados podem ser baixados ao computador do usuário e visualizados. Programas tais como RasMol (SAYLE AND MILNER-WHITE, 1995) e WebLab ViewerPro (<http://www.msi.com>) para visualização e manipulação de estruturas estão disponíveis para PC, Macintosh e estações de trabalho UNIX.

### **3. MATERIAIS E MÉTODOS**

#### **3.1 Cristalização**

Lectina de sementes de *Canavalia gladiata* purificada foi dissolvida (10 mg/mL) em Tris-HCl 10 mM, pH 7,5 contendo 5 mM de CaCl<sub>2</sub> e MnCl<sub>2</sub>. Foi feito o *screening* das condições de cristalização utilizando-se Screen I (Hampton Research Inc.) segundo o método da matriz esparsa (JANCARIK & KIM, 1991), por difusão de vapor, sistema de gota suspensa.

Para a obtenção de cristais da lectina complexada com açúcar, foram adicionados 3 µL de solução de alfa-metil-manosídeo 1 mM em poços com cristais de lectina de sementes de *C. gladiata* 12 horas antes da coleta de dados.

#### **3.2 Coleta e processamento de dados de difração de Raios X**

Dados de difração de raios X de comprimento de onda de 1,431 Å foram coletados utilizando-se fonte de luz síncrotron na estação PGr do Laboratório Nacional de Luz Síncrotron, LNLS (POLIKARPOV *et al.*, 1998a e POLIKARPOV *et al.*, 1998b), utilizando-se placa de imagens MARCCD e uma distância cristal/detector de 65 mm, à temperatura de 100k.

Foram utilizados os programas MARCCD para aquisição de imagens, MOSFLM (LESLIE, 1993) para integração e SCALA (EVANS, 1997) para análise dos dados.

#### **3.3 Substituição Molecular**

Utilizando o programa CCP4 foi obtido o volume de Mathews, a quantidade possível de moléculas presentes na unidade assimetria e a porcentagem de solvente.

A estrutura da molécula de lectina de sementes de *C. gladiata* foi determinada por substituição molecular usando-se lectina de *C. ensiformis* (PDB código 5CNA (NAISMITH *et al.*, 1994)) como modelo de busca. Foram retirados do modelo moléculas de água e o açúcar ligante.

O programa MolRep (VARGIN & TAPLYAKOV, 1998) foi utilizado para fazer a busca

dos monômeros na unidade assimétrica. O raio de integração utilizado foi de  $34,26^\circ$  com resolução máxima de  $3,00 \text{ \AA}$  e mínima de  $42,73 \text{ \AA}$ . Foram utilizados 10 picos de máxima intensidade para a busca das funções de rotação e translação.

Para obtenção da estrutura da CGL complexada com alfa-metil-manosídeo (AMM) a substituição molecular foi feita utilizando-se a estrutura obtida da lectina nativa como modelo de busca. O raio de integração utilizado foi de  $33,66^\circ$  com resolução máxima de  $3,00 \text{ \AA}$  e mínima de  $52,26 \text{ \AA}$ . Foram utilizados 10 picos de máxima intensidade para a busca das funções de rotação e translação.

### 3. 4 Refinamento

Várias opções de refinamento foram testadas, utilizando-se os programas X-PLOR (BRÜNGER, 1992) e REFMAC (MURSHODOV, DODSON AND VAGIN, 1996; MURSHODOV *et al.*, 1999).

Tanto X-PLOR como REFMAC possuem implementados em suas rotinas cálculos de  $R_{\text{factor}}$  e  $R_{\text{free}}$ . Como o cálculo destes fatores se faz sobre conjuntos de dados distintos e X-PLOR e REFMAC utilizam-se dos dados de difração de raios X compilados em formatos distintos, cuidado especial foi tomado na construção dos arquivos de reflexões específicos, de maneira a manter os mesmos conjuntos de dados para o cálculo de  $R_{\text{factor}}$  e  $R_{\text{free}}$ , em ambos os programas.

Refinamentos completos utilizando-se cada um dos programas e combinações dos mesmos foram realizados. A seguir é descrito o procedimento que apresentou melhor resultado, considerando-se os valores de R e o aspecto geral dos mapas de densidade eletrônica.

A partir do arquivo de reflexões original, no formato gerado pelo programa MOSFLM (arquivo MTZ), foi feito o escalonamento e indexação utilizando-se a rotina SCALA do pacote de programas CCP4 (arquivo MTZ), e gerado também um conjunto completo de reflexões pela rotina selecionando 5% dos dados para o cálculo do  $R_{\text{free}}$ .

O procedimento de refinamento da estrutura de CGL gerada pela substituição

molecular iniciou-se com um refinamento de restrições usando REFMAC5 (CCP4, 1994) e colocando-se os quatros monômeros como elementos de simetria não-cristalográfica e mantendo-se quaisquer ligações peptídicas *cis*. Todos os passos de refinamento foram feitos adicionando-se todos os átomos de hidrogênio, mas sem gravação no arquivo (PDB) de saída.

O segundo passo do refinamento da estrutura de CGL foi a substituição da seqüência primária, originalmente da lectina de *C.ensiformis*, pela seqüência da lectina de *C.gladiata* (YAMAUCHI *et al*, 1989).

Para tanto primeiramente foi gerado um mapa da densidade eletrônica  $2F_o-F_c$ , utilizando o resultado da substituição molecular como modelo através da rotina FFT do pacote CCP4. A troca das seqüências primárias foi realizada utilizando-se o programa XTALVIEW (McRee, 1992). Os resíduos Ser<sup>21</sup>, Asp<sup>58</sup>, Ala<sup>70</sup> e Glu<sup>192</sup> foram substituídos por Asn<sup>21</sup>, Gly<sup>58</sup>, Gly<sup>70</sup> e Asn<sup>192</sup>, e as cadeias laterais foram ajustadas em sua conformação de modo a satisfazer o mapa de densidade eletrônica.

O modelo da estrutura da CGL foi submetido a outro procedimento de refinamento utilizando-se o programa REFMAC5. Foram feitos 10 ciclos de refinamento de restrições utilizando novamente os monômeros como elementos de simetria não-cristalográfica.

Neste ponto, a segunda checagem do modelo da estrutura de CGL foi realizada utilizando-se o programa XFIT do pacote CCP4. Um mapa de densidade eletrônica  $2F_o-F_c$  foi calculado e o modelo final destas etapas de refinamento foi inspecionado para a realização das correções manuais necessárias.

A seguir, outro ciclo de refinamento com REFMAC5 e novo mapa de densidade eletrônica  $F_o-F_c$  foi gerado para ser utilizado no procedimento automático de procura de moléculas de água, através do mesmo programa XFIT.

Nova checagem foi realizada no programa XFIT, excluindo-se as moléculas de água localizadas em posições sem as correspondentes densidades eletrônicas, além daquelas que apresentaram fator de temperatura B superior a  $60 \text{ \AA}^2$ . Correções em cadeias laterais e outros ajustes na cadeia para coincidência com a densidade eletrônica foram feitos chegando-se assim

ao modelo refinado da estrutura de CGL.

O procedimento de refinamento da estrutura de CGL complexada com alfa-metil-manosídeo foi feito inicialmente com 20 ciclos de refinamento de corpo rígido e posteriormente com 10 ciclos de refinamento por restrições. Foi feito o mapa de densidade eletrônica para análise da existência do açúcar nos sítios ativos e seu posicionamento

Foram utilizados em todos os ciclos de refinamentos posteriores duas estruturas de simetria não-cristalográfica, a cadeia polipeptídica e a molécula de açúcar.

Após mais um ciclo de refinamento foram gerados os mapas de densidade eletrônica e foram feitas correções manuais nas cadeias laterais dos resíduos de modo a satisfazer o mapa  $2F_o-F_c$ . O mapa  $F_o-F_c$  foi utilizado para a adição de 400 moléculas de água utilizando o programa XFIT.

Outros 10 ciclos de refinamento de restrição foram efetuados na estrutura chegando-se assim ao modelo refinado da estrutura de CGL complexada com alfa-metil-manosídeo.

### **3. 5 Avaliação da estrutura**

A avaliação das estruturas obtidas foi feita utilizando-se a rotina Sfcheck (VAGUINE, RICHELLE, WODAK, 1999) do pacote de programas CCP4.

## 4. RESULTADOS

Os melhores cristais foram obtidos com 4 dias em gotas contendo 3  $\mu\text{L}$  de solução de lectina de *C.gladiata* numa concentração de 10 mg/mL, misturada com igual volume de solução Tris-HCl, pH 8,5, 1M de sulfato de amônio. Foram obtidos cristais de volume médio aproximado de 2-3  $\text{mm}^3$ .

### 4.1 Lectina de sementes de *C.gladiata* nativa

Foi testado e utilizado como crio-protetor do cristal para a coleta de dados a solução do poço com 30% de glicerol.

Usando-se oscilações de 1°, 160 imagens de padrão de difração de CGL nativa foram coletadas no LNLS com resolução de 42.728 Å - 1.99 Å. Os dados foram integrados usando-se o programa MOSFLM (LESLIE, 1993).

Os cristais pertencem ao grupo espacial ortorrômbico C222<sub>1</sub>. A cela unitária tem dimensões  $a = 100,90$  Å,  $b = 115,35$  Å e  $c = 241,08$  Å e com  $\alpha = \beta = \gamma = 90^\circ$ .

O processamento e aferimento dos dados foram feitos à resolução de 40,71 Å - 2,30 Å, utilizando SCALA (EVANS, 1997). Os dados não foram processados à resolução de coleta porque a última camada de resolução, embora tivesse melhor completeza (97,8%), apresentou  $R_{\text{merge}}$  de 34,4%.

O número total de reflexões gravadas foi de 519778, sendo 60164 reflexões únicas. A completeza dos dados na faixa de resolução é de 95,9%.

O volume da cela unitária é de 2805885,5 Å<sup>3</sup>. O coeficiente de Matthews (MATTHEWS, 1968) indicou maior probabilidade de um tetrâmero na unidade assimétrica (Tabela 2).

Após cálculos das funções de rotação e translação utilizando o programa MolRep (VARGIN & TAPLYAKOV, 1997) os melhores resultados indicaram um coeficiente de correlação de 56,9% e  $R_{\text{int}}$  de 42,1% para a estrutura de CGL.

Tabela 2 – Valores de  $V_m$  para a lectina de sementes de *C. gladiata*. O valor de maior probabilidade encontra-se na linha sombreada.

No. unid.	mol/ assim.	Coefficiente de Matthews	% solvente
1		13,7	91,0
2		6,9	81,9
3		4,6	72,9
4		3,4	63,8
5		2,7	54,8
6		2,3	45,7
7		2,0	36,7
8		1,7	27,7
9		1,5	18,6
10		1,4	9,6
11		1,2	0,5

O primeiro ciclo de refinamento utilizando REFMAC5 foi feito utilizando as quatro cadeias polipeptídicas como unidades de simetria não-cristalográfica. Foram encontradas ligações peptídicas *cis* entre Ala207-Asp208 em todos os quatros monômeros. Após 10 ciclos de refinamento foi obtido um modelo onde  $R_{\text{fator}}$  e  $R_{\text{free}}$  convergiram a 21,24 % e 24,99 %, respectivamente.

Após a substituição dos resíduos de aminoácidos mais uma etapa de refinamento foi feita com REFMAC5 com  $R_{\text{fator}}$  chegando a 19,83 % e  $R_{\text{free}}$  a 23,87 %. O coeficiente de correlação chegou a 93,0 %.

Foram adicionadas 432 moléculas de água com XtalView (McREE, 1999) ao modelo, e o último refinamento por restrição foi feito resultando em  $R_{\text{fator}}$  de 18,21 % e  $R_{\text{free}}$  de 23,05 %. O coeficiente de correlação  $F_o-F_c$  chegou a 94,0 %. Os resultados finais do refinamento da CGL estão apresentados na Tabela 3.

Tabela 3 - Estatísticas finais do refinamento da lectina nativa de sementes de *C. gladiata*.

Grupo Espacial	C <sub>222</sub> <sub>1</sub>
Parâmetros de Célula Unitária	a = 100,90 Å b = 115,35 Å c = 241,08 Å
Faixa de Resolução	10 - 2,3 Å
Conteúdo da Unidade Assimétrica	Um tetrâmero
Reflexões Únicas	60164
Completeza (%)	95,9 (97,8)
Rmerge (%)	8,8 (34,4)
$\langle I/\sigma(I) \rangle$	7,2 (2,2)
Rfator (%)	18,21
Rfree (%)	23,05

A estrutura obtida foi avaliada apresentando 85,9 % de resíduos de aminoácidos em posições muito favoráveis no gráfico de Ramachandran e 14,1 % em posições adicionalmente favoráveis. Nenhum resíduo se posicionou em áreas generosamente permitidas ou proibidas. A média dos desvios padrões de comprimento de ligações foi de 0,027 Å e dos ângulos de ligação foi de 2,221 ° (Tabela 4).

Tabela 4 - Detalhes estereoquímicos dos resíduos de aminoácidos da lectina nativa de sementes de *C. gladiata*.

<b>Distribuição dos resíduos no gráfico de Ramachandran</b>	
<b>Resíduos em posições</b>	
Muito favoráveis	85,9 %
Adicionalmente favoráveis	14,1 %
Generosamente permitidas	0,0 %
Proibidas	0,0 %
<b>Desvio Médio Quadrático</b>	
Comprimento de ligações	0,027 Å
Ângulos de ligação	2,221 °

#### **4. 2 Lectina de sementes de *C.glabrata* complexada com alfa-metil-manosídeo**

Usando-se oscilações de  $1^\circ$ , 120 imagens de padrão de difração de CGL complexada com alfa-metil-manosídeo foram coletadas no LNLS com resolução de  $52,262 \text{ \AA} - 2,219 \text{ \AA}$ . Os dados foram processados utilizando-se o programa MOSFLM (LESLIE, 1993).

Os cristais pertencem ao grupo espacial ortorrômbico  $C222_1$ . A cela unitária tem dimensões  $a = 101,241 \text{ \AA}$ ,  $b = 115,861 \text{ \AA}$  e  $c = 242,375 \text{ \AA}$  e com  $\alpha = \beta = \gamma = 90^\circ$ .

Os dados foram processados e aferidos à resolução de  $52,26 \text{ \AA} - 2,32 \text{ \AA}$ , utilizando-se o programas SCALA (EVANS, 1997) em razão do  $R_{\text{merge}}$  para a última camada de resolução ser muito alto (32,3 %).

O número total de reflexões gravadas foi de 428157, sendo 58145 reflexões únicas. A completude dos dados na faixa de resolução é de 96,2 % e da última camada, 97,4 %.

O volume da cela unitária é de  $2843030,5 \text{ \AA}^3$ .

Após cálculos das funções de rotação e translação utilizando o programa MolRep (VARGIN & TAPLYAKOV, 1997) os melhores resultados indicaram um coeficiente de correlação de 64,2 % e  $R_{\text{fator}}$  de 37,3 % para a estrutura de CGL complexada.

O primeiro ciclo de refinamento utilizando REFMAC5 foi feito utilizando as quatro cadeias polipeptídicas como corpos rígidos. Foram encontradas ligações peptídicas *cis* entre Ala207-Asp208 em todos os quatros monômeros. Após 20 ciclos de refinamento de corpo rígido foi obtido um modelo onde  $R_{\text{fator}}$  e  $R_{\text{free}}$  convergiram a 33,53 % e 34,36 %, respectivamente e correlação de 78,6 % com o modelo de busca.

O ciclo de refinamento seguinte feito foi com restrições utilizando as quatro cadeias polipeptídicas e os açúcares como unidades de simetria não-cristalográfica. Após 10 ciclos de refinamento foi obtido um modelo onde  $R_{\text{fator}}$  e  $R_{\text{free}}$  convergiram a 20,97 % e 26,00 %, respectivamente e apresentando correlação de 92,2 % com o modelo de busca.

Após a correção das cadeias laterais dos resíduos de aminoácidos foram adicionadas 400 moléculas de água com XtalView (McREE, 1999) ao modelo, e o último refinamento por

restrição foi feito resultando em  $R_{\text{fator}}$  de 18,59 % e  $R_{\text{free}}$  de 24,57 %. O coeficiente de correlação chegou a 93,9 %.

Tabela 5 - Estatísticas finais do refinamento da lectina de sementes de *C. gladiata* complexada com alfa-metil-manosídeo.

Grupo Espacial	$C222_1$
Parâmetros de Célula Unitária	a = 101,24 Å b = 115,86 Å c = 242,38 Å
Faixa de Resolução	10 – 2,32 Å
Conteúdo da Unidade Assimétrica	Um tetrâmero
Reflexões Únicas	58145
Completeza (%)	96,2 (97,4)
$R_{\text{merge}}$ (%)	8,2 (32,3)
$\langle I/\sigma(I) \rangle$	7,8 (2,2)
$R_{\text{fator}}$ (%)	18,59
$R_{\text{free}}$ (%)	24,57

A estrutura obtida foi avaliada apresentando 86,0 % de resíduos de aminoácidos em posições muito favoráveis no gráfico de Ramachandran e 14,0 % em posições adicionalmente favoráveis. Nenhum resíduo se posicionou em áreas generosamente permitidas ou proibidas. A média dos desvios padrões de comprimento de ligações foi de 0,027 Å e dos ângulos de ligação foi de 2,221 ° (Tabela 6).

Tabela 6 - Detalhes estereoquímicos dos resíduos de aminoácidos da lectina de sementes de *C.glabrata* complexada com alfa-metil-manosídeo.

<b>Distribuição dos resíduos no gráfico de Ramachandran</b>	
Resíduos em posições	
Muito favoráveis	86,0 %
Adicionalmente favoráveis	14,0 %
Generosamente permitidas	0,0 %
Proibidas	0,0 %
<b>Desvios Quadráticos</b>	
Comprimento de ligações	0,027 Å
Ângulos de ligação	2,221 °

## 5. DISCUSSÃO

Na tribo Diocleinae, somente as lectinas de *Canavalia ensiformis*, *C. brasiliensis*, *Dioclea grandiflora*, *D.guianensis* e *Cratylia mollis* tiveram suas estruturas tridimensionais resolvidas por difração de raios X. A similaridade de seqüências primárias entre estas lectinas e a CGL é de 99%, 98%, 83%, 88% e 78%, respectivamente (Figura 46). Embora estas lectinas apresentem alta similaridade na seqüência primária, têm pequenas diferenças nas estruturas terciárias que se refletem em alterações em suas estruturas quaternárias. Esta parece ser a razão de tantas diferenças na potência de várias atividades biológicas apresentadas por estas lectinas.

	1	10	20	30	40	50	60	70
ConA	ADTIVAVELDT	YPNTDIGDPS	SYPHIGID	IKSVRSKKTAK	KNMONGKVGTAHI	IYNSVVKRRLSAVVSYPHA		
ConBr	ADTIVAVELDT	YPNTDIGDPS	SYPHIGID	IKSVRSKKTAK	KNMONGKVGTAHI	IYNSVVKRRLSAVVSYPHA		
DGL	ADTIVAVELNS	YPNTDIGDEN	SYPHIGID	IKSIRSESTAR	KNMONGKVGTAHI	IYNSVVKRRLSAVVSYPHA		
Dgui	ADTIVAVELDS	YPNTDIGDPS	SYPHIGID	IKSIRSESTAR	KNMONGKVGTAHI	IYNSVVKRRLSAVVSYPHA		
Cra	ADTIVAVELDT	YPNTDIGDPS	SYPHIGID	IKSIRSESTAR	KNMONGKVGTAHI	IYNSVVKRRLSAVVSYPHA		
CGL	ADTIVAVELDT	YPNTDIGDPS	SYPHIGID	IKSVRSKKTAK	KNMONGKVGTAHI	IYNSVVKRRLSAVVSYPHA		

	80	90	100	110	120	130	140
ConA	DSATVSYDVID	DNVLPWVRVGLSA	STGLYKETNTIL	SWSETSKLKS	NSSTHE	TNALHFP	QPSKDC
ConBr	DSATVSYDVID	DNVLPWVRVGLSA	STGLYKETNTIL	SWSETSKLKS	NSSTHE	TNALHFP	QPSKDC
DGL	SSSTVSYDVID	DNVLPWVRVGLSA	STGLYKETNTIL	SWSETSKLKT	NSSTHE	TNALHFP	QPSKDC
Dgui	SSSTVSYDVID	DNVLPWVRVGLSA	STGLYKETNTIL	SWSETSKLKT	NSSTHE	TNALHFP	QPSKDC
Cra	SSATVSYDVID	DNVLPWVRVGLSA	STGLYKETNTIL	SWSETSKLKS	NSSTHE	TNALHFP	QPSKDC
CGL	DSATVSYDVID	DNVLPWVRVGLSA	STGLYKETNTIL	SWSETSKLKS	NSSTHE	TNALHFP	QPSKDC

	150	160	170	180	190	200	210
ConA	ILQGDAT	TGTDGNLELTRVS	SHGSP	QSSVGRALF	YAPVHIK	ESSAVVASFE	ATFTFLIKSPDR
ConBr	ILQGDAT	TGTDGNLELTRVS	SHGSP	QSSVGRALF	YAPVHIK	ESSAVVASFE	ATFTFLIKSPDR
DGL	ILQGDAT	TGTDGNLELTRVS	SHGSP	QSSVGRALF	YAPVHIK	ESSAVVASFE	ATFTFLIKSPDR
Dgui	ILQGDAT	TGTDGNLELTRVS	SHGSP	QSSVGRALF	YAPVHIK	ESSAVVASFE	ATFTFLIKSPDR
Cra	ILQGDAT	TGTDGNLELTRVS	SHGSP	QSSVGRALF	YAPVHIK	ESSAVVASFE	ATFTFLIKSPDR
CGL	ILQGDAT	TGTDGNLELTRVS	SHGSP	QSSVGRALF	YAPVHIK	ESSAVVASFE	ATFTFLIKSPDR

	220	230
ConA	APFISNI	DSSIFSGSTGRLLGLFFDAH
ConBr	APFISNI	DSSIFSGSTGRLLGLFFDAH
DGL	TEFIANT	DTSIFSGSGGRLLGLFFDAH
Dgui	TEFIANT	DTSIFSGSGGRLLGLFFDAH
Cra	TEFIANT	DSSIFSGSGGRLLGLFFDAH
CGL	APFISNI	DSSIFSGSTGRLLGLFFDAH

Figura 46 - Alinhamento das seqüências primárias das lectinas de *Canavalia ensiformis* (ConA), *C. brasiliensis* (ConBr), *Dioclea grandiflora* (DGL), *D. guianensis* (Dgui), *Cratylia mollis* (Cra) e *Canavalia gladiata* (CGL).

As substituições de aminoácidos que foram feitas na seqüência da estrutura modelo (5CNA, ConA) foram feitas e o mapa de densidade eletrônica foi satisfeito nestas regiões (Figuras 47, 48, 49 e 50).



Figura 48 - Substituição do resíduo Ser<sup>21</sup> por Asn no monômero da CGL.



Figura 47 - Substituição do resíduo Asp<sup>56</sup> por Gly no monômero da CGL.



Figura 49 - Substituição do resíduo Ala<sup>70</sup> por Gly no monômero da CGL.

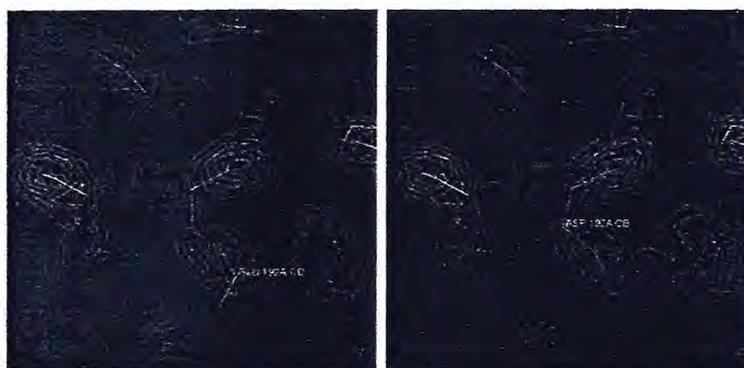


Figura 50 - Substituição do resíduo Glu<sup>192</sup> por Asp no monômero da CGL.

A estrutura tridimensional de todas as lectinas de sementes de leguminosas pode ser descrita como sendo cadeias de cerca de 30 kDa (massa molecular aparente) que formam uma folha-beta pregueada posterior quase plana, constituída por cinco cadeias; uma outra folha-beta dianteira que é curva e formada por sete cadeias e uma folha-beta curta, de cinco cadeias, que se situa no topo da molécula. Entre estas folhas-beta, vários "loops" fazem a conexão (BANERJEE *et al.*, 1996).

A estrutura completa da CGL nativa é mostrada na Figura 51. O tetrâmero é formado pela associação de dois dímeros canônicos típicos de leguminosas associados através de contatos pelas folhas-beta para formar a estrutura quaternária (Figura 52). A cadeia peptídica

formadora do monômero tem 237 resíduos.

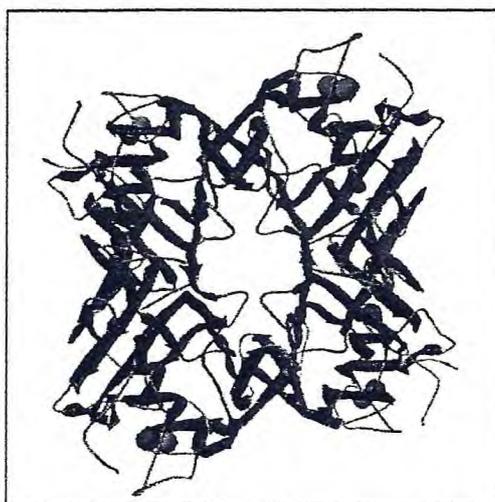


Figura 51 - Estrutura tridimensional completa do tetrâmero de CGL. As bolas verdes representam íons  $\text{Ca}^{2+}$  e as bolas vermelhas,  $\text{Mn}^{2+}$ . A figura foi feita com o programa PdbViewer (GUEX AND PEITSCH, 1997).

Cada monômero apresenta 16 trechos em conformação secundária do tipo beta interligadas por resíduos que não apresentam nenhuma estrutura secundária regular e quatro pequenos trechos que se aproximam de uma conformação de alfa-hélice. O primeiro trecho em alfa-hélice compreende os resíduos Thr<sup>15</sup>-Ile<sup>17</sup>, o segundo Leu<sup>81</sup>-Asn<sup>83</sup>, o terceiro Thr<sup>150</sup>-Gly<sup>152</sup> e o quarto Gly<sup>227</sup>-Leu<sup>229</sup>.

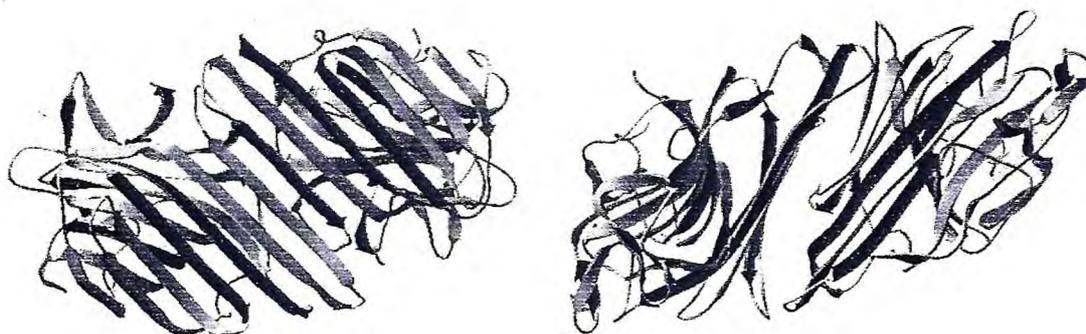


Figura 52 - Dímeros da DGL, à esquerda, dímeros canônicos; à direita, não-canônicos.

Os trechos de conformação tipo beta são Ile<sup>4</sup>-Asp<sup>10</sup> (7 resíduos), His<sup>24</sup>-Ile<sup>29</sup> (6 resíduos), Lys<sup>36</sup>-Lys<sup>39</sup> (4 resíduos), Val<sup>47</sup>-Asn<sup>60</sup> (14 resíduos), Arg<sup>65</sup>-Ser<sup>71</sup> (7 resíduos), Ser<sup>77</sup>-Asp<sup>83</sup> (7 resíduos), Trp<sup>88</sup>-Ser<sup>96</sup> (7 resíduos), Thr<sup>105</sup>-Ser<sup>117</sup> (13 resíduos), Thr<sup>123</sup>-Phe<sup>130</sup> (8 resíduos), Leu<sup>140</sup>-Gly<sup>144</sup> (5 resíduos), Thr<sup>147</sup>-Thr<sup>148</sup> (2 resíduos), Leu<sup>154</sup>-Glu<sup>155</sup> (2 resíduos), Val<sup>170</sup>-Phe<sup>175</sup> (6 resíduos), Val<sup>179</sup>-His<sup>180</sup> (2 resíduos), Ala<sup>186</sup>-Leu<sup>198</sup> (13 resíduos) e Gly<sup>209</sup>-Asn<sup>216</sup> (8 resíduos).

Na prática, o que se vê na estrutura são duas grandes folhas-beta pregueadas. A face posterior ou dorso da molécula é composta por uma folha-beta formada por 6 segmentos antiparalelos (Leu<sup>140</sup>-Gly<sup>144</sup>, Val<sup>170</sup>-Phe<sup>175</sup>, Trp<sup>88</sup>-Ser<sup>96</sup>, Gly<sup>209</sup>-Asn<sup>216</sup>, Ile<sup>4</sup>-Asp<sup>10</sup> e His<sup>24</sup>-Ile<sup>29</sup>) e a face anterior composta por 7 segmentos (Thr<sup>123</sup>-Phe<sup>130</sup>, Thr<sup>105</sup>-Ser<sup>117</sup>, Ala<sup>186</sup>-Leu<sup>198</sup>, Val<sup>47</sup>-Asn<sup>60</sup>, Arg<sup>65</sup>-Ser<sup>71</sup>, Ser<sup>77</sup>-Asp<sup>83</sup> e Lys<sup>36</sup>-Lys<sup>39</sup>) também antiparalelos (Figura 53). Estas folhas-beta são mantidas juntas pelos segmentos sem estrutura secundária regular que os ligam.

Como em outras lectinas ConA-like, o dímero canônico é formado com as folhas-beta

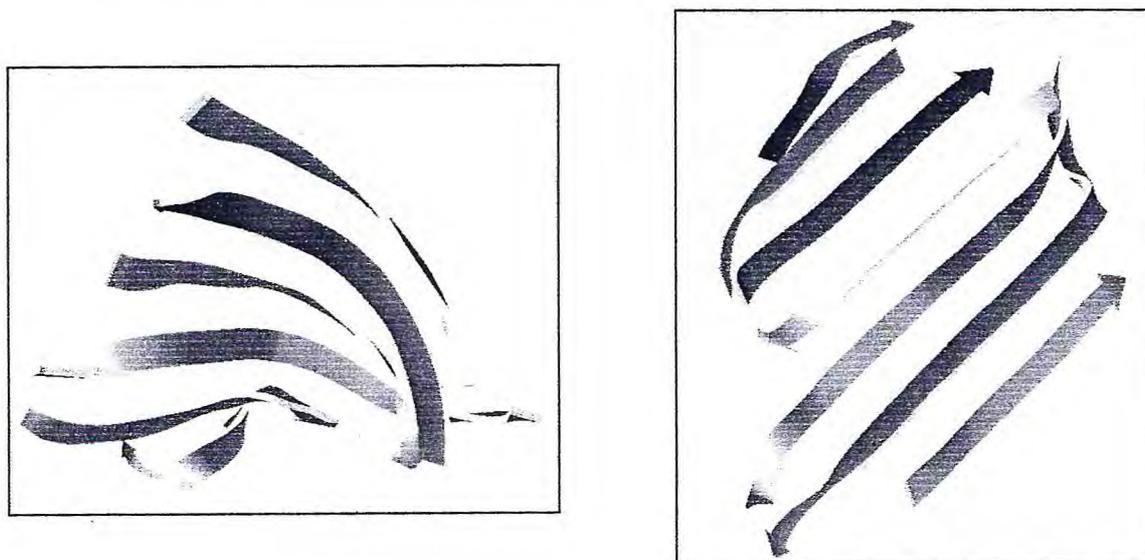


Figura 53 - Folhas-beta de 6 segmentos e de 7 segmentos do monômero de CGL.

de sete segmentos de cada monômero associadas através de contatos beta-beta de maneira a formar uma folha única de 14 segmentos. As folhas Asn<sup>124</sup>-Gln<sup>132</sup> de cada monômero estão associadas antiparalelamente. Além dos contatos beta-beta, ligações de hidrogênio também estabilizam o tetrâmero como a existente entre os resíduos Asn<sup>131</sup> e Gln<sup>132</sup> de um monômero e Thr<sup>123</sup> e Asn<sup>124</sup> do outro monômero que têm entre si, em média, a distância de 2,66 Å (Figura 54).

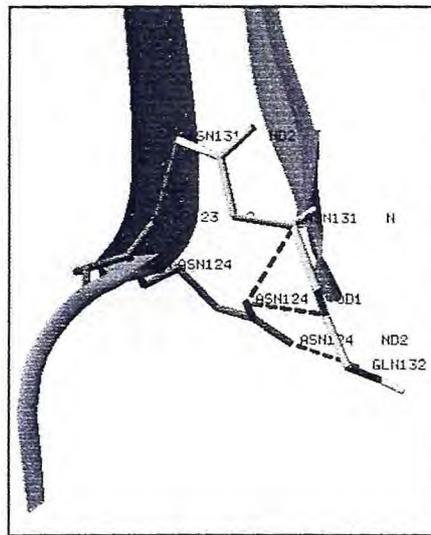


Figura 54 - Ligações de hidrogênio entre os resíduos Asn<sup>131</sup> e Gln<sup>132</sup> de um monômero e Thr<sup>123</sup> e Asn<sup>124</sup> do outro monômero de CGL. Em tracejado verde estão representadas as ligações de hidrogênio entre as duas folhas  $\beta$ .

Os contatos entre os dímeros não canônicos são feitos através de contatos frente a frente dos conjuntos de folhas- $\beta$  de 6 segmentos. A distância entre eles varia em torno de 10,05 Å (entre os resíduos His<sup>127</sup> dos dois monômeros) a 5,46 Å (entre os resíduos Gly<sup>58</sup>) (Figura 55).

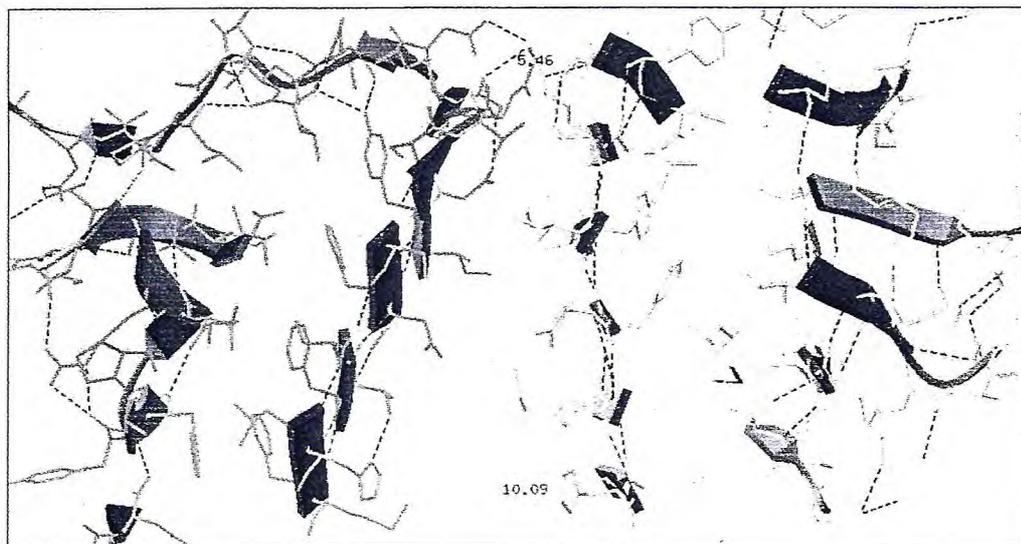


Figura 55 – Corte longitudinal de dímeros não-canônicos de CGL mostrando os contatos  $\beta$ - $\beta$  entre eles. A distância entre a folhas  $\beta$  varia de 10,09 a 5,46 Å.

Nas estruturas de ConA já publicadas, alguns "loops" superficiais (118-121, 149-151, 160-163) (NAISMITH, 1994; DEACON *et al*, 1997; BOUCKAERT *et al*, 1999) e resíduos da região N-terminal mostraram pouca densidade eletrônica. No caso da CGL os "loops" 149-151 e 160-163 apresentaram forte densidade eletrônica, a exceção foi apenas do "loop" superficial 118-121 (Figura 56).

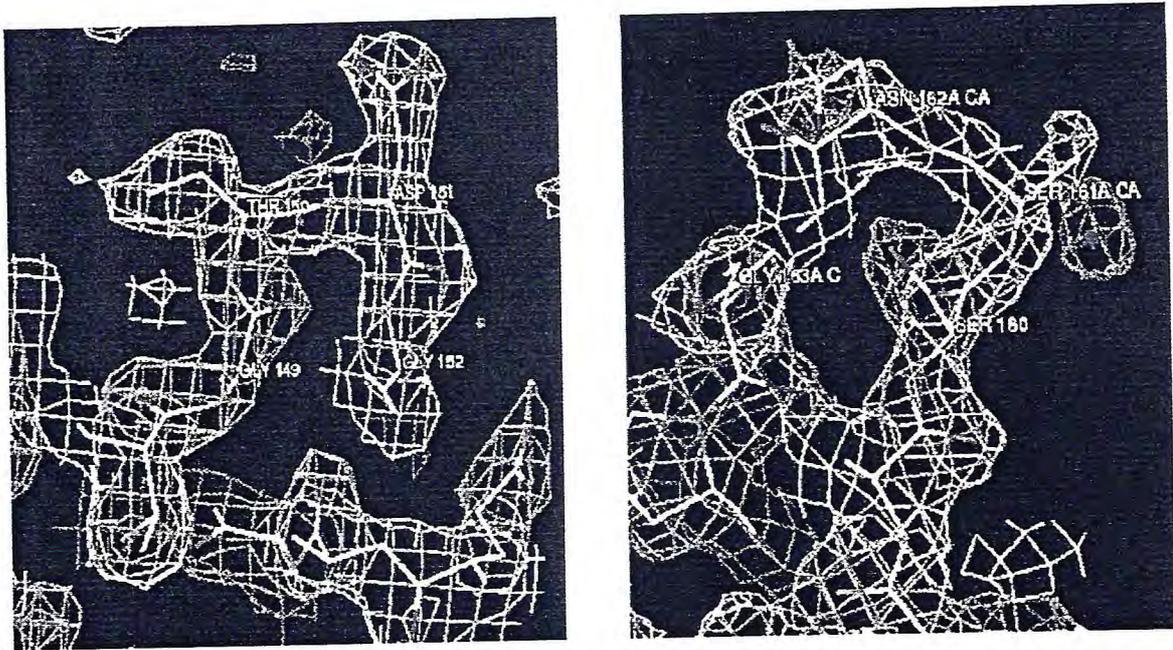


Figura 56 – “Loops” da CGL. à esquerda o “loop” 149-151 e à direita “loop” 160-163.

Na CGL nativa os aminoácidos envolvidos no sítio de ligação a metais são conservados e suas estruturas mostram similaridades com aquelas determinadas em outras lectinas de leguminosas (CALVETE *et al.*, 1999). (Figura 57)

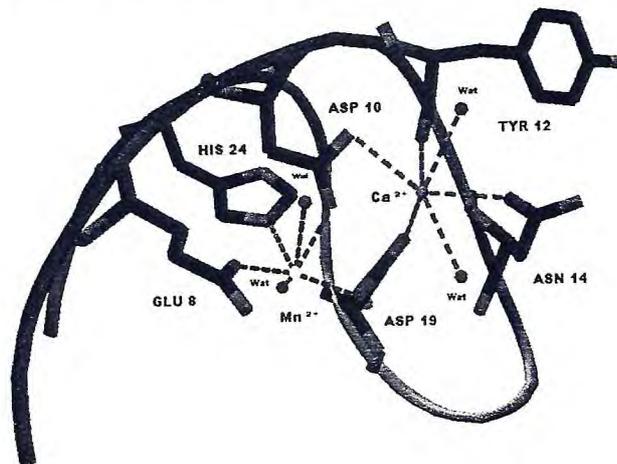


Figura 57 - Sítio de ligação de metais. Com vermelho estão representados os átomos de carbono, em azul os de nitrogênio. O íon Cálcio está representado em magenta e o íon manganês em amarelo. Moléculas de água são representadas como esferas vermelhas. O esqueleto carbônico protéico é representado pela fita em azul.

Os monômeros contêm íons manganês e cálcio na vizinhança do sítio de ligação a carboidrato (Fig. 53). O íon metálico cálcio provoca a isomerização trans-cis da ligação peptídica Ala<sup>207</sup>-Asp<sup>208</sup> que é conservada em todas as estruturas de lectinas de leguminosas conhecidas. A ligação cis-peptídica contribui para a estabilização do sítio de ligação orientando as posições dos resíduos Asn<sup>14</sup> e Arg<sup>228</sup>. Cada um dos íons metálicos forma ligações de coordenação com quatro cadeias laterais de aminoácidos, e duas moléculas de água, Glu<sup>8</sup>, Asp<sup>10</sup>, Asp<sup>19</sup> e His<sup>24</sup> com Mn<sup>2+</sup>; Asp<sup>10</sup>, Tyr<sup>12</sup>, Asn<sup>14</sup> e Asp<sup>19</sup> com Ca<sup>2+</sup>. Na tabela 7 são dadas as distâncias entre átomos de alguns destes resíduos e os íons metálicos. No caso do íon cálcio, uma das moléculas de água forma uma ponte entre o metal e o grupo carbonila da cadeia principal do Asp<sup>208</sup> também estabilizando a ligação peptídica cis Ala<sup>207</sup>-Asp<sup>208</sup>.

Tabela 7 - Distância entre os átomos de resíduos que fazem coordenação com os íons metálicos na molécula da lectina de sementes de *C.glabrata*.

Átomo	Resíduo	Metal	Distância
OE2	8 GLU	MN	2.297
OD2	10 ASP	MN	1.974
OD1	19 ASP	MN	2.151
O	12 TYR	CA	2.271
OD2	19 ASP	CA	2.341

No modelo final, 85,9 % dos resíduos estão em regiões amplamente favoráveis no gráfico de Ramachandran e 14,1 % em regiões adicionalmente permitidas.

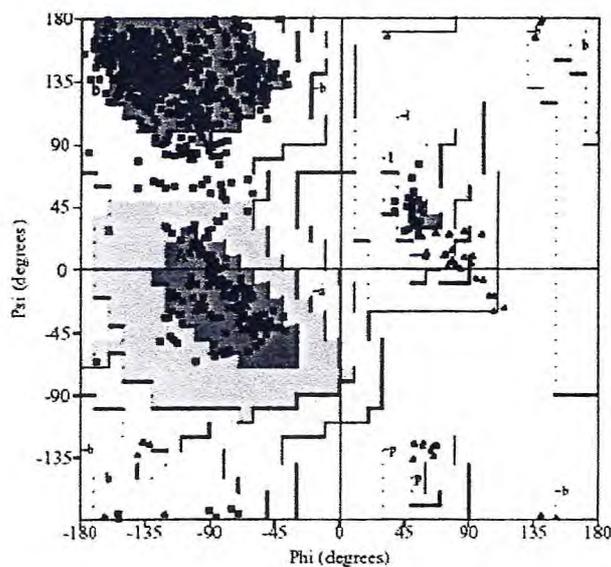


Figura 58 - Gráfico de Ramachandran dos resíduos da lectina de sementes de *C. gladiata*. Feito com o programa PROCHECK.

Promovendo-se a sobreposição das cadeias A dos tetrâmeros de CGL e da ConA (5CNA) utilizando o programa SPDBViewer, obtêm-se um desvio médio quadrático de 1,95 Å e entre estas cadeias. Na Figura 59 fica evidente a diferença entre CGL e ConA na angulação dos monômeros para a formação do tetrâmero.



Figura 59 – Sobreposição das cadeias A dos tetrâmeros de CGL (amarelo) e da ConA (5CNA, azul) utilizando o programa SPDBViewer.

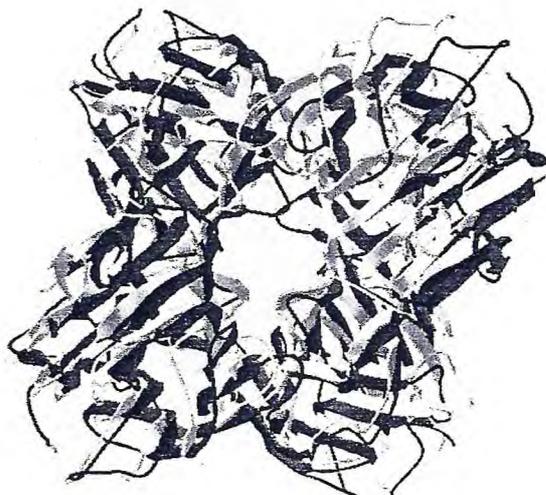


Figura 60 – Sobreposição dos tetrâmeros de CGL (amarelo) e da ConA (5CNA, azul) utilizando o programa SPDBViewer.

Quando a sobreposição é feita levando-se em conta o tetrâmero como um todo, o desvio padrão quadrático entre as posições dos átomos de carbono alfa das estruturas de ConA e CGL passa a 4,66 Å (Figura 60).

Cerca de 49,2 % dos resíduos de aminoácidos da estrutura de CGL estão na superfície da molécula, em contato com o solvente (resultados obtidos através do programa PROTEIN VOLUME EVALUATOR (PONTIUS, RICHELLE AND WODAK, 1996)).

Na estrutura da lectina de CGL complexada com o açúcar  $\alpha$ -metil manosídeo, a densidade eletrônica deste último foi fortemente visível indicando que o processo de soaking utilizado foi altamente eficiente na ocupação dos sítios de ligação de carboidratos.

Entre outras, as ligações de hidrogênio entre o açúcar e Asn<sup>14</sup>, Asp<sup>208</sup>, Gly<sup>89</sup> e Leu<sup>99</sup> ficaram evidenciadas (Figura 61).

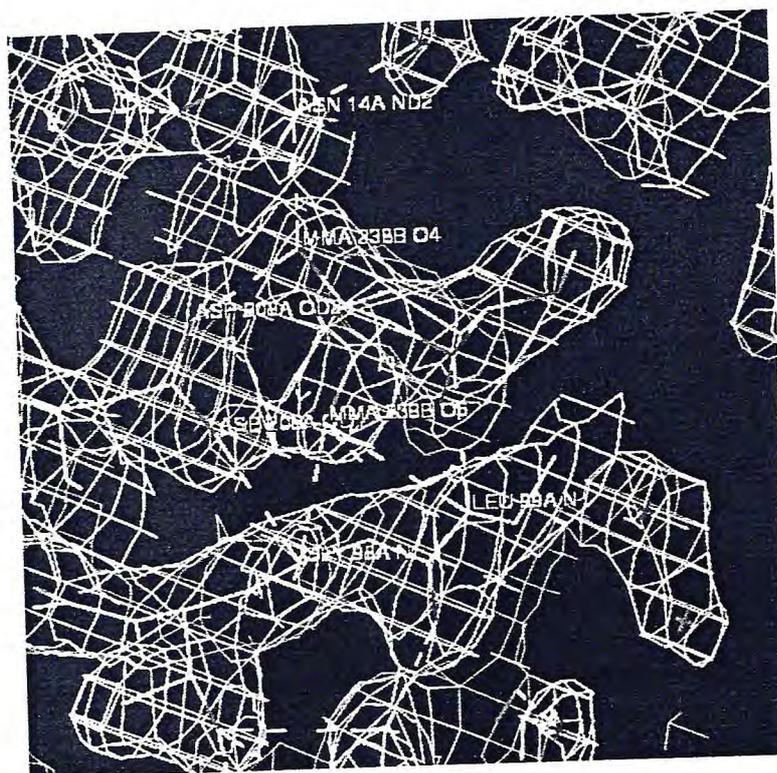


Figura 61 – Posicionamento do açúcar  $\alpha$ -metil-manosídeo no sítio de ligação de carboidratos e as ligações de hidrogênio estabelecidas com a cadeia polipeptídica de CGL. Fica evidenciada a densidade eletrônica das estruturas.

## 6. CONCLUSÕES

Esta tese mostra a estrutura tridimensional da lectina de sementes de *Canavalia gladiata*, uma lectina ConA-like, a 2,3 Å de resolução, que foi depositada no Protein Databank com o código 1WUV e a estrutura da lectina complexada com alfa-metil-manosídeo, ainda não depositada.

As subunidades de CGL mostram um alto grau de conservação na estrutura terciária e nos sítios de ligação. Arquitetura de lectinas de leguminosas é composta de folhas-beta pregueadas, seus monômeros são formados por 16 folhas-beta antiparalelas interconectadas por voltas e “loops”. A estrutura quaternária é relacionada à atividade biológica, uma vez que a distribuição espacial dos sítios de ligação a carboidratos nas lectinas oligoméricas determina a habilidade destas proteínas distinguirem e fazerem ligações cruzadas com sacarídeos ligantes multivalentes ou glicoconjugados de superfície celular. O exame de estruturas adicionais de lectinas de *Canavalias* indubitavelmente contribuirá significativamente para posterior consideração da estrutura molecular destas lectinas, no que concerne a suas atividades biológicas tão distintas.

Embora algumas estruturas de ConA já sejam conhecidas, as informações sobre estruturas tridimensionais de outras lectinas de Diocleinae, particularmente de outras espécies do gênero *Canavalia*, são restritas. Este grupo de lectinas apresenta um alto grau de similaridade, mas mostra significantes diferenças no potencial de várias propriedades biológicas e estudos mais profundos devem ser feitos para estabelecer as relações estrutura/função neste grupo de proteínas.

## REFERÊNCIAS

- BANERJEE, R. *et al.* Conformation, Protein-Carbohydrate Interactions and a Novel Subunit Association in the Refined Structure of Peanut Lectin-Lactose Complex. **Journal of Molecular Biology**, v. 259, n. 2, p. 281-296. 1996.
- BARRAL-NETTO, M. *et al.* Human lymphocyte stimulation by legume lectins from the Diocleae tribe. **Immunol. Invest.**, v. 21, p. 297-303. 1992.
- BARRAL-NETTO, M. *et al.* *In vivo* protective effect of the lectin from *Canavalia brasiliensis* on BALB/c mice infected by *Leishmania amazonensis*. **Acta Trop.** v. 60, n. 4, p. 237-250. 1996.
- BENTO, C.A.M. *et al.* Rat paw edema and leukocyte immigration induced by plant lectins. **Agent Actions**, v. 38, p. 48-54. 1993.
- BERGFORS, T.M. **Protein Crystallization**. La Jolla, CA: International University Line, 1999.
- BERMAN, H.M. *et al.* The Protein Data Bank. **Nucleic Acids Research**, v. 28, p. 235-242. 2000.
- BERNSTEIN, F.C. *et al.* The Protein Data Bank: a computer-based archival file for macromolecular structures. **J. Mol. Biol.**, v. 112, p. 535-542. 1977.
- BINSTED, N.; STRANGE, R.W.; HASNAIN, S.S. Constrained and restrained refinement in EXAFS data analysis with curved wave theory. **Biochemistry**, v. 31, n. 48, p. 12117-12125. 1992.
- BLUNDELL, T.L., JOHNSON, L.N. **Protein Crystallography**. New York, USA: Academic Press, 1976.
- BOUCKAERT, J. *et al.* The crystal structures of Man( $\alpha$ 1-3)Man( $\alpha$ 1-O)Me and Man( $\alpha$ 1-6)Man( $\alpha$ 1-O)Me in complex with concanavalin A. **J Biol Chem.**, v. 274, n. 41, p. 29188-29195. 1999.
- BOUTRON, P. Non-equilibrium formation of ice in aqueous solutions: efficiency of polyalcohol solutions for vitrification. In: Pegg, D.E. & Karow, A.M. Jr. (eds). **The biophysics of organ preservation**. New York: Plenum Press, 1987. p. 201-228.
- BOYD, W.C.; SHAPLEIGH, E. Diagnosis by subgroups of blood groups A and AB by use of plant agglutinins (lectins). **Journal of Laboratory Clinical Medicine**, v. 44, n. 2, p. 235-237. 1954.
- BRANDEN, C.I., JONES, T.A. Between objectivity and subjectivity. **Nature**, v. 343, p. 687-689. 1990.
- BRUENGER, A.T. Free R value: a novel statistical quantity for assessing the accuracy of crystals structures. **Nature**, v. 355, p. 472-474. 1992.

BRUENGER, A.T., ADAMS, P.D., RICE, L.M. New applications of simulated annealing in X-ray crystallography and solution NMR. **Structure**, v. 5, p. 325-336. 1997a.

BRUENGER, A.T. *et al.* Solution conformations of human growth hormone releasing factor: comparison of the restrained molecular dynamics and distance geometry methods for a system without long-range distance data. **Protein Eng.**, v. 1, n. 5, p. 399-406. 1987.

BUTS, L. *et al.* Weak Protein-Protein Interactions in Lectins: The Crystal Structure of a Vegetative Lectin from the Legume *Dolichos Biflorus*. **J.Mol.Biol.**, v. 309, p. 193. 2001.

CALVETE, J. J. *et al.* Molecular characterization and crystallization of Diocleinae lectins. **Biochim Biophys Acta**, v.1430, n. 2, p. 367-375. 1999.

Canavalia gladiata. In: Tibetan Medicine Encyclopaedia. [http://www.doctorgarma.com/rst/canavalia\\_gladiata/](http://www.doctorgarma.com/rst/canavalia_gladiata/)

CHAYEN, N.E. A novel technique to control the rate of vapour diffusion, giving larger protein crystals. **J. Appl. Cryst.**, v. 30, n. 2, p.198-202. 1997.

Collaborative Computational Project, Number 4. The CCP4 Suite: Programs for Protein Crystallography. **Acta Cryst. D**, v. 50, p. 760-763. 1994.

CREIGHTON, T.E. **Proteins – Structures and Molecular properties**. 2 ed. USA: W.H.Freeman and Company, 1993. p. 202-205.

CUDNEY, B. AND D'ARCY, A. RAMC 1999 Round Table Notes. Hampton Research Corp. Disponível em <http://www.hamptonresearch.com/stuff/rtn99.html>. Acesso em 10 de novembro de 2004.

CUDNEY, B. *et al.* Screening and optimization strategies for macromolecular crystal growth. **Acta Crystallogr. D**, v. 50, p. 414–423. 1994.

DAUTER, Z.; LAMZIN, V.S.; WILSON, K.S. Proteins at atomic resolution. **Curr. Opin. Struct. Biol.**, v. 5, p. 784-790. 1995.

DEACON, A. *et al.* The structure of concanavalin A and its bound solvent determined with small-molecule accuracy at 0.94Å resolution. **Journal Of The Chemical Society-Faraday Transactions**, v.93, n.24, p.4305-4312. 1997.

DELBAERE, L.T.J. *et al.* Structures of the lectin IV of *Griffonia simplicifolia* and its complex with the Lewis b human blood group determinant at 2.0 Å **Journal of Molecular Biology**, v. 230, p. 950-965. 1993.

DeLUCAS, L.J. *et al.* Efficient protein crystallization. **Journal of Structural Biology**, v. 142, n.1, p. 188-206. 2003.

DRENTH, J. **Principles of Protein X-ray Crystallography**. New York, USA: Springer-Verlag New York, 1994.

- DURCRUIX, A. AND GIEGE, R. **Crystallization of Nucleic Acids and Proteins, A Practical Approach**. EUA: Oxford University Press, 1992. p. 78-82.
- EDELMAN, G.M. *et al.* The covalent and three-dimensional structure of concanavalin A. **Proc. Natl. Acad. Sci. U.S.A.**, v. 69, p.2580-2584. 1972.
- EKANAYAKE, S.; JANSZ, E. R.; NAIR ,B. M. Nutritional evaluation of protein and starch of mature *Canavalia gladiata* seeds. **International Journal of Food Sciences and Nutrition**, v. 51, n. 4, p. 289-294. 2000.
- EKANAYAKE, S.; JANSZ, E.R.; NAIR, B.M. Literature review of an underutilized legume: *Canavalia gladiata* L. **Plant Foods Human Nutritional**, v.55, n. 4, p. 305-21. 2000.
- ENGH, R.A., HUBER, R. Accurate bond and angle parameters for X-ray structure refinement. **Acta Crystallogr. A**, v. 47, p. 392-400. 1991.
- ERIKSSON, A.E.; LILJAS, A. Refined structure of bovine carbonic anhydrase III at 2.0 Å resolution. **Proteins**, v. 16, n. 1, p. 29-42. 1993.
- EVANS, P. R. **Proceedings of the CCP4 Study Weekend. Data Collection and Processing**, Warrington: Daresbury Laboratory. L. Sawyer, N. Isaacs & S. Bailey, pp. 114-122. 1993.
- FLOHIL, J.A.; VRIEND, G.; BERENDSEN, H.J. Completion and refinement of 3-D homology models with restricted molecular dynamics: application to targets 47, 58, and 111 in the CASP modeling competition and posterior analysis. **Proteins**, v. 48, n. 4, p. 593-604. 2002.
- FUJIHARA, S.; NAKASHIMA, T.; KUROGOCHI, Y. Occurrence of a new polyamine, canavamine, in the sword bean *Canavalia gladiata*. **Biochem Biophys Res Commun.**, v. 107, n. 1, p. 403-410. 1982.
- FUJINAGA, M.; GROS, P.;VAN GUNSTEREN, W.F. Testing the method of crystallographic refinement using molecular dynamics. **J. Appl. Crystallogr.**, v. 22, p. 1-8. 1989.
- GARMAN, E.F., & MITCHELL, E.P. Glycerol concentrations required for cryoprotection of 50 typical protein crystallization conditions. **J. Appl. Cryst.**, v. 29, p. 584-587. 1996.
- GIACOVAZZO, C. *et al.* **Fundamentals of Crystallography**. New York, USA: Oxford University Press, 1992.
- GILLILAND, G.L. *et al.* The Biological Macromolecule Crystallization Database, Version 3.0: New Features, Data, and the NASA Archive for Protein Crystal Growth Data. **Acta Crystallogr. D**, v. 50, p. 408-413. 1994.
- GOLDSTEIN, I.J. AND PORETZ, M. R.D. Isolation, physicochemical characterization, and carbohydrate-binding specificity of lectins. In: *The lectins, Properties, Functions, and Applications in Biology and Medicine*. Leirner, IE, Sharon, N. and Goldstein, I.J. Eds. Academic Press, Orlando, FL, 1986. p. 33-247.

- GONZALES, A., AND C. NAVE. Radiation damage in protein crystals at low temperature. **Acta Crystallogr. D**, v. 50, p. 874-877. 1994.
- GREEN, A. A. Studies in the physical chemistry of the proteins. VIII. The solubility of hemoglobin in concentrated salt solutions. A study of the salting out of proteins. **J. Biol. Chem.**, v. 93, p. 495-516. 1931.
- GREER, J. Three-dimensional pattern recognition: an approach to automated interpretation of electron density maps of proteins. **J. Mol. Biol.**, v.82, n. 3, p. 279-301. 1974.
- GUEX, N. AND PEITSCH, M.C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. **Electrophoresis**, v. 18, p. 2714-2723. 1997.
- HAMANA, K.; MATSUZAKI, S. Natural occurrence of guanidinoxypropylamine in *Wistaria floribunda* and the sword bean *Canavalia gladiata*. **Biochem Biophys Res Commun.**, v. 129, n. 1, p. 46-51. 1985.
- HAMELRYCK, T.W. *et al.* Carbohydrate binding, quaternary structure and a novel hydrophobic binding site in two legume lectin oligomers from *Dolichos biflorus*. **Journal of Molecular Biology**, v. 286, n. 4, p. 1161-1177. 1999.
- HAMELRYCK, T.W. *et al.* Crystal structure of arcelin-5, a lectin-like defense protein from *Phaseolus vulgaris*. **J Biol Chem.**, v. 271, n. 51, p. 32796-32802. 1996.
- HARDMAN, K.D. AND AINSWORTH, C.F. Structure of concanavalin A at 2.4 Å resolution. **Biochemistry**, v. 11, p. 4910-4919. 1972.
- HENDERSON, R. Cryo-protection of protein crystals against radiation damage in electron and x-ray diffraction. **Proc. R. Soc. Lond. Biol.**, v. 24, p. 6-8. 1990.
- HENDRICKSON, W. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. **Science**, v. 254, p. 51-58. 1991.
- HENDRICKSON, W.A. *et al.* Crystallographic structure analysis of lamprey hemoglobin from anomalous dispersion of synchrotron radiation. **Proteins**, v. 4, p. 77-88. 1988.
- HENDRICKSON, WA. Stereochemically restrained refinement of macromolecular structures. **Methods Enzymol.**, v. 115, p. 252-70. 1985.
- HOLM, L., SANDER, C. Evaluation of proteins models by atomic solvations preference. **J. Mol. Biol.**, v. 225, p. 93-105. 1992.
- HOPE, H. Cryocrystallography of biological macromolecules: a generally applicable method. **Acta Cryst. B**, v. 44, p. 22-26. 1988.
- Horticultural Sciences Department, Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, University of Florida. Endereço BBS: <http://edis.ifas.ufl.edu/MV020>. Acesso em: 16 dezembro 2004.

International Tables for X-Ray Crystallography: The Kynoch Press. 1969.

IP, W.K.; LAU, Y.L. Role of mannose-binding lectin in the innate defense against *Candida albicans*: enhancement of complement activation, but lack of opsonic function, in phagocytosis by human dendritic cells. **J Infect Dis.**, v. 190, n. 3, p. 632-640. 2004.

JANCARIK, J. *et al.* Optimum solubility (OS) screening: an efficient method to optimize buffer conditions for homogeneity and crystallization of proteins. **Acta Crystallogr D Biol Crystallogr.**, v. 60, n. 9, p.1670-1673. 2004.

JANCARIK, J.; KIM, S. Sparse matrix sampling: a screening method for crystallization of proteins. **J. Appl. Crystallogr.**, v. 24, p. 409-411. 1991.

JONES, S. AND THORNTON, J.M. Protein-protein interactions: a review of protein dimer structures. **Prog. Biophys. Mol. Biol.**, v. 63, p. 31-65. 1995.

JONES, T.A. Interactive computer graphics: FRODO. **Methods Enzymol.**, v. 115, p. 157-171. 1985.

JONES, T.A., KJELDGAARD, M. '**O**': the Manual, version 5.9. Upsala. 1993.

JONES, T.A., THIRUP, S. Using known substructures in protein model building and crystallography. **EMBO J.**, v. 5, n. 4, p. 819-822. 1986.

KARLE, J. Determination of molecular formula and stereoconfiguration of unique steroids by X-ray diffraction analysis. **Lipids.**, v. 15, n. 9, p. 793-7. 1980.

KARPUNINA, L.V.; MEL'NIKOVA, U.Y.; KONNOVA, S.A. Biological role of lectins from the nitrogen-fixing *Paenibacillus polymyxa* strain 1460 during bacterial-plant-root interactions. **Curr Microbiol.**, v. 4, n. 5, p. 376-378. 2003.

KOKOUREK, J. Historical background. In: **The Lectins, Properties, Functions, and Applications in Biology and Medicine**. Liener, I.E., Sharon, N. and Goldstein, I.J., Eds, Academic Press, Orlando, FL. 1986. p. 1-32.

KRISHNAMURTHI, H.M., *et al.* Crystal structure of Clostridium acidic-uric Ferredoxin at 5 Å resolution based on measurements of anomalous X-ray scattering at multiple wavelengths. **J. Biol. Chem.**, v. 263, p. 18430-18436. 1988.

KWONG, P.D. LIU, Y. Use of cryoprotectants in combination with immiscible oils for flash cooling macromolecular crystals. **J. Appl. Cryst.**, v. 32, p. 102-105. 1999.

LASKOWSKI, R.A. *et al.* PROCHECK: a program to check the stereochemical quality of protein structures. **J. Appl. Crystallogr.**, v. 26, p. 283-291. 1993.

LESLIE, A. Data collection and processing. In: Sawyer L, Isaac N, Bailey S, eds. **Proceedings of the CCP4 study weekend**. Warrington, UK: SERC Daresbury Laboratory, 1993. p. 44-51.

LORIS, R. *et al.* Legume lectin structure. **BBA**, v. 1383, p.9-36. 1998.

- LORIS, R. Principles of structures of animal and plant lectins. **Biochim Biophys Acta**, v. 19; n. 2-3, p.198-208. 2002.
- LUCA, G; REIS, B. Sistema em fluxo para determinação espectrofotométrica de uréia em plasma de sangue animal empregando leguminosa como fonte natural da enzima urease. **Quimica Nova**, v. 24, n. 2, p. 191-194. 2001.
- LUTHY, R., BOWIE, J.U., EISENBERG, D. Assessment of protein models with three-dimensional profiles. **Nature**, v. 356, p. 83-85. 1992.
- MAKELA, O.; MAKELA, P. An erythrocyte sedimentation method for detecting incomplete antibodies. **Ann Med Exp Biol Fenn**, v. 32, n. 3, p. 276-281. 1954.
- MANOJ, N. AND SUGUNA, K. Signature of quaternary structure in the sequences of legume lectins. **Protein Engineering**, v. 14, n.10, p. 735-745. 2001.
- MATTHEWS, B.W. Solvent content of proteins crystals. **J. Mol. Biol.**, v. 33, p. 491-497. 1968.
- MCPHERSON, A. **Crystallization of Biological Macromolecules**. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 1998.
- McPHERSON, A. **Crystallization of Biological Macromolecules**. New York, EUA: Cold Spring Harbor Press, 1999. 424 p.
- McPHERSON, A. Current approaches to macromolecular crystallization. **Eur. J. Biochem.**, v. 189, p. 1-23. 1990.
- MCPHERSON, A. **Preparation and Analysis of Protein Crystals**. Alemanha: Krieger Publishing, 1992. p. 88-91.
- McREE, D.E. **Practical Protein Crystallography**. USA: Academic Press, inc. 1993.
- McREE, D.E. XtalView/Xfit - A versatile program for manipulating atomic coordinates and electron density. **J Struct Biol.**, v. 125, n. 2-3, p.156-165. 1999.
- MEHL, P. Experimental dissection of devitrification in aqueous solutions in 1,3-butanediol. **Cryobiology.**, v. 26, p. 567-568. 1989.
- MURAKAMI, T. *et al.* Medicinal foodstuffs. XIX. Absolute stereostructures of canavalioides, a new Ent-kaurane-type diterpene glycoside, and gladiatosides A1, A2, A3, B1, B2, B3, C1, and C2, new acylated flavonol glycosides, from sword bean, the seeds of *Canavalia gladiata*. **Chem Pharm Bull**, v. 48, n.11, p. 673-80. 2000.
- MUROFUSHI, N. *et al.* **Agricultural Biological Chemistry**, v. 33, p. 598-609. 1969.
- MURSHODOV, G. N., DODSON, E. J. AND VAGIN, A. A. Macromolecular refinement. In: Dodson, E., Moore, M., Ralph, A. and Bailey, S., eds. **Proceedings of the CCP4 study weekend**. Daresbury Laboratory, Warrington, UK. 1996. p. 93-104.

- MURSHUDOV, G.N. *et al.* Efficient anisotropic refinement of macromolecular structures using FFT. **Acta Crystallogr D Biol Crystallogr.** v. 55, n. 1, p. 247-55. 1999.
- NAISMITH, J.H. Refined structure of Concanavalin A complexed with methyl alpha-D-mannopyranoside at 2.0 Å resolution and comparison with the saccharide-free structure. **Acta Crystallogr D Biol Crystallogr.**, v. 50, n. 6, p. 847-58. 1994.
- NAVAZA, J., 1994. AmoRe: an automated package for molecular replacement. **Acta Crystallogr. A**, v. 50, p.157-163.
- PANNU, N.S., READ, R.J. Improved structure refinement through maximum likelihood. **Acta Crystallogr. A**, v. 52, p. 659-668. 1996.
- PAPET, M. P. *et al.* Characterization of a high molecular mass cytokinin-like compound extracted from wheat germ (*Triticum vulgare* L.). **Plant Science**, v. 68, n. 2, p. 175-182. 1990.
- PARKIN, S., & HOPE, H. Macromolecular cryocrystallography: Cooling, mounting, storage and transportation of crystals. **J. Appl. Cryst.**, v. 31, p. 945-953. 1998.
- PETCOCK, J.M. *et al.* Effects of different post-crystallization soaking conditions on the diffraction of Mtcp1 crystals. **Acta Cryst. D**, v. 57, p. 763-765. 2001.
- PEUMANS, W.J.; VAN DAMME, E.J. Lectins as plant defense proteins. **Plant Physiol.**, v. 109, n. 2, p. 347-352. 1995.
- POLIKARPOV, I. *et al.* Set-up and experimental parameters of the protein crystallography beamline at the Brazilian national synchrotron laboratory. **J. Synchrotron Rad.**, v. 5, p. 72-76. 1998a.
- POLIKARPOV, I. *et al.* The protein crystallography beamline at LNLS, the Brazilian National Synchrotron Light Source. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, v. 405, n. 1, p.159-164. 1998b.
- PONDER, J.W., RICHARDS, F.M. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. **J. Mol. Biol.**, v. 193, n. 4, p. 775-791. 1987.
- PONTIUS, J.; RICHELLE, J.; AND WODAK, S.J. Quality assessment of protein 3D structures using standard atomic volumes. **J. Mol. Biol.**, v. 264, p. 121-136. 1996.
- PRABU, M.M.; SUGUNA, K.; VIJAYAN, M. Variability in quaternary association of proteins with the same tertiary fold: a case study and rationalization involving legume lectins. **Proteins**, v. 35, n. 1, p. 58-69. 1999.
- Protein Crystallography Course. Disponível em: <http://www-structmed.cimr.cam.ac.uk/course.html>. Acesso em 10 dezembro 2004.

- RAY, W.J. AND BRACKER, C.E. Polyethylene glycol: catalytic effect on the crystallization of phosphoglucosyltransferase at high salt concentration. **J. Crystal Growth**, v. 76, p. 562-576. 1986.
- RICE, L.M., BRUNGER, A.T. Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement. **Proteins**, v. 19, p. 277-290. 1994.
- ROBERTS, D.D. AND GOLDSTEIN, I.J. Adenine Binding Sites of the Lectin from Lima Beans (*Phaseolus lunatus*). **The Journal Of Biological Chemistry**, v. 258, n. 22, p. 13820-13824. 1983.
- ROBERTS, D.D.; GOLDSTEIN, I.J. Binding of hydrophobic ligands to plant lectins: titration with arylaminonaphthalenesulfonates. **Arch Biochem Biophys**. v. 224, n. 2, p. 479-484. 1983.
- RODRIGUEZ, D. *et al.* Differences in macrophage stimulation and leukocyte accumulation in response to intraperitoneal administration of glucose/mannose-binding plant lectins. **Brazilian Journal of Medicine Biological Research**, v. 25, p. 823-826. 1992.
- ROSSMANN, M.G. *Ab initio* phase determination and phase extension using non-crystallographic symmetry. **Curr Opin Struct Biol.**, v. 5, p. 650-655. Review. 1995.
- ROSSMANN, M.G., BLOW, D.M. The detection of sub-units within the crystallographic asymmetric unit. **Acta Crystallogr.**, v. 15, p. 24-31. 1962.
- RUDIGER H, *et al.* Medicinal chemistry based on the sugar code: fundamentals of lectinology and experimental strategies with lectins as targets. **Curr Med Chem.**, v. 7, n. 4, p. 389-416. 2000.
- SANZ-APARICIO J. *et al.* The crystal structure of *Canavalia brasiliensis* lectin suggests a correlation between its quaternary conformation and its distinct biological properties from Concanavalin A. **FEBS Letters**, v. 17, n.405(1), p. 114-118. 1997.
- SAYLE, R., MILNER-WHITE, E.J. RASMOL: biomolecular graphics for all. **TiBS**, v. 20, p. 374-376. 1995.
- SCHNEIDER, T.R. Cryocrystallography of biological macromolecules. **Acta Physica Polonica A.**, v. 91, p. 739-744. 1997.
- SEGELKE, B. W. **Efficiency analysis of sampling protocols used in protein crystallization screening and crystal structure from two novel crystal forms of PLA2**. 1995. Tese de doutorado, University of California, San Diego, EUA.
- SEGELKE, B. W. Efficiency Analysis of Screening Protocols Used in Protein Crystallization. **Journal of Crystal Growth**, v. 232, p. 553-562. 2001.
- SENGUPTA S, *et al.* Phytolectins: natural molecules with immense biotechnological potential. **Indian J Exp Biol**, v. 35, n. 2, p. 103-110. 1997.

- SHAANAN, B; LIS, H.; SHARON, N. Structure of a legume lectin with an ordered *N-linked* carbohydrate in complex with lactose. **Science**, v. 254, p. 862-866. 1991.
- SOUSA, R. Use of glycerol, polyols and other protein structure stabilizing agents in protein crystallization. **Acta Crystallogr D Biol Crystallogr.**, v. 51, n. 3, p. 271-277. 1995.
- SOUZA, D.H.; SELISTRE-DE-ARAUJO, H.S.; GARRATT, R.C. Determination of the three-dimensional structure of toxins by protein crystallography. **Toxicon.**, v. 38, n. 10, p. 1307-1353. Review. 2000.
- SUMMER, J.B.; AND HOWELL, S.F. The identification of the hemagglutinin of the Jack bean with concanavalin A. **J. Bacteriol.**, v. 32, p. 227-237. 1936.
- SUSSMAN, J.L.; PODJARNY, A.D. The use of a constrained-restrained least-squares procedure for the low-resolution refinement of a macromolecule, yeast tRNA<sub>f Met</sub>. **Acta Crystallogr. B**, v. 39, p. 495. 1983.
- TAKAHASHI, N. *et al.* Structures of gibberellins A26 and A27 in immature seeds of *Pharbitis nil*. **Tetrahedron Letters**, v. 10, n. 25, p. 2077-2080. 1969.
- TAKAHASHI, N. *et al.* Structures of new gibberellins in immature seeds of *Canavalia gladiata*. **Tetrahedron Letters**, v. 8, n. 48, p. 4861-4865. 1967.
- TAMURA, S. Historical aspects of gibberellins. In: **Gibberellins**. Eds Takahashi N, Phinney BO and Macmillan J. Springer-Verlag. New York, 1990. p. 1-8.
- TERWILLIGER, T.C. Maximum-likelihood density modification. **Acta Crystallogr D Biol Crystallogr.**, v. 56, n. 8, p. 965-972. 2000.
- TERWILLIGER, T.C. Reciprocal-space solvent flattening. **Acta Crystallogr D Biol Crystallogr.**, v. 55, n.11, p. 1863-1871. 1999.
- TONG, L. How to take advantage of non-crystallographic symmetry in molecular replacement: 'locked' rotation and translation functions. **Acta Crystallogr D Biol Crystallogr.**, v. 57, n. 10, p.1383-1389. 2001.
- TRAKHANOV, S. AND QUIOCHO, F.A. Influence of divalent cations on protein crystallization. **Protein Science**, v. 4, n. 9, p. 1914-1919. 1995.
- UMEKAWA, H. *et al.* **Agric. Biol. Chem.**, v. 54, p. 3295-3299. 1990.
- UVX parameter list, L. Lin, Editor LNLS-CT 17/94.
- VAGIN A, TEPLYAKOV A. An approach to multi-copy search in molecular replacement. **Acta Crystallogr D Biol Crystallogr.**, v.56, n.12, p.1622-1624. 2000.
- VAGUINE, A.A.; RICHELLE, J.; WODAK, S.J. SFCHECK: a unified set of procedure for evaluating the quality of macromolecular structure-factor data and their agreement with atom model. **Acta Cryst D**, v. 55, p. 191-205. 1999.

- VAN DAMME, E.J.M. *et al.* Cytoplasmic/nuclear plant lectins: a new story. **Trends Plant Sci.**, v. 9, n.10, p. 484-489. 2004.
- VAN DAMME, E.J.M. *et al.* Plant lectins: A composite of several distinct families of Structurally and Evolutionary related proteins with diverse biological roles. **Critical Reviews in Plant Sciences**, v. 17, n.6, p. 575-692. 1998<sup>b</sup>.
- VAN DAMME, E.J.M. *et al.* **Handbook of Plant Lectins: Properties and Biomedical Applications**. John Wiley & Sons, Chichester, UK. 1998<sup>a</sup>.
- VARGIN, A.; TEPLYAKOV, A. A translation-function approach for heavy-atom location in macromolecular crystallography. **Acta Crystallogr D Biol Crystallogr.**, v. 54, n., p. 400-402. 1998.
- VODKIN, L.O.; RHODES, P.R. AND GOLDBERG, R.B. A lectin gene insertion has the structural features of a transposable element. **Cell**, v. 34, p. 1023-1031. 1983.
- VRIEND, G. WHAT IF: a molecular modeling and drug design program. **J. Mol. Graph.**, v. 8, p. 52-56. 1990.
- VRIEND, G., SANDER, C., 1993. Quality control of protein models: directional atomic contact analysis. **J.Appl. Crystallogr.**, v. 26, p. 47-60. 1990.
- WATKINS, W.M. AND MORGAN, W.T.J. Neutralization of the anti-H agglutinin in eel by simple sugars. **Nature**, v. 169, p. 825-826. 1952.
- WINICK, H. **Synchrotron radiation sources. A primer**. Singapore: World Scientific, 1994.
- WOOH, J.W. *et al.* Comparison of three commercial sparse-matrix crystallization screens. **Acta Crystallogr D Biol.**, v. 59, n. 4, p. 769-72. 2003.
- YAMAUCHI, D. *et al.* Nucleotide sequence of cDNA for concanavalin A from *Canavalia gladiata* seeds. **Plant Cell Physiol.**, v. 30, p.147-150. 1989.
- YANG W, *et al.* Structure of ribonuclease H phased at 2 Å resolution by MAD analysis of the selenomethionyl protein. **Science**, v. 249, n. 4975, p. 1398-1405. 1990.
- YOKOTA, T.; TAKAHASHI, N. **Agricultural Biological Chemistry**, v. 45, p. 1251-1254. 1981.